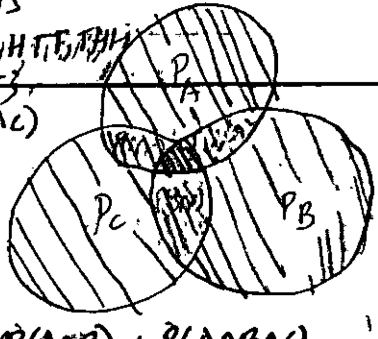
Chapter 1: 1a. Sample Space: {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

b. 1) {HHH, HHT, THH, HTH} 2) ~~HHH, HHT~~ 3) HHT, THT, TTT

c. A^c = "complement": the elements in the space which are not A. {~~HHT, TTT, HTH, THT~~}

$A \cap B$ = "intersection": the event both A and B occur: {HHT, HHH}

$A \cup B$ = "Union": events of A and B, and A or B. {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}



2. a) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A\cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

$= P(A \cup B) \cup P(C) = [P(A) + P(B) - P(A\cap B)] \cup P(C)$

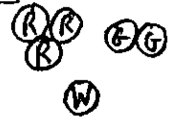
Addition Law

$= P(A \cup C) + P(B \cup C) - P(A \cap B) \cup P(C)$

$= P(A) + P(C) - P(A \cap C) + P(B) + P(C) - P(B \cap C) - P(C) - P(A \cap B) + P(A \cap B \cap C)$

$\boxed{= P(A) + P(B) + P(C) - P(A\cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)}$

3. Semi



3 draws; RRR RRG RRW RGG GGW

RGR RWR GRG GWG WGR

GRR WRR GGB WGG WRG

$n = 6$
$k = 3$

$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{6!}{3!(3!)} = \frac{6\cdot5\cdot4}{6} = \frac{120}{6} = \frac{20}{6\cdot 6} = \frac{5}{9}$

$\frac{6!}{1!(5!)} = 6$

Event A: 1 Draw

$\frac{P(R)+P(G)+P(W)}{P(G \cap R \cap W)} = \frac{\binom{3}{1}+\binom{2}{1}+\binom{1}{1}}{\binom{6}{1}} = \frac{\frac{3!}{1!(2!)} + \frac{2!}{1!\,1!} + \frac{1!}{1!\,0!}}{\frac{6!}{1!\,5!}} = \left(\frac{3}{6}\right) + \left(\frac{2}{6}\right) + \left(\frac{1}{6}\right)$

Event B: 2 Draw

$\frac{P(R)+P(G)+P(W)}{P(G \cap R \cap W)}$ Should write Unions and intersection instead.

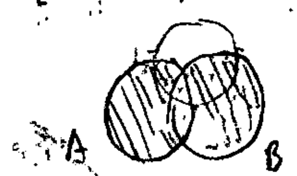
4. Prove $P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$; $P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1) + P(A_2) + \cdots + P(A_n) - P(A_1 \cap A_2) - \cdots - P(A_1 \cap A_n) - P(A_2 \cap A_3) - P(A_2 \cap A_n)$

$\sum_{i=1}^{n} P(A_i) = P(A_1) + P(A_2) + \cdots + P(A_n)$

$\boxed{P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)}$

5. Let (A, not B; or B, not A) be C

$\boxed{C = A \triangle B = (A \cap \neg B) \vee (\neg A \cap B) = A + B - A \cup B = A^c \cap B \vee A \cap B^c}$



6. Two six-sided dice are thrown: A) Sample space:

Dice 1:	Dice 2:
1	1
1	2
1	3
1	4
1	5
1	6
2	1
2	2
2	3
2	4
2	5
2	6
3	1
3	2
3	3
3	4
3	5
3	6
4	1
4	2
4	3
4	4
4	5
4	6
5	1
5	2
5	3
5	4
5	5
5	6
6	1
6	2
6	3
6	4
6	5
6	6

B)(1) A = sum of the two value is at least 5.

(1,4) (5,1) (1,6) (6,1) (4,4)
(2,4) (4,2) (2,6) (6,2) (4,5)
(3,3) (3,3) (3,6) (6,3) (5,4)
(2,5) (5,2) (4,6) (6,4)
(3,4) (4,3) (5,6) (6,5)
(3,5) (5,3) (6,6) (6,6)

(2) B = the value of the first dice greater than the second.

(2,1) (3,2) ~~(4,3)~~ (5,4) (6,5)
(3,1) ~~(4,2)~~ (5,3) (6,4)
~~(4,1)~~ (5,2) (6,3)
(5,1) (6,2)
(6,1)

(3) C = the first value is 4

(4,1) (4,4)
(4,2) (4,5)
(4,3) (4,6)

c) $A \cap C = (4,2),(4,3),(4,4),(4,5),(4,6)$

$B \cup C = (2,1),(3,1),(5,1),(6,1),(3,2),(5,2),(6,2)$
$(5,3),(6,3),(5,4),(6,4),(6,5),(4,2),(4,3)$
$(4,4),(4,5),(4,6)$

$A \cap (B \cup C) = (4,2),(4,3),(4,4),(4,5),(4,6).$

7. Bonferroni's equality: $P(A \cap B) \geq P(A) + P(B) - 1.$

Addition Law: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(A \cap B) = P(A) + P(B) - P(A \cup B)$

Therefore, $P(A \cap B) = P(A) + P(B) - P(A \cup B) \geq P(A) + P(B) - 1$

$\boxed{P(A \cup B) \leq 1}$

8. De Morgan's Law: $(A \cup B)^c = A^c \cap B^c$ 
$(A \cap B)^c = A^c \cup B^c$ 

9. Probability of rain on Saturday (25%)
Probability of rain on Sunday (25%)
The probability of consecutive events would be the multiplication of the probability of the events $(\frac{1}{4} \cdot \frac{1}{4}) = \frac{1}{16} = 12.5\%$, and not 50% proposed

10. n balls into k urns, What's the probability the last urn contains j balls? $\dfrac{\binom{i}{j}}{\binom{n}{k}} = \dfrac{\frac{i!}{i!(j-1)!}}{\frac{n!}{k!(n-k)!}} = \boxed{\dfrac{k!(n-k)!}{n!(j-1)!}}$

11. Telephone with seven total digits; 432 are the first three digits; four digits remain with 10 potential digits each; 10^4

Total possibilities: $10 \cdot 10 \cdot 10 \cdot 10^4$.

Chances of four more distinct digits are

$\dfrac{7 \cdot 6 \cdot 5 \cdot 4}{10^4} = \dfrac{840}{10000} = \boxed{8.4\%}$

12. 26-letter English Alphabet ~~into 8-bit binary words~~.

$\binom{2^8}{26} = \dfrac{2^8!}{26!(2^8-26)!}$

$= 2.74 \times 10^{35}$

13. a) Straight five cards in unbroken sequence: 4 suits $/ \binom{13}{5} = 4 / [13!/(13-5)!] = \boxed{4\left(\frac{1}{13}\right)\left(\frac{1}{12}\right)\left(\frac{1}{11}\right)\left(\frac{1}{10}\right)\left(\frac{1}{9}\right)}$

b) Four of a kind: $\binom{13}{1}\binom{4}{1}\binom{12}{1}\binom{4}{2} = \dfrac{13!}{1!\,12!} \cdot \dfrac{4!}{1!\,3!} \cdot \dfrac{12!}{1!\,11!} \cdot \dfrac{4!}{2!\,2!} = 13 \cdot 4 \cdot 12 \cdot 4 \cdot 3/2 = 3744$

$\dfrac{3744}{\binom{52}{5}} = \dfrac{3744}{2598960} = \boxed{0.05\%}$

c) A full house (three cards of one value and two of another)

(Probability of three cards of fifty-two) x (Probability of two cards of fifty-two)

$= \dfrac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}} = \dfrac{6}{4165} = \boxed{0.144\%}$

14. Prove if $P(A|E) \geq P(B|E)$ and $P(A|E^c) \geq P(B|E^c)$

then $P(A) \geq P(B)$

$P(A|E) \geq P(B|E)$ and $P(A|E^c) \geq P(B|E^c)$

E 600

A X 600

B Y 600

50 50

$\dfrac{X}{50} \geq \dfrac{Y}{50}$; $\dfrac{1-\frac{X}{100}}{50} \geq \dfrac{1-\frac{Y}{100}}{50}$

$\boxed{X \geq Y}$; $1 - \dfrac{X}{100} \geq 1 - \dfrac{Y}{100}$

$\boxed{X \geq Y}$

15. 4 meats, 6 vegetables, three starches

$\binom{1}{4} \cdot 4 \cdot 6 \cdot 3 = \boxed{72 \text{ meals}}$

16. Simpsons Paradox:

Black Urn: {3 red and 6 green balls}
White Urn: {5 red and 4 green balls} } Set #1

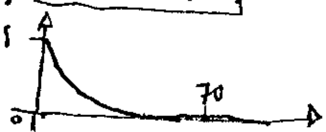
First trial: Black Urn $\left(\frac{3}{9}\right)$; $\boxed{\text{White Urn} \left(\frac{5}{9}\right)}$

Black Urn: {2 red and 6 green balls}
White Urn: {15 red and 1 green balls} } Set #2

Second Trial: Black Urn $\left(\frac{2}{8}\right)$; $\boxed{\text{White Urn} \left(\frac{15}{16}\right)}$

Black Urn: {5 red and 12 green balls}
White Urn: {20 red and 5 green balls} } Set #3

Third Trial: Black Urn $\left(\frac{5}{12}\right)$; $\boxed{\text{White Urn} \left(\frac{20}{25}\right)}$



17. Accepts: 4 items of 100
Rejects: 1 item is defective

$P(A) = \dfrac{\binom{100-K}{4}\binom{K}{0}}{\binom{100}{4}} = \dfrac{\frac{(100-K)!}{4!(100-K-4)!}}{\frac{100!}{4!(96!)}}$

$= \dfrac{\frac{(100-K)(100-K-1)(100-K-2)(100-K-3)(100-K-4)!}{(100-K-4)!}}{(100 \times 99 \times 98 \times 97)}$

$= \dfrac{(100-K)(99-K)(98-K)(97-K)}{100 \times 99 \times 98 \times 97} = \left(1-\frac{K}{100}\right)\left(1-\frac{K}{99}\right)\left(1-\frac{K}{98}\right)\left(1-\frac{K}{97}\right)$

19. $\dfrac{\text{Player one choice}}{\text{Total choices}} = \dfrac{\left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{1}{6}\right)}{} = \boxed{\dfrac{1}{1296}}$;

19. Five Chicanos, two Asians, three African Americans.

a) $\dfrac{\binom{5}{1}\cdot\binom{2}{1}\cdot\binom{3}{1}}{\binom{10}{4}4!} = \dfrac{\left(\frac{5!}{1!4!}\right)\left(\frac{2!}{1!1!}\right)\left(\frac{3!}{1!2!}\right)}{\frac{10!}{4!6!}}$

$= 3\dfrac{5\cdot 2\cdot 3\,(4!)}{10\cdot 9\cdot 8\cdot 7}$

$= \boxed{1/7}$

b. $\dfrac{\binom{5}{1}\cdot\binom{2}{1}\cdot\binom{3}{1}}{\binom{10}{5}} = \dfrac{5\cdot 2\cdot 3\cdot 5!}{10\cdot 9\cdot 8\cdot 7\cdot 6} = \boxed{\dfrac{5}{42}}$

20. Arrangements Statistically

s's = 2 ; t's = 3; a's = 2;

i's = 2; c's = 1 ; u's = 1; l's = 2

Total Amount = 13 ; y's = 1

$\dfrac{\binom{2}{1}\binom{3}{1}\binom{2}{1}\binom{2}{1}\binom{1}{1}\binom{1}{1}\binom{2}{1}\binom{1}{1}}{\binom{13}{13}} = 2\cdot 3\cdot 2\cdot 2\cdot 1\cdot 1\cdot 2\cdot 1$ $\boxed{= 48}$

21. $\dfrac{2^2+2^2+2^2}{2^5} = \boxed{\dfrac{12}{32}}$

22. $\left(\frac{1}{4}\right)\left(\frac{1}{3}\right)\left(\frac{1}{2}\right)\left(\frac{1}{1}\right) = \boxed{\frac{1}{24}}$

23. $\underbrace{\binom{n}{1}}_{1^{st}\text{ player}}\underbrace{\binom{n}{2}}_{2^{nd}\text{ players}}\underbrace{\binom{n-1}{1}}_{\text{Remaining one}}(n-2)!$

$\dfrac{n!}{1!(n-1)!}\cdot\dfrac{n!}{2!(n-2)!} = \dfrac{(n-1)!}{1!(n-2)!}(n-2)!$

$\binom{n}{2}\cdot n! = \binom{n-1}{2}\binom{n}{1}$

24. 52 cards ; $\dfrac{\text{Probability Aces next to each other}}{\text{Total Arrangements}} = \dfrac{\binom{13}{4}4!}{\binom{52}{4}}$

$= \dfrac{13\cdot 12\cdot 11\cdot 10\cdot 4!}{4!\cdot 52\cdot 51\cdot 50\cdot 49} = 11/4165 = \boxed{26\%}$

25. $\boxed{3!}$

26. n items with K defects ; m are selected and inspected.

value of m to be below a probability of 0.90

a) n = 1000 ; K = 10 $\dfrac{\text{Probability of defect}}{\text{Total outcomes}} = \dfrac{\binom{n-K}{m}\binom{K}{0}}{\binom{n}{m}} = \dfrac{\binom{m+n}{n}\binom{n}{0}}{\binom{n}{n}} = \dfrac{\frac{(n-m)!}{n!(n-m-k)!}}{\frac{n!}{k!(n-k)!}} = \dfrac{(1000-m)!\,(n-k)!}{1000!\,(n-m-k)!}$

b) n = 1000 ; K = 10 ; $0.9\,\dfrac{(1000!)}{(900)!} = \dfrac{(1000-m)!}{(900-m)!}$

27. $\dfrac{\text{Probability of no letters occuring}}{\text{Total Probability}} = \dfrac{26C5}{26^5} = \dfrac{\binom{26}{5}}{26^5} = \dfrac{\frac{26!}{5!(21!)}}{26^5} = \dfrac{\frac{26\cdot 25\cdot 24\cdot 23\cdot 22}{5\cdot 4\cdot 3\cdot 2\cdot 1}}{26^5}$

28. 5 players with five cards from 52-card deck. $= \dfrac{(52)!}{25!(27!)}$ $\boxed{= 4.775\times 10^{14}\text{ ways}}$ $= 5.53\times 10^{-3} = 0.55\%$

29. Three spades and Two Hearts:

a) Discards two hearts and draws two more cards.

$= \left(\frac{11}{47}\right)\left(\frac{10}{46}\right) = \dfrac{110}{2162} = \dfrac{55}{1081}$ $\boxed{5.1\%}$

30. 60 2nd graders into two classes of 30 each. Probability of five chosen into same class

$\dfrac{\binom{2}{1}}{\binom{60}{5}} = \dfrac{2!}{\frac{60!}{5!55!}} = \dfrac{2\cdot 5!}{60\cdot 59\cdot 58\cdot 57\cdot 56} = 0.00004\% = \binom{60}{30}\ \binom{60}{5}\ \dfrac{60!}{}$

$(1000-m)(1000-(m+1))(1000-(m+2)) \cdots (1000-(m+10))\cdots(1000-m)$

four students: $\frac{\binom{2}{1}}{\binom{60}{4}} = \frac{2}{\frac{60!}{4!\,56!}} = \frac{2\cdot 4\cdot 3\cdot 2}{60\cdot 59\cdot 58\cdot 57} = \boxed{0.0004\%}$

Morelle in one class and her friends in another: $\frac{\binom{2}{1}\binom{58}{29}}{\binom{60}{30}} = \frac{\frac{2\cdot 58!}{29!\,(29)!}}{\frac{60!}{30!\,30!}} = 6\%$ 4%

31. Six Male and Six Female Dancers. ${}_6P_6 \cdot {}_6P_6 = \frac{6!}{0!} \cdot \frac{6!}{0!} = 720^2$

32. $\frac{4\cdot\binom{13}{n}}{\binom{52}{n}} = \frac{4\cdot\frac{13!}{n!(13-n)!}}{\frac{52!}{n!(52-n)!}}$

When is the value 0.5?

$= \frac{4\cdot 13!\,(52-n)!}{(13-n)!\,52!} = \boxed{\frac{1.86\times 10^{-57}(52-n)!}{(13-n)!}} = \frac{3.7\times 10^{-57}\,\Gamma(53-n)}{\Gamma(14-n)} = 1;\ n=3.$

Gamma Identity: $\Gamma(n) = (n-1)!$

$\Gamma(n+1) = n\Gamma(n)$

33. Five people and five floors. $\frac{\text{Probability of a person per floor}}{\text{Probability of choosing five}} = \frac{\binom{5}{1}}{\binom{7}{5}} = \frac{\frac{5!}{1!(5-1)!}}{\frac{7!}{5!(7-5)!}} = \frac{5!}{4!}\frac{5!(2!)}{7!} = \frac{5!\,5!\,2!}{4!\cdot 7!} = \boxed{23.8\%}$

34. Prove the following identity:

$$\sum_{k=0}^{n}\binom{n}{k}\binom{m-n}{n-k} = \binom{m}{n}$$

$$\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\frac{(m-n)!}{(n-k)!(m-n-n+k)!} = \sum_{k=0}^{n}\frac{n(n-1)(n-2)(n-3)\cdots(n-k)!\,(m-n)!}{k!\,(n-k)!\,(n-k)!\,(m+k)!}$$

$$= \sum_{k=0}^{n}\frac{n(n-1)(n-2)(n-3)\cdots n!\,(m-n)!}{k!\left(\frac{n!}{n(n-1)(n-2)(n-3)\cdots}\right)(m+k)!} = \sum_{0}^{n}\frac{[n(n-1)(n-2)(n-3)\cdots]^2(m+n)!}{k!\;n!\;(m+k)!}$$

Two methods: 1. Pascal's Identity: $\binom{n+1}{r} = \binom{n}{r} + \binom{n}{r-1}$; $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$

$k! = k(n-1)(n-2)$

$$\sum_{k=0}^{n}\binom{n}{k} = \sum\left[\binom{n+1}{k+1} - \binom{n}{k+1}\right] = \sum^{n}\binom{n+1}{k+1} - \sum^{n}\binom{n}{k+1}$$

$$= \sum_{k+1}^{n+1}\binom{n}{k+1} - \sum^{n}\binom{n}{k+1} = \binom{n+1}{k+1} - \binom{n}{k+1} = \binom{n+1}{k+1}$$

2. Binomial Theorem: $\sum_{k=0}^{n}\binom{n}{k}\binom{m-n}{n-k} = \sum_{k=0}^{n}(1+k)^n(1+n-k)^{m-n} = ({}^nC_0 k^n + {}^nC_1 k^{n-1} + {}^nC_2 k^{n-2} + \cdots)$
$({}^{m-n}C_0 (n-k)^{m-n} + {}^{m-n}C_1 (n-k)^{m-n-1} + {}^{m-n}C_2 (n-k)^{m-n-2} \cdots)$

$$\text{"}(1+x)^n\text{"} \quad = \left(1 + nk + \frac{n(n-1)}{2!}k^2 + \frac{n(n-1)(n-2)}{3!}k^3 + \cdots\right)\left(1 + (m-n)(n-k) + \frac{(m-n)(m-n-1)}{2!}(n-k)^2 + \cdots\right)$$

$$= \left(1 + mn + \frac{m(m-1)}{2!}n^2 + \cdots\right) = \binom{m}{n}$$

35. Prove the following identities.

a) $\binom{n}{r} = \binom{n}{n-r} = \frac{n!}{r!(n-r)!} = \frac{n!}{(n-n+r)!\,(n-r)!} = \boxed{\binom{n}{n-r}}$; An expansion of a binomial is telescoping

b) $\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{(n-1)!\,n}{r!(n-r)!} = \frac{(n-1)!}{r!}\left[\frac{n-k}{(n-r)!} + \frac{k}{(n-r)!}\right] = \frac{(n-1)!}{r!(n-1-k)!} + \frac{r(n-1)!}{(r-1)!(n-k)!} = \boxed{\binom{n-1}{k} + \binom{n-1}{k-1}}$

36. $\boxed{R|R|R|G|G|G}$ Arrangements ; Combinations per type – $\binom{6}{3}\binom{3}{3} = \frac{6!}{3!3!}\frac{3!}{3!0!} = \frac{6\cdot5\cdot4}{3\cdot2}$

6 Blocks $= \frac{120}{6} = \boxed{20}$

$\boxed{R|R|R|W|W|W|G|G|G}$ Combinations per type – $\binom{9}{3}\binom{6}{3}\binom{3}{3}$

9 blocks $= \frac{9!}{3!\,6!}\frac{6!}{3!\cdot3!} = \frac{9!}{3!3!3!} = \boxed{1680}$

37. Coefficient of $x^2y^2z^3$ in $(x+y+z)^7$;

Multinomial: $(X_1+X_2+X_3)^n = \sum \binom{n}{n_1 n_2 n_3} X_1^{n_1} X_2^{n_2} X_3^{n_3}$; $(x+y+z)^7 = \sum \binom{7}{2\,2\,3} x^2y^2z^3$

coefficient : $\binom{7}{223} = \frac{7!}{2!2!3!} = \frac{7\cdot6\cdot5\cdot4\cdot}{4}$

$\boxed{=210}$

38. coefficient of x^3y^4 for $(x+y)^7$;

$(x+y)^7 = \sum \binom{n}{k} x^k y^{7-k} = \binom{7}{0} x^7 y^0 + \binom{7}{6} x^6 y^1 + \cdots$

coefficient : $\binom{7}{34} = \frac{7!}{3!\,4!} = \frac{7\cdot6\cdot5}{3\cdot2}$

$= \frac{210}{6} = \boxed{35}$

39. a. 26 letters choose 6.

$\frac{\text{Probability Hamlet}}{\text{Total outcomes}} = \frac{\binom{6}{1}}{\binom{26}{6}} = \frac{6!}{5!}\left(\frac{6!(20)!}{26!}\right)$

$= \frac{3}{115115} = \boxed{0.0026\%}$

b. $0.90 = n\left(\frac{3}{115115}\right)$

$\boxed{n = 34534 \text{ monkeys}}$

40. 12 people into three groups : $\binom{12}{4}\binom{8}{4}\binom{4}{4} = \frac{12!}{4!(8)!}\frac{8!}{4!\,4!}\frac{4!}{4!\,1!} = \frac{12!}{4!4!4!} = 34650.$

6 pairs of partners : $\binom{6}{2}\binom{4}{2}\binom{2}{2} = \frac{6!}{2!\,4!}\frac{4!}{2!\,2!}\frac{2!}{2!\,0!} = \frac{6!}{2!2!2!} = 90.$

41. Seven black socks, eight blue socks, and nine green socks. Total : 24.

a) Probability of Matching: $\frac{\binom{7}{2}}{\binom{24}{2}} + \frac{\binom{8}{2}}{\binom{24}{2}} + \frac{\binom{9}{2}}{\binom{24}{2}} = \frac{7!}{5!}\frac{2!\,22!}{24!} + \frac{8!}{6!}\frac{2!\,22!}{24!} + \frac{9!}{7!}\frac{2!\,22!}{24!}$

$= \frac{7}{92} + \frac{7}{69} + \frac{3}{32} = \boxed{27\%}$

b. $7/92 = \boxed{7.61\%}$

42. Number of ways to choose : 11 boys grouped into 4 forwards, 3 midfielders, 3 defenders, 1 goalie.

$\binom{11}{4}\binom{7}{3}\binom{4}{3}\binom{1}{1} = \frac{11!}{4!(7!)}\frac{7!}{3!\,3!}\frac{4!}{3!\,1!}\frac{1!}{1!\,1!} = \boxed{46,200}$

43. Three jobs ; Two jobs require 3 programmers, the third requires four.

Total of ten programmers. $\binom{10}{3}\binom{7}{3}\binom{4}{4} = \boxed{4200}$

44. Combinations in Tentacles x Shaking Hands : $\sum_{i=1}^{n=8} 8(8-i)\binom{8}{i+1} =$: Many

$= 8\cdot7\binom{8}{2} + 8\cdot6\binom{8}{3} + 8\cdot5\binom{8}{4} + 8\cdot4\binom{8}{5} + 8\cdot3\binom{8}{4} + 8\cdot2\binom{8}{3} + 8\cdot1\binom{8}{1}$

$= 1568 + 2688 + 2800 + 1792 + 672 + 128 + 8 = 9656$

45. Prove $P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$

Multiplication Law: A & B be events and assume $P(B) \neq 0$, then $P(A \cap B) = P(A|B)P(B)$

$$P(A_n \cap A_{n-1} \cap \cdots \cap A_2 \cap A_1) = P(A_n | A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \cdot P(A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

$$= P(A_n | A_1 \cap A_2 \cap \cdots \cap A_{n-1}) P(A_{n-1} | A_1 \cap A_2 \cap A_3 \cdots) P(A_1 \cap A_2 \cap A_3 \cdots)$$

$$\boxed{= P(A_n | A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \cdots P(A_3|A_1 \cap A_2) \cdot P(A_2|A_1) \cdot P(A_1)}$$

48. cont. [Read wrong]

Urn A	Urn B
3xR, 2xW	2xR, 5xW

A ball from Urn A into B, then a ball is drawn from Urn B.

A) Probability of a red ball: $\frac{\binom{3}{1}}{\binom{5}{1}} \cdot \frac{\binom{3}{1}}{\binom{6}{1}} = \left(\frac{3!}{2!}\right)\left(\frac{4!1!}{5!}\right)\left(\frac{3!}{2!}\right)\left(\frac{5!}{6!}\right) = \frac{3}{10}$

$\frac{\binom{2}{1}}{\binom{5}{1}} \cdot \frac{\binom{2}{1}}{\binom{6}{1}} = \frac{2}{15}$

46 a)

Urn A	Urn B
3xR, 2xW	2xR, 5xW

coin [50/50]:
Heads = Urn A
Tails = Urn B.

$P(\text{coin} \cap \text{Urn A}) = P(\text{Coin}|\text{Urn A}) \cdot P(\text{Urn A}) = \frac{1}{2} \cdot \left(\frac{3}{5}\right) = \frac{3}{10}$

$P(\text{coin} \cap \text{Urn B}) = P(\text{coin}|\text{Urn B}) \cdot P(\text{Urn B}) = \frac{1}{2}\left(\frac{2}{5}\right) = \frac{2}{10}$

$P(R) = P(\text{coin} \cap \text{Urn A}) + P(\text{coin} \cap \text{Urn B}) = \frac{3}{10} + \frac{2}{10} = \frac{5}{10} = \boxed{\frac{1}{2}}$

b) $P(R) = P(\text{Heads}) P(\text{Urn A}) + P(\text{Tails}) P(\text{Urn B}) = 1 = P(\text{Heads})\left(\frac{3}{5}\right) + (1 - P(\text{Heads}))\left(\frac{2}{5}\right)$

$1 = P(\text{Heads})\left[\left(\frac{3}{5}\right) - \left(\frac{2}{5}\right)\right] + \frac{2}{5}$; $\boxed{P(\text{Heads}) = \frac{3}{5}}$

47.

Urn A	Urn B
4xR, 3xB, 2xG	2xR, 3xB, 4xG

a) $P(R) = P(R|\text{Urn B} \cap R) P(\text{Urn A} \cap R) + P(R|\text{Urn B} \cap B) P(\text{Urn A} \cap B) + P(R|\text{Urn B} \cap G) P(\text{Urn A} \cap G)$

Multiplication Law

$= \left(\frac{3}{10}\right)\left(\frac{4}{9}\right) + \left(\frac{2}{10}\right)\left(\frac{3}{9}\right) + \left(\frac{2}{10}\right)\left(\frac{2}{9}\right) = \left(\frac{12}{90}\right) + \left(\frac{6}{90}\right) + \left(\frac{4}{90}\right) = \frac{22}{90} = \boxed{\frac{11}{45}}$

b) ~~$1 = P(R|\text{Urn B} \cap R) P(\text{Urn A} \cap R) + P(R|\text{Urn B} \cap B) P(\text{Urn A} \cap B) + P(R|\text{Urn B} \cap G) P(\text{Urn A} \cap G)$~~

~~$1 = \left(\frac{3}{x}\right)\left(\frac{4}{9}\right) + \left(\frac{3}{10}\right)\left(\frac{3}{9}\right) + \left(\frac{2}{10}\right)\left(\frac{2}{9}\right)$; $\frac{10}{3}\left(1 - \left(\frac{2}{10}\right)\left(\frac{2}{9}\right) - \left(\frac{2}{10}\right)\left(\frac{2}{9}\right)\right) = x$~~

b) Bayes Formula: $P[\text{Urn A}(\text{Red}) | \text{Urn B}(R)] = \frac{P[\text{Urn A}(R), \text{Urn B}(\text{Red})]}{P[\text{Urn B}(\text{Red})]} = \frac{P[\text{Urn B}(\text{Red})|\text{Urn A}(\text{Red})] P[\text{Ur}}{P(R|\text{Urn B} \cap B) P(\text{Urn A} \cap B) + P(R|\text{Urn B} \cap R) P(R) + P(R|\text{Urn B} \cap G) P(G)}$

$\boxed{= 6/11}$

48.

Urn A
3xR, 2xW

1 Draw + 1 Return + Same color Ball. → 2nd Draw

Multiplication Law ... $P(\ldots A)$

a) Probability of White?

$P(W|\text{Draw \#2}) = P(W|\text{Draw\#2} \cap W) P(W \cap \text{Draw\#1}) + P(W|\text{Draw\#2} \cap R) P(R \cap \text{Draw\#1})$

$= \left(\frac{3}{6}\right)\left(\frac{2}{5}\right) + \left(\frac{2}{6}\right)\left(\frac{3}{5}\right) = \left(\frac{6}{30}\right) + \left(\frac{6}{30}\right) = \frac{12}{30} = \frac{4}{10} = \boxed{\frac{2}{5}}$

b). Bayes Theorem!

$P(\text{Draw\#2} \cap W | \text{Draw \#2} \cap W)$

$= \frac{P(\text{Draw \#2} \cap W | \text{Draw\#1 W}) P(\text{Draw\#1 W})}{P(W|\text{Draw\#2} \cap W) P(W \cap \text{Draw\#1}) + P(W|\text{Draw \#2} \cap R) P(R \cap \text{Draw \#1})}$

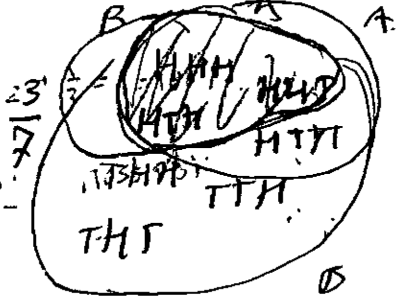
$= \frac{(3/5)(2/5)}{\left[\left(\frac{3}{6}\right)\left(\frac{2}{5}\right) + \left(\frac{2}{6}\right)\left(\frac{3}{5}\right)\right]} = \boxed{1/2}$

49. 3 tosses of a coin

a) $P(A|B)$ $\frac{P(A\cap B)}{P(B)} = \frac{P(HHH + HHT + HTH)}{P(HHH + HHT + HTH + HTT + THH + TTH + THT)} = \frac{3}{7}$

b) $P(T|H) = \frac{HHT + HTH + HTT + THH + THT + TTH}{HHH + HHT + HTH + HTT + THH + TTH + THT} = \boxed{\frac{6}{7}}$

$P(T) = \cancel{(Prob\ T)\cdot(Prob\ T)(Prob\ T)}$ $\frac{1}{2}$



50. Two dice; sum total $= 6$: $P(6) = \frac{P(6\cap 3)}{P(6)} = \frac{1/36}{5/36} = \frac{1}{36}\cdot\frac{36}{5} = \boxed{\frac{1}{5}}$

$\boxed{\text{Law of Independent Events}}$

51. Two dice; sum total $= 6$; $P(<6) = \frac{P(<6\cap 3)}{P(<6)} = \frac{4}{10} = \boxed{\frac{2}{5}}$

52. $P(G|G) = \frac{P(G\cap AL)}{P(L)} = \boxed{\frac{1}{4}}$; $P(G|G) = \frac{P(G\cap G)}{P(G)} = \boxed{\frac{1}{4}}$

53. High-Risk $[0.02]$ $[0.10]$ 2×10^{-3} High R $P(\text{High Risk}|\text{People}) = \frac{P(\text{High Risk}\cap\text{People})}{P(\text{Risk}\cap\text{People})} = \boxed{\frac{8}{23}}$

Medium Risk $[0.01]$ $[0.20]$ 2×10^{-3}

Low Risk $[0.0025]$ $[0.70]$ 1.75×10^{-3}

(Filing) (People)

0.0325 1.00

54. R_i = event of rain on day i;

$P(R_i|R_{i-1}) = \alpha$; $P(R_i^c|R_{i-1}^c) = \beta$

$P(R_i|R_{i-1}\cap R_{i-2}\cap\cdots\cap R_0) = P(R_i|R_{i-1})$

a) p = probability of rain, What about tomorrow?

$P(R_{i+1}|R_0) = P(R_i|R_{i0}) = P = x$

b) $P(R_2|R_{i1}\cap R_0) = P(R_2|R_{i1}) = x$

c) $P(R_i|R_{i-1}\cap R_{i-2}\cap\cdots\cap R_0) = P(R_i|R_{i-1})$

$\boxed{\lim_{i\to\infty} = P\cdot\prod_{i=1}^{n}\alpha_i}$

55. Upper (U), middle (M), and lower (L)

1 = Father occupation: 2 = Son's occupation.

	U_2	M_2	L_2
U_1	0.45	0.48	0.07
M_1	0.05	0.70	0.25
L_1	0.01	0.50	0.49

$P(U_2|U_1) = 0.45$

a) $P(M_1|M_2) = 0.70$; $P(L_1|L_2) = 0.49$

b)

	U_3	M_3	L_3
U_2	$P(U_3|U_2)P(U_2)$	$P(M_3|U_2)P(U_2)$	$P(L_3|U_2)P(U_2)$
M_2	$P(U_3|M_2)P(M_2)$	$P(M_3|M_2)P(M_2)$	$P(L_3|M_2)P(M_2)$
L_2	$P(U_3|L_2)P(L_2)$	$P(M_3|L_2)P(L_2)$	$P(L_3|L_2)P(L_2)$

	U_3	M_3	L_3
	V_2	42/25	81/100
U_2	0.225	0.3064	0.0567
M_2	0.025	1.176	0.2025
L_2	0.005	0.84	0.3969

$P(U_2) = P(U_2|U_1)P(U_1) + P(U_2|M_1)P(M_1) + P(U_2|L_1)P(L_1) = 0.050$

$P(M_2) = 0.5785$

$P(L_2) = 0.3652$

$P(U_3) = 0.064$

$P(M_3) = 0.614$

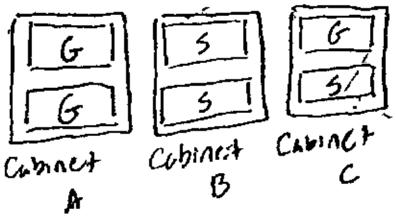
$P(L_3) = 0.322$

56. 5 cards of 52 card Deck

1^{st} = King. $\boxed{\text{Law of Independent Events}}$

$\frac{3\binom{51}{3}}{\binom{52}{4}} = \frac{\frac{13\cdot 51}{11,881}}{\frac{51}{41,491}} = \frac{1}{2} = \boxed{50\%}$

57. Cabinet A, B, C with two drawers each, inside a coin. $\boxed{\text{Multiplicative Law}}$ $P(E) = P(S|S_1)P(S_1) + P(S_2|G_1)P(G_1)$



$P(\text{Draw \#2} | \text{Draw \#1}) = \ldots$

$\left(\frac{2}{3}\right)\left(\frac{1}{2}\right) = \ldots P(B) + P(\text{Draw \#}|A) \cdot P(A)\ \frac{1}{2} + P(\text{Draw \#1}|C)\,P(C)$

$$\frac{1}{2} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{3} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right)} = \frac{\frac{1}{3}}{3/6} = \frac{1}{3} \times \frac{6}{3} = \boxed{\frac{2}{3}}$$

58. Drew, Chris, Jason; Two must stay home; one leaves

$\boxed{\text{Drew, chris, Jason}}$ → 2 remain, 1 leave $\left(\frac{1}{3}\right)$

3 slots

The possibilities following Drew asking the teacher contain a relationship known as the multiplicative law. If Chris is chosen $\left(\frac{1}{3}\right)$ to remain $\left(\frac{1}{3}\right)$, then there are $\left(\frac{1}{2}\right)$ possible outcomes. If Jason is chosen $\left(\frac{1}{3}\right)$ then there is $\left(\frac{1}{2}\right)$-outcomes. While if Drew is chosen 1 outcome. Thus,

$P(\text{Drew Asking}) = P(\text{Outcomes}|\text{Drew})P(\text{Drew}) + P(\text{outcomes}|\text{Chris}) \times P(\text{Chris}) + P(\text{outcomes}|\text{Jason})P(\text{Jason})$

$= \frac{1}{3}(1) + \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) = \frac{1}{3}$

Asking is no better than writing

59. | HH |
|---|
| HT |
| T:T |

Random selection
Flip
Heads

BOX

a) Probability of a two-headed coin

$P(HH|H) = \frac{P(HH|HH)\,P(H)}{P(HH|HH)P(H) + P(HT|HT)P(HT)}$

$$P(HH|H) = \frac{P(H|HH)P(HH)}{P(H|HH)P(HH) + P(H|HT)P(HT) + P(H|TT)P(T)} = \frac{1 \cdot 1/3}{1 \cdot \frac{1}{3} + \frac{1}{2}\left(\frac{1}{3}\right) + 0\left(\frac{1}{3}\right)} = \frac{1/3}{\frac{1}{3} + \frac{1}{6}} = \frac{\frac{2}{6}}{\frac{2}{6} + \frac{1}{6}} = \boxed{\frac{2}{3}}$$

b) $P(H) = P(H|HH)P(HH) + P(H|HT)P(HT) + P(H|TT)P(T) = \frac{3}{6} = \boxed{\frac{1}{2}}$; $P(T) = P(T|HT)P(HT) + P(T|TT)P(TT)$
$= \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + 1 \cdot \frac{1}{3} = \frac{3}{6} = \boxed{\frac{1}{2}}$

c) $P(H_2) = P(H|HH)\,P(HH_1) + P(H|HH_1)P(HT_1)$
$= 1 \cdot \left(\frac{2}{3}\right) + \frac{1}{2}\left(\frac{1}{3}\right) = \frac{2}{3} + \frac{1}{6} = \frac{4}{6} + \frac{1}{6} = \boxed{\frac{5}{6}}$

60. $P(B) > 0$; $Q(A) = P(A|B)$; $\boxed{\text{Addition Law}}$ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\boxed{P(A \cup B|B) = P(A|B) + P(C|B) - P(A \cap C|B)}$

61. Defect Accuracy $= 0.95$ Sound Accuracy $= 0.97$

If 0.5% are faulty, what is the probability faulty is sound?

	Defect	Sound
T	0.95	0.97
F	0.05	0.03
	1	1

1.03

$$P(F|\text{Defect}) = \frac{P(F|\text{Defect})\,P(\text{Defect})}{P(F|\text{Defect})P(\text{Defect}) + P(F|\text{Sound})P(\text{Sound})} \ldots \frac{1}{800}$$

$$= \frac{0.05\,(0.5/200)}{0.05\,(5/200) + 0.03\,(195/200)}; \quad P(\text{Sound}|\text{Defect}) = \frac{P(F \cap \text{Defect})}{P(\text{Defect})} = \frac{P(F|\text{Sound})\,P(\text{Sound})}{P(F|\text{Sound})P(\text{Sound}) + P(T|\text{Defect})\,P(\text{Defect})}$$

$$= \frac{0.03 \cdot 0.995}{0.03\,0.995 + 0.95\,(0.005)} = 86\%$$

62. Four players [13 cards each] $\frac{52!}{4(13!)} = \boxed{8 \times 10^{67}} =$

63. Age > 70 = 0.6 ; Age > 80 = 0.2. $P(Age>80 \mid 70) = \frac{P(Age>80 \cap >70)}{P(70)} = \frac{P(>80)}{0.6} = \boxed{\frac{0.2}{0.6} = \frac{1}{3}}$

64. Three Shifts: 1% of Shift 1 are defective; 2% of Shift 2 are defective; 5% of shift 3 are defective;

$P(\text{Defective}) = P(\text{Defective} \mid \text{Shift}\#1)P(\text{Shift}\#1) + P(\text{Defective} \mid \text{Shift}\#2)P(\text{Shift}\#2) + P(\text{Defective} \mid \text{Shift}\#3)P(\text{Shift}\#3)$

$= 1\%\left(\frac{1}{3}\right) + 2\%\left(\frac{1}{3}\right) + 5\%\left(\frac{1}{3}\right) = \boxed{2.667\%}$; $P(\text{Shift}\#3) \Rightarrow \frac{P(\text{Defective} \cap \text{Shift}\#3)}{\sum P(\text{Defective} \mid \text{Shift}\#i)} = \frac{5\%\left(\frac{1}{3}\right)}{2.667\%} = \boxed{62\%}$

65. A and B are independent; A and B^c, A^c and B^c are too.

$P(A \cap B) = P(A)P(B)$; $P(A \cap B^c) = P(A)P(B^c)$; $P(A^c \cap B^c) = P(A^c)P(B^c)$

66. $\emptyset$ independent of A for any A. $P(A \cap \emptyset) = \boxed{P(A) \cdot P(\emptyset) = 0}$

67. If $P(A \cap B) = P(A)P(B)$; then $P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B)$

Addition Law   Law of Independence

68. If $p(A \cap B) = p(A)p(B)$ and $p(B \cap C) = p(B)p(C)$; $p(A \cap C) = p(A \cap B \cap C) = p(A \cap B)p(C)$
$= p(A)p(B \cap C) = p(A)p(B)p(C)$

$P(A \cap B) = p(A)\frac{p(B \cap C)}{p(C)} \Rightarrow \boxed{\frac{p(A \cap B)}{p(B \cap C)} = p(A \cap C) = \frac{p(A)}{p(C)}}$

69. If $A \cap C = \emptyset$ "Disjoint", $p(A) = \emptyset \vee p(C) = \emptyset$; then independent

70. If $A \subset B$; then they are not independent.

71. If A, B, C are mutually independent, then $A \cap B$ and C are independent along with $A \cup B$ and C.

$p(A \cap (B \cap C)) = p(A)p(B \cap C) = p(A)p(B)p(C)$

$p(A \cup (B \cap C)) = P(A) + P(B \cap C) - P(A \mid B \cap C)$

$\boxed{= P(A) + P(B)P(C) - P(A \mid B \cap C)}$

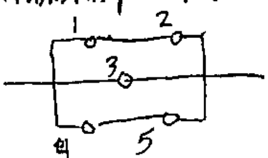
72. $(t = 0, 1, 2 \cdots)$; p, then q. @ $t=0, n=1$;

Probability of 0, 1, 2, 3 people at $t=2$.

73. n independent units, each with ~~P(Fail)~~ p; $P(R^c)$ = System failure. $P(\text{System}) = (1-p)^n$; $P(\text{Fail}|\text{System}) = 1-(1-p)^n$

$\boxed{= 1-(1-p)^n}$

74. Probability of failure

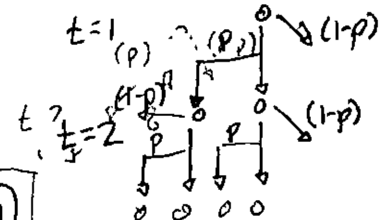


$P(F) = P(A_1 \cap A_2) + P(A_3) + P(A_4 \cap A_4)$

$+ P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5)$

$\boxed{= p + 2p^2 - p^5}$

75.



$0.5 = -3p + 2p^2$

$2p^2 - 3p - 0.5 = 0$

$\boxed{p_1 = 1.65}$; $p_2 = -0.15$

~~$P(\text{Dead}) = 1-(1-p)^3$~~ $= 1^3 - a^3 - 3a^2b + 3ab^2 - b^2 = 1-(a-b)^3$

$= 1 - 1 - 3p + 3p^2 - p^2 = \boxed{-3p + 2p^2}$

$t = 3$

0.5

76. $-o-||-o-||-o-$ n-components $P(\text{Success}) = \boxed{(1-P(E))^n}$; $P(\text{Success}) = (1-0.05)^{10} = 0.9975 = \boxed{99.75\%}$

77. $5\% = P(\text{Bulls-Eye})$ $0.5 = (1-0.05)^n$; $\log\frac{1}{2} = n\log(1-0.05)$; $\boxed{n = 13.51}$

78. Pairs of (a or A); AA, Aa, aa, or (Aa or aA). a) Parent #1: AA; Parent #2: Aa; Offspring: AA, AA, Aa, Aa $\boxed{AA\ (50\%),\ Aa\ (50\%)}$

b) $AA\ (p)$; $Aa\ (2q)$; $aa\ (r)$; $1 = (p+2q+r)^n$; $\boxed{n=2;\ 1=(p+2q+r)^2}$; $n=3$; $\boxed{1=(p+2q+r)^3}$

c) $\boxed{1 = (u+v+w)^2}$; $\boxed{1 = (u+v+w)^3}$ Hardy-Weinberg Law

79. a. aa = Dead; Aa = carrier, alive; AA = not carrier, not diseased. $Aa \times Aa = AA + 2Aa + aa$

$\boxed{AA\ (25\%);\ Aa\ (50\%);\ aa\ (25\%)}$

b. $P(\text{Not Diseased} \cap \text{Carrier}) = 50\%$

c. $P(\text{Offspring}) = \cancel{P(AA) + P(Aa)}$ $p\boxed{\frac{1}{3}} + \boxed{\frac{2}{3}}$

$P(AA) = \frac{1}{3}(1-p) + \frac{1}{6}p + \frac{2}{6}(1-p) + \frac{p}{6}$

$P(AA) = \boxed{\frac{3}{3} - \frac{p}{3}}$

d.

Offspring $P(Aa)$	AA-AA	AA-Aa	Aa-AA	Aa-Aa
AA	$\frac{1}{3}(1-p)$	$\frac{1}{2}(\frac{1}{3})p$	$\frac{1}{2}(\frac{2}{3})(1-p)$	$\frac{1}{4} \times \frac{2}{3} \times p$
Aa	0	$\frac{1}{2}(\frac{1}{3}) \times p$	$\frac{1}{2}(\frac{2}{3})(1-p)$	$\frac{1}{2} \times (\frac{2}{3}) \times p$
Aa	0	0	0	$\frac{1}{4} \times \frac{2}{3} \times p$

$P(Aa) = \frac{p}{6} + \frac{2}{6}(1-p) + \frac{2}{6}p$

$= \frac{p}{6}$

$P(AA)/P(AA) = \boxed{\dfrac{(2-p)\frac{1}{3}}{\frac{p}{6}}}$

80. Parents Aa (50%); A = child #1 & #2 have the same gene

B = child #1 and #3 have the same gene C = child #2 & #3 have same gene.

Mutually Independent: $P(A_{i_1} \cap \cdots \cap A_{i_m}) = P(A_{i_1}) \cdots P(A_{i_m})$

Pairwise Independence: $P(A \cap B \cap C) = 0 \neq P(A)P(B)P(C)$.

$P(B) = P(\text{child \#1} \cap \text{child \#3})$

$P(A) = P(\text{child \#1} \cap \text{child \#2})$

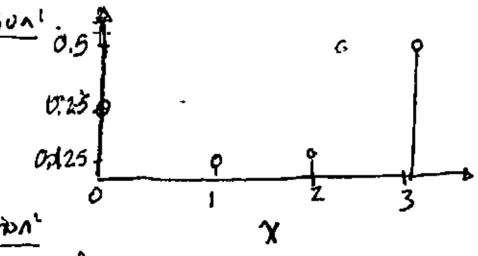
$P(C) = P(\text{child \#2} \cap \text{child \#3})$

$P(\#1 \cap \#2 \cap \#3) = P(\#1 \cap \#2)P(\#3) = \boxed{P(\#1)P(\#2)P(\#3)}$

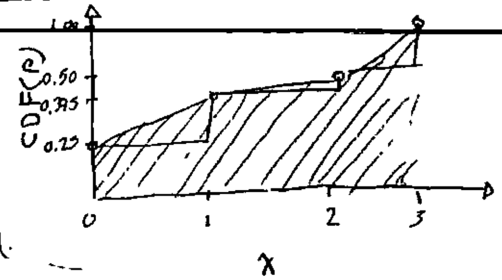
$\boxed{P(A \cap B \cap C) \neq P(A) \cdot P(B) P(C)} = P(\#1 \cap \#3)P(\#1 \cap \#2)P(\#2 \cap \#3)$

Chapter 2: 1. $P(X=0)=0.25,\ P(X=1)=0.125,\ P(X=3)=0.5$ Frequency Function:

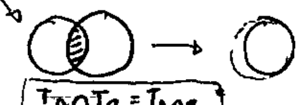
$P(X=2)=0.125$



Cumulative Distribution Function:



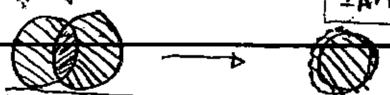
2. $I_{A\cap B}=I_A I_B=\min(I_A, I_B)$



"Perfectly overlapped"

$\boxed{I_A \cap I_B = I_{A\cap B}}$

$I_{A\cup B}=\max(I_A; I_b)$



"Perfectly overlapped"

$\boxed{I_{A\cup B}=\max(I_A, I_B) = I_A + I_B - I_A \cdot I_B}$

3. CDF → Frequency (Δ)

k	F(k)
0	0
1	0.1
2	0.3
3	0.7
4	0.8
5	1.0

CDF

k	P(X)
0	0
1	0.1
2	0.2
3	0.5
4	0.1
5	0.2

Frequency

4. Prove $p(k)=F(k)-F(k-1)$

$CDF = \sum_0^n p(x) = p(k) + \sum_1^n p(k)$

$F(k) = p(k) + F(k-1)$

$\boxed{p(k)=F(k)-F(k-1)}$

5. Show $P(u<X\le v)=F(v)-F(u)$ a) Discrete $F(k)=\sum_0 p(k) = p(k) + \sum_1 p(k)$

$F(k) = p(k) + F(k-1)$

$\boxed{P(k)=F(k)-F(k-1)}$

b. $F(k)=\int_0^n p(k)\,dk = p(k) + \int_1^n p(k)\,dk = p(k) + F(k-1)$

$\boxed{p(k)=F(k)-F(k-1)}$

6. a) Four tosses: ~~$P(T)=(\frac{1}{2})^4(1-\frac{1}{2})^{1-1}=\frac{1}{2}$, $P(T|2)=(\frac{1}{2})^2(1-\frac{1}{2})^{1-2}=\frac{1}{4}(2)$; $P(T|3)=(\frac{1}{2})^3(1-\frac{1}{2})^{1-3}=(\frac{1}{8})4$; $P(T|4)=(\frac{1}{4})^4(1-\frac{1}{2})^{1-4}$~~ $=\frac{8}{16}$

① Sample space (n=16)

Outcome	HHHH	HHHT	HHTH	HHTT	HTHH	HTHT	HTTH	HTTT	T x 8
Heads before Tails	4	3	2	2	1	1	1	1	0

② Frequency Function (n(X))

X	0	1	2	3	4
n(X)	8	4	2	1	1
$P(X)=\frac{n(X)}{16}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
F(X)	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{7}{8}$	$\frac{15}{16}$	$\frac{16}{16}$

① Sample Space: HHHH HHHT HHTH HHTT HTHH HTHT HTTH HTTT THHH THHT THTH THTT TTHH TTHT TTTH TTTT

b) the number of Heads following first tail

② Frequency Function

X	0	1	2	3	4
n(x)	1	1	2	4	8
$P(x)=\frac{n(x)}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$
F(x)	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1

c) (the number of heads) − (number of tails)

② Frequency Function

X	−4	−2	0	2	4
n(x)	1	4	6	4	1
$P(x)=\frac{n(x)}{16}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$
F(X)	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{11}{16}$	$\frac{15}{16}$	$\frac{16}{16}$

d) Frequency Function = (c)

7. Bernoulli Random Variable: $p(1)=p;\ p(0)=1-p;\ p(x)=0$; $p(x)=\begin{cases} p^x(1-p)^{1-x}; & x=0 \text{ or } x=1 \\ 0 & ;\text{ otherwise.}\end{cases}$

CdF of Bernoulli Random Variable: $F(x)=\sum_0^{x+y+z} p(x) = \sum_0^x p(0) + \sum_0^y p(1) + \sum_0^z p(x)$

$= \sum_0^x p + \sum_0^y 1-p + \sum_0^z 0 = \boxed{\sum_0^x p + \sum_0^y 1-p}$

8. P(9 Heads in 10 Tosses) $= \frac{\text{chance}}{\text{outcome}} = \frac{10}{2^{10}} = \frac{10}{1024} = \boxed{\frac{5}{512}}$ ✓

P(18 Heads in 20 Tosses) $= \frac{\binom{20}{2}}{2^{20}} = \frac{201}{2^{!}\,181} = \boxed{\frac{95}{524288}}$

1048576

9. $n=3; p=\sum_0^3 \binom{n}{k} p^k (1-p)^{n-k} = \binom{3}{0}p^0(1-p)^3 + \binom{3}{1}p^1(1-p)^2 + \binom{3}{2}p^2(1-p)^1 + \binom{3}{3}p^3(1-p)^0 = (1-p)^3 + 3p(1-p)^2 + 3p^2(1-p)^1 + bp^3 = 1; p=\frac{n+1}{n-1}$ when

$n=3; p=\sum_0^2 \binom{3}{k} p^k (1-p)^{n-k} = \binom{3}{0}p(1-p)^3 + \binom{3}{1}p(1-p)^2 + \binom{3}{2}\ldots(1-\ldots) = (1-p)^3 + 3p(1-p)^2 + (1-p)^3 + 3p(1-p)^2 > 1-p$

$p<0.5$

10. Player A: $p_1 = P(\text{success})$; Player B: $p_2 = P(\text{success})$ a) $P(X) = \ldots$ [illegible] ; $r \cdot n$ [illegible]

b. $P(\text{Player A Win}) = \dfrac{p_1}{1-(1-p_1)(1-p_2)} = \boxed{p_1 \sum [(1-p_1)(1-p_2)]^i}$ odd: $p(k) = (1-p_1)^{k/2} \cdot (1-p_2)^{\frac{k}{2}-1} \cdot p_2$

11. Binomial Distribution: $P(X) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$; Mode: $p'(x) = 0 = k p^{k-1}(1-p)^{n-k} + (n-k) p^k (1-p)^{n-k-1}$

$= k p^{k-1}(1-p)^{n-k} + n \cdot p^k (1-p)^{n-k} - k p^k (1-p)^{n-k}$

$= k\left(\frac{1}{p}\right) \ldots (n-k)$; $k\left(\frac{1}{p} - 1\right) + n - k = 0$

$\frac{k}{p} - k + n - k = 0$

$\frac{k}{p} + \left(\frac{n}{k}\right) - 2 = 0$

$k \ldots = n - 2k$

$p = \dfrac{k}{n + 2k}$ [illegible]

12. Prove $p(x) = \sum_0^1 \binom{1}{k} p^k (1-p) = 1$

$= \binom{1}{0} p^0 (1-p)^1 + \binom{1}{1} p^1 (1-p)^0$

$= (1-p) - p = 1$

13. 20 items [4-choices] - Elimination of one, remainder of three
- Passing is 12 or more, correct.

a. $P(\text{Pass}) = \dfrac{P(\text{correct})}{\text{Total Outcomes}} = \dfrac{1/20}{\binom{3}{1}/\binom{20}{1}} = \dfrac{1}{3 \cdot 20} = \boxed{\dfrac{1}{60}}$

b. $P(\text{Pass}) = \dfrac{P(\text{correct})}{\text{Total outcomes}} = \dfrac{1}{\binom{2}{1}/\binom{20}{1}} = \dfrac{1}{2 \cdot 20} = \boxed{\dfrac{1}{40}}$

14. 4 bits + 3 bits $P(\text{change}) = 0.05$; Mutually Independent.

$P(\text{change} \mid 7 \text{ bits}) = \prod^{7} p(\text{change}_i) = p(\text{change})^7 = \boxed{0.05^7}$

$P(\text{change} \mid 4 \text{ bits}) = \prod_i^4 p(\text{change}_i) = p(\text{change})^4 = \boxed{0.05^4}$ $P(\text{correct}) = 1 - P(\text{change}) = 1 - 0.05^4$

$\boxed{P(\text{correct}) = 1 - 0.05^4}$

15. $P(\text{Winning Team A}) = 0.4$; Better advantage at 3 of 5 games or 4 of 7 games. 0.6 P correct + P(fail)

$\boxed{P(3 \text{ of } 5) = \prod_1^3 p(\text{Winning Game}_i) = (0.4)^3}$; $P(4 \text{ of } 7) = \prod_1^4 p(\text{Winning Game}_i) = (0.4)^4$

16. $n \to \infty$ and $r/n \to p$ and $m = \text{constant}$. Hypergeometric Function: $P(X=k) = \dfrac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$

$\lim_{n\to\infty} \lim_{r/n \to p} P(X=k) = \lim_{n\to\infty} \lim_{r/n\to p} \dfrac{\frac{r!}{k!(r-k)!} \frac{(n-r)!}{(m-k)!(n-r-m+k)!}}{\frac{n!}{m!(n-m)!}} = \binom{m}{k} \dfrac{[r(r-1)\cdots(r-k+1)][(n-r)\cdots(n-r-m+k+1)]}{[n(n-1)\cdots(n-m+1)]}$

$= \binom{m}{k} \left(\frac{1}{q}\right)^k \left(\frac{q-1}{q}\right)^{m-k}$; where $q = \frac{1}{p} = \frac{n}{r}$

$= \binom{m}{k} p^k (1-p)^{m-k}$

17. Bernoulli Trials ; $p(\text{success}) = p$; Failures till first round are counted.

Frequency Function: $\boxed{p(\text{Failure}) = p \prod_1^n (1-p)}$

18. Frequency Function $p(\text{Failure}) = p \prod^{r} (1-p) = p(1-p)^n$

19. CDF: $\sum p(1-p)^{x-1} = 1$; [illegible] CDF

(20) Minimum Trials: $n+1$; Maximum trials: $2n+1$

$P(\text{left}) = \frac{1}{2}$; $P(k \mid \text{other box}) = \binom{n+1}{n+1} p^{n+1} q^{0} + \binom{n+1}{n+1} p^{0} q^{n+1} = p^{n+1} + q^{n+1} = 2\left(\frac{1}{2}\right)^{n+1} = \left(\frac{1}{2}\right)^{n}$

$P(k-1 \mid \text{other box}) = \binom{n+1}{n} p q^{n} + \binom{n+1}{n} p q^{n} = 2\binom{n+1}{n}\left(\frac{1}{2}\right)^{n+1} = \binom{n+1}{n}\left(\frac{1}{2}\right)^{n}$

$P(k = n-r) = \binom{n+r}{n}\left(\frac{1}{2}\right)^{n+r}$; $P = \binom{n+n-k}{n}\left(\frac{1}{2}\right)^{2n-k}$

21. X = Geometric Random Variable $P(X > n+k-1 \mid X > n-1) = P(X > k)$

$$P(\text{Hypergeometric Function}) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

$P(X > n+k-1 \mid X > n-1)$

$= P(X+1-n > k \mid X+1 > n)$

$= P(X-n > k \mid X-n > 0)$

$= \boxed{P(X > K \mid X > 0) = P(X > k)}$

(22). X: Geometric random variable

$p = 0.5$; $P(X \le k) \approx 0.99 = 1-(1-p)^{k}$

$(0.5)^{k} = 0.01$

$k \log(0.5) \approx \log(0.01)$; $k = 6.6438$

$\boxed{k = 7.}$

(23) p(success) ; r successes before k^{th} failure ; Binomial $\boxed{p(k) = \binom{k+r-1}{k-1} p^{r}(1-p)^{k}}$

Total Number of trials : $(k+r)$

Last trial probability : $(1-p)$

Binomial : $\boxed{\binom{k+r-1}{k-1} p^{r}(1-p)^{k-1}}$

$p(\text{success}) = \binom{K+r-1}{K-1} p^{r}(1-p)^{k-1}(1-p)$

$\boxed{= \binom{K+r-1}{K-1} p^{r}(1-p)^{k}}$

(24) $H = 2^{3}$; $P(\text{Same}) = P(HHH) + P(TTT)$

$= \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$

$P(k) = (1-p)^{k-1}(p) = (0.75)^{k-1}(0.25)$

$\sum_{k=1}^{\infty} (0.75)^{k-1}(0.25) = 1$

$P(X > 3) = 1 - P(X \le 3)$

$= 1 - \{P(X=1) + P(X=2) + P(X=3)\}$

$= 1 - \{(0.75)^{1-1}(0.25) + (0.75)^{2-1}(0.25) + (0.75)^{3-1}(0.25)\}$

$= 1 - \{0.25 + 0.1875 + 0.140625\}$

$\boxed{\approx 0.4219}$

25. $P(\text{Royal Straight Flush}) = 1.3 \times 10^{-8}$

$n = 100$ hands/week ; 52 weeks/year ; 20 years $= 1.04 \times 10^{5}$

a) Time Sequence: Poisson Distribution

$\lambda = n \cdot p = 1.04 \times 10^{5} \cdot 1.3 \times 10^{-8} = 1.35 \times 10^{-3}$

$p(k) = \frac{\lambda^{k} e^{-\lambda}}{k!}$; $P(0) = \frac{\lambda^{0} e^{-\lambda}}{0!} = e^{-1.35 \times 10^{-3}}$; $\boxed{9.986 \times 10^{-1}}$

① State Space

② Frequency Function

③ $1 - p(1)p(2)p(3)$

b). $2 = \prod p^{k}(1-p)^{n-k}$; $P(2) = \frac{\lambda^{2} e^{-\lambda}}{2!} = \frac{(1.35 \times 10^{-3})^{2} e^{-1.35 \times 10^{-3}}}{2!} = \boxed{9.10 \times 10^{-7}}$

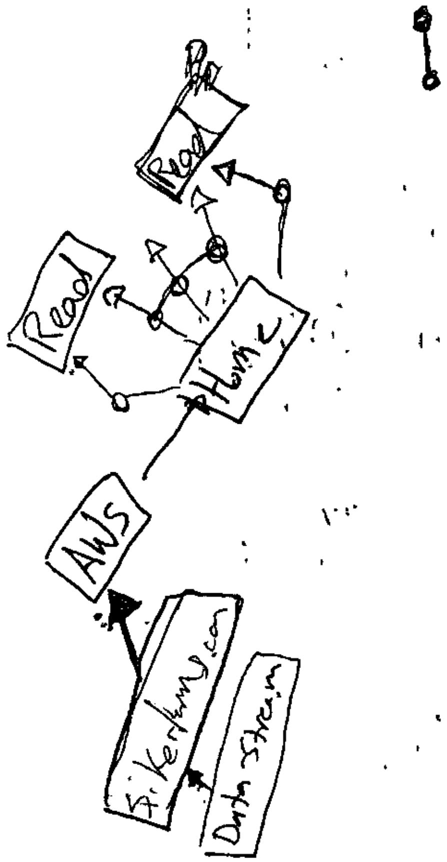
26. $\frac{1}{10{,}000}$ chance of being trapped. $n = 5 \frac{\text{days}}{\text{Week}} \cdot \frac{52\ \text{weeks}}{\text{Year}} \cdot 10$ years $= 2{,}600$; $\lambda = np = \boxed{\frac{13}{50}}$

$P(0) = \frac{\lambda^{0} e^{-\lambda}}{0!} = e^{-13/50} = \boxed{7.71 \times 10^{-1}}$; $P(1) = \frac{\left(\frac{13}{50}\right)^{1} e^{-13/50}}{1!} = \boxed{2.0 \times 10^{-1}}$; $\boxed{p(2) = 0.026}$

- Randomize Router [IP] ☐
- Randomize Mac Address ✓
- Randomize external IP. ✓
- VPN through aws
- Tor for browsing

  ↳ IP Scan - ping

- Tor guard -



27. $P(\text{Disease}) = \frac{1}{10{,}000}$; $n = 100{,}000$ people

Poisson Distribution

$k = 0$ cases ; $P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$; $p(0) = \frac{\lambda^0 e^{-10}}{0!}$; $\boxed{4.54 \times 10^{-5}}$

$k = 1$ cases : $p(1) = \frac{\lambda^1 e^{-\lambda}}{1!} = \frac{(10)^1 e^{-10}}{1!} = \boxed{4.54 \times 10^{-4}}$

$k = 2$ cases : $p(2) = \frac{\lambda^2 e^{-\lambda}}{2!} = \frac{(10)^2 e^{-10}}{2!} = \boxed{2.26 \times 10^{-2}}$

28. CDF $= F(k) = p_0, p_1, \ldots, p_n$; n, and p. $q = 1 - p$

Prove the binomial probabilities by $p_0 = q^n$.

$$p_k = \frac{(n-k+1)p}{kq} p_{k-1} \quad ; \quad k = 1, 2 \ldots n$$

$\boxed{p_0 = (1-p)^n}$; $p_1 = n \frac{p}{(1-p)} (1-p)^n$; $p_2 = \frac{(n-1)p}{2(1-p)} p_1 = \frac{(n-1)n}{2} p^2 (1-p)^n$

$p_1 = \frac{(n - 1 + 1)p}{1(1-p)} p_0 = \boxed{= n(p(1-p)^{n-1})}$; $p_2 = \frac{(n-1)p}{2(1-p)} p_1 \ldots \boxed{= \frac{(n^2 - n)}{2} p^2 (1-p)^{n-2}}$

Recursive Binomial Distribution $= \frac{n}{(n-k)} p (1-p)^n$

$$\boxed{= \prod \frac{(n-k+1)!}{k!} p^n (1-p)^{n-k}}$$

$p(X \leq 4)$ for $n = 9000$ and $p = 0.0005$

$$= \frac{(9000-3)(9000-2)(9000-1)(9000)}{4!} (0.0005)^{9000} (1 - 0.0005)^{9000-4}$$

$\boxed{= 1.89 \times 10^{-1}}$ $\cong 0.00$

As a Poisson

$n = 9000$; $p = 0.0005$; $np = 9/2$

$P(4) = \frac{(9/2)^4}{4!} e^{-9/2} = \boxed{1.89 \times 10}$

29. $p_0 = \exp(-\lambda)$

$$p_k = \frac{\lambda}{k} p_{k-1} \; ; \; k = 1, 2 \ldots$$

$p_0 = \exp(-\lambda)$; $p_1 = \lambda \cdot p_0 = \lambda \exp(-\lambda)$; $p_2 = \frac{\lambda^2}{2} \exp(-\lambda)$; $p_k = \frac{\lambda^k}{k!} \exp(-\lambda)$

$P(X \leq 4)$; $\lambda = 4.5$; $p_k = \frac{(4.5)^4}{4!} \exp(-4.5) = \boxed{1.89 \times 10^{-1}}$

30. Poisson Frequency Function ; $p(k) = \frac{\lambda^k \exp(-\lambda)}{k!}$; $p'(k) = \frac{k \cdot \lambda^{k-1}(-\lambda)\exp(-\lambda)}{k!} = 0$

$-\lambda = 1$

$\boxed{\lambda = 1}$

Sources cite' ratio: $\frac{P(X=k+1)}{P(X=k)} = \frac{\lambda^{k+1}\exp(-\lambda)/(k+1)!}{\lambda^k \exp(-\lambda)/k!}$

$= \lambda / k+1$

"Not logical because mix of Poisson changes its pr-changes

There are maximum and minimums to this problem set. $\lambda < 1$, $\lambda > 1$ (int), $\lambda > 1$ (Rational) $\lambda = np = 1.5$

Probability Decreases

Max: $k=0$

Fractions and Factorial approx to one

31. $\lambda = 2$ per hour a) 10-min shower; $p(\text{phone rings}) = \frac{(2)^{1/6}\exp(-2)}{1/6} = \boxed{0.277}$

b) $p(\text{phone rings}) = 0.5 = \frac{\delta^0 e^{-\delta}}{0!}$; $\exp(2) \; 69.3$ min

Time $= \frac{60 \text{ min}}{2 \text{ phone calls}} \cdot 0.693$

$= \boxed{20.79 \text{ min}}$

32. $\lambda = 0.33$ per month

a) $k=0$; $p(0) = \frac{(1/3)^0 \exp(-1/3)}{0!} = \exp(-1/3) = \boxed{0.716}$

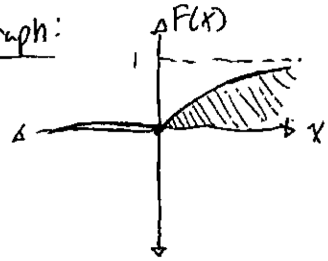
$k=1$; $p(1) = \frac{(1/3)^1 \exp(-1/3)}{1!} = \boxed{0.239}$

$k=2$; $p(2) = \frac{(1/3)^1 \exp(-1/3)}{2!} = \boxed{0.004}$

The most probable number of suicides would be at $k=0$ because $\lambda < 1$ and demonstrates a decreasing probability.

33. $F(x) = 1 - \exp(-\alpha x^\beta)$ for $x \geq 0, \alpha > 0, \beta > 0$, $F(x) = 0$ for $x < 0$.

Graph:

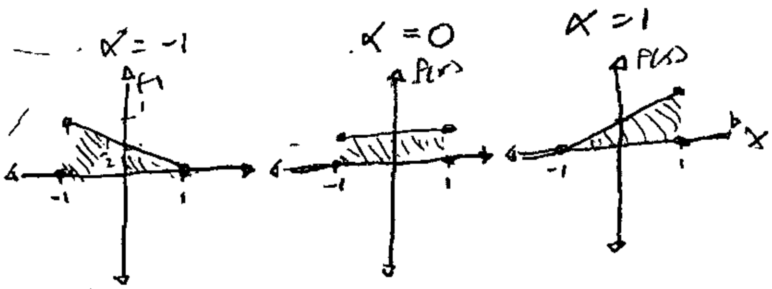


$p(x) = \frac{d}{dx} F(x) = \boxed{\alpha \beta x^{\beta-1} \exp(-\alpha x^\beta)}$

Cumulative Density Function:

$\lim_{x \to -\infty} F(x) = 0$; $\lim_{x \to \infty} F(x) = 1$

34. $f(x) = (1 + \alpha x)/2$ for $-1 \leq x \leq 1$ and $f(x) = 0$



Probability Density Function Requires

$P(X = x_i) = p(x_i)$; $\sum_i p(x_i) = 1$

$F(x) = \sum_{-1}^{1} (1 + \alpha x)/2 = \int_{-1}^{1} (1 + \alpha x)/2 \, dx = \frac{x}{2}\Big|_{-1}^{1} + \frac{\alpha x^2}{4}\Big|_{-1}^{1} = \frac{1}{2} + \frac{1}{2} + \frac{\alpha x^2}{4} - \frac{\alpha x^2}{4} = \boxed{1}$

$\boxed{F(x) = \frac{x}{2} + \frac{\alpha x^2}{4}}$

35. $p(x) = \begin{cases} -1 & -1 \leq x \leq 0 \\ 2 & 0 < x < 1 \end{cases}$

$F(x) = \int_{-1}^{0} -1 \, dx + \int_{0}^{1} 2 \, dx = 3$

36. U is uniform [0, 1].

$X = [nU]$, where [t]

"Probability Mass Function"

37. $p(X \leq 1/3) = \frac{1}{3}$;

$p(X \geq 2/3) = 2/3$

38. $\alpha f + (1 - \alpha) g$; $\sum_i P(x_i) = \alpha f + (1 - \alpha) g$

where $0 \leq \alpha \leq 1$

max $\frac{d}{d\alpha} \alpha f + (1 - \alpha) g = f - g$

$F(\alpha) = \frac{\alpha^2}{2} f + (\alpha - \frac{\alpha^2}{2}) g$

39. Cauchy Cumulative Distribution: $F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x), \ -\infty < x < \infty$

a. Cumulative Distribution Requirements: $\lim_{x\to-\infty} F(x) = 0$; $\lim_{x\to\infty} F(x) = \boxed{1}$

$\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x) = \frac{1}{2} - \frac{1}{2} = \boxed{0}$; $\lim_{x\to\infty} \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x) = \frac{1}{2} + \frac{1}{2} = \boxed{1}$

b. $p(x) = F'(x) = \frac{d}{dx} \ldots \boxed{\frac{1}{(1+x^2)\pi}}$ — $\cos^2 x$; $\sec^2$

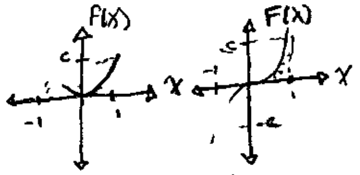
c. $P(X > x) = 0.1$; $0.1\pi(1+x^2)\pi x = 1$

$(1+x^2) = \frac{10}{\pi} - 1$

$x^2 = \frac{10-\pi}{\pi}$

$\boxed{x = \sqrt{\frac{10-\pi}{\pi}}}$

40. $f(x) = cx^2$ for $0 \le x \le 1$ and $f(x) = 0$ otherwise



a) $F(1) = c$; $F(0) = 0$; $c = \frac{F(1) - F(0)}{1 - 0} = \Delta F(x)$

b) $F(x) = \frac{cx^3}{3}$ c) $P(0.1 \le X < 0.5) = \frac{[F(0.5) - F(0.1)]/(0.5 - 0.1)}{F(1) - F(0)/(1-0)} = \left[c\left(\frac{1}{4}\right) - c\left(\frac{1}{10}\right)\right]/0.4$

$= \frac{6}{40} \cdot \frac{10}{4} = \frac{6}{16} = \boxed{\frac{3}{8}}$

41. Find the upper and lower quartiles of an exponential distribution.

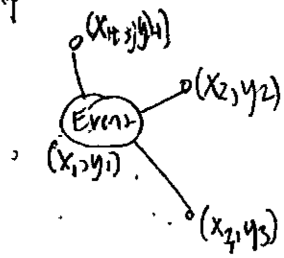
Exponential Distribution: $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$ Lower quartile: $P(X) = \frac{1}{4} = \lambda e^{-\lambda x}$; $-\log 4\lambda = -\lambda x$

$\boxed{x = \frac{+\log 4\lambda}{\lambda}}$

Upper quartile: $p(x) = \frac{3}{4} = \lambda e^{-\lambda x}$

$\boxed{x = \log\left(\frac{4\lambda}{3}\right)/\lambda}$

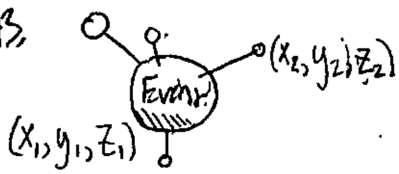
42.



$\boxed{p(x) = \sqrt{\left(\frac{\lambda^{x_2} e^{-\lambda}}{x_2!} - \frac{\lambda^{x_1} e^{-\lambda}}{x_1!}\right)^2 + (x_2 - x_1)^2}}$

$\boxed{P(x) = 2\pi\lambda x\, e^{2-\pi\lambda x^2}}$ ← volume lambda

43.



$F(x) = \frac{d}{dx} e^{-4\lambda\pi x^3/3}$ $\boxed{= 4\lambda\pi x^3 e^{-4\lambda\pi x^3/3}}$

Multivariate Poisson Distribution:

$P(X) = \exp\left(-\sum_{i=1}^{m} \theta_i\right) \prod_{i}^{m} \frac{\theta_i^{x_i}}{x_i!} \sum_{i=0}^{s} \prod_{j=1}^{m} \binom{x_j}{i} i! \left(\frac{\theta_0}{\prod \theta_i}\right)^i$

44. T: Exponential Random Variable with λ $T = \lambda e^{-\lambda x}$; $k = \lambda e^{-\lambda x}$; $k+1 = \lambda e^{-\lambda x}$

$\boxed{x = \log\left(\frac{\lambda}{T}\right)/\lambda}$

Exponential Random Distribution:

$p(x) = \lambda e^{-\lambda x}$

X: Discrete Random variable; $X = K$; $K < T < K+1$

for $K = 0, 1, \ldots$

45. Exponential Distribution: $p(x)=\lambda e^{-\lambda x}$; $\lambda=0.1$ a) Probability lifetime < 10 years.

$P(\text{lifetime})+P(\text{death})=P(\text{lifetime})+\frac{e^{-10/10}}{10}=1$

$\boxed{P(\text{lifetime})=1-\frac{e^{-1}}{10}}$

b. $\frac{e^{-5/10}}{10}-\frac{e^{-15/10}}{10}$ c) $0.01=\frac{e^{-t/10}}{10}$; $-1=-t/10$; $\boxed{t=10}$

46. Gamma Density: $g(t)=\frac{\lambda^\alpha}{\Gamma(\alpha)}t^{\alpha-1}e^{-\lambda t}$, $t\geq 0$

where $\Gamma(x)=\int_0^\infty u^{x-1}e^{-u}\,du$; $x>0$

$$\int_0^\infty g(t)\,dt=\int_0^\infty \frac{\lambda^\alpha t^{\alpha-1}e^{-\lambda t}}{\Gamma(\alpha)}dt=\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty t^{\alpha-1}e^{-\lambda t}dt \;;\; t=x/\lambda;\; x=t\lambda;\; =\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty\left(\frac{x}{\lambda}\right)^{\alpha-1}e^{-x}\frac{dx}{\lambda}$$

$$=\frac{1}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}e^{-x}dx=\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda}=\frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)\lambda}=\boxed{\frac{\alpha}{\lambda}}$$; $\lambda=1$ and $\alpha=1$.

47. $\alpha>1$, Show maximum of Gamma Density: $(\alpha-1)/\lambda$; $\frac{d}{dt}g(t)=0$

$=\frac{\lambda^\alpha}{\Gamma(\alpha)}\left[(\alpha-1)t^{\alpha-2}-\lambda t^{\alpha-1}\right]e^{-\lambda t}=0$; $(\alpha-1)t^{\alpha-2}=\lambda t^{\alpha-1}$; $\boxed{\frac{(\alpha-1)}{\lambda}=t^{\alpha-1-\alpha+2}=t}$

48. T is an exponential Random variable: $p(x)=\lambda e^{-\lambda x}$, and $P(T<1)=0.05$.
What is λ? $0.05=\lambda e^{-\lambda T}=\lambda\left(1+\lambda+\frac{\lambda^2}{2!}-\dots\right)=\lambda-\lambda^2+\frac{\lambda^3}{2!}$; $\boxed{\lambda=0.0527}$

"Third order Quadratic"

49. a) $\Gamma(1)=1$; Gamma Density: $g(t)=\frac{\lambda^\alpha}{\Gamma(\alpha)}t^{\alpha-1}e^{-\lambda t}$

Gamma Function: $\Gamma(x)=\int_0^\infty u^{x-1}e^{-u}du$; $\Gamma(1)=\int_0^\infty u^{1-1}e^{-u}du=\int_0^\infty e^{-u}du=-e^{-u}\Big|_0^\infty=\boxed{0-(-1)=1}$

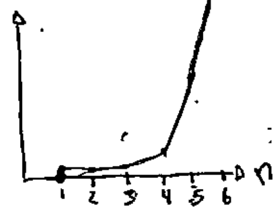
b) $\Gamma(x+1)=x\Gamma(x)$; $\Gamma(x+1)=\int_0^\infty u^x e^{-u}du$; Integration by Parts: $\int u\,dv=uv-\int v\,du$; $u=u^x$, $dv=e^{-u}$, $du=(x)u^{x-1}du$, $v=-e^{-u}$ $=-u^xe^{-u}\Big|_0^\infty+\int(x)u^{x-1}e^{-u}=\boxed{x\Gamma(x)}$

c) Conclude $\Gamma(n)=(n-1)!$; $n=1,2,3\dots$

Table:

n	$\Gamma(n)$	$(n-1)!$
1	0	0
2	1	1
3	2	2
4	6	6
5	24	24
6	120	120

Graph:



51. Normal Distribution: $p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$ Standard Normal: $\mu=0,\sigma=1$ $p(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$

$2\int_0^\infty e^{-x^2/2}dx=2\sqrt{\frac{\pi}{2}}=\sqrt{2\pi}$; $\int_{-\infty}^\infty e^{-x^2/2}dx\cdot\int_{-\infty}^\infty e^{-y^2/2}dy=2\pi$

50. $\Gamma(x)=2\int_0^\infty t^{2x-1}e^{-t^2}dt=\int_{-\infty}^\infty e^{xt}\cdot e^{-e^t}dt$; e

$=\int_{-\infty}^\infty t^{2x-1}e^{-t^2}dt=\int_{-\infty}^\infty (e^{t/2})^{2x-1}e^{-e^t}dt$; $e^{xt}=t^{2x-1}$

$e^t=t^2$; $\int_{-\infty}^\infty e^{xt-t/2}e^{-e^t}dt=\int_{-\infty}^\infty e^{xt}e^{-e^t}dt$; $\int_{-\infty}^\infty\int_{-\infty}^\infty e^{-(x^2+y^2)/2}dx\,dy=\int_0^\infty e^{-r^2/2}\dots dr$

$dr=\frac{1}{2}(x^2+y^2)\dots$

52. $\mu = 70$ and $\sigma = 3$ in. a) What proportion of the population is over 6 ft tall?

$$P(78) = \frac{1}{3\sqrt{2\pi}} e^{-(78-70)^2/2\cdot 3^2} = 0.10654 ; \boxed{0.35\% \text{ over the height of 6ft tall.}}$$

$$P(>78) = \frac{1}{3\sqrt{2\pi}} \int_{78}^{\infty} e^{-(x-70)^2/18} dx = 0.00038 = \boxed{0.38\%}$$

b) Cm: $70 \text{in} \times \frac{2.54\text{cm}}{1\text{inch}} = 179\text{cm}$; $3\text{in} \times \frac{2.54\text{cm}}{1\text{inch}} = 7.5\text{cm}$; m: $70 \text{inches} \times \frac{2.54\text{in}}{1\text{inch}} \times \frac{1\text{m}}{100\text{cm}} = 1.75\text{cm}$; $7.5\text{in} \times \frac{1\text{m}}{100\text{cm}} = \boxed{0.075}$

53. $\mu = 5$, and $\sigma = 10$. a) Find $P(X>10) = \frac{1}{10\sqrt{2\pi}} \int_{10}^{\infty} e^{-(x-5)^2/2\cdot 10^2} dx = 0.3085 = \boxed{30.85\%}$

Z-table: $\frac{10-5}{10} = \frac{1}{2} = Z(30.85\%)$

b) Find $P(-20<X<15) = \frac{1}{10\sqrt{2\pi}} \int_{-20}^{15} e^{-(x-5)^2/2\cdot 10^2} dx = 0.8351 = \boxed{83.51\%}$ $\left(\frac{15-5}{10} - \frac{-20-5}{10}\right)$

$Z(0.9938) - Z(0.8413)$

$\boxed{84.75\%}$

54. $X \sim N(0,\sigma^2)$; $Y = |X|$

$$= \frac{1}{\sigma^2\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x)^2/2\cdot\sigma^2} dx = \frac{\sqrt{\pi \cdot 2 \cdot \sigma^2}}{\sigma \sqrt{2\pi}} = \boxed{1}$$

55. $X \sim N(\mu,\sigma^2)$ Find c in terms of σ, such that, $P(\mu - c \le X \le \mu + c) = 0.95$

$$0.95 = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-c}^{\mu+c} e^{-(x-\mu)^2/2\sigma^2} dx ; \int e^{-ax^2} = \sqrt{\frac{\pi}{a}} ; \boxed{c \approx 1.95996\sigma}$$

56. $X \sim N(\mu,\sigma^2)$; $P(|X-\mu| \le 0.675\sigma) = 0.5$; $= 0.5 = \frac{1}{0.675\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-0.675\sigma^2/2\sigma^2} dx = \boxed{0.5}$

57. $X \sim N(\mu,\sigma^2)$; $Y = aX + b$, where $a<0$, show $Y \sim N(a\mu + b, a^2\sigma^2)$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = P\left(\frac{y-b}{a}\right) = \frac{1}{a\sigma\sqrt{2\pi}} e\left[-\frac{1}{2}\left(\frac{y-b-a\mu}{a\sigma}\right)^2\right] = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2}\frac{(y-(a\mu+b))^2}{a^2\sigma^2}\right]$$

$\boxed{Y \sim (a\mu + b, a^2\sigma^2)}$

58. $Y = e^Z$, where $Z \sim N(\mu,\sigma^2)$; $\boxed{\text{Lognormal Density}}$

$Y = e^z = e^{N(\mu,\sigma^2)}$; $\log Y = N(\mu,\sigma^2) = 0.5$ $\boxed{Y = e}$

59. $U[-1,1]$; Find density function of U^2; $F_u = P(-\sqrt{x} \le z \le \sqrt{x})$

$= \Phi(\sqrt{x}) - \Phi(-\sqrt{x})$

$= \frac{1}{2} x^{-1/2} \phi(\sqrt{x}) + \frac{1}{2} x^{-1/2} \phi(-\sqrt{x})$

$= x^{-1/2} \phi(\sqrt{x})$; $\boxed{f_x(x) = \frac{x^{-1/2} e^{-x/2}}{\sqrt{2\pi}}}$

60. $U[0,1]$; Find density function of $\sqrt{U}$.

$F_u = P(x^2 \le z \le x^2) = \Phi(x^2) - \Phi(0)$

$= 2x\,\phi(x^2)$; $\boxed{f(x) = \frac{2x e^{-x/2}}{\sqrt{2\pi}}}$

61. Density of cX When X follows gamma distribution ; Gamma Distribution: $g(t) = \frac{\lambda^{\kappa}}{\Gamma(\kappa)} t^{\kappa-1} e^{-\lambda t}$

$g(ct \le c\lambda) = \boxed{g(t \le \frac{\lambda}{c}) = \frac{(\frac{\lambda}{c})^{\kappa}}{\Gamma(\kappa)} t^{\kappa-1} \cdot e^{-\lambda t/c}}$

$\Gamma(x) = \int_0^{\infty} u^{x-1} \cdot e^{-u} du$

62. m = mass ; V = random velocity ; $\mu = 0$ and σ. Find the density function of Kinetic Energy ; $E = \frac{1}{2} m V^2$:

Normally Distributed: $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$

$= \frac{1}{\sigma\sqrt{2\pi}} e^{-E^2/2\sigma^2} = \boxed{\frac{1}{\sigma\sqrt{2\pi}} e^{-(\frac{1}{2}mV^2)^2/2\sigma^2}}$

63. Suppose Θ is a uniform distribution ; Interval or domain $[-\pi/2, \pi/2]$: Find the CDF and density of tan

$\tan\Theta = x$; $\Theta = \arctan(x)$; $P(\arctan(-X) \le \Theta \le \arctan(X)) = \phi(\arctan(X)) + \phi(\arctan(+X))$

$= \frac{a^2}{x^2+a^2}\phi(\arctan(x)) + \frac{a^2}{x^2+a^2}\phi(\arctan(+x))$

64. f_x = "Density Function" and $Y = aX + b$, then.

$f_y(y) = \frac{1}{|a|} f_x\left(\frac{y-b}{a}\right)$

$\boxed{F_u = \frac{1}{\pi(1+x^2)} \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}}}$

$f_y(y) = P(Y \le y) = P(aX + b \le y) = P(X \le \frac{y-b}{a}) = f_x(\frac{y-b}{a}) = P'(X) = \boxed{f_x(\frac{y-b}{a})\frac{1}{|a|}}$

65. $f(x) = \frac{1+\alpha x}{2}$ from $-1 \le x \le 1$ and $-1 \le \alpha \le 1$.

$F_x(x) = \int_{-1}^{x} \frac{1+\alpha t}{2} dt = \int_{-1}^{x} \frac{1}{2} + \frac{\alpha t}{2} t \, dt$

$F_x(x) = \frac{1}{2}[x+1] + \frac{\alpha}{4}[x^2-1]$

$4F_x(x) = 2[x+1] + \alpha[x^2-1]$

$4X = 2[F^{-1}(x)+1] + \alpha[F^{-1}(x)^2 - 1]$

$4X = 2F^{-1}(x) + 2 + \alpha F^{-1}(x)^2 - \alpha$

$\alpha F^{-1}(x)^2 + 2F^{-1}(x) + (2 - \alpha - 4X) = 0$

$\frac{-2 \pm \sqrt{4 - 4(\alpha)(2-\alpha-4X)}}{2\alpha} = F^{-1}(X)$

$\frac{-2 \pm \sqrt{4 - 8\alpha + 4\alpha^2 + 8X}}{2\alpha} = F^{-1}(X)$

$\boxed{\frac{-1 \pm \sqrt{1 - 2\alpha + \alpha^2 + 2X}}{\alpha} = F^{-1}(X)}$

66 $f(x) = \alpha x^{-\alpha-1}$ for $x \ge 1$ and $f(x) = 0$

$\int_1^{\infty} \alpha x^{-\alpha-1} dx = \frac{\alpha x^{-\alpha}}{(-\alpha)}\Big|_1^{\infty} = -\frac{1}{x^{\alpha}}\Big|_1^{\infty} = -\frac{1}{\infty} + \frac{1}{1^{\alpha}} = 1 = F(x)$

$F^{-1}(1) = X$ $\boxed{\text{"Cumulative Distribution Functions are pseudo random number generators"}}$

67. Weibull Cumulative Distribution Function:

$F(x) = 1 - e^{-(x/\alpha)^{\beta}}$, $x \ge 0$, $\alpha > 0$, $\beta > 0$

a) Find the density function. $p(x) = F'(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-(x/\alpha)^{\beta}}$

b) If Weibull is W, then $X = (W/\alpha)^{\beta}$ is an exponential distribution.

$W = \alpha^{\beta} x^{1/\beta}$; $\frac{dW}{dX} = \frac{1}{\beta} \alpha x^{1/\beta - 1}$

$F_x(W) = f_x(\alpha \cdot x^{1/\beta}) \cdot \left|\frac{\partial W}{\partial x}\right| = \frac{\beta}{\alpha} e^{-W} \cdot \left(\frac{x^{1/\beta}}{\alpha}\right)^{\beta-1} \cdot \frac{\alpha}{\beta} x^{1/\beta - 1} = \boxed{e^{-W}}$

c). $U = e^{-W}$; $\ln U = -W$; $\boxed{W = -\ln V}$;

69. U = Uniform Random variable. Find $V = U^{-\alpha}$ for : $\alpha > 0$ $U^{\alpha} = V$; $U = V^{1/\alpha}$

$P(V^{1/\alpha}) = P(V^{1/\alpha}) \cdot \left|\frac{d}{dy} V^{1/\alpha}\right| = P(V^{1/\alpha}) \cdot \left(\frac{1}{\alpha}\right) \cdot V^{(1-\alpha)/\alpha}$

$\boxed{\text{Proposition B}}$

The rates of decrease for the density function is decreased as α takes greater rates.

69. $p(x) = \lambda e^{-\lambda x}$; $V = \frac{4}{3}\pi R^3$; $R = \sqrt[3]{\frac{3V}{4\pi}} = \lambda\, p(R) = \lambda e^{-\lambda \sqrt[3]{\frac{3V}{4\pi}}} \cdot \frac{d}{dV}\left|\lambda e^{-\lambda\sqrt[3]{\frac{3V}{4\pi}}}\right|$

$= \lambda e^{-\lambda\sqrt[3]{\frac{3V}{4\pi}}} \cdot \underbrace{\lambda\left(-\lambda \frac{1}{3}\left(\frac{3V}{4\pi}\right)^{-2/3} \cdot \frac{3}{4\pi}\right) e^{-\lambda\sqrt[3]{\frac{3V}{4\pi}}}}$

$\boxed{\text{"Density Function"}}$

70. $p(X) = \lambda e^{-\lambda r}$; $A = \pi r^2$; $r = \sqrt{\frac{A}{\pi}}$;

$f(y) = f_X(g^{-1}(y))\left|\frac{d}{dy} g^{-1}(y)\right|$; $f(y) = \lambda e^{-\lambda r} = \lambda e^{-\lambda\sqrt{\frac{A}{\pi}}} \cdot \left|\frac{d}{dA}\lambda e^{-\lambda\sqrt{\frac{A}{\pi}}}\right|$

$= \lambda e^{-\lambda\sqrt{\frac{A}{\pi}}} \underbrace{\left| -\frac{\lambda}{2}\left(\frac{A}{\pi}\right)^{-1/2} e^{-\lambda\sqrt{\frac{A}{\pi}}}\right|}$

$\boxed{\text{"Density Function"}}$

71. F = cdf of random variable.
U is uniform from $[0,1]$.
Define, $Y = K$ if $F(k-1) < U \le F(k)$;

$\boxed{\int_{k-1=0}^{k=1} F(k)\,dk = \int_0^1 \dots F}$

$m=3$;

X	0	2	2	0	2	1

$c=2$

72. $X_n = (a X_{n-1} + c) \bmod m$ a) $a=1$ b)

X	0.064069	2	13813	944	—

Chapter 3: Joint Distributions:

1. Joint Frequency Function:

y \ X	1	2	3	4	$p_Y(y)$
1	0.10	0.05	0.02	0.02	0.19
2	0.05	0.20	0.05	0.02	0.32
3	0.02	0.05	0.20	0.04	0.31
4	0.02	0.02	0.04	0.10	0.18
$p(x)$	0.19	0.32	0.31	0.18	

= Joint Frequencies

Marginal Frequencies

A) Find the marginal frequency function of X and Y. i.e. $p(X)$, and $p(y)$.

b) Find the conditional frequency of X given $Y=1$ and Y given $X=1$.

~~$P(X|Y=1) = \frac{P(X, Y=1)}{P(y)} = \frac{0.10}{0.19} = \boxed{\frac{10}{19}}$; $P(Y|X=1) = \frac{P(Y, X=1)}{P(X)} = \frac{0.10}{0.19} = \boxed{\frac{10}{19}}$~~

$\boxed{p(1|1) = \frac{10}{19};\ p(2|1) = \frac{5}{19};\ p(3|1) = \frac{2}{19};\ p(4|1) = \frac{2}{19}}$

2. p - black balls, q - white balls, r - red balls (Urn); n chosen

a) Find the joint distribution of black, white and red balls.
Multinomial Distribution: $p(n_1, \dots, n_r) = \binom{n}{n_1, \dots, n_r} p_1^{n_1} p_2^{n_2} \cdots p_3^{n_r}$

$p(\text{red, white, black}) = \frac{\binom{p}{x}\binom{q}{y}\binom{r}{z}}{\binom{p+q+r}{n}}$

b) Joint Distribution of black and white balls.

$P(\text{black}(X), \text{white}(Y)) = \frac{\binom{p}{x}\binom{q}{y}\binom{r}{n-x-y}}{\binom{p+q}{n-r}}$

where $n = X+Y+Z$

Total selection outcomes

c) $P(Y) = \frac{\binom{q}{y}\binom{p+r}{n-y}}{\binom{p+q+r}{n}}$

outcomes problem

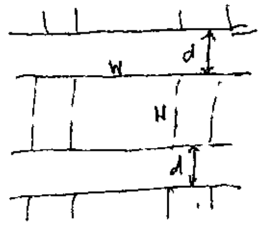
Outcome
Black + White outcomes

Total selection

3. Three players play 10 independent rounds of a game. $P(\text{Player}\mid\text{winning}) = \frac{1}{3}$

$P(\text{Player } A, B, C \mid \text{Win}) = \boxed{\binom{10}{x\,y\,z}\left(\frac{1}{3}\right)^x\left(\frac{1}{3}\right)^y\left(\frac{1}{3}\right)^z}$. Multinomial Distribution: $P(X, Y, Z) = \boxed{\binom{n}{x\,y\,z} p_1^x p_2^y p_3^z}$

4 Wire diameter $= d$, Holes side length $= W$, Spherical particle radius $= r$. What is probability of passing?



Area Circle $= \pi r^2$ Probability passing per hole $= \frac{\text{Area particle}}{\text{Area square}} = \boxed{\frac{\pi r^2}{W^2}}$

Area Square $= W^2$, Probability passing per mesh $=$ Probability passing $\cdot \frac{\text{Area square}}{\text{Area mesh}}$

$= \frac{\pi r^2}{W^2} \frac{W^2}{(nW + (n+1)d)^2} = \boxed{\frac{\pi r^2}{(nW+(n+1)d)^2}}$

Fails to pass through if dropped n-times:

$P(\text{Failing to pass through}) = \boxed{\left(1 - \frac{\pi r^2}{(nW+(n+1)d)^2}\right)^n}$ $\binom{n}{1}$

$n \frac{1}{2} + W$

5. $\vdash L \dashv$ $a \geq L$

Probability needle crosses line $= 2\left(\frac{\text{Length of Needle}}{\text{Distance of lines}}\right) \times \frac{\text{1 full rot}}{\pi \text{ rad}} = \boxed{\frac{2L}{\pi D}}$

$\leftarrow a \rightarrow \leftarrow a \rightarrow$

6. $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$; Marginal Density of X and Y coordinates inside an ellipse: Area $= \pi(a\cdot b)$

$f_y(y) = \int_{-\infty}^{\infty} f(x,y)\,dy = \frac{1}{\pi a b}\int_{-\sqrt{a^2(1-y^2/b^2)}}^{\sqrt{a^2(1-y^2/b^2)}} dx = \frac{1}{\pi a b} 2\sqrt{a^2(1-y^2/b^2)} = \boxed{\frac{2\sqrt{1-y^2/b^2}}{\pi b}}$

$\boxed{f_x(x) = \frac{2\sqrt{1-x^2/a^2}}{\pi a}}$

7. CDF: $F(X,y) = (1-e^{-\alpha x})(1-e^{-\beta y})$; $x \geq 0$; $y \geq 0$; $\alpha > 0$; $\beta > 0$

Joint Density: $f(x,y) = \frac{d}{dx}\frac{d}{dy}(1-e^{-\alpha x})(1-e^{-\beta y})$

$\boxed{= (1+\alpha e^{-\alpha x})(1+\beta e^{-\beta y})}$

Marginal Density: $F_x(x) = \int_{-\infty}^{\infty} (1-e^{-\alpha x})(1-e^{-\beta y})\,dy = 2(1-e^{-\alpha x})\left(1+\frac{e^{-\beta y}}{\beta}\right)\Big|_{0}^{\infty}$

$F_y(y) = \int_{-\infty}^{\infty} (1-e^{-\alpha x})(1-e^{-\beta y}) \quad (= 2(1-e^{-\beta y})(1/\beta)$

$\boxed{= 2(1-e^{-\beta y})(1/\alpha)}$

8. $f(x,y) = \frac{6}{7}(x+y)^2$, $0 \leq x \leq 1$, $0 \leq y \leq 1$

a) i) Find $P(X > Y) = \frac{6}{7}\int_0^1\int_0^x (x^2+xy)\,dy\,dx = \frac{6}{7}\int_0^1 \left(x^3 + \frac{x^3}{2}\right)dx = \frac{6}{7}\left(\frac{x^4}{4} + \frac{x^4}{8}\right)\Big|_0^1 = \frac{6}{7}\left(\frac{1}{4} + \frac{1}{8}\right) = \frac{6}{28} + \frac{6}{48} = \boxed{\frac{19}{56}}$

ii) Find $P(X+Y \leq 1) = \frac{6}{7}\int_0^1\int_0^{1-Y}(x^2+xy)\,dx\,dy = \frac{6}{7}\int_0^1\left(\frac{x^3}{3} + \frac{x^2}{2}y\right)\Big|_0^{1-Y} = \frac{6}{7}\int_0^1\left(\frac{(1-Y)^3}{3} + \frac{(1-Y)^2}{2}y\right)dy$

$X \leq 1-Y$

$= \frac{6}{7}\left[\frac{-(u)^4}{12} + \frac{y^2}{4} - \frac{2y^3}{2\cdot 3}\right]\Big|_0^1 = \frac{6}{7}\left[-\frac{(1-1)^4}{12} + \frac{1}{4} - \frac{1}{3} + \frac{1}{12}\right] = \boxed{0}$

iii) $P(X \leq \frac{1}{2}) = \frac{6}{7}\int_0^1\int_0^{1/2}(x^2+xy)\,dx\,dy = \frac{6}{7}\int_0^1\left(\frac{x^3}{3} + \frac{x^2}{2}y\right)\Big|_0^{1/2} dy = \frac{6}{7}\int_0^1 \frac{1}{18} + \frac{1}{8}y\,dy = \frac{6}{7}\left[\frac{1}{18} + \frac{1}{16}\right] = \boxed{17/168}$

$u = 1-y$, $du = dy$

b) Marginal Densities of X & Y:

$f_X(x) = F_x'(x) = \int_0^1 (x^2+yx)\,dy = \boxed{x^2 + \frac{x}{2}}$

$(1-y)^2 y$

$f_Y(y) = F_y'(x) = \int_0^1 (x^2+yx)\,dx = \boxed{\frac{1}{3} + \frac{y}{2}}$

$y - 2y + y^3$ $(1-2y+y^2)y$

c. Conditional Density: $f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \boxed{\frac{\frac{6}{7}(x^2+xy)}{x^2+\frac{x}{2}}}$; $f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(x)} = \boxed{\frac{\frac{6}{7}(x^2+xy)}{\frac{1}{3}+\frac{y}{2}}}$

9. (X,Y) uniformly distributed over: $0 \le y \le 1-x^2$; $-1 \le x \le 1$ Assuming Bivariate Normal Density:

a) Find the marginal distribution: $F_X(x) = \int_0^{1-x^2} \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(\frac{[x-\mu_x]^2}{\sigma_x^2} + \frac{[y-\mu_y]^2}{\sigma_y^2}\right)$; $f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left(\frac{[x-\mu_x]^2}{\sigma_x^2} + \frac{[y-\mu_y]^2}{\sigma_y^2}\right)$

b) Find the two conditional densities:

$f_{Y|X}(y|x) = \frac{f_{xy}(x,y)}{f_x(x)}$

$= \frac{1}{2\pi\sigma_x\sigma_y}\exp\left(\frac{[x-\mu_x]^2}{\sigma_x^2} \frac{[y-\mu_y]^2}{\sigma_y^2}\right)\left[\frac{1}{\sqrt{2}\sigma_y}\Gamma\left(\frac{1}{2}\right)\right]\left[\frac{2\sigma_y^2}{\mu_y^2}\right]\Big|^{1-x^2}$

$= \boxed{\frac{\mu_y}{\sqrt{2}\sigma_y\Gamma(\frac{1}{2})}\exp\left[\frac{-1}{2}\right]\frac{[y-\mu_y]^2-[1-x^2-\mu_y]^2+\mu_y^2}{\sigma_y^2}}$

$f_x(x) = \boxed{\frac{\Gamma(\frac{1}{2})}{\sqrt{2\pi}\sigma_x\mu_y}\exp\left(\left[\frac{[x-\mu_x]^2}{\sigma_x^2} + \frac{[1-x^2-\mu_y]^2}{\sigma_y^2} - \frac{[\mu_y]^2}{\sigma_y^2}\right]\left(-\frac{1}{2}\right)\right)}$; $= \frac{-1}{2\sigma_y^2}$

$f_{X|Y}(x|y) = \frac{f_{xy}(x,y)}{f_y(y)}$

$f_y(y) = \int_{-1}^{1} \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[\frac{-1}{2}\right]\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]dy$

$= \boxed{\frac{\mu_x}{\sqrt{2}\sigma_x\Gamma(\frac{1}{2})}\exp\left[\frac{-1}{2}\right]\frac{(x-\mu_x)^2-(1-\mu_x)^2-\mu_x^2}{\sigma_x^2}}$

$= \boxed{\frac{2\Gamma(\frac{1}{2})}{\sqrt{2\pi}\sigma_x\mu_x}\exp\left[\frac{[1-\mu_x]^2-\mu_x^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]}$

10. Suppose $f(x,y) = xe^{-x(y+1)}$

$0 \le x \le \infty$; $0 \le y < \infty$

a) Find the marginal density of X and Y. $f(x) = \int_0^\infty xe^{-x(y+1)}dy = \frac{\cancel{x}e^{-x(y+1)}}{-x}\Big|_0^\infty = -e^{-x(\infty+1)} + e^{-x(1)} = \boxed{e^{-x}}$

b) Find the conditional densities.

$f_{Y|X}(y|x) = \frac{f_{xy}(x,y)}{f_x(x)} = \frac{xe^{-x(y+1)}}{e^x} = \boxed{xe^{-x[y+2]}}$

$f(y) = \int_0^\infty xe^{-x(y+1)}dx$; $\int u\,dv = uv - \int v\,du$; $u = x$, $dv = e^{-x(y+1)}$, $du = dx$, $v = \frac{-e^{-x(y+1)}}{y+1}$

$= \frac{-xe^{-x(y+1)}}{y+1}\Big|_0^\infty + \int_0^\infty \frac{e^{-x(y+1)}}{y+1}dx$

$f_{X|Y}(x|y) = \frac{f_{xy}(x,y)}{f_y(y)} = \frac{xe^{-x(y+1)}}{\frac{1}{(y+1)^2}} = \boxed{x(y+1)^2e^{-x(y+1)}}$

Independent

$= 0 - \frac{e^{-x(y+1)}}{(y+1)^2}\Big|_0^\infty = \boxed{\frac{1}{(y+1)^2}}$

11. U_1, U_2, and U_3 independent from $[0,1]$

Find the probability the roots of the quadratic $U_1x^2+U_2x+U_3$ are real.

$0 = U_1x^2+U_2x+U_3$; $x_1, x_2 = \frac{-U_2 \pm \sqrt{U_2^2-4(U_1)(U_3)}}{2(U_1)}$; $P(U_1) = \int_0^1\int_0^1 u_1x^2+u_2x+U_3\,dU_2\,dU_3 = \int_0^1 u_1x^2+U_2x+\frac{U_2^2}{2}\,dU_2$

$= U_1x^2+\frac{x}{2}+\frac{1}{2}$

~~Extrema and Order Statistics~~

$\cancel{f(U_{(1)}, U_{(2)}, U_{(3)} \ldots) = n!\prod_{i=1}^{n} f(U_i)}$; for $i=3$; $\cancel{f(U_1, U_2, U_3) = 3!\prod_{i=1}^{3} f(U_i) : 6f(U_1)f(U_2)f(U_3)}$

$P(U_2^2 > 4U_1U_3) = P(|U_2| \ge 2\sqrt{u_1U_3}) = \int_0^1\int_0^1\int_{2\sqrt{U_1U_3}}^1 f(u_1, U_2, U_3)\,du_2\,du_3\,du_1 = \int_0^1\int_0^1 (1-2\sqrt{u_1u_3})$

$= 1 - \frac{4}{3}\,\frac{1^{3/2}}{3/2} = \boxed{1/9}$

12. $f(x,y)=c(x^2-y^2)e^{-x},\ 0\le x<\infty;\ -x\le y<x$ a) Find c: $P(x,y)=1=\int_0^\infty\int_{-x}^{x}c(x^2-y^2)e^{-x}dy\,x$

$=\int_0^\infty c(x^2-\frac{y^3}{3})e^{-x}\Big|_{-x}^{x}dx$

$=\int_0^\infty c(x^2-\frac{x^3}{3})e^{-x}-c(x^2+\frac{x^3}{3})e^{-x}dx$

$=\int_0^\infty c(-2\frac{x^3}{3})e^{-x}dx$

$1=-4c;\ \boxed{c=\frac{-1}{4}}$

b) $f(x)=\int_{-x}^{x}c(x^2-y^2)e^{-x}dy=\boxed{\frac{-4x^3}{3}e^{-x}}$ / $f(y)=\int_0^\infty c(x^2-y^2)e^{-x}dx=\boxed{\frac{1}{4}(2-y^2)}$

c) $f_{Y|X}(Y|X)=\frac{f(x,y)}{f(x)}=\frac{-2(x^2-y^2)e^{-x}}{\frac{4}{3}x^3e^{-x}}=\boxed{\frac{-3}{2}\frac{(x^2-y^2)}{x^3}}$; $f_{X|Y}(x|y)=\frac{f(x,y)}{f(y)}=\boxed{\frac{(x^2-y^2)e^{-x}}{(2-y^2)}}$

13. Sample Space: Throws

0	1	2
1/2	1/2	1/4

1	2
H	H
H	T
T	~~H~~
T	~~T~~

$P(0)=1/2$ $P(2)=\frac{1}{4}$
$P(1)=1/2$

a) $P(x)=\int_{-\sqrt{y^2+z^2}}^{\sqrt{y^2+z^2}}$

$f(x)=\int_{-\sqrt{r^2-x^2-z^2}}^{\sqrt{r^2-x^2-z^2}}\int_{-\sqrt{r^2-x^2-y^2}}^{\sqrt{r^2-x^2-y^2}}\sqrt{\left(\frac{V}{4/3\pi}\right)^2-y^2-z^2}\,dz\,dy$

$=\int_{-\sqrt{r^2-x^2-z^2}}^{\sqrt{r^2-x^2-z^2}}\frac{1}{2}\left(z\sqrt{\left(\frac{V}{4/3\pi}\right)^2-y^2-z^2}+(a^2-y^2)\tan^{-1}\left(\frac{z}{\sqrt{a^2-y^2-z^2}}\right)\right)\Bigg|_{-\sqrt{r^2-x^2-y^2}}^{\sqrt{r^2-x^2-y^2}}$ $u=$

14. Point in a unit sphere $(\mathbb{R}_3)$

$V=\frac{4}{3}\pi r^3=\frac{4}{3}\pi\sqrt{x^2+y^2+z^2}$

Density Function of a unit sphere:

$f(x,y,z)=\begin{cases}k & 0\le x^2+y^2+z^2\le 1\\ 0 & \text{otherwise}\end{cases}$

To find the value of k, such that:

$\iiint_{0\le x^2+y^2+z^2\le 1} k\,dx\,dy\,dz=1$; let $x=\rho\sin\phi\cos\theta,\ y=\rho\sin\phi\sin\theta,\ z=\rho\cos\phi$; $\rho^2=x^2+y^2+z^2=1$

$0<\rho<1;\ 0<\phi<\pi;\ 0<\theta<2\pi$

Volume $=\int_0^{2\pi}\int_0^{\pi}\int_0^1\rho^2\sin\phi\,d\rho\,d\phi\,d\theta=\frac{4\pi}{3}$; $k\frac{4\pi}{3}=1$; $k=\frac{3}{4}\pi$

The density function becomes: $\boxed{f(x,y,z)=\begin{cases}\frac{3}{4\pi} & 0<x^2+y^2+z^2\le 1\\ 0 & \text{otherwise.}\end{cases}}$

Marginal Densities: 1) Joint Densities:

$f_{x,y}(x,y)=\int_{-\infty}^{\infty}f(x,y,z)\,dz=\int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}}\frac{3}{4\pi}dz=\frac{3}{2\pi}\sqrt{1-x^2-y^2}$; $f_{xy}(x,y)\begin{cases}\frac{3}{2\pi}\sqrt{1-x^2-y^2}; & 0\le x^2+y^2\le 1\\ 0, & \text{otherwise}\end{cases}$

2) $f_X(x)=\int_{-\infty}^{\infty}f_{xy}(x,y)dy=\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}\frac{3}{2\pi}\sqrt{1-x^2-y^2}dy$; $y=\sqrt{1-x^2}\sin u$, $dy=\sqrt{1-x^2}\cos u\,du$; $=\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}\frac{3}{2\pi}\sqrt{(1-x^2)-(\sqrt{1-x^2})^2\sin^2u}\,(\sqrt{1-x^2}\cos u\,du)$

$\boxed{=\frac{3}{4}(1-x^2)}$

$\boxed{f_y(y)=\begin{cases}\frac{3}{4}(1-y^2) & -1\le y\le 1\\ 0, & \text{otherwise}\end{cases};\ f_z(z)=\begin{cases}\frac{3}{4}(1-z^2) & -1\le z\le 1\\ 0, & \text{otherwise}\end{cases}}$

@ $z=0$ $f_{XY|Z}(x,y|z)=\frac{f_{xyz}(x,y,z)}{f_z(z=0)}=\frac{3/4\pi}{\frac{3}{4}(1-(0)^2)}=\boxed{\frac{1}{\pi}}$

15. Suppose the joint density $f(x,y) = c\sqrt{1-x^2-y^2}$; $x^2+y^2 \le 1$

a) Find c: $x = r\cos\theta$; $y = r\sin\theta$; $\iint f(x,y)\,dA = \int_0^{2\pi}\int_0^1 c\sqrt{1-r^2}\,r\,dr\,d\theta = c\left(\frac{2}{3}\right)\pi = 1$; $\boxed{c = \left(\frac{3}{2\pi}\right)}$

b) $P(x^2+y^2) \le 1/2$ is a half of the disk;

~~$\iint f(x,y)\,dA;\ x^2+y^2 \le 1/2 = \int_0^{\pi/4}\int_0^1 \left(\frac{3}{2\pi}\right)\sqrt{1-r^2}\,r\,dr\,d\theta = \iint_0^{\pi/4,1}\left(\frac{3}{2\pi}\right)\sqrt{1-r^2}\,r\,dr\,d\theta = \int_0^{\pi/4}\left(\frac{1}{2}\pi\right)d\theta = \boxed{1/8}$~~

b) The Joint density is an area of decreasing size.

d) $f(x) = \int_{-1}^{1}\sqrt{1-x^2-y^2}\left(\frac{3}{2\pi}\right)dy \doteq \boxed{\left(\frac{3}{4}\right)x^2}$; $f(y) \doteq \int_{-1}^{1}\sqrt{1-x^2-y^2}\left(\frac{3}{2\pi}\right)dx = \boxed{\left(\frac{3}{4}\right)x^2}$

To check independence, $f(x,y) = f_x(x)f_y(y) = \left(\frac{9}{16}\right)x^2y^2 \ne$ independence.

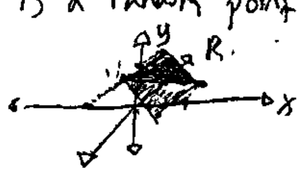
e) Conditional Densities: $\boxed{f(y|x) = f(x,y)/f_x(x);\ f(x|y) \doteq f(x,y)/f(y)}$

16. X_1 is uniform on $[0,1]$, and, conditional on X_1, X_2, from $[0, X_1]$

Find the joint Distributions: $\boxed{f(x_1,x_2) = \int_0^1\int_0^{x_1} dx_2\,dx_1}$ = Marginal Distributions: $\boxed{f(x_1) = \int_0^{x_1} dx_2}$ $\boxed{f(x_2) = \int_0^1 dx_1}$

17. (X,Y) is a random point of the region $R = \{(x,y) : |x|+|y| \le 1\}$

a)



b) $f_x(x) = R$, $f_y(y) = R$

c) $f_{Y|X}(Y|X) = \frac{f(y,x)}{f(x)} = \frac{|x|+|y|}{R} = \boxed{1}$

18. $f(x,y) = k(x-y)$, $0 \le y \le x \le 1$ and zero elsewhere.

a)

b) $1 = \int_0^1\int_0^x k(x-y)\,dy\,dx = \int_0^1 k\left(x - \frac{y^2}{2}\right)\Big|_0^x = \int_0^1 k\left[\cancel{x} - \frac{x^2}{2}\right) - \cancel{k}(x)\Big]\,dx$

$= \int_0^1 k\frac{x^2}{2}\,dx = -\frac{k(1)^3}{6} = 1$

$\boxed{k = 6}$

c) $f_x(x) = \int_0^x k(x-y)\,dy = \boxed{k\left(1 - \frac{x^2}{2}\right)}$; $\boxed{= 3x^2}$; $f_y(y) = \int_0^1 k(x-y)\,dx = k\left(\frac{x^2}{2} - y\right) + ky$

$= 6\left(\frac{1}{2}\right) = \boxed{3}$

d) $f_{Y|X}(Y|X) = \frac{f(x,y)}{f(x)} = \frac{k(x-y)}{-3x^2}$; $f_{X|Y}(X|Y) = \frac{f(x,y)}{f(y)} = \boxed{\frac{k(x-y)}{3}}$

19. a) Exponentially Distributed lifetimes means: $\lambda e^{-\lambda x} = f(\lambda)$; $f(T_1) = \alpha e^{-\alpha T_1}$; $f(T_2) = \beta e^{-\beta T_2}$

Find $P(T_1 > T_2) = \int_{-0}^{T_1} \beta e^{-\beta T_2}\,dT_2 = -e^{-\beta T_2}\Big|_{-\infty}^{T_1} = -\left[e^{-\beta T_1} - 1\right] = \boxed{1 + e^{-\beta T_1}}$ Let $S = T_1 + T_2$

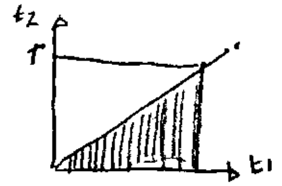
b) $\boxed{P(T_2 > 2T_1) = \frac{1 \pm e^{-\alpha(2T_1)}}{2}}$ $P(T_1 > T_2) = \int_0^{T_1} \alpha e^{-\alpha T_1} \cdot \beta e^{-\beta T_2}\,dT_2 = -\int_{-0}^{T_1} \alpha e^{-\alpha T_1} \cdot e^{-\beta T_1} = -\alpha e^{-(\alpha+\beta)T_1}$

$P(T_1 > T_2) = \int_{T_2}^{\infty}\int_0^{\infty} \alpha e^{-\alpha T_1} \beta e^{-\beta T_2}\,dT_2\,dT_1 = \int_0^{\infty} \beta e^{-\beta T_2} - \left[e^{0} e^{-\alpha T_2}\right] dT_2 = \int_0^{\infty} \beta e^{-(\beta+\alpha)T_2}\,dT_2 \quad \boxed{\frac{\beta}{\alpha+\beta}}$

$P(T_1 > 2T_2) = \int_0^{\infty}\int_{2T_2}^{\infty} \kappa e^{-\kappa T_1} \beta e^{-\beta T_2}\, dT_1\, dT_2 = \frac{\beta}{(2\kappa+\beta)}$

20. Probability of packet collision: $f(t_1, t_2) = \frac{1}{T^2}$ $\boxed{\text{Joint Density}}$ from $[0,T]$

Time between arrivals:



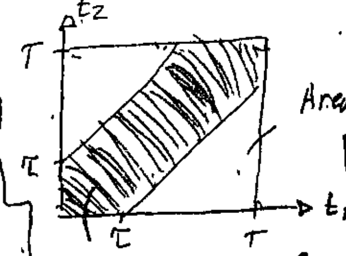
Integral is: $f(t_1,t_2) \times \frac{1}{2}T \cdot T = \boxed{\frac{T}{2}}$

$\int_0^{T}$ Integral is: $f(t_1,t_2) \times \text{Area}$

$= \frac{1}{T^2}\left(T^2 - (T-\tau)^2\right)$

$\boxed{= 1 - (1 - \tau/T)^2}$

21.



Area $(T-\tau)^2/2$

Area $T^2 - (T-\tau)^2$

$f(x)$ = Present chemical

$R(x)$ = Probability Detected

Y = concentration of a chemical in soil

If $g(y)$ is uniform, then $g(y) = \frac{\text{Probability of Detection} \times \text{Density Function}}{\text{Total outcomes of concentration}}$

$\boxed{= R(y) \cdot f(y) \Big/ \int_0^{\infty} R(x) f(x)\, dx}$

22. Poisson Distribution: $\frac{\lambda^k}{k!}e^{-\lambda} = p(x)$; $N(t_1, t_2)$ = Number of events.

If $t_0 < t_1 < t_2$; find the conditional distribution of $N(t_0, t_1)$ given $N(t_0, t_2) = n$

$N(t_0|t_1) = \frac{N(t_0, t_1)}{N(t_1)}$; $N(t_0|t_1) = \frac{N(t_0 t_2)}{N(t_2)} = \frac{n}{N(t_2)}$; $N(t_1, t_2) \cdot p = \lambda$; $p(x) = \frac{[N(T_1, T_2)p]^k}{k!} e^{-N(T_1,T_2)\cdot p}$

$\boxed{= \frac{[N(T_1)N(T_2)\cdot P_1 \cdot P_2]^k}{k!} e^{-N_{T_1} \cdot N_{T_2} \cdot P_1 \cdot P_2}}$

$P(N(t_i, t_j)) = e^{-\lambda(t_i - t_j)} \frac{[\lambda(t_j - t_i)]^x}{x!}$

$P(N(t_0, t_1) = x \mid N(t_0, t_2) = n) = \frac{P(N(t_0, t_1) = x, N(t_0, t_1) + N(t_1, t_2) = n)}{P(N(t_0, t_2) = n)}$

$= \frac{P(N(t_0, t_1) = x, N(t_1, t_2) = n - N(t_0, t_1))}{P(N(t_0, t_2) = n)} = \frac{P(N(t_0, t_1) = x, N(t_1, t_2) = n - x)}{P(N(t_0, t_2) = n)}$

$= \frac{e^{-\lambda(t_1 - t_0)} \frac{[\lambda(t_1 - t_0)]^x}{x!} \times e^{-\lambda(t_2 - t_1)} \frac{[\lambda(t_2 - t_1)]^{n-x}}{(n-x)!}}{e^{-\lambda(t_2 - t_0)} \frac{[\lambda(t_2 - t_0)]^n}{n!}}$ $\boxed{= \frac{n!}{x(n-x)!} \frac{(t_1 - t_0)^x (t_2 - t_1)^{n-x}}{(t_2 - t_0)^n}}$

23. $p(N|X) = \frac{p(N, X)}{p(X)}$

Binomial Distribution:

$p(k) = \binom{n}{k} p^k (1-p)^{n-k}$

N = Trials, p = probability of successs

= Binomial random variable with m trials and probability r

$p(X) = \frac{p(N, X)}{p(N|X)} = \frac{\binom{N}{k} p^k (1-p)^{N-k}}{\binom{m}{k} r^k (1-r)^{m-r}}$

$\boxed{= \binom{N}{m} (pr)^m (1 - pr)^{N-m}}$

24.

Section 3.5.2

Baysian Inference: Conditional: $f_{X|\theta}(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ $x = 0, 1, \dots n$

Joint Density: $f_{\theta,X}(\theta, x) = f_{X,\theta}(x|\theta) \cdot f(\theta)$

$= \binom{n}{x}\theta^x(1-\theta)^{n-x}$ $x = 0, 1, \dots n$; $0 \le \theta \le 1$

Marginal Density: $f_\theta(\theta) = \int_0^1 \binom{n}{x}\theta^x(1-\theta)^{n-x} d\theta = \int_0^1 \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x} d\theta$ by the fact $\Gamma(k) = (k-1)!$

...becomes the beta density $= \int_0^1 \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)} \theta^x(1-\theta)^{n-x} d\theta = \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)} \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)} = \frac{1}{n+1} \frac{\Gamma(n+1)}{\Gamma(n+1)} = \boxed{\frac{1}{n+1}}$

Conditional: $f_{\theta|X}(\theta|x) = \frac{f_{\theta X}(\theta, x)}{f_X(x)} = (n+1)\binom{n}{x}\theta^x(1-\theta)^{n-x} = (n+1)\frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x} = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x}$

Posterior Density is a β-density with $a = x+1, b = n-x+1$; $g(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u)^{b-1}$; $g'(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left[(a-1)u^{a-2}(1-u)^{b-1} + (b-1)u^{a-1}(1-u)^{b-2}\right] = 0$

$\frac{(a-1)}{(b-1)}(1-u) = u$

$\frac{(a-1)}{(b-1)} = u\left[1 + \frac{(a-1)}{(b-1)}\right]$

$\boxed{\theta = \frac{(a-1)(b-1)}{(b-1)[(b-1)+(a-1)]}}$

26. P is uniform from $[0,1]$, and conditional on $P = p$.

Let X be a Bernoulli Distribution with parameter p.

Find the conditional distribution of P given X

Bernoulli Distribution: $p(k) = \frac{n!}{k!} p^k (1-p)^{n-k}$

$f(x=0) = \int_0^1 p^0 (1-p)^1 dp = 1/2$

$f(x=1) = \int_0^1 p^1 (1-p)^0 dp = 1/2$

Find $f_{P|X}(P|X) = \frac{f(p, x)}{f(x)} = \frac{p^x (1-p)^{1-x}}{1/2} = \boxed{2p(1-p)^{1-x}}$

25 $f(x)$; $P(Y=x)=1/2$; $P(Y=-x)=1/2$; Show Y is symmetric about zero. Bernoulli Distribution: $P(k)=p^{x}(1-p)^{1-x}$

$$P(0)=\int_0^1 p^{0}(1-p)^{1}dp = 1-1/2 = 1/2$$

$$P(1)=\int_0^1 p^{1}(1-p)^{0}dp = \frac{1}{2}$$

Conditional Density of a random variable is expressed as:

$$f_{Y|X}(y|x)=f_{Y|X}(x|x)=1/2 \; ; \; f_{Y|X}(y|x)=f_{Y|X}(-x|x)=1/2$$

$$f_{Y|X}(y|x)=\frac{f_{YX}(x,y)}{f(x)} \; ; \; f(x,y)=f_{Y|X}(y|x)f(x)=\boxed{\frac{1}{2}f_X(x)}$$

$$f(x,x)=\frac{1}{2}f_X(x);$$

$$F_{Y|X}(y|x)=\frac{F(x,y)}{f_X(x)} \; ; \; F(x,y)=f_{Y|X}(y|x)f_X(x)=\boxed{\frac{1}{2}f_X(-x)} \qquad f_Y(y)=\frac{1}{2}f(x)+\frac{1}{2}f(x) = \boxed{f_Y(-y)}$$

27. Prove X and Y are independent if $\boxed{f_{X|Y}(x|y)=f_X(x)=\frac{f(x,y)}{f(y)}=\frac{f(x)f(y)}{f(y)}=f(x)}$

28. Show if $C(u,v)=uv$ is a copula. Why is it called "the independence copula?"
Copula: a joint cumulative distribution of random variables that have uniform marginal distributions.
The function $C(u,v)=uv$ is known by the independence copula because independent variables, and marginals, are seperable.

29. Marginal Density $=\lambda e^{-\lambda x}$

Farlie-Morgenstern Copula:

$$H(x,y)=F(x)G(y)\{1+\alpha[1-F(x)][1-G(y)]\}=\lambda_1 e^{-\lambda_1 x}\lambda_2 e^{-\lambda_2 y}\{1+\alpha[1-\lambda_1 e^{-\lambda_1 x}][1-\lambda_2 e^{-\lambda_2 y}]\}$$

$$h(x,y)=\frac{\partial^2}{\partial x\partial y}H(x,y)=\frac{\partial}{\partial x}\lambda_1 e^{-\lambda_1 x}\lambda_2^2 e^{-\lambda_2 y}\{1+\alpha[1-\lambda_1 e^{-\lambda_1 x}][1-\lambda_2 e^{-\lambda_2 y}]\}$$
$$+\lambda_1 e^{-\lambda_1 x}\lambda_2 e^{-\lambda_2 y}\{1+\alpha[1+\lambda_1 e^{-\lambda_1 x}][1+\lambda_2 e^{-\lambda_2 y}]\}$$

$$\boxed{\begin{aligned}&=\lambda_1^2 e^{-\lambda_1 x}\cdot\lambda_2^2 e^{-\lambda_2 y}\{1+\alpha[1-\lambda_1 e^{-\lambda_1 x}][1-\lambda_2 e^{-\lambda_2 y}]\}\\&+\lambda_1 e^{-\lambda_1 x}\lambda_2 e^{-\lambda_2 y}\{1+\alpha[1+\lambda_1^2 e^{-\lambda_1 x}][1-\lambda_2 e^{-\lambda_2 y}]\}\\&+\lambda_1^2 e^{-\lambda_1 x}\lambda_2 e^{-\lambda_2 y}\{1+\alpha[1-\lambda_1 e^{-\lambda_1 x}][1+\lambda_2 e^{-\lambda_2 y}]\}\\&+\lambda_1 e^{-\lambda_1 x}\lambda_2 e^{-\lambda_2 y}\{1+\alpha[1+\lambda_2 e^{-\lambda_2 x}][1+\lambda_2 e^{-\lambda_2 y}]\}\end{aligned}}$$

30. For $0\le\alpha\le1$ and $0\le\beta\le1$
Show $C(u,v)=\min(u^{1-\alpha}v, uv^{1-\beta})$
is a copula (Marshall-Olkin)

$$\boxed{\lim_{\alpha\to1}\lim_{\beta\to0} C(u,v)=\min(v,u)=1}$$

Joint Density:

$$h(x,y)=\frac{\partial^2}{\partial u\partial v}H(x,y)=\boxed{\min[(1-\alpha)u^{-\alpha},(1-\beta)v^{-\beta}]}$$

31. (X,Y) is a uniform disc of radius of 1. $f(x,y)=\begin{cases}\frac{1}{\pi} & x^2+y^2\le1\\ 0, & \text{otherwise}\end{cases}$ X and y are not independent because of the constraint $x^2+y^2=1$.

32. $f_R(r)$; Probability of passing per mesh: Probability Passing = $\frac{\text{Area Square}}{\text{Area mesh}}=\boxed{\frac{\pi f_R^2(r)}{(nW+(n+1)d)^2}}$

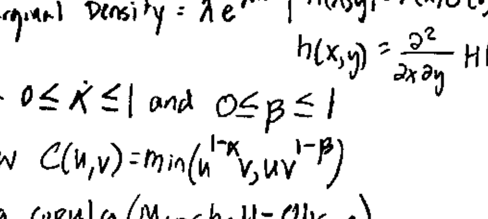
33 a) Posterior Density [Beta Density]: $f(\theta)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$; $a=x+1$; $b=n-x+1$

$$=\frac{\Gamma(N+2)}{\Gamma(2)\Gamma(N)}\theta(1-\theta)^{N-1}=(n+1)(n)\frac{\Gamma(n)}{\Gamma(2)\Gamma(n)}\theta(1-\theta^{N-1})$$

$$=\frac{(n+1)(n)}{\Gamma(2)=1}\theta(1-\theta^{N-1})$$

$$\boxed{=(n+1)(n)\,\theta(1-\theta^{N-1})}$$

b)

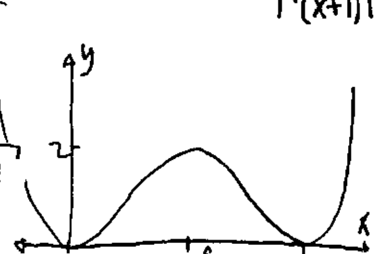


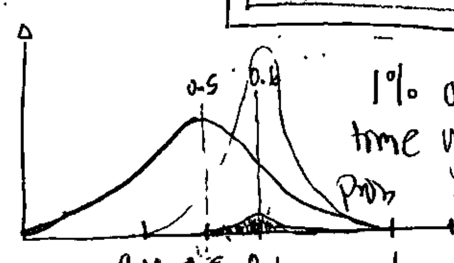
Graph this function:

$$\frac{\Gamma(n-x+1)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^{x}(1-\theta)^{n-x}$$

34. Beta Density: $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$; Where $a=b=3$.

Graph the Function: $\frac{\Gamma(6)}{\Gamma(3)\Gamma(3)}\theta^{3-1}(1-\theta)^{3-1}$



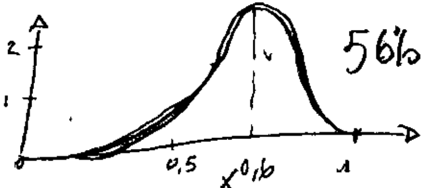


1% of the time your average prob exists

35. $n=20$ $x = \text{卌 卌 ||||}$

Beta Distribution: $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$ $a = x+1 = 13$, $b = n-x+1 = 8$

$\frac{\Gamma(21)}{\Gamma(13)\Gamma(8)}\theta^{12}(1-\theta)^{7}$ $20-14+2$

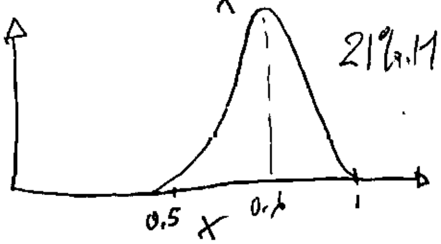


56% H:T

The amount of spins should arise to 50:50.

$n = 40$
$x = \text{卌 卌 卌 卌 卌 |}$
$a = 27$
$b = 14$

$\frac{\Gamma(41)}{\Gamma(14)\Gamma(27)}\theta^{26}(1-\theta)^{13}$



21% H : 79% T

36. Rejection Method $m(x) = \frac{M(x)}{\int_a^b M(x)\,dx}$; $P(X) = \frac{\sum_{i=1} f(Y)U}{M(x)} = 1$

Assuming a geometric random variable: $p\{X=i\} = (1-p)^{i-1}\cdot p$; $\sum P(X=i) = 1 - P(X > j-1) = 1-(1-p)^{j-1}$

$\boxed{M(X) = \sum f(j)\cdot U}$

By simulating a random variable: $1-(1-p)^{j-1} < U < 1-(1-p)^j$; $(1-p)^j \le 1-U < (1-p)^{j-1}$

$X = \min\left\{j : j \ge \frac{\log U}{\log(1-p)}\right\}$; $\boxed{X = 1 + \frac{\log U}{\log(1-p)}}$

37. $f(x) = 6x^2(1-x)^2$; $-1 \le x \le 1$ a) Rejection method:

$f'(x) = 12x(1-x)^2 + 12x^2(1-x)$
$0 = (1-x) - x$
$2x = 1$
$x = \frac{1}{2}$

$\boxed{\frac{\text{Part}}{\text{Total}}}$

$P(\text{accept}) = P(U \le f(T)/M(T)) = \frac{6x^2(1-x)^2}{\int_{-1}^{1} \frac{1}{2}\,dx} = \frac{6x^2(1-x)^2}{1}$

$P(x \le X \le x+dx) = \frac{P(\text{accept} \mid x \le T \le x+dx)\,P(x \le T \le x+dx)}{P(\text{accept})}$

$\boxed{= 6x^2(1-x)^2 \cdot \int_{-1}^{1} \frac{1}{2}\,dt}$

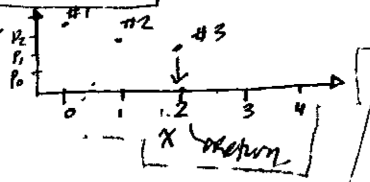
38. $f(x) = (1+\alpha x)/2$ for $-1 \le x \le 1$ and $-1 \le \alpha \le 1$

"Columbia - Accept Reject Method"

a) 1) Generate Y distributed as G.
2) Generate U (independent) from Y
3) If $U \le \frac{f(Y)}{cg(Y)}$, then set $X = Y$ ("accept") otherwise return ("reject")

b) int val = rand()
int U = rand()
if ($U \le f(Y)/g(Y)$); then "accept"; break
else return "reject"

39. $X = \{0,1,2\ldots\}$ with p_0, p_1, p_2. U is a random variable. If $U < p_0$, return $X=0$.
Else $U = U - p_0$.
If $U < p_1$, return $X=1$.
Else $U = U - p_1$



40. Suppose X & Y are Discrete random variables: 42. $P_{XY}(x,y)$ a) Generate $X \sim p_X(x) = \int_{-\infty}^{\infty} P_{XY}(x,y)\,dy$ a)

b) Accept X with $p(y \mid X) = \frac{P(x,y)}{P(x)}$

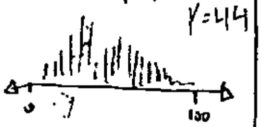
b) Else A.

$\boxed{\text{If X and Y should be continuous random variables, then the process would be rather of functions}}$

$X = \{1,2\ldots100\}$; $X = X, Y$
$Y = \{1,2,\ldots X\}$.

If $Y = 44$, then $P(44 \mid X) = \frac{P(X,44)}{P(X)}$

$P(X) = \frac{P(X,44)}{P(44 \mid X)} \Rightarrow$



41. T = exponential Random Distribution: $P(X) = \lambda e^{-\lambda x}$; $W = \pm 1$ with $p(1) = \frac{1}{2}$; $p(0) = \frac{1}{2}$.
$X = WT$. Show that $f_X(x) = \frac{\lambda}{2}e^{-\lambda|x|}$ is the double exponential distribution

$f_X(x) = p(W)\cdot p(X) = \boxed{\frac{1}{2}\lambda e^{-\lambda x}}$

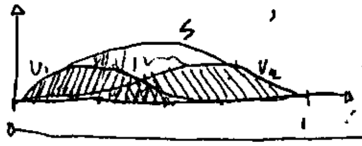
b) Prove $\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \le ce^{-|x|}$; $e^{-x^2/2+|x|} \le c\sqrt{2\pi}$; $-\frac{x^2}{2} + |x| \le \log c\sqrt{2\pi}$

$-\frac{x^2}{2} + |x| - \log c\sqrt{2\pi} \le 0$; $x^2 - 2|x| + \log c \le 0$

$x_1, x_2 = \frac{+2 \pm \sqrt{4-4(1)(\log c)}}{2(1)}$

Real $|x|$ implies C is between 0 to 1.

43. U_1 & U_2 from $[0,1]$; $S = U_1 + U_2$



44. X & Y $\{0,1,2\}$ $p(0)=\frac{1}{3}$; $p(1)=\frac{1}{3}$; $p(2)=\frac{1}{3}$; Frequency function of $X+Y$.

X	0	0	0	1	1	1	2	2	2
Y	0	1	2	0	1	2	0	1	2
$X+Y$	0	1	2	1	2	3	2	3	4

$X+Y$	0	1	2	3	4
N	1	2	3	2	1
$P(X+Y)$	$1/9$	$2/9$	$3/9$	$2/9$	$1/9$

45 Poisson Distribution:

$p(X) = \frac{\lambda^k e^{-\lambda}}{k!}$

$p_A + p_B = 1$

Prove N_A is a poisson with parameter $p_A\lambda$; $P(N_A = n) = \sum_{i=n}^{\infty} P[N_A = n \mid X_i = i]\,P[X=i] = \sum_{i=n}^{\infty}\binom{i}{n} p_A^{\,n}(1-p_A)^{i-n}\cdot\frac{\lambda^{i}}{i!}e^{-\lambda}$

$\boxed{\text{Law of Total Probability}}$ $= \frac{e^{-\lambda}\cdot\lambda^n p_A^{\,n}}{n!}\sum_{i=n}^{\infty}\frac{[(1-p_A)\lambda]^{i-n}}{(i-n)!} = \frac{e^{-\lambda}\lambda^n p_A^{\,n}}{n!}\sum_{j=0}^{\infty}\frac{[(1-p_A)\lambda]^{j}}{j!} = \frac{e^{-\lambda}\lambda^n p_A^{\,n}}{n!}e^{(1-p_A)\lambda}$ $= \boxed{e^{-p_A\lambda}\frac{(p_A\lambda)^n}{n!}}$

46. Let T_1 and T_2 be independent exponentials with λ_1 and λ_2. Find $T_1 + T_2$.

$P(T_1) = \lambda_1 e^{-\lambda_1}$; $P(T_2) = \lambda_2 e^{-\lambda_2}$; $T_1 + T_2 = \lambda_1 e^{-\lambda_1} + \lambda_2 e^{-\lambda_2}$;

$J = \begin{vmatrix} \frac{\partial T_1}{\partial r} & \frac{\partial T_1}{\partial s} \\ \frac{\partial T_2}{\partial r} & \frac{\partial T_2}{\partial s}\end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1\end{vmatrix} = 1$

$\boxed{F(T_1,T_2) = \lambda_1\lambda_2 e^{-(\lambda_1+\lambda_2)}}$

47. $P(X) = \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu)^2}{2\sigma_1^2}}$; $P(Y) = \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-\mu)^2}{2\sigma_2^2}}$; $Z = X+Y = \frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{(X-\mu_x)^2}{2\sigma_1^2}-\frac{(y-\mu_y)^2}{2\sigma_2^2}}$

Sums and Quotients: $X, Y, Z = X+Y$, then $Y = Z - X$; $p(z) = \sum_{x=-\infty}^{\infty} p(x, z-x)$; $p(x,y) = p_X(x)\cdot p_Y(y)$

$= \sum_{-\infty}^{\infty} p_X(x)\,p_Y(z-x)$ $\boxed{\text{Convolution}}$ $\boxed{\text{Discrete}}$

$f(z) = \int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-x^2/2}\cdot e^{-\frac{(z-x)^2}{2}}dx$

$= \int_{-\infty}^{\infty}\frac{1}{2\pi}e^{\frac{-x^2+z^2-2xz+x^2}{2}}dx = \frac{e^{-\frac{z^2}{2}}}{2\pi}\int_{-\infty}^{\infty}e^{-\left[\frac{2x^2-2xz}{2}\right]}dx$

$= \frac{e^{-z^2/2}}{2\pi}\times\sqrt{\pi}\,e^{z^2/4} = \boxed{\frac{1}{\sqrt{2}\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{z}{\sqrt{2}}\right)^2}}$

$F(z) = \iint f(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)\,dy\,dx =$

$f(z) = \int_{-\infty}^{\infty} f(x, z-x)\,dx = \int_{-\infty}^{\infty} f_x(x)\,f_y(z-x)\,dx$

48. $f(N_1) = \frac{\lambda^k}{k!}e^{-\lambda N_1}$; $f(N_2) = \frac{\lambda^k}{k!}e^{-\lambda N_2}$; $f(N) = \int_{-\infty}^{\infty} f(N_1, N-N_1)\,dN_1 = \int_{-\infty}^{\infty} f_x(N_1)\cdot f(N-N_1)\,dN_1 = \int_{-\infty}^{\infty}\frac{\lambda_1^k}{k!}e^{-\lambda_1 N_1}\cdot\frac{\lambda_2^k}{k!}e^{-\lambda_2(N-N_1)}\,dN_1$

$= \frac{(\lambda_1\lambda_2)^k}{(k!)^2}\int_{-\infty}^{\infty} e^{-\lambda_1 N_1 - \lambda_2 N + \lambda_2 N_1}\,dN_1 = \frac{(\lambda_1\lambda_2)^k}{k!^2}e^{-\lambda_2 N}\int_{-\infty}^{\infty}e^{-(\lambda_2-\lambda_1)N_1}\,dN_1$

$\boxed{= \frac{(\lambda_1\lambda_2)^k}{(k)!^2}e^{-\lambda_2 N}\cdot\frac{-1}{(\lambda_1-\lambda_2)}}$

49. $f(x,y) = \begin{cases}\lambda^2 e^{-\lambda y}; & 0 \le x \le y,\ \lambda > 0 \\ 0 & \text{elsewhere}\end{cases}$

$F(z) = \int_{0}^{s/2}\lambda^2 e^{-\lambda y}\,dy = \boxed{\lambda e^{2-\lambda s/2} - \lambda e^{-\lambda s}}$

50. X & Y are jointly continuous variables. Find $Z = X - Y$

$\boxed{P(z) = \int_{-\infty}^{\infty} f(x, x-z)\,dx = \int_{-\infty}^{\infty} f(x)\cdot f(x-z)\,dx}$

51. $Z = XY$; $f(x,y) = f(x, \frac{z}{x}) = f(\frac{z}{y}, y)$

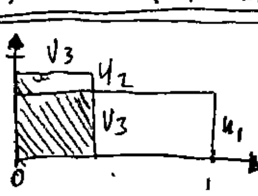
$F(z) = \int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_{-\infty}^{\infty} f(\frac{z}{y}, y)\frac{1}{y}\,dz\,dy$

$\boxed{f(z) = \int_{-\infty}^{\infty} f(\frac{z}{y}, y)\frac{1}{y}\,dy}$

52. $Z = X/Y$

$\boxed{F(z) = \int f(x,y)\,dx\,dy = \int f(yz, y)\,y\,dz\,dy;\ f(z) = \int f(yz, y)\,y\,dy}$

53.



$1 \ge \frac{\int_0^1\int_0^1 p(U_1)p(U_2)}{\int_0^1\int_0^1 p(U_3)^2}$

$\text{Area}_{12} = U_1\cdot U_2$; $P(U_1, U_2) = P(U_1)\cdot P(U_2)$;

$\text{Area}_{33} = U_3^2$; $P(U_3, U_3) = P(U_3)^2$;

$P(U_3^2 > U_1 U_2)$; $f(u_1,u_2,u_3) = \begin{cases} 1 & 0 \le u_i \le 1,\ i=1,2,3 \\ 0 & \text{Otherwise} \end{cases}$

To find the required probability consider,

$$P(U_3^2 \ge U_1 U_2) = \int_0^1 \int_0^1 \int_{\sqrt{u_1 u_2}}^1 f(u_1,u_2,u_3)\, du_3\, du_2\, du_1$$

$$= \int_0^1 \int_0^1 (1 - \sqrt{u_1}\sqrt{u_2})\, du_2\, du_1 = \int_0^1 \left(1 - \tfrac{2}{3}\sqrt{u_1}\right) du_1 \quad \boxed{= 5/9}$$

54. X, Y, and Z be independent $N(0,\sigma^2)$. Let Θ, Φ, R be Spherical coordinates.

$x = r\sin\phi\cos\theta$; $y = r\sin\phi\sin\theta$; $z = r\cos\phi$; $0 \le \phi \le \pi$; $0 \le \theta \le 2\pi$

Find $f(\Theta,\Phi,R)$ - $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\Theta, y, z)\, dx\, dy\, dz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x) f(y) f(z)\, dx\, dy\, dz$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{3/2}\sigma^3} e^{-(x^2+y^2+z^2)/2}\, dx\, dy\, dz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{3/2}\sigma^3} e^{-r^2/2\sigma^2} r^2 \sin\phi\, dr\, d\theta\, d\phi$$

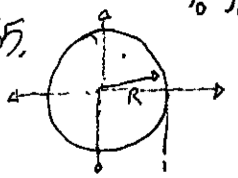
$$= \int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} \frac{1}{(2\pi)^{3/2}\sigma^3} e^{-r^2/2\sigma^2} r^2 \sin\phi\, dr\, d\theta\, d\phi = \frac{2\sigma^2}{2^2(2\pi)^{3/2}\sigma^3}\sqrt{\pi 2\sigma^2}\int_0^{2\pi}\int_0^{\pi} \sin\phi\, d\theta\, d\phi$$

$$= \frac{2\pi}{2(2\pi)^3}\left[-\cos(2\pi) + \cos(0)\right] = \frac{4\pi}{16\pi^2} = \boxed{\frac{1}{4\pi}}$$

$$f(\theta) = \int_0^{\pi}\int_0^{\infty} \frac{1}{(2\pi)^{3/2}\sigma^3} e^{-r^2/2\sigma^2} r^2 \sin\phi\, dr\, d\phi = \int_0^{\pi} \frac{1}{2(2\pi)} \sin\phi\, d\phi = \boxed{\frac{1}{2\pi}}$$

$$f(r) = \int_0^{\pi}\int_0^{2\pi} \frac{1}{(2\pi)^{3/2}\sigma^3} e^{-r^2/2\sigma^2} r^2 \sin\phi\, d\theta\, d\phi = \boxed{\frac{4\pi}{(2\pi)^{3/2}\sigma^3} e^{-r^2/2\sigma^2} \cdot r^2}$$

$$f(\phi) = \int_0^{2\pi}\int_0^{\infty} \frac{1}{(2\pi)^{3/2}\sigma^3} e^{-r^2/2\sigma^2} r^2 \sin\phi\, d\theta\, d\phi = \frac{2\pi \cdot 2\sigma^2}{(2\pi)^{3/2}\sigma^3}\sqrt{2\pi\sigma^2}\, \sin\phi = \boxed{\frac{1}{2\pi^2}\sin\phi}$$

55. 

$R\,[0,1]$ a) $X = R\cos\Theta$, $Y = R\sin\Theta$ a) Find $f(R,\Theta)$ $f(R\cos\Theta, R\sin\Theta)$

$\Theta\,[0,2\pi]$ b) $f(x) = \int_0^{\infty} \frac{1}{2\pi} e^{-(x^2+y^2)/2}\, dx\, dy = \frac{\sqrt{2\pi}}{2\pi} e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ $\boxed{= \frac{r}{2\pi} e^{-r^2/2}}$

$A = \pi R^2$; $A = \pi \sqrt{x^2+y^2}^{\,2}$

c) The density is uniform over the disk.

$\boxed{f(y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}}$

$\frac{A}{x^2+y^2}$; $x^2+y^2 \le 1$

56. Exponential Random variables: $\lambda e^{-\lambda}$; $x = \lambda_x e^{-\lambda|x|}$; $y = \lambda_y e^{-\lambda|y|}$

$$f(x,y) = \lambda_x e^{-\lambda x} \cdot \lambda_y e^{-\lambda y} = \lambda_x \lambda_y e^{-\lambda(x+y)} \boxed{= \lambda^2 e^{-\lambda r(\cos\theta + \sin\theta)}}$$

r and θ are not independent.

57. $Y_1 = N(0,1)$; $Y_2 = N(0,2)$; $\rho = 1/\sqrt{2}$; Find $X_1 = a_{11}y_1 + a_{12}y_2$ and $X_2 = a_{21}y_1 + a_{22}y_2$

$J(y_1,y_2) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}\cdot a_{22} - a_{12}a_{21} = 1$

Example C: (Section 3.6.2)

$Y_1 = X_1$; $Y_2 = X_1 + X_2$; $J(x,y) = \det\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = 1$

$$f_{Y_1Y_2}(y_1,y_2) = \frac{1}{2\pi}\exp\left[-\frac{1}{2}\left[y_1^2 + (y_2 - y_1)^2\right]\right]$$

$$= \frac{1}{2\pi}\exp\left[-\frac{1}{2}(2y_1^2 + y_2^2 - 2y_1y_2)\right]$$

$\sigma_{Y_1}\sigma_{Y_2}\sqrt{1-\rho^2} = 1$

$1\cdot(2)\sqrt{1-\rho^2} = 1$; $1 - \frac{1}{4} = \rho^2$

$\boxed{X_1 = y_1 ;\ X_2 = y_2 - y_1 ;}$

If X_1 & X_2 are $N(\mu,\sigma^2)$ then $f_{Y_1Y_2}(y_1,y_2)$ is bivariate normal.

$Y_1 = a_1X_1 + b_1$; $Y_2 = a_2X_2 + b_2$

58. $J(x,y)=\det\begin{bmatrix}1&0\\0&1\end{bmatrix}=1$; $\boxed{f_{Y_1Y_2}(y_1,y_2)=\frac{1}{2\pi}e\left[-\frac{1}{2}\left([(y_1-b_1)/a_1]^2+[(y_2-b_2)/a_2]^2\right)\right]}$ $\frac{1}{\ }\times\gamma\left[\frac{1}{2}\left(\right.\right.$

59. $Y_1=a_{11}X_1+a_{12}X_2+b_1$; $Y_2=a_{21}X_1+a_{22}X_2+b_2$

① $f(x_1,x_2)=\frac{1}{2\pi}e^{-\frac{1}{2}(x_1^2+x_2^2)}$

② Linear Transformation: $\left(\frac{-a_{22}}{a_{12}}\right)Y_1=\left(\frac{-a_{22}}{a_{12}}\right)a_{11}X_{11}+\left(\frac{-a_{22}}{a_{12}}\right)a_{12}X_2+\left(\frac{-a_{22}}{a_{12}}\right)b_1$ $\quad\boxed{X\left(\frac{-a_{22}}{a_{11}}\right)}$

$\left(\frac{-a_{22}}{a_{12}}\right)Y_1+Y_2=\left[\left(\frac{-a_{22}}{a_{12}}\right)a_{11}+a_{21}\right]X_1+\left[\cancel{a_{22}}-\cancel{a_{22}}\right]X_2+\left(\frac{-a_{22}}{a_{12}}\right)b_1+b_2$ $\quad\boxed{①+②}$

$\boxed{\text{Solve for } X_1}$

$$X_1=\frac{\left(\frac{-a_{22}}{a_{11}}\right)Y_1+Y_2+\left(\frac{a_{22}}{a_{12}}\right)b_1+b_2}{\left[\left(\frac{-a_{22}}{a_{12}}\right)a_{11}+a_{21}\right]}=\frac{+a_{22}(Y_1-b_1)-a_{12}(Y_2-b_2)}{(a_{22}a_{11}-a_{21}\cdot a_{12})}$$

③ Solve the Jacobian: $$|J|=\begin{vmatrix}\frac{\partial x_1}{\partial y_1}&\frac{\partial x_1}{\partial y_2}\\\frac{\partial x_2}{\partial y_1}&\frac{\partial x_2}{\partial y_2}\end{vmatrix}=\begin{vmatrix}\frac{a_{22}}{(a_{22}a_{11}-a_{21}a_{12})}&\frac{-a_{12}}{(a_{22}a_{11}-a_{21}a_{12})}\\\frac{a_{21}}{(a_{22}a_{11}-a_{21}a_{12})}&\frac{a_{11}}{(a_{22}a_{11}-a_{21}a_{12})}\end{vmatrix}=\frac{1}{(a_{22}a_{11}-a_{21}a_{12})}$$

④ Solve for the new bivariate density:
$$f(y_1,y_2)=\frac{1}{2\pi\sigma_{Y_1}\sigma_{Y_2}\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1-\mu_{Y_1}}{\sigma_{Y_1}}\right)^2-2\rho\frac{(y-\mu_{Y_1})(y-\mu_{Y_2})}{\sigma_{Y_1}\sigma_{Y_2}}+\left(\frac{y_2-\mu_{Y_2}}{\sigma_{Y_2}}\right)^2\right]\right]$$

⑤ Evaluate $X_1^2+X_2^2$
$$x_1^2+x_2^2=\frac{\{(y_1-b_1)^2(a_{22}^2+a_{21}^2)+(y_2-b_2)^2(a_{12}^2+a_{11}^2)-2(y_1-b_1)(y_2-b_2)(a_{22}a_{12}+a_{11}a_{21})\}}{(a_{22}a_{11}-a_{21}a_{12})^2}$$

⑥ Since $X_1,X_2\sim N(0,1)$, then $a_{11}X_1+a_{22}X_2+b_1\sim N(b_1,a_{11}^2+a_{12}^2)$
$a_{21}X_1+a_{22}X_2+b_2\sim N(b_2,a_{21}^2+a_{22}^2)$ $\quad\boxed{\begin{aligned}\mu_{Y_1}&=b_1 ; \sigma_{Y_1}^2=a_{11}^2+a_{12}^2\\\mu_{Y_2}&=b_2 ; \sigma_{Y_2}^2=a_{21}^2+a_{22}^2\end{aligned}}$

⑦ Reviewing $x_1^2+x_2^2$, ... first term $\left(\frac{(y_1-\mu_{Y_1})}{\sigma_{Y_1}}\right)^2\Big/\frac{(\sigma_{Y_2}\sigma_{Y_1})}{(a_{22}a_{11}-a_{21}a_{12})^2}=\left(\frac{y_1-b_1}{\sqrt{a_{11}^2+a_{12}^2}}\right)^2$

⑧ Solving for $\frac{1}{1-\rho^2}=\frac{\sigma_{Y_1}\sigma_{Y_2}}{(a_{22}a_{11}-a_{21}a_{12})^2}$; $\boxed{\rho=\frac{(a_{22}^2a_{12}^2+a_{21}^2a_{11}^2)}{\sqrt{(a_{22}^2+a_{21}^2)(a_{11}^2+a_{12}^2)}}}$

⑨ $\boxed{\sigma_{Y_1}\sigma_{Y_2}\sqrt{1-\rho^2}=|(a_{22}a_{11}-a_{21}a_{12})|=J^{-1}}$

60. Pseudorandom variables occur from the previous bivariate normal by a cumulative sum distribution from $-\infty$ to X.

61. X & Y are continuous random variables. $U=a+bX$; $V=c+dY$. $f(U,V)=f(x,y)\cdot J^{-1}$
$=f\left(\frac{U-a}{b},\frac{V-c}{d}\right)\left|\frac{1}{\begin{vmatrix}b&0\\0&d\end{vmatrix}}\right|$

$\boxed{\text{Proposition A}}$

$\boxed{=f\left(\frac{U-a}{b},\frac{V-c}{d}\right)\frac{1}{bd}}$

62. X & Y are $N(0,1)$; $P(X^2+Y^2\le1)=\frac{1}{2\pi}e^{-(x^2+y^2)/2}$; $\le\frac{1}{2\pi}e^{-1/2}$

$\boxed{\frac{1}{2\pi}\le P(X^2+Y^2\le1)\le\frac{1}{2\pi}e^{-1/2}}$

63. a) $\boxed{\begin{aligned}&X+Y=U ; f(u,v)=f_{XY}\left(\frac{v+z}{2},\frac{v-z}{2}\right)\cdot\frac{1}{2}\\&X-Y=V\end{aligned}}$

b) $\boxed{\begin{aligned}&XY=Z ; f(U,V)=f_{XY}\left(\sqrt{ZV},\sqrt{Z/V}\right)\frac{1}{2|V|}\\&Y/X=V\end{aligned}}$

c) $X\sim N(0,1)$, $Y\sim N(0,1)$

64. $X+Y=Z$, $X|Y=V$; $f_{XY}\left(\frac{VZ}{(V+1)},\frac{Z}{(V+1)}\right)\cdot\frac{1}{\begin{vmatrix}1&1\\\frac{1}{Y}&-\frac{X}{Y^2}\end{vmatrix}}=f_{XY}\left(\frac{VZ}{(V+1)},\frac{Z}{(V+1)}\right)\frac{-Y^2}{X+Y}$

$=f_{XY}\left(\frac{VZ}{(V+1)},\frac{Z}{(V+1)}\right)\frac{-Z^2/(V+1)^2}{\frac{VZ}{(V+1)}+\frac{Z}{(V+1)}}$ $\boxed{f_{XY}\left(\frac{VZ}{(V+1)},\frac{Z}{(V+1)}\right)\frac{-Z}{(V+1)^2}}$

65. Exponential random variable: $\lambda e^{-\lambda x}$; $f_{xy}($

67. n-chips; $P(\text{failure} \mid \text{chips} \geq 2)$; Exponential Dist. $p(x) = \lambda e^{-\lambda x}$

Order Statistics:

$f_u(u) = n[F(u)]^{n-1} f(u)$; $u \leq U \leq u + du$

K^{th}-order Statistic:

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} f(x) F^{k-1}(x)[1-F(x)]^{n-k}$$

$f(x) = \lambda e^{-\lambda x}$ $\quad F(x) = \int_0^x f(x) = 1 - e^{-\lambda x}$

$$f_k(x) = \frac{n!}{(2-1)!(n-2)!}\left[\lambda e^{-\lambda x}\right]\left[1-e^{-\lambda x}\right]^{2-1}\left[1-1+e^{-\lambda x}\right]^{n-2}$$

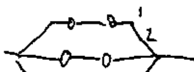
$$= \frac{n!}{(n-2)!}\left[\lambda e^{-\lambda x}\right]\left[1-e^{-\lambda x}\right]\left[e^{-\lambda x}\right]^{n-2}$$

$$= n(n-1)\lambda e^{-(n-1)\lambda x}\left[1-e^{-\lambda x}\right]$$

$$\boxed{= n(n-1)\lambda\left[e^{-(n-1)\lambda x} - e^{-n\lambda x}\right]}$$

65. Exponential Random variable: $p(x) = \lambda e^{-\lambda x}$;

o-o-o-o-o-o, $p(x_1, \dots, x_n) = \prod_{i=1}^{n} \lambda_i e^{-\lambda x} = \lambda_1 \lambda_2 \cdots \lambda_n \boxed{e^{-\sum \lambda_i}}$

66. 

$F(t) = p(1A)p(1B) + p(2A)p(2B) + p(3A)p(3B)$

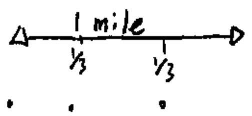
$= 3\lambda^2 e^{-2\lambda t}$

$f(t) = \frac{d}{dt} F(t) = \boxed{-6\lambda^3 e^{-2\lambda t}}$

Notes about order statistics: Multinomial + Differential Argument.

68. U_1, U_2, and U_3 be independent uniform random variables.

a) Find $f(U_1, U_2, U_3) = n! \prod_{i=1}^{n} f(U_i) = 3 \cdot 2 \cdot 1 f(U_1) \cdot f(U_2) \cdot f(U_3)$

b) 

$\int\int f(u)\, du = \int_0^{1/3} u\, du = \frac{(1/3)^2}{2} = \frac{1}{18}$

P(no stations 1/3 apart) = 1 − P(any two stations 1/3)

$= 1 - \frac{1}{18}$

$= 0.94.$

70.

69. Weibull Random Variables:

$f(t) = \beta \alpha^{-\beta} t^{\beta-1} e^{-(t/\alpha)^\beta}$

$f_k(x) = \frac{n!}{(k-1)!(n-k)!} f(x) F^{k-1}(x)[1-F(x)]^{n-k}$

if $k = 1, n = n$

$$f_k(t) = \frac{n!}{(1-1)!(n-1)!} f(t) \cdot F(t)^{0} \cdot [1-F(t)]^{n-1}$$

$$F(t) = \int_0^t \beta \alpha^{-\beta} t^{\beta-1} e^{-(t/\alpha)^\beta} dt \quad \left(\frac{t}{\alpha}\right)^\beta = u$$

$\beta\left(\frac{t}{\alpha}\right)^{\beta-1} \frac{1}{\alpha} dt = du$

$$= \int_0^{(t/\alpha)^\beta} \beta \alpha^{-\beta}\left(\frac{\alpha^\beta}{\beta}\right) e^{-u}\, du$$

$\frac{\beta}{\alpha^\beta} t^{\beta-1} dt = du$

$t^{\beta-1} dt = \frac{\alpha^\beta}{\beta} du$

$= 1 - e^{-(t/\alpha)^\beta}$

$$f_k(t) = \frac{n\beta}{\alpha^\beta} t^{\beta-1} e^{-(t/\alpha)^\beta} \cdot e^{-(n-1)(t/\alpha)^\beta}$$

$$\boxed{= \frac{n\beta}{\alpha^\beta} t^{\beta-1} e^{-n\left(\frac{t}{\alpha}\right)^\beta}}$$

Joint Distribution of Order Statistics

$x \leq X_{(1)} \leq x + dx$; $y \leq X_{(n)} \leq y + dy$

$V = X_{(1)}$; $U = X_{(n)}$

$f(u, v) = n(n-1) f(v) f(u)[F(u) - F(v)]^{n-2}$ $\quad u \geq v$

Uniform case: $f(u,v) = n(n-1)(u-v)^{n-2}$, $1 \geq u \geq v \geq 0$

$$F(x,y) = \int_{F(y)-F(x)}^{F(y)} F(y)^{n-1} = \boxed{F(y)^n - [F(y) - F(x)]^n}$$

71. $X_1, \dots, X_n$: $f_{x_1} \dots f_{x_n}$: $f(r) = \dfrac{\int_{-\infty}^{X_{(n)}} f(v+r, v)^{n-1} dv}{\int_{-\infty}^{\infty} f(v+r, v)^{n-1} dv} = \dfrac{\int_{-\infty}^{X_{(n)}} f(v+r, v)^{n-1} dv}{\int_{-\infty}^{\infty} f(v+r, v)^{n-1} dv} = \dfrac{f(X_{(n)}) - F(-\infty)}{F(\infty) - F(-\infty)}$

72. Five numbers $[0,1]$; probability $\left[\frac{1}{4} \leq X_1, X_2, X_3, X_4, X_5 \leq \frac{3}{4}\right] = \int\int\int\int\int f(u) = \int\int\int\int u\, du = \int\int\int \frac{u^2}{2} du = \int\int \frac{u^3}{6} = \int_{1/4}^{3/4} \frac{u^4}{24} = \frac{\left(\frac{3}{4}\right)^5 - \left(\frac{1}{4}\right)^5}{120}$

$= \frac{242/4^5}{120}$

4 16 64 256 1024

73. Definition of a independent random variable:

$F(x_1 \dots x_n) = F_{x_1}(x_1) \cdot F_{x_2}(x_2) \cdots F_{x_n}(x_n)$

Definition k^{th} order Statistic:

$f_k(x) = n! f(x_1) f(x_2) f(x_3) \cdots f(x_n)$

74. n-servers; $T = \lambda e^{-\lambda t}$; $P(\text{Service Time} \geq t) = P(\text{No departure during } t) = P_N(t)$; $P_N(t) = e^{-\mu t}$

$$S(t) = P(T \leq t) = 1 - P(T \geq 1) = 1 - e^{-\mu t}$$

75. $\frac{d}{dt} S(t) = \mu e^{-\mu t}$; $S(t) = \begin{cases} \mu e^{-\mu t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$ || Distribution of waiting time $\frac{1}{\mu}$ with a variance $\frac{1}{\mu^2}$.

Find the joint density of $X_{(i)}$ and $X_{(j)}$, $i<j$.

$f_{ij}(u,v) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \times F(u)^{i-1} \times f(u) [F(v) - F(u)]^{j-i-1} \times f(v) [1-F(v)]^{n-j}$ ... continued...

76. Prove Theorem A: $f_k(x) = \frac{n!}{(k-1)!(n-k)!} f(x) F^{k-1}(x) [1-F(x)]^{n-k}$ is derived from interdependent paths, $f(x) F^{k-1}(x) [1-F(x)]^{n-k} \times$ Multinomial theorem

$$\boxed{f_k(x) = \frac{n!}{(k-1)!(n-k)!} f(x) F^{k-1}(x) [1-F(x)]^{n-k}}$$

75. $(i-1)$, $(j-i-1)$; x, $x+dx$

$$f(x,y) = \lim_{dx \to 0} \lim_{dy \to 0} \frac{P(x \leq X_i \leq x+dx, y \leq X_j \leq y+dy)}{dx\, dy}$$

Multinomial Probability Law:

$$\frac{n!}{(i-1)!(j-i-1)!(n-j)!} P_1^{i-1} P_2 \cdot P_3^{j-i-1} P_4 \cdot P_5 \cdots = F(X)$$

$$\boxed{P_1 = P(X \leq X_i \leq x+dx)}$$

$P_2 = P(x \leq X_i \leq x+dx) = F(x+dx) \cdot F(X)$

$P_3 = P(x+dx < X_i \leq y) = F(y) \cdot F(x+dx)$

$P_4 = P(y < X_i < y+dy) = F(y+dy) \cdot F(y)$

$P_5 = P(X_i > y+dy) = 1 - P(X \leq y+dy) = 1 - F(y+dy)$

$$f(x,y) = \lim_{dx \to 0} \lim_{dy \to 0} \frac{P(E)}{dx\, dy}$$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \times F^{(i-1)} \times \lim_{dx \to 0} \frac{F(x+dx) - F(x)}{dx} \times \lim_{dx \to 0} F(y) - F(x+dx)^{j-i-1} \times \lim_{dy \to 0} \frac{F(y+dy) - F(y)}{dy} \times \lim_{dy \to 0} F(y) - F(x+dx)^{j-i-1} \times \lim_{dy \to 0} \frac{F(y+dy) - F(y)}{dy} \times \lim_{dy \to 0} 1 - F(y+dy)$$

77. $U_{(k)} - U_{(k-1)}$; $U_{i=1,\ldots,n}$.

$f(U_k) = n! \prod_{i=1}^{n} f(U_i)$; $f(U_{k-1}) = n! \prod_{i=1}^{n-1} f(U_i)$

$$\boxed{f(U_k) - f(U_{k-1}) = n!\, f(U_1) \cdot f(U_2) \cdots f(U_{n-1}) (f(U_n) - 1)}$$

78. Show $\int_0^1 \int_0^y (y-x)^n \, dx\, dy = \frac{1}{(n+1)(n+2)}$; $\int_0^1 \frac{-1}{n+1} (y-x)^{n+1} \Big|_0^y dy = \frac{-1}{(n+1)} \int_0^1 \left[(y-1)^{n+1} - y^{n+1}\right] dy \cdot \frac{-1}{(n+1)(n+2)} (y-1)^{n+2} - (y)^{n+2} \Big|_0^1$

$$= \frac{-1}{(n+1)(n+2)} \left[0 - 1^{n+2} - 1^{n+2} + 0\right] = \boxed{\frac{1}{(n+1)(n+2)}}$$

79. T_1 & T_2 are exponential random variables; $R = T_{(2)} - T_{(1)}$

$$F(R, T_2) = \iint f(T_1, T_2)\, dT_1\, dT_2 = \iint f(T_1, R - T_1)\, dT_1\, dR = \iint \lambda e^{-\lambda T_1} \cdot \lambda e^{-\lambda(R - T_1)}\, dT_1\, dR$$

$$\boxed{\begin{matrix} T_2 = R - T_1 \\ dT_2 = dR \end{matrix}}$$

$$f(z) = \int_0^\infty \lambda_1 \lambda_2 e^{-(\lambda_1 + \lambda_2) T_1 - \lambda_2 R} dT_1$$

$$\boxed{= \frac{-\lambda_1 \lambda_2}{(\lambda_1 - \lambda_2)} e^{-(\lambda_2 R)}}$$

80. $U_1 \ldots U_n$, V = uniform, U_i = independent.

a) Find $P(V \leq U_{(n)}) = \int_0^V n! \prod U_1 \cdot U_2 \cdots U_n \, dU = n! \prod f(U_i)^2 / 2 \Big|_0^V = \boxed{n! \prod_{i=1}^{n} f(V)^2 / 2}$

b) $P(U_{(1)} < V < U_{(n)}) = \int_{U_{(1)}}^{U_{(n)}} n! \prod f(V_i)\, dV_i = n! \left[\prod_{i=1}^{n} f(V_n)^2 - \prod f(U_1)^2\right] / 2$

$$P(V \leq U_{(n)}) = \int_0^{U_{(n)}} n! \prod f(V_i)\, dV = \boxed{n! \prod f(U_n)^2 / 2}, \quad P(U_{(1)} < V < U_{(n)}) = \int_{U_{(1)}}^{U_{(n)}} n! \prod f(U_n)\, dV = \boxed{\frac{n! \prod f(V)^2 - f(V_1)^2}{2}}$$

Chapter 4: 1) Prove if $|X| < M < \infty$, then $E(X)$ exists. $\mu = \sum x_i p(x_i)$; $M_1 + M_2 + \cdots + M_\infty$

$= x_1 p(x_1) + x_2 p(x_2) + \cdots + x_\infty p(x_\infty)$

2) $F(X) = 1 - X^{-\kappa}$, $x \geq 1$; a) $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-\infty}^{\infty} x \frac{d}{dx}(1 - x^{-\kappa})\,dx$

$= \int_1^{\infty} x(\kappa x^{-(\kappa+1)})\,dx = \int_1 \kappa x^{-(\kappa)} dx = \frac{-\kappa x^{-(\kappa+1)}}{(\kappa+1)}\Big|_1^{\infty} = \frac{-\kappa}{(\kappa+1)}(\infty)^{-(\kappa+1)} \frac{1}{\ } = \boxed{\frac{\kappa}{(\kappa+1)}} = E(X)$; $\boxed{\text{Definition of Expectation}}$

b) $Var(X) = E\{[X - E(X)]^2\}$

$= \int_1^{\infty} \left[X - \frac{\kappa}{\kappa+1}\right]^2 dx = \int_1 \left[x^2 - 2x\frac{\kappa}{(\kappa+1)} + \left(\frac{\kappa}{\kappa+1}\right)^2\right] dx = \frac{x^3}{3} - x^2\left(\frac{\kappa}{\kappa+1}\right) + \left(\frac{\kappa}{\kappa+1}\right)^2 x \Big|_1^{\infty} = \infty.$

$= E(X^2) - \mu^2 = \int_1^{\infty} x^2(\kappa x^{-(\kappa+1)})\,dx = \int_1^{\infty} \kappa x^{-\kappa+1}\,dx = \kappa \cdot x^{-(\kappa+2)}/(\kappa+2)\Big|_1^{\infty} = \frac{\pm\kappa}{(\kappa+2)} + \left(\frac{\kappa}{\kappa+1}\right)^2 = \boxed{\frac{\kappa}{(\kappa+2)} - \left(\frac{\kappa}{\kappa+1}\right)^2}$

$-\left(\frac{\kappa}{\kappa+1}\right)^2 \qquad -\left(\frac{\kappa}{\kappa+1}\right)^2$

3. Find E(X) and Var(X) for Chapter 2: Problem #3.

K	F(K)
0	0
1	0.1
2	0.3
3	0.7
4	0.8
5	1.0

$E(X) = \sum x p(x) = \sum x f(x) = \sum x[F(x_i) - F(x_{i-1})] = 5[1.0 - 0.8] + 4[0.8 - 0.7] + 3[0.7 - 0.3] + 2[0.3 - 0.1] + 1[0.1 - 0]$

$= \frac{5\cdot 2}{10} + \frac{4\cdot 1}{10} + \frac{3\cdot 4}{10} + \frac{2\cdot 2}{10} + \frac{1\cdot 1}{10}$

$= \frac{10 + 4 + 12 + 4 + 1}{10} = \frac{31}{10} = \boxed{3.1}$

$Var(X) = (0.2 - 3.1)^2(0.2) + (0.1 - 3.1)^2(0.1) + (0.4 - 3.1)^2 0.4 + (0.2 - 3.1)^2 \cdot 0.2 + (0.1 - 3.1)^2 \cdot 0.1$

$\boxed{= 1.62}$

4. $P(X = k) = 1/n$ for $k = 1, 2, \ldots, n$: Find $E(X)$ and $Var(X)$; $E(X) = 1\cdot\left(\frac{1}{n}\right) + 2\cdot\left(\frac{1}{n}\right) + 3\left(\frac{1}{n}\right) + \cdots + n\left(\frac{1}{n}\right) = \frac{n(n+1)}{2}\cdot\frac{1}{n}$

$Var(X) = \left[\left(1 - \frac{(n+1)}{2}\right)^2 + \left(2 - \frac{(n+1)}{2}\right)^2 + \cdots + \left(n - \frac{(n+1)}{2}\right)^2\right]\frac{(n+1)n}{2}$

5. $f(x) = \frac{1 + \alpha x}{2}$; $-1 \leq x \leq 1$; $-1 \leq \alpha \leq 1$

$E(X) = \int_{-1}^{1} \frac{(1 + \alpha x)}{2} x\,dx = \frac{x^2}{4} + \frac{\alpha x^3}{6}\Big|_{-1}^{1} = -\frac{1}{4} + \frac{\alpha}{6} + \frac{1}{4} + \frac{\alpha}{6} = \frac{\alpha}{3}$; $E(x^2) = \int_{-1}^{1} x^2 \frac{(1 + \alpha x)}{2}\,dx = \frac{1}{2}\left[\frac{x^3}{3} + \frac{\alpha x^4}{4}\right]_{-1}^{1} = \boxed{\frac{1}{3}}$

$Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{3} - \frac{\alpha^2}{9} = \boxed{\frac{3 - \alpha^2}{9}}$

6. $f(x) = 2x$; $0 \leq x \leq 1$

a) $E(X) = \int_0^1 2x^2 dx = \boxed{\frac{2}{3}}$ b) $Y = X^2$; Find $E(Y) = E(X^2) = \int_0^1 2x^3 dx = \boxed{\frac{1}{2}}$

c) $E(X^2) = \int_0^1 2x^3 dx = \frac{2}{3}$; d) $Var(X) = E\{[X - E(X)]^2\} = \int_0^1 (x - \frac{2}{3})^2 2x\,dx = \boxed{\frac{1}{18}}$

Theorem B: $Var(X) = E(X^2) - E(X)^2 = \frac{2}{3} - \left(\frac{2}{3}\right)^2 = \boxed{\frac{1}{18}}$

8. Average $= \sum$ weight $\times x_i$

$= \boxed{\sum p_i \cdot x_i}$

7.

X	0	1	2
P(X)	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{8}$

a) $E(X) = \sum x f(x) = 0\left(\frac{1}{2}\right) + 1\left(\frac{3}{8}\right) + 2\left(\frac{1}{8}\right) = \boxed{5/8}$

b) $Y = X^2$. $E(Y) = 0^2\left(\frac{1}{2}\right) + 1^2\left(\frac{3}{8}\right) + 2^2\left(\frac{1}{8}\right) = \boxed{7/8}$

c) Theorem A: a) $E(Y) = \sum g(x)p(x) = 0^2\left(\frac{1}{2}\right) + 1^2\left(\frac{3}{8}\right) + 2^2\left(\frac{1}{8}\right) = \boxed{7/8}$

b) $E(Y) = \int_{-\infty}^{\infty} g(x) f(x)\,dx$;

9. c = \$ to stock an item

s = \$ to sell an item.

$P(k)$ = Number of items by customer.

$s\sum_{k=n}^{\infty} p(k) > c\sum_{k=1}^{n-1} p(k)$, and $s\sum_{k=n+1}^{\infty} p(k) < c\sum_{k=1}^{n} p(k)$

Selling should be greater than cost.

Future sales should be cheaper than cost.

$n + 1$

d) $Var(X) = E[(X - E(X))^2] = \frac{1}{2}(0 - 5/8)^2 + \frac{3}{8}(1 - 5/8)^2 + \frac{1}{8}(2 - 5/8)^2 = \frac{31}{64}$

$Var(X) = E(X^2) - E(X)^2 = 7/8 - (5/8)^2 = \frac{31}{64}$

10. n-items "Random" Search(n) $E(X)=\sum_{i=1}^{n} p(x_i)x_i=\sum\left(\frac{1}{n}\right)x_i=\boxed{\frac{\sum x_i}{n}}$

11. $E(X)=\sum p_i x_i$; $E(X)=\sum^{n} x_i(1-p_1)(1-p_2)\cdots(1-p_{x-1})p_x$

$p_1+2(1-p_1)p_2+3(1-p_1)(1-p_2)p_3+\cdots+n(1-p_1)(1-p_2)(1-p_n)$

12. Suppose $E(X)=\mu$ and $Var(X)=\sigma^2$. Let $Z=(X-\mu)/\sigma$. Show $E(Z)=0$ and $Var(Z)=1$

$E(Z)=E\left(\frac{X-\mu}{\sigma}\right)=\frac{E(X)-E(\mu)}{\sigma}=\frac{\mu-\mu}{\sigma}=0$; $Var(Z)=Var\left(\frac{X-\mu}{\sigma}\right)=\frac{Var(X)}{\sigma^2}=\frac{\sigma^2}{\sigma^2}=\boxed{1}$

13. $E(X)=\int_0^{\infty} x\cdot f(x)\,dx$; Product Rule: $[xF(x)]'=F(x)+xf(x)$; $1-F(x)=xf(x)$: $E(X)=\int_0^{\infty}[1-F(x)]\,dx$

$E(X)=\int_0^{\infty}[1+Ae^{-\lambda x}]\,dx=x+\frac{e^{-\lambda x}}{\lambda}\Big|_0^{\infty}=\boxed{\frac{1}{\lambda}}$

14. $f(x)=2x$; $0\le x\le 1$; $E(X)=\int_0^1 2x^2\,dx=\frac{2}{3}x^3\Big|_0^1=\boxed{\frac{2}{3}}$ b) $E(X^2)=\int_0^1 2x^3\,dx=\boxed{\frac{1}{2}}$

15. Lottery A n-possible #s; Lottery B n-possible #'s; Payoff A = Payoff B

$E(A)=\sum_{n=1}^{n}\frac{1}{n!}x$; $E(B)=\sum_{1}^{n}\frac{1}{n!}x$; $E(A+B)=\frac{E(A)+E(B)}{2}=E(A)=E(B)$ No difference!

$Var(X)=E(X^2)-E(X)^2=E[(X-E(X))^2]$

$=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\boxed{5/9}$

16. $Var(+X)=Var(-X)$; $f_X(5+s)=f_X(5-s)$; $Y=X-5$; $f_Y(y)=f_X(y+5)$; $f_Y(y)=f_Y(-y)$

$Y=X-5$; $E(Y)=E(X)-E(5)$; $E(X)=E(5)$

17. k^{th}-order Statistic:

$\frac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k}$; $0\le x\le 1$; $E(X)=\int_0^1\frac{n!}{(k-1)!(n-k)!}x^{k}(1-x)^{n-k}\,dx=\frac{n!}{(k-1)!(n-k)!}\frac{k!(n-k)!}{(n+1)!}$

$=\frac{k!}{(k-1)!(n+1)}=\boxed{\frac{k}{(n+1)}}$

$E(X^2)=\frac{n!}{(k-1)!(n-k)!}\int_0^1 x^{k+1}(1-x)^{n-k}\,dx=\frac{n!}{(k-1)!(n-k)!}\frac{(k+1)!(n-k)!}{(n+2)!}=\frac{k(k+1)}{(n+1)(n+2)}$

$=k(k+1)$; $Var(X)=\frac{k(k+1)}{(n+1)(n+2)}-\frac{k^2}{(n+1)^2}=\frac{k(k+1)(n+1)-k^2(n+2)}{(n+1)^2(n+2)}$

$=\frac{(k^2+k)(n+1)-k^2n+k^2(2)}{(n+1)^2(n+2)}$

$=\frac{k^2n+k^2+kn+k-k^2n-k^2 2}{(n+1)^2(n+2)}$

$=\boxed{\frac{k(n+1-k)}{(n+1)^2(n+2)}}$

18. $U_1,\ldots,U_n$; $E(U_{(n)}-U_{(1)})$; $E(U_{(n)}-U_{(1)})=\sum_{i=1}^{n}(U_{(n)}-U_{(1)})f(u)$

19. $E(U_{(k)}-U_{(k-1)})=\sum_{i=1}^{n}[U_{(k)}-U_{(k-1)}]f(u)$

20. $E[1/(X+1)]$; $X=\frac{\lambda^k}{k!}e^{-\lambda x}$; $E\left[\frac{1}{(x+1)}\right]=\int_0^{\infty}\frac{\lambda^k}{k!}\frac{1}{(x+1)}e^{-\lambda x}\,dx$

$=\frac{\lambda^k}{k!}\int_0^{\infty}\frac{e^{-\lambda x}}{(1+x)}\,dx$; $u=1+x$, $du=dx$, $x=(u-1)$; $\frac{\lambda^k}{k!}\int_0^{\infty}\frac{e^{-\lambda(u-1)}\,du}{u}=\frac{\lambda^k}{k!}e^{\lambda}\int_0^{\infty}\frac{e^{-\lambda u}}{u}\,du=\frac{\lambda^k}{k!}e^{\lambda}E_1(\lambda)$ "Exponential Integral"

21. Expected Area $(X^2)=\int_0^1 x^2\,dx=\boxed{\frac{1}{3}}$

22. Expected Area $(X^2)=\int_0^1 x^2\,dx=\boxed{\frac{1}{3}}$

23. Expected Area $(X^2)=\lambda\int_0^{\infty}x^2e^{-\lambda x}\,dx$

$\left(\frac{1}{2}\right)^2+1^2$

$E(X)E(Y)=\lambda\int_0^{\infty}xe^{-\lambda x}\,dx\cdot\lambda\int_0^{\infty}ye^{-\lambda y}\,dy=\boxed{1/\lambda^2}$

24. Prove Theorem A of Section 4.1.2 ; $Y = a_0 + \sum_{i=1}^{n} b_i X_i$; $E(Y) = a + \sum_{1}^{n} b_i E(X)$

Proof: $E(Y) = \iint (a + b_1 x_1 + b_2 x_2) f(x_1, x_2)\,dx_1 dx_2 = a \iint f(x_1, x_2)\,dx_1 dx_2 + b_1 \iint x_1 f(x_1, x_2)\,dx_1 dx_2$
$+ b_2 \iint x_2 f(x_1, x_2)\,dx_1 dx_2$

$\iint x_1 f(x_1, x_2)\,dx_1 dx_2 = \int x_1 \left[\int f(x_1, x_2)\,dx_2\right] dx_1 = \int x_1 f(x_1)\,dx_1 = E(X_1)$

$E(Y) = a + b_1 E(X_1) + b_2 E(X_2)$; $\iint |a + b_1 x_1 + b_2 x_2| f(x_1, x_2)\,dx_1 dx_2 < \infty$

$|a + b_1 x_1 + b_2 x_2| \leq |a| + |b_1||x_1| + |b_2||x_2|$

25. Gamma Distribution:

$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}$ Find $R^2 = X_1^2 + X_2^2$; $E(R^2) = \int (x_1^2 + x_2^2) \Gamma(\alpha)\,dx_1 dx_2 = \int_0^\infty x_1^2 \frac{\lambda^\alpha}{\Gamma(\alpha)} x_1^{\alpha-1} e^{-\lambda x_1} dx + \int_0^\infty x_2^2 \frac{\lambda^\alpha}{\Gamma(\alpha)} x_2^{\alpha-1} e^{-\lambda x} dx$

$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \left[\left(\int_0^\infty x_1^{(\alpha+2)-1} e^{-\lambda x} dx \right) + \left(\int_0^\infty x_2^{(\alpha+2)-1} e^{-\lambda x} dx \right) \right]$

$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+k}} = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\lambda^k} = \frac{(\alpha+2+1)\Gamma(\alpha+2-1)}{\Gamma(\alpha)\lambda^k} + \cdots$

$= \frac{(\alpha+2-1)\cdots\alpha}{\lambda^k} = \frac{(\alpha+1)\alpha}{\lambda^k}$

$\boxed{E(R^2) = \frac{2(\alpha+1)\alpha}{\lambda^k}}$

26. $\frac{\text{Expectation Long Piece}}{\text{Expectation (Shorter Piece)}} = \frac{E(L)}{E(S)} = \frac{\sum L \cdot f(x)}{\sum S \cdot f(x)}$

27. n-men → pile $E(\#\text{ of matches}) = \sum \frac{x}{n} = \frac{n}{n} = 1$

28. $X_i = \begin{cases} 1 : \text{Aircraft } i \text{ is hit} \\ 0 : \text{Otherwise} \end{cases}$; $Z_{ij} = \begin{cases} 1 & \text{if gunner } j \text{ hits } i \\ 0 & \text{otherwise} \end{cases}$

$P(Z_{ij} = 1 | B_i) = p$; $P(Z_{ij} = 1 | \bar{B}_i) = 0$; $P(B) = \frac{1}{n}$

$P(\bar{B}) = 1 - P(B) = 1 - \frac{1}{n}$ "Every gunner selects i"

Addition Rule:

$P(Z_{ij} = 1) = P(Z_{ij} = 1 | B_i) P(B_i) + P(Z_{ij} = 1 | \bar{B}_i) P(\bar{B})$

$= p\left(\frac{1}{n}\right) + 0\left(\frac{1}{n}\right) = \frac{p}{n}$

Since the total probability is equal to 1, this implies:

$P(Z_{ij} = 0) = 1 - P(Z_{ij} = 1) = 1 - \frac{p}{n}$

$E(X) = \sum x_i P(x) = 0(P(X=0)) + 1 \cdot P(X=1)$

$= 1 - P(X=1) = \ldots$

$= 1 - \{P(Z_1 = 0) P(Z_{12} = 0) \cdots P(Z_{1m} = 0)\}$

$\boxed{= 1 - \left(1 - \frac{p}{n}\right)^m}$

29. Prove Corollary 4.1.1

$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

$E[g(X) \cdot h(Y)] = \iint g(x) \cdot h(y)\,dx\,dy = \boxed{\int g(x)\,dx \int h(Y)\,dY}$

30. Coupon Collection: n-distinct types of coupons. P(coupon) are equivalent.

$X_1 = 1, X_2 = 2, X_3 = 3 \cdots X_n = n$; $P_r(\text{success}) = (n-r+1)/n$

$E(X_r) = n/(n-r+1)$

$E(X) = \sum_{r=1}^{n} E(X_r) = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{1} = n \sum_{r=1}^{n} \frac{1}{r}$

$\boxed{r = 2r \sum \frac{1}{r} = r}$

31. $X[1, 2]$

$E[1/X] = \int_1^2 x\left(\frac{1}{x}\right) dx$

$= \int_1^2 dx = 1$

$\frac{1}{E(X)} = \frac{1}{\int_1^2 x^2 dx} = \frac{3}{8-1} = \frac{3}{7}$; They are not equal.

32. Gamma Distribution.

$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}$

$E\left(\frac{1}{X}\right) = \int_0^\infty \left(\frac{1}{x}\right) \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} dx = \frac{\lambda^a}{\Gamma(a)} \int_0^\infty x^{(a-1)-1} e^{-\lambda x} dx$

$= \frac{\lambda^a}{\Gamma(a)} \frac{\Gamma(a-1)}{\lambda^{a-1}} = \frac{\lambda\,\Gamma(a-1)}{\Gamma(a)} = \boxed{\frac{\lambda}{a}}$

33. Prove Chebyshev's Inequality:

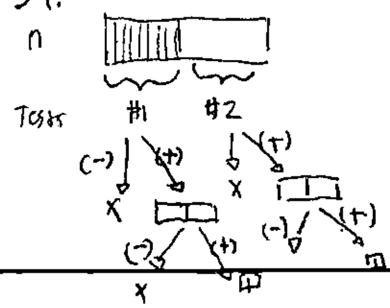
X, μ, σ^2. For $t > 0$,

$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$

$E(y) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (X-\mu)^2 f(x) \geq \int_{-\infty}^{\mu-\varepsilon} (x-\mu)^2 f(x)\,dx + \int_{\mu+\varepsilon}^{\infty} (x-\mu)^2 f(x)\,dx = \int_{-\infty}^{\mu-\varepsilon} \varepsilon^2 f(x)\,dx + \int_{\mu-\varepsilon}^{\infty} \varepsilon^2 f(x)\,dx = \boxed{\varepsilon^2 P(|X-\mu| \geq \varepsilon)}$

34\. n

Tests #1 #2



n = Group Samples
m = Groups
k = samples
$n = mk$

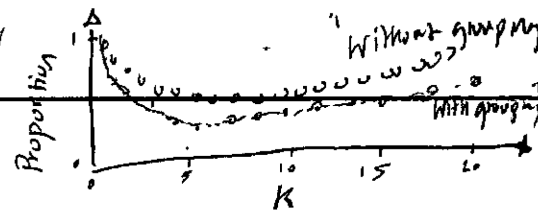
$E(N) = \sum_{1}^{m} \hat{E}(X) = \sum^{m} p^k + (k+1)(1-p^k) = \sum^{m} k - kp^k + 1$

$\boxed{= m(k+1) - mkp^k = n\left(1 + \frac{1}{k} - p^k\right)}$

"Average number of tests performed"

"With no grouping"

$\boxed{E(X) = k + 1 - kp^k = k\left(1 + \frac{1}{k} - p^k\right)}$



35\. Mean of Negative Binomial Random Variable:

$$E(K) = \sum_{k=1}^{\infty} k \binom{k-1}{r-1} p^r (1-p)^{k-r} = \sum \frac{k(k-1)!}{(r-1)!(r-k)!} p^r (1-p)^{k-r}$$

$$= \frac{k\,T(k)}{T(r)\,T(k-\cancel{k}+1)} \cdot \frac{T(r+1)\,T(k-\cancel{r}+1)}{T(k+2)} = \frac{k\,T(k)\; r\,T(r)}{T(r)(k+1)\,k\,T(k)} = \boxed{\frac{r}{k+1}}$$

36\. $X[0,1], Y = \sqrt{X}$; $E(Y)$ a) $E(Y) = P(\sqrt{X}) = \sum_{0}^{1} \sqrt{X}\, f(x) \neq \boxed{\sqrt{x}\cdot p^x + \sqrt{x}\,p(1-p)^{1-x}} = \sqrt{\;\;} \neq 1$

b) $F(n\,k(\cdot|x)$ $\int_0^1 \sqrt{x}\,dx$ $\frac{2}{3}(1)^{3/2}$

37\. Example C "Section 4.1.2-

$E(N) = n\left(1 + \frac{1}{k} - p^k\right)$; $E(X) = n!X$ $E(X) = E(N)$

$n = n\left(1 + \frac{1}{k} - p^k\right)$

$1 = 1 + \frac{1}{k} - p^k$

$\frac{1}{k} = p^k$

$\boxed{p = \left(\frac{1}{k}\right)^{1/k}}$

38\. $E(Y) = \sum_{k=0}^{n} \binom{n}{k} k p^k (1-p)^{n-k} = np$

a) $Y = \sum_{i=1}^{n} X_i$; Length of DNA $= G$, Fragments $= N$ of length $= L$.
$G \sim > 100{,}000$; $L \sim > 500$

Probability of left end is $1, 2 \ldots G-L+1$.

What is the probability a particular location $x \in \{L, L+1 \ldots G\}$

What is the probability a particular location $\in \{1, 2 \ldots L-1\}$

How many fragments are expected to cover a particular location

What is the chance of $\neq$ covering the left end of L locations $\{x-L+1, x-L, \ldots x\}$

$p = \frac{L}{G-L+1} \approx \frac{L}{G}$; The binomial probability formula, $p(W>0) = 1 - P(W=0) = 1 - \left(1 - \frac{L}{G}\right)^N$

$\lambda = Np = NL/G$

a. Probability that a fragment is the leftmost member of a contig $\frac{L}{G-L+1}$

b. Expected number of fragments left of contig: $E(K) = \sum_{N=0}^{N} K(P(W=0)) = \left[1 - \left(1 - \frac{L}{G}\right)^N\right] L$

c. Expected number of contigs: $\boxed{E(X) = L\,e^{-NL/G}}$

39\. DNA Length $= 10^6$, fragment length $= 100$

a) $P(W>0) = 0.99 = 1 - \left(1 - \frac{100^2}{10^6}\right)^N$; $(1 - 10^{-4})^N = 0.01$; $N \geq \frac{10^{-2}}{\log(0.9999)} = 4.60 \times 10^4$ fragments.

b) The expected misses: $E(I) = e^{-4.60\times10^4 \cdot 100/10^6}$ $\boxed{= 0.01}$ 0.999

40\. Q, W, E, R, T, Y produces 1000 letters in all. $E(QQQQ) = \sum_{n=1}^{N-q+1} E(I_n) = (N-q+1)\left(\frac{1}{5}\right)^q$

$\boxed{n = 1000,\ q = 4 \quad = (997)\left(\frac{1}{5}\right)^4 = 15.95 \text{ times}}$

41\. $E(I_{TOY}) = \sum_{n=1}^{N-q+1} E(I_n) = (1000-3+1)\left(\frac{1}{5}\right)^3 = 998\left(\frac{1}{5}\right)^3$
$= 79.84$ times.

Markov Inequality:

$\frac{79.84}{1000} \approx$ Yes, the author would be surprised for the answer to occur.

42. Exponential Random Variable: $p(x) = \lambda e^{-\lambda x}$; $P(|X-E(X)|>k\sigma)$ for $k=2,3,4\ldots$ compare results to the bounds from Chebyshev's inequality

Chebyshev Inequality: $P(|X-\mu|>t) \le \frac{\sigma^2}{t}$

$P(|X-E(X)|>k\sigma) \le \frac{\sigma^2}{k\sigma^2} = \frac{1}{k}$ for $k=2,3,4\ldots$

43. Show Var(X−Y)

$Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$

$= E([X-Y]^2) - E(X-Y)^2 = E(X^2) - 2E(X \cdot Y) + E(Y^2) - E(X)^2 + E(Y)^2$

$\boxed{= Var(X) + Var(Y) - 2Cov(X,Y)}$

covariance: $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_X)]$ provided the expectation exists

44. X&Y have equal variance. Find $Cov(X+Y, X-Y)$

$Cov(X+Y, X-Y) = E[(X+Y-E(X+Y))(X-Y-E(X-Y))] = E[(X-E(X)+Y-E(Y))(X-E(X)-Y+E(Y))]$

45. Find the covariance of N_j and N_i where $N_1, N_2, \ldots N_r$ are multinomial random variables.

$= E[(X-E(X))^2] - E[(X-E(X))(Y+E(Y))] + E[(Y-E(Y))(X-E(X))] + E[[Y-E(Y)]^2]$

$= E[(X-E(X))^2] - 2E[(X-E(X))(Y+E(Y))] - E[(X-E(Y))^2]$

Multinomial Random Variable: $p(n_1, \ldots, n_r) = \binom{n}{n_1 \ldots n_r} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$

$Cov(N_i, N_j) = E[(N_i - E(N_i))(N_j - E(N_j))]$

$= E(N_i N_j) - E(N_i)E(N_j)$

$= Pr(N_i = N_j) - Pr(N_i = 1) Pr(N_j = 1) = 0 - p_i p_j p_k = \boxed{p_i p_k p_r}$

46. U & V; μ and σ^2; $Z = \alpha U + V\sqrt{1-\alpha^2}$. Find $E(Z)$ and ρ_{UZ}.

$E(Z) = E(\alpha U + V\sqrt{1-\alpha^2}) = E(\alpha U) + E(V\sqrt{1-\alpha^2}) = \alpha E(U) + \sqrt{1-\alpha^2}E(V) = \boxed{(\alpha + \sqrt{1-\alpha^2})\mu}$

Correlation coefficient: If X and Y are jointly distributed random variables and the variances and covariances of both X and Y are nonzero, then the correlation of X and Y, denoted by ρ,

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E(VU) - E(V)E(V)}{\sqrt{Var(V)Var(V)}} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u \cdot v \, e^{-\frac{(u-\mu)^2}{2\sigma^2} - \frac{(v-\mu)^2}{2\sigma^2}} du\,dv - \left(\frac{1}{2}\right)^2}{\sqrt{\sigma^2 \cdot \sigma^2}}$$

$$= \frac{\left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2}{\sigma^2} = \phi$$

47. $Z = Y - X$; Find

$-Cov(X,Z) = Cov(X, Y-X) =$

$= E[(X-E(X))(Z-E(Z))] = E[XZ] - E[X]E(Z) = E(X(Y-Z)) - E(X)E(Y-Z)$

$\boxed{= E(XY) - E(XZ) - E(X)[E(Y) - E(Z)] =}$

$corr(X,Z) = \rho_{XZ} = \frac{Cov(X,Z)}{\sqrt{Var(X)Var(Z)}} = \boxed{\frac{E(XY) - E(XZ) - E(X)[E(Y) - E(Z)]}{\sqrt{Var(X)Var(Z)}}}$

48. $U = a + bX$; $V = c + dY$ show that $|\rho_{UV}| = |\rho_{XY}|$;

$$\rho = \frac{Cov(U,V)}{\sqrt{Var(U)Var(V)}} = \frac{Cov(a+bX, c+dY)}{\sqrt{Var(a+bX)Var(c+dY)}} = \frac{E[(a+bX)(c+dY)] - E(a+bX)E(c+dY)}{\sqrt{[E[(a+bX)^2] - E(a+bX)^2][E[(c+dY)^2] - E(c+dY)^2]}}$$

$$= \frac{E(ac) + E(adY) + E(bXc) + E(bXdY) - (E(a) + E(bX))(E(c) + E(dY))}{\sqrt{[E(a^2) + 2E(abX) + E(b^2X^2) - [E(a) - E(bX)]^2][E(c^2) + 2E(cdY) + E(dY)^2 - [E(c) + E(dY)]^2]}}$$

$\boxed{= \emptyset}$

49. $E(X)=E(Y)=\mu$, but $\sigma_X \neq \sigma_Y$; $Z=\kappa X+(1-\kappa)Y$; where $0 \leq \kappa \leq 1$

a) Show $E(Z)=\mu$; $E(Z)=E(\kappa X+(1-\kappa)Y)=\kappa E(X)+(1-\kappa)E(Y)=\kappa\mu+(1-\kappa)\mu=\boxed{\mu}$

b) Find κ in terms of σ_X, σ_Y to minimize $Var(Z)$

$Var(Z)=Var(\kappa X+(1-\kappa)Y)=\kappa^2 E[X^2]-\kappa^2 E[X]^2+(1-\kappa)^2 E[Y^2]-(1-\kappa)^2 E[Y]^2=\kappa^2[Var(X)]+(1-\kappa)^2[Var(Y)]$

$\frac{d}{d\kappa}Var(Z)=0=(\kappa^2 Var(X)-(1-\kappa)^2 Var(Y))=(2\kappa\, Var(X)+2(1-\kappa)(-1)Var(Y))=0$

$\boxed{\kappa=\frac{Var(Y)}{Var(X)+Var(Y)}}$ || $Var\left(\frac{X+Y}{2}\right) \leq Var(X)$

$\frac{1}{4}[Var(X)+Var(Y)] \leq Var(X)$; $\boxed{Var(Y) \leq 3\,Var(X)}$

c) When is the average $(X+Y)/2$ better to use than X or Y alone?

When the variance of the average is less than variance of X or Y alone.

50. X_i; $i=1\ldots n$; $E(X_i)=\mu$ and $Var(X_i)=\sigma^2$ $\bar{X}=n^{-1}\sum_{i=1}^{n}X_i$; Show $E(\bar{X})=\mu$ and $Var(\bar{X})=\sigma^2/n$

$E(\bar{X})=E(n^{-1}\sum_{i=1}^{n}X_i)=E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=\mu$; $Var(\bar{X})=E(\bar{X}^2)-E(\bar{X})^2=\frac{1}{n}E(\sum X_i^2)-\frac{1}{n}E(\sum X_i)^2$

$=\sigma^2/n$

51. Example E: Section 4.3; $\mu_1=\mu_2=\mu$; $\sigma_{ij}=Cov(R_i,R_j)=0$; Portfolio $(\pi, 1-\pi)$

Expected Return: $E(R(\pi))=\pi\mu+(1-\pi)\mu=\mu$: Risk of Return: $Var(R(\pi))=\pi^2\sigma_1^2+(1-\pi)^2\sigma_2^2$

Minimizing Risk with respect to π: $\pi_{opt}=\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$; $Var(R(\frac{1}{2}))=\frac{\sigma^2}{2}$

When considering unique returns: $E(R(\pi))=\pi\mu_1+(1-\pi)\mu_2$

$Var(R(\pi))=\pi^2\sigma_1^2+2\pi(1-\pi)\rho\sigma_1\sigma_2+(1-\pi)^2\sigma_2^2$

When considering n-total investments: $E(R(\pi))=\sum\pi_i\mu_i$; $Var(R(\pi))=\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\pi_j\sigma$

Problem: n-securities (μ, σ^2)

uncorrelated. $E(R(\pi))=\sum_{i=1}^{n}\pi_i\mu$; $Var(R(\pi))=\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\pi_j\sigma$; $\boxed{Var=\frac{1}{n}\sigma^2}$; $S.D=\frac{1}{\sqrt{n}}\sigma$

$\mu=n\pi_i\mu$; $\boxed{1=n\pi_i}$ $\boxed{\frac{1}{n}=\pi_i}$

Risk of one security $=\boxed{\frac{\sigma}{1}}$

52. Two securities $(\mu_1=1, \sigma_1=0.1)$

$(\mu_2=0.8, \sigma_2=0.12)$; $\rho=-0.8$;

a) Return #1 $=\mu_1\pm0.1=0.1\pm0.1$; $\boxed{(\mu_1/\sigma_1)=10}$ greater dollar per risk.

Return #2 $=\mu_2\pm0.12=0.8\pm0.12$ $(\mu_2/\sigma_2)=6.75$

b) 50% into each stock

$E(R(\pi))=0.5\cdot1+0.8\cdot0.5=0.9$

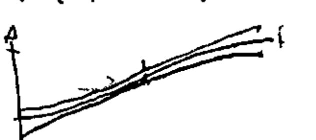
$Var(R(\pi))=0.5^2(0.1)^2+2\cdot0.5(1-0.5)(-0.8)\,0.1\cdot0.12+(1-0.5)^2(0.12)^2$

$\boxed{\sigma_R=0.04}$

c) $E(R(\pi_1=80\%, \pi_2=20\%))=0.8\cdot1+0.2\cdot0.8=\boxed{0.96}$; $Var(R(\pi_1=80\%, \pi_2=20\%))$

$=0.8^2(0.1)^2+2\cdot0.8(0.2)(-0.8)\cdot0.1\cdot0.12+(0.2)^2(0.12)^2$

$\boxed{\pm0.06}$

d) 

e) 

54. X, Y, Z with $\sigma_X^2; \sigma_Y^2; \sigma_Z^2$

Let $U=Z+X$; $V=Z+Y$

$Cov(U,V)=E(UV)-E(U)E(V)=E[(Z+X)(Z+Y)]-E[Z+X]E[Z+Y]$

$=E[Z^2]+E[XZ]+E[ZY]+E[XY]-E[Z^2]-E[Z]-E[XZ]-E[X]E[Y]$

$=0$

$Corr(U,V)=\rho_{UV}=\boxed{0}$

53. $Cov(X,Y) \leq \sqrt{Var(X)Var(Y)}$;

$Cov(X,Y)=E(XY)-E(X)E(Y) \leq \sqrt{[E(X^2)-E(X)^2][E(Y^2)-E(Y)^2]}$

$\leq\sqrt{E(X^2)E(Y^2)-E(X^2)E(Y)^2-E(X)^2E(Y^2)-E(X)^2E(Y)^2}$

55. $T=\sum_{k=1}^{n} kX_k$; X_k are independent random variables with μ, σ^2. Find $E(T)$ and $Var(T)$.

$E(T)=E\left(\sum_1^n kX_k\right)=\frac{n(n+1)}{2}\mu$; $Var(T)=Var\left(\sum_1^n kX_k\right)=E\left[\left(\sum_1^n kX_k\right)^2\right]-E\left[\sum kX_k\right]^2=\frac{n(n+1)(2n+1)\sigma^2}{6}$

56. $S=\sum_{k=1}^{n} X_k$; $Cov(S,T)=E(ST)-E(S)E(T)=E\left(\sum kX_k \sum X_k\right)^2-E\left(\sum kX_k\right)E\left(\sum X_k\right)$

$=\frac{n(n+1)(2n+1)}{6}\mu^2-\frac{n(n+1)}{2}\mu\cdot\mu=\left[\frac{n(n+1)(2n+1)}{6}-\frac{n(n+1)}{2}\right]\mu^2$

$Corr(S,T)=\rho_{ST}=\frac{Cov(S,T)}{\sqrt{Var(S)Var(T)}}=\frac{\left[\frac{n(n+1)(2n+1)}{6}-\frac{n(n+1)}{2}\right]\mu^2}{\sqrt{\sigma^2\cdot\frac{n(n+1)(2n+1)}{6}\sigma^2}}=\boxed{\frac{\left[\frac{n(n+1)(2n+1)}{6}-\frac{n(n+1)}{2}\right]\mu^2}{\sqrt{\frac{n(n+1)(2n+1)}{6}}\sigma^2}}$

57. $Var(XY)=E\left[(XY-E(XY))^2\right]$

$=E\left(X^2Y^2-2XYE(XY)+E(XY)^2\right)=E(X^2Y^2)-2E(X)E(Y)E(XY)+E(XY)^2$

$=E(X^2Y^2)-2E(XY)^2+E(XY)^2=E(X^2Y^2)-E(XY)^2$

$=E(X^2)E(Y^2)-\mu_x^2\mu_y^2=\left[Var(X)+E(X)^2\right]\left[Var(Y)+E(Y)^2\right]-\mu_x^2\mu_y^2$

$=Var(X)Var(Y)+Var(X)E(Y)^2+E(X)^2Var(Y)+E(X)^2E(Y)^2-\mu_x^2\mu_y^2$

$\boxed{=\sigma_x^2\sigma_y^2+\sigma_x^2\mu_y^2+\mu_x^2\sigma_y^2}+\cancel{\mu_x^2\mu_y^2-\mu_x^2\mu_y^2}$

58. $X_1=f(x)+\epsilon_1$; $X_2=f(x+h)+\epsilon_2$; $\epsilon_1 \,\&\, \epsilon_2(\mu=0,\sigma^2)$; $\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$; $Z=\frac{X_2-X_1}{h}$

a) Find $E(Z)=E\left(\frac{X_2-X_1}{h}\right)=E\left(\frac{f(x+h)+\epsilon_2-f(x)-\epsilon_1}{h}\right)=\frac{1}{h}\left[E(f(x+h))+E(\epsilon_2)-E(f(x))-E(\epsilon_1)\right]$

$=\frac{f(x+h)-f(x)}{h}$ $\quad=\frac{1}{h^2}\left[Var(f(x+h))+Var(\epsilon_2)-Var(f(x))-Var(\epsilon_1)\right]$ Mean Squared Error

Find $Var(Z_1)=Var\left(\frac{f(x+h)+\epsilon_2-f(x)-\epsilon_1}{h}\right)=\frac{(-1)^2}{h^2}\sigma^2+\frac{\sigma^2}{h^2}=\boxed{\frac{2\sigma^2}{h^2}}$

$E\left[(X-X_0)^2\right]$ Mean Squared Error

$=Var\left[(X-X_0)\right]+E\left[(X-X_0)\right]^2$

$\boxed{=\sigma^2+\beta^2}$

In the limit of $E(Z)=\lim_{h\to 0}E(Z)=f'(x)$; $\lim_{h\to 0}Var(Z)=\boxed{\lim_{h\to 0}\frac{2\sigma^2}{h^2}=0}$

b) Mean Squared Error of Z:

$MSE(Z)=E\left[(Z-Z_0)^2\right]=Var(Z-Z_0)+E\left[(Z-Z_0)\right]^2=\frac{2\sigma^2}{h^2}+\frac{f(x+h)-f(x)}{h}$

$\lim_{h\to 0}MSE(Z)=f'(x)$ | $X_3=f(x+h+k)+\epsilon_3$; $E(\epsilon_1)=E(\epsilon_2)=E(\epsilon_3)=0$; $Var(\epsilon_1)=Var(\epsilon_2)=Var(\epsilon_3)=\sigma^2$

c) $X_1=f(x)+\epsilon_1$; $X_2=f(x+h)+\epsilon_2$; $Z_1=\frac{1}{h}\left[X_2-X_1\right]$; $Z_2=\frac{1}{h}\left[X_3-X_2\right]$; $Z_3=\frac{1}{h}\left[Z_2-Z_1\right]=\frac{1}{h}\left(\frac{X_3-X_2}{k}-\frac{X_2-X_1}{h}\right)$

$Z_3=\frac{1}{h^2}X_1-\left(\frac{1}{hk}+\frac{1}{h^2}\right)X_2+\frac{1}{hk}X_3$; $\boxed{E(Z_3)=\frac{1}{h^2}f(x)-\left(\frac{1}{hk}+\frac{1}{h^2}\right)f(x+h)+\frac{1}{hk}f(x+h+k)}$

$\boxed{Var(Z_3)=2\sigma^2\left(\frac{1}{h^4}+\frac{1}{h^2k^2}+\frac{1}{h^3k}\right)}$

61. $f(x,y)=2$, $0\le x\le y$; a) $Cov(X,Y)$, $Corr(X,Y)$

59. (X,Y) is a random point on a disk.

Show the $Cov(X,Y)=0$: $E(XY)-E(X)E(Y)$

$=\frac{4}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x\sqrt{1-x^2}\,y\sqrt{1-y^2}\,dx\,dy-\frac{2}{\pi}\int_{-\infty}^{\infty}x\sqrt{1-x^2}\int_{-\infty}^{\infty}y\sqrt{1-y^2}=0$

60. Y is symmetric about zero. $X=SY$

$S=\pm 1$; $P(S=1)=P(S=-1)=\frac{1}{2}$; Show $Cov(X,Y)=0$

$f(x,y)=\begin{cases}\frac{1}{\pi r} & x^2+y^2\le r\\ 0 & \text{otherwise}\end{cases}$

$Cov(X,Y)=E(XY)-E(X)E(Y)=E(SY^2)-E(SY)E(Y)=\boxed{SE(Y^2)-SE(Y^2)=0}$

$Cov(X,Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)f(x,y)\,dx\,dy$

$=2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)\,dx\,dy$

$E(X)=E(SY)=SE(Y)$; $\frac{E(X)}{E(Y)}=S$ | $=2\int_{-\infty}^{\infty}\left(\frac{x^2}{2}\right)(y-\mu_y)\,dy=\infty$

X & Y from $[0,1]$; $f(x,y)=2$; $0\le x\le y\le 1$

$E(x)=\int_0^1 x f(x)\,dx=\int_0^1 x\left[\int_x^1 f(x,y)\,dy\right]dx=\int_0^1 x\left[\int_x^1 2\,dy\right]dx$

$=\int_0^1 x\,[2-2x]\,dx=2\int_0^1 (x-x^2)\,dx=2\left[\frac{y^2}{2}-\frac{y^3}{3}\right]=\boxed{\frac{1}{3}}$

$E(y)=\int_0^1 y f(y)\,dy=\int_0^1 y\left[\int_0^y f(x,y)\,dx\right]dy=\int_0^1 y\,[2y]\,dy=\int_0^1 2y^2\,dy$

$\boxed{=\frac{2}{3}}$

a) $Cov(X,Y)=E(XY)-E(X)E(Y)$

$=\frac{1}{4}-\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)=\frac{1}{4}-\frac{2}{9}$

$=\frac{9}{36}-\frac{8}{36}=\boxed{\frac{1}{36}}$

$E(xy)=\int_0^1\int_0^y xy\,f(x,y)\,dx\,dy=2\int_0^1 \frac{y^3}{2}\,dy=\frac{y^4}{4}\Big|_0^1=\boxed{\frac{1}{4}}$

$Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$

$=\frac{(1/36)}{\sqrt{(E[X^2]-E[X]^2)(E[Y^2]-E[Y]^2)}}=\frac{1/36}{\sqrt{(2/12-1/9)(\frac{1}{2}-4/9)}}=\frac{1/36}{1/18}=\boxed{\frac{1}{2}}$

$x^2[2-2x]$

$2[x^2-x^3]$

$2\left[\frac{x^3}{3}-\frac{x^4}{4}\right]$

$\frac{2}{3}-\frac{2}{4}=\frac{8}{12}-\frac{6}{12}=2$

b. Find $E(X|Y=y)$ and $E(Y|X=x)$

Conditional Expectations:

if $P_{Y|X}(y|x)$, then $E(Y|X=x)=\sum_y y\,P_{Y|X}(y|x)$

$=\int y\,f_{Y|X}(y|x)\,dy$

More Generally,

$E[h(Y)|X=x]=\int h(y)\,f_{Y|X}(y|x)\,dy$

$E(Y|X=x)=\sum_y y\,\frac{f(x,y)}{f(x)}\Rightarrow\int_x^1 y\left(\frac{2}{2-2x}\right)dy\ \ 2\ \frac{y^3}{4}$

$=\int_x^1 y\cdot 2\left(\frac{1}{1-x}\right)dy=\frac{1}{2}[1-x^2]\left(\frac{1}{1-x}\right)\boxed{\frac{(x+1)}{2}}$

$E(X|Y=y)=\sum_x x\cdot P_{X|Y}(x|y)\Rightarrow\int_0^y x\,\frac{f(x,y)}{f(y)}\,dx=\int_0^y x\,\frac{2}{2y}\,dx$

$=\frac{y^2}{2y}=\boxed{\frac{y}{2}}$

c) Find $E(X|Y)$ and $E(Y|X)$

$=y/2$ $\qquad=(x+1)/2$

$f_{W_1}(W_1)=2f_Y(2W_1)$ $\qquad f_{W_2}(W_2)=4(2W_2-1)$

$=2(2(2W_1))$ $\qquad\boxed{=8(1-W_2)}$

$\boxed{=8W_1}$

d) $\hat{Y}=a+bX$; $\min(E((Y-\hat{Y})^2)$ $\boxed{\text{Predictor}}$

$E(\hat{Y})=a+bE(\hat{X})$; $\mu_y=a+b\mu_x$

$a=\mu_y-b\mu_x=\frac{2}{3}-\rho_{XY}\left(\frac{1}{3}\right)=\frac{2}{3}-\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)=\frac{2}{3}-\frac{1}{6}$

$=\frac{12}{18}-\frac{3}{18}=\boxed{\frac{9}{18}}$

Mean Squared Error : $E\left(Y-\frac{1}{2}-\frac{1}{2}x\right)^2=\sigma_y^2(1-\rho^2)$

$=\frac{1}{18}\left(1-\frac{1}{4}\right)=\boxed{1/24}$

e) Predictor of Y in terms of X

MSE = $\boxed{\text{Mean Squared Error}}$ $=E(Y-\hat{Y})^2=E(Y-E(Y|X))^2$

$=E(Y^2)-E(E^2(Y|X))=\frac{1}{2}-E\left(\frac{(x+1)^2}{2}\right)=\frac{1}{2}-\int_0^1\frac{(x+1)^2}{4}(1-x)\,dx$

$\boxed{=1/24}$

$\boxed{\text{Mean Squared Predictor Error}}$

$E(Y|X=x)=\frac{1}{2}-E\left(\frac{(x+1)^2}{4}\right)$

$=\frac{1}{2}-\int_0^1\frac{(x+1)^2}{4}(x-1)\,dx$

$=\boxed{1/24}$

62.

X & Y joint random variables with correlation ρ_{XY},

Define the Standardized random variables $\tilde{X}$ and $\tilde{Y}$

$\tilde{X}=(X-E(X))/\sqrt{Var(X)}$

$\tilde{Y}=(Y-E(Y))/\sqrt{Var(Y)}$

Show that $Cov(\tilde{X},\tilde{Y})=\rho_{XY}$

$Cov(\tilde{X},\tilde{Y})=E((\tilde{X}-\bar{X})(\tilde{Y}-\bar{Y}))$

$=E(\tilde{X}\tilde{Y})-E(\bar{X}\tilde{Y})-E(\bar{Y}\tilde{X})+E(\bar{X})E(\bar{Y})$

$=E(\tilde{X}\tilde{Y})-E(\tilde{X})E(\tilde{Y})$

$\boxed{=E\left(\frac{[X-E(X)]}{\sqrt{Var(X)}}\frac{[Y-E(Y)]}{\sqrt{Var(Y)}}\right)}-\frac{\cancel{E(X)-E(E(X))}}{\sqrt{Var(X)}}\frac{\cancel{[E(Y)-E(E(Y))]}}{\sqrt{Var(Y)}}\boxed{=\rho_{XY}}$

63.

$f(x,y)=\frac{6}{7}(x+y)^2$

$0\le x\le 1$; $0\le y\le 1$

a) $Cov(X,Y)=E[(X-\mu_X)(Y-\mu_Y)]=E[XY]-E[X]E[Y]$

$=\int_0^1\int_0^1 xy\,f(x,y)\,dx\,dy-\int_0^1 x f(x)\,dx\int_0^1 y f(y)\,dy$

$=\int_0^1\int_0^1 xy\,\frac{6}{7}(x+y)^2\,dx\,dy-\int_0^1 x\left[\int_0^1\frac{6}{7}(x+y)^2\,dy\right]dx\int_0^1 y\left[\int_0^1\frac{6}{7}(x+y)^2\,dx\right]dy$

$= \int_0^1\int_0^1 xy\,\frac{6}{7}(x^2+2xy+y^2)\,dx\,dy - \int_0^1 x\,\frac{6}{7}(x^2+x+\frac{1}{3})\,dx \int_0^1 y\,\frac{6}{7}(\frac{1}{3}+y+y^2)\,dy$

$= \int_0^1 y\,\frac{6}{7}(\frac{1}{3}+y+y^2)\,dy - \frac{6}{7}\left[\frac{1}{4}+\frac{1}{3}+\frac{1}{6}\right]\frac{6}{7}\left[\frac{1}{6}+\frac{1}{3}+\frac{1}{4}\right] = \frac{6}{7}\left[\frac{1}{6}+\frac{1}{3}+\frac{1}{4}\right] - \frac{36}{49}\left(\frac{3}{4}\right) = \boxed{-0.0085}$

$\doteq 0.0085$

$\mathrm{Corr}(X,Y) = \rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{-19/34}{\sqrt{(E[X^2]-E[X]^2)(E[Y^2]-E[Y]^2)}} = \frac{}{\sqrt{\left(\int_0^1 \frac{6}{7}[x^4+x^3+\frac{x^2}{3}]dx - \frac{81}{196}\right)\left(\int_0^1 \frac{6}{7}[\frac{y^2}{3}+y^3+y^4]dy - \frac{81}{196}\right)}}$

$= \frac{-19/34}{\sqrt{\left(\frac{6}{7}\left[\frac{1}{5}+\frac{1}{4}+\frac{1}{9}\right]\right)^2}} = \boxed{0.1\,2\,516}$; $\boxed{\text{Miscalculated}}$

b.

Find $E(Y|X=x)$ for $0 \le x \le 1$

Conditional Expectation:

$E(Y|X=x) = \sum x\,p_{Y|X}(Y|X)$

$= \int_0^1 y\,\frac{f(x,y)}{f(x)}\,dx$

$= \int_0^1 \frac{y}{(x^2+x+1/3)}(x^2+2xy+y^2)\,dy$

$\boxed{= \frac{6x^2+8x+3}{4(3x^2+3x+1)}}$

64.

y \ x	1	2	3	4	f(y)
1	0.10	0.05	0.02	0.02	0.19
2	0.05	0.20	0.05	0.02	0.32
3	0.02	0.05	0.20	0.04	0.31
4	0.02	0.02	0.04	0.10	0.18
E(x)	0.19	0.32	0.31	0.18	

a. $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$

$= \sum xy\,f(x,y) - \sum x f(x) \sum y f(y)$

$= 1^2 \times 0.1 + 2\cdot1\times0.05 + 3\times1\times0.02 + 4\times1\times0.02$
$+ 2\cdot1\times0.05 + 4^2\cdot0.2 + 6\times0.05 + 8\times0.02$
$+ 3\cdot1\times0.02 + 6\times0.05 + 5^2\cdot0.20 + 12\times0.04$
$+ 4\cdot1\times0.02 + 8\times0.02 + 3\cdot4\times0.04 + 4^2\times0.10$

$= 6.66$

$E(X) = 1\cdot0.19 + 2\cdot0.32 + 3\times0.31 + 4\times0.18$
$= 2.48 = E(Y)$

$E(X^2) = 1^2\times0.19 + 2^2\cdot0.32 + 3^2\times0.31 + 4^2\times0.18$
$= 7.14 = E(Y^2)$

$\mathrm{Var}(X) = E(X^2) - E(X)^2$
$= 7.14 - 2.48^2 = 0.9896$

$\mathrm{Var}(Y) = 0.9896$

$\boxed{\mathrm{Cov} = 0.5096\ ;\ \mathrm{Corr} = 0.514955}$

b)

Find $E(Y|X=x)$ for $x = 1,2,3,4$;

$E(Y|X=1) = \sum y f(y|x) = \sum y\,\frac{f(x,y)}{f(x)} = \frac{1\times0.1}{0.19} + \frac{2\times0.05}{0.19} + \frac{3\times0.02}{0.19} + \frac{4\cdot0.02}{0.19}$ $\boxed{= 1.78}$

$E(Y|X=2) = \sum y f(y|2) = \sum y\,\frac{f(x,y)}{f(y)} = \frac{1\cdot0.05 + 2\times0.2 + 3\cdot0.05 + 4\cdot0.02}{0.32}$

$E(Y|X=3) = \sum y f(y|3) = \frac{1\cdot0.02 + 2\cdot0.05 + 3\cdot0.2 + 4\cdot0.04}{0.31}$ $\boxed{= 2.13}$

$E(Y|X=4) = \frac{1\times0.02 + 2\cdot0.02 + 3\times0.04 + 4\cdot0.1}{0.18} \boxed{3.22}$ $\boxed{= 3.87}$

$E(Y|x) = \{1.78, 2.13, 2.87, 3.22\}$

65. Random Sums:

$T = \sum_{i=1}^{N} X_i$

$E(T) = E[E(T|N)] = E[N E(X)] = E(N)E(X) \leftarrow \boxed{\text{Independence}}$

66.

$\boxed{\text{Fast}}$ 1 min, $\boxed{\text{Slow}}$ 3 min ; $E(T) = \sum E(T|E_i)p(E_i) = 1\text{min}\left(\frac{2}{3}\right) + 3\text{min}\left(\frac{1}{3}\right) \boxed{\frac{5}{3}\text{min}}$

$p(F) = 2/3 \quad p(S) = 1/3$

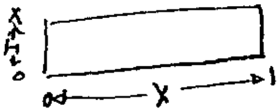
$E(XH) = E[E(XH|X)] = E[XE(H|X)] = E(X)\,E(H|X)$
$= \frac{1}{2}E(X^2)$
$= \frac{1}{2}\int_0^1 x^2\,dx = \boxed{\frac{1}{6}}$

$E(2(X+H)) = E[E(2(X+H)|X)]$
$= E[2X + 2E(H|X)]$

Note:
$= E[2X] + E[2E(H|X)] = 2\left(\frac{a+b}{2}\right) + \left(\frac{a+b}{2}\right) = 1 + \frac{1}{2} = \boxed{\frac{3}{2}}$

$E(Z) = \int_a^b z f(z)\,dz = \frac{1}{b-a}\int_a^b z\,dz = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$

67.



$E_{Area} = \sum E(Y|X=x)p_X(x)$

$= \sum_y \sum_x y\,p_{Y|X}(y|x)\,p_X(x)$

68. Show $E[\mathrm{Var}(Y|X)] \le \mathrm{Var}(Y)$

$\boxed{E[E(Y^2|X)] - E[E(Y|X)]^2 \le E[Y^2] - E[Y]^2}$

69. $\mu_X = \mu_Y = 0$: Sketch $E(Y|X=x)$ and $E(X|Y=y)$ for $\rho = 0, 0.5, 0.9$

$\sigma_X = \sigma_Y = 1$;

$E(Y|X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X)$

@ $\rho = 0$ $= 0 + 0 = 0$

@ $\rho = 0.5$ $= 0 + \frac{1}{2}\left(\frac{1}{1}\right)(X-\mu_X) = \boxed{(X-\mu_X)/2} = X/2$

@ $\rho = 0.9$ $= 0 + \boxed{(X-\mu)/9} = X/9$

$E(X|Y) = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(Y-\mu_Y)$

@ $\rho = 0$ $E(X|Y) = 0$

@ $\rho = 0.5$ $E(X|Y) = Y/2$

@ $\rho = 0.9$ $E(X|Y) = Y/9$

70. Show $E(X|Y=y) = E(X)$; $E(X|Y=y) = \sum x P_{X|Y}(x|y) = \sum x \frac{f(x,y)}{P(y)} = \sum \frac{x}{P(y)} f(x)f(y) = \sum x f(x) = \boxed{E[X]}$

71. Binomial Random Variable: n-trials, Y-success for m-trials; m<n

$p(1) = p$; $p(0) = 1-p$; $p(x) = 0$

$p(x) = \begin{cases} p^x(1-p)^{1-x}, & \text{if } x=0 \text{ or } x=1 \\ 0 & \text{otherwise} \end{cases}$

Find the conditional frequency function of Y given X=x. and conditional mean.

$p(x) = \binom{n}{x} p^x q^{n-x}$; $p(y) = \binom{m}{y} p^y q^{m-y}$

$P(Y|X=x) = \frac{\binom{n}{x} p^x q^{n-x} \binom{m}{y} p^y q^{n-y}}{\binom{n}{x} p^x q^{n-x}} p(x) = \frac{\binom{n-m}{x-y} p^{x-y} q^{n-m-x+y}}{\binom{n}{x} p^x q^{n-x}} p(y)$

Conditional Mean:

$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \frac{nK}{N}$ $\boxed{\text{Hypergeometric Fn}}$

$\boxed{P(Y|X=x) = \frac{\binom{m}{y}\binom{n-m}{x-y}}{\binom{n}{x}}}$

$= m\frac{x}{n} = E[Y|X=x]$

$\boxed{\text{Distribution of a Hypergeometric Fn}}$

$\frac{\binom{x}{k}\binom{n-x}{x-k}}{\binom{n}{m}}$

73. n-tosses, N-heads, +N more tosses.

Binomial Random Variable

First N-tosses:

$E[\text{Heads}] = E[\text{Heads}|n] + E[\text{Heads}|E[\text{Heads}|n]]$

$= np + np \cdot p = \boxed{np(1+p)}$

72. μ, σ^2 $\boxed{\begin{array}{l} E[X] = \mu; E[\mu] = \mu^2; E[\mu^2] = \mu^3 \\ Var(X) = \sigma^2; Var(\sigma^2) = \sigma^4; Var(\sigma^4) = \sigma^6 \end{array}}$

74. $\leftarrow n \rightarrow$ $\leftarrow p(n) \rightarrow$ $E[X] = \sum_{i=1}^{n} X_i p(x) = \boxed{\frac{n(n+1)}{2} p(x)}$

75. $P(T) = \lambda e^{-\lambda T}$; $E[UT] = E[U]E[T] = \frac{1}{2}\cdot\boxed{\frac{1}{2}\cdot\frac{1}{\lambda}}$; $Var(U \cdot T) = Var(U)\cdot Var(T) = \boxed{\frac{1}{12}\left(\frac{1}{\lambda^2}\right)}$

76. $(X,Y): x^2+y^2 \le 1$; $y \ge 0$.

$f_{XY}(x,y) = \frac{|A|}{|R|} = \frac{1}{\pi/2} = \frac{2}{\pi}$

$f_X(x) = \int_0^{\sqrt{1-x^2}} f_{XY}(x,y)\,dy = \frac{2}{\pi}\sqrt{1-x^2}$

$f_Y(y) = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} f(x,y)\,dx = \frac{4}{\pi}\sqrt{1-y^2}$

$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{2/\pi}{\frac{2}{\pi}\sqrt{1-x^2}} = \frac{1}{\sqrt{1-x^2}}$

$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{2\sqrt{1-x^2}}$

$E_{Y|X}(y|x) = \int_0^{\sqrt{1-x^2}} y\, f_{Y|X}(y|x)\,dy = \boxed{\frac{\sqrt{1-x^2}}{2}}$

$E_{X|Y}(x|y) = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} x\, f(x|y)\,dx$

$= \varnothing$

77. $f(x,y) = e^{-y}$; $0 \le x \le y$ a) Find $Cov(X,Y) = E[XY] - E[X]E[Y] = \iint xy f(x,y)\,dx\,dy - \int x\left[\int f(x,y)\,dy\right]dx \int y\left[\int f(x,y)\,dx\right]$

$= \int_0^{\infty}\left[\frac{x^2}{2}\right]_0^y y e^{-y}\,dy - \int_0^y x\left[\int_0^{\infty} e^{-y}\,dy\right]\int_0^{\infty} y\left[\int_0^y e^{-y}\,dx\right]dy$

$\boxed{\int_0^{\infty}\frac{y^3}{2}e^{-y}\,dy + \int_0^y x e^{-x}\,dx \int y^2 e^{-y}\,dy = 3}$

b) $E(X|Y=y) = \int x f_{X|Y}(x|y)\,dx = \int_0^y x\frac{f(x|y)}{f(y)}\,dx$

$= \int_0^y x\frac{e^{-y}}{y e^{-y}}\,dx = \boxed{\frac{y}{2}}$

$E(Y|X=x) = \int y f_{Y|X}(Y|X) = \int_x^{\infty} y\frac{e^{-y}}{1}\,dy = \boxed{X+1}$

uv-vdu

$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}} = \frac{3}{\sqrt{(2-1)(6-2)}}$

$= 1/\sqrt{2}$

c) Density Function $E(X|Y)$ and $E(Y|X)$

$E[X] = E[E[X|Y]] = E[Y/2] = E[Y]/2$

$E[Y] = E[E[Y|X]] = E[X+1] = E[X]+1$

Skewness - third central moment - asymmetry of a density or frequency about a mean.

78. $M(t) = \sum e^{tx} p(x) = e^{tx}/t$ } 1^{st} Moment

$M'(t) = E[X] = e^{tx} x$

$M''(t) = E[X^2] = t e^{tx}$ } 2^{nd} Moment

$M'''(t) = e^{tx} + t^2 e^{tx}$
$= (1+t^2)e^{tx}$ } 3^{rd} Moment

$\boxed{X = \lim_{t\to\infty} M'''(t) - \lim_{t\to\infty} M''(t)}$

79. $p(0)=\frac{1}{2}; p(1)=\frac{3}{8}; p(2)=\frac{1}{8}$; Find $M(t)$; $M(t)=\sum_x e^{tx}p(x) = e^{tx}\left[\frac{1}{2}+\frac{3}{8}+\frac{1}{8}\right] = e^{tx}$

80. $f(x)=2x; 0\le x\le 1$

$M(t)=\int_0^1 e^{tx}\cdot 2x\,dx = \frac{2e^{tx}(tx-1)}{t^2}\Big|_0^1 = \frac{2e^t(x-1)}{t^2}+\frac{2}{t^2}$

$= \frac{2e^t(x-1)}{t^2} - \frac{2e^0(0-1)}{t^2}$

$M'(0)=xe^{0x}=E[x]$; $M''(0)=x^2e^{0\cdot x}=E(x^2)$; $M''(0)=0$

$M'(0)=\frac{d}{dt}\left[\left[1+t+\frac{t^2}{2!}+\frac{t^3}{3!}+\cdots\right]\left[\frac{2}{t}-\frac{2}{t^2}\right]+\frac{2}{t^2}\right]$

$=2(1)\left(\frac{1}{2!}-\frac{1}{3!}\right)+2(2\times 0)\left(\frac{1}{3!}-\frac{1}{4!}\right)+2(3\times 0)\left(\frac{1}{4!}-\frac{1}{5!}\right)+\cdots$

$=\boxed{\frac{2}{3}}$

81. Bernoulli Random Variable.

$p(1)=p$

$p(0)=1-p$

$p(x)=0$

$p(x)=\begin{cases}p^x(1-p)^{1-x}\\ 0\end{cases}$

$M(t)=\sum_0^1 e^{tx}p(x) = 1-p+e^t p$

$M'(0)=p=E(x)$ ... $M'''(0)=p$

$M''(0)=p=E(x^2)$

$\boxed{Var(x)=M''(0)-M'(0)^2=p-p^2}$

82. MGF of a Binomial is $\sum$ Bernoulli

Binomial Random Variable:

$M(t)=\prod_{i=1}^n M_{x_i}(t)=(1-p+e^tp)^n$

$M'(0)=n(1-p+p)^{n-1}\cdot p = n\cdot p = E[x]$

$M''(0)=n(n-1)(1-p+p)^{n-2}p^2+n(1-p+p)^{n-1}\cdot p = n(n-1)p^2+np=(n^2-n)p^2+np = n^2p^2-np^2+np$

$Var(x)=E[x^2]-E[x]^2 = n^2p^2-np^2+np-n^2p^2=\boxed{np(1-p)}$

83. Binomial Distribution:

$p(x)=\binom{n}{x}p^x(1-p)^{n-x}$; $P(X)=\sum_{i=1}^n X_i = X_1+X_2+\cdots+X_n = \binom{n}{1}p^1(1-p)^{n-1}+\binom{n}{2}p^2(1-p)^{n-2}+\cdots+\binom{n}{n}p^n(1-p)^{n-n} = \boxed{\sum_{k=1}^n\binom{n}{k}p^k(1-p)^{n-k}}$

84. $=\binom{n}{1}p_1(1-p_1)^{n-1}+\binom{n}{2}p_2(1-p_2)^{n-2}+\cdots+\binom{n}{n}p_n(1-p_n)^{n-n}\neq$ Binomial Distribution

85. Moment Generating Function: Geometric Random Variable $p(k)=p(1-p)^{k-1}$, $k=1,\ldots$

$M(t)=\sum_{k=1}^{\infty}e^{tk}\cdot p(1-p)^{k-1} = \frac{e^{tk}p}{1-(1-p)e^{tk}}$; Binomial Relationship: $e^t(1-p)^{k-1}\to\frac{1}{1-qe^t}$;

$M'(t)=\frac{ke^{tk}p[1-(1-p)e^{tk}]-e^{tk}p\,k(1-p)e^{tk}}{[1-(1-p)e^{tk}]^2} = \frac{ke^{tk}p-ke^{tk}p(1-p)e^{tk}+e^{tk}pk(1-p)e^{tk}}{[1-(1-p)e^{tk}]^2} = \frac{ke^{tk}p[1+2(1-p)e^{tk}]}{[1-(1-p)e^{tk}]^2} = \frac{pe^t}{(1-qe^t)^2}$

$M'(0)=\frac{1\cdot p}{p^2}=\frac{1}{p}$; $M''(t)=\frac{pe^t(1-qe^t)^2-pe^t\cdot 2(1-qe^t)\cdot(-q)e^t}{(1-qe^t)^4}=\frac{pe^t(1-qe^t)+2pe^t(1-qe^t)qe^t}{(1-qe^t)^4}$

$M''(0)=\frac{p(1-q)+2p(1-q)q}{p^4}=\frac{p-pq+2p+2pq^2}{p^4}=\frac{3-q-2q^2}{p^3}=\frac{3-(1-p)-2(1-p)^2}{p^3}$

$=\frac{2+p-2(1+2p+p^2)}{p^3}=\frac{2+p-2-4p-2p^2}{p^2}$

$Var(x)=E[x^2]-E[x]^2$

$=\frac{1+q}{p^2}-\frac{1}{p^2}=\boxed{\frac{1+p}{p^2}}$ $\boxed{=\frac{1+q}{p^2}}$

86. MgF of a Negative Binomial:

$p(X=k)=\binom{k-1}{r-1}p^r(1-p)^{k-r}$

$M(t)=\prod_{i=1}^n M_x(t)=p^n(1-qe^t)^{-n}$

$M'(t)=-np^n(1-qe^t)^{-(n+1)}(-qe^t)$; $M'(0)=+np^n(1-q)^{-(n+1)}q$

$=np^n\cdot p^{-(n+1)}\cdot q = np^{n-n-1}q = \boxed{\frac{nq}{p}}$

$M''(t)=\frac{d}{dt}\left[\frac{+np^nqe^t}{(1-qe^t)^{n+1}}\right]$

$=\frac{nq}{p}+\frac{n^2q^2}{p^2}+\frac{nq^2}{p^2}=\boxed{\frac{nq}{p^2}+\frac{n^2q^2}{p^2}}$

$Var(x)=E(x^2)-E[x]^2=\boxed{\frac{nq}{p^2}}$

87. When is the sum of independent random variables of a binomial also a negative binomial?

Binomial Distribution: $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$

Negative Binomial Distribution: $p(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$

When $k-1=n$, and $r-1=k$. When the total trials is the n and successes $(t \cdot a) \cdot k!x$

88. $X \sim N(0, \sigma_i^2)$; Prove odd moments of $M_X t$ are zero and even moments are $\mu_{2n} = \frac{(2n)!\sigma^{2n}}{2^n(n!)}$

$$\frac{-x^2}{2\sigma^2} + tx = \frac{-1}{2\sigma^2}(x^2 - 2\sigma^2 t x)$$

$$= \frac{-1}{2\sigma}\left(x^2 - 2\sigma^2 t x + \left(\frac{2\sigma^2 t}{2}\right)^2\right) - (\sigma^2 t)^2$$

$$= \frac{-1}{2\sigma^2}(x - \sigma^2 t)^2 - (\sigma^2 t)^2$$

Even Moments: $M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tX} N(0,\sigma^2)$

$$= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2\sigma^2} + tx} dx = \frac{e^{\sigma^2 t^2}}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma}(x-\sigma^2 t)^2} dx; \quad u = x - \sigma t, \quad \frac{du}{dt} = -\sigma$$

$$= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{\frac{-1}{2\sigma}(x-\sigma^2 t)^2 - (\sigma^2 t)^2} dx = \frac{e^{-(\sigma^2 t)^2}}{\sqrt{2\pi\sigma}(t\sigma)} \int_{-\infty}^{\infty} e^{-u^2/2\sigma^2} du = \frac{e^{+\frac{(\sigma^2 t)^2}{2}}}{\sqrt{2\pi}\,\cancel{\sigma}} \sqrt{2\pi\sigma} \cdot \frac{t!}{0} = \frac{t\, e^{+\frac{1}{2}\sigma^2 t^2}}{-\sigma^2}$$

$$\mu_{20} = \frac{(2 0)!\sigma^{2\cdot 0}}{2^0(0!)} = 1$$

$$M'(t) = +\sigma^2 t e^{-\sigma^2 t^2/2} \qquad \frac{2!\sigma^{2\cdot 1}}{2^1(1)} \quad \sigma^2$$

89. $$M''(t) = +\sigma^2 e^{-\sigma^2 t^2/2} + \sigma^2 t(-\sigma^2 t)e^{-\sigma^2 t^2/2} = -\sigma^2 e^{-\sigma^2 t^2/2}\left[1 - \sigma^2 t^2 e^{-\sigma^2 t^2/2}\right] = \boxed{+\sigma^2} @ t = 0$$

$$\frac{(2\cdot 1)!\sigma^{2n}}{2^1(1!)} = \boxed{=\sigma^2}$$

$X_1, X_2, \ldots X_n = N(\mu_i, \sigma_i^2)$; Prove $Y = \sum_{i=1}^{n} \alpha_i X_i$ where α_i = scalar. Find μ_Y, σ_Y^2

$Y = \sum_{i=1}^{n} \alpha_i X_i = \sum_{i=1}^{n} \alpha_i N(\mu_i, \sigma_i^2) \div$ MGF of Y: $M(t) = \sum_{i=1}^{n} e^{tx} \cdot \sum \alpha_i X_i$

$$= \sum_{X=1}^{n} e^{tX}\left(\sum \alpha_i N(\mu_i, \sigma_i^2)\right)$$

$$\sum_{X=1}^{n} \sum_{i=1}^{n} \alpha_i e^{\frac{-1}{2\sigma^2}(X_i - (\mu + \sigma^2 t)^2)^2} e^{-\frac{(2\mu + \sigma^2 t^2)}{2}}$$

$$= e^{-\frac{(2\mu + \sigma^2 t^2)}{2}} \sum_{X=1}^{n} \sum_{i=1}^{n} \alpha_i e^{\frac{-1}{2\sigma^2}(X_i - (\mu_i + \sigma_i^2 t)^2)^2}$$

$$= \sum_{i=n}^{n} \alpha e^{-(2\mu + \sigma^2 t)/2} \sum_{X=1}^{n} e^{\frac{-1}{2\sigma^2}(X_i - (\mu_i + \sigma_i^2 t)^2)^2}$$

$$= M_Y(t) = \sum_{X=1}^{n} e^{tX} e^{-y^2/2} = e^{t^2/2}$$

$$= \sum_{X=1}^{n} e^{tX} \sum_{i=1}^{n} \alpha_i N(\mu_i, \sigma_i^2) = e^{t^2/2} \sum_{i=1}^{N} \alpha_i N(\sigma_i^2 t, \sigma_i^2)$$

$$\boxed{= e^{\mu_i \sigma_i^2 t^2/2}}$$

Completing the Square:

$$tx - \frac{(x-\mu)^2}{2\sigma^2} = tx - \frac{x^2 + 2\mu x + \mu^2}{2\sigma^2}$$

$$\frac{-x^2 + (2\mu + 2\sigma^2 t)x}{2\sigma^2} + \frac{\mu^2}{2\sigma^2}$$

$$\frac{-1}{2\sigma^2}\left(x + (\mu + \sigma^2 t)^2\right)^2 - \frac{(\mu + \sigma^2 t)^2}{2\sigma^2} + \frac{\mu^2}{2\sigma^2}$$

$$\frac{-1}{2\sigma^2}\left(x - (\mu + \sigma^2 t)\right)^2 - \frac{(2\mu + \sigma^2 t^2)}{2} + 2\mu\sigma^2 t + (\sigma^2 t)^2 + \mu^2$$

90. $Z = \alpha X + \beta Y$; $M_Z(t) = E(e^{tZ})$

$$= E\left(e^{tX\alpha + t\beta Y}\right) = E\left(e^{t\alpha X} \cdot e^{t\beta Y}\right)$$

$$= E\left(e^{t\alpha X}\right) E\left(e^{t\beta Y}\right)$$

$$\boxed{= M_X(t) M_Y(t)}$$

91. Exponential Distribution:

$$p(X) = \lambda e^{-\lambda x}$$

$$\cancel{M_X(t) = \lambda e^{(t-\lambda)x}}$$

$$M_{cX}(t) = \lambda e^{(t - c\lambda)X}$$

92. Gamma Distribution: ; Poisson Distribution ~~With Example~~:

$g(t) = \frac{\lambda^\kappa}{\Gamma(\kappa)} t^{\kappa-1} e^{-\lambda t}$ $\quad p(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ $\quad f(\theta) = \frac{\lambda^\kappa}{\Gamma(\kappa)} \theta^{\kappa-1} e^{-\lambda\theta}$; $X|\theta = \frac{\theta^x}{x!} e^{-\theta}$

$$P(X=x) = \int_0^\infty P(X|\theta) f(\theta) d\theta = \int_0^\infty e^{-\theta} \frac{\theta^x}{x!} \frac{\lambda^\kappa \theta^{\kappa-1} e^{-\lambda\theta}}{\Gamma(\kappa)} d\theta = \int_0^\infty \frac{\lambda^\kappa \theta^{x+\kappa-1} e^{-(\lambda+1)\theta}}{x!(\kappa-1)!} d\theta$$

$$= \frac{(x+\kappa-1)!}{x!(\kappa-1)!} \frac{\lambda^\kappa}{(\lambda+1)^{x+\kappa}} \int_0^\infty \frac{(\lambda+1)^{x+\kappa} \theta^{x+\kappa-1} e^{-(\lambda+1)\theta}}{\Gamma(x+\theta)} d\theta \text{ ; Rate } = (\lambda+1) \quad \text{Shape} = \kappa+x$$

$M_{X|\theta}(t) = E[e^{tX}|\theta] = \exp((e^t-1)\theta)$ and $M_\theta(m) = E[e^{m\theta}] = (1-m/\lambda)^{-\alpha}$

$$M_X(t) = E[e^{tX}] = E[E[e^{tX}|\theta]] = E[M_{X|\theta}(t)] = E[\exp((e^t-1)\theta)] = M_\theta(e^t-1) = \left(1-\frac{e^t-1}{\lambda}\right)^{-\kappa}$$

93. Geometric Sum: Exponential Random Variable $\quad M_X(t) = \left(\frac{\lambda/(1+\lambda)}{1-e^t(1-\lambda/[1+\lambda])}\right)^\kappa \propto \frac{\lambda}{\lambda+1}$

Negative Binomial

$X_r = X_1 + X_2 + \cdots + X_n \quad p(X) = \lambda e^{-\lambda x}$

$$M(t) = \sum_{x=0}^{\infty} \sum_{i}^{n} X_{j,i}(x)_{j,i} e^{tx_i} = \sum_{x=0}^{\infty} \lambda e^{-(\lambda+t)x_i} = \boxed{\frac{-\lambda}{(\lambda+t)} e^{-(\lambda+t)x}}$$

94. Probability-Generating Function: $G(s) = \sum_{k=0}^{\infty} s^k p_k$; where $p_k = P(X=k)$

a) Show $p_k = \frac{1}{k!} \frac{d^k}{ds^k} G(s)\Big|_{s=0}$; Fundamental theorem of Calculus: $\int_0^b f(x)dx = F[b]-F(a) = \frac{d}{dx}F(x)$

b). Show $\frac{dG}{ds} = E(X)$ $\quad \frac{dG}{ds} = k s^{(k-1)} p(k) = E(X) = \boxed{k \cdot p(k)}$

$\frac{d^2G}{ds^2}\Big|_{s=1} = E[X(X-1)]$ $\quad \frac{d^2G}{ds^2} = k(k-1)s^{(k-2)} p(k) \sim \boxed{k(k-1)p(k)}$

c) $M(t) = \sum_{s=0}^{\infty} \sum_{s=0}^{\infty} e^{tk} G(s) = \sum_0^\infty \sum_0^\infty e^{tk} s^k p_k = \boxed{\frac{\sum e^{tk} s^{k+1} p(k)}{k+1}}$

d) $G(s) = \int_0^\infty s^k \frac{\lambda^k}{k!} e^{-\lambda} dk = e^{-\lambda} \sum_0^\infty \frac{(s\lambda)^k}{k!} = e^{-\lambda} e^{\lambda s} = \boxed{e^{(\lambda s-\lambda)}}$

95. Joint Moment Generating Function. $M(t) = \sum_{-\infty}^{\infty} e^{t(x+y)} xy = \sum e^{tx} x e^{ty} y = \boxed{M_X(t) M_Y(t)}$

96. $E(XY) = M'(0) = \frac{d}{dt}[\sum e^{tx} x e^{ty} y] = \boxed{xy\, p(x,y)}$ $\quad M(t) = \sum e^{tX+tY} p(x)$; $M'(t) = \sum (x+y) e^{tX+tY} p(x)$

97. $Var(aX+bY) = a^2 Var(X) + b^2 Var(Y)$; $\quad M''(t) = \sum (x+y)^2 e^{tX+tY} p(x)$

$M''(0) = (x+y)^2 p(x) = E[X^2]$; $M'(0) = (x+y) p(x) = E[X]$

$= E[X^2] - E[X]^2 = (x+y)^2 p(x) - (x^2+y)^2 p(x)^2 = (x^2+2xy+y^2)(p(x)-p(x)^2)$

$= \boxed{Var(X) + Var(Y)}$

98. Compound Poisson Distribution

$M_S(t) = \exp[\mu(e^{\lambda(e^t-1)}-1)]$

$M_S'(0) = [\mu(e^{\lambda(e^t-1)}-1)]' \exp[\mu(e^{\lambda(e^t-1)}-1)]$

$= [\lambda(e^t-1)]'[\mu(e^{\lambda(e^t-1)}-1)] \exp[\mu(e^{\lambda(e^t-1)}-1)]$

$= \lambda e^t[\mu(e^{\lambda(e^t-1)}-1)] \exp[\mu(e^{\lambda(e^t-1)}-1)] = \emptyset \neq E[X]$

$M_S''(0) = \lambda[\mu(e^{\lambda(e^t-1)}-1)] \exp[\mu(e^{\lambda(e^t-1)}-1)] + \lambda e^{2t}[\mu(e^{\lambda(e^t-1)}-1)] \exp[\mu(e^{\lambda(e^t-1)}-1)]$
$+ (\lambda e^t[\mu(e^{\lambda(e^t-1)}-1)] \exp[\mu(e^{\lambda(e^t-1)}-1)]$

$= \emptyset$

99. $Y = g(X)$ a) $g(x) = \sqrt{X}$

$E[Y] = \int_0^\infty x\sqrt{x}\,dx = x(\frac{2}{3})x^{3/2}\Big|_0^\infty - \int_0^\infty \sqrt{x}\,dx = \infty$

$Var[Y] = \infty$

b) $E[X] = \int_0^{\infty} x \log x \, dx = x\left(\frac{1}{x}\right)\Big|_{10}^{\infty} - \int_0^{\infty} \log(x)\, dx = 1 - \frac{1}{x}\Big|_0^{\infty} = \text{Undefined}$

Var(X) = Undefined

c) $g(x) = \sin^{-1}(x)$; $E[X] = \int_{-\pi/2}^{\pi/2} x \sin^{-1}(x)\, dx$ = Does not converge Var(X) = Does not converge

100. $X[10, 20]$; $Y = 1/X$; $E[X] = \int_{10}^{20} \frac{1}{10}\left(\frac{1}{x}\right) dx = \frac{\ln 20 - \ln 10}{10}$; $= 0.0693$; $E[Y^2] = \int_{10}^{20} \frac{1}{10}\left(\frac{1}{x^2}\right) dx = \int_{10}^{20} \frac{1}{10x^2} dx = \frac{-x^{-1}}{10}\Big|_{10}^{20} = 0.005$

Exact Method

$Var(X) = E[X^2] - E[X]^2 = 0.005 - 0.0693^2 = 0.000198$

Approximate Method

$Y(X) = \frac{1}{X}$; $Y'(X) = \frac{-1}{X^2}$; $Y''(X) = \frac{-2}{X^3}$; $E(Y) \sim g(\mu_X) + \left(\frac{1}{2}\right)\sigma_X^2 g''(\mu_X) = \frac{1}{15} + \left(\frac{1}{2}\right) 8.33(0.0006)$ $\boxed{= 0.0249}$

$Var(Y) \approx \sigma_X^2 [g'(\mu_X)]^2 = 0.000161$

$\sigma_X^2 = \frac{(b-a)^2}{12} = 8.33$ $\sigma_Y = 0.0127$

101. $Y = \sqrt{X}$; X = Poisson Distribution

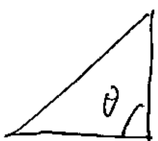
$\frac{\lambda^k}{k!} e^{-\lambda}$; $Y'(X) = \frac{1}{2}(X)^{-1/2}$; $Y''(X) = \frac{-1}{4}(X)^{-3/2}$; $Y'''(X) = \frac{-3}{8}(X)^{-5/2}$

$\mu_X = \lambda, \sigma_X^2 = \lambda$

$E(Y) \sim Y(\mu_X) + \frac{1}{2}(\sigma_X^2) g''(\mu_X)$
$\sim \lambda^{1/2} - \frac{1}{2}\lambda \cdot \frac{1}{4}\lambda^{-3/2}$
$\boxed{\sim \sqrt{\lambda} - \frac{1}{8\sqrt{\lambda}}}$

$Var(Y) \approx \sigma_X^2 [g'(\mu_X)]^2$
$\approx \lambda\left[\frac{1}{2}(\lambda)^{-1/2}\right]^2$
$\boxed{\approx \frac{1}{4}}$

102.



$X_0 = X$
$E(X) = X_0$

$y_0 = Y$; $E(Y) = y_0$; $Var(X) = Var(Y) = \sigma^2$

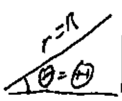
$\theta = \tan^{-1}\left(\frac{Y}{X}\right)$; $E(\theta) \sim$

$E(\theta) \sim \tan^{-1}\left(\frac{E(Y)}{E(X)}\right) = \boxed{\tan^{-1}\left(\frac{y_0}{x_0}\right)}$

$Var(\theta) \sim \tan^{-1}\left(\frac{Var(Y)}{Var(X)}\right) = \tan^{-1}(1) = \boxed{45°}$

103. $V = \frac{\pi}{6} D^3$; $D = 2\text{mm}$; $\sigma_X = 0.01\text{mm}$; $V' = \frac{\pi}{2} D^2$; $\sigma_Z^2 \approx \sigma_X^2 [g'(\mu_X)]^2$; $\sigma_Z \approx \sigma_Z g'(\mu_X)$
$\approx 0.01\text{mm} \frac{\pi}{2} 2\text{mm}^2$
$\boxed{\approx 6.28 \times 10^{-2} \text{mm}^3}$

104.



$Y = R\sin\theta$; a) $Var(Y) \sim \sigma_X^2 [g'(\mu_Y)]^2 \sim \sigma_\theta^2 [R\cos\theta]^2$

b) $\frac{dVar(Y)}{d\theta} = \sin\theta = 0$; $\boxed{90° = \theta}$

Chapter 5: 1) $X_1, X_2 \ldots$; $E(X_i) = \mu$; $Var(X_i) = \sigma_i^2$, Show if $n^{-2}\sum_{i=1}^{n}\sigma_i^2 \to 0$; $\bar{X} \to \mu$

Law of Large Numbers: $P(|\bar{X} - \mu| > \epsilon) \leq \frac{Var(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$

$P(|\bar{X} - \mu| > \epsilon) \leq E[X^2] - E[X]^2$; $\bar{X}^2 = \mu^2$; $\boxed{\bar{X} = \mu}$

2. $E(X_i) = \mu_i$; if $E(\bar{X}_i) = \frac{1}{n}\sum_1^n E(X_i) = \mu$
$= \frac{1}{n}(n E(X_i)) = E(X_i) = \mu$

3. Number of Insurance claims, N, is a Poisson Distribution: $p(x) = \frac{\lambda^k}{k!} e^{-\lambda}$
E(N) = 10,000.
$E(X) = \lambda$
$Var(X) = \lambda$
Apply a normal approximation to the Poisson to approximate $P(N > 10,200)$.

Standardizing Random Variables

$Z_n = \frac{X_n - E(X_n)}{\sqrt{Var(X_n)}} = \frac{10,200 - 10000}{\sqrt{10000}} = 102 - 100 = 2$

$P(Z_n) = 1 - 0.9772 = \boxed{0.028}$

4. Number of Traffic Accidents (N) is $E(N) = 100$
Find Δ if a poisson $P(100 - \Delta < N < 100 + \Delta) \approx 0.9$

$Z_n = \frac{X_n - E(X_n)}{\sqrt{Var(X)}}$; $P(100 - \Delta) = \frac{100 - \Delta - 100}{\sqrt{100}} = 10 - 10 - \Delta = -\Delta$

$P(N) = \frac{X_n - 100 + \Delta}{\sqrt{100}} = \Delta$; $\boxed{\Delta = 1.3}$

Means 1.3 cars more or less for probability of 90%

5. $n \to \infty, p \to \infty$, and $np = \lambda \to \infty$ Binomial Distribution: $p(x) = \sum_{1}^{n} \binom{n}{k} p^k (1-p)^{n-k}$; n and p tend poisson $p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$

Moment Generating Function of Binomial: $M(t) = \int_0^{\infty} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} dk$; if $np = \lambda$

$= \binom{n}{k} \int_0^{\infty} e^{tk} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} dk$; $\boxed{\text{Continuity Theorem}}$ $\lim_{n\to\infty} M_n(t) \to M(t)$; $\lim_{n\to\infty} F_n(x) = F(x)$ $\boxed{\text{Law of Large Numbers}}$

$\lim_{n\to\infty} M_n(t) = \lim_{n\to\infty} \int_0^{\infty} \binom{n}{k} e^{tk} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} dk = \lim_{n\to\infty} \int_0^{\infty} e^{tk} \frac{\lambda^k}{k!} \underbrace{\frac{n!}{(n-k)!} \frac{1}{n^k}}_{=1} \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{=e^{-\lambda}} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{=1} dk = \int_0^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} dk$

$$\underbrace{= \int_0^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} dk}_{\text{Poisson Distribution}} = \int_0^{\infty} \frac{(\lambda e^t)^k}{k!} e^{-\lambda} dk = e^{\lambda e^t} \cdot e^{-\lambda} = \boxed{e^{\lambda(e^t - 1)}}$$

6. $\alpha \to \infty$ Gamma Distribution: $\frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}$ $\to M_\alpha(t) = \int_0^{\infty} e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$; $\lim_{\alpha\to\infty} M_\alpha(t) = \lim_{\alpha\to\infty} \int_0^{\infty} e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$

$= \lim_{\alpha\to\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{x(t-\lambda)} dx = \lim_{\alpha\to\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{\Gamma(\alpha)}{(\lambda-t)^\alpha}\right) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha = \infty$

If $\mu = \lambda/t$, $\sigma^2 = k/t^2$? $= \left(\frac{\mu t}{\mu t - k/t^2}\right)^k$

$= \left(1 - \frac{t}{\lambda n}\right)^{-\lambda}$

7. $X_n \to c$; g is continuous, then $g(X_n) \to g(c)$

Continuity Theorem $\boxed{\lim_{X_n \to c} g(X_n) = g(c)}$

8. Poisson Cumulative Distribution: a) $\lambda = 10$; $cdf_{Poisson} = \int_0^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} dk = e^{\lambda} \cdot e^{-\lambda}$; 1 b) 1 c) $1 \approx$ Normal Standard

9. Binomial Cumulative Distribution: a) $n=20$, $p=0.2$ $CDF_{Binomial} = \int_0^{\infty} \binom{n}{k} p^k (1-p)^{n-k} dk = \int_0^{\infty} \frac{20!}{k!(n-k)!} 0.2^k (0.8)^{n-k} dk = 20! \, 0.8^{20} \int_0^{\infty} \frac{1}{4^k} \frac{dk}{k!(20-k)!}$

$\boxed{8.9 \times 10^{-5}}$

b) $n=40$, $p=0.5$ $= 40! \, 0.5^{40} \int_0^{\infty} \frac{dk}{k!(40-k)!} = 0.99$

$\boxed{\text{The binomial converges to the Normal Standard when correct run}}$

$\left(\frac{40 - 40 \cdot 0.5}{\sqrt{40 \cdot 0.5(1-0.5)}}\right) x$?

Normal Approximation $\left(\frac{\lambda - E(\lambda)}{\sqrt{Var}}\right) = \frac{0.2 - 20 \cdot 0.2}{\sqrt{20 \cdot 0.2(1-0.2)}} = \varnothing$

10. Six sided dice; $n=100$; $\left(\frac{X - E(X)}{\sqrt{Var}}\right) = P(x)$; $P(15 < X < 20) = \left(\frac{15 - 100(1/6)}{\sqrt{100(\frac{1}{6})(1-1/6)}} < X < \frac{20 - 100(1/6)}{\sqrt{100(1/6)(1-1/6)}}\right)$ ~~35% to 23%~~

$\approx P(15.5 < X < 19.5) = P\left(\frac{15.5 - 100/6}{\sqrt{100/6(1-1/6)}} < X < \frac{19.5 - 100/6}{\sqrt{100/6(1-1/6)}}\right)$

$\approx P(-0.31 < Z < 0.76) = P(Z < 0.76) - P(Z < -0.31)$

$= P(Z < 0.76) - 1 + P(Z < 0.31) = 0.774 - 1 + 0.6217 = 0.4$

$E[X] = \frac{6+1}{2} = 3.5$; $Var(X) = \frac{1}{12}(6^2 - 1) = 2.917$

$E[S] = 100 E[X] = 100(3.5) = 350$; $Var(S) = 100(2.917) = 291.67$

$P(S < 300) \approx P(S < 299.5) = P\left(Z < \frac{299.5 - 350}{\sqrt{291.67}}\right)$

$= P(Z < -2.96) = 1 - P(Z < 2.96) = \boxed{0.00154}$

Central Limit Theorem:

$S_n = \sum_{i=1}^{n} X_i$; $\lim_{n\to\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \le x\right) = \Phi(x)$; $-\infty < x < \infty$

Proof: $Z_n = S_n / (\sigma\sqrt{n})$

$M_{S_n}(t) = [M(t)]^n$; $M_{Z_n}(t) = \left[M\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$

$M(s) = M(0) + sM'(0) + \frac{1}{2}s^2 M''(0) + \epsilon_s$

$\epsilon_s / s^2 \to 0$ as $s \to 0$; $E[X] = 0$, $M'(0) = 0$, $M''(0) = \sigma^2$

As $n \to \infty$, $t/(\sigma\sqrt{n}) \to 0$

$M\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{1}{2}\sigma^2 \left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \epsilon_n$

$M_Z(t) = \left(1 + \frac{t^2}{2n} + \epsilon_n\right)^n$; $\lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$

$M_{Z_n}(t) = e^{t^2/2}$ $n \to \infty$

11. The argument suffices to say $\lambda = 1 = np = \infty \cdot p \Rightarrow \frac{1}{p} = \infty$; The probability must approach $\emptyset$ as $n \to \infty$.

12. Uniform Random Variable $[-\frac{1}{2}, \frac{1}{2}]$. $n = 100$; $P(X > 1) = \text{Prob}\left(\frac{X - \sum E(X_i)}{\sqrt{\sum Var(X)}}\right) = P\left(\frac{X - 100 \cdot 0}{\sqrt{100 \cdot 1/12}} > \frac{1-0}{\sqrt{100/12}}\right) = P\left(\frac{\sqrt{3}}{25} Y > \frac{\sqrt{3}}{25}\right)$

$E[X] = \frac{a+b}{2}$; $Var(X) = \frac{(b-a)^2}{12} = \frac{(\frac{1}{2} - -\frac{1}{2})^2}{12} \cdot \frac{1}{12}$

$= 1 - \phi\left(\frac{\sqrt{3}}{5}\right)$

$= 0.3645$.

b) $P(Y > 2) = P\left(\frac{Y-0}{\sqrt{\frac{25}{3}}} > \frac{2-0}{\sqrt{\frac{25}{3}}}\right) = 1 - \phi\left(\frac{2\sqrt{3}}{5}\right) = 0.2442$

c) $P(Y > 5) = P\left(\frac{Y-0}{\sqrt{\frac{25}{3}}} > \frac{5-0}{\sqrt{\frac{25}{3}}}\right) = 1 - \phi\left(\frac{5\sqrt{3}}{5}\right) = \boxed{0.0416}$

13. $P(\text{North}) = \frac{1}{2}$; Step Length $= 50$ cm $E(X) = \sum X \cdot P(X)$; $Var(X) = E[X^2] - E[X]^2$

$\underbrace{P(\text{South}) = \frac{1}{2}}_{\text{per minute}}$; Approximate probability after 1 h. $= 50(\frac{1}{2}) + (-50)(\frac{1}{2})$ $= \emptyset$

$= \sum x^2 P(x) + \emptyset^2$

$= 50^2(\frac{1}{2}) + (-50)^2(\frac{1}{2})$

$= 2500$

$E(S) = \sum^{60} E(X) = 60 \cdot \emptyset = \emptyset$ $Var(S) = \sum^{60} Var(X) = 60 \cdot 2500$

$= 150{,}000$

Standard Deviation $(\sqrt{Var(s)}) = \sigma_s \cdot \sqrt{150\,000}$

$= 387$

14. $P(\text{North}) = \frac{2}{3}$

$P(\text{South}) = \frac{1}{3}$

$E[X] = \frac{1}{3}(-50) + \frac{2}{3}(+50) = \frac{50}{3}$; $E[X^2] = \frac{1}{3}(-50)^2 + \frac{2}{3}(50)^2$

$= 2500$; $Var[X] = E[X^2] - E[X]^2 =$

$E(S) = \sum_0^{60} E[X] = \frac{3000}{3} = 1000$;

$= 2500 - 1000 = 1500$

$E[S^2] = \sum_0^{60} E[X^2]$

$Var(S) = \sum_0^{60} Var[X] = 60 \cdot 1500$

$= 90{,}000$

$= 60 \cdot 2500 = 150{,}000$

Standard Deviation $(\sqrt{Var}) = \sigma = \boxed{300}$

15. $n = 50$; Amount $= \$5$

Probability (Loss > 75).

$P(\text{Win}) = \frac{1}{2}$; $P(\text{Loss}) = \frac{1}{2}$

$\sum^{50} X_i < -75$; $\bar{X} < \frac{-75}{50} = -1.5$; $E(X) = 0$; $Var(X) = E(X^2) - E(X)^2$

$= \sum x^2 p(x) - 0 = 5^2(\frac{1}{2}) + (-5)^2(\frac{1}{2}) = 25$

$P(\bar{X} < -1.5) = P\left(Z < \frac{-1.5}{\sqrt{1/2}}\right) = P(Z < -2.12) = \phi - \phi(2.12) = 0.017$

16. $X_1 \dots X_{20}$; $f(x) = 2x$; $0 \le x \le 1$; $S = X_1 + \dots + X_{20}$; $P(S \le 10)$; $E[X] = \int_0^1 x f(x)\,dx = \int_0^1 x \cdot 2x\,dx = \frac{2x^3}{3}\Big|_0^1 = \frac{2}{3}$

$P(S \le 10) = P\left(\frac{S - E[S]}{\sqrt{Var(S)}} < \frac{10 - E[S]}{\sqrt{Var(S)}}\right) = P\left(\frac{X - 40/3}{\sqrt{1.11}} < \frac{10 - 40/3}{\sqrt{1.11}}\right)$

$E[S] = \sum^{20} \frac{2}{3} = \frac{40}{3}$; $E[X^2] = \int_0^1 x^2 2x\,dx = \frac{2x^4}{4}\Big|_0^1 = \frac{1}{2}$

$Var(\frac{S}{n}) = \frac{1}{n} Var$

$Var(X) = E[X^2] - E[X]^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$

$= P\left(\frac{S - 40/3}{\sqrt{1.11}} < -3.16\right) = 1 - (1 + \phi(3.16)) = \boxed{0.9992}$

$Var(S) = \sum^{20} Var[X] = \frac{20}{18} = 1.11$.

17. $\mu, \sigma^2 = 25$; $P(|\bar{X} - \mu| < 1) = P(-1 < \bar{X} - \mu < 1) = P\left(-\sqrt{n}/\sigma < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \sqrt{n}/\sigma\right) = 0.95$

$\underbrace{P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \sqrt{n}/\sigma\right) = \frac{1 - 0.95}{2} = 0.025}_{\text{"Describes size of range"}}$; $-\frac{\sqrt{n}}{\sigma} = -1.97$; $n = 97$.

18. $\mu = 15$ lbs ; $P(F.S) < 1700$ lbs$) = P\left(\frac{S - E[S]}{\sigma/\sqrt{n}} < \frac{1700 - n\mu}{\sigma/\sqrt{n}}\right) = P\left(\frac{S - n\mu}{\sigma/\sqrt{n}} < \frac{1700 - 1500}{10/\sqrt{100}}\right)$

$\sigma = 10$;

$n = 100$

$= P(2) = 0.9772$

19. a) $n=100$, $n=1000$; $f(x)=\int_0^1 \cos(2\pi x)\,dx = \frac{\sin(2\pi x)}{2\pi}\Big|_0^1 = \emptyset$; $\hat{I}_{100}(f)=\frac{1}{n}\sum_{i=0}^{1} f(X_i) = \frac{1}{100}\sum_{i=0}^{1}\cos(2\pi x) = \frac{1}{100}+\frac{1}{100}=\frac{2}{100}$

$\hat{I}_{1000}(f)=\frac{1}{1000}\sum_{i=0}^{1} f(X) = \frac{1}{1000}\sum_{0}^{1}\cos(2\pi x) = \frac{1}{1000}+\frac{1}{1000}=\frac{2}{1000}$

b) $I(f)=E[X]=\int_0^1 \cos(2\pi x^2)\,dx =$

$\hat{I}(f)=\frac{1}{n}\sum_{i=0}^{1}\cos(2\pi x^2) = \frac{1}{10^3}\sum_{i=0}^{10^3}\cos(2\pi x^2) = 1 \ldots > 0$

Exact Solution: cosine integral: $\int \cos(u)\,dx$; $u=2\pi x^2$, $\frac{du}{dx}=2\pi x$; $\int \frac{\cos(u)}{2\pi x}dx = \int \frac{\cos(u)}{2\pi\sqrt{u}/2\pi}du = \int_0^1 \frac{\cos(u)}{\sqrt{u}}du = \int_0^1 \frac{\cos(2u)}{u}du$

$= \frac{C(2)}{2} = \boxed{0.24}$

20. $E(\hat{I}^2(f)) = \left(\frac{1}{1000^3}\right)\left[\left(\frac{1}{\sqrt{2\pi}}\right)\sum_{i=1}^{100} e^{-x^2/2}\right]^2$

$\approx \frac{1}{1000}\,\frac{1}{2\pi}\sum_{i=1}^{100}\left(e^{-x_i^2}\right) = \frac{0.386}{2\pi\times 10^3} = 6.14\times 10^{-5}$

$Var(\hat{I}(f)) = E[\hat{I}^2(f)] - E[\hat{I}(f)]^2 = \left(\frac{1}{1000}\sum_{i=1}^{100}\left(e^{-x^2/2}\right)^2 - \frac{1}{1000}\left[\sum_{i=1}^{1000} e^{-x^2/2}\right]^2\right)\frac{1}{2\pi}$

21. $I(f)=\int_a^b f(x)\,dx$; $\hat{I}(f)=\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)}{g(X_i)}$ a) Show $E(\hat{I}(f))=I(f) = \int_a^b f(x)/g(x)\,dx$ $\boxed{\frac{1}{n}\sum_{i=1}^{n} f(x)/g(x)}$

b) $Var(\hat{I}(f)) = E[\hat{I}(f)^2] - E[\hat{I}(f)]^2$; $n=100$; $n\to\infty$

c) $E(\hat{I}(f)) = \hat{I}(f) = \int_0^1 \frac{f(x)}{g(x)}dx = \int_0^1 h(x)\,dx = \int_0^1 e^{-x^2/2}dx \to$ Example A: Section 5.2.

22.

Find Δ; $P(|\hat{I}(f)-I(f)|\le\Delta)=0.05$, where $\hat{I}(f)$ is the Monte Carlo Estimate of $\int_0^1\cos(2\pi x)\,dx$ based upon $n=1000$.

$P\left(\left|\frac{1}{1000}\sum_{i=1}^{100}\cos(2\pi x) - \int_0^1\cos(2\pi x)\,dx\right|\le\Delta\right)\approx 0.05$

$P(|1|\le\Delta)=0.05$; $P(\Delta)=0.05-P(1)=0.05-0.84=-0.79$; $\boxed{\Delta=0.81}$

$P(\Delta)-P(1)=0.005$

23. $0\le x\le 1$; $0\le y\le 1$; Random (x,y); $Z=1$ if so, $Z=0$ otherwise. Prove $E(Z)=A=\sum_0^1\sum_0^1 xy\cdot Z(1) + \sum_0^1\sum_0^1 xy\,Z(0)$

$\boxed{=1.1}A$

24. $\hat{A}$; $E(S)=\sum_{i=1}^{n}E[Z]=\boxed{nE[Z]}$; $\hat{A}=nE[Z]=nA$; $P(|\hat{A}-A|<0.1)\approx 0.99$

$P(|nA-A|<0.1)\approx 0.99$; $P(|(n-1)0.2<0.1)\approx 0.99$; $P(|n|<3/2)\approx 0.99$

$P(3/2)-P(|n|)\approx 0.99$

$P(|n|)\approx P(3/2)-0.99$

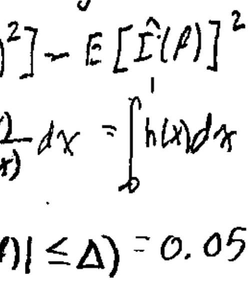
$\approx 0.9332-0.99=-0.0568$

$P(n)=0.0568$

$n=1.58$

25. $f(x)=\frac{3}{2}x^2$, $-1\le x\le 1$;

$\frac{2}{3}\le x\le 1$

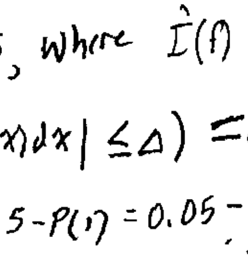


$f(x)=\begin{cases}\frac{3}{2}x^2; & -1\le x\le 1\\ 0\end{cases}$

$S=X_1+\ldots+X_{50}$

$E[S]=\sum_{i=1}^{50}S = \sum_{i=1}^{50}X_i$

$=50E[X]=50\cdot\int_1 x\,dx$



$E[X]=\int_{-1}^{1}x\,\frac{3}{2}x^2\,dx$

$=0$

$E[X^2]=\int_{-1}^{1}x^2\,\frac{3}{2}x^2\,dx = 3/5$; $\sigma^2=\frac{3}{5}$

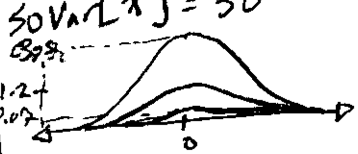
$E[S]=\sum_{i=1}^{50}X_i = 50E[X]=\phi$

$Var[S]=\sum Var(X) = 50Var[X]=30$

$\frac{3}{2}x^2-1$

$=50\left[\frac{x^2}{2}-\frac{(\frac{3}{2}x^2-1)^2}{2}\right]$

$=50\left[\frac{x^2}{2}-\frac{9}{8}x^4+\frac{3}{2}x^2-\frac{1}{2}\right] = 50\left[-\frac{9}{8}x^4+2x^2-\frac{1}{2}\right]=$



26. $P(\text{shot}) = 0.3$; $E[S] = \sum_{i=1}^{25} x\,p(\text{success} \mid \text{shot}) \ldots 25 \cdot p(x) = 25\left(\frac{3}{10}\right)$; $\underbrace{\frac{5}{25} \geq p(x)}$, $\underbrace{\frac{7}{25} \geq p(x) \quad \frac{9}{25} \geq p(x) \quad \frac{11}{25} \geq p(x)}$

$\leq \frac{3}{10} \leq$ | 30.

27. Prove $a_n \to a$, then $(1+a_n/n)^n \to e^a$; $\lim_{a_n \to a} (1+a_n/n)^n = \boxed{1 + \frac{1}{2^2}\left(\frac{a^2}{2}\right) + \frac{1}{2}\left(\frac{a^2}{2^2}\right) = e^a}$

28. $f_n(x) \begin{cases} 1/2 & \text{if } x = \pm(1/2)^n \\ 0 & \text{otherwise} \end{cases}$; $E[X] = \sum_{-\infty}^{\infty} x f(x) = -\frac{1}{2} - \frac{1}{4} \cdots 0 + \frac{1}{4} + \frac{1}{2}$

$F_n(x) = \begin{cases} 1/2 & (1/2)^n/n \\ 0 & \text{otherwise} \end{cases}$ $\boxed{\lim_{n} F_n(x) = F_n(x)}$

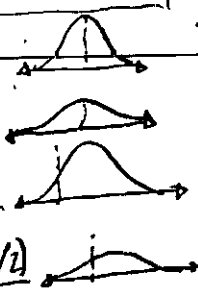
$U_1, U_2, \ldots U_{1000}$; $S_n = \sum_{i}^{n} U_i$; $n = 1 \ldots 1000$

a) S_n

b) $\frac{S_n}{n}$

c) $S_n - n/2$

d) $\frac{(S_n - n/2)}{n}$



e) $(S_n - n/2)/\sqrt{n}$

29. $U_1 \ldots U_n$ from $[0,1]$; $U_{(n)}$ = maximum

$\int_0^1 U_{(n)}\,du \doteq \Phi = \frac{U - E[U]}{\sqrt{Var(U)}} = \varphi(n)$

Chapter 6: Distributions derived from Normal Distribution

χ^2 = Chi-squared $V = Z^2$

$X \sim N(\mu, \sigma^2)$; $(X-\mu)/\sigma \sim N(0,1)$; $\left[\frac{X-\mu}{\sigma}\right]^2 \sim \chi_1^2$

$\boxed{\text{Chi-squared Distribution}}$ $[\chi_n^2]$ "n-degrees of freedom"

$V = U_1 + U_2 + \cdots + U_n$

sum of independent Gamma $\alpha = \frac{n}{2}$; $\lambda = \frac{1}{2}$

$f(v) = \frac{1}{2^{n/2}\Gamma(n/2)} v^{(n/2)-1} e^{-v/2}$; $v \geq 0$

$\boxed{\text{t-distribution}}$ if $Z \sim N(0,1)$ and $U \sim \chi_n^2$

then $Z/\sqrt{U/n}$ is a t-distribution:

Density Function: $f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$

Moment Generating Function $M(t) = (1-2t)^{-n/2}$

$E(V) = n$; $Var(V) = 2n$

$\boxed{\text{F-distribution}}$ $W = \frac{U/m}{V/m}$

Density Function: $f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} w^{m/2-1}\left(1 + \frac{m}{n}w\right)^{-(m+n)/2}$

$E(W) = \frac{n}{(n-2)}$

Sample Mean and Variance

Sample: independent random variables. Which are normally distributed

$\bar{X} = \frac{1}{n}\sum X_i$; $S^2 = \frac{1}{n-1}\sum_{i}^{n}(X_i - \bar{X})^2$

$\underbrace{}_{\text{Sample mean}}$ $E(\bar{X}) = \mu$; $\underbrace{}_{\text{Sample variance}}$ $Var(\bar{X}) = \frac{\sigma^2}{n}$

If $(X_1 - \bar{X}, X_2 - \bar{X} \ldots X_n - \bar{X})$ are independent

$M(s, t_1 \ldots t_n) = E\{\exp[s\bar{X} + t_1(X_1 - \bar{X}) + \cdots + t_n(X_n - \bar{X})]\}$

$\sum_{i=1}^{n} t_i(X_i - \bar{X}) = \sum_{i=1}^{n} t_i X_i - n\bar{X}\bar{t}$

$s\bar{X} + \sum t_i(X_i - \bar{X}) = \sum_{i=1}^{n}\left[\frac{s}{n} + (t_i - \bar{t})\right]X_i = \sum_{i=1}^{n} a_i X_i$

$M(s, t_1 \ldots t_n) = \exp\left(\mu s + \frac{\sigma^2}{2n}s^2\right)\exp\left[\frac{\sigma^2}{2}\sum_{i}^{n}(t_i - \bar{t})^2\right]$

1. Prove Proposition A of Section 6.2

$\frac{Z}{\sqrt{U/n}} = \frac{N(0,1)}{\sqrt{\chi^2/n}} = \frac{N(0,1)}{\sqrt{\frac{1}{2^{n/2}\Gamma(n/2)} v^{(n/2)-1} e^{-v/2}/n}}$

$f(t^2) = \int \sqrt{\frac{v}{n}}\, N(\sqrt{v/n} \cdot t) f(\chi_n^2)\, dv$

$x = (1 + t^2/n)v/2$

$= \frac{1}{\sqrt{2\pi n}\, 2^{n/2}\Gamma(n/2)} \int_0^\infty v^{(n-1)/2} e^{-(1+t^2/n)v/2}\, dv$

$= \frac{(1+t^2/n)^{-(n+1)/2}}{\sqrt{\pi n}\,\Gamma(n/2)} \int_0^\infty x^{(n+1)/2-1} e^{-x}\, dx$ $\boxed{= \frac{(1+t^2/n)^{-(n+1)/2}}{\sqrt{\pi n}\,\Gamma(n/2)}\Gamma[(n+1)/2]}$

2. Prove Proposition B of Section 6.2

$W = \frac{U/m}{V/n}$

$f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} w^{m/2-1}\left(1 + \frac{m}{n}w\right)^{-(m+n)/2}$

$$f(w)=\int_0^{\infty}\frac{1}{2^{n/2}\Gamma(n/2)}x^{(n/2)-1}e^{-x/2}\cdot\frac{m}{2^{m/2}\Gamma(m/2)}(xz)^{m/2-1}\cdot e^{-xz/2}dx=\frac{z^{m/2-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})2^{(m+n)/2}}\underbrace{\int_0^{\infty}x^{(m+n)/2-1}e^{-x(z+1)/2}dx}$$

"similar to $\Gamma(x)=\int_0^{\infty}t^{x-1}e^{-t}dt$. where $t=x\left(\frac{z+1}{2}\right)$

$$=\frac{z^{m/2-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})2^{(m+n)/2}}\frac{1}{\left(\frac{z+1}{2}\right)^{(m+n)/2}}\int_0^{\infty}t^{(m+n)/2-1}e^{-t}dt$$

$$=\frac{\Gamma(\frac{m+n}{2})z^{m/2-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})(z+1)^{(m+n)/2}}\Rightarrow\boxed{\frac{\Gamma(\frac{m+n}{2})\left(\frac{m}{n}\right)^{m/2}x^{m/2-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})\left(1+\frac{m}{n}x\right)^{(m+n)/2}}}$$

3. $n=16$; $\bar{X}=\mu=0=\frac{1}{16}\sum_{i=1}^{16}X_i$

$P(|\bar{X}|\le c)=P(-c\le\bar{X}\le c)$
$=P(\bar{X}\le c)-P(\bar{X}\le -c)$
$=P(\bar{X}\le c)-[1-P(\bar{X}\le c)]$
$=2\cdot P(\bar{X}\le c)-1=2\cdot\Phi\left(\frac{c-0}{\sqrt{1/16}}\right)-1$
$=2\cdot\Phi(4c)-1=0.5$
$\Phi(4c)=0.75$
$4c=\Phi^{-1}(0.75)$
$c=0.7734/4$
$=0.19335$

4. T follows a t_7-distribution.

Find t_0 such that a) $P(|T|\le t_0)=0.9$ of a t_7 distribution

$$t_7=f(t)=\frac{\Gamma[(7+1)/2]}{\sqrt{7\pi}\,\Gamma(7/2)}\left(1+\frac{t^2}{7}\right)^{-(7+1)/2}$$
$$=\frac{\Gamma(4)}{\sqrt{7\pi}\,\Gamma(7/2)}\left(1+\frac{t^2}{7}\right)^{-4}$$

$=P(-t_0\le T\le t_0)$
$=P(T\le t_0)-P(-t_0\le T)$
$=P(T\le t_0)-[1-P(t_0\le T)]$
$=2P(T\le t_0)-1=0.9$
$P(T\le t_0)=0.95$
$\boxed{T=\Phi^{-1}(0.95)=1.895}$

b) $P(T>t_0)=0.05$
$1-P(T\le t_0)=0.05$; $P(T\le t_0)=0.95$
$T=\Phi^{-1}(0.95)=1.895$

5. If $X\sim F_{n,m}$, then $X^{-1}\sim F_{m,n}$.

$F_{n,m}=W=\frac{U/m}{V/n}=X$; $X^{-1}=\frac{1}{F}=\frac{V/n}{U/m}\boxed{F_{m,n}}$

6. $T\sim t_n$, then $T^2\sim F_{1,n}$; $T=t_n=Z/\sqrt{V/n}$; $T^2=Z^2/V/n=\boxed{nZ^2/U=F_{1,n}}$

7. Cauchy Distribution:
$$\boxed{f(x)=\frac{1}{\pi}\left(\frac{1}{1+x^2}\right)}$$

t-Distribution:
$$f(t)=\frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}$$
@ df=1, $f(t_1)=\frac{\Gamma(1)}{\sqrt{\pi}\,\Gamma(1/2)}(1+t^2)^{-1}$
$$\boxed{=\frac{1}{\pi}\left(\frac{1}{1+t^2}\right)}$$

8 Exponential Random Variable:
$f(x)=\lambda e^{-\lambda x}$; $\lambda=1$; $\frac{X}{Y}=\left(\frac{1}{1}\right)e^{-x+y}=e^{y-x}$

F-Distribution:
$$f(w)=\frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2}w^{m/2-1}\left(1+\frac{m}{n}w\right)^{-(m+n)/2};$$
if $m,n=1$; $f(w)=\frac{\Gamma(1)}{\Gamma(1/2)\Gamma(1/2)}\left(\frac{1}{1}\right)^{1/2}(1+w)^{-1}$
$=1+\frac{w^2}{2}+\frac{w^3}{8}+\cdots$
$=e^{w}=\boxed{e^{y-x}}$

9. Find mean and variance of $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$

$\bar{S}^2=\frac{1}{n}\sum\frac{1}{(n-1)}(X_i-\bar{X})^2=\boxed{\sum_{i=1}^{n}X_i\ \sum_{i=1}^{n}(X_i-\bar{X})^2}$

$[S^2]^2=\frac{1}{(n-1)^2}\sum(X_i-\bar{X})^{2}$

$P(a<S^2/\sigma^2<b)=P(S^2/\sigma^2<b)-P(a<S^2/\sigma^2)$
$$=\int_a^b f(v)\,dv=1=\frac{1}{2^{n/2}\Gamma(n/2)}[\Gamma(b+1)-\Gamma(a+1)]$$
$$=\frac{1}{2^{n/2}\Gamma(n/2)}(b-a)\Gamma(1)=1;\ \boxed{(b-a)=2^{n/2}\sqrt{\pi}}$$

10. Chi-Squared Distribution
$$f(v)=\frac{1}{2^{n/2}\Gamma(n/2)}v^{(n/2)-1}e^{-v/2}$$

10. $X_1 \dots X_n ; N(\mu_x, \sigma^2)$; $Y_1 \dots Y_N ; N(\mu_y, \sigma^2)$; Show how a F-distribution can find $P(S_x^2/S_y^2 > c)$

F Distribution:

$f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \cdot \left(\frac{m}{n}\right)^{m/2} \cdot w^{m/2-1} \cdot \left(1+\frac{m}{n}w\right)^{-(m+n)/2}$; $\int_{-c}^{c} f(w)\,dw = \int_{-c}^{c} \frac{S_x^2}{S_y^2}\,dx\,dy = 2\int_0^c\int_0^c \frac{S_x^2}{S_y^2}\,dx\,dy$

$= \frac{2 \cdot e^{(-2c^2+2\mu_x c+2\mu_y c)/\sigma^2}}{+(1/\sigma^2)^2 \mu_x \mu_y (7)^2} = 1$

$= \frac{+\sigma^4 e^{-2(c^2+\mu_x c+\mu_y c)/\sigma^2}}{2\mu_x\mu_y} = 1$

$\frac{-2(c^2+\mu_x c+\mu_y c)}{\sigma^2} = \ln\left[\frac{2\mu_x\mu_y}{\sigma^4}\right]$

$c^2+\mu_x c+\mu_y c = \sigma^2 \ln\left[\frac{\sigma^4}{2\mu_x\mu_y}\right]$

$c^2+(\mu_x+\mu_y)c-\sigma^2\ln\left[\frac{\sigma^4}{2\mu_x\mu_y}\right] = 0$

$$\boxed{c_1,c_2 = \frac{-(\mu_x+\mu_y) \pm \sqrt{(\mu_x+\mu_y)^2+4\sigma^2\ln\left[\frac{\sigma^4}{2\mu_x\mu_y}\right]}}{2}}$$

Chapter 7: Survey Sampling!

1. Sample: 1, 2, 2, 4, and 8;

$E[X] = \bar{X} = \frac{1}{5}\sum_{i=1}^{5} X_i = \frac{(1+2+2+4+8)}{5} = 17/5$

$E[X^2] = \frac{1}{5}\sum_{i=1}^{5} X_i^2 = \frac{(1^2+2^2+2^2+4^2+8^2)}{5} = \frac{89}{5}$

$Var(X) = \frac{89}{5} - \left(\frac{17}{5}\right)^2 = \frac{89-(10+7)(10+7)/5}{5} = \frac{89-57.8}{5} \approx \frac{21.00}{5}$ $\boxed{4.20}$

Sampling Distribution

Sample size = 2

$T = N\bar{X} = 2(17/5) = \boxed{1.7}$

$E(T) = E(N\bar{X}) = 1.7$

$Var(\bar{X}) = \frac{4.20}{5}\left(\frac{5-2}{5-1}\right) = \frac{12.60}{20} = \frac{1.26}{2} = \boxed{0.63}$

(2) $n = 2$; $X = 4, 8$; $T = N\bar{X} = NE(X) = 1.7$; $Var(\bar{x}) = 0.63$

(3) a) A population mean [No] b) Population Size: [No] c) Sample Size [No] d) The Sample Mean [Yes] e) Variance of Sample mean [No] f) The largest value of Data [Yes] g) Population Variance [No] h) Estimated variance of sample mean [Yes]

(4) Population I: n_1, σ_1 Population II: $n_2 = 2n_1$, $\sigma_2 = 2\sigma_1$

Accuracy is better approximated by a large n-value. The law of large numbers states a sequence of independent values converge to $E(\bar{X}_n)$ as $n \to \infty$.

(5) A random variable is defined as a variable which can take on only a finite number of values. The sample mean is a random variable because of the finite form.

(6) $N_1 = 100{,}000$; $N_2 = 10{,}000{,}000$; $n = 25$. $var(\bar{X}) = \frac{\sigma^2}{25}\left(\frac{100{,}000-25}{100000-1}\right) = 0.04\sigma^2$

Yes, it is substantially easier to measure the smaller size because of finite solution to sample mean. $var(\bar{x}) = \frac{\sigma^2}{25}\left(\frac{10^7-25}{10^7-1}\right) = 0.04\sigma^2$

(7). Standard Error: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$; $Var(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{15}{100}\right) \Rightarrow \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\frac{\sqrt{15}}{10} = 0.02$; $\frac{\phi}{\sqrt{n}} = \frac{0.2}{\sqrt{15}}$; $\frac{15}{0.04} = n = 375$

8. $n=100$; $p=1/5$ a) Find δ such that $P(|\hat{p}-p|\geq\delta)=0.025$

Sample proportion $\hat{p}$; standard error $\sigma_{\hat{p}}=\sqrt{\frac{p(1-p)}{n}}=\sqrt{\frac{0.2(1-0.2)}{50}}=0.057$ ‖ $P\left(\frac{|p-\hat{p}|}{\sigma_p}\geq\frac{\delta}{\sigma_p}\right)=0.025$

$P\left(\frac{|p-\hat{p}|}{\sigma_p}\geq\frac{\delta}{\sigma_p}\right)=1-0.025$

$2P\left(z\leq\frac{\delta}{0.057}\right)-1=0.975$

$P\left(z\leq\frac{\delta}{0.04}\right)=\frac{1.975}{2}$

$\Phi\left(\frac{\delta}{0.04}\right)=0.9875$

$\frac{\delta}{0.04}=\Phi^{-1}(0.9875)$

$\delta=0.08964$

b) $\hat{p}=0.25$; 95% ;

$p=\hat{p}\pm z(0.025)s_{\hat{p}}=0.25\pm1.96\sqrt{\frac{0.25(1-0.25)}{100-1}}$

$=0.25\pm0.0853$

$=(0.1647, 0.3353)$

The original $p=0.2$ is within the range of proportion and at the true population.

9. $n=1,500$ voters, 55% planned to vote a particular proposition. $\boxed{\text{Standard Error}}$ $\sigma_{\bar{x}}\cong\frac{\sigma}{\sqrt{n}}$

45% planned to vote against a proposition.

Margin of victory [10%]. $\boxed{\text{Confidence Interval}}$

$\bar{X}=50\%$; $\bar{X}\pm z(0.025)\sigma_{\bar{x}}=10\%\pm1.96(2.66\%)=(9.49\%, 10.51\%)\cong\frac{DA\%}{\sqrt{1500}}=0.26\%$

$\boxed{=(4.90\%, 15.16\%)}$

10. False, as a population grows($n\to\infty$), then a distinct mean(μ) and standard deviation(σ) become more distinct, and possibly less normal and more dichotomous.

11. $n=4$; X_1, X_2, X_3, X_4 a) $\binom{n}{k}=\binom{4}{2}=\frac{4!}{2!(4-2)!}=\frac{4\cdot3}{2}=\boxed{6}$ b) $\{X_1,X_2\},\{X_2,X_3\},\{X_3,X_4\},\{X,X_4\}$

Mean Square error = variance + bias^2 ; $E[X]=\frac{1}{6}$; $E[X^2]=\frac{1}{6}$; $\sigma=\sqrt{\frac{1}{6}-\frac{1}{6^2}}=\sqrt{\frac{5}{36}}=\frac{\sqrt{5}}{6}$

$\frac{1}{n}\sum_i(X-\mu)^2=\sigma^2+\beta^2$ ‖ This case shows the sample mean is unbiased because the choices are

$\frac{1}{8}\sum X=\frac{12}{8}=\frac{1+2+3+1+4+4}{4}$

12. Random Sampling with replacement

$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ is the unbiased parameter of σ^2. $\boxed{\text{Variance of a Biased Sample}}$

$Var(\bar{X})=\frac{1}{n^2}\sum_i\sum_j Cov(X_i,X_j)=\frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}Cov(X_i,X_j)=\frac{\sigma^2}{n}-\frac{1}{n^2}n(n-1)\frac{\sigma^2}{N-1}$

$=\frac{\sigma^2}{n}\left[1-\frac{(n-1)}{N-1}\right]$

$=\frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right]$

$\boxed{\text{Expected Variance of a population}}$

$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2$; $E(\hat{\sigma}^2)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^2)-E(\bar{X}^2)$

$=\frac{1}{n}\sum_i\left[Var(X_i)+E(X)^2\right]-\left[Var(\bar{X})+E(\bar{X})^2\right]$

$=\frac{1}{n}\sum_i\left[\sigma^2+\mu^2\right]-\left[\frac{\sigma^2}{n}\left(1-\frac{n-1}{N-1}\right)+\mu^2\right]$

$=\frac{\sigma^2}{n}\sum_i\left[1-\left(1-\frac{n-1}{N-1}\right)\frac{1}{n}\right]=\frac{\sigma^2}{n}\left[1-\frac{1}{n}+\frac{(n-1)}{n(N-1)}\right]=\frac{\sigma^2}{n}\left[\frac{n(N-1)-(N-1)+(n-1)}{n(N-1)}\right]$

$=\frac{\sigma^2}{n}\left[\frac{nN-\cancel{n}-N+1+\cancel{n}-1}{n(N-1)}\right]=\boxed{\frac{\sigma^2}{n}\left[\frac{(n-1)N}{n(N-1)}\right]}$

$$s_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{n}{n+1}\right)\left(\frac{N-1}{N}\right)\left(\frac{N-n}{N-1}\right) + \left[\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)\right]^2 \left(\frac{1}{n}\right)\left(\frac{n}{n-1}\right)\left(\frac{N-1}{N}\right)\left(\frac{N-n}{N-1}\right)$$

$$= \frac{\sigma^4}{n^2}\left(\frac{N-n}{N-1}\right)^2\left(\frac{1}{n}\right)\left(\frac{n}{n-1}\right)\left(\frac{N-1}{N}\right)\left(\frac{N-n}{N-1}\right) = \frac{\sigma^4}{n^3}\left(\frac{N-n}{N-1}\right)\left(\frac{1}{n-1}\right)\left(\frac{N-n}{N}\right)$$

12 ?

(12) $s^2 = \frac{1}{n-1}\sum(x_i-\bar{x})^2$; $E(s^2) = \frac{1}{n-1}E\left(\sum^n x_i^2 - n\bar{x}^2\right) = \frac{n}{n-1}\left[\sum E(x^2) - E(x)^2\right]$

$$= \frac{1}{n-1}\left[\sum\left[Var(x_i) + E(y_i)^2\right] - n\left[Var(y) + E(y)^2\right]\right]$$

$$= \frac{1}{n-1}\left[n\sigma^2 + n\bar{y}^2 - \sigma^2 - ny^2\right] = \frac{(n-1)}{(n-1)}\sigma^2 = \sigma^2$$

b) $E(s) = E\left(\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}\right) = \frac{1}{\sqrt{n-1}}E\left(\sqrt{\frac{\sum(x-\bar{x})^2}{\sigma}}\,\sigma\right) = \frac{\sigma^2}{\sqrt{n-1}}E\left(\sqrt{\frac{\sum(x-\bar{x})^2}{\sigma^2}}\right) = \frac{\sigma^2}{\sqrt{n-1}}E(\sqrt{y})$

$\boxed{E(s) \neq \sigma \text{ ; It is not an unbiased estimate of } \sigma}$

c) Show $\frac{s^2}{n}$ is an unbiased estimate of $\sigma_{\bar{x}}$;

$$E\left(\frac{s^2}{n}\right) = E\left(\frac{1}{n(n-1)}\left(\sum x_i^2 - n\bar{x}^2\right)^2\right) = \frac{1}{n(n-1)}\sum E(x^2) - E(x)^2 =$$

$$= \frac{1}{n(n-1)}\left[\sum\left[Var(X_i) + E(y_i)^2\right] - n\left[Var(y) + E(y)^2\right]\right]$$

$$= \frac{1}{n(n-1)}\left[n\sigma^2 + n\cancel{\bar{y}^2} - \sigma^2 - \cancel{n\bar{y}^2}\right] = \frac{\sigma^2(n-1)}{n(n-1)}$$

Yes, $E\left(\frac{s^2}{n}\right)$ is an unbiased estimate, being $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$; $\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$

d) $E\left(\frac{N^2 s^2}{n}\right) = \frac{N^2\sigma_T^2}{n}$

$\boxed{\sigma_T^2 = N^2\sigma_x^2}$

These are expected estimates of the sample and population, separately.

e) $E\left(\frac{\hat{p}(1-\hat{p})}{(n-1)}\right) = \frac{1}{n-1}\left[E(\hat{p}) - E(\hat{p}^2)\right] = \frac{1}{n-1}\left[p - \left[Var(\hat{p}) + E(\hat{p})^2\right]\right]$

$$= \frac{1}{n-1}\left[p - \left[\frac{p(1-p)}{n} + \hat{p}^2\right]\right] = \frac{1}{n-1}\left[p(1-p) - \frac{p(1-p)}{n}\right] = \frac{n-1}{n-1}\,\frac{p(1-p)}{n}$$

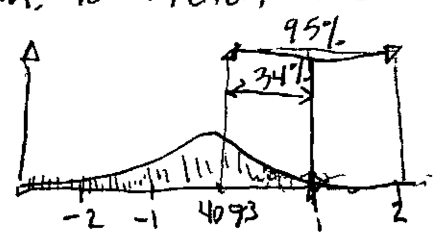
13.

Example A.7.2:

Herksen (1976); $N = 393$; x_i = number of patients discharged from i^{th} hospital January 1968. $\boxed{\neq \sigma_p^2}$

Suppose Total $[\tau]$ is an estimate of size 50.

Denoting estimate T by the Central Limit theorem to sketch the probability density of the error $T-\tau$



$T = N\bar{X}$, $\sigma_T^2 = N^2\sigma_x^2$ $s_T = N^2 s_x^2$

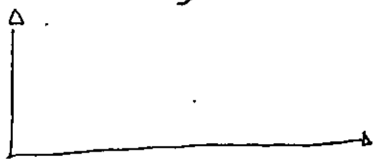
$\bar{x} = 814.6$; $\sigma_x^2 = \frac{\sigma^2}{n} = \frac{590^2}{393} = 886$

$s_x = 81.19$

$T = 50 \cdot 814.6$; $\sigma_T^2 = 221,500$; $s_T = 1647,954$
$= 4073$

Huge Standard error with small total population

14. $p = 0.654$ Total Number (<1000) discharges is from $n = 25$. Apply central limit theorem to the distribution



$n = 393$, $\mu = 814.6$, $\sigma = 590$

$\sigma = \sqrt{\dfrac{0.654(1-0.654)}{25}} = \boxed{0.095}$

15. a) n = simple random sample.

a) Sketch $P(|\bar{X}-\mu| > 200)$, $20 < n \leq 100$

$P(-200 < \bar{X}-\mu < 200) = 1 - P(|\bar{X}-\mu| \leq 200)$

$= 1 - P(-200 \leq \bar{X}-\mu \leq 200)$

$= 1 - P\left(\dfrac{-200}{\sigma_{\bar{X}}} \leq \dfrac{\bar{X}-\mu}{\sigma_{\bar{X}}} \leq \dfrac{200}{\sigma_{\bar{X}}}\right)$

$= 1 - \left[\Phi\left(\dfrac{200}{\sigma_{\bar{X}}}\right) - \Phi\left(\dfrac{-200}{\sigma_{\bar{X}}}\right)\right]$

$= 1 - \left[\Phi\left(\dfrac{200}{\sigma_{\bar{X}}}\right) - 1 + \Phi\left(\dfrac{200}{\sigma_{\bar{X}}}\right)\right]$

$= 2 - \left[2\Phi\left(\dfrac{200}{\sigma_{\bar{X}}}\right)\right] = 2\left[1 - \Phi\left(\dfrac{200}{\sigma_{\bar{X}}}\right)\right]$

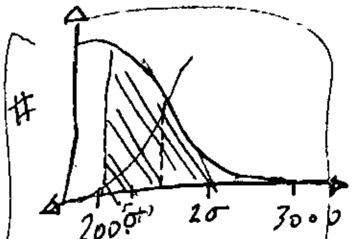
$= 2\left[1 - \Phi\left(\dfrac{200}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-1}{N-n}}}\right)\right]$

$= 2\left[1 - \Phi\left(6.71\sqrt{\dfrac{n}{393-n}}\right)\right]$

@ $n = 20$; $= 2\left[1 - \Phi\left(6.71\sqrt{\dfrac{20}{393-20}}\right)\right]$

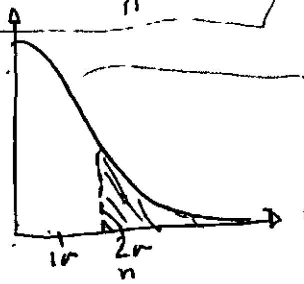
$= 2[1 - \Phi(1.55)] = 0.12$

@ $n = 100$ $= 2[1 - \Phi(3.92)] = 0.002$



#

200 σ 2σ 300
n



#

1σ 2σ
n

$P(-200 < \bar{X}-\mu < 200)$

b) For $n = 20, 40$, and 80. Find Δ such that $P(|\bar{X}-\mu| > \Delta) \approx 0.10$

$n = 20$; $P(|\bar{X}-\mu| > \Delta) \approx 0.1$

$P(\Delta < \bar{X}-\mu < \Delta) \approx 0.1$

$P(\bar{X}-\mu < \Delta) - [1 - P(\bar{X}-\mu < \Delta)] \approx 0.1$

$2\left[1 - \Phi\left(\dfrac{\Delta}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-1}{N-n}}}\right)\right] \approx 0.1$

$\Phi\left(\dfrac{\Delta}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-1}{N-n}}}\right) \approx 0.95$

$\dfrac{\Delta}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-1}{N-n}}} = 1.65$

$\Delta_{0.1} = \dfrac{590}{\sqrt{20}}\sqrt{\dfrac{393-1}{393-20}}(1.65) = 222$

$\Delta_{0.5} = \dfrac{590}{\sqrt{20}}\sqrt{\dfrac{393-1}{393-20}}(0.68) = 91.9$

$n = 40$ $P(|\bar{X}-\mu| > \Delta) \approx 0.10 = 1 - P(|\bar{X}-\mu| < \Delta)$

$= 1 - P(-\Delta < |\bar{X}-\mu| < \Delta) = 1 - [\Phi(\bar{X}-\mu < \Delta) - \Phi(-\Delta < \bar{X}-\mu)] = 1 - [\Phi(\bar{X}-\mu < \Delta) - 1 + \Phi(-\Delta < \bar{X}-\mu)]$

$= 2 - [2\Phi(\bar{X}-\mu < \Delta)] = 2[1 - \Phi(\bar{X}-\mu < \Delta)] = 0.1$; $\Phi\left(\dfrac{\bar{X}-\mu}{\sigma_{\bar{X}}} < \dfrac{\Delta}{\sigma_{\bar{X}}}\right) = 0.95$

$\Delta_{0.1} = \dfrac{590}{\sqrt{40}}\sqrt{\dfrac{393-1}{393-40}}(1.65) = 162$

$\Delta_{0.5} = \dfrac{590}{\sqrt{40}}\sqrt{\dfrac{393-1}{393-40}}(0.68) = 66.8$

$n = 80$

$\Delta_{0.1} = 122$
$\Delta_{0.5} = 50$

16.
a) The mean is a random variable.
b) True, a 95% confidence interval contains the mean and has a total probability of 0.95
c) True, a 95% confidence interval contains 95% of the population
d) True, 95% of a 100 is 95.

17. A 90% confidence interval $(1-\alpha)$ for average number of children per household is (min = 0.7, max = 2.1). Yes a confidence interval describes a random interval from a lower and upper bound that contains the mean.

18. 90% Confidence Interval: $P(\text{Mean} \mid \text{Interval}) = \sum_{0.9}^{n} p^{n} = 90\%^2 = \boxed{81\%}$

$1 - P(\text{Mean} \mid \text{Interval}) = P(\text{No mean} \mid \text{Interval}) = 1 - 0.81 = \boxed{19\%}$

19. One-sided Confidence Interval

k be chosen $(-\infty, \bar{X} + k s_{\bar{X}})$ that a 90% confidence interval for μ

$P(-\infty \le \mu \le \bar{X} + k s_{\bar{X}}) = 90\%$; $P(\mu \le \bar{X} + k s_{\bar{X}})$; $1.28 = \bar{X} + k s_{\bar{X}}$; $\boxed{k = \frac{1.28 - \bar{X}}{s_{\bar{X}}}}$

$P(\bar{X} - k s_{\bar{X}} \le \mu) = 0.95$; $\bar{X} - k s_{\bar{X}} = 1.65$; $k = \frac{\bar{X} - 1.65}{s_{\bar{X}}}$

20. $N = 8000$ condominium units; $n = 100$ sample size; $\bar{X} = 1.6$ motor vehicles; $\sigma_{\bar{X}} = 0.8$

$s_{\bar{X}} = \frac{s}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} = \frac{0.8}{10} \sqrt{1 - \frac{100}{8000}} = 0.08$; confidence interval $\bar{X} \pm z(0.025) s_{\bar{X}} = \bar{X} \pm 1.96(0.08) = (1.44, 1.76)$

- Total Number of Motor Vehicles $T = 8000 \times 1.6 = 12,800$; $s_T = N s_{\bar{X}} = 640$ $= (11546, 14054)$

- 12% respondents planned $\hat{p} = 0.12$ with a proportion p.

Standard Error: $s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}} \sqrt{1 - \frac{100}{1000}} = 0.03$

$\hat{p} \pm 1.96 s_{\hat{p}} = (0.06, 0.18)$

$T = N\hat{p} = 960$; $s_T = N s_{\hat{p}} = 240$

$T \pm 1.96 s_T$; $(490, 1490)$

At 95% confidence interval suggests another sample size of 100 would contain a mean between (1.44 and 1.76

21. To halve the width of a 95% confidence interval

$\frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = 95\%$; $\frac{\bar{X} - \mu}{\sigma/\sqrt{4}} = 95\%$; $\frac{\bar{X} - \mu}{\sigma/2} = \frac{\bar{X} - \mu}{\sigma}\left(\frac{1}{2}\right) = 95\%/2 = 47.5\%$

22. $\bar{X} \pm s_{\bar{X}} = \bar{X} \pm 1 s_{\bar{X}} \Rightarrow z(\alpha/2) = 1$; $\alpha = 2(1 - 0.8413)$; Confidence Interval: $\boxed{1 - \alpha = 0.682}$

23 a) Show $\sigma_{\bar{X}}$ is largest when $p = \frac{1}{2}$; $\frac{d}{dp} s_{\hat{p}} = \frac{d}{dp} \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}} \sqrt{1 - \frac{100}{8000}} = 0$

$= \frac{1}{2} \cdot \frac{1 - 2p}{\sqrt{\hat{p}(1-\hat{p})/(n-1)}} = \frac{1}{2}(1 - 2p) = \frac{1}{2} - p$

$\boxed{p = 1/2}$

b) $s_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{n-1}\left(1 - \frac{n}{N}\right)$ "Unbiased Estimate of $Var(\hat{p})$"

$s_{\hat{p}}^2 = \frac{\frac{1}{2}(1 - 1/2)}{n-1}\left(1 - \frac{n}{N}\right)$

$s_{\hat{p}}^2 = \frac{1}{4}\left(\frac{N-n}{N(n-1)}\right)$; $s_{\hat{p}} = \sqrt{\frac{1}{4}\left(\frac{N-n}{N(n-1)}\right)}$ $\boxed{\frac{1}{2}\sqrt{\left(\frac{N-n}{N(n-1)}\right)}}$

c) $\hat{p} \pm \sqrt{\frac{N-n}{N(n-1)}} \to \Phi(z) = \Phi(0.05)$

$P\left(\hat{p} - \sqrt{\frac{N-n}{N(n-1)}} < \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} < \hat{p} + \sqrt{\frac{N-n}{N(N-1)}}\right) \to \Phi(z)$

$\lim_{n \to \infty} P\left(\hat{p} - \sqrt{\frac{N-n}{N(n-1)}} < \frac{\bar{X} - \mu}{\sigma \cdot \sqrt{n}} < \hat{p} + \sqrt{\frac{N-n}{N(n-1)}}\right)$

$P(\hat{p} < 0 < \hat{p}) = \Phi(z) = 2P(0 \le \hat{p}) - 1 = \Phi(z)$; $z = \boxed{0.999}$

24. Sample size $= n$; Population size $= N$; Estimate of $\mu = \bar{X}_c = \sum_{i=1}^{n} c_i X_i$

a) Find the condition $[c_i]$ such that the estimate is unbiased.

$$\bar{X} = E\left[\sum_{i=1}^{n} c_i X_i\right] = \sum_{i=1}^{n} c_i E(X_i) = \mu \sum_{i=1}^{n} c_i = \mu(1) \quad \boxed{\sum c_i = 1}$$

b) $Var(\bar{X}_c) = Var(\sum c_i X_i) = \sum c_i^2 Var(X) = \sum c_i^2 \sigma^2 = \sigma^2 \sum c_i^2$

Applying a LaGrangian Multiplier: $L(c_1 \ldots c_n, \lambda) = \sigma^2 \sum_{i=1}^{n} c_i^2 + \lambda(\sum c_i - 1)$

$\frac{\partial L}{\partial c_i} = 0$; $\frac{\partial}{\partial c_i}\left[\sigma^2 \sum_{i=1}^{n} c_i^2 + \lambda(\sum_{i=1}^{n} c_i - 1)\right] = 0$; $\frac{d}{dc_i}\left[\sigma^2 \sum c_i^2\right] + \frac{d}{dc_i}\left[\lambda(\sum c_i - 1)\right] = 0$

$\sigma^2 \frac{\partial}{\partial c_i}\left[\sum c_i^2\right] + \lambda \frac{\partial}{\partial c_i} \sum c_i = 0$

Therefore, $\sum c_i = 1 \sum \frac{-\lambda}{2\sigma^2} = -\frac{n\lambda}{2\sigma^2} = 1$ $\quad 2\sigma^2 \sum c_i^{2-1} + \lambda = 0$; $2\sigma^2 c_i + \lambda = 0$

$\boxed{\lambda = \frac{-2\sigma^2}{n}}$ $\quad 2\sigma^2 c_i = -\lambda$

$c_i = \frac{-1}{2\sigma^2}\left(\frac{-2\sigma^2}{n}\right)$ $\quad Cov(X_i, X_j) = -\sigma^2/(N-1)$ $\quad c_i = \frac{-\lambda}{2\sigma^2}$

$= E(X_i X_j) - E(X_i)E(X_j)$

25. Lemma B $\boxed{= 1/n}$

Section 7.3.2:

$$E(X_i X_j) = \sum_{k=1}^{m} \sum_{l=1}^{m} \zeta_k \zeta_l P(X_i = \zeta_k \text{ and } X_j = \zeta_l) = \sum_{1}^{m} \zeta_k P(X_i = \zeta_k) \sum_{1}^{m} \zeta P(X_j | X_i)$$

where $P(X_j | X_i) = \begin{cases} n_l/(N-1) & k \neq l \\ (n_l - 1)/(N-1) & k = l \end{cases}$

$$\sum \zeta_l P(X_j | X_i) = \sum_{l \neq k} \zeta_l \frac{n_l}{N-1} + \zeta_k \frac{n_k - 1}{N-1}$$

$$= \sum_{l=1}^{m} \zeta_l \frac{n_l}{N-1} - \zeta_k \frac{1}{N-1}$$

$$E(X_i X_j) = \sum_{k=1}^{m} \zeta_k \frac{n_k}{N}\left(\sum \zeta_l \frac{n_l}{N-1} - \frac{\zeta_k}{N-1}\right) = \frac{1}{N(N-1)}\left(\tau^2 - \sum_{k=1}^{m} \zeta^2 n_k\right)$$

$$= \frac{\tau^2}{N(N-1)} - \frac{1}{N(N-1)} \sum_{k=1}^{m} \zeta^2 n_k = \frac{N\mu^2}{N-1} - \frac{1}{N-1}(\mu^2 + \sigma^2)$$

$$= \mu^2 - \frac{\sigma^2}{N-1}$$

26. $U_i = 1$; i^{th} population member $\quad Cov(X_i X_j) = \mu^2 - \frac{\sigma^2}{N-1} - \mu^2 = \frac{-\sigma^2}{N-1}$ $\quad \| \; Cov(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i) \cdot E(Y_j)$

$= E(Y_i Y_j) - \sqrt{E(Y_i^2) - Var(Y_i)} \sqrt{E(Y_j^2) - Var(Y_j)}$

a) show $\bar{X} = n^{-1} \sum_{i=1}^{N} U_i X_i$; $E(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} B$ $\boxed{\frac{1}{n} \sum_{i=1}^{N} U_i X_i}$

b) $P(U_i = 1) = n/N$; $E(U_i) = \frac{1}{N} \sum_{i=1}^{n} U_i P(U_i = 1) = \frac{n}{N}(1) = \boxed{\frac{n}{N}}$

c) $Var(U_i) = E[U_i^2] - E[U_i]^2 = \frac{1}{N^2} \sum_{i=1}^{n} U_i^2 P(U_i = 1) - \left[\frac{n}{N} \sum U_i P_0\right]^2$

$= \frac{n}{N} - \left(\frac{n}{N}\right)^2 = \boxed{\frac{n}{N}\left(1 - \frac{n}{N}\right)}$

d) $E(U_i U_j) = \frac{\sum_i^n \sum_j^n U_i U_j P(U_i, U_j)}{N^2} = \frac{n^2}{N^2}$

e) $Cov(U_i, U_j) = E(U_i U_j) - E(U_i)E(U_j) = \frac{n^2}{N^2} - \left(\frac{n}{N}\right)\left(\frac{n}{N}\right) = 0$

f) $Var(\bar{X}) = E(\bar{X}^2) - E(\bar{X})^2 = \frac{n}{N^2} - \frac{n}{N^2} = 0$

27. Population size (N) is unknown ; $n \le N$. Show will generate a simple random sample.

a) List $\{X_1, X_2 \ldots X_n \ldots X_N\}$

$n = \{X_1, X_2 \ldots X_n\}$

$N = \{X_1, X_2 \ldots\ldots X_n\}$

b) For $k = 1, 2 \ldots$; $k < N$

i) $X_{(n+k)} = X_{i-1+k} + X_i$; $X_{i=1} = X_1$

ii) $\dfrac{n}{(n+k)} = \dfrac{X_1, X_2 \ldots X_n}{X_1, X_2 \ldots\ldots X_{n+k}}$

28. Randomized Response: $\underbrace{\text{Spin an Arrow - Draw ball from urn.}}_{\text{Action}}$ $\underbrace{\text{Choice A} \mid \text{choice B}}_{\text{Result}}$ } Interviewee

Records Response } Interviewer

R = Proportion Yes ; $p = P(\text{Response}_{\text{Yes}} \mid \text{Statement \#1})$ "Randomized Device"

$r = P(\text{Yes})$ $\quad q$ = proportion of Characteristic A. $= P(\text{Statement \#1})$

a) Show $r = (2p-1)q + (1-p)$; Hint: $P(\text{yes}) = P(\text{yes} \mid \text{Statement \#1}) \times P(\text{Statement \#1}) + P(\text{yes} \mid \text{Statement \#2}) P(\#2)$

$r = p \cdot q + (1-p)(1-q)$

b) If r, what is q?

$$\boxed{q = \frac{r-1+p}{(2p-1)}}$$

$2pq - q + 1 - p = (1-p)(1-q)$

$1 + pq - p - q$

c) $E(R) = r$ and propose Q, for q. Show the expected estimator unbiased.

$E(R) = \sum_{i=1}^{2} P(\text{yes} \mid \text{Statement \#}i) P(\text{Statement \#}i) = P(\text{yes} \mid \text{Statement \#1}) P(\text{Statement \#1}) + P(\text{yes} \mid \text{Statement \#2}) P(\#2)$

$= r$

$E(Q) = E\left(\dfrac{R-(1-p)}{2p-1}\right) = \dfrac{E(R) - (1-p)}{2p-1} = \dfrac{r-1+p}{2p-1} = q$

d) Show $\text{Var}(R) = \dfrac{r(1-r)}{n} = E[R^2] - E[R]^2 = \dfrac{1}{(2p-1)^2} \text{Var}(R - (1-p)) = \dfrac{1}{(2p-1)^2} \dfrac{r(1-r)}{n}$

e) $\text{Var}(Q) = E[Q^2] - E[Q]^2 = \dfrac{1}{(2p-1)^2} \dfrac{r(1-r)}{n}$

29. a. $P(\text{yes} \mid \text{Statement \#3}) \quad P(\text{Statement \#2})$

$r = qp + t(1-p)$

b) $E(Q) = \boxed{\dfrac{E(R) - t(1-p)}{p} = q}$

c) $\text{Var}(Q) = \dfrac{r(1-r)}{np^2} = \dfrac{[qp + t(1-p)](1 - qp + t(1-p))}{np^2} = \dfrac{qp - q^2p^2 + qpt(1-p) + t(1-p) - qpt(1-p) + t^2(1-p)^2}{np^2}$

30. Problem #28 ; $\text{Var}(Q) = \dfrac{r(1-r)}{(2p-1)^2 n}$ } Difference

~~Problem #29~~ $\text{Var}(Q) = \dfrac{r(1-r)}{np^2}$

31. $N = 8000$ condominium units ; $n = 100$ sample size ; $\bar{X} = 1.6$ motor vehicles

$s_{\bar{X}} = \frac{s}{\sqrt{n}}\sqrt{1-\frac{n}{N}} = \frac{0.8}{\sqrt{100}}\sqrt{1-\frac{100}{8000}} = 0.08$ } Standard error $\sigma = 0.8$ motor vehicles

confidence interval { $\bar{X} \pm 1.96 s_{\bar{X}} = (1.44, 1.76)$; $T = 8000 \times 1.6 = 12{,}800$ } Total

$\underbrace{s_T = N s_{\bar{X}} = 640}_{\text{Total standard error}}$; $T \pm 1.96 s_T$ $(11546, 14054)$ } Interval of total

12% planned to sell their cars $[\hat{p} = 0.12]$; $s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\sqrt{1-\frac{100}{8000}} = 0.03$

$\hat{p} \pm 1.96 s_p (0.06, 0.18)$; $T = N\hat{p} = 960$; $s_T = N s_p = 240$; $T \pm 1.96 s_T$ $(490, 1430)$

What is the sample size for 95% confidence interval to have 500 width (ΔT) of 500?

$T + 1.96 s_T - T + 1.96 s_T = 2 \cdot 1.96 s_T = 500$; $s_T = 500/2 \cdot 1.96 = 127.55$

$= N s_p = 8000 s_p$; $s_p = \frac{127.55}{8000} = 1.59 \times 10^{-2} = \sqrt{\frac{0.12(1-0.12)}{n-1}}\sqrt{1-\frac{n}{8000}}$

$2.53 \times 10^{-4} = \frac{0.12(0.88)}{n-1}\left(\frac{8000-n}{8000}\right) \Rightarrow 2.53\times10^{-4} n - 2.53\times10^{-4} = \frac{66}{625}\left(\frac{8000-n}{8000}\right)$

$2.53\times10^{-4} n - 2.53\times10^{-4} = \frac{66}{625} - \frac{33}{250000} n$; $2.66\times10^{-4}(n) = 0.1058$; $\boxed{n = 397.9}$

32. $N = 12{,}000$ units ; $n = 200$ sample size ; $\hat{p} = 0.18$

a) What is $s_{\hat{p}}$? $s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\sqrt{1-\frac{n}{N}} = \sqrt{\frac{0.18(1-0.18)}{200-1}}\sqrt{1-\frac{200}{12000}} = 0.027$

$P\left(\frac{|p-\hat{p}|}{s_p} < z(\alpha/2)\right) = 1-\alpha = 0.90$; $\alpha = 0.05$; $z(0.1/2) = z(0.05) = z(1-0.05) = z(0.95) = 1.65$

$P(|p-\hat{p}| \le s_{\hat{p}} \cdot 1.96) = 0.95$; $P(-1.96 s_{\hat{p}} \le |p-\hat{p}| \le 1.96 s_{\hat{p}}) = P(\hat{p} - 1.96 s_p \le p \le \hat{p} + 1.96 s_{\hat{p}})$

$= P(p \le \hat{p} + 1.96 s_{\hat{p}}) - P(\hat{p} - 1.96 s_p \le p) = P(p \le \hat{p} + 1.96 s_{\hat{p}}) - (1 - P(p \le \hat{p} + 1.96 s_p))$

$= 2P(p \le \hat{p} + 1.96 s_p) - 1 = 0.95$; $P(p \le \hat{p} + 1.96 s_p) = \frac{1.95}{2} = 0.975 \rightarrow$ solve

... or ... $(\hat{p} - 1.65 s_p, \hat{p} + 1.65 s_p) = (0.18 - 1.65 \cdot 0.027, 0.18 + 1.65 \cdot 0.027)$
$\boxed{= (0.135, 0.225)}$

b) $\hat{p}_1 = 0.12$, $\hat{p}_2 = 0.18$; $\hat{d} = \hat{p}_1 - \hat{p}_2$

$Var(\hat{d}) = s_p^2 = E[\hat{d}^2] - E[\hat{d}]^2 = 1$

$= E[(\hat{p}_1 - \hat{p}_2)^2] - E[\hat{p}_1 - \hat{p}_2]^2$

$= E[\hat{p}_1^2] - 2E[\hat{p}_1\hat{p}_2] + E[\hat{p}_2^2] - E[\hat{p}_1 - \hat{p}_2]^2$

$= \hat{p}_1 - 2\hat{p}_1\hat{p}_2 + \hat{p}_2 - (\hat{p}_1 - \hat{p}_2)^2$

$= \hat{p}_1 - 2\hat{p}_1\hat{p}_2 + \hat{p}_2 - \hat{p}_1^2 + \cancel{2\hat{p}_1\hat{p}_2} - \hat{p}_2^2$

$\boxed{= p_1(1-\hat{p}_1) + p_2(1-\hat{p}_2)}$

Standard Error (unbiased):

$s_{\bar{X}}^2 = \frac{1}{n}\sum E(X_i^2) - E(\bar{X}^2) = \frac{1}{n}\sum[Var(X) + E(X)^2] - [Var(\bar{X}) + E(\bar{X})^2]$

$= \frac{1}{n}\left[p_1(1-p_1) + p_2(1-p_2) + E(\hat{d})^2\right] - \left[\frac{1}{n^2}\sum Var(X_i) + \frac{1}{n^2}\sum\sum Cov(X_i X_j) + E(X)^2\right]$

$= \frac{p_1(1-p_1) + p_2(1-p_2)}{n}\left[1 - \frac{1}{n^2}[E(P_1P_2) - E(P_1)E(P_2)]\right]$

$= \frac{p_1(1-p_1) + p_2(1-p_2)}{n}\left[1 - \frac{1}{n^2}\frac{N}{n^2}\frac{(n-1)}{(N-1)}\right]$

$\boxed{= \frac{p_1(1-p_1) + p_2(1-p_2)}{n}\left[1 - \frac{N}{n^4}\frac{(n-1)}{(N-1)}\right]}$

①. 99% confidence Interval: $\hat{d} + \hat{z}\left(\frac{1+0.01}{2}\right) s_x < \hat{d} < \hat{d} + \hat{z}\left(\frac{1-0.01}{2}\right) s_x$

$\hat{d} - \hat{z}\,2.57\left(\sqrt{\frac{0.12(1-0.12)+0.18(1-0.18)}{100}}\right)\left[1 - \frac{4000}{100^4}\frac{(100-1)}{(4000-1)}\right]$

$-0.06 - 2.57(5.03\times10^{-2}) < d < \frac{3}{50} + 2.57(5.03\times10^{-2})$

$-\frac{3}{25} < d < 6.927\times10^{-2}$

95% confidence Interval:

$-0.06 - 1.96(5.03\times10^{-2}) < d < -0.06 + 1.96(5.03\times10^{-2})$

$-0.159 < d < 3/25$

90% confidence Interval:

$-0.06 - 1.65(5.03\times10^{-2}) < d < -0.06 + 1.65(5.03\times10^{-2})$

$-0.143 < d < 0.023$

No, there is little difference for a 99% confidence interval ranging from $\left(\frac{-3}{25}, \frac{6.9}{100}\right)$

33. n = simple random samples, two proportions, $\hat{p}_1$ and $\hat{p}_2 \approx 0.5$

What should the sample size be for $\hat{p}_1 - \hat{p}_2 < 0.02$?

standard error of proportions: $s_{\hat{p}_1,\hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} < 0.02$

$p_1(1-p_1) + p_2(1-p_2) = 0.0004 \cdot n$

$\frac{2\,p_1(1-p_1)}{0.0004} = \frac{2\cdot0.5\cdot0.5}{0.0004} = \boxed{1.25\times10^3}$

34. P(Problem #1) = 3% population

P(Problem #2) = 40% population

a) $s_{\hat{p}_1\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_2} + \frac{p_2(1-p_2)}{n_2}}$

Same population size

$s_{\hat{p}_1,\hat{p}_2} = 0.01 > \sqrt{\frac{0.03(0.97)+0.4(0.6)}{n}}$

2 $n > \frac{2.691\times10^{-1}}{1\times10^{-4}}$

$\boxed{n > 2691}$

b) $n > \frac{2.691\times10^{-1}}{0.01}$

$n > 26.91$

c)

35. N = Population Size = 2000

with n = 25 values.

104	109	111	109	87
86	80	119	88	122
91	103	99	108	96
104	98	98	83	107
79	87	94	92	97

a) Calculate unbiased estimate of population mean.

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{25}\left[104 + \cdots + \cdots + 97\right] = \boxed{98}$

b) Calculate unbiased estimate of population variance

$\sigma_{\bar{x}}^2 = \left(1 - \frac{1}{N}\right)s^2 = \left(1 - \frac{1}{2000}\right)\frac{x}{2}\left\{E[X^2] - E[X]^2\right\}$

$= \frac{1999}{2000}\left[\frac{1}{n}\sum_{i=1}^{25} x_i^2 + \left[\frac{1}{n}\sum(x_i)\right]^2\right] \cdot \frac{1999}{48000}\left[\frac{243505}{25} - \left(\frac{2450}{25}\right)^2\right]$

$= \boxed{136}$ ‖ $\frac{\sigma_{\bar{x}}^2}{n}\left(1 - \frac{n}{N}\right) = \boxed{5.37}$

C) Approximate a 95% confidence ~~interval~~.

$P\left(\left|\frac{X - \bar{X}}{\sigma_s}\right| < z\left(1 - \frac{\alpha}{2}\right)\right) = 0.95$; $X - \sigma_s z\left(1 - \frac{\alpha}{2}\right)$

$2P\left(x - \bar{x} < \sigma_s z\left(1 - \frac{\alpha}{2}\right)\right) = 0.95$; $= \sigma_s\,1.96$

$(\bar{X} \pm 1.96\sqrt{5.37}\,X - \bar{X}\,N\,\sigma_s) = 196{,}000$

$\boxed{98 \pm 4.54}$, $\boxed{196{,}000 \pm 23{,}323}$

36. Simple Random Sampling: $\bar{x}^2$ is unbiased estimate of μ^2. Simple random sampling is an unbiased estimate of $\bar{\mu}^2$ when there is true random sampling, for example, each value has equal probability. Otherwise, the $\bar{x}^2$ is not random, and biased.

37. Population Mean $= \mu$; Survey #1 · Survey #2

$\bar{X}_1 =$ Mean, $\bar{X}_2 =$ Mean } Unbiased. $X = \alpha \bar{X}_1 + \beta \bar{X}_2$

$\sigma_{\bar{X}_1} =$ Standard Error $\quad \sigma_{\bar{X}_2} =$ Standard Error

a) Find conditions for α and β which are an unbiased combination.

$$Var(\mu) = E[\mu^2] - E[\mu]^2 = \frac{(\alpha \bar{X}_1 + \beta \bar{X}_2)^2}{n} - \left[\frac{(\alpha \bar{X}_1 + \beta \bar{X}_2)}{n}\right]^2 = \frac{(\alpha \bar{X}_1)^2 n + 2(\alpha\beta \bar{X}_1 \bar{X}_2) n + (\beta \bar{X}_2)^2 n - (\alpha \bar{X}_1)^2 - 2(\alpha\beta \bar{X}_1 \bar{X}_2) - (\beta \bar{X}_2)^2}{n^2}$$

$$= \frac{(\alpha \bar{X}_1)^2 (n-1) + 2(\alpha\beta \bar{X}_1 \bar{X}_2)(n-1) + (\beta \bar{X}_2)^2 (n-1)}{n^2} = \frac{n-1}{n^2}\left[\alpha \bar{X}_1 + \beta \bar{X}_2\right]^2$$

$$Var(\mu) = Var(\alpha \bar{X}_1 + \beta \bar{X}_2) = \alpha^2 Var(\bar{X}_1) + \beta^2 Var(\bar{X}_2) \quad \frac{(n-1)^2}{n^2}\left[\alpha \bar{X}_1 + \beta \bar{X}_2\right]^2$$

$$E(X) = \alpha E(\bar{X}_1) + \beta E(\bar{X}_2) \qquad = \frac{n-1}{n^2}\left[(\alpha \bar{X}_1)^2 + 2\alpha\beta \bar{X}_1 \bar{X}_2 + (\beta \bar{X}_2)^2\right]$$

$$= (\alpha + \beta)\mu$$

$\boxed{(\alpha + \beta) = 1}$ "unbiased"

$$\alpha^2 Var(\bar{X}_1) + \beta^2 Var(\bar{X}_2) = (\alpha^2[\sigma_{\bar{X}_1}^2] + (1-\alpha)^2[\sigma_{\bar{X}_2}^2]) (1 \cdot \bar{X}_2)^2$$

$$\boxed{= \alpha^2[\sigma_{\bar{X}_1}^2] + \sigma_{\bar{X}_2}^2 + \alpha^2[\sigma_{\bar{X}_2}^2] - 2\alpha[\sigma_{\bar{X}_2}^2]}$$

b) $\frac{d}{d\alpha}[\alpha^2[\sigma_{\bar{X}_1}^2] + \sigma_{\bar{X}_2}^2 + \alpha^2[\sigma_{\bar{X}_2}^2] - 2\alpha[\sigma_{\bar{X}_2}^2]]$

$$\left| \alpha = \frac{\sigma_{\bar{X}_2}^2}{[\sigma_{\bar{X}_1}^2 + \sigma_{\bar{X}_2}^2]} \right| ; \left| \beta = \frac{\sigma_{\bar{X}_1}^2}{\sigma_{\bar{X}_1}^2 + \sigma_{\bar{X}_2}^2} \right|$$

38. $X_1 \ldots X_n$ be random sample. Show $\frac{1}{n}\sum_{i=1} X_i^3$ is an unbiased estimate of $\frac{1}{N}\sum_{i=1}^{N} X_i^3$

$E[X^3] = \frac{1}{N}\sum_{i=1}^{N} X_i^3$ ← No β to incorporate, no bias and no shift.

39. $N =$ Population of items. How large should a sample be to find a defective item?

$k =$ Defective items.

Assuming $p = 0.95$; $n =$ sample size; $k = 1$; $P\left(1 < \frac{X-M}{\sigma_x} < k\right) = 0.95$

$$1 - \frac{N-k}{N} \times \frac{N-k-1}{N-1} \times \cdots \times \frac{N-k-n+1}{N-n+1} > 0.95 \quad \text{if } k = 1, 95$$

$$1 - \frac{N-1}{N} \times \frac{N-1-1}{N-1} \times \cdots \times \frac{N-n}{N-n+1} > 0.95;$$

40. i^{th} member has $P(i) = \frac{n - n_i}{N - i + 1}$

$$= \frac{(N-n)n!}{N!} = \boxed{\frac{N-n}{N-(i+1)}} = \frac{1}{\binom{N}{n}}$$

$$1 - \left(\frac{N-n}{N}\right)^k > 0.95$$

$$\left(\frac{N-n}{N}\right)^k < 0.05$$

$$\log(N-n) - \log(N) \approx \frac{\log(0.05)}{k}$$

$$n \approx N - e^{\left(\frac{\log(0.05)}{k} + \log(N)\right)}$$

≈ 501

41. $D =$ Audited value − Book value

$N\bar{D}$, where $N =$ Population size.

Inventory value.

a) Prove unbiased estimate

$$\boxed{E[N\bar{D}] = \frac{N}{n}\sum^{N} D = NE[D]}$$

b.) Variance of Estimate: $Var(N\bar{D}) = E[N\bar{D}^2] - E[N\bar{D}]^2 = N^2E[\bar{D}^2] - N^2E[\bar{D}]^2 = N^2[E[\bar{D}^2] - E[\bar{D}]^2]$

c) Population Parameter $[\tau]$; Estimate $T = N\bar{X}$; Variance of Estimate $\sigma_T^2 = N^2\sigma_{\bar{X}}^2 = N^2[E[X^2] - E[D]^2]$
$= N^2\sigma_D^2 \pm N^2[E[D^2] - E[D]^2]$

The proposed method would be as accurate.

d) Estimation of Ratio: $r = \frac{\sum y_i}{\sum x_i} = \frac{\mu_D}{\mu_X}$

A ratio estimate would provide advantage to a differently sized pool of populations. In the listed case of part a, b, or c, then there would be no difference.

42. Population Correlation Coefficient: $\rho = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{\frac{1}{N}\sum_i (x_i-\mu_x)(y_i-\mu_y)}{\frac{1}{N}\sum_i (x-\mu_x)^2 \frac{1}{N}\sum (y-\mu_y)^2}$ $= \boxed{\frac{1}{N}\left(\frac{1}{\sum(x-\mu_x)\sum(y-\mu_y)}\right)}$

43. Example D: Section 7.3.3.

44. $\bar{X} = 2.2$, $\sigma = 0.7$, $\rho[\text{\#occupancy}|\text{\#Motor vehicles}] = 0.85$

Estimate Population ratio of # Motor-Vehicles per occupant + S.E.

Population Ratio: $\frac{\mu_2}{\mu_1} = \frac{\bar{X}_2}{\bar{X}_1} = \frac{2.2 \text{ motor vehicle per car}}{2.2 \text{ motor vehicle per car}} = 0.727$

Standard Error: $Var(r) = \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\frac{1}{\mu_1^2}\left(r^2\sigma_1^2 + \sigma_2^2 - 2r\sigma_{12}\right)$

$= \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\frac{1}{\mu_1^2}\left(r^2\sigma_1^2 + \sigma_2^2 - 2r\rho\sigma_x\sigma_y\right)$

$= \frac{1}{100}\left(1 - \frac{100-1}{8000-1}\right)\frac{1}{2.2^2}\left(0.727^2 \cdot 0.7^2 + 0.8^2 - 2(0.727)(0.85)(0.8)(0.7)\right)$

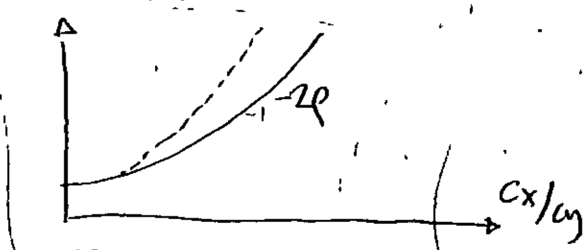
$= 4.25 \times 10^{-4}$; $\boxed{SE = 2.05 \times 10^{-2}}$

Confidence Interval (95%): $0.95 = P\left(\frac{x-\bar{x}}{\sigma_{\bar{x}}} < z(1-\frac{\alpha}{2})\right)$; $= 0.727 \pm 1.96(\sigma_{\bar{x}})$

$2.0727 \pm 1.96(2.05 \times 10^{-2})$

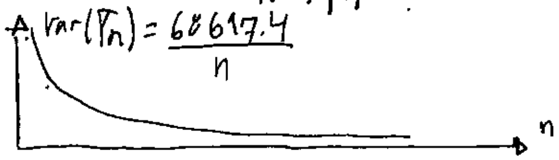
$\boxed{0.727 \pm 0.04}$

44. $\frac{Var(\bar{Y}_R)}{Var(\bar{Y})} \approx 1 + \frac{c_x}{c_y}\left(\frac{c_x}{c_y} - 2\rho\right) = 1 + x^2 + x$



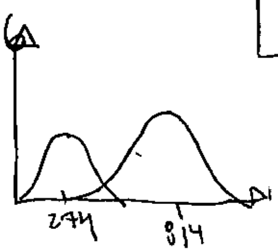
45. $\rho = 0.91$ $Var(\bar{Y}_R)$; Plot $Var(\bar{Y}_R)$ for $n = 64$
$-1 < \rho < 1$

$Var(\bar{Y}_R) = \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\frac{1}{\mu_1^2}\left(r^2\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_2\sigma_1\right)$



46. $\sigma_{\bar{Y}_R} \approx 32.76$

$\sigma_{\bar{Y}} = 66.3$



47. $n=64$. Corollary B of Section 7.4 : Approximate Bias of the ratio estimate of μ_y

$E(\bar{Y}_R)-\mu_y \approx \frac{1}{64}\left(1-\frac{64-1}{393-1}\right)\frac{1}{274.8}(2.96\cdot 213.2^2-0.91\times 213.2\times 589.7)$

$$E(\bar{Y}_R)-\mu_y \approx \frac{1}{n}\left(1-\frac{n-1}{N-1}\right)\frac{1}{\mu_x}(r\sigma_x^2-\rho\sigma_x\sigma_y)$$

$n=128$ $= 0.96$

$\approx \frac{1}{128}\left(1-\frac{64-1}{393-1}\right)\frac{1}{274.8}(2.96\times 213.2^2-0.91\times 213.2\times 559.7)$

48. ≈ 0.39.

$n=100$ Households; # people in Household [X] ; Weekly Expenditure for Food [Y.]

Total Number of Households = 100,000

a) Estimate the ratio $r=\frac{\mu_y}{\mu_x} = \frac{(\sum Y_i)/N}{(\sum X_i)/N}$

$\sum X_i = 320$: Total sum # people in Household

$\sum Y_i = 10,000$: Total Weekly Expenditure for Food

$$\boxed{= \frac{10,000}{3,200} = \frac{125}{4}}$$

b) Confidence Interval (95%): $r \pm 1.96\,\sigma_x$

$\sum X_i^2 = 1250$

$\frac{125}{4} \pm 1.96 \cdot \left(\frac{1}{n}\left[1-\frac{n-1}{N-1}\right]\frac{1}{\mu_x^2}(r^2\sigma_x^2+\sigma_y^2-2r\rho\sigma_x\sigma_y)\right)$

$\sum Y_i^2 = 1,100,000$

$\frac{125}{4} \pm 1.96\left(\frac{1}{10^2}\left[1-\frac{10^2-1}{10^5-1}\right]\frac{1}{\left(\frac{320}{100}\right)^2}\left(\left[\frac{125}{4}\right]^2\left(\frac{1250}{100}-\left(\frac{320}{100}\right)^2\right)+\left(\frac{1,100,000}{100}-\left(\frac{10000}{100}\right)^2\right)-2\left(\frac{125}{4}\right)\left(\frac{36,000}{100}-\frac{32,0000}{100^2}\right)\right)\right)$

$\sum X_i Y_i = 36,000$

$$\boxed{\frac{125}{4} \pm 1.34}$$

c. $\tau = N\bar{Y} = 100000 \cdot \frac{10000}{100} = 10^7$; $\left(1-\frac{1}{10^5}\right)s^2 \cdot \left(\frac{1}{99995}\right)\frac{1}{100-1}\left(\frac{1,100,000^2}{100}-\left(\frac{10,000}{100}\right)^2\right) = 1.00\times 10^{-4}$

$P\left(\left|\frac{\bar{X}-\bar{Y}}{\sigma_{\bar{x}}}\right| < z\left(1-\frac{\alpha}{2}\right)\right) = 0.90$

$100 \pm 1.65 \cdot 1.66\times 10^{-4}$

$$\boxed{100 \pm 0.00027}$$

49. $N=1000$ squares

$n=50$ sampled

$Y=$ # of birds

$X=$ Area covered by vegetation

a) $r = \frac{\sum Y_i/n}{\sum X_i/n} = \frac{150}{3000} = \frac{1}{20}$

$\sum X_i = 3,000$

b) Standard Error:

$\sum Y_i = 150$

$s_R = \sqrt{Var(R)} = \sqrt{\frac{1}{n}\left(1-\frac{n-1}{N-1}\right)\frac{1}{\mu_x^2}(r^2\sigma_x^2+\sigma_y^2-2r\rho\sigma_x\sigma_y)}$

$\sum X_i^2 = 225,000$

$\sigma_x = \sqrt{\frac{1}{50-1}[225,000-50\cdot 60^2]} = 30$; $\sigma_{xy} = \frac{1}{50-1}[11,000-50\cdot 60\cdot 3] = 40.8$

$\sum Y_i^2 = 650$

$\sigma_y = \sqrt{\frac{1}{50-1}[650-50\cdot 3^2]} = 2.1$ $\rho = 0.68$

$\sum X_i Y_i = 11,000$

$s_R = \sqrt{\frac{1}{50}\left(1-\frac{50-1}{1000-1}\right)\frac{1}{\left(\frac{3000}{50}\right)^2}\left(\left(\frac{1}{20}\right)^2 30^2+2.1^2-2\left(\frac{1}{20}\right)\cdot 30\cdot 2.1\cdot 0.68\right)}$

$= 7.34\times 10^{-3}$

~~95% Confidence Interval~~

90% Confidence Interval: $r \pm 1.65\,s_R = \frac{1}{20} \pm 1.65\cdot 7.34\times 10^{-3} = 0.05 \pm 0.012$

c) Total Number of Birds: 95% Confidence Interval

$\tau = N\cdot\bar{Y} = 1000\cdot\frac{150}{50} = 3,000$

Standard Error:

$\bar{T} \pm 1.96\, s_T = 3000 \pm 1.96\cdot 275 = 300 \pm 540$

$s_T = s_T\sqrt{\frac{N(N-n)}{n}} = N s_{\bar{x}} = 1000\sqrt{\frac{s_y^2}{n}\left(1-\frac{50}{1000}\right)} = 1000\sqrt{\frac{4}{50}\left(1-\frac{50}{1000}\right)} = 2.75$

d) $T_R = \frac{\bar{Y}}{\bar{X}} \tau_X = R \tau_X$

$= \frac{150}{3000} 1000 \cdot \left(\frac{3000}{50}\right) = \boxed{3000}$

$S_{TR} = \sqrt{\frac{N^2}{n}\left(1 - \frac{n-1}{N-1}\right)\left(R^2 s_x^2 + s_y^2 - 2R s_{xy}\right)} = N S_{Rn}$

$= \sqrt{\frac{1000^2}{50^2}\left(1 - \frac{50-1}{1000-1}\right)\left(\left(\frac{1}{20}\right)^2 30^2 + 2^2 - 2\left(\frac{1}{20}\right) 40.8\right)}$

$= 28.7$ the standard error is much worse.

50. Standard Error of a Ratio Estimate:

$\frac{|E(R) - r|}{\sigma_R} \le \frac{\sigma_{\bar{X}}}{\mu_X} = \frac{\sigma_X}{\mu_X} \sqrt{\frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)}$

a) $Cov(R, \bar{X}) = E\left[\frac{\bar{Y}}{\bar{X}} \bar{X}\right] - E\left(\frac{\bar{Y}}{\bar{X}}\right) E(\bar{X})$

$= E[R][E(\cdots)]\cdots(\bar{Y}) E(\bar{X})]$

$= \cdots (R - \cdots)(\bar{X} - \mu_X)$

$= \frac{Cov(\bar{X}, \bar{Y})}{\bar{X}}$

$\frac{\sigma_{XY}}{\bar{X}} = E[\bar{Y}] - E(R) E[\bar{X}]$

$-\frac{\sigma_y \sigma_x \rho}{\bar{X}} = \frac{E(R) E[\bar{X}] - E[\bar{Y}]}{\rho \cdot \sigma_y}$

$-\frac{\sigma_x}{\bar{X}} = \frac{E[R] E[\bar{X}] - E[\bar{Y}]}{\sigma_y \rho}$

$\boxed{\left|\frac{\sigma_x}{\bar{X}} = \frac{E(R) E[\bar{X}] - E[\bar{Y}]}{\sigma_y \cdot \rho}\right| = \frac{|E(R) - E\left[\frac{\bar{Y}}{\bar{X}}\right]|}{\sigma_R}}$

b $\frac{Var(\bar{Y}_R)}{Var(\bar{Y})} = 1 + \frac{C_x}{C_y}\left(\frac{C_x}{C_y} - 2\rho\right)$

$\boxed{\left|\frac{E(R) - \bar{Y}_R}{\sigma_{\bar{X}} / \mu_X}\right| = 1 + \frac{C_x}{C_y}\left(\frac{C_x}{C_y} - 2\rho\right)}$

51. $E(\hat{\theta}) = \hat{\theta} + \frac{b_1}{n} + \underbrace{\frac{b_2}{n^2}}_{\theta_j} + \cdots$; $\hat{\theta}$ = estimate of θ

$\frac{1}{n} \cdots \frac{1}{n}$ $\boxed{m} \cdots \boxed{m}$ $n = mp$ For $(j = 1 \cdots p)$

p Estimate $\hat{\theta}_j$ from $m(p-1)$

$E(\hat{\theta}) = \hat{\theta} + \frac{b_1}{m(p-1)} + \frac{b_2}{[m(p-1)]^2} + \cdots$

p-pseudovalues: $V_j = p\hat{\theta} - (p-1)\hat{\theta}_j$ Prove $\theta_j = \frac{1}{p}\sum_{j=1}^{p} V_j$ $\sigma_R^2 = \cdots E[R]\frac{1}{n^2}$

$\hat{\theta}_j = \frac{1}{p}\sum_{j=1}^{p} V_j$ $= E\left[\frac{\bar{Y}}{\bar{X}}\right] - E\left[\frac{\bar{Y}}{\bar{X}}\right]$?

$E(\hat{\theta}_j) = \hat{\theta} + \frac{b_1}{m(p-1)} + \frac{b_2}{[m(p-1)]^2} + \cdots = \frac{1}{p}\sum_{j=1}^{p} V_j + \frac{b_1}{m(p-1)} + \frac{b_2}{[m(p-1)]^2} + \cdots = p\hat{\theta} - (p-1)\theta_j + \frac{b_1}{m(p-1)} + \frac{b_2}{[m(p-1)]^2} z^{+\cdots}$

$\frac{dE(\hat{\theta}_j)}{dp} = \hat{\theta} - \theta_j - \frac{b_1}{m(p-1)^2} - \frac{2b_2}{[m(p-1)]^3} + \cdots$

$\frac{d^2E(\hat{\theta}_j)}{d^2p} = \frac{+2b_1}{m(p-1)^2} + \frac{6b_2}{m(p-1)^4} + \cdots = 0$

$b_1 + \frac{3b_2}{m(p-1)^4} + \cdots = 0$

$b_1 = \frac{-3b_2}{m(p-1)^4}$

$E(\hat{\theta}) = \hat{\theta} - \frac{3b_2}{n(p-1)^4} + \frac{b_2}{n^2} + \cdots$

$= \hat{\theta} + \frac{b_2}{n}\left(\frac{1}{n^2} - \frac{3}{n(p-1)^4}\right) + \cdots$

52. $N_1 = N_2 = 1000$

$N_3 = 500$

10 observations

Stratum #1: 94 99 106 106 101 102 122 104 97 97

Stratum #2: 183 183 179 211 178 179 192 192 201 177

Stratum #3: 343 302 286 317 289 284 357 288 314 276

$\bar{X}_1 = 103$ $\sigma_1 = 7.4$
$\bar{X}_2 = 188$ $\sigma_2 = 11$
$\bar{X}_3 = 278$ $\sigma_3 = 38$

$\boxed{\tau_1 = N \cdot \bar{X}_1 = 103000;\ \tau_2 = N\bar{X}_2 = 188000;\ \tau_3 = N\bar{X}_3 = 139,000}$

$\boxed{103 \pm 12}$ $\boxed{188 \pm 18}$ $\boxed{278 \pm 63}$

53. a. n=100 Sample size Methods of Allocation. Sample size $\left(n_\ell = n \frac{W_\ell \sigma_\ell}{\sum_{k=1} W_k \sigma_k}\right)$

$\sum W_k \sigma_k = \frac{394}{2010}8.3 + \frac{461}{2000}\cdot 13.3 + \frac{391}{2010}\cdot 15.1 + \frac{334}{2010}\cdot 19.8 + \frac{164}{2010}\cdot 24.5 + \frac{113}{2010}26 + \frac{140}{2010}35.2$

$= 1.63 + 3.05 + 2.94 + 3.29 + 2.06 + 1.46 + 2.59$

$\boxed{=17}$ $(Var(\bar{X}_{so}), Var(\bar{X}_{sp}), Var(\bar{X}_{SRS})$

Optimal Allocation

Farm Size	n_ℓ
0-40	9.631
41-80	17.9
81-120	17.3
121-160	19.4
161-200	12
201-240	8.6
241+	15.2

"A scaled sample size to the true value."

b. $Var(\bar{X}) = \frac{1}{n_\ell}\left(1 - \frac{n_\ell - 1}{N_\ell - 1}\right)\sigma_\ell^2$ "a scaled variance"

Farm Size	$Var(\bar{X})$
0-40	7.01
41-80	9.51
81-120	12.62
121-160	19.09
161-200	43.23
201-240	73.27
241+	73.64

c.

Farm Size:	$E(\bar{X}_s) = \sum W_\ell E(\bar{X}_\ell)$
0-40	1.06
41-80	3.74
81-120	4.73
121-160	5.73
161-200	3.54
201-240	2.81
241+	4.70

"Population Expectation"

$Var(\bar{X}_s) = \sum_{\ell=1}^{L} W_\ell^2 \left(\frac{1}{n_\ell}\right)\left(1 - \frac{n_\ell - 1}{N_\ell - 1}\right)\sigma_\ell^2$

0-40	1.98×10^{-2}
41-80	7.30×10^{-2}
81-120	6.44×10^{-2}
121-160	7.61×10^{-2}
161-200	1.74×10^{-2}
201-240	2.48×10^{-3}
240+	2.19×10^{-2}

"Population Variance"

Proportional Allocation

d. n = 10 Farms

Farm Size:	$Var(\bar{X}_s)$:
0-40	2.59×10^{-1}
41-80	9.12×10^{-1}
81-120	8.42×10^{-1}
121-160	1.05
161-200	4.02×10^{-1}
201-240	1.96×10^{-1}
241+	6.30×10^{-1}

"smaller sample size, variance"

e) n = 70 samples

Farm Size:	$Var(\bar{X}_s)$:
0-40	2.69×10^{-1}
41-80	5.00×10^{-1}
81-120	4.78×10^{-1}
121-160	5.27×10^{-1}
161-200	3.06×10^{-1}
201-240	2.32×10^{-1}
241+	3.99×10^{-1}

b) $Var(\bar{X}_{so}) = \frac{\left(\sum_{\ell=1} W_\ell \sigma_\ell\right)^2}{n}$ "Stratified Optimal Allocation" to stratified population

$Var(\bar{X}_{sp}) = \sum_{\ell=1}^{L} W_\ell^2 Var(\bar{X}_\ell) = \sum_{\ell=1}^{L} W_\ell^2 \frac{\sigma_\ell^2}{n_\ell}$ "Stratified Proportional Allocation" to total stratified population

$Var(\bar{X}_{SRS}) = \sum_{\ell=1}^{L} W_\ell^2 \frac{\sigma_\ell^2}{n_\ell} + \frac{\sum W_\ell^2 (\mu_\ell - \mu)^2}{n_\ell}$ "Stratified Random Sampling" increases precision for diverse values of population.

$= \sum_{\ell=1}^{L} \frac{W_\ell^2\left(\sigma_\ell^2 + (\mu_\ell - \mu)^2\right)}{n_\ell}$

$Var(\bar{X}_{so}) =$

54a) $C = C_0 + C_1 n$; L Strata ; Find a function which minimizes the variance.

Start-up Cost (C_0), cost per observation (C_1)

Lagrangian Multiplier: $L(n_1, \dots n_L, \lambda) = \sum_{l=1}^{L} \frac{W_l^2 \sigma_l^2}{n_l} + \lambda\left(\sum_{l=1}^{L} n_l - n\right)$

$L(n_1, \dots n_L, \lambda) = \sum_{l=1}^{L} \frac{W_l^2 \sigma_l^2}{n_l} + \lambda\left(\sum C_l n_l - C_1 n\right)$

$\frac{\partial L}{\partial n_l} = -\frac{W_l^2 \sigma^2}{n_l^2} + \lambda$; $n_l = \frac{W_l \sigma_l}{\sqrt{\lambda}}$

$n = \frac{1}{\sqrt{\lambda}} \sum_{l=1}^{L} W_l \sigma_l$

$\frac{1}{\sqrt{\lambda}} = \frac{n}{\sum_{l=1}^{L} W_l \sigma_l}$

$n_l = n \frac{W_l \sigma_l}{\sum_{l=1}^{L} W_l \sigma_l}$

$\frac{\partial L}{\partial n_l} = -\frac{W_l^2 \sigma_l^2}{n_l^2} + \lambda C_l = 0$; $n_l = \frac{W_l \sigma_l}{\sqrt{\lambda C_l}}$; $n = \frac{1}{\sqrt{\lambda}} \sum_{l=1}^{L} \frac{W_l \sigma_l}{\sqrt{C_l}}$

$\frac{1}{\sqrt{\lambda}} = n \sum_{l=1}^{L} \frac{\sqrt{C_l}}{W_l \sigma_l}$

$n_l = n W_l \sigma_l \sum_{l=1}^{L} \frac{\sqrt{C_l}}{W_l \sigma_l}$

$Var(\bar{X}_{st}) = \sum_{l=1}^{L} W_l^2 \left(\frac{1}{n_l}\right)\left(1 - \cancel{\frac{n_l - 1}{N_l - 1}}\right)\sigma_l^2$

Neglecting finite population correction

$= \sum_{l=1}^{L} \frac{W_l^2 \sigma_l^2}{n_l} = \frac{\sum W_l^2 \sigma_l^2}{n W_l \sigma_l} \sum \frac{W_l \sigma_l}{\sqrt{C_l}}$

$= \frac{\sum_{l=1}^{L} (W_l \sigma_l)^2 / \sqrt{C_l}}{n}$

c) $\boxed{n_l = n W_l \sigma_l \sum \frac{\sqrt{C_l}}{W_l \sigma_l}}$

b) $\boxed{Var(\bar{X}_{st}) = \frac{\sum (W_l \sigma_l)^2 / \sqrt{C_l}}{n}}$

55. a) Proportional Allocation of a population mean $\bar{X}_{sp} = \sum_{l=1}^{L} W_l \bar{X}_l = \sum W_l \left(\frac{1}{n_l} \sum X_{il}\right)$ $= \frac{1}{n} \sum_{l=1}^{L} \sum_{i=1}^{n_l} X_{il}$

is utilized when W_L is large representation.

Optimal Allocation of a population mean $\bar{X}_{st} = \sum_{l=1}^{L} W_l \bar{X}_l$

is utilized when ~~a sample of each stratum is taken.~~

Being that ($H_s = 100$, $H_T = 100{,}000$) and ($L_s = 200$, $L_T = 500{,}000$) then optimal allocation of a population ~~seems best for model~~ $\frac{1}{6}\bar{X}_H + \frac{5}{6}\bar{X}_L = \bar{X}_T$

b) $\sigma_H = 20$, $\sigma_L = 10$; Standard Error ; $S_{\bar{X}H} = \sqrt{\frac{20^2}{100}\left(1 - \frac{100}{100{,}000}\right)} = \boxed{2.00}$ $S_T = 0.6775$

c) A 95% confidence interval $S_{\bar{X}L} = \sqrt{\frac{10^2}{200}\left(1 - \frac{200}{500{,}000}\right)} = \boxed{0.71}$

would be best to determine allocation error. The current allocation ($H_s = 100$, $L_T = 200$) provides an interval of (± 3.92, ± 1.39) by increasing the allocation to ($H_s = 200$, $L_s = 100$) would shift the interval to (± 2.77, ± 1.96). Also, the standard error of the population would shift from 0.6775 to 0.8866 and be of greater error.

d) Proportional Allocation provides a standard error of:

$\sqrt{\sigma_{st}^2} = \sigma_{sp}$ ... [illegible] worse, $\frac{1}{4}$ ... $S_T =$

56. a) A survey of household expenditures in a city. [Stratification of expenditure type or district]
b) Examination of lead concentration in a large plot of land [concentration ranges]
c) Surveying the number of people who use elevators in a large building [Time of day]
d) Surveying television by time of day [Stratification of seasons]

57. Sample Pool: $\{1,2,2,4, \text{ and } 8\} \rightarrow (1,2,2)$ and $(4,8)$: $\bar{X}_{s1} = \frac{1}{3}(1)+\frac{1}{3}(2)+\frac{1}{3}(2) = \frac{5}{3}$; $\bar{X}_s = \frac{3}{5}\left(\frac{5}{3}\right)+\frac{2}{5}(6)$
$= 1 + \frac{12}{5} = 1+2.4$ $\boxed{=3.4}$

$\bar{X}_{s2} = \frac{1}{2}(4)+\frac{1}{2}(8) = 6$.

$\sigma_{\bar{X}_{s1}} = \sqrt{\frac{(1-5/3)^2+(2-5/3)^2+(2-5/3)^2}{3}} = 0.47$

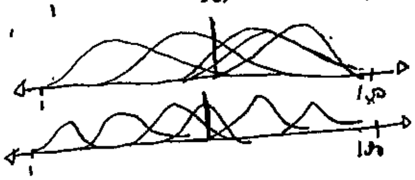
$\sigma_{\bar{X}_{s2}} = \sqrt{\frac{(4-6)^2+(8-6)^2}{2}} = 2.00$

$Var(\bar{X}_s) = \left(\frac{3}{5}\right)^2\left(\frac{1}{3}\right)\left(1-\frac{3-1}{5-1}\right)\cdot 0.47^2 + \left(\frac{2}{5}\right)^2\left(\frac{1}{2}\right)\left(1-\frac{2-1}{5-1}\right)2^2$

$\boxed{=1.02}$

58. $[n_1, n_2 \cdots\cdots n_{12}]$
$[n_1, n_2 \cdots\cdots n_{12}]$
$[n_1, n_2 \cdots\cdots n_{12}]$
$\vdots$
$[n_1, n_2 \cdots\cdots n_{12}]$ $\Big\}$ x100

59.



60. $n_H = 100{,}000$, $\sigma_{EH} = 20$
$n_L = 500{,}000$ $\sigma_{EL} = 12$
Sample size $= 100$

a) Optimal Allocation
Population Mean:
$\bar{X}_{so} = \sum_{L=1}^{L} W_L \cdot \bar{X}_L$

b) $\mu_H - \mu_L = W_H \bar{X}_H - W_L \bar{X}_L$
$= \frac{100}{100{,}000}\bar{X}_H - \frac{100}{500{,}000}\bar{X}_L$

61. $\mu(t) = \alpha + \beta t$
Sample size $= n$
@ times $= 1, 2$ and 3
a) Find W_1, W_2 and W_3, such that $\hat{\beta} = w_1\bar{X}_1 + w_2\bar{X}_2 + w_3\bar{X}_3$ is unbiased estimate of β. X_i denotes sample mean.

$\boxed{\frac{d}{dX}\hat{\beta} = w_1 + w_2 + w_3 = 0}$
and
$X \propto$ time: $\boxed{w_1 + w_2(2) + w_3(3) = 1}$

b) $w_1 = -w_2 - w_3$; $w_1 = 1 - 2w_2 - 3w_3$; $-w_2 - w_3 = 1 - 2w_2 - 3w_3$
$w_2 = 1 - 2w_3$
$w_1 + 1 - 2w_3 + w_3 = 0$.
$w_1 + 1 - w_3 = 0$; $w_1 - w_3 = -1$; $\boxed{w_1 = -\frac{1}{2} \text{ ; } w_3 = \frac{1}{2} \text{ ; } w_2 = 0}$

62. Example B: Section 7.5.2

$\bar{X}_1 = 240.6$ $\quad s_1^2 = 6827.6$
$\bar{X}_2 = 507.4$ $\quad s_2^2 = 23790.7$
$\bar{X}_3 = 865.1$ $\quad s_3^2 = 42573.0$
$\bar{X}_4 = 1716.5$ $\quad s_4^2 = 152{,}099.6$

$\bar{X}_s = 832.5$
$s_{\bar{X}}^2 = \frac{1}{10}\sum_{\ell=1}^{4} W_\ell^2\left(1-\frac{n-1}{N-1}\right)s_\ell^2 = 1282.0$

$s_{\bar{X}} = 35.8$

$T_s = 393\bar{X}_s = 327{,}172.$
$s_{T_s} = 393\, s_{\bar{X}} = 14{,}069$

The estimate of $s_{\bar{X}} = 35.8$ has a standard error of

$\sigma_{\bar{X}} = \sqrt{\frac{s_{\bar{X}}^2}{10}\left(1-\frac{10}{393}\right)} = 11.2$

which demonstrates the sample size of $n=10$ is small compared to the population size of $N=393$.

63.
$X = [0,1]$
$f(x) = X_0$: Vertical line test.
$\hat{I}(f) = \frac{1}{n}\sum_{i=1}^{n} f(x)$: Mean.
Standard Error:
$\sigma^2 = E[f(x)^2] - E[f(x)]^2$
$= \frac{1}{n}\left[\sum f(x)^2\right] - \left[\sum f(x)\right]^2$
$= \frac{1}{n^2}\left[n\sum f(x)^2 - \left(\sum f(x)\right)^2\right] =$

The actual

$\sigma_X = \sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{\frac{1}{1000^2}\left[\left(\frac{1}{2\pi}\right)\left[2\pi\sum_{i=1}^{1000} e^{-X_i^2} - \sum e^{-X_i^2/2}\right]\right]}{n}\left(1-\frac{n}{N}\right)}$ where $n = 1000$

The population stratification would be best exampled as σ-intervals of separation.

64.

Stratum	N_ℓ	μ_ℓ	σ_ℓ
\$1000+	70	3000	1250
\$200-1000	500	500	100
\$1-200	10,000	90	30

a) Proportional Allocation Mean S.D.

Proportional Allocation $\bar{X} = \sum_{\ell=1}^{L} W_\ell X_\ell$ $\quad Var(\bar{X}_s) = \frac{1}{n}\sum W_\ell\left(1 - \frac{n_\ell - 1}{N-1}\right)\sigma_\ell^2$

Optimal Allocation: $\mu = \frac{1}{N}\sum_{\ell=1}^{L} N_\ell \mu_\ell$ $\quad var(\bar{X}_s) = \sum_{\ell=1}^{L} W_\ell^2\left(\frac{1}{n_\ell}\right)\left(1 - \frac{n_\ell - 1}{N_\ell - 1}\right)\sigma_\ell^2$

Relative Sampling: $\frac{n_\ell}{N_\ell} = \frac{n}{N}$; $n_\ell = N_\ell \frac{n}{N} = n W_\ell$ $\boxed{\text{Proportional Allocation}}$

\$1-200: $n_\ell = n\left(\frac{10000}{10570}\right) = 0.945\, n$

\$200-1000: $n_\ell = n\left(\frac{500}{10570}\right) = 0.47n$

\$1000: $n_\ell = n\left(\frac{70}{10570}\right) = 0.66n$

b) Two methods exist to compare the differences of population mean based upon proportional allocation and optimal allocation.

$n_\ell = n \frac{W_\ell \sigma_\ell}{\sum_{\ell=1}^{L} W_\ell \sigma_\ell}$ $\boxed{\text{Optimal Allocation}}$

\$1-200: $\frac{28.5n}{900.5} = 0.03n = n_\ell$

\$200-1000: $0.05n = n_\ell$

\$1000: $0.09n = n_\ell$

$$Var(\bar{X}_{sp}) - Var(\bar{X}_{so}) = \frac{1}{n}\sum_{\ell=1}^{L} W_\ell(\sigma_\ell - \bar{\sigma})^2$$

$$= \frac{1}{n}\left[\sum_{\ell=1}^{L} W_\ell \sigma_\ell^2 - \left(\sum_{\ell=1}^{L} W_\ell \sigma_\ell\right)^2\right] \text{ and } \frac{Var(\bar{X}_{sp})}{Var(\bar{X}_{s})} = 1 + \frac{\sum W_\ell(\sigma_\ell - \bar{\sigma})^2}{\left(\sum W_\ell \sigma_\ell\right)^2}$$

65.

Time: 1950-1960

Adult White Female Population (160,301 counties, North Carolina, South Carolina, Georgia.

a) Histogram: Bins by year b) Mean $= \frac{1}{N}\sum X_i$; Total Cancer Mortality: $T = N\bar{x}$

variance: $\frac{\sum(X_i - \bar{X})^2}{N}$; SD $= \sqrt{\frac{\sum(X_i - \bar{X})^2}{N}}$

c) $n = 25$

d) Mean $= \frac{1}{n}\sum X_i$; $T = N\bar{X}$ e) $Var(\bar{X}) = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}}$

f) $\bar{X} \pm 1.96 s_{\bar{x}}$, $T = N s_{\bar{x}}$ g) See (d-f) h) Ratio Estimator $(r) = \frac{\sum_{i=1}^{n} \bar{X}}{\sum_{i=1}^{N} \bar{X}}$ A ratio estimator would be effective to compare [illegible]

i) see c) j) see d) k) --- l) Seperate and stratify by

66. The sampling procedure would be regions of the beach vs # of people. Stratified means would be calculated with proportional allocation for variance.

$n_\ell = n \frac{W_\ell \sigma_\ell}{\sum W_\ell \sigma_\ell}$ $\boxed{\text{Optimal Allocation}}$ or $n_\ell = n W_\ell$ $\boxed{\text{Proportional Allocation}}$

n) Further stratum may resolve or represent the data better.

67. a) i) Proportion of female-headed Families [$n = 500$]; $\bar{X} = \sum_{\ell=1}^{L} W_\ell X_\ell$; $Var(\bar{X}_s) = \frac{1}{500}\sum(W_\ell)\left(1 - \frac{500-1}{43086-1}\right)\sigma^2$

ii) The average number of children per family. $s_x = \sqrt{\frac{\sigma_{xs}^2}{n}\left(1 - \frac{n}{N}\right)}$; $\bar{X} \pm 1.96 s_x$

representation would be best demonstrated by region of Cyberville. $\bar{X}_R = \sum_{\ell=1}^{L} W_\ell X_\ell$

$Var(\bar{X}_{sR}) = \frac{1}{500}\sum(W_\ell)\left(1 - \frac{500-1}{43086-1}\right)\sigma^2$

$\bar{X} \pm 1.96 s_x$

iii) The proportion of heads of household who did not receive a high scool diploma

$\bar{X} \pm 1.96\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}$

iv) Average Family income with a sample size of 500 could be represented as a proportional distribution by region. $\bar{X} = \frac{1}{n}\sum_i X_i$; $Var(\bar{X}_{sp}) = \frac{1}{n}\sum W_e \left(1 - \frac{n_e - 1}{N - 1}\right)\sigma^2$

b) i) 100 samples of $n = 400$; Average Family income: $\mu = \frac{1}{N}\sum_{e=1}^{L}\sum_{i=1}^{N_e} X_{ie}$

ii) SO $E[\sigma] \cdot \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}}$ iv) $CDF = \int_{-\infty}^{n} f(x)\,dx = n\sum f(x)^2 - (\sum f(x))^2$

iii) Not applicable

vi) $\bar{X} \pm 1.96 s_{\bar{x}}$ ... vii) ... c) Boxplot:

Histogram:

d) i) see (c) ii) iii) see c) e) f) ... see c)

Chapter 8: Estimation of Parameters and Fitting of Probability Distribution

1. Poisson $\left[f(\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}\right]$; $\lambda = \frac{\text{Counts}}{\text{Total}} = \frac{10,220}{12,169} = 0.8398$

n	Observed	Poisson
0	5267	5254
1	4436	4413
2	1800	1853
3	534	519
4	111	109
5+	21	957
	12,169	

The expected counts do match the observed counts.

2. $\lambda = \frac{1168}{300} = 3.89$

n	Frequency	Expected
0	14	6
1	20	24
2	36	46
3	68	60
4	43	59
5	43	46
6	30	30
7	14	16.5
8	10	8.05
9	6	3.5
10	4	1.36
11	1	0
12	1	
13+	0	
	300	

3. a) Estimate λ for each dataset.

Total 1: ~~400~~ Total 2: ~~400~~ Total 3: ~~400~~ Total 4: 480

$$\boxed{\lambda_1 = 0.6825 \quad \lambda_2 = 1.3225 \quad \lambda_3 = 1.80 \quad \lambda_4 = 4.65}$$

b) Concentration #1: $E_1(X) = \lambda_1$; $Var_1(X) = E_1[X^2] - E_1[X]^2 = 1.28 - 0.46 = 0.82$; $SE = \sqrt{\frac{\sigma^2}{n}} = 0.045$

95% confidence Interval: $P\left(\left|\frac{X - \bar{X}}{\sigma_{\bar{x}}}\right| \le Z(\alpha/2)\right) = 0.95$; $\bar{X} \pm 1.96\sigma_{\bar{x}}$ $\boxed{0.6825 \pm 0.088}$

Concentration #2: $E_2(X) = \lambda_2$; $Var_2(X) = E_2[X^2] - E_2[X]^2 = 3.03 - 1.3225 = 1.71$; $SE = \sqrt{\frac{\sigma^2}{n}} = 0.064$

95% Confidence Interval: $P\left(\left|\frac{X - \bar{X}}{\sigma_{\bar{x}}}\right| \le Z(\alpha/2)\right) = 0.95$; $\bar{X} \pm 1.96\sigma_{\bar{x}}$ $\boxed{1.33 \pm 0.125}$

Concentration #3: $E_3(X) = \lambda_3$; $Var_3(X) = E_3[X^2] - E[X]^2 = 5.20$; $SE = \sqrt{\frac{\sigma^2}{n}} = 0.114$

95% Confidence Interval: $P\left(\left|\frac{X - \bar{X}}{\sigma_{\bar{x}}}\right| \le Z(\alpha/2)\right) = 0.95$; $\bar{X} \pm 1.96\sigma_{\bar{x}}$ $\boxed{1.80 \pm 0.223}$

Concentration #4: $E_4(X) = \lambda_4$; $Var_4(X) = E_4[X^2] - E_4[X]^2 = 34.5 - 4.65^2 = 12.86$; $SE = \sqrt{\frac{\sigma^2}{n}} = 0.179$

95% Confidence Interval: $P\left(\left|\frac{X - \bar{X}}{\sigma_{\bar{x}}}\right| \le Z(\alpha/2)\right) = 0.95$; $\bar{X} \pm 1.96\sigma_{\bar{x}}$ $\boxed{4.65 \pm 0.351}$

c)

n	Observed	Expected	Observed	Expected	Observed	Expected	Obs.	Exp.
0	213	214	103	106	75	66	0	4
1	128	134	143	141	103	119	20	18
2	37	42	98	93	121	109	43	41
3	13	9	42	41	54	64	53	65
4	3	1.36	8	14	30	29	86	75
5	1	0	4	4	13	10	70	69
6	0	0	2	1	2	3	54	54
7	0	0	0	0	1	1	37	36
8	0	0	0	0	0	0	18	21
9	0	0	0	0	1	0	10	11
10	0	0	0	0	0	0	5	5
11	0	0	0	0	0	0	2	2
12	0	0	0	0	0	0	2	1

The expected and observed counts are fitting.

1. Suppose X is a discrete Random Variable : $P(X=0)=\frac{2}{3}\theta$ where $0\le\theta\le 1$

$P(X=1)=\frac{1}{3}\theta$ $n=10$ observations

$P(X=2)=\frac{2}{3}(1-\theta)$ $(3,0,2,1,3,2,1,0,2,1)$

$P(X=3)=\frac{1}{3}(1-\theta)$

a) Find the method of moments estimator of θ

(1) $\boxed{\text{Parameter Estimate}}$

$\mu_k=E(X^k)$; $\mu_1=E(X^1)=\frac{1}{10}[3+0+2+1+3+2+1+0+2+1]$

$=\frac{3}{2}$

$\mu_2=E(X^2)=\frac{1}{10}[3^2+0^2+2^2+1^2+3^2+2^2+1^2+0^2+2^2+1^2]=3.3$

(2) $\boxed{\text{Moment Estimators}}$

$E(X^1)=\frac{1}{10}[3\cdot 1p(X=1)+3\cdot 2p(X=2)+2\cdot 3\cdot p(X=3)]=\frac{1}{10}[\theta+4(1-\theta)+2(1-\theta)]$

$=\frac{1}{10}[\theta+4-4\theta+2-2\theta]=\frac{6+5\theta}{10}=\frac{3}{2}$

$E(X^2)=\frac{1}{10}[3\cdot 1^2p(X=1)+3\cdot 2^2p(X=2)+2\cdot 3^2p(X=3)]$

(3) $\boxed{\text{Estimate in terms of Moments}}$

$=\frac{1}{10}[\theta+8(1-\theta)+6(1-\theta)]=\frac{1}{10}[\theta+8-8\theta+6-6\theta]$

$=\frac{1}{10}[14-13\theta]=3.3$

Therefore,

$\mu_1=\frac{3}{2}=\frac{6+5\theta}{10}$

$\mu_2=\frac{33}{10}=\frac{14-13\theta}{10}$

b) $\boxed{\text{Standard Error of Estimates}}$

$\sigma_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}}=\sqrt{\frac{E[X^2]-E[X]^2}{n}}=\sqrt{\frac{33/10-9/4}{10}}=\sqrt{\frac{\frac{132}{40}-\frac{90}{40}}{10}}=\sqrt{\frac{11/10}{10}}=\sqrt{0.11}$

c) $\boxed{\text{Maximum Likelihood of Estimate}}$ $lik(\theta)=f(X_1,X_2,\ldots X_n|\theta)=\prod_{i=1}^{n}P(X_i|\theta)$

$lik(\theta)=\prod^{3}P(X=i)=\frac{2}{3}\theta\cdot\frac{1}{3}\theta\cdot\frac{2}{3}(1-\theta)\cdot\frac{1}{3}(1-\theta)=\boxed{\frac{4}{81}[\theta(1-\theta)]^2}$; $\frac{d\,lik(\theta)}{d\theta}=\frac{4}{81}[2\theta(1-\theta)^2+\theta^2(2[1-\theta])(-\theta)]$

d) $\boxed{\text{Standard Error of Likelihood Estimate}}$

$=\frac{8}{81}[\theta(1-2\theta+\theta^2)+\theta^3-\theta^4]$

$=\frac{8}{81}[\theta-2\theta^2+\theta^2+\theta^3-\theta^4]$

$0=\frac{8}{81}[\theta-\theta^2+\theta^3-\theta^4]$

$0=1-\theta+\theta^2-\theta^3$

iv) Average Family income with a sample size of 500 could be represented as a proportional distribution by region. $\bar{X} = \frac{1}{n}\sum_i X_i$; $Var(\bar{X}_{sp}) = \frac{1}{n}\sum W_e\left(1 - \frac{n_e - 1}{N-1}\right)\sigma^2$

b) i) 100 samples of $n = 400$; Average Family income: $\mu = \frac{1}{N}\sum_{e=1}^{L}\sum_{i=1}^{N_e} X_{ie}$

ii) $\$0\,[\sigma] \cdot \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$ iv) $CDF = \int_{-\infty}^{n} f(x)\,dx = n\sum f(x)^2 - (\sum f(x))^2$

iii) Not applicable

vi) $\bar{x} \pm 1.96 s_{\bar{x}}$ ... vii) ... c) Boxplot:

Histogram:

d) i) see (c) ii) iii) see c) e) f) ... see c)

Chapter 8: Estimation of Parameters and Fitting of Probability Distributions

1.

n	Observed	Poisson
0	5267	5259
1	4436	4413
2	1800	1853
3	534	519
4	111	109
5+	21	957
	12,169	

$\left[f(\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}\right]$; $\lambda = \frac{\text{counts}}{\text{total}} = \frac{10{,}220}{12{,}169} = 0.8398$

The expected counts do match the observed counts.

2. $\lambda = \frac{1168}{300} = 3.89$

n	Frequency	Expected
0	14	24
1	30	46
2	36	60
3	68	59
4	43	46
5	43	30
6	30	16.5
7	14	8.05
8	10	3.5
9	6	1.36
10	4	0
11	1	0
12	1	
13+	0	
	300	

3. a) Estimate λ for each data set.

Total 1: 400 Total 2: 400 Total 3: 400 Total 4: 400

$\boxed{\lambda_1 = 0.6825 \quad \lambda_2 = 1.3225 \quad \lambda_3 = 1.80 \quad \lambda_4 = 4.65}$

b) Concentration #1: $E_1(X) = \lambda_1$; $Var_1(X) = E_1[X^2] - E_1[X]^2 = 1.28 - 0.46 = 0.82$; $SE = \sqrt{\frac{\sigma^2}{n}} = 0.045$

95% confidence Interval: $P\left(\left|\frac{X - \bar{X}}{\sigma_x}\right| \le Z(\alpha/2)\right) = 0.95$; $\bar{x} \pm 1.96\sigma_x$ $\boxed{0.6825 \pm 0.088}$

Concentration #2: $E_2(X) = \lambda_2$; $Var_2(X) = E_2[X^2] - E_2[X]^2 = 3.03 - 1.3225 = 1.71$; $SE = \sqrt{\frac{\sigma^2}{n}} = 0.064$

95% Confidence Interval: $P\left(\left|\frac{X - \bar{X}}{\sigma_x}\right| \le Z(\alpha/2)\right) = 0.95$; $\bar{x} \pm 1.96\sigma_x$ $\boxed{1.33 \pm 0.125}$

Concentration #3: $E_3(X) = \lambda_3$; $Var_3(X) = E_3[X^2] - E[X]^2 = 5.20$

$SE = \sqrt{\frac{\sigma^2}{n}} = 0.114$

95% Confidence Interval: $P\left(\left|\frac{X - \bar{X}}{\sigma_x}\right| \le Z(\alpha/2)\right) = 0.95$; $\bar{x} \pm 1.96\sigma_x$ $\boxed{1.80 \pm 0.223}$

Concentration #4: $E_4(X) = \lambda_4$; $Var_4(X) = E_4[X^2] - E_4[X]^2 = 34.5 - 4.65^2 = 12.86$; $SE = \sqrt{\frac{\sigma^2}{n}} = 0.179$

95% Confidence Interval: $P\left(\left|\frac{x - \bar{x}}{\sigma_x}\right| \le Z(\alpha/2)\right) = 0.95$; $\bar{x} \pm 1.96\sigma_x$ $\boxed{4.65 \pm 0.351}$

c)

k	Observed	Expected	Observed	Expected	Observed	Expected	Ob.	Exp.
0	213	214	103	106	75	66	0	4
1	128	134	143	141	103	119	20	18
2	37	42	98	93	121	109	43	41
3	13	9	42	41	54	64	53	65
4	3	1.36	8	14	30	29	86	75
5	1	0	4	4	13	10	70	69
6	0	0	2	1	2	3	54	54
7	0	0	0	0	1	1	37	36
8	0	0	0	0	0	0	18	21
9	0	0	0	0	1	0	10	11
10	0	0	0	0	0	0	5	5
11	0	0	0	0	0	0	2	2
12	0	0	0	0	0	0	2	1

The expected and observed counts are fitting.

4. Suppose X is a discrete Random Variable: $P(X=0)=\frac{2}{3}\theta$ where $0\le\theta\le1$

$P(X=1)=\frac{1}{3}\theta$ $\quad n=10$ observations

$P(X=2)=\frac{2}{3}(1-\theta)$ $\quad (3,0,2,1,3,2,1,0,2,1)$

$P(X=3)=\frac{1}{3}(1-\theta)$

a) Find the method of moments estimator of θ

① $\boxed{\text{Parameter Estimate}}$

$\mu_k=E(X^k)$; $\mu_1=E(X^1)=\frac{1}{10}[3+0+2+1+3+2+1+0+2+1]=\frac{3}{2}$

$\mu_2=E(X^2)=\frac{1}{10}[3^2+0^2+2^2+1^2+3^2+2^2+1^2+0^2+2^2+1^2]=3.3$

② $\boxed{\text{Moment Estimators}}$

$E(x^1)=\sum_{i=0}^{3} xP(x)=0(\frac{2}{3}\theta)+1(\frac{1}{3}\theta)+2(\frac{2}{3}(1-\theta))+3(\frac{1}{3}(1-\theta))=\frac{1}{3}\theta+\frac{7}{3}(1-\theta)=\boxed{-2\theta+7/3}$

$E(X^2)=\sum_{i=0}^{3} x^2P(x)=0^2(\frac{2}{3}\theta)+1^2(\frac{1}{3}\theta)+2^2(\frac{2}{3}(1-\theta))+3^2(\frac{1}{3}(1-\theta))=\frac{1}{3}\theta+\frac{8}{3}(1-\theta)+3(1-\theta)=\boxed{\frac{-16\theta}{3}+\frac{17}{3}}$

③ $\boxed{\text{Estimates in terms of Moments}}$

If $\frac{3}{2}=-2\theta+7/3$; $\boxed{\theta=\frac{5}{12}}$; then $\boxed{\hat\theta_1=\frac{7}{6}-\frac{\hat\mu_1}{2}}$

If $\frac{33}{10}=\frac{-16\theta}{3}+\frac{17}{3}$; $\theta_2=\frac{71}{240}$; then $\boxed{\hat\theta_2=\frac{171}{160}+\frac{3}{16}\hat\mu_2}$

b) $\boxed{\text{Standard Error of Estimates}}$

$Var(\hat\theta)=Var(\frac{7}{6}-\frac{1}{2}\hat\mu_1)=(-\frac{1}{2})^2Var(\hat\mu_1)=\frac{1}{4}\frac{Var(X)}{n}=\frac{1}{4}\frac{1}{10}[\frac{-16\theta}{3}+\frac{17}{3}-(-2\theta+7/3)^2]$

$=\frac{1}{40}[\frac{-16\theta}{3}+\frac{17}{3}-4\theta^2+\frac{28}{3}\theta-49/9]=\frac{1}{40}[-4\theta^2+4\theta+2/9]$

$\sigma_x=\sqrt{\frac{\sigma^2}{n}}=\sqrt{\frac{E[X^2]-E[X]^2}{n}}=\sqrt{\frac{1}{40}[-4\theta^2+4\theta+\frac{2}{9}]}=\sqrt{\frac{1}{40}[-4(\frac{5}{12})^2+4(\frac{5}{12})+\frac{2}{9}]}=\sqrt{\frac{43/3}{1440}}=\boxed{0.073}$

c) $\boxed{\text{Maximum Likelihood of Estimate}}$

Multinomial because multiple X and $p(x)$; $f(x_1\ldots x_m|p_1\ldots p_m)=\frac{n!}{\prod_{i=1}^{m}x_i!}\prod_{i=1}^{m}p_i^{x_i}$

d) $\boxed{\text{Standard Error of Likelihood Estimate}}$

$\log[f(x_1\ldots x_n|p_1\ldots p_n)]=\log n!-\sum_{i=1}^{m}\log x_i!+\sum_{i=1}^{m}x_i p_i$

$F(x_0\ldots x_3|p_0(\theta)\ldots p_3(\theta))=\left[\prod_{i=0}^{3}p(X=i)^{x_i}=(\frac{2}{3}\theta)^2(\frac{1}{3}\theta)^3(\frac{2}{3}(1-\theta))^3(\frac{1}{3}(1-\theta))^2\right]\frac{n!}{3!}$

$\log f(x_0\ldots x_3|p_0(\theta)\ldots p_3(\theta))=2\log(\frac{2}{3}\theta)+3\log(\frac{1}{3}\theta)+2\log(\frac{2}{3}(1-\theta))+\frac{2}{3}\log\frac{1}{3}(1-\theta)+\log(\frac{n!}{3!})$

$\frac{d\log f(x_0\ldots x_3|p_0(\theta)\ldots p_3(\theta))}{d\theta}=\frac{2(\frac{3}{2})(\frac{2}{3})}{\theta}+\frac{3(\frac{3}{1})(\frac{1}{3})}{\theta}+\frac{2(\frac{3}{2})(\frac{-2}{3})}{(1-\theta)}+\frac{\frac{2}{3}(\frac{3}{1})(\frac{-1}{3})}{(1-\theta)}=0$; $\frac{5}{\theta}-\frac{8}{3}\frac{1}{1-\theta}=0$

$5(1-\theta)=\frac{8}{3}\theta$

$15(1-\theta)=8\theta$

$15-15\theta=8\theta$

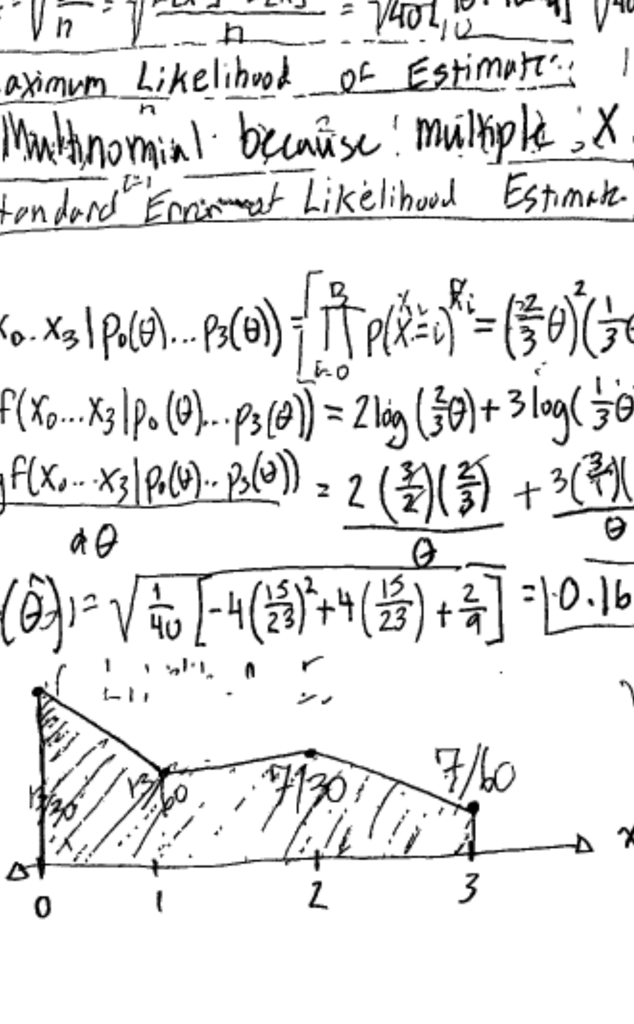
$15=23\theta$

$\frac{15}{23}=\theta$

$\boxed{=0.65}$

$\sigma_{\hat x}(\hat\theta)=\sqrt{\frac{1}{40}[-4(\frac{15}{23})^2+4(\frac{15}{23})+\frac{2}{9}]}=\boxed{0.168}$

e) PE



The mode of the posterior seems to be at $x=1$

5. $P(X=1)=\theta$; $P(X=2)=1-\theta$ || $x_1=1$, $x_2=2$, $x_3=2$ || a) Find the method of moment estimators.

a) $E[X']=\sum_{i=1}^{2} x_i P(X=i) = 1\theta + 2(1-\theta) = 2$; $E[X_i] = \frac{1}{3}\sum x_i = (\frac{1}{3}) + (\frac{2}{3})2 = \boxed{\frac{5}{3}}$

$E[X^2]=\sum_{i=1}^{2} x_i P(X=i) = 1^2\theta + 2^2(1-\theta) = \theta + 4 - 4\theta = 4-3\theta$; $E[X_i^2] = \frac{1}{n}\sum x_i^2 = \frac{1}{3}(1) + (\frac{2}{3})2^2 = \boxed{9/3}$; $4-3\theta = \frac{9}{3}$; $\boxed{\theta = \frac{1}{3}}$

b) $l(x_1, x_2 | p_1, p_2) = \frac{n!}{\prod_{i=1}^{2} x_i} \prod_{i=1}^{2} p_i^{x_i} = \frac{2!}{2!}(\theta)^1(1-\theta)^2 = \boxed{\theta(1-\theta)^2}$

c) $\frac{dl(x_1,x_2|p_1,p_2)}{d\theta} = (1-\theta)^2 - 2\theta(1-\theta) = (1-\theta)(1-\theta-2\theta) = (1-\theta)(1-3\theta) = 0$; $\boxed{\hat{\theta} = 1 \text{ or } 1/3}$

d) $a=2, b=3$, $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$; $f(x) = \frac{\Gamma(5)}{\Gamma(2)\Gamma(3)} \cdot x^1(1-x)^2 = \frac{123}{\Gamma(2)} x^1(1-x)^2$

$5 = \frac{12}{\Gamma(2)}$ $\boxed{B(2,3)}$

6. Suppose $X \sim bin(n,p)$ a) show mle of p is $\hat{p} = X/n$

a) $p(x) = \binom{n}{x} \prod_{i=1}^{2} p_i^{x_i}$ $\log p(x) = \log n! - \sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} x_i \log p_i$; $1 = \sum_{i=1}^{2} p_i$; $\boxed{\text{Lagrangian}}$

b) Cramér-Rao Lower Bound: $\log(n!) - \sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} x_i \log p_i + \lambda(\sum_{i=1}^{2} p_i - 1)$

Measure of concentration: $\frac{d \log p(x)}{dp} = \sum_{i=1}^{2} \frac{x_i}{p_i} + \lambda = 0$; $\lambda = -\sum_{i=1}^{2} \frac{x_i}{p_i}$; $\sum_{i=1}^{2} p_i = \frac{-\sum x_i}{\lambda}$; $1 = \frac{-n}{\lambda}$

$MSE(\theta) = E(\hat{\theta} - \theta_0)^2$ Cramer-Rao Inequality: $var(T) \geq \frac{1}{n I(\theta)}$ $\lambda = -n$

$= var(\hat{\theta}) + (E(\hat{\theta}) - \theta_0)^2$ $\boxed{p_j = \frac{x_j}{n}}$

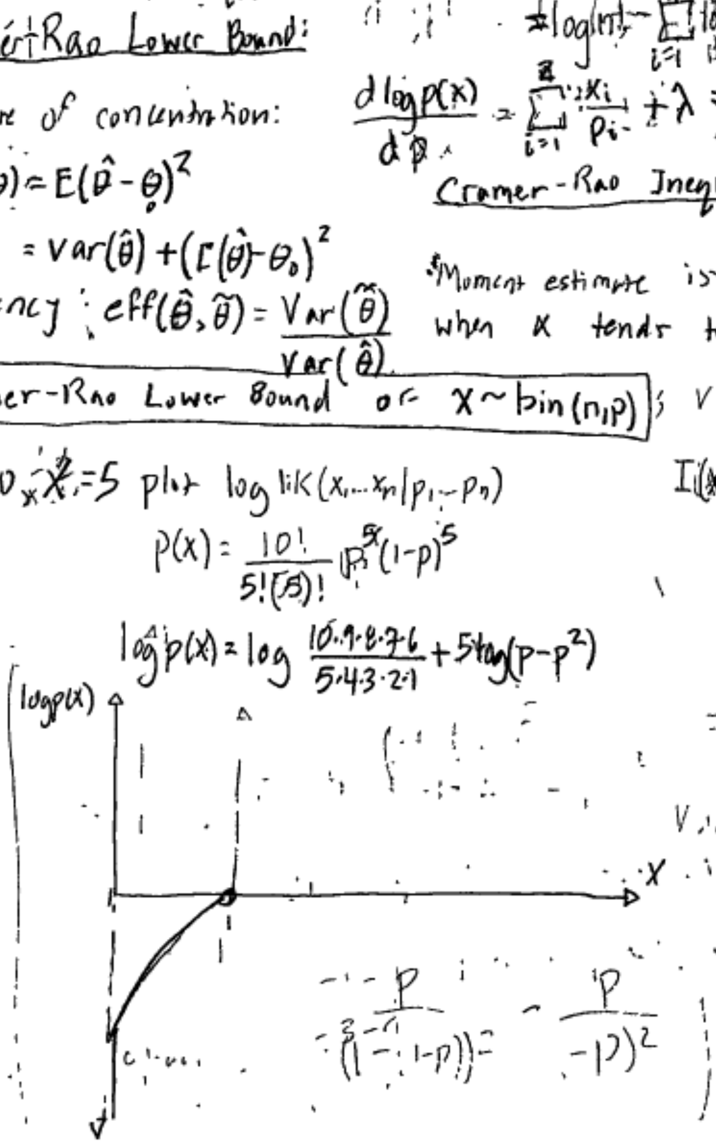
Efficiency: $eff(\hat{\theta}, \tilde{\theta}) = \frac{Var(\tilde{\theta})}{Var(\hat{\theta})}$ *Moment estimate is similar to mle when α tends to 1.

$\boxed{\text{Cramer-Rao Lower Bound of } X \sim bin(n,p)}$; $var(T) \geq \frac{1}{n I(\theta)}$ where $I(\theta) = E\left[\frac{\partial}{\partial\theta} \log bin(n,p)\right]^2$

c. $n = 10$, $x = 5$ plot $\log lik(x_1 \ldots x_n | p_1 \ldots p_n)$ $I(x) = E\left[\frac{\partial}{\partial p} \log\left[\frac{n!}{\prod x_i} \sum_{i=1}^{n} p_i^{x_i}\right]\right]^2$

$p(x) = \frac{10!}{5!(5)!} p^5 (1-p)^5$ $= E\left[\frac{x}{p} - \frac{(n-x)}{(1-p)}\right]^2 = E\left[\frac{x(1-p) - p(n-x)}{p(1-p)}\right] = E\left[\frac{x - pn}{p(1-p)}\right]^2$

$\log p(x) = \log \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} + 5\log(p - p^2)$ $= \frac{E[X - pn]^2}{p^2(1-p)^2} = \frac{Var(X)}{p^2(1-p)^2} = \boxed{\frac{n}{p(1-p)}}$



$Var(T) \geq n^2$

7a. Geometric Distribution $p(X=k) = p(1-p)^{k-1}$ $\mu_2 = \frac{p}{n}\sum_{k=1}^{\infty} (1-p)^{2(k-1)} =$

Methods of Moments:

$\mu_1 = \frac{1}{n}\sum_{k=1}^{\infty} p(1-p)^{k-1} = \frac{p}{n}\sum_{k=1}^{\infty}(1-p)^{k-1} = \frac{p}{n}\sum_{k=0}^{\infty}(1-p)^{k} = \frac{p^2}{n}\frac{d}{dp}\sum_{k=0}^{\infty}(1-p)^k = \frac{p^2}{n}\frac{d}{dp}\left(\frac{1-p}{p}\right) = \frac{p^2}{p^2}\frac{-1}{(1-p)^2}$

b. Maximum likelihood of p: $P(X=k) = p(1-p)^{k-1}$; $\frac{dp(x=k)}{dp} = (1-p)^{k-1} + (k-1)p(1-p)^{k-2} = 0$

$(1-p)^{k-1} = (k-1)(p)(1-p)^{k-2}$

$(1-p) = (k-1)(p)$; $1-p(1+(k-1)) = 0$

$1 - pk = 0$

$\boxed{\frac{1}{k} = p}$

c. Asymptotic variance of mle:

$I(\theta) = E\left[\frac{\partial}{\partial\theta}\log f(X|\theta)\right]^2$; $I(\theta) = E\left[\frac{d}{dp}\log p + (x-1)\log(p-1)\right]$

$= E\left[\frac{1}{p} - \frac{x-1}{(1-p)}\right]^2 = E\left[\frac{1}{p^2} - \frac{2(x-1)}{p(1-p)} + \frac{(x-1)^2}{(1-p)^2}\right] = \boxed{1/p^2(1-p)}$

$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right]$; $I(x) = -E\left[\frac{\partial^2}{\partial p^2}\log p(1-p)^{k-1}\right] = -E\left[-\frac{1}{p^2} - \frac{x-1}{(1-p)^2}\right]$

$nI(\theta)$ [illegible]

d) Let p be uniform from $[0,1]$. What is the posterior distribution? What is the posterior mean?

$\boxed{\text{Posterior Distribution}}$: $f_{\theta|X}(\theta|X) = \frac{f_{X,\theta}(x,\theta)}{f_X(x)} = \frac{f_{X|\theta}(x|\theta)f(\theta)}{\int f_{X|\theta}(x|\theta)f_\theta(\theta)d\theta}$; Posterior $\propto$ likelihood $\times$ prior.

Posterior Mean: Most probable value of the posterior mode [Requires calculation]

Prior of a Geometric Function $[f(\theta)]$: $\beta eta(1,1) = \frac{1}{\beta(1,1)}p^{1-1}(1-p)^{1-1} = 1$

Likelihood of a Geometric Function $[f_{X|\theta}(x|\theta)]$: $p(x|\theta) = (1-\theta)^{x-1}\theta$

Posterior Distribution $[f_{\theta|X}(\theta|X)]$: $X=1$, $\boxed{f_{\theta|X}(\theta|X) \propto \theta(1-\theta)^{x-1}}$

$X=2$, Prior: $\beta eta(2,x) = \theta^{2-1}(1-\theta)^{x-1}$

likelihood: $p(x|\theta) = \cancel{1-\theta}(1-\theta)^{x-1} = \theta^2(1-\theta)^{2-1}$

Posterior: $\theta(1-\theta)^1\,\theta(1-\theta)^{x-1} = \theta^3(1-\theta)^x = \boxed{\beta(a+2, b+2)}$

Posterior Mean: Expectation of a Beta Distribution - $\boxed{E(X) = \frac{a}{a+b}}$;

8.

Number of Hops	Frequency	$P(\hat{p})$	[illegible]		[illegible]	σ_x
1	48	0.37	0.37		0.62	0.18
2	31	0.24	0.18		1.05	0.303
3	20	0.15	0.11		1.77	0.51
4	9	0.07	0.06		3.97	1.15
5	6	0.05	0.04		5.62	1.62
6	5	0.04	0.03		7.09	2.04
7	4	0.03	0.02		9.47	2.73
8	2	0.02	0.02		14.29	4.13
9	1	0.01	0.01		28.72	8.28
10	1	0.01	0.01		28.72	8.28
11	2	0.02	0.02		14.29	4.13
12	1	0.01	0.01		28.72	8.28

Total: 130

$f(X=k) = p(1-p)^{k-1}$

confidence interval for p.

[illegible] $+ 4 \cdot 0.07 + \ldots + 6 \cdot \ldots + 7 \cdot 0.0 \ldots + 1 \cdot 0.01 + \ldots = |1.27|$

$(x-k) - 2.17$

a) Fit a geometric distribution to the data

b) Find a 95% confidence interval

$E(X) = \frac{1}{130}[1 \cdot 48 + 2 \cdot 31 + 3 \cdot 20 + 4 \cdot 9 + 5 \cdot 6 + 6 \cdot 5 + 7 \cdot 4 + 8 \cdot 2 + 9 + 10 + 11 \cdot 2 + 12 \cdot 1]$

$= 2.77$

$\bar{p} = \frac{1}{\bar{x}} = 0.36$

b) Approximate 95% confidence Interval: $\boxed{\text{Likelihood}}$ $L(p) = \prod_{i=1}^{n} p(1-p)^{x_i - 1}$

$\boxed{\text{Log likelihood}}$ $\log L(p) = \sum_{i=1}^{n} \log p + (x_i - 1)\log(1-p) = n\log p + n\log(1-p) + \log(1-p)\sum x_i$

$\boxed{\text{Maximum log Likelihood}}$ $\log L(p)' = \frac{n}{p} + \frac{n}{1-p} - \frac{\sum x_i}{(1-p)}$; $\bar{p} = \frac{1}{\sum x_i}$ $\boxed{\text{Mean}}$

$$\log L(p)'' = \frac{-n}{p^2} + \frac{n}{(1-p)^2} - \frac{\sum x_i}{(1-p)^2} \quad @ \ \bar{p} = \frac{1}{\sum x_i}$$

$$= \frac{n}{p^2(1-p)^2}\left[-(1-p)^2 + p^2 - \bar{x}p^2\right] = \frac{-n}{\bar{p}^2(1-p)}$$ Variance $V = \frac{1}{\log L(p)''} @ \bar{p}$ $\boxed{= \frac{\bar{p}^2(1-p)}{17}}$

$\boxed{\text{95\%-Confidence Interval}}$ $\boxed{95\%\ CI = \bar{p} \pm 1.96\sigma = 0.36 \pm 0.5}$

c) Goodness of Fit (Comparison)

# Hops	Frequency	Fit
1	48	47
2	31	30
3	20	19
4	9	12
5	6	8
6	5	5
7	4	3
8	2	2
9	1	1
10	1	1
11	2	1
12	1	0

$P(x) = n\bar{p}(1-\bar{p})^{k-1}$

$\sum x = 327$

$n = 129$

$\bar{x} = 2.53$

$\bar{p} = \frac{1}{\bar{x}} = 0.394$

$(O-E)^2$	$(O-E)^2/E$
0.154	0.21
0.318	0.33
0.573	0.33
0.86	0.75
0.71	0.25
0.0	0
0.257	0.33
0.0	0
0.01	0
0.0	0
0.01	1
Undefined	Undefined

χ^2 cdf: 1.71

Degree Freedom: 5

$p \sim 0.8$

d) Posterior Distribution, Posterior Mean, and Standard Deviation.

$\boxed{\text{Prior}}$ $Beta(1,1) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} = x(1-x)$

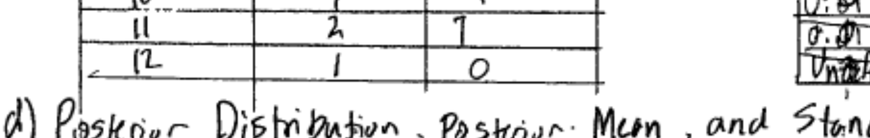
$\boxed{\text{Likelihood}}$ $p(x|\theta) = (1-\theta)^{x-1}\theta$ $\boxed{\text{Posterior}}$ $\pi(\theta|x,a,b) = (1-\theta)^{x-1}\theta \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$

$\boxed{\text{Posterior Mean}}$ $\frac{a}{a+b}$ $\boxed{\text{Posterior S.D.}}$ $\sqrt{\frac{ab}{(a+b)^2(a+b+1)}}$ $= (1-0.394)^{x-1}(0.394)\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$

9 $\lambda = E(X)$; $\hat{\mu}_1 = \hat{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\hat{\lambda} = \bar{X}$ ... The mean of a probability distribution is considered a random variable while the standard deviation is not.

10. Normal Approximation of a Poisson Distribution. $P(\hat{\lambda} = v) = \frac{(n\lambda_0)^{nv} e^{-n\lambda_0}}{(nv)!}$

What is $P(|\lambda_0 - \hat{\lambda}| > \delta)$ for $\delta = 0.5, 1, 1.5, 2,$ and 2.5.



$P(|\lambda_0 - \hat{\lambda}| > \delta) = P(|\lambda_0 - \bar{x}| > \delta) = P\left(\frac{\sqrt{n}|\lambda_0 - \bar{x}|}{\sqrt{\lambda_0}} > \sqrt{n}\frac{\delta}{\sqrt{\lambda_0}}\right)$

$\approx P\left(|N(0,1)| > \sqrt{n}\frac{\delta}{\sqrt{\lambda}}\right)$

$\cong P\left(|N(0,1)| > \frac{\sqrt{23}}{\sqrt{24.9}}\delta\right)$

Standard Error: $\hat{\sigma}_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{\bar{\sigma}^2}{n}} = 1.04$ $= 2(1 - \Phi(\sqrt{23/24.9} \times \delta))$

$\delta = 0.5$: 63% ; $\delta = 1$: 33% ; $\delta = 1.5$: 28% ; $\delta = 2$: 2.8% $\delta = 2.5 = 0.4\%$

11. $s_{\hat{\lambda}} = \sqrt{\frac{\hat{\lambda}}{n}}$; Poisson Distribution : $\frac{\lambda^k}{k!} \cdot e^{-\lambda}$ $n = 23$, $\lambda = 24.9$

The bootstrap method of sampling a large population, then averaging the estimates will approach s_λ because $\boxed{\lim_{B \to \infty} \frac{1}{B} \sum_{i=1}^{B} [\sum_{i=1}^{n} x_i / n] = \frac{\sum^{B} x_i}{N} = \bar{x}}$

12. The method of moments is best when n is small, but approaches maximum likelihood estimates at $n \to \infty$. The answer of choice depends on n-amounts.

13. Example D : 8.4 : $f(x|\alpha) = \frac{1+\alpha x}{2}$, $-1 \le x \le 1$, $-1 \le \alpha \le 1$; $x = \cos\theta$ $|\alpha| \le \frac{1}{3}$

a) $\mu = \int_{-1}^{1} x \frac{1+\alpha x}{2} dx = \frac{\alpha}{3}$; Thus, $\hat{\alpha} = 3\bar{X}$ $E[\hat{x}^2] = \int_{-1}^{1} x^2 \frac{1+\alpha x}{2} dx$

$= \int_{-1}^{1} \frac{x^2}{2} + \frac{\alpha x^3}{2} dx = \frac{x^3}{6}\Big|_{-1}^{1} + \frac{\alpha x^4}{8}\Big|_{-1}^{1}$

Show $E(\hat{\alpha}) = \alpha$; $E(\hat{\alpha}) = E(3\bar{x}) = 3E(\bar{x}) = 3 \cdot \frac{\alpha}{3} = \boxed{\alpha}$

b) Show $Var(\hat{\alpha}) = (3-\alpha^2)/n$; $Var(\hat{\alpha}) = Var(3\bar{x})$

$= \frac{1}{6} + \frac{1}{6} + \frac{\alpha}{8} - \frac{\alpha}{8} = \frac{1}{3}$

$= 9 \, var(\bar{x}) = \frac{9}{n}\left[E[X^2] - E[X]^2\right] = \frac{9}{n}\left[\frac{1}{3} - \alpha^2\right]$

$\boxed{= (3-\alpha^2)/n}$

c) $n = 25$, $\alpha = 0$. $P(\frac{|x-\bar{x}|}{\sigma_{\bar{x}}} > 0.5) = P(\frac{|x-\bar{x}|}{(3-\alpha^2)/n^{1/2}} > 0.5) = P(\frac{|x-\bar{x}|}{3/25^{1/2}} > 0.5) = P(|x-\bar{x}| > 0.012) = 0.5040$ $= 0.012$

14. Example C : Section 8.5 :

a) $P(|\hat{\alpha} - \alpha_0| > 0.05) = 0.5160$ ... 7.

Through comparing the probability to a normal table,

By comparing the expectation, variance, and standard error.

c) $P(|\hat{\alpha} - \alpha_0| > \Delta) = 0.5$ ... 7

By comparing this probability of the norm to $= \Delta$

15. $F(q_{0.25}) = 0.75$. Gamma Distribution : $g(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}$; Upper quantile is dependent on $q(\alpha, \lambda)$.

Example C of Section 8.5:

To estimate the standard error of $q(\alpha, \lambda)$ with the bootstrap method, then $q(\hat{\alpha} - \alpha_0, \hat{\lambda} - \lambda)$ should be evaluated.

16. $f(x|\sigma) = \frac{1}{2\sigma} \exp\left(\frac{-|x|}{\sigma}\right)$ a) Find the method of moments of σ

$\mu_2 = E[x|\sigma] = \int_{-\infty}^{\infty} \frac{|x|^2}{2\sigma} \exp\left(\frac{-|x|}{\sigma}\right) dx = \int_0^\infty \frac{|x|^2}{\sigma} \exp\left(\frac{-|x|}{\sigma}\right) dx = 2\sigma^2$

b) Find the maximum liklihood estimate σ

$2\sigma^2 = \frac{1}{n} \sum_i^n (x_i - \bar{x})^2$

$\boxed{\hat{\sigma} = \sqrt{\frac{1}{2n} \sum_i^n (x_i - \bar{x})^2}}$

$\log \prod f(x_i|\sigma) = \log \prod \frac{1}{2\sigma} \exp\left(\frac{-|x_i|}{\sigma}\right) = \sum \log \frac{1}{2\sigma} \sum \exp\left(\frac{-|x_i|}{\sigma}\right)$

$= n \log\left(\frac{1}{2\sigma}\right) + \sum_i^n \frac{|x_i|}{\sigma} = -n \log 2\sigma - \sum_i^n \frac{|x_i|}{\sigma}$

$\frac{\partial \log L(\sigma|x)}{\partial \sigma} = -\frac{2n}{2\sigma} + \frac{\sum_i^n |x_i|}{\sigma^2} = 0$; $\boxed{\hat{\sigma} = \frac{\sum_i^n |x_i|}{n}}$ "Asymptotic Unbiased"

c) Find the asymptotic variance of mle.

$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} \log f(x|\theta)\right]^2 = -E\left[\frac{\partial}{\partial\sigma}\left[\frac{n}{\sigma} + \frac{\sum_i^n |x_i|}{-\sigma^2}\right]\right] = -E\left[\frac{-n}{\sigma^2} - \frac{2\sum_i^n |x_i|}{\sigma^3}\right] = -2\int_0^\infty \frac{-n}{\sigma^2}\left(\frac{1}{2\sigma}\right)\exp\left(\frac{-|x|}{\sigma}\right) + \frac{2}{\sigma^3}\int_0^\infty |x| \frac{\exp\left(\frac{-|x|}{\sigma}\right)}{2\sigma}$

$= \frac{n}{\sigma^2} \ldots \frac{1}{nI(\theta_0)} = \boxed{\frac{\sigma^2}{n2}}$ ... $\frac{1}{nI(\theta_0)} = \frac{1}{n2[\ldots]}$

d. Find the sufficient statistic. Sufficient statistic $[T]$ is the limit of knowledge for $x_1 \ldots x_n$

$\boxed{\text{Possibly when the variance of } Var(\theta) = \frac{\sigma^2}{n} \text{ is set to } 1 \text{ (e.g. } n = \sigma^2).}$

17. $f(x|\alpha) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}[x(1-x)]^{\alpha-1}$ where $\alpha > 0$. $E(x) = \frac{1}{2}$; $Var(x) = \frac{1}{4(2\alpha+1)}$

a. The shape of the density depends on the α-variable through adjustment of the community.

b. The method of moments will aid the estimation of α through the variance of the second-moment.

c. Maximum Likelihood estimates denote the equation $\boxed{\log f'(x|\alpha) = \frac{(\alpha-1)(2x-1)}{x(x-1)} + 2\Psi^{(0)}(2\alpha)}$

d. The asymptotic variance $I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\alpha)\right] = -E\left[\frac{\partial}{\partial\alpha}\left[\frac{(\alpha-1)(2x-1)}{(\alpha-1)(x)} + 2\Psi^{(0)}(2x)\right]\right]$

e). The sufficient statistic is:

$$= -E\left[4\Psi^{(1)}(2x) - \frac{(\alpha-1)(2x^2-2x+1)}{(x^2)(x-1)^2}\right]$$

$$\frac{f(x,\alpha)}{f(\alpha)} = \frac{\Gamma(2\alpha)/\Gamma(\alpha)^2[x(1-x)]^{\alpha-1}}{\Gamma(2\alpha)/\Gamma(\alpha)^2\binom{n}{s}[x(1-x)]^{\alpha-1}} \quad = -2\underbrace{\left[\int_0^\infty 4\Psi^{(1)}(2x)\frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}x(x-1)^{\alpha-1}dx - \int_0^\infty \frac{(\alpha-1)(2x^2-2x+1)}{x^2(x-1)^2}\frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}x(x-1)^{\alpha-1}dx\right]}_{b}$$

$\boxed{= 1/\binom{n}{s}}$

$$\boxed{Var(\hat\theta - \theta_0) = \frac{1}{4(2(\hat\alpha - \alpha_0) - 1)} \cong \frac{1}{nI(\theta)} = \boxed{\frac{1}{nb}}}$$

18. Suppose

$$f(x|\alpha) = \frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)}x^{\alpha-1}(1-x)^{2\alpha-1}; \alpha > 0$$

$E(X) = \frac{1}{3}$; $Var(X) = \frac{2}{9(3\alpha+1)}$

a) Method of Moments Estimate for α.

$\mu_1 = \frac{1}{3}$; $\mu_2 = \sigma^2 + \mu_1^2 = \frac{2}{9(3\alpha+1)} + [1/3]^2$

b) What is the maximum log likelihood?

$\mu_2 - \frac{1}{9} = \frac{2}{9(3\alpha+1)}$; $9(3\alpha+1) = \frac{2 \cdot 9}{9\mu_2 - 1}$

$l(\alpha) = \log\prod f(x|\alpha) = \sum_i \log f(x|\alpha) = \sum \log \frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)}x^{\alpha-1}(1-x)^{2\alpha-1}$ $\quad 27\alpha + 9 = \frac{18}{9\mu_2 - 1}$

$= n\log\frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)} + (\alpha-1)\sum_i \log x_i + (2\alpha-1)\sum_i \log(1-x)$ $\quad \boxed{\alpha = \frac{2}{3}\left(\frac{1}{9\mu_2 - 1}\right) - \frac{1}{3}}$

$\frac{\partial l(\alpha)}{d\alpha} = \left[\frac{n\Gamma(3\alpha)'}{\Gamma(3\alpha)}(3) - \frac{n\Gamma(\alpha)'}{\Gamma(\alpha)} - \frac{n\Gamma(2\alpha)'}{\Gamma(2\alpha)}(2)\right] + \sum_i \log x_i + 2\sum_i \log(1-x_i) = 0$ $\quad \boxed{\hat\alpha = \frac{2}{3}\left[\frac{1}{9(\hat\sigma^2 + \bar{x}^2) - 1}\right] - \frac{1}{3}}$

$\boxed{\alpha = \text{Solve for alpha.}}$

c) Compute the asymptotic variance. $I(\alpha) = -E\left[\frac{\partial^2}{\partial\alpha^2}l(\alpha)\right] = \alpha_0$; $\frac{1}{nI(\alpha_0)} = \boxed{\frac{1}{\alpha_0}}$

d) Find the sufficient statistic for α. $L = \prod^n P(x|\alpha) = \left[\frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)}\right]^n\left[\prod x_i\right]^{\alpha-1}\left[\prod[1-x_i]\right]^{2\alpha-1}$

19. The sufficient statistics are $\frac{f(x|\alpha)}{f(\alpha)} = \frac{\boxed{\sum_i \log x_i + \sum_i \log(1-x_i)}}{\sum_i x_i + \sum_i \log(1-x_i)}$

a) Suppose $N(\mu, \sigma^2)$. If μ is unknown, what is the mle of σ

$l(\mu,\sigma^2) = \prod_{i=1}^n \log\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x_i-\mu)^2/2\sigma^2) = -\frac{n}{2}\log 2\pi\sigma^2 + \sum_{i=0}^n (x-\mu)^2/2\sigma^2$

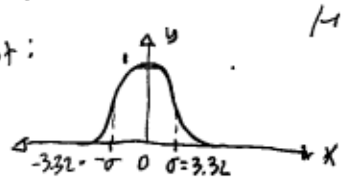
$\frac{dl(\mu,\sigma^2)}{d\sigma^2} = -\frac{n4\pi\sigma}{2(2\pi\sigma^2)} + \sum_{i=1}^n (x_i-\mu)^2/\sigma^3 = -n + \sum_{i=1}^n (x-\mu)^2/\sigma^2 = 0$; $\boxed{\hat\sigma^2 = \sqrt{\frac{1}{n}\sum_i (x-\mu)^2}}$

b. $l(\mu,\sigma^2) = \prod_{i=1}^{n} \log \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$; From part a) $\hat{\sigma}^2 = \sqrt{\frac{1}{n}\sum (x-\mu)^2}$

c. An unbiased estimate of μ would be $\hat{\mu} = \frac{\sum_i X_i}{n} = \bar{X}$ $n\hat{\sigma}^2 = \sum (x-\mu) = \sum X_i - \sum \mu_i$

$$\hat{\mu} = \frac{\sum X_i - n\hat{\sigma}^2}{n} = \boxed{\bar{X} - \sigma^2}$$

20. $X_1, X_2 \ldots X_{25} \to N(\mu,\sigma^2)$, $\mu = 0$, $\sigma = 10$. Plot:



21. $f(x|\theta) = e^{-(x-\theta)}$, $x \geq \theta$, $f(x|\theta) = 0$ otherwise.

a) Method of Moments Estimate of θ.

$$\mu_1 = E(X) = \int_\theta^\infty x e^{-(x-\theta)} dx = \int_\theta^\infty x^{2-1} e^{-(x-\theta)} dx = e^\theta \int_\theta^\infty x^{2-1} e^{-x} dx = e^\theta(\theta+1)e^{-\theta} = \boxed{\theta+1}; \boxed{\hat{\theta} = \mu_1 - 1}$$

$$\mu_2 = E(x^2) = \int_\theta^\infty x^2 e^{-(x-\theta)} dx = e^\theta \int_\theta^\infty x^2 e^{-x} dx = \boxed{(\theta^2 + 2\theta + 2)}$$

b) Maximum Likelihood Estimate: $f(x|\theta) = e^{-(x-\theta)}$; $\frac{dl(x|\theta)}{d\theta} = 1$; $\boxed{\text{Undefined solution. Analytically search for } \min(X_i)}$

c) The sufficient statistic for the function $f(x|\theta)$ is the minimal value of $\boxed{\frac{f(x|\theta)}{f(\theta)} = \min(X_i)}$

22. Weibull Distribution

Cumulative: $F(x) = 1 - e^{-(x/\alpha)^\beta}$, $x \geq 0$, $\alpha > 0$, $\beta > 0$

Weibull Distribution: $f(x) = \frac{\beta}{\alpha} e^{-(x/\alpha)^\beta}$

The Weibull distribution would fit the decay a lifetime by approximating the decay of a process, where α is the decay constant and β is the scaling factor.

To find the standard error of the Weibull fitting, then one must solve the second-moment (or variance), and compare observed to predicted model.

23. $1 \ldots N$, where N = # objects manufactured. A random object is selected with serial number 888.

3:36
3:40
3:45

$$\mu_1 = \sum_{k=1}^{N} k P(N) = \sum_{k=1}^{N} k \frac{1}{N} = \frac{1}{N} \cdot \frac{N(N+1)}{2} = \boxed{\frac{N+1}{2}}$$

$\hat{N} = 2\mu_1 - 1 = \boxed{1775}$

MLE $= \frac{1}{N} = \boxed{\frac{1}{888}}$

Trial	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Face	T	T	T	H	T	H	T	T	H	H	T	H	H	H	T	T	H	T	H	T

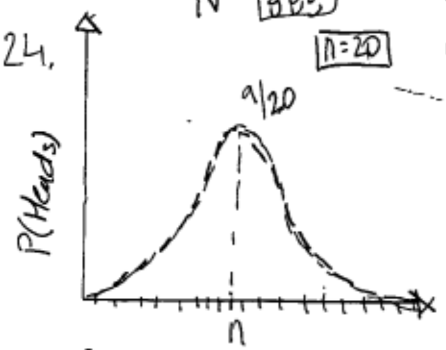
n-heads = 9, n-tails = 11, $\pi = \frac{9}{20}$

25.

Log likelihood of a thumbtack

$\log(P(k)) = \log\binom{n}{k} + n\log p + (n-k)\log(1-p)$

24.



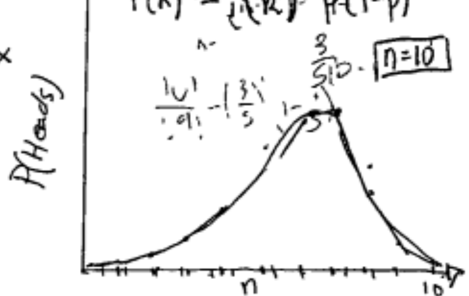
Predicted max at $n = 9/20$, where $p = 9/20$

$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$

Trial	1	2	3	4	5	6	7	8	9	10
Face	H	H	H	T	H	T	T	H	T	H

n-heads = 6, n-tails = 4, $\pi = \frac{3}{5}$

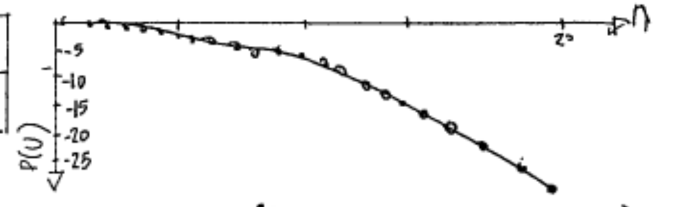
$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$



$n=20$

Trial	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Side	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	U

$P(U)=\frac{1}{20}$, $P(D)=\frac{19}{20}$, $\pi=0.05$

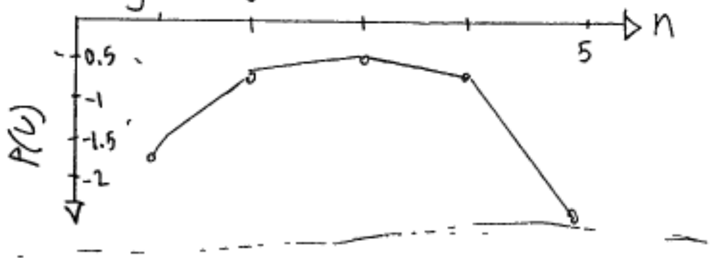


$\log P(k) = \log\binom{20}{k} + k\log p(u) + (n-k)\log(1-p(u))$

$n=5$

Trials	1	2	3	4	5
Side	D	D	D	D	U

$P(u)=\frac{1}{5}, P(D)=\frac{4}{5}, \pi=0.2$



Posterior Distribution of a Binomial:

Posterior $\propto$ Likelihood $\times$ Prior where Prior

$$P(\theta)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

$$\boxed{\propto \binom{n}{k}\theta^{k}(1-\theta)^{n-k}\theta^{a-1}(1-\theta)^{b-1}}$$

likelihood

$$p(k|\theta)=\binom{n}{k}\theta^{k}(1-\theta)^{n-k}$$

1:6:12
1:6:02
1:6:14

Posterior Mean: $E[\theta|k]=\int_0^1 \theta\, p(k|\theta)\,d\theta = \int_0^1 \theta\binom{n}{k}\theta^{k}(1-\theta)^{n-k}d\theta = \int_0^1 Beta(k+1, n+1-k)\,\theta^{k}(1-\theta)^{n-k}d\theta$

26.

$=\frac{k+1}{n+2}$; $k=1, E[\theta|k]=\frac{2}{7}$, $k=2, E[\theta|k]=\frac{3}{7}$, $k=3, E[\theta|k]=\frac{4}{7}$, $k=4, E[\theta|k]=\frac{5}{7}$, $k=5, E[\theta,k]=\frac{6}{7}$

$n_1=100$

$Var[\theta|k]=\frac{(1+k)(n+1-k)}{(n+2)^2(n+3)}$; $k=5, n=20$, $\boxed{Var[\theta|k]=0.0081}$

$n_2=50$, $p(tagged)=\frac{20}{50}$

If tossed 20 more times, then $\theta=p$ would change and shift the distribution.

$$P(tagged)=\frac{\binom{n_1}{t}\binom{n-n_1}{n_2-t}}{\binom{n}{n_2}}$$

Maximum Likelihood

$$=\frac{\binom{100}{20}\binom{n-100}{30}}{\binom{n}{20}};\quad \frac{L_n}{L_{n-1}}=\frac{\binom{n_1}{t}\binom{n-n_1}{n_2-t}}{\binom{n}{n_2}}\cdot\frac{\binom{n-1}{n_2}}{\binom{n_1}{t}\binom{n-n_1-1}{n_2-t}}=\frac{(n-n_1)(n-n_2)}{n(n-n_1-n_2+t)};\quad \frac{n_2 n_1}{n_2-t}=\frac{5000}{30}=\boxed{166}$$

The assumptions about the capture and recapture process include dependent data sets, and no bias to solution.

27.

$$f(t|\tau)=\frac{e^{-t/\tau}}{\tau};\ t\geq 0$$

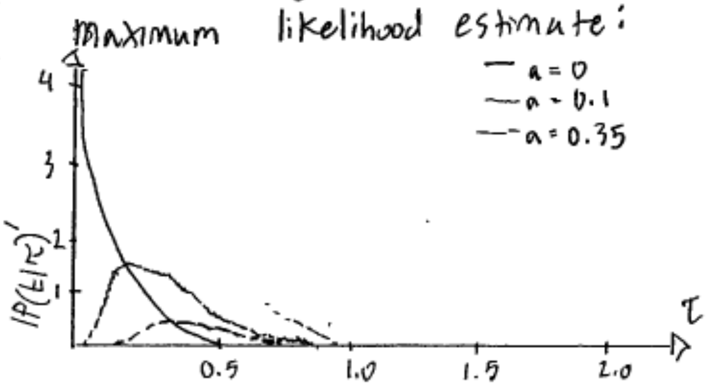
Five components are tested, the first fails in 100 days.

a) The maximum likelihood function of τ is:

$$\boxed{l(f(t|\tau))=\ln\frac{e^{-t/\tau}}{\tau}=\frac{-t}{\tau}-\ln\tau}$$

b) $\frac{d\,l(t|\tau)}{d\tau}=\frac{t}{\tau^2}-\ln\tau=0$; $\boxed{t=\ln\tau^{2\tau}}$

c) The sampling distribution of the maximum likelihood estimate:



d) Standard Error of Maximum Likelihood Estimate:

$$\sigma_{\tau}=\sqrt{\frac{\sigma^2}{n}}=\sqrt{\frac{E[\tau^2]-E[\tau]^2}{n}}$$; seemingly impossible to solve.

Asymptotic variance $Var(\tau)\cong\frac{1}{nI(\tau)}=\frac{1}{-nE\left[\frac{\partial^2}{\partial\tau^2}l(t|\tau)\right]}$

$$I(\tau)=-E\left[-\frac{3t}{\tau^3}-\frac{1}{\tau}\right]=\frac{3\Gamma(2)}{\tau^3}-\frac{1}{\tau}$$

$$\boxed{\sigma_{\tau}=\frac{1}{n}\sqrt{\frac{\tau^{3}}{3\Gamma(2)-\tau^2}}}$$

28. The intervals on the left panel represent 20-trials of a sample size of $n=11$, each with a unique μ and confidence intervals.

29. Yes, variance estimates of 20 trials are represented in Figure 8.8b from a sample sizes of $n=11$. The intervals are short and long because of individual trials, which each have their own confidence interval. Variance with smallest span is trial #4, and largest trial #10.

30. $f(x;\lambda)=\lambda e^{-\lambda x}$ and $E(X)=\lambda^{-1}$; $F(x)=P(X\le x)=1-e^{-\lambda x}$; $x_1=5$, $x_2=3$, $x_3>10$.

a) The likelihood function is: $\boxed{\ln f(x;\lambda)=\ln\lambda-\lambda x}$ b) The mle is $\frac{d\ln f(x;\lambda)}{d\lambda}=\frac{1}{\lambda}-x$; $\boxed{\lambda=\frac{1}{\bar{x}}}$

31.

Trial	#1	#2	#3
Side	T	T	T

George

Trial	#1	#2	#3	#4
Side	T	T	T	H

Hilary

a) $p(x\mid\theta)=\binom{n}{x}\theta^{x_i}(1-\theta)^{n-x}$

$\boxed{p(0\mid\theta)=\binom{3}{0}\theta^{0}(1-\theta)^{3-0}=(1-\theta)^3}$

$p(4\mid\theta)=(1-\theta)^{3}\theta$

$\boxed{p(4\mid\theta)=(1-\theta)^3\theta}$

$p(X,Y\mid\theta)=P(X=0\mid\theta)\cap P(Y=4\mid\theta)=(1-\theta)^3(1-\theta)^3\theta=\boxed{(1-\theta)^6\theta}$

b) The mle of $\hat{\theta}$ is $\frac{d\ln p(x,y\mid\theta)}{d\theta}=\frac{-6}{(1-\theta)}+\frac{1}{\theta}=0$

$6(1-\theta)=\theta$

$\boxed{\theta=\frac{1}{7}}$

32.

a. mean $[\mu]=\frac{\sum x_i}{n}=\frac{57.77}{16}=\boxed{3.61}$; variance $[\sigma^2]=E[x^2]-E[x]^2=\frac{260}{16}-3.61^2=\boxed{3.21}$

b)
$90\%: 3.61\pm1.65\sigma_{\bar{x}}=3.61\pm0.74$

$95\%: 3.61\pm1.96\sigma_{\bar{x}}=3.61\pm0.88$

$99\%: 3.61\pm2.58\sigma_{\bar{x}}=3.61\pm1.16$

$\frac{(n-1)s^2}{\chi^2_{\alpha/2,df}}<\sigma^2<\frac{(n-1)s^2}{\chi^2_{1-(\alpha/2),df}}$

$\frac{(16-1)\sigma^2}{\chi^2_{\frac{0.10}{2},16-1}}=\frac{15\times3.21}{7.26}=6.63<\sigma^2<\frac{(16-1)\sigma^2}{\chi^2_{1-\frac{0.10}{2},16-1}}=\frac{15\times3.21}{25}=1.92$

$\boxed{p=0.10}$ $\boxed{\sigma^2\pm1.29}$

$\frac{(16-1)\sigma^2}{\chi^2_{1-\frac{0.05}{2},16-1}}=\frac{15\times3.21}{27.49}=1.79<\sigma^2<\frac{(16-1)\sigma^2}{\chi^2_{\frac{0.05}{2},16-1}}=\frac{15\times3.21}{6.26}=7.69$

$\boxed{p=0.05}$ $\boxed{\sigma^2\pm4.48}$

$\frac{(16-1)\sigma^2}{\chi^2_{1-\frac{0.01}{2},16-1}}=\frac{15\times3.21}{32.80}=1.47<\sigma^2<\frac{(16-1)\sigma^2}{\chi^2_{\frac{0.01}{2},16-1}}=\frac{15\times3.21}{4.60}=10.47$

$\boxed{p=0.01}$ $\boxed{\sigma^2\pm7.25}$

c)
$90\%: \sigma\pm1.34$

$95\%: \sigma\pm2.12$

$99\%: \sigma\pm2.69$

d) $\frac{1}{2}0.74=1.65\sigma_{\bar{x}}$

$=1.65\sqrt{\frac{\sigma^2}{n}}$

$n=\frac{1.65^2\sigma^2\cdot2^2}{0.74^2}$

$\boxed{n=205}$

33. $X_1,X_2..X_n$ i.i.d $N(\mu,\sigma^2)$. How should c be chosen for $(-\infty,\bar{X}+c)$ to be 95% confidence interval for μ; so that $P(-\infty<\mu\le\bar{X}+c)=0.95$;

$\boxed{c=z(1-\frac{\alpha}{2})\sigma=1.96\sigma}$

34. $X_1,X_2...X_n$ i.i.d. $N(\mu,\sigma^2)$

a) The bootstrap estimate would be $N(\hat{\mu},\frac{\hat{\sigma}^2}{n})$, because this methods distribution is representative of the sampling distribution

b) The bootstrap estimate of $\hat{\mu}-\mu_0$ is $N(0,\hat{\sigma}^2/n)$, due to the fact $\hat{\mu}=\bar{X}$ and $\mu_0=\bar{X}$, therefore $\hat{\mu}-\mu_0=0$.

c) $\boxed{\hat{\mu}\pm z(1-\alpha/2)\hat{\sigma}=\pm z(1-\alpha/2)\hat{\sigma}}$

$\boxed{\mu\pm z(1-\alpha/2)\hat{\sigma}}$

35. $U_1, U_2, \dots, U_{1029}$. $X_1 = U_i \le 0.331$; $X_2 = 0.331 < U_i \le 0.820$; $X_3 = 0.820 < U_i$; Why X_1, X_2 and X_3 multinomial with probabilities 0.331, 0.489, and 0.180 and $n = 1029$? Multinomial Distribution:

$$p(n_1 \dots n_r) = \binom{n}{n_1 \cdots n_r} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

The example of Section 8.5.1 described gene frequencies modeled with Hardy-Weinberg Equations; $1 = [\theta + (1-\theta)]^2 = \theta^2 + 2\theta(1-\theta) + (1-\theta)^2$. If the maximum likelihood estimate is $\hat{\theta} = 0.4247$, then each probability of M, MN, and N representing θ, $2\theta(1-\theta)$, and $(1-\theta)^2$, respectively, show probabilities of existence 0.331, 0.489, and 0.180.

36. The 90% confidence intervals of Example E: Section 8.5.3 were determined to be $(\hat{\alpha} - \bar{\delta}, \hat{\alpha} - \underline{\delta}) = (0.404, 0.523)$, along with $\lambda_{\alpha i} = (1.462, 2.321)$; however, a normal approximation generated $S\alpha = (0.408, 0.443)$ and $S\lambda = \left(\frac{\sqrt{(n-1)s}}{\chi^2_{0.05,df}}, \frac{\sqrt{(n-1)s^2}}{\chi^2_{0.95,df}} \right) = (?, ?)$?

37. Lower Bound $\underline{\theta}$, Upper Bound $\bar{\theta}$ } Quantiles of Distribution θ^*: Prove the bootstrap confidence interval is $(2\hat{\theta} - \bar{\theta}, 2\hat{\theta} - \underline{\theta})$.

$\underline{\theta} = \underline{\theta}^* - \hat{\theta}$; $\bar{\theta} = \bar{\theta}^* - \hat{\theta}$;

$$(\hat{\theta} - \underline{\theta}, \hat{\theta} - \bar{\theta}) = (\hat{\theta} - \underline{\theta}^* + \hat{\theta}, \hat{\theta} - \bar{\theta}^* + \hat{\theta}) = (2\hat{\theta} - \underline{\theta}^*, 2\hat{\theta} - \bar{\theta}^*)$$

38. $P(\underline{\theta} \le \theta^* \le \bar{\theta}) \Rightarrow (\underline{\theta}, \bar{\theta})$

39. If the distribution were considered $\hat{\theta}/\theta$, then the argument would proceed with the bootstrap interval of $\left(\frac{\hat{\theta} - \underline{\theta}}{\theta}, \frac{\hat{\theta} - \bar{\theta}}{\theta} \right)$

40. a) $P(|\hat{\theta} - \theta_0| > 0.01) = 1 - P(-0.01 < \theta - \theta_0 < 0.01) = 0.5080$

Sample the probability distribution 1000 times and determine if 50.80% of the values are found outside the interval -0.01 to 0.01.

b). $E(|\hat{\theta} - \theta_0|) = \frac{1}{n}\sum_i^n (|\hat{\theta} - \theta_0|)\, p(|\hat{\theta} - \theta_0|) = \frac{1}{n}\sum (|\hat{\theta} - \theta_0|)(|\hat{\theta} - \theta_0| - (1 - |\hat{\theta} - \theta_0|))^2$ If sampled 1000 times would generate a substitute expectation for the mle estimate

c) $P(|\hat{\theta} - \theta_0| > \Delta) = 0.5$ would be determined by sampling 1000 times and checking if the points fit between 0.25 and 0.75.

41. Efficiency: Given two estimates, $\hat{\theta}$ and $\tilde{\theta}$, the ratio of variances $var(\tilde{\theta})/var(\hat{\theta})$.

8.4 Example C: Gamma Distribution:

$\mu_1 = \frac{\alpha}{\lambda}$; $\mu_2 = \frac{\alpha(\alpha+1)}{\lambda^2} = \mu_1^2 + \frac{\mu_1}{\lambda}$

$\lambda = \frac{\mu_1}{\mu_2 - \mu_1^2}$; $\alpha = \lambda\mu_1 = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$

$\hat{\lambda} = \frac{\bar{X}}{\hat{\sigma}^2}$ $\hat{\alpha} = \frac{\bar{X}^2}{\hat{\sigma}^2}$

Method of Moments

8.5: Example C: Gamma Distribution

$f(x|\alpha,\lambda) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x}$

$l(\alpha,\lambda) = \sum_{i=1}^n [\alpha \log \lambda + (\alpha-1)\log X_i - \lambda X_i - \log \Gamma(\alpha)]$

$\frac{\partial l}{\partial \alpha} = n \log \lambda + \sum \log X_i - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$

$\frac{\partial l}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum X_i$

$\hat{\lambda} = \frac{\hat{\alpha}}{\bar{X}}$; $\hat{\alpha} = \bar{X}\hat{\lambda}$

Maximum Likelihood Estimate

Efficiency

$eff(\hat{\lambda}) = 0.85$

$eff(\hat{\alpha}) = 0.44$

42. Poisson Distribution:

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$\lambda = 0.004207$ counts/sec

$\sigma_\lambda = \sqrt{\frac{\lambda}{n}} = 0.00004209$ counts/sec

An informal determination that emission rate is constant would be sampling the dataset for similar values.

43.

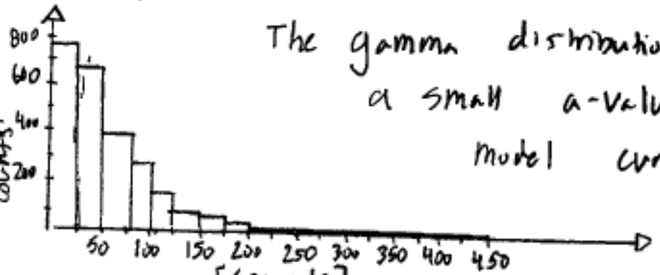
The posterior distribution = likelihood x prior = poisson x gamma.

$$= \prod_{k=0}^{n} \frac{\lambda^k}{k!} e^{-\lambda} \cdot \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

$$= \frac{\lambda^{\Sigma k}}{\prod k!} e^{-\lambda} \cdot \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

$$= \frac{\lambda^{(a-1)+\Sigma k} \cdot b^a}{\prod k! \, \Gamma(a)} e^{-(b+1)\lambda}$$

a) Histogram of Interarrival Times:



The gamma distribution would fit if a small a-value represented model curve.

$\bar{X} = 79.3522$

$\sigma^2 = 6313.291$

b) Method of Moments

Gamma Distribution: $P(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$

$$M(t) = \int_0^\infty e^{tx} \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} dx = \frac{b^a}{\Gamma(a)} \left(\frac{\Gamma(a)}{(b-t)^a}\right) = \left(\frac{b}{b-t}\right)^a$$

$M'(0) = E[X] = \frac{a}{b}$

$M''(0) = E[X^2] = \frac{a(a+1)}{b^2}$

$Var(x) = \frac{a(a+1)}{b^2} - \frac{a^2}{b^2}$

$= a/b^2$

$a = b \cdot E[X]$

$var(x) = \frac{b \cdot E[X]}{b^2}$

$b = \frac{E[X]}{Var(x)} = 1.012$

$a = \frac{E[X]^2}{Var(x)} = 78.980$

The method of moments does not fit, and decays to zero by sight.

Maximum Likelihood

$\ln P(x|a,b) = \ln b^a - \ln\Gamma(a) + (a-1)\ln x - bx$

$\prod \ln P(x|a,b) = n \ln b^a - n\ln\Gamma(a) + (a-1)\sum \ln x - b\sum x$

$\frac{d(\prod \ln P(x|a,b))}{db} = \frac{na}{b} - \sum x = 0; \; b = \frac{a}{\bar{X}}$

$\frac{d(\prod \ln P(x|a,b))}{da} = n\ln a - \ln\bar{x} - n\frac{\Gamma'(a)}{\Gamma(a)} + \sum \ln x = 0$

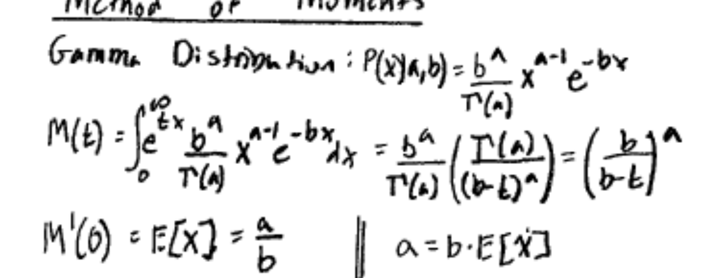
"Solving for roots"

$a = 0.72898, \; 3.93466$

$b = 0.00912, \; 0.049223$

The MLE shows two solutions because of two roots, when graphically evaluating shape(a). The fit is of the wrong scale.

c) Plot of the fittings:



The fits are of wrong scale, but $a = 0.729$, $b = 3.935$ models at half-height.

d) Bootstrap Estimate of S.E

Method of Moments S.E.

$Shape(a) = \bar{x}^2/\sigma^2$

$Scale(b) = \bar{x}/\sigma^2$

The variance does not change for parameter. Ba's $S.E = \sqrt{\frac{Var(x)}{n}} = 17.24$

Maximum Likelihood S.E.

Shape(a) = Solved numerically

Scale(b) = solved via bootstrap.

The variance of parameters depends on precision of mean and standard deviation. S.E. = 1.27

e Confidence Interval of Method of Moments

$P(c_1 \le \frac{n\bar{x}}{a} \le c_2) = 0.95$

$P(\frac{1}{c_1} \le \frac{a}{n\bar{x}} \le \frac{1}{c_2}) = 0.95$

$P(\frac{n\bar{x}}{c_1} \le a \le \frac{n\bar{x}}{c_2}) = 0.95$

$\min\left[\int_{\frac{n\bar{x}}{c_1}}^{\frac{n\bar{x}}{c_2}} Gam(x|a,b)dx = 0.95\right]$ for c_1 & c_2

Confidence Interval of Maximum Likelihood Estimate:

$a = 0.72898, b = 3.93466$ $\quad \min\left[\int_{\frac{n\bar{x}}{c_1}}^{\frac{n\bar{x}}{c_2}} Gam(x|a,b)dx = 0.95\right]$

$a = 3.93466, b = 0.049223$ $\quad \min\left[\int_{\frac{n\bar{x}}{c_1}}^{\frac{n\bar{x}}{c_2}} Gam(x|a,b)dx = 0.95\right]$

F. See plot of (part c)

44.

Gender	Mean Temperature (°F)	Standard Deviation (°F)	Mean Rate (beats/min)	Standard Deviation (beats/min)
Male	98.1	0.69	73.37	5.83
Female	98.39	0.74	74.15	8.04

95% – Confidence Interval

Male: 98.1°F ± 0.18

Female: 98.39°F ± 0.18

Male: 73.37 beats/min ± 1.4

Female: 74.15 beats/min ± 1.97

Folklore of 98.6°F does not fit inside the confidence interval for males or females

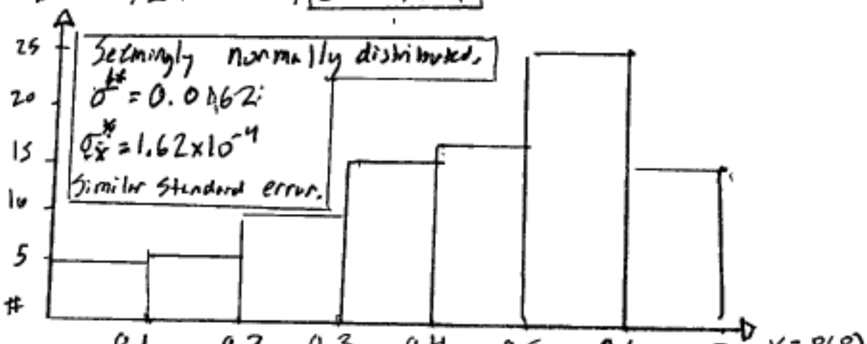
45. Begin on otherside:

h. Proposition D: Section 2.3: Of the domain [0,1] let $X = F^{-1}(U)$, then the cdf of X is F.

Pseudorandom variables were generated of the domain [0,1] with $\Delta r = 0.1$

$B = 100$, $\Delta r = 0.01$, $\boxed{\theta^* = 0.32}$

x	$\theta = 1$, $P(x)$	CDF
0.0	0	0
0.1	0.17	0.17
0.2	0.39	0.56
0.3	0.46	1.02
0.4	0.29	1.31
0.5	0.60	1.91
0.6	0.12	2.03
0.7	0.55	2.58
0.8	0.56	3.14
0.9	0.24	3.38
1.0	0.60	3.99



i. $B = 1000$; $\theta^* = 0.43$. A value generated from large sample theory.

45. a. Maximum Likelihood Estimate: Rayleigh Distribution: $f(r|\theta) = \frac{r}{\theta^2}\exp\left(\frac{-r^2}{2\theta^2}\right)$

Log Rayleigh Distribution: $\ln f(r|\theta) = \sum \ln r_i - 2n\ln\theta - \frac{\sum r^2}{2\theta^2}$

MLE Estimate: $\frac{d\ln f(r|\theta)}{d\theta} = -\frac{2n}{\theta} + \frac{\sum r_i^2}{\theta^3} \Rightarrow \boxed{\hat{\theta} = \left(\frac{1}{2n}\sum r_i^2\right)^{1/2}}$

b. Method of Moments Estimate:

$E[r] = \int_0^\infty \frac{r^2}{\theta^2}\exp\left(\frac{-r^2}{2\theta^2}\right)dr = \frac{1}{\theta^2}\int_0^\infty r^2\exp\left(\frac{-r^2}{2\theta^2}\right)dr = \frac{1}{\theta^2}\int_0^\infty u\exp\left(\frac{-u}{2\theta^2}\right)\frac{du}{2\sqrt{u}} = \frac{1}{2\theta^2}\int_0^\infty u^{\frac{3}{2}-\frac{2}{2}}\exp\left(\frac{-u}{2\theta^2}\right)du$; $u = r^2$; $dr = \frac{1}{2\sqrt{u}}du$

$= \frac{\Gamma(3/2)}{2\theta^2}(2\theta^2)^{3/2} = \sqrt{2}\,\theta\,\Gamma\left(\frac{3}{2}\right) = \sqrt{2}\,\theta \times \left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right) = \theta\left(\frac{\sqrt{\pi}}{\sqrt{2}}\right)$

$\boxed{\hat{\theta} = \frac{\sqrt{2}}{\sqrt{\pi}}E[r]}$

c. Approximate Variance of MLE:

$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f(r|\theta)\right]$

$= -E\left[\frac{\partial}{\partial\theta}\left[-2\left(\frac{1}{\theta}\right) + \frac{r^2}{\theta^3}\right]\right] = -E\left[\frac{2}{\theta^2} - \frac{3r^2}{\theta^4}\right] = \frac{3E[r^2]}{\theta^4} - \frac{2}{\theta^2} = \frac{6\theta^2}{\theta^4} - \frac{2}{\theta^2} = \boxed{\frac{4}{\theta^2}}$ $\quad var(\theta) \approx \frac{1}{nI(\theta)} = \boxed{\frac{\theta^2}{4n}}$

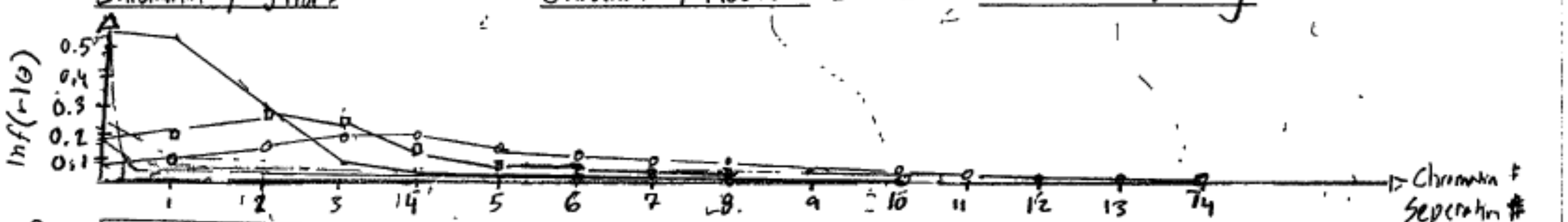
$\boxed{\frac{\sum r_i^2}{8n^2}}$

Approximate variance of Method of Moments:

$I(\theta) = \frac{4}{\theta^2}$; $Var(\hat{\theta}) = \frac{1}{nI(\theta)} = \frac{\hat{\theta}^2}{4n} = \boxed{\frac{\sum r_i^2}{2n^2\pi}}$

d. Plot of likelihood Function

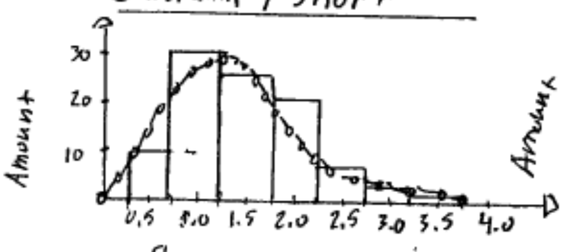
Chromatin / Short ——● Chromatin / Medium □——□ Chromatin / Long



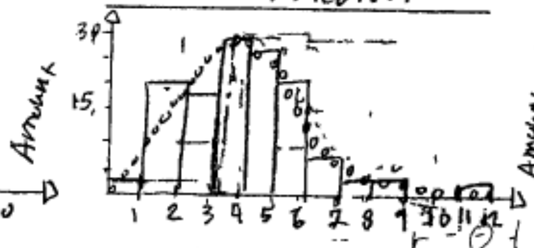
e.

Chromatin	MLE	MLE Asymptotic Variance	MoM	MoM Asymptotic Variance
Short	1.12	3.27×10^{-3}	1.17	3.04×10^{-3}
Medium	3.31	2.09×10^{-2}	3.39	2.18×10^{-2}
Long	2.08	4.34×10^{-3}	2.07	4.30×10^{-3}

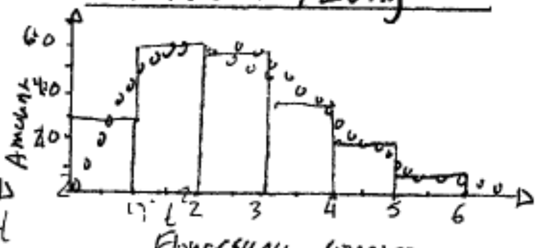
f. Chromatin / Short



Chromatin / Medium



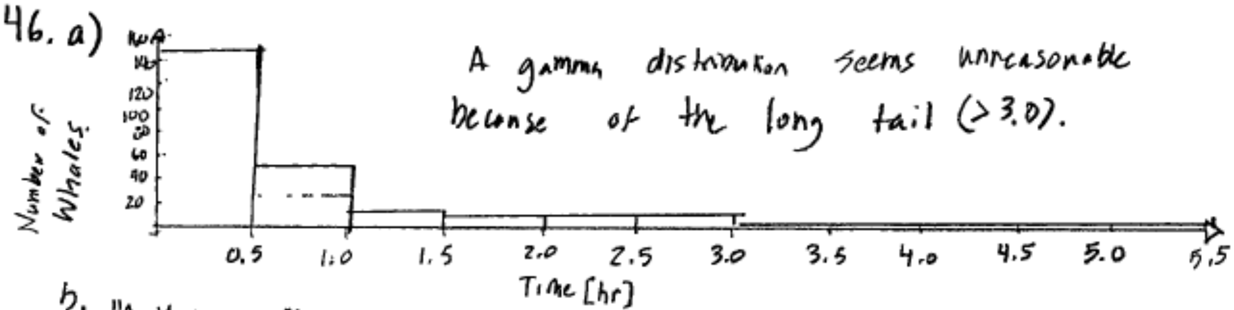
Chromatin / Long



The distributions for the domain [0, 12] fit similarly to data sets.

g. Both short and long strands of DNA show flourescence signals with a mle of ~2 and medium strands ~3.01.

46. a)



A gamma distribution seems unreasonable because of the long tail (>3.0).

b. Method of Moments

$$E[x]=\int_0^\infty x\frac{\beta^\kappa}{\Gamma(\alpha)}x^{\kappa-1}e^{-x\beta}dx=\frac{\beta^\kappa}{\Gamma(\kappa)}\int_0^\infty x^{\kappa}e^{-x\beta}dx=\frac{\beta^\kappa}{\Gamma(\kappa)}\frac{\Gamma(\kappa+1)}{\beta^{\alpha+1}}=\frac{\kappa}{\beta};\ \kappa=\beta E[x];$$

$\kappa=n-1$ $\Gamma(\kappa+1)=\kappa\Gamma(\kappa)$ $n=\beta-1$

$$E[x^2]=\int_0^\infty x^2\frac{\beta^\kappa}{\Gamma(\kappa)}x^{\kappa-1}e^{-x\beta}dx=\frac{\beta^\kappa}{\Gamma(\kappa)}\int_0^\infty x^{\kappa+1}e^{-x\beta}dx=\frac{\beta^\kappa}{\Gamma(\kappa)}\frac{(\kappa+1)\kappa\Gamma(\kappa)}{\beta^{\kappa+2}}=\frac{(\kappa+1)\kappa}{\beta^2};\ Var[x]=E[x^2]-E[x]^2$$

$\kappa+1=n-1$

$$=\frac{(\kappa+1)\kappa}{\beta^2}-\frac{\kappa^2}{\beta^2}=\frac{\kappa}{\beta^2}$$

From data:

Mean = 0.6060,	Variance = 0.4595
Shape(κ) = 0.7991	Rate = 1.3198

$$Var[x]=\frac{\beta E[x]}{\beta^2};\ \boxed{\beta=\frac{E[x]}{Var[x]}}\quad \boxed{\kappa=\frac{E[x]^2}{Var[x]}}$$

c. Maximum Likelihood Estimate:

$$\ln f(x|a,b)=\kappa\ln\beta-\ln\Gamma(\kappa)+(\kappa-1)\ln x-\beta x$$

$$\sum\ln f(x|a,b)=n\kappa\ln\beta-n\ln\Gamma(\alpha)+(\kappa-1)\sum\ln x_i-\beta\sum x_i$$

$$\frac{\sum\ln f'(x|a,b)}{d\kappa}=n\ln\beta-n\frac{\Gamma'(\kappa)}{\Gamma(\kappa)}+\sum\ln x_i=0$$

Solve with roots

$n=210$

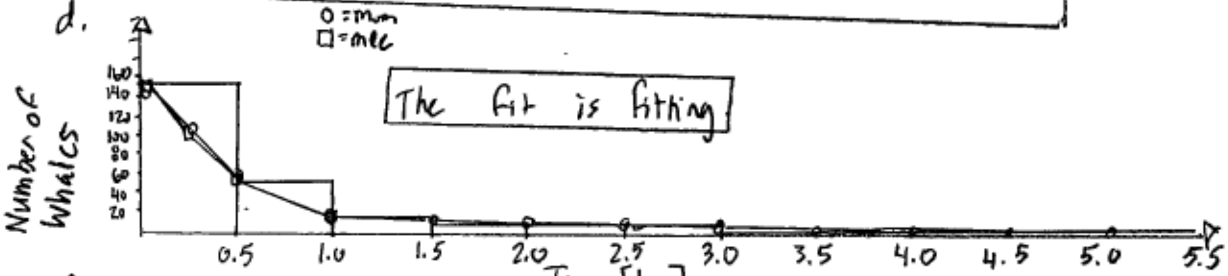
$E[x]=0.6060$

$$\frac{\sum\ln f'(x|a,b)}{d\beta}=\frac{n\kappa}{\beta}-\sum x_i=0;\ \boxed{\frac{\kappa}{E[x]}=\beta};\ \boxed{n\ln\kappa-n\ln E[x]-n\frac{\Gamma'(\kappa)}{\Gamma(\kappa)}+\sum\ln x_i=0}$$

$\boxed{\kappa=0.7569;\ \beta=1.2491}$

The values from M.O.M in part b) are similar to the mle of scale(κ) and shape(β).

d.



The fit is fitting

e. Method of Moments:

Variance: 0.4595; Standard Deviation: 0.6779; $n=210$; Standard Error: $\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}=4.68\times10^{-2}$

Sampling Distribution: $P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}}<Z_{upper}\right)=0.95$, $\boxed{Z_{upper}=1.556}$

f. Maximum Likelihood/Mom:

Variance $=\frac{E[x]^2}{\kappa}=\frac{0.6060^2}{0.7569}=0.4852$; Standard Deviation = 0.6966; Standard Error $=\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}=4.81\times10^{-2}$

Sampling Distribution: $P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}}<Z_{upper}\right)=0.95$; $\boxed{Z_{upper}=1.66}$ Similar results to part c

g. Confidence Interval: 95%; $\boxed{0<\bar{x}<0.700}$

47. Pareto Distribution:

$f(x|x_0,\theta) = \theta x_0^{\theta} x^{-\theta-1}, \; x \geq x_0, \; \theta > 1$

a) Method of Moments Estimate: for θ

$$E[x] = \int_{x_0}^{\infty} x\theta x_0^{\theta} x^{-\theta-1} dx = \frac{\theta x_0^{\theta} x^{-\theta+1}}{-\theta+1}\Big|_{x_0}^{\infty} = \left(\frac{\theta}{1-\theta}\right) x_0^{\theta} x_0^{-\theta+1} = \boxed{\left(\frac{\theta}{1-\theta}\right) x_0}$$

$$E[x^2] = \int_{x_0}^{\infty} x^2 \theta x_0^{\theta} x^{-\theta-1} dx = \frac{\theta x_0^{\theta} x^{-\theta+2}}{-\theta+2}\Big|_{x_0}^{\infty} = \frac{\theta x_0^{\theta}}{-\theta+2} x_0 x^{-\theta+2} = \left(\frac{\theta}{1-\theta}\right) x_0^2$$

$$Var[x] = \left(\frac{\theta}{1-\theta}\right) x_0^2 - \left(\frac{\theta}{1-\theta}\right)^2 x_0^2 = \left(\frac{\theta-\theta^2-\theta^2}{(1-\theta)^2}\right) x_0^2 = \frac{(1-2\theta)\theta}{(1-\theta)^2} x_0^2$$

$$\boxed{\hat{\theta} = \frac{E[x]}{(x_0 + E[x])}}$$

b) Maximum Likelihood Estimate:

$\ln f(x|x_0,\theta) = \ln\theta + \theta\ln x_0 - (\theta+1)\ln x$

$\sum \ln f(x|x_0,\theta) = n\ln\theta + n\theta\ln x_0 - (\theta+1)\sum \ln x_i$

$\frac{d\sum \ln f(x|x_0,\theta)}{d\theta} = \frac{n}{\theta} + n\ln x_0 - \sum \ln x_i = 0$

$$\boxed{\hat{\theta} = \frac{n}{\sum \ln x_i - n\ln x_0}}$$

c) Asymptotic Variance: $I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f(x|x_0,\theta)\right] = -E\left[\frac{\partial}{\partial\theta}\left(\frac{n}{\theta} + n\ln x_0 - \sum \ln x_i\right)\right] = +\frac{n}{\theta^2}$

$Var(\hat{\theta}) \approx \frac{1}{nI(\theta)} = \frac{\hat{\theta}^2}{n^2} \boxed{\frac{1}{(\sum \ln x_i - n\ln x_0)^2}}$

d) Sufficient Statistic $\boxed{X \geq X_0, \text{ and } \theta > 2.}$

48. Observation: $p_0 = P(X=0) = e^{-\lambda}$

Method of Propagation Error: ① Expansion of $F(x)$ about the mean ② First term is the mean ③ Second term is variance

Poisson Distribution: Notes: $Y \sim bin(n, p_0)$

$P(x) = \frac{\lambda^x}{x!}e^{-\lambda}$ $\quad \hat{\lambda} = -\log(Y/n)$; $p_0 = e^{-\lambda} = \frac{Y}{n}$; $E[Y] = \boxed{np_0}$

Approximate Expression of Variance: $\sigma_F^2 = \sum_{i=1}^{n}\left|\frac{df}{dx}\right|^2 \sigma_i^2 = \sum \frac{1}{y^2}\lambda = \boxed{\frac{n\lambda}{y^2}}$

Bias: $\sum \frac{\partial e}{\partial x_i}E[x_i] - \lambda = \boxed{-\frac{n\lambda}{y} - \lambda}$;

Maximum Likelihood Estimate: $\frac{d\sum \ln p(\lambda)}{d\lambda} = \frac{\sum x_i}{\lambda} - n = 0$; $\boxed{\hat{\lambda}_{MLE} = \frac{\sum x_i}{n} = \bar{X}}$

Variance Maximum likelihood Estimate: $Var(\hat{\lambda}_{MLE}) = \frac{\lambda}{n}$ $\quad$ Efficiency: $\frac{Var(\hat{\lambda}_{MLE})}{Var(\hat{\lambda}_{MOP})} = \frac{\frac{\lambda}{n}}{\frac{n\lambda}{y^2}} = \boxed{\frac{1}{1}\frac{y^2}{n^2}}$

49. Muon Decay $\quad$ Binomial Distribution $\quad$ Method of Moments:

a) $f(x|\alpha) = \frac{1+\alpha x}{2}$ $\quad P(k) = \binom{n}{k}p^k(1-p)^{n-k}$

$-1 \leq x \leq 1$

$-1 \leq \alpha \leq 1$

$\int_0^1 f(x|\alpha)dx = \int_0^1 \frac{1+\alpha x}{2}dx = \left[\frac{1}{2} + \frac{\alpha}{4}\right] = \frac{2+\alpha}{4} = p$

$E[x] = \mu = \frac{\alpha}{3}$

Binomial(n,p) = Bin$\left(n, \frac{2+\alpha}{4}\right)$

$E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{2+\alpha}{4} = p$; $Var(4p-2) = 4^2 Var(p) = \boxed{16p(1-p)}$

$\boxed{\alpha = 4p - 2}$

$Var\left(4\frac{X}{n} - 2\right) = \frac{4^2}{n^2}\boxed{Var(p) = \frac{16}{n^2}p(1-p)}$

b) Binomial Variance: $Var[k] = np(1-p)$; Muon Decay variance

K	Val. Bin	Var Bin	Var Mom /n²
0	~~.5~~	3.99	3.99/n²
0.1	0.53	3.96	3.96/n²
0.2	0.55	3.91	3.91/n²
0.3	0.58	3.84	3.84/n²
0.4	0.60	3.75	3.75/n²
0.5	0.63	3.64	3.64/n²
0.6	0.65	3.51	3.51/n²
0.7	0.68	3.51	3.51/n²
0.8	0.70	3.36	3.36/n²
0.9	0.73	3.19	3.19/n²

Maximum Likelihood Estimate

$\frac{d\sum \ln p(k)}{dp} = \frac{\sum x}{p} - \frac{n - \sum x}{1-p} = 0$

$(1-p)\sum x = (n - \sum x)p$

$\sum x - p\sum x = np - p\sum x$

$\boxed{p = \frac{\sum x}{n}}$

$Var(4p-2)$ $\boxed{= 16p(1-p)}$

Method of Moments was more efficient than Binomial Bootstrap or Maximum Likelihood.

50. Rayleigh Distribution

a) $f(x|\theta) = \frac{x}{\theta^2} e^{-x^2/2\theta^2}$

a) Method of Moments Estimate: $x^2 = u$; $dx = \frac{1}{2} u^{-1/2} du$

$$E[X] = \int_0^\infty \frac{x^2}{\theta^2} e^{-x^2/2\theta^2} dx = \frac{1}{\theta^2} \int_0^\infty u e^{-u/2\theta^2} \frac{du}{2\sqrt{u}} = \frac{1}{2\theta^2} \int_0^\infty u^{1/2} e^{-u/2\theta^2} du = \frac{1}{2\theta^2} \int_0^\infty u^{3/2 - 1/2} e^{-u/2\theta^2} du$$

$$= \frac{\Gamma(3/2)}{2\theta^2} (2\theta^2)^{3/2} = \sqrt{2}\theta\, \Gamma(3/2) = \sqrt{2}\theta \cdot \frac{1}{2} \cdot \Gamma(1/2) = \theta \sqrt{\frac{\pi}{2}} ; \boxed{\hat{\theta} = \sqrt{\frac{2}{\pi}} E[X]}$$

b) Maximum Likelihood Estimate:

$$\frac{d \sum \ln f(x|\theta)}{d\theta} = \frac{-2n}{\theta} + \frac{\sum x^2}{\theta^3} = 0$$

$$-2n + \frac{\sum x^2}{\theta^2} = 0 ; \frac{\sum x^2}{2n} = \theta^2 ; \boxed{\hat{\theta} = \sqrt{\frac{1}{2n} \sum x_i}}$$

c) Asymptotic Variance of Maximum Likelihood:

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} \ln f(x|\theta)\right] = -E\left[\frac{\partial}{\partial\theta}\left[-2n + \frac{x^2}{\theta^2}\right]\right]$$

$$= +\frac{2E[x^2]}{\theta^3} ; \text{Var}(\hat{\theta}) \approx \frac{1}{n I(\hat{\theta})} = \boxed{\frac{\hat{\theta}^3}{2E[x^2]}} = \boxed{\frac{\hat{\theta}}{4}}$$

51. Double Exponential Distribution

$$f(x|\theta) = \frac{1}{2} e^{-|x-\theta|} ; -\infty < x < \infty \quad ; n = 2m+1$$

$$\frac{d \sum \ln f(x|\theta)}{d\theta} = \sum \frac{x-\theta}{|x-\theta|} \Rightarrow 0 ; \boxed{\hat{\theta} = \sum x_i}$$

$$\hat{\theta} = \sum_{i=1}^{n} x_i = \sum_{i=1}^{2m+1} x_i = \boxed{\underbrace{\sum_{i=1}^{m+1} x_i}_{1/2} + \underbrace{\sum_{m+1}^{2m+1} x_i}_{1/2}}$$

52. $f(x|\theta) = (\theta+1)x^\theta$; $0 \le x \le 1$ a) Method of Moments Estimate:

$$E[X] = \int_0^1 (\theta+1) x^{\theta+1} dx = \frac{(\theta+1)}{(\theta+2)} x^{\theta+2} \Big|_0^1 = \frac{(\theta+1)}{(\theta+2)} ; \theta(E[x]-1) = 1 - 2E[x]$$

$$\boxed{\hat{\theta} = \frac{1-2E[x]}{E[x]-1}}$$

b) Maximum Likelihood Estimate:

$$\frac{d \sum \ln f(x|\theta)}{d\theta} = \frac{1}{\theta+1} + \sum \ln x_i = 0 ; \boxed{\hat{\theta} = \frac{1 - \sum \ln x_i}{\sum \ln x_i}}$$

c) Asymptotic Variance of MLE:

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} \ln f(x|\theta)\right] = -E\left[\frac{\partial}{\partial\theta}\left[\frac{1}{\theta+1} + \ln x_i\right]\right] = \frac{1}{(\theta+1)^2} ; \text{Var}(\hat{\theta}) = \frac{1}{n I(\hat{\theta})} = \frac{(1+\hat{\theta})^2}{n} = \frac{\left(1 + \frac{1-\sum \ln x_i}{\sum \ln x_i}\right)^2}{n}$$

$$\boxed{= \frac{1}{n(\sum \ln x_i)^2}}$$

d) Sufficient Statistic: $\boxed{x^\theta}$

53. $X_1 \ldots X_n$ uniform on $[0,\theta]$ a) Find the method of moments estimate of θ, mean, and variance.

Uniform Distribution:

$$P(x) = \frac{1}{b-a} \quad [a,b]$$

$$= \frac{1}{\theta - 0} \quad [0,\theta]$$

$$E[x] = \int_0^\theta \frac{x}{\theta} dx = \frac{\theta^2}{2\theta} = \frac{\theta}{2} ; \boxed{\hat{\theta} = 2E[X]} ; E[x^2] = \frac{\theta^2}{3} ; \text{Var}(\hat{\theta}) = \frac{\hat{\theta}^2}{12} ; \boxed{\frac{\hat{\theta}^2}{3}}$$

b) Maximum Likelihood Estimate: $\frac{d \sum \ln p(x)}{d\theta} = \frac{-1}{\theta} = 0 ; \boxed{\hat{\theta} = \infty}$

Thus $\boxed{\theta_{max} = \max(X_1 \ldots X_n)}$

c) $F(x_n|\theta) = n \frac{x^{n-1}}{\theta^n}$; $E[X] = \int_0^\theta \frac{n x^n}{\theta^n} dx = \frac{n}{\theta^n} \int_0^\theta x^n dx = \frac{n}{(n+1)} \theta$

$$E[x^2] = \int_0^\theta \frac{n}{\theta^n} x^{n+1} dx = \boxed{\frac{n\theta^2}{n+2}} ; \text{var}(x) = \boxed{\frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}}$$

$$= \frac{n\theta^2}{(n+1)^2(n+2)}$$

Bias of Maximum Likelihood Estimate:

$$\text{Bias} = E[\hat{X} - \theta] = \theta\left[\frac{n}{n+1} - 1\right] = \boxed{\frac{-\theta}{(n+1)}}$$

Bias of Method of Moments:

$$\text{Bias} = E[x - \theta] = \boxed{0} ; \boxed{\text{Bias of MOM} < \text{MLE}}$$

d) $\hat{\theta} = \max(X_i)$

$$E[\hat{\theta}] = E[\max(x_i)] = \frac{n\theta}{(n+1)}$$

$$\boxed{\theta = \frac{n+1}{n} \max x_i}$$

54. $n = 15$; $\mu = 10$; $s^2 = 25$

90% confidence Interval

$P(-|\frac{X-\mu}{\sigma_X}| < z(\alpha/2)) = 0.90$

$P(0.05 < \frac{x-\mu}{\sigma_x} < 0.95) = 0.90$

$P(-1.65 < \frac{x-\mu}{\sigma_x} <$

$\boxed{\mu \pm 2.13}$

$$P\left(\chi^2_{n-1}(1-\alpha/2) < \frac{n\hat{\sigma}^2}{\sigma^2} < \chi^2_{n-1}(\alpha/2)\right) = 0.90$$

$$P\left(\frac{n\hat{\sigma}^2}{\chi^2_{14,0.95}} < \sigma^2 < \frac{n\hat{\sigma}^2}{\chi^2_{14,0.05}}\right) = 0.90$$

$\boxed{(15.7 < \sigma^2 < 57)}$

55.

Type	Count	Probability
Starchy Green	1997	$0.25(2+\theta)$
Starchy White	906	$0.25(1-\theta)$
Sugary Green	904	$0.25(1-\theta)$
Sugary White	32	0.25θ

Linkage Factors.

a) Multinomial Distribution: $P(\theta) = \frac{n!}{X_1 \cdot X_2 \cdot X_3 \cdot X_4} 0.25(2+\theta)^{X_1} \cdot 0.25(1-\theta)^{X_2+X_3} \cdot 0.25\theta^{X_4}$

$P(X) = \frac{n!}{\prod X_i} \prod p(X_i)^{n_i}$

$\ln P(\theta) = \ln n! - \sum X_i + X_1 \ln\frac{(2+\theta)}{4} + (X_2+X_3)\ln\frac{(1-\theta)}{4} + X_4 \ln\frac{\theta}{4}$

$\ln' P(\theta) = \frac{4X_1}{2+\theta} - \frac{4(X_2+X_3)}{1-\theta} + \frac{4X_4}{\theta} = 0$; $[4X_1\theta + 4X_4(2+\theta)](1-\theta) = 4(X_2+X_3)(2+\theta)\theta$

$(4X_1\theta + 8X_4 + 4X_4\theta) - 4X_1\theta^2 - 8X_4\theta + 4X_4\theta^2$

$= (4X_2\theta + 4X_3\theta)(2+\theta) = 8X_2\theta + 8X_3\theta + 4X_2\theta^2 + 4X_3\theta^2$

$-4(X_1+X_2+X_3+X_4)\theta^2 + 4(X_1 + 2X_2 + 2X_3 + X_4)\theta - 2X_4 = 0$

$3839\theta^2 + 1650\theta + 64 = 0$

$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\ln p(\theta)\right]$

$= -E\left[\frac{-4X_1}{(2+\theta)^2} - \frac{4(X_2+X_3)}{(1-\theta)^2} + \frac{4X_4}{\theta^2}\right]$

$= \frac{-4n(2+\theta)}{4(2+\theta)^2} - \frac{4(1-\theta)+4(1-\theta)n}{4(1-\theta)^2} - \frac{4n\theta}{4\theta^2} = \frac{+n_1}{(2+\theta)} + \frac{n_2+n_3}{2(1-\theta)} + \frac{n_4}{\theta} \frac{\sum n_1+n_2+n_3+n_4}{(2+\theta)2(1-\theta)\theta} \cdot n$ $\boxed{\hat\theta = 0.0357}$

$Var(\hat\theta) = \frac{1}{nI(\theta)} = \frac{1}{5670} = \boxed{8.24\times10^{-5}}$

b) 95% confidence Interval: $\boxed{0.0357 \pm 1.81\times10^{-2}}$

c) Actual $\theta = 4\left(\frac{32}{3839}\right) = 3.33\times10^{-2}$; $SD = \sqrt{np(1-p)} = \sqrt{32 \cdot 0.25(3.33\times10^{-2})(1 - 0.25(3.33\times10^{-2}))} = \boxed{5.15\times10^{-3}}$

2_1

$\boxed{3.33\times10^{-2} \pm 1.01\times10^{-2}}$

56. 1) $\bar X = n(2+\theta_1)/4$ Bias:

$\boxed{\hat\theta_1 = \frac{4\bar X}{n} - 2}$

$E[\hat\theta_1] = E\left[\frac{4\bar X}{n} - 2\right] = \frac{4}{n}E[\bar X] - 2 = \boxed{\frac{4\bar X}{n} - 2}$

Variance: $\boxed{Var(\hat\theta_1) = \frac{1}{n}\sum(X_i - \bar X) - \frac{4\bar X}{n} - 2}$

Standard Error: $\boxed{\sigma_{\hat\theta_1} = \frac{\sqrt{\frac{1}{n}\sum(X_i-\bar X) - \frac{4\bar X}{n} - 2}}{\sqrt{n}}}$

2) $\bar X = n\theta_2/4$

$\boxed{\hat\theta_2 = \frac{4\bar X}{n}}$

$E[\hat\theta_2] = E\left[\frac{4\bar X}{n}\right] = \frac{4}{n}E[\bar X] = \boxed{\frac{4}{n}\bar X}$

$\boxed{Var(\hat\theta_2) = \frac{1}{n}\sum(X_i - \bar X) - \frac{4}{n}\bar X}$

$\boxed{\sigma_{\hat\theta_2} = \frac{\sqrt{\frac{1}{n}\sum(X_i - \bar X) - \frac{4}{n}\bar X}}{\sqrt{n}}}$

σ_x σ_x

57. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ a) Which of the following is unbiased? $S^2 = \frac{1}{n-1}\sum(X_i - \bar X)^2$; $\hat\sigma^2 = \frac{1}{n}\sum(X_i - \bar X)^2$

$MSE = Var + Bias^2[Y^2] \cdot Var(X) + E[X]^2$

$\boxed{Bias_{S^2} = E[S^2] = E\left[\frac{n}{n-1}\sigma^2\right] = \frac{n}{n-1}E[\hat\sigma^2] = \frac{n}{n-1}\sigma^2}$

b) $MSE_{S^2} = Var(S^2) + Bias(S^2)^2 = \boxed{\frac{2\sigma^4}{(n-1)}}$ - Var

b) $MSE_{\hat\sigma^2} = Var(\hat\sigma^2) + Bias(\hat\sigma^2)^2 = Var\left(\frac{n-1}{n}S^2\right) + \left[E(\hat\sigma^2) - \sigma^2\right]^2$

$Bias_{\hat\sigma^2} = E[\hat\sigma^2] = E\left[\frac{(n-1)}{n}S^2\right] = \frac{(n-1)}{n}E[S^2] = \frac{n-1}{n}\sigma^2$

$= Var\left(\frac{n-1}{n}S^2\frac{\sigma^2}{\sigma^2}\right) + \frac{(n-1)}{n}\sigma^2 - \sigma^2 = \frac{\sigma^4}{n^2}Var\left(\frac{n-1}{\sigma^2}S^2\right) - \frac{1}{n}\sigma^2 = \boxed{\frac{(2n-1)\sigma^4}{n^2}}$ $\boxed{\text{Smallest}}$

$MSE'[Y] = 2(n-1)\sigma^4(\rho n + \rho - 1) = 0$; $\rho = \frac{1}{n+1}$

c) $Y = \rho\sum(X_i - \bar X)^2$; $E[Y] = E[\rho(n-1)S^2] = \rho(n-1)E[S^2] = \rho(n-1)\sigma^2$

$Var(Y) = V(\rho(n-1)S^2) = \rho^2(n-1)^2V(S^2) = 2\rho^2(n-1)\sigma^4$; $MSE[Y] = V[Y] + Bias^2(Y) = \sigma^4[2\rho^2(n-1) + (\rho n - \rho - 1)^2]$ $MSE'[Y] = 2(n-1)\sigma^4(n+1) = 0$

$\boxed{MSE \text{ is minimized for } \frac{1}{n+1}\sum(X_i - \bar X)^2}$

58. $P(AA) = (1-\theta)^2$; $P(Aa) = 2\theta(1-\theta)$; $P(aa) = \theta^2$ a) Maximum Likelihood Estimate:

Haptoglobin Type		
Hp1-1	Hp1-2	Hp2-2
10	68	112

Total: 200

a) $P(\theta) = \frac{n!}{\prod X_i!}\prod_i p(X_i)^{X_i}$; $\ln P(\theta) = \ln n! - \sum \ln X_i + \sum X_i\, p(X_i)$

$\ln P(\theta) = \ln n! - \sum_{i=1}^{3}\log X_i! + X_1\log((1-\theta)^2) + X_2\log(2\theta(1-\theta)) + X_3\log\theta^2$

$\ln' P(\theta) = \frac{-2X_1}{(1-\theta)} + \frac{2X_2}{2\theta} + \frac{X_2}{1-\theta} + \frac{2X_3}{\theta} = \frac{(X_2+2X_3)}{\theta} + \frac{(X_2+2X_1)}{(1-\theta)} = 0$

b) $+(X_2+2X_3)(1-\theta) = (X_2+2X_1)\theta$; $\frac{+X_2+2X_3}{X_2+2X_1+X_2+2X_3} = \boxed{\frac{+X_2+2X_3}{2X_1+2X_3+2X_2}} = \frac{+68+2(112)}{2(10)+2(112)+2(68)} = 0.768$

Asymptotic Variance:

$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}[\ln n! - \sum\log X_i! + X_1\log((1-\theta)^2) + X_2\log(2\theta(1-\theta)) + X_3\log\theta^2]\right]$

$= -E\left[\frac{\partial}{\partial\theta}\left[\frac{-2X_1}{(1-\theta)} + \frac{2X_2}{2\theta} - \frac{2X_2}{1-\theta} + \frac{2X_3}{\theta}\right]\right] = -E\left[\frac{-2X_1}{(1-\theta)^2} - \frac{2X_2}{\theta^2} - \frac{2X_2}{(1-\theta)^2} - \frac{2X_3}{\theta^2}\right] = \left[\frac{2n(1-\theta)^2}{(1-\theta)^2} - \frac{n2\theta(1-\theta)}{\theta^2} - \frac{2 \cdot n2\theta(1-\theta)}{(1-\theta)} - \frac{2\theta^2 n}{\theta^2}\right]$

$= \frac{2n}{\theta(1-\theta)}$ $= -2n + 2n$

99% Confidence Interval: $\hat\theta \pm 2.575\cdot\sqrt{Var(\hat\theta)}$

$\boxed{0.768 \pm 3.85\times10^{-3}}$

d) $\hat\theta^2 = \frac{112}{200}$; $\boxed{\hat\theta = 0.748}$ $\boxed{SD = 1.53\times10^{-3}}$

e) 99% Confidence: $\boxed{0.748 \pm 3.94\times10^{-3}}$

vs mle : $SD = 1.43\times10^{-3}$

$Var(\hat\theta) = \frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{2n^2} = \boxed{2.23\times10^{-6}}$

59. $P(M|M) = 50\%$; $P(X|F) = 50\%$; $P(a|M) = P(b|M) = \alpha$: $P(a|F) = P(b|F) = \alpha$

	M	F
$\hat{M}$	b $1/2$	$a + 2b + \alpha^2$
$\hat{F}$	α	$1/2$

$\boxed{\text{Maternal}}$ α

a) $P(MM) = P(I) \cdot P(I|M) + P(F)P(F|FF) = \frac{1}{2} \cdot \alpha + \frac{1}{4}(1-\alpha)$

$= \frac{1}{2}\alpha + \frac{1}{4} - \frac{\alpha}{4} = \frac{1}{4}\alpha + \frac{1}{4} = \boxed{\frac{1+\alpha}{4}}$

$P(MF) = 1 - P(MM) - P(FF) = \boxed{\frac{(1-\alpha)+\alpha}{2}}$

Identical

M	$1/4\,\alpha$	$1/4$
F	$1/4$	$1/4$

$\boxed{\text{Fraternal}}$ $(1-\alpha)$

b) $n_1 = MM$; $n_2 = FF$; $n_3 = MF$

Maximum Likelihood Estimation: $\boxed{\text{Multinomial}}$

$\ln P(x|M/F) = \ln n! - \ln \prod x_i + (n_1+n_2)\log(1+\alpha) + n_3 \log(1-\alpha)$

$\ln' P(x|M/F) = \frac{n_1+n_2}{1+\alpha} - \frac{n_3}{1-\alpha} = 0$; $\boxed{\alpha = \frac{n_1+n_2-n_3}{n_1+n_2+n_3}}$

Variance of a Multinomial: $Var(\theta) = n\,\alpha(1-\alpha) = (n_1+n_2-n_3)(1 - \frac{n_1-n_2+n_3}{n_1+n_2+n_3}) = \boxed{\frac{(n_1+n_2-n_3)(2n_3)}{n_1+n_2+n_3}}$

60. Exponential Distribution: $f(x|\tau) = \frac{1}{\tau} e^{-x/\tau}$

a) Maximum Likelihood Estimate: $\ln f(x|\tau) = -\ln\tau - \frac{x}{\tau}$; $\frac{d\ln f(x|\tau)}{d\tau} = -\frac{1}{\tau} + \frac{x}{\tau^2} = 0$; $1 = \frac{x}{\tau}$; $\boxed{\tau = x}$

b) Sampling Distribution of the mle: $f(x|\hat{\tau}) = \frac{1}{\hat{\tau}} e^{-x/\hat{\tau}} = \frac{1}{x} e^{-1} = \boxed{\frac{1}{ex}}$

$u-1 = -1$; $\frac{x}{\tau} = y$; $x = \tau y$; $dx = \tau\, dy$

c) Central Limit Theorem: $\lim_{n\to\infty} P\left(\frac{S}{\sigma\sqrt{n}} \le x\right) = \Phi(x)$

where $S = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \frac{1}{e}$; $\boxed{\lim_{n\to\infty} P\left(\frac{1}{e\sigma\sqrt{n}} < \frac{1}{e}\right) = \Phi(x)}$

d) Bias: $E[\hat{\tau}] = \int_0^\infty \frac{x}{\tau} e^{-x/\tau} dx = \frac{1}{\tau}\int_0^\infty x^{2-1} e^{-x/\tau} dx = \frac{1}{\tau}\int_0^\infty (\tau y)^{u-1} e^{-y} \tau\, dy = \tau\Gamma(2)$

$\to$ Bias $= E[\hat{\tau}] - \tau = \tau\Gamma(2) - \tau = \boxed{\emptyset}$

e) Asymptotic Variance: $I(\theta) = -\frac{1}{n}E\left[\frac{\partial^2}{\partial\tau^2}[\ln f(x|\tau)]\right] = -E\left[\frac{\partial}{\partial\tau}\left[-\frac{1}{\tau} + \frac{x}{\tau^2}\right]\right] = -E\left[\frac{1}{\tau^2} - \frac{2x}{\tau^3}\right] = \frac{2E[x]}{\tau^3} - \frac{1}{\tau^2} = \frac{2\tau}{\tau^3} - \frac{1}{\tau^2} = \frac{1}{\tau^2}$; $\boxed{Var(\theta) = \frac{\tau^2}{n}}$

Method of Moments: $E[x] = \tau$; $E[x^2] = 2\tau^2$; $\boxed{\text{The method of moments estimate shows a similar unbiased estimate of variance.}}$

f) Confidence Interval for τ:

$\hat{\tau} \pm 1.96\sqrt{Var(\hat{\tau})} = \boxed{\hat{\tau} \pm \frac{1.96\tau}{\sqrt{n}}}$

g) The exact confidence interval for $\hat{\tau}$ would be $\boxed{\bar{x} \pm 1.96\tau}$

61. $\lim_{n\to\infty} \frac{(n+1)}{(n+2)} = 1$; Laplace's rule of succession suggests the probability approaches 100% success.

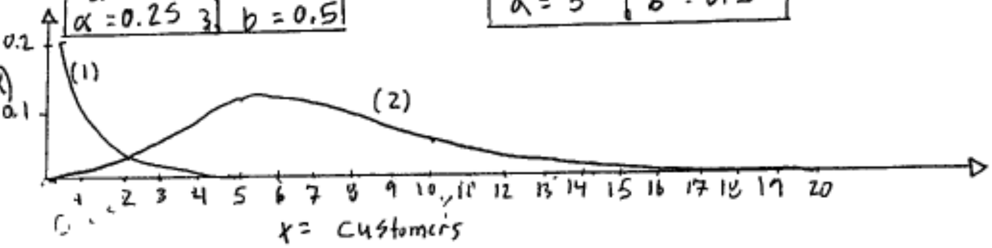
62. Gamma Distribution: Exponential Distribution: Average time to serve = 5.1 minutes ; $\lambda = 20/5.1 = 3.92 \frac{\text{customers}}{\text{min}}$

$n = 20$ customers

$\underbrace{\Gamma(x) = \int_0^\infty \frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)} dx}_{\text{Conjugate prior}}$ $\quad \underbrace{f(x|\lambda) = \frac{1}{\lambda} e^{-x/\lambda}}_{\text{Posterior}}$ $\quad f(x|\tau) \cdot \Gamma(x)$

1) $\boxed{\begin{array}{c|c} \bar{X}_{Gam} = 0.5 & \sigma_{Gam} = 1 \\ \alpha = 0.25 & b = 0.5 \end{array}}$

2) $\boxed{\begin{array}{c|c} \bar{X}_{Gam} = 10 & \sigma_{Gam} = 20 \\ a = 5 & b = 0.5 \end{array}}$



Posterior Means: $\frac{x}{\lambda_{post}} = bx + \frac{x}{\lambda_{prior}}$

$\lambda_{post} = \frac{\lambda_{prior}}{b + \lambda_{prior}}$

1) $\lambda_{post} = 0.887$

2) $\lambda_{post} = 0.887$

The posterior means represent exact average customers per minute, although with different priors, waiting times for 1-2 or 4-8 customers in the restaurant are shifted.

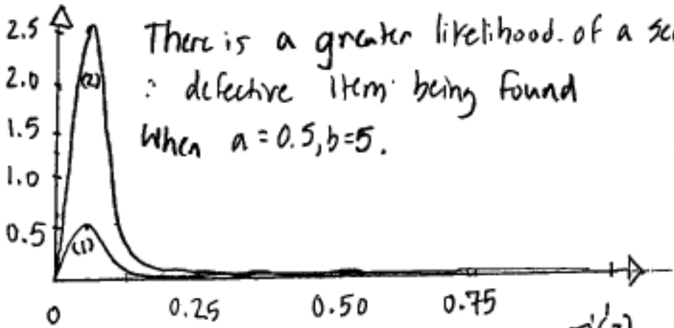
63. $n=100$; $N=3$ defective items ; Beta Distribution: $f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}x^{a-1}(1-x)^{b-1}$ $E[x]=\frac{a}{a+b}$

1) $a=b=1$ 2) $a=0.5,\ b=5$

$Var[X]=\frac{ab}{(a+b)^2(a+b+1)}$

Prior

Likelihood



There is a greater likelihood of a second ∴ defective item being found When $a=0.5, b=5$.

$f(x)=\sum_{i=1}^{n}\binom{n}{k}p^{k}(1-p)^{n-k}$

Posterori $f(x|\alpha,\beta)=\binom{n}{k}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{(k+\alpha-1)}(1-x)^{(n-k+\beta-1)}$.

64. $X=0$ or 1 ; 1) $a=b=1$ $f(x|\alpha,b)=\binom{100}{3}\frac{\Gamma(2)}{\Gamma(1)\Gamma(1)}\int_0^1 x^{(3)-1)}(1-x)^{(99)}\,dx$; $f(x|a,b)=\binom{99}{1}\frac{\Gamma(2)}{\Gamma(1)\Gamma(1)}\int_0^1 x^{4}(1-x)^{98}dx$

2) $a=0.5, b=5$ $f(x|a,b)=\binom{100}{3}\frac{\Gamma(11/2)}{\Gamma(\frac12)\Gamma(5)}\int_0^1 x^{5/2}(1-x)^{108}\,dx$; $f(x|a,b)=\binom{99}{1}\frac{\Gamma(11/2)}{\Gamma(\frac12)\Gamma(5)}\int_0^1 x^{7/2}(1-x)^{102}dx$

#1 draw #2 Draw

65. $n=20$
$\mu=? \to \bar{x}=10$
$\sigma^2=1$

Normal Distribution: $f(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Prior: $\mu=?$ $\sigma^2=?$

Posterior: $\sigma^2=0.1$ $\mu=15$

Prior: $\mu \sim N(\mu_0,\sigma_0^2)\propto exp\left(\frac{-\sum_i(x-\mu)^2}{2\sigma^2}\right)$

$=\frac{1}{\sqrt{2\pi\,0.1^2}}exp\left(\frac{-(\theta-\bar{x})^2}{2\sigma^2/n}\right)$

$\sum(x_i-\mu)^2=\sum[x_i-\bar{x}-(\mu-\bar{x})]^2=\sum(x-\bar{x})^2+n(\theta-\bar{x})^2$

Posterior = Likelihood × Prior:

$\boxed{\propto e^{\frac{-\sum(x-\mu)^2}{2\sigma^2}}\cdot e^{\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}}}$

$\propto e^{\frac{-(\mu-\bar{x})^2}{2\sigma^2/n}-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}$

$\propto e^{\frac{-\mu^2+2\mu\bar{x}-\bar{x}^2}{2\sigma^2/n}-\frac{\mu^2+2\mu_0\mu+\mu_0^2}{2\sigma_0^2}}$

$\propto e^{-\left(\frac{1}{2\sigma^2/n}-\frac{1}{2\sigma_0^2}\right)\mu^2+\frac{2}{2}\left(\frac{\mu_0}{\sigma_0^2}+\frac{\bar{x}}{\sigma^2/n}\right)\mu-\frac{\bar{x}^2}{2\sigma^2/n}-\frac{\mu_0^2}{2\sigma_0^2}}$

$\propto e^{-\frac{1}{2}\left(\frac{1}{\sigma^2/n}-\frac{1}{\sigma_0^2}\right)\left[\mu-\frac{\frac{\mu_0}{\sigma_0^2}+\frac{\bar{x}}{\sigma^2/n}}{\left(\frac{1}{\sigma^2/n}-\frac{1}{\sigma_0^2}\right)}\right]^2+\left[\frac{\frac{\mu_0}{\sigma_0^2}+\frac{\bar{x}}{\sigma^2/n}}{(1/\sigma^2/n-1/\sigma_0^2)}\right]^2-\frac{\bar{x}^2}{2\sigma^2/n}-\frac{\mu_0^2}{2\sigma_0^2}}$

$\boxed{\propto e^{-\frac{1}{2}\frac{(\mu-\bar{\mu})^2}{\sigma_0^2}}}$

Posterior $p(x)\propto exp\left\{\frac{-\sum(x-\mu)^2}{2\sigma^2}-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}$

$\propto exp\left\{-\frac{(\mu-\bar{x})^2}{2\sigma^2/n}-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}$

$\propto exp\left\{\frac{-(\mu-\mu_n)^2}{2\sigma_n^2}\right\}$

$\mu_n=\frac{\frac{1}{\sigma_0^2}\mu_0+\frac{n}{\sigma^2}\bar{X}}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}$

$\frac{1}{\sigma_n^2}=\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}$

The $\mu_n=$ $\frac{1}{\sigma_n^2}=\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}=\frac{1}{0.1^2}=\frac{1}{\sigma_0^2}+\frac{1}{1/20}$

$\frac{1}{\sigma_0^2}=100+20=120$; $\boxed{\sigma_0=\sqrt{1/120}}$

$\mu_n=\frac{\frac{1}{1/120}\cdot\mu_0+\frac{20}{0.1^2}10}{\frac{1}{1/120}+\frac{20}{0.1^2}}=15=\frac{120\mu_0+20{,}000}{2120}$; $\mu_0=\frac{15\cdot 2120-20000}{120}=98.33$

66 θ is uniform $[0,1]$

$P(x)=\frac{1}{b-a}$

a) Posterior Density

$P(\text{success})=\theta$

$\boxed{P(x)=\frac{1}{b-a}\left(\frac{1}{b-a}\right)=\theta}$

b) Probability of a third shot would be θ.

67. Negative Binomial Distribution:

$$P(X=k)=\binom{k-1}{r-1}p^r(1-p)^{k-r}$$

Poisson Distribution:

$$P(x)=\frac{\lambda^x}{x!}e^{-\lambda}$$

Frequency 1st Data Set: 500 contiguous $20cm^2$ quad

Frequency 2nd Data Set: 48 rows wide and 96ft long. 2304 sampling units of 2ft length.

1st Data Set: Glaux maritima			
Count	Frequency	Poisson	Negative Binomial
0	1	1.77	2
1	15	9.3	11
2	27	26.4	29
3	42	50.8	51
4	77	72.6	70
5	77	83.5	77
6	89	79.2	73
7	57	65	60
8	48	47	44
9	24	30	29
10	14	17	18
11	16	9	10
12	9	4	5
13	3	2	2
14	1	1	1
Total	500		
Mean	5.76		
S.D.	2.53		
λ	5.76		
r	50.29		
p	0.1028		

The negative binomial shows a goodness of fit with greater accuracy.

2nd Data Set: Potato Beetles			
Count	Frequency	Poisson	Negative Binomial
0	190	20	135
1	264	95	254
2	304	226	~~57~~
3	260	357	52
4	294	423	44
5	219	401	36
6	183	316	29
7	150	214	23
8	104	126	17
9	90	67	13
10	60	32	10
11	46	14	7
12	29	5	5
13	36	2	4
14	19	1	3
15	12	0	2
16	11	0	2
17	6	0	1
18	10	0	0
19	2	0	0
20	4	0	0
21	1	0	0
22	3	0	0
23	4	0	0
24	1	0	0
25	1	0	0
26	0	0	0
27	0	0	0
28	1	0	

Method of Moments:

Negative Binomial: $y=k-r$

$$E[x]=\sum_{k=1}^{\infty}k\binom{k-1}{r-1}p^r(1-p)^{k-r}=\sum_{k=r}^{\infty}\binom{y+r-1}{r-1}k\,p^r(1-p)^{k-r}$$

$$E[y]=\sum_{k=1}^{\infty}\binom{y+r-1}{y}y\,p^r(1-p)^y=\sum\frac{(y+r-1)!}{(y-1)!(r-1)!}p^r(1-p)^y$$

$$=\sum_{k=1}^{\infty}\frac{r(1-p)}{p}\frac{p}{r(1-p)}\frac{(y+r-1)!}{(y-1)!(r-1)!}p^r(1-p)^y$$

$$=\frac{r(1-p)}{p}\sum_{k=1}^{\infty}\frac{(y+r-1)!}{(y-1)!\,r!}p^{r+1}(1-p)^{y-1}$$

Let $y-1=z$; $\therefore y=z+1$

$y=1$; $z=0$

$$=\frac{r(1-p)}{p}\sum_{k=1}^{\infty}\underbrace{\frac{(z+1+r-1)!}{z!\,r!}p^{r+1}(1-p)^z}_{\text{Probability Mass Function}}$$

$$=\frac{r(1-p)}{p}\sum_{k=1}^{\infty}\binom{r+1+z-1}{z}p^{r+1}(1-p)^z \quad 1$$

$$\boxed{E[y]=\frac{r(1-p)}{p}}\ ;\ \boxed{p=1-\frac{E[y]}{Var[y]}}$$

$\therefore r=\frac{p}{(1-p)}E[y]$

Poisson: $E[x]=\int_0^\infty x\frac{\lambda^x}{x!}e^{-\lambda}dx=e^{-\lambda}\int_0^\infty\frac{\lambda^x}{(x-1)!}dx$; $x-1=t$

$$=e^{-\lambda}\int_1^\infty\frac{\lambda^{t+1}}{t!}dt=\lambda e^{-\lambda}\int_0^\infty\frac{\lambda^t}{t!}dt=\lambda e^{-\lambda}e^{\lambda}=\boxed{\lambda}$$

The goodness of fit for the second data set did not accurately represent the data. χ^2-values described by $\frac{(x-E)}{\sigma}$ were large for both Poisson and Negative Binomial data.

68 λ = mean ; $T=\sum_{i=1}^{n} X_i$ a) $\mathbb{P}(x)=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\frac{\lambda^{\sum x_i}}{\prod x_i!}e^{-n\lambda}$; $\frac{\mathbb{P}(x)}{\mathbb{P}(y)}=\frac{\lambda^{\sum x_i}}{\prod x_i!}e^{-n\lambda}\frac{\prod y_i!}{\lambda^{\sum y_i}}e^{n\lambda}=\lambda^{\sum x_i-\sum y_i}\frac{\prod y_i!}{\prod x_i!}$

Poisson Distribution:

$P(x)=\frac{\lambda^x}{x!}e^{-\lambda}$

Independent if and only if $\sum x_i-\sum y_i=0$

$\boxed{\sum x_i=\sum y_i=T}$

b) $\frac{P(x_1)}{P(y_1)}=\lambda^{x_1-y_1}\frac{y_1}{x_1}$; $\boxed{x_1=y_1 \text{ which is not independent.}}$

c) Theorem A : Section 8.8.1:

$f(x_1\ldots x_n|\theta)=g[T(x_1\ldots x_n),\theta]h(x_1\ldots x_n)$; $\boxed{g[T(x_1\ldots x_n),\theta]=\lambda^T=\lambda^{\sum x_i}}$ $\boxed{h(x_1\ldots x_n)=\frac{e^{-\lambda}}{\prod x_i}}$ ok.

Social! Physically! Mentally Solved! Provided!

69. Geometric Distribution: Theorem A : Section 8.8.1 ; $f(x_1\ldots x_n|\theta)=g[T(x_1\ldots x_n),\theta]h(x_1\ldots x_n)$

$P(X=k)=(1-p)^{k-1}p$

$\mathbb{P}(x)=\prod_{i=1}^{n}(1-p)^{n(k_i+1)}p=(1-p)^{\sum k_i-n}\cdot p^n=(1-p)^{\sum x_i-n}\cdot p^n$

$f(x_1\ldots x_n|\theta)=(1-p)^{T-n}\cdot p^n$; $\boxed{g[T(x_1\ldots x_n),\theta]=(1-p)^{T-n}\text{ ; }h(x_1\ldots x_n)=p^n}$

70. Factorization Theorem

$f(x_1\ldots x_n|\theta)=g[T(x_1\ldots x_n),\theta]h(x_1\ldots x_n)$; Exponential Distribution: $P(x)=\lambda e^{-\lambda x}$; $\mathbb{P}(x)=\prod\lambda e^{-\lambda x_i}=\lambda e^{-\lambda\sum x_i}$

$f(x_1\ldots x_n|\theta)=\lambda e^{-\lambda\sum x_i}$; $\boxed{g[T(x_1\ldots x_n),\theta]=e^{-\lambda\sum x_i}}$ $\boxed{h(x_1\ldots x_n)=\lambda}$

71. $F(x|\theta)=\frac{\theta}{(1+x)^{\theta+1}}$:

$\frac{P(x|\theta)}{P(y|\theta)}=\frac{\cancel{n\theta}}{\prod(1+x)^{\theta+1}}\cdot\frac{\prod(1+y)^{\theta+1}}{\cancel{n\theta}}$; $\boxed{\prod_{i=1}^{n}(1+x_i)=\prod_{i=1}^{n}(1+y_i)=T}$

72. Gamma Distribution: $\mathbb{P}(x)=\prod\frac{b^a}{\Gamma(a)}x_i^{a-1}e^{-bx_i}$; $\frac{\mathbb{P}(x)}{\mathbb{P}(y)}=\frac{b^a}{\Gamma(a)}\prod x_i^{a-1}e^{-bx_i}\cdot\frac{\Gamma(a)}{b^a}\frac{1}{\prod y_i^{a-1}e^{-by_i}}$

$P(x)=\frac{b^a}{\Gamma(a)}x^{a-1}e^{-bx}$;

$\prod_{i=1}^{n}x_i^{a-1}e^{-bx_i}=\prod_{i=1}^{n}y_i^{a-1}e^{-by_i}$

73. Rayleigh Density: $f(x|\theta)=\frac{x}{\theta^2}e^{-x^2/(2\theta^2)}$ $\boxed{\prod x_i \text{ or } \sum x_i}$

$\mathbb{P}(x|\theta)=\frac{\prod x_i}{\theta^2}e^{-\sum x_i^2/(2\theta^2)}$; $\frac{\mathbb{P}(x|\theta)}{\mathbb{P}(y|\theta)}=\frac{\prod x_i}{\theta^2}e^{-\sum x_i^2/(2\theta^2)+\sum y_i^2/(2\theta^2)}\left(\frac{\theta}{\prod y_i}\right)$

$\prod x_i e^{-\sum x_i^2/2\theta^2}=\prod x_i e^{-\sum y_i^2/2\theta^2}$; sufficient statistic $\boxed{\prod x_i \text{ or } \sum x_i}$

74. Binomial Distribution:

$P(k)=\sum_{i=1}^{n}\binom{n}{k}p^k(1-p)^{n-k}$

$\mathbb{P}(k)=\exp[\log\sum\binom{n}{k}+\sum k\log p+(n-\sum k)\log(1-p)]$

$\boxed{\approx\exp[\sum_i d(k)+\sum T_i(k)c_i(\theta)+S(k)]}$

75. Gamma Distribution

$P(x)=\frac{b^a}{\Gamma(a)}x^{a-1}e^{-bx}$

$\mathbb{P}(x)=\exp[n\log\frac{b}{\Gamma(a)}+(a-1)\sum\log x-b\sum x]$

$\boxed{\approx\exp[d(\theta)+\sum c_i(\theta)T_i(x)+S(x)]}$

Crying in Goo

Chapter 9: Goodness of Fit

1. $P(H|X)=0.50$; $P(T|X)=0.50$

a) $P(\text{reject}|0)+P(\text{reject}|10)=$
$\binom{10}{0}0.5^0 0.5^{10}+\binom{10}{10}0.5^{10}0.5^0=0.002$

b) Power of Test $[1-\beta]=\binom{10}{0}(0.1)^0 0.9^{10}+\binom{10}{10}0.9^{10}0.9^0$ $\boxed{0.9344}$

$\boxed{=0.3487}$

2. a. X is uniform on [0,1]

$\boxed{\text{Simple}}$

Simple Hypothesis: only null and alternative hypothesis.

Composite Hypothesis: when a probability distribution is not specified.

b. A die is unbiased

$\boxed{\text{Composite}}$

c. X follows a normal with mean 0 and var $\sigma^2>10$

$\boxed{\text{Simple}}$ $\boxed{\text{simple}}$

d. X follows a normal with mean $\mu=0$.

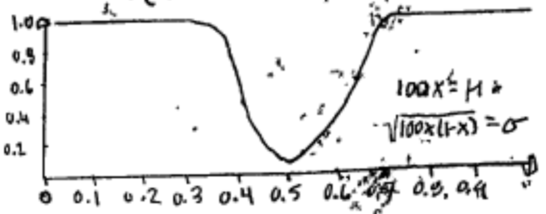
3. $X\sim bin(100,p)$; H_0 is rejected when: $p=0.5$

H_A is favored when: $p\neq 0.5$ for $|X-50|>10$

a) $\alpha=P(|X-50|>10)=P\left(\frac{|X-50|}{5}>2\right)=2P\left(\frac{X-50}{5}<-2\right)\approx 2\Phi(-2)$
$=2(1-\Phi(2))$
$\boxed{=0.0456}$

b) Power as a function of $p=1-\beta=P(|X-50|>10)$
$=1-P(40<X<60)$.

[1-β]



[X]

4.

X	H_0	H_A
x_1	0.2	0.1
x_2	0.3	0.4
x_3	0.3	0.1
x_4	0.2	0.4

a) Likelihood Ratio:

X_1	X_2	X_3	X_4
2.0	0.75	3	0.5

	X_3	X_1	X_2	X_4
Λ	3.0	2.0	0.75	0.5

b) $\alpha=0.2$: 3 degrees of freedom

$\Lambda=\frac{P(x|H_0)}{P(x|H_1)}=\alpha$;

$\boxed{P(\Lambda\leq\lambda_0|H_0)=0.2}$ or $\Lambda>\chi^2_{df3,\,0.2}=4.642$

$\boxed{P(\Lambda\leq\lambda_0|H_0)=0.5}$ $\Lambda>\chi^2_{df3,\,0.5}=2.366$

c) Prior $P(H_0)=P(H_A)$;

d) Λ_{X_3} corresponds to the decision rules for prior probabilities

$\boxed{H_0 \text{ is accepted for } \alpha=0.5, \text{ but not } \alpha=0.2}$

5. a) False, the significance level of a statistical test is equal to the probability the likelihood is less than a threshold.

b) False, the power $[1-\beta]$ is described by the null hypothesis rejection, while significance level is denoted as the threshold α of rejection β if the null hypothesis is true is $[1-\beta]$

c) False, the probability that the null hypothesis is falsely rejected is not rejected when is false.

d) False, the probability the null hypothesis is falsely rejected is not rejected when is false.

d) False, the probability the null hypothesis crosses the significance level,

e) False, a type I error occurs when the statistic crosses the significance level,

The test statistic relates the hypothesis likelihood.

F) False, a type II error tends to be less important than type I.

g) False, the power of a test is determined by the alternative hypothesis.

h) True, the likelihood is a random variable.

6. a)

X	0	1	2	3	4	5	6	7	8	9	10
$C_{10,n}$ z	0.001	0.0098	0.0439	0.1172	0.2051	0.2461	0.2051	0.1172	0.0439	0.0098	0.0010
$C_{10,n}$ z	0.000	0.001	0.0014	0.0090	0.0368	0.1029	0.2001	0.2668	0.2335	0.1211	0.0282
$\frac{P(x|H_1)}{P(x|H_0)}=\frac{Gen(x|p)}{Gen(x|p)}$		1.4	0.24	0.50	0.30	0.19	0.11	0.09	0.04	0.02	0.01

$P(x|H_0)=0.5$

$P(x|H_1)=0.7$

$\boxed{\frac{P(X|H_1)}{P(x|H_0)}<1 : \text{Favors } H_0 \quad \frac{P(X|H_1)}{P(X|H_0)}>1 : \text{Favors } H_1}$

b) If $P(H_0)|P(H_1)=10$ then each of the outcomes favor H_0.

c) Significance Level of H_0 if $X\geq 8$

$\alpha=P(\text{Reject } H_0|H_0)=P(X\geq 8|H_0)$
$=1-P(X\leq 7|0.5)=1-\sum_1^7(1-0.5)^{i-1}(0.5)$
$=1-\sum_1^7 0.5^i=\boxed{0.0078}$

d) The power of the test $[1-\beta]=P(\text{Reject } H_0|H_1)$
$=P(X\geq 8|H_1)=1-P(X\leq 7|0.7)=1-\sum_1^7(1-0.7)^{i-1}(0.7)$

$\boxed{=0.0002}$

7. Poisson Distribution

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Likelihood Ratio: $\Lambda_o = P(\lambda=\lambda_1|H_o)/P(\lambda=\lambda_o|H_1) = \left(\frac{\lambda_o}{\lambda_1}\right)^x e^{-(\lambda_o-\lambda_1)}$

$\Lambda_n = \left(\frac{\lambda_o}{\lambda_1}\right)^{\sum x} e^{-n(\lambda_o-\lambda_1)}$; $\ln\Lambda_n = \sum x \ln\left(\frac{\lambda_o}{\lambda_1}\right) - n(\lambda_o-\lambda_1)$

$-2\ln\Lambda = -2\sum x\ln\left(\frac{\lambda_o}{\lambda_1}\right) - \cancel{\phi} = 2\sum x\ln\left(\frac{\lambda_1}{\lambda_o}\right)$

$$-2\ln\Lambda = \sum_{i=0}^{n}\frac{(\lambda-\lambda_o)^2}{\lambda_o} = \chi^2$$

$\boxed{\text{Pearson's Chi-square statistic}}$

Taylor Series: $\ln(x) = \sum_{i=1}^{n}\frac{(x-x_o)^i}{i}x^i(0) = (x-x_o)\ln(0) + \frac{1}{2}(x-x_o)^2\frac{1}{x_o}$

$= (\lambda_1-\lambda_o) + \frac{1}{2}(\lambda_1-\lambda_o)^2\frac{1}{\lambda_o}$

$$\boxed{P\left(\frac{\sum(\lambda_1-\lambda_o)^2/\lambda_o - \nu}{\sqrt{2\nu}} \geq Z(\alpha/2)\right) = \alpha \text{ ; } \nu = \text{degrees of Freedom}}$$

$\boxed{\text{Normal Z-table}}$

$-2\ln\Lambda = 2\underbrace{\sum(\lambda-\lambda_o)}_{=0} + \sum\frac{(\lambda_1-\lambda_o)^2}{\lambda_o}$

8. $\lambda = \lambda_o$ $P\left(\frac{\ldots}{\sqrt{2\nu}} > Z(\alpha/2)\right) = \alpha$ $\boxed{\text{Simple hypothesis}}$ vs.

$\lambda > \lambda_o$ $P\left(\frac{\sum(\lambda-\lambda_o)^2 - \lambda_o}{\lambda_o\sqrt{2\nu}} > Z(\alpha/2)\right) = \alpha$ $\boxed{\text{composite hypothesis}}$

$$-2\ln\Lambda = \sum\frac{(\lambda_1-\lambda_o)^2}{\lambda_o}$$

$\boxed{\text{Poisson Dispersion Test}}$

"Measure of clustering"

9. Normal Distribution

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\sigma^2 = 100$ $\mu_o = 0.0$ $\alpha = 0.10$ $\mu_A = 1.5$

$\Lambda = \frac{P(x|H_o)}{P(x|H_A)} = e^{\frac{-\sum(x-\mu_o)^2 + \sum(x-\mu_1)^2}{2\sigma^2}}$

$= e^{\frac{2n\bar{x}(\mu_o-\mu_1) + n\mu_1^2 - n\mu_o^2}{2\sigma^2}}$; $\Lambda \gg 1$: $\mu_o > 0$; $\bar{x} \ll 1$

$\Lambda \ll 1$: $\mu_o < 0$; $\bar{x} \gg 1$

$P(\bar{X} \geq x_o) = P\left(\frac{\bar{X}-\mu_o}{\sigma/\sqrt{n}} > \frac{x_o-\mu_o}{\sigma/\sqrt{n}}\right) = \frac{x_o-\mu_o}{\sigma/\sqrt{n}} = Z(\alpha)$; $x_o = \frac{10}{\sqrt{n}}Z(0.1) = \frac{10 \cdot 1.28}{\sqrt{n}} = \boxed{2.56}$

$\boxed{\text{Unfavorable for } H_o}$

$\alpha = 0.01$ $\Lambda = \frac{P(x|H_o)}{P(x|H_1)} = e^{\frac{2n\bar{x}(\mu_o-\mu_1) + n\mu_1^2 - n\mu_o^2}{2\sigma^2}}$

$Z = \frac{x_o-\mu_1}{\sigma/\sqrt{n}} = \frac{2.56-1.5}{10/5} = 0.53$; $P(Z > 2.56) = 1-\beta$; $= 0.7019$

β $\boxed{=0.2981}$

10. $P(\bar{X} > x_o) = P\left(\frac{\bar{X}-\mu_o}{\sigma/\sqrt{n}} > \frac{x_o-\mu_o}{\sigma/\sqrt{n}}\right) = \frac{x_o-\mu_o}{\sigma/\sqrt{n}} = Z(\alpha)$; $x_o = \frac{10}{5}Z(0.01) = Z(0.5040) = \boxed{1.080}$

Suppose $X_1 \ldots X_n$

$f(x|\theta)$, T = sufficient statistic

$Z = \frac{x_o-\mu_1}{\sigma/\sqrt{n}} = \frac{1.09-1.5}{2} = 0.21$; $P(Z > 1.23) = 1-\beta = 0.8907$

β $\boxed{=0.1093}$

Likelihood: $\Lambda = \frac{f(x|\theta_o)}{f(x|\theta_1)}$

$\boxed{\text{The rejection region is determined from normalizing the threshold mean of the numerator } [f(x|\theta_o)] \text{ such that rejection value is greater than the significance level}}$

11. $\mu = 25$

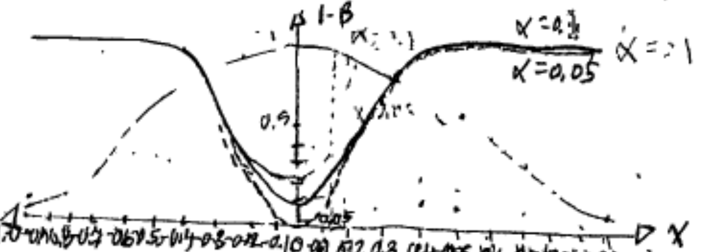
$\sigma^2 = 100$

Normal Distribution

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

$H_o: \mu = 0$

$H_A: \mu \neq 0$



12. Let $X_1, \ldots X_{n5}$

Exponential Distribution:

$$f(x) = \theta e^{-\theta x}$$

$$\Lambda = \frac{f(x|H_o)}{f(x|H_A)} = \frac{f(x|\theta_o)}{f(x|\theta \neq \theta_o)}$$

$$= \left(\frac{\theta_o}{\theta_A}\right) e^{(\theta_A-\theta)x}$$

$\Lambda = (\theta\bar{x})^n e^{-n(\bar{x}\theta-1)}$; $\theta_o = 1$, $n = 10$, $\alpha = 0.05$

$$\boxed{= \theta e^n\left[\bar{x}e^{-n\bar{x}\theta}\right]}$$

a) Show $\boxed{\{\bar{X} \leq x_o\} \cup \{\frac{1}{\bar{X}} \geq x_1\} = \{\bar{X}e^{-10\bar{X}} \leq \chi^2_{?}\}}$

b) The rejection region is chosen by a threshold value of c at a significance level.

c) The exponential distribution relates to the gamma through n summation and $\theta = \frac{1}{\theta_o}$

d) Generating outputs from an exponential may be graphically similar to a gamma.

14. $P(X|H_0) = N(0,\sigma^2)$

$P(X|H_1) = N(1,\sigma^2)$

$P(H_0) = 2 \times P(H_1)$

$$\frac{P(H_0|X)}{P(H_1|X)} = \frac{P(H_0)}{P(H_1)}\frac{P(X|H_0)}{P(X|H_1)} = 2e^{\frac{(x-1)^2 - x^2}{2\sigma^2}} = 2e^{\frac{-2x+1}{2\sigma^2}} ; \ln(\tfrac{1}{2}) = -\frac{2x+1}{2\sigma^2} ; x = \frac{2\sigma^2\ln(\frac{1}{2})-1}{-2}$$

a)

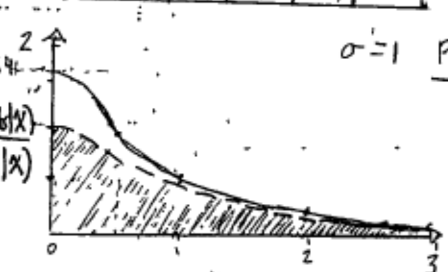
σ^2	0.1	0.5	1.0	5.0
$X: H_0 > H_1$	0.57	0.85	1.19	3.96

b) $\boxed{2/3}$

15. $P(X|H_0): N(0,\sigma^2)$

$P(X|H_1) = N(1,\sigma^2)$

$P(H_0) = P(H_1)$



$\sigma^2 = 1$ $\frac{P(H_0|X)}{P(H_1|X)} = \frac{P(H_0)}{P(H_1)} \cdot \frac{P(X|H_0)}{P(X|H_1)} = e^{\frac{-2x+1}{2\sigma^2}}$

The p-value of $P(H_0|X)/P(H_1|X)$ and $P(H_0)$ show similarly symmetric graphs, so their p-values are similarly symmetric. Another σ will show equivalent results because of scaled proportions.

16. $\alpha = 0.05$

$$\frac{P(H_0|X)}{P(H_1|X)} = e^{\frac{-2x+1}{2\sigma^2}} > 1 ; \boxed{x > \tfrac{1}{2}} ; \frac{P(H_0|X)}{P(H_1|X)} = e^{\frac{-2x+1}{2\sigma^2}} < 1 ; \boxed{x < \tfrac{1}{2}}$$

17. $P(X|H_0) = N(0,\sigma_0^2)$

$P(X|H_1) = N(0,\sigma_1^2)$

$\sigma_1 > \sigma_0$

a) $\Lambda = \frac{P(X|H_0)}{P(X|H_1)} = \frac{N(0,\sigma_0^2)}{N(0,\sigma_1^2)} = e^{-x^2/2\sigma_0^2 + x^2/2\sigma_1^2} \boxed{= e^{-(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2})\frac{x^2}{2}} \cdot \sigma_1/\sigma_0}$ H_0 Favored: $\boxed{X > \sqrt{\frac{2\ln\sigma_1/\sigma_0}{(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2})}}}$

The rejection region of a level α test: $\ln\Lambda = -(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2})\frac{x^2}{2} + \ln\sigma_1/\sigma_0$

$P(X^2 > X^2(\alpha)) = \alpha$

$P((\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2})\sum X_i^2 - 2n\ln\sigma_1/\sigma_0) > X^2(\alpha)) = \alpha$

Assuming largest term is $\frac{1}{\sigma_0^2}\sum X_i^2$ because $\sigma_1 >> \sigma_0$.

$P(\frac{1}{\sigma_0^2}\sum X_i^2 > X^2(\alpha)) = \alpha ; \boxed{\sigma_0^2 X^2 < X^2}$

$-2\ln\Lambda = (\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2})x^2 + 2\ln\sigma_1/\sigma_0$

$-2\ln\Lambda = \sum(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2})x_i^2 - 2n\ln\sigma_1/\sigma_0$

$X^2 < (\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2})\sum X_i^2 - 2n\ln\sigma_0/\sigma_1$

$\left|\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right| > \frac{2z(\alpha/2) + 4n\ln\sigma_0/\sigma_1}{\sum X_i^2}$

b. $X_1, X_n \ldots X_n$

$P((\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2})\sum X_i^2 - 2n\ln\sigma_1/\sigma_0) > X^2(\alpha)) = \alpha ; \boxed{\sigma_0^2 X_n^2(\alpha) \leq \sum X_i^2}$

c. Yes, because the rejection level is a simple hypothesis.

18. $X_1, X_2 \ldots X_n$ i.i.d.

Double Exponential Distribution

$f(x) = \frac{1}{2}\lambda\exp(-\lambda|x|)$

Maximum Likelihood:

$\frac{d\ln f(x)}{d\lambda} = \frac{n}{\lambda} - \sum|x|$

$\hat{\lambda} = \frac{1}{\bar{x}}$

Likelihood Ratio Test: $\Lambda = \frac{f(x|\lambda_0)}{f(x|\lambda_1)} = (\frac{\lambda_0}{\lambda_1})\exp^{(\frac{1}{x_i}-\frac{1}{\bar{x}})|\bar{x}-x_i|}$

$-\ln\Lambda = \sum\ln(\frac{\lambda_0}{\lambda_1}) + \frac{\sum(\bar{x}-x_i)^2}{x_i\bar{x}} \overset{=0}{}, \lambda$

$\ln\Lambda = \sum\ln(\frac{x_i}{\bar{x}_i}) = \sum(x_i-\bar{x}) + \frac{1}{2}\sum(x_i-\bar{x})^2\frac{1}{\bar{x}}$

$\boxed{2\ln\Lambda = 2\frac{\sum(x_i-\bar{x})^2}{\bar{x}} = 2X^2}$

a) $P(H_0) = P(X|H_1)$

$$\frac{P(H_0|X)}{P(H_1|X)} = \frac{P(H_0)}{P(H_1)}\frac{P(X|H_0)}{P(X|H_1)} = \frac{F_0(x)'}{F_1(x)'} = \frac{2x}{3x^2} = \frac{2}{3x} > 1 ; \boxed{x < \tfrac{2}{3}}$$

b) $\boxed{\Lambda = \frac{2}{3x}}$

The test is uniformly most powerful because of the squared terms. $\lambda > \lambda_0$ vs $\lambda_1 > \lambda_0$ are equivalent outcomes.

19. $H_0: F_0(x) = x^2$

$0 \leq x \leq 1$

$H_1: F_1(x) = x^3$

$0 \leq x \leq 1$

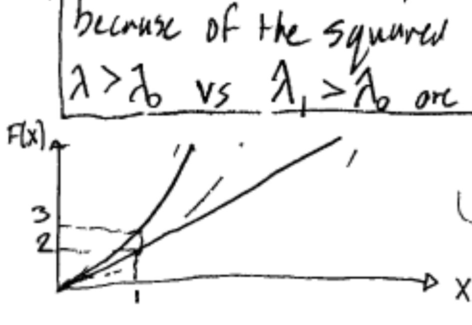
c) $P(x^2 \cdot 2/3 > z(\alpha/2)) = 1-\alpha$

$x^2 = 1-\alpha$

$\boxed{x = \sqrt{1-\alpha}}$

d) $P(x \cdot 2/3 > z(\alpha/2)) = 1-\beta$

$\boxed{\beta = 1 - \frac{(1-\alpha)^3}{2}}$



20. $[0,1]$: $\underbrace{f_0(x)=1}_{H_0}$, $\underbrace{f_1(x)=2x}_{H_1}$; $\alpha=0.10$; $P\left(\frac{f_0(x)}{f_1(x)}\le c\right)=P\left(\frac{1}{2x}\le c\right)=P\left(x\ge\frac{1}{2c}\right)$

$\boxed{\text{Null Hypothesis}}=\int_{1/2c}^{1} f_0(x)\,dx=\int_{1/2c}^{1} 1\,dx=1-\frac{1}{2c}=0.1$; $c=5/9$

$P\left(\frac{f_0(x)}{f_1(x)}<c\right)=P\left(\frac{1}{2x}\le c\right)=P\left(x\ge\frac{1}{2c}\right)=P(x\ge 0.9)=\int_{0.9}^{1} f_1(x)\,dx$

$=1-0.9^2=\boxed{0.19}$

21. $[0,\theta]$; $H_0:\theta=1$ vs $H_1:\theta=2$ Uniform:

a. $\alpha=0$. What is its power? $f(\theta)=\frac{1}{\theta}$

$P\left(\frac{f_0(\theta|H_0)}{f_1(\theta|H_1)}<c\right)=P(2<c)=\int_c^\theta d\theta=\theta-c=0$ $\quad c=\theta$

$\bar{P}\left(\frac{f_0(\theta|H_0)}{f_1(\theta|H_1)}<c\right)=P(2<c)=P(2>c)=P(2>1)=\int_1^\theta \frac{1}{2}d\theta=\frac{1}{2}(\theta-1)=\boxed{\frac{1}{2}}$

b. $0<\alpha<1$; $X\in[0,\alpha]$; $P\left(\frac{f_0}{f_1}<\alpha\right)=P(X<\alpha)-P(0<X)=\int_X^\alpha d\theta-\int_0^X d\theta=\alpha-X-X=\boxed{\alpha}$

$P(1<X<1-\alpha)=P(X<1-\alpha)-P(1<X)=\int_X^{1-\alpha}\frac{1}{2}d\theta-\int_1^X\frac{1}{2}d\theta=\frac{1}{2}(1-\alpha-X)-\frac{1}{2}(X-1)=\boxed{1-\frac{\alpha}{2}}$

c. $X\in[1-\alpha,1]$; $P(1-\alpha<X<1)=P(X<1)-P(1-\alpha<X)$

$=\int_1^{X} d\theta-\int_{1+\alpha}^{X} d\theta=X-1-X+1+\alpha=\boxed{\alpha}$; $P(\alpha<X<1)=P(X<1)-P(X<\alpha)=\int_\alpha^X\frac{1}{2}d\theta-\int_\alpha^X\frac{1}{2}d\theta$

d $\boxed{X\in[(1-\alpha)/2\le X\le(1+\alpha)/2]}$ $\quad \frac{X}{2}\boxed{1-\frac{\alpha}{2}}\alpha$

e. $\boxed{\text{The likelihood ratio test determines unique rejection for } \alpha>0}$

f. $H_0=\theta=2$; $H_1:\theta=1$; $P\left(\frac{f_0(\theta|H_0)}{f_1(\theta|H_1)}<c\right)=P\left(\frac{1}{2}<c\right)=\int_c^\theta\frac{1}{2}d\theta=\frac{\theta-c}{2}=0$; $c=\theta$

$P\left(\frac{f_0(\theta|H_0)}{f_1(\theta|H_1)}<c\right)=P\left(\frac{1}{2}<c\right)=P\left(\frac{1}{2}<1\right)=\int_1^\theta d\theta=\theta-1=0$

$\boxed{\text{The rejection region is not capable of being determined.}}$

22. Example A: Section 8.5.3

$\left(\frac{n\hat{\sigma}^2}{\chi^2_{n-1}(\alpha/2)},\frac{n\hat{\sigma}^2}{\chi^2_{n-1}(1-\alpha/2)}\right)$ $\quad P(X|H_0)=P(X|\sigma_0^2)$ $\quad A(\sigma_0^2)=\{X|\sigma_0^2\in C(X)\}$

Theorem B: Section 9.3 $\quad$ Significance level: α

$=\left\{X\,\middle|\,\frac{n}{n-1}\frac{\sum_1^n(X_i-\bar{X})^2}{\chi^2_{n-1}(\alpha/2)}>\sigma_0^2>\frac{n}{n-1}\frac{\sum_1^n(X_i-\bar{X})^2}{\chi^2_{n-1}(1-\alpha/2)}\right\}$

$P[\theta_0\in C(X)|\theta=\theta_0]=1-\alpha$ $\quad X_1,X_2,\ldots,X_n$

$A(\theta_0)=\{X|\theta_0\in C(X)\}$ $\quad \sigma_0=1, n=15, \alpha=0.05$

$=\left\{X\,\middle|\,\frac{15}{14}\frac{\sum_1^{15}(X_i-\bar{X})^2}{\chi^2_{14}(0.05/2)}>\sigma_0^2>\frac{15}{14}\frac{\sum_1^{15}(X_i-\bar{X})^2}{\chi^2_{14}(1-0.05/2)}\right\}$

$P[\sigma_0^2\in C(x)|\sigma^2=\sigma_0^2]=1-\alpha$

$\boxed{P\left(\frac{19}{100}\sum_1^{15}(X_i-\bar{X})^2>1>\frac{4}{100}\sum_1^{15}(X_i-\bar{X})^2\right)=0.95}$ $\quad \boxed{=\left\{X\,\middle|\,\frac{19}{100}\sum_1^{15}(X_i-\bar{X})^2>\sigma_0^2>\frac{4}{100}\sum_1^{15}(X_i-\bar{X})^2\right\}}$

23. 99% Confidence Interval: μ

Normal Distribution:

$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}$; $\mu\in(-2.0,3.0)=P(-\sigma^2 z(\alpha/2)<\mu<\sigma^2 z(\alpha/2))=1-\alpha=0.99$

$\alpha=0.01$; $z(0.005)=2.57$ $\sigma^2=0.195$;

$\Lambda=e^{\frac{+(X-\mu_1)^2-(X-\mu_0)^2}{2\sigma^2}}=e^{\frac{-2\mu_1X+\mu_1^2+2\mu_0X+\mu_0^2}{2\sigma^2}}=e^{\frac{-2\mu_1X+\mu_1^2-6X+9}{2(0.195)}}=e^{\frac{-(2\mu_1+6)X+\mu_1^2+9}{0.39}}$

$P\left(\frac{-(2\mu_1+6)X+\mu_1^2+9}{0.39}>0\right)=P\left(X>\frac{\mu_1^2+9}{(2\mu_1+6)}\right)$; $\boxed{\text{Yes, the rejection region demonstrates } H_A:\mu_1\neq-3 \text{ does not suffice.}}$

24. n = trials, p = prob. success

Binomial Random Variable $\quad$ a) $H_0:p=0.5$ $\quad$ b. Rejection Region: $\Lambda=\frac{(1/2)^n}{\left(\frac{X+n/2}{n}\right)^{X+n/2}\left(1-\frac{X+n/2}{n}\right)^{n/2-X}}$

$P(1)=p$ $\quad H_A:p\neq0.5$

$P(0)=1-p$ $\quad \Lambda=\frac{P(X|H_0)}{P(X|H_A)}=\frac{P_0^n(1-P_0)^{n-X}}{P_A^n(1-P_A)^{n-X}}=\boxed{\frac{e(0.5)^n}{\left(\frac{X}{n}\right)^X\left(1-\frac{X}{n}\right)^{n-X}}}$ $\quad =\frac{(1/2)^n}{\left(\frac{1}{2}+\frac{X}{n}\right)^{X+n/2}\left(\frac{1}{2}-\frac{X}{n}\right)^{n/2-X}}$

$P(X)=0$; $X\neq0, X\neq1$ $\quad$ K; $c<<\alpha<0$ $\rightarrow$

25. Example B: Section 9.5

Number per Square	0	1	2	3	4	5	6	7	8	9	10	19
Frequency	56	104	80	62	42	27	9	9	5	3	2	1

Likelihood Ratio for a Poisson

$$\Lambda = \frac{\prod \hat{\lambda}^{x_i} e^{-\bar{X}}/x_i!}{\prod \tilde{\lambda}^{x_i} e^{-\tilde{\lambda}_i}/x_i!} = \prod_{i=1}^{n} \left(\frac{\bar{X}}{x_i}\right)^{x_i} e^{x_i - \bar{X}} \; ; \; \bar{X} = \frac{0\times56 + 1\times104 + 2\times80 + \cdots + 19\times1}{400} = 2.44$$

Number Per Square	0	1	2	3	4	5	6	7, 8, 9, 10	19
log likelihood	[illegible]	[illegible]	30	[illegible]	[illegible]	[illegible]	[illegible]	27.5726	30

$= 2.8239$

$$-2\log\Lambda = -2\sum x_i \log\left(\frac{\bar{X}}{x_i}\right) + (x_i - \bar{x}) = -2\sum x_i \log\left(\frac{x_i}{\bar{x}}\right) \approx 2\sum (x_i - \bar{X}) - \frac{(x_i-\bar{X})^2}{2\bar{X}} + \cdots$$

$$\approx \sum_{i=1}^{n} \frac{(x_i - \bar{X})^2}{\bar{X}} = 58$$

$\boxed{\Lambda = 2.54\times10^{-13}}$

The similarities of likelihood $[\Lambda]$ and chi-square $[\chi^2]$ are within 23% of each other indicating error of measurement from maximum likelihood estimate applied to observed vs expected values.

26. a) False, the generalized likelihood ratio statistic favors or rejects at a boundry of less than or greater than 1

b) True, the corresponding test would reject at $\alpha = 0.02$

c) False, the p-value is 0.06 and not less than.

d) False, p-value is the smallest value at which the test would be rejected.

e) False, p-value is a threshold for rejection and simple hypothesis depend on a single value, w.

f) False, the p-value would be greater than 0.05

27. df = 7 $\boxed{{}_4\chi^2_{0.90} = 12.02}$; $\boxed{{}_4\chi^2_{0.95} = 14.07}$; $\boxed{{}_4\chi^2_{0.975} = 16.01}$

$\boxed{{}_4\chi^2_{0.99} = 18.48}$; $\boxed{{}_4\chi^2_{0.995} = 20.28}$

28 a) $P(T > t_0 | H_0) = \alpha$; $P(1.5 > t_0 | H_0)$; $1.5 = z(\alpha)$

$\boxed{0.9332 = \alpha}$

b) $\boxed{\alpha > 0.9901}$

29. Yes, the monotone increasing function $S > g(t_0)$ is a test.

30 a) Show $V = 1 - F(T)$; $F(t) = \int_{-\infty}^{\alpha} f(t)\,dt = P(T > \alpha | H_0) = 1 - V = 1 - \alpha$

b) $P(V \le z) = P(T \le F^{-1}(z)) = F(F^{-1}(z)) = z$

c) 0.1

d) $P(V < \alpha) = P(F(T) < \alpha) = P(T < F^{-1}(\alpha)) = P(F^{-1}(1-\alpha) < T)$

$= F(F^{-1}(1-\alpha)) = z = \alpha$

31. ${}_1\chi^2_{0.1} = 0.0158$

$$-2\log\Lambda = \frac{n}{\sigma^2}(\bar{X} - \mu_0)^2 > {}_1\chi^2_{0.1} = 0.0158$$

${}_n\chi^2_p$	${}_1\chi^2_{0.1}$	${}_5\chi^2_{0.1}$	${}_{10}\chi^2_{0.1}$	${}_{20}\chi^2_{0.1}$
Λ	2.6×10^{-1}	9.4×10^{-3}	3.3×10^{-4}	6.8×10^{-7}

24. continued...

c) $n = 100$, $k = 2$

$P(X - 5 > 2) = P(X = 0, 1, 2, 8, 9, 10)$
$= \frac{7}{64} = 0.1094$

d) $n = 100$, $k = 2$

$P(X - 50 > 10) = P(X - 50 \ge 11)$

$E(X) = 100 \times 0.5 = 50$

$V(X) = 100 \times 0.5 \times 0.5 = 25$

32. Object A: Object B: $X = 120$

$\mu_A = 100$; $\mu_B = 125$

$\sigma = 25$; $\sigma = 25$

a) $\Lambda = \dfrac{e^{-(x-\mu_A)^2/2\sigma^2}}{e^{-(x-\mu_B)^2/2\sigma^2}} = e^{\frac{-(120-100)^2+(120-125)^2}{2\cdot 25^2}} = \boxed{0.74}$

b) $P(A) = P(B) = \frac{1}{2}$; $\dfrac{P(H_A|X)}{P(H_B|X)} = \dfrac{P(H_A)}{P(H_B)}\dfrac{P(X|H_A)}{P(X|H_B)} = \boxed{0.74}$

c) $P(X>125|H_B) = P(X>125) = \alpha$; $.20 = z(\alpha/2)\sigma$; $z(\alpha/2) = 0$; $\boxed{\alpha = 0.5}$

d) Power of Test; $\beta = \int_{125}^{\infty} \dfrac{e^{-(x-\mu_B)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}\,dx = \boxed{0.5}$

e) $X = 120$; $P\frac{X-125}{\sigma} = z(\alpha)$; $\frac{-5}{25} = z(\alpha)$; $\frac{-1}{5} = z(\alpha)$; $\frac{1}{5} = 1 - z(\alpha)$; $z(\alpha) = 0.8$; $\boxed{\alpha = 0.7881}$

33.

	Jewish	Chinese & Japanese
Deaths Before Holiday	922	418
Deaths After Holiday	997	434
Pearsons $[X^2]$	2.93	3.00
Likelihood $[\Lambda]$	2.93	3.004
p-value $_{(df=1)}$	0.05-0.10	0.05-0.10

No evidence for capability of postponing death.

34. Problem #55: Chapter 8 0.0357

Type	O_i	E_i	Likelihood $[\Lambda]$	Chi-Squared $[X^2]$
Starchy Green	1997	1953		
Starchy White	906	925	5.98	1.97
Sugary Green	904	925		
Sugary White	32	34		

p value ~0.1 pvalue ~0.9

$_3X^2_{0.9} = 6.25$ $_3X^2_{0.1} = 0.584$

There is consideration to reject null hypothesis.

35.

	Hp I-1	Hp 1-2	Hp 2-2
O_i	10	68	112
E_i	$(1-\theta)^2$	$2\theta(1-\theta)$	θ^2
E_i	10	68	112

Haptoglobin Type; $n = 190$; $\hat{\theta} = 0.768$

$\Lambda = 0$ $X^2 = 0$

$df = 2$; p-value < 0.01

$_2X^2_{0.995} = 7.88$

Cannot reject null hypothesis.

36.

Month	# Suicides	Days/Month	Probability
Jan	1867	31	0.085
Feb	1789	28	0.077
Mar	1944	31	0.085
Apr	2094	30	0.082
May	2097	31	0.085
June	1981	30	0.082
July	1887	31	0.085
Aug	2024	31	0.085
Sep	1928	30	0.082
Oct	2032	31	0.085
Nov	1978	30	0.082
Dec	1859	31	0.085

$X^2_{(12)} = \dfrac{(1867-1957)^2}{1957} + \dots + \dfrac{(1859-1957)^2}{1957}$; $P(\theta|H_0) = \frac{1}{12}$; $p(\theta|H_0) = \sum p(\theta_i) = 1$

$n = 23480$; $E = np = 1957$

$= 51.79$

$_{11}X^2_{0.995} = 26.76$; p-value < 0.005

Since $X^2 > {_{11}X^2_{0.995}}$ a rejection of the null hypothesis occurs and suicide is not constant. at p-value $= 0.005$.

Multinomial: $f(\theta) = \dfrac{n!}{x_1!\cdots x_m!} p_1(\theta)^{x_1}\cdots p_m(\theta)^{x_m}$

mle $= x = np_i$; $p_i = \frac{x}{n}$

Likelihood Ratio:

$$\Lambda = \frac{\frac{n!}{x_1!\cdots x_m!} p_1(\theta)^{x_1}\cdots p_m(\theta)^{x_m}}{\frac{n!}{x_1!\cdots x_m!} p_1^{x_1}\cdots p_m^{x_m}} = \prod_{i=1}^{n}\left(\frac{p_i(\theta)}{p_i}\right)^{x_i}$$

$$-2\log\Lambda = -2n\sum x_i\left(\frac{p_i(\theta)}{p_i}\right) = -2n\sum p_i\left(\frac{p_i(\theta)}{p_i}\right) = 2n\sum p_i\left(\frac{p_i}{p_i(\theta)}\right)$$

37.

Month	# of Deaths
Jan	1668
Feb	1407
Mar.	1370
Apr	1309
May	1341
June	1338
July	1406
Aug	1446
Sep.	1332
Oct	1363
Nov	1410
Dec	1526

$p(H_0)$ = Same "rate" of deaths $= \frac{1}{12}$

$p(H_1)$ = Different "rate" of Deaths $= \sum_i p(x_i) = 1$

$$\chi^2 = \frac{(1668-1410)^2 + \cdots + (1526-1410)^2}{1410}$$

$\boxed{= 79}$

$\chi^2_{11,\,0.995} = 26.76$; $\boxed{\chi^2 > \chi^2_{11,\,0.995} \text{ which is a rejection of null hypothesis at p-value} \approx 0.005.}$

38 ~~38~~

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
~~Male~~	~~3755~~	~~3251~~	~~3777~~									
~~Female~~												

38.

Month	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
Male	3755	3251	3777	3706	3717	3660	3669	3626	3481	3590	3605	3392
Female	1362	1244	1496	1452	1448	1376	1370	1301	1337	1351	1416	1226

Male: $n = 43229$; $E[x] = \bar{x} = 3602$; $\chi^2 = 74.6$; $\chi^2_{11,\,0.995} = 26.76$; Reject H_0

Female: $n = 16379$; $E[x] = \bar{x} = 1364.9$; $\chi^2 = 20.4$; $\chi^2_{11,\,0.975} = 21.92$: Accept H_0 at p-value of 0.025.

$p(H_0)$ = Same "rate" of sickness $= 1/12$

$p(H_1)$ = Different "rate" of sickness $= \sum_i p(x_i) = 1$

39.

Lunar Day	16,17,18	19,20,21	22,23,24	25,26,27	28,29,1	2,3,4	5,6,7	8,9,10	11,12,13	14,15
# of Bites	137	150	163	201	269	155	142	146	143	110

$E[x] = \bar{x} = 162.1$; $\chi^2 = 107$; $\chi^2_{9,\,0.995} = 23.59$; $\boxed{\text{Reject } H_0}$ 0.005

$P(H_0)$ = Same "rate" of bites $= 1/10$

$\boxed{\text{No, temporal trends}}$

$P(H_1)$ = Different "rate" of bites $= \sum_i p(x_i) = 1$

42, $n = 230$

40. Multinomial Distribution:

$$f(\theta) = \frac{n!}{x_1! \cdots x_n!} p_1(\theta)^{x_1} \cdots p_n(\theta)^{x_n}$$

Observation
X_1
X_2

Pearson Chi-Squared:

$$\sum_{i=1}^{2} \frac{(X_i - np_i)^2}{np_i} = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2}$$

$$= \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - n(1-p_1))^2}{n(1-p_1)}$$

$$= \frac{(X_1 - np_1)^2 n(1-p_1)}{np_1(1-p_1)} + \frac{(X_2 - n(1-p_1))^2 p_1}{np_1(1-p_1)}$$

$$= \frac{X_1^2 - 2X_1np_1 + n^2p_1^2 - X_1^2 + 2np_1^2 - np_1^3 + X_2^2p_1 - 2n(1-p_1)p_1 + n^2(1-p_1)^2p_1}{np_1(1-p_1)}$$

$$\boxed{= \frac{(X_1^2 - np_1)^2}{np_1(1-p_1)}}$$

Relationships: $p_1 X_T = X_1$; $x_1 np = 2\frac{x_1 + x_2}{x_1 + x_2} = np_1^3 + X_2^2 p_1 - 2np_1 + 2np_1^2 + n^2p_1 - 2n^2p_1^2 + n^2p_1^3$

$p_2 X_T = X_2$; $+ 4np_1^2$

42. a) mle of p: $f(\theta) = \frac{n!}{\prod_i x_i!} \prod_i p_i^{x_i}$; $\ln f(\theta, \lambda) = \ln(n!) - \sum_i \ln x_i! + \sum_i x_i \log p_i + \lambda(\sum_i p_i - 1)$

$$\frac{d \ln f(\theta, \lambda)}{dp} = \sum_i \frac{x_i}{p_i} + \lambda \; ; \; \boxed{\hat{p}_i = \frac{-x_i}{\lambda}} \; ; \; \lambda = n$$

$$\boxed{p_i = \frac{x_i}{n}}$$ $\boxed{= 0.71}$

b)

Breaks	Frequency
0	157
1	69
2	35
3,4,5	19

$$\chi^2 = \frac{(0-0.71)^2 + (1-0.71)^2 + (2-0.71)^2 + (3-0.71)^2}{0.71}$$

$= 1.06$; $\chi^2_{3,\,0.85} = 1.2504$ $\boxed{\text{Reject above p-value 0.15.}}$

c) $$\Lambda = \frac{\hat{p}^{\sum n_i p_i}(1-\hat{p})^{\sum n_i(1-\hat{p}_i)}}{\prod \hat{p}_i^{n_i p_i}(1-\hat{p}_i)^{n_i(1-\hat{p}_i)}} \; ; \; -2\log\Lambda = \sum_i \frac{(X_i - n_i\hat{p})^2}{n_i\hat{p}(1-\hat{p})} = 2701$$

$\boxed{\Lambda \cong \emptyset}$

41. $X_i = \text{bin}(n_i, p_i)$; $i = 1 \ldots m$

$H_0: p_1 = p_2 = \cdots = p_m$

$H_1: p_1 \neq p_2 \neq \cdots \neq p_m$; $\sum p_i = 1$

$$\Lambda = \frac{P(x|H_0)}{P(x|H_1)} = \frac{P(x|H_0)}{(1 - P(x|H_0))} = \boxed{\frac{\sum_i \binom{n}{m} p^m (1-p)^{n-m}}{1 - \sum_i \binom{n}{m} p^m (1-p)^{n-m}}}$$

43. a) 9207 heads; 8743 tails in 17,950 coin tosses. $\hat{p} = \frac{x}{n} = 8,975$; $\chi^2 = \frac{(9207-8975)^2 + (8743-8975)^2}{8975} = 11.99$

b) n = 5

$T_1 = 3590$

#Heads	Freq.
0	100
1	524
2	1080
3	1126
4	655
5	105

$P(x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{5}{x}(0.5)^x (0.5)^{5-x}$; $x = 0,1,2,3,4,5$ $\quad \chi^2_{0.995} = 7.88$; Reject Ho above p = [illegible]

$X^2 = 21.6$

$\chi^2_{0.999} = 5.893$; $\boxed{\text{Reject } H_0 \text{ above p-value} = 0.001}$

c) $P(x|H_0) = P(\text{coin } \#1 | H_0) = p(\text{coin } \#2 | H_0) = \cdots = p(\text{coin } \#5 | H_0) = \frac{1}{2}$

$X^2 = \sum \frac{[X_i - np_i(\theta)]^2}{np_i(\theta)} = 8.75$; ${}_5\chi^2_{0.9995} = 6.8$; $\boxed{\text{Reject } P(x|H_0) \text{ above p-value} = 0.0005}$

44.

	Haploglobin Type		
	Hp1-1	Hp1-2	Hp2-2
O_i	10	68	112
E_i	12	72	106
$-2\ln\Lambda$	[illegible] 0.91		

$\hat{\theta} = 0.748$

Multinomial Distribution:

$f(\theta) = \frac{n!}{x_1! \cdots x_n!} P_1(\theta)^{x_1} \cdots P_n(\theta)^{x_n}$

Likelihood Ratio: $\Lambda = \prod_{i=1}^{m} \left(\frac{p_i(\theta)}{\hat{p}_i}\right)^{x_i}$; $-2\log\Lambda = 2\prod O_i \log\left(\frac{O_i}{E_i}\right)$

$P(\theta = 1/2, H_0) \neq P(\theta \neq 1/2; H_1)$; ${}_2\chi^2_{0.95} = 2.92$; $\boxed{\text{Accept } P(\theta = 1/2, H_0) \text{ at p-value of } 0.05}$

45. n = 6115 families

#	Frequency	Expected Freq.
0	7	4
1	45	36
2	181	180
3	478	511
4	829	980
5	1112	1336
6	1343	1327
7	1033	969
8	670	516
9	286	195
10	104	50
11	24	8
12	3	1

$\bar{X} = \sum x_i$; $P(X|H_0)$ = Binomial

$p(\bar{x}) = 0.46$ $\quad P(x|H_1) \neq$ Binomial

$X^2 = 551$

${}_{12}\chi^2_{0.9995} = 4.43$

$\boxed{\text{Reject } P(x|H_0) \text{ at p-value} = 0.0005}$

Binomial Random Variable

$P(1) = P$

$P(0) = 1-p$

$P(x) = 0$

$P(x) = \binom{n}{k} p^k (1-p)^{n-k}$

46. $Y = \sin^{-1}\sqrt{\hat{p}}$; $\hat{p} = \frac{X}{n}$; $X \sim bin(n,p)$

$\frac{dY}{d\hat{p}} = \frac{1}{\sqrt{1-\hat{p}^2}} = \frac{1}{\sqrt{(1-\hat{p})\hat{p}}}$

$Var(\sin^{-1}\sqrt{\hat{p}}) \approx n\sqrt{\hat{p}(1-\hat{p})} = \sqrt{n}\, X(1-X)$

47. λ = mean; $Y = \sqrt{X}$; $Y' = \frac{1}{2\sqrt{X}}$

Poisson Distribution

$P(x) = \frac{\lambda^x}{x!} e^{-\lambda x}$;

$\lim_{n\to\infty} \sqrt{n}(P(x) - \theta) \to N(0, \theta)$

$\left(2\sqrt{P(x)} - \frac{5}{\sqrt{n}}, 2\sqrt{P(x)} + \frac{5}{\sqrt{n}}\right)$

$\left(\left(\sqrt{P(x)} - \frac{5}{2\sqrt{n}}\right)^2, \left(\sqrt{P(x)} + \frac{5}{2\sqrt{n}}\right)^2\right)$

48. $E(X) = \mu$; $Var(X) = c\mu^2$

$Y = \int_0^x \frac{d\mu}{Var(x)} = \int_0^x \frac{d\mu}{c\mu^2} = \frac{1}{c}\ln(x)$; $\boxed{c = \frac{\ln(x)}{Y}}$

49.

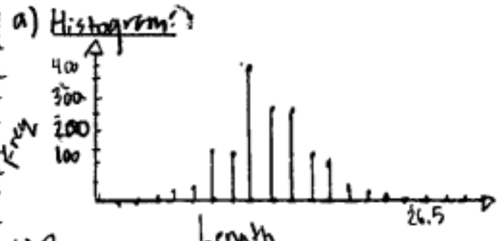
Len	Freq.
18.5	0
19.0	1
19.5	3
20.0	33
20.5	39
21.0	156
21.5	152
22.0	392
22.5	288
23.0	286
23.5	100
24.0	86
24.5	21
25.0	12
25.5	2
26.0	0
26.5	1

n = 1572

E[X] = 22.3

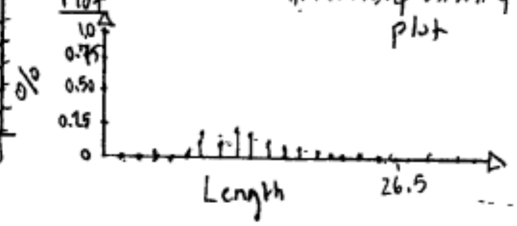
$\hat{n} = 87$

a) Histogram

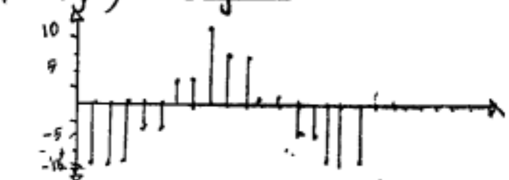


b) Probability Plot

* First was probability plot



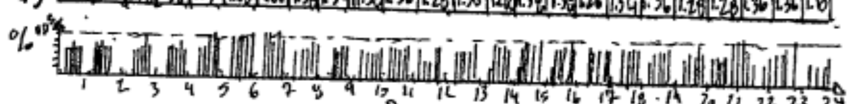
c) Hanging Rootogram

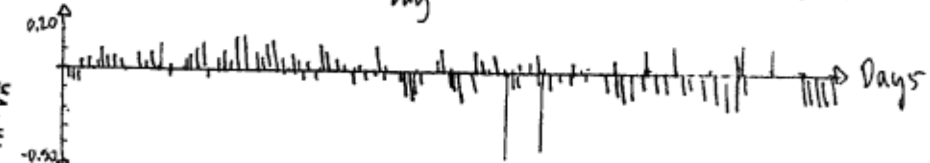


50.

	Days											
	1	2	3	4	5	6	7	8	9	10	11	12
#1	1.40	1.40	1.80	1.54	1.52	1.62	1.58	1.62	1.60	1.38	1.34	1.50
#2	1.28	1.34	1.44	1.50	1.46	1.58	1.64	1.46	1.44	1.34	1.26	1.46
#3	1.36	1.54	1.46	1.48	1.42	1.62	1.62	1.38	1.46	1.36	1.08	1.28
#4	1.38	1.44	1.50	1.52	1.58	1.76	1.72	1.42	1.34	1.58	1.03	1.18
#5	1.44	1.46	1.30	1.56	1.38	1.68	1.60	1.38	1.34	1.32	1.36	1.28

	Days											
	13	14	15	16	17	18	19	20	21	22	23	24
#1	1.26	1.52	1.50	1.48	1.32	1.16	1.24	1.30	1.30	1.35	1.32	[illegible]
#2	1.50	1.50	1.42	1.32	1.40	1.54	1.22	1.48	1.52	1.46	1.22	1.20
#3	1.52	1.48	1.38	1.48	1.40	1.40	1.20	1.28	1.76	[illegible]	1.22	1.10
#4	1.32	1.34	1.36	1.30	1.26	1.16	1.30	1.18	1.16	1.42	1.18	1.06
#5	1.50	1.40	1.34	1.38	1.26	1.54	1.36	1.28	1.28	1.36	1.34	1.10



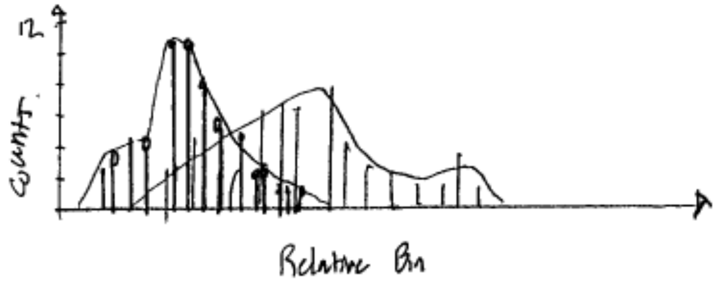


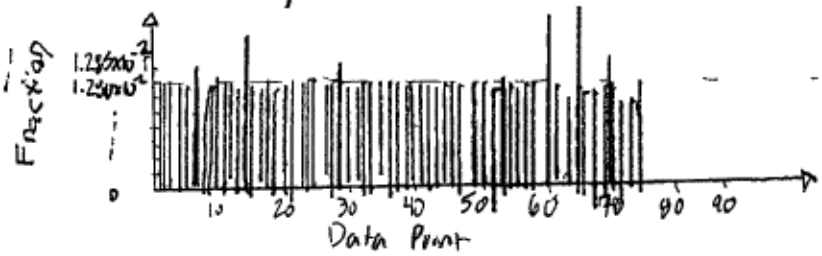
Note: Probability plots were correctly modelled by Problem #64

The horizontal bands of Figure 9.6 represent groupings of data with similar observations

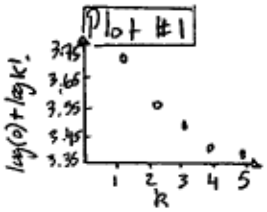
52. See chapter 9: Problem 52.

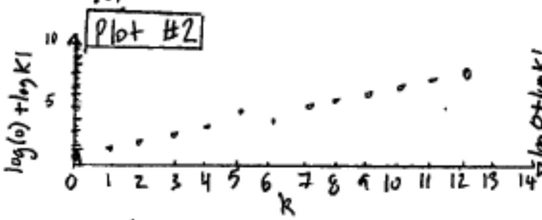
Probability Plot.

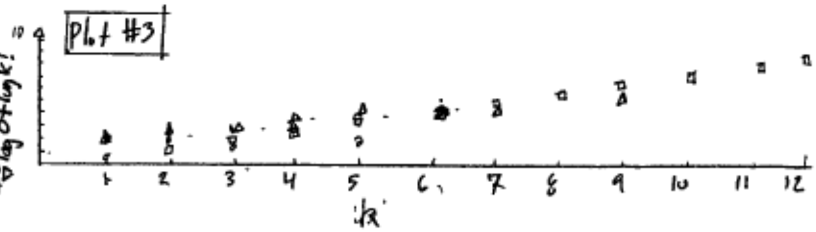




53. $E_K = nP(X=K) = ne^{-\lambda}\dfrac{\lambda^K}{K!}$; $\log E_K = \log n - \lambda + K\log\lambda - \log K!$





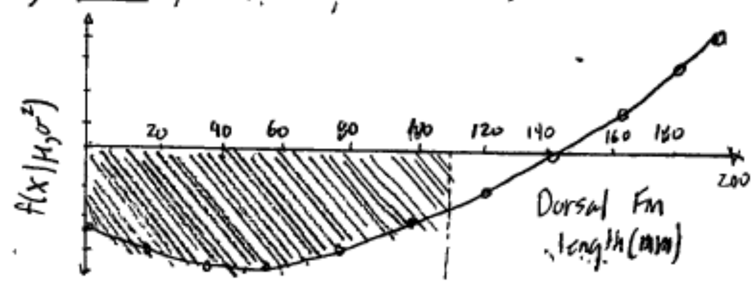


$\boxed{\text{The plots contain regions of linearity.}}$

54. $Y = \log(X)$ a) Log Normal Distribution

$$f(x|\mu,\sigma^2) = \ln\phi(x|\mu,\sigma^2) \boxed{-\frac{(X-\mu)^2}{2\sigma^2} - \ln\sqrt{2\pi\sigma^2}}$$

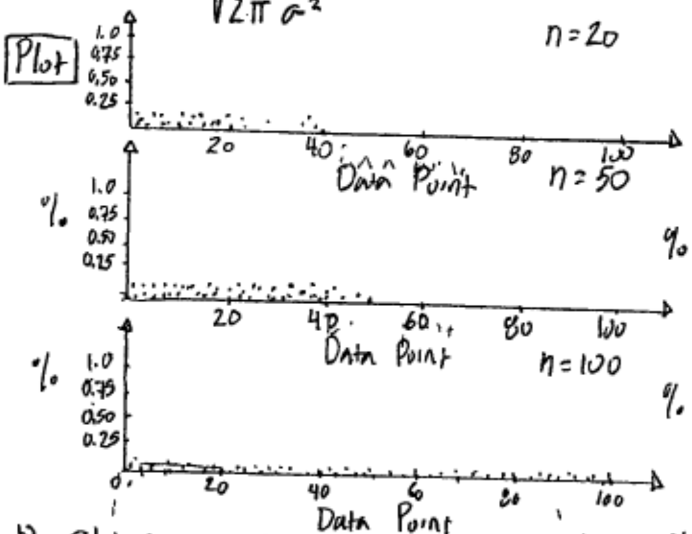
b) Plot $\mu = 114.67$ mm ; $\sigma = 35.87$ mm



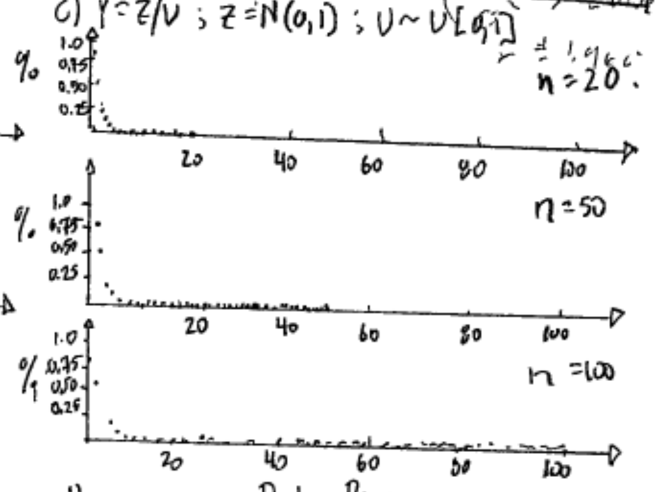
The data fits lognormal until length 115mm which is above a p-value of 0.05.

55. a) Normal Distribution:

$$f(x|\mu,\sigma^2) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$$



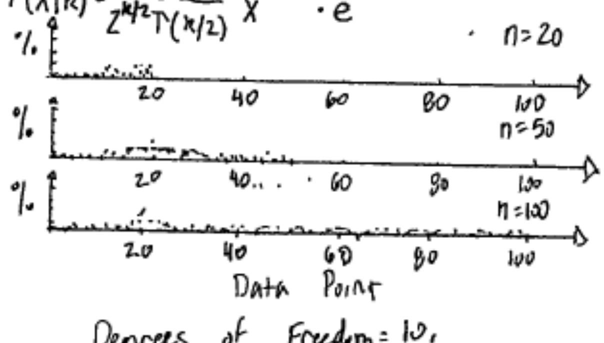
c) $Y = Z/V$; $Z = N(0,1)$; $V \sim U[0,1]$



Note: Probability plots were correctly modelled by problem #64

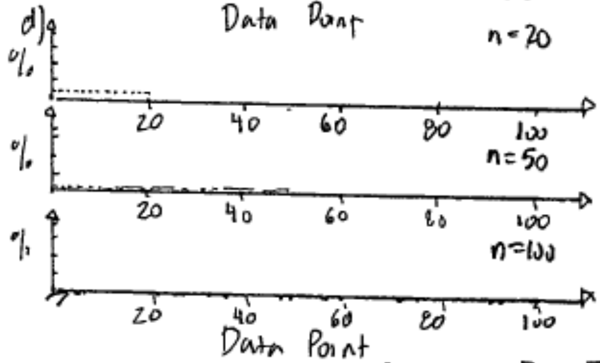
b) Chi-Square Distribution

$$P(X|k) = \frac{1}{2^{k/2}\Gamma(k/2)} X^{k/2-1} \cdot e^{-x/2}$$



Degrees of Freedom = 10.

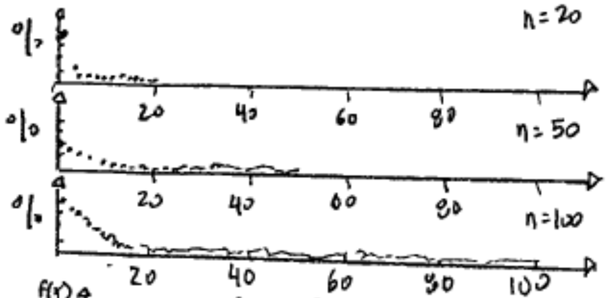
d)



Uniform Distribution

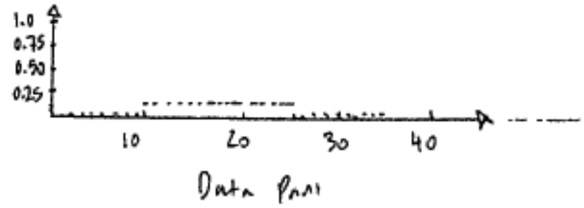
$$f(x) = \frac{1}{b-a}$$

e.

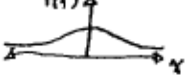


f. The distributions plotted are seperable from each other, and a normal distribution

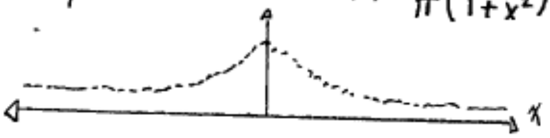
58. $F(x) = 1 - e^{-\lambda x}$; $\lambda = \frac{1}{\bar{x}}$



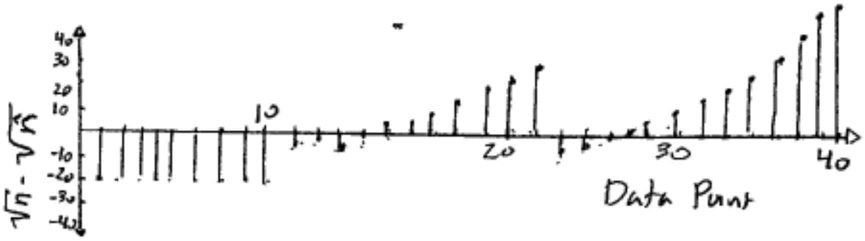
56.



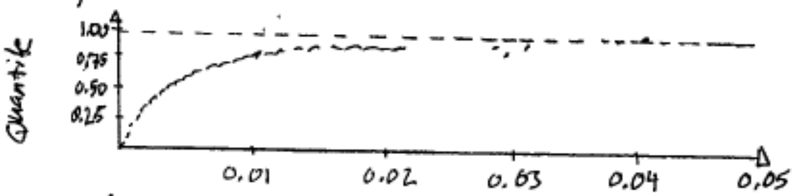
57. Cauchy Distribution $f(x) = \frac{1}{\pi}\left(\frac{1}{1+x^2}\right)$



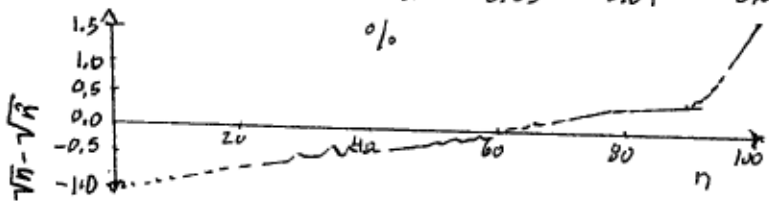
59.



59. 60, $n = 76$ a)

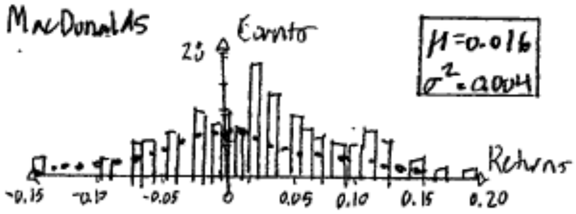


b)

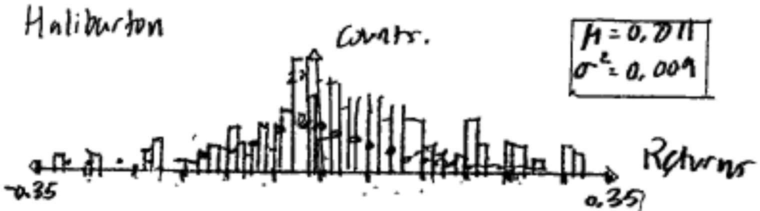


The appearence of the plot represents a convergence of % stress failure for Kevlar 49/epoxy against quantile grouping of an exponential distribution.

61. a) MacDonalds



Haliburton



b) The more volatile stock company is Haliburton

Note: Probability plots were correctly modelled by problem #64.

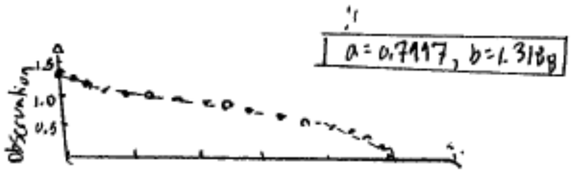
62. Poisson Dispersion Test

$$\Lambda = \prod \left(\frac{\bar{x}}{x_i}\right)^{x_i} e^{x_i - \bar{x}}$$

$$-2\log\Lambda = 2\sum x_i \log\left(\frac{x_i}{\bar{x}}\right) \approx \frac{1}{\bar{x}}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

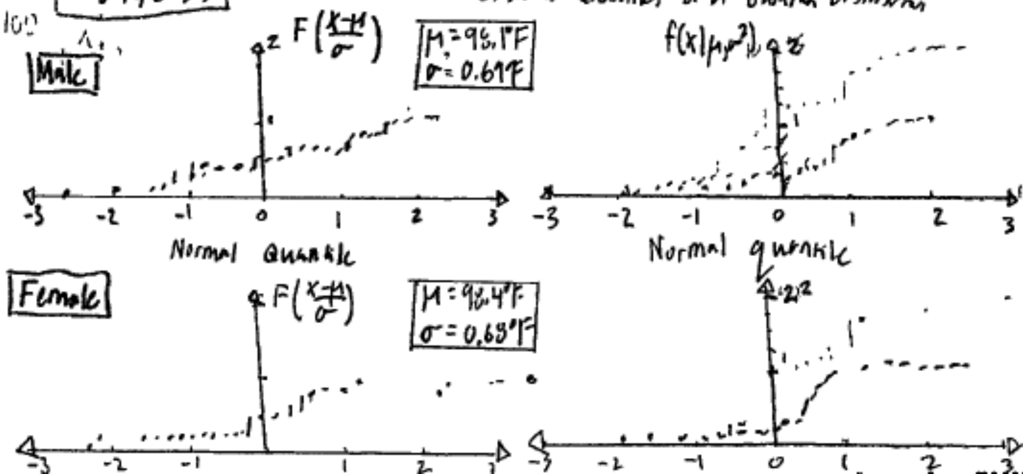
$$\boxed{= 3476.5}$$

63.



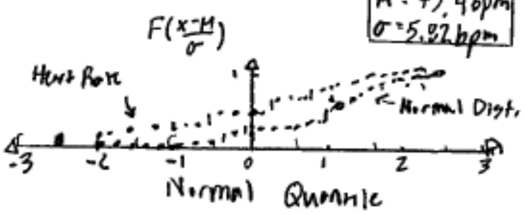
Ordered Quantiles of a Gamma Distribution

64. a)

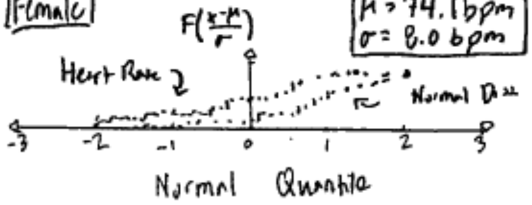


The assessment of body temperatures for both male and female demonstrate a higher proportion near the mean than normally distributed.

b Male



Female



Male and Female heart rate both contain a larger proportion of reading near the mean than a Normal distribution.

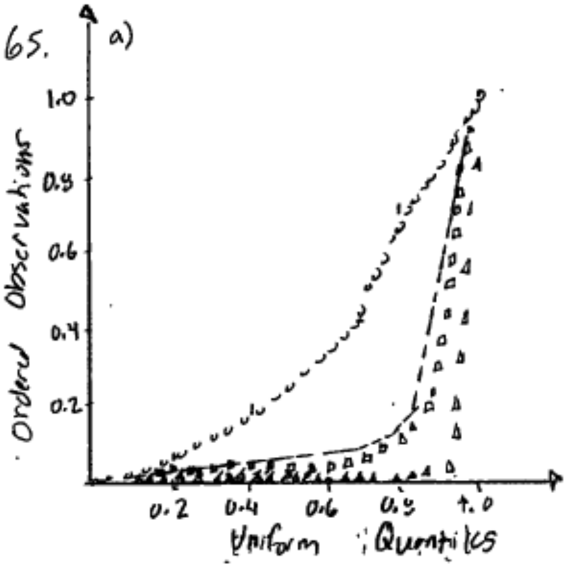
c. $P(x = 98.6°F \mid H_0)$ vs $P(x \neq 98.6°F \mid H_1)$ Male

$\chi^2 = 0.3135$; $\chi^2_{65\ 0.005} = 39.38$; Accept $P(x \mid H_0)$ at p-value > 0.995

$P(x = 98.6°F \mid H_0)$ vs $P(x \neq 98.6°F \mid H_1)$

$\chi^2 = 0.3956$; $\chi^2_{65\ 0.005} = 39.38$; Accept $P(x \mid H_0)$ at p-value > 0.995

65. a)



□ = Small Chromatin
$\mu =$ $\sigma =$
○ = Medium Chromatin
$\mu = 426$ $\sigma = 1.94$
Δ = Long Chromatin
$\mu = 260$ $\sigma = 1.38$

---- = Rayleigh Dist ($\sigma^2 = 1$)

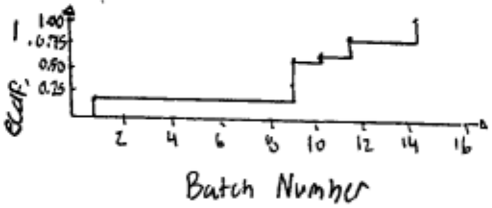
b) $\chi^2_{short} = 24.43$ $\chi^2_{med} = 117.4$ $\chi^2_{long} = 182.0$

df = 95 df = 131 df = 248

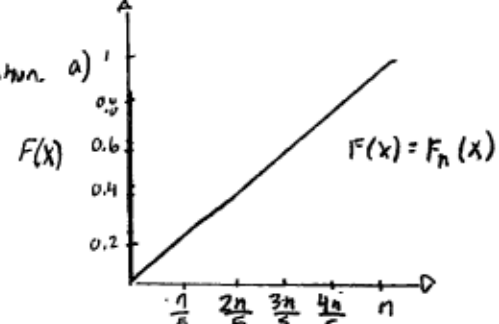
$\chi^2_{95\ 0.995} = 63.25$ $\chi^2_{131\ 0.75} = 119.750$ $\chi^2_{248\ 0.995} = 194.391$

Accept at p-value > 0.005 | Reject at p-value > 0.25 | Accept at p-value > 0.005

Chapter 10:

1. 

2. $X_1, X_2 \ldots X_n$ with Uniform Distribution a)

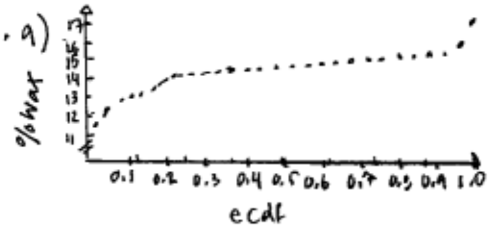


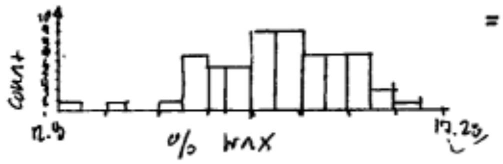
b) $F_n(x) - F(x)$ monotonically increasing function, while $F_n(x) - F(x)$ is a decreasing function

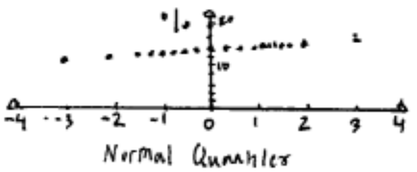
3. Lower quartile ($T < 63.9°F$)
Upper quartile ($T > 63.8°F$)

4. $I_{(-\infty, x]}(X_i)$ are independent random variables because the data are independent.

5. $Cov[F_n(u), F_n(v)] = \frac{1}{n}[F(m) - F(u)F(v)]$; $m = \min(u,v)$; $Cov(F_n(u), F_n(v)) = E(F_n(u)F_n(v)) - E(F_n(u))E(F_n(v))$
$= \frac{1}{n} F(\min(u,v)) - F(u)F(v)$
$= \frac{1}{n}[F(m) - F(u)F(v)] = (+)$

6. a) 





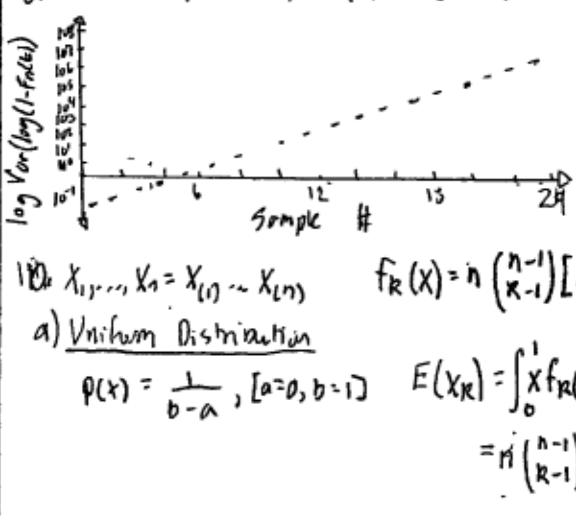
$q_{0.90} = 16.6\%$ $q_{0.50} = 14.7\%$ $q_{0.10} = 12.8\%$

$q_{0.75} = 15.9\%$ $q_{0.25} = 13.5\%$

b) Average (%) Wax = 85% ; At dilution of 1%, 3% and 5% microcrystalline wax are measurable quantities with a standardized average (%) wax.

7. The 10% weakest guinea pigs die within 300 days, and the 10% strongest survive until 400 days. While the median population live until 350 days.

8. $n = 100$; $\lambda = 1$ a) $S(t) = P(T > t) = 1 - F(t)$; $S_n(t) = 1 - F_n(t) \cdot 1 - e^{-t}$; $Var[\log(1 - F_n(t))] = \frac{1}{n}\left(\frac{F(t)}{1 - F(t)}\right)$



b) Several plots show exponentially growing standard deviation for the survival function

9. Method of Propagation Error: $Y = g(X) \approx g(\mu_X) + (X - \mu_X) g'(\mu_X)$
$E[Y] \approx g(\mu_X) + \frac{1}{2}\sigma_X^2 g''(\mu_X)$
$g(\mu_X) = \log S_n(\mu_X)$
$= \log(1 - F(\mu_X))$
$g'(\mu_X) = \frac{-F(\mu_X)}{1 - F(\mu_X)} = \frac{1}{n}\left(\frac{-F(x)}{1 - F(x)}\right)$

$$\boxed{E[Y] \approx \log(1 - F(\mu_X)) - \frac{1}{2n}\frac{F(x)}{1 - F(x)}}$$
$\boxed{(-), \text{Error}}$

10. $X_1, \ldots, X_n = X_{(1)} \ldots X_{(n)}$ $\quad f_k(x) = n\binom{n-1}{k-1}[F(x)]^{k-1}[1 - F(x)]^{n-k} f(x)$

a) Uniform Distribution

$p(x) = \frac{1}{b-a}$, $[a=0, b=1]$ $\quad E(X_k) = \int_0^1 x f_k(x)\,dx = n\binom{n-1}{k-1}\int_0^1 x\left[\int_0^x f(x)\right]^{k-1}\left[1 - \int_0^x f(x)\right]^{n-k} \cdot 1\,dx$

$= n\binom{n-1}{k-1}\int_0^1 x^k (1-x)^{n-k}\,dx = n\binom{n-1}{k-1} \cdot Beta(k+1, n-k+1)$

$= n\frac{(n-1)!}{(k-1)!(n-k)!} \cdot \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \frac{n\,(n-1)!}{(k-1)!(n-k)!}$

$= \frac{(k!\,n-k)!}{(n+1)!} \frac{1}{\ } \boxed{\frac{k}{n+1}}$ $\quad \Gamma(p) = (p-1)\Gamma(p-1) = (p-1)!$

$E[X_k^2] = \int_0^1 x^2 f(x)\,dx = n\binom{n-1}{k-1}\int_0^1 x^{k+1}(1-x)^{n-k}\,dx = n\binom{n-1}{k-1} \times Beta(k+2, n-k+1)$; $Var(X) = \frac{k(k+1)}{(n+1)(n+2)} - \frac{k^2}{(n+1)^2}$

$= n\frac{(n-1)!}{(k-1)!(n-k)} \times \frac{\Gamma(k+2)\Gamma(n-k+1)}{\Gamma(n+3)} = n\frac{(n-1)!}{(k-1)!(n-k)!} \cdot \frac{(k+1)(n-k)!}{(n+2)!} = \frac{k(k+1)}{(n+1)(n+2)}$

$$\boxed{= \frac{1}{(n+2)}\left(\frac{k}{n+1}\right)\left(1 - \frac{k}{n-1}\right)}$$

b. $X_i = F(Y_i)$; $Y_i = F^{-1}(X_i)$; $F(Y_i) = \int_0^y f(y)dy = y$; $\boxed{X_i = Y_i}$; $Y_{(k)} = F^{-1}(X_{(k)}) \approx F^{-1}\left(\frac{k}{k+1}\right) + \left(X_{(n)} - \frac{k}{n+1}\right) \frac{d}{dx} F^{-1}(x)\Big|_{k/n+1}$

$= E[Y_k] + \frac{1}{2}\sigma_{Y_k}^2 F''(Y_k)$

$\boxed{E[Y_k] = F^{-1}\left(\frac{k}{k+1}\right)}$

$\boxed{Var(Y_k) = \frac{k}{n+1}\left(1 - \frac{k}{n+1}\right) \frac{1}{f(F^{-1}[k/(n+1)])^2} \left(\frac{1}{n+2}\right)}$

c. If $p = \frac{k}{n+1}$; $var(Y_k) = p(1-p) \cdot \frac{1}{f(x)^2} \cdot \frac{k+1}{n+2} \approx \boxed{\frac{p(1-p)}{n f(x)^2}}$

$\frac{1}{n}$ $\frac{k}{n+2}$

d. $N(\mu, \sigma^2)$; Median $= x = 1/2$

$Var(1/2) = \boxed{\frac{1}{n} \frac{p(1-p)}{f(1/2)^2}}$

11. $F(t) = 1 - e^{-\kappa t^{\beta}}$; Hazard Function : $P(t \le T \le t+\delta \mid T \ge t) = \frac{P(t \le T \le t+\delta)}{P(T \ge t)}$

$= \frac{F(t+\delta) - F(t)}{1 - F(t)} = \frac{-e^{-\kappa(t+\delta)^{\beta}} + e^{-\kappa(t)^{\beta}}}{e^{-\kappa(t)^{\beta}}}$

$\boxed{= 1 - e^{-\kappa(\delta)^{\beta}}}$ -or- $\frac{\frac{d}{dt}F(t)}{1-F(t)} = \frac{\kappa\beta t^{\beta-1} \cancel{e^{-\kappa t^{\beta}}}}{\cancel{e^{-\kappa t^{\beta}}}}$

12. $f(t) = h(t) e^{-\int_0^t h(s)ds}$; $h(t) = \frac{F(t)}{1-F(t)}$

$-\int_0^t \frac{F(s)}{1-F(s)} ds = -\int_0^t \frac{u-u}{u} du = \int_0^t \frac{u+1}{u} du = -\int_0^t 1 + \frac{1}{u} du$

$= F(s) \Big|_0^t - \ln(|F(s) - 1|) \Big|_0^t = F(t) - F(0) - \ln(|F(t) - 1|) + \ln(|F(0) - 1|)$

$= F(t) - \ln(F(t) - 1)$; $f(t) = \frac{F(t)}{1-F(t)} e^{-F(t) - \ln(F(t)-1)} = \boxed{-F(t) e^{-F(t)}}$

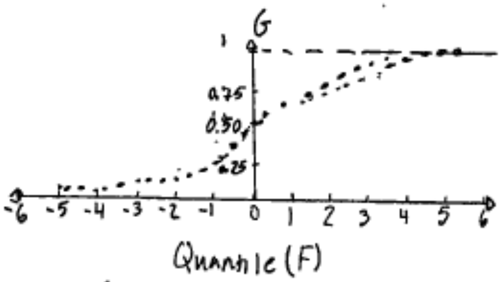
13. A probability distribution with increasing failure rate would be a Rayleigh distribution because $h(t)$ is positive. The uniform distribution has form e^{-u} too.

14. A probability distribution with decreasing failure rate would be the exponential distribution with the form $h F(t) e^{-F(t)}$.

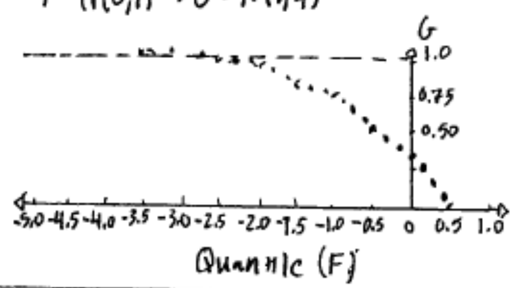
15. T = Time of release. $h(t) = \frac{f(t)}{1-F(t)} = \frac{\frac{1}{24}}{1 - \frac{t}{24}} = \boxed{\frac{1}{24-t}}$; The smallest of t is 0 hours while largest 24 hours.

$h(5) = \frac{1}{19}$ vs $h(1) = \frac{1}{24}$; The prisoner has greater likelihoods.

16. $F = N(0,1)$; $G = N(1,1)$



$F = N(0,1)$; $G = N(1,4)$

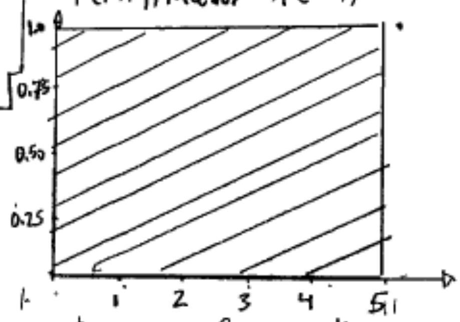


18. ill $\propto$ longer lifetime

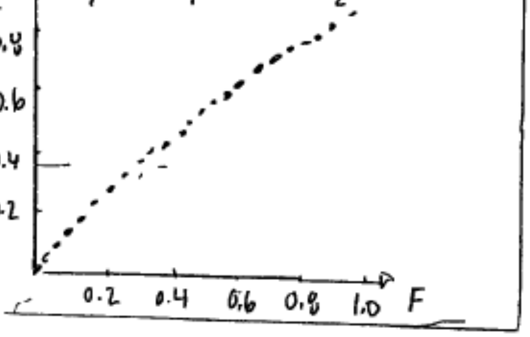
healthy $\propto$ shorter lifetime

P(healthy | Placebo) = P(healthy)

P(ill | Placebo) = P(ill)

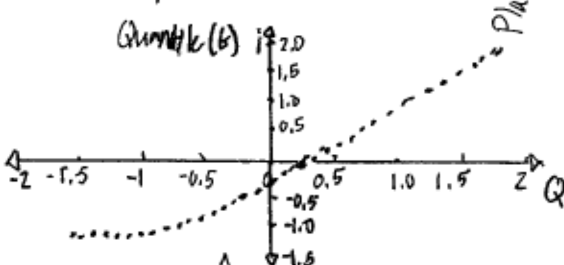


17. $F = \lambda_1 e^{-\lambda_1 x}$; $\lambda_1 = 1$; $G = \lambda_2 e^{-\lambda_2 x}$; $\lambda_2 = 2$



19. $F(x) = x$; $0 \le x \le 1$

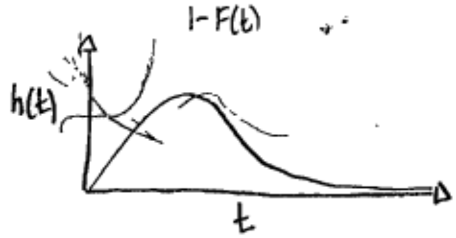
$G(x) = x^2$; $0 \le x \le 1$



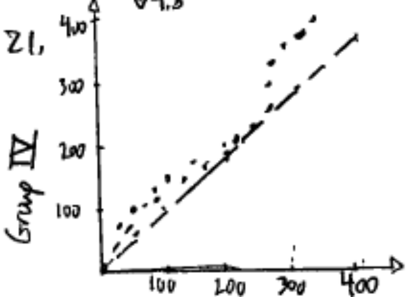
20. Hazard Function : $h(t) = \frac{f(t)}{1-F(t)}$

$f(t) = (-)$

$1 - F(t) = (\to)$



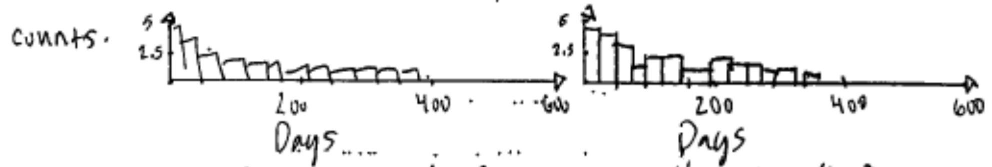
21.



$\boxed{\text{Group IV is living longer than Group V}}$

22. Survival Function: $S(t) = 1 - F(t)$; $S_n(t) = 1 - F(t)$

23. $X_{(k)}$ & $Y_{(k)} = \frac{k}{n+1}$;



Linear Interpolation Function:

$y = y_0 \left(\frac{x_1 - x}{x_1 - x_0}\right) + y_1 \left(\frac{x - x_0}{x_1 - x_0}\right)$; $\frac{k}{(n+1)} \le p \le \frac{(k+1)}{(n+1)}$; $\frac{-k}{(n+1)} \ge -p \ge -\frac{(k+1)}{(n+1)}$; $1 - \frac{k}{(n+1)} \ge 1 - p \ge 1 - \frac{(k+1)}{(n+1)}$

$\boxed{f(X) = X_{(k)} (n+1)\left(\frac{k+1}{n+1} - p\right) + X_{(k+1)} (n+1)\left(p - \frac{k}{n+1}\right)}$; 1, 2, 3, 4, 5, 6.

$\boxed{L.Q. = \frac{(n+1)}{4} = 2}$ $\boxed{\text{Median} = 3.5}$ $\boxed{U.Q. = \frac{3(n+1)}{4} = 5}$

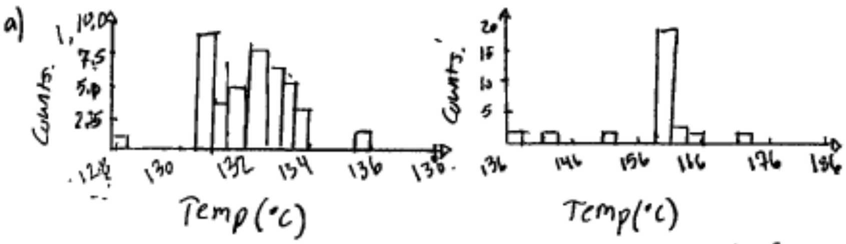
24. Empirical Distribution: F_n = [illegible]
Theoretical Distribution: F = [illegible]

25. $y_p = c x_p$; $F(x) = p$; $x = F^{-1}(p) = \frac{y_p}{c}$; $\boxed{F\left(\frac{y_p}{c}\right) = p = G(y)}$

26.

Rhodium: Irridium:

a)

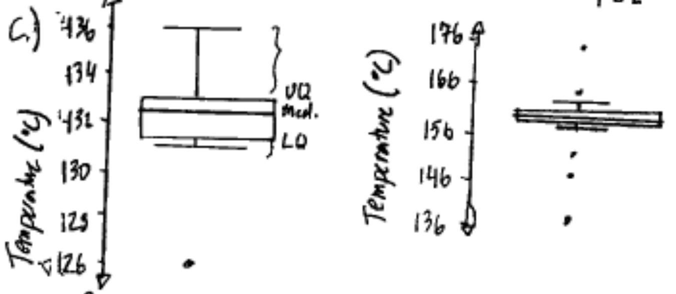


b)

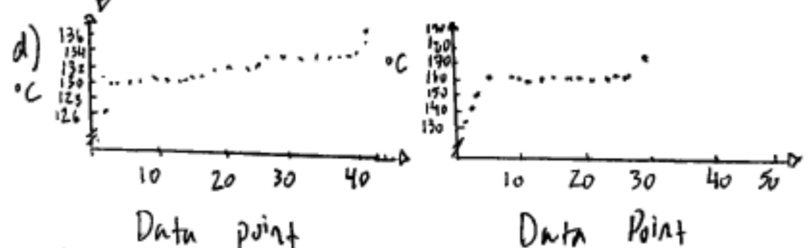
		Stem	Leaf
1	1	126	: 4
13	12	131	: 1234569
25	12	132	: 123456789
37	12	133	: 03458
39	2	134	: 12
40	1	135	: 7

		Stem	Leaf
1	1	136	: 6
2	1	145	: 2
3	1	151	: 5
14	11	159	: 1235678
24	10	160	: 123468
25	1	161	: 7
26.	1	162	: 9

c)



d)



e) Yes, independence and identically distributed for the experimental measurements, in addition to plots of data.

f. Rhodium Irridium:

Rhodium	Irridium
$\mu = 132.42$ °C	$\mu = 158.81$ °C
$\mu_{90\%} = 132.46$ °C	$\mu_{90\%} = 159.86$ °C
$\mu_{80\%} = 132.47$ °C	$\mu_{80\%} = 159.84$ °C

g. $\boxed{SE = 0.24 °C}$ $\boxed{132.42 \pm 0.40 °C}$ $\boxed{SE = 1.17 °C}$ $\boxed{158.81 \pm 1.93 °C}$

h. $\boxed{CI = (132.02 °C, 132.82 °C)}$ $\boxed{CI = (156.88 °C, 160.74 °C)}$

i. 10% trim $\boxed{SE = 0.25 °C}$ 10% trim $\boxed{SE = 1.33 °C}$

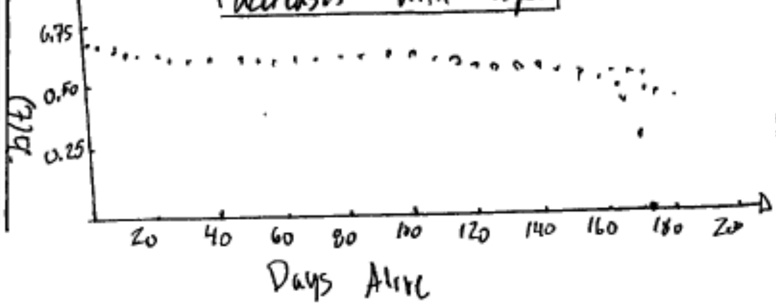
20% trim $\boxed{SE = 0.29 °C}$ 20% trim $\boxed{SE = 1.58 °C}$

j. $\boxed{\text{Median} = 132.65 °C}$ $\boxed{\text{Median} = 159.8 °C}$

k. 10% trim $\boxed{CI = (132.05 °C, 132.87 °C)}$ 10% trim $\boxed{CI = (157.67 °C, 162.05 °C)}$

20% trim $\boxed{CI = (132.00 °C, 132.93 °C)}$ 20% trim $\boxed{CI = (157.23 °C, 162.45 °C)}$

27. No, the hazard function decreases with age.



28. $n=3$; $P(X_{(1)} < \eta < X_{(2)}) = 1 - P(\eta < X_{(1)} \text{ or } \eta > X_{(2)}) = 1 - P(\eta < X_{(1)}) - P(\eta > X_{(2)})$
$= 1 - 1/2 = 1 - \frac{1}{2^3}\sum_{j=0}^{1}\binom{4}{j} = 1 - \frac{1}{8}(1+4) = \boxed{= 37.50\%}$

$P(X_{(1)} < \eta < X_{(3)}) = 1 - P(\eta < X_{(1)} \text{ or } \eta > X_{(3)}) = 1 - P(\eta < X_{(1)}) - P(\eta > X_{(3)}) = 1 - \frac{1}{2^{n-1}}\sum_{j=0}^{1}\binom{5}{j} = \boxed{62.50\%}$

29. a) The distribution is binomial because probability for success and failure exist when considering an outlier distribution. $= 1 - 1/2 \ldots (1+5) \ldots 31.25\%$

b) $P(N \geq 10) = \boxed{B}(n \geq 16/34, p = 5/26) = \boxed{0.018}$

c) The probability of 1000 bootstrap samples would be $1000 \cdot B(n \geq 10, p\, 5/26) = \boxed{= 13.80}$

d) The probability that every sample is an outlier would be $(5/26)^{26} = \boxed{= 2.42 \times 10^{-19}}$

30. By sampling 1000 times without replacement the bootstrap standard deviation was 0.64 vs. the actual standard deviation of 0.37.

31. a) n = number of samples ; n = probability ; $\boxed{n^n}$

b) $n=3$, $X_1 = 1$, $X_2 = 3$, $X_3 = 4$; $\boxed{x \in \{(1,1,1), (1,1,3), (1,1,4), (1,3,1), (1,4,1), (3,1,1), (4,1,1), (1,3,4), (1,4,3), (3,3,3), (3,1,3), (3,4,3), (1,3,3), (4,3,3), (3,3,1), (3,3,4), (3,1,4), (3,4,1), (4,4,4), (4,4,1), (4,4,3), (4,1,4), (4,3,4), (1,4,4), (3,4,4), (4,1,3), (4,3,1)\}}$

c)

Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Mean	1.00	1.67	2.00	1.67	2.00	1.67	2.00	2.67	2.67	3	2.33	2.00	2.33	2.00	2.33	2.33	2.67	2.67	4.00	3.00	3.67	3.33	3.67	3.00	3.67	2.67	2.67

$\bar{X}$	1.00	1.67	2.00	2.33	2.67	3.00	3.33	3.67	4.00
$P(\bar{x})$	1/27	3/27	3/27	3/27	6/27	3/27	3/27	3/27	1/27

d) $S_{\hat{\theta}} = \sqrt{\frac{1}{n-1}(\bar{X}_i - \bar{\bar{X}})^2} = \boxed{0.5366 \text{ vs } 0.5469}$

32. The Median Absolute Deviation from the median (MAD) as defined by the median of $|X_i - \bar{X}|$ is approximated by a bootstrap through sampling the set produced by the definition.

33. The mean and standard deviation.

34. $\psi(x) = |x|$; $\psi'(x) = 0$ @ $x=0$; Median of $f(x) = |x|$: is 0.

35. The proportion of points marked by an asterisk would be <5% of sample because of the inner quartile range (IQR)

36. The IQR is divided by 1.35 because for a double sided distribution each tail is subdivided by 0.675σ, which each represent Q_1 and Q_3 of the sample.

Median Absolute Deviation from the median (MAD) contains 0.675 because of the quartile range being 0.675σ.

37. a) $\boxed{\text{Mean} = 14.58,\ \text{Median} = 14.57\quad \text{Mean}(10\%) = 14.59\quad \text{Mean}(20\%) = 14.59}$

b) $\boxed{14.58 \pm 1.27\%\ \text{wax}}$

c) $\boxed{CI = (13.31;\ 15.85)}$

d) $SE_{10\%} = \frac{\sigma}{\sqrt{n}}\sqrt{1 - \left(\frac{n-1}{N-1}\right)} = \boxed{0.056}$

$SE_{20\%} = \boxed{0.08\%}$

f. $\boxed{SD = 0.769\% \text{ wax}; IQR = 1.09\% \text{ wax}; MAD = 0.58\% \text{ wax}}$

g. SE = 0.17% wax; Sampling Distribution

15.13, 15.15, 15.18, 15.21, 15.22, 15.31, 15.38, 15.4, 15.47
15.47, 15.49, 15.56, 15.63, 15.91, 17.09.

38. Cauchy Distribution:

$$f(x) = \frac{1}{\pi}\left(\frac{1}{1+x^2}\right); -\infty < x < \infty$$

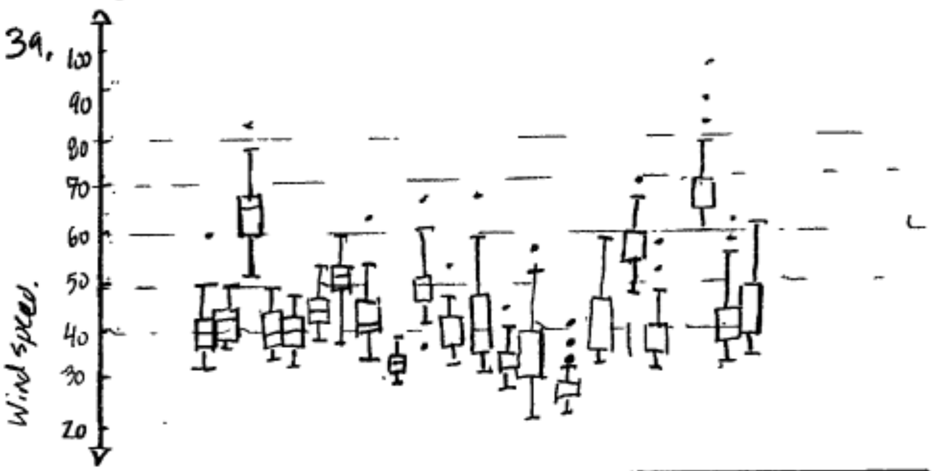
$\mu = 0.009813$	SD = 0.039019.
Median = 0	

Distribution

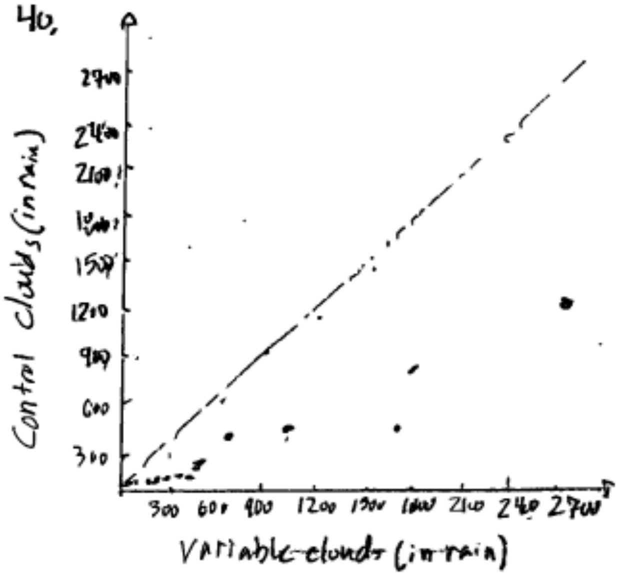
$\mu = 0.133712$	SD = 16.52375
Median = 0.000506	

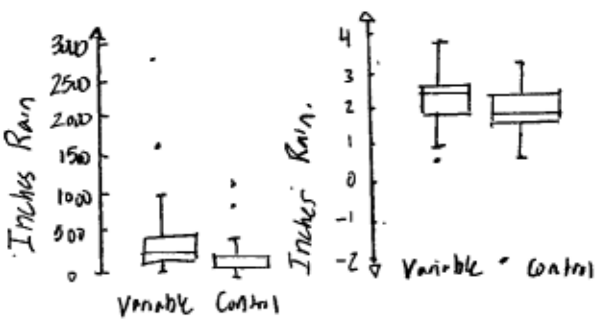
Sample.

39.

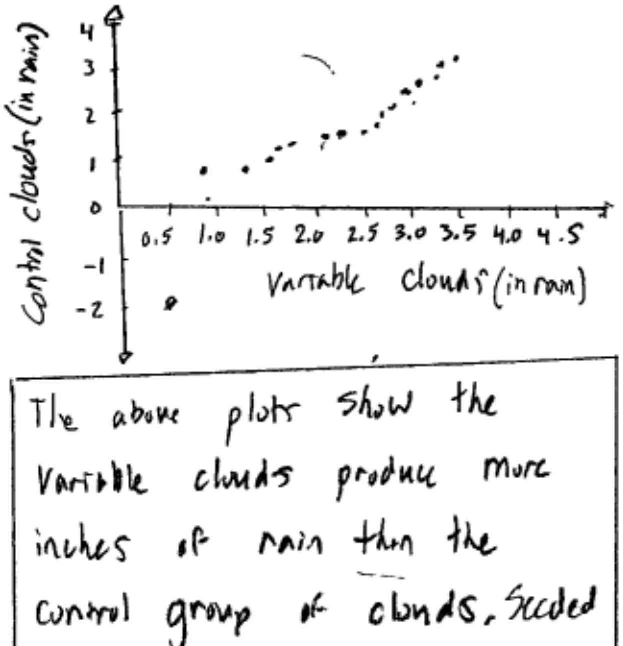


40.





The variable or "seeded" clouds produced more inches of rain than the control group of clouds



41. $P(X > x_p) = 1 - p$; $p(X < x_p) = p$;

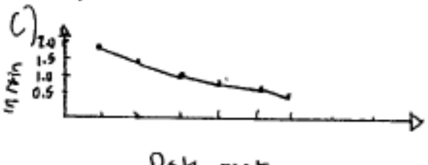
$$= 1 - P(X < p) = \boxed{1 - \sum_{i=0}^{n}\binom{n}{i}p^i(1-p)^{n-i}}$$

42. a) Skewed b)

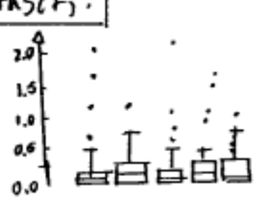
Year	'60	'61	'62	'63	'64
Average	0.22	0.27	0.18	0.26	0.19
Median	0.015	0.075	0.02	0.11	0.055

The median is different from average because of the large skew of the datasets.

c)



d)

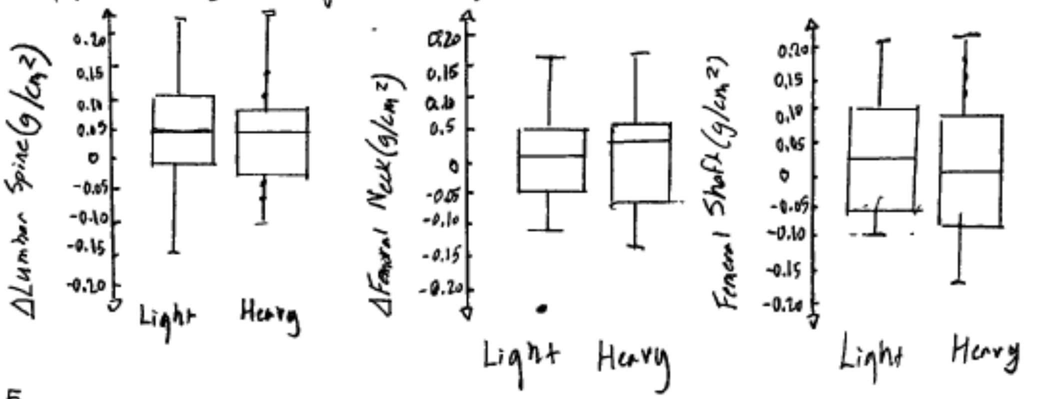


e) The wet years were '61 and '63, while dry years '62 and '64. The reason for wet years were because of storms not daily rain fall.

The above plots show the variable clouds produce more inches of rain than the control group of clouds. Seeded clouds produced more rain exponentially vs control set.
The box plots of each graph would be much different because of mean and standard deviation differences.

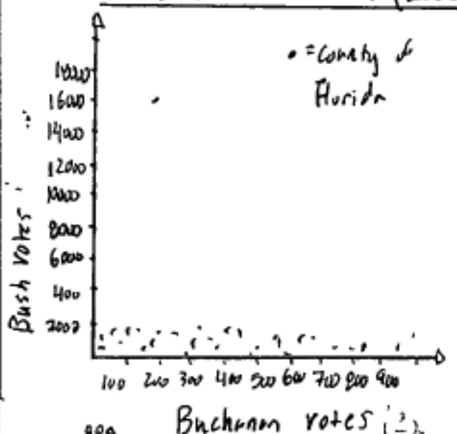
43. When evaluating Kevlar [70%, 80%, 90%], the data's mean, median, and standard deviation demonstrate increasing skew for increasing stress levels.

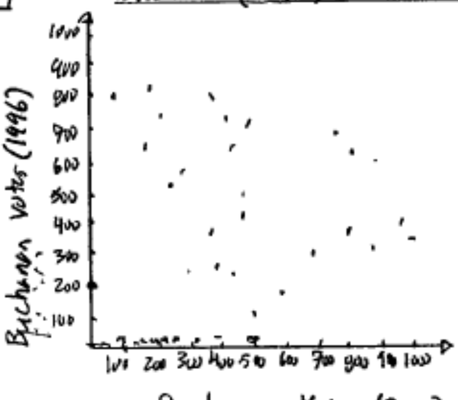
44. Light Smoking (≤ 7 cigarettes/day) ; Heavy Smoking (> 7 cigarettes/day)



The bone density with light vs heavy smoking twins led to different conclusions. Although both light and heavy smoking twins had larger change to bone mass than the nonsmoking group. Data not shown.

45.

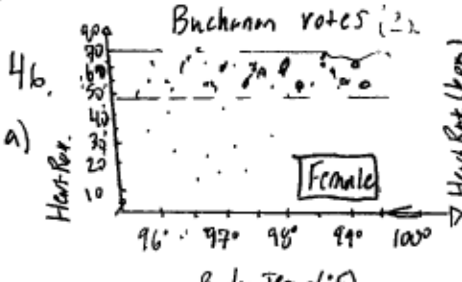


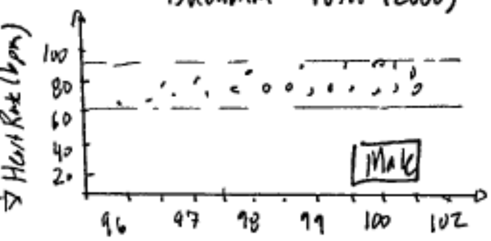


The Buchanan (1996) vs. Buchanan (2000) shows similar amount of popular votes, while Bush's entry into presidency was led with 2x amount of votes, and in Monroe count 16,000+ votes.

46.

a)





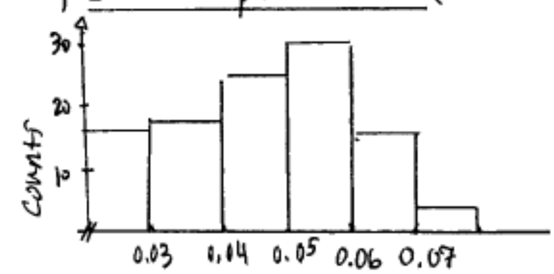
Heart rate for males and females remained within the same range, but male temperatures did contain maximum upwards of 100°F.

b) $r_{Male} = 1.72$; $r_{Female} = 2.61$; $\rho_{Male} = 0.034$ $\rho_{Female} = 0.077$

Heart rate and body temperature show positive correlation.

c) The female body temperature shows greater linear correlation than males.

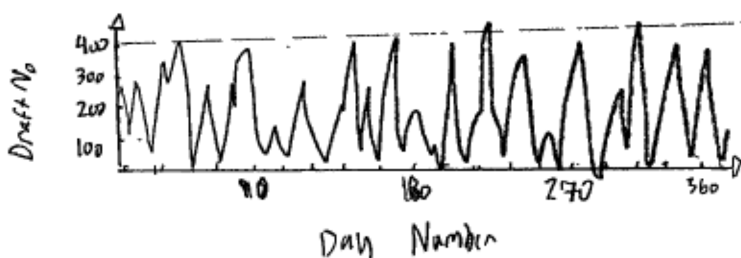
47. a) Duration per Interval (min/interval)



b) Duration per Interval (min/interval)

Old Faithful when measured tends to erupt for 3/6 seconds per "gush".

48. a) Draft Number vs Day Number.



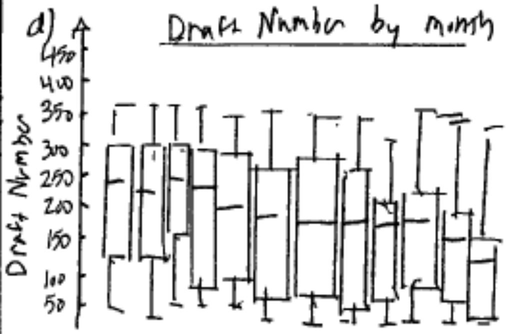
No trend

b) Pearson Rank Coefficient (r) = -2.0×10^{-5} | No linear correlation or trend.

Spearman Rank Correlation Coefficient (ρ) = -0.23. | Little negative correlation to justify

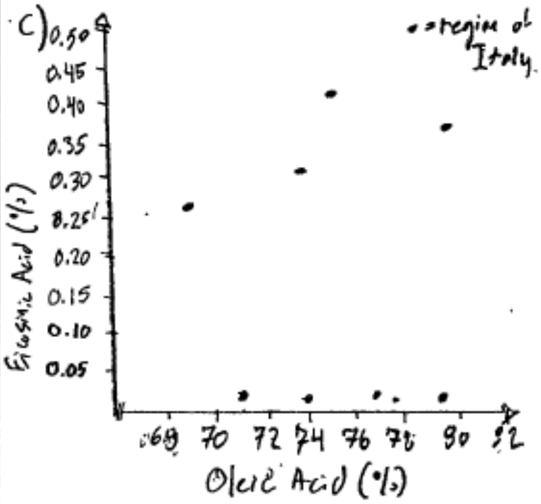
c) Statistical Significance via method listed in the book justified no correlation

d)



No pattern for 1970 lottery because max/min were equivalent and mean within 50 drafts.

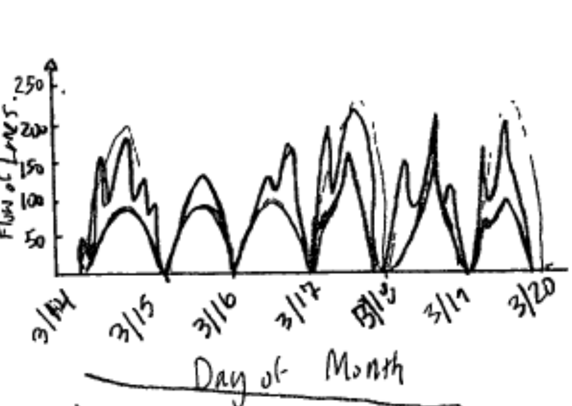
c)



2% separation as a pairwise analysis

d) The regions are distinguishable with simple tools and scale.

e) Completed.



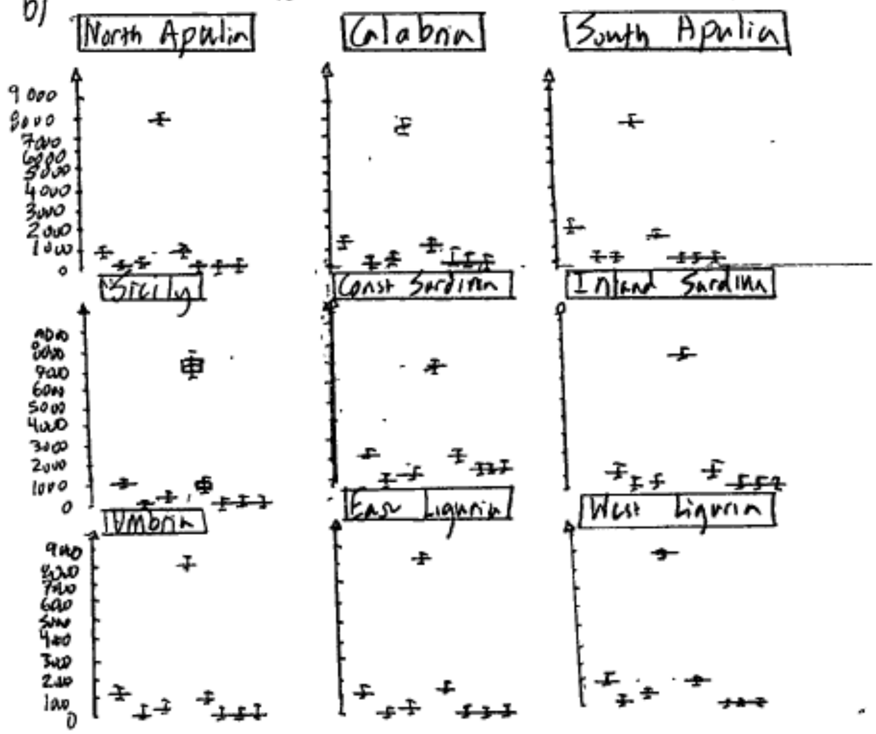
3/15 and 3/16 were weekends

49. a.

North Oil Type Calabria		Palmitic	Palmitol	Stearic	Oleic	Linoleic	Linolenic	Arachidic	Eicosenoic
North Apulia	Mean	10.3%	0.6%	2.3%	78.2%	7.1%	0.4%	0.7%	0.3%
Sicily	Median	10.5%	0.6%	2.4%	78.2%	7.0%	0.5%	0.8%	0.3%
Calabria	Mean	13.0%	1.2%	2.6%	73.1%	8.2%	0.5%	0.6%	0.2%
Coast Sardinia	Median	13.0%	1.2%	2.6%	73.0%	8.3%	0.5%	0.7%	0.3%
South Apulia	Mean	14.0%	1.8%	2.1%	69.1%	11.7%	0.3%	0.6%	0.2%
Liguria	Median	13.7%	1.8%	2.1%	69.1%	11.7%	0.3%	0.6%	0.2%
Sicily	Mean	12.3%	1.0%	2.7%	73.6%	8.3%	0.4%	0.8%	0.3%
	Median	12.2%	1.0%	2.7%	73.6%	8.3%	0.4%	0.8%	0.4%
Coast Sardinia	Mean	11.4%	1.0%	2.4%	70.9%	13.4%	0.2%	0.7%	0.0%
	Median	11.4%	1.0%	2.4%	70.9%	13.4%	0.2%	0.7%	0.0%
Inland Sardinia	Mean	11.0%	0.9%	2.2%	73.6%	11.5%	0.3%	0.7%	0.0%
	Median	11.0%	1.0%	2.2%	73.7%	11.2%	0.3%	0.7%	0.0%
Umbria	Mean	10.9%	0.6%	1.9%	79.6%	6.0%	0.3%	0.4%	0.0%
	Median	10.9%	0.6%	2.0%	79.6%	6.0%	0.4%	0.4%	0.0%
East Liguria	Mean	11.5%	0.8%	2.4%	77.5%	6.9%	0.3%	0.6%	0.0%
	Median	11.6%	0.8%	2.4%	77.4%	6.8%	0.3%	0.7%	0.0%
West Liguria	Mean	10.5%	1.1%	2.6%	76.8%	9.0%	0.0%	0.0%	0.0%
	Median	10.4%	1.0%	2.5%	77.0%	9.13	0.0%	0.0%	0.0%

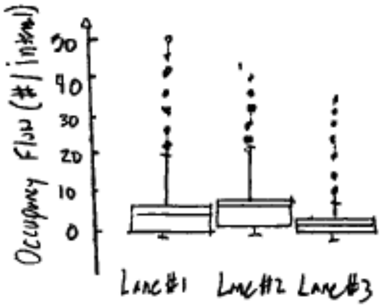
Note: Dataset inconsistent with %-values. A modification to demonstrate %-values was fraction of total.

b)



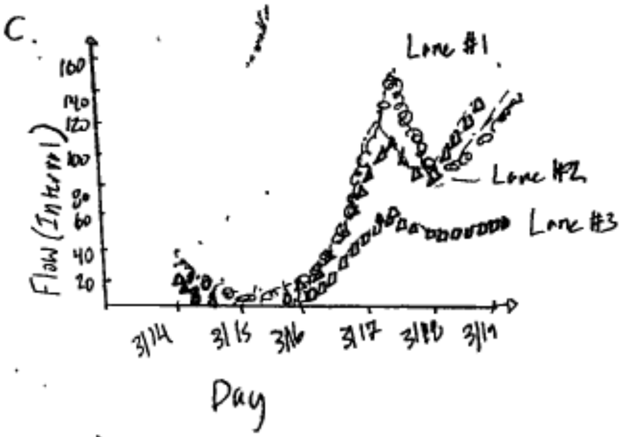
c) Oleic and eicosenoic fatty acids demonstrate a paired measurement of Italy's regions with a composite of ±2% differences.

b.



Lane #2 is busiest

c.



In terms of solely flow and not occupancy, Lane #1 is the busiest lane.

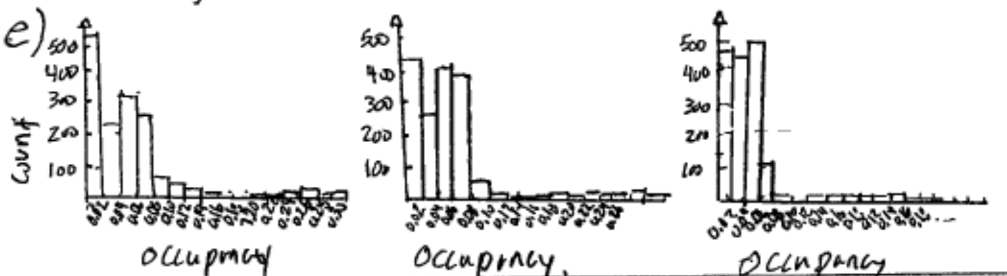
The flow of lane #2 is twice that of lane #3.

d)

	Mean	Median
Lane #1	0.061	0.048
Lane #2	0.061	0.055
Lane #3	0.051	0.041

The distributions within each lane contribute to higher average because of the proportion of higher occupancy days.

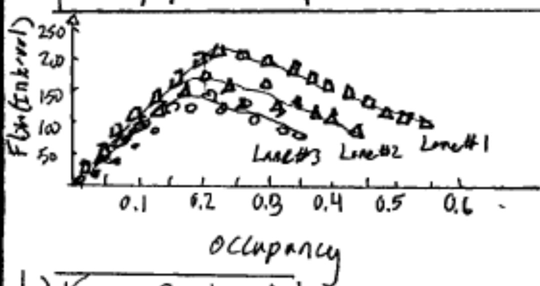
e)



The bins from 0-0.16 occupants seem to be good representation of the weekly activity. Although, a smaller distribution exists from 0.16-0.22 occupants.

f. The scatter plot of part c is evidence to argue "when one lane is busy the others are busy."

g. Throughput depends on congestion because of the scatterplot generated for occupancy vs. flow. Below 0.02 occupants flow increases proportionally, while more occupants decreases flow



h) See part A.

i) The data is tabulated with ggobi.org.

j) March 14th aided with finding the maximum of part g's plot of occupancy vs. flow.

k) In the higher dimensional scatterplot, the points generate a surface.

i) I) The points are scattered over three dimensions because occupancy vs flow vs day of the week is 3-dimensional.

II) Again, 3-dimensions because of triplet per datapoint.

III) The differences begin to occur near 3/17.

l) The right lane has lowest mean occupancy, so the taxi driver must merge right.

Chapter 11: Comparing Two Samples:

1. $X \in \{1.1650, 0.6268, 0.0751, 0.3516\}$; $Y \in \{0.3035, 2.6961, 1.0591, 2.7971, 1.2641\}$

a) $\boxed{\mu_x = 0.55 ; \mu_y = 1.62 ; \mu_x - \mu_y = -1.07}$ b) Pooled Sample Variance : $S_p^2 = \dfrac{(n-1)S_x^2 + (m-1)S_y^2}{m+n-2}$ $\boxed{= 11.71}$; $S_x^2 = (n-1)\sum_{i=1}^{n}(X_i - \bar{X})^2$ $\boxed{= 1.95}$

$S_y^2 = (m-1)\sum_{i=1}^{m}(Y_i - \bar{Y})^2$ $\boxed{= 19.03}$

c) Pooled Standard Error: $S_{\bar{X}-\bar{Y}} = S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$ $\boxed{= 2.30}$

d) Pooled Confidence Interval: $(\bar{X} - \bar{Y}) \pm t_{m+n-2}(\alpha/2)S_{\bar{X}-\bar{Y}}$

$1.07 \pm t_7(0.05) \cdot 2.30$ $\boxed{= 1.07 \pm 3.25}$

e) A two sided test "seems appropriate because of the statement "normal dist.""

f) The p-value of a two-sided test of a null hypothesis represents the probability an alternative hypothesis is accepted.

g) Yes because the model for ~~a~~ 90% confidence interval is 90% = 100% ($1-\alpha$) when $\alpha = 0.1$.

h) The argument may change by refining the confidence interval to an $\alpha < 0.1$

2. The standard error of the mean $S_{\bar{X}-\bar{Y}} = S_p\sqrt{\frac{1}{n}+\frac{1}{m}}$ will halve by doubling the sample size and reduce difference of mean error of sampling.

3. $Var(\bar{X}-\bar{Y}) ; S_p^2\left(\frac{1}{n}+\frac{1}{m}\right) ; \frac{S_x^2}{n} + \frac{S_y^2}{m} = \frac{(n-1)\sum(X-\bar{X})^2}{n} + \frac{(m-1)\sum(Y-\bar{Y})^2}{m} = S_x^2\left[\frac{1}{n}+\frac{1}{m}\right] = \boxed{S_p^2\left[\frac{1}{n}+\frac{1}{m}\right]}$

4. The t distribution is valid when sample sizes are small and standard deviation is not known, or both

5. The expected measurement of any two methods can equal each other; By testing $\mu_x = \mu_y$, a comparison of method accuracy occurs and is beneficial to scientists.

6. A test certifies foundational reasoning with others and when alone.

7. 1) $X_1, X_2, \ldots, X_n$ are independent random variables → when the normal distribution is drawn

2) $Y_1, Y_2, \ldots, Y_n$ are independent random variables → when the samples are drawn from a normal

3) X's and Y's are independent → When analyzing and making inferences about the data

8. a) Yes, because the sample size is <30 total.

b) Yes, because the sample size is <30 for each group.

9.

Concentration	$\bar{X}-\bar{Y}$	S_p	$S_{\bar{X}-\bar{Y}}$	df	t	t_{n+m-2}
10.2 mM	2.23	41.64	13.86	17	0.2612	-1.69
0.3 mM	0.89	111.72	37.24	17	0.0239	-1.69

Accept till significance level of $\alpha = 0.05$ for a one-sided distribution

Accept till significance level of $\alpha = 0.05$ for a one-sided distribution.

10. $\boxed{t = \dfrac{\mu_x - \mu_y}{S_{\bar{X}-\bar{Y}}} = 0 < t\ t_{m+n-1} \text{ rejects } 0 \text{ and } 0 > t_{n+m-2} \text{ rejects}}$

11. $H_0 : \mu_x = \mu_y + \Delta$ vs $H_A : \mu_x \neq \mu_y + \Delta$

$t = \dfrac{\bar{X}-\bar{Y}}{S_{\bar{X}-\bar{Y}}} = \dfrac{\Delta}{S_{\bar{X}-\bar{Y}}}$: Reject $H_0 \begin{cases} |t| > t_{n+m-2}(\alpha/2) \\ t > t_{n+m-2}(\alpha) \\ t < -t_{n+m-2}(\alpha) \end{cases}$

12. $P(X_{(k)} \leq \eta_0 = \eta \leq X_{(n-k+1)}) = 100\%(1-\alpha)$

if $\eta_0 = 0$; $P(\eta_0 = 0 < k-1) = \frac{1}{2^n}\sum_{j=0}^{k-1}\binom{n}{j} = 1/2$

From Section 11.3.3.

$S = \sum D_+ = 14 = Bin(n = 24, p = 0.5) = P(S \leq 14)$ $\boxed{= 0.8463}$

$P(S \geq 14) = 1 - P(S \leq 13) = 1 - Bin(n = 24, p = 0.5)$ $\boxed{= 0.2706}$

P-value = min(0.2706, 0.8463)

13. $X_1, \dots X_{25}$ i.i.d. $N(0.3,1)$ $P(\mu=0|H_0)$ vs $P(\mu>0|H_A)$ at $\alpha=0.05$

$H_1: \mu_x - \mu_y = -\mu_y = -0.3 = \Delta$; $\left|\frac{(\bar{X}=0)-\mu_X}{\sigma/\sqrt{n}}\right| > z(\alpha/2)$; $\frac{\mu_X}{1/\sqrt{25}} \geq 1.96$; $\mu_X > 0.392$; $1-\beta = P(\bar{T} \geq k) = 1-\Phi\left(\frac{0.392-0.3}{1/\sqrt{5}}\right)$; $\boxed{= 0.384}$

$T \sim Bin(25, 1/2) = \sum_k^{25} \binom{25}{k}(1/2) \sim N\left(\frac{25}{2}, \frac{25}{4}\right)$; $0.05 = P(T \geq k-1) \approx 1-\phi\left(\frac{k-0.5-12.5}{4}\right) = 1-\left(\frac{k-13}{2.5}\right)$

$k = 17.9$; $1-\beta = P(T \geq 17) = 1-\Phi\left(\frac{17-0.5-15.5}{2.4}\right) = 1-\Phi(0.41) = 0.34$

14. $X_1, \dots, X_n$ are i.i.d. $N(\mu, \sigma^2)$; $H_0: \mu = \mu_0$, the test is often $t = \frac{\bar{X}-\mu_0}{S_X}$; $df = n-1$. $\Lambda = \prod N(\mu, \sigma^2) \prod N(\mu_0, \sigma^2)$

$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1/2)(X_i-\mu_0)^2/\sigma^2} \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1/2)(Y_i-\mu_0)^2/2\sigma^2}$ $\quad X Y - X\mu_0 - \mu_0 - \mu_0^2$

$l(\mu, \mu_0, \sigma^2) = -\frac{(m+n)}{2}\log 2\pi - \frac{(m+n)}{2}\log\sigma^2 - \frac{1}{2\sigma_0^2}\left[\sum(X_i-\mu_a)^2 + \sum(Y_i-\mu_0)^2\right]$

$\frac{dl(\mu, \mu_0, \sigma^2)}{d\mu_0} = \frac{1}{(m+n)}\left(\sum X_i + \sum Y_i\right) + \sum \mu_0$; $\frac{dl(\mu, \mu_0, \sigma^2)}{d\sigma^2} = \frac{1}{(m+n)}\left[\sum(X-\mu_X) + \sum(Y-\mu_0)^2\right] - \sum \sigma_0^2$

$\hat{\mu}_0 = \frac{1}{(m+n)}\left(\sum X_i + \sum Y_i\right)$; $\hat{\sigma}_0^2 = \frac{1}{(m+n)}\left[\sum(X-\mu_X) + \sum(Y-\mu_0)^2\right]$

$l(\hat{\mu}_0, \hat{\sigma}_0^2) = -\frac{(m+n)}{2}\log 2\pi - \frac{(m+n)}{2}\log\hat{\sigma}_0^2 - \frac{(m+n)}{2}$; $l(\mu_X, \mu_Y, \sigma_1^2) = -\frac{(m+n)}{2}\log 2\pi - \frac{(m+n)}{2}\log\hat{\sigma}_1^2 - \frac{m+n}{2}$

$\Lambda = \frac{l(\mu_X, \mu_Y, \sigma^2)}{l(\hat{\mu}_0, \hat{\sigma}_0^2)} = \frac{m+n}{2}\log\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2}\right)$; $\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} = \frac{\sum(\bar{X}-\mu_0)^2 + \sum(Y-\mu_0)^2}{\sum(X_i-\bar{X})^2 + \sum(Y_j-\bar{Y})^2} = \frac{\sum(\bar{X}-\mu_0)^2}{\sum(X_i-\bar{X})^2}$; if $Y = X$, $\bar{X} = \bar{Y}$

$$\boxed{t = \Lambda = \frac{\sigma_0}{\sigma_1} = \frac{\bar{X}-\mu_0}{S_X}}$$

15. $n = m$ = treatment = control

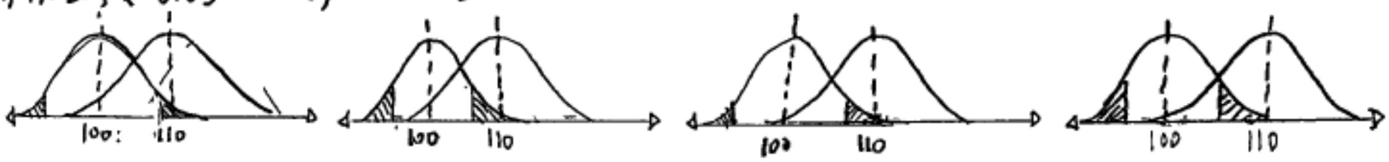
$\sigma_t = \sigma_c = 10$; n? 95% confidence Interval for $\mu_X - \mu_Y = 2$

$P\left(\left|\frac{\bar{X}-\bar{Y}}{\sigma\sqrt{2/n}}\right| > z(\alpha/2)\right) = P\left(\bar{X}-\bar{Y} > z(\alpha/2)\sqrt{\tfrac{2}{n}}\sigma\right) - P\left(\bar{X}-\bar{Y} < z(\alpha/2)\sqrt{\tfrac{2}{n}}\sigma\right) = 2$

$2 \cdot 1.96 \sqrt{\tfrac{2}{n}} \cdot 10 = 2$; $\boxed{n = 768}$

16. $H_0: \mu_X = \mu_Y$; $H_A: \mu_X > \mu_Y$; $\Delta = 0.5$ if $\mu_X - \mu_Y = 2$; $\sigma = 10$. $\quad 2 \cdot 1.64 \sqrt{\tfrac{2}{n}} \cdot 10 = 2$; $\boxed{n = 538}$

17. a) $n=20$, $\alpha=0.05$ $\quad$ b) $n=20$, $\alpha=0.10$ $\quad$ c) $n=40$, $\alpha=0.05$ $\quad$ d) $n=40$, $\alpha=0.10$



$H_0: \mu_X \cdot \mu$

18. m = subjects a) $|\mu_x - \mu_y| \pm z(\alpha/2)\sigma\sqrt{\frac{2}{m}}$; The total subject allocation is independent of confidence interval and can be allocated in random proportions.

b) $H_0: \mu_x = \mu_y$; $\Delta = \mu_x - \mu_y$ is already as powerful as possible being $H_A: \Delta = \mu_x - \mu_y = 0$. The sample proportions are independent of the argument.

19. $n = 25$; $m = 25$; Normal Distribution

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} ; \sigma = 5.$$

a) Pooled Standard Error: $S_{\bar{X}-\bar{Y}} = S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$; where $S_p = \sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{m+n-2}} = \sqrt{\frac{2 \cdot 24 \cdot 10^2}{48}} = 10$

$\boxed{S_{\bar{X}-\bar{Y}} = 2\sqrt{2}}$

b) $\alpha = 0.05$; $H_0: \mu_x = \mu_y$ vs $H_A: \mu_y > \mu_x$; $\mu_y = \mu_x + 1$; $t = \frac{\bar{X}-\bar{Y}}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{1}{10\sqrt{\frac{2}{25}}} = \frac{1}{2\sqrt{2}} = 0.35$

${}_{0.975}t_{24} = 2.064$ · is the rejection region.

c) Power of Test if $\mu_y = \mu_x + 1$

$1 - \beta = t/2 = 0.17$

d) p-value is 0.07 ; Would H_0 reject if $\alpha = 0.10$; The test would reject because for a two-sided normal distribution $\alpha/2 = 0.05$ and the test arrived to a p-value greater than $\alpha/2$.

20. Example A: 11.3.1 Bayes $= P(x|A) = \frac{P(x|B) \cdot P(B)}{P(A)}$

$$f(x|\mu,\xi^{-1}) = f(x|\theta,\xi) \cdot f(\theta) \cdot f(\xi) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-\sum(x_i-\theta)^2}{2\sigma^2}\right)}_{\text{Likelihood}} \cdot \underbrace{\exp\left(-\frac{\xi_{prior}}{2}(\theta-\theta_0)^2\right)}_{\text{Priors}} \cdot \underbrace{\xi^{\alpha-1}\exp(-\lambda\xi)}_{\text{gamma prior}}$$

$\boxed{= 9.9,996 \text{ Positive}}$

21. a) $t = \frac{\bar{X}-\bar{Y}}{S_{\bar{X}-\bar{Y}}} = \frac{10.693 - 6.75}{1.88} = 2.09$;

1.05 133

If p-value = 0.05, $\alpha = 0.05$, for a one-sided distribution. $df = 9$; ${}_{9}t_{0.95} = 1.833$. The null hypothesis of $H_0: \mu_x = \mu_y$ is rejected.

b) Mann-Whitney Test [Nonparametric]

Type I	Rank	Type II	Rank
3.03	1	3.14	2
5.53	8	4.26	3
5.60	9	4.47	4
9.30	11	4.53	5
9.92	13	4.67	6
12.51	14	4.69	7
12.95	17	12.78	16
15.21	18	6.79	10
16.04	19	9.37	12
16.84	20	12.75	15
R	130	R	80
R'	80	R'	130
$R^*(0.05)$	78	$R^*(0.05)$	78

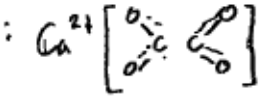
The null hypothesis (H_0) is rejected at a significance level of $\alpha = 0.05$, because $R' > R^*$.

c) Either the t-test or Mann-Whitney Test is applicable to determining the null hypothesis point of rejection. The key qualifiers for each test are $(m,n) < 30$, and are chosen by the case of extreme outliers which are not representative of this data set.

d) π is the probability that a component of one type will last longer than the component of another type (effect), or the probability that an observation from one distribution is smaller than the independent observation from another distribution.

$$\pi = \frac{1}{mn}\sum_{i}^{10}\sum_{j}^{10} Z_{ij} \text{ ; where } Z_{ij} = \begin{cases} 1 & \text{if } X_i < Y_j \\ 0 & \text{otherwise} \end{cases}$$

$= \frac{1}{100} \cdot 25 = \boxed{\frac{1}{4}}$ e) $\hat{\pi} = \frac{1}{m \cdot n}\sum_{i}^{10}\sum_{j}^{10} Z_{ij} = \boxed{\frac{1}{4}}$ f. CI $= \left\{\frac{9}{40}, \frac{5}{19}\right\}$

22. Calcium oxalate $[CaC_2O_4]$: Ca^{2+} 

Parametric Test:

oxalate	flame	
$\mu = 2.39$	$\mu = 2.35$	$s_p = 2.05$
$\sigma = 1.45$	$\sigma = 1.42$	$S_{\bar{x}-\bar{y}} = 0.19$
		$t = 0.43$
		$t_{117,0.995} = 3.38$

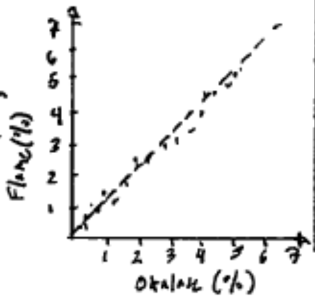
Nonparametric Test:

oxalate:	flame:
$R = 14146.5$	$R = 13819.5$
$R' = 6790.5$	$R' = 7125.5$

Z-statistic = 0.31

p-value = 0.62

Graphical:



The parametric test t-statistic advises to accept the null hypothesis, Nonparametric advises to not reject the null hypothesis and the graphs do look similar.

23. $X_1, \dots X_n$ i.i.d. with cdf F ; $Y_1, \dots Y_m$ i.i.d. with cdf G. $H_0: F = G$ vs $H_A: F \neq G$. $(m+n)\%2 = 0$

a) Hypergeometric Distribution:

$$p(X=k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

k = amount of total without replacement.

r = Total of type

n-r = Total of second type

m = total chosen

$$P(T=t) = \frac{\binom{(m+n)/2}{t}\binom{(m+n)/2}{n-t}}{\binom{m+n}{n}} \; ;$$

r = (m+n)/2 or total of values below median

n-r = (m+n)/2 or total of values greater than median.

$(m+n)/2 < n$

$(m+n)/2 > n$

k = amount of X's chosen without replacement below the median.

n = total X's chosen.

b. CI for $G = P(X_{(i)} < \eta < X_{(j)})$ where η = median

The hypergeometric distribution is approximated by a binomial distribution.

A rejection region would be discovered by $P(T<t) = 1 - \dots$

$$P(T=t) = \frac{\binom{(m+n)/2}{t}\binom{(m+n)/2}{n-t}}{\binom{m+n}{n}} = \binom{n}{t}\prod_{k=1}^{t}\frac{(\frac{m+n}{2} - t + k)}{(m+n-x+k)} \cdot \prod_{j=1}^{n-t}\frac{(m+n-r-(j-k)-n)}{(m+n+j)}$$

c. $P(T<t) = 1 - P(T>t)$

$$= 1 - \sum_{i=1}^{T}\frac{\binom{(m+n)/2}{t}\binom{(m+n)/2}{n-t}}{\binom{m+n}{n}}$$

$$= 1 - \sum_{i=1}^{5}\frac{\binom{10}{t}\binom{10}{10-t}}{\binom{20}{10}} = \boxed{32.8\%}$$

$$\lim_{N \to \infty} P(T=t) = \binom{n}{t}p(1-p) \; ; \; \lim_{N\to\infty} P(X_i < \eta < X_j) = \sum_{k=i}^{j}\binom{n}{k}p^k(1-p)^{n-k}$$

$$= Bin(j-1, n, p) - Bin(i-1, n, p)$$

With $G(x) = F(x - \Delta)$, then $\boxed{= Bin(j-1-\Delta, n, p) - Bin(i-1-\Delta, n, p)}$

24. Mann-Whitney Statistic

$R' = n(m+n+1) - R$

$E(U) = \frac{mn}{2}$

$Var(U) = \frac{mn(m+n+1)}{12}$

$m=3, n=2$

$$\frac{U - E(U_y)}{\sqrt{Var(U_y)}} = N(0,1) \; ; \; U = N(0,1)\sqrt{Var(U_y)} + E(U_y) = \boxed{\left(\frac{e^{-x^2/2}}{\sqrt{2\pi}}\dots + 1\right)\dots 3}$$

25. a) If 79.94 is arbitrarily small, say 0.001, then the new mean is 69.98J and standard deviation 26.45J. The pooled variance become 9.75J, with a standard error of 1.40J. The final t-statistic was evaluated to 7.16. At $t_{19}(0.005) = 2.861 < 7.16$ and would reject with a significance of $\alpha = 0.01$

25. If method B (80.03 J) became 10,000 J, $\mu = 1319.93$, $\sigma = 3230.74$, $S_p^2 = 1208.691$, $S_{\bar{x}-\bar{y}} = 15.62$

C.

Method A	Method B
79.98 (7.5)	80.03 (15)
80.04 (19)	79.94 (1)
80.02 (11)	79.98 (9.5)
80.04 (19)	79.97 (4.5)
80.03 (15)	79.97 (4.5)
80.03 (15)	80.03 (15)
80.04 (19)	79.95 (2)
79.97 (4.5)	79.97 (4.5)
80.05 (21)	
80.03 (15)	
80.02 (11)	
80 (9)	
80.02 (11)	

R 177 54

R' 232

Z-Statistic: 13.04

$t_{20}(0.005) = 2.85$

Null hypothesis would reject

Modifying 79.94 to an arbitrarily low value had no effect on the data because 79.94 is the lowest rank. While raising 80.03 to a larger value of 10,000 changed the Z-statistic to 12.64 with a null hypothesis also rejecting.

t-statistic = -79.37
The null hypothesis would reject

26. $X_1, \ldots X_n$ be from $N(0,1)$: $Y_1 \ldots Y_n$ be from $N(1,1)$.

a) $T_X = U_X + \frac{n(n+1)}{2}$; $U_X = \sum\sum Z_{ij}$; $Z_{ij} = \begin{cases} 1 & X_i < Y_j \\ 0 & \text{otherwise.} \end{cases}$

$$E[T_X] = E[U_X] + E\left[\frac{n(n+1)}{2}\right] = E\left[\sum\sum Z_{ij}\right] + \frac{n(n+1)}{2} = \sum\sum E[Z_{ij}] + \frac{n(n+1)}{2}$$

$$= \sum\sum P(X_i < Y_i) + \frac{n(n+1)}{2} = n^2 P(X_i - Y_i > 0) + \frac{n(n+1)}{2}$$

$$= n^2 P\left(\frac{X_1 - X_2 - E(X_1 - X_2)}{\sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2}} > \frac{0 - E(X_1 - X_2)}{\sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2}}\right) + \frac{n(n+1)}{2}$$

$$= n^2 P\left(Z > \frac{1}{\sqrt{2}}\right) + \frac{n(n+1)}{2} = n^2\left[1 - \Phi\left(\frac{1}{\sqrt{2}}\right)\right] + \frac{n(n+1)}{2} = n^2(1 - 0.7611) + \frac{n(n+1)}{2}$$

$$\boxed{= 0.2389n^2 + \frac{n(n+1)}{2}}$$

b. $Var(T_X) = Var\left(U_X + \frac{n(n+1)}{2}\right) = Var(U_X) = \sum\sum Var(Z_{ij}) + \sum Cov(Z_{ij}, Z_{kl})$

$$= n^2 p(1-p) + \sum\left(E(Z_{ij} Z_{kl}) - E[Z_{ij}]E[Z_{kl}]\right) = n^2 p(1-p) + \sum\left[E(Z_{ij} Z_{kl})^{0} - p^2\right] = \boxed{n^2 p(1-p)}$$

uncorrelated

27. Exact Null Distribution of W where $n = 4$

$W_+ = \sum_{k=1}^{n} k I_k$ where $I_k = \begin{cases} 1 & \text{if the } k^{th} \text{ largest } |D_i| \text{ has } D_i > 0 \\ 0 & \text{otherwise} \end{cases}$

$W_+ = \sum_{k=1}^{4} k I_k$ ∴ Total values: $\frac{n(n+1)}{2} = 10$;

"Above Diagonal"

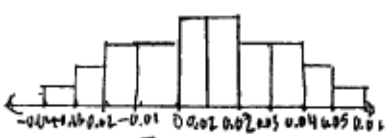
W	0	1	2	3	4	5	6	7	8	9	10
P(x)	1/16	1/16	1/16	2/16	2/16	2/16	2/16	2/16	1/16	1/16	1/16

	1	1	1	-1	1	-1	-1	-1	1	1	1	-1	-1	1	-1	-1
	2	2	-2	2	2	-2	2	2	-2	-2	2	-2	2	-2	-2	-2
	3	3	3	-3	3	3	-3	3	-3	3	-3	3	-3	-3	-3	-3
	4	4	4	4	-4	4	4	-4	4	-4	4	-4	-4	-4	4	-4
Total (W_+)	10	9	8	7	6	7	6	5	5	4	3	3	2	1	4	0

28. $n = 10, 20, 30$ find $\alpha = 0.05$ and 0.01 for a two sided rank test

n	10	20	30
0.05	8	52	137
0.01	3	37	109
$E(W_+)$	27.5	105	233
$Var(W_+)$	96.25	718	2364
$Z(0.05)$	-1.98	-1.97	-1.97
$Z(0.01)$	-2.49	-2.54	-2.55

29. a)



b) The process of randomizing two samples, then comparing the difference is similar to the Mann-Whitney test, but more commonly the Wilcoxon signed rank test. Each of these methods compares to a randomized distribution.

30. $X_{\bar{A}} - X_{\bar{B}}$; Standard Error : $\sigma_{\bar{Y}} = \sqrt{Var(U_y)} / nm = \boxed{\sqrt{\frac{(m+n+1)}{12}}}$

31. Section 11.2.3: A Nonparametric Method: The Mann-Whitney Test ; $F=G$; $E(\hat{\pi}) = E\left[\frac{1}{mn}\sum_{i=1}^{\hat{n}}\sum_{j=1}^{\hat{m}} Z_{ij}\right]$

$= \frac{n \cdot m}{n \cdot m \cdot 2} = \boxed{\frac{1}{2}}$ $\frac{r(n)}{2}$

32. $X \sim N(\mu_x, \sigma_x^2)$; $Y \sim N(\mu_y, \sigma_Y^2)$; $\pi = P(X<Y)$

$\hat{\pi} = P(X-Y) = \text{sign}(|N(x_i|\mu_x,\sigma_x^2) - N(y_i|\mu_{y},\sigma_y^2)|) = n(n \cdot$

$\boxed{= \text{sign}\left(\frac{1}{2\pi\sigma_x^2\sigma_i^2} e^{-(x_i-\mu_x)^2/2\sigma_x^2 - (y_i-\mu_y)^2/2\sigma_y^2}\right)}$

$Var(\hat{\pi}) = \boxed{\frac{(m+n+1)}{12mn}}$ $4^2 \cdot 12$

$n \neq m$ or $n=m$ does not influence outcome of $E(\hat{\pi})$ or $Var(\hat{\pi})$.

33. $N(\mu_x, \sigma_x^2)$; $N(\mu_y, \sigma_y^2)$; $H_0: \sigma_x = \sigma_y$; $\frac{S_x^2}{S_y^2} \sim F_{n-1,m-1}$

$F_{n-1,m-1} : \frac{X}{Y} = \frac{N(\mu_x,\sigma_x^2)}{N(\mu_y,\sigma_y^2)} = \frac{\sigma_x}{\sigma_y}$ a) $H_0: \sigma_x = \sigma_y$; $H_1: \sigma_x \neq \sigma_y$

One-sided: $\boxed{\frac{\sigma_x^2}{\sigma_y^2} > F_{n-1,m-1}(\alpha)}$

Two-sided: $\boxed{F_{n-1,m-1}(\alpha/2) < \frac{\sigma_x^2}{\sigma_y^2} < F_{n-1,m-1}(1-\frac{\alpha}{2})}$

b) CI : $\left(\frac{S_x^2}{S_y^2} F_{n-1,m-1}(\alpha/2) < \sigma_x^2/\sigma_y^2 < \frac{S_x^2}{S_y^2} F_{n-1,m-1}(1-\frac{\alpha}{2})\right)$

$n\left(-\frac{n}{2}+\frac{1}{2}\right)$

c) onesided: Two-sided: CI

$0.59 > \frac{1}{F_{8,7}(0.05)}$ $\frac{\sigma_x^2}{\sigma_y^2}(F_{12,7}(0.025))^{-1} < 0.59 < F_{12,7}(0.975)\frac{\sigma_x^2}{\sigma_y^2}$

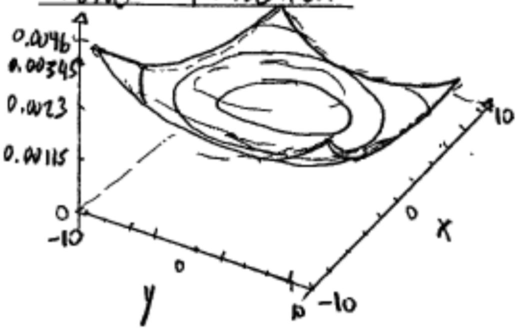
$0.59 > 0.5219$ $\boxed{0.4264 < 0.59 < 22.1237}$

$n\left(\frac{n}{2}-\frac{1}{2}\right)$

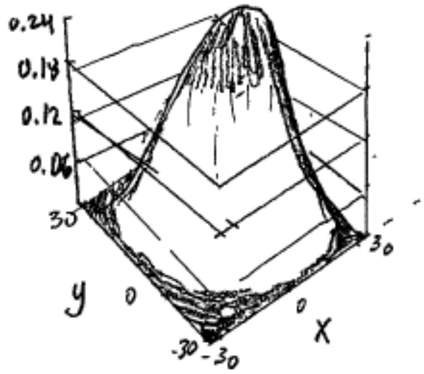
34. $H_0: \mu_x = \mu_y$ a) Paired: $Cov(X_i, Y_i) = 50$, $\sigma_x = \sigma_y = 10$, $i = 1 \ldots 25$ if $H_1: \mu_x \neq \mu_y$; $\rho = \frac{cov(X,Y)}{\sqrt{\sigma_x^2\sigma_y^2}} = \frac{1}{2}$ $\frac{n(n-1)}{2}$

$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 10^2 & 50 \\ 50 & 10^2 \end{bmatrix}$; $\mu = \begin{pmatrix}\mu_x \\ \mu_y\end{pmatrix}$; $N(\mu, \Sigma)$

Power Function:

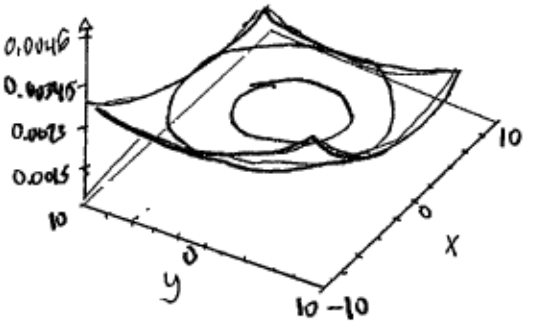


$N(\mu, \Sigma)$



$H_0: \mu_x = \mu_y$

b) Power Function



$N(\mu_x - \mu_y, \frac{\sigma_x^2}{n^2} + \frac{\sigma_y^2}{n^2}) = N(\mu_x - \mu_y, 8)$

$\boxed{\text{Power function is the only requirement.}}$

35. $n = 22$; $t = 70$ days, exposure time $= 7$ days to ozone

$n = 23$; $t = 70$ days, exposure time $= 0$ days to ozone.

$H_0: \mu_{control} = \mu_{ozone}$

$H_1: \mu_{control} \neq \mu_{ozone}$

	Mean	Median	SD
Control(23)	22.43g	22.7g	10.54g
Ozone(22)	11.01g	11.1g	18.60g

Parametric

$S_p = 225.43g$

$S_{\bar{x}-\bar{y}} = 67.22g$

t-statistic : 2.55

$t_{43}(0.05) : 1.68$

Nonparametric

$R(\text{control}) = 661$; $R(\text{Ozone}) = 374$

$R'(\text{Ozone}) = 638$

Z-statistic : 8.74

$Z(0.05) : 1.96$

For both parametric and nonparametric analysis, the null hypothesis (H_0) is rejected. The parametric pooled study shows a t-statistic greater than a significance of $\alpha = 0.05$. While the nonparametric test, Mann-Whitney, demonstrates a Z-statistic above the rejection level of 1.96, purporting a rejection of the null hypothesis.

36.

	E[X]	SD[X]	$\bar{X}-\bar{Y}$	S_p	$S_{\bar{x}-\bar{y}}$
Microbiological Method	85.26%	20.48%	0.44%	20.65%	7.54%
Hydroxylamine Method	84.82%	20.82%			

$H_0: \bar{X} = \bar{Y}$; $H_1: \bar{X} \neq \bar{Y}$

Outcomes would be similar by randomizing the sample sets.

t-statistic : 0.06 ; $t_{30}(0.05) : 2.04$.

The null hypothesis is not rejected.

37 a)

	Ward A ΔDose	Ward A ΔPlacebo	Ward B ΔDose	Ward B ΔPlacebo
	0.8	-0.4	-0.45	0.01
	0.1	0.4	0.15	0.44
	0.55	-0.1	-0.19	-0.55
	0.6	-0.9	0.12	-0.05
	0.34	0.2	0.03	0.04
	1.42	0.78	-0.15	-1.16
	1.74	0.3	0.48	-0.16
	-0.29	0.64		
	0.53	0.42		
E[x]	0.64	0.15	-0.001	-0.25
SD[x]	0.59	0.51	0.28	0.43
Sp	0.55		0.36	
$S_{\bar{x}-\bar{y}}$	0.26		0.19	
Z	1.92		1.28	

$H_{0A}: \mu_{\Delta Dose} = \mu_{\Delta Placebo}$

$H_{1A}: \mu_{\Delta Dose} \neq \mu_{\Delta Placebo}$

H_{0A} is below the significance level of $\alpha = 0.05$, and is not rejected.

$H_{0B}: \mu_{\Delta Dose} = \mu_{\Delta Placebo}$

$H_{1B}: \mu_{\Delta Dose} \neq \mu_{\Delta Placebo}$

H_{0B} is below the threshold of a $Z(0.05) = 1.96$ and accepted as a hypothesis.

b) Pooled Variance of ΔDose of Ward A and Ward B:

$$S_{Pd}^2 = \frac{(10-1)\,0.59^2 + (7-1)\,0.28^2}{(10+7-2)} = \boxed{0.24}$$

Pooled variance of ΔPlacebo of Ward A and Ward B:

$$S_{Pp}^2 = \frac{(10-1)\cdot 0.51^2 + (7-1)\,0.43^2}{(10+7-2)} = \boxed{0.23}$$

Standard Error of ΔDose of Ward A and Ward B

$$S_{\bar{A}-\bar{B}} = S_p\sqrt{\left(\frac{1}{10} + \frac{1}{7}\right)} = \boxed{0.24}$$

Standard Error of ΔPlacebo of Ward A and Ward B.

$$S_{\bar{A}-\bar{B}} = S_p\sqrt{\left(\frac{1}{10} + \frac{1}{7}\right)} = \boxed{0.24}$$

Z-statistic of ΔDose of Ward A and Ward B.

$$= \frac{0.64 + 0.001}{0.24} = \boxed{2.67}$$

Z-statistic of ΔPlacebo of Ward A and Ward B.

$$H_0: \overline{\Delta Placebo}(A) = \overline{\Delta Placebo}(B) = \frac{0.15 + 0.25}{0.24} = \boxed{1.67}$$

$H_1: \overline{\Delta Placebo}(A) \neq \overline{\Delta Placebo}(B)$

$\boxed{H_0 \text{ rejected} : Z(0.05) < Z = 1.67}$

Part B:

$H_0: \overline{\Delta Dose}(\text{Ward A}) = \overline{\Delta Dose}(\text{Ward B})$ Rejected

$H_1: \overline{\Delta Dose}(\text{Ward A}) \neq \overline{\Delta Dose}(\text{Ward B})$ Accepted

$\boxed{Z(0.05) < Z = 2.673}$

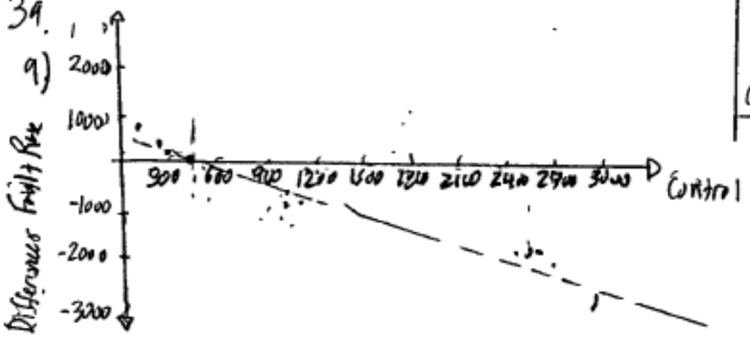
38. Δ = Added(%) - Found(%)

	$E[x]$	$SD[x]$	S_p	$S_{\bar{x}-\bar{y}}$	t
Δ Sulfonic Acid	0.007	0.015	0.014	0.006	0.19
Δ Pyrazolone-T	0.006	0.013			

$H_0: \overline{\Delta \text{Sulfonic Acid}} = \overline{\Delta \text{Pyrazolone-T}}$
$H_1: \overline{\Delta \text{Sulfonic Acid}} \neq \overline{\Delta \text{Pyrazolone-T}}$ Measurements on HPLC.

The null hypothesis (H_0) is not rejected because the t-statistic = 0.19 is below $t_{29}(0.005) = 2.845$. Bailey, Cox, and Spangers HPLC measurements were consistent for two sets of data.

39. a)



b)c)

	$E[x]$	$SD[x]$	Median	S_p	$S_{\bar{x}-\bar{y}}$
Test	434.21	161.87	442.50	550.60	208.106
Control	895.50	761.78	616.00		

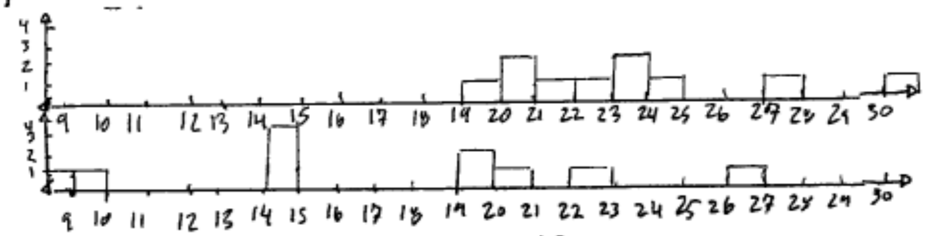
$\bar{X}-\bar{Y} = 523.3$ $\Delta\eta = ~~-1988.5~~$

$CI = \boxed{\{115, 931\}}$ $CI = \{nq - z(0.05)\sqrt{nq(1-q)},\ nq + z(0.05)\sqrt{nq(1-q)}\}$
$= \{R(8.46), R(12.54)\}$
$\boxed{= \{-587, -163\}}$

d) A t test seems more appropriate because of the sample size, but skew of the control demonstrates for one of the datasets suggests examination with a ranked test.

40. n = 10 cages, t = 30-day old mice Treated ; Treat 12 days 80 Oe/cm.
1 cage = 3 rats. Control : 30 mice.

a)



Field Absent $\mu = 25g$, $\sigma = 5.4g$
Field Present $\mu = 17g$ $\sigma = 3.0g$

b)

	$E[x]$	$SD[x]$	S_p	$S_{\bar{x}-\bar{y}}$	Lower CI	Upper CI
Field Absent	25g	5.4g	4.4g	1.9g	4.3g	11.7g
Field Present	17g	3.0g				

c) $\boxed{t_{18}\text{-score} = 3.94}$ H_0: Mean Field Absent = Mean Field Present
p-value = 8×10^{-5} H_1: Mean Field Absent ≠ Mean Field Present

H_0 is rejected at a significance of $\alpha = 0.05$ and alternative hypothesis accepted.

d)

	R	R'	U
Field Absent	143		
Field Present	67	153	6.07

p-value = 0.0 ; null hypothesis is rejected.

e)

	Median	ΔMedian
Field Absent	25g	5g
Field Present	17g	

f) SE of Mean = 1.9g : Bootstrap. $\boxed{\sigma_{Median} = 1.253\sigma_{\bar{x}} = 2.4g}$

g) $\boxed{CI = \{R(1.9), R(8.1)\} = \{5, 11.8\}}$

41. a) $E(\hat{\Delta}) = E(\text{median}(X_i - Y_j)) = \boxed{\mu_x - \mu_y}$ b) $\boxed{\hat{\Delta} \text{ is robust to outliers because the method is rank based.}}$

c) $\boxed{\sigma_{Median}^{4} < \hat{\Delta} = 4g < \Delta \text{Median}}$ d) $\hat{\Delta} = 1.253\sigma_{\bar{x}}$; $\sigma_{\bar{x}} = 3.19g$. $\boxed{f\left(x \mid \mu, \frac{1}{8f(\tilde{\mu})^2 \cdot 10}\right)}$

e. $\boxed{CI = \{-2.3g, 10.3g\}}$

42. a) $\pi = \frac{1}{mn}\sum\sum Z_{ij}$; Where $Z_{ij} = \begin{cases} 1 & X_i > Y_i \\ 0 & \text{Otherwise} \end{cases}$; $\pi = \frac{229}{26 \cdot 9} \doteq \boxed{0.3388}$

b) SE=0 For π because matrix was generated with exact values. Alternative methods include generating a sampling distribution for $S_x = \sqrt{\frac{1}{B}\sum(\bar{\pi}_i - \bar{\pi})^2}$

c) The confidence interval = {0.3388}, but with a sampling distribution $CI = \{\pi - 1.96 S_x, \pi + 1.96 S_x\}$

43. $X_1 \ldots X_n, Y_1 \ldots Y_m$, $\mu_{20\%}$; A bootstrap could be used to estimate the standard error of a 20% trimmed mean vs. 0% trimmed mean by producing a sampling distribution x [illegible] $f(\bar{X}-\bar{Y} \mid \bar{X}-\bar{Y}, S_p^2)$, [illegible] with those parameters.

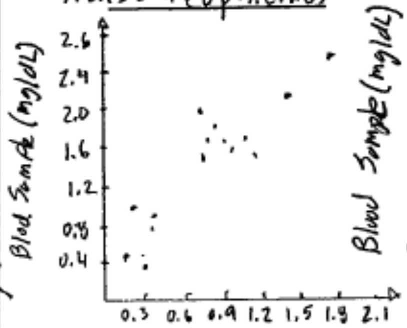
44. n=20 ; m=15 ; 2 months. Volume=2mL of blood at breakfast and urination.

Post-urination: 1g ascorbic acid.

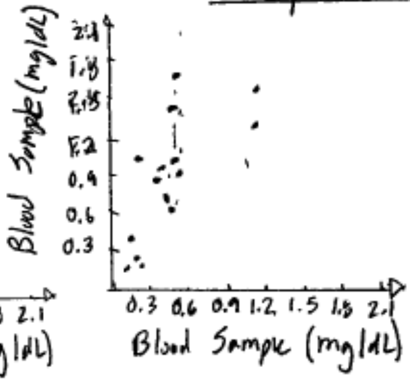
a)

t_1 = 6 hours urine collection ; t_2 = 2 hours after dose of Vitamin C.

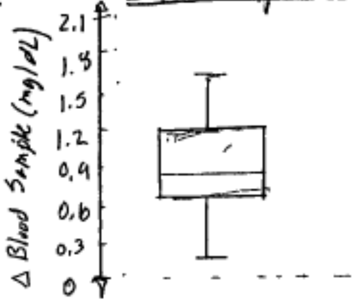
Nonschizophrenics:



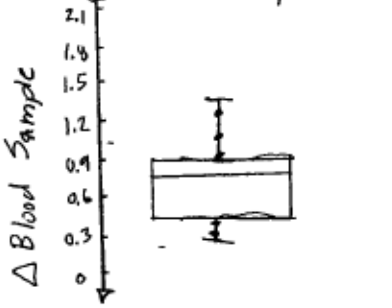
Schizophrenics:



Δ Non Schizophrenic



Δ Schizophrenic



b)

	E[X]	SD[X]	S_p	$S_{\bar{x}-\bar{y}}$	t	df	t
Nonschizo (0hr)	0.62	0.42	0.34	0.12	2.45	15	1.69
Schizo (0hr)	0.33	0.26				20	
Nonschizo (2hr)	1.45	0.59	0.58	0.17	2.92	15	1.69
Schizo (2hr)	0.95	0.42				20	
Δ Nonschizophren	0.61	0.33	0.33	0.17	1.99	15	1.69
Δ Schizophrenic	0.33	0.34				20	

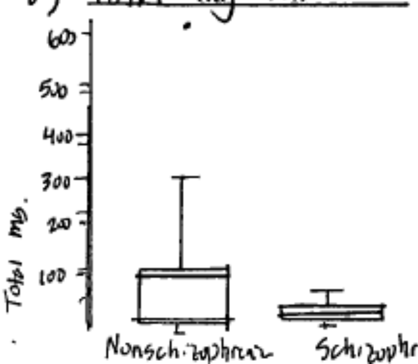
H_0: $\overline{\text{Nonschizophrenic (0hr)}} = \overline{\text{Schizophrenic (0hr)}}$ — Reject at α=0.05
H_1: $\overline{\text{Nonschizophrenic (0hr)}} \neq \overline{\text{Schizophrenic (0hr)}}$ — Accept at α=0.05
H_0: $\overline{\text{Nonschizophrenic (2hr)}} = \overline{\text{Schizophrenic (2hr)}}$ — Reject at α=0.05
H_1: $\overline{\text{Nonschizophrenic (2hr)}} \neq \overline{\text{Schizophrenic (2hr)}}$ — Accept at α=0.05
H_0: $\overline{\Delta\text{Nonschizophrenic}} = \overline{\text{Schizophrenic (2hr)}}$ — Reject at α=0.05
H_1: $\overline{\Delta\text{Nonschizophrenic}} \neq \overline{\Delta\text{Schizophrenic}}$ — Accept at α=0.05

c)

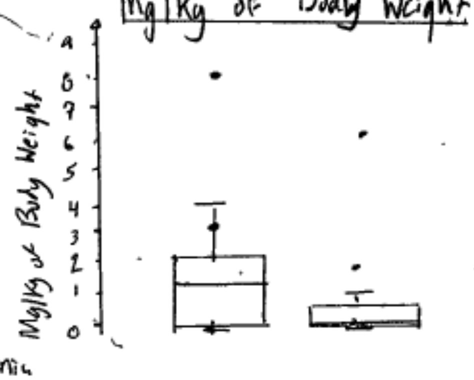
	R	R'	U
Nonschizo (0hr)	324		
Schizo (0hr)	306	234	218
Nonschizo (2hr)	345		
Schizo (0hr)	285	255	315
ΔNonschizophrenic	319		
ΔSchizophrenic	311	229	2.06

H_0: $\overline{\text{Nonschizophrenic (0hr)}} = \overline{\text{Schizophrenic (0hr)}}$ — Reject at α=0.05
H_1: $\overline{\text{Nonschizophrenic (0hr)}} \neq \overline{\text{Schizophrenic (0hr)}}$ — Accept at α=0.05
H_0: $\overline{\text{Nonschizophrenic (2hr)}} = \overline{\text{Schizophrenic (2hr)}}$ — Reject at α=0.05
H_1: $\overline{\text{Nonschizophrenic (2hr)}} \neq \overline{\text{Schizophrenic (2hr)}}$ — Accept at α=0.05
H_0: $\overline{\text{Nonschizophrenic}} = \overline{\text{Schizophrenic}}$ — Reject at α=0.05
H_1: $\overline{\text{Nonschizophrenic}} \neq \overline{\text{Schizophrenic}}$ — Accept at α=0.05

d) Total mg Vitamin C



Mg/Kg of Body Weight



	E[X]	SD[X]	S_p	$S_{\bar{x}-\bar{y}}$	df	t
Nonschizophrenic (Total mg)	122.57	153.7	116.9	39.9	33	-2.2
Schizophrenic (Total mg)	35.8	94.6				
Nonschizophrenic (mg/Kg)	1.78	2.0	1.6	0.5	33	-1.2
Schizophrenic (mg/Kg)	0.53	1.3				

The data shows a hypothesis of mean weights are not equivalent.

e) Assuming Normality is standard for Z-statistics, but the mean and median show normal distribution may not best fit the data.

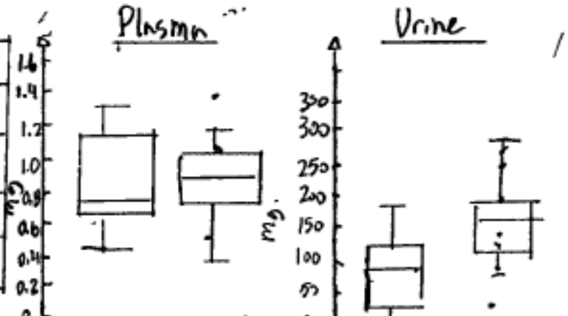
f)

	R	R'	U
Nonschizophrenic (Total)	344		
Schizophrenic (Total)	286	254	3.47
Nonschizophrenic (mg/kg)	312		
Schizophrenic (mg/kg)	238	207	2.34

If a hypothesis (H_{0A}: Nonschizophrenic (Total mg) = Schizophrenic (Total mg)) or (H_{0B}: Nonschizophrenic (mg/Kg) = Schizophrenic (mg/Kg)) were proposed, then they would be rejected at $\alpha = 0.05$.

g)

	E[X]	SD[X]	s_p	$s_{\bar{x}-\bar{y}}$	t	df	t(0.05)
Schizophrenic (Plasma)	0.39	0.26	0.24	0.09	0.14	28	2.04
Control (Plasma)	0.98	0.22					
Schizophrenic (Urine)	80.86	52.29	61.09	22.31	-3.56	28	2.04
Control (Urine)	160.31	65.54					



The hypothesis (H_0: Schizophrenic (Plasma) = Control (Plasma)) is accepted with the alternative (H_1: Schizophrenic (Plasma) ≠ Control (Plasma)) because the t-statistic for 28 degrees of freedom is less than a standard curve at the significance level. While the urine samples show an argument or rejection of hypothesis (H_0: Schizophrenic (Urine) = Control (Urine)).

h) The normality is reasonable because the mean is within 10% of the median for each set of data.

i)

	R	R'	U
Schizo (Plasma)	161.5	303.5	7.42
Control (Plasma)	233.2		
Schizo (Urine)	96.0	369	10.63.
Control (Urine)	365.1		

Unlike part G), the Mann-Whitney shows reason to reject the hypothesis that the means are equivalent, which is argued against by the table of part G. Urine samples tested with Mann-Whitney list a similar outcome to part g.

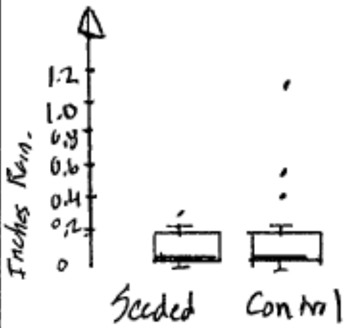
45. a)

Year	Experiment	E[X]	SD[X]	s_p	$s_{\bar{x}-\bar{y}}$	t	df	t(0.05)
1957	Seeded	0.07	0.07	0.08	0.03	0.31	30	2.04
	Unseeded	0.06	0.08					
1958	Seeded	0.06	0.08	0.09	0.03	0.55	30	2.04
	Unseeded	0.04	0.11					
1959	Seeded	0.02	0.04	0.16	0.06	-1.28	38	2.02
	Unseeded	0.09	0.21					
1960	Seeded	0.02	0.04	0.05	0.02	-1.13	30	2.04
	Unseeded	0.03	0.05					

Years	Experiment	E[X]	SD[X]	s_p	$s_{\bar{x}-\bar{y}}$	t	df	t(0.05)
57-'60	Seeded	0.04	0.06	0.11	0.02	-0.95	134	1.98
	Unseeded	0.06	0.14					

Tabulated data of each experiment individually demonstrate seeding did not influence the rainfall mid-day ($H_0: \mu_x = \mu_y$). The pooled years of seeding was tested to $t < t(0.05)$ and had an outcome of accepting the null hypothesis.

Pooled Years of Cloud Seeding Experiment:



b) The day on which seeding should be done should be chosen at random because daily parameters cycle throughout the month. Days are paired in the experiment because of similar conditions.

46.

Type	Experiment	E[x]	SD[x]	s_p	$s_{\bar{x}-\bar{y}}$	t	df	t(0.05)
I	Seeded	0.14	0.08	0.09	0.03	-0.38	33	2.03
I	Control	0.12	0.10					
II	Seeded	0.13	0.10	0.01	0.03	-0.96	33	2.03
II	Control	0.10	0.10					

Hypothesis:

$H_{0,\text{Type I}}$: Seeded Mean (Type I) = Control Mean (Type I)

$H_{1,\text{Type I}}$: Seeded Mean (Type I) ≠ Control Mean (Type I)

$H_{0,\text{Type II}}$: Seeded Mean (Type II) = Control Mean (Type II)

$H_{0,\text{Type II}}$: Seeded Mean (Type II) ≠ Control Mean (Type II)

The analysis accepts the null hypothesis for each type (I/II) at a significance level of $\alpha = 0.05$. TACN-cloud seeding project had no effect on outcomes of rain, and including analysis of cloud formations rather than years.

47. a)

Experiment	Variable	E[x]	SD[x]	s_p	$s_{\bar{x}-\bar{y}}$	t	df	t(0.05)
Seeded	Target	11.72	12.11	11.24	3.06	0.48	50	2.01
	Control	10.24	10.29					
Unseeded	Target	13.46	17.12	14.13	3.72	0.96	54	2.00
	Control	9.89	10.44					

$t < t(0.05)$ for each Experiment suggesting the null hypothesis cannot be rejected with the supplied information.

b) The square root transformation should play no effect on the data set or analysis because of the 1:1 relationship of input to output.

c) A control area provides a control variable to test against. Comparing seeded to unseeded limits experiment variability and requires a control variable to reference.

48. Parametric

Experiment	R	R'	U
Before	165		
After	106	134	3.58

U > Z-statistic (0.05, Two-tailed)
3.58 > 1.96 ; Reject $H_0: \mu_x = \mu_y$

Nonparametric:

Experiment	Mean (9/24)	S.d. (9/24)	s_p (9/24)	$s_{\bar{x}-\bar{y}}$ (9/24)	t (9/24)	df	t
Before	9.27	4.04	4.26	1.74	2.69	22	2.07
After	4.58	2.14					

$t > t_{22}(0.05)$; Reject $H_0: \mu_x = \mu_y$
2.69 > 2.07

49. $n = 126$ police officers ; $\bar{X}_1$ = Blood concentration of Lead ($\mu g/dL$) = $29.2 \mu g/dL$

$\sigma_{\bar{x}_1} = 7.5 \mu g/dL$.

$m = 50$ police officers ; $\bar{y} = 18.2 \mu g/dL$; $\sigma_y = 5.8 \mu g/dL$.

$H_0: \bar{x} = \bar{y}$; $H_1: \bar{x} \neq \bar{y}$

$s_p = 7.1 \mu g/dL$; $s_{\bar{x}-\bar{y}} = 1.2 \mu g/dL$

$df = 174$; $t = 9.3 > t_{174}(1.05)$

Rejection of null hypothesis (H_0).

The officers from Cairo do not test to be of the same sample set as Abbasia

$CI_{Cairo-Abbasia} = \{8.69 \mu g/dL , 13.3 \mu g/dL\}$

50. a. $CI_{Male-Female\ Temp} = \{-0.53°F, -0.04°F\}$

The use of normal approximation is reasonable because the mean ≈ median, and indicates low skew.

b. $CI_{Male-Female\ Heart\ rate} = \{-2.7 \text{ bpm}, 1.2 \text{ bpm}\}$

Application of a normal approximation better fits the male heart rate because of little skew, but the females mean heart rate (74 bpm) is low to the median (76 bpm) and may indicate the need for a parametric test,

c. Parametric:

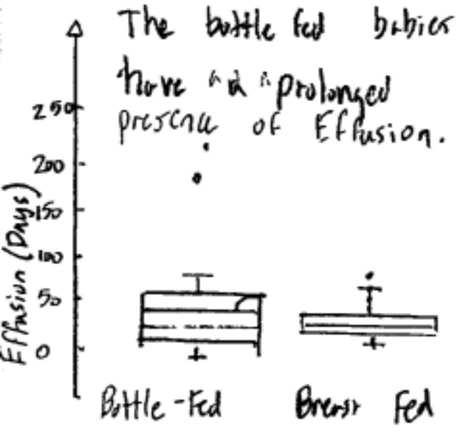
Experiment	R	R'	U
Temperature (Male)	4732	4733	12.2
Temperature (Female)	4733		
Heart Rate (Male)	4433	4442	10.9
Heart Rate (Female)	4423		

Nonparametric:

Experiment	E[x]	SD[x]	s_p	$s_{\bar{x}-\bar{y}}$	df	t
Temperature (Males)	98.1	0.69	0.731	0.13	120	2.30
Temperature (Females)	98.4	0.77				
Heart Rate (Males)	73	6.99	7.72	1.23	128	-0.83
Heart Rate (Females)	74	8.00				

The nonparametric test is showing rejection of the null hypothesis for male and female mean temperatures while acceptance of the alternative hypothesis for heart rate means. Although, since the sample size is large (>30), the parametric test should be established as the leading indicator, and demonstrates corresponding p-values greater than a significance of $\alpha = 0.05$

51. a.



b) A parametric test seems applicable because of the large skew between means and median values.

$H_0: \overline{\text{Bottle Fed (days)}} = \overline{\text{Breast Fed (Days)}}$

$H_1: \overline{\text{Bottle Fed (days)}} \neq \overline{\text{Breast Fed (Days)}}$

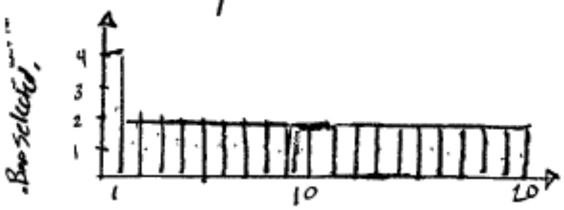
Experiment	R	R'	U
Bottle Fed	515	660	7.69
Breast Fed	660		

The model suggest rejection of the null hypothesis in favor of the alternative that the Breast Fed babies do not have prolonged Effusion

52. a) "Recover faster" does not indicate which disease.
b) Insufficient evidence to conclude the wife does or does not smoke.
c) How does breakfast relate to industrial accidents?
d) Did the student scores compare to the majority or minority school?
e) Would a questionnaire better been prepared if other days were tested?
f) A comparator would help determine if beer or alcohol should be reported.
g) Did the 15-year study have a controlled variable?
h) What about the other 35% of married couples?
i) Were the elderly of the same age group?

53. Both lettuce leaves and unlit cigarettes represent placebos to the experimental design. Lettuce leaves contain no amount of nicotine, while unlit cigarettes do, but are not inhaled and solely behaviorly considered.

54. The length of bar was not randomized, and yet the error would be detectable if not randomized over time because of selection bias being time independent.



Chapter 12: Analysis of Variance:

1.

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{m+n-2} = \frac{\sum_{i=1}^{2}(J-1)s_i^2}{2(J-1)} = \boxed{\frac{(J-1)s_1^2 + (J-1)s_2^2}{2J-2}}$$

3. $I=2$ treatment groups ; F-statistic

$$F = \frac{SS_B/(I-1)}{SS_W/[I(J-1)]} \quad ; \text{ where } SS_B = J\sum_{i=1}^{I}\alpha_i^2 + (I-1)\sigma^2 ; \; SS_W = \sum_{i=1}^{I}(J-1)s_i^2$$

t-statistic:

$$t = \frac{(\bar{X}-\bar{Y}) - (\mu_x - \mu_y)}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \; ; \quad = \frac{J\sum_{i=1}^{2}\alpha_i^2 + (1)\sigma^2}{(2-1)} \cdot \frac{2(J-1)}{\sum_{i=1}^{2}(J-1)s_i^2} = \frac{[J(\alpha_1^2+\alpha_2^2)+\sigma^2]\,2(J-1)}{(J-1)s_1^2 + (J-1)s_2^2} = \frac{J(\alpha_1^2+\alpha_2^2)+\sigma^2}{s_p^2}$$

$\sum \alpha_i^2 = 0$

$$t^2 = \frac{(\bar{X}-\bar{Y})^2 - 2(\bar{X}-\bar{Y})(\mu_x-\mu_y) + (\mu_x+\mu_y)^2}{s_p^2\left(\frac{1}{n}+\frac{1}{m}\right)} = \boxed{\frac{\sigma^2}{s_p^2} = \Leftrightarrow = \frac{\sigma^2}{s_p^2}}$$

$$\boxed{t^2 = F} \;@\; I=2.$$

4. Theorem A:

$$E(SS_W) = \sum_{i=1}^{I}\sum_{j=1}^{J}E(Y_{ij}-\bar{Y}_{i.})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(\mu_i-\bar{\mu})^2 + \frac{J-1}{J}\sigma^2 = \boxed{\frac{1}{J}\sum_{i=1}^{I}\sum_{j=1}^{J}(\mu_i-\bar{\mu})^2 + I(J-1)\sigma^2}$$

From section 12.2.1 : $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

$$E(SS_B) = J\sum_{i=1}^{I}E(\bar{Y}_{i.}-\bar{Y}_{..})^2 = J\sum_{i=1}^{I}(\mu_i-\bar{\mu})^2 + \frac{(I-1)}{IJ}\sigma^2 = \boxed{J\sum_{i=1}^{I}\alpha_i^2 + \frac{(I-1)\sigma^2}{IJ}}$$

Theorem B: $\mu = 0$; σ^2; $SS_W/\sigma^2 = \chi^2_{I(J-1)}$

$$\boxed{= \frac{\sum_{i=1}^{I}(J-1)s_i^2}{\sigma^2} = (J-1)\sum^{I}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \chi^2_{I(J-1)}}$$

$$SS_B/\sigma^2 = \frac{J\sum_{i=1}^{I}\alpha_i^2 + (I-1)\sigma^2}{\sigma^2} = \frac{\sum \alpha_i^2}{Var(\bar{Y}_i)} + \frac{(I-1)}{(I-1)}\left(\frac{X-\mu}{\sigma}\right)^2 = \boxed{\chi^2_{I-1}}$$

5. F-Statistic

$$F = \frac{SS_B/(I-1)}{SS_W/[I(J-1)]}$$

Null hypothesis:

$H_0: \alpha_1 = \alpha_2 = \cdots \alpha_I = 0$ means $F_1 \approx 1$, while $H_1: \alpha_1 \neq \alpha_2 \neq \cdots \alpha_I \neq 0$ variation of groups and within groups exist.

Assuming normal distribution, likelihood (Λ): $\dfrac{W(\ldots, \sigma^2, Z^2)/m}{V(X\mid\mu,\sigma^2)\sum V_i/n} = \dfrac{N(X\mid\mu,\sigma^2)^2/m}{\sum N(X_i\mid\mu,\sigma^2)^2/n}$

if $m = I(J-1)$, $n = (I-1)$; $= \boxed{\dfrac{\chi^2_{I(J-1)}/I(J-1)}{\chi^2_{I-1}/(I-1)}}$

Bonferroni Inequality:

$P(\bigcap_{i=1}^{n} A_i) \geq 1 - \sum P(A_i^c)$;

$1-\sum P(A_i^c)$; A_i is the inphasi'ty three probability space.

A_i^c is the complement inputs to the probability space.

1 $P(\bigcap_{i=1} A_i)$

7. Show Theorem B of Section 12.2.1: $SS_B/\sigma^2 \sim \chi^2_{I-1}$

$$SS_B/\sigma^2 = \frac{\sum_j \sum_{i=1}^{I} x_i^2 + (I-1)\sigma^2}{\sigma^2} = \frac{\sum x_i^2}{Var(\bar{Y}_i)} + (I-1) = \frac{\sum x_i^2}{Var(\bar{Y}_i)} + \frac{(I)}{Var}$$

$$= (I-1)\sum_{i=1}^{I}\left(\frac{X-\mu}{\sigma^2}\right)^2 = \boxed{\chi^2_{I-1}}$$

8. Lob1 $CI_i(L_m)$ $CI(\ldots)$

2

3

4

9. The columns of a t Distribution relate to Studentized ranges because of similar derivation Tukey's method on the probability of random variable led from a max of mean difference per pooled standard deviation per degrees of freedom for a normal distribution to $\frac{s_p}{\sqrt{5}}$; nevertheless, t distribution was closely derived from mean differences, but per standard deviation per degrees freedom of two normal distributions. Thus t distribution columns here a $\sqrt{2}$ for two distributions, like

10. $I=7$, $J=10$. Tukeys Method: $|\bar{Y}_{i1}-\bar{Y}_{i2}| \pm q_{7,63}(\alpha)\frac{s_p}{\sqrt{J}}$ $= 1 \pm \frac{q_{7,63}(\alpha)}{t_{63}(\alpha/n)}\sqrt{\frac{5}{J}} = 1 \pm \frac{4.31}{3.16}\sqrt{\frac{5}{10}}$ $\boxed{= 1 \pm 0.96}$

Bonferroni Method: $(\bar{Y}_{i1}-\bar{Y}_{i2}) \pm s_p \frac{t_{63}(\alpha/n)}{\sqrt{5}}$

Regular t-statistic: $(\bar{Y}_{i1}-\bar{Y}_{i2}) \pm t_{63}(0.025) s_p \sqrt{\frac{2}{J}} = 1 \pm 0.053$

11.

Factors				
I	A	B	C	D
II	A	B	C	D
III	A	B	C	D

Diagonal

Factors / Levels

$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$	$\bar{X}_5$
$\bar{X}_3$	$\bar{X}_4$	$\bar{X}_5$	$\bar{X}_6$

Mean Space: $\bar{X}_1, \bar{X}_2, \bar{X}_3, \bar{X}_4, \bar{X}_5$

12.

A B C Sample Space

I $\bar{X}_1$ $\bar{X}_2$ $\bar{X}_3$ $\quad \bar{X}_1, \bar{X}_2, \bar{X}_3, \bar{X}_4$

II. $\bar{X}_2$ $\bar{X}_3$ $\bar{X}_4$

$\boxed{\text{Variance} = \frac{12}{N(N+1)} = K}$

13. Kruskal-Wallis test:

$\boxed{\text{Mann-Whitney: Average Rank } \bar{R} = \frac{(N+1)}{2}}$

R_{ij} = the rank of Y_{ij} in the combined sample

$\bar{R}_i = \frac{1}{J_i}\sum_{j=1}^{J_i} R_{ij}$ = Average rank in i^{th} group

$\bar{R}_{..} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{J} R_{ij} = \frac{N+1}{2}$ = Average total rank

$SS_B = \sum_{i=1}^{I} J_i(\bar{R}_i - \bar{R}_{..})^2$ = Variance between ranks

$K = \frac{12}{N(N+1)} SS_B$ = Chi-squared Distribution with $(I-1)$ degrees of freedom.

$= \frac{12}{N(N+1)}\left(\sum_{i=1}^{I} J_i \bar{R}_i^2\right) - 3(N+1)$

14. Friedman's Test: $SS_A = J\sum_{i=1}^{I}(\bar{R}_i - \bar{R}_{..})^2$

$$Q = \frac{12J}{I(I+1)}\sum(\bar{R}_i - \bar{R}_{..})^2$$

Sign Test: $p = P(X>Y)$; $H_0: p = 0.50$

m pairs: $\{x_1, y_1\} \cdots \{x_n, y_n\}$

15. $W = \#$ pairs $y_i - x_i > 0 \sim bin(m, p)$

The tested probability of a signed rank test ($p = P(X>Y)$) is Kin to Friedman's probability ($P(\chi^2_{k-1} \geq Q)$) when two categories are presented ($k=2$)

$$K = \frac{12}{N(N+1)}\sum_{i=1}^{I} J_i(\bar{R}_i - \bar{R}_{..})^2$$

$$= \frac{12}{N(N+1)}\sum_{i=1}^{I} J_i(\bar{R}_i^2 - 2\bar{R}_i\bar{R}_{..} + \bar{R}^2_{..})$$

$$= \frac{12}{N(N+1)}\sum_{i=1}^{I} J_i\bar{R}_i^2 - 2\sum_{i=1}^{I} J_i \frac{(N+1)}{2}\frac{1}{J_i}\sum_{j}^{J} R_{ij} + \sum_{i=1}^{I} J_i \frac{(N+1)^2}{2\cdot 2}$$

$$= \frac{12}{N(N+1)}\sum_{i=1}^{I} J_i\bar{R}_i^2 - \frac{12\cdot \cancel{2}}{\cancel{N}(N+1)}\cdot\frac{\cancel{N}(N+1)}{2}\frac{(N+1)}{\cancel{2}} + \frac{12}{N(N+1)}\frac{(N+1)^2}{4} = \boxed{\frac{12}{N(N+1)}\sum_{i=1}^{I} J_i\bar{R}_i^2 - 3(N+1)}$$

16. Prove $SS_{TOT} = SS_A + SS_B + SS_{AB} + SS_E = JK\sum^{I}(\bar{Y}_{i..} - \bar{Y}_{...})^2 + IK\sum^{J}(\bar{Y}_{.j.} - \bar{Y}_{...})^2 + K\sum^{I}\sum^{J}(\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 + \sum^{I}\sum^{J}\sum^{K}(Y_{ijk} - \bar{Y}_{ij})^2$

$= \sum^{J}\sum^{K}\sum^{I}(\bar{Y}_{i..} - \bar{Y}_{...})^2 + \sum^{I}\sum^{J}\sum^{K}(\bar{Y}_{.j} - \bar{Y}_{...})^2 + \sum^{K}\sum^{I}\sum^{J}(\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{j.} - \bar{Y}_{...})^2 + \sum^{I}\sum^{J}\sum^{K}(Y_{ijk} - \bar{Y}_{ij})^2$

$= \sum^{I}\sum^{J}\sum^{K}\left[(\bar{Y}_{i..} - \bar{Y}_{...})^2 + (\bar{Y}_{.j} - \bar{Y}_{...})^2 + (\bar{Y}_{ij} - \bar{Y}_{i..} - Y_{j.} - \bar{Y}_{...})^2 + (Y_{ijk} - \bar{Y}_{ij})^2\right]$

$\boxed{= \sum^{I}\sum^{J}\sum^{K}\left[(Y_{ijk} - \bar{Y}_{ij})^2\right]}$

17. Find mle's of $\alpha_i, \beta_j, \delta_{ij}$, and μ.

$$l = -\frac{IJK}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum^{I}\sum^{J}\sum^{K}(Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij})^2$$

$$\frac{dl}{d\alpha_i} = \frac{\alpha_i}{\sigma^2}\sum^{I}\sum^{J}\sum^{K}(Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij}) = 0$$

$\boxed{\hat{\alpha}_i = \bar{Y}_{i..} - \bar{Y}_{...}}$

$$\frac{dl}{d\beta} = \frac{\beta_j}{\sigma^2}\sum^{I}\sum^{J}\sum^{K}(Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij}) = 0$$

$\boxed{\beta_j = \bar{Y}_{.j.} - \bar{Y}_{...}}$

$$\frac{dl}{d\delta_{ij}} = \frac{\delta_{ij}}{\sigma^2}\sum^{I}\sum^{J}\sum^{K}(Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij})$$

$\boxed{\delta_{ij} = \bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{ij.} + \bar{Y}_{...}}$

$$\frac{dl}{d\mu} = \frac{\mu}{2\sigma^2}\sum^{I}\sum^{J}\sum^{K}(Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij}) = 0$$

$\boxed{\mu = \bar{Y}_{...}}$

18.

19. Fischers Exact Test:

$$p(n_{11}) = \frac{\binom{n_{1.}}{n_{11}}\binom{n_{2.}}{n_{21}}}{\binom{n_{..}}{n_{.1}}}$$

n_{11}	$p(n_{11})$
5	0.03
6	0.06
7	0.09
8	0.12
9	0.12
10	0.10
11	0.06
12	0.03
13	0.01
14	0.00
15	0.000

The ~~Fisher~~ Exact ~~Test~~ presents no significant difference between the high and low anxiety groups at $n_{11} = 12$.

20. Random Sample / Prospective Study / Retrospective Study

a) Random Sample:

$$P(D|X) = \frac{\pi_{11}}{\pi_{10}+\pi_{11}}$$

$$P(D|\bar{X}) = \frac{\pi_{01}}{\pi_{00}+\pi_{01}}$$

$$\Delta|X) = \frac{odds(D|X)}{odds(D|\bar{X})} = \frac{P(D|X)}{1-P(D|X)}\frac{1-P(D|\bar{X})}{P(D|\bar{X})}$$

$$\boxed{= \frac{\pi_{11}\pi_{00}}{\pi_{01}\pi_{10}}}$$

Prospective Study:

$$P(D|X) = \frac{n_{11}}{n_{10}+n_{11}}$$

$$P(D|\bar{X}) = \frac{n_{01}}{n_{00}+n_{01}}$$

$$\Delta = \frac{odds(D|\bar{X})}{odds(D|\bar{X})} = \frac{P(D|X)}{1-P(D|X)}\frac{1-P(D|\bar{X})}{P(D|X)}$$

$$\boxed{= \frac{n_{11}}{n_{01}}\left(\frac{n_{00}}{n_{10}}\right)}$$

Retrospective Study:

$$P(X|D) = \frac{n_{10}}{n_{00}}$$

$$P(X|\bar{D}) = \frac{n_{10}}{n_{00}}$$

$$\Delta = \frac{odds(D|X)}{odds(D|\bar{X})} = \frac{P(D|X)}{1-P(D|X)}\frac{1-P(D|\bar{X})}{P(D|\bar{X})}$$

$$\boxed{= \frac{n_{11}}{n_{01}}\cdot\frac{n_{00}}{n_{10}}}$$

b) Method of Propagation of Error: $Y = g(x) \approx g(\mu_x) + (X-\mu_x)g'(\mu_x)$

$\pi_{11} + \pi_{10} \cdot \pi_{11} \quad \pi_{00}$

$$\mu_y = g(\mu_x)$$

$$\sigma_y^2 = \sigma_x^2\left[g'(\mu_x)\right]^2$$

Random Sample:

$$\log\Delta = \log\frac{\pi_{11}\pi_{00}}{\pi_{01}\pi_{10}}$$

$$Var\log(\Delta) = \frac{\sum_i^n(X-\mu_x)^2}{n}\circ\log'(\Delta)$$

$$= \frac{\sum\sum(\pi_{ij}-\mu)^2}{m\cdot n}\circ\frac{1}{\Delta^2}$$

$$\boxed{= \frac{\sigma^2}{\Delta^2}} = \sigma^2\left(\frac{\pi_{01}\pi_{10}}{\pi_{11}\pi_{00}}\right)^2$$

Prospective Study:

$$\log\Delta = \log\frac{n_{11}}{n_{01}}\cdot\frac{n_{00}}{n_{10}}$$

$$Var\log\Delta = \frac{\sum_i^n(X-\mu_x)^2}{n}\circ\log'\Delta$$

$$= \frac{\sum\sum(\pi_{ij}-\mu)^2}{m\cdot n}\circ\frac{1}{\Delta}$$

$$\boxed{= \frac{\sigma^2}{\Delta^2}} = \sigma^2\left(\frac{n_{01}n_{10}}{n_{11}n_{00}}\right)^2$$

Retrospective Study:

$$\log\Delta = \log\frac{n_{11}}{n_{01}}\cdot\frac{n_{00}}{n_{10}}$$

$$Var\log\Delta = \frac{\sum_i^n(X-\mu_x)^2}{n}\log\Delta$$

$$= \frac{\sum\sum(\pi_{ij}-\mu)^2}{m\cdot n}\circ\frac{1}{\Delta}$$

$$\boxed{= \frac{\sigma^2}{\Delta} = \sigma^2\left(\frac{n_{01}n_{10}}{n_{11}n_{00}}\right)^2}$$

21.

	Diabetic	Normal
Bb or bb	12	4
BB	39	49

$P(D|X) = \frac{49}{39+49}$; $odds(D|\bar{X}) = \frac{49}{39}$; $\Delta = \frac{49\cdot 12}{39\cdot 4}$ $\boxed{= 3.77}$

$P(D|\bar{X}) = \frac{12}{12+4}$; $odds(D|\bar{X}) = \frac{4}{12}$

The odds of being normal (Dominant) vs Diabetic (Dominant-recessive) is 3.77:1.

22. a

	Advised	Not Advised
Male	48	47
Female	80	136

Proportion of males advised to quit: $\frac{48}{48+47} = 50.6\%$

Proportion of females advised to quit: $\frac{80}{80+136} = 37.7\%$

Standard error of male proportion: 0.05

Standard error of female proportion: 1.91

Standard error of their difference: 0.0 ; Significant with t-test

	Advised	Not Advised
White	26	34
African-American	102	149

Proportion of Whites asked to quit: 43%

Proportion of African-Americans asked to quit: 41%

Standard error of White proportion: 0.6

Standard error of African-American proportion: 0.0 ; Significant t-test

19. One-Way Layout

Y_{ij} = j^{th} observation of i^{th} treatment

$= \mu + \alpha_i + \varepsilon_{ij}$ ← random error of j^{th} treatment

overall mean; Differential Effect of i^{th} treatment

$\sum_i^I \alpha_i = 0$: Normalized

$\sum_i^I \sum_j^J (Y_{ij} - \bar{Y}_{..})^2 = \sum_i^I \sum_j^J (Y_{ij} - \bar{Y}_{i.})^2 + J\sum_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$

Where $\bar{Y}_i = \frac{1}{J}\sum_j^J Y_{ij}$: Average observation of i^{th} treatment

$\bar{Y}_{..} = \frac{1}{IJ}\sum_i^I \sum_j^J Y_{ij}$: overall Average

$SS_{TOT} = SS_W + SS_B = \sum_i^I (J-1)s_i^2 + \sum_i^I \alpha_i^2 + \sigma^2$

Two-Way Layout

$\bar{Y}_{..} = \frac{1}{IJ}\sum^I \sum^J Y_{ij}$: Grand Average

$\bar{Y}_{i.}$: Average over rows

$\bar{Y}_{.j}$: Average over columns

$\hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}_{..}$: Differential row Average

$\hat{\beta}_i = \bar{Y}_{.j} - \bar{Y}_{..}$: Differential column Average.

$\hat{\delta}_{ij}$ = Interaction of rows x columns.

$Y_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\delta}_{ij}$: Additive Model

$Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}$: Balanced.

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\varepsilon_i)^2/2\pi^2$

$SS_{TOT} = SS_A + SS_B + SS_{AB} + SS_E$

Three-Way Layout

$Y_{ijkl} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\delta}_{ijk} + \ldots$

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\varepsilon_{ijkl}^2/2\sigma^2)$

$l = \ldots \sum \varepsilon_i^2/2\sigma^2$

$\frac{dl}{d\mu} = Y_{ijkl} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k - \hat{\delta}_{ijk} - \mu = 0$

$\boxed{\hat{\mu} = \bar{Y}_{....}}$

$\frac{dl}{d\alpha} = Y_{ijkl} - \alpha_i - \beta_j - \gamma_k - \delta_{ijk} - \mu = 0$

$\boxed{\hat{\alpha}_i = \bar{Y}_{i...} - \bar{Y}_{....}}$

$\frac{dl}{d\beta} = Y_{ijkl} - \mu - \alpha_i - \beta_j - \delta_{ijkl} = 0$

$\boxed{\hat{\beta}_j = \bar{Y}_{.j..} - \bar{Y}_{....}}$

$\frac{dl}{d\gamma} = Y_{ijkl} - \mu - \alpha_i - \beta_i - \delta_{ijkl} = 0$

$\boxed{\hat{\gamma} = \bar{Y}_{..k.} - \bar{Y}_{....}}$

$\frac{dl}{d\delta} = Y_{ijkl} - \mu - \alpha_i - \beta_i - \delta_{ijkl} = 0$

$\boxed{\hat{\delta}_{ijk} = \bar{Y}_{ijk.} - \bar{Y}_{i...} - \bar{Y}_{.j..} - \bar{Y}_{..k.} + \bar{Y}_{....}}$

The parametrization of a balanced three-way layout includes two-factor and three-factor interactions. A two-factor interaction $(\alpha_i, \beta_i, \gamma_i)$ represents the mean of a row, column, or depth of the matrix. While the three-factor interaction $(\hat{\delta}_{ijkl})$ is interpreted as the the residual within the cell.

20. $Y_{ij} = \mu + A_i + \varepsilon_{ij}$; A_i = Random ; $E(A_i) = 0$; $Var(A_i) = \sigma_A^2$; ε_{ij} independent of A_i

$E(\varepsilon_j) = 0$; $Var(\varepsilon_j) = \sigma_\varepsilon^2$

a) Show $E(MS_W) = \sigma_\varepsilon^2$; $E(MS_B) = \sigma_\varepsilon^2 + J\sigma_A^2$; $E(MS_W) = \frac{E(SS_{AB})}{(I-1)(J-1)} = \sigma^2 + \frac{1}{(I-1)(J-1)} \sum_i^I \sum_j^J \delta_{ij}^2 = \boxed{\sigma_\varepsilon^2}$; $=0, \sum\sum \delta_{ij} = 0$

$E(MS_B) = \frac{E(SS_B)}{(J-1)} = \sigma_\varepsilon^2 + \frac{I}{(J-1)} \sum^J \beta_j^2 = \boxed{\sigma_\varepsilon^2 + J\sigma_A^2}$

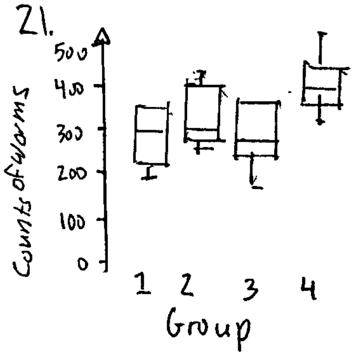
For dataset: Dye

Source	dF	SS	MS	F
Samples	5	111.74	22.35	5.65
~~Error~~	5	113.65	22.73	
Total	10	225.39		

$F = \frac{SS_B/(J-1)}{SS_W/[I(J-1)]} = 5.90.$

b) The parameters of the model are estimated by the F-statistic because the stat incorporates mean square error of multiple categories. The hypothesis $(H_0 = \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0)$ fails at a p-value of 0.05.

21.



Parametric Technique

Source	dF	SS	MS	F
Groups	3	27234	9078	2.27
Larval	16	63953	3997	
Total	19	91187		

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$

$H_1: \alpha_1 \neq \alpha_2 \neq \alpha_3 \neq \alpha_4 \neq 0$

p-value > 0.10 $\quad F_{3,16} = 2.24$

Rejection of null hypothesis.

Nonparametric Technique:

Group	1	2	3	4
R	42	53	36	79

$H = 6.2$, $df = 3$, ${}_{3}\chi^2_{0.90} = 7.78$

Rejection of null hypothesis at p-value > 0.10.

Both parametric and nonparametric tests demonstrate unequal amounts of larvae per group.

22. Manufacturer #1

Source	df	SS	MS	F
Labs	6	0.125	0.021	5.60
Amount	63	0.231	0.004	
Total	69	0.356		

$H_0: \alpha_1 = \alpha_{\dots} = \alpha_7 = 0$

$H_1: \alpha_1 \neq \alpha_{\dots} \neq \alpha_7 \neq 0$

$F_{6,60} = 3.12$

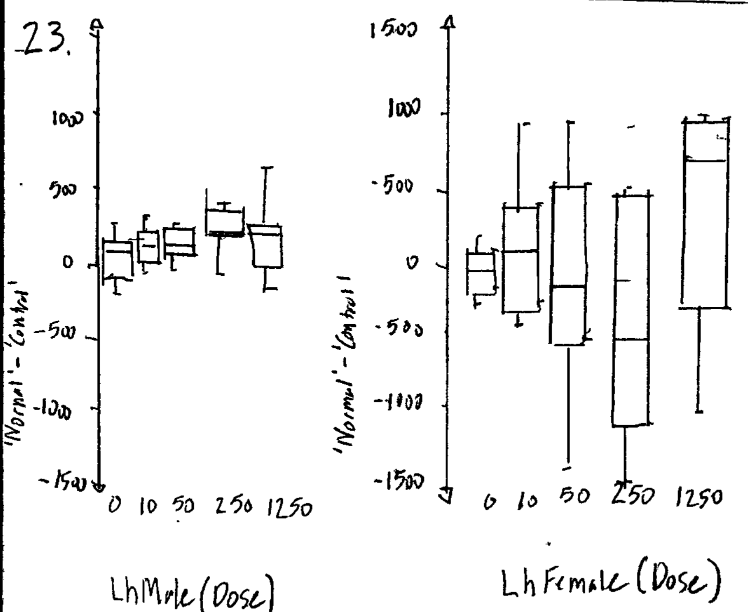
Rejection of the null hypothesis that lab data contains similar means and variances.

Manufacturer #2

Source	df	SS	MS	F
Labs	6	0.153	0.026	11.90
Amount	63	0.135	0.002	
Total	69	0.288		

Also, rejection of the null hypothesis with a p-value > 0.01.

23.



${}_{28}\chi^2(0.95) = 41.337$, $\chi^2 = 1040$ Females 'Normal' - 'Constant'

$H_0: \bar{X} = \mu_0$; $H_1: \bar{X} \neq \mu_0$; Rejection of null hypothesis for females. There is indication of threshold dosage for luteinizing hormone.

Source	dF	SS	MS	F
Dosage	4	6603.9701	1659925	21.4
Light	1	9506947	9506447	122.4
Interaction	4	4641765	1160441	14.9
Error	50	3883247	77664	
Total	59	24671660		

24.

Source	SS	dF	MS	F	p-value
Labs	33.1	11	3.0	6.7	3×10^{-8}
Cereal	3556.3	5	711.3	1453.5	2×10^{-121}
Interaction	30.3	55	0.55	1.1	0.29
Error	70.5	144	0.41		
Total	3690.2	215			

The null hypothesis is rejected that the effect from luteinizing hormone generates a balanced result with increasing dosage for both men and women ($H_A: \alpha_1 = \alpha_i$)

The first successful Two-Way ANOVA by hand presents inconsistent data per lab and bran vs. cereal! because of a p-value < 0.05. The error is consistent between data sets at 0.29.

b. $H_0: \pi_R = 1/2$ $\chi^2 = 0.30$; $\chi^2 = 7.81$
$H_1: \pi_{SR} \neq 1/2$ $df = 3$ Not significant between sports

c. Both of the tests are an equivalent argument with a chi-squared statistic.

d. There is little evidence to suggest red is favorable to the wins of sports.

e.

Sport	Red	Blue
Boxing	128	105
Free Wrestling	28	39
GR Wrestling	25	34
TaeKwonDo	27	26

$\chi^2 = 8.59$
$df = 3$
$\chi^2_{0.95} = 7.81$

Examination of larger datasets show a different conclusion; more specifically, color does effect the outcome of a match. The means are significant across wrestling.

25. a. Individual analysis of aspirin's effects were conducted for myocardial infarction, and stroke. Myocardial disease ($\chi^2 = 1.38, df = 1$) results were not significant, but aspirin did effect a stroke ($\chi^2 = 30.65, df = 1$).

b. Aspirin analysis significant results for lowering myocardial mortalities, but overall had no influence on death ($\chi^2 = 12.61$; $df = 4$).

26. McNemars Test evaluated DKA relationship to therapy. A chi-squared of 5.53, with one degree of freedom lists significant results for the relationship between side effects before and after therapy.

27. Upon examination, the defendants race does not influence death penalty. By including the victims race, the results could change the outcome, but not the analysis. The chi-square statistic ($\chi^2 = 15.92, df = 3$) does list significant results for a relationship of race to death penalty.

28. The chi-square test is independent to counts, frequency, or percents and presents similar outcomes

29. Model: Satisfied Somewhat Dissatisfied Very Dissatisfied
A statistician should carry out a test of homogeneity, rather than a test for independence because of the workers being already independent to each other.

25. Null Hypothesis: $H_0: \alpha_1 = \cdots = \alpha_i = 0$

Alternative Hypothesis $H_1: \alpha_1 \neq \cdots \neq \alpha_i \neq 0$

Source	SS	dF	MS	F	p-value	Fcrit
Between	201.8	9	22.4	2.84	0.005	1.99
Within	710.4	90	7.9			
Total	912.2	99				

The data from magnesium samples, at glance, presents large chemical content in column #1. An One-way Anova describes error between the groups and within at a p-value of 0.005, rejecting the null hypothesis of similar chemical content across samples and portions of the bar.

26. Parametric

Source	SS	dF	MS	F	p-value	F-crit
Dogs	0.52	9	0.057	0.349	0.95	2.39
Chemical	3.29	20	0.16			
Total	3.81	29				

Nonparametric

Dog	1	2	3	4	5	6	7	8	9	10
R	37	60.5	67.5	38.5	45	43	46.5	55.5	44.5	25

$H = 0.87$, $\chi^2_{9,\,0.005} = 1.73$

The F-statistic concludes no differences per dog or chemical exposed, with a p-value > 0.95. A nonparametric Kruskill-Wallace test confirmed no differences as compared to a 9-degree chi-squared distribution. Blood plasma did not change as compared to exposure.

27. Bonferroni Method:

Comparison	Mean Diff.	CI
C7 - AJ	177.3	69.4
C7 - F2	101.4	69.4
AJ - F2	-75.9	69.4

The Bonferroni method confirms each group is drastically different as represented by their means; while, confidence intervals show errors of nearly double the comparator.

28. Parametric:

Source	SS	dF	MS	F	p-value	Fcrit
Types (Between)	446.6	2	223.3	0.497	0.61	3.59
Types (Within)	7632.3	17	449.0			
Total	8078.9	19				

Nonparametric:

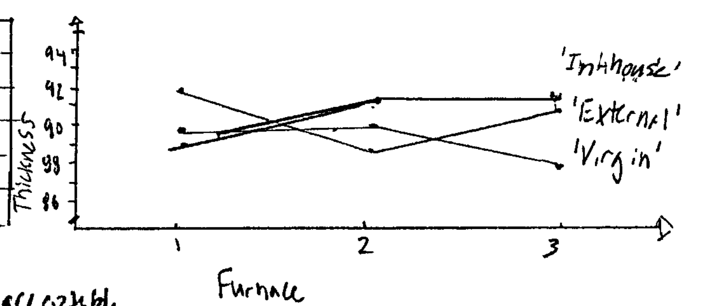
	Type I	Type II	Type III
R	76.5	78	55.5
n	9	6	5
H		2.15	

$\chi^2_{18,\,0.95} = 28.87$

∴ Both the parametric and nonparametric analysis confirmed the hypothesis ($H_0: \Pi_1 = \Pi_2 = \Pi_3$).

29.

Source	SS	dF	MS	F	p-value	F-crit
Furnace	4.1	2	2.05	1.45	0.26	3.55
Wafer Type	5.9	2	2.95	2.07	0.16	3.55
Furnace and Wafer	21.3	4	5.34	3.76	0.02	2.93
Residual	25.6	18	1.42			
Total	56.9	26				



Noticeable differences exist between the furnace and wafer thickness, but the type and furnace individually are acceptable.

30. Tukey's Method:

$q_{6,40}(0.05) = 4.23$

$S_p = 216.07$

$\sqrt{J} = 3.16$

Tukey $CI = 289.02$

Plot	Average
'3'	1266
'2'	1272.6
'5'	1306.4
'4'	1316.4
'6'	1339.4
'1'	1768.5

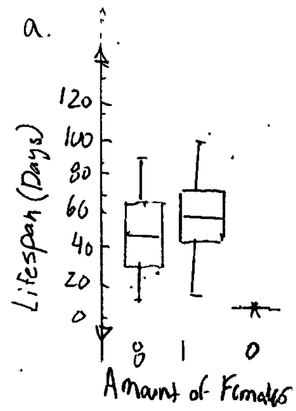
$Max(\bar{Y}_i - \bar{Y}_i) = $ Plot 3 - Plot 1

$= 500.5$

Tukey's method's comparison of maximal average differences is not significant.

31. An additive model would provide a good fit to the dataset for secondary indication maximum windspeeds are variable

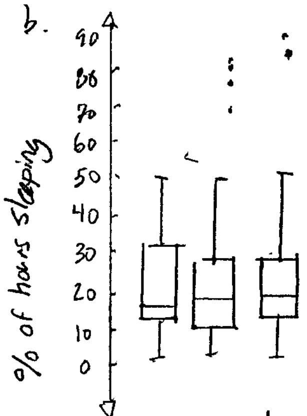
32. a.



Amount of Females	Mean	Variance
8	63.6	259.8
1	60.8	240.73
0	51.0	323.6

The lifespan without females was drastically for male fruitflies.

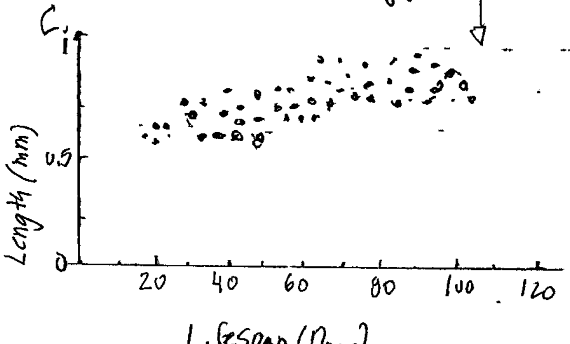
b.



Amount of Females	Mean	Variance
8	21.6	149.0
1	24.9	297.7
0	23.0	249.0

The hours slept were equivalent with or without a female fruitfly present.

c.



Lifespan of a fruitfly is correlated to the length of the thorax, with a positive relationship.

d. Thorax

Source	df	MS	F	p-value	Fcrit
Between	4	0.006	1.15	0.33	2.45
Within	120	6.006			
Total	124				

Lifespan

Source	df	MS	F	p-value	Fcrit
Between	4	2985	13.6	3.6×10^{-9}	2.45
Within	120	219			
Total	124				

Sleep

Source	df	MS	F	p-value	F-crit
Between	4	121.7	0.47	0.75	2.45
Within	120	256.5			
Total	124				

Bonferroni Method

Lifespan : CI = 0.02

Comparison	9/10	9/11	9/80	9/81	10/11	10/80	10/81	11/80	11/81	80/81
Mean Difference	-4.28	6.8	0.2	24.8	8.0	1.4	26.1	-6.6	18.0	24.6

Thorax : CI = 3.21

Comparison	9/10	9/11	9/80	9/81	10/11	10/80	10/81	11/80	11/81	80/81
Mean Difference	0.01	0.00	0.00	0.04	-0.01	-0.01	0.08	0.01	0.04	0.04

Comparison	9/10	9/11	9/80	9/81	10/11	10/80	10/81	11/80	11/81	80/81
Mean Difference	-2.5	-4.2	-3.6	0.8	1.7	-1.1	3.3	0.6	5	4.4

Sleep : CI = 0.02.

Tukey Method:

Category	$q_{5,100}$	S_p	$\sqrt{J}$	CI	$\max(\bar{Y}_i - \bar{Y}_i)$
Thorax	5.32	3.05	2.3	7.2	0.04
Lifespan	5.32	0.02	2.3	0.04	26.08
Sleep	5.32	0.07	2.3	0.03	4.4

The lifespan is concluded as significant.

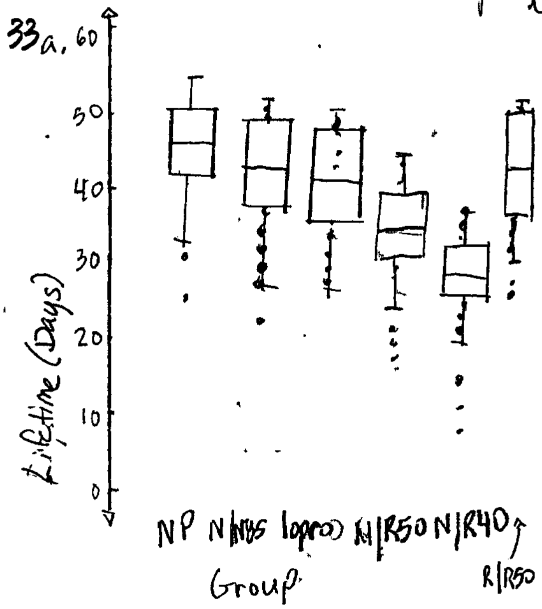
e. Kruskal-Wallis test:

Group	H	df	$\chi^2_{0.95}$
Lifespan	37.9	5	11.07
Thorax	5.6	5	11.07
Sleep	0.3	5	11.07

With another nonparametric analysis, lifespan of the fruitfly is significant, and an outlier to the information collected.

f. The dataset of one, eight (pregnant, or virgin) females provides no change on sleep.

33a.



Kruskal-Wallis Test:

$H = 159.01$; $\chi^2_{0.95} = 11$ $H_0: \pi_1 = \pi_2 = \pi_3 = \pi_4 = \pi_5 = \pi_6$

df: 5 $H_1: \pi_1 \neq \pi_2 \neq \pi_3 \neq \pi_4 \neq \pi_5 \neq \pi_6$

The Kruskall-Wallis test rejected the null hypothesis at a p-value = 0.05, and accepted diet effects lifespan.

Preweaning (R/R50) does not lower lifespan of a mouse when the life is of restricted calories.

b. The reduction of protein does effect lifespan.

c. Reduction to 40 KCal per week does lower lifespan.

34. a.

Source	SS	df	MS	F	p-value	F-crit
Treatment	91.9	3	30.6	11.8	5.92×10^{-5}	3.0
Poison	105.1	5	21.0	8.1	0.0001	2.6
Treatment x Poison	39.2	15	2.6	1.01	0.48	2.1
Error	62.1	24	2.6			
Total	298.4	47				

The treatment and poison category show significance, but the interaction of the treatment x poison does not accept alternatives.

b.

Source	SS	df	MS	F	p-value	F crit.
Treatment	0.20	3	0.07	23.6	2.41×10^{-7}	3.0
Poison	0.40	5	0.07	24.9	8.99×10^{-9}	2.6
Treatment x Poison	0.02	15	0.00	0.5	0.89	2.1
Error	0.07	24				
Total	0.65	47				

Bob and Cox (1964) analysis of reciprical data conclude similar outcomes to partial a. Treatment x Poisons is not significant, and individual treatments or poisons are significant

35. The interaction of dosage and serum were not important, indicating the culture produced similar amounts of estrogen relative to primary categories. Although, estrogen production did have significant differences for type of serum, in addition to increased dosage.

Source	SS	df	MS	F	p-value	F crit.
Dosage	2.79×10^{9}	11	2.54×10^{8}	23.85	0.00	2.06
Treatment	6.30×10^{8}	2	3.15×10^{8}	29.60	0.00	3.26
Dosage x Treatment	3.98×10^{8}	22	1.81×10^{7}	1.70	0.07	1.85
Error	3.83×10^{8}	36	1.1×10^{7}			
Total	4.21×10^{9}	71				

Chapter 13: The Analysis of Categorical Data.

1.

	Diabetic	Normal	
Bb bb	12	4	16
BB	39	49	88
	51	53	104

$$p(Bb \text{ or } bb \mid \text{Diabetic}) = \frac{\binom{P(\text{Diabetic})}{P(Bb \text{ or } bb \mid \text{Diabetic})}\binom{P(BB)}{P(BB \mid \text{Diabetic})}}{\binom{\text{Total}}{P(\text{Diabetic})}}$$

Fishers Exact Test

$$= \frac{\binom{51}{12}\binom{88}{39}}{\binom{104}{51}} = 7.48 \times 10^{-5}$$

.1690

Yes, the probability of Allele (BB or bb) is less than a significance of $\alpha = 0.05$, which suggests an imbalance between the categories and acceptance of the alternative hypothesis ($p(Bb \text{ or } bb) \neq p(BB)$).

Chi-Square Test

	Diabetic	Normal
BB bb	24.99	62.3
BB	0.399	0.384

$$\chi^2 = \sum^{I}\sum^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum^{I}\sum^{J} \frac{(n_{ij} - n_{i.}n_{.j}/n_{..})^2}{n_{i.}n_{.j}/n_{..}} = 5.08,\ df = 1$$

$\chi^2_{0.975} = 5.02$

The outcome of comparing χ^2-statistic was a p-value < 0.025, which agrees with Fishers exact statistic.

2

Week	Chinese	Jewish	
-2	55	141	196
-1	33	145	178
1	70	139	209
2	49	161	210
	207	586	793

-2	0.29	0.10
-1	3.90	1.38
1	4.37	0.58
2	0.62	0.22

$$\chi^2 = 0.29 + 0.10 + 3.90 + 1.38 + 4.37 + 0.58 + 0.62 + 0.22 = 11.46,\ df = 3;\ {}_3\chi^2_{0.995} = 11.34$$

By leaving the data alone, a p-value < 0.005 was evaluated as a rejection of the null hypothesis. Although, by combining the information before ceremony and weeks after into two groups. χ^2 became 2.55 with df = 1, an acceptance of the null hypothesis (H_0: Weeks Prior to Ceremony = Weeks after Ceremony) with a p-value > 0.10.

-2/-1	88	286	374
1/2	119	300	419
	119	300	

-2/-1	0.949	0.458	793
1/2	0.847	0.299	
	0.949	0.458	

3.

Severity	
Moderate-Advanced	0.847 0.299
Minimal	$\chi^2 = 2.55$

ABO System

Severity	O	A	AB	B
Moderate-Advanced	7	5	3	13
Minimal	27	32	8	18
Not Present	55	50	7	24

$df = 6$, $X^2 = 20.20$, $\chi^2(0.995) = 18.55$

O	A	AB	B
1.13	2.64	0.38	11.33
0.31	0.81	0.98	1.07
0.37	0.05	1.03	0.10

MN System

Severity	MM	MN	NN
Moderate Advanced	21	6	1
Minimal	54	27	5
Not Present	74	51	11

$df = 4$, $X^2 = 12.05$, $\chi^2(0.995) = 14.86$

MM	MN	NN
[illegible]	[illegible]	0.14
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

$H_0: \pi_{ij} = \pi_{i.}\pi_{.j}$ (Independent Datasets) $H_1 = \pi_{ij} \neq \pi_{i.} \cdot \pi_{.j}$ (Dependent)

Each of the datasets test $X^2 > \chi^2_{0.995}$, which implies disease severity is dependent on blood type.

4.

	Male	Female	
Mentioned	86	55	141
Not Mentioned	283	360	643
	369	415	784

$df = 1$; $X^2 = 13.38$; $\chi^2(0.995) = 7.88$

5.81	5.17
1.27	1.13

$H_0: \pi_{i1} = \pi_{i2}$; $H_1: \pi_{i1} \neq \pi_{i2}$

The null hypothesis is rejected at a p-value < 0.005, that implies males and females do attribute differently

5.

Ethnic Origin	Yes	No	
Italian	78	47	125
Northern European	56	29	85
Other European	43	29	72
English	53	32	85
Irish	43	30	73
French Canadian	36	22	58
French	42	23	65
Portuguese	29	7	36
	380	219	599

0.02	0.04
0.03	0.01
0.02	0.01
0.02	0.03
0.02	0.03
0.02	0.03
0.01	0.03
0.02	0.03

$df = 7$, $X^2 = 0.42$

$\chi^2_{7, 0.005} = 0.989$

$H_0 = \pi_{i1} = \pi_{i2}$

$H_1 = \pi_{i1} \neq \pi_{i2}$

The hypothesis is accepted at a p-value > 0.95, suggesting each ethnic origin equally responded 'yes' and 'no' to the questionnaire

6.

Father's Activity	Female Offspring	Male Offspring	
Flying Fighters	51	38	89
Flying Transports	14	16	30
Not Flying	38	46	84
	103	100	203

0.756	0.779
0.931	0.101
0.501	0.526

$X^2 = 3.64$, $df = 2$, $\chi^2_{0.90} = 4.61$

$H_0: \pi_{i1} = \pi_{i2}$; $H_1: \pi_{i1} \neq \pi_{i2}$; The null hypothesis was accepted with this information at a p-value > 0.10, implying the offspring genders were independent to the Father's activity.

7.

Grade	Psychology	Biology	Other	
A	8	15	13	36
B	14	19	15	48
C	15	4	7	26
D-F	3	1	4	8
	40	39	39	118

1.45	0.81	0.10
0.321	0.62	0.05
4.34	2.45	0.30
0.03	1.02	0.695

$df = 6$, $X^2 = 13.19$

${}_6\chi^2_{0.975} = 14.45$

$H_0: \pi_{i1} = \pi_{i2} = \pi_{i3}$; $H_1: \pi_{i1} \neq \pi_{i2} \neq \pi_{i3}$; The analysis of the data suggests a rejection of the null hypothesis in favor of the alternative, and grades dependendeny to major.

8.

	# Patients	Incidence	
Placebo	165	95	260
Chlorpromazine	152	52	204
Dimenhydrinate	85	52	137
Pentobarbital (100mg)	67	35	102
Pentobarbital (150mg)	85	37	122
	554	271	825

Tabulated Abimation tests to be a null hypothesis positive. The expected cases of patients is independency of Incidence, implying none of the medications were effective to to Placebo.

0.53	1.09
1.64	3.36
0.53	1.09
0.03	0.07
0.12	0.24

$df = 4$

$X^2 = 8.69$

${}_4\chi^2(0.95) = 9.49$

9.

Austen	Imitator
24	2
273	81
26	1
424	153
14	17
486	204
1247	458

$df = 5$, $X^2 = 3.04$

1.31	3.56
0.77	2.09
1.98	5.39
0.009	0.03
3.32	9.03
0.69	1.99

${}_5\chi^2(0.995) = 16.75$

$H_0: \pi_{i1} = \pi_{i2}$

$H_1: \pi_{i1} \neq \pi_{i2}$

The null hypothesis is rejected at a p-value of 0.005. Austen's Imitator did not write to similar standards as her original works.

10. Chi-Square Test of Independence:

Contingency Table:

$\pi_{i.} = \sum_j \pi_{ij}$; $\pi_{ij} = \pi_{i.}\pi_{.j}$; $H_0: \pi_{ij} = \pi_i \pi_j$; MLE H_0: $\hat{\pi}_{ij} = \hat{\pi}_{i.}\hat{\pi}_{.j} = \frac{n_{i.}}{n} \times \frac{n_{.j}}{n}$

$\pi_{.j} = \sum^{I} \pi_{ij}$

McNemars Test: $\hat{\pi}_{12} = \hat{\pi}_{21} = \frac{n_{12}+n_{21}}{2n}$

$$X^2 = \frac{(n_{12}-n_{21})^2}{n_{12}+n_{21}}$$

Derived From Multinomial:

$$lik(\pi_1, \pi_2 \ldots \pi_\ell) = \prod_{j=1}^{J} \binom{n_j}{n_{1j} n_{2j} \ldots n_{ij}} \pi_1^{n_{1j}} \pi_2^{n_{2j}} \ldots \pi_\ell^{n_{\ell j}}$$

$$l(\pi, \lambda) = \sum \log \binom{n_{.j}}{n_{1j} n_{2j} \ldots n_{ij}} + \sum_i^\ell n_i \log \pi_i + \lambda\left(\sum_i^\ell \pi_i - 1\right)$$

$\frac{dl}{d\pi_i} = \frac{n_i}{\pi_i} + \lambda$; $i = 1 \ldots \ell$

$\hat{\pi}_i = \frac{-n_i}{\lambda}$; $\sum \pi_i = 1$; $\sum n_i = n$

$\lambda = -n$

11. a) Likelihood Ratio Test for Homogeneity:

$$\Lambda = \frac{P(\theta|H_0)}{P(\theta|H_1)} = \frac{P(\hat{\theta})}{\hat{P}(\theta)} = \frac{\frac{n!}{x_1! \cdots x_m!} P_1(\hat{\theta})^{x_1} \cdots P_m(\hat{\theta})^{x_m}}{\frac{n!}{x_1! \cdots x_m!} \hat{P}_1^{x_1} \cdots \hat{P}_m^{x_m}}$$

$$= \prod^{m} \left(\frac{P_i(\hat{\theta})}{\hat{P}_i}\right)^{x_i} ; \quad -2\log\Lambda = -2n \sum p_i \log\left(\frac{P_i(\hat{\theta})}{\hat{P}_i}\right) = 2\sum\sum O_i \log\left(\frac{O_i}{E_i}\right) \quad \boxed{\pi_i = \frac{n_i}{n}}$$

b) $-2\log\Lambda = 2\sum\sum O_i \log\left(\frac{O_i}{E_i}\right) = 2\sum\sum n_{ij} \cdot \log\left(\frac{n_{ij}^2}{n_{.j} \cdot n_{i.}}\right)$

c) Likelihood Ratio test for Independence:

$$\Lambda_{ij} = \Lambda_i \Lambda_j = \frac{P(\theta_i|H_0)}{P(\theta_i|H_1)} \cdot \frac{P(\theta_j|H_0)}{P(\theta_j|H_1)} = \frac{\frac{n!}{x_1! \ldots x_m!} P_1(\hat{\theta}_i)^{x_1} \cdots P_m(\hat{\theta}_i)^{x_m}}{\frac{n!}{x_1! \ldots x_m!} \hat{P}_1^{x_1} \cdots P_m^{x_m}} \cdot \frac{\frac{n!}{y_1! \ldots y_m!} P_1(\hat{\theta}_j)^{y_1} \cdots P_m(\hat{\theta}_j)^{y_m}}{\frac{n!}{y_1! \ldots y_m!} \hat{P}_1^{y_1} \cdots P_m^{y_m}}$$

$$\Lambda_{ij} = \prod^{m} \left(\frac{P_i(\hat{\theta})}{\hat{P}_i}\right)^{x_i} \cdot \prod^{m} \left(\frac{P_i(\hat{\theta})}{\hat{P}_i}\right)^{y_i}$$

$$-2\log\Lambda = 2\sum\sum O_i \cdot \log\frac{O_i}{E_i} \ldots \quad \hat{\pi}_{ij} = \frac{n_i}{n} \cdot \frac{n_j}{n} = \boxed{\frac{n_{ij}}{n^2}}$$

d) $2\log\Lambda = 16.52$; $\boxed{X^2_{11} , 16.52}$

12. McNemar's Test:

$\hat{\pi}_{11} = \frac{n_{11}}{n}$; $\hat{\pi}_{22} = \frac{n_{22}}{n}$; $\hat{\pi}_{12} = \hat{\pi}_{21} = \frac{n_{12}+n_{21}}{2n}$

$X^2 = \frac{(n_{12}-n_{21})^2}{n_{12}+n_{21}}$; $H_0: \pi_{12} = \pi_{21}$

Example Dataset representing marker of 0/1

0	1
1	0

Test	value
McNemars	0
t-test	0

13. a) A test of independence would entail a hypothesis ($H_0: \pi_{ij} = \pi_2 \frac{n_{ij}}{n}$) and a chi-squared statistic of columns of younger vs older with rows of children. If $X^2 < X^2_{1,.95}$ then the age of the sister is independent to number of children.

b) A test of family size distribution for sisters is described by ($H_0: \pi_{.1} = \pi_{.2}$) with column of potentially number of children.

14. No high school Education.

Degrees of Interest	Under 45	Over 45
Great	71	217
Little	305	652

Some high school or more

Degrees of Interest	Under 45	Over 45
Great	305	180
Little	869	259

a. No highschool Education: $\chi^2 = 5.54$; $\chi^2_{0.95} = 3.84$

Some high school Education: $\chi^2 = 34.58$; $\chi^2_{1,0.95} = 3.84$

Each analysis of political interest generated results to reject the null hypothesis and accept the alternative that age is an indicator for political interests.

b. H_1: Given educational level, age, and degree of interest are unrelated.

H_2: Given age, educational level and degree of interest are unrelated.

By adding the columns of educational level;

	No school	Some school
Great	288	485
Little	957	1128

$\chi^2 = 17.12$
Reject $H_0: \pi_1 = \pi_2$
Accept H_1.

-By adding the rows of age,

	Under 45	Over 45
No School	376	869
Some School	1174	439

$\chi^2 = 513$
Reject H_0, Accept H_2.

15.

Source	χ^2	df	$\chi^2_{.95}$	p-value
Incidence Among Males	13.54	4	9.49	0.008
Incidence Among Females	13.471	4	9.49	0.02

The number of females who grew a tumor was ~~highly~~ significant because the expected values of predicted tumors correlated to the observed outcome. Male tumor data has a p-value of 0.008 and does present significant investigation at level of 0.95. There is

16. A Chi-squared analysis ($\chi^2 = 27.29$, df = 4) confirms rejection of the null hypothesis ($H_0: \pi_{1.} = \pi_{2.} = \pi_{3.}$) and accepts the alternative hypothesis ($H_1: \pi_{1.} \neq \pi_{2.} \neq \pi_{3}$) at a significance level of 0.95. There is a relationship between personality type and attitude toward small cars, explorers have an unfavorable attitude.

17. London ($\chi^2 = 42.4$, df = 1) and Manchester ($\chi^2 = 5.5$, df = 1) data conclude different results upon the relationship between blood type and peptic ulcers. The city of London contains a relationship between blood type to disease, while Manchester does not. A p-value of <0.00 for London [illegible] reason for investigation to the source

18.

Test	χ^2	df	$\chi^2_{0.95}$	H_0	H_1	Suggestion	p-value
Matched Pair Design	75.03	1	3.84	$\pi_{12} = \pi_{21}$	$\pi_{12} \neq \pi_{21}$	Reject H_0	0.14
Homogeneity	0.03	11	3.84	$\pi_{.1} = \pi_{.2}$	$\pi_{.1} \neq \pi_{.2}$	Accept H_0	1.00

Estrogen is correlated to cancer.

19. Fischers Exact Test:

$$p(n_{11}) = \frac{\binom{n_{1.}}{n_{11}}\binom{n_{2.}}{n_{21}}}{\binom{n_{..}}{n_{.1}}}$$

n_{11}	$p(n_{11})$
5	0.03
6	0.06
7	0.09
8	0.12
9	0.12
10	0.10
11	0.06
12	0.03
13	0.01
14	0.00
15	0.000

The ~~Fisher~~ Exact ~~Test~~ presents no significant difference between the high and low anxiety groups at $n_{11} = 12$.

20. Random Sample | Prospective Study | Retrospective Study

a) $P(D|X) = \frac{\pi_{11}}{\pi_{10}+\pi_{11}}$ $\quad P(D|X) = \frac{n_{11}}{n_{10}+n_{11}}$ $\quad P(X|D) = \frac{n_{10}}{n_{00}}$

$P(D|\bar{X}) = \frac{\pi_{01}}{\pi_{00}+\pi_{01}}$ $\quad P(D|\bar{X}) = \frac{n_{01}}{n_{00}+n_{01}}$ $\quad P(X|\bar{D}) = \frac{n_{10}}{n_{00}}$

$\Delta|X) = \frac{odds(D|X)}{odds(D|\bar{X})}$ $\quad \Delta = \frac{odds(D|\bar{X})}{odds(D|\bar{X})}$ $\quad \Delta = \frac{odds(D|X)}{odds(D|\bar{X})}$

$= \frac{P(D|X)}{1-P(D|X)}\frac{1-P(D|\bar{X})}{P(D|\bar{X})}$ $\quad = \frac{P(D|X)}{1-P(D|X)}\frac{1-P(D|\bar{X})}{P(D|X)}$ $\quad = \frac{P(D|X)}{1-P(D|X)}\frac{1-P(D|\bar{X})}{P(D|\bar{X})}$

$\boxed{= \frac{\pi_{11}\pi_{00}}{\pi_{01}\pi_{10}}}$ $\quad \boxed{= \frac{n_{11}}{n_{01}}\left(\frac{n_{00}}{n_{10}}\right)}$ $\quad \boxed{= \frac{n_{11}}{n_{01}} \cdot \frac{n_{00}}{n_{10}}}$

b) Method of Propagation Error: $Y = g(x) \approx g(\mu_x) + (X-\mu_x)g'(\mu_x)$

$\pi_{11}+\pi_{10} \quad \pi_{11} \quad \pi_{00}$

$\mu_y = g(\mu_x)$

$\sigma_y^2 = \sigma_x^2 [g'(\mu_x)]^2$

Random Sample | Prospective Study | Retrospective Study

$\log\Delta = \log\frac{\pi_{11}\pi_{00}}{\pi_{01}\pi_{10}}$ $\quad \log\Delta = \log\frac{n_{11}}{n_{01}}\cdot\frac{n_{00}}{n_{10}}$ $\quad \log\Delta = \log\frac{n_{11}}{n_{01}}\cdot\frac{n_{00}}{n_{10}}$

$Var\log(\Delta) = \frac{\sum_i^1(X-\mu_X)^2}{n}\circ\log'(\Delta)$ $\quad Var\log\Delta = \frac{\sum_i^1(X-\mu_X)^2}{n}\circ\log'\Delta$ $\quad Var\log\Delta = \frac{\sum_i^1(X-\mu)^2}{n}\log\Delta$

$= \frac{\sum\sum(\pi_{ij}-\mu)^2}{m\cdot n}\circ\frac{1}{\Delta^2}$ $\quad = \frac{\sum\sum(\pi_{ij}-\mu)^2}{m\cdot n}\circ\frac{1}{\Delta}$ $\quad = \frac{\sum\sum(\pi_{ij}-\mu)^2}{m\cdot n}\circ\frac{1}{\Delta}$

$\boxed{= \frac{\sigma^2}{\Delta^2}} = \sigma^2\left(\frac{\pi_{01}\pi_{10}}{\pi_{11}\pi_{00}}\right)^2$ $\quad \boxed{= \frac{\sigma^2}{\Delta^2}} = \sigma^2\left(\frac{n_{01}n_{10}}{n_{11}n_{00}}\right)^2$ $\quad \boxed{= \frac{\sigma^2}{\Delta} = \sigma^2\left(\frac{n_{01}n_{10}}{n_{11}n_{00}}\right)^2}$

21.

	Diabetic	Normal
Bb or bb	12	4
BB	39	49

$P(D|X) = \frac{49}{39+49}$; $odds(D|\bar{X}) = \frac{49}{39}$; $\Delta = \frac{49\cdot12}{39\cdot4}$

$P(D|\bar{X}) = \frac{12}{12+4}$; $odds(D|\bar{X}) = \frac{12}{4}$ (12+14) ; $\boxed{= 3.77}$

The odds of being normal (Dominant) to Diabetic (Dominant-recessive) is 3.77:1.

22. a

	Advised	Not Advised
Male	48	47
Female	80	136

Proportion of males advised to quit: $\frac{48}{48+47} = 50.6\%$

Proportion of females advised to quit: $\frac{80}{80+136} = 37.7\%$

Standard error of male proportion: 20.05

Standard error of female proportion: 1.91

Standard error of their difference: 90 ; Significant with t-test

	Advised	Not Advised
White	26	34
African-American	102	149

Proportion of Whites asked to quit: 43%

Proportion of African-Americans asked to quit: 41%

Standard error of White proportion: 0.6

Standard error of African-American proportion: 60 ; Significant t-test

	Advised	Not Advised
<15	64	112
15-25	39	54
>25	25	16

Proportion of smokers who smoke <15, advised to quit: 36%
Proportion of smokers, 15-25, advised to quit: 42%
Proportion of smokers who smoke >25, advised to quit: 61%
Standard error of smokers (<15): 1.8
Standard error of smokers (15-25): 0.77
Standard error of smokers (>25): 0.70

The difference in proportion are significant at an F-stat, one-sided.

b)

	Advised	Not Advised
Male	78	94
Female	50	89

Proportion of male physicians who advised: 45%
Proportion of female physicians who advised: 36%
Standard error of male physicians who advised: 0.61
Standard error of female physicians who advised: 0.65
Standard error of the difference: ∞ ; $t < t(0.95)$
$F < F_{crit}$ $\boxed{\text{Significant results}}$

	Advised	Not Advised
Smoker	13	37
Nonsmoker	115	146

Proportion of smokers advised: 26%
Proportion of nonsmokers advised: 44%
Standard error of smokers advised: 1.70
Standard error of smokers not advised: 0.96
Standard error of the difference: ∞ ; $t < t(0.95)$
$F < F_{crit}$ $\boxed{\text{significant results}}$

	Advised	Not Advised
<30	88	128
30-39	28	37
>39	12	18

Proportion of physicians age (<30): 41%
Proportion of physicians age (30-39): 43%
Proportion of physicians age (>39): 40%
Standard error of physicians (<30): 1.36
Standard error of physicians (30-39): 0.56
Standard error of physicians (>39): 0.55
Standard error of the difference: F-statistic - 17.07
F-crit : 9.55
$\boxed{\text{Not significant}}$

23.

Previous Day	Day of Infarction		
	Exertion	No Exertion	Total
Exertion	4	9	13
No exertion	50	1165	1215
Total	54	1174	1228

McNemar's Test:

$H_0: \pi_{12} = \pi_{21}$

$H_1: \pi_{12} \neq \pi_{21}$

$$\chi^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} = 28.49, \; df = 1; \; \chi^2_{0.95} = 3.84$$

The results by McNemar's Test are significant. Exertion is associated to infarction.

24.

Sport	Red	Blue
Boxing	148	120
Freestyle Wrestling	27	24
Greco Roman Wrestling	25	23
Tae Kwon Do	45	35

a. π_R = contestant in red wins

$= \frac{n_{15}}{n} \pm 0.55$

$\chi^2 = 0.30$

df = 3

$H_0: \pi_R = 1/2$

$H_1: \pi_R \neq 1/2$

$\chi^2 = 7.81$

There appears to be no significant relationship between sport and color. The means of red and blue clothes are roughly equivalent.

b. $H_0: \pi_R = 1/2$ $\chi^2 = 0.30$; $\chi^2 = 7.81$

$H_1: \pi_{SR} \neq 1/2$ $df = 3$ $\boxed{\text{Not significant between sports}}$

c. Both of the tests are an equivalent argument with a chi-squared statistic.

d. There is little evidence to suggest red is favorable to the wins of sports.

e.

Sport	Red	Blue
Boxing	123	105
Free Wrestling	28	39
GR Wrestling	25	34
TaeKwonDo	27	26

$\chi^2 = 8.59$
$df = 3$
$\chi^2_{0.95} = 7.81$

Examination of larger datasets show a different conclusion; more specifically, color does effect the outcome of a match. The means are significant across wrestling.

25. a. Individual analysis of aspirin's effects were conducted for myocardial infarction, and stroke. Myocardial disease ($\chi^2 = 1.38, df = 1$) results were not significant, but aspirin did effect a stroke ($\chi^2 = 30.65, df = 1$).

b. Aspirin analysis significant results for lowering myocardial mortalities, but overall had no influence on death ($\chi^2 = 12.61$; $df = 4$).

26. McNemars Test evaluated DKA relationship to therapy. A chi-squared of 5.53, with one degree of freedom lists significant results for the relationship between side effects before and after therapy.

27. Upon examination, the defendants race does not influence death penalty. By including the victims race, the results could change the outcome, but not the analysis. The chi-square statistic ($\chi^2 = 15.92, df = 3$) does list significant results for a relationship of race to death penalty.

28. The chi-square test is independent to counts, frequency, or percents and presents similar outcomes

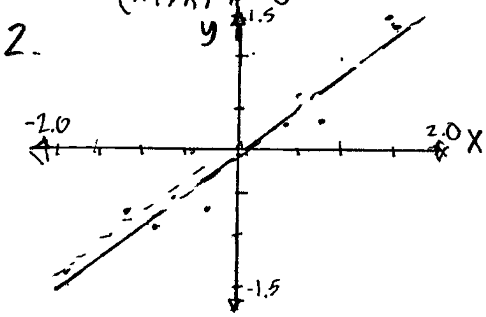
29. Model: Satisfied Somewhat Dissatisfied Very Dissatisfied
A statistician should carry out a test of homogeneity, rather than a test for independence because of the workers being already independent to each other.

Chapter 14: Linear Least Squares

1. a. $y = \frac{a}{(b+cx)}$; $\log y = \log a - \log(b+cx)$ b. $y = ae^{-bx}$; $\log y = -bx + \log a$ c. $y = ab^x$; $\log y = x \log b + \log a$

d. $y = \frac{x}{(a+bx)}$; $\frac{1}{y} = \frac{a}{x} + b$; $\frac{1}{y} = k$; $\frac{1}{x} = j$; $k = aj + b$ e. $y = \frac{1}{1+e^{bx}}$; $\log k - 1 = bx$ where $k = \frac{1}{y}$

2.



a. $\beta_0 = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - (\sum x)^2} = -0.03$

$\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = 0.904$

b. $x = c + dy$; $c = \frac{-\beta_0}{\beta_1}$; $d = \frac{1}{\beta_1}$; c. The lines are equivalent graphically because the inverse is being plotted.

3. $y = \mu + e_i$; $i = 1 \ldots n$; where e_i = independent errors $[\mu_e = 0, \sigma_e^2]$

$= \beta_0 + \beta \cancel{x}^{0} + e_i = \beta_0 + e_i$; $\beta_0 = \frac{(\sum x^2)(\sum y) - (\sum x)(\sum xy)}{n\sum x^2 - (\sum x)^2} = \frac{\mu(\sum x^2 - (\sum x)^2)}{\sum x^2 - (\sum x)^2}\frac{n}{n} = \mu$

4. GPA: $Y_i = \beta_0 + \beta_1 X_i + e_i$, $i = 1, 2 \ldots n$

Model: $Y_F = I_F(i)\beta_F + I_M(i)\beta_M + \beta_i x_i + e_i$; where $I_F(i)$ and $I_M(i)$ are indicator variables that are 0/1 for male or female.

Design Matrix: $\boxed{X = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}}$

5. $P_1 < P_2 < P_3$ a. $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$ b. $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix}$ c. $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix}$ $X = \begin{array}{c} (P_1 \quad P_2 \quad P_3)^T \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \end{array}$ $\beta = (X X)^{-1} X^T Y$

6. w_1 and w_2

1) Object 1 weighed by itself, 3g ; a) $X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ $Y = \begin{bmatrix} 3 \\ 3 \\ 1 \\ 7 \end{bmatrix}$ $e = Y - \hat{Y} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$

2) Object 2 weighed by itself, 3g

3) Object 1 and object 2, $\Delta w = 1g$.

4) $\sum w = 7g$

4 x 3 3 x 1

b) $S(\beta_0, \beta_1) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x)^2$; $S(w_1, w_2) = \sum_{i=1}^{4}(y_i - w_1 - w_2 x_i)^2 = (3-w_1)^2 + (3-w_2)^2 + (1-w_1+w_2)^2 + (7-w_1-w_2)^2$

$\frac{dS}{dw_1} = -2(3-w_1) - 2(1-w_1+w_2) - 2(7-w_1-w_2) = 0$; $11 - 3w_1 = 0$; $w_1 = \boxed{\frac{11}{3}}$

$\frac{dS}{dw_2} = -2(3-w_2) + 2(1-w_1+w_2) - 2(7-w_1-w_2) = 0$; $9 - 3w_2 = 0$; $w_2 = \boxed{3}$

c) $s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$; $s^2 = \frac{1}{4-2}\left[(3-\frac{11}{3})^2 + (3-3)^2 + (1-\frac{11}{3}+3)^2 + (7-\frac{11}{3}-3)^2\right]$; $\boxed{s^2 = 1/3}$

d. $\sigma_s^2 = (w_1 - w_2) s^2 = \frac{2}{3} \cdot \frac{1}{3} = \boxed{\frac{2}{9}}$

F. $H_0: w_1 = w_2$; $t_{obs} > t$; $2.353 > \sqrt{2}$; Accept null hypothesis.
$H_1: w_1 \neq w_2$

e. $t = \frac{w_1 - w_2}{\sqrt{\sigma_s^2}} = \frac{2/3}{\sqrt{2/9}} = \boxed{\sqrt{2}}$

7. $y_i = \beta_0 + \beta_1 x_i + e_0$; $Var(e_i) = \rho_i^2 \sigma^2$; $\rho_i^{-1} y_i = \rho_i^{-1}\beta_0 + \rho_i^{-1}\beta_1 x_i + \rho_i^{-1} e_i$ or $z_i = u_i\beta_0 + v_i\beta_1 + \delta_i$
where $u_i = \rho_i^{-1}$; $v_i = \rho_i^{-1} x_i$; $\delta_i = \rho_i^{-1} e_i$

a. $z_i = u_i\beta_0 + v_i\beta_1 + \delta_i$; $\frac{y_i}{\rho_i} = \frac{\beta_0}{\rho_i} + \frac{x\beta_1}{\rho_i} + \frac{e_i}{\rho_i}$; $y_i = \beta_0 + x\beta_1 + e_i$

b. $S(\beta_0, \beta_1) = \sum_{i=1}^{n} (z_i - u_i\beta_0 + v_i\beta_1 + \delta_i)^2$; $\frac{dS}{d\beta_0} = -2\sum u_i (z_i - u_i\beta_0 + v_i\beta_1 + \delta_i) = 0$; $\sum z_i = \beta_0 \sum u_i + \beta_1 \sum v_i$

$\frac{dS}{d\beta_1} = -2\sum v_i (z_i - u_i\beta_0 + v_i\beta_1 + \delta_i) = 0$; $\sum v_i z_i = (\beta_0 \sum u_i + \beta_1 \sum v_i) v_i$

$\sum z_i = \frac{\beta_0 n}{\rho_i} + \frac{\beta_1}{\rho_i} \sum x_i$; $\rho_i \sum z_i = n\beta_0 + \beta_1 \sum x_i$

$\sum v_i z_i = \beta_0 \sum v_i + \beta_1 \sum v_i^2$; $\rho_i \sum v_i z_i = \beta_0 \sum x + \beta_1 \sum x^2$

$$\boxed{\beta_0 = \frac{\rho_i \sum x_i^2 \cdot \sum y_i - \rho_i \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2}}$$

$$\boxed{\beta_1 = \frac{n\rho_i \sum x_i y_i - \rho_i (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2}}$$

c. $\frac{d}{d\beta_0} \sum (y_i - \beta_0 - \beta_1 x_i)^2 \rho_i^{-2} = -2\sum (y_i - \beta_0 - \beta_1 x_i) \cdot \rho_i^{-2} = 0$; $\beta_0 = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2}$

$\beta_1 = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2}$

d. $\boxed{Var(\beta_0) = \frac{\sigma^2 \sum x_i^2 \cdot \rho_i^2}{n\sum x_i^2 - (\sum x_i)^2}}$; $\boxed{Var(\beta_1) = \frac{n\sigma^2 \rho_i^2}{n\sum x_i^2 - (\sum x_i)^2}}$

8. $\underset{(n\times p)}{\mathbb{X}} = \underset{(n\times p)}{Q} \cdot \underset{(p\times p)}{\mathbb{R}}$ where $Q^T Q = I$; $\mathbb{R} = (r_{ij} = 0, i > j)$; Show $\hat\beta = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{Y}$ is $\hat\beta = \mathbb{R}^{-1} Q^T \mathbb{Y}$

$\mathbb{Y} = \mathbb{X}\beta$; $\mathbb{X}^T\mathbb{Y} = (\mathbb{X}^T\mathbb{X})\beta$; $\boxed{(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{Y} = \hat\beta}$

$= (Q^T R^T Q R)^{-1} Q^T R^T Y$

$= (Q^T R^T Q)^{-1} \cdot R^{-1} Q^T \cdot R^T \mathbb{Y}$

$\boxed{= \mathbb{R}^{-1} Q^T \mathbb{Y}}$

The last equation ($\mathbb{R}\hat\beta = Q^T\mathbb{Y}$) may be solved by back-substitution because there are p variables for p equations

9. Cholesky Decomposition $\mathbb{X}^T\mathbb{X} = \mathbb{R}^T\mathbb{R}$ where $R = (r_{ij} = 0, i > j)$

Show $R^T v = \mathbb{X}^T Y$; $\mathbb{R}\hat\beta = v$; Show $\mathbb{R}^T v = \mathbb{X}^T Y$ is solved with Back-substitution because of the dimension $(p\times p)(p\times 1)$, indicating p variables and p equations, which is similar to $\mathbb{R}\hat\beta = v$.

$\mathbb{R}^T\mathbb{R}\hat\beta = \mathbb{X}^T\mathbb{Y}$

$\boxed{\hat\beta = \mathbb{X}^T\mathbb{Y}}$

10. $\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$ and $\beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$; $\bar{y} = \beta_0 + \beta_1 \bar{x}$; $\boxed{\beta_0 = \bar{y} - \beta_1 \bar{x}}$

11. $Corr(\beta_0, \beta_1) = \frac{Cov(\beta_0, \beta_1)}{\sqrt{Var(\beta_1) \cdot Var(\beta_0)}}$

$$= \frac{\frac{-\sigma^2 \sum x_i}{n \sum x_i^2 - (\sum x_i)^2}}{\sqrt{\frac{\sigma^2 \sum x_i^2}{n \sum x_i^2 - (\sum x_i)^2} \frac{n\sigma^2}{n \sum x_i^2 - (\sum x_i)^2}}}$$

@ $\bar{x} = 0$; $Corr(\beta_0, \beta_1) = 0$

$\beta_1 = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$

$\boxed{= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}}$ $(xy - \bar{x}y + y^2$

$x y + x_i \cancel{\bar{y}_i} + \bar{x}\bar{y}, \cancel{\bar{x}\bar{y}}$

12. $\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$

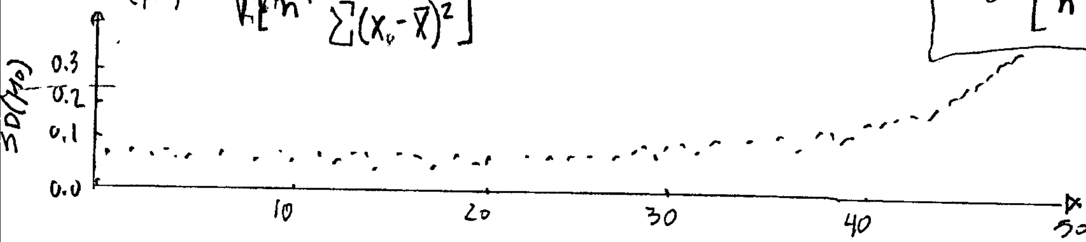
$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

@ $(\bar{x}, \bar{y})$; $\boxed{\bar{y} = \beta_0 + \beta_1 \bar{x} = \beta_0 + \beta_1 \bar{x} = \bar{y} - \beta_1 \bar{x} + \beta_1 \bar{x} = \bar{y}}$

13. a) $\mu_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$; $Var(\mu_0) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_0) = Var(\beta_0) + Var(\hat{\beta}_1 x_0) = \frac{\sigma^2 \sum (x_0 - \bar{x}_0)^2}{n \sum (x_0 - \bar{x})^2} + \frac{\sigma^2 (x_0 - \bar{x})^2}{\sum (x_0 - \bar{x})^2}$

$$\boxed{= \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_0 - \bar{x})^2} \right]}$$

b) $SD(\mu_0) = \sigma \sqrt{\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_0 - \bar{x})^2} \right]}$



c. 95% confidence interval! $\boxed{\mu_0 \pm SD(\mu_0) \cdot t_{n-2}(\alpha/2)}$ Point number

14. $Y_0 = \beta_0 + \beta_1 x_0 + e_0$ by the estimate $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$

a) $Var(\hat{Y}_0 - Y_0) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_0 - \beta_0 - \beta_1 x_0 - e_0) = Var(\hat{\beta}_0 - \beta_0) + Var[(\hat{\beta}_1 - \beta_0) x_0] - Var(e_i)$

$= \cancel{Var(\hat{\beta}_0 - \beta_0) + Var[(\hat{\beta}_1 - \beta_0) x_0]}$

$$\boxed{= (\hat{\sigma}^2 - \sigma_0^2) \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_0 - \bar{x})^2} \right]}$$

b) $\boxed{(\hat{Y} - Y_0) \pm \sqrt{(\hat{\sigma}^2 - \sigma_0^2) \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_0 - \bar{x})^2} \right]} \cdot t_{n-2}(\alpha/2)}$

15. $y = \beta x$ to points (x_i, y_i) where $i = 1 \ldots n$

$\beta = (X^T X)^{-1} X^T y$ -or- $S = \sum (y_i - \beta x)^2$; $\frac{dS}{d\beta} = -2 \sum x_i (y - \beta x) = 0$; $\sum x_i y_i - \beta \sum x_i^2 = 0$; $\boxed{\beta = \frac{\sum x_i y_i}{\sum x_i^2}}$

16. a $y = \beta_0 x + \beta_1 x^2$; $X^T X = \begin{pmatrix} x \\ x^2 \end{pmatrix} (x \; x^2)$; $(X^T X)^{-1} = \frac{1}{\sum x^2 \sum x^4 - (\sum x^3)^2} \begin{bmatrix} \sum x^4 & -\sum x^3 \\ -\sum x^3 & \sum x^2 \end{bmatrix}$

$y = \beta x^2 \ldots x^3$; $\beta_1 = \begin{pmatrix} x^2 & x^3 \\ x^3 & x^4 \end{pmatrix}$

$= \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} (x \; x^2)$

$\frac{\ldots x}{\sum x}$

$= \begin{pmatrix} \sum x^2 & \sum x^3 \\ \sum x^3 & \sum x^4 \end{pmatrix}$

$(X^T Y) = \begin{pmatrix} \sum x^2 y & \sum x^3 y \\ \sum x^3 y & \sum x^4 y \end{pmatrix}$

$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \frac{\begin{pmatrix} \sum x^4 & -\sum x^3 \\ -\sum x^3 & \sum x^2 \end{pmatrix} \begin{pmatrix} \sum x^2 y & \sum x^3 y \\ \sum x^3 y & \sum x^4 y \end{pmatrix}}{\sum x^2 \sum x^4 - (\sum x^3)^2}$

$$\frac{\begin{pmatrix} \sum x^4 \sum x^3 y - \sum x^3 y \sum x^3 & \sum x^4 \sum x^3 y - \sum x^3 \sum x^4 y \\ \sum x^4 \sum x^3 y - \sum x^3 \sum x^4 y & \sum x^3 \sum x^3 y + \sum x^2 \sum x^4 y \end{pmatrix}}{\sum x^2 \sum x^4 - (\sum x^3)^2} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$\beta_0 = \frac{\sum x^4 \sum x^3 y - \sum x^3 y \sum x^3 + \sum x^4 \sum x^3 y - \sum x^3 \sum x^4 y}{\sum x^2 \sum x^4 - (\sum x^3)^2} = \boxed{\frac{2\sum x^4 \sum x^3 y - \sum x^3 (\sum x^3 y - \sum x^4 y)}{\sum x^2 \sum x^4 - (\sum x^3)^2}}$$

$$\beta_1 = \frac{\sum x^4 \sum x^3 y - \sum x^3 \sum x^4 y + \sum x^3 \sum x^3 y + \sum x^2 \sum x^4 y}{\sum x^2 \sum x^4 - (\sum x^3)^2} = \boxed{\frac{2\sum x^3 y (\sum x^4 + \sum x^3) + (\sum x^2 - \sum x^3) \sum x^4 y}{\sum x^2 \sum x^4 - (\sum x^3)^2}}$$

b. $\boxed{Cov(\beta_0, \beta_1) = \frac{\sum x^4 \sum x^3 y - \sum x^3 \sum x^4 y}{\sum x^2 \sum x^4 - (\sum x^3)^2} \sigma^2}$

17. $E(X) = \mu_X$ $\quad E(Y) = \mu_y$ $\quad$ a) $E(Y-\hat{Y})^2 = E(Y)^2 - 2E(Y\hat{Y}) + E(\hat{Y})^2 = \mu_y^2 - 2E(Y \cdot (\alpha + \beta X)) + E((\alpha + \beta X)^2)$

$Var(X) = \sigma_X^2$ $\quad Var(Y) = \sigma_y^2$

$$= \mu_Y^2 - 2\alpha \mu_y - 2\beta Cov(X,Y) + E(\alpha)^2 + 2E(\alpha \beta X) + E(\beta^2 X^2)$$

$$= \mu_Y^2 - 2\alpha \mu_y - 2\beta \sigma_{xy} + \alpha^2 + 2\alpha\beta\mu_x + \beta^2 \sigma_x^2$$

$$\frac{dE(Y-\hat{Y})^2}{d\alpha} = -2\mu_y + 2\alpha + 2\beta\mu_x = 0; \quad \boxed{\alpha = \mu_y - \beta\mu_x}$$

$$\frac{dE(Y-\hat{Y})^2}{d\beta} = -2\sigma_{XY} + 2\beta\sigma_x^2; \quad \beta = \frac{\sigma_{xy}}{\sigma_x^2}$$

b. Prove $\frac{Var(Y) - Var(Y-\hat{Y})}{Var(Y)} = r_{xy}^2$; $\frac{Var(Y) - Var(Y-\hat{Y})}{Var(Y)} = \frac{\sigma_y^2 - \sigma_y^2 + \cancel{\emptyset} + \beta^2 \sigma_x^2}{Var(Y)} = \boxed{\frac{\sigma_{xy}^2}{Var(Y)\, Var(X)}}$

18. $Y_i = \beta_0 + \beta_1 x_i + e_i$ where $i = 1 \ldots n$, $e_i = N(\mu = 0, \sigma^2)$; find the mle's of β_0 and β_1.

$S(\beta_0, \beta_1) = \sum (y_i - \beta_0 - \beta_1 x_i + e_i)^2$; $\frac{dS}{d\beta_1} = -2\sum (y_i - \beta_0 - \beta_1 x_i + e_i) \sum x_i = 0$; $\sum x_i y_i - \beta_0 \sum x_i - \beta_1 (\sum x_i^2) = 0$

$$\sum x_i y_i = \beta_0 \sum x_i + \beta_1 (\sum x_i^2)$$

$$\frac{dS}{d\beta_0} = -2\sum (y_i - \beta_0 - \beta_1 x_i + e_i) = 0; \quad \sum y_i = n\beta_0 + \beta_1 \sum x_i$$

$$n \sum x_i y_i = \sum x \sum y - \beta_1 (\sum x_i)^2 + n\beta_1 \sum x_i^2 = \sum x \sum y + \beta_1 (\sum x_i^2 - (\sum x_i)^2)$$

$$\boxed{\beta_1 = \frac{n\sum x_i y_i - \sum x \sum y}{n \sum x_i^2 - (\sum x_i)^2}}$$

$$\sum y_i = n\beta_0 + \frac{n\sum x \sum x_i y_i - (\sum x_i)^2 \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$\sum y_i = \frac{n\beta_0 [n\sum x_i^2 - (\sum x_i)^2] + n\sum x \sum xy - \sum x_i^2 \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$\frac{n\sum y_i \sum x_i^2 - \sum y_i (\sum x_i)^2}{} = \beta_0 [n\sum x_i \sum x_i^2 - (\sum x)^3] + n\sum x \sum xy - \sum x_i^2 \sum y_i$$

$$\boxed{\frac{n\sum x_i^2 \sum y_i - n\sum x \sum xy}{n\sum x_i^2 - (\sum x)^2} = \beta_0}$$

19. a. Vector of Residuals: $\hat{e} = Y - \hat{Y} = (I-P)Y$

$\hat{e}XX^T = \begin{vmatrix} Y_1 - X_1\beta \\ Y_2 - X_2\beta \end{vmatrix} [X_1 \; X_2]^T = Y_1X_1 - \cancel{X_1\beta X_1} + Y_2X_2 - \cancel{X_2\beta X_2}$

$= Y_1X_1 - Y_1X_1 + Y_2X_2 - Y_2X_2$

$\boxed{=0}$

b. $\sum_{i=1}^{2} e_i = Y_1 - X_1\beta + Y_2 - X_2\beta = Y - \hat{Y} = 0$

20. Assume X, X_1, X_p are orthogonal; that is $X_i^T X_j = 0$ for $i \neq j$

If $X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}$; Covariance matrix of a least squares estimate:

$\Sigma_{\beta\beta} = \sigma^2 (X^T X)^{-1} = \sigma^2 \left(\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \right)^{-1} = \sigma^2 \begin{bmatrix} X_1^2 & 0 \\ 0 & X_2^2 \end{bmatrix}^{-1}$

$\boxed{= \sigma^2 \begin{bmatrix} \frac{1}{X_1^2} & 0 \\ 0 & \frac{1}{X_2^2} \end{bmatrix}}$

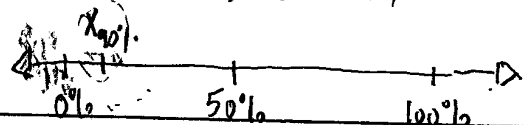
$AI = A$

$\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} =$

21. $Var(\beta_1) = \dfrac{\sigma^2}{\sum_i^n (X_i - \bar{X})^2}$

To minimize the variance of β_1, the ~~average~~ value $\sum_i^n (X_i - \bar{X})^2$ should be close to zero.

22. Family Income $\propto$ consumption;



Less than 50% of families would be at the 90% percentile of consumption when the total proportion is above the 90% of family income.

23a. $\rho = \dfrac{cov(Exam, midterm)}{\sqrt{Var(Exam)}\sqrt{Var(Midterm)}} = 0.5$; $\bar{X} = \dfrac{Exam + midterm}{2} = 75$; $\sigma = \sqrt{\dfrac{(Exam-75)^2 + (midterm-75)^2}{2}} = 10$

If student scores 95 on the midterm, what is the final exam?

$\frac{95+75}{2} = \boxed{85}$

b. If a student scored 85 on the final, what would you guess her score on the midterm was?

$\frac{85+75}{2} = \boxed{80}$

24. $\hat{Y} = X\beta = KX\beta = [K_1 \; K_2 \cdots K_n] \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_n \end{bmatrix} [\beta_1 \; \beta_2 \cdots \beta_n] = \begin{bmatrix} K_1X_1 & 0 & 0 \\ 0 & K_2X_2 & 0 \\ 0 & 0 & K_nX_n \end{bmatrix} [\beta_1 \; \beta_2 \cdots \beta_n] = U\beta$

The linear least squares constant β does not change

25. Y_{i1}, Y_{i2} are regressive mean responses to a linear least squares because the process is independent of grouping data points.

26. Z_1, Z_2, Z_3, Z_4 are random variables with $Var(Z_i) = 1$ and $Cov(Z_i, Z_j) = \rho$ for $i \neq j$

Prove $Z_1 + Z_2 + Z_3 + Z_4$ is uncorrelated with $Z_1 + Z_2 - Z_3 - Z_4$

If $Y_1 = Z_1 + Z_2 + Z_3 + Z_4$ and $Y_2 = Z_1 + Z_2 - Z_3 - Z_4$

$COV(Y_1, Y_2) = var(Z_1) + \cancel{cov(Z_1, Z_2)} - \cancel{cov(Z_1, Z_3)} - \cancel{cov(Z_1, Z_4)} + var(Z_2) - \cancel{cov(Z_2, Z_3)}$

$- \cancel{cov(Z_2, Z_4)} - var(Z_3) - cov(Z_3, Z_4) + \cancel{cor(Z_4, Z_1)} + \cancel{cov(Z_4, Z_2)}$

$- cov(Z_4, Z_3) - var(Z_4) + \cancel{Cov(Z_2, Z_1)} + \cancel{cov(Z_3, Z_1)} + \cancel{cov(Z_3, Z_2)}$

$= 1 + 1 - 1 - 2\rho - 1 = \boxed{2\rho}$

27. $\sigma^2 I = \Sigma_{\hat{Y}\hat{Y}} + \Sigma_{\hat{e}\hat{e}}$; Prove $n\sigma^2 = \sum_{i=1}^{n} Var(Y_i) + \sum_{i=1}^{n} Var(e_i)$ (ωω

p4? temp? prove [illegible]

$\Sigma_{YY} = \sigma^2 I$ $(\hat{z}\ \Sigma_{ee}\ z = \sigma^2 I$; $\boxed{\Sigma_{YY} + \Sigma_{ee} = \sigma^2 I}$

$n\sigma^2 = n\left(\Sigma_{YY} + \Sigma_{ee}\right) = \boxed{\sum_{i=1}^{n} Var(Y_i) + \sum_{i=1}^{n} var(e_i)}$

28. $Y = \sum a_i X_i$; $Var(Y) = \sigma^2$; $\bar{X} = \mu_i$

a) If $Z = \sum b_i X_i$; Use theorem D ; $Y = AX$; $Z = BX$; $\Sigma_{YZ} = A\Sigma_{XX}B^T$

Find $Cov(Y,Z) = \Sigma_{YZ} = A\Sigma_{XX}B^T = \sum a_i \Sigma_{XX} \sum b_i = \boxed{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} b_{ij} \sigma_{ij}}$

b) Theorem C of 14.4.1

$E(X^TAX) = trace(A\Sigma) + \mu^T A\mu$

$E(X_iX_j) = \sigma_{ij} + \mu_i\mu_j$

$E\left(\sum\sum X_iX_ja_{ij}\right) = \sum\sum \sigma_{ij}a_{ij} + \sum\sum \mu_i\mu_ja_{ij}$

$\boxed{= trace(A\Sigma) + \mu^T A\mu}$

$E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_iX_j\right) = \sigma_{ij} + \mu_i\mu_j$
$= \sigma + \mu^2$

29. $Cov(X_1, X_2) = 0$; $Var(X_1) = Var(X_2) = \sigma^2$; Use matrix methods to show $Y = X_1 + X_2$, $Z = X_1 - X_2$

Find $\Sigma_{YZ} = \cancel{Cov}(Y, Z) = Cov(X_1 + X_2, X_1 - X_2)$
$= Var(X_1) - Cov(X_1, X_2) + Cov(X_1, X_2) - Var(X_2) \boxed{= 0}$

30 $X_1 \ldots X_n$; $Var(X_i) = \sigma^2$; $Cov(X_i, X_j) = \rho\sigma^2$ for $i \neq j$; Find $Var(\bar{X})$;

$Var(\bar{X}) = Var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}Var(X_1 + \cdots + X_n) = \frac{1}{n^2}\left[Var(X_1) + \cdots + Var(X_n)\right] = \boxed{\frac{\sigma^2}{n}}$

31. $Z_{1\times 4}$; $\Sigma_{ZZ} = \sigma^2 I$; Let $V = Z_1 + Z_2 + Z_3 + Z_4$ and $Z = (Z_1 + Z_2) - (Z_3 + Z_4)$

Find $Cov(V, Z) = Cov(Z_1 + Z_2 + Z_3 + Z_4, (Z_1 + Z_2) - (Z_3 + Z_4))$
$= Var(Z_1) + Cov(Z_1, Z_2) - Cov(\cancel{Z_1}, Z_3) - Cov(\cancel{Z_1}, Z_4) + Var(Z_2) + Cov(Z_1, Z_2)$
$- Cov(\cancel{Z_2}, Z_3) - Cov(\cancel{Z_2}, Z_4) + Cov(\cancel{Z_3}, Z_1) + Cov(\cancel{Z_3}, Z_2) - Var(Z_3)$
$- Cov(Z_3, Z_4) + Cov(\cancel{Z_4}, Z_1) + Cov(\cancel{Z_4}, Z_2) - Cov(Z_4, Z_3) - Var(Z_4)$
$= 2Cov(Z_1, Z_2) - 2Cov(Z_3, Z_4) = \boxed{0}$

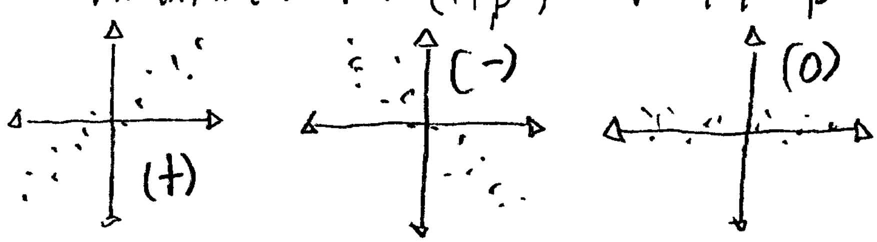
32. $Y_1 = X_1$, $Y_i = X_i - X_{i-1}$, $i = 1, 2, \ldots n$

a. $\Sigma_{YY} = \sum_{i=1}^{n} X_i - X_{i-1} = X_n = \boxed{\sigma^2 I}$ b. $\Sigma_{XX} = \sum Y_i = \boxed{\sigma^2 I}$

33. $X \sim N(0,1)$ and $E \sim N(0,1)$, let $Y = X + \beta E$

a. $r_{xy} = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{Var(X-\bar{X})(Y-\bar{Y})}{\sqrt{1\cdot(1+\beta^2)}} = \frac{Var((X-\bar{X})(X+\beta E - 0 + 0))}{\sqrt{1+\beta^2}} = \boxed{\frac{Var(X) + \beta Var(XE) - Var(\bar{X}X) - Var(\bar{X}\beta E)}{\sqrt{1+\beta^2}}} = \frac{1}{\sqrt{1+\beta^2}}$

b.



34

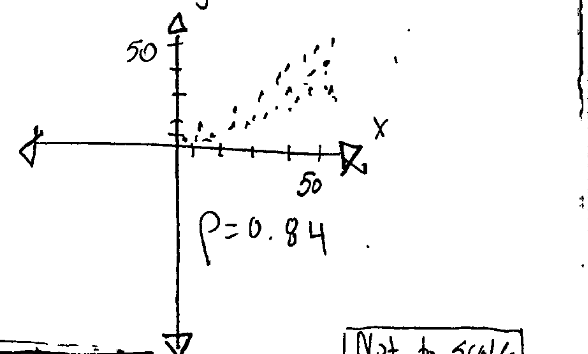
x	y
1	0.99
2	2.09
3	2.05
4	3.13
5	7.03
6	5.96
7	4.09
8	9.78
9	6.78
10	5.84
11	19.42
12	11.77
13	13.80
14	13.77
15	24.25
16	8.72
17	11.24
18	18.36
19	18.04
20	10.88
21	23.49
22	15.37
23	19.51
24	11.56
25	21.99
26	32.89
27	20.71
28	13.83
29	23.76
30	20.54
31	40.26
32	30.81
33	36.15
34	25.18
35	24.26
36	53.2
37	23.70
38	23.65
39	36.69
40	37.53
41	40.62
42	14.44
43	63.42
44	36.17
45	41.64
46	47.46
47	70.14
48	43.41
49	29.17762
50	56.80
X	1275
y	1205
X^2	42925
xy	41073

β_0	-1.22
β_1	0.99
ρ_{xy}	0.84

$y = -1.22 + 0.99x$

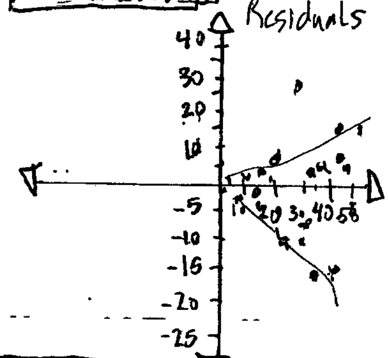
X	$Y - \hat{Y}$
1	0.16
2	1.96
3	2.80
4	1.02
5	5.20
6	7.39
7	4.54
8	9.48
9	11.76
10	10.01
11	13.98
12	10.06
13	10.04
14	0.84
15	14.00
16	16.56
17	4.45
18	20.68
19	26.74
20	32.59
21	14.77
22	16.37
23	16.48
24	21.67
25	32.37
26	14.22
27	24.29
28	35.09
29	71.17
30	20.33
31	14.95
31	19.95
33	22.19
34	55.78
35	25.27
36	21.97
37	28.99
38	34.31
39	22.23
40	20.97
41	30.70
42	36.51
43	48.14
44	58.55
45	42.29
46	49.99
47	27.06
48	24.55
49	61.49
50	41.57

Scatterplot

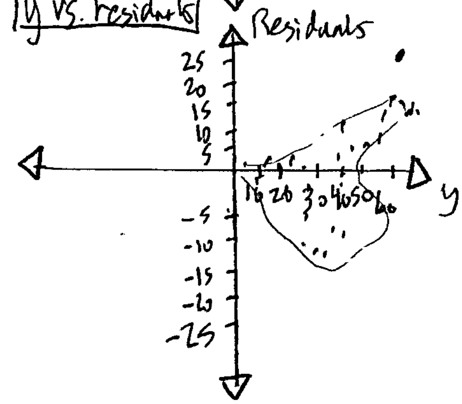


Not to scale
-Lost my ruler-

X vs. Residuals



y vs. residuals



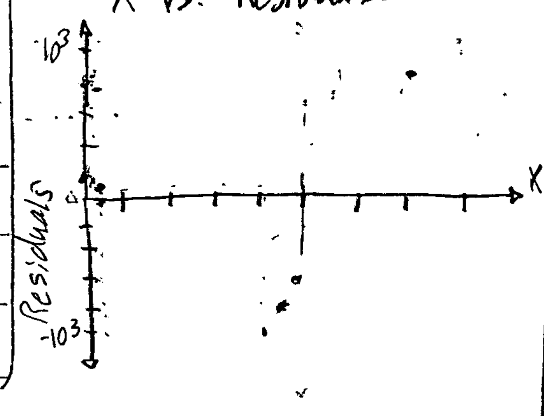
35. $X_3 = X_1 + X_2$;

$$\mathbb{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_2 - X_1 \end{pmatrix} \xrightarrow{R_3 + R_1} \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 2 & X_2 \end{pmatrix} \xrightarrow{R_3 - R_2} \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & 0 \end{pmatrix} \xrightarrow{-1} \begin{pmatrix} 0 & X_1 \\ 0 & X_2 \\ 0 & 0 \end{pmatrix} \mapsto \boxed{\begin{pmatrix} 0 & X_1 \\ 0 & X_2 \end{pmatrix}}$$

36.

$\sum x$	14383
$\sum y$	448764
$\sum xy$	19512852
$\sum x^2$	39764174
β_0	24818
β_1	-8.49

X vs. Residuals



The residuals demonstrate the fit wasn't an exact approximation. Separately, an analysis of variance demonstrated an F-statistic that accepted the alternative.

37. $\ln(\text{Pressure}) = A + \frac{B}{T}$; $\beta_0 = 6.51$; $\beta_1 = 91.66$; $A = 6.51$; $B = 91.66$

$$\mathrm{var}(\beta_0) = \mathrm{Var}(A) = \frac{\sigma^2 \sum X_i^2}{n\sum X_i^2 - (\sum X_i)^2} = 0.2 \quad ; \ \mathrm{var}(\beta_1) = \mathrm{var}(B) = \frac{n\sigma_1^2}{n\sum X_i^2 - (\sum X_i)^2}$$

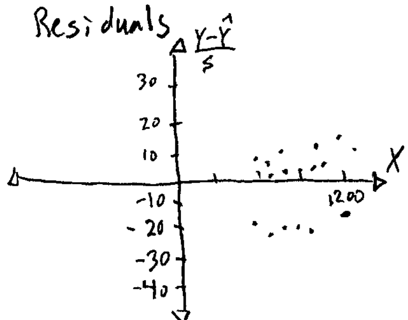
$$\hat{A} = 6.51 \pm t\left(\tfrac{0.05}{2}\right)_{23-2} \cdot \sigma \qquad \hat{B} = 91.66 \pm t\left(\tfrac{0.05}{2}\right)_{23-1} \cdot \sigma = 1.4 \quad \boxed{= 180.4}$$

$$= 6.51 \pm 2.056 \cdot 0.4 \qquad = 91.66 \pm 2.056 \cdot 13.4$$

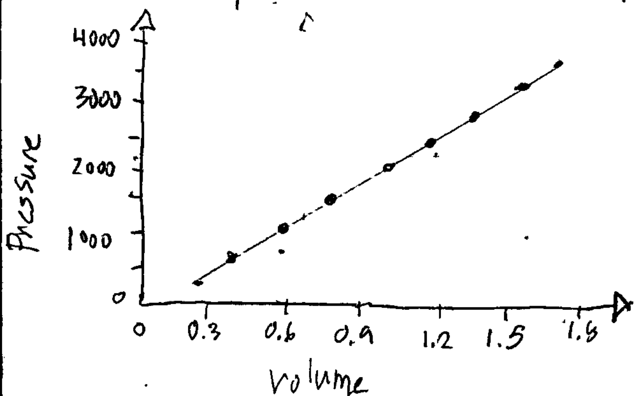
$$\boxed{= 6.51 \pm 0.822} \qquad \boxed{= 91.66 \pm 27.551}$$

38.

Coefficient	Estimate	Standard Error	Confidence Interval
β_0	628.63	13.428	± 27.78
β_1	-0.61	0.018	± 0.037

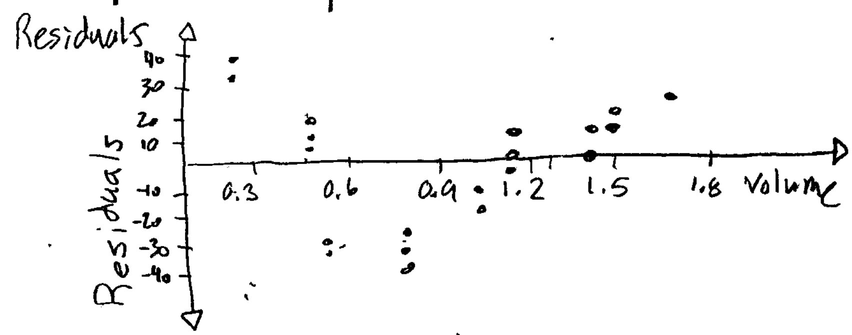


39. a. The plot of volume vs. pressure is linear



b.

Coefficient	Estimate	Standard Error
β_0	-257.301	9.43
β_1	2316.469	9.24



c. $S = \sum(y - \beta_0 - \beta_1 X - \beta_2 X^2)^2$

$$\frac{dS}{d\beta_0} = -2\sum(y - \beta_0 - \beta_1 X - \beta_2 X^2) = 0$$

$$\sum y = n\beta_0 + \beta_1 \sum X + \beta_2 \sum X^2$$

$$\frac{dS}{d\beta_1} = -2\sum X(y - \beta_0 - \beta_1 X - \beta_2 X^2) = 0$$

The residual plot is multiple measurement at a given volume.

$$\sum Xy = \beta_0 \sum X_i + \beta_1 (\sum X)^2 + \beta_2 \sum X_i^3$$

$$\frac{dS}{d\beta_2} = -2\sum X^2(y - \beta_0 - \beta_1 X - \beta_2 X^2) = 0$$

$$\sum X^2 y = \beta_0 \sum X^2 + \beta_1 \sum X_i^3 + \beta_2 (\sum X)^4$$

$$\beta_0 = \frac{D_0}{D} = \frac{\begin{bmatrix} \sum y & \sum X & \sum X^2 \\ \sum Xy & (\sum X)^2 & \sum X^3 \\ \sum X^2 y & \sum X^3 & \sum X^4 \end{bmatrix}}{\begin{bmatrix} n & \sum X & \sum X^2 \\ \sum X & (\sum X)^2 & \sum X^3 \\ \sum X^2 & \sum X^3 & \sum X^4 \end{bmatrix}} \quad \beta_1 = \frac{D_1}{D} = \frac{\begin{bmatrix} n & \sum y & \sum X^2 \\ \sum X & \sum Xy & \sum X^3 \\ \sum X^2 & \sum X^2 y & \sum X^4 \end{bmatrix}}{\begin{bmatrix} n & \sum X & \sum X^2 \\ \sum X & (\sum X)^2 & \sum X^3 \\ \sum X^2 & \sum X^3 & \sum X^4 \end{bmatrix}} \quad \beta_2 = \frac{D_2}{D} = \frac{\begin{bmatrix} n & \sum X & \sum y \\ \sum X & (\sum X)^2 & \sum Xy \\ \sum X^2 & \sum X^3 & \sum X^2 y \end{bmatrix}}{\begin{bmatrix} n & \sum X & \sum X^2 \\ \sum X & (\sum X)^2 & \sum X^3 \\ \sum X^2 & \sum X^3 & \sum X^4 \end{bmatrix}}$$

Cramer's Rule did not work

$$\beta_2 = \frac{[\sum x^2y - \frac{1}{n}\sum x^2 \sum y][\sum x^2 - \frac{1}{n}(\sum x)^2] - [\sum xy - \frac{1}{n}\sum x \sum y][\sum x^3 - \sum x \sum x^2/n]}{[\sum x^2 - (\sum x)^2/n][\sum x^4 - \frac{1}{n}(\sum x^2)^2] - (\sum x^3 - \sum x \sum x^2/n)^2}$$

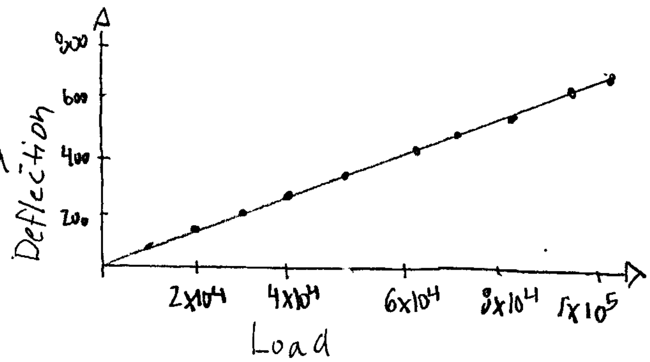
$$\beta_1 = \frac{[\sum xy - \sum x \sum y/n][\sum x^4 - (\sum x^2)^2/n] - [\sum x^2y - \sum x^2 \sum y/n][\sum x^3 - \sum x \sum x^2/n]}{[\sum x^2 - (\sum x)^2/n][\sum x^4 - \frac{1}{n}(\sum x^2)^2] - (\sum x^3 - \sum x \sum x^2/n)^2}$$

$$\beta_0 = \frac{\sum y}{n} - \beta_1\left[\frac{\sum x}{n}\right] - \beta_2\left[\frac{\sum x^2}{n}\right]$$

Coefficient	Estimate
β_0	-305
β_1	2164
β_2	83.19

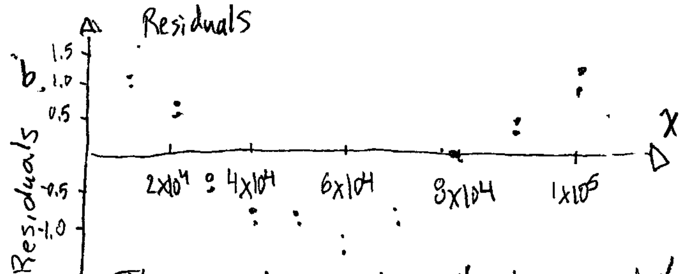
The fit is slightly better but negligible to linear.

40. a



The plot resembles a linear function.

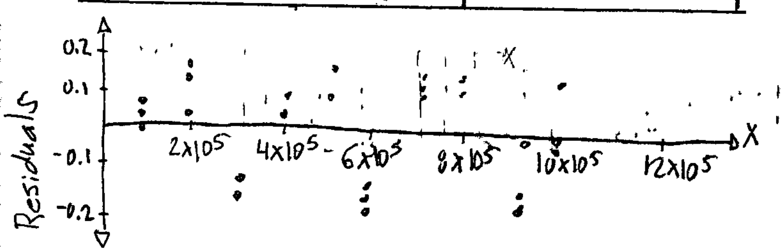
b.



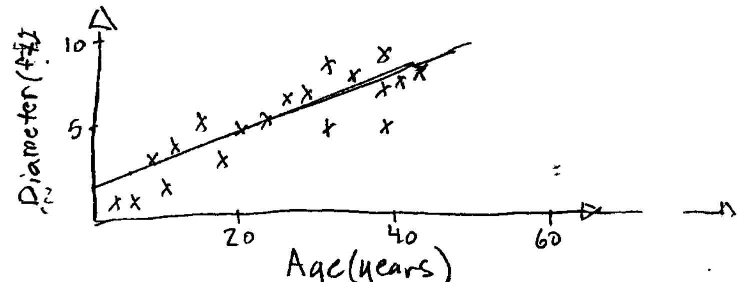
The residual plot helps aid the recognition of regions that over or under fit.

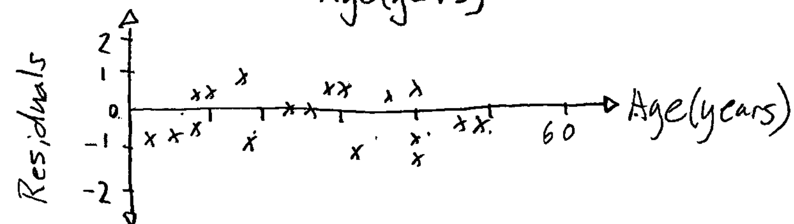
c.

Coefficient	Estimate	Standard Error
β_0	0.1362	0.073
β_1	0.0068	3.04×10^{-6}
β_2	7.29×10^{-10}	2.69×10^{-11}



41.





I thought a better fit of Age vs. Diameter required Age vs. $(\text{Diameter})^2$, but the book reasons $\sqrt{\text{Age}}$ vs Diameter.

42. The residuals of velocity vs $\sqrt{\text{Distance}}$ were 10x smaller than residuals 'extracted' from velocity vs. Distance. A reason to explain the difference is the second-order differential equation $m\ddot{x} = mg - kv^2$, in which re-arranged is

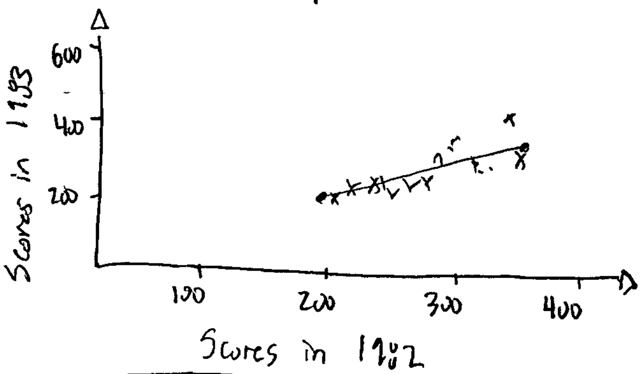
$$v = \sqrt{\frac{m}{k}\left(g = \frac{\Delta\Delta\,\text{Distance}}{\Delta\Delta\,\text{time}}\right)}$$

43. The information provided of cyst diameters as a function of temperature looks to be quadratic.

Coefficient	Estimate	Standard Error
β_0	23.86	0.83
β_1	-0.17	0.13
β_2	0.07	0.15
β_3	-0.45	0.18

44. A correlation coefficient ($\rho = -0.37$) signifies weak correlation for the asthma dataset, and similar insignificance for the cystfibr dataset having a correlation of ($\rho = -0.27$). The F-statistic for asthma ($F_{0.95}(40,1) = 0.01 < 6.57$) and cystfibr ($F_{0.95}(22,1) = 0.20 < 1.75$) also say no significance. A plot of either datasets fits were unresolved, having no confident, "yes!"

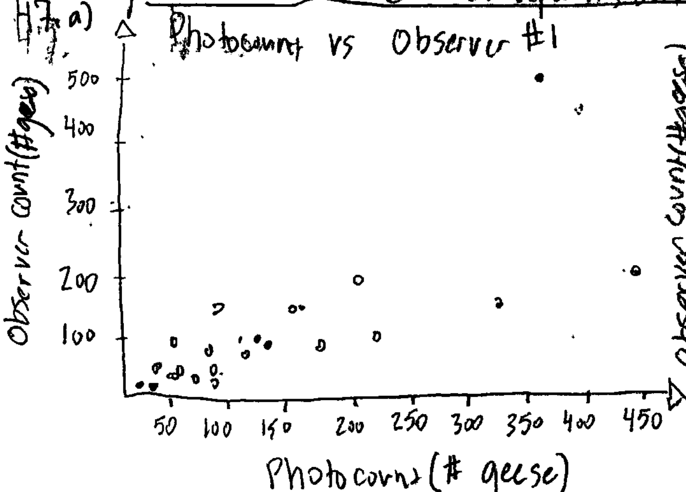
45. The relationship of student reading scores exists from year-to-year.

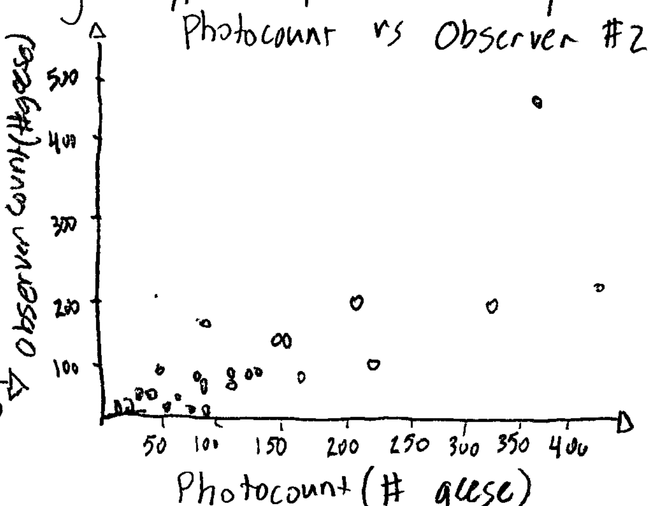


Coefficient	Estimate	Standard Error
β_0	57.54	24.39
β_1	0.80	0.09

46. Phase Contrast Microscopy (PCM) Fit Scanning Electron Microscopy (SEM) with an R^2 of 0.37. The residuals of data points were as large as the magnitude of the datapoint, but generally do present a positive correlation.

47. a)





Each plot comparison of the true photocount to the observer looks to be linear.

b)

Coefficient	Estimate	Standard Error
β_0	-4.93	9.31
β_1	0.85	0.07

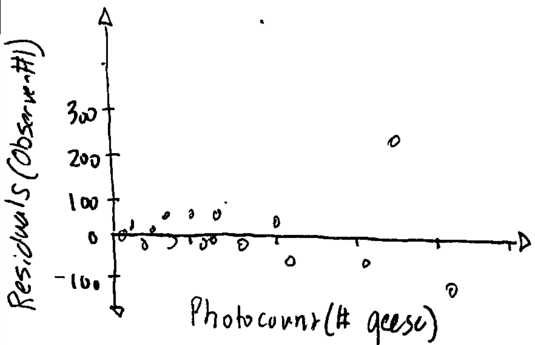
Standard Residual Error: 18.15

Coefficient	Estimate	Standard Error
β_0	-4.15	8.71
β_1	1.11	0.07

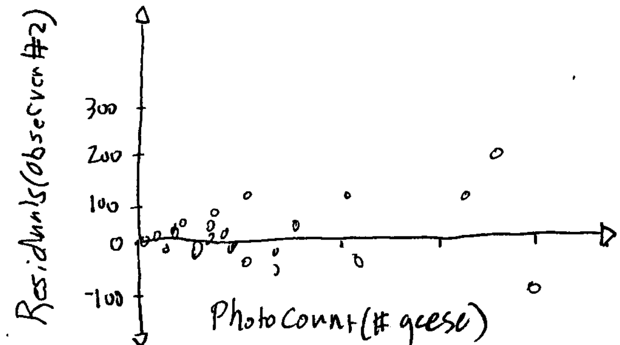
Standard Residual Error: 15.84

The coefficients and standard residual error for each plot demonstrate similarities.

Residuals of Observer #1:



Residuals of Observer #2:



c. Square rooting the photocounts did not stabilize variance, and increased the errors across sampled datasets.

d. Again, the original fit was appropriate to the analysis.

e. Both observers either over-, or under-estimated the number of geese by 15%.

48.

Coefficients	Estimate	Standard Error
β_0	-57.99	8.64
β_1	4.71	0.26
β_2	0.34	0.13

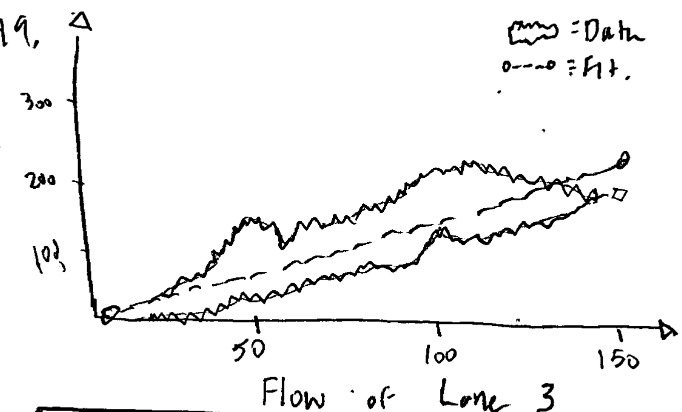
Model : $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

where X_1 = height, X_2 = Diameter.

Y = Volume.

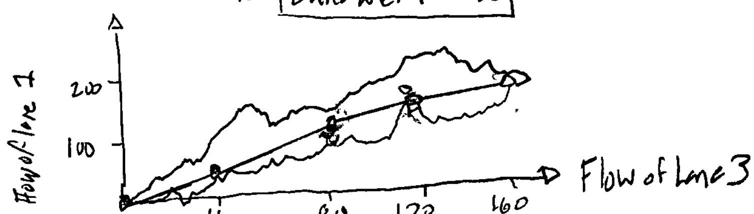
The relationship $\text{volume} = \pi\left(\frac{\text{Diameter}}{2}\right)^2 \times \text{Height}$ did not fit well.

49.



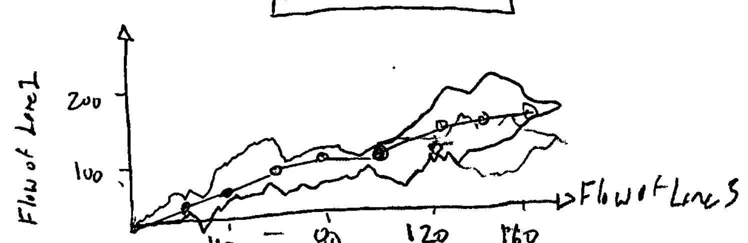
A linear regression of Lane 1 vs Lane 3 flow is justified by observation and tabulated data.

Bandwidth = 40

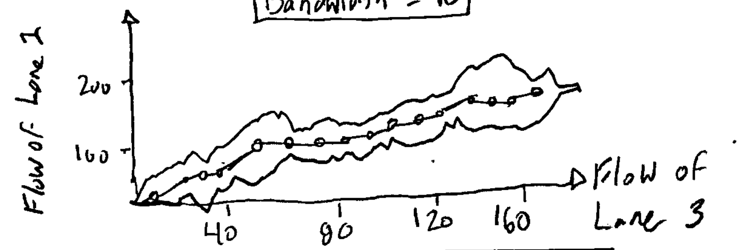


The best balance between a small bandwidth (10) and larger (160) became a $w(x_i - \bar{x}) = 40$. My perspective is only by observation and not numerical determinations.

Bandwidth = 20

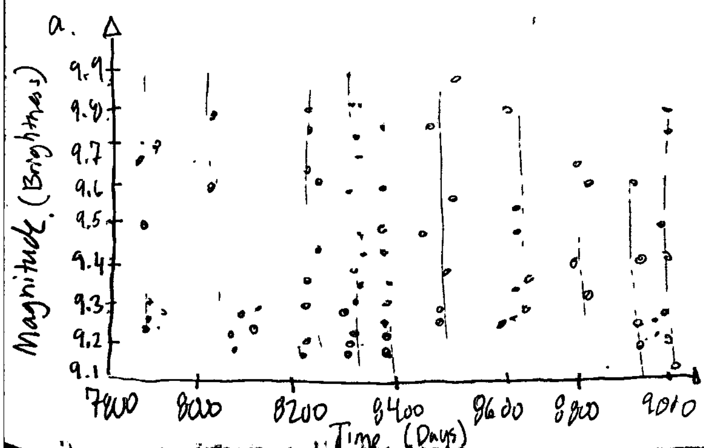


Bandwidth = 10



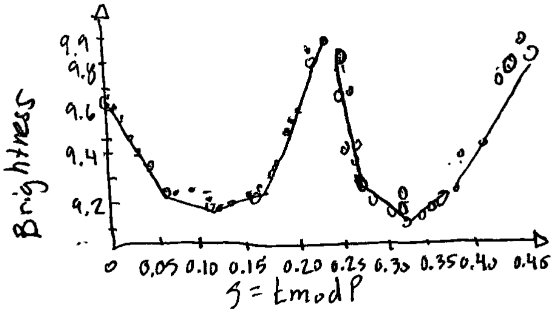
56.

a.



The structure of time vs brightness seems to be measurements near the same time.

b. $P = 0.407528$ days ; $s = t \bmod P$ c. see part b)



d) As the modular period changes, the graph becomes less periodic. The true period could be estimated with a sinusoidal wave.

The structure seems to be periodic from the $s = t \bmod P$ transformation.

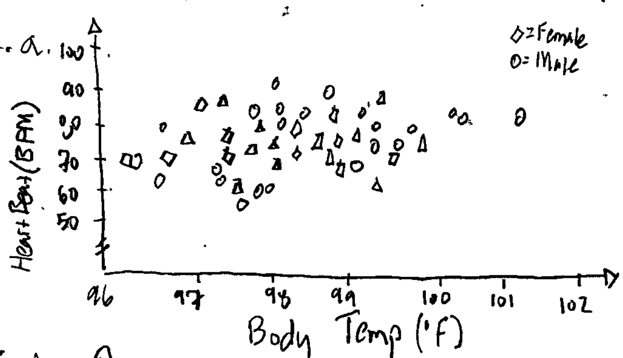
51. Model : Disney = $\beta_0 + \beta_1 \cdot$ MacDonalds $+ \beta_2 \cdot$ Schumberger $+ \beta_3 \cdot$ Haliburton.

Coefficient	Estimate	Standard Error
β_0	0.0938	0.0981
β_1	-0.8812	1.1045
β_2	1.3151	1.9877
β_3	-0.1717	1.3442

Standard Deviation of Residuals = 0.056
R^2 of Fit = 0.6298.
The fitted errors of multiple linear regression
$$S = \sum (y_i - \beta_0 - \beta_1 X_1 - \beta_2 X_2 - \beta_3 X_3)^2$$
has a standard deviation of 0.002.

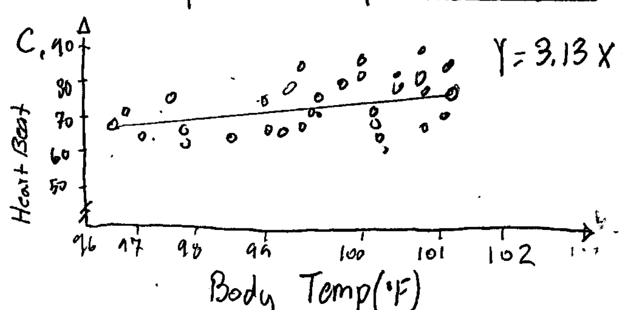
The comparison of monthly returns from 1998 to 1999 coefficient estimates of different values. The fundamental model has three independent variables that represent individual stock price.

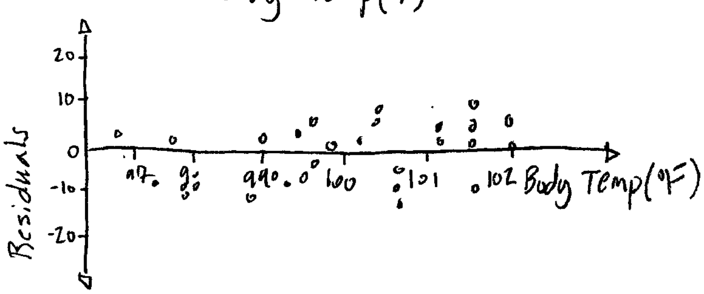
52. a.



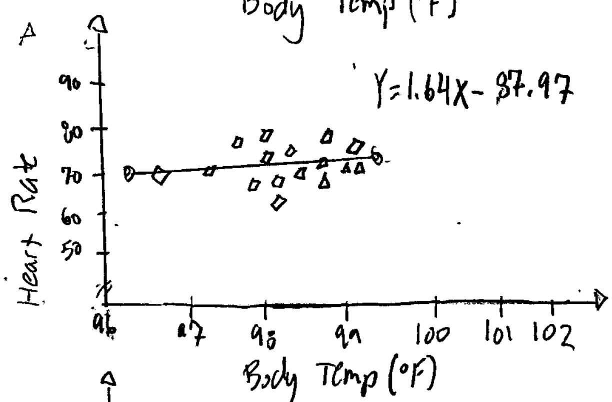
b. Male vs female data do relate through a positive correlation value ($\rho_{male} = 0.28$, $\rho_{female} = 0.20$).

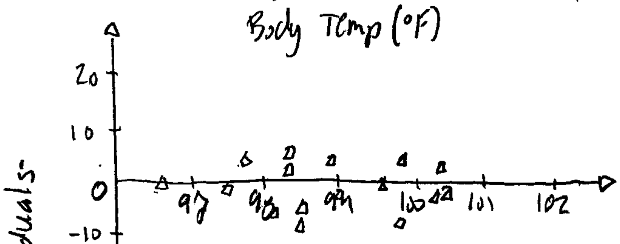
c.





d.





The R^2 of 0.02 demonstrates increasingly minor correlation to a line.

Male body temperature is regressed to a linear fit with a standard error of 7.70 for residuals. The R^2 of 0.08 may indicate a slightly positive relationship to bodily measurements.

e. $H_0: \beta_{1,Male} = \beta_{2,Female}$

$H_1: \beta_{1,Male} \neq \beta_{2,Female}$

$t = \dfrac{\beta_{1,Male} - \beta_{2,Female}}{s_{\beta_1 - \beta_1}}$; $s_{\beta_M - \beta_F} = s_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}} = \sqrt{\dfrac{(n-1)s_x^2 + (m-1)s_y^2}{m+n-2}\left[\dfrac{1}{n} + \dfrac{1}{m}\right]}$

$= \dfrac{3.13 - 1.64}{0.209} = 7.13$

$= \sqrt{\dfrac{(65-1)1.639^2 + (64-1)1.316^2}{65+64-2}\left[\dfrac{1}{65} + \dfrac{1}{64}\right]}$

$\boxed{= 0.209}$

Degrees of Freedom $= \boxed{127}$

$t_{120}(0.95) < t$; An alternative hypothesis is acceptable, indicating male and female temperature vs heart rate are not relatable.

F. $H_0: \beta_{0,Male} = \beta_{0,Female}$

$H_1: \beta_{0,Male} = \beta_{0,Female}$

$t = \dfrac{|\beta_{0,Male} - \beta_{0,Female}|}{s_{\beta_M - \beta_F}}$; $s_{\beta_M - \beta_F} = \sqrt{\dfrac{(65-1)101.9^2 + (64-1)129.52^2}{65+64-2}\left[\dfrac{1}{65} + \dfrac{1}{64}\right]}$

$\boxed{= 20.4}$

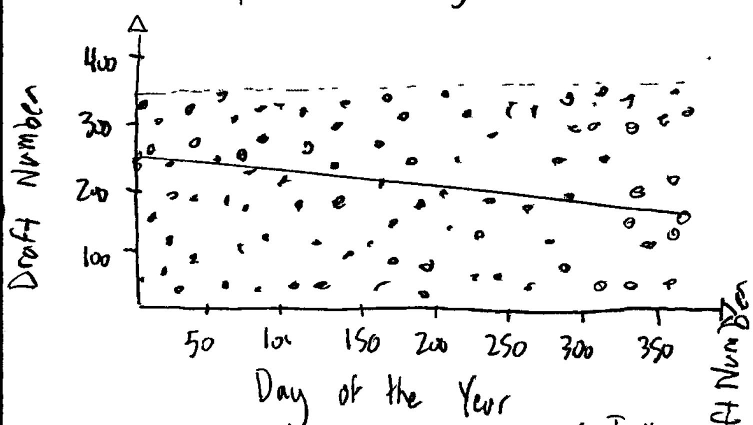
$= \dfrac{|-233 + 87|}{20.4} = \boxed{7.16}$

$t_{120}(0.95) < t$; An alternative hypothesis is acceptable.

53.
a) Old faithfuls interval between explosion is related to duration by a linear fit having an R^2 of 0.73

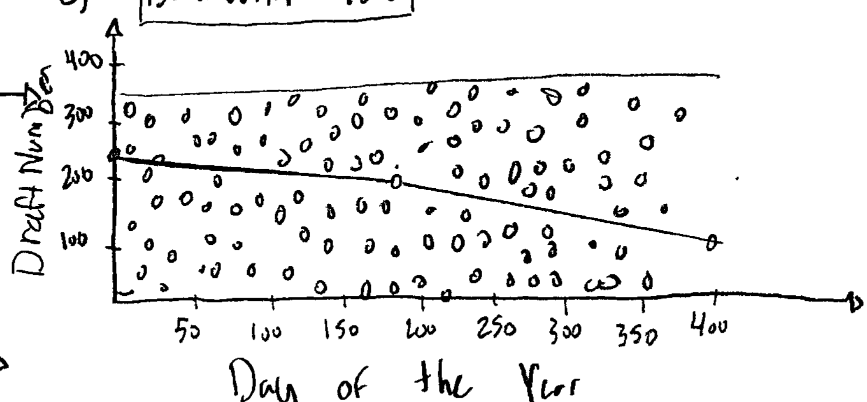
b) If the duration is 2 minutes, to predict a time until next eruption would require a function. Linear regression of $Y = 10.74 \cdot \text{Duration} + 33.83$ minutes is best fit and describes an interval till next eruption to be 55 min.

54. a) The plot of 'Day of the Year' vs 'Draft Number' shows lack of trend

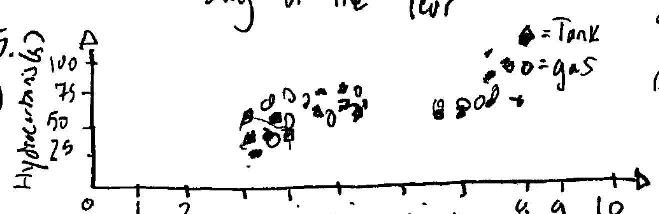


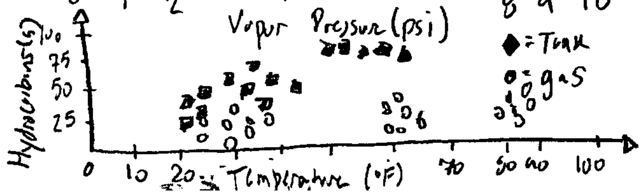
b) The linear regression model $Y = -0.22X + 225$ is plotted in part a), but has an R^2 of 0.05.

c) Bandwith = 180



55.
a)





The groupings (vapor pressure, temperature) are associated to hydrocarbon production by visual observations, especially vapor pressure.

b) $\boxed{\text{The best fits are tank vapor pressure vs. hydrocarbon weight and gass vapor pressure vs hydrocarbon Weight.}}$

c) Root mean square Prediction: $RMSPE = \sqrt{\frac{1}{40}\sum_{i=1}^{40}(Y_i - \hat{Y}_i)^2}$

Model: Gas Vapor Pressure vs. Hydrocarbon Weight

Coefficient	Estimate
β_0	6.62
β_1	5.56

$$RMSPE = \sqrt{\frac{1}{40}\sum^{40}(Y_i - 6.62 - 5.56X_i)}$$

$\boxed{= 0.667}$

Model: Tank Vapor Pressure vs Hydrocarbon weight

Coefficient	Estimate
β_0	3.59
β_1	6.34

$$RMSPE = \sqrt{\frac{1}{40}\sum^{40}(Y_i - 3.59 - 6.34X_i)}$$

$\boxed{= 0.634}$

56. a) After examination of different combinations of data, insolation fits the upper ranges of oxidation, while wind speed, temperature, and humidity better ~~fit oxidation~~ when lesser than 30 ppm. Oxidation level maximums are modeled with insolation to a regression of $Y = 0.23X - 1.43$, but do have a squared multiple correlation coefficient (R^2) of 0.22.

b) Serial Correlation (or experimental drift) appears in time-resolved data as increasing (or decreasing) residual error.