Midterm 1: Chapters 1 to 4

1	2	3	4	5	6	7	8	9	10
22.43	10.25	23.71	21.77	22.11	18.71	19.77	20.33	20.17	21.12

(a) $\bar{x} = \frac{\Sigma_{i=1}^{10} x_i}{10} = 20.037$. $\tilde{x} = \frac{20.33+21.12}{2} = 20.725$
(b) $s_x^2 = \frac{\Sigma(x_i-\bar{x})^2}{n-1} = 13.935$. Split into 4 sections, numbers seperating are quartiles. Last minus first is IQR $= 22.11 - 19.77 = 2.34$
(c) Trimmed mean of 10%: remove 10% from highest and lowest. $= 20.801$. Close to median but more than mean; data is slightly skewed to the left.
(d) Set decimal point to |.
(e) Away from Q1 and Q3 by 1.5·IQR are outliers. Left dot is minimum, start of box is Q1, middle line is median, end of box is Q3, last dot is maximum.

A: polluted, B: test detects pollution, $P(A) = 0.2$, $P(B \mid A) = 0.60$, $P(B \mid A') = 0.3$
(a) $P(A \cap B) = P(A)P(B \mid A) = 0.2 \cdot 0.6 = 0.12$
(b) $P(B) = P(B \cap A) + P(B \cap A') = 0.12 + P(A')P(B \mid A') = 0.12 + 0.8 \cdot 0.3 = 0.36$
(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2 + 0.36 - 0.12 = 0.44$
(d) $P(A' \mid B') = \frac{P(A' \cap B')}{P(B')} = \frac{1-0.44}{0.64} = 0.875$ (e) $P(P(A \cap B) \neq 0)$ so are not mutually exclusive. $P(A \cap B) \neq P(A)P(B)$ so are not independent.

X number of cash registers being used for location 1, Y the number used at the same time for location 2.

$f(x,y)$		$y=0$	$y=1$	$y=2$
x	0	0.10	0.05	0.05
	1	0.10	0.20	0.05
	2	0.05	0.10	0.30
$h(y)$		0.25	0.35	0.40

(a) Marginal probability mass functions, add the rows for X; columns for Y. $h(y)$ example.
(b) Cumulative distribution function of X: 0 if $x < 0$, 0.2 if $0 \le x < 1$. 0.55 if $1 \le x < 2$. 1 if $x \ge 2$. So $F(1.5) = 0.55$.
(c) Conditional distribution of Y given $X = 2$, $f(y \mid X = 2)$: $\frac{0.05}{0.45}$ when $y = 0$, $\frac{0.10}{0.45}$ when $y = 1$, $\frac{0.30}{0.45}$ when $y = 2$.
(d) Mean of X: $\mu_X = E(x) = \Sigma_{x=0}^2 x g(x) = 0.35 + 2 \cdot 0.45 = 1.25$ Variance of X: $\sigma_X^2 = E(X^2) - \mu_X^2 = \Sigma_{x=0}^2 x^2 g(x) - \mu_X^2 = 0.35 + 4 \cdot 0.45 - 1.25^2 = 0.5875$
(e) $\sigma_{XY} = E(XY) - \mu_X\mu_Y = (0.20 \cdot 1 \cdot 1 + 0.10 \cdot 2 \cdot 1 + 0.05 \cdot 1 \cdot 2 + 0.30 \cdot 2 \cdot 2) - 1.25 \cdot 1.15 = 0.2625 \neq 0$ so X and Y are not independent.

X be a continuous random variable with probability density function $f(x) = Cx^2$ if $-2 < x < 1$ and zero otherwise.
(a) Find C, it must make the function equal to 1 for the interval. $1 = \int_{-2}^1 Cx^2 dx$, $C = \frac{1}{3}$
(b) $\mu = E[X] = \int_{-2}^1 x(\frac{1}{3}x^2)dx = -\frac{5}{4}$. $\sigma^2 = E[X^2] - \mu^2 = \int_{-2}^1 x^2(\frac{1}{3}x^2)dx - (-\frac{5}{4})^2 = 0.6375$
(c) $P[X < -1] = P(-2 < x < -1) = \int_{-2}^{-1} \frac{1}{3}x^2 dx = \frac{7}{9}$
(d) $P[-1 < X \le 3] = P(-1 < X < 1) + P(1 < X < 3) = \int_{-1}^1 \frac{1}{3}x^2 + \int_1^3 \frac{1}{3}x^2 = \frac{2}{9} + 0 = \frac{2}{9}$
(e) $g(X) = 4X - 3$. $\mu_{g(X)} = \int_{-2}^1 g(x)f(x)dx = \int_{-2}^1 (4X-3)(\frac{1}{3}x^2) = -8$. $\sigma_{g(X)}^2 = E(g(X)^2) - \mu_{g(X)}^2 = \int_{-2}^1 (4X-3)^2(\frac{1}{3}x^2) - 64 = 10.2$

<u>Combinations</u> Different order is still same set. $\binom{n}{r} = \frac{n!}{r!(n-r)!}$
<u>Permutations</u> Different order is different set. $({}_nP_r) = \frac{n!}{(n-r)!}$
<u>Chebyshev's Theorem</u> The probability that a random variable X will be within k standard deviations of the mean is $P(\mu - k\sigma < X < \mu + k\sigma) \ge 1 - \frac{1}{k^2 y}$.

Midterm 2: Chapters 5 to 9

There are on average, 3 potholes in a section of 1 mile. Over time, so use Poisson Distribution $\frac{e^{-\lambda t}(\lambda t)^x}{x!}$, $\lambda = 3$
(a) Probability at least 2 potholes appear in a section of 1 mile. $P(X < 2) = P(X = 0) + P(X = 1) = \frac{e^{-3}3^0}{0!} + \frac{e^{-3}3^1}{1!} = 0.20$
(b) Probability at least 2 but less than 4 appear in a section of 1 mile. $P(2 \le X < 4) = P(X = 2) + P(X = 3) = \frac{e^{-3}3^2}{2!} + \frac{e^{-3}3^3}{3!} = 0.45$
(c) Exactly 5 potholes occur in a section of 2 miles. $P(X = 5) = \frac{e^{-(3\cdot2)}\cdot(3\cdot2)^5}{5!} = 0.161$
(d) Probability of having less than 2 potholes in exactly 2 of the next 5 miles. Binomial, either succeed or fail. $b(2; 5, 0.2) = \binom{5}{2} \cdot 0.2^2 \cdot 0.8^3 = 0.20$
(e) Probability the first section of 1 mile with less than 2 potholes is the 4th mile. Geometric, xth trial is the first success. $g(4, 0.20) = (0.20)(0.8)^{4-1} = 0.1024$

Life of a electrical switch has an exponential distribution with a mean of 1.45 ($\beta = 1.45$) and parameter β given by $f(x; \beta) = \frac{1}{\beta}e^{-x/\beta}$ when $x > 0$ otherwise 0. β is the mean time between failures.
(a) Probability a randomly selected switch fails within a year. $P(X < 1) = 1 - e^{(-1/1.45)\cdot 1} = 0.50$
(b) Switch functions more than a year but less than 2.9 years. $P(1 < X < 2.9) = \int_1^{2.9} \frac{1}{1.45}e^{-x/1.45} = 0.37$
(c) Switch was functioning for 2 years, probability that it will function for another 3.48 years. $P(x > 3.48 + 2 \mid X > 2) = P(X > 3.48)$ memoryless $= e^{-3.48/1.45} = 0.09$
(d) If 10 switches are installed what is the probability exactly 3 fail during the first year? $Y \sim b(10, 0.5)$, $P(Y = 3) = \binom{10}{3}0.5^3 \cdot 0.5^7 = 0.117$
(e) If 100 switches are installed what is the probability that at least 40 but less than 55 fail within the first year? $Y \sim b(100, 0.5)$, $\mu_Y = n \cdot p = 50$, $\sigma_Y = n \cdot p \cdot q = 25$. Because $np = nq = 50 > \sqrt{npq} = 5$, use Normal distribution to approximate. $Y \approx Normal(50, 5)$. $P(40 \le Y < 55) = P(39.5 \le Y < 54.5)$ to standard normal $Z = \frac{X-\mu}{\sigma}$, $= P(\frac{39.5-50}{5} \le \frac{Y-50}{5} < \frac{54.5-50}{5}) = P(-2.1 \le Z < 0.9)$ Check table values. $= 0.8159 - 0.0179 = 0.798$.

Approximately normally distributed with a mean of 175cm and a standard deviation of 6cm. $X \sim N(175, 6)$
(a) Probability that a randomly selected student is at least 170 cm. $P(X \ge 170) = P(\frac{X-175}{6} \ge \frac{170-175}{6}) = P(Z \ge -0.83) = 1 - 0.2033 = 0.7967$
(b) Probability the student's height is between 170 and 180. $P(170 < X < 180) = P(-0.83 < Z < 0.83) = 1 - P(Z < -0.83) - P(Z > 0.83) = 1 - 0.2033 - (1 - 0.7967) = 0.59$
(c) Probability sample of 16 students have an average height is less than 170. Using CLT, $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is a standard normal distribution. $\frac{170-175}{6/\sqrt{16}} = -3.3\bar{3}$. $P(Z < -3.3\bar{3}) = 0.0004$
(d) Continuing from (c), would the probability the average height is between 170 and 180 be bigger, smaller, or the same as (b)? The larger the sample size, the smaller the variance of sample mean. So it will be bigger.
(e) Calcualte a 95% confidence interval of the mean height. $n = 25$, $\bar{x} = 170$, $s = 5$. μ known, σ^2 unknown. CI $= \bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$. Calculate α: $95 = 100(1 - \alpha)$, $\alpha = 0.05$. $170 \pm t_{0.025,24}\frac{5}{\sqrt{25}} = 167.9 < \mu < 172.1$. CI does not include 175, it is likely the average height has changed in the last 10 years.

---

16 batteries for both type 1 and type 2. $\bar{x}_1 = 37.3$, $\bar{x}_2 = 40.5$, $s_1 = 1.9$, $s_2 = 2.1$. Assume batteries are normally distributed and assum $\sigma_1 = \sigma_2$.
(a) Find 95% confidence interval for μ_1. μ known, σ unknown. $37.3 - 2.131\frac{1.9}{\sqrt{16}} < \mu_1 < 37.3 + 2.131\frac{1.9}{\sqrt{16}} = 36.3 < \mu_1 < 38.3$.
(b) Find 90% confidence interval for μ_2. μ known, σ unknown. $40.5 - 1.753\frac{2.1}{\sqrt{16}} < \mu_2 < 40.5 + 1.753\frac{2.1}{\sqrt{16}} = 39.6 < \mu_2 < 41.4$
(c) From (a) and (b), which CI has a higher probability of containing the true mean? Since we don't know what the true mean is, it is impossible to tell which CI it is in.
(d) Find a 95% confidence interval for $\mu_1 - \mu_2$. $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = 4.01$ $(\bar{x}_1 - \bar{x}_2) - t_{0.025,30}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{0.025,30}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = -4.6 < \mu_1 - \mu_2 < -1.8$.
(e) From (d), what conclusions can you draw for the two types of batteries?. The CI only contains negative values, which suggests that battery type 2 might be better.

<u>Hypergeometric Distribution:</u> Choosing successful items.

$h(x;N,n,k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$. $max(0, n-(N-k)) \le x \le min(n,k)$. x: num of successes. N: num of items. n: num of selection. k: num of total successes. $\mu = \frac{nk}{N}$, $\sigma^2 = \frac{N-n}{N-1} n \frac{k}{N}(1-\frac{k}{N})$

Negative Binomial Distribution: Prob. the kth success will happen by the xth trial. $b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k}$. x: trial number. k: success number. p: prob. success. q: prob. failure.

Normal Distribution: Bell curve. $n(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$. x: select time. μ: mean. σ: standard deviation.

CI on μ, σ^2 known: $\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

CI for $\mu_1 - \mu_2$, σ_1^2 and σ_2^2 known: $(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

CI for $\mu_1 - \mu_2$, $\sigma_1^2 \ne \sigma_2^2$ and both unknown: $(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Final Practice: Chapters 1 to 10

Student	1	2	3	4	5	6	7	8	9	10
Test Score x	65	50	55	65	55	70	65	70	55	70
Chemistry Grade y	85	74	76	90	85	87	94	98	81	91

$\bar{x} = 62$, $s_x^2 = 56.7$, $\Sigma_{i=1}^{10} = y_i^2 = 74653$.

(e) The Test Score (x) is linearly related to IQ score (w) and the relationship is $w_i = \frac{27}{22}x_i + 40$. Calculate the sample mean and the sample variance of IQ score. $\bar{w} = \frac{27}{22}\bar{x} + 40 = \frac{27}{22} \cdot 62 + 40 = 116.1$. $s_w^2 = (\frac{27}{22})^2 \cdot s_x^2 = (\frac{27}{22})^2 \cdot 56.7 = 85.4$

(f) Construct the 95% two-sided confidence interval for the mean Chemistry Grade (μ_Y). $\bar{y} - t_{0.025,9}\frac{s_y}{\sqrt{10}} < \mu_y < \bar{y} + t_{0.025,9}\frac{s_y}{\sqrt{10}} = 86.1 - 2.262\sqrt{\frac{57.9}{10}} < \mu_y < 86.1 + 2.262\sqrt{\frac{57.9}{10}} = 80.7 < \mu_y < 91.5$.

$P(B_1) = 0.3$, $P(B_2) = 0.45$, $P(B_3) = 0.25$. $P(D \mid B_1) = 0.02$, $P(D \mid B_2) = 0.03$, $P(D \mid B_3) = 0.02$.

(a) What is the probability a randomly selected product is defective? $P(D) = P(D \cap B_1) + P(D \cap B_2) + P(D \cap B_3) = P(D \mid B_1)P(B_1) + P(D \mid B_2)P(B_2) + P(D \mid B_3)P(B_3) = 0.02 \cdot 0.3 + 0.03 \cdot 0.45 + 0.02 \cdot 0.25 = 0.0245$

(b) What is the probability a randomly selected defective product was made by machine B_3? Is defectiveness independent from the machine produced? $P(B_3 \mid D) = \frac{P(B_3 \cap D)}{P(D)} = \frac{P(D|B_3)P(B_3)}{P(D)} = \frac{0.02 \cdot 0.25}{0.0245} = 0.204 \ne 0.25 = P(B_3)$ Defective status is not independent of which machine it was produced.

(c) What is the probability a randomly selected product is made by machine B_3 or defective, or both? $P(D \cup B_3) = P(D) + P(B_3) - P(D \cap B_3) = 0.0245 + 0.25 - 0.02 \cdot 0.25 = 0.2695$.

(d) From a sample of 35 products, what is the exact probability that exactly 3 are defective? $X \sim b(3; 35, 0.0245) = \binom{35}{3} 0.0245^3 (1 - 0.0245)^3 2 = 0.0435$

(e) From (d), What is the approximated probability that exactly 3 of them are defective? Is it a good approximation? $n = 35$ is large and $p = 0.0245$ is very small. Can use Poisson with $\lambda = np = 35 \cdot 0.0245 = 0.8575$. $X \approx p(3; 0.8575) = \frac{e^{-0.8575} \cdot 0.8575^3}{3!} = 0.0446$ Very close approximation.

(f) Machines have been modified. From a sample of 500, 21 were defective. Does the data suggest the defective rate increased? Calculate the approximate P-value. H_0: $p \le 0.0245$ versus H_1: $p > 0.0245$. $np_0 = 500 \cdot 0.0245 = 12.25$ and $n(1-p_0) = 500 \cdot 0.9755 = 487.75$. Large enough for Normal approximation. Check H_0, $X \sim b(500, 0.0245) \approx N(12.25, \sqrt{11.95}), x = 21$. P-value $= P(X \ge 21) \approx P(\frac{X - 12.25}{\sqrt{11.95}} \ge \frac{20.5 - 12.25}{\sqrt{11.95}}) = P(Z \ge 2.39) = 1 - 0.9916 = 0.084$ P-value is very small, reject H_0 and conclude that modifications may have increased defective rate.

X is a random variable from a population with probability mass functions $P(X = 0) = 0.2$, $P(X = 1) = 0.4$, $P(X = 2) = 0.3$, $P(X = 4) = 0.1$ and 0 otherwise

(a) $\mu_X = E(X) = 0 \cdot 0.2 + 1 \cdot 0.4 + 2 \cdot 0.3 + 4 \cdot 0.1 = 1.4$. $\sigma_X^2 = 0^2 \cdot 0.2 + 1^2 \cdot 0.4 + 2^2 \cdot 0.3 + 4^2 \cdot 0.1 - 1.4^2 = 3.2 - 1.4^2 = 1.24$

(b) Derive the cumulative distribution functions of X. $F(x) = P(X \le x) = 0$ if $x < 0$, 0.2 if $0 \le x < 1$, 0.6 if $1 \le x < 2$, 0.9 if $2 \le x < 4$, 1 if $4 \le x$.

(c) Calculate $P(X > 1.9)$ $P(X > 1.9) = 1 - P(X \le 1.9) = 0.4$

(d) Let X_1 and X_2 be two independent random variables with the same probability mass function above. Derive the probability mass function of $\bar{X}_2 = \frac{X_1 + X_2}{2}$. 0.04 if $x = 0$, 0.16 if $x = 0.5$, 0.28 if $x = 1$, 0.24 if $x = 1.5$, 0.13 if $x = 2$, 0.08 if $x = 2.5$, 0.06 if $x = 3$, 0.01 if $x = 4$, 0 otherwise.

(e) $E(\bar{X}_2 = E(X_i) = 1.4$, $Var(\bar{X}_2) = \frac{Var(X_i)}{2} = \frac{1.24}{2} = 0.62$

(f) Suppose a random sample size 100 is selected from the above population, let $\bar{X} = \frac{X_1 + \ldots + X_{100}}{100}$. Based on the CLT, what is the approximate probability that $\bar{X}$ is less than 1.3? $Var(\bar{X}) = \frac{1.24}{100}$, $\bar{X} \approx N(1.4, \sqrt{\frac{1.24}{100}})$. $P(\bar{X} < 1.3) \approx P(\frac{\bar{X} - 1.4}{\sqrt{\frac{1.24}{100}}} < \frac{1.3 - 1.4}{\sqrt{\frac{1.24}{100}}}) = P(Z < -0.90) = 1 - P(Z > -0.90) = 1 - 0.8159 = 0.1841$

Comparing two types of concrete, Type A and Type B.

(a) X is the compressive strength of a randomly selecter type A concrete. Assume normal distribution with mean $\mu_X = 3350$(psi) and s.d. $\sigma_X = 350$(psi) What is the probability that X is less than 3850? $X \sim N(3350, 350)$, $P(X < 3850) = P(\frac{X - 3350}{350} < \frac{3850 - 3350}{350}) = P(Z < 1.43) = 0.9236$.

(b) From (a), what is the probability that X is between 2900 and 3850? $P(2900 < X < 3850) = P(\frac{2900 - 3350}{350} < \frac{X - 3350}{350} < \frac{3850 - 3350}{350} = P(Z < 1.43) - P(Z > 1.29) = P(Z < 1.43) - (1 - P(Z < 1.29)) = 0.9236 - (1 - 0.9015) = 0.8251$

(c) From (a), let Y be the compressive strength of randomly selected type B concrete. Assume normal distribution. $\mu_Y = 3130$(psi), $\sigma_Y = 130$(psi). Calculate $E(\frac{3}{2}X - 2Y + 1500)$ and $\mathrm{Var}(\frac{3}{2}X - 2Y + 1500)$. Specify the distribution of $\frac{3}{2}X - 2Y + 1500$. $Y \sim N(3130, 130)$, $E(\frac{3}{2}X - 2Y + 1500) = \frac{3}{2}E(X) - 2E(Y) + 1500 = \frac{3}{2} \cdot 3350 - 2 \cdot + 1500 = 265$. $\mathrm{Var}(\frac{3}{2}X - 2Y + 1500) = (\frac{3}{2})^2 Var(X) + (-2)^2 Var(Y) = \frac{9}{4} \cdot 350^2 + 4 \cdot 130^2 = 343225$. $\frac{3}{2}X - 2Y + 1500 \sim N(265, \sqrt{343225})$.

(d) Supposed Type A and Type B changed. Randome samples of size 11 and 6, sample means of 3414 and 3173, sample s.d. of 394 and 159. Are the population variance for Type A and B equal? Test at $\alpha = 0.05$. H_0: $\sigma_X^2 = \sigma_Y^2$ versus H_1: $\sigma_X^2 \ne \sigma_Y^2$. $\frac{394^2}{159^2} = 6.14$. $F_{0.025,10,5} = 6.62$. $F_{0.975,10,5} = \frac{1}{F_{0.025,5,10}} = \frac{1}{4.24} = 0.24$. Because $0.24 < 6.14 < 6.62$ do not reject H_0 at $\alpha = 0.05$. Reasonable to assume equal variances.

(e) From (d), suppose equal variances. Construct a 95% two-sided confidence interval for $\mu_X - \mu_Y$. What can be concluded. $s_p^2 = \frac{10 \cdot 394^2 + 5 \cdot 159^2}{15} = 111918$. $95 = 100(1 - \alpha), \alpha = 0.05$. $((x - y) - t_{0.025,15}\sqrt{s_p^2(\frac{1}{11} + \frac{1}{6})}, (x - y) + t_{0.025,15}\sqrt{s_p^2(\frac{1}{11} + \frac{1}{6})}) = ((3414 - 3173) - 2.131\sqrt{s_p^2(\frac{1}{11} + \frac{1}{6})}, (3414 - 3173) + 2.131\sqrt{s_p^2(\frac{1}{11} + \frac{1}{6})}) = (-120.8, 602.8)$ There is no difference between μ_X and μ_Y based on the data (includes zero).

(f) From (d), suppose equal variances. Do the types of concrete have equal means? Decide based on classical approach with $\alpha = 0.05$. Same conclusion as (e)? Why or why not? H_0: $\mu_X = \mu_Y$ versus H_1: $\mu_X \ne \mu_Y$. $t_{obs} = \frac{3414 - 3173 - 0}{\sqrt{s_p^2(\frac{1}{11} + \frac{1}{6})}} = 1.42$. $t_{0.025,15} = 2.131$. Because $-2131 < 1.42 < 2.131$ Do not reject H_0 at $\alpha = 0.05$. Same conclusion as in (e), since 95% CI includes 0, data is consitent with H_0.