MATH 351

1.1 Introduction

Data Types

- Qualitative data
  - ① normminal data: catagorical data, put into catagories (named) no intrinsic order

ranking	majors	f	relf (%f)
2	Bio	5	.25
1	chem	10	.50
3	math	3	.15
4	Physics	2	.10
		$\Sigma = 20$	$\Sigma = 1$ (100%)

  - ③ dichotomous var./data: happen or not (in science), Yes or no, A or B
  - ② ordinal data: data have meanings (ranked)
    ranking 1-4 for majors

- quantitative data (scale variable)
  - ① Interval data: can't ×, ÷ (distance)
    temp of 0°C is not absolute zero (absense of sth), don't have abs zero
    time, etc. (≠ no temperature) val. ≠ 0
  - ② ratio data: any numerical computation (abs. zero)
    absolute zero is meaningful
    i.e. weight

Data Types

- Discrete: data can be counted, countable
  eg. # of cars go thru chick-fil-a

specific values

- Continuous: any value in continuem

Populations, Sample, Simple Random Sample (SRS)

<u>Population</u> parameters

μ = mean of population — point estimater of μ → <u>Sample</u> Statistics: $\bar{x}$ = mean of sample

σ = Standard dev of pop — S = ST Dev. of Sample

N = size of population (infinite pop = too many to count) — n = sample size (A subset of population)

<u>SRS</u>

Items have equally chance of being selected
= sampling w/ replacement
- sampling variation
  2 diff. samples from same population vary
  → SRS not gurranteed to reflect population perfectly

<u>Descriptive Stats</u>

- margin of error
- point est. from sample 92% ± ?
- sample size
- Variation from sample (st. dev, variance) mean, IQR, percentiles
- probablity willing to accept

<u>Sample of convenience</u>

sample not drawn by well-defined random methods
{ random assignment
{ random sampling
= biased sample
i.e. drug tests for volunteers

<u>Conceptual population</u>

all values that might possibly have been observed from a population

why sampling?
Sample representative of population

Research Studies

Outcome: dependent variable (DVAR)
one

Control: independent variable(s)
multiple variables could expain

Research Plans

- Experimental Study
  can explain causation (IV → DV)
  Confounding variables: not part of exp. causing the results
- Observational Study
  not causation, only relationship
  obs. → hypothesis: cross-sectional = short-time
  e.g. doctors asking patients past medical history
  - case-control (retrospective study)
    probs: ppl can't remember things
  - Cohort study: a group of ppl, long period of time
    i.e. retention rate, graduation rate

Journal → lead to expriment.

1.2

Sampling Methods   random sampling, random assignment

(1) Simple Random Sampling (SRS)
each participants. has equal chance of being selected and they're independent of one another
replacement.
eg. 1 0 0 1   one selection doesn't affect another one
50% select 1.
{ independent selection
  random

(2) Stratified Sampling
Stratify population: strata

eg	Strata (age categories)	pop rel f
pop	<18	
Student 30%	19-29	.20
faculty 10%	30-39	.10
Staff 60%	⋮	.15
		⋮

pop of 20,000 = N
n = 1000
rel f × n   then do
(.20)(1000) ← SPS
(.10)(1000)
(.15)(1000)

sample representative of pop.

(3) Systematic Random Sampling
e.g. $N=1000$, $n=100$ $\xrightarrow{N/n}$ $k=10$ (sample every 10th item)
random starting place.
$P = 3, 13, 23, \ldots$
① randomly selected # between 1 ~ k as a start
② random #, ⋯+k, ⋯+2k, ⋯

(4) Cluster Sampling
e.g.

40 counties
randomly select counties
1. one stage clustering
   n = 12 counties (random)
2. Sample every single family (ask) in that county
Second stage: SRS in all streets in the county
Stage 2 clusters: after pick SPS, cluster again

(5) Convenience Sampling
Volunteer, self-select for the study

Descriptive Statistics

Sample Statistics

mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ pt estimator of μ

Median n is odd, $\frac{n+1}{2}$; n is even $\frac{n}{2}$ and $\frac{n}{2}+1$
50% percentile

used w/ data: outliers, influencial data points

Mode

pth percentile   P% at or less
75th percentile   75% ppl at that value or lower
25% ppl above it

location $i = (P\%)(n+1)$ n = sample size
= not integer e.g. 6.25
avg 6th. 7th number (two sides)

deciles (10, 20, ⋯)
25th, 50th, 75th. ..

Trimmed Mean   extreme values
better idea of avg. based on middle population (P%)

$\frac{nP}{100}$ items take off

n = 24 items
20% trimmed mean
$\frac{nP}{100} = 4.8 \approx 5$
round the nearest whole # and trim that amt

Sample Statistics : Measures of dispersion

range

IQR 75th percentile − 25th percentile

Variance : Squared units of $(x-\bar{x})$

avg squared deviation from the mean

(sample) $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1 \text{ (df)}}$ degrees of freedom

amt of freedom you have when making estimate

e.g. avg of 3#s = 100 , sum of 3#s = 300

50, 150, 100 $n=3$, $df=2$

(no freedom to pick the last one)

x	$x-\bar{x}$	$(x-\bar{x})^2$
1	−2	4
2	−1	1
3	0	0
4	1	1
5	2	4

$\frac{\Sigma = 10}{df: n-1} = \frac{10}{4} = 2.5$

ways to correct) to estimate correctly

σ^2 pop variance

otherwise it's a biased estimator

Standard dev. $\sqrt{\text{variance}}$ $s = \sqrt{s^2}$ Sample σ std dev. Population

1.3 Graphical Summaries : Visualing the data

(1) Stem-and-leaf plot

visual rep

last digit

3 | 2
4 | 5
4 | 0

Stem	Leaf
4	2 5 9
5	0 1 1 1 1 3 ...
6	
7	

(2) Dot plot

(3) Histogram

Qualitative data : pie, bar chart

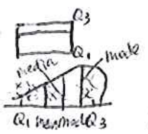
Quantitave data : classes (5-20)

real class limits (so bars merge together)

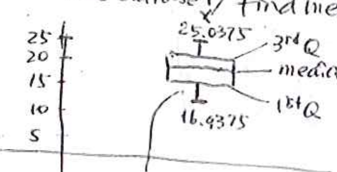
40-49 ⟶ 39.5-49.5

(4) Box plot

take dataset / find median



Skew to the left



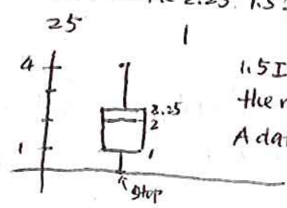
to see central location of your data

$1.5(Q_3-Q_1) = 1.5\ IQR$

can't determine mean

i.e. medium 2 range 4 (0-4)

75 percentile 2.25 1.5IQR 1.875



1.5IQR goes beyond the range ⟶ whiskers stop at min/max value

A data point goes beyond 1st Q − 1.5 IQR or 3rd Q + 1.5 IQR ⟶ × outlier

(5) Scatterplot : correlation. 2 variables

2.1 Counting Methods (discrete probability)

1. Permutations : order matters

5 items

A B C D E

↓ n=3

ABC
ACB
CBA
...

$_nP_r = \frac{n!}{(n-r)!} = \frac{5!}{2!} = 5 \cdot 4 \cdot 3$

$= 60$ ways

n: total, r: items to pick

note: $0! = 1$

2. Combinations : no order

A B C D E

↓ n=3

ABC BCD CDE
D E
E

AC

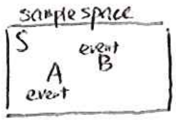
$_nC_r = \frac{n!}{r!(n-r)!}$

$= \frac{5!}{3!\,2!} = 10$

Experiment

any process that yields outcomes

classical or empirical



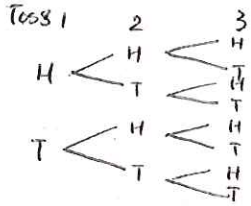
Event

subset of experiment

Sample Spaces

all possible outcomes from exp.

Classical: Toss a coin 3 times

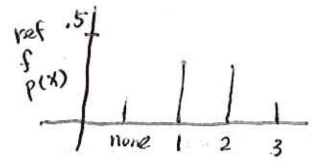


$S = \{HHH, HHT, HTH, \dots\}$ sample spaces

x	$P(x)$
0	$\frac{1}{8} = 0.125$
1	$\frac{3}{8} = .375$
2	$3/8 = .375$
3	$1/8 = .125$

discrete probability distribution

let x = # heads in toss of coin 3 times



like normal distribution

Empirical: base upon observations

1000 manufacturing items

Defective/Not

random var.

x	f	$P(x)$
D	5	5%
N	95	95%

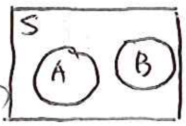
$n = 100$

Sampling plan

Pick items at random

Stop when get defective item

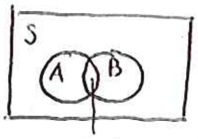
$\{D, ND, NND, \dots\}$



(disjoint event)

Event A, B nothing in common

mutually exclusive



$A \cap B$

not mutually exclusive

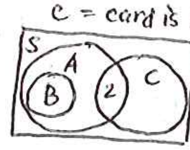
$P(A \overset{or}{\cup} B) = P(A) + P(B)$ (if M.E.)

$P(A \cup B) = P(A) + P(B) - P(A \overset{and}{\cap} B)$ (if not M.E.)

Example: Event A = red card

B = J, Q, K of diamonds

C = card is Ace



Probability Rule

S = Sample Space $P(S) = 1$ $\begin{cases} \text{Continuous} & \text{Area under curve} = 1 \\ \text{discreate} & \text{adds up to } 1 \end{cases}$

A = Event $0 \le P(A) \le 1$

A_1, A_2, A_3, M.E. $P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$

2.3 Conditional probability: based on part of sample space

$P(A/B) = \frac{P(A \cap B)}{P(B)}$ A given B, given

dependent Events or $\boxed{\begin{array}{l} P(B/A) \cdot P(A) = P(A \cap B) \\ P(A/B) \cdot P(B) = P(A \cap B) \end{array}}$ if dependent Events

$P(B/A) = \frac{P(A \cap B)}{P(A)}$

$\boxed{P(A \cap B) = P(A) \cdot P(B)}$ if independent

Complement of event A $P(A \text{ not happening}) = 1 - P(A) = A'$ A^c $\sim A$

Ex.

	Non-Sm.	Mod Sm.	Heav Sm	Tot.
Hypertesion	15	25	20	60
Non-Hyp.	10	15	15	40
total	25	40	35	100

Marginal Probability $\rightarrow$ P(non-smokers) = 25/100

(margin out some discarded Variables) $P(\text{mod smk} \cap H) = 25/100$

$P(H/\text{mod smk}) = \frac{P(H \cap S)}{P(MS)} = \frac{.25}{.40}$

Ex. 2.3/17

Sampling of 100 ppl

		Gene 2 D	Gene 2 R	
Gene 1	D	56	24	80
	R	14	6	20
		70	30	$\boxed{100}$

(1) $P(\text{Gene1}=D) = 80\%$ } marginal prob.

(2) $P(\text{Gene2}=D) = 70\%$

(3) $P(\text{gene2}=D/\text{gene1}=D) = \frac{56}{80} = 70\%$ Conditi. prob.

$P(\text{gene1}=D/g2=D) = \frac{56}{70} = 0.8$

$P(\text{gene1}=D) = \frac{80}{100} = p(\text{gene1}=D/g2=D) \Rightarrow$ independent

$\begin{cases} P(A\cap B) = P(A)\cdot P(B) \\ P(A/B) = P(A), \text{ if } A; B \text{ independent.} \end{cases}$ $A \xrightarrow[\text{defect}]{X} B$ indp. and M.E. have no relations!

$P(A\cap B) = 0$, if A, B. Mutually exclusive

$A \rightarrow \times B/A^c$

Bayes' Theorm

$P(A\cap B) = P(A/B)\cdot P(B)$

$P(A\cap B) = P(B/A)\cdot P(A)$

$P(A/B)\cdot P(B) = P(B/A)\cdot P(A)$

$\boxed{P(A/B) = \frac{P(B/A)\cdot P(A)}{P(B)}}$ prior prob.

Posterior prob.

Law of Total Probability

A_1, A_2 mutually exclusive, exhausive

$P(B) = P(B\cap A_1) + P(B\cap A_2)$ general,

$= P(B/A_1)\cdot P(A_1) + P(B/A_2)\cdot P(A_2)$

Bayes' Theorm

$A_1, A_2, \cdots A_n$ mutually exclusive, exhaustive

$$\boxed{P(A_k/B) = \frac{P(B/A_k)P(A_k)}{\sum_{i=1}^{n} P(B/A_i)P(A_i)}}$$

Ex. 1. 4 suits

A → 10 JQK face

52 cards

$P(\text{King}/\text{Face}) = \frac{\overset{1}{P(F/K)}\cdot \overset{1/13}{P(K)}}{P(F)\ \frac{3}{13}} = \frac{1}{3}$

Ex 2 Measles + Flu = Rash

① all event: exhausted → Bayes' Theorm

② mutally exclusive

$P(R/M) = 0.95$ $P(R/F) = 0.08$

$P(R) = ?$ $P(M\cap R) + P(F\cap R) = P(R/M)\cdot P(M) + P(R/F)\cdot P(F)$

$P(M/R) = ?$ $= 0.95\cdot 0.1 + 0.08\cdot 0.9 = 0.167$

$P(M/R) = \frac{P(M\cap R)}{P(R)} = \frac{P(R/M)\cdot P(M)}{P(R)}$

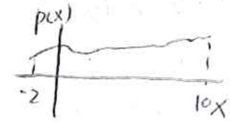
$= \frac{0.95\cdot 0.1}{0.167} = 0.5689$

2.4

Random Variables = assign each outcome w/ a numerical value { discrete / continuous

Probability Distribution (Prob { density / Mass function)

- discrete distribution: counting variable
  only certain outcome can happen
  i.e. binomial, discrete, over an event (Poisson)
- Continuous distribution
  when $n \to$ large, we can proximate discrete to continuous

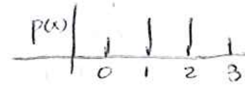


toss a coin 3 times

$X = \#H$	$p(x)$
0	
1	
2	
3	

$\Sigma p(x) = 1$

can get exact val. for a pt



looking for Area under the Curve.
Sum of area = 1

can't get exact val. for a pt but can get in a range (random var)

Discrete Dist.

create distribution → function → graph

Cumulative Distribution Function $F(X) = P(X \le x)$

pmf: prob. mass function

ex. 1

x = # of flaws	$P(x)$
0	.48
1	.39
2	.12
3	.01

2 conditions each $p(x)$ has to be
$\begin{cases} 0 \le P(x) \le 1 \\ \sum P(x) = 1 \end{cases}$

→ discrete → Cumulative

$F(0) = P(X \le 0) = 0.48$

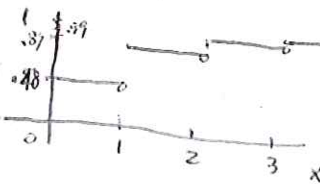
$F(1) = P(X \le 1) = 0.87$

$F(2) = 0.99$

$F(3) = 1$

x	cumulative probabilities
$X \le 0$	.48
$X \le 1$	.87
$X \le 2$	.99
$X \le 3$	1

$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ .48, & 0 \le x < 1 \\ .87, & 1 \le x < 2 \\ .99, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases}$



$x^2 p(x)$	x^2	x	$p(x)$	$xp(x)$ x-1.6
0	0	0	0.48	0
.39	1	1	0.39	0.39
.48	4	2	0.12	0.24
.09	9	3	0.01	0.03

$\Sigma = .96$

$\boxed{\mu = \sum x \cdot p(x)}$ mean

$\sum x \cdot p(x) = 1.6$

$\boxed{\begin{cases} \sigma^2 = \sum (x-\mu)^2 \cdot p(x) \\ \sigma^2 = \left[\sum x^2 \cdot p(x)\right] - \mu^2 \end{cases}}$ variance

ex2.

x	$p(x)$
0	0.16
1	0.48
2	0.36

$\mu = \sum x \cdot p(x)$
$= 0.48 + 0.72 = 1.2$

$\sigma^2 = \left[\sum x^2 p(x)\right] - \mu^2$
$= (0.48 + 1.44) - 1.44$
$= 0.48$

$\sigma = 0.693$

prove $\quad \Sigma p(x) = 1$

$\sigma^2 = \sum (x-\mu)^2 p(x)$
$= \sum\left[x^2 p(x) + \mu^2 p(x) - 2x\mu p(x)\right]$
$= \sum\left[x^2 p(x) + \mu^2 p(x) - 2\mu^2\right]$
$= \sum\left[x^2 p(x)\right] + \mu^2 - 2\mu^2 = \sum (x^2 p(x)) - \mu^2$

Continuous Dist.

pdf: probability density function

can't get probability for one value only.

$f(x) \ge 0$

$\int_{-\infty}^{\infty} f(x)dx = 1$

$P(a < x < b) = \int_a^b f(x)dx$

ex. $f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{else} \end{cases}$

$\int_{-1}^{2} \frac{x^2}{3} dx$

$= \left.\frac{x^3}{9}\right|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1$

verify $f(x)$ is density func.

$$\boxed{\begin{aligned} &\text{mean} \quad \mu_x = \int_{-\infty}^{\infty} x f(x)\,dx \\ &\text{variance} \quad \sigma_x^2 = \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x)\,dx \end{aligned}}$$

or $\sigma_x^2 = \left[\int_{-\infty}^{\infty} x^2 f(x)dx\right] - \mu_x^2$

probability or percentile

$x = 6$

$f(x)$ prob. density func.

$P(x \le 6) = P(x < 6)$

can't find exact value $P(x=6)$

→ Continuous → cumulative

$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$

in seperate parts

refer to P105 Example 2.42

pth percentile

$P(X \le x_p) = \int_{-\infty}^{x_p} f(x)dx = \frac{P}{100}$

median $P = 50$

1.1 SAMPLING

Central tendency = mean, median, mode, percentiles
Dispersion: range, IQR, Variance, Stan. Dev.

Types of Measurement

Nominal, ordinal, interval, ratio

1.2-1.3 Graphical Summaries

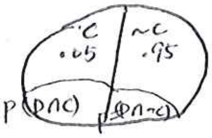
Quantitative
· histogram, dot plot, Stem/leaf, box plot, scatterplot

Qualitative
bar/Pie

2.1-2.3 Prob. Concepts

Union, intersec, conditional, Bayes Thm.

Bayes Theorm Review:

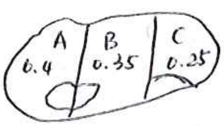


C = Cancer = Event C
D = Diagnosis = Event D

$P(C) = .05$
$P(\sim C) = .95$

$$P(D) = P(D\cap C) + P(D\cap \sim C)$$
$$= P(D/C)\cdot P(C) + P(D/\sim C)\cdot P(\sim C)$$
$$= (.78)(.05) + (.06)(.95)$$
$$= .096$$

2.



$P(A) = 0.4$
$P(B) = 0.35$
$P(C) = 0.25$

$P(O/A) = 0.05$
$P(O/B) = 0.03$
$P(O/C) = 0.15$

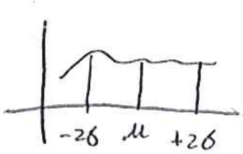
(1) $P(C/O) = \frac{P(C\cap O)}{P(O)} = \frac{P(O\cap C)}{P(O)} = \frac{P(O/C)\cdot P(C)}{P(O\cap A) + P(O\cap B) + P(O\cap C)}$

$= \frac{(0.15)(0.25)}{(0.05)(0.4) + (0.03)(0.35) + (0.15)(0.25)} \approx .55$

(2) $P(A/O) = \frac{P(A\cap O)}{P(O)} = \frac{P(O/A)\cdot P(A)}{P(O)} = \frac{(.05)(.4)}{P(O)} \approx .294$

Chebyshev's Inequality

any cont. prob. dist., at least $1-\frac{1}{k^2}(\%)$ of all data/ of the observations, is within k St. dev of mean, $k > 1$.



$1 - \frac{1}{2^2} = \frac{3}{4} = 75\%$
data

$1 - \frac{1}{3^2} = \frac{8}{9}$

normal dis. ±3σ = 99%

Ex. 1. a. mean = 28 min
σ = 8 min

$1 - \frac{1}{2.5^2} = 84\%$ lies in ±2.5σ

range: 8-48 min

b. $1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9} = 89\%$
$28 \pm (3)(8) = 4 \pm 52$

c. 12-44 min
2σ
(2 St. dev)

$28 \pm (x)(8)$
$x = 2$

2.5 Random Variables : adding constants

- If a constant b is added to a random Var. x

$$\mu_{x+b} = \mu_x + b$$
$$\sigma^2_{x+b} = \sigma^2_x$$

- multiple a constant a

$$\mu_{ax} = a\mu_x$$
$$\sigma^2_{ax} = a^2\sigma^2_x$$
$$\sigma_{ax} = |a|\sigma_x$$

x	$x-\mu$	$(x-\mu)^2$	$5x$	$(5x-\mu_{5x})^2$
1	-2	4	5	100
2	-1	1	10	25
3	0	0	15	0
4	1	1	20	25
5	2	4	25	100

$\mu_x = 3$

$\sigma^2 = \frac{10}{5} = 2$ $\mu_{5x} = 15$ $\sigma^2_{ax} = \frac{250}{5} = 50$

Population $\sigma^2 = \frac{\sum(x-\mu)^2}{N}$ $N \rightarrow$ population size

- altogether

$$\mu_{ax+b} = a\mu_x + b$$
$$\sigma^2_{ax+b} = a^2\sigma^2_x$$
$$\sigma_{ax+b} = |a|\sigma_x$$

- more

$$\mu_{ax+by} = \mu_{ax} + \mu_{by} = a\mu_x + b\mu_y$$

- $X_1, \dots X_n$ independent $P(A\cap B) = P(A|B)\cdot P(B) = P(A)\cdot P(B)$ (random/replacement) Prob. of one event doesn't affect another

$$\sigma^2_{X_1+X_2+\dots+X_n} = \sigma^2_{X_1} + \sigma^2_{X_2} + \dots + \sigma^2_{X_n}$$
$$\sigma^2_{C_1X_1+C_2X_2+\dots+C_nX_n} = C_1^2\sigma^2_{X_1} + C_2^2\sigma^2_{X_2} + \dots + C_n^2\sigma^2_{X_n}$$

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$
$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$$

2.5/1 $\mu_x = 9.5$, $\mu_y = 6.8$, $\sigma_x = 0.4$, $\sigma_y = 0.1$

mean and std. dev of

(1) $3x$

$\mu_{3x} = 3\mu_x = 3(9.5) = 28.5$

$\sigma_{3x} = \cancel{\sqrt{3}\sigma_x}$

$= \sqrt{3^2\sigma^2_x} = 3(0.4) = 1.2$

(2) $y-x$

$\mu_{y-x} = \mu_y - \mu_x$

$= 6.8 - 9.5 = -2.7$

$\sigma_{y-x} = \sqrt{\sigma^2_y + \sigma^2_x} = \sqrt{0.01+0.16} = \sqrt{0.17} = 0.4123$

(3) $x+4y$ $\mu_{x+4y} = \mu_x + 4\mu_y$

$= 9.5 + 4(6.8) = 9.5 + 27.2 = 36.7$

$\sigma_{x+4y} = \sqrt{\sigma^2_x + 16\sigma^2_y}$

$= \sqrt{0.16 + 16(0.01)}$

$= \sqrt{0.32} \approx 0.566$

2.5/4 $V = V_1 + V_2$

$\mu_{V_1} = 12V$ $\mu_{V_2} = 6V$

$\sigma_{V_1} = 1V$ $\sigma_{V_2} = 0.5V$

find: μ_V, σ_V

$\mu_V = \mu_{V_1} + \mu_{V_2} = 18V$

$\sigma_V = \sqrt{\sigma^2_{V_1} + \sigma^2_{V_2}} = \sqrt{1+0.25} = \sqrt{1.25} = 1.118V$

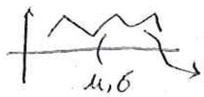
Quality mean of sample
Unbiased estimator

Sampling Distribution

population N

Sampling distribution of the means

sample size > 30 $\xrightarrow{\text{normal distribution}}$

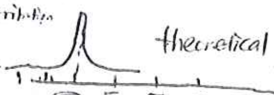


sample size n_1, $\bar{x}_1$, S_1 — point estimator, Unbiased, Consistent

n_2, $\bar{x}_2$, S_2

⋮

forever w/ replacement



n=30 or more ⇒ normal distrib.

Central Limit Theorm

$\mu_{\bar{x}_1}$ = mean of the mean dist. = μ = mean of pop.

$\sigma_{\bar{x}_1}$ = std dev of means/sampling distribution

Central Limit Theorm

$\mu_{\bar{x}} = \mu$ mean of sampling distribution = mean of pop.

Stand. Error = std. dev. of samp distrib. of the means $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ = Standard error

σ pop std dev (biased estimator of pop)

n sample size you pick

n increases

2.5/8

pop

$\mu = 20.01$

$\sigma = 0.02$

sample
$n = 24$

a. mean of total volume $\mu_{TV} = (20.01)(24) = 480.24$

b. Std dev. $\sigma_{TV} = \sqrt{24(0.02)^2}$

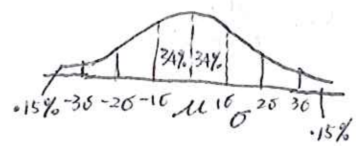
c. $\mu_{\bar{x}} = \mu = 20.01$

d. $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{0.02}{\sqrt{24}}$

e. $\sigma_{\bar{x}} = 0.0025 = \dfrac{\sigma}{\sqrt{n}} = \dfrac{0.02}{\sqrt{n}}$ $\quad \sqrt{n} = \dfrac{0.02}{0.0025} = \dfrac{200}{25} = 8$

$n = 64$

Normal Distribution



≈ 68% data is within 1σ of mean

≈ 95.4% data is within 2σ of mean

≈ 99.7% data is within 3σ

STANDARD NORMAL DISTRIBUTION

convert normal dist. thru z. score

random vari.

$z = \dfrac{x - \mu}{\sigma}$ # std dev from mean

$\sigma = 1$

$\mu = 100$, $\sigma = 10$

$z = \dfrac{90 - 100}{10} = -1$

Converted Value.

↓

percentile rank

4.2 Binomial Distribution (discrete)

- Binomial Prob. Distb. is a Bernouilli Process. (The Binomial Experiment)

$x \sim BIN(n, p)$ — n: # of trial / samples; p: prob. of success (in population)

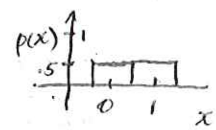
Bernouilli Process

1. Exp. fixed # of trial, n — 1 = success, 0 = failure
2. Only two possible Outcomes $P(1) = p$, $P(0) = 1-p$

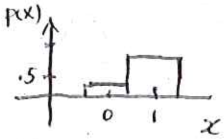
trials independent, mutually exclu. outcome

i.e. Biology research: diseases / No diseases, etc.

e.g. $n = 1$ (trial) — flip coin: 1 = get a head/succ., 0 = get a tail/failure, $\mu = 0.5 = p$



e.g. $P(\text{pass}) = .8$, $P(\sim\text{pass}) = .2$



x	P(x)
1	p
0	$1-p$

$\mu_x = \sum x\,p(x) = p$

$$\sigma_x^2 = \sum (x-\mu)^2 p(x) = (1-p)^2 p + (0-p)^2(1-p) = (1-p)^2 p + p^2(1-p) = p + p^3 - 2p^2 + p^2 - p^3 = p - p^2 = p(1-p)$$

Binomial = $n \to \infty$

$P(1) = p$

$P(0) = 1-p$

$$\boxed{\begin{array}{l}\text{For } n=1 \text{ trial, } \mu_x = p \text{ and } \sigma_x^2 = p(1-p) \\ \text{For } n \text{ trials, } \mu_x = np \text{ and } \sigma_x^2 = np(1-p)\end{array}}$$

$$\boxed{P(x \text{ successes in } n \text{ trials}) = {}_nC_x\, p^x (1-p)^{n-x}}$$ (not cumula.)

$${}_nC_x = \frac{n!}{x!(n-x)!}$$

prob. $P(\text{recover}) = 0.4$, 8 have disease

(a) $P(x \geq 5) = P(5) + P(6) + P(7) + P(8)$

$= \frac{8!}{5!3!}(0.4)^5(0.6)^3$

or use the table: $P(x \geq 5) = 1 - P(4)$ cumulative in table

x	P(x)
0	
1	
2	
3	
4	
5	
6	
7	
8	

(b) $P(6) = P(6) - P(5)$

(c) $P(x \leq 3) = 0.594$

(d) $\mu = np = (8)(0.4) = 3.2$

$\sigma^2 = n \cdot p(1-p) = 3.2(0.6)$

ex.

B	1	A B C
A	2	A B C
C	3	A B C
B	4	A B C

x = # correct/success

x	f	P(x)
0	3	
1	11	
2	6	
3	2	
4	2	
	24	

$P(4 \text{ correct}/4) = (\frac{1}{3})^4 = \frac{1}{81}$

Sample $\xrightarrow{\text{estimate}}$ population

use the sample proportion to est. success prob.

Pop

proportion of successes in a pop. $= p$

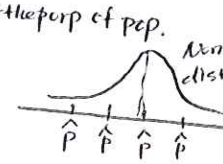
sample → $\hat{p}$: est. of the porp of pop.

$\hat{p} = \frac{x}{n}$ — x: # of success; n: # of trials

μ { mean; p { proportion of population

not quantitative

$E(\hat{p}) = p$



Std. error: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

$E(\hat{p}) = p$

Ex. 4.11, P209

See P210 for summary

Ex. Prob. $n=100$, 7 flaws

$\hat{p} = \frac{7}{100}$

$n = 10{,}000$

$P_{decay} = 0.0002/min$

4.3 Poisson Distribution

Poisson Distrib $\approx$ approx. $\Leftarrow$ when $n \to$ large, success $\to$ small

$$\left\{ \begin{array}{l} P(x=3) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ = \frac{10000!}{3!\,9997!} (0.0002)^3 (0.9998)^{9997} \end{array} \right.$$

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

λ is the mean of the distrib

$\lambda = np = (10{,}000)(0.0002) = 2$

$\approx e^{-2}\left(\frac{2^3}{3!}\right) = .1805$

prob. $n = 1000$

$p = 0.03$ $\quad \lambda = np = 30$

$P(2) = e^{-30} \frac{30^2}{2!}$

$X \sim poisson(\lambda)$

$$\boxed{\begin{array}{l} \mu_x = \lambda \\ \sigma_x^2 = \lambda \end{array}}$$

- Descriptive Stats
- Probability $\longrightarrow$ Prob distribution
  normal distrib.

test $\left\{ \begin{array}{l} \text{parametric test (parameter of pop.)} \\ \text{non-parametric test} \end{array} \right.$

Inferencial Stats (Confidence interval)

Pt estimator prob. willing to accept (99, 95, 90% precision)

4.5

Normal distribution

2 continuous distrib.

$\left\{ \begin{array}{l} \text{uniform distrib.} \\ \text{normal distrib} \end{array} \right.$

Probability Density Func.

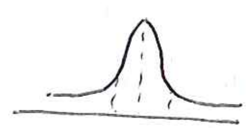
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

based on σ, μ

Kurtosis: how tall/high dist goes

Platy kurtic

σ large



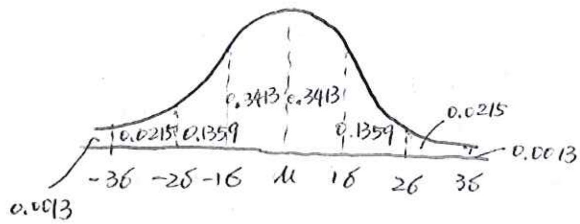
leptokurtic (Gaussian Dist.)

σ small

- Symmetric μ, median, mode
- Inflection points $\mu - \sigma$ and $\mu + \sigma$ (concave up / down)

68.3 %	$\mu \pm \sigma$
95.4 %	$\mu \pm 2\sigma$
99.7%	$\mu \pm 3\sigma$

1. Calculating Probabilities

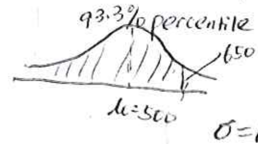


z score: convert into standard normal distrib.

$$z = \frac{x-\mu}{\sigma}$$

mean $\mu = 0$, $\sigma = 1$

e.g.



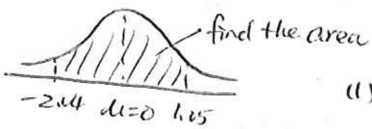
$\sigma = 100$

$$z = \frac{650-550}{100} = 1.5$$

look on table $\to$ 93.3%

table = area to the left

E.g. 1



(1) $A_{z=1.05} - A_{z=-2.24} = .8324$

(2) $P(-1.67 < z < 2) = .9297$

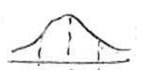
(3) Area to the left of $z = -0.24$ OR to the right of $z = 1.20$

$A_{z=.24} = .4452$

$A = 1 - A_{z=1.20} = A_{z=-1.20} = .5203$

(2 ways)

2. x, normal distrib. $\mu = 8$, $\sigma = 2$



(1) $P(x<5)$

$$z = \frac{5-8}{2} = -\frac{3}{2} = -1.5$$

$P(x<5) = A_{z=-1.5} = 0.0668$

(2) $P(5<x<10) = A_{z=1} - A_{z=-1.5} = 0.7745$

$$z = \frac{10-8}{2} = 1$$

(3) percentile score for $x = 10.5$

$$z = \frac{10.5-8}{2} = \frac{2.5}{2} = 1.25 \quad A_{z=1.25} =$$

(4) $P(x>14)$

0.0013

Eg 3. $\mu = 10$, $\sigma = 3$

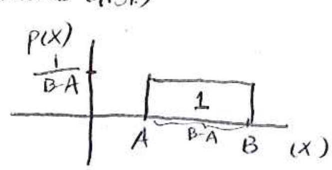
(a) $P(x) = 93\% = 0.93$ $\quad z = 1.48 = \frac{x-\mu}{\sigma} = \frac{x-10}{3}$ $\quad x = 14.44$

(b) $P(X \geq x) = .826$ $\quad P(\cancel{X} < \cancel{x}) = 1 - 0.826 = 0.174$ $\quad z = -0.94 = \frac{x-10}{3}$ $\quad x = 7.18$

See HW 4.5 3(e)

4.8 Uniform Probability Distribution (continuous dist)

$$f(x) = \begin{cases} \frac{1}{B-A} & \text{for } A \leq x \leq B \\ 0 & \text{else} \end{cases}$$



$$\boxed{\mu = \frac{A+B}{2} \qquad \sigma^2 = \frac{(B-A)^2}{12}}$$

derivation

$$\mu = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_{-\infty}^{+\infty} (x)\left(\frac{1}{B-A}\right)dx = \frac{1}{B-A}\frac{x^2}{2}\Big|_A^B = \frac{B+A}{2}$$

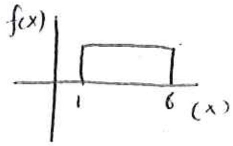
$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 f(x)\,dx - \mu^2 = \int_{-\infty}^{+\infty} x^2\left(\frac{1}{B-A}\right)dx - \frac{(A+B)^2}{4} = \frac{1}{B-A}\frac{x^3}{3}\Big|_A^B - \frac{(A+B)^2}{4}$$

$$= \frac{B^3-A^3}{3(B-A)} - \frac{(A+B)^2}{4}$$

$$= \frac{\cancel{(B-A)}(B^2+AB+A^2)}{3\cancel{(B-A)}} - \frac{(A+B)^2}{4} = \frac{4B^2+4A^2+4AB-3A^2-3B^2-6AB}{12} = \frac{(B-A)^2}{12}$$

Eg. rxn time x, $1 \le x \le 6$

(a) density func.



(b) $P(2 < x < 6) = \frac{4}{5} = 0.8$

(c) $\mu = \frac{6+1}{2} = 3.5$ $\quad \sigma^2 = \frac{(5)^2}{12} = \frac{25}{12}$

$\sigma = 1.44$

4.5 cont'

4.11

Calculating Binomial Prob.

$P(x \text{ succ. in } n \text{ trials}) = {}_nC_x \, p^x (1-p)^{n-x}$

$n = 100$
$p = .2$
$1-p = .8$

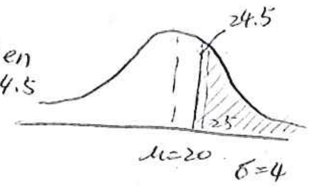
$P(x \le 25)$ use 25.5

$P(x \ge 25) = 1 - (P(x=25) + P(x=24) + P(x=23) + \cdots + P(x=0))$ using z when $x = 24.5$

$P(x > 25) = ?$ use $x = 25 \Rightarrow z$

$P(x = 18)$: Area between 17.5 and 18.5 or use binomial

normal approx. of binomial



Using cont. to approx discrete dist.

- when $np(1-p) \ge 10$
- find μ, σ of binomial $\quad \mu = np = 20$
- find z-score w/ correction factor and look up probabilities $\quad \sigma = \sqrt{np(1-p)} = 4$

Correction factor = 0.5

$z_{24.5} = \frac{24.5 - 20}{4} = 1.125 \approx 1.13$

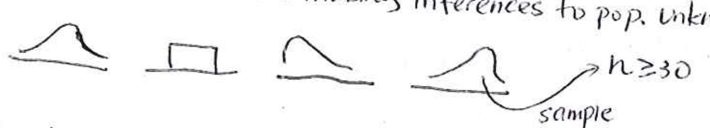
$A = 0.1292$

4.11

CLT

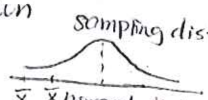
Hypothesis Testing

Inferential Statistics = making inferences to pop. unknown



Sampling distrib.

sampling distrib.



normal dist. data

$\mu_{\bar{x}}$: mean of means

$$\begin{cases} \mu_{\bar{x}} = \mu \\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \end{cases}$$ sampling error

Sampling w/ displacement random sample

$\Downarrow$

independent

$\underline{\underline{n}}$ $\underline{Pop}$ $\underline{\text{parameters}}$

take sample info $\xrightarrow{\text{estimate}}$ pop parameters

binomal dist.
normal distri
skewed
...

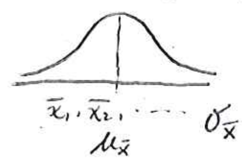
μ, σ (quantitive)
p (qualitative, proportion of success)

Setup Sampling distribution (of the means)

$n \geq 30$ large sample.

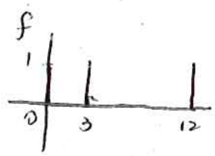
n : $\bar{x}_1, s_1$
$\bar{x}_2, s_2$
...
pop. μ

unbiased
point estimator
if random + replacement



$$\boxed{\begin{aligned}\mu_{\bar{x}} &= \mu \\ \sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}}\end{aligned}}$$

$\underline{ex}$. $n=3$, value: $0, 3, 12$.



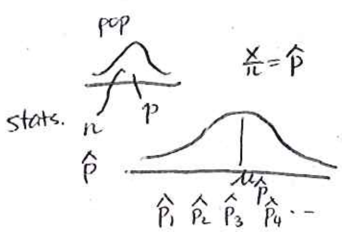
$\mu = 5$
$\sigma = 5.1$

samples of size 3
$\mu_{\bar{x}} = 5 = \mu$
$\sigma_{\bar{x}} = 3 \sim \sigma = \frac{5.1}{\sqrt{3}} = 2.94$

finite population n close to pop size
$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}} \cdot \frac{\sigma}{\sqrt{n}}$ (rare)
if $n < 0.05N$, correction factor ignored

Central Limit theorem : random sample of size n w/ pop. mean μ
- Shape of samping distribution of $\bar{x}$ becomes normal as n increases
- $\mu_{\bar{x}} = \mu$
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

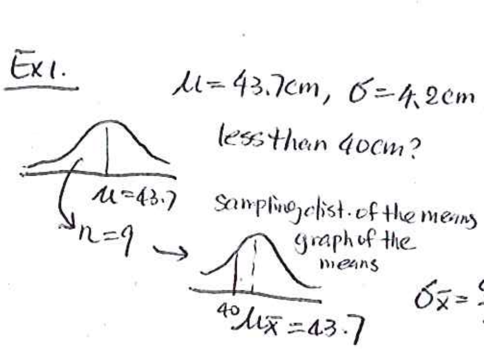
Sampling distribution of the proportions : random sample of size n w/ pop proportion, p



$\frac{x}{n} = \hat{p}$

- shape is normal as long as $np(1-p) \geq 10$
- $\mu_{\hat{p}} = p$
- $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

$\underline{Ex1.}$ $\mu = 43.7$cm, $\sigma = 4.2$cm, probability less than 40cm?



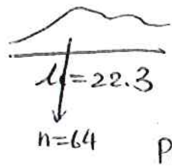
$\sigma_{\bar{x}} = \frac{4.2}{\sqrt{9}} = 1.4$cm

$$\boxed{z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}}$$ z for samp. dist.

$z = \frac{40 - 43.7}{1.4} = -2.64$

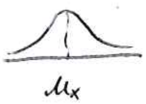
$P = 0.0041$

Ex 2.



Probability of avg age > 23.

$P(\bar{x} > 23)$



$\mu_{\bar{x}} = \mu = 22.3$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{64}} = 0.5$

$z = \frac{23 - 22.3}{0.5} = \frac{0.7}{0.5} = 1.4$

$P(\bar{x} > 23) = 1 - P_{z=1.4} = P_{z=-1.4} = 0.0808$

5.1 Confidence Intervals for the mean

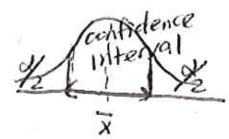
$\bar{x}$

large samples ($n \geq 30$)

$\bar{x} \pm Z_{\frac{\alpha}{2}} \left(\frac{s}{\sqrt{n}}\right)$

point est. for σ

68% Confidence $Z_{\frac{\alpha}{2}} = \pm 1$
90% $= \pm 1.645$
95% $= \pm 1.96$
99% $= \pm 2.58$
99.7% $= \pm 3$



$\alpha = 1 - CI$

α: level of significance

90% confidence $\alpha = 1 - 0.9 = 0.1$

$\frac{\alpha}{2} = 0.5$

$Z_{\frac{\alpha}{2}} = 1.645$

Sample

CI for μ
where the mean lies

Ex3 $n = 123$, $\bar{x} = 136.9$, $s = 22.6$

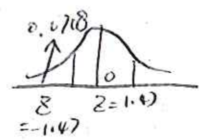
(a) 95% CI

$\bar{x} \pm Z_{\frac{0.05}{2}} \left(\frac{s}{\sqrt{n}}\right) = 136.9 \pm 1.96 \left(\frac{22.6}{\sqrt{123}}\right) = 136.9 \pm 3.99$

(b) Level of Confidence for 133.9 - 139.9.

$136.9 \pm Z_{\frac{\alpha}{2}} (2.04)$

$Z_{\frac{\alpha}{2}} = \frac{3}{2.04} = 1.47$



$1 - (0.0708)(2) = 85.84\%$

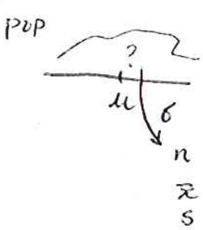
(c) $n = ?$ so that CI = 99%, Specify the mean to ±3 min?

$\bar{x} \pm 3$ min.

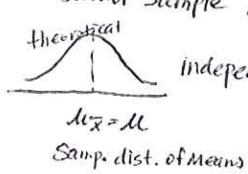
$\bar{x} \pm (2.58)\left(\frac{s}{\sqrt{n}}\right) = 3 \pm \bar{x}$ ($s \to 22.6$)

$\Rightarrow n \approx 380$

based on sampling distribution { means / proportions



μ = large sample $n \geq 30$
small sample $n < 30$ (req. pop. normal)



independent: replacement & random

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (error)

If σ unknown $\frac{s}{\sqrt{n}}$ as point estimator

C.I.

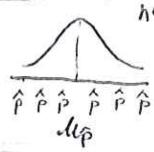
μ { large n. → $Z_{\frac{\alpha}{2}}$ (normal)
small n. → $t_{\frac{\alpha}{2}}$, df, σ must be unknown, pop ~ N

quanti.

p { always z $np(1-p) \geq 10$ if met
qualit. nothing else

5.2 Confidence Interval of Proportion

$\rightarrow n, \hat{p}, (1-\hat{p})$



$\mu_{\hat{p}} = p$

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ standard error of proportion

$\boxed{\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ for $np(1-p) \geq 10$

Ex 4 $\quad n = 100, \hat{p} = 73\%$

(a) $CI = 95\%$

$\hat{p} \pm z_{(\frac{1-.95}{2})} \left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$ $\quad 0.73 \pm 1.96 \sqrt{\frac{(0.73)(0.27)}{100}}$ $\quad 0.73 \pm 0.087$

$0.643 - 0.817$

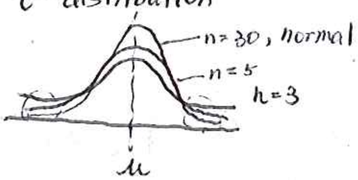
large $n \rightarrow$ estimate μ

small n ($n < 30$) (σ is unknown) (Pop $\sim N$) – 3 criteria

1) $n < 30$
2) σ unknown
3) Pop $\sim N$
   - know in advance
   - plot, if data $\approx$ symmetric and no outliers

5.3 Small Sample ~~confidence~~ Intervals of the mean

t-distribution



every distrib. is different depending on n

$CI = \bar{x} \pm t_{n-1, \frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

Ex. $\quad n = 8, \bar{x} = 3410.14, s = 1.018$

95% C.I.?

$\bar{x} \pm \underbrace{t_{\frac{\alpha}{2} = 0.025} (7df) \left(\frac{s}{\sqrt{n}}\right)}_{\text{margin of error}}$

$3,410.14 \pm (2.365)\left(\frac{1.018}{\sqrt{8}}\right)$

5.3 / 1, 3, 8

1. (a) 1.796 (b) 2.447 (c) 63.657
   (d) 2.048

3. (a) 95% (b) 98% (c) 99% (d) 80% (e) 90%

8. (b) $\bar{x} + t_{\frac{\alpha}{2} = 0.01, 7df} \left(\frac{1.018}{\sqrt{8}}\right)$ $\quad 3410.14 \pm 2.998\left(\frac{1.018}{\sqrt{8}}\right)$

   (c) outliers – shouldn't use t-dist.

4.5 Normal Distribution

any μ, σ / Standard normal dist.

x z

$$z = \frac{x-\mu}{\sigma}$$

4.11 Sampling distrib for μ, p

Central Limit theorem

($n\uparrow$, Samp. dist. of mean becomes normally distributed)

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

5.1-5.3 Large n

Small n

$\hat{p}$ if $np(1-p) = 10$

5.4

2 populations (treatment, control)

1) random sampling, Independent
2) Population Normally distributed (not a critical assumption)
3) Variances are approx. equal (large n, small s; small n, large s ⇒ trouble)

critical! (homogeneity of variance)

If not, do non-parametric test

1 pop.

1) random sampling w/ replacement

2 pops $\mu_1 - \mu_2$

pop 1 → sample: n_1, $\bar{x}_1$, $\bar{s}_1$

pop 2 → sample: n_2, $\bar{x}_2$, $\bar{S}_2$

Confidence Intervals for the difference $(\mu_1 - \mu_2)$ between 2 populations

$n_1 \geq 30$, $n_2 \geq 30$, large n

$\bar{x}_1 - \bar{x}_2$ $\bar{x}_1 - \bar{x}_2$ $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

Central Limit Theorem for mean differences

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Standard error

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\left(\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right)$$

ex 1

	n	$\bar{x}$	s^2
pgm 1	40	78.7	201.6
pgm 2	40	75.3	259.2

(drop-out $n \Rightarrow$ different)

99% C.I. ?

$$(78.7 - 75.3) \pm z_{0.05}\left(\sqrt{\frac{201.6}{40} + \frac{259.2}{40}}\right)$$

(2.58)

3.4 ± 8.76

$-5.36 \rightarrow 12.16$

has 0 in it ⇒ Can't reject H_0 no difference betw Programs

5.6 $n_1 < 30$ or $n_2 < 30$ small sample (any one)

1) equal variance 2) not equal variances

1) assume variances are approx. equal

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)}$$

weighted avg of 2 variances df in small sample
pooled variance

$\bar{x}_1 - \bar{x}_2 = \mu_1 - \mu_2$ $(\bar{x}_1 - \bar{x}_2) \pm t_{df,\,\alpha/2} \sqrt{Sp^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{sp^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ $df = n_1 + n_2 - 2$

2) assume variances are not equal

$\nu = df$

$$\nu = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\left(\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right)$$

Ex 2

	n	$\bar{x}$	s	s^2
Sample 1	15	7.8	3.5	12.25
Sample 2	10	10.3	5.4	29.16

find 95% CI for $(\mu_1 - \mu_2)$

unequal vari. method 1

alter $df = \nu$

$$\nu = \frac{\left(\frac{12.25}{15}+\frac{29.16}{10}\right)^2}{\frac{\left(\frac{12.25}{15}\right)^2}{14} + \frac{\left(\frac{29.16}{10}\right)^2}{9}} \approx 14$$

$$(7.8 - 10.3) \pm 2.145\sqrt{\left(\frac{12.25}{15}\right)+\left(\frac{29.16}{10}\right)}$$

$-2.5 \pm 2.145\,(1.932)$

-6.64 to 1.64

no difference bet samples

equal variance method 2

$$(\bar{x}_1 - \bar{x}_2) \pm t_{23df,\ \alpha=.025}\sqrt{18.87\left(\frac{1}{10}+\frac{1}{15}\right)}$$

$df = n_1 + n_2 - 2$

$$Sp^2 = \frac{14(12.25) + 9(29.16)}{23} = 18.87$$

$$(7.8 - 10.3) \pm 2.069\sqrt{18.87\left(\frac{1}{10}+\frac{1}{15}\right)}$$

-2.5 ± 3.68

5.7 Dependent Samples

usually used on small samples. Ex: twins, husband & wifes
Pull sample from pop, results of one are affected by another (correlated)

μ_1 $\bar{x}_1$

μ_2 $\bar{x}_2$

Differences $\bar{x}_1 - \bar{x}_2$

only care about the difference

Ex	before	After	D
1)	90	93	3
2)	94	96	2
3)	91	92	1
4)	85	88	3
5)	88	90	2

$\bar{x}_d =$ avg diff. $= 2.2$

$S_d =$ st. dev. of diff. $= 0.837$

$df = 4$

$2.2 \pm 2.132\left(\frac{.837}{\sqrt{5}}\right)$

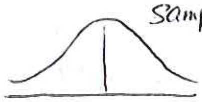
$$\boxed{\bar{x}_d \pm t_{\alpha/2,\,df}\frac{S_d}{\sqrt{n_d}}}$$

proportions

5.5 2 pops

p_1

p_2



sample distrib. of $\hat{p}_1 - \hat{p}_2$

$n_1\hat{p}_1(1-\hat{p}_1) \geq 10$ and $n_2\hat{p}_2(1-\hat{p}_2) \geq 10$

estimating discrete prob. distribution!

$\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$

$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

$(p_1 - p_2) \pm z_{\alpha/2}\left(\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$

Ex 95% C.I. for $p_1 - p_2$?

	n	# Successes
1)	100	78
2)	100	87

$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2 = 0.025}\left(\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$

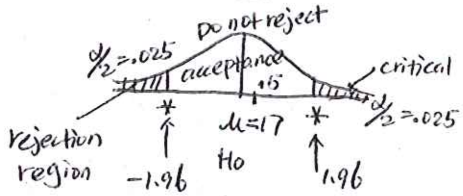
$(0.87 - 0.78) \pm 1.96\sqrt{\frac{(.87)(.13)}{100} + \frac{(.78)(.22)}{100}}$

6.1-6.2 Hypothesise testing (make a guess about sth.)

$H_0: \mu = 17$ assume to be true

$H_1: \mu \neq 17$ alternative hypothesis

2-tail test



Confidence level (95%)

$\alpha = 1 - \text{confidence level}$ (0.05)

Pick sample $n = 100$, $\bar{x} = 18$, $S = 20$

$Z_{obs} = \frac{\bar{x} - \mu_{H_0}}{S/\sqrt{n}} = \frac{18-17}{20/\sqrt{100}} = \frac{1}{2} = 0.5$

observed value

$\Rightarrow$ DO NOT reject Ho

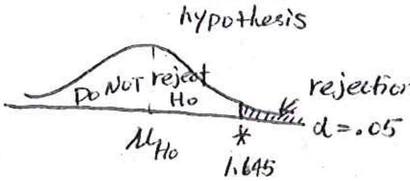
1-tail Test

$H_0: \mu \leq 17$ must have equality

$H_1: \mu > 17$ research hypothesis

$n = 100$, $\bar{x} = 18$, $S = 25$

$Z_{obs} = \frac{\bar{x} - \mu_{H_0}}{S/\sqrt{n}} = \frac{18-17}{2.5} = 0.4$



$\Rightarrow$ cannot reject Ho

Hyp. Test – Mean

Ho : null hypothesis $\mu = a$

H_1 : Alternate hypothesis, research hypothesis $\mu \neq a$

researcher in rejecting Ho

(2) one-tail (right)

$H_0: \mu \leq a$

$H_1: \mu > a$

1) 2 tailed test of μ (one population)

(3) one-tail (left)

$H_0: \mu \geq a$

$H_1: \mu < a$

Types of error

Type I error = reject Ho when Ho's true

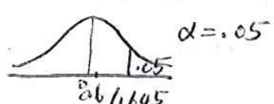
P(Type I error) = α (make α small to ↑ sure to make right decision)

Type II error = D.N. reject Ho when H_1 is true

P(Type II error) = β (every β will change based on true, much worse than Type I)

6.3-6.4 One Pop. Hype Test

1. Test μ or p?
2. Ho, H_1 $\quad H_0: \mu \leq 10$, $H_1: \mu > 10$
3. assumption = n, x̄, s $\quad n = 49, \bar{x} = 11.2, s = 14$
4. critical value $\quad \alpha = .05$



$Z_{crit} = 1.645$

5. Test Stats: $Z_{obs} = \dfrac{\bar{x} - \mu_{H_0}}{\sigma/\sqrt{n}}$ (σ or s)

$= \dfrac{11.2 - 10}{14/\sqrt{49}} = \dfrac{1.2}{2} = 0.6$

2 methods

(1) traditional

compare Z_{crit}, Z_{obs}

$0.6 < 1.645$

cannot reject Ho

huge stand. dev.!

(2) p-value

(P)

actual proportion of area in the right tail for this test

If $P > \alpha$, cannot reject H_0 (p-value)

If $P < \alpha$, reject Ho

Prob. 1

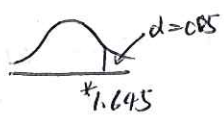
1 pop with large n

Z

$\bar{x} = 714.2$, $n = 40$, $S = 83.2$

$H_0: \mu \leq 703.5$

$H_1: \mu > 703.5$



$Z_{obs} = \dfrac{\bar{x} - \mu_{H_0}}{S/\sqrt{n}} = \dfrac{714.2 - 703.5}{83.2/\sqrt{40}} = \dfrac{10.7}{13.16} = 0.81$

p-value = area to right of 0.81 (obs. z) $\quad p = 0.2090 = \text{Sig} > .05$ (significance)

cannot reject Ho

Prob 2

1 pop with small n

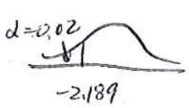
t

$H_0: \mu \geq 80$

$H_1: \mu < 80$

$n = 22$, normally dist

$\alpha = 0.02$ $\quad df = 21$



$t = -2.189$

If $\bar{x} = 76.9$, $S = 8.5$ $\quad t_{obs} = \dfrac{\bar{x} - \mu_{H_0}}{S/\sqrt{n}} = \dfrac{76.9 - 80}{8.5/\sqrt{22}} = -1.71$

$0.05 < P\text{ value} < 0.10$

can't get exact unless on computer

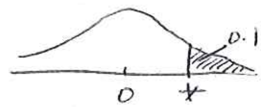
P value > .02

Cannot reject Ho

Prob 3 — 1 pop for p

$\alpha = 0.1$

$np(1-p)$ check both $p, \hat{p}$.



$H_0: p \le 0.85$

H_1: (researchers believe to be true) $p > 0.85$

$Z_{crit} = 1.28$

sample data $n = 200$

$\hat{p} = \frac{171}{200} = 0.855$

$$Z_{obs} = \frac{\hat{p} - p_{H_0}}{\sqrt{\frac{(0.85)(0.15)}{200}}} = \frac{0.005}{0.025} = 0.2$$

$A_{z=0.2 \text{ right}} = 0.4207 > \alpha$ cannot rej. H_0

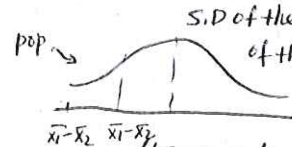
6.5-6.7 two populations

Difference between means

Pop 1	Pop 1
n_1	n_2
$\bar{x}_1$	$\bar{x}_1$
$S_1(\sigma_1)$	$S_2(\sigma_2)$

- $n_1, n_2 > 30$, z-test
- n_1 or $n_2 < 30$, σ unknown, t-test

assumption
- * draw independently (independence sampling)
- * N.D. population
- * population variances are appro. equal (Homogeneity of variance)



$\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$

$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ $(n_1, n_2 \ge 30)$

- $n_1, n_2 \ge 30$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$H_0: \mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$ — H_0 must have equality

$H_1: \mu_1 \ne \mu_2$ or $\mu_1 - \mu_2 \ne 0$

- n_1, or $n_2 < 30$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

Levene's Test

test of equality of variances

$H_0: \sigma_1^2 = \sigma_2^2$ (we don't want to reject) Homogeneity of variance

$H_1: \sigma_1^2 \ne \sigma_2^2$

Difference between proportions

$n_1\hat{p}_1(1-\hat{p}_1) \ge 10$ and $n_2\hat{p}_2(1-\hat{p}_2) \ge 10$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

prob 4 — diff. 2 means small n

$H_0 = \mu_{female} = \mu_{male}$

$H_1: \mu_F \ne \mu_M$

$\alpha = 0.05$

$n_F = 20$, $n_M = 15$

Review for Exam 2

- Normal Distribution

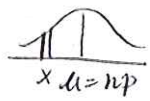
$\mu = 100,\ \sigma = 10,\ P(x > 95) = ?$

$z = \frac{95-100}{10} = \frac{-5}{10} = -0.5$

85th percentile ?

- Normal Approx. of the Binomial as long as $np(1-p) \geq 10$

$n,\ x,\ p,\ 1-p$



$\sigma = \sqrt{np(1-p)}$

$P(x \geq a) = ?$ $\quad \sigma^2 = np(1-p)$

by finding $p(x \geq (a-0.5))$

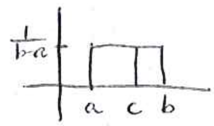
$P(x \leq a) \approx P(x \leq (a+0.5))$

$P(x = a) \approx P(a-0.5 \leq x \leq a+0.5)$

$P(a < x < b) \approx P(a+0.5 < x < b-0.5)$

- Confidence Intervals

- Uniform Prob Distrib.



$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{elsewhere} \end{cases}$

$P(x \leq c) = ?$

$a = 1,\ b = 6$

$p(x) = \frac{1}{5} \quad P(x > 4) = \frac{2}{5}$

4.11 - Sampling distrib.

1 pop - mean

$\mu,\ \sigma \longrightarrow n,\ \bar{x}$

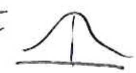
as long as $n \geq 30$

$P(\bar{x} > a)$ from samp. distrib.

$\mu_{\bar{x}} = \mu$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

P (qualitive variable) $= \frac{x}{n}$



$\mu_{\hat{p}} = p$

$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

2 pops.

Pop. 1	Pop 2
μ	μ
$\sigma \to n_1,\ \bar{x}_1,\ s_1$	$\sigma \to n_2,\ \bar{x}_2,\ s_2$
p_1	p_2

$\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$

$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

as long as $np(1-p) \geq 10$ for both pop.s.

$\mu_{\hat{p}_1-\hat{p}_2} = \hat{p}_1 - \hat{p}_2$

$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

6.8/10 dependent samples t test

paired t test

	Before	After	difference = After−Before
1	283	215	−68
2	299	206	−93
3	274	187	−87
4	284	212	−72
5	248	178	−70
6	275	212	−63
7	293	192	−101
8	277	196	−81

$H_0: \mu_D \geq 0$

$H_1: \mu_D < 0$

$\alpha = .05$

$^{*}t_{7df} = 1.895$ — critical value

$$t_{obs} = \frac{\bar{x}_D - D_{H_0}}{S_D/\sqrt{n}} = \frac{-79.375 - 0}{13.3838/\sqrt{8}} = -16.775$$

∴ reject H_0

6.10 Chi-Square

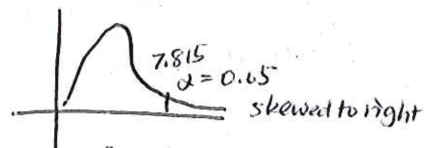
	public speaking	bugs insects	heights	small spaces	
F	30 $E = \frac{(100)(45)}{200} = 22.5$	20 E=30	30 E=27.5	20 E=20	100
M	15 E=22.5	40 E=30	25 E=27.5	20 E=20	100
(Categories)	45	60	55	40	200

If indep. of F, M, should be 22.5 each

Variable column — Variable in row

H_0: Greatest fear is independent of gender

H_1: Greatest fear is dependent on gender



χ^2 distribution

df = (# rows − 1)(# column − 1)

= (1)(4−1) = 3

greater 7.815, reject

$$\chi^2_{obs} = \sum \frac{(O-E)^2}{E} = 12.12 > 7.815$$

∴ fear dep on gender

$$E = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

Final Exam Review

contingency table

	Choc	vala.	Peach	
M	50	10	40	100
F	20	50	30	100

$p(V/M) = ?$

$p(V \cap M) = ?$

Prob.

A + B are M.E.

$P(A) = .4 \quad P(B) = .3$

$\boxed{(A) \quad (B)}$

$P(A \cap B) = 0$

$\left.\begin{array}{l} P(A/B) = P(A) \\ P(A \cap B) = P(A) \cdot P(B) \end{array}\right\}$ independent

discrete P.D.

assign #s to outcome

x	$P(x)$
1	.2
2	.4
3	.1
4	.3

$\Sigma = 1$

$\mu = \Sigma\, x \cdot p(x)$

$\sigma^2 = \Sigma[x^2 p(x)] - \mu^2$

most important P.D.

① binominal

$\mu = np$

$\sigma^2 = np(1-p)$

② Poisson Dist.

measure time, event

occurences per unit of time

λ = mean

use to est/approx. binominal when $n \to$ large, $p \to$ small

Prob. $\{1, 4, 7, 10, 15\}$, $\{1, 1, 3, 5\}$, $\{1, 1, 1, 1\}$ s

sampling, non-response, response bias

$\bar{x}$ unbiased

test where $\mu \geq 10$ = H_0: $\mu \geq 10$

H_1: $\mu < 10$

μ has changed from 36 $\quad H_0: \mu = 36$

$H_1: \mu \neq 36$

- Hypo 1 pop
  - 2 pop
- X^2
- hypothesis

Hypotheses Tests for μ

One Population

σ known; n ≥ 30

$$Z_{obs} = \frac{\bar{x} - \mu_{H_0}}{\frac{\sigma}{\sqrt{n}}}$$

σ unknown; n ≥ 30

$$Z_{obs} = \frac{\bar{x} - \mu_{H_0}}{\frac{s}{\sqrt{n}}}$$

σ unknown; n < 30

$$t_{obs} = \frac{\bar{x} - \mu_{H_0}}{\frac{s}{\sqrt{n}}}$$

Hypotheses Tests for p

One Population

$$(n)(p_{H_0})(1 - p_{H_0}) \geq 10$$

$$Z_{obs} = \frac{\hat{p} - p_{H_0}}{\sqrt{\frac{(p_{H_0})(1 - p_{H_0})}{n}}}$$

Hypotheses Tests for $\mu_1 - \mu_2$

Two Populations; Independent Samples

If $n_1 \geq 30$ and $n_2 \geq 30$

- $$Z = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

If σ_1 and σ_2 are unknown, use the s_1 and s_2

If $n_1 < 30$ or $n_2 < 30$ or both sample sizes are less than 30 and σ's are unknown

- $$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{Where } S_p^2 = \frac{(n_1-1)S_1{}^2+(n_2-1)S_2{}^2}{(n_1+n_2-2)}$$

Hypotheses Tests for $p_1 - p_2$

Two Populations; Independent Samples

$n_1\ \hat{p}_1 (1- \hat{p}_1) \geq 10$ and $n_2\ \hat{p}_2 (1- \hat{p}_2) \geq 10$

- $$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}\ (1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

- $$\text{Where } \hat{p} = \frac{(x_1+x_2)}{(n_1+n_2)}$$