

① Please write the followings in your class notes:

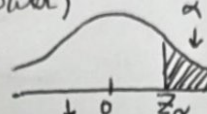
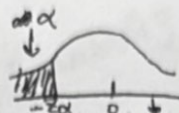
03/12

Null hyp. $H_0: \mu = \mu_0$ VS

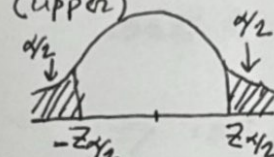
(Reject H_0)
Rejection Region

c.e: chocolate

① $H_a: \mu < \mu_0 \rightarrow$ left tailed test $\rightarrow$
(c.e) (Decrease) (lower)



② $H_a: \mu > \mu_0 \rightarrow$ right tailed test $\rightarrow Z > Z_\alpha$
(c.e) (Increase) (upper)



③ $H_a: \mu \neq \mu_0 \rightarrow$ Two-tailed test $\rightarrow |Z| > \frac{Z_\alpha}{2}$

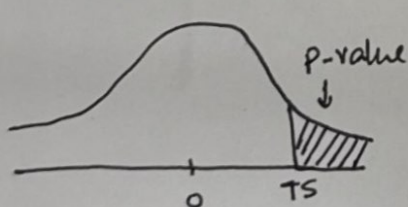
= Test statistic

② P-value $\rightarrow$ Area further into the tail(s) from test statistic.

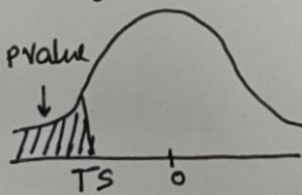
$\rightarrow$ Also known as the "observed significant level".

$\rightarrow$ measures how significant the data is

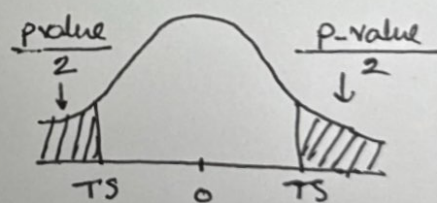
$\rightarrow$ The closer the p-value to zero, the more evidence we have against H_0



$H_a: \mu > \mu_0$



$H_a: \mu < \mu_0$



$H_a: \mu \neq \mu_0$

on a two-tailed test, p-value is doubled!

$\rightarrow H_0$ is rejected iff p-value $< \alpha$

H_0 is not rejected (fail to reject) iff p-value $> \alpha$

where, α = level of significance

② If σ is unknown, then a one-sample T-test for μ

(σ is unknown) is on a t-curve with $(n-1)$ degrees of freedom and a test statistic of

$$T = \frac{\bar{Y} - \mu_0}{s/\sqrt{n}}, \text{ under } H_0.$$

$$\mu = 8?$$

One Sample Z-test for μ :

- 1) **Assumption:** Population Standard Deviation σ is known and sample mean $\bar{y}$ is normally distributed or approximately normal.
- 2) **Hypothesis:** Choose H_0 , H_a , and α = Level of significance

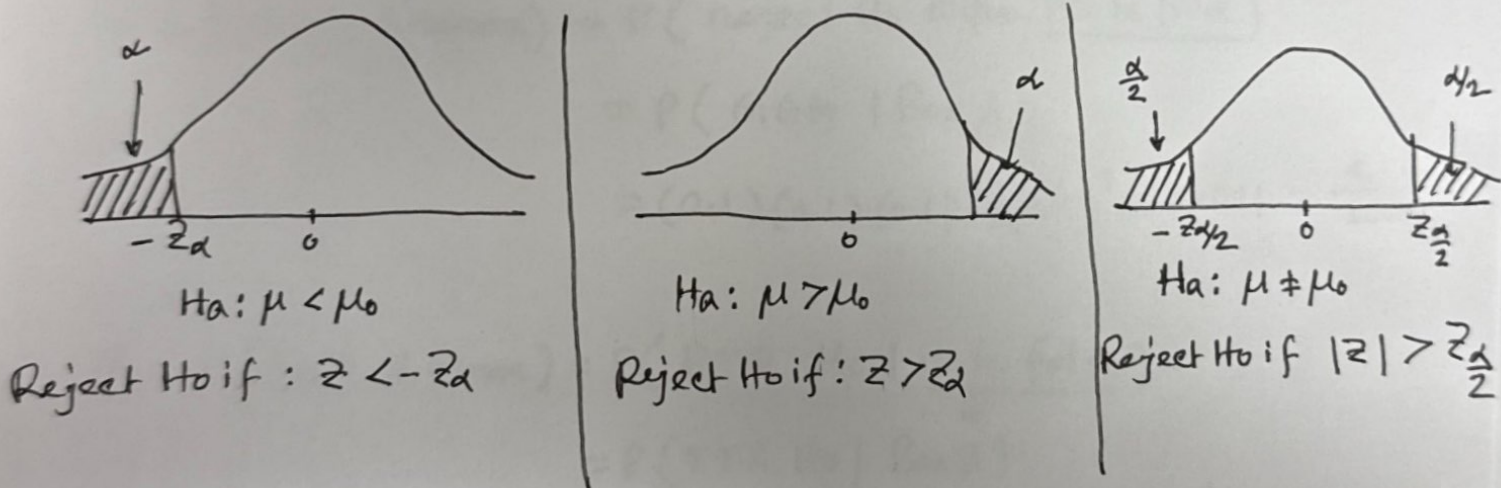
$$H_0: \mu = \mu_0 \text{ vs}$$

Three options for H_a :

$\mu < \mu_0$	Left-tailed or lower-tailed	CC: Decrease ($\downarrow$)
$\mu > \mu_0$	Right-tailed or upper-tailed	CC: Increase ($\uparrow$)
$\mu \neq \mu_0$	Two-tailed	CC: Change/Different

3) Test Statistic: $Z = \frac{(\bar{y} - \mu_0)}{\sigma/\sqrt{n}}$ [Assuming H_0 is true] μ_0

4) Critical/Rejection Region: (α = area, H_a = Location)



5) Conduct Experiment & find Value of the test statistic.

6) **Conclusion:** If test statistic falls in the rejection region, then reject H_0 , otherwise, fail to reject H_0 .

Reject H_0 : "We have significant evidence at level α that H_0 is false"

Fail to reject H_0 : "We do not have significant evidence at level α that H_0 is false; i.e. data is not significant".

Sometimes, we also compute a confidence interval for μ after a rejection of H_0 to report the location of the plausible values for μ .

Example 2:

Purchasing shipments of iron ore pellets:

 μ = mean weight among all shipments being purchased (unknown) σ = 10 pounds (known)

Supplier claims that "the average shipment weighs 80 pounds". You suspect you may be getting shortchanged.

$$H_0: \mu = 80 \text{ vs}$$

Three options for H_a :

$\mu < 80$	Left-tailed or lower-tailed	CC: Decrease ($\downarrow$)
$\mu > 80$	Right-tailed or upper-tailed	CC: Increase ($\uparrow$)
$\mu \neq 80$	Two-tailed	CC: Change/Different

Choose the appropriate

alternative hypotheses (H_a) and perform a hypothesis test. $H_a: \mu < 80$

Plan: Randomly select $n=30$ shipments, weight them, compute $\bar{y}$
 reject H_0 if $\bar{y}$ is too low.

⊛ willing to make a type-1 error with prob. $\alpha = 0.05$
 [5% level of significance]

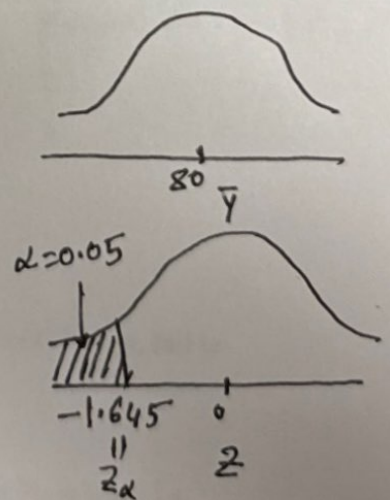
⊛ Temporarily, assume H_0 is true

$$\bar{y} \sim N(\mu_0 = 80, \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{30}})$$

Test Statistic:

$$Z = \frac{\bar{y} - \mu_0}{\underbrace{\frac{\sigma}{\sqrt{n}}}_{10/\sqrt{30}}}$$

(RR) Rejection Region: $Z < -Z_\alpha$
 (reject H_0)



⊛ Given: Experiment is done,

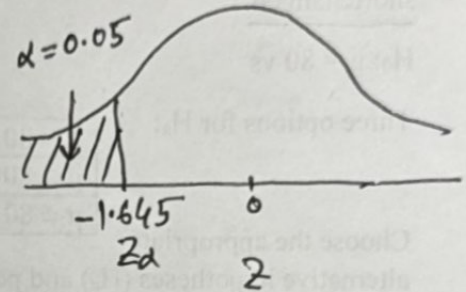
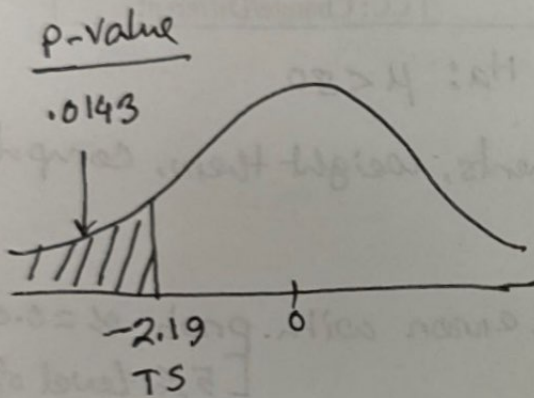
$$\bar{y} = 76 \text{ pounds}$$

$$Z = \frac{76 - 80}{10/\sqrt{30}} = -2.191$$

Since, $Z = -2.191$ is in the RR, we reject H_0 .

Therefore, we have significant evidence at $\alpha = 0.05$ significance level that the average shipment weights less than 80 pounds.

$$Z = -2.19$$



$$Z = \frac{\bar{y} - \mu_0}{\frac{s}{\sqrt{n}}}$$

(RR) Rejection Region: $Z < -Z_\alpha$
(reject H_0)

$$Z = \frac{76 - 80}{\frac{10}{\sqrt{130}}} = -2.191$$

4.3: Inference for a single mean

03/12

Ex.

$$H_0: \mu = 50$$

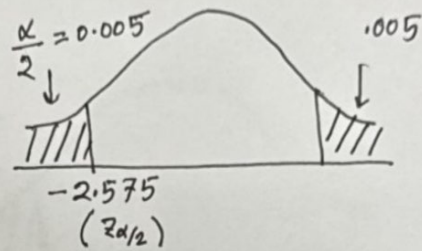
$$vs H_a: \mu \neq 50$$

$$\alpha = 0.01$$

$$\bar{y} = 53$$

$$n = 64$$

$$s = 16 \text{ (known)}$$

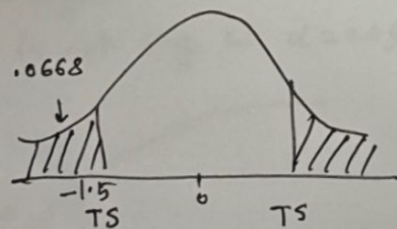


$$RR: |z| > z_{\alpha/2}$$

$$TS: z = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{53 - 50}{16/\sqrt{64}} = 1.5$$

Fail to reject H_0 at $|z| < 2.575$

Data is not significant at $\alpha = 0.01$



$$p\text{-value}: 2(0.0668) = 0.1336 > \alpha = 0.01$$

FTR H_0 at $\alpha = 0.01$

Ex

$$H_0: \mu = 70$$

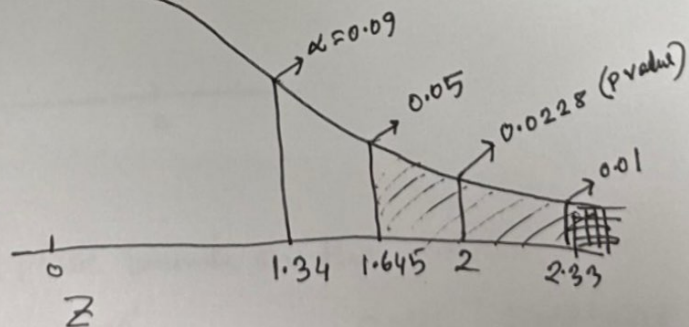
$$vs H_a: \mu > 70$$

$$\alpha = 0.09$$

$$\bar{y} = 74$$

$$n = 36$$

$$s = 12 \text{ (known)}$$



$$RR: z > z_{\alpha}$$

$$z = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{74 - 70}{12/\sqrt{36}} = 2$$

Reject H_0

Data is significant at $\alpha = 0.09$

Data is still significant at $\alpha = 0.05$

Data is NOT significant at $\alpha = 0.01$

$$\alpha = \frac{0.05}{25050}$$

Ex. μ : pop mean cholesterol among 70 year old American men

Assume levels are normally distributed

$$H_0: \mu = 45.5$$

$$\text{vs } H_a: \mu < 45.5$$

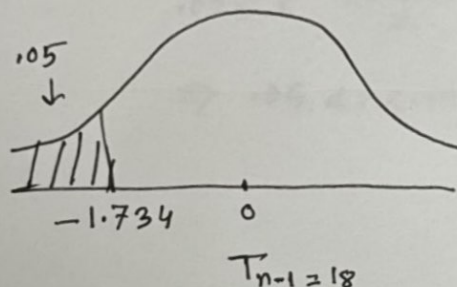
$$\alpha = 0.05$$

$$n = 19$$

$$\bar{y} = 44.2$$

σ unknown

$$s = 3.8$$

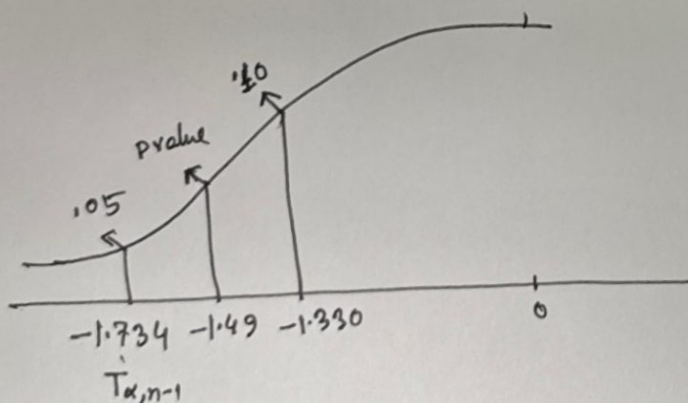


$$RR: T < -T_{\alpha, n-1}$$

$$T = \frac{\bar{y} - \mu}{s/\sqrt{n}} = \frac{44.2 - 45.5}{3.8/\sqrt{19}} = -1.49$$

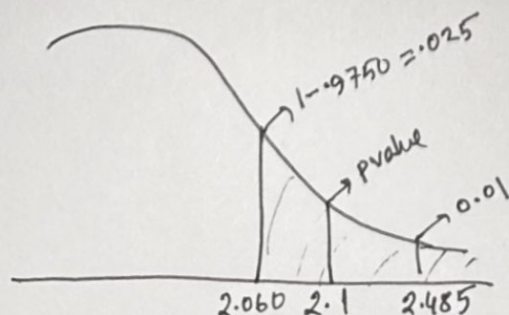
Fail to reject H_0 .

Data is not sig. at $\alpha = 0.05$



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

Ⓐ Given the TS, place bounds on the p-value:

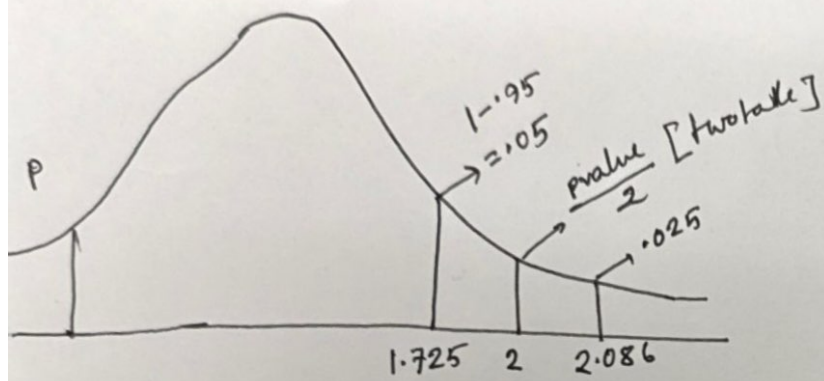


$n = 26$, Right tailed

$$T = 2.1$$

$$n-1 = \text{df} = 25$$

$$0.01 < P\text{-value} < 0.025$$



$$n=21, \text{ Two-tailed}$$

$$T=2$$

$$n-1 = \text{Dof} = 20$$

$$.025 < \frac{P\text{-value}}{2} < .05$$

$$\Rightarrow .05 < P < .10$$

Topic: Computing Correlation Coefficient (r)

Example 3: An Engineering student has designed a new electric clothes dryer. Maybe there is something special about this dryer, maybe it is energy efficient. However, the student is having a problem with the temperature setting (maybe there is a knob to select the temperature from the setting menu) and the actual temperature inside the dryer.

X: Temperature setting (F)

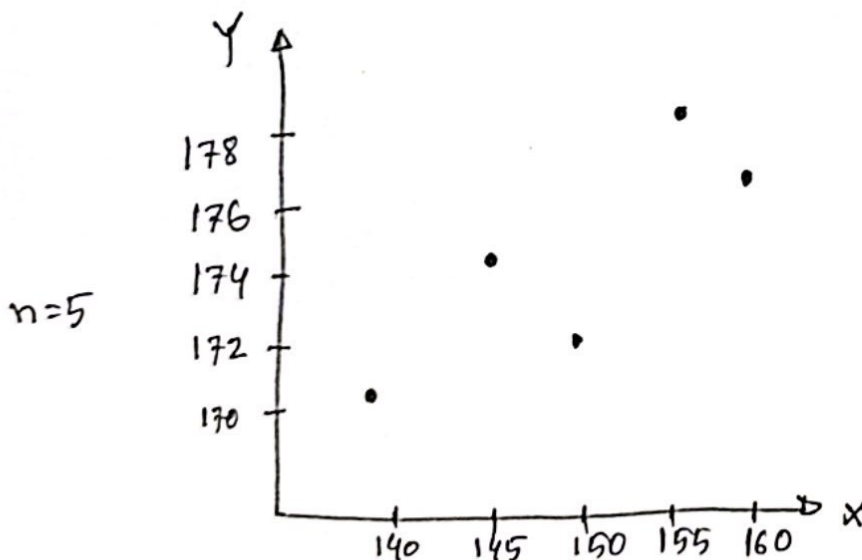
cc y depends on x

Y: Actual interior temperature during cycle (F)

The student has carefully measured several readings for X along with the values of Y. The dataset is below:

X	140	145	150	155	160
Y	170	174	172	178	176

a) Draw a scatterplot with the above dataset and interpret it.



i trend: 😊 Linear

ii Association: +ve

iii Strength: 😊

iv Outlier: N/A

- b) Compute SS_{XX} , SS_{YY} , SS_{XY} , and r . Interpret the correlation coefficient. Does the interpretation of r match with the interpretation of the scatterplot you have drawn in part (a)?

X	Y	X^2	Y^2	XY
140	170	$140^2 = 19600$	$170^2 = 28900$	$140 \times 170 = 23800$
145	174	✓	✓	✓
150	172	✓	✓	✓
155	178	✓	✓	✓
160	176	✓	✓	✓
$\Sigma X = 750$	$\Sigma Y = 870$	$\Sigma X^2 = 112750$	$\Sigma Y^2 = 151420$	$\Sigma XY = 130580$

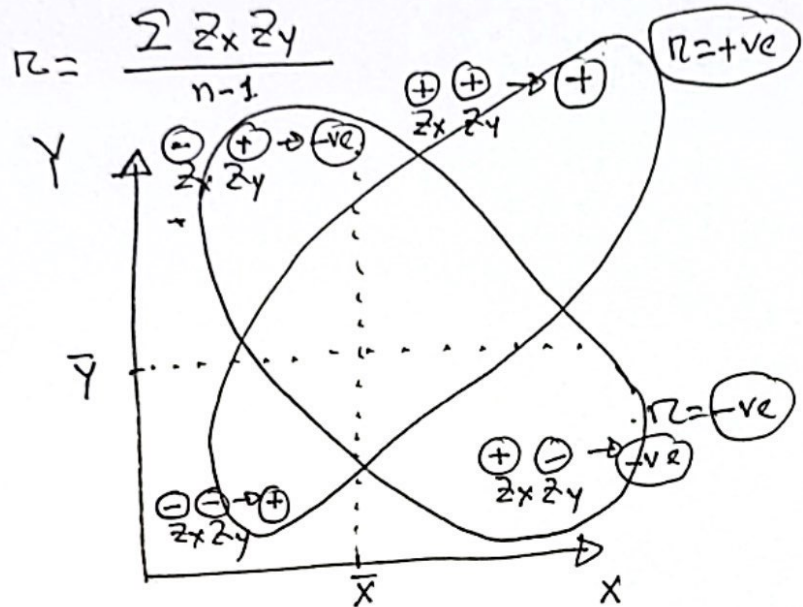
$$SS_{XX} = \Sigma X^2 - \frac{(\Sigma X)^2}{n} = 112750 - \frac{(750)^2}{5} = 250$$

$$SS_{YY} = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n} = 151420 - \frac{(870)^2}{5} = 40$$

$$SS_{XY} = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n} = 130580 - \frac{(750)(870)}{5} = 80$$

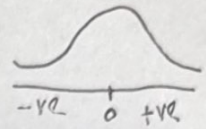
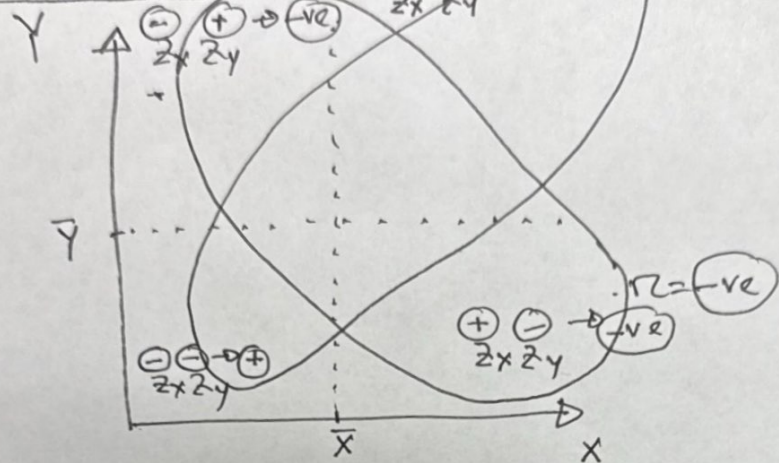
$$r = \frac{SS_{XY}}{\sqrt{SS_{XX}} \sqrt{SS_{YY}}} = \frac{80}{\sqrt{250} * \sqrt{40}} = 0.80$$

04-02-25

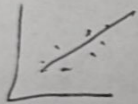


04-02-25

$$r = \frac{\sum z_x z_y}{n-1}$$



(Ex) 1 Switch $\leftrightarrow$ X: Actual Interior temp during cycle of
 Y: temp setting of



$$r = .80$$

Ex. 2

X: Temp. setting in $^{\circ}\text{C}$

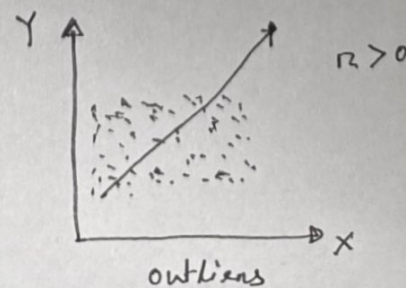
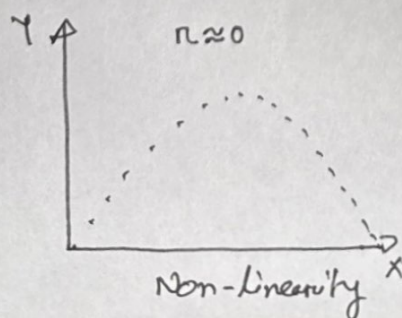
Y: Actual . . . $^{\circ}\text{C}$

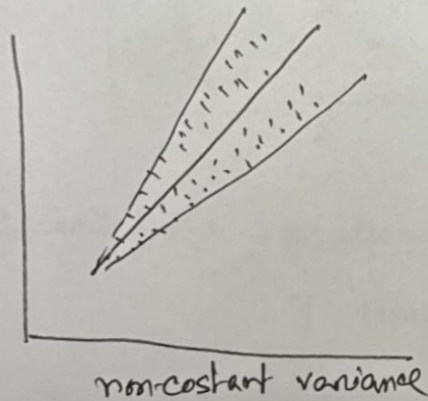
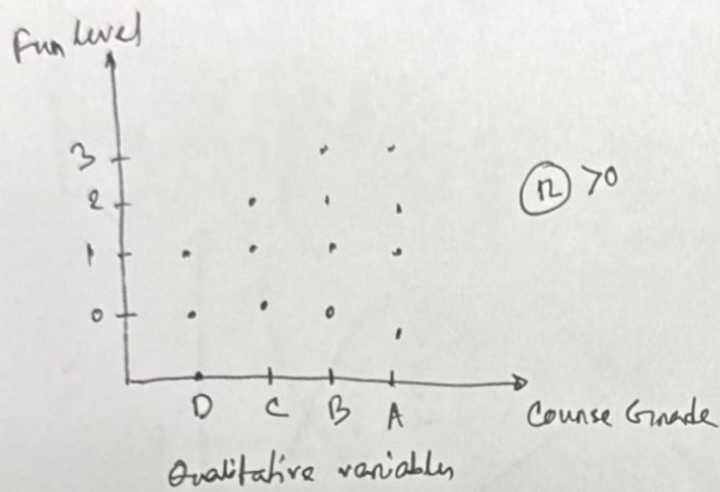
$$R = .80$$

Properties of $n!$

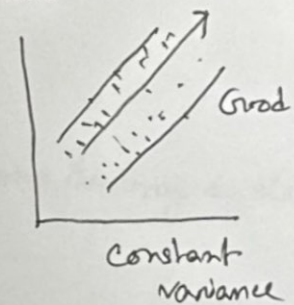
The value of r_z does not change if X & Y are switched or if the units of measurement are changed

⚠ Danger with n_0 watch out for -





Cone shapes are bad



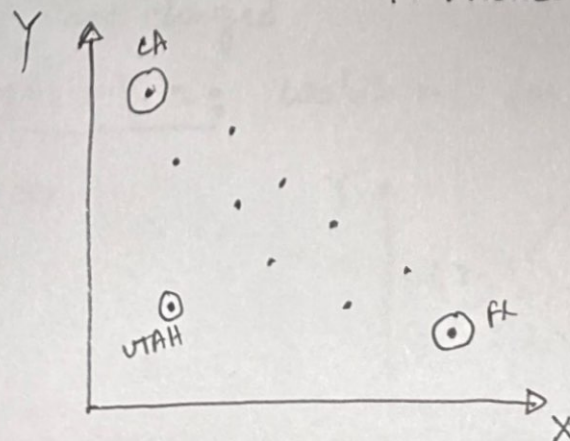
Association is NOT causation!

Association: one variable tends to act along with the other

+ve -ve

Causation: One variable directly causes the other to behave in a certain manner.

Ex. 2020 census data. X: Death Rate
Y: Divorce Rate



Age

↓

lurking variable!

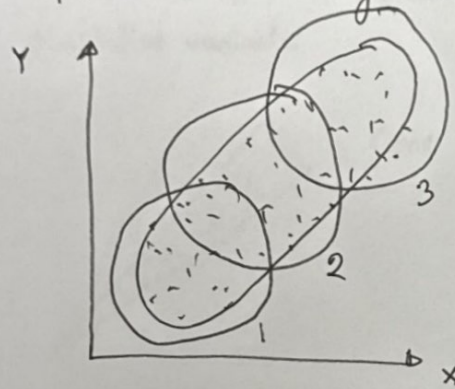
- ④ Linking Variable: Is NOT directly measured but is responsible for the association between two other variables.

Elementary School

Ex 1)

X: Height

Y: Score on reading test



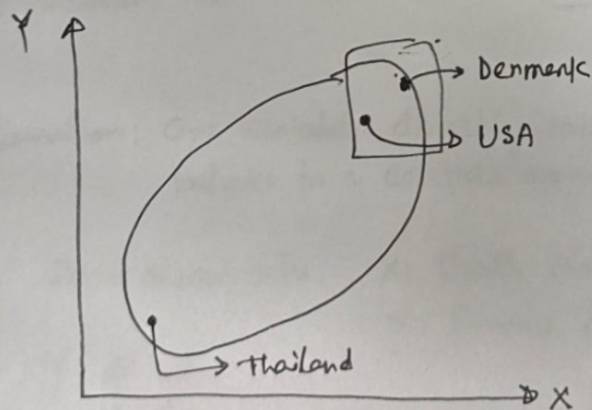
(Age)
Grade: linking variable

- ④ mystery correlation! (??)

Nation of the world

X: Fat intake per capita

Y: Death rate from cancer



Sugar?
Air pollution?
Genetics/Ethnicity?
Life expectancy?

□ If X and Y have a linear relationship, then,

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Intercept

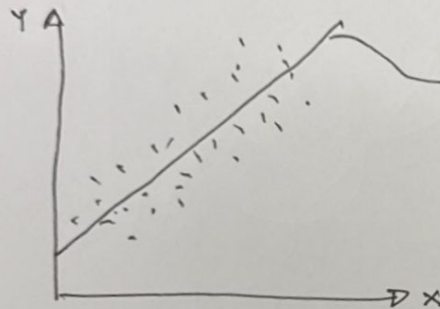
slope

Error term (Residual)

β : Beta

ε : Epsilon $\sim N(0, \sigma^2)$

Constant Variance



True Regression Line (Least square reg. line)

we estimate LS reg line with

$$\hat{Y} = b_0 + b_1 X$$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

unknown parameter

β_0

β_1

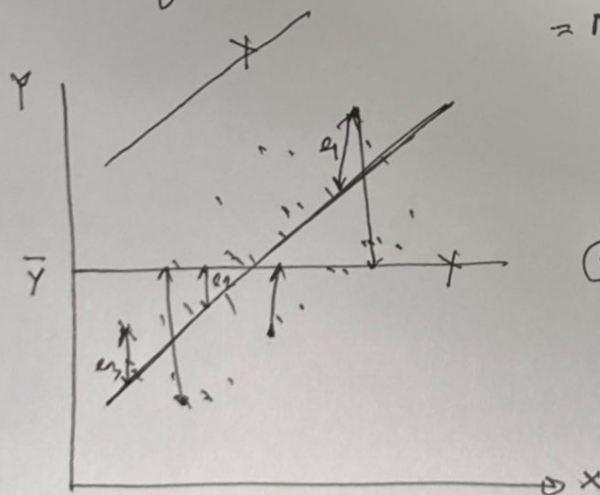
σ^2

Estimates

$b_0 / \hat{\beta}_0$

$b_1 / \hat{\beta}_1$

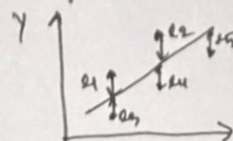
$s^2 = \text{mean squared Residual}$
 $= MS_{res}$



$$\varepsilon = Y - \hat{Y}$$

The least-square regression line has two properties:

① Sum of errors/Residuals is zero; $\sum \varepsilon_i = 0$



$$\sum \varepsilon_i = 0$$

② Sum of the squared errors is minimized
 $\sum \varepsilon_i^2 = SS_{res}$ [Sum of squared of residual]

Simple Linear Regression

04/09/25

* If X and Y have a linear relationship, then

True Reg. Line Eq: $Y = \beta_0 + \beta_1 X + e$; $e \sim N(0, \sigma^2)$ →

Intercept Slope Error/Residual

Y : Response/Dependent variable
 X : Independent variable

(*) Least ^(LS) Square regression equation:

LS Reg. Line Equ: $\hat{Y} = b_0 + b_1 X$ where, $\rightarrow$

with the interpretations:

b_1 : on average, we expect for every 1 unit increase ~~in~~ in X will make a b_1 degree increase/decrease in Y (depending on it's sign)

b_0 : on average, we expect an actual y if we set $x=0$; $\hat{y}$: predicted mean y that has/have x unit.

$\hat{y}$ = predicted Response

$$b_0 = \hat{\beta}_0 = \text{Estimate of } \beta_0 = \bar{Y} - b_1 \bar{X}$$
$$b_1 = \hat{\beta}_1 = \text{Estimate of } \beta_1 = \frac{r S_y}{S_x} = \frac{SS_{xy}}{SS_{xx}}$$
$$SS_{XX} = \sum X^2 - \frac{(\sum X)^2}{n}$$
$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$
$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$

* Evaluating a linear model

$$\begin{array}{lcl} \text{Total variation in } y & = & \text{variation due to } X + \text{variation due to } e \\ \downarrow & & \downarrow \qquad \qquad \downarrow \\ SS_{\text{total}} & = & SS_{\text{reg}} + SS_{\text{res}} \end{array}$$

$e \sim N$

where, $SS_{\text{total}} = \sum (Y - \bar{Y})^2 = SS_{YY}$

$$SS_{reg} = \sum (\hat{Y} - \bar{Y})^2 = \frac{(SS_{xy})^2}{SS_{xx}}$$
$$SS_{res} = \sum \underbrace{(y - \hat{y})}_e^2 = \sum e^2 = SS_{total} - SS_{reg}$$
$$e \sim N(0, \sigma^2)$$

⊛ Estimate σ^2 with s^2 :

$$S^2 = MS_{nes} = \frac{SS_{nes}}{n-2}$$

(*) Testing the slope Hypothesis test for β_1

$H_0: \beta_1 = 0$ (X & Y are independent; X is a useless predictor of Y)

vs $H_1: \beta_1 \neq 0$ (X & Y are dependent)

Test statistic: $T = \frac{b_1 - 0}{\sqrt{(MS_{res}/SS_{xx})}}$ with $(n-2)$ d.o.f.

prev. page

Correlation coefficient, r :

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}} \sqrt{SS_{yy}}} \quad ; -1 \leq r \leq 1$$

Coefficient of determination: r^2

r^2 : the proportion / ^{percentage} of variation in y that is explained by x

$$SS_{total} = SS_{reg} + SS_{res} \quad ; 0 \leq r^2 \leq 1$$

$$r^2 = \frac{SS_{reg}}{SS_{total}} \quad (\text{the bigger } r^2 \text{ is, the better})$$

Ex. If $r^2 = (R^2) = 0.95$

95% of the variation in y that is explained by x

Go to handout (Ex. 3)

(e') what percentage of the variation in dryer temp. (y) is explained by setting (x)?

$$r^2 = \frac{SS_{\text{reg}}}{SS_{\text{total}}}$$

$$= \frac{25.6}{40}$$

$$r^2 = 0.64$$

$$\hat{r}^2 = (.80)^2 = 0.64$$

$$SS_{\text{reg}} = \frac{(SS_{xy})^2}{SS_{xx}} = \frac{(80)^2}{250}$$

$$= 25.6$$

$$SS_{\text{total}} = SS_{yy} = 40$$

$$SS_{\text{res}} = SS_{\text{total}} - SS_{\text{reg}}$$

$$= 40 - 25.6 = 14.4$$

64% of the variation in dryer temp (y) is explained by setting (x).

$$\mu \rightarrow \bar{x}$$

$$\sigma^2 \rightarrow s^2$$

(*) Estimate σ^2 with s^2

$$s^2 = MS_{\text{res}} = \frac{SS_{\text{res}}}{n-2} = \frac{14.4}{5-2} = 4.8$$

$$Y = \beta_0 + \beta_1 X + \epsilon$$

1) Perform the hypothesis testing for β_1 .

$$H_0: \beta_1 = 0$$

$$\text{vs } H_1: \beta_1 \neq 0$$

$$n = 5$$

$$SS_{XX} = 250$$

$$b_1 = 0.32$$

$$MS_{res} = 4.8$$

$$\alpha = 0.05$$

$$T_{\frac{\alpha}{2}, n-2} = 3.182$$

$$\alpha = 0.05$$

$$1 - 0.025 = 0.975$$

$$\frac{\alpha}{2} = 0.025$$

$$0.025$$

$$T_{n-2} = T_{5-2} = T_3$$

$$T = \frac{b_1}{\sqrt{\frac{MS_{res}}{SS_{XX}}}} = \frac{0.32}{\sqrt{\frac{4.8}{250}}} = 2.31$$

Rejection Region: Reject if $|T| > T_{\frac{\alpha}{2}, n-2}$

$$2.31 < 3.182$$

Don't reject H_0 .

Data is not significant at $\alpha = 0.05$

i.e. X is not a significant independent variable to predict the dependent variable y (on average) at $\alpha = 0.05$

$$\text{cancer}(y) \sim \text{Gene}_1(x)$$

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$\alpha = 8 \times 10^{-8}$$

Analysis of variance

ANOVA table

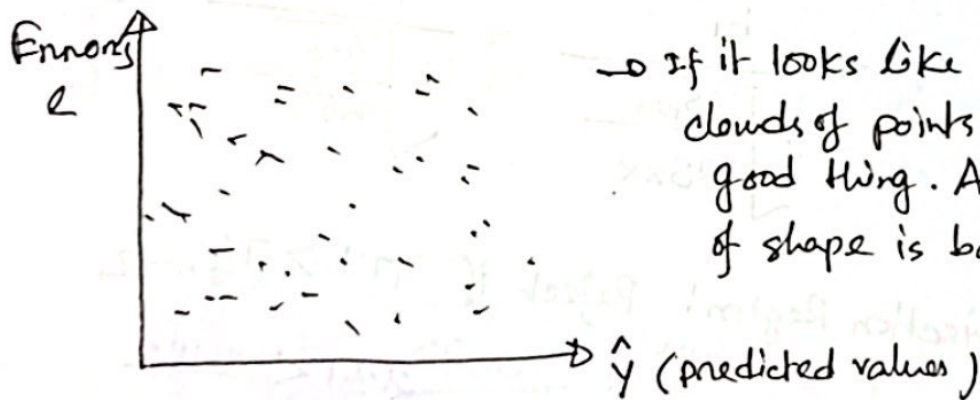
Source	dof	SS	MS	F
Regression	1	SS_{reg}	$MS_{reg} = \frac{SS_{reg}}{dof} = \frac{SS_{reg}}{1}$	$\frac{MS_{reg}}{MS_{res}}$
Residual	$n-2$	SS_{res}	$MS_{res} = \frac{SS_{res}}{n-2}$	
Total.	$n-1$	SS_{total}		

In testing $\beta_1 = 0$

$$F \approx T^2$$

⊗ Residual/ Error plots:

→ can reveal problems with your data set.



* If X and Y have a linear relationship, then

True Reg. Line: $Y = \beta_0 + \beta_1 X + e$; $e \sim N(0, \sigma^2)$

$\uparrow$ Intercept $\nwarrow$ slope $\swarrow$ Error/Residual

Y : Response/Dependent variable

X : Independent variable

$e = Y - \hat{Y}$; $\sum e = 0$

* Least Square regression equation:

LS Reg. Line: $\hat{Y} = b_0 + b_1 X$ where,

$\hat{Y}$ = predicted Response

$b_0 = \hat{\beta}_0 = \text{Estimate of } \beta_0 = \bar{Y} - b_1 \bar{X}$

$b_1 = \hat{\beta}_1 = \text{Estimate of } \beta_1 = \frac{r S_y}{S_x} = \frac{SS_{xy}}{SS_{xx}}$

$SS_{xx} = \sum X^2 - \frac{(\sum X)^2}{n}$

$SS_{xy} = \sum XY - \frac{(\sum X)(\sum Y)}{n}$

with the interpretations:

b_1 : on average, we expect for every 1 unit increase/decrease in X will make a b_1 degree increase/decrease in Y (depending on its sign)

b_0 : on average, we expect an actual Y if we set $X=0$

* Evaluating a linear model There is variation in Y .

Total Variation in Y = variation due to X + variation due to e

$\downarrow$ $\downarrow$ $\downarrow$
 SS_{total} SS_{reg} SS_{res}

where, $SS_{total} = \sum (Y - \bar{Y})^2 = SS_{yy}$

$SS_{reg} = \sum (\hat{Y} - \bar{Y})^2 = \frac{(SS_{xy})^2}{SS_{xx}}$

$SS_{res} = \sum (Y - \hat{Y})^2 = \sum e^2 = SS_{total} - SS_{reg}$

$e \sim N(0, \sigma^2)$

* Estimate σ^2 with S^2 :

$S^2 = MS_{res} = \frac{SS_{res}}{n-2}$

* Testing the slope Hypothesis test for β_1

$H_0: \beta_1 = 0$ (X & Y are independent; X is a useless predictor of Y)

vs $H_1: \beta_1 \neq 0$ (X & Y are dependent)

Test statistic: $T = \frac{b_1 - 0}{\sqrt{(MS_{res} / SS_{xx})}} = \frac{b_1}{\sqrt{(MS_{res} / SS_{xx})}}$ with $(n-2)$ d.o.f.

$MS_{res} = \frac{SS_{res}}{n-2}$

correlation coefficient, r %

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}} \sqrt{SS_{yy}}} \quad ; \quad -1 \leq r \leq 1$$

Coefficient of determination: r^2

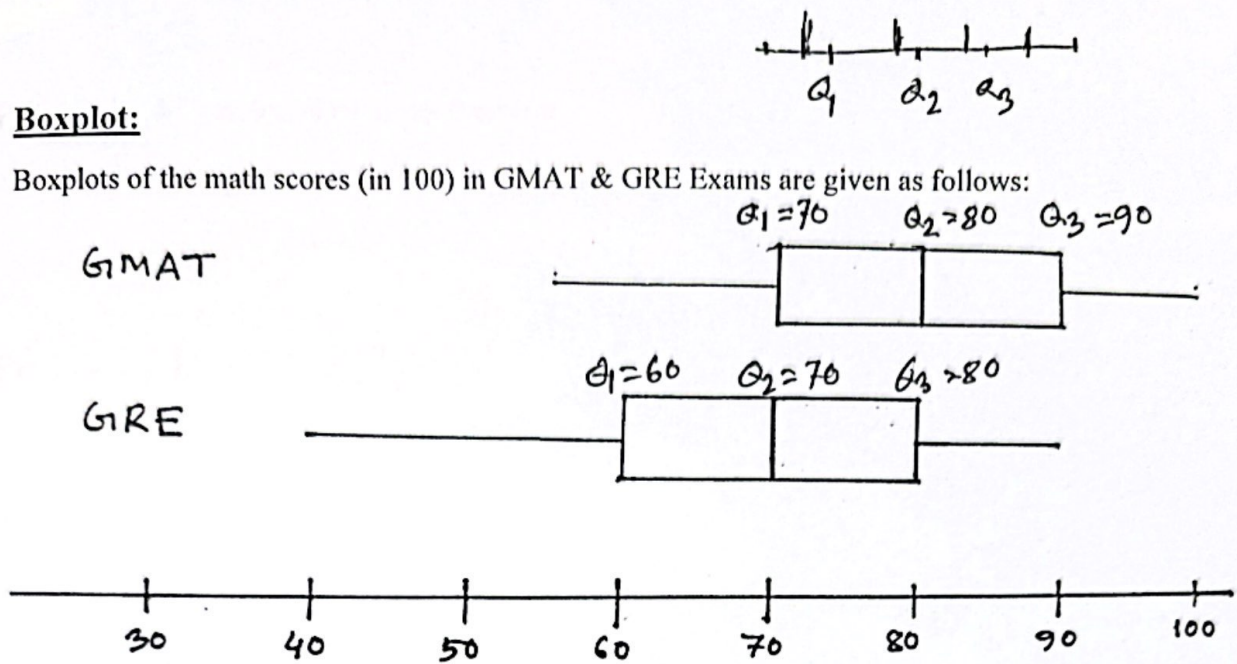
r^2 : the proportion of variation in y that is explained by x

$$SS_{total} = SS_{reg} + SS_{res}$$

$$r^2 = \frac{SS_{reg}}{SS_{total}} \quad (\text{the bigger } r^2 \text{ is, the better})$$

Boxplot:

Boxplots of the math scores (in 100) in GMAT & GRE Exams are given as follows:



Suppose student A got a score 85 in GMAT, Verify the following statements, can we say –

- a) Over 50% of the students in GRE scored higher than student A? *[No; Boxplot doesn't say anything abt # of students]*
- b) Student A's GMAT score is in between the median & 75-th percentile? *Yes*
- c) Some of the students who took the GRE, had a better score than student A? *No*
- d) Student A's GMAT score is less than 80-th percentile? *Yes*
- e) Student A's GMAT score is greater than 75-th percentile? *No*

Binomial Distribution:

A food corporation is introducing a new brand of cereal. To promote its new cereal, 37% of boxes found at grocery stores will have a free toy inside. The probability that any particular cereal box contains a toy equals .37. Assume that consecutive cereal boxes are independent of each other. Let, $Y_1 = \#$ of toys found in a sample of 25 cereal boxes

$$p = 0.37$$

(a) Under which conditions a discrete random variable follows binomial distribution:

n is fixed, each trial has two outcomes, outcomes of each trial are independent of each other, & probability of success remains same for each trial.

$$Y_1 \sim \text{BIN}(n=25, p=0.37)$$

(b) What kind of discrete random variable is Y_1 ?

(c) What are the possible values of Y_1 ?

$$\{0, 1, 2, \dots, 25\}$$

(d) Find the probability that at least two of the cereal boxes will contain a free toy.

$$\begin{aligned} P(Y_1 \geq 2) &= 1 - P(Y_1 \leq 1) \\ &= 1 - P(Y_1 = 1) - P(Y_1 = 0) \\ &= 1 - \binom{25}{1} 0.37^1 (0.63)^{25-1} - \\ &\quad \binom{25}{0} 0.37^0 (0.63)^{25-0} \end{aligned}$$

✓

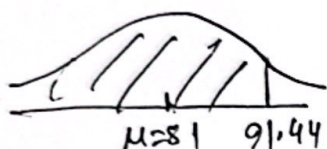
Normal Distribution:

Biologists are studying the weights of 11-year-old American boys. This population of weights is normally distributed with a mean of $\mu = 81$ pounds and a standard deviation of $\sigma = 9$ pounds. Let Y = the weight of a randomly selected 11-year-old American boy.

$$Y \sim N(\mu = 81, \sigma = 9)$$

- (a) What is the probability the weight of a randomly selected boy is less than 91.44 pounds?

$$P(Y < 91.44) \rightarrow = P\left(\frac{Y - \mu}{\sigma} < \frac{91.44 - 81}{9}\right) = P(Z < 1.16) = 0.8770$$



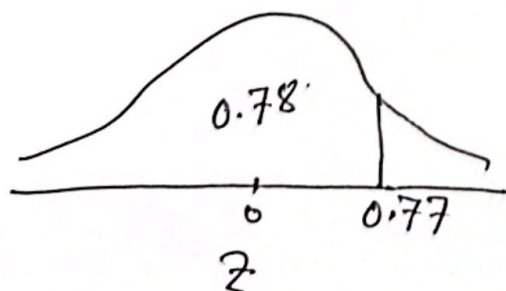
- (b) What is the probability the weight of a randomly selected boy is greater than 85.86 pounds?

$$P(Y > 85.86)$$

- (c) If 50 boys are randomly selected, then what is the probability that the sample mean weight among them is less than 70 pounds? $Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$

$$P(\bar{Y} < 70) = P\left(\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} < \frac{70 - 81}{9/\sqrt{50}}\right) = P(Z < -1.22) = 0.112$$

- (d) Find the 78-th percentile of Y



$$Z = \frac{Y - \mu}{\sigma} \Rightarrow 0.77 = \frac{Y - 81}{9} \Rightarrow Y = 87.9$$

Confidence Interval:

A book publishing company has recently produced a batch of several thousand copies of a hardcover book. They have hired a team of engineers to inspect the machines they use because too many of the books have been assembled with missing pages. p = the proportion of defective books in the batch. 96 of 160 randomly selected books from the batch are defective.

- (a) Use this data to find a 90% confidence interval for p .

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

90% $\rightarrow z_{\frac{\alpha}{2}} = 1.645$

95% $\rightarrow z_{\frac{\alpha}{2}} = 1.96$

- (b) The book publishing company decides that the error margin in the confidence interval is too big, and they would like a 90% confidence interval for p to have an error margin no larger than .01. What sample size is necessary to achieve this? Use the estimate of p from the original sample (96 defective books out of 160).

$$n = \hat{p}(1-\hat{p}) \left(\frac{z_{\frac{\alpha}{2}}}{B} \right)^2$$

Error margin,
 $B = 0.01$

Control chart:

Problem 01: Suppose, adidas soccer ball manufactures $n = 2000$ soccer balls every day. Before shipment to retailers, every soccer ball is subjected to a pressure test. Balls that crack during the pressure test are considered failures and then discarded. Engineers use data from the last 70 days to make an np chart. The first $m = 20$ days are used as the base period.

During the base period, a total of 800 soccer balls failed the pressure test.

- 1) Compute UCL and LCL.

$$\bar{p} = \frac{1}{mn} \sum_{i=1}^m Y_i = \frac{1}{20(2000)} [800] = .02$$

UCL and LCL for np chart:

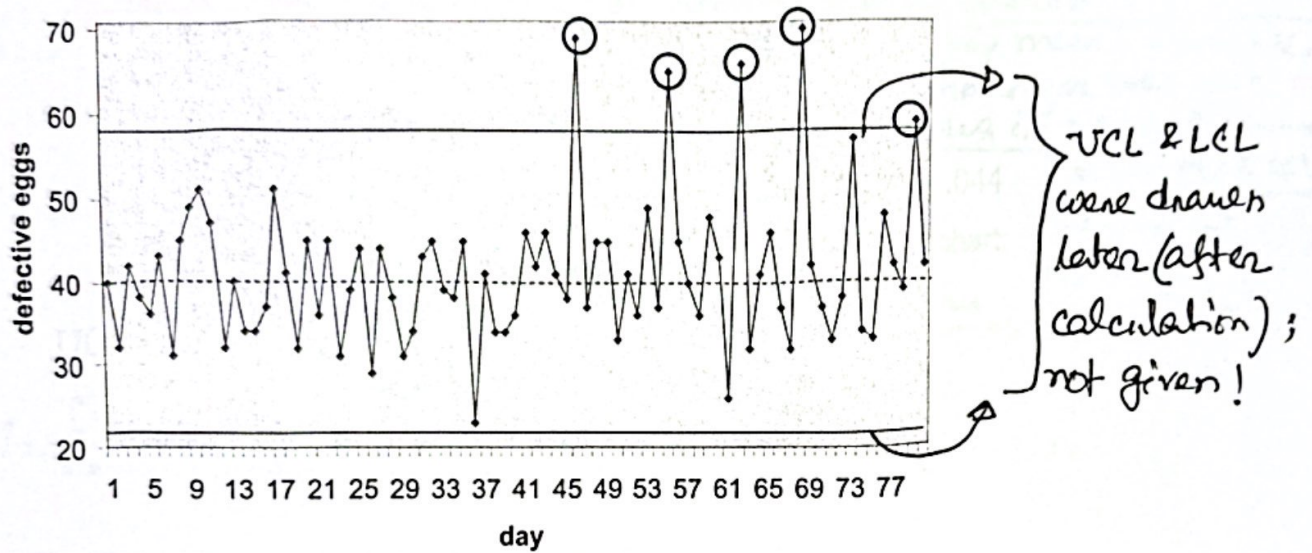
$$\begin{aligned} n\bar{p} \pm 3\sqrt{n\bar{p}q} &= 2000(.02) \pm 3\sqrt{2000(.02)(.98)} \\ &= 40 \pm 18.783 = \boxed{\{21.217, 58.783\}} \end{aligned}$$

- 2) What assumptions about the data are you required to make?

The process was in control during the base period. The observations came from a binomial distribution.

3) Look at the following chart, does this process appear to be in control? Comment.

Problem 5.3 np chart



The np chart appears to be out of control. There are five observations above the UCL.

$$\{UCL, LCL\} = \bar{\bar{y}} \pm 3 \sqrt{\frac{s^2}{n}} = \mu \pm 3 \sqrt{\frac{\sigma^2}{n}}$$

$3 \frac{\sigma}{\sqrt{n}}$

Problem 02: Since the hardness of the maple bats contributes to their popularity, Sweet Club Inc. would like to monitor the density of the bats they produce. Of the dozens of bats produced per day, five are selected at random and the sample mean density is measured (mg/cm³). Some people have collected the data for the last 88 days. Using the first 22 samples as the base period, we have found the base period average mean as $\bar{\bar{Y}} = 702$, and base period average variance as 24.044.

- 1) Compute UCL and LCL for $\bar{X}$ chart.

$$\bar{\bar{Y}} = 702 \quad \bar{s}^2 = 24.044$$

UCL and LCL for $\bar{X}$ chart:

$$\bar{\bar{Y}} \pm 3 \sqrt{\frac{s^2}{n}} = 702 \pm 3 \sqrt{\frac{24.044}{5}} = \{695.42, 708.58\}$$

Note: we used base period info as we don't know pop mean & variance. If we know those, we no longer need base period info & directly plug in pop mean & variance in the UCL & LCL formula.

$$\bar{\bar{Y}} = 702 \quad \bar{s}^2 = 24.044$$

UCL and LCL for s^2 chart:

$$UCL = \frac{(\chi^2_{n-1, 0.0135}) \bar{s}^2}{n-1} = \frac{(17.8) 24.044}{4} = 106.996$$

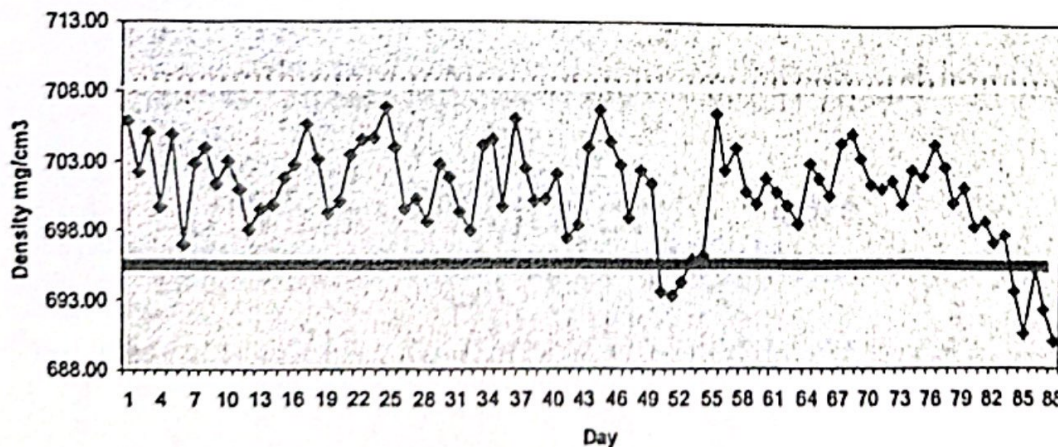
$$LCL = \frac{(\chi^2_{n-1, 0.9865}) \bar{s}^2}{n-1} = \frac{(0.106) 24.044}{4} = 0.637$$

- 2) What assumptions about the data are you required to make?

The process was in control during the base period. The data is normally distributed. The process variance remains constant

- 3) Look at the following chart, does this process appear to be in control? Comment.

Problem 5.1 - Density of Baseball Bats: X-Bar chart



The X-bar chart dips between the LCL several times. We recommend that the process be inspected immediately.

Computation of Regression Analysis:

Given the followings:

$$\sum X = 252, \sum Y = 483, \sum X^2 = 9202, \sum Y^2 = 33367, \sum XY = 17435, n = 7$$

Where, X: Independent Variable, Y: Dependent Variable.

(a) Find the correlation coefficient.

$$r = \frac{SS_{XY}}{\sqrt{SS_{XX}} \sqrt{SS_{YY}}} \\ = 0.6518$$

$$\begin{aligned} SS_{XY} &= \sum XY - \frac{(\sum X)(\sum Y)}{n} \\ &= 17435 - \frac{(252)(483)}{7} \\ &= 47 \end{aligned}$$

$$SS_{XX} = \sum X^2 - \frac{(\sum X)^2}{n} = 130$$

$$SS_{YY} = \sum Y^2 - \frac{(\sum Y)^2}{n} = 40$$

(b) Fit the least square regression line.

$$\hat{Y} = b_0 + b_1 X$$

$$\begin{aligned} b_0 &= \bar{Y} - b_1 \bar{X} \\ &= 55.98 \end{aligned}$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}} = \frac{47}{130} = 0.3615$$

$$\begin{aligned} \bar{X} &= \frac{\sum X}{n} = 36 \\ \bar{Y} &= \frac{\sum Y}{n} = 69 \end{aligned}$$

$$\hat{Y} = 55.98 + 0.3615 X$$

(c) What is the predicted value of Y (on average) if X=35?

$$\hat{Y} = 55.98 + 0.3615(35) = 68.64$$

(d) Find the coefficient of determination.

$$r^2 = \frac{SS_{\text{reg}}}{SS_{\text{total}}}$$

$$= \frac{16.99}{40}$$

$$= 0.4258$$

OR

$$r^2 = (r)^2 = (0.6518)^2$$

$$= 0.4258$$

; 42.58% variation in y explained by x.

(e) Test $H_0: \beta_1 = 0$ vs $H_a: \beta_1 \neq 0$, consider the level of significance, $\alpha = 0.05$.

$$RR: |T| > T_{\frac{\alpha}{2}, n-2} = 2.571$$

Test statistic:

$$T = \frac{b_1}{\sqrt{\frac{MS_{\text{res}}}{SS_{xx}}}}$$

$$= 1.92 < 2.571$$

Fail to reject H_0 ; X is not a sig. variable to predict y

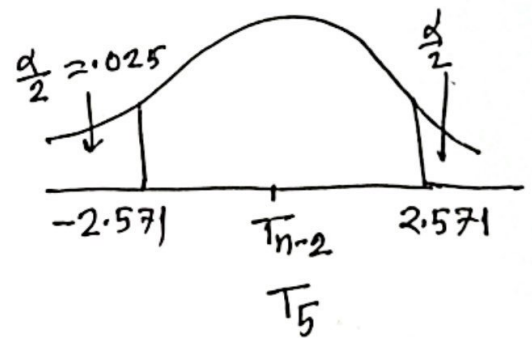
$$SS_{\text{reg}} = \frac{(SS_{xy})^2}{SS_{xx}} = \frac{(47)^2}{130}$$

$$= 16.99$$

$$SS_{\text{total}} = SS_{yy} = 40$$

$$SS_{\text{res}} = SS_{\text{total}} - SS_{\text{reg}}$$

$$= 40 - 16.99 = 23.01$$



$$MS_{\text{res}} = \frac{SS_{\text{res}}}{n-2}$$

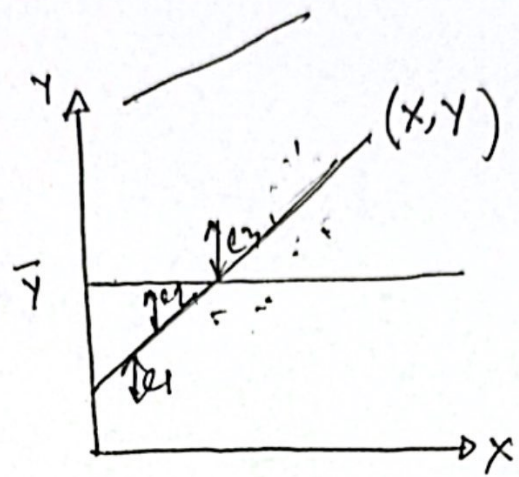
$$= \frac{23.01}{5} = 4.602$$

$$b_1 = 0.3615$$

Least square regression line: Residual/Error:

Properties of residual/error:

- 1) Sum of the errors/residuals is zero. $\sum e = 0$
- 2) Sum of the squared errors is minimized.



Residual/error plot Reveals problem with the dataset; random cloud of points is good; any pattern is bad.

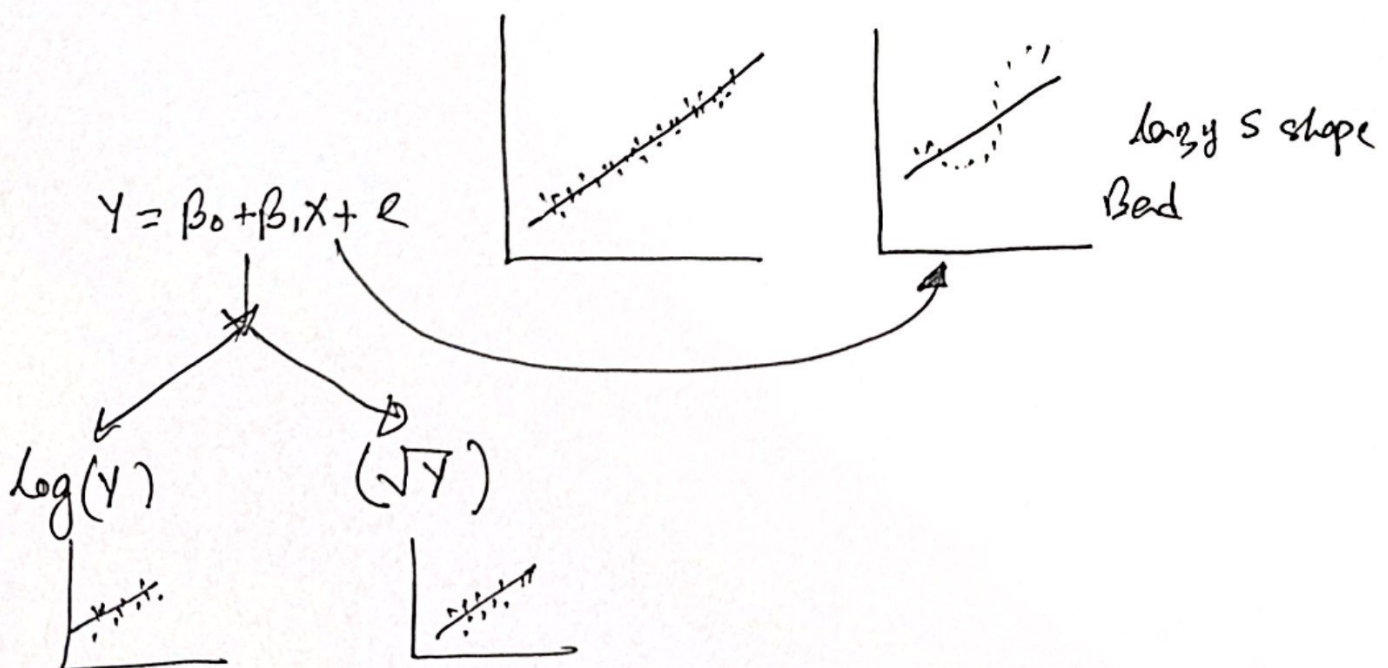


Scatterplot:

- 1) Shows relationship between paired data (x,y).
- 2) Examine by trend (linear, curved, cluster, or no pattern), association/direction (positive, negative, or no direction), strength (how closely the points fit the trend/least square regression line), and outlier.
- 3) Can be used to guess the value of correlation coefficient (r). [Reg by eye!]

Normal Q-Q plot:

Checks the normality assumption; if violated, the first thing we do is the transformation of the response/dependent variable (Y) by taking logarithm or square root transformation.



Stem-and Leaf Plot:

<u>Stem</u>	<u>Leaf</u>
51	3
52	8
53	4 5
54	1 6 6
55	2 7 8

Properties of probability distribution:

Probability Laws:

Expected value of a discrete random variable:

There is an unfair die with unusual numbers on the six faces. If Y = the outcome of one roll of the die, then the probability distribution is given below. Suppose that every time you roll this die, you win the number of dollars equal to the number that appears. There is a catch, of course, and that is you must pay 12 dollars for each roll. In the long run, do you expect to win money, lose money, or break even? Why? If you do any calculations to answer this question, show your work.

Y	$P(Y=y)$
4	.30
6	.10
9	.10
11	.10
15	.20
30	.20

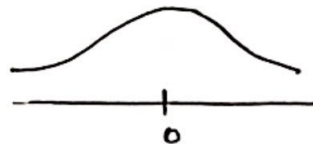
Poisson Distribution:

A cashier works the midnight shift at a grocery store. The number of customers that come to her register per hour follows a Poisson distribution with $\lambda = 5.4$. Let, $Y_4 = \#$ of customers per hour a customer comes to her register

- (a) What is the probability that exactly 3 customers will come to her register in the next hour?
- (b) What is the probability that four or more customers will come to her register in the next hour?

Properties of standard normal distribution:

- 1) Z-distribution with $\mu = 0$ & $\sigma = 1$
- 2) Symmetric about mean
- 3) Bell-shaped
- 4) Never touches the x-axis.



Hypothesis Testing:

A construction company is making a huge batch of thousands of concrete blocks for a new project. The mean weight among all the blocks should be 25 kg. It is too time consuming to conduct a census and weigh every single block, so the company randomly selects 15 blocks to weigh. The sample mean weight among the 15 blocks is 25.5 kg and the sample standard deviation is $s = 0.6$ kg. σ is unknown. Assume the population of weights is normally distributed.

Do you have significant evidence that μ is different from 25 kg? Report H_0 , H_a , the rejection region (use $\alpha = .10$), test statistic, conclusion, and place bounds on the p-value.

Chapter 3.4 (Continuous Random Variables):

Example 1: Uniform Distribution

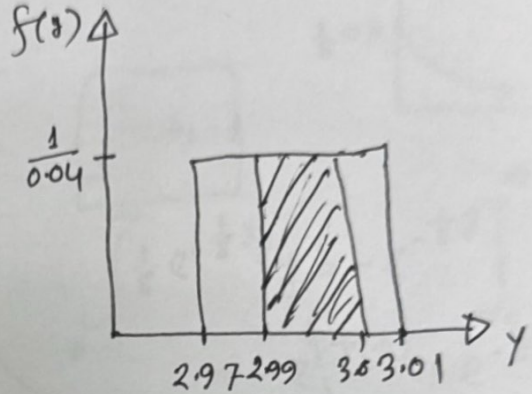
Suppose, Y = weight (grams) of a randomly selected car part.

$Y \sim \text{UNIF}(2.97, 3.01)$

$$f(y) = \frac{1}{b-a} = \frac{1}{3.01-2.97} = \frac{1}{0.04}$$

(a) Find $P(2.99 < Y < 3.00)$

$$\begin{aligned} &= \int_{2.99}^{3.00} f(y) dy = \int_{2.99}^{3.00} \frac{1}{0.04} dy \\ &= 0.25 \end{aligned}$$



(b) Find μ_y & σ_y

$$\mu_y = \frac{a+b}{2} = \frac{2.97+3.01}{2} = 2.99$$

$$\sigma_y^2 = \frac{(b-a)^2}{12} = \frac{(3.01-2.97)^2}{12} = \frac{1}{7500}$$

$$\sigma_y = \sqrt{\frac{1}{7500}}$$

Exponential r.v. - a special continuous r.v.

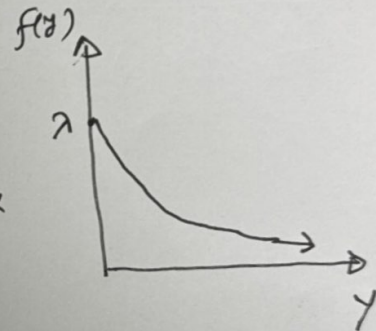
→ Closely Related to the discrete r.v. named Poisson.

→ Recall: A poisson r.v. counts the # of occurrences of an event in a specified time period.

$$Y_{\text{poi}} = \{0, 1, 2, \dots\} \quad \lambda = \text{mean rate of occurrences}$$

→ Here, the exponential r.v. is the waiting time until the next occurrence.

$$f(y) = \begin{cases} \lambda e^{-\lambda y} & ; y > 0 \\ 0 & ; \text{otherwise} \end{cases}$$



$$\begin{aligned} \mu = E(Y) &= \int_0^{\infty} y f(y) dy \\ &= \int_0^{\infty} y \lambda e^{-\lambda y} dy = \frac{1}{\lambda} \end{aligned}$$

$$\sigma^2 = E[Y^2] - \mu^2$$

$$E(Y^2) = \int_0^{\infty} y^2 f(y) dy = \int_0^{\infty} y^2 \lambda e^{-\lambda y} dy = \frac{2}{\lambda^2}$$

$$\sigma^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda} = \frac{1}{\lambda^2}$$

CH 3.4 Ex. 02 handout

$$Y_{Exp} \sim \text{Exp}(\lambda = \frac{1}{8})$$

Example 2: Exponential Distribution

$$f(y) = \lambda e^{-\lambda y}; y > 0$$

Suppose, Y_{Poi} = # of lightning strikes to ME-EM per year. $Y_{Poi} \sim \text{Poisson}(\lambda = 0.125 = 1/8)$.

$$\mu_{Y_P} = \frac{1}{8}$$

Let, Y_{Exp} = # of years until the next strike (i.e. waiting time until next strike)

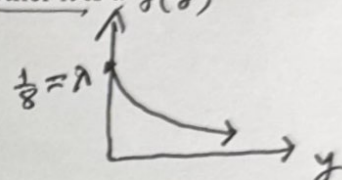
- (a) Find $E(Y_{Exp})$ & construct the probability density function of Y_{Exp} . Check whether it is a true p.d.f.

$$E(Y_{Exp}) = \mu_{Y_E} = \frac{1}{\lambda} = 8.$$

$$f(y) = \begin{cases} \frac{1}{8} e^{-\frac{1}{8}y} & ; y > 0 \\ 0 & ; \text{otherwise} \end{cases}$$

$$\int_{-\infty}^{\infty} f(y) dy = 1$$

$$\int_0^{\infty} \frac{1}{8} e^{-\frac{1}{8}y} dy = -e^{-\frac{1}{8}y} \Big|_0^{\infty} = -e^{-\frac{1}{8}(\infty)} + e^{-\frac{1}{8}(0)} = 0 + 1 = 1$$

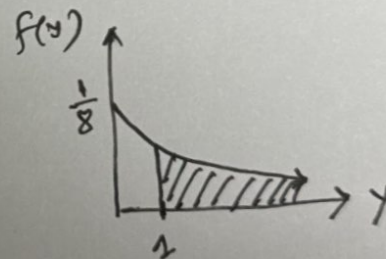


- (b) Find
P(no lightning strikes for the next year) =

$$P(Y_{Poi} = 0) = \frac{(\frac{1}{8})^0 e^{-1/8}}{0!} = e^{-1/8}$$

- (c) Find
P(waiting time until next strike is one year or more) =

$$P(1 < Y_{Exp} < \infty) = \int_1^{\infty} \frac{1}{8} e^{-\frac{1}{8}y} dy = e^{-1/8}$$



- (d) Find
P(we will observe at least one lightning strike within the next 66 months) =

$$P(0 < Y_{Exp} < 5.5) = \int_0^{5.5} \frac{1}{8} e^{-\frac{1}{8}y} dy \approx 0.497$$

MA 3710 Engineering Statistics EXAM 01 Announcement

The exam will cover the following sections:

Chapter 1

Population, sample

All of Chapter 2

Histogram, histogram shapes, dot plot, stem-and-leaf plot, percentile, median, quartiles, interquartile range (IQR), & boxplot.

3.1, 3.2, 3.3, 3.4

Set, empty set, subset, sample space, Venn diagram, intersection, union, complement, mutually exclusive, conditional probability, independence, mutually exclusive, multiplicative laws, the law of total probability, additive law, complement rule, tree diagram, probability table, Bayes Rule, drawing with or without replacement, random variable, discrete random variable, probability function, population mean (expected value) and variance and standard deviation of a discrete random variable, combination/choosing symbol (Binomial coefficient), Binomial random variable, Geometric random variable, Poisson random variable, continuous random variable, the population mean (expected value) of a continuous random variable, Uniform distribution, & Exponential distribution.



Michigan
Technological
University

Department of Mathematical Sciences

Engineering Statistics

Write your full name at the top of your answer paper.

Read each question carefully and *answer all parts of each question*. Show all your works for any computations and formulas.

For MCQ or T/F/N question, just write the answer shortly. For instance, write down the answer of the MCQ question like: 1(b), 2(c) & T/F/N question like: 3 – T, 4 – F, 5 – N.

Even if you don't know the final answer, providing some work is gradable, **REMEMBER, for essay questions, NO WORK MEANS NO POINTS!!!**

Cheating or Plagiarism:

Proper professional and ethical behavior is expected of all students in this class. ***This is taken very seriously.*** If cheating is suspected (in this course cheating means typing anywhere except your calculator) a score of zero will be given for the exam/course. At the discretion of the instructor, the Dean of Students will be notified.

Sign the honor code

"On my honor, I have neither given nor received unauthorized aid on this examination."

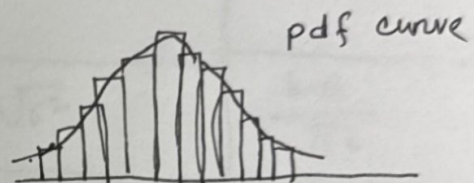
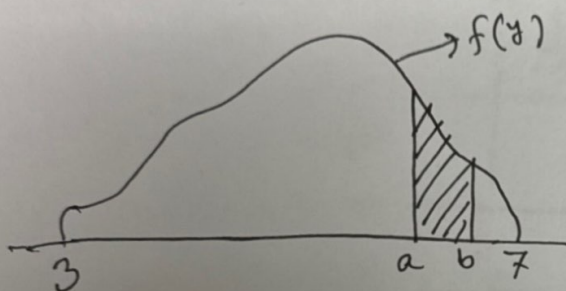
Write your Full Name here: _____

Good Luck!

Continuous Random Variable

→ Recall, cont. R.V. have an uncountable (unlistable/squeezable) set of possible value.

→ probability Density function (p.d.f.) = $f(y)$



$$\text{shaded area} = \int_a^b f(y) dy = P(a < Y < b)$$

properties:

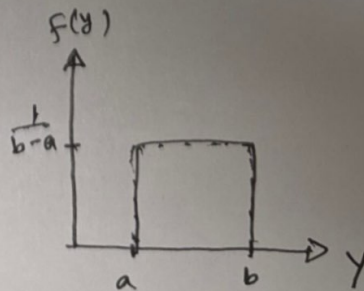
(i) $f(y) \geq 0$

(ii) $\int_{-\infty}^{\infty} f(y) dy = 1$

☐ Uniform R.V. - a famous cont. R.V.

If $Y \sim \text{UNIF}(a, b)$, then

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



MA 3710: Md Mutasim Billah
EXAM 01 Review

HW CH 1 & 2: Problem 2.3

Biologists are studying a woman who has given birth to twenty children (ten boys and ten girls). When each child reached 10 years of age, their height (in inches) was recorded:

Boys:

51.3, 52.8, 53.4, 53.5, 54.1, 54.6, 54.6, 55.2, 55.7, 55.8

Girls:

50.5, 51.0, 51.4, 52.7, 52.8, 53.1, 53.6, 54.4, 54.4, 55.3

Use this data to draw side-by-side stem-and-leaf plots. Have your stems be 50, 51, 52, 53, 54, 55.

Stem	Boys	Girls
50	-	5
51	3	0 4
52	8	7 8
53	4 5	1 6
54	1 6 6	4 4
55	2 7 8	3

Symmetric

Range: MAX - MIN

Boys: $55.8 - 51.3 = 4.5$

Girls: $55.3 - 50.5 = 4.8$ *higher*

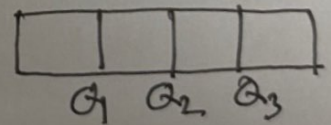
(Boys) Locating median Q_2 : $l_m = \frac{n+1}{2}$

$n = 10$; $l_m = \frac{10+1}{2} = 5.5$

median: avg of 5th & 6th value: $\frac{54.1 + 54.6}{2} = 54.35$

Locating Q_1 :

$$l_q = \begin{cases} \frac{n+3}{4} & \text{if } n \text{ is odd} \\ \frac{n+2}{4} & \text{if } n \text{ is even.} \end{cases}$$



Boys, $n = 10$ (even) $l_q = \frac{n+2}{4} = \frac{10+2}{4} = 3$

MA 3710: Md Mutasim Billah
EXAM 01 Review

HW CH 1 & 2: Problem 2.3

Biologists are studying a woman who has given birth to twenty children (ten boys and ten girls). When each child reached 10 years of age, their height (in inches) was recorded:

Boys:

51.3, 52.8, 53.4, 53.5, 54.1, 54.6, 54.6, 55.2, 55.7, 55.8

Girls:

50.5, 51.0, 51.4, 52.7, 52.8, 53.1, 53.6, 54.4, 54.4, 55.3

Use this data to draw side-by-side stem-and-leaf plots. Have your stems be 50, 51, 52, 53, 54, 55.

Stem	Boys	Girls
50	-	5
51	3	0 4
52	8	7 8
53	<u>4</u> 5	1 6
54	<u>6</u> <u>6</u> 6	4 4
55	<u>2</u> 7 8	3

left

Symmetric

Q_1

Q_3

Range: MAX - MIN

Boys: $55.8 - 51.3 = 4.5$

Girls: $55.3 - 50.5 = 4.8$ *height*

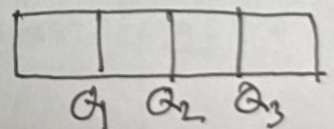
(Boys) Locating median Q_2 : $L_m = \frac{n+1}{2}$

$n = 10$; $L_m = \frac{10+1}{2} = 5.5$.

median: avg of 5th & 6th value: $\frac{54.1 + 54.6}{2} = 54.35$

Locating Q_1 :

$$L_q = \begin{cases} \frac{n+3}{4} & \text{if } n \text{ is odd} \\ \left\lfloor \frac{n+2}{4} \right\rfloor & \text{if } n \text{ is even.} \end{cases}$$



Boys, $n = 10$ (even)

$L_q = \frac{n+2}{4} = \frac{10+2}{4} = 3$.

$Q_1 = Y_{(L_q)} = 53.4$

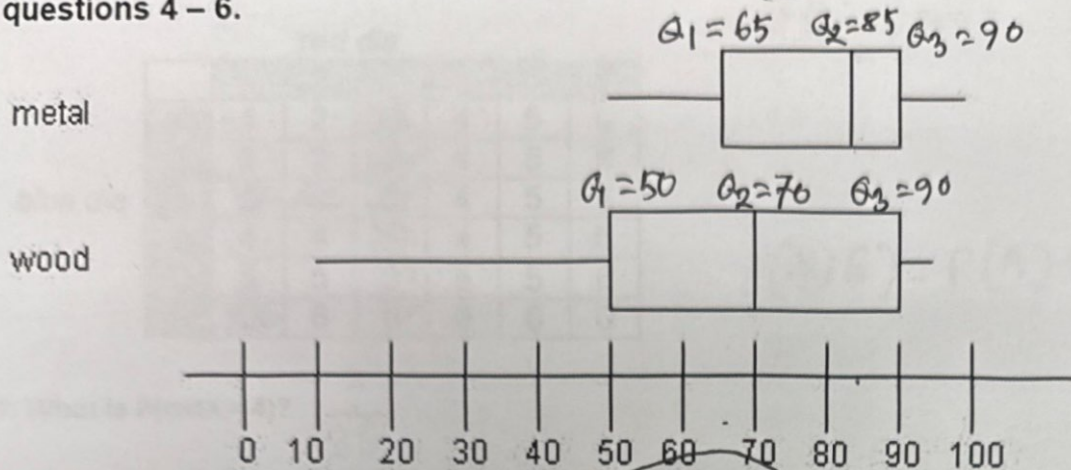
Q_3 : $n+1-L_q = 10+1-3 = 8$

11 Page

$Q_3 = Y_{(n+1-L_q)} = 55.2$ IQR: $Q_3 - Q_1 = 1.8$

Spr 09 Exam 01: Problem 02

Every semester, an engineering professor assigns a project to her students to construct a catapult. Each catapult receives a score out of 100. The professor notices that catapults built out of metal tend to be of higher quality and earn higher scores than those built out of wood. Side-by-side boxplots were made with data she has collected over the years. Answer questions 4 – 6.



Histogram of metal: Left skewed

Bell shaped:

④ metal: $IQR = Q_3 - Q_1 = 90 - 65 = 25$

Wood: $IQR = Q_3 - Q_1 = 90 - 50 = 40$ (greater)

④ There are more wood staffs than metal staffs.
[T / F / Not enough information] $n = ?$

④ Upper Inner fence UIF: $Q_3 + \text{step}$ | $\text{step} = 1.5 (IQR)$
Lower Inner fence LIF: $Q_1 - \text{step}$ | $= 40 (\text{wood})$

Wood: $UIF = 90 + 60 = 150$; $LIF: Q_1 - \text{step} = 50 - 60 = -10$

④ Lowest value of the obs. ~~(A)~~ + metal? (wood)?

④ 15th percentile of wood score is less than 50
[T / F / N] [Q1 = 50]

Spr 09 Exam 01: Problem 03

When we roll two dice, we usually add the two numbers that appear. Suppose instead we pay attention to the maximum. For example, if red = 3 and blue = 4, then the maximum equals 4. The table below has 36 cells corresponding to 36 equally likely outcomes. The maximum corresponding to each outcome appears in the appropriate cell. Answer questions 7 – 10.

	red die					
	1	2	3	4	5	6
blue die	1	1	2	3	4	5
	2	2	2	3	4	5
	3	3	3	3	4	5
	4	4	4	4	4	5
	5	5	5	5	5	5
	6	6	6	6	6	6

max (blue, red)

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

7. What is $P(\max = 4)$? $\frac{7}{36}$

8. What is $P(\max = 6 \cup \text{red} = 1)$? $\frac{16}{36}$

$$P(\max = 6) + P(\text{red} = 1) - P(\max = 6 \cap \text{red} = 1)$$

$$= \frac{11}{36} + \frac{6}{36} - \frac{1}{36} = \frac{16}{36}$$

9. What is $P(\max = 3 \mid \text{blue} = 3)$?

$$= \frac{P(\max = 3 \cap \text{blue} = 3)}{P(\text{blue} = 3)} = \frac{3/36}{6/36} = \frac{3}{6} = \frac{1}{2}$$

HW CH 3a: Problem 3.8

3.8

There is an unfair die with unusual numbers on the six faces. If Y = the outcome of one roll of the die, then the probability distribution is given below. Suppose that every time you roll this die, you win the number of dollars equal to the number that appears. There is a catch, of course, and that is you must pay 12 dollars for each roll. In the long run, do you expect to win money, lose money, or break even? Why? If you do any calculations to answer this question, show your work.

Y	P(Y=y)
4	.30
6	.10
9	.10
11	.10
15	.20
30	.20

$$M = E(Y) = \sum y P(Y=y)$$

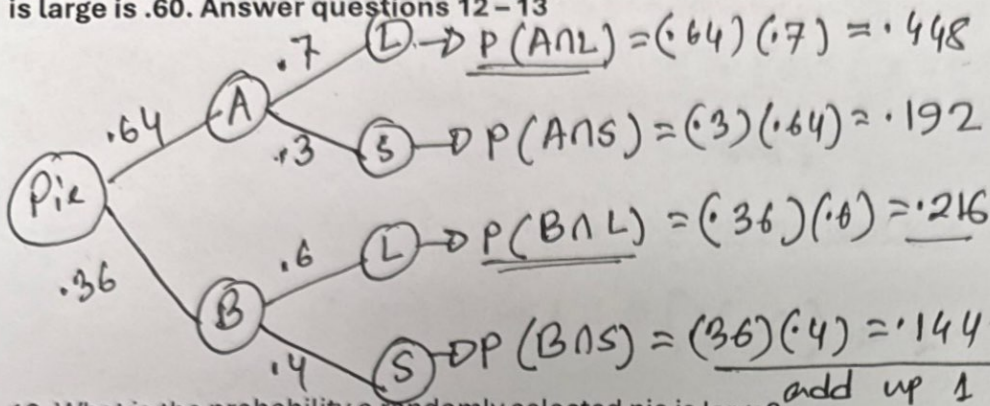
$$= 4(.30) + 6(.10) + 9(.10) + \dots + 30(.20)$$

$$= 12.80$$

(Win)

Spr 09 Exam 01: Problem 05

Bob's Bakery makes two kinds of pies: apple and blueberry. If you select a pie at random, the probability it is apple is .64. Bob's Bakery also makes two sizes of pies: large and small. The probability a randomly selected apple pie is large is .70. The probability a randomly selected blueberry pie is large is .60. Answer questions 12-13



A: Apple

B: Blue berry

$$P(A) = .64$$

$$P(B) = 1 - P(A)$$

$$= 1 - .64 = .36$$

L: large, S: Small

$$P(L|A) = .70$$

$$P(A|L) =$$

$$P(L|B) = .60$$

12. What is the probability a randomly selected pie is large?

$$P(L) = P(AL) + P(BL) = .448 + .216 = .664$$

13. Given that a randomly selected pie is small, what is the probability it is an apple pie?

$$P(A|S) = \frac{P(AS)}{P(S)}$$

$$P(S) = 1 - P(L) = .336$$

Spr 09 Exam 01: Problem 06

Some balls will be drawn at random one at a time out of a basket that contains 100 balls (53 are orange, 41 are yellow, and 6 are purple). For questions 14-17, identify each variable as "binomial", "geometric" or "neither". If your answer is "binomial", identify the values of n and p. If your answer is "geometric", identify the value of p.

14. A = # of purple balls that appear out of 5 draws (without replacement) (N)

15. B = # of yellow balls that appear out of 11 draws (with replacement). Find P(B=5).

$$B \sim \text{BIN}(n=11, p=\frac{41}{100})$$

$$n=11$$

$$p(s) = \frac{41}{100}$$

16. C = # of draws needed to observe the first orange ball (without replacement) (N)

17. D = # of draws needed to observe the first purple ball (with replacement).

$$D \sim \text{Geo}(p) ; p = \frac{6}{100}$$

18. Find probability of getting at least one yellow out of 4 draws.

$$1 - P(\text{no yellow}) = 1 - \left(\frac{59}{100}\right)^4$$

Spr 09 Exam 01: Problem 08

WebRock, an online music site, has become very popular and its server is overloaded 3.2 times a day, on average. If $Y = \#$ of times in a day that the server is overloaded, then Y is Poisson with $\lambda = 3.2$. Answer questions 20 – 21.

20. What is the probability that the server will be overloaded exactly 2 times tomorrow?

$$Y \sim \text{Pois}(\lambda = 3.2) \quad P(Y=y) = \frac{e^{-\lambda} \lambda^y}{y!}$$

$$P(Y=2) = \frac{e^{-3.2} (3.2)^2}{2!} = .2087$$

21. What is the probability that the server will be overloaded at least once tomorrow?

$$P(Y \geq 1) = 1 - P(Y=0) = 1 - \frac{e^{-3.2} (3.2)^0}{0!} = 0.9595$$

Uniform Distribution

Spr 09 Exam 01: Problem 09

You decide to purchase a used truck from your friend Paul. Paul warns you that the truck is very old and he believes that the waiting time (in months) until the transmission fails is an exponential random variable with $\lambda = 0.05$. Answer questions 22 – 24.

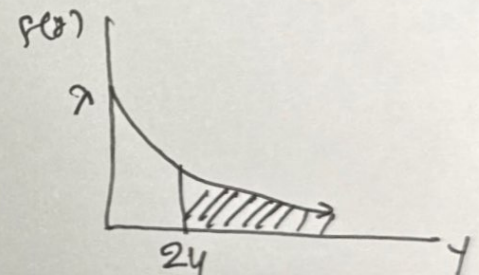
Y : waiting time (months) until it fails

22. If Paul is correct, what is the probability that the transmission on the truck will survive for at least the next 24 months?

$$Y \sim \text{Exp}(\lambda = 0.05)$$

$$P(24 < Y < \infty) = \int_{24}^{\infty} 0.05 e^{-0.05y} dy$$

$$24 \Rightarrow e^{-1.2} = .3012$$



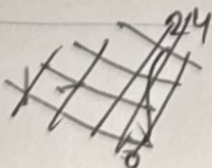
23. What is the expected waiting time until the transmission fails?

$$\mu_{\text{Exp}} = \frac{1}{\lambda} = \frac{1}{0.05} = 20 \text{ months.}$$

24. What discrete random variable is closely related with the exponential distribution? Can you define a discrete random variable following that for this problem?

Poisson.

Y_{pois} : # of failure in a given time





Michigan
Technological
University

Department of Mathematical Sciences

Engineering Statistics

MA 3710 R02

Solution

Exam 1 Spring 2025

Time: 2:00pm-02:50pm

Read each problem carefully and *answer all parts of each problem*. Show all your works for any computations and formulas.

There are 8 problems and 24 parts to solve.

Even if you don't know the final answer, providing some work is gradable. Remember, for *essay questions, NO WORK MEANS NO POINTS!!!*

Cheating or Plagiarism:

Proper professional and ethical behavior is expected of all students in this class. *This is taken very seriously*. If cheating is suspected a score of zero will be given for the exam/course. At the discretion of the instructor, the Dean of Students will be notified.

Sign the honor code

"On my honor, I have neither given nor received unauthorized aid on this examination."

Full Name: md mutasim Billah

Good Luck

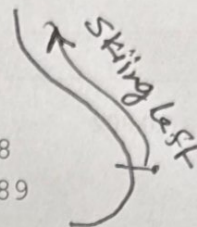
Problem 01: Fifteen boys and twenty girls at a high school make side-by-side stem-and-leaf plots with their heights (in inches, rounded off to the nearest tenth of an inch). For example, the shortest boy is 65.4 inches tall and the tallest girl is 69.9 inches tall. Answer questions 1 – 3.

Boys:

stem	leaf
65	488
66	068
67	389
68	348
69	245

Girls:

stem	leaf
65	36
66	026
67	4669
68	02478
69	355789



1. Which data set is left-skewed? Circle one answer.

- ☒ a) girls
- b) Neither data set is left-skewed.
- c) boys

2. Which data set has a larger range? Circle one answer.

- a) girls
- b) Both data sets have the same range.
- c) boys

$$\text{Boys: } 69.5 - 65.4 = 4.1$$

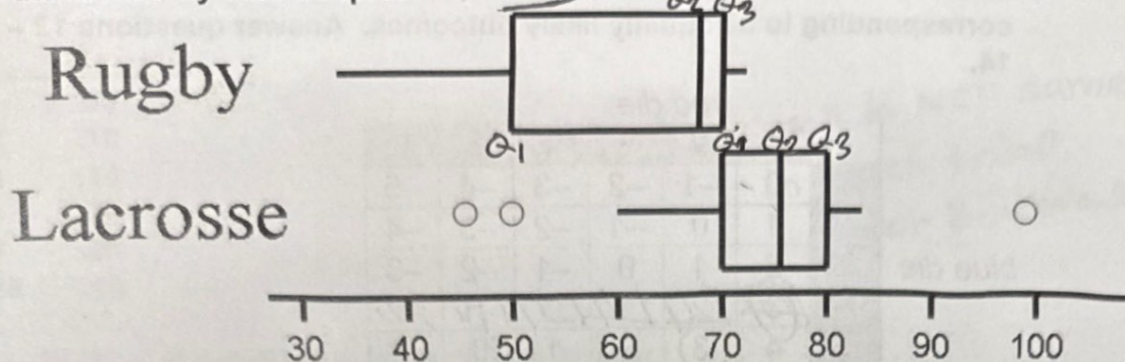
$$\text{Girls: } 69.9 - 65.3 = 4.6$$

3. What is the median height among the boys?

$$L_m = \frac{n+1}{2} = \frac{15+1}{2} = 8$$

8th value $\rightarrow$ 67.8 inches

Problem 02: A course titled "Statistics for Athletes" is populated by Rugby players and Lacrosse players. Scores from the last midterm exam are used to make side-by-side boxplots.



For problems 4 – 11, circle T (true), F (false), or N (not enough info) after each statement.

4. A histogram of the rugby player's scores would most likely be ~~right~~ *left*-skewed.

T ☒ F N

5. The rugby players have a ~~smaller~~ *greater* IQR than the lacrosse players.

T ☒ F N

6. There are more rugby players than lacrosse players.

T F ☒ N $n = ?$

7. The 85th percentile among the rugby players must be greater than 70.

☒ T F N *since $Q_3 = 70$*

8. The lower inner fence for the lacrosse players is located at 55.

☒ T F N $70 - 1.5(10) = 55$

9. The lower inner fence for the rugby players is located at ~~70~~ *20*.

T ☒ F N

10. The lowest score in the statistics class belongs to a rugby player.

☒ T F N

11. The 35th percentile among the lacrosse players equals 60.

T ☒ F N *since $Q_1 = 70$*

Problem 03: When we roll two dice, we usually add the two numbers that appear. Suppose instead we pay attention to D = the value on the blue die minus the value on the red die. The table below has 36 cells corresponding to 36 equally likely outcomes. Answer questions 12 – 14.

		red die					
		1	2	3	4	5	6
blue die	1	0	-1	-2	-3	-4	-5
	2	1	0	-1	-2	-3	-4
	3	2	1	0	-1	-2	-3
	4	3	2	1	0	-1	-2
	5	4	(3)	2	1	0	-1
	6	5	4	(3)	2	1	0

12. Find $P(D = 3)$

$$\frac{3}{36}$$

13. Find $P(D \geq 0 | \text{blue} = 4) = P(D \geq 0 | \text{blue} = 4) = \frac{P(D \geq 0 \cap \text{blue} = 4)}{P(\text{blue} = 4)}$

$$= \frac{4/36}{6/36} = \frac{2}{3}$$

14. The events $\text{red} = 5$ and $D = 0$ are _____. Circle one answer.

- a) independent and mutually exclusive
- ☒ b) independent but not mutually exclusive
- c) mutually exclusive but not independent
- d) not independent and not mutually exclusive

$$P(\text{red} = 5 \cap D = 0) \stackrel{?}{=} P(\text{red} = 5) P(D = 0)$$

not m.e. $0 \neq \frac{1}{36} \stackrel{?}{=} \frac{1}{6} * \frac{6}{36}$

yes Independent

Problem 04: There is an unfair die with unusual numbers on the six faces. If Y = the outcome of one roll of the die, then the probability function is given below. Answer question 15.

Y	$P(Y=y)$
1	.30
2	.10
3	.10
6	.20
9	.20
25	.10

$p(\text{success})$ is NOT same
in each trial
not Binomial

15. What statement is true? Circle one answer.

- a) Y is a Geometric random variable.
- ☒ b) Y is discrete, but it isn't Binomial, Poisson, or Geometric.
- c) Y is a Binomial random variable.
- d) Y is a Poisson random variable

Problem 05: Y = the number of fluorescent light bulbs that need to be replaced (per day) in an office building. Y is Poisson with $\lambda = 3.2$. Answer question 16.

16. What is the probability at least one light bulb will need to be replaced tomorrow?

$$P(Y \geq 1) = 1 - P(Y=0) = 1 - \frac{(3.2)^0 e^{-3.2}}{0!}$$

$$\approx \boxed{.9592}$$

$$P(Y \neq 1)$$

Problem 06: For questions 17 - 21, assume balls will be drawn at random (one at a time with replacement) out of a basket that contains 13 balls (7 are green, 4 are yellow, and 2 are red).

17. Are the consecutive draws out of the basket independent of each other? Circle one answer.

- a) no
- b) Not enough information has been given to answer this question.
- ☒ c) yes

18. Suppose you draw balls out of the basket until you see a yellow ball for the first time. Let A = the number of draws required. What kind of a random variable is A ? Circle one answer.

- ☒ a) Geometric
- b) Binomial
- c) Poisson
- d) none of the above

19. Suppose you draw balls out of the basket until you see a green ball for the third time. Let B = the number of draws required. What are the possible values of B ? Circle one answer.

- ☒ a) $\{3, 4, 5, 6, \dots\}$
- b) $\{3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$
- c) $\{3, 4, 5, 6, 7, 8, 9\}$
- d) $\{0, 1, 2, 3\}$

20. Suppose you draw 18 balls out of the basket. Let C = the number of green balls that appear. What kind of a random variable is C ? Circle one answer.

- a) Geometric
- ☒ b) Binomial
- c) Poisson
- d) none of the above

Binomial ($n=18, p=\frac{7}{13}$)

21. Suppose you draw 3 balls out of the basket. What is the probability that a yellow ball appears at least once in the 3 draws?

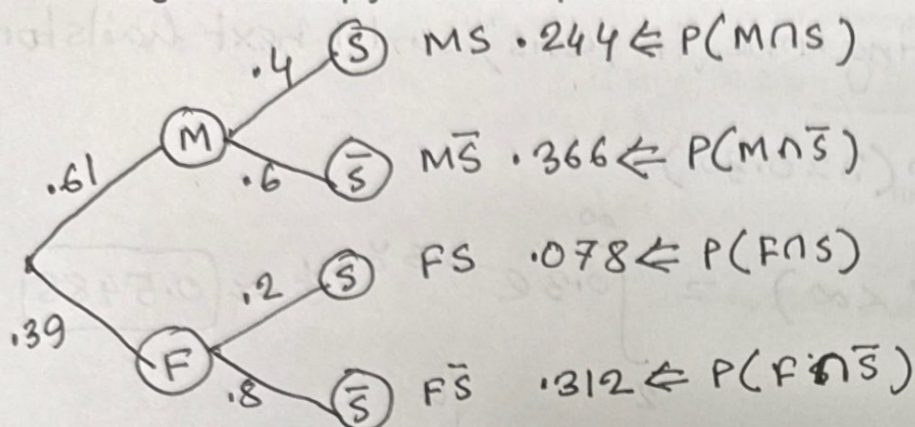
$$P(\text{at least one yellows}) = 1 - P(\text{no yellow})$$

$$= 1 - \left(\frac{9}{13}\right)^3$$

$$\approx \boxed{.6682}$$

Total: 13
yellow: 4
Not yellow: 9

Problem 07: All students at Winterland College fill out a survey that asks questions about the vehicles they drive. First of all, if you select a student at random, the probability they are male is .61 and the probability they are female is .39. If you randomly select a male student, the probability he drives a vehicle with snow tires is .40. If you randomly select a female student, the probability she drives a vehicle with snow tires is .20. A tree diagram will help you answer questions 22-23.



22. What is the probability a randomly selected student drives a vehicle with snow tires?

$$\begin{aligned}
 P(S) &= P(\overset{M \cap S}{MS}) + P(\overset{F \cap S}{FS}) \\
 &= .244 + .078 = \boxed{.322}
 \end{aligned}$$

23. Given that a randomly selected student drives a vehicle without snow tires, what is the probability they are female?

$$\begin{aligned}
 P(F|\bar{S}) &= \frac{P(F \cap \bar{S})}{P(\bar{S})} = \frac{.312}{.678} = \boxed{.460} & \left| \begin{aligned} P(\bar{S}) &= 1 - P(S) \\ &= 1 - .322 \\ &= .678 \end{aligned} \right.
 \end{aligned}$$

Problem 08: The waiting time (in years) until the next severe hailstorm hits Kirksville, Missouri is an exponential random variable with $\lambda = 0.30$.
Answer question 24.

24. What is the probability that Kirksville, Missouri will not have any severe hailstorms for at least the next 2 years? $\Leftrightarrow$ (next hailstorm after 2 years)

Y : waiting time (in years) until next hailstorm

$$Y \sim \text{Exp}(\lambda = 0.30)$$

$$P(2 < Y < \infty) = \int_2^{\infty} 0.3 e^{-0.3y} dy \approx \boxed{0.5488}$$

$\downarrow$
0

Section 3.5 The Normal Dist

	Score	μ	σ	Z-score
You:	100 200	80	20	$z = \frac{100-80}{20} = 1$
Friend:	75 100	80	10	$z = \frac{75-80}{10} = -0.5$

▣ "standardizing" an observation (finding its z-score) is measuring how many standard deviation above/below the mean it is.

→ Greater than mean: +ve z-score

→ Less than mean: -ve z-score.

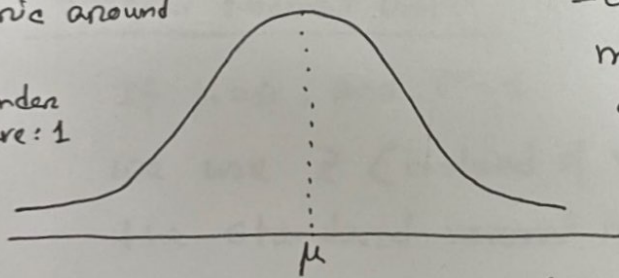
▣ Normal distribution

- The most popular & important continuous dist.

Bell curve:

① symmetric around mean.

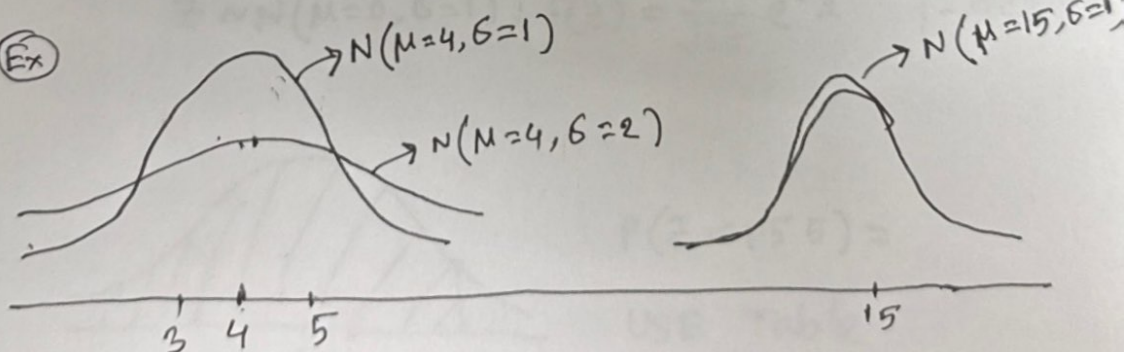
② Area under the curve: 1



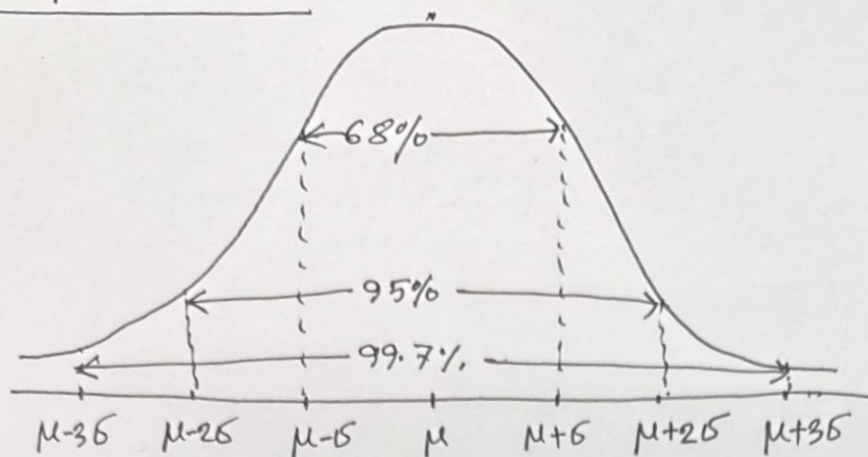
- centered around the mean μ & its width depends on σ .

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}; \quad -\infty < y < \infty$$

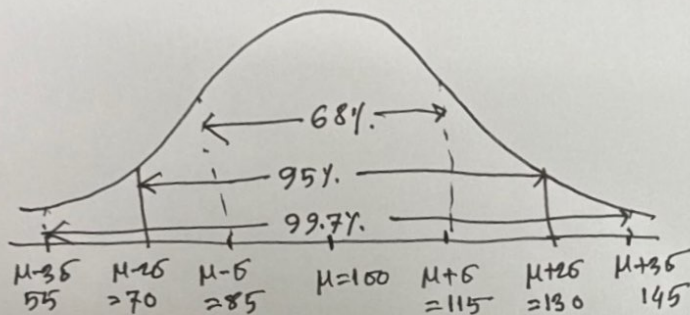
(Ex)



Empirical Rule



Example: IQ test: $\mu = 100$ $\sigma = 15$



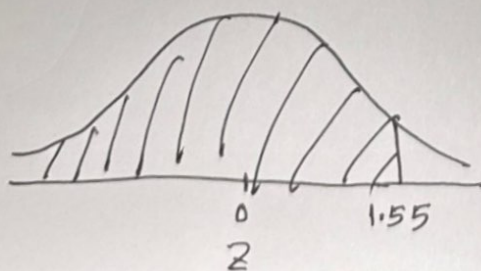
Special Normal Dist

If $\mu = 0$, and $\sigma = 1$

we use Z (instead of Y) & call it the standard normal r.v.

$$Y \sim N(\mu, \sigma) \quad f(Y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(Y-\mu)^2}{\sigma^2}}; -\infty < Y < \infty$$

$$Z \sim N(\mu=0, \sigma=1); f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} Z^2}; -\infty < Z < \infty$$



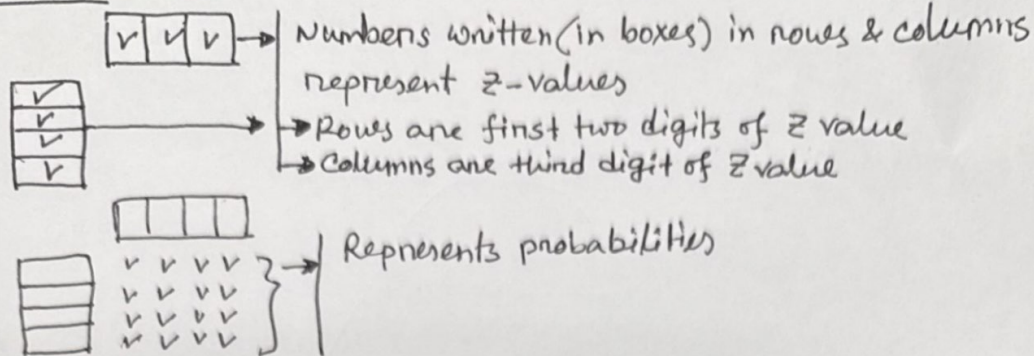
$$P(Z < 1.55) =$$

USE Table

Z table:

Please write these down

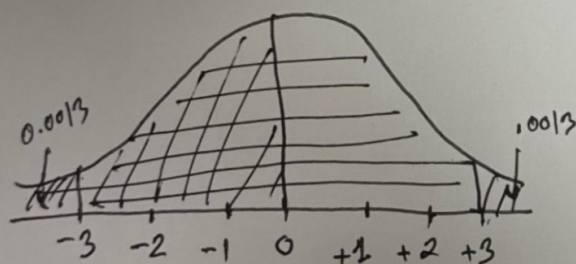
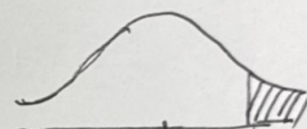
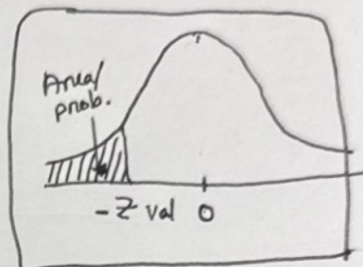
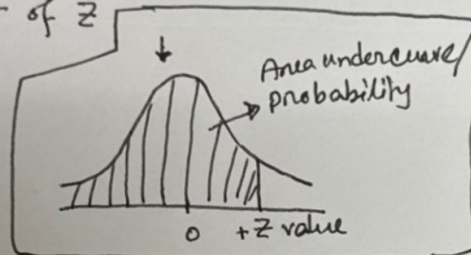
02/14/25



Remember: (i) Area under the curve = 1

(ii) Curve is symmetric around zero

(iii) Table gives you area to the LEFT of Z



$$\mu = 0$$

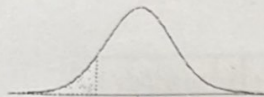
$$\sigma = 1$$

Go to handout ch 3.5 Ex. 1

md mutasi Billah

Z-Table

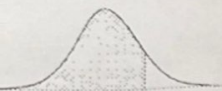
negative Z values



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

-Table

stive Z values



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

[If n is odd, median has a dual citizenship]

Chapter 3.5 (The Normal Distribution): Handout

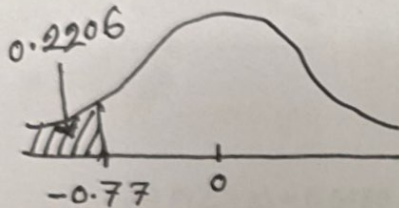
Md Mutasim Billah

1st scenario: Z-value is given; find probability

Example 1:

If $Z \sim N(\mu=0, \sigma=1)$, then find the following probabilities:

(a) $P(Z < -0.77)$

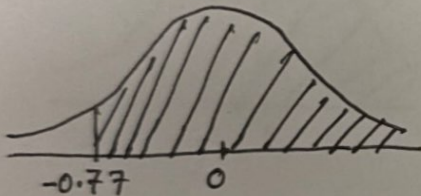


0.07 column
↓

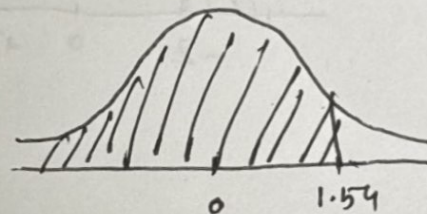
-0.7 row
→ 0.2206

(b) $P(Z > -0.77)$

$$1 - 0.2206 = 0.7794$$



$$P(Z < 1.54) = 0.9382$$

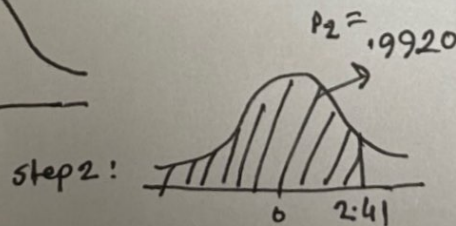
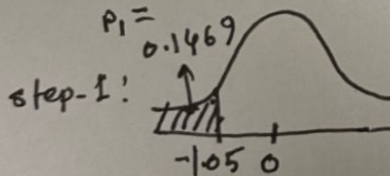
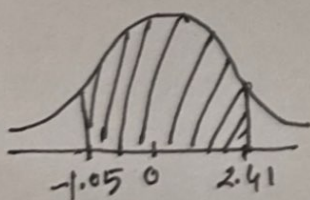


$$P(Z < 1.548) \approx 0.9394$$

$$P(Z < 1.541) \approx 0.9382$$

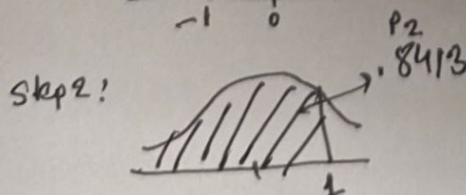
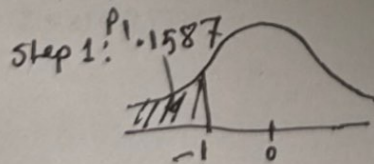
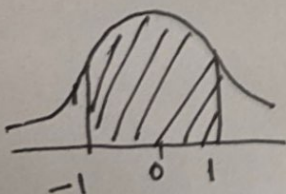
$$P(Z < 1.545) = \frac{0.9394 + 0.9382}{2}$$

(c) $P(-1.05 < Z < 2.41)$

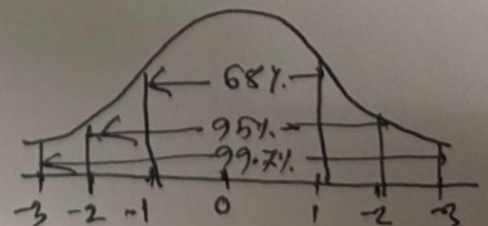


$$P = P_2 - P_1 = 0.9920 - 0.1469 = 0.8451$$

(d) $P(-1 < Z < 1)$; $P(-2 < Z < 2)$; $P(-3 < Z < 3)$



$$P = P_2 - P_1 = 0.8413 - 0.1587 = 0.6826$$

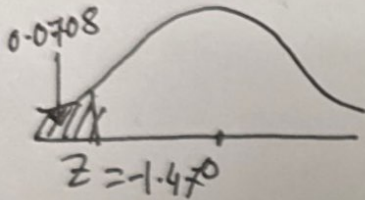


2nd scenario: probability (area under the curve) given,
Find z -value.

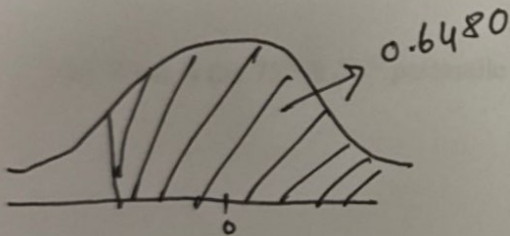
Example 2:

Suppose, $Z \sim N(\mu=0, \sigma=1)$, then find the value of Z if:

(a) $P(Z < z) = 0.0708$



(b) $P(Z > z) = 0.6480$



(c) $P(-z < Z < z) = 0.95$

(d) $P(-1.21 < Z < z) = 0.7330$

02/17/25

▣ Please write these on your Notes:

(*) If $Y \sim N(\mu, \sigma)$, then the Z-score:

$$Z = \frac{Y - \mu}{\sigma}$$

Recall: $\mu = \text{Population Mean}$
 $= E(Y)$

$\text{Var}(Y) = \sigma^2 = \text{population variance}$

$\text{SD}(Y) = \sigma = \sqrt{\sigma^2} = \text{pop. standard deviation (SD)}$

So, $Y \sim N(\mu = E(Y), \sigma = \text{SD}(Y))$

3.6

Random Behavior of Means

→ Instead of the r.v. Y , we are interested in sample mean $\bar{Y}$.

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$$

$\bar{Y}$ is a random variable.

Chapter 3.5 (The Normal Distribution): Handout

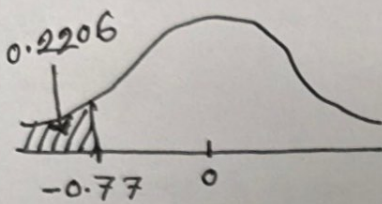
Md Mutasim Billah

1st Scenario: Z-value is given; find probability

Example 1:

If $Z \sim N(\mu=0, \sigma=1)$, then find the following probabilities:

(a) $P(Z < -0.77)$

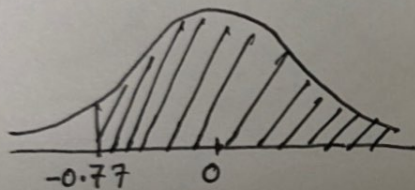


0.07 column
↓

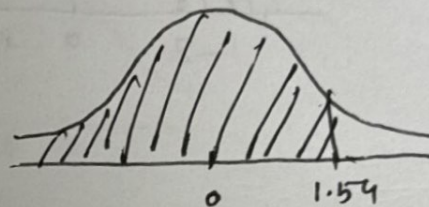
-0.7 row
→ 0.2206

(b) $P(Z > -0.77)$

$$1 - 0.2206 = 0.7794$$



$$P(Z < 1.54) = 0.9382$$

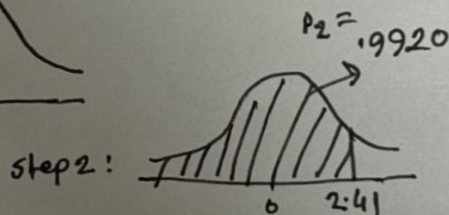
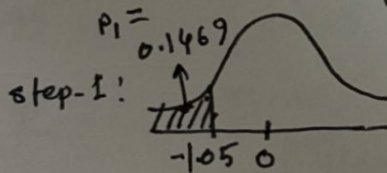
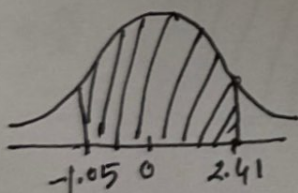


$$P(Z < 1.548) \approx 0.9394$$

$$P(Z < 1.541) \approx 0.9382$$

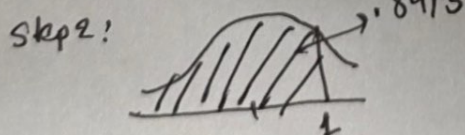
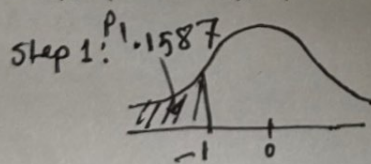
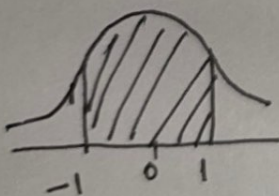
$$P(Z < 1.545) = \frac{0.9394 + 0.9382}{2}$$

(c) $P(-1.05 < Z < 2.41)$

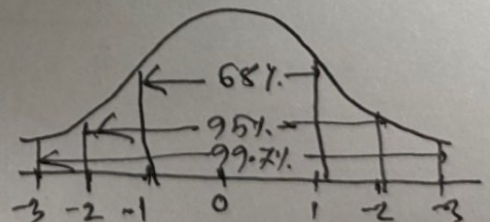


$$P = P_2 - P_1 = 0.9920 - 0.1469 = 0.8451$$

(d) $P(-1 < Z < 1); P(-2 < Z < 2); P(-3 < Z < 3)$



$$P = P_2 - P_1 = 0.8413 - 0.1587 = 0.6826$$

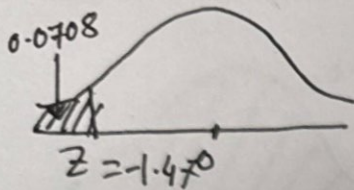


2nd scenario: probability (area under the curve) given,
Find z -value.

Example 2:

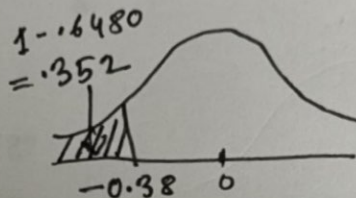
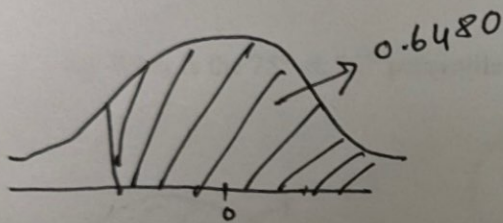
Suppose, $Z \sim N(\mu=0, \sigma=1)$, then find the value of Z if:

(a) $P(Z < z) = 0.0708$



① table gives prob. from left to right

(b) $P(Z > z) = 0.6480$

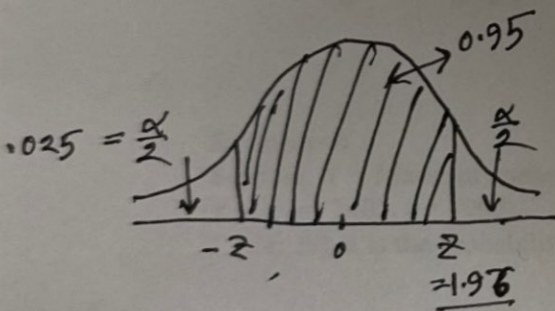
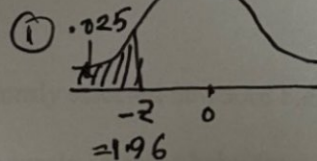


②

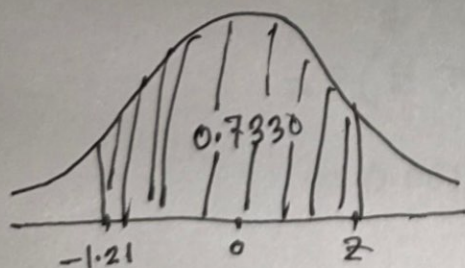
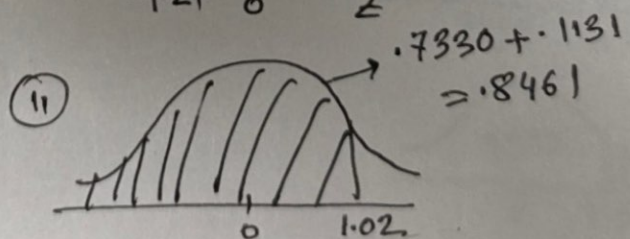
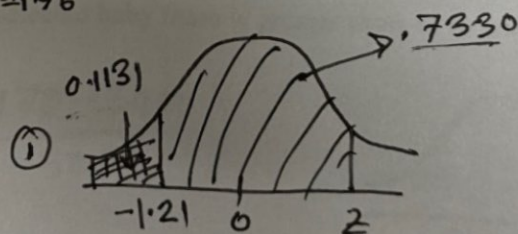
(c) $P(-z < Z < z) = 0.95$

Let, $\alpha = 1 - 0.95 = 0.05$

$\frac{\alpha}{2} = 0.025$



(d) $P(-1.21 < Z < z) = 0.7330$



Recall: 90th percentile is greater than or equal to 90% of data

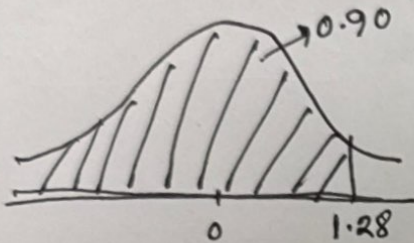
Ex. your score was greater than 90% of all the scores among all the people who took the exam!

Example 3:

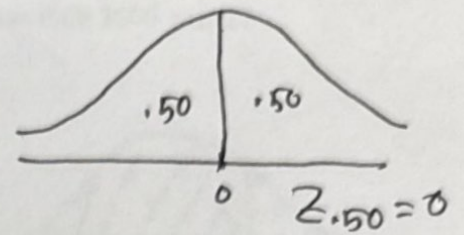
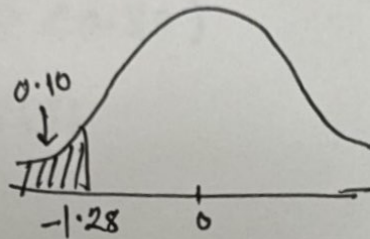
Suppose, $Z \sim N(\mu=0, \sigma=1)$. Answer the following questions:

(a) What is the 90th & 10th percentile of Z?

90th percentile:



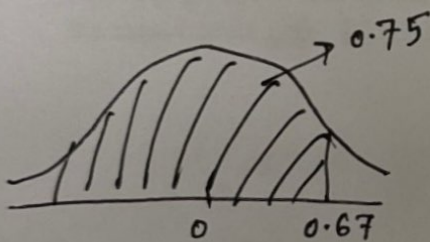
10th percentile



$$Z_{\alpha} = (1-\alpha) \text{ 100th percentile}$$

$$Z_{.10} = (1-.10) * 100\text{th percentile} \\ = 90\text{th percentile}$$

(b) What is the 75th & 25th percentile of Z?



$$\cancel{Z_{.25}} = 75\text{th percentile} = 0.67$$

$$25\text{th percentile} = -0.67$$

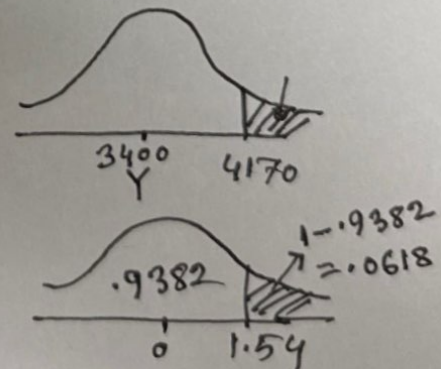
Example 4:

Suppose, $Y = \text{Mass (in grams) of a randomly selected newborn baby}$, where, $Y \sim N(\mu=3400, \sigma=500)$.

$$Z = \frac{Y - \mu}{\sigma}$$

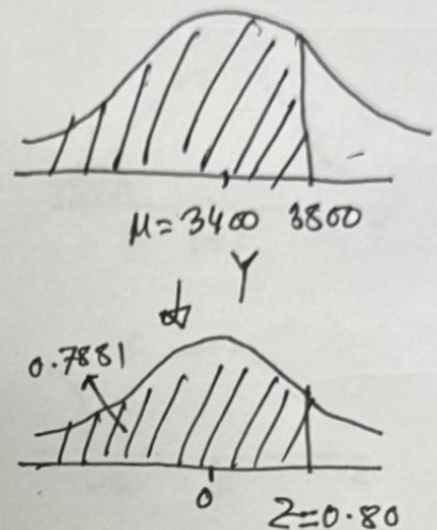
(a) What is the probability that a randomly selected baby mass is greater than 4170 grams?

$$P(Y > 4170) = P\left(\frac{Y - \mu}{\sigma} > \frac{4170 - 3400}{500}\right) \\ = P(Z > 1.54) \\ = 0.0618$$



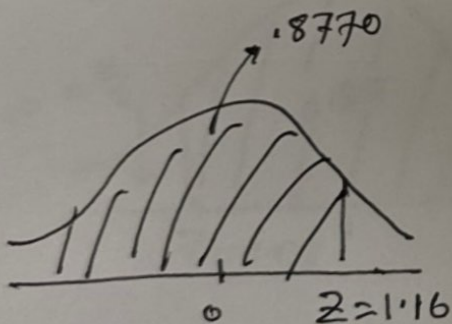
(b) What is the probability that a randomly selected baby mass is less than 3800 grams?

$$\begin{aligned}
 P(Y < 3800) &= P\left(\frac{Y - \mu}{\sigma} < \frac{3800 - 3400}{500}\right) \\
 &= P(Z < 0.80) \\
 &= 0.7881
 \end{aligned}$$



Example 5: Fill in the blanks! ☺

87.7% of all babies are lighter than Mary. Mary's mass is the 87.7th percentile and the mass equals 3980 grams.



$$Z = \frac{Y - \mu}{\sigma}$$

$$\Rightarrow 1.16 = \frac{Y - 3400}{500}$$

$$\Rightarrow Y = 3980$$

Example 6:

Suppose, $Y = \text{Gallons of milk placed in a plastic jug}$, where, $Y \sim N(\mu=1.02, \sigma=.01)$.

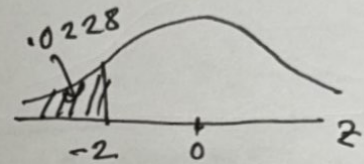
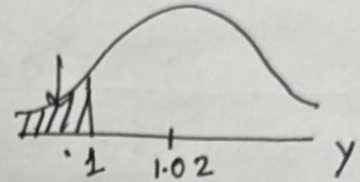
(a) Find

$P(\text{a randomly selected jug will be under filled}) =$

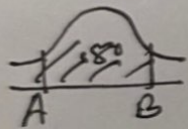
$$P(Y < 1) = P\left(\frac{Y - \mu}{\sigma} < \frac{1 - 1.02}{.01}\right)$$

$$= P(Z < -2)$$

$$= .0228$$



(b) 80% of all milk jugs contain a volume of milk in an interval (A, B) centered around $\mu = 1.02$. What is (A, B) ?



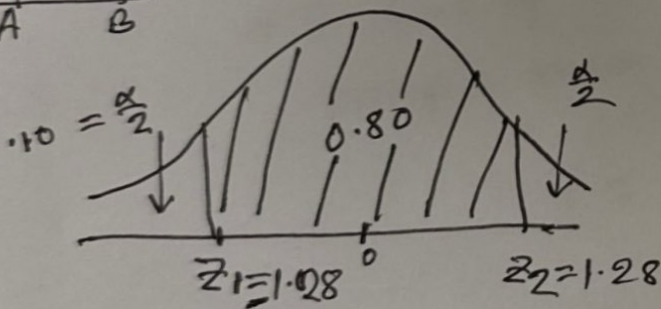
$$\alpha = 1 - .80 = .20$$

$$\frac{\alpha}{2} = .10$$

$$Z = \frac{Y - \mu}{\sigma}$$

$$-1.28 = \frac{A - 1.02}{.01} \Rightarrow A = 1.0072$$

$$1.28 = \frac{B - 1.02}{.01} \Rightarrow B = 1.0328$$



① Please write the followings in your notes: 02/19/25

①

* If $Y \sim N(\mu = E(Y), \sigma = SD(Y))$ then $Z = \frac{Y - E(Y)}{SD(Y)} = \frac{Y - \mu}{\sigma}$

	Pop. mean	Pop var.	Pop. SD
About 'population' ← Parameter	μ	σ^2	σ
	sample mean	sample var.	Sample SD
About 'Sample' ← Statistic	$\bar{y}$	s^2	s

pop. $\Rightarrow$ population

var. $\Rightarrow$ Variance

SD $\Rightarrow$ standard deviation

[HWCH302 Discussion Due
03/05/25 at 11:59 pm
on canvas]

$\rightarrow E(\text{statistic}) = \text{parameter}$

then that statistic is an unbiased estimator of the parameter.

$$E(\bar{y}) \stackrel{?}{=} \mu$$

Let, $y_1, y_2, \dots, y_n$ are i.i.d (identically & independently distributed r.v.)

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{y_1 + y_2 + \dots + y_n}{n}$$

$\bar{y}$ is a random variable

$$\mu_{\bar{y}} = E(\bar{y}) = E\left[\frac{\sum_{i=1}^n y_i}{n}\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^n y_i\right]$$

$$= \frac{1}{n} \left(\sum_{i=1}^n E(y_i)\right)$$

$$= \frac{1}{n} \times n \mu$$

$$E(\bar{y}) = \mu$$

$\bar{y}$ is an unbiased estimate of μ .

$$\bullet E(aX) = a \cdot E(X)$$

↑
Constant

$$\bullet E\left(\sum_{i=1}^n y_i\right) = \sum_{i=1}^n E(y_i)$$

$$\bullet E(y_1) = \mu; E(y_2) = \mu$$

$$\dots E(y_n) = \mu$$

(2)

$$\begin{aligned}
 \sigma_{\bar{y}}^2 &= \text{var}(\bar{y}) = \text{var}\left(\frac{\sum_{i=1}^n y_i}{n}\right) \\
 &= \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n y_i\right) \\
 &= \frac{1}{n^2} \underbrace{\sum_{i=1}^n \text{var}(y_i)}_{n\sigma^2} \\
 &= \frac{1}{n^2} * n\sigma^2 = \frac{\sigma^2}{n}
 \end{aligned}$$

constant

$$\begin{aligned}
 &\bullet \text{Var}(aX) = a^2 \text{Var}(X) \\
 &\bullet \text{Var}\left(\sum_{i=1}^n y_i\right) = \sum_{i=1}^n \text{Var}(y_i) \\
 &\text{Var}(y_1) = \sigma^2, \dots \\
 &\text{Var}(y_n) = \sigma^2
 \end{aligned}$$

$$SD(\bar{y}) = \sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

What kind of r.v. is $\bar{y}$?

Central Limit Theorem

- ① If the population is normal [$Y \sim N(\mu = E(Y), \sigma = SD(Y))$]
then $\bar{y}$ is also normally distributed.
- ② Even if the population is NOT normal, $\bar{y}$ is
still approximately normally distributed if
the sample size is large enough ($n \geq 30$)

$$Y \sim N(\mu, \sigma) ; Z = \frac{Y - \mu}{\sigma}$$

$$\begin{aligned}
 &\bar{y} \sim N(E(\bar{y}) = \mu, SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}) \\
 &\rightarrow \boxed{Z = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}}} \quad \left[Z = \frac{\bar{y} - E(\bar{y})}{SD(\bar{y})} \right]
 \end{aligned}$$

Go to CH 3.6 Handout Ex. ①

Chapter 3.6 (Random Behavior of Means): Handout
Md Mutasim Billah

Example 1:

Suppose, Y = Gallons of milk placed in a plastic jug, where, $Y \sim N(\mu=1.02, \sigma=.01)$.

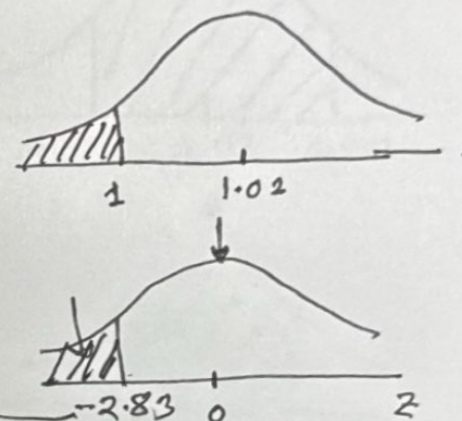
(a) If we randomly select $n=2$ jugs and compute the mean $\bar{Y}$, then find $P(\bar{Y} < 1)$.

Before: $P(Y < 1) = 0.0228$; $z = \frac{Y - \mu}{\sigma}$
 $n=1$

Now: $P(\bar{Y} < 1) = P\left(\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} < \frac{1 - 1.02}{.01/\sqrt{2}}\right)$ $\left[z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}\right]$

$= P(z < -2.83)$

$= .0023$



(b) What happens to $P(\bar{Y} < 1)$ as sample size increases?

n	z	$P(\bar{Y} < 1)$
1	-2	.0228
2	-2.83	.0023
3	-3.46	$\approx \frac{1}{3759}$
4	-4	$\approx \frac{1}{31574}$
5	-4.47	$\approx \frac{1}{258257} !$

for larger n

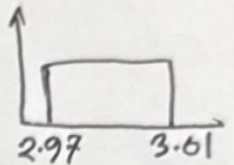
$\bar{Y} \approx \mu$

→ what is the prob that the average weight of the car part will lie between 2.987 & 2.991

Example 2:

Suppose, $Y = \text{weight (grams)}$ of a randomly selected car part, where, $Y \sim \text{UNIF}(2.97, 3.01)$. a, b

(a) If you select $n=30$ car parts at random & compute $\bar{Y}$, then find $P(2.987 < \bar{Y} < 2.991)$



$$\mu = \frac{1}{2}(a+b) = \frac{1}{2}(2.97 + 3.01) = 2.99$$

$$\sigma^2 = \frac{1}{12}(b-a)^2 = \frac{(3.01-2.97)^2}{12} = \frac{1}{7500}$$

$$\sigma = \frac{1}{\sqrt{7500}}$$

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

By CLT,

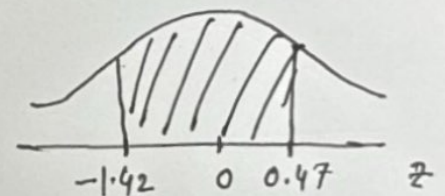
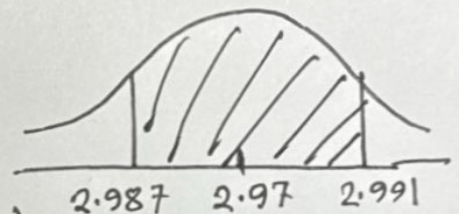
$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

$$\bar{Y} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$P(2.987 < \bar{Y} < 2.991)$$

$$= P\left(\frac{2.987 - 2.99}{\frac{1}{\sqrt{7500}}/\sqrt{30}} < \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{2.991 - 2.99}{\frac{1}{\sqrt{7500}}/\sqrt{30}}\right)$$

$$= P(-1.42 < Z < 0.47)$$



(b) If you select $n=50$ car parts at random & compute $\bar{Y}$, then find $P(2.987 < \bar{Y} < 2.991)$

Try at home.

$$P(-1.84 < Z < 0.61) = 0.6942.$$

① Please write the followings in your notes: 02/19/25

①

* If $Y \sim N(\mu = E(Y), \sigma = SD(Y))$ then $Z = \frac{Y - E(Y)}{SD(Y)} = \frac{Y - \mu}{\sigma}$

	Pop. mean	Pop var.	Pop. SD
About 'population' ← Parameter	μ	σ^2	σ
	sample mean	sample var.	Sample SD
About 'sample' ← Statistic	$\bar{y}$	s^2	s

pop. $\Rightarrow$ population

var. $\Rightarrow$ Variance

SD $\Rightarrow$ standard deviation

$\rightarrow E(\text{statistic}) = \text{parameter}$

then that statistic is an unbiased estimator of the parameter.

$$E(\bar{y}) \stackrel{?}{=} \mu$$

Let, $y_1, y_2, \dots, y_n$ are i.i.d (identically & independently distributed r.v.)

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{y_1 + y_2 + \dots + y_n}{n}$$

$\bar{y}$ is a random variable

$$\mu_{\bar{y}} = E(\bar{y}) = E\left[\frac{\sum_{i=1}^n y_i}{n}\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^n y_i\right]$$

$$= \frac{1}{n} \left(\sum_{i=1}^n E(y_i)\right)$$

$$= \frac{1}{n} \times n \mu$$

$$E(\bar{y}) = \mu$$

$\bar{y}$ is an unbiased estimate of μ .

$$\bullet E(aX) = a \cdot E(X)$$

↑
Constant

$$\bullet E\left(\sum_{i=1}^n y_i\right) = \sum_{i=1}^n E(y_i)$$

$$\bullet E(y_1) = \mu; E(y_2) = \mu$$

$$\dots E(y_n) = \mu$$

(2)

$$\begin{aligned}
 \sigma_{\bar{y}}^2 &= \text{var}(\bar{y}) = \text{var}\left(\frac{\sum_{i=1}^n y_i}{n}\right) \\
 &= \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n y_i\right) \\
 &= \frac{1}{n^2} \underbrace{\sum_{i=1}^n \text{var}(y_i)}_{n\sigma^2} \\
 &= \frac{1}{n^2} * n\sigma^2 = \frac{\sigma^2}{n}
 \end{aligned}$$

constant

$$\begin{aligned}
 &\bullet \text{Var}(aX) = a^2 \text{Var}(X) \\
 &\bullet \text{Var}\left(\sum_{i=1}^n y_i\right) = \sum_{i=1}^n \text{Var}(y_i) \\
 &\text{Var}(y_1) = \sigma^2, \dots \\
 &\text{Var}(y_n) = \sigma^2
 \end{aligned}$$

$$SD(\bar{y}) = \sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

What kind of r.v. is $\bar{y}$?

☞ Central Limit Theorem

- ① If the population is normal [$Y \sim N(\mu = E(Y), \sigma = SD(Y))$]
then $\bar{y}$ is also normally distributed.
- ② Even if the population is NOT normal, $\bar{y}$ is
still approximately normally distributed if
the sample size is large enough ($n \geq 30$)

$$Y \sim N(\mu, \sigma) ; Z = \frac{Y - \mu}{\sigma}$$

$$\begin{aligned}
 &\bar{y} \sim N(E(\bar{y}) = \mu, SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}) \\
 &\rightarrow \boxed{Z = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}}} \quad \left[Z = \frac{\bar{y} - E(\bar{y})}{SD(\bar{y})} \right]
 \end{aligned}$$

Go to CH 3.6 Handout Ex. ①

Chapter 3.6 (Random Behavior of Means): Handout
Md Mutasim Billah

Example 1:

Suppose, Y = Gallons of milk placed in a plastic jug, where, $Y \sim N(\mu=1.02, \sigma=.01)$.

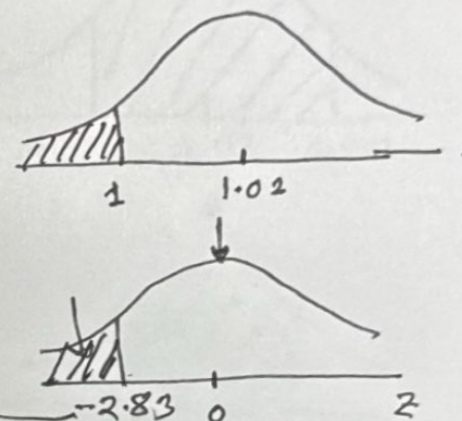
(a) If we randomly select $n=2$ jugs and compute the mean $\bar{Y}$, then find $P(\bar{Y} < 1)$.

Before: $P(Y < 1) = 0.0228$; $z = \frac{Y - \mu}{\sigma}$
 $n=1$

Now: $P(\bar{Y} < 1) = P\left(\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} < \frac{1 - 1.02}{.01/\sqrt{2}}\right)$ $\left[z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}\right]$

$= P(z < -2.83)$

$= .0023$



(b) What happens to $P(\bar{Y} < 1)$ as sample size increases?

n	z	$P(\bar{Y} < 1)$
1	-2	.0228
2	-2.83	.0023
3	-3.46	$\approx \frac{1}{3759}$
4	-4	$\approx \frac{1}{31574}$
5	-4.47	$\approx \frac{1}{258257}!$

for larger n

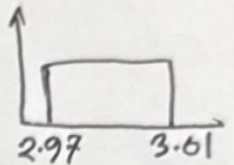
$\bar{Y} \approx \mu$

→ what is the prob that the average weight of the car part will lie between 2.987 & 2.991

Example 2:

Suppose, $Y = \text{weight (grams)}$ of a randomly selected car part, where, $Y \sim \text{UNIF}(2.97, 3.01)$.

(a) If you select $n=30$ car parts at random & compute $\bar{Y}$, then find $P(2.987 < \bar{Y} < 2.991)$



$$\mu = \frac{1}{2}(a+b) = \frac{1}{2}(2.97 + 3.01) = 2.99$$

$$\sigma^2 = \frac{1}{12}(b-a)^2 = \frac{(3.01-2.97)^2}{12} = \frac{1}{7500}$$

$$\sigma = \frac{1}{\sqrt{7500}}$$

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

By CLT,

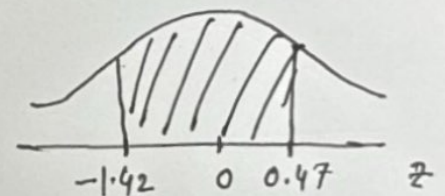
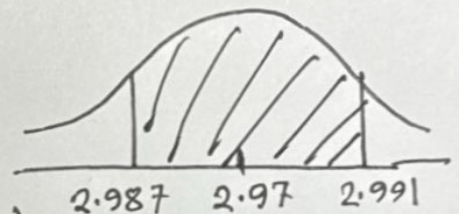
$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

$$\bar{Y} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$P(2.987 < \bar{Y} < 2.991)$$

$$= P\left(\frac{2.987 - 2.99}{\frac{1}{\sqrt{7500}}/\sqrt{30}} < \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{2.991 - 2.99}{\frac{1}{\sqrt{7500}}/\sqrt{30}}\right)$$

$$= P(-1.42 < Z < 0.47)$$



(b) If you select $n=50$ car parts at random & compute $\bar{Y}$, then find $P(2.987 < \bar{Y} < 2.991)$

Try at home.

$$P(-1.84 < Z < 0.61) = 0.6942.$$

Section 3.7
② Please write down the followings in your class notes

Recall: population variance, $\sigma^2 = \frac{\sum (Y - \mu)^2}{n}$

, where, μ = pop. mean.

Sample variance, $s^2 = \frac{\sum (Y - \bar{Y})^2}{(n-1)}$ [Regular formula]

$$= \frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{(n-1)} \text{ [Convenient]}$$

$$= \frac{n \sum Y^2 - (\sum Y)^2}{n(n-1)} \text{ [Book]}$$

sample standard deviation, $s = \sqrt{s^2}$

Goto Handout ch 3.7 Ex.3

By CLT (Central Limit th^m)

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

If σ is unknown; it must be estimated with s .

$$T = \frac{\bar{Y} - \mu}{s/\sqrt{n}}$$

Properties of T-curve & T-table

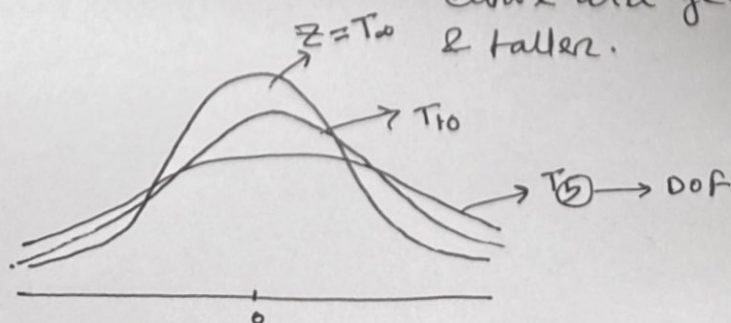
✓ (i) symmetric around zero

✓ (ii) Total area under the curve = 1

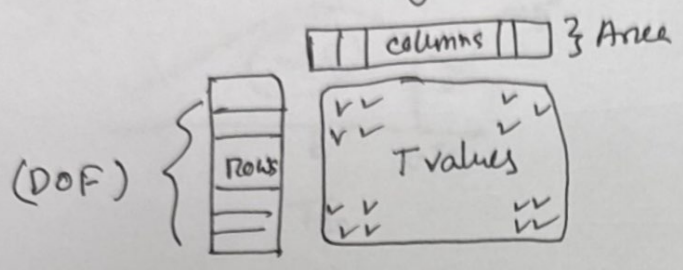
(iii) shorter & wider than Z

(iv) Degrees of Freedom: always positive integer;
(DOF)

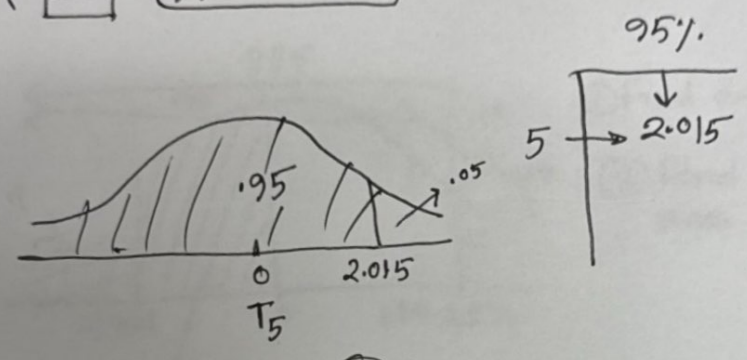
as DOF increases, T curve will get thinner & taller.



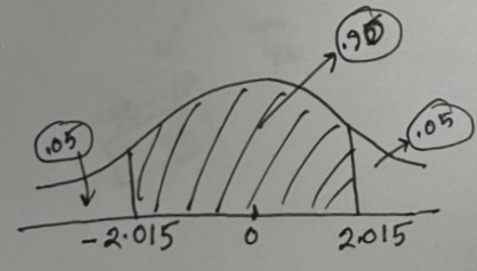
T-table gives us a t-value with an area to the with degrees of freedom (DOF)



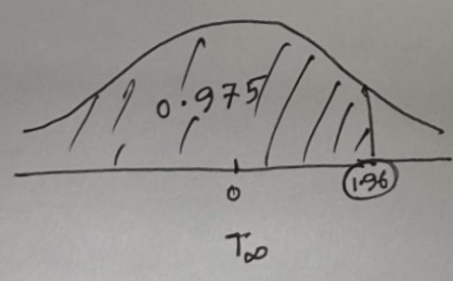
Ex. 1



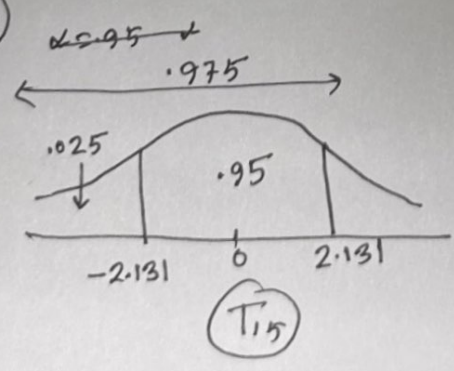
Ex. 2



Ex. 3



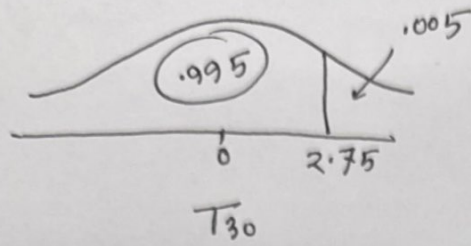
Ex. 4



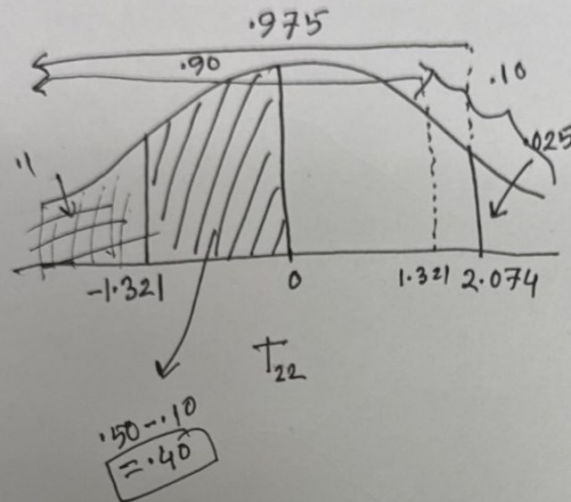
Find DOF

Ex. 5

Find the dof.



Ex. 6



① Find dof

② Find shaded area

Chapter 3.7 (Random Behavior of means when the variance is unknown)

Example 3: Suppose, we have collected a random sample of $n=5$ observations as follows:

{3,7,10,11,14}. Find the sample variance and sample standard deviation.

$$\bar{y} = \frac{\sum y}{n}$$

Y	Y - $\bar{Y}$	(Y - $\bar{Y}$) ²	Y ²
3	-6	36	9
7	-2	4	49
10	1	1	100
11	2	4	121
14	5	25	196
Total 45	0	70	475
$\sum y$		$\sum (y - \bar{y})^2$	$\sum y^2$

Regular:

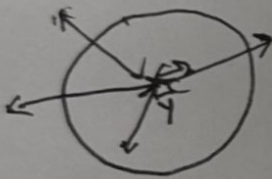
$$s^2 = \frac{\sum (y - \bar{y})^2}{n-1} = \frac{70}{5-1} = 17.5$$

Convenient:

$$s^2 = \frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n-1} = \frac{475 - \frac{45^2}{5}}{5-1} = 17.5$$

Book:

$$s^2 = \frac{n \sum y^2 - (\sum y)^2}{n(n-1)} = \frac{5(475) - 45^2}{5(5-1)} = 17.5$$



$$\text{Standard deviation, } S = \sqrt{s^2} = \sqrt{17.5} = 4.18$$

Example 4: Without calculation, answer the following questions:

(a) {10,10,10,12,12,12}; Find $\bar{Y}$ & s.

$$\bar{Y} = 11, s \approx 1 (1.095)$$

(b) {0,0,0,0,100,100,100,100}; Find $\bar{Y}$ & s.

$$\bar{Y} = 50, s \approx 50 (53.45)$$

(c) Given, $\{3, 7, 10, 11, 14\}$; $\bar{Y} = 9$, $s^2 = 17.5$, $s = 4.18$. Then, for $\{13, 17, 20, 21, 24\}$, find $\bar{Y}$, s^2 , & s .

$$\bar{Y} = 19 \quad ; \quad s^2 = 17.5 \quad s = 4.18$$

Rule: If you add 'k' to a data set:

$$\bar{Y}_{\text{new}} = \bar{Y}_{\text{old}} + k \quad ; \quad S_{\text{new}}^2 = S_{\text{old}}^2 \quad ; \quad S_{\text{new}} = S_{\text{old}}$$

(d) Given, $\{3, 7, 10, 11, 14\}$; $\bar{Y} = 9$, $s^2 = 17.5$, $s = 4.18$. Then, for $\{30, 70, 100, 110, 140\}$, find $\bar{Y}$, s^2 , & s .

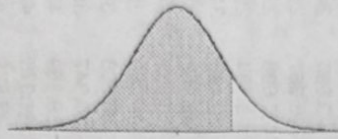
$$\bar{Y} = 90 \quad ; \quad s = 4.18(10) = 41.8 \quad ; \quad s^2 = \frac{17.5}{10^2} (10^2) = 1750$$

Rule: If you multiply a data set by k

$$\bar{Y}_{\text{new}} = k * \bar{Y}_{\text{old}} \quad ; \quad S_{\text{new}} = |k| S_{\text{old}}$$

$$S_{\text{new}}^2 = k^2 (S_{\text{old}}^2)$$

T-Table



	90%	95%	97.50%	99%	99.50%	99.90%
1	3.078	6.314	12.706	31.821	63.656	318.289
2	1.886	2.920	4.303	6.965	9.925	22.328
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.894
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
80	1.292	1.664	1.990	2.374	2.639	3.195
100	1.290	1.660	1.984	2.364	2.626	3.174
200	1.286	1.653	1.972	2.345	2.601	3.131
∞	1.282	1.645	1.960	2.326	2.576	3.090

Section 3.7
② Please write down the followings in your class notes

Recall: population variance, $\sigma^2 = \frac{\sum (Y - \mu)^2}{n}$

, where, μ = pop. mean.

Sample variance, $s^2 = \frac{\sum (Y - \bar{Y})^2}{(n-1)}$ [Regular formula]

$$= \frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{(n-1)} \text{ [Convenient]}$$

$$= \frac{n \sum Y^2 - (\sum Y)^2}{n(n-1)} \text{ [Book]}$$

sample standard deviation, $s = \sqrt{s^2}$

Goto Handout ch 3.7 Ex.3

By CLT (Central Limit th^m)

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

If σ is unknown; it must be estimated with s .

$$T = \frac{\bar{Y} - \mu}{s/\sqrt{n}}$$

Properties of T-curve & T-table

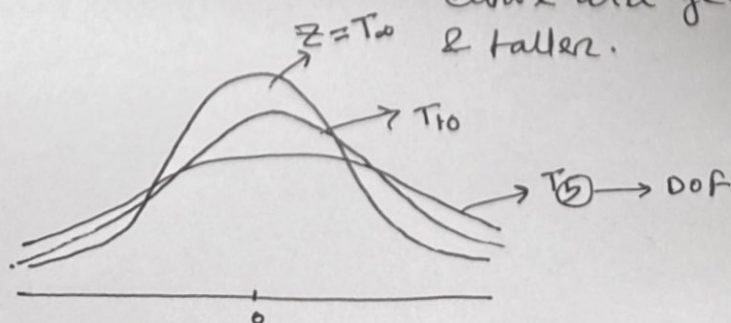
✓ (i) symmetric around zero

✓ (ii) Total area under the curve = 1

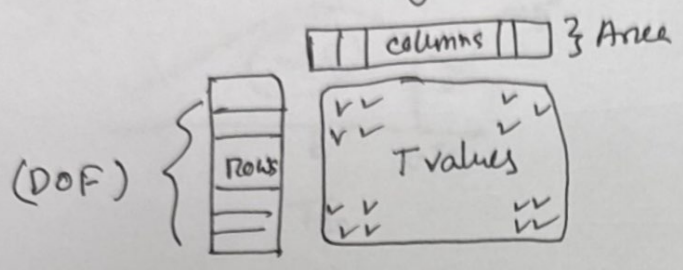
(iii) shorter & wider than Z

(iv) Degrees of Freedom: always positive integer;
(DOF)

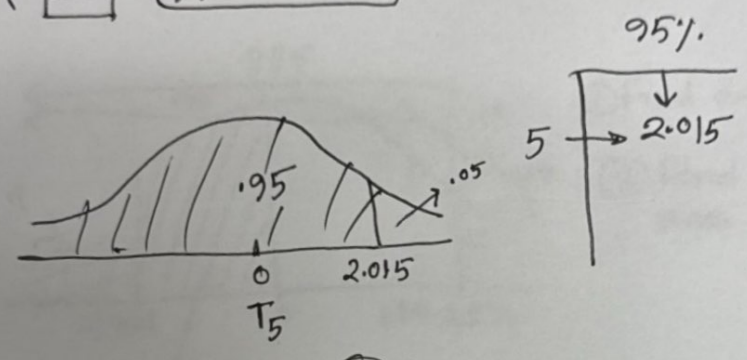
as DOF increases, T curve will get thinner & taller.



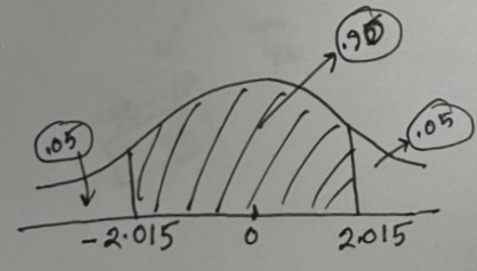
T-table gives us a t-value with an area to the with degrees of freedom (DOF)



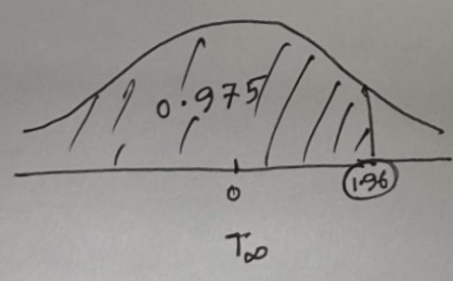
Ex. 1



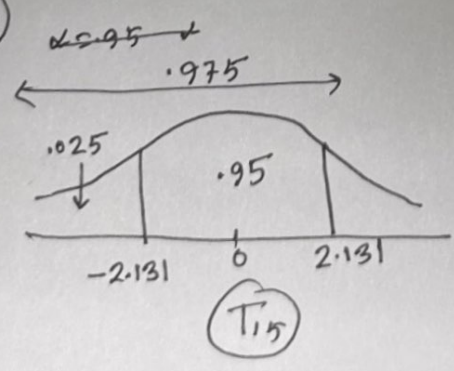
Ex. 2



Ex. 3



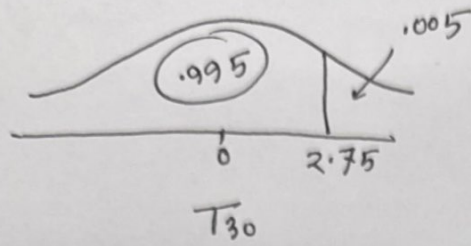
Ex. 4



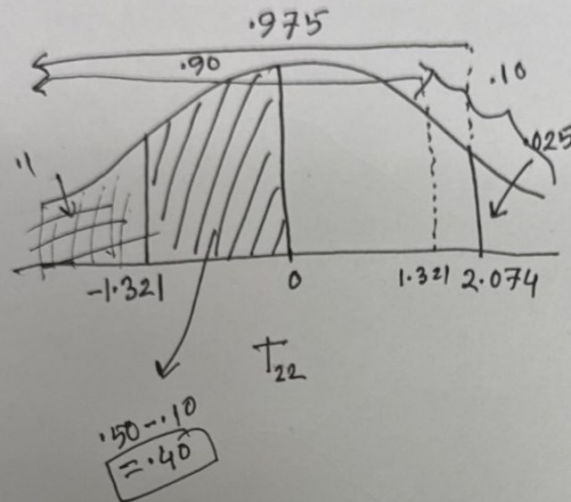
Find DOF

Ex. 5

Find the dof.



Ex. 6



① Find dof

② Find shaded area

Chapter 3.7 (Random Behavior of means when the variance is unknown)

Example 3: Suppose, we have collected a random sample of $n=5$ observations as follows:

{3,7,10,11,14}. Find the sample variance and sample standard deviation.

$$\bar{y} = \frac{\sum y}{n}$$

Y	Y - $\bar{Y}$	(Y - $\bar{Y}$) ²	Y ²
3	-6	36	9
7	-2	4	49
10	1	1	100
11	2	4	121
14	5	25	196
Total 45	0	70	475
$\sum y$		$\sum (y - \bar{y})^2$	$\sum y^2$

Regular:

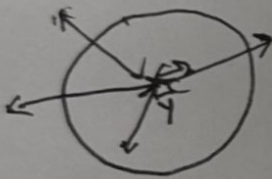
$$s^2 = \frac{\sum (y - \bar{y})^2}{n-1} = \frac{70}{5-1} = 17.5$$

Convenient:

$$s^2 = \frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n-1} = \frac{475 - \frac{45^2}{5}}{5-1} = 17.5$$

Book:

$$s^2 = \frac{n \sum y^2 - (\sum y)^2}{n(n-1)} = \frac{5(475) - 45^2}{5(5-1)} = 17.5$$



$$\text{Standard deviation, } S = \sqrt{s^2} = \sqrt{17.5} = 4.18$$

Example 4: Without calculation, answer the following questions:

(a) {10,10,10,12,12,12}; Find $\bar{Y}$ & s.

$$\bar{Y} = 11, s \approx 1 (1.095)$$

(b) {0,0,0,0,100,100,100,100}; Find $\bar{Y}$ & s.

$$\bar{Y} = 50, s \approx 50 (53.45)$$

(c) Given, $\{3, 7, 10, 11, 14\}$; $\bar{Y} = 9$, $s^2 = 17.5$, $s = 4.18$. Then, for $\{13, 17, 20, 21, 24\}$, find $\bar{Y}$, s^2 , & s .

$$\bar{Y} = 19 \quad ; \quad s^2 = 17.5 \quad s = 4.18$$

Rule: If you add 'k' to a data set:

$$\bar{Y}_{\text{new}} = \bar{Y}_{\text{old}} + k \quad ; \quad s_{\text{new}}^2 = s_{\text{old}}^2 \quad ; \quad s_{\text{new}} = s_{\text{old}}$$

(d) Given, $\{3, 7, 10, 11, 14\}$; $\bar{Y} = 9$, $s^2 = 17.5$, $s = 4.18$. Then, for $\{30, 70, 100, 110, 140\}$, find $\bar{Y}$, s^2 , & s .

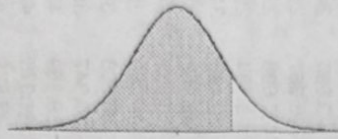
$$\bar{Y} = 90 \quad ; \quad s = 4.18(10) = 41.8 \quad ; \quad s^2 = \frac{17.5}{10^2} (10^2) = 1750$$

Rule: If you multiply a data set by k

$$\bar{Y}_{\text{new}} = k * \bar{Y}_{\text{old}} \quad ; \quad s_{\text{new}} = |k| s_{\text{old}}$$

$$s_{\text{new}}^2 = k^2 (s_{\text{old}}^2)$$

T-Table



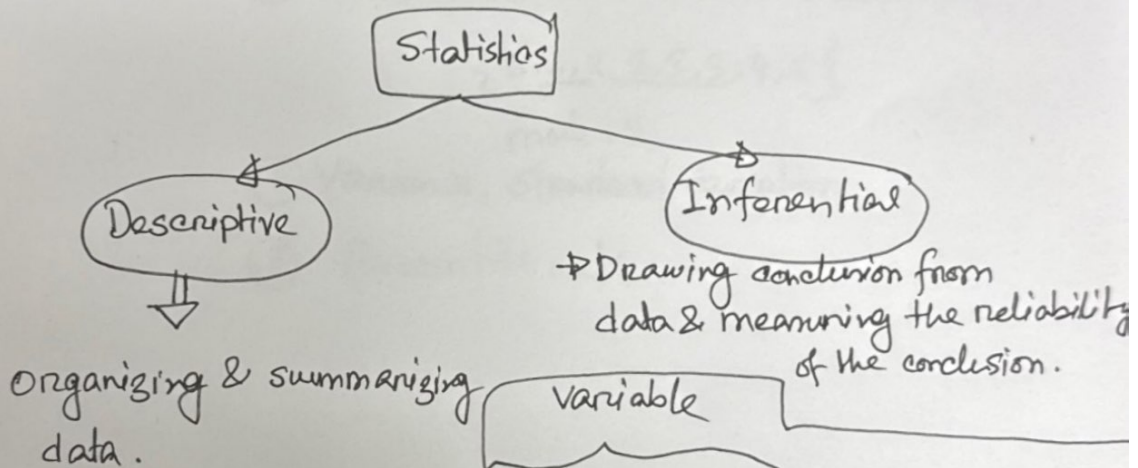
	90%	95%	97.50%	99%	99.50%	99.90%
1	3.078	6.314	12.706	31.821	63.656	318.289
2	1.886	2.920	4.303	6.965	9.925	22.328
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.894
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
80	1.292	1.664	1.990	2.374	2.639	3.195
100	1.290	1.660	1.984	2.364	2.626	3.174
200	1.286	1.653	1.972	2.345	2.601	3.131
∞	1.282	1.645	1.960	2.326	2.576	3.090

Engineering Statistics

Chapter 2

Statistics : $\rightarrow$ Collecting, analyzing, & interpreting data.

$\downarrow$
Information we gather with experiments & surveys



Ex: Investigation of the length of bolts being manufactured in a factory.

Suppose, we measure the length of 100 randomly selected bolts.

$n = 100$
bolt 1 = 2.001 inch
bolt 2 = 1.999 inch
 $\vdots$
bolt 100 = 2.001 inch

more convenient or meaningful:

Average length = Sample mean length

$$= \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{\sum_{i=1}^{100} Y_i}{100}$$

Visualization: Graphs/plots like

Histogram, Dotplot, Stem & Leaf Plot,
Boxplot, etc.

Numerical
values

① Mean (Average)

② Median: $\{1, \textcircled{3}, 4\}$

middle value of
a dataset

$\{1, \textcircled{3}, 4\}$

3.5

③ Mode: value with most occurrence.

$\{1, \underline{2}, \underline{2}, \underline{3}, \underline{3}, \underline{3}, 4, 5\}$

mode: 3

④ Variance, Standard deviation.

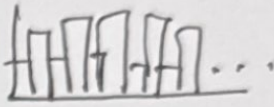
⑤ Percentile, etc.

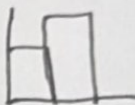
Histogram

→ Useful for large dataset

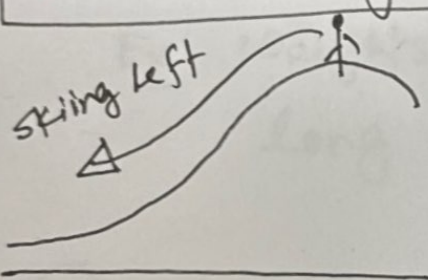
→ Ex. 1 from CH 2 Handout

→ Limitation: Not good if we have

i) too many classes. 

ii) too few classes. 

Popular Histogram Shapes

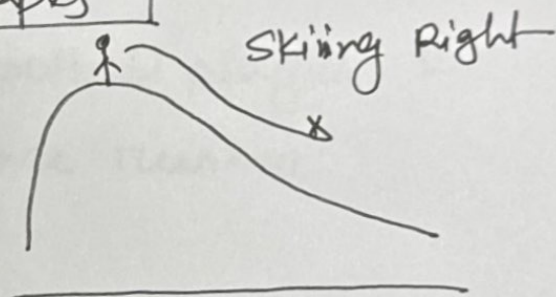


Left-skewed

$$\bar{y} < \text{median.}$$

(mean

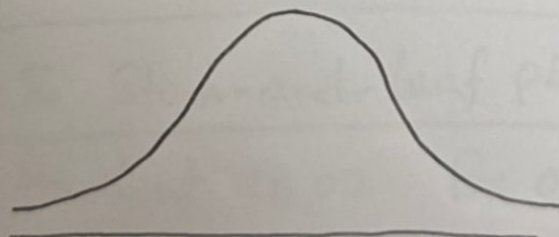
Ex. Age at death.



Right-skewed.

$$\bar{y} > \text{median.}$$

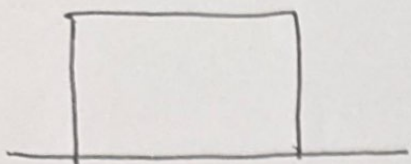
Ex. Annual Income.



Bell-shaped

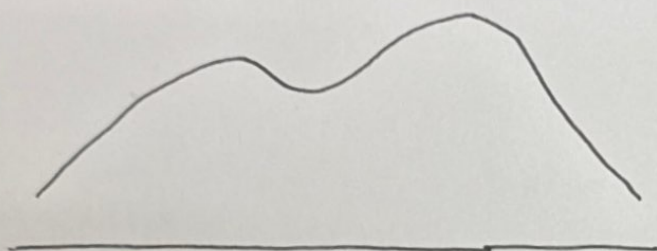
$$\bar{y} \approx \text{median.}$$

Ex: Exam Score.



uniform.

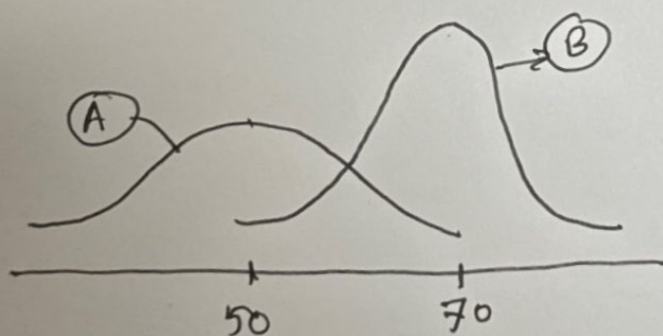
Ex. Winning lottery numbers.



Bimodal.

Ex. weights of football players & long distance runner

Ex



Higher mean? (B)
Higher variance? (A)

Higher

Dot plot & Stem-and-leaf plot

→ Handout ch 02 Ex 02 & 03.

CH 182 Hw due: 01/15/25
on Canvas
(at 11:59 PM)
CH 182 Hw discussion due: Same

MA 3710: Chapter 2 Handout Md Mutasim Billah

Example 1: Histogram: large Set.

Considering the following example, draw the histogram:

Suppose, 50 students run a mile, record times in minutes:

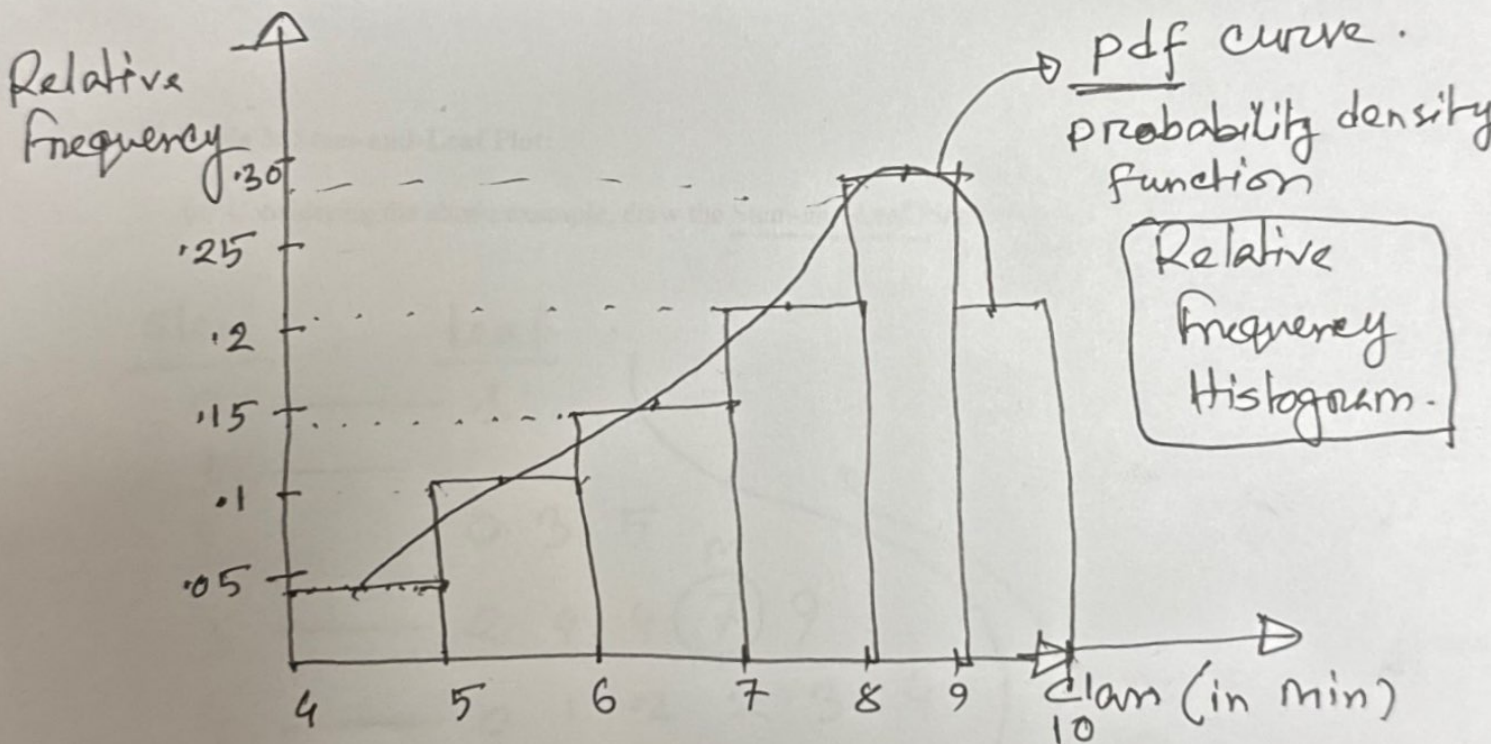
4:55.21 Jake, 5:28.20 Alyssa, 7:15.03 Hannah, ..., 9:31.07 Tyler.

Exclude.

Includes

Class	Frequency	Relative frequency
[4,5)	2	$2/50 = 0.04$
[5,6)	5	$5/50 = 0.1$
[6,7)	7	$7/50 = 0.14$
[7,8)	11	$11/50 = 0.22$
[8,9)	14	$14/50 = 0.28$
[9,10)	11	$11/50 = 0.22$
Total	50	1

Ratio of actual # of favourable events to the total possible # of events



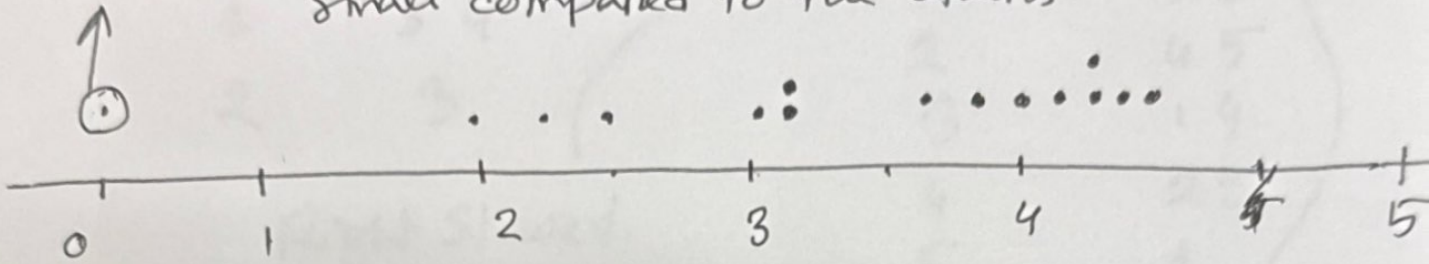
Example 2: Dot Plot

Considering the following example, draw the Dot Plot:

Suppose, we are given the breaking strength of carbon fibers in GPa (gigapascals) as follows: $n=15$ observations

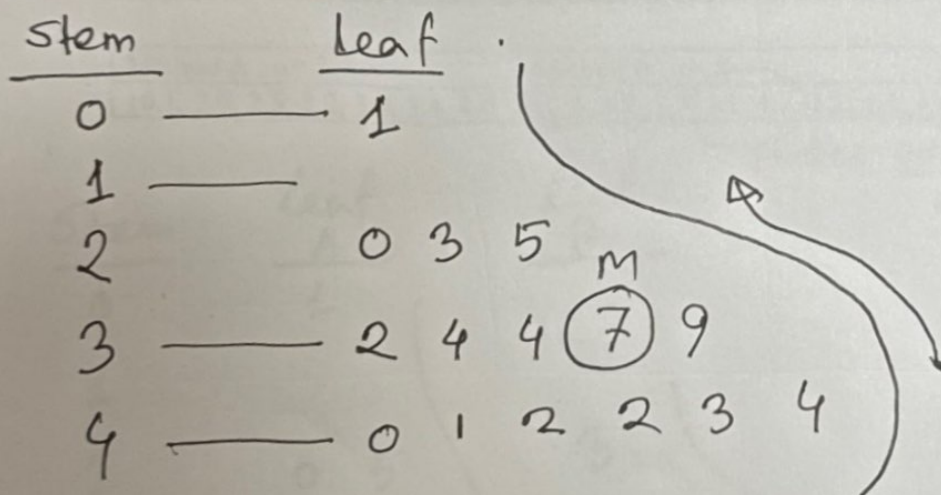
{0.1, 2.0, 2.3, 2.5, 3.2, 3.4, 3.4, 3.7, 3.9, 4.0, 4.1, 4.2, 4.2, 4.3, 4.4}

outlier: An observation that is extremely large or small compared to the others.



Example 3: Stem-and-Leaf Plot:

(a) Considering the above example, draw the Stem-and-Leaf Plot:



left-skewed.

Option 1

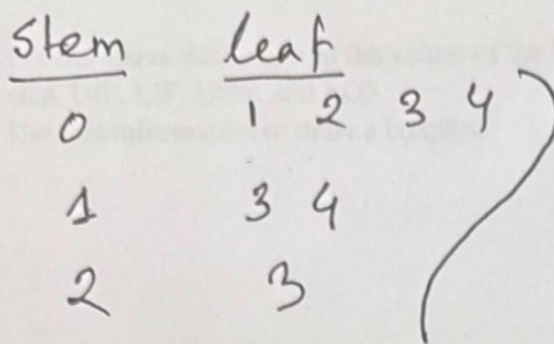
stem	leaf	
2	37	89
3	12	

{2.37, 2.89, 3.12}

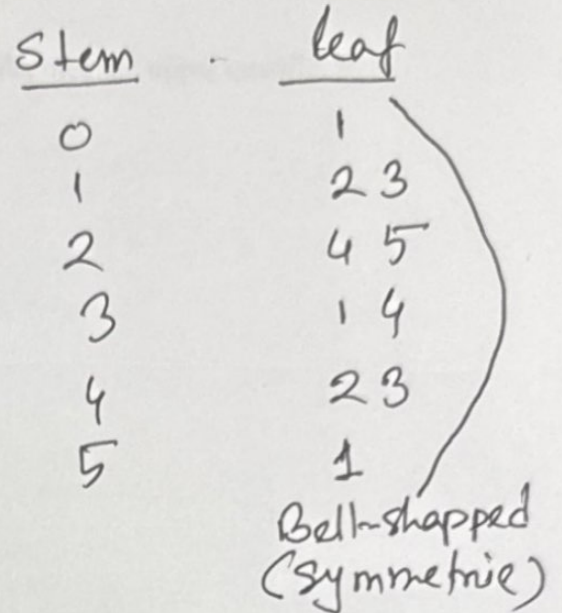
Option 2

stem	leaf	
2.3	7	2
2.8	9	
3.1	2	

~~{0.1, 2.0, 2.3, 2.5, 3.2, 3.4, 3.4, 3.7, 3.9, 4.0, 4.1, 4.2, 4.2, 4.3, 4.4}~~

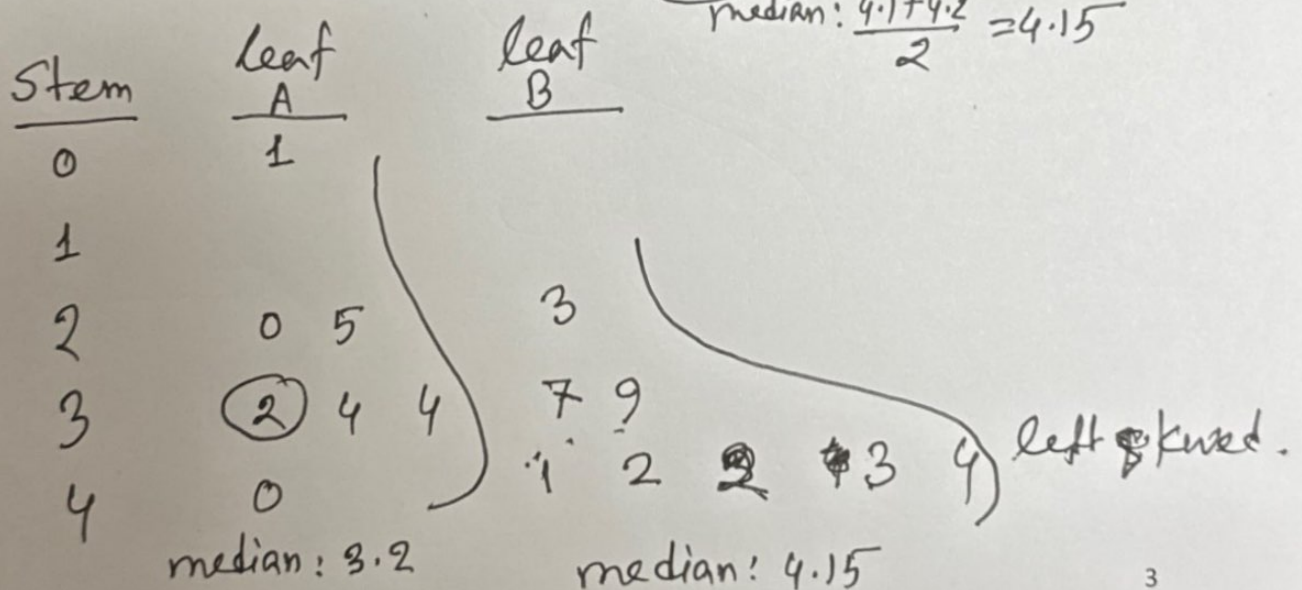


Right Skewed.



- (b) Suppose, we have collected a random sample of 15 observations, 7 were from Method A and 8 were from Method B. Considering the following data, draw a side-by-side Stem-and-Leaf Plot. Describe the plots as right-skewed, left-skewed, or symmetric. Also, compare the mean and median of the observations of method A & B.

Method A : n=7	Method B : n=8
{0.1, 2.0, 2.5, 3.2, 3.4, 3.4, 4.0}	{2.3, 3.7, 3.9, 4.1, 4.2, 4.2, 4.3, 4.4}



Percentile

→ The p -th percentile is greater than or equal to $p\%$ of the data.

Ex. 65-th percentile is greater than or equal to 65% of the data (and less than or equal to 35% of the data)

⑤ median $\sim$
 $\tilde{Y} = 50$ -th percentile

Steps:

- i) Order data from smallest to largest
- ii) If n is odd, median = middle value
If n is even, median = mean (avg.) of the two middle values

$n=3$

$\{ 2, \boxed{4}, 6 \}$
↓
median

$n=4$

$\{ 2, \boxed{4, 6}, 8 \}$

Notation:

data set = $\{ Y_1, Y_2, \dots, Y_n \}$

ordered data = $\{ \underset{\substack{\uparrow \\ \text{minimum}}}{Y_{(1)}}, Y_{(2)}, \dots, Y_{(n)} \underset{\substack{\uparrow \\ \text{maximum}}}{\}$

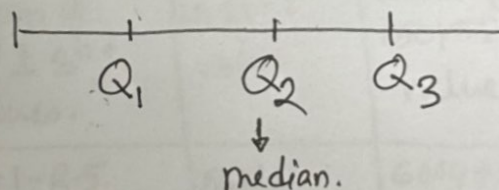
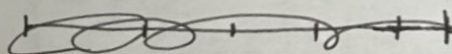
Location of the median: $L_m = \frac{n+1}{2}$

- n is odd; median = $Y_{(L_m)}$
- n is even; median = $\frac{Y_{(L_m - \frac{1}{2})} + Y_{(L_m + \frac{1}{2})}}{2}$

n	6	7	5681	6004
l_m	$\frac{6+1}{2} = \frac{7}{2} = 3.5$	$\frac{7+1}{2} = 4$	2841	3002.5
median	mean of 3rd & 4th value.	4th value.	2841th value.	mean of 3002nd & 3003rd values.

$x \ x \ x \ x \ x \ x$ $x \ x \ x \ x \ x \ x \ x$

Quantiles



Q_1 : 1st/lowen quantile (25th percentile)

: median of the lower half of the data set
(include median if n is odd)

Q_2 : 2nd/middle quantile = median (50th percentile)

Q_3 : 3rd/upper quantile (75th percentile)

: median of the upper half of the dataset
(include median if n is odd)

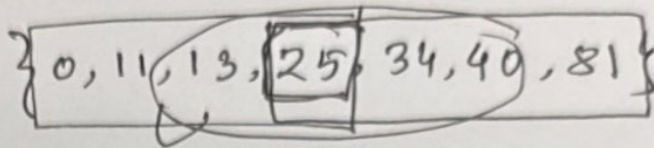
location of quartiles:

$$l_q = \begin{cases} \frac{n+3}{4} & \text{if } n \text{ is odd} \\ \frac{n+2}{4} & \text{if } n \text{ is even} \end{cases}$$

$$Q_1 = Y_{(l_q)}$$

$$Q_3 = Y_{(n+1-l_q)}$$

n	6	7	5681	6004
l_q	$\frac{6+2}{4} = 2$	$\frac{7+3}{4} = 2.5$	$\frac{(n+3)}{4} = \frac{5681+3}{4} = 1421$	$\frac{(n+2)}{4} = \frac{6004+2}{4} = 1501.5$
Q_1	2 nd value	mean of 2 nd & 3 rd values.	1421 st value	mean of 1501 st & 1502 nd value
$n+1-l_q$	$6+1-2 = 5$	$7+1-2.5 = 5.5$	$5681+1-1421 = 4261$	$6004+1-1501.5 = 4503.5$
Q_3	5 th value.	mean of 5 th & 6 th values.	4261 st value.	mean of 4503 rd & 4504 th value.



Example 4: Box-Plot

Suppose, we have the following $n = 7$ observations:

{0, 11, 13, 25, 34, 40, 81}

For the above data set, find the values of the Q_1 , Q_2 , Q_3 , IQR, step, UIF, LIF, UOF, and LOF.

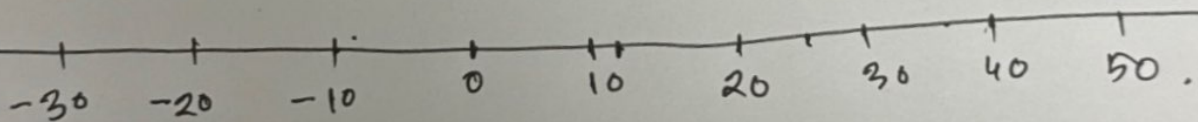
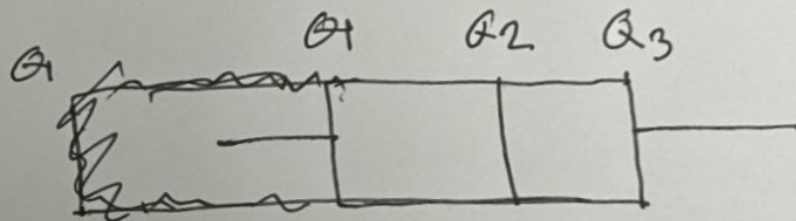
Use this information to draw a boxplot.

$$Q_1 = \frac{11+13}{2} = 12$$

$$\text{median, } Q_2 = 25$$

$$Q_3 = \frac{34+40}{2} = 37$$

$$IQR = Q_3 - Q_1 = 37 - 12 = 25$$



$\{0, 11, 13, 25, 34, 40, 81\}$

$(n = \text{odd}; \text{Dual citizenship})$

Example 4: Box-Plot

Suppose, we have the following $n = 7$ observations:
 $\{0, 11, 13, 25, 34, 40, 81\}$

$Q_1 \quad Q_2 \quad Q_3$

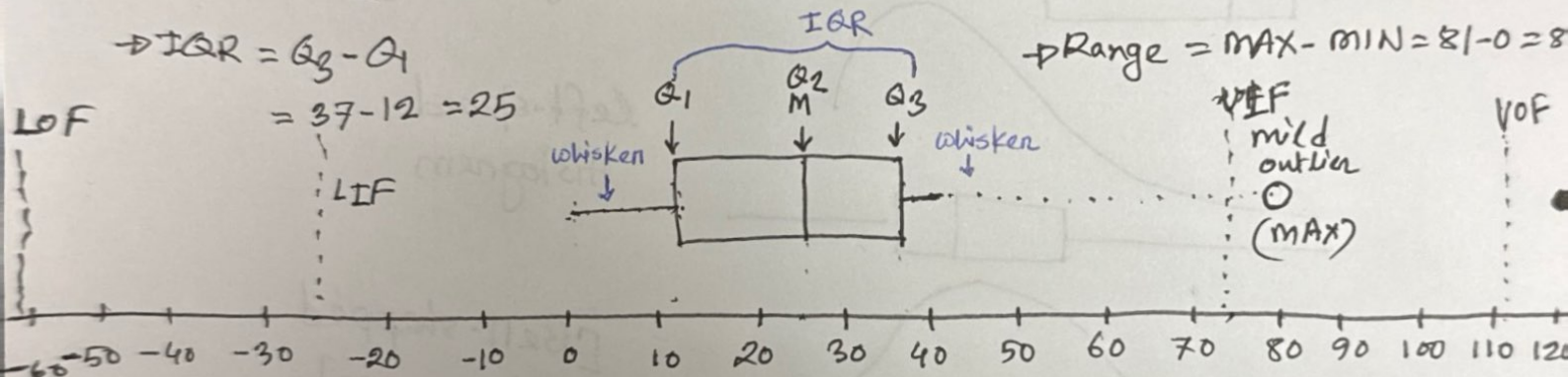
For the above data set, find the values of the lower quartile, median, upper quartile, IQR, step, UIF, LIF, UOF, and LOF.

Use this information to draw a boxplot.

$$\rightarrow Q_1 = \frac{11+13}{2} = 12 \quad \rightarrow Q_2 \text{ (median)} = 25 \quad \rightarrow Q_3 = \frac{34+40}{2} = 37$$

$$\rightarrow \text{IQR} = Q_3 - Q_1 = 37 - 12 = 25$$

$$\rightarrow \text{Range} = \text{MAX} - \text{MIN} = 81 - 0 = 81$$



$$\text{Step} : 1.5 (\text{IQR}) = 1.5 (25) = 37.5$$

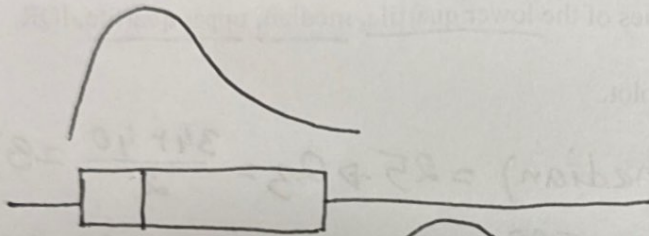
$$\text{Upper Inner fence (UIF)} = Q_3 + \text{Step} = 37 + 37.5 = 74.5$$

$$\text{Lower Inner fence (LIF)} = Q_1 - \text{Step} = 12 - 37.5 = -25.5$$

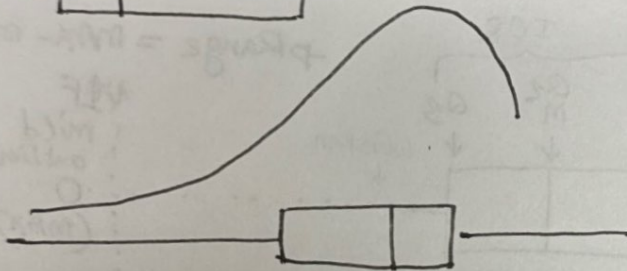
$$\text{Upper Outer fence (VOF)} = Q_3 + 2 * \text{Step} = 37 + 2(37.5) = 112$$

$$\text{Lower Outer fence (LOF)} = Q_1 - 2 * \text{Step} = 12 - 2(37.5) = -63$$

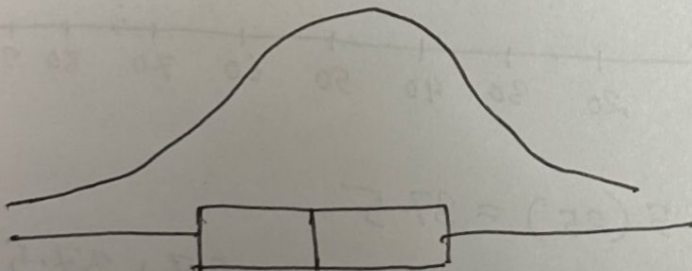
Example:



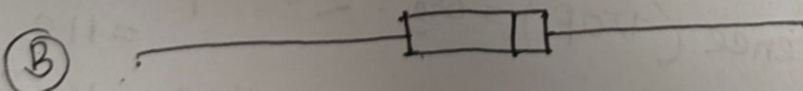
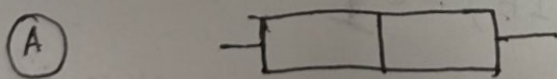
Right-skewed
Histogram



Left-skewed
Histogram



[Bell-shaped
Histogram]



larger median? B

" Range? B

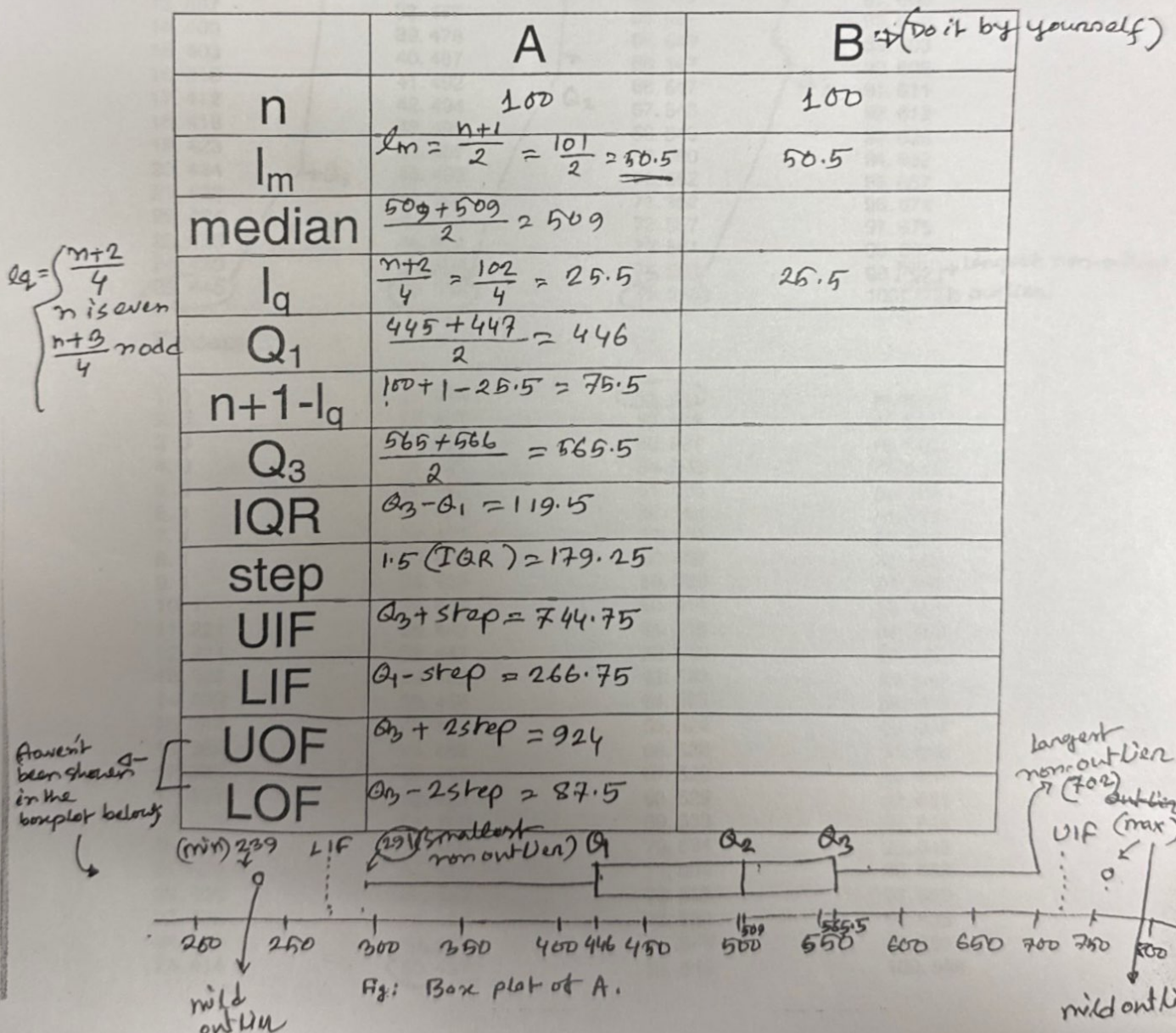
" IQR? A

" Q_1 ? B

" Q_3 ? A.

Brookman Electronics manufactures a new type of electronic component and is examining the lifetime of the components. Engineers are studying two different manufacturing processes named A and B. Engineers are concerned that Process B has about a 10% defective rate; about one out of every ten components made with Process B has a lifetime of less than two hours. Last year, one-hundred components were produced with each process. Engineers recorded the lifetime of each component and rounded off to the nearest hour. The ranked data is on the last page.

For each data set find Q1, median, Q3, IQR, inner fences, and outer fences. Then draw side-by-side boxplots.



$$L_m = 50.5$$

$$Q_2 (\text{median}) =$$

$$L_q = 25.5$$

$$Q_1 =$$

$$n+1 - L_q = 75.5$$

$$Q_3 =$$

Process A:

1. 2391 → outlier	26. 447	51. 509	76. 566
2. 291 → smallest non-outlier	27. 453	52. 511	77. 567
3. 309	28. 455	53. 511	78. 569
4. 312	29. 456	54. 511	79. 575
5. 331	30. 463	55. 516	80. 576
6. 336	31. 464	56. 517	81. 580
7. 356	32. 467	57. 522	82. 585
8. 367	33. 467	58. 523	83. 586
9. 369	34. 473	59. 525	84. 590
10. 376	35. 473	60. 526	85. 592
11. 378	36. 475	61. 528	86. 594
12. 385	37. 475	62. 532	87. 600
13. 397	38. 477	63. 537	88. 603
14. 400	39. 478	64. 540	89. 603
15. 403	40. 487	65. 547	90. 605
16. 410	41. 492	66. 547	91. 611
17. 412	42. 494	67. 548	92. 613
18. 418	43. 495	68. 549	93. 635
19. 423	44. 497	69. 550	94. 652
20. 424	45. 499	70. 552	95. 657
21. 429	46. 500	71. 552	96. 674
22. 434	47. 503	72. 557	97. 675
23. 435	48. 504	73. 561	98. 697
24. 439	49. 506	74. 563	99. 702 → largest non-outlier
25. 445	50. 509	75. 565	100. 772 → outlier

Process B:

1. 0	26. 414	51. 491	76. 551
2. 0	27. 420	52. 499	77. 554
3. 0	28. 423	53. 501	78. 560
4. 0	29. 425	54. 502	79. 560
5. 0	30. 425	55. 505	80. 563
6. 1	31. 426	56. 506	81. 565
7. 1	32. 427	57. 507	82. 567
8. 1	33. 437	58. 508	83. 580
9. 1	34. 438	59. 509	84. 580
10. 1	35. 439	60. 514	85. 581
11. 221	36. 443	61. 515	86. 582
12. 321	37. 447	62. 519	87. 587
13. 325	38. 457	63. 523	88. 587
14. 339	39. 458	64. 523	89. 598
15. 348	40. 460	65. 524	90. 601
16. 352	41. 462	66. 528	91. 613
17. 390	42. 462	67. 528	92. 634
18. 392	43. 465	68. 529	93. 641
19. 394	44. 468	69. 533	94. 644
20. 400	45. 476	70. 534	95. 648
21. 403	46. 477	71. 534	96. 659
22. 408	47. 484	72. 535	97. 663
23. 408	48. 485	73. 537	98. 665
24. 410	49. 486	74. 543	99. 691
25. 414	50. 487	75. 549	100. 698

CH 3: Probability



Set : An unordered collection of zero or more elements.

$$\begin{aligned}\text{Ex.} \rightarrow \text{Alphabet} &= \{A, B, \dots, Z\} \\ &= \{Q, Z, \dots, K\}\end{aligned}$$

$$\rightarrow \{5, \alpha, \text{John}, \text{Alicia}\}$$

$$\rightarrow \text{positive integers} = \{1, 2, 3, \dots\}$$

$$\rightarrow \text{people over 500 years old} = \{\} = \underline{\underline{\phi}}$$

↓
Empty Set

"
the set of zero elements

Subset : A set where each element is also a member of another set.

$A \subseteq B$ if every element in A has also in B .

$$\text{Ex.} \therefore \text{vowels} = \{A, E, I, O, U\} \subseteq \{A, B, \dots, Z\}$$

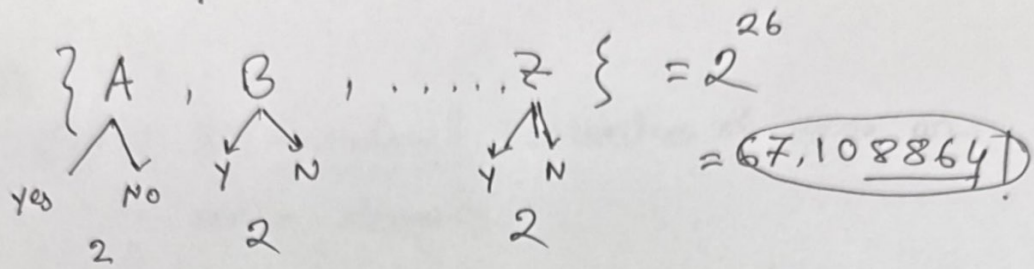
$$\text{prime numbers} = \{2, 3, 5, 7, \dots\} \subseteq \{1, 2, 3, \dots\}$$

$$\rightarrow A \subseteq A \text{ (Every set is a subset of itself)}$$

$$\rightarrow \phi \subseteq A \text{ (An empty set is a subset of all sets)}$$

Ex.

Ex. How many subset of $\{A, B, \dots, Z\}$ exists?



Rule: If a set has k elements, there are 2^k subsets.

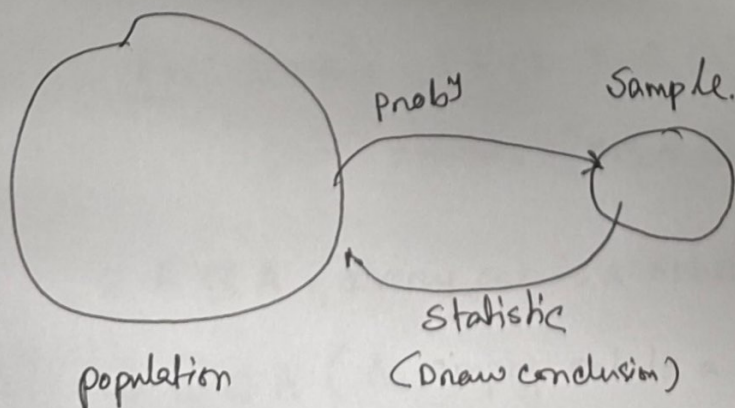
Ex. Possible outcomes of one coin flip = $\{H, T\}$

Subsets: $\emptyset, \{H\}, \{T\}, \{H, T\}$

2 elements
 $\Downarrow$
 $2^k = 2^2 = 4$
 events.

Experiment: A process with variation in its outcome.

Sample Space: S : The set of all possible outcomes of an experiment.



Jan 15
Experiment $\rightarrow$ A process with variation in its outcome

Sample Space $\rightarrow S$: The set of all possible outcome of an experiment

$\rightarrow$ Ex.: Flip a coin once $= \{H, T\} = S$ $\sim$

Event $\rightarrow$ A subset of S

$P(A)$ = The probability of event A is a value between 0 and 1 - representing how likely it is and the proportion of time it will occur in the long run.

$$= \frac{\text{'size' of event } A}{\text{'size' of } S}$$

Alternatively,

$$\frac{\text{\# of occurrences of } A}{\text{\# of experiments}} \xrightarrow{n \text{ large}} P(A)$$

Properties: $0 \leq P(A) \leq 1$

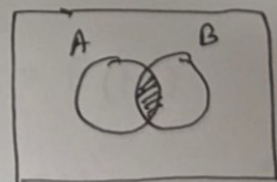
$\rightarrow$ If $P(A) = 1$, it will occur, $P(S) = 1$

$\rightarrow$ If $P(A) = 0$, it won't occur, $P(\emptyset) = 0$.

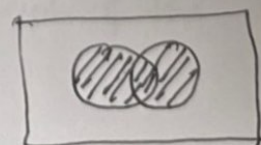
Go to CH 3.1 Handout Ex. 01.

Intersection: $A \cap B = \text{"A and B"}$

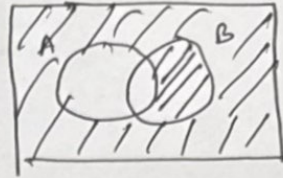
c.c. $A \rightarrow$ And.



Union: $A \cup B = \text{"A or B or Both"}$



Complement : $\bar{A} = \text{"not A"}$



$$\bar{A} = A^c$$

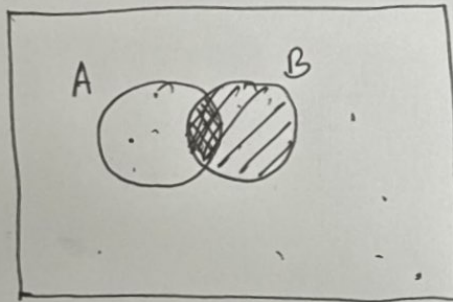
Go to Ex. 1 again.

Conditional Probability :

$P(A | B)$ = The prob of A given B

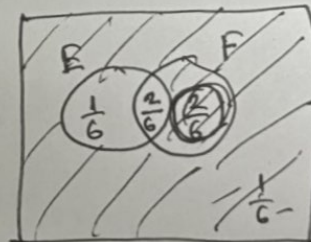
↑
"given"

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



Continue Ex. 1:

$$P(F | \bar{E}) = \frac{P(F \cap \bar{E})}{P(\bar{E})}$$



$$= \frac{2/6}{3/6} = \frac{2}{3}$$

$$P(F) = \frac{4}{6} = \frac{2}{3} ; P(F | \bar{E}) = P(F)$$

⊛ A & B are independent if the occurrence of one event does not effect the prob of the other.

→ A & B are indep if & only if

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

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

→ A & B are mutually exclusive (disjoint) if
it is impossible for both to occur.

→ A & B are mutually exclusive if & only if

$$P(A \cap B) = 0.$$

Chapter 3.1 (Probability): Handout
Md Mutasim Billah

$$\text{Ex. 1) } S = \{1, 2, 3, 4, 5, 6\}$$
$$E = \{2, 4, 6\}$$
$$F = \{1, 2, 3, 4\}$$

Example 1: Let's we have a 6-sided die and we flip the die one time. Suppose,

E : An Even Number appears

F : A number less than or equal to 4 appears. Find $P(E)$, and $P(F)$.

$$P(E) = \frac{\text{size of event 'E'}}{\text{size of 'S'}} = \frac{3}{6} = \frac{1}{2} ; \text{ etc.}$$

$$P(F) = \frac{4}{6} = \frac{2}{3} ; P(E \cap F) = \frac{2}{6} = \frac{1}{3} ; P(E \cup F) = \frac{1}{6} + \frac{2}{6} + \frac{2}{6}$$

$$P(E') = \frac{3}{6} \quad P(E) + P(F') = \frac{3}{6} + \frac{3}{6} = 1 ; P(F') = \frac{2}{6} = \frac{1}{3} \quad \left| = \frac{5}{6} \right.$$

Example 2: Suppose, we select a freshman at random at a big state University.

$P(\text{enrolled in Calculus}) = P(C) = .3$

$P(\text{enrolled in History}) = P(H) = .5$

$P(\text{enrolled in both classes}) = P(H \cap C) = .2$

Draw a Venn diagram & find the following:

(a) $P(H \cup C)$:

(b) $P(\overline{H \cup C})$:

(c) $P(H \cap \bar{C})$:

(d) $P(\bar{H} \cup C)$:

(e) $P(C/H)$:

(f) Are C & H independent? Justify your answer.

(g) Are C & H mutually exclusive?

Chapter 3.1 (Probability): Handout
Md Mutasim Billah

Ex. 1) $S = \{1, 2, 3, 4, 5, 6\}$
 $E = \{2, 4, 6\}$
 $F = \{1, 2, 3, 4\}$

Example 1: Let's we have a 6-sided die and we flip the die one time. Suppose,

E : An Even Number appears

F : A number less than or equal to 4 appears. Find $P(E)$, and $P(F)$.

$$P(E) = \frac{\text{size of event 'E'}}{\text{size of 'S'}} = \frac{3}{6} = \frac{1}{2} ; P(F) = \frac{4}{6} = \frac{2}{3}$$

$$P(E \cap F) = \frac{2}{6} = \frac{1}{3} ; P(E \cup F) = \frac{1}{6} + \frac{2}{6} + \frac{2}{6} = \frac{5}{6}$$

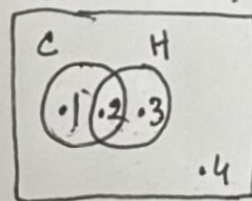
$$P(E') = \frac{3}{6} ; P(F') = \frac{2}{6} ; P(E) + P(F') = \frac{3}{6} + \frac{2}{6} = 1 ; P(F') = \frac{2}{6} = \frac{1}{3}$$

Example 2: Suppose, we select a freshman at random at a big state University.

$P(\text{enrolled in Calculus}) = P(C) = .3$

$P(\text{enrolled in History}) = P(H) = .5$

$P(\text{enrolled in both classes}) = P(H \cap C) = .2$



$$P(S) = 1$$

Draw a Venn diagram & find the following:

(a) $P(H \cup C)$: $.1 + .2 + .3 = .6$

(b) $P(\overline{H \cup C})$: $P(\text{enrolled in neither course}) = 0.4 [P(H \cup C) + P(\overline{H \cup C}) = 1]$

(c) $P(H \cap \bar{C})$: 0.3

(d) $P(\bar{H} \cup \bar{C})$: $.4 + .1 + .2 = .7$

(e) $P(C|H)$: $\frac{P(C \cap H)}{P(H)} = \frac{.2}{.5} = .4$ [40% History students are in calculus]

(f) Are C & H independent? Justify your answer.

$$P(C) = .3, P(H) = .5 \neq P(C) * P(H) = .3 * (.5) = .15 \neq P(C \cap H)$$

(g) Are C & H mutually exclusive?

$$P(H \cap C) = .2 \neq 0 \text{ No,}$$

H & $\bar{H}$ are mutually exclusive.

Jan 17

⑧ Multiplicative Laws:

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

$$\Rightarrow P(A \cap B) = P(B) * P(A|B)$$

$$\Rightarrow P(A \cap B) = P(A) * P(B|A)$$

$$P(A \cap B) = P(B \cap A) = P(B) * P(A|B)$$

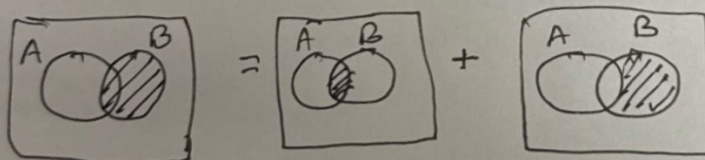
$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

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

⑨ Laws of total probability:

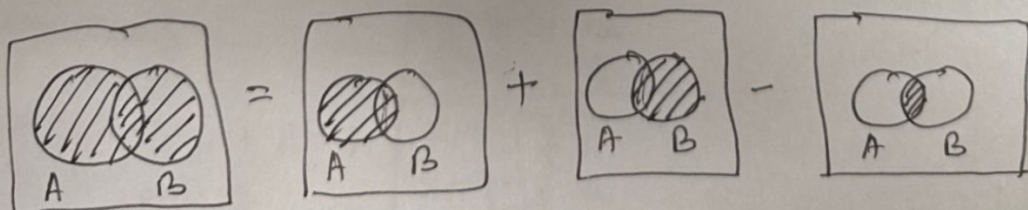
$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$

$$= P(AB) + P(\bar{A}B)$$



⑩ Additive Law:

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



⑪ Complement Rule:

$$P(A) + P(\bar{A}) = 1$$

$$\Rightarrow P(A) = 1 - P(\bar{A})$$

Handout Ex.03

Bayes Rule

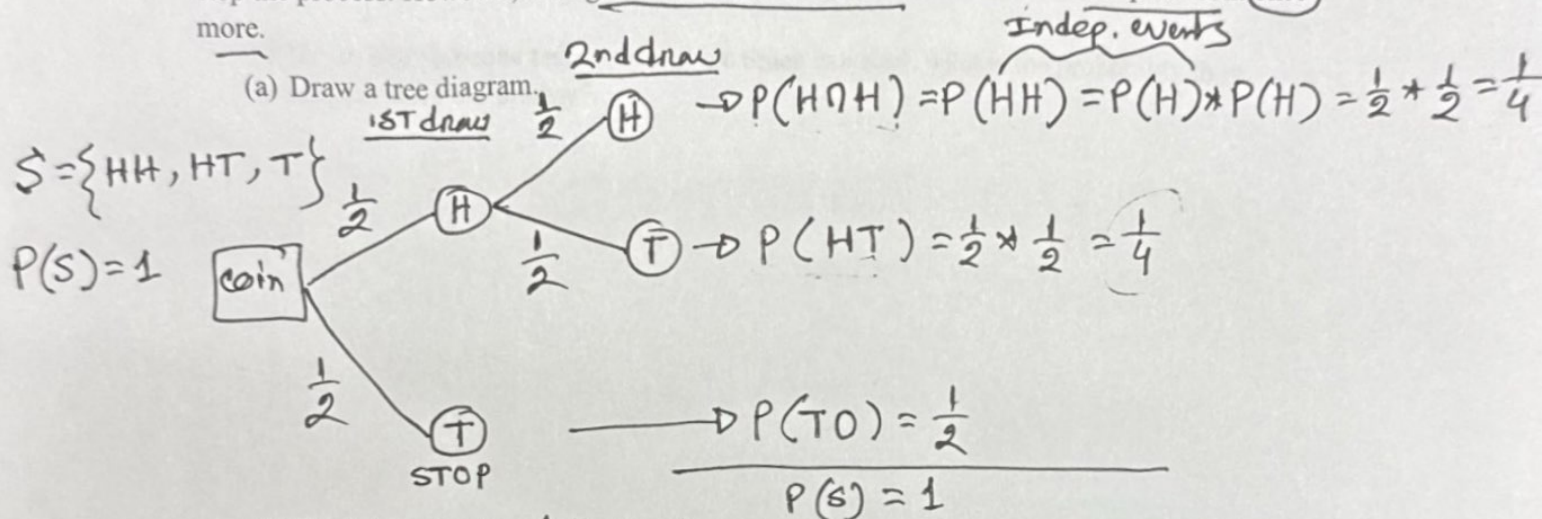
$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \begin{array}{l} \text{law} \\ \text{Total Prob.} \end{array} \quad \left[P(B) = P(AB) + P(\bar{A}B) \right]$$

$$\begin{aligned} &= \frac{P(AB)}{P(AB) + P(\bar{A}B)} \\ &\quad \text{(multiplicative Law)} \rightarrow \frac{P(A) * P(B|A)}{P(A) * P(B|A) + P(\bar{A}) * P(B|\bar{A})} \end{aligned}$$

Ex. 04. Handout

(h) H and $\bar{H}$ are mutually exclusive.

Example 3: Suppose, John is flipping a 2-sided coin. If he gets a tail in the first draw, he could stop the process. However, if he gets a head in the first draw, he is needed to flip the coin once more.



(b) Find $P(\text{Head appears} | \text{tail appears})$.

$$P(H|T) = \frac{P(HT)}{P(T)} = \frac{\frac{1}{4}}{P(HT) + P(TO)} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{2}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

By law of total prob.

Example 4: 1% of the people on an island have Disease X. A test for the disease is 99% accurate.

(a) Given that someone on the island tests positive, what is the probability that they actually have the disease?

Define: D: Disease, $\bar{D}$: No disease; T: Test is positive
 $\bar{T}$: Test is negative.

complement (or) Rule $\rightarrow$

$$P(D) = 0.01$$

$$P(\bar{D}) = 0.99$$

1st

$$P(T|D) \text{ OR } P(D|T) = 0.99$$

2nd

complement $\rightarrow$

$$P(T|\bar{D}) = 0.01$$