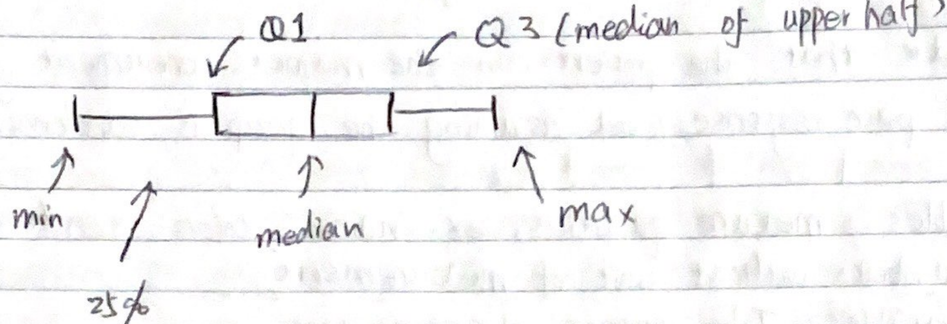
AP Stats

• Calculator Notes.

Conclusion: Focus on what's being measured.
Observation: Focus on ~~is~~ facts. lists the facts, Randomized.



• Using Graphing Calculator for data information:

(#) Given data set: 0, 0, 0, 0, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 6, 7, 7, 9, 10, 10, 10, 10.

Go to Stats, add values to the list, Click "Calc," perform 1-Var Stats, then we can view the information.
$\bar{x}$ (mean) med (median) Q1 (median of first 50%)
Q3 (median of the second 50%)

Causation: If $A \xrightarrow{cause} B$, and $\bar{A} \xrightarrow{cause} \bar{B}$, then causation is established.

$A \xrightarrow{cause} B$ ① Variables consistent.

$\bar{A} \rightarrow \bar{B}$ (control group)

Create Frequency Table:
Sample: 4 6s, 3 7s, 3 8s, 6 9s

L_1	L_2
6	4
7	3
8	3
9	6

Calc → 1st Var → L_1, L_2
Set Freq List L_2.

Conclusion:

① What is measured? ② What is the treatment?

↳ Response to pain levels. ↳ The magnet treatment.

※ we can conclude that the patients in the magnet treatment group yield a lower pain response level assuming the group is randomized.

- Categorical Variables: measure qualities, ex. names, colors, genders. can be numbers without unit of measurement.
- Quantitative Variables: Take numeric values, ex ages.

Individuals: a single person/things we measure.
Population: An accumulation of individuals.
Sample: A small group taken from population.

Confounding Variables: two variables that have the same effect on a response and can not be distinguished.

Ie: The effect of exercising on health. Those who exercise also change their diet, which also helped improve health. Diet becomes a confounding variable.

In other words: there are other factors in the experiment that can lead to the same result, then those factors are known as confounding variables.

Part 2) Remember to randomly assign.

- Surveys: • Select a sample of the population
  - Ask questions, record answers
  - Potential dishonest response, false result.

- Observational Study: • Observe individual and do no impact to influence response, cannot establish causation
  - Do nothing to ~~grou~~ observation group and do nothing, no groups assign
- Experimental: • Assign treatments to groups, indicate treatment when needing to identify.

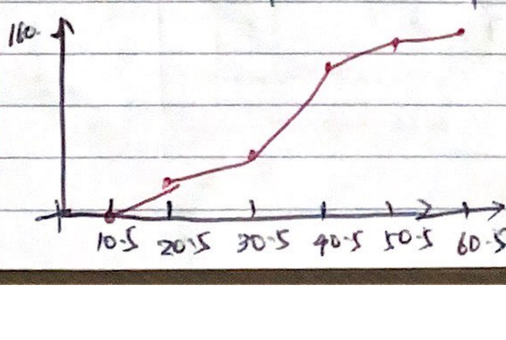
Measuring-variable: the variable that ~~是~~ is measured and

Additions for Studies: Consider Ethical issues, realistic scenarios.
- Experimental: Assign study groups, the topic is cause/effect issue, assign treatment to groups.

• Graphs: ① Ogive: - Table of class boundaries and cumulative frequencies.
- Make a dot over upper class boundaries at the height of the cumulative class frequency.
- Begins on Horizontal axis at the lower class boundary of first class.

ex. → Cumulative Frequency Table.

Intervals	Frequency	Class Boundaries	Cumulative Frequency
11 - 20	23	10.5 - 20.5	23
21 - 30	43	20.5 - 30.5	66
31 - 40	51	30.5 - 40.5	117
41 - 50	27	40.5 - 50.5	144
51 - 60	7	50.5 - 60.5	151



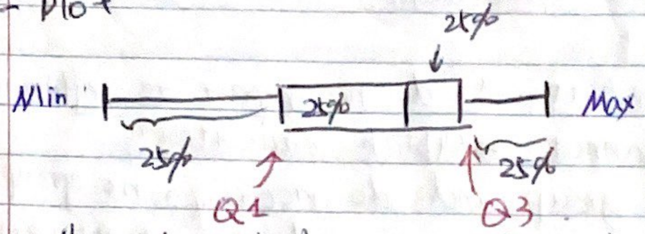
~~(10.5, 0), (20.5, 66), (30~~
~~(20.5, 23)~~ (10.5, 0), (20.5, 23)
~~(30.5, 43)~~ (30.5, 66), (40.5, 117)
~~(40.5~~ (50.5, 144), (60.5, 151)
coordinates ↑

Graph Box - Plot using Ti - 84.

2nd Y=, Plot, → choose box - plot, set window (x-max).

Section 1.1

Box - Plot



the median of the first 50%

The median of the second 50%

Interquartile Range (IQR): distance between the first and third quartile.

(IQR = Q3 - Q1)

Values are considered an "Outlier" if it's greater than Q3 by 1.5IQR, or less than Q1 by 1.5IQR.

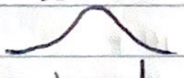
Outlier > 1.5IQR + Q3 or outlier < Q1 - 1.5IQR

when dealing with similar data, we want to show the calculations for IQR etc, list observations outliers are usually used against the target

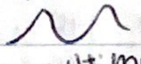
~~cout~~ | count column | 10s, | unit digit | (stem-leaf plot)

Describing Distributions:

- Shape
  - bell curve



  - skewed
  - symmetrical
  - Bimodal

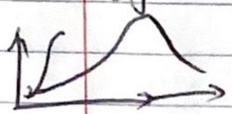


  - multimodal

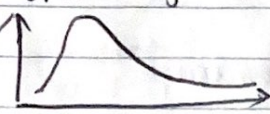
- Center:
  - Mean
  - Median
  - Middle 50% (where is)
- Spread / Range.
  - Find Q1
  - Median
  - Q3,
  - IQR?
- Outliers:
  - If yes, justify it.

- A graph is symmetrical if mode is in the middle, equal numbers of either side of the graph.
- A graph is skewed when data points that are either bigger or smaller than normal range of values.

skew - left



skew - right



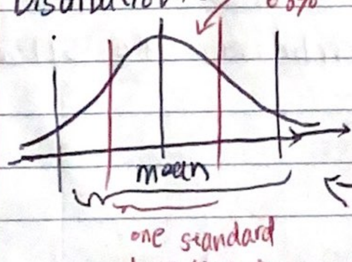
Making distribution discription.
The distribution is centered around at x, describe shape. The IQR is between $x \sim y$, it range is from $x - y$. Describe outliers.

- Histograms:
- the x-axis is the quantitative variables (numbers).
  - usually split 6 - 10 boxes. max - min / # desired groups
- the y axis can be labeled with
  - Frequency (count)
  - Relative Frequency (percentage)
  - Cumulative - total cout for the current and all previous groups.
- Ogive (Cumulative Relative Frequency graph)

Drawing Histograms on Graphig Calc.

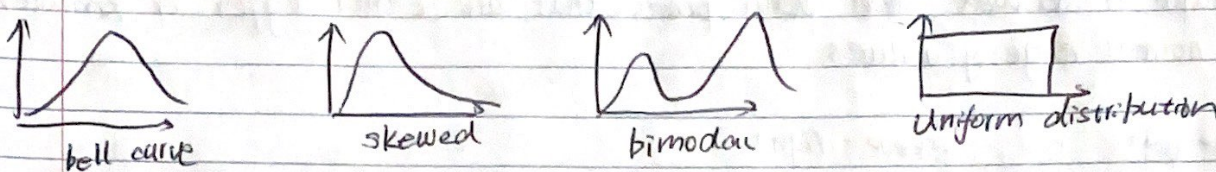
1.2

Normal Distribution:

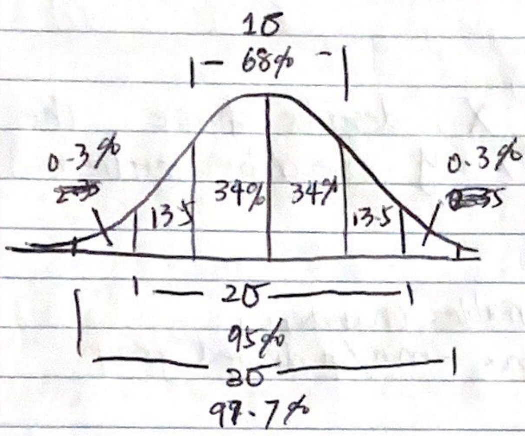


Standard Deviation: A measure how spread the data is from the mean.

Distribution refer how data is spread out from a range of values.
Distribution can be both categorical or quantitatice



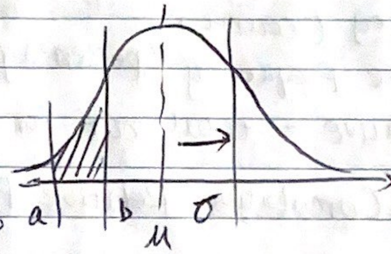
• Normalization: σ for standard deviation



z-score, the z-score tells you how many standard deviations your data is from the mean.

The mean of a "normalized" distributer will become 0, the σ will become 1.

Calculator- normal cdf (a, b, μ, σ)
lower upper mean σ

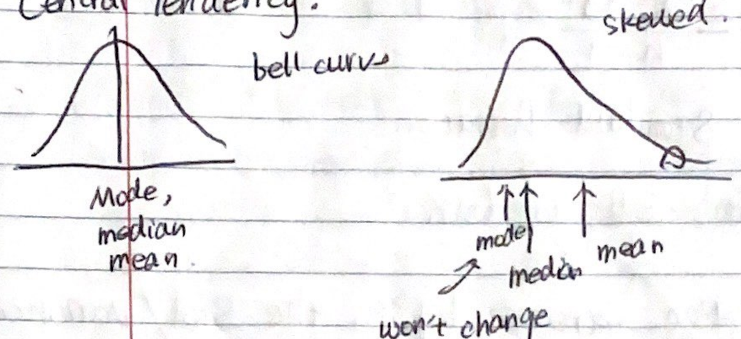


this calculates the percentage of data in range a-b

• If a distribution is skewed, choose the median to describe it, use box-plots. Find Q1, Q3 and IQR, look for outliers.

• If Distribution is symmetrical, choose mean to describe it, use STD and varicnce of distribution to describe the spread.

Central Tendency.

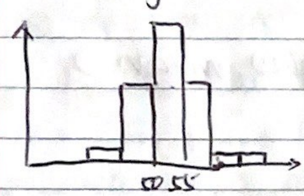


Mid Range : $MR = \dfrac{\text{Lowest} + \text{Highest}}{2}$

~~Ex: The test scores of a Math 12 exam~~

- Key Concepts:
  - Describing Distribution (Relate to context, center, spread, outliers, shape)

ex.



The distribution of yield (relate to context) is roughly symmetrical, the median is between 50-55 bushels (relate to context), the range of yield is approximately 25 bushels, there're no potential outliers (outliers are usually distant from cluster)

  - Comparing Distributions (varied noticeably more, higher on average etc)

Variance(s^2) & standard deviation(s) should only be used when the distribution is symmetrical with the mean in the center and unimodal.

- Large Variance / std deviation → data is spread out
- If variance is small, datas are close to the center.

Standard Deviation measures the average distance of observations from their mean.

Std / Variance are heavily affected by outliers.

$$S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} \qquad S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

variance Standard Deviation.

When population goes up, the Std will go down.

If all values in a normal distribution increase by x, the Std/variance remains the same.

2.1. Z-score

Compare a data relative to all other scores or data points.

Z-score and normal distributions can be used to compare how a data point is relative to every other data point in a distribution.

$$Z\text{-score} = \frac{x_i - \mu}{\sigma}$$ (how many std above/below mean)

Z-score ≥ 0, better or mean,

Z-score < 0, worse than mean.

When answering questions related to z-score / better performance, measure z-score (# of STD above/below mean)

x_i = the value of the data point

μ = population mean σ = standard deviation (population)

s: sample standard deviation.

InvNorm function on Calc, inputs the probability / area, which outputs the z-score.

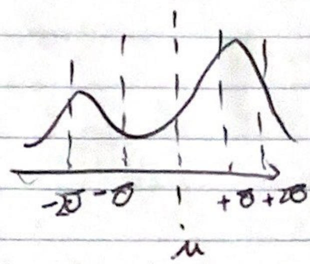
Inv (0.99, 70.5, 6.5)

Inv Norm (x_i, $\bar{x}$, μ) ⇒ the actual value

Chebyshev's Inequality:

percentage within $= 100\left(1-\frac{1}{K^2}\right)$

- chebyshev's inequality gives a minimum percentage of observations that will be within "K" std of the mean, where "$K>1$".
- "K" is the # of std. from mean. (For any shape of distribution)



ex. $K=2$

$100\left(1-\frac{1}{2^2}\right) = 0.75 = 75\%$

At least 75% of the data is within 2 standard deviation.

- A data is considered as majority for above 50%

☆ How do you tell that a set of data is normally distributed

Choose distribution graph (6th) → zoomstat (straight ⇒ then normally distributed

Add on to chebyshev's Inequality:

If we use chebyshev's Inequality using 1σ, and the value is not $= 68\%$, then the graph is not normally distributed

2.2.

Telling a set of data is normally distributed

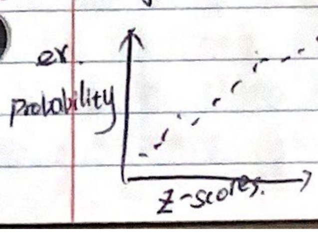
Method 1: Draw a histogram → bell-shape, 68 - 95 - 99.7.

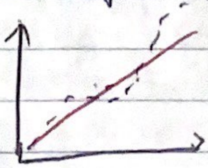
Method 2: Draw a "Normal Probability Plot".

- Most data will be on a straight line.
- If the largest observations deviate from this line → skewed right.
- If the smaller observations deviate from the line → data is skewed left.

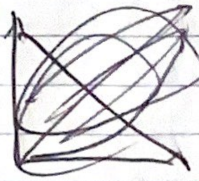
To be tested

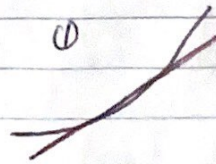
ex.

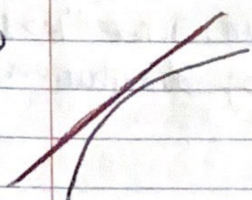


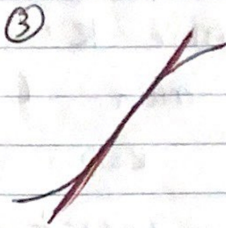


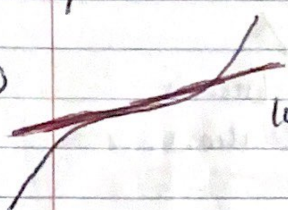
larger observations deviate, skew-right.



①  curve up - Right skewed

②  Left skewed

③  short - tails.

④  long - tails.

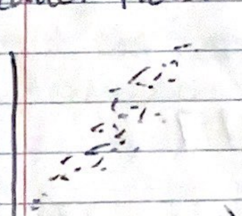
Shade Norm ⇒ Draws a picture of the normal distribution 

Chapter 3. Bi - Variable Correlation.

Correlation does not imply causation -☆

☆ Draw Scattered Plots, 2nd 0, → diagnostic On, type in correlation data in L1, L2. L1 and L2 must be correspansive.

Scatter Plots.



positive correlation,
somewhat linear
Large clusters in the middle
Strong Correlation, Outliers

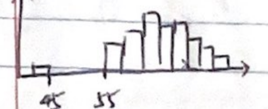
key points.

- positive, negative
- Strength / strong, weak, moderate.
- Shape
- Outliers
- clustered, scattered.

Concepts Not Included (On College Board)

- Continuous / data / variable : a data that has infinite recordings, different values between an interval
- Discrete data / variable : A numerical value is determined by counting, fixed data, infinite possibe values, however, values are $\boxed{\text{limited}}$ (ex. number of times a person smokes)
- Describing Histograms: You can touch on gaps, clusters and potential outliers

ex.



The distribution displays a gap with a potential outlier located between 40% and 45%.

- You can describe a dot plot as Quantitative and Discrete (whole num)

→ Draw Scatter plot: turnDignosticOn ⇒ done ⇒ Stats ⇒ Calc → Linear Reg (ax+b)
⇒ Reg (ax+b, L1, L2, Vars ⇒ Function (Y1), Graph (zoom stat)
$-1 < r < 1$, no units.

※ Correlation Data (r) : the closer it is to 1, the stronger a positive linear relationship
the closer to −1, the stronger a negative linear relationship.
The closer the value is to zero, there would be no correlation.

Using regresion, Y is the "predicted value" $\hat{x}$

(r^2) Represents the percentage of data can be explained by the x-value
x - variable

- If we swap the variables of x and Y, the correlation remains same.

ex. Be precise about the x, y variables (ex. the number of words in the SAT essay, the score for your SAT essay)

The number of words written in the SAT essay and score in your SAT essay has a strong positive, linear correlation (0.88), thus we can conclude there is a positive association between the two variables. No guarantees that an ~~large~~ essay with more words would grant you high score.

3.2 : Least Square Regression Line.

- r^2 is used to determine how well the LSRL does at predicting values of the response variable 'y'.
- r^2 indicate the percentage of the variation in the response variable that can be explained by using the LSRL of "y" on "x".
- If r^2 is close to 1, the LSRL is better model for using "x" to predict "y";

LSRL always intersect at $(\bar{x}, \bar{y})$

In general, LSRL is compared against an "average"

- This line use "x" to predict "$\hat{y}$"

$\hat{y} = a + bx$

predicted response

- The LSRL will have errors (residuals), the difference between the actual y-value and predict $\hat{y}$ value.

Equation of LSRL:

$\hat{y} = a + bx$ $\boxed{b = r\left(\frac{S_y}{S_x}\right)}$

b: Slope of LSRL

'r' is the correlation

S_x and S_y are standard deviations of each data set "x" and "y".

mean x and y

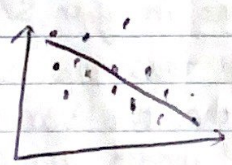
$(\bar{x}, \bar{y})$

Y intercept:

$a = \bar{y} - b\bar{x}$ "a" = Y-intercept of the LSRL;

You cannot use L_1 to L_2 Regression line to [use L_2 predict L_1].

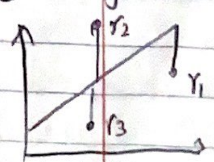
~~Residute~~ Residuals : Residual = $y - \hat{y}$



"+" Residual : Underestimate.

"−" Residual : Overestimate

Line of best fit (LSRL)



Minimize $(r_1)^2+(r_2)^2+(r_3)^2$

We want to turn all the $(y-\hat{y})$ positive.

Interpret the slope:

ex. For every additional year that a person smokes, it is predicted that their lung damage increases by 1.85%.

Using r^2:

ex. 97% of the variation in percentage can be explained by using the LSRL on years of smoking.

X Interpret the coefficient of determination "r"

___% of the variation in response variable can be explained by the linear relationship with explanatory variable

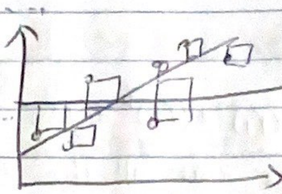
ex. 82% of the variation in child's weight can be explained by the linear relationship with the child's height.

Definition of r^2.

$SSE = \Sigma(y-\hat{y})^2 = \square\square\square\square$

$SST = \Sigma(y-\bar{Y})^2 = \square\square\square\square$

$r^2 = \frac{SST - SSE}{SST}$



3.3. Residual Plot.

Reading Data: "ax+b"

- Constant means Y-intercept (b)
- Coef : Slop (a).

SE : Standard error. (ex. 51.03 ± 51.03)

ex. 63.2% of the variation in the back-pack weight can be explained by the linear relationship with the person's body weight.

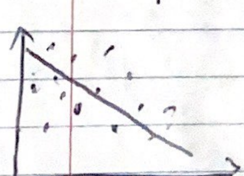
r^2 intepretation.

Assessing LSRL (Residual Plot)

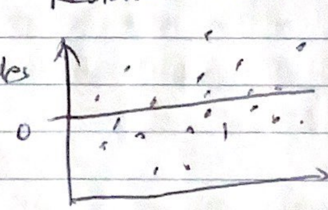
3 methods determine if a LSRL is a fit of the data:

- r^2, coefficient of determination.
- Graph Residual Plot, shows difference between predicted and actual value.
- STD of ~~a~~ residuals, a large variation in residual std, larger prediction error. Small variation in residual std, smaller prediction error.

ex. Scatterplot.



Residual Plot.

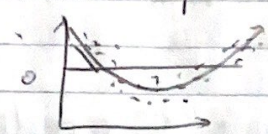


amount of the data on both sides of the Residual plot gives a good indication of the regression line

STD of residual tells you the average value that is off by the true value using the LSRL.

- A LSRL is a good model if half the residuals are above and half below, random, spread out.
- Indicates a linear model and regression is useful for predicting y-values.
- Positive residual → model predictions underestimate.
- Negative residual → model predictions overestimates.

A curve pattern in residual plot indicates a linear regression line is a poor model ex.



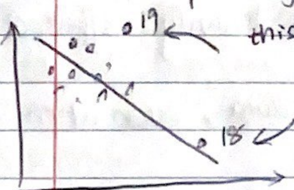
Graph Residual plot:

Enter L_1 and L_2, L_3, vars → Y-var, Y_1 ⇒ $(Y_1(L_1))$ ⇒ $L_4 = L_2 - L_3$

Outlier and Influential Points:

- All influential points are outlier, yet not all outliers are influential points.
- An observation is considered influential if it affects slope / y-intercept when added or removed.
- Outliers are point far from other data on the x-axis.

ex:



Lurking Variable:

- Lurking variable Z causes A and B to increase, however there is no causation between A and B thus Z is a lurking variable.
- They are usually unnoticed, they are neither explanatory or response variable.
- Make others seem to have a causation, yet it causes both to happen.

Confounding Variable:

Confounding occurs when effect of variables cannot be distinguished from one another.

$x \xrightarrow{?} y$ Both causes, yet ignorant about which ~~causes~~ impacts more.

$z \xrightarrow{?} y$ ex. Being active tends to live longer.

z confounding. However exercise, health can be other variables that lead to the same result.

4.1

- your regression model may only be good for using "x" to predict "y" within the domain of your data points.

Extra Notes:

① Correlation requires that both variables to be quantitative, so that it makes sense to do the arthmetic indicated by the formula r. We can't calculate r between incomes of a group of people and what city they live in, because city is categorical.

② Correlation measures only the strength of a linear relationship between 2 variables.

③ Like μ and σ, the correlation is not resistant. r is heavily affected by outliers

④ Correlation is not a complete summary of the 2-variable data, even when relationship is linear. Should provide other relevant info.

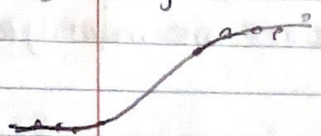
- Extrapolation is the use of a regression line for prediction outside the range of values of the explanatory variable x used to obtain the line. Such prediction are often not accurate.

Standard Deviation of Residules:

$$s = \sqrt{\frac{\sum \text{residuals}^2}{n-2}}$$

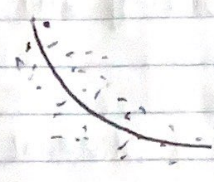
- If the role of explanatory and response variable is reversed, they would result in a different LSRL.

- Logistic Regression:



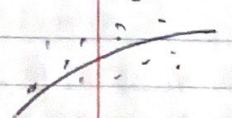
$$y = \frac{c}{1+ae^{-bx}}$$

- Exponential Regression



$$y = a(b)^x$$

- Power Regression



$$y = a(x)^b$$

We take S_x as the STD of residuals for a set of data.

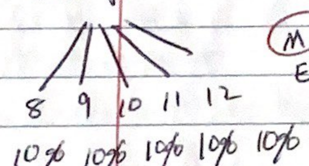
Chapter 5.1 Sampling and Surveys

Objective: Get a sample that's representative of the entire population?

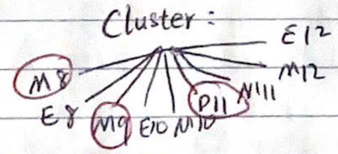
▲ Tyes of Sampling:

- SRS – Sample random sample
- Stratified Sampling (Taking a set number in each ~~fair~~ group, ex. groups grades)
- Cluster Sampling (Divide into small fair groups, then take a number of small groups)
- Multistage

Stratified:



Cluster:



▲ Bad Sampling Methods:

- Convinience Sampling / Voluntary Sampling

→ Problem of Voluntary group: Overrepresent extreme views, not representitive of the entire population.
→ problem of Convinience: Overepresent certain groups, people around will share similar ideologies and qualities.
- Garbage in, garbage out; If data is collected badly, outcomes are meaningless.

- Population and Samples:
population: entire group of idividual (not necessairly people) being observed.
sample: part of population in the study;
census: Information is collected regarding everyone;
Individual: one sample from the population.

- Data Collection:
The method to select sample is called " sample design".

- SRS:
- All individuals are equally likely to be selected for samples
- ✗ Any group of "n" individuals are equally likely to be chosen.
- It's important to describe how you would implement SRS.
- Use Table of random digits to select the sample
  - (Table B) random Digits — back of the book.
  - rand Int (lower, upper, [# of trials])
- Not good enough.

↳ Use of Table of Random Digits
- The left most column is the line number;
- You read the number horizontally;

ex. 19223 95034 05756

↑ this is an example of taking every group of 3.

iv). Amongst 100 AP students, 3 sample groups of 10 students will be selected.
→ Assign a random 3 digit number to the 100 AP students, ~~the first~~ starting with line 101, the first 10 students with the matching number will be group 1,

※ If a treatment is agiven, then it's an experiment.
If there're no control group, we can't establish causation.

the next 10 is group 2, next 10 will be group 3.

— Stratified: It's not a type of SRS. / considered ratio

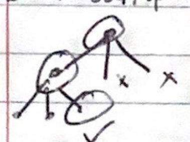
• In SRS, it's possible that an entire strata can be selected. In stratified sampling, this is Not possible.

— Cluster:

• Taking one group out of all random groups:

— Multistage:

• In stats, multistage sampling is taking of samples in stages using smaller and smaller sampling units at each stage.



When using stratified, we should us SRS for each individual group.
When designing an experiment, make sure specify the group, how you select data from each data (percentage, number), explain why. Touch on aspects such as overrepresentation, underrepresentation

• Convinience sampling: Survey people around you;
• Voluntary sampling: Survey people who volunteer to complete survey

— Undercoverage:

◦ Some groups are left out in the process of choosing a sample
◦ Excluding certain groups

— Non-Response:

◦ Individuals chosen cannot be contacted or don't cooperate.

— Response Bias:

◦ Respondents may lie or try to guess what the interviewer wants to hear.

- selection Bias: the selected participants tend to give a specific direction of answer

- Confusing / Misleading Questions:
  - questions that implies or intends to lead the audience to think or answer in a certain way.

5.2 Designing Experiments

Experiment: has a control group, has a treatment.
※ Control used to establish causation, $A \to B$, $\bar{A} \to \bar{B}$
Surveys can never be used to establish causation

- Experimental units: The individuals on which the experiment is done.
- Subject: If the units are people. Specify number.
- treatment: The experimental condition we apply to the unit (indicate levels)
- Factor: The explanatory variables
  - Factors may be applied in different levels.
- Placebo: a fake treatment
- placebo Effect: individuals of an experiment receiving a fake treatment have the same result as those recieving the actual treatment.

3 Principles of Experiment Design.
- There must be a control group, this is to establish Causation
- The control group usually receives a placebo treatment.
- Randomization used to assign experimental unit to each treatment, eliminates variation between groups, keeping variable the only variation.
- Replication: Repeat the experiments enough times to so that the results you see due to effects of the treatment and not by a chance.

Dividing Sample into groups, assign treatments to each group, and also how results are interpreted.

A statistically significant result is defined as a result that could not have happened by chance.

Separating Based on characteristics

- Blocking: separating subjects into different groups that have similar traits (age, hair colour, fitness level).

- Matching: Experiment with only 2 treatments, subjects are paired based on some variable, then randomly assigned to each treatment.

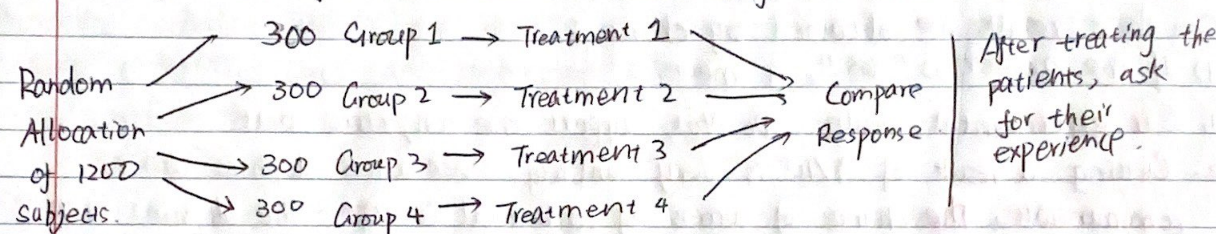
- Randomization: Randomly assign units to each group.

X

Ex.: Assign Four Treatment groups: → explain the groups.

Each individual is assigned to a 4 digit ranging from 0000 - 1200, Randomly select the first 300 for 1st treatment, next 300 for 2nd treatment, so on.

X



Example of Blocking: By using blocking to form equivalent groups, we can control this variable. Each instructor will teach one class using multimedia and one using an overhead. We can number the classes from 1 to 12. Using the random number table, the first lesson of each instructor will be the one taught by multimedia and the other by an overhead. (In this scenario, blocks are each professer, treatments are the 2 teaching methods)

Other Types of Experiments:

Double-blinded Experiment: Neither the subjects nor the people who have contact with them know which treatment a subject has received.

Matched-Pair Design: An experimental design where participants are matched in pairs based on shared characteristics before they are assigned to groups; One participant from the pair is randomly assigned to the treatment group while the other is assigned to control group.

Section 6.1 . Simulations and Probability Experiments.

• Why do we do simulations ?

~~Simulations~~ can be ~~If you think something is true~~

RandInt NoRep (Initial, Last)
↳ it only arranges order, so exclude last parameter

Doing Simulation on Ti - 83
↳ Math → Ran Prob → RandInt 2nd → 0, Alpha S .

Do simulations based on probability ex. 50% 50% , we can use flipping coin.

ex. Conjecture: Can dophins actually communicate with each other.
Counter Argument: If dophins can't talk to each other, the probablity of success will be 50%)

Can these results be obtained by chance ;

1) If the answer is "Yes", it means =
1. Our experimental results Do Not support the conjecture that
2. Getting a result of 9/10 is very ~~un~~likely, even when dophins can't communicate. The chance of success of each trial is 50%. The dophins were just lucky in this experiment.
3. If conduct experiment agains, resules can be completely different. Since they can't communicate the results won't likely repeat.

2) If the answer is "NO", it means :
1. The experimental results Do support the conjecture that they can communicate.
2. Getting a result of 9/10 is very unlikely, if dophins can't communicate these results are so rare that it is impossible for dophins to Not be able to communicate.
3. If we are to conduct this experiment again, It's very likely that we will obtain similar results because dophins CAN communicate with each other.

X We do simulations many times, we only do the actual experiment once.

Binopdf (trials, probability, number)

6.2 Probabilities Rules.

Rule 1) When A and B are independent, the probability of A and B happening at same time.

$P(A \cap B) = P(A) \times P(B)$, when you can use formula to prove independency.

Rule 2) When A and B are dependent, then we have

$P(A \cap B) > P(A) \times P(B)$ when A occurs, B is likely to occur. Positive association

$P(A \cap B) < P(A) \times P(B)$ when A occurs, B is less likely to occur. Negative Association

Rule 3) When A and B are mutually exclusive, they cannot happen at same time!

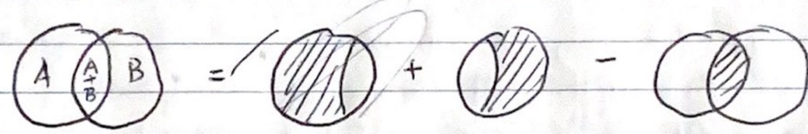
Ie. when "A" occure, "B" can't happen.

$P(A \cap B) = 0$ Two are mutually exclusive

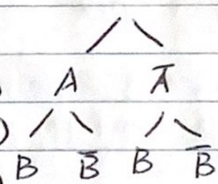
- Mutually exclusive events are events that one outcome supersedes the other. Mutually exclusive events are dependent

二、Join Events:

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$



Use Tree Diagram For Probability.



True: somth happening

False: somth not happening

Case 1: Both A and B are true $= P(A) \times P(B)$

Case 2: Both A true and B false $= P(A) \times P(\bar{B})$

Case 3: Event A is false B true $= P(\bar{A}) \times P(B)$

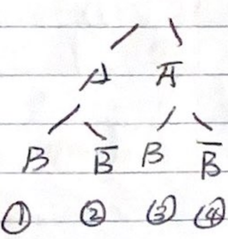
Case 4: A and B false $= P(\bar{A}) \times P(\bar{B})$! We only use this when A & B are independent!!

三、Conditional Probability.

Conditional probability is used when we know certain event occured. When this occure, certain branches in the tree will be ignored.

ex. Given B already occured, whis is $P(A|B)$

$P(A|B)$ Given B occurre, ~~P(A~~ probability of A occur.



$$P(A|B) = \frac{1}{1 + 3}$$

$1 + 2 + 3 + 4 = 100\%$

X) If A and B are dependent

then $\boxed{P(A|B) = \frac{P(A \text{ and } B)}{P(B)}}$ $P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$

False Positive: You don't have HIV but tests say Positive

False Negative: You have HIV but tests say Negative.

8.1 Binomial Distribution

- Binomial Distribution are used when there're only 2 outcomes: ex. HT, TF.
- Formula for Binomial Distribution:

$$\boxed{P(X) = {}_nC_x (p)^x (q)^{n-x}}$$

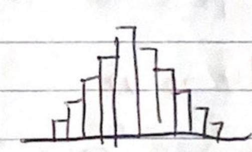
The probability of "x" number of success, "n" is the # of events.
Out of "N" trials, we chose "x" number of success.

Conditions for using binomial distribution:
Condition 1: Only 2 outcomes in scenario
2: Probability of success in each are constant, P(success) + P(faliure) = 1.
3. Outcome of any individual trial doesn't affect other trials.
4: # of trials is set, conditions separates Binomial vs Geometric Dist.

Binomial PDF vs CDF
2nd - Vars / 0 = Binopdf.
binompdf (n, p, x)
- the probability of having exactly 4 correct answers on a MC test with 10 Questions and 4 choices each: binompdf (10, 0.25, 4)

BinomCdf: Binomial Cumulative Distribution Function:
Calculate the probability from zero to "x" number of successes.
- 2nd / Vars / A
Binomcdf (n, p, x)
- The probability of having up to 4 correct answers (0, 1, 2, 3, 4)
binocdf (10, 0.25, 4) = 0.92187

Drawing a histogram using probability:



L1	L2
0	P(0)
1	P(1)
2	P(2)
3	P(3)
⋮	

- set L2 using binompdf (15, 0.45, L1)
- Draw histogram using (L1) (x-list) L2 (freq)
- Adjust window "xscl" to 1
- Ymax = 0.5 (less than 1

Characteristics of A Binomial Distribution:

i) The probability of Binomial Distribution is "Symmetrical".

ii) The distribution is symmetrible at its Mean, give equation: Mean $\mu_{Binom} = n \times P$

iii). The distribution also has a standard deviation and variance, given by the equations: Standard Deviation

$$\sigma_x = \sqrt{\mu_x(q)} = \sqrt{np(q)} = \sqrt{np(1-p)}$$

iv) Changin increasing / decreasing doesn't change the shape of Dist. It only shifts the dist horizontally; (smaller "p" → left or larger "p" → Right)

V) If the number of trials "n" is large, use a "Normal Distribution" to approximate the probabilities of a Binomial Distribution.

Normal Approximation.

normalcf $(a, b, np, \sqrt{np(1-p)}) =$

vi). Three conditions for using a normal Dist to approximate a Binomial Distribution:

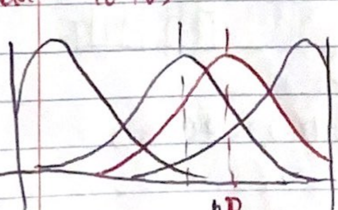
1. Binomial: There're only 2 outcomes.
2. Independence: Each trial must be independent
3. "N": There must be a fixed number of trials.
4. "P" The probability of sucess in each trial must be consistent
5. $n \times p \geq 10$ and $n \times (1-p) \geq 10$ (Ensure enough population)

- Calculating complements:

ex. At least 9 females in 13 with probability = 0.75.

$P(x=4) =$ ~~$_{13}C_4 (\ldots)$~~ $_{13}C_4 (\frac{3}{4})^9 (1/4)^4$

X add on to iv)



Changing "p" will not change the shape. If "p" is extreme, it may seem scaled.

• When answering Questions: Make sure to be specific

Sample: ex 1. Random Variable X ?

The # of people who (p×n) would ag

• Chapter 7.1

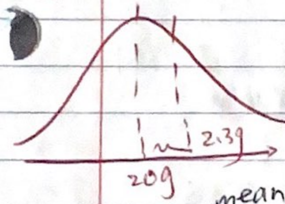
ex. Suppose the mean weight of an egg is 20 g with std of 2.3g, and the mean weight of a basket is 5lbs with a std of 1.2 lbs. What is the mean weight and std of basket with 12 eggs. What is p total weight is more than 300 grams.

Break Down

① Suppose weight of one egg is normally distributed.

② Weight of one basket in grams.

convert grams to pounds.



Then what is the mean weight of 12, and std ?

X Rule 1: When adding / subtracting random variables, you add or subtract their mean respectively.

X Rule 2: You can't combint standard deviations, Instead, combine $\boxed{\text{variance}}$

ex. $\sigma_{12} \neq \sigma_1 + \cdots \sigma_{12}$ but $\sigma_{12}^2 = \sigma_1^2 + \cdots \sigma_{12}^2$

$\sqrt{\sigma_{12}^2} = 2.3^2 \times 12$

$\sigma_{12} = 2.3 \times \sqrt{12}$ $\quad \sigma_{combine} = \sigma \times \sqrt{n}$

X Rule 3: It doesn't matter whether you're adding or subtracting two random variables, you always ADD them.

If $A = B + C$ $\quad \sigma_A^2 = \sigma_B^2 + \sigma_C^2$

If $A = B - C$ $\quad \sigma_A^2 = \sigma_B^2 + \sigma_C^2$

Random variables "A", "B", "C" must be independent.

Rule 4: When converting random variables using a linear equation, convert the mean using the same equation.

Let. $F = a + bX$

$\mu_F = a + b\mu_F$

grams = 453.592 x pounds.

Rule 5: When converting ~~variance~~ random variable using a linear equation, convert the variance by multiplying by (slope^2) or multiple std by slope.

$\sigma_F^2 = b^2 \sigma_F^2$ $\sigma_F = b\sigma_F$

$\sigma_{grams} = 453.592 \times 12$

Conditions For using a normal Distribution.

1. Each of the random variables are independent of one another.

2. Each random variable comes from a normal distribution

Finding Mean & STD For discrete random variable.

$\boxed{\text{Discrete}}$: values that are not continuous, set values.

When finding μ and σ, Use "1-Var stat test" Calculate with Ti 83.

ex.

Random	1	2	3	4	5	6
Probability	0.15	0.2	0.1	0.15	0.3	0.1

$\boxed{\text{expected Value}}$ $E(x) = \Sigma(x_i P_i) = \mu_x$

Average value is also the expected value)

STD of "X" $\sigma = \sqrt{\Sigma(x - \mu_x)^2 \times P_i}$

- When Finding Expected value, 1-var Stats Calc, L1 as List, L2 as Frequency. This gives your average earn or loss

- $\boxed{\text{Continuous}}$: Values are continuous, such as all values from 3 to 4. if only 3 and 4, then it's discrete.

Important Example:

Y = units sold	200	300	500	1000
P	0.1	0.2	0.5	0.2

X = units sold				

g) suppose profit is given by $Z = 200X + 350Y$.

Find the mean and standard deviation for profit.

$\mu_Z = 200\mu_x + 350\mu_y$

$\sigma_Z^2 = (200\sigma_x)^2 + (350\sigma_y)^2$

- Rules For means and Variance.

If X and Y are dependent, "ρ" is the correlation between X and Y

$\sigma^2_{x+y} = \sigma^2_x + \sigma_y^2 + 2\rho\sigma_x\sigma_y$

$\sigma^2_{x-y} = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$

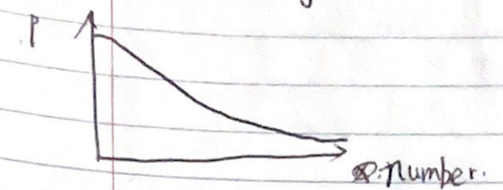
Geometric Distribution.

Random Variable X - number of trials required to obtain the "First Choice"

"X" is a geometric random ~~when~~ variable when $\boxed{\text{(4 conditions)}}$

- There're only 2 outcomes: Success or faliure
- The variable of interest is the number of trials required to obtain the first success.
- Each observation is independent
- The probability of success p for each observation is the same

- Since "n" is not fixed there could be an infinite number of "x"



- Geometcdf (p, x) the probability of getting the first success within the first "x" attempts
- Geometpdf (p, x) – the p of getting first success with exactly "x" attempts.

Not Getting any success in the first "x" attempts =
$P(x) = 1 - \text{Geometcdf}(p, x)$

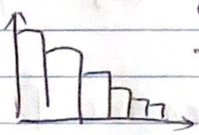
X Geocdf(1/9, 6) = $\dfrac{\text{Geocdf}(1/9, 10) - \text{Geocdf}(1/9, 4)}{1 - \text{Geocdf}(1/9, 4)}$

Mean & Standard Deviation of Geo

$\mu = \frac{1}{p}$ $\sigma = \sqrt{\frac{q}{p^2}}$

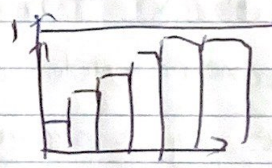
$(PX = n) = (1-p)^{n-1} \times p$

- Graphing Geo Dist (PDF) in Ti 83
Enter 2nd → Vars → D: geometpdf (1/6, L2)



Change the scale
XList: L1
Freq: L2

- Graphing Geo CDF in Ti 83
In L2, Enter 2nd → Vars → Geocdf (1/6, L1)



Infinitely close to 1.

※ Chapter 9. (most important Chapter)

9.1.

"Inference" Intro

- Use a sample to find the 'true μ" of the population.
- Rather than focusing on one sample, you will use theories on mean of many samples.

∘ Parameters: (Information about the population)

- Sample averages will be different every time you select a new sample
- The word "Statistics" is used for describing a sample.

μ ← population mean	$\bar{x}$ ← sample mean
σ ← population STD	S ← sample STD

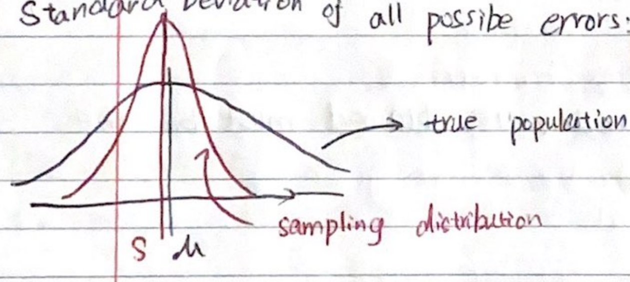
Assume the true population of has a μ of 175cm

(▯ represents a sample average)

Then we take a number of samples and find their $\bar{x}$ (averages).
Eventually, as more samples we take, we are going to have a distribution close to Normal.

This is called the distribution of sample averages.

Standard Deviation of all possible errors: standard error $= \frac{SD}{\sqrt{n}}$



S and μ are very close as long as the sample size is large.

△ Notations:

Parameter:		Statistics.	
mean: μ	Pop size $= N$	mean: $\bar{x}$	sample size: n
SD: σ	Proportion: p	SD: s	Proportion: $\hat{p}$
Variation σ^2		Variation s^2	

In order to use our theories. on distribution of sample mean

1. The samples you take Must be a Random sample from population.
   A random sample prevents bias, over/under representation
2. The sample size must be same size. The sample size will affect Variation of Distribution of sample means.

About Dist of sample means,

1. Depending on the population Distribution or sample size, the distribution of sample mean is "Normally Distributed" Central Limit Theorem.
2. The variation in the Dist of sample mean is smaller than the variation of means in population
3. The mean from the Dist of sample means is Equal to the population mean.

- Central Limit Theorem: If distribution is skewed (not normal), your sample size needs to be at least $\boxed{30}$ for the sampling distribution to be normal.

- The distribution of your sample means should look normally distributed if certain conditions are met [SRS, $n \geq 30$, $N \geq 10n$]
- The variation the distribution of sample means will ↓ as your sample sizes get bigger.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \leftarrow [\text{Given } N \geq 10n]$$

- The sample sizes where all the samples means are obtained must be the same [consistent sample sizes].

— Population Distribution Matter?

- If population dist is normal, dist of sample mean is normal
- If population dist is not normal, sample size 30 or more to be normally Dist

- The Dist of sample means theoretically includes all possible samples.

※

3 Conditions to use Dist of sample means for probability purposes.

(1) Population is normally distributed or sample size is 30 or more.

(2) The population size is 10 times or more than the sample size.

(3) The sample MUST be a random sample.

- The mean of the sampling distribution of $\bar{x}$ is equal to the population mean 'μ' even if the sample size 'n' is small.

Useful Functions:

Math → Prob.

- RandNorm (μ, σ, n) μ is mean, σ is std, n is sample size.
  ↪ returns 15 values from the distribution.
- List - math - mean.
  mean (RandNorm (μ, σ, n)) gives mean of 15 sample means.

Means vs Proportions:

- Proportion is often used when the variable of interest is binary (2 outcomes)
- If the variable of interest is not binary [count], we could use population mean and sample mean.

population proportion p, & Sample population $\hat{p}$.

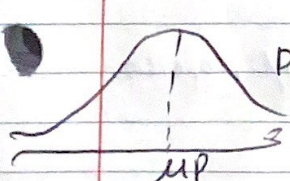
$\hat{p} = \frac{x}{n}$ (x = sample num ~~that conforms~~) meet condition
(n = ~~population~~ sample size)

Dist of sample Proportions.

1) The μ of the dist of all the statistic $\hat{p}$ is equal to population proportion p.

2). The STD of the distribution of all the statistic $\hat{p}$ is given by the equation: $\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}}$ [if P not given, We assum P = 50% if binary]

↑ Dist of sample proportion?,



Conditions for using Dist and Sampling Proportions.

#1) The population size must be 10 times or more than the sample size.

$N > 10n$ ↑ this must be met to use $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

#2). The sample sizes n must be big enough to satisfy two inequalities.

$np \geq 10$ $\quad$ $n(1-p) \geq 10$

#3) Must be SRS.

	$\vec{x}$	$\vec{p}$
1.	SRS	SRS
2.	$N \geq 10n$	$N \geq 10n$
3.	Normality $n \geq 30$	$\begin{cases} np \geq 10 \\ nq \geq 10 \end{cases}$

- Estimator:
  - You use p to estimate $\hat{p}$
  - You use σ to estimate $\vec{x}$

Section 10.1 Estimating with Confidence & Confidence Int.

Calculator Process: STAT → "Tests" →

- The "Tests" are for performing Hypothesis Testing.
- The Intervals are for finding confidence intervals. [
- Zint is used when the population dev is given.
- Tint is used when population std dev is not given.
- "Prop Int" refers to proportion intervals.
- Use 1-sample when there's dependence between samples
- 2-sample is used when there're two independent samples.

Main idea "Confidence Interval".

- It's a range of values centered with your sample mean/proportion with the goal of "capturing the true population mean".

$\bar{x}$ ∘ When we want 95% Confidence Interval, we're saying that our "method used for creating an interval [range]" will emcompass the population mean μ, 95% of the time.

- Ie. a 90% C.I means that the method we used for creating this interval will capture population mean 90% of the time.

∘ $$C.I = \bar{x} \pm z^{*}(\sigma_{\bar{x}}) = \bar{x} \pm z^{*}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Based on z-interval (10.1)

∘ $$C.I = \bar{x} \pm t^{*}\left(\frac{s}{\sqrt{n}}\right)$$

Use t-distribution (10.2)

∘ $$C.I = \hat{p} \pm z^{*}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

10.3. Use this when we have sample proportion

A: 1-Prop Interval.

∘ Your. C.I. is your sample mean $\bar{x}$ or sample proportion $\hat{p}$ +/- "margin of error"

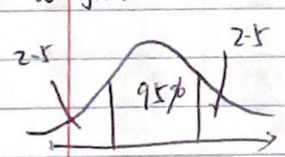
∘ If we ↑ the percentage of confidence, the margin of error will ↑ the range gets bigger.

$$\text{Confidence Interval} = \bar{x} \pm \boxed{z^{*}\left(\frac{\sigma}{\sqrt{n}}\right)}$$ Margin of Error

z^{*} : z-score (Depends on % of C.I)

$z^{*}\left(\frac{\sigma}{\sqrt{n}}\right)$: margin of error. $\frac{\sigma}{\sqrt{n}}$: standard error.

How to find z^{*}



inuNorm(0.025) → z^{*}

Z-Interval → Stat →

Say finding 95% C.I, we use invNorm (0.025) for z^{*}

ex. C.I.(95%) = invNorm(0.025) = -1.96 ∴ z^{*} = -1.96

※ About Confidence Interval

- Concluding sentence: Our 95% Confidence Interval betwee [a, b]. In context, we are confident that method if repeatedly used with sample size "n" will capture the true population mean 95% of the time.

- How to increase a confidence Interval?
  - Bigger z value → Bigger

Concluding sentence in a question:

- We are confident that the method we used to create the confidence Interval of 102.99 to 107.01 will capture the true mean IQ score of Moscrop students 99% of the time.

$\mu_m = \frac{1}{4}\mu_A = \frac{1}{4}(\mu_B + \mu_C + \mu_D + \mu_E)$

$\sigma_m = \frac{1}{4}\sigma_A = \frac{1}{4}\sqrt{(\sigma_B^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2)}$

10.2 Using T- Distribution

3 Conditions that must be met to construct a confidence Interval.

(SRS) The data must come from a Random Sample or Randomized

(Independences)

(Normality)

T- Interval : When population σ is unknown;

If σ is unknown, we use "t-distribution"

$C.I = \bar{x} \pm t^* \frac{s}{\sqrt{n}}$ → standard Error.

t^* — t - critical Value

- Replace σ with Sample STD (s)
- The standard deviation $\sigma_{\bar{x}}$ is replaced with standard Error.

• How to Find the T^* (T Critical Value) using Ti-84.

When looking t^* critical value, you need 2 things:

• Left area from the level of confidence.

ie: a confidence of 95% will have left area of (0.975) ← right area 0.025

• Degrees of freedom (df): this is your sample size − 1

$$\boxed{DF = n - 1}$$

• Use t^* value table, find left-area + df to find t^* value

Always round down Df using t^* table.

• A t-distribution is symmetrical, bell-shaped, centered at 0. but has larger areas on the tails than the normal distribution.

• When sample size "n" gets larger, the (df) gets larger too and t-distribution will look closer to a normal distribution.

• ~~As df gets larger.~~ We can use T- interval to check

ex. Matching the 3 conditions:

① Assuming the population distribution is normal, since sample size ≥30)

② Assuming that ~~there~~ the population is 10 times larger than sample size.

③ Assuming the sample of 15 test tubes are selected at random.

Concluding Sentence:

• We're confident that the method we used to create our range of (1020, 1440) will capture the true population mean of how many times a test tube can be used 99% of the time.