Stats Medic Important Ideas for CED Unit 1: Exploring One-Variable Data

Analyzing Categorical Data (Activity: How Are Your Favorite Classes Related?)

- Identify the individuals and variables in a set of data.
- Classify variables as categorical or quantitative.
- Make and interpret bar graphs for categorical data.
- Identify what makes some graphs of categorical data misleading.
- Calculate marginal and joint relative frequencies from a two-way table.
- Calculate conditional relative frequencies from a two-way table.
- Use bar graphs to compare distributions of categorical data.
- Describe the nature of the association between two categorical variables.

Important Ideas:

LT#2
Categorical data: Report frequencies (counts) or relative frequencies (%) in bar graphs, pie charts, side-by-side bar graphs & segmented bar graphs.

LT#3
Misleading Graphs:
- Watch out for vertical axis. Should start at 0!
- Be careful of pictograph

LT#4 Two way table

Variable 2 \ Variable 1			Total
	A		B
Total			C

marginal rel. freq: B/C
joint rel. freq: A/C
Conditional rel freq: A/B
* Association b/w 2 variables means knowing one affects the other.

Mosaic Plot (Activity: What Will Be the Mascot?)

- Make and interpret bar graphs and mosaic plots for categorical data.
- Use bar graphs and mosaic plots to compare distributions of categorical data.
- Describe the nature of the association between two categorical variables.

Important Ideas:

LT#1 & LT#2 Displays for categorical data

BAR GRAPH: Each bar represents frequency or relative frequency for each category.

SEGMENTED BAR GRAPH: Stack up the bars to make 100%.

MOSAIC PLOT: Segmented bar graph where width of bars is proportional to group size.

LT#3 Association
If knowing the value of one variable helps us to predict the value of the other variable.
→ segmented bar graphs are different.

Displaying Quantitative Data (Activity: How Many Pairs of Shoes Do You Own?)

- Make and interpret dotplots, stemplots, and histograms of quantitative data.
- Identify the shape of a distribution from a graph.
- Describe the overall pattern (shape, center, and variability) of a distribution and identify any major departures from the pattern (outliers).

Important Ideas:

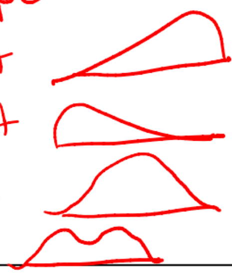
LT#2 Shape

skewed left

skewed right

symmetric

Bimodal:



LT#3 Describing Distributions

Shape *Use -ly words

Outliers

Center: sym → mean skewed/outlier → med.

Variability: how spread out the data is.

+ Context!

Describing Quantitative Data (Activity: How Many Colleges Are You Applying To?)

- Calculate measures of center (mean, median) for a distribution of quantitative data.
- Calculate and interpret measures of variability (range, standard deviation) for a distribution of quantitative data.
- Explain how outliers and skewness affect measures of center and variability.

Important Ideas:

$\boxed{\text{LT\#1}}$ Measures of center

Mean: average $\bar{x} = \frac{\Sigma x_i}{n}$

Median: middle value or the average of middle two values

$\boxed{\text{LT\#2}}$ Measures of variability

Range = max − min

$SD = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$ IQR → next lesson

"The context typically varies by SD from the mean of $\bar{x}$."

$\boxed{\text{LT\#3}}$

-Mean & SD are greatly affected by outliers (nonresistant)

* outliers or skewed dist. → use median & IQR

* symmetric → use mean & SD

Describing Quantitative Data (Activity: Where Do I Stand?)

- Identify outliers using the 1.5×IQR rule.
- Make and interpret boxplots of quantitative data.
- Use boxplots and numerical summaries to compare distributions of quantitative data.

Important Ideas:

LT#1 Outliers

1.5 IQR Rule

low outlier < $Q_1 - 1.5\,IQR$

high outlier > $Q_3 + 1.5\,IQR$

LT#2 Boxplots

5# Summary

min, Q_1, med, Q_3, max

Variable

LT#3 Comparing Distributions

Shape

Outlier - 1.5 IQR Rule

Center - mean, median

Variability - SD, IQR, range

+ Context

+ Comparative language

Percentiles and Cumulative Relative Frequency Graphs (Activity: Where Do I Stand?)

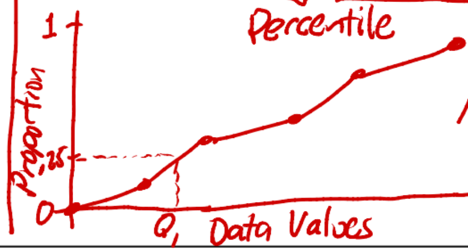
- Find and interpret the percentile of an individual value within a distribution of data.
- Estimate percentiles and individual values using a cumulative relative frequency graph.

Important Ideas:

$\boxed{\text{LT\#1}}$ Percentile: The percent of values less than or equal to a given value.

* "at", not "in"

$\boxed{\text{LT\#2}}$ Cumulative Relative Frequency Graph



$Q_1 = 25^{th}$ %ile
Med $= 50^{th}$ %ile
$Q_3 = 75^{th}$ %ile

z-scores and Transforming Data (Activity: How Did I Do?)

- Find and interpret the standardized score (z-score) of an individual value in a distribution of data.
- Describe the effect of adding, subtracting, multiplying by, or dividing by a constant on the shape, center, and variability of a distribution of data.

Important Ideas:

LT#1 Standardized values (z-scores)

$$z = \frac{\text{value} - \text{mean}}{\text{SD}}$$

"___context___ is ___z___ standard deviations ___above / below___ the mean."

LT#1 Given any distribution,

- Add / Subtract each value by a constant a
  → shape & variability stay same
  → Center shifts up / down by a
- Multiply / divide each value by a constant b
  → shape stays the same
  → center & variability ×/÷ by b.

Density Curves, 68-95-99.7 Rule (Activity: Exploring Density Curves)

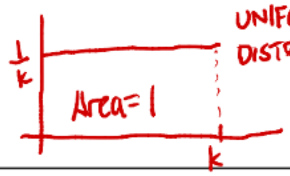
- Use a density curve to model distributions of quantitative data.
- Identify the relative locations of the mean and median of a distribution from a density curve.
- Use the 68-95-99.7 Rule to estimate the proportion of values in a specified interval.

Important Ideas:

$\boxed{\text{LT\#1}}$ Density curves

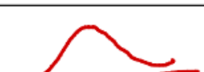
- Total area = 1

UNIFORM DISTRIBUTION



$\boxed{\text{LT\#2}}$

Symmetric



mean ≈ median

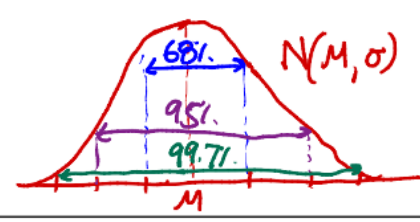
Skewed right

mean > median

Skewed left

mean < median

$\boxed{\text{LT\#3}}$ Empirical Rule



$N(\mu, \sigma)$

Normal Distribution Calculations (Activity: Will Marty Make it Back to the Future?)

- Find the proportion of values in a specified interval in a Normal distribution using Table A or technology.
- Find the value that corresponds to a given percentile in a Normal distribution using Table A or technology.

Important Ideas:

LT#1 Value → %

$$z = \frac{\text{Value} - \text{mean}}{\text{SD}}$$

z → Table A

Table A

z

% to left

LT#2 % → Value

① Find proportion in center of table A.

② match % to z-score

③ Plug in numbers & solve

$$z = \frac{\text{Value} - \text{mean}}{\text{SD}}$$

* Table A gives proportion to the LEFT.

Assessing Normality (Activity: Do We Have Normal Test Scores?)

- Determine whether a distribution of data is approximately Normal from graphical and numerical evidence.

Important Ideas:

To check for Normality

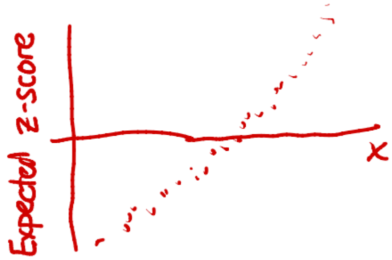
① GRAPH

② COMPARE MEAN & MEDIAN

③ 68-95-99.7 RULE

④ NORMAL PROBABILITY PLOT

Normal Probability Plot



If the Normal probability plot is fairly linear, then the distribution of X is approximately Normal.

Stats Medic Important Ideas for CED Unit 2: Exploring Two-Variable Data

Scatterplots (Activity: How Many Rubber Bands Does Barbie Need?)

- Distinguish between explanatory and response variables for quantitative data.
- Make a scatterplot to display the relationship between two quantitative variables.
- Describe the direction, form, and strength of a relationship displayed in a scatterplot and identify unusual features.

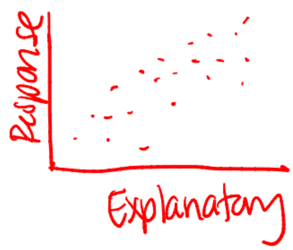
Important Ideas:

LT#1

Explanatory variable: used to predict

Response variable: outcome responds to explanatory

LT#2 & 3



Describing a relationship.

Direction - +/-/None

Unusual Features

Form (Linear/Non linear)

Strength

Correlation (Activity: How Safe is Barbie?)

- Interpret the correlation.
- Understand the basic properties of correlation, including how the correlation is influenced by unusual points.
- Distinguish correlation from causation.

Important Ideas:

$\boxed{\text{LT\#1}}$ Interpret r

Direction: (+/-)

Form: Always linear!

Strength:

-1 → strong negative
-0.5 → moderate
0 → weak
0.5 → moderate
+1 → strong positive

use "-ly" words

$\boxed{\text{LT\#2}}$ Properties of r

Unusual value in pattern
→ strengthens r (closer to 1 or -1)

Unusual value not in pattern
→ weakens r (closer to 0)

- r does not have units
- Changing units for x, y does not change r.

$\boxed{\text{LT\#3}}$

Correlation does <u>NOT</u> equal causation.

Regression Line, Predictions & Residuals (Activity: How Good are the Predictions for Barbie?)

- Make predictions using regression lines, keeping in mind the dangers of extrapolation.
- Calculate and interpret a residual.
- Interpret the slope and y intercept of a least-squares regression line.

Important Ideas:

LT#1 Predictions

$\hat{y} = a + bx$

predicted, y-int, slope

*Be careful of extrapolation.

LT#2 Residuals

Residual = Actual − Predict.

R = A − P

"The actual <u>context</u> was <u>residual</u> above/below the predicted value for <u>x=#</u>."

LT#3 y-int & slope

y-int: "When <u>x=0 context</u> the predicted <u>y-context</u> is <u>y-int</u>."

slope: "With each additional <u>x-context</u> the predicted <u>y-context</u> <u>increases/decreases</u> by <u>slope</u>."

Least Squares Regression & Residual Plots (Activity: How Many iPhones Will be Sold?)

- Determine the equation of a least-squares regression line using technology or computer output.
- Construct and interpret residual plots to assess whether a regression model is appropriate.

Important Ideas:

$\boxed{\text{LT\#1}}$ Equation LSRL

Applet: TWO QUANTITATIVE VARIABLES

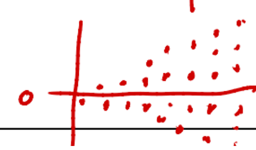
Calculator: → put values into lists
→ STAT/CALC/LinReg(a+bx)

$\boxed{\text{LT\#2}}$ Residual Plots

Good:

No leftover pattern

Bad:



Standard Deviation of Residuals & r-squared (Activity: Can You Guess My IQ?)

- Interpret the standard deviation of the residuals and r-sq and use these values to assess how well a least-squares regression line models the relationship between two variables.

Important Ideas:

LT#1 Standard deviation of the residuals (s)

Interpretation: "The actual <u>y-context</u> is typically about <u>s</u> away from number predicted by the LSRL."

LT#1 Coefficient of Determination

Interpretation: "About <u>r^2</u>% of the variability in <u>y-context</u> is accounted for by the LSRL."

Outliers for Scatterplots (Activity: How do Outliers Affect the LSRL?)

- Describe how the least-squares regression line, standard deviation of the residuals, and r2 are influenced by outliers.
- Find the slope and y-intercept of the LSRL from the means and standard deviations of x and y and their correlation.

Important Ideas:

Outliers: out of pattern (large residuals)

High leverage: very large or very small x-values.

Influential: if removed, big changes to slope, y-intercept, r

$\boxed{\text{LT\#1}}$ Outliers and the LSRL

Horizontal outliers → tilt the line

Vertical outliers → shift line up or down

* A good LSRL has low s & high r^2 (close to 1)

$\boxed{\text{LT\#2}}$ Formulas for LSRL ($\hat{y} = a + bx$)

$b = r\frac{s_y}{s_x}$ $\quad \bar{y} = a + b\bar{x} \rightarrow a = \bar{y} - b\bar{x}$

Transforming Non-linear Data (Activity: How Many iPhones Will be Sold?)

- Use transformations involving powers, roots, or logarithms to create a linear model that describes the relationship between two quantitative variables and use the model to make predictions.

Important Ideas:

LT#1

Function	Plot
Linear	x vs. y
Power	log x vs. log y
Exponential	x vs. log y

Predictions

- Plug in x & solve for y.

* may need to undo a root or log in order to get y *

Ex: $\cancel{10^{\log}}\, y = 10^{(a + bx)}$

ln ⟷ $e^{\wedge}$ $\quad y^3 \leftrightarrow \sqrt[3]{\ }$

Choosing the Best Regression (Activity: How Close to the Finish Line Can You Get?)

- Determine which of several models does a better job of describing the relationship between two quantitative variables.

Important Ideas:

$\boxed{\text{LT\#1}}$ Choosing the best regression model.

1. Check the scatterplot for a linear pattern.
2. Check the residual plot for no leftover pattern.
3. Check for the r^2 that is closest to 1.

Stats Medic Important Ideas for CED Unit 3: Collecting Data

Sampling Methods (Activity: Does Beyonce Write Her Own Lyrics?)

- Identify the population and sample in a statistical study.
- Identify voluntary response sampling and convenience sampling and explain how these sampling methods can lead to bias.
- Describe how to select a simple random sample using slips of paper, technology, or a table of random digits.

Important ideas:

LT #1,2

POPULATION

- Convenience sample — easy to reach
- Voluntary response — people choose
- SRS

can lead to bias

LT#3 Simple random sample (SRS)

① Label individuals → assign numbers → write names on slips of paper

② Randomize → random number generator (no repeats) → names in a hat (shuffle)

③ Select

More Sampling Methods Day 1 (Activity: How Much Do Fans Love Justin Timberlake?)

- Describe how to select a sample using stratified random sampling, cluster sampling, and systematic random sampling, and explain whether a particular sampling method is appropriate in a given situation.

Important Ideas:

- Simple Random Sample
- Choosing a group from the population so that every individual and group of individuals is equally likely to be chosen.

Steps: ① Label ② Randomize ③ Select

- Stratafied Random Sample: splits population into groups (strata) and chooses an SRS from each strata.
- Cluster sampling: splits population into groups based on location. Randomly select groups and sample everyone in the group (cluster).

More Sampling Methods Day 2 (Activity: How Much Do Fans Love Justin Timberlake? Day 2)

- Describe how to select a sample using stratified random sampling, cluster sampling, and systematic random sampling, and explain whether a particular sampling method is appropriate in a given situation.

POPULATION

Important Ideas: STRATIFIED RANDOM SAMPLE → SRS → SRS → SRS sample some from all groups	CLUSTER SAMPLE sample all from some groups	SYSTEMATIC RANDOM SAMPLE • Choose a random starting point. • Use equal intervals.

Problems with Sample Surveys (Activity: What is Wrong with These Surveys?)

- Explain how undercoverage, nonresponse, question wording, and other aspects of a sample survey can lead to bias.

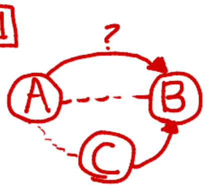
Important Ideas: LT #1 Undercoverage: when some members of a population cannot or are less to be included in a sample. Ex: Calling Landlines.	Nonresponse: When an individual is part of a sample but chooses not to respond or they cannot be reached. * This is different from Voluntary Response.	Response Bias: Pattern of inaccurate results. Ex: Wording of question, interviewer, lying, etc.

Observational Studies vs. Experiments (Activity: Does SAT Prep Produce Higher Scores?)

- Explain the concept of confounding and how it limits the ability to make cause-and-effect conclusions.
- Distinguish between an observational study and an experiment, and identify the explanatory and response variables in each type of study.
- Identify the experimental units and treatments in an experiment.

Important Ideas:

LT#1



A → explanatory variable: helps explain/predict response

B → response variable: outcome being measured

C → confounding variable: influences both the explanatory & response variables

LT#2 Observational study: no treatment

Experiment: treatments imposed, allow us to show causation

LT#3 Experimental units: what/who treatment is imposed on.

Treatments: what is done (or not done) to experimental units

Designing Experiments (Activity: Would You Fall For That?)

- Describe the placebo effect and the purpose of blinding in an experiment.
- Describe how to randomly assign treatments in an experiment using slips of paper, technology, or a table of random digits.
- Explain the purpose of comparison, random assignment, control, and replication in an experiment.

Important Ideas:

LT#1
- Control group: used to provide baseline data for comparison.
- Blinding: when subjects (single-blind) and/or experimenter (double-blind) who interact are unaware of what treatment is given.
- Placebo Effect: when a fake treatment (placebo) works.

LT#2 Conducting Random Assignment

① Label ② Randomize ③ Assign.

LT#3 4 key principles of Experiments

① Comparison → 2 or more treatments

② Random Assignment

③ Control: keep all other variables besides treatments constant.

④ Replication: using enough exp. units to distinguish differences.

Randomized Block Design (Activity: Does the Type of SAT Prep Matter?)

- Describe a completely randomized design for an experiment.
- Describe a randomized block design and a matched pairs design for an experiment and explain the purpose of blocking in an experiment.

Important Ideas:	
<u>Block</u> - group of experimental units that are similar <u>Randomized Block Design</u> Seperate subjects into blocks and then randomly assign treatments within each block.	<u>Matched Pairs Design</u> block of size 2 - subjects are paired and then randomly assigned to a treatment - each subject receives two treatments. *order of the treatments must be randomized.

Inference for Sampling & Experiments (Activity: What's in a Name?)

- Explain the concept of sampling variability when making an inference about a population and how sample size affects sampling variability.
- Explain the meaning of statistically significant in the context of an experiment and use simulation to determine if the results of an experiment are statistically significant.

Important Ideas:	
LT#1 Sampling Variability • Larger sample sizes decrease the variability of estimates. • Larger sample sizes make it easier to determine statistical significance.	LT#2 Statistical Significance: when results are too unusual to have occurred purely by chance. Simulation:  sample results prop of dots < 5% → statistically significant prop of dots > 5% → not stat. sig.

Score of Inference (Activity: Does SAT Prep Improve Scores? Part 2)

- Understand why a result can be extended to the population from which the sample was randomly selected.
- Given a description of a study with or without random assignment, determine whether there is evidence for a causal relationship.

Important Ideas:

LT#1 Population → Random sample

<u>Random sample</u> allows us to <u>generalize</u> our conclusions to the population from which we sampled.

LT#2 30 students → Random Assignment → Group 1 / Group 2

<u>Random assignment</u> allows us to say a treatment <u>causes</u> changes in the response variable.

Stats Medic Important Ideas for CED Unit 4: Probability, Random Variables and Probability Distributions

Randomness & Probability (Activity: How Good is Mrs. Gallas at Free Throws?)

- Interpret probability as a long-run relative frequency.

Important ideas: LT#1 Probability Long run relative frequency • It is always between 0 & 1. • Short term → unpredictable • Long term → predictable	Law of Large Numbers. If we do something many many times the proportion of desired outcomes will approach its probability.

Simulation (Activity: Are Soda Contests True?)

- Use simulation to model a random process in order to estimate a probability.

Important ideas: Simulation: imitation of chance behavior based on a model that accurately reflects the situation. Examples: dice, flip coin, applet, random number generator	Simulation process ① Describe how you will simulate one trial (one repetition) ② Perform many trials (repetitions) ③ Use the results to answer the question.

Probability Rules (Activity: The Last Banana)

- Give a probability model for a random process with equally likely outcomes and use it to find the probability of an event.
- Use basic probability rules, including the complement rule and the addition rule for mutually exclusive events.

Important ideas: Probability Model: List showing all possible outcomes and their probabilities. – must add to 1 – Each probability is between 0 & 1.	• Complement: Probability of an event NOT happening. $P(A^c) = 1 - P(A)$ • Mutually Exclusive: Events cannot occur together. • General Addition Rule: "OR" $P(A \text{ OR } B) = P(A) + P(B)$

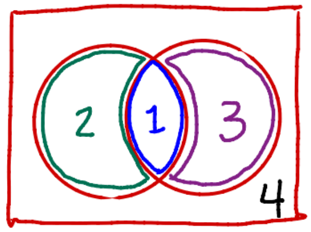
General Addition Rule (Activity: Can You Taco Tongue and Evil Eyebrow?)

- Use a two-way table or a Venn Diagram to model a random process and calculate probabilities involving two events.
- Apply the general addition rule to calculate probabilities.

Important Ideas:

$\boxed{\text{LT\#1}}$ TWO-WAY TABLE & VENN DIAGRAM

	B	B^c
A	1	2
A^c	3	4



$\boxed{\text{LT\#2}}$ General Addition Rule

$P(\underbrace{A \text{ or } B}_{A \cup B \text{ "union"}}) = P(A) + P(B) - P(\underbrace{A \text{ and } B}_{\text{"intersection" } A \cap B})$

If events A and B are mutually exclusive, they can't occur together;

$P(A \text{ and } B) = 0$, so $P(A \text{ or } B) = P(A) + P(B)$

Independent & Dependent Events (Activity: Can You Taco Tongue and Evil Eyebrow? Day 2)

- Use a two-way table or Venn diagram to model a random process and calculate probabilities involving two events.
- Calculate and interpret conditional probabilities.
- Determine whether two events are independent.

Important Ideas:

LT#2 Conditional Probability

$P(A \mid B) = \dfrac{P(A \text{ and } B)}{P(B)}$

"A given B"

LT#3 Independent: when knowing one event has occurred or has not occurred does not affect the probability of the second event.

If ...

$P(A) = P(A \mid B) = P(A \mid B^c)$

then A and B are independent.

Conditional Probability & Tree Diagrams (Activity: Can You Geta Pair of Aces or a Pair of Kings?)

- Use the general multiplication rule to calculate probabilities.
- Use a tree diagram to model a chance process involving a sequence of outcomes and to calculate probabilities.
- When appropriate, use the multiplication rule for independent events to calculate probabilities.

Important Ideas:

LT#1 General Multiplication Rule

$P(A \text{ and } B) = P(A) \times \underbrace{P(B|A)}_{\text{conditional probability}}$ "given that"

If A, B are independent

→ $P(B|A) = P(B)$, so

→ $P(A \text{ and } B) = P(A) \cdot P(B)$

LT#2 Tree Diagrams

$P(A)$ → A: $P(B|A)$ → B → $P(A \text{ and } B)$; $P(C|A)$ → C → $P(A \text{ and } C)$

$P(A^c)$ → A^c: $P(B|A^c)$ → B → $P(A^c \text{ and } B)$; $P(C|A^c)$ → C → $P(A^c \text{ and } C)$

Add to 1

LT#3 $P(\text{at least } 1) = 1 - P(\text{none})$

Discrete Random Variables (Activity: How Many Children Are in Your Family?)

- Use the probability distribution of a discrete random variable to calculate the probability of an event.
- Make a histogram to display the probability distribution of a discrete random variable and describe its shape.
- Calculate and interpret the mean (expected value) of a discrete random variable.

Important ideas:

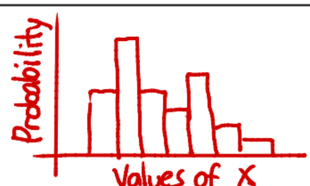
LT#1 Probability distribution

X	_	_	_	_	_	_
P(x)	_	_	_	_	_	_

Discrete random variable

Takes a fixed number of values with gaps between values.

LT#2 Histogram



LT#3 Mean (expected value)

$\mu_x = \sum x_i p_i$

Continuous Random Variables (Activity: How Much Do You Get Paid?)

- Calculate and interpret the standard deviation of a discrete random variable.
- Use the probability distribution of a continuous random variable (uniform or Normal) to calculate the probability of an event.

Important ideas:

LT#1 Standard Deviation of a discrete prob. dist.

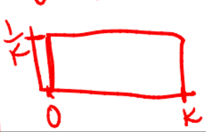
$\sigma = \sqrt{\sum (x_i - \mu)^2 \times p_i}$

$\sigma^2 = \text{variance}$

LT#2 Probability for continuous random variables

- find area under the curve.

Uniform



Normal:

$N(\mu, \sigma)$ $z = \dfrac{x - \mu}{\sigma}$

Table A or Normalcdf

Transforming Random Variables (Activity: Time for a Raise)

- Describe the effect of adding or subtracting a constant or multiplying or dividing by a constant on the probability distribution of a random variable.

Important ideas:	
Adding/subtracting a constant c	Multiplying/dividing by constant c
Shape: stays the same	Shape: stays the same
center: add/subtract c	center: multiply/divide by c
variability: stays the same	variability: multiply/divide by c

Combining Random Variables (Activity: What Will You Make Next Year?)

- Calculate the mean and standard deviation of the sum or difference of random variables
- Find probabilities involving the sum or difference of independent Normal random variables.

Important ideas:

LT#1 Adding & subtracting Random Variables X & Y

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

$\sigma_{x+y} = \sqrt{\sigma_x^2 + \sigma_y^2}$ | $\sigma_{x-y} = \sqrt{\sigma_x^2 + \sigma_y^2}$

*Be careful! Always add Variance

LT#2 Normal Probability Distributions

Find new mean & SD then find z-score.

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

Binomial Distributions (Activity: Is it Smart to Foul at the End of the Game?)

- Determine whether the conditions for a binomial setting are met.
- Calculate and interpret probabilities involving binomial distributions

Important ideas:

LT#1 Conditions for binomial setting

Binary: each trial is a success or failure
Independent: each trial is independent
Number of trials: is fixed (n=)
Same probability of success for each trial (p=)

LT#2 Binomial formula

$$P(X=k) = {}_{\underset{\text{number of trials}}{n}}C_{\underset{\text{number of successes}}{k}}\, \underset{\text{probability success}}{p}^{\overset{\text{number of successes}}{k}} \underset{\text{probability failure}}{(1-p)}^{\overset{\text{number of failures}}{n-k}}$$

STATS MEDIC

Binomial Distributions (Activity: Pop Quiz!)

- Determine whether the conditions for a binomial setting are met.
- Calculate the mean and standard deviation of a binomial random variable. Interpret these values.

Important ideas:

LT#1 Describing Binomial Dist.

BINS:

Shape: Make Histogram
Center: mean
Variability: Standard deviation

LT#2 Binomial Distributions

mean: $\mu_x = n \cdot p$ "After many many trials the average # of successes context is $\underline{\mu_x}$ out of $\underline{n}$."

SD: $\sigma_x = \sqrt{np(1-p)}$ "The number of successes typically varies by $\underline{\sigma_x}$ from the mean of $\underline{\mu_x}$ out of $\underline{n}$."

Binomial Distributions (Activity: Where Are All the Green Skittles?)

- Check the 10% condition to be able to assume independence of observations.
- Check the Large Counts condition to us the Normal approximation to the binomial distribution.

Important ideas:

$\boxed{\text{LT\#1}}$ 10% condition

When taking a random sample (without replacement) of size n from a population of size N we can use a binomial distribution if $n \leq .10N$

$\boxed{\text{LT\#2}}$ Large Counts condition

Use a Normal distribution to model a binomial distribution if $\underbrace{np \geq 10}_{\text{expected number of successes}}$ and $\underbrace{n(1-p) \geq 10}_{\text{expected number of failures}}$

Geometric Distributions (Activity: How Many Bottle Flips to go Viral?)

- Calculate and interpret probabilities involving geometric random variables.
- Calculate the mean and standard deviation of a geometric distribution. Interpret these values.

Important ideas:

Geometric Setting:

B: Binary
I: Independent
T: Trials until success
S: Same probability

Geometric Formula

# of failures

$$P(X=x) = \underbrace{(1-p)}_{p(\text{failure})}{}^{(x-1)} \cdot \underbrace{p}_{p(\text{success})}$$

Describing Geometric Distribution

Shape: Right skewed
Center: $\mu_x = 1/p$
Variability: $\sigma_x = \frac{\sqrt{1-p}}{p}$

Stats Medic Important Ideas for CED Unit 5: Sampling Distributions

Sampling Distributions (Activity: What was the Average for the Chapter 6 Test?)

- Distinguish between a parameter and a statistic.
- Create a sampling distribution using all possible samples from a small population.
- Determine if a statistic is an unbiased estimator of a population parameter.

Important ideas:

LT#1
Parameter: # describing a population
Statistic: # describing a sample

	mean	Prop.	SD
para.	μ	p	σ
stat.	$\bar{x}$	$\hat{p}$	s

LT#2 Sampling Distribution: Shows the statistic found in all possible samples of size n.

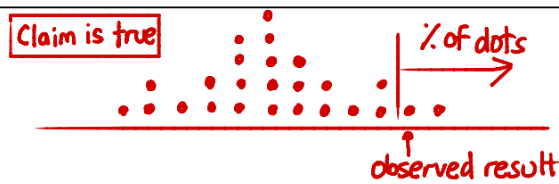
LT#3 A statistic is an unbiased estimator if the mean of sampling dist. is equal to parameter.
*when increasing sample size, the samp. dist. variability decreases.

Sampling Distributions (Activity: What was the Average for the Chapter 6 Test?)

- Use the sampling distribution of a statistic to evaluate a claim about a parameter.

Important ideas:

LT#1 Evaluating a Claim
① Assume the claim is true.
② Create simulated sampling distribution.
③ Find % chance of getting observed result.



If < 5% → convincing evidence against claim.
If ≥ 5% → not convincing evidence against claim.

Sample Proportions (Activity: What is the Proportion of Orange Reese's?)

- Calculate the mean and standard deviation of the sampling distribution of a sample proportion and interpret the standard deviation.
- Determine if the sampling distribution of a sample proportion is approximately Normal.
- If appropriate, use a Normal distribution to calculate probabilities involving a sample proportion.

Important ideas:

LT#1 mean & SD

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

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

*If the 10% condition is met

LT#2 Approx Normal
If Large Counts is met:
$n \cdot p \geq 10$
$n(1-p) \geq 10$

LT#3 If the sampling dist of $\hat{p}$ is approx. Normal

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

Difference Between Two Sample Proportions (Activity: Do Skittles or M&Ms Have More Orange?)

- Calculate the mean and standard deviation of the sampling distribution of a difference between sample proportions.
- If appropriate, use a Normal distribution to calculate probabilities involving a difference between two proportions.

Important ideas:

$\boxed{\text{LT\#1}}$ Sampling Distribution of $\hat{p}_1-\hat{p}_2$

SHAPE: Approximately Normal if

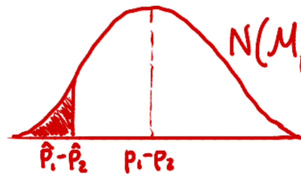
$n_1p_1 \geq 10$ $\quad n_2p_2 \geq 10$

$n_1(1-p_1) \geq 10$ $\quad n_2(1-p_2) \geq 10$

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

VARIABILITY: $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$

$\boxed{\text{LT\#2}}$ Normal Calculations



$N(\mu_{\hat{p}_1-\hat{p}_2}, \sigma_{\hat{p}_1-\hat{p}_2})$

$$z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$$

Sample Means (Activity: How Tall to be in the NBA? Part 1)

- Calculate the mean and standard deviation of the sampling distribution of a sample mean and interpret the standard deviation.
- If appropriate, use a Normal distribution to calculate probabilities involving sample means.

Important ideas:

LT#1 Sampling distribution of $\bar{x}$

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

Variability: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

LT#2

Shape: If the population is approx. Normal, the sampling distribution of $\bar{x}$ is also approx. normal.

Samp. dist. of $\bar{x}$

$N(\mu_{\bar{x}}, \frac{\sigma}{\sqrt{n}})$ $\quad z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Central Limit Theorem (Activity: How Tall to be in the NBA? Part 2)

- Explain how the shape of the sampling distribution of a sample mean is affected by the shape of the population distribution and the sample size.
- If appropriate, use a Normal distribution to calculate probabilities involving sample means.

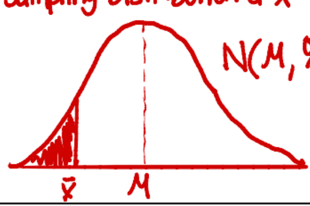
Important ideas:

$\boxed{\text{LT\#1}}$ Central Limit Theorem (CLT)

The sampling distribution of $\bar{x}$ is approximately Normal when the sample size is large ($n \geq 30$).

$\boxed{\text{LT\#2}}$ Normal Distribution Calculations

Sampling Distribution of $\bar{x}$



$N(\mu, \sigma/\sqrt{n})$ $\quad z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$\Rightarrow$ Table A

Difference Between Two Sample Means (Activity: ACT Scores – Which School is Better?)

- Calculate the mean and standard deviation of the sampling distribution of a difference between sample means.
- If appropriate, use a Normal distribution to calculate probabilities involving a difference between two means.

Important ideas:

LT#1 Sampling Distribution of $\bar{x}_1 - \bar{x}_2$

Shape: Check both sampling distributions

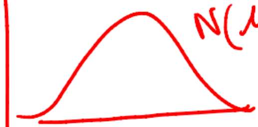
① Population is approx. Normal

OR ② $n \geq 30$, Central Limit Theorem

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

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

LT#2 Normal Calculations



$N(\mu_{\bar{x}_1 - \bar{x}_2}, \sigma_{\bar{x}_1 - \bar{x}_2})$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Stats Medic Important Ideas for CED Unit 6: Inference for Categorical Data: Proportions

What is a Confidence Interval? (Activity: Guess the Mystery Proportion)

- Identify an appropriate point estimator and calculate the value of a point estimate.
- Interpret a confidence interval in context.
- Determine the point estimate and margin of error from a confidence interval.
- Use a confidence interval to make a decision about the value of a parameter.

Important ideas:

LT#1 Point Estimate:
A statistic that provides a reasonable estimate about the population parameter.

Pt. Est → Parameter
$\hat{p} \rightarrow p$
$\bar{x} \rightarrow \mu$

LT#2 & 3
For interval (A, B)

$\text{Point Est.} = \frac{A+B}{2}$ $\text{margin of error} = \frac{B-A}{2}$

"We are ___% confident that interval from A to B captures the true parameter context."

LT#4 Decisions:
Confidence intervals contain a plausible values.

What is a Confidence Level? (Activity: What Does "95% Confidence" Mean?)

- Interpret a confidence level in context.
- Describe how the sample size and confidence level affect the margin of error.
- Explain how practical issues like nonresponse, undercoverage, and response bias can affect the interpretation of a confidence interval.

Important ideas:

$\boxed{\text{LT\#1}}$ Interpret confidence level

If we take many, many samples and calculate a confidence interval for each, about ___% of them will capture the true $\underbrace{p/\mu}_{\text{parameter}}$ (context)

$\boxed{\text{LT\#2}}$ Margin of error

confidence level ↑ ME ↑
(wider interval)

n ↑ ME ↓
(narrower interval)

$\boxed{\text{LT\#3}}$

Margin of error does account for sampling variability.

DOES NOT ACCOUNT FOR:
- nonresponse
- undercoverage
- response bias

Estimating a Population Proportion (Activity: Which Way Will the Hershey's Kiss Land?)

- State and check the Random, 10%, and Large Counts conditions for constructing a confidence interval for a population proportion.
- Determine the critical value for calculating a C% confidence interval for a population proportion using a table or technology.

Important ideas: LT#1 Conditions	LT#2 Critical Values	Formulas for CI for p
Random 10%: $n < \frac{1}{10}N$ Normal: Large Counts $n \cdot \hat{p} \geq 10$ $n(1-\hat{p}) \geq 10$	90% $z^* = 1.645$ 95% $z^* = 1.960$ 99% $z^* = 2.576$ To find any %, use InvNorm (tail %) InvNorm $\left(\frac{1 - C\%}{2}\right)$	Point Estimate ± margin of Error $\hat{p} \pm \underbrace{z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ $SE_{\hat{p}}$

Estimating a Population Proportion (Activity: What Proportion of the Earth is Covered by Water?)

- Construct and interpret a confidence interval for a population proportion.
- Determine the sample size required to obtain a C% confidence interval for a population proportion with a specified margin of error.

Important ideas: $\boxed{\text{LT\#1}}$ Four Step Process	$\boxed{\text{LT\#2}}$ Choosing a Sample Size
STATE: Parameter and confidence level PLAN: Name the procedure Check conditions DO: General and specific formulas Plug numbers in, calculate interval CONCLUDE: Interpret interval in context "We are 95% confident...."	$ME = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ Solve for n • When $\hat{p}$ is unknown, use $\hat{p}=0.5$ for conservative calculation. • If n has a decimal, always round up.

Confidence Intervals for a Difference in Proportions (Activity: Which Grade is More Likely to Go to Prom?)

- Determine whether the conditions are met for constructing a confidence interval about a difference between two proportions.
- Construct and interpret a confidence interval for a difference between two proportions.

Important ideas:	Convincing Evidence?
Confidence interval for $p_1 - p_2$ State $p_1 - p_2 \rightarrow$ true difference in proportions Plan Two Sample Z interval for $p_1 - p_2$ ① Independent Random samples ② 10% Condition ③ Large Counts Do PE ± MoE $\Rightarrow (\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ Conclude "We are 95% ...."	$(+,+) \rightarrow$ 1st proportion is greater $(-,-) \rightarrow$ 2nd proportion is greater $(-,+) \rightarrow$ No convincing evidence of a difference. Interval contains 0.

What is a Significance Test? (Activity: Is This Gender Discrimination?)

- State appropriate hypotheses for a significance test about a population parameter.
- Interpret a P-value in context.
- Make an appropriate conclusion for a significance test.

Important ideas:		
LT#1 Hypotheses NULL H_0: p / μ = null value ALTERNATIVE H_a: p / μ $<$ / $>$ / $\neq$ null value	LT#2 Interpret P-value Assuming H_0 is true ($\underline{\quad}_{H_0}$), there is a <u>P-value</u> probability of getting the observed result or more extreme, purely by chance.	LT#3 Conclusion Because P-value $<$ / $>$ α we do / do not have convincing evidence for Ha (context). P-value $< \alpha \Rightarrow$ Reject H_0 P-value $> \alpha \Rightarrow$ Fail to reject H_0

Tests About a Proportion (Activity: Are You Sure Mrs. Gallas Isn't a Good Free Throw Shooter?)

- State and check the Random, 10%, and Large Counts conditions for performing a significance test about a population proportion.
- Calculate the standardized test statistic and P-value for a test about a population proportion.

Important ideas:

LT#1 Conditions:

① Random:

② 10% $n < \frac{1}{10}N$

③ Normal: Large Counts

$n \cdot p \geq 10$

$n(1-p) \geq 10$

LT#2 Standardized Test Statistic (z-score) & P-value

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

$N(\mu_{\hat{p}}, \sigma_{\hat{p}})$

p

$z = \dfrac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \rightarrow$ P-value

P-value < 5% → Statistically sig. & have convincing evidence in H_a.

Tests About a Proportion (Activity: Can You Taste the Rainbow?)

- Perform a significance test about a population proportion (4-step)

Important ideas: $\boxed{\text{LT\#1}}$ 4-step process

STATE: parameter, statistic, hypotheses, α

PLAN: name procedure, check conditions

DO: General & specific formulas, picture, work, answer

CONCLUDE: Interpret P-value, conclusion

test statistic = $\dfrac{\text{statistic} - \text{parameter}}{\text{SD statistic}}$

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

* for a two-sided test ($H_a: p \neq \#$) double the P-value *

Tests about a Difference in Proportions Intro (Activity: Is Yawning Contagious?)

- State appropriate hypotheses for a significance test about a difference between two proportions.
- Determine whether the conditions are met for performing a test about a difference between two proportions.

Important ideas:

$\boxed{\text{LT\#1}}$ Hypotheses

$H_0: p_1 - p_2 = 0 \quad (p_1 = p_2)$

$H_a: p_1 - p_2 < 0 \quad (p_1 < p_2)$

$H_a: p_1 - p_2 > 0 \quad (p_1 > p_2)$

$H_a: p_1 - p_2 \neq 0 \quad (p_1 \neq p_2)$

* Define $p_1 - p_2$ with direction

$\boxed{\text{LT\#2}}$ Conditions

① Random: random samples (generalize to populations) or random assignment (show causation)

② 10% condition: only check when sampling without replacement. $n_1 \leq \frac{1}{10}N_1 \quad n_2 \leq \frac{1}{10}N_2$

③ Large Counts:

$n_1\hat{p}_c \geq 10, n_1(1-\hat{p}_c) \geq 10$

$n_2\hat{p}_c \geq 10, n_2(1-\hat{p}_c) \geq 10$

$\hat{p}_c = \dfrac{X_1 + X_2}{n_1 + n_2}$ ← p-combined

Significance Tests for a Difference in Proportions (Activity: Which Grade is More Likely to go to Prom?)

- Calculate the standardized test statistic and P-value for a test about a difference between two proportions.
- Perform a significance test about a difference between two proportions (4-step).

Important ideas:

STATE $p_1 - p_2 \rightarrow$ true difference in proportions

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \; >, <, \neq \; 0$ $\quad \alpha =$

PLAN Two-sample z-test for $p_1 - p_2$
Check conditions.

DO
$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$$

$$\hat{p}_c = \frac{x_1 + x_2}{n_1 + n_2}$$ *use for Large Counts*

CONCLUDE Interpret P-value
Decision + Conclusion

Type 1 and Type 2 Error (Activity: Should Rockford Switch to Bottled Water?)

- Interpret a Type I and a Type II error in context. Give a consequence of each error in a given setting.

Important ideas:

CONCLUSION / TRUTH	H_0 true	H_a true
Reject H_0	Type I	:)
Fail to reject H_0	:)	Type II

Type I Error H_0 true, Reject H_0 $\quad P(\text{Type I}) = \alpha$
"The null hypothesis (context) is true, but we find convincing evidence for Ha (context)."

Type II Error H_a true, Fail to reject H_0
"The alternative hypothesis (context) is true, but we don't find convincing evidence for Ha (context)."

Power (Activity: Will Mrs. Gallas Prove Herself?)

- Interpret the power of a significance test and describe what factors affect the power of a test.

Important ideas:

Decision / Truth	H_0 true	H_a true
Reject H_0	Type 1	Power
Fail to reject H_0	:)	Type 2

$P(\text{Type 1}) = \alpha$
$P(\text{Type 2}) = 1 - \text{Power}$

Power $P(\text{Reject } H_0 \mid H_a \text{ is true})$

"If Ha is true (at a specific value in context) there is a power probability of finding convincing evidence to reject the null (context)."

- To increase power: ↑ sample, ↑ alpha, ↑ distance to Ha

Stats Medic Important Ideas for CED Unit 7: Inference for Means

Estimating a Population Mean (Activity: How Much Does an Oreo Weigh?)

- Determine the critical value for calculating a C% confidence interval for a population mean using a table or technology.
- State and check the Random, 10%, and Normal/Large Sample conditions for constructing a confidence interval for a population mean.

Important ideas:	
LT#1 Critical values $t^* \rightarrow$ use for estimating means degrees of freedom (df) = n − 1 Table B → use df & confidence level to find t^* (always round df down if it is not on Table B)	LT#2 Conditions ① Random: random sample or random assignment ② 10%: $n < \frac{1}{10}N$ ③ Normal/Large Sample: • Population distribution is approximately Normal • $n \geq 30$, Central limit theorem • sample data shows no strong skew or outliers

Estimating a Population Mean (Activity: How Many States Can You Name?)

- Construct and interpret a confidence interval for a population mean.

Important ideas:	
LT#1 Construct & Interpret: State: Parameter & Confidence level Plan: One sample t interval for μ Conditions: Random: 10% Normal: ① Pop is approx. ② $n \geq 30$ CLT ③ Sample shows no strong skew or outliers.	Do Formulas, work, answer Pt. Est ± margin of error $\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$ df = n − 1 Conclude "We are ___% confident ..."

Confidence Interval for a Difference of Mean (Activity: Which Cookie Has the Most Chocolate Chips?)

- Determine whether the conditions are met for constructing a confidence interval for a difference between two means.
- Construct and interpret a confidence interval for a difference between two means.

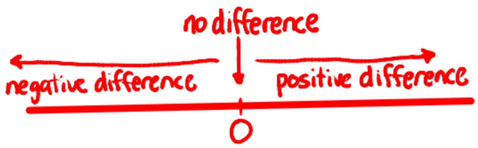
Important ideas:	
LT#1 Conditions Random: 10%: Normal/Large Counts: check for each population	LT#2 4-step process STATE: Parameter + confidence level PLAN: two sample t interval for $\mu_1 - \mu_2$ DO: General formula, specific formula, numbers, answer CONCLUDE: "We are ___% confident ..."

Confidence Intervals for a Mean Difference (Activity: Is Climate Change Real?)

- Analyze the distribution of differences in a paired data set using graphs and summary statistics.
- Construct and interpret a confidence interval for a mean difference.

Important ideas:

$\boxed{\text{LT\#1}}$ Distribution of differences



Statistics: $\bar{x}_{diff}$, s_{diff}

$\boxed{\text{LT\#2}}$ Confidence interval for paired data

Name: one sample t interval for μ_{diff}

Specific formula: $\bar{x}_{diff} \pm t^{*} \cdot \frac{s_{diff}}{\sqrt{n_{diff}}}$

Significance Test for a Mean (Activity: Are You Getting Enough Sleep?)

- State and check the Random, 10%, and Normal/Large Sample conditions for performing a significance test about a population mean.
- Calculate the standardized test statistic and P-value for a test about a population mean.
- Perform a significance test about a population mean.

Important ideas:

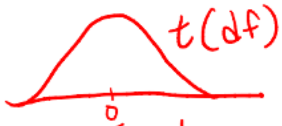
LT#1 Conditions

Random

10%

Normal: ① Pop. is approx normal ② n≥30 CLT ③ Sample shows no strong skew or outliers.

LT#2 Test Statistic & P-value



$t = \frac{\bar{x} - \mu}{s_x / \sqrt{n}}$

P-value

- Use table B

or

- tcdf (lower t, upper t, df)

Significance Test for a Mean (Activity: What is Normal Body Temperature?)

- Perform a significance test about a population mean. (4-step)
- Understand the connection between confidence intervals and significance tests.

Important ideas:

$\boxed{\text{LT\#1}}$ 4-step

STATE: Parameter, hypotheses, statistic, α

PLAN: Name procedure, check conditions

① Random ② 10% condition ③ Normal/Large sample

DO: General, specific, work, test statistic, P-value

CONCLUDE: Interpret P-value

Decision, conclusion about H_a (context)

$\boxed{\text{LT\#2}}$ Confidence Interval + Two-sided tests

If H_0 value in interval → H_0 value plausible, Fail to reject H_0

If H_0 value not in interval → H_0 value not plausible, Reject H_0

A C% confidence interval will make the same decision as a two-sided significance test with $\alpha = 1 - C\%$ level.

Significance Test for a Difference of Means (Activity: Is One Form of the AP Exam Harder?)

- State appropriate hypotheses for a significance test about a difference between two means.
- Determine whether the conditions are met for performing a test about a difference between means.

Important ideas:

LT#1 Hypotheses

$H_0: \mu_1 - \mu_2 = 0 \quad (\mu_1 = \mu_2)$

$H_a: \mu_1 - \mu_2 < 0 \quad (\mu_1 < \mu_2)$

$H_a: \mu_1 - \mu_2 > 0 \quad (\mu_1 > \mu_2)$

$H_a: \mu_1 - \mu_2 \neq 0 \quad (\mu_1 \neq \mu_2)$

* Define $\mu_1 - \mu_2$ with direction

LT#2 Conditions

① Random: random samples (generalize to populations) or random assignment (show causation)

② 10% condition: only check when sampling without replacement. $n_1 \leq \frac{1}{10}N_1 \quad n_2 \leq \frac{1}{10}N_2$

③ Normal/Large Sample
- both populations ≈ Normal
- $n_1 \geq 30$ and $n_2 \geq 30$ CLT
- sample data shows no strong skew or outliers (×2)

Significance Test for a Difference of Means (Activity: Does Labeling Menus Reduce Calories?)

- Perform a significance test about a difference between two means. (4-step)

Important ideas:

LT#1-4

State $\mu_1 - \mu_2 \Rightarrow$ true difference in means *Order!

Plan Two sample t test for $\mu_1 - \mu_2$
① Independent random samples OR random assignment
② 10% if sampling
③ Normal - Check for both samples

Do $$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

t(df)

Conclude P-value < α Reject the null & have convincing evidence

*Context!

Significance Test for a Mean of Differences (Activity: Climate Change Part 2)

- Perform a significance test about a mean difference.
- Determine when it is appropriate to use paired *t* procedures versus two-sample *t* procedures.

Important ideas:

LT#1

one sample t test for μ_{diff}
= one sample t test for μ

$$t = \frac{\bar{x}_{diff} - \mu_{diff}}{s_{diff}/\sqrt{n}}$$

* Use T-test on calculator

LT#2

Paired t
- one sample (data are paired)
- mean of differences (μ_{diff})
- subtract, then average

Two sample t
- two independent samples (or groups)
- difference of means ($\mu_1 - \mu_2$)
- average, then subtract

Stats Medic Important Ideas for CED Unit 8: Inference for Categorical Data: Chi-Square

Chi-Square Goodness of Fit (Activity: Does Harvard Discriminate Against Asian Applicants? Part 1)

- State appropriate hypotheses and compute the expected counts and chi-square test statistic for a chi-square test for goodness of fit.
- State and check the Random, 10%, and Large Counts conditions for performing a chi-square test for goodness of fit.
- Calculate the degrees of freedom and P-value for a chi-square test for goodness of fit.

Important ideas:

$\boxed{\text{LT\#1}}$ Hypotheses

H_0: The claimed distribution (context) is true.

H_a: The claimed distribution (context) is not true.

Expected Count = np

$\boxed{\text{LT\#2}}$ Test statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

* larger χ^2 = more convincing evidence for Ha

$\boxed{\text{LT\#3}}$ df + P-value

df = k - 1 (k = # of categories)

P-value → Table C

Chi-Square Goodness of Fit (Activity: Does Harvard Discriminate Against Asian Applicants? Part 2)

- Perform a chi-square test for goodness of fit.
- Conduct a follow-up analysis when the results of a chi-square test are statistically significant.

Important ideas:

$\boxed{\text{LT\#1}}$ Conditions

① Random

② 10%

③ Large Counts: All expected counts ≥ 5.

$\boxed{\text{LT\#2}}$ χ^2 distribution

* right skewed
* starts at 0

$\boxed{\text{LT\#3}}$ Follow up Analysis

If your test is statistically significant, find the largest component of χ^2 and explain.

Chi-Square Test of Homogeneity (Activity: Does Gummy Bear Brand Matter?)

- State appropriate hypotheses and compute the expected counts and chi-square test statistic for a chi-square test based on data in a two-way table.
- State and check the Random, 10%, and Large Counts conditions for a chi-square test based on data in a two-way table.
- Calculate the degrees of freedom and P-value for a chi-square test based on data in a two-way table.
- Perform a chi-square test for homogeneity.

Important ideas:		
LT#1 Hypotheses: H_0: There is no difference in cat. variable distribution for population 1 & population 2. H_a: There is a difference in cat. variable distribution for population 1 & population 2 Expected Counts	LT#2 Conditions -Random -10% -Large Counts Expected Counts ≥ 5 LT#3 χ^2 & P-value df = (Rows-1)(Columns-1) χ^2cdf(lower, upper, df)	LT#4 χ^2 for Homogeneity χ^2 GOF → 1 sample 1 variable χ^2 for Homogeneity → 2 samples 1 variable 4 Step: State, Plan, Do Conclude.

$$\text{Exp. Count} = \frac{(\text{Row Total})(\text{Column Total})}{\text{Table Total}}$$

Chi-Square Test for Independence (Activity: Are Taco Tongue and Evil Eyebrow Independent?)

- Perform a chi-square test for independence.
- Choose the appropriate chi-square test in a given setting.

Important ideas:	
$\boxed{\text{LT\#1}}$ χ^2 test for independence H_0: There is not an association between ______ & ______. H_a: There is an association between ______ & ______.	$\boxed{\text{LT\#2}}$ χ^2 GOF 1 sample, 1 variable χ^2 test for homogeneity 2+ samples, 1 variable χ^2 test for independence 1 sample, 2 variables

Stats Medic Important Ideas for CED Unit 9: Inference for Quantitative Data: Slopes

Sampling Distribution of Slopes (Activity: Does Seat Location Matter? Part 1)

- Identify statistics and parameters for linear regression.
- Use the sampling distribution of slopes to evaluate a claim.

Important ideas:

POPULATION
$\mu_y = \alpha + \beta x$

sample
$\hat{y} = a + bx$

$\boxed{\text{LT\#1}}$

	Statistic	Parameter
y-intercept	a	α
Slope	b	β
SD of residuals	s	σ
SD slope	SE_b	σ_b

$\boxed{\text{LT\#2}}$

Sampling Distribution of b

Confidence Intervals for Slope (Activity: Does Seat Location Matter? Part 2)

- Construct and interpret a confidence interval for the slope of the population (true) regression line. (4-step)

Important ideas:

$\boxed{\text{LT\#1}}$ CI for slope (4-step)

$\boxed{\text{STATE:}}$

$\beta \rightarrow$ true slope of the population LSRL for _x_ (context) and y (context).

___% confidence level

$\boxed{\text{PLAN:}}$

one sample t interval for slope

Linear:
Independent:
Normal:
Equal SD:
Random:

} See previous page

$\boxed{\text{DO:}}$

POINT ESTIMATE $\pm$ MARGIN OF ERROR

$b \pm t^* \cdot SE_b$

df = n-2

SE_b from computer output

$\boxed{\text{CONCLUDE}}$

Significance Tests for Slope (Activity: How Does GPA Relate to ACT Score?)

- Perform a significance test about the slope of the population (true) regression line. (4-step)

Important ideas:

$\boxed{\text{STATE:}}$

$\beta \rightarrow$ true slope of population LSRL for _x_ and _y_

$H_0: \beta = 0$ No linear relationship

$H_a: \beta < 0$ Negative

$\beta > 0$ Positive $\alpha =$ ____

$\beta \neq 0$ Something

$\boxed{\text{PLAN:}}$

one sample t test for slope

L
I
N
E
R

} see previous lesson

$\boxed{\text{DO:}}$

$$\text{STANDARDIZED TEST STATISTIC} = \frac{\text{STATISTIC} - \text{PARAMETER}}{\text{SD}}$$

$$t = \frac{b - \beta}{SE_b} \qquad df = n - 2$$

$\boxed{\text{CONCLUDE:}}$

- Interpret P-value
- Decision + conclusion