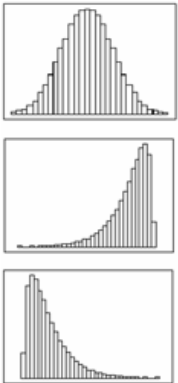


<p><i>In General:</i> B.S.: Be Specific; include numbers where possible and refer back to the context of the problem.</p> <p>Define what you're talking about and show your work</p> <p>Read the whole question first. And ask: "what is this question about?"</p> <p>Check your answer when you're done to ensure it makes sense.</p>	<p><i>Describing/Comparing Distributions:</i> C: Center-the median or mean U: Unusual features-gaps or outliers S: Spread- interquartile range (IQR), standard deviation, or range S: Shape-symmetrical, bimodal, or skewed</p> <div data-bbox="638 483 955 863">  <p>Symmetric Bell shaped</p> <p>Skewed to the Left</p> <p>Skewed to the Right</p> </div>	<p><i>Binomial Distribution:</i> B: Binary-either success or failure I: Independent trials N: Number of trials is fixed S: Success probability stays the same</p> <p>Defining a Binomial Distribution: X has a binomial distribution with $n = \#$ and $p = \#$</p> <p>NSpire: Menu, 6: Stats, 5: Distributions, A or B</p>	<p><i>Geometric distribution:</i> B: Binary-either success or failure I: Independent trials T: Trials until success S: Success probability stays the same</p> <p>Defining a Geometric Distribution: X has a geometric distribution with $p = \#$</p> <p>NSpire: Menu, 6: Stats, 5: Distributions, H or I</p>
<p><i>Constructing a Confidence Interval:</i> P: Define parameters, including confidence level A: State assumptions/conditions N: Name the interval I: Calculate the interval C: State the conclusion in context</p>	<p><i>Performing a Hypothesis Test:</i> P: Define parameters, including significance level H: Write your hypotheses A: State assumptions/conditions N: Name the test T: Calculate the test statistic O: Obtain the p-value M: Make a decision S: State the conclusion in context</p>	<p><i>Describing a Relationship:</i> STD uses correlation coefficient, r S: Strength-strong, moderate, or weak T: Trend- linear or nonlinear D: Direction-positive or negative</p>	<p><i>Describing a Scatterplot:</i> D: Direction- positive or negative U: Unusual features-outliers F: Form- linear or nonlinear S: Strength- strong, moderate, or weak</p>

Sentence Stems: Replace what is underlined

Interpret slope: As the x-variable increases by 1 unit, the predicted y-variable increases/decreases by slope units.

Interpret y-intercept: When the x-variable is 0 units, the predicted y-variable is y-intercept units. IF THIS NUMBER DOES NOT MAKE SENSE (i.e. is negative), YOU MUST COMMENT ON THAT.

Describe the relationship: There is a strength, trend, direction relationship between x-variable and y-variable based on the graph or correlation coefficient r.

Coefficient of Determination: R²% of the variability in the y-variable can be explained by a linear relationship with x-variable.

Interpret the mean: If many, many context are randomly selected, the average context will be about mean value.

Interpret standard deviation: The context typically vary from the mean value by about standard deviation value.

Interpret Confidence Interval (Conclusion): We are confidence level% confident that the true parameter in context is between lower bound and upper bound in context.

Interpret Confidence Level: If we constructed many, many confidence intervals from random samples of size n, about confidence level% of the intervals would capture the true population parameter in context.

Interpret the p-value: Assuming the H₀ in context is true, there is a p-value% chance of getting a sample proportion/mean of sample value or something more extreme by chance in random samples of size n.

$p > \alpha$ Fail to Reject H₀: We do not have convincing evidence for H_a in context.

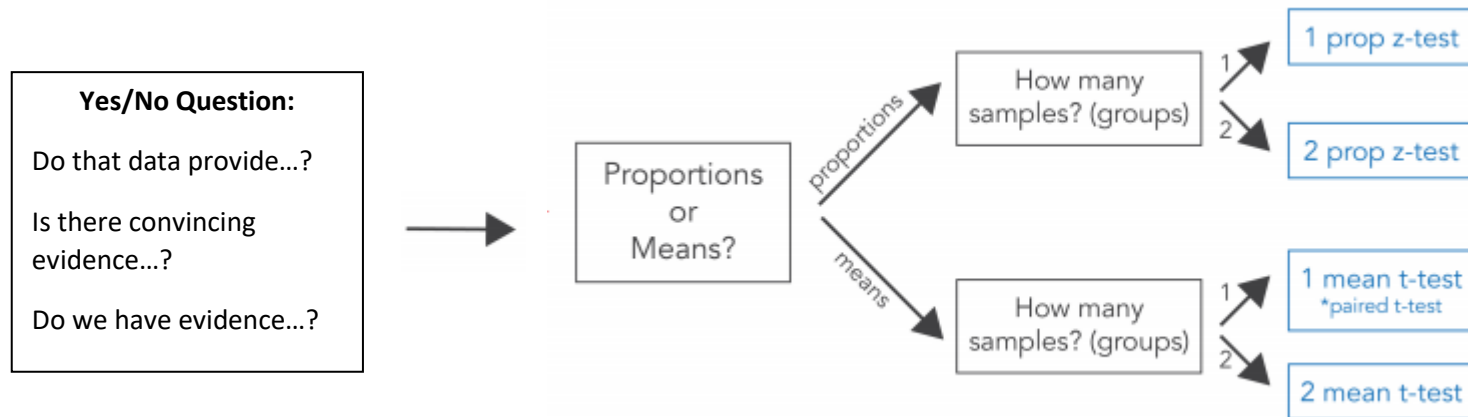
$p \leq \alpha$ Reject H₀: We do have convincing evidence for H_a in context.

Interpret the power: If the true population parameter in context is H_a, there is a power% probability of finding convincing evidence to reject H₀ in context.

Inference Table:

Name	Assumptions	Interval	Test	Hypothesis
1 Proportion z	<ul style="list-style-type: none"> Random: sample must be randomly selected or randomly assigned Independence: $n < \frac{1}{10}N$ Normality: stated OR $np \geq 10$ $n(1 - p) \geq 10$ 	$CI = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$ <p>NSpire: Menu, 6: Stats, 6: CI, 5: 1-Prop Z Interval</p>	$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$ <p>NSpire: Menu, 6: Stats, 7: Stats Tests, 5: 1-Prop Z Test</p>	$H_0: p =$
2 Proportion z	<p>MUST CHECK FOR BOTH SAMPLES</p> <ul style="list-style-type: none"> Random: sample must be randomly selected or randomly assigned Independence: $n < \frac{1}{10}N$ Normality: stated OR $np \geq 10$ $n(1 - p) \geq 10$ 	$CI = (\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$ <p>NSpire: Menu, 6: Stats, 6: CI, 6: 2-Prop Z Interval</p>	$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$ <p>NSpire: Menu, 6: Stats, 7: Stats Tests, 6: 2-Prop Z Test</p>	$H_0: p_1 = p_2$ OR $H_0: p_1 - p_2 = 0$
1 Sample t	<ul style="list-style-type: none"> Random: sample must be randomly selected or randomly assigned Independence: $n < \frac{1}{10}N$ Normality: stated OR graph has no unusual features OR $n \geq 30$ 	$CI = \bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$ <p>NSpire: Menu, 6: Stats, 6: CI, 2: t Interval</p>	$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ <p>NSpire: Menu, 6: Stats, 7: Stats Tests, 2: t Test</p>	$H_0: \mu =$
2 Sample t	<p>MUST CHECK FOR BOTH SAMPLES</p> <ul style="list-style-type: none"> Random: sample must be randomly selected or randomly assigned Independence: $n < \frac{1}{10}N$ Normality: stated OR graph has no unusual features OR $n \geq 30$ 	$CI = (\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{(S_1)^2}{n_1} + \frac{(S_2)^2}{n_2}}$ <p>NSpire: Menu, 6: Stats, 6: CI, 4: 2 Sample Z Interval</p>	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ <p>NSpire: Menu, 6: Stats, 7: Stats Tests, 4: 2-Samp t Test</p>	$H_0: \mu_1 = \mu_2$ OR $H_0: \mu_1 - \mu_2 = 0$

Is it a Test? /What Type of Test?



Proportions vs. Means	
Proportions: Categorical Data	Means: Quantitative Data
"proportion"	"mean" or "average"
Yes/No Answers	Numerical Answers
Data given as fractions, includes phrase "out of"	Data points

*Note: A question may only ask you to state the conclusion of a test. Look out for a significance level α for this.

Random Formulas not on the Formula Sheet:

$$\text{Interquartile range (IQR)} = Q_3 - Q_1$$

$$\text{Low/Left Outliers} = Q_1 - 1.5(\text{IQR})$$

$$\text{High/Right Outliers} = Q_3 + 1.5(\text{IQR})$$

$$z\text{-score} = \frac{\text{value} - \text{mean}}{\text{standard deviation}} = \frac{x - \mu}{\sigma}$$

$$\text{Combining Means of Random Variables: } \mu_{x \pm y} = \mu_x \pm \mu_y$$

$$\text{Combining Standard Deviations of Random Variables: } \sigma_{x \pm y} = \sqrt{\sigma_x^2 \pm \sigma_y^2}$$

$$\text{Transforming the Mean of Random Variables: } \mu_{a+bx} = a + b\mu_x$$

$$\text{Transforming the Standard Deviation of Random Variables: } \sigma_{a+bx} = |b|\sigma$$

$$\text{Residual} = \text{actual} - \text{predicted} = y - \hat{y}$$

$$\text{"Given"/Conditional Probability: } P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\text{Least Squares regression line: } LSRL = \hat{y} = mx + b = a + bx$$

$$\text{Degrees of freedom (d.f.)} = n - 1$$

To check if events are independent:

Check the formula: $P(A) = P(A|B)$

- If they are equal, they are independent.
- If they are NOT equal, they are NOT independent.

If A and B are independent events:

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

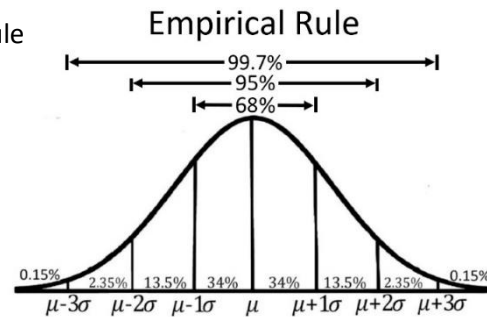
$\cup = \text{OR}$

$\cap = \text{AND}$

General Things to Remember:

- **Labeling a graph:** Title, Label x-axis, Label y-axis, Scale x-axis, Scale y-axis
- PDF vs CDF
 - **PDF:** used when looking for the probability of one specific number; uses equal signs; $P(X=x)$
 - **CDF:** used when looking for the probability of a range of numbers; uses inequalities; $P(X>x)$

- Percentiles are read from left to right and represent the % of values that are less than or equal to a given value.
- Empirical Rule/68-95-99.7% Rule



- Type I and Type II Errors

	H_0 rejected	Fail to reject H_0
H_0 false	Correct	Type II error
H_0 true	Type I error	correct

Alpha (α) = Prob (Type I error)

Beta (β) = Prob (Type II error)

Power = $1 - \beta$

- Use normalcdf to find the probability/percent; use invNorm to find the number given the percent/probability
 - Make sure to draw a curve with labels to show work.

When a problem uses a normal distribution but wants you to find the probability given a sample size n or says "sample proportion/mean": it is a sampling distribution.

Use the correct mean and standard deviation!

The Power of a Test – is the probability that the test will reject the null hypothesis when the null hypothesis is false assuming the null is true.

If you increase	Type I error	Type II error	Power
α	Increases	Decreases	Increases
n	Same	Decreases	Increases
$(\mu_0 - \mu_a)$	Same	Decreases	Increases

Type of Graph	Pie Graph	Bar Graph	Segmented Bar Graph	Histogram
Picture				
Type of Graph	Stemplot/Stem-and-leaf plot	Boxplot/Box-and-whisker plot	Dotplot	Ogive/Cumulative Frequency
Picture	<p>Stemplot of Data Set</p> <pre> 0 4 6 1 2 4 8 2 3 3 4 4 5 5 7 8 4 2 2 5 5 0 1 8 6 8 7 2 </pre> <p>Key: 1 0 = 10</p>			

Labeled Computer Output:

LSRL Equation: $\hat{y} = 2.544 + 0.164(\text{caffeine})$

Predictor	Coef	SE Coef	T	P
Constant	2.544	0.134	18.955	0.000
Caffeine (mg)	0.164	0.057	2.862	0.005

$S = 1.532$ $R\text{-Sq} = 60.032\%$ $R\text{-Sq}(\text{adj}) = 58.621\%$

$r = \sqrt{r^2} = \sqrt{0.60032} = 0.774803$

r is positive since the slope is positive

Sampling Method	Description	Pros	Cons
Simple Random (SRS)	One chooses the sample so that each unit (and every set of units) has an equal chance of being selected <u>Examples:</u> names in a hat, random number generator/table	Easy and unbiased	Large variation, and must have a known population
Stratified (Random)	Divide the population into groups (strata) based on a similar characteristic, then use an SRS to choose from EACH group	More precise than an SRS and can be cheaper if the groups are already available	Difficult to divide into groups, more complex than SRS, and must have known population
Cluster	Divide the population into groups (usually by location), randomly select a group and sample everything in THAT group	Cost is reduced, is unbiased, and don't need to know entire population	Sample may not be representative of overall population
Systematic	Use a system (every n th number) after choosing randomly where to begin	Unbiased, the sample is evenly distributed across the population, and don't need to know entire population	Large variation and can be affected by trends
Voluntary	Sample is selected in a way that people do not have to respond	Easy	Highly unrepresentative of overall population
Convenience	Sample people who are easy or comfortable to collect information from	Easy	Highly unrepresentative of overall population

Bias in Sampling	Definition
Voluntary Response	People choose themselves to participate
Undercoverage	Some groups are left out of the sample selection process
Non-response	Someone cannot or does not participate in sample
Response	False/incorrect answers (can be intentional or not)
Wording of Questions	Question is worded so that a certain response is given

General Vocab	Definition
Observational Study	Treatment IS NOT randomly assigned
Experiment	Treatment IS randomly assigned
Blocking	Reduces variation; subjects are grouped based on a characteristic, then treatments are randomly assigned within the groups