AP STATISTICS STUDY GUIDE

created by shayla huynh (*note: not everything from the course has been added yet*)

contact me if you have any questions

EMAIL FACEBOOK TWITTER INSTAGRAM

HEY FOLKS. [HERE'S THE ORG POST WITH SOME INFO ABOUT THE AP TEST AND MORE RESOURCES]
IF YOU HAVE ANY QUESTIONS, OR WANT TO SEE THE NOTES I TOOK AND/OR EXAMPLES OF PROBLEMS, CONTACT ME
GOOD LUCK ON THE TEST. DRINK WATER AND SLEEP WELL.

SOCS		Use to describe a distribution	* A statistic is *robust* if it resists extreme values (When there are outliers use: median & IQR instead of mean & range)	
	Item	**About**	**Explanation**	**Sentence Structure**
S	Shape	**SHAPE** • Left skewed ○ Most of the data is on the right • Right skewed ○ Most of the data is on the left • Roughly symmetrical **PEAKS** • Unimodal • Bimodal • Multimodal		The distribution of [title] is [shape] is [mode] with a peak at [value of peak(s)]. There is [# of outliers] at [value of outlier(s)] because the data point .... The center of distribution is best described by the [mean/median], [statistic value]. The spread of distribution is best described by the [range/I.Q.R.], [statistic value]. **EXAMPLE** The distribution of sibling data in AP Statistics is right skewed with a peak at one sibling. There is 1 outlier at 5 siblings because the data point exceeds the upper outlier fence of 3.5 siblings. The center of distribution is best described by the median, 1.5 siblings. The spread of distribution is best described by the interquartile range, 1 sibling. **NOTES** • Points are given on the AP test for mentioning whether or not the distribution is linear. • Always provide numbers to support your claim whenever
O	Outliers	Are there any? If so, use robust statistics. Interquartile Range • $Q_3 - Q_1$ Upper Outlier Fence • $Q_3 + (1.5)(I.Q.R.)$ Lower Outlier Fence • $Q_1 - (1.5)(I.Q.R.)$ **CALCULATOR** Insert Numbers into a list • *1-Var Stats*	There are three ways to talk about outliers. No Outliers • There are no outliers. The entire data set is within the L.O.F. of [value] and U.O.F. of [value] Upper Outlier Fence • There is/are [# of outliers] at [value of outlier(s)] because the data point exceeds the upper outlier fence of [value] Lower Outlier Fence • There is/are [# of outliers] at [value of outlier(s)] because the data point is below the lower outlier fence of [value]	
C	Center	Would the mean or median be a better	If there are outliers, use median.	

		description?		possible.
S	Spread	Would the standard deviation or I.Q.R. be the better description?	If there are outliers, use I.Q.R.	

INTERPRETING

Item	Variable	Explanation	Example
Slope	b	For every 1 [explanatory unit] of [explanatory variable], [response variable] increases by [b value + units].	B ≈ 0.385 For every 1 additional year of subject's age, their glucose level increases by 0.385 units.
Y-intercept	a	When the [explanatory variable] is zero [explanatory units], our model predicts that [response variable] will be [y-intercept value] [response units]. **NOTE:** Is this an extrapolation in the data set?	a ≈ 65.141 When the subject's age is zero years, our model predicts that glucose level will be 65.141 unit
Correlation Coefficient	r	[Explanatory Variable] and [Response Variable] have a [strength] [direction] [linear/nonlinear] relationship. **NOTE:** • **Strength** ○ Weak: |r| = 0.0 - 0.3 ○ Moderate: |r| = 0.4 - 0.6 ○ Strong: |r| = 0.7/+ • **Direction (Depends on Slope)** ○ Positive = Increasing ○ Negative = Decreasing	SLOPE = Increasing; r ≈ 0.52 Subject's age and glucose level have a positive moderate linear relationship.
Coefficient of Determination	r^2	[r^2 converted to a percentage] of the variability in [response variable] is explained by [explanatory variable].	r^2 ≈ 0.2806 28.06% of the variability in glucose level is explained by subject's age.

Standard Deviation of Residuals	s	The average prediction error is [s value].	S ≈ 10.861 The average prediction error is approximately 10.861

SAMPLING STRATEGIES

TYPE	NAME	ABOUT	
VALID	Simple Random Sample (SRS)	Every possible sample of a given size can be chosen	
VALID	Stratified Random Sample	Divide population into stratas then perform a SRS in each strata proportionally	
VALID	Cluster Sample	Divide population into clusters then randomly select a cluster. All individuals in the cluster are sampled	
VALID	Systematic Random Sample	Choose a random location, take every n^{th} person to be in the sample	
INVALID	Voluntary Response Sample	Respondents choose themselves	These are invalid methods of collecting a sample from a population because it may not represent the entire population accurately. The proportion may be under or overestimated. ------------------------ On the AP test, be able to explain *in context* why a certain strategy would be better than another or *in context* why an invalid strategy is flawed.
INVALID	Convenience Sample	Respondents are easy to sample	
INVALID	Undercoverage Sample	Missing members of the population	
INVALID	Nonresponse bias	Respondents refuse to answer	
INVALID	Response bias	Respondents are “lead” to answer or respond correctly	

INFERENCES

Central Limit Theorem (C.L.T.) - if a population distribution is not normal, the C.L.T. states that when n is large, the sampling distribution of x bar is normal.

IMPORTANT TIP: A lot of tests have the same structure/set-up to them. (I.E. Anything proportions have the same conditions, etc.)

NAME	STATE	CONDITION	DO	CONCLUDE
1 Proportion z Interval	**DESCRIPTION** Confidence level Context	**Random:** Data came from random sample **Normal:** • $n\hat{p} \geq 10$	**EQUATIONS**	I am __% confident that the *population proportion* of [context] is

AP STATISTICS STUDY GUIDE

created by shayla huynh (*note: not everything from the course has been added yet*)

contact me if you have any questions

EMAIL FACEBOOK TWITTER INSTAGRAM

<table>
<tr>
<td></td>
<td></td>
<td>• $n(1-\hat{p}) \geq 10$
Independent:
Either
• Individual observations are independent when sampling without replacement
• Sample is less than 10% of its respective population ($n \leq 0.10N$)</td>
<td>$z^* = \text{invNorm}[\frac{(1 + C.L.)}{2}, 0, 1]$
$\hat{p} \pm z^* \sqrt{(\frac{\hat{p}(1-\hat{p})}{n})}$
DEFINITIONS
$\hat{p}$ = Point estimate
z^* = Critical value
Standard Deviation =
• $\sqrt{\{ [\hat{p}(1-\hat{p})] \div n \}}$
Margin of Error =
• $z^* \sqrt{\{ [\hat{p}(1-\hat{p})] \div n \}}$

CALCULATOR
STAT ⇒ TESTS ⇒ 5A: 1-PropZInterval</td>
<td>within the interval (__ , __)</td>
</tr>
<tr>
<td>1 Sample z Interval for Means</td>
<td rowspan="2">DESCRIPTION
Confidence level
Degrees of Freedom = n-1
Context

NOTES
Identifying
• σ is known (rate)
DESCRIPTION
Confidence level
Degrees of Freedom = n-1
Context

NOTES
Identifying between z & t
• z test
  ○ σ is known (rare)
  ○ comes from the population
• t test
  ○ σ is unknown (common)
  ○ Comes from a sample</td>
<td rowspan="2">Random: Data came from random sample
Normal:
For z
• $n \geq 30$
For t
• $n \geq 30$
• $n \geq 15$
  ○ Without strong skewness or outliers
• $n < 15$
  ○ Without strong skewness or outliers
  ○ Distribution is approx. normal
Independent:
Either
• Individual observations are independent when sampling without replacement
• Sample is less than 10% of its respective population ($n \leq 0.10N$)</td>
<td>EQUATIONS
$z^* = \text{invNorm}[\frac{(1 + C.L.)}{2}, 0, 1]$
$\bar{x} \pm z^*(\sigma_{\bar{x}} / \sqrt{n})$
$\sigma_{\bar{x}} = \sigma_x / \sqrt{n}$
Delta x bar equals delta x divided by the square root of n

CALCULATOR
STAT ⇒ TESTS ⇒ 7: ZInterval</td>
<td rowspan="2">I am __% confident that the population mean of [context] is within the interval (__ , __)</td>
</tr>
<tr>
<td>1 Sample t Interval for Means</td>
<td>EQUATIONS
$t^* = \text{invT}[\frac{(1 + C.L.)}{2}, d.f.]$
$\bar{x} \pm t^*(S_x \div \sqrt{n})$
$\sigma_x = S_x / \sqrt{n}$

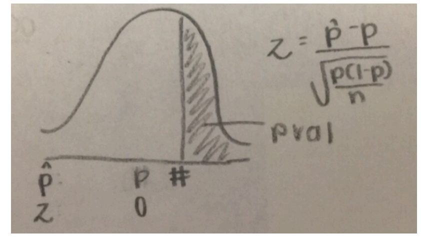
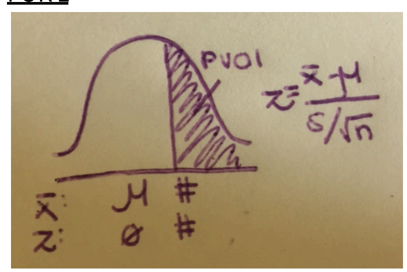
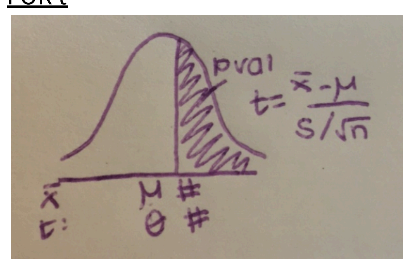
CALCULATOR
STAT ⇒ TESTS ⇒ 8: TInterval</td>
</tr>
</table>

AP STATISTICS STUDY GUIDE

created by shayla huynh (*note: not everything from the course has been added yet*)

contact me if you have any questions

EMAIL FACEBOOK TWITTER INSTAGRAM

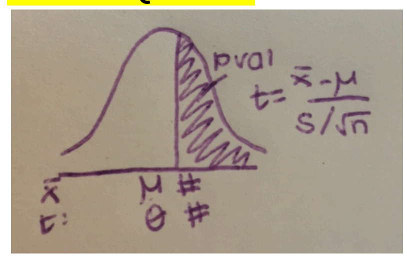
1 Proportion z Test	**DESCRIPTION** Significance Level • Use α=0.05 if none given Context **HYPOTHESIS** **H_o:** p = #. **H_a:** p ? # • Replace ? with either ≠, >, or <	**Random:** Data came from random sample **Normal:** • $n\hat{p} \geq 10$ • $n(1-\hat{p}) \geq 10$ **Independent:** Either • Individual observations are independent when sampling without replacement • Sample is less than 10% of its respective population ($n \leq 0.10N$)	**GRAPH & EQUATIONS**  **CALCULATOR** STAT ⇒ TESTS ⇒ 5: 1-PropZTest **P-VAL** DISTR (2ND VARS) ⇒ DISTR ⇒ 2: normalcdf(• *Shading left:* Lower = -1×10^{99} • *Shading right:* Upper = 1×10^{99}	I reject/fail to reject the H_o because the p-val, [value], is less/greater than α = [value]. There is sufficient/insufficient evidence to suggest [H_a in words].
1 Sample z Test 1 Sample t Test	**DESCRIPTION** Significance Level • Use α=0.05 if none given For t: Degrees of Freedom = n-1 Context **HYPOTHESIS** **H_o:** μ = #. **H_a:** μ ? # • Replace ? with either ≠, >, or < **NOTES** Identifying between z & t • *z test* ○ σ is known (rare) ○ comes from the population • *t test* ○ σ is unknown (common) ○ Comes from a sample	**Random:** Data came from random sample **Normal:** For z • $n \geq 30$ For t • $n \geq 30$ • $n \geq 15$ ○ Without strong skewness or outliers • $n < 15$ ○ Without strong skewness or outliers ○ Distribution is approx. normal **Independent:** Either • Individual observations are independent when sampling without replacement • Sample is less than 10% of its respective population ($n \leq 0.10N$)	**GRAPH & EQUATIONS** FOR z  FOR t  **CALCULATOR**	I reject/fail to reject the H_o because the p-val, [value], is less/greater than α = [value]. There is sufficient/insufficient evidence to suggest [H_a in words].

AP STATISTICS STUDY GUIDE

created by shayla huynh (*note: not everything from the course has been added yet*)

contact me if you have any questions

EMAIL FACEBOOK TWITTER INSTAGRAM

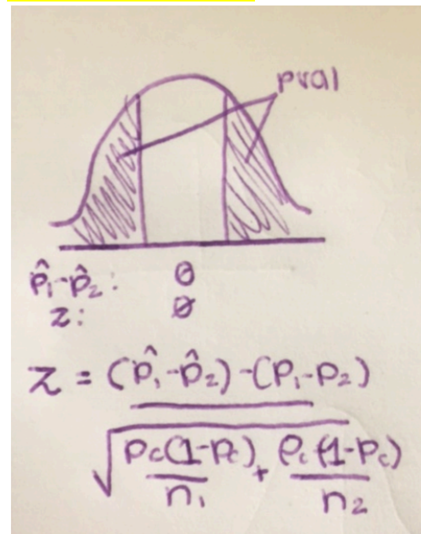
			Choose the according one STAT ⇒ TESTS ⇒ 1: Z-Test STAT ⇒ TESTS ⇒ 2: T-Test **P-VAL** DISTR (2ND VARS) ⇒ DISTR ⇒ 2: normalcdf(● *Shading left:* Lower = -1 x 10^{99} ● *Shading right:* Upper = 1 x 10^{99}	
Matched Pairs t Test	**DESCRIPTION** Significance Level ● Use α=0.05 if none given ● Degrees of Freedom = n-1 Context **HYPOTHESIS** **H_o:** μ_D = #. **H_a:** μ_D ? # ● Replace ? with either ≠, >, or < **NOTES** Identifying ● Two extremely similar subjects ● Before/after on the same subjects The test is identical to 1 sample t Test except for the following: ● Testing against a population mean *difference* ● Mark capital D subscripts on μ_D, $\bar{x}_D$, S_D	**Random:** Data came from random sample **Normal:** ● n ≥ 30 ● n ≥ 15 ○ Without strong skewness or outliers ● n < 15 ○ Without strong skewness or outliers ○ Distribution is approx. normal **Independent:** Either ● Individual observations are independent when sampling without replacement ● Sample is less than 10% of its respective population (n ≤ 0.10N)	**GRAPH & EQUATIONS**  ***NOTE: BE SURE TO ADD SUBSCRIPT D ON MU, X BAR, AND S*** **CALCULATOR** STAT ⇒ TESTS ⇒ 2: T-Test **P-VAL** DISTR (2ND VARS) ⇒ DISTR ⇒ 2: normalcdf(● *Shading left:* Lower = -1 x 10^{99} ● *Shading right:* Upper = 1 x 10^{99}	I reject/fail to reject the H_o because the p-val, [value], is less/greater than α = [value]. There is sufficient/insufficient evidence to suggest [H_a in words].
2 Proportion z Interval	**DESCRIPTION** Confidence Level Use "difference" Context	**Random:** Data came from random sample **Normal:** ● $n_1\hat{p}_1 \geq 10$ ● $n(_1 1-\hat{p}_1) \geq 10$	**EQUATIONS** $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2})}$ *Margin of error is everything after the ±*	**NOTES** Make sure to use "difference"

AP STATISTICS STUDY GUIDE

created by shayla huynh (*note: not everything from the course has been added yet*)

contact me if you have any questions

EMAIL FACEBOOK TWITTER INSTAGRAM

	PARAMETER Let p_1 = population proportion of … p_2 = population proportion of …	• $n_2\hat{p}_2 \geq 10$ • $n_2(1-\hat{p}_2) \geq 10$ "Both sample sizes are sufficiently large." **Independent:** Either • Both samples are independent • Both samples are less than 10% of its respective population	symbol **CALCULATOR** STAT ⇒ TESTS ⇒ 2-PropZTest	I am __% confident that the *difference* in *population proportion* of [context] is within the interval (__ , __)
2 Proportion z Test	**DESCRIPTION** Significance Level • Use α=0.05 if none given Use "difference" Context **PARAMETER** Let p_1 = population proportion of … p_2 = population proportion of … **HYPOTHESIS** H_o: $p_1 - p_2 = 0$ H_a: $p_1 - p_2$? 0 • Replace ? with either ≠, >, or <	**Random:** Data came from random sample **Normal:** • $n_1\hat{p}_1 \geq 10$ • $n(_1 1-\hat{p}_1) \geq 10$ • $n_2\hat{p}_2 \geq 10$ • $n_2(1-\hat{p}_2) \geq 10$ "Both sample sizes are sufficiently large." **Independent:** Either • Both samples are independent • Both samples are less than 10% of its respective population	**GRAPH & EQUATIONS**  • In a two tail test, the difference is the same distance on each side • Shade left if less than • Shade right if greater than • Parameter is equal to zero **POOLED SAMPLE STATISTIC (COMBINED)** $\hat{p}_c = \frac{x_1 + x_2}{n_1 + n_2}$	I reject/fail to reject the H_o because the p-val, [value], is less/greater than α = [value]. There is sufficient/insufficient evidence to suggest [H_a in words].

AP STATISTICS STUDY GUIDE

created by shayla huynh (*note: not everything from the course has been added yet*)

contact me if you have any questions

EMAIL FACEBOOK TWITTER INSTAGRAM

<table>
<tr><td></td><td></td><td></td><td>CALCULATOR
STAT ⇒ TESTS ⇒ B: 2-PropZInt</td><td></td></tr>
<tr><td>2 Sample z Interval for Means

2 Sample t Interval for Means</td><td>DESCRIPTION
Confidence level
<u>For t:</u>
Degrees of Freedom =
• Use d.f. on calculator
Context

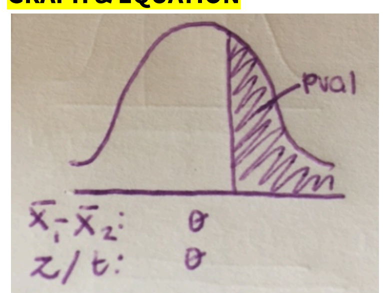
PARAMETER
Let μ_1 = population mean __ of …
μ_2 = population mean __ of …

NOTES
Identifying between z & t
• <u>z test</u>
○ σ is known (rare)
○ comes from the population
• <u>t test</u>
○ σ is unknown (common)
○ Comes from a sample</td><td>Random: Data came from random sample
Normal: (do for both n_1 and n_2)
• $n \geq 30$
• $n \geq 15$
○ Without strong skewness or outliers
• $n < 15$
○ Without strong skewness or outliers
○ Distribution is approx. normal
Independent:
Either
• Individual observations are independent when sampling without replacement
• Sample is less than 10% of its respective population ($n \leq 0.10N$)</td><td>EQUATION
FOR z
$(\bar{x}_1 - \bar{x}_2) \pm z^* \sqrt{(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}$
FOR t
$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})}$
Margin of error is everything after the ± symbol

CALCULATOR
Choose the according one
STAT ⇒ TESTS ⇒ 0: 2-SampleZInt
STAT ⇒ TESTS ⇒ 0: 2-SampleTInt</td><td>I am __% confident that the population mean difference of [context] is within the interval (__ , __)</td></tr>
<tr><td>2 Sample z Test for Means

2 Sample t Test for Means</td><td>DESCRIPTION
Significance Level
• Use α=0.05 if none given
<u>For t:</u>
Degrees of Freedom =
• Use d.f. on calculator
Context

PARAMETER
Let μ_1 = population mean __ of …
μ_2 = population mean __ of …

HYPOTHESIS
H_o: $\mu = 0$
H_a: μ ? 0</td><td>Random: Data came from random sample
Normal: (do for both n_1 and n_2)
• $n \geq 30$
• $n \geq 15$
○ Without strong skewness or outliers
• $n < 15$
○ Without strong skewness or outliers
○ Distribution is approx. normal
Independent:
Either
• Individual observations are independent when sampling without replacement</td><td>GRAPH & EQUATION


FOR z</td><td>I reject/fail to reject the H_o because the p-val, [value], is less/greater than α = [value]. There is sufficient/insufficient evidence to suggest [H_a in words].

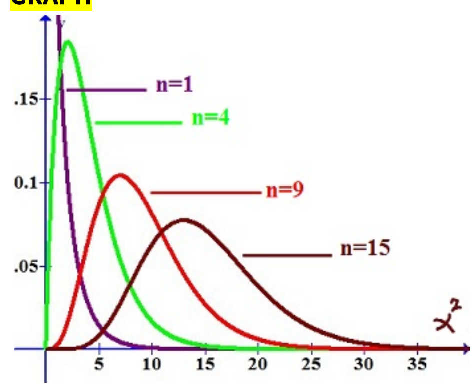
EXAMPLE
I fail to reject the H_o because the p-val, 0.08, is greater than α = 0.05. There is insufficient evidence to suggest that</td></tr>
</table>

AP STATISTICS STUDY GUIDE

created by shayla huynh (*note: not everything from the course has been added yet*)

contact me if you have any questions

EMAIL | FACEBOOK | TWITTER | INSTAGRAM

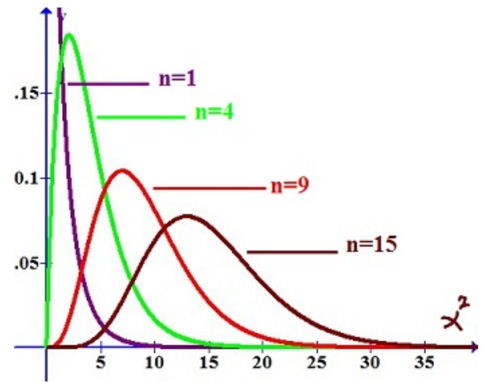
	● Replace ? with either ≠, >, or < **NOTES** Identifying between z & t ● *z test* ○ σ is known (rare) ○ comes from the population ● *t test* ○ σ is unknown (common) ○ Comes from a sample	● Sample is less than 10% of its respective population ($n \le 0.10N$)	$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}}$ *FOR t* $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$ **CALCULATOR** *Choose the according one* STAT ⇒ TESTS ⇒ 3: 2-SampleZTest STAT ⇒ TESTS ⇒ 4: 2-SampleTTest	there is a difference in mean response time between the Northern and Southern fire stations..
Chi Squared Goodness of Fitness Test	**DESCRIPTION** Significance Level ● Use α=0.05 if none stated Degrees of Freedom = k - 1 ● k = # of categories Context **HYPOTHESIS** **H_o =** P_1=, P_2=, P_3=, ... **H_a =** At least one P_i is incorrect	**Random** Normal, 10% condition when necessary **Expected Counts:** All expected counts exceed 5 ● E[x] = np ● Always in parentheses ● Must show expected counts in the answers **Independence:** Sample is less than 10% of its population	**GRAPH**  $x^2 = \sum [(\text{Observed-Expected})^2 / \text{Expected}]$ **CALCULATOR** STAT ⇒ TESTS ⇒ D: x^2GOF-Test *TIP:* Press draw to see what the graph should look like	I reject/fail to reject the H_o because the p-val, [value], is less/greater than α = [value]. There is sufficient/insufficient evidence to suggest [H_a in words].
Chi Squared Test for Homogeneity	**DESCRIPTION** Significance Level ● Use α=0.05 if none stated	**Random** Normal, 10% condition when necessary **Expected Counts:** All expected counts	**GRAPH**	I reject/fail to reject the H_o because the p-val, [value], is less/greater

AP STATISTICS STUDY GUIDE

created by shayla huynh (*note: not everything from the course has been added yet*)

contact me if you have any questions

EMAIL FACEBOOK TWITTER INSTAGRAM

Chi Squared For Association/ Independence	Degrees of Freedom = • (# of Rows)(# of Columns) Context **HYPOTHESIS** Homogeneity **H_o:** There *is no* difference in __ when __ **H_a:** There *is a* difference in __ when __ Independence **H_o:** There *is no* association between __ and __ **H_a:** There *is an* association between __ and __ **NOTES** Identifying • Homogeneity ○ Multiple samples and treatment groups • Independence ○ Cut into groups	exceed 5 • E[x] = (Row Total ∗ Column Total) ÷ Grand Total • Always in parentheses • Must show expected counts in the answers **Independence:** Sample is less than 10% of its population	 $x^2 = \sum [(Observed-Expected)^2 / Expected]$ **CALCULATOR** MATRIX (2nd x^{-1}) ⇒ EDIT ⇒ 1: [A] • Adjust dimensions accordingly • Insert in observed count values STAT ⇒ TESTS ⇒ (C) x^2-Test • Observed: [A] • Expected: [B] MATRIX (2nd x^{-1}) ⇒ EDIT ⇒ 1: [B] • Get expected count values *TIP:* Press draw to see what the graph should look like	than α = [value]. There is sufficient/insufficient evidence to suggest [H_a in words].
Linear Regression t Interval for Slope	**DESCRIPTION** Confidence level Degrees of Freedom = n-2 Context	**"LINEAR"** **Linear:** No trend in the residual **Independent:** Normal, 10% condition when necessary **Normal:** Histogram of residuals appear normal **Equal Variance:** Same amount of scatter above and below the residual plot **Random:** Normal	**EQUATIONS** $b \pm t^*SE_b$ $SE_b = S/[S_x\sqrt{(n-1)}]$ **DEFINITIONS** **SE_b** = Standard Error of the Slope **S** = Standard deviation of the residuals **CALCULATOR** STAT ⇒ TESTS ⇒ G: LinRegTInt	I am __% confident that the *population slope* between [explanatory variable] and [response variable] is within the interval (__ , __)
Linear Regression t	**DESCRIPTION**		**EQUATIONS**	I reject/fail to reject the

AP STATISTICS STUDY GUIDE

created by shayla huynh (*note: not everything from the course has been added yet*)

contact me if you have any questions

EMAIL FACEBOOK TWITTER INSTAGRAM

Test for Slope

Significance Level
- Use α=0.05 if none stated

Degrees of Freedom = n-2

HYPOTHESIS

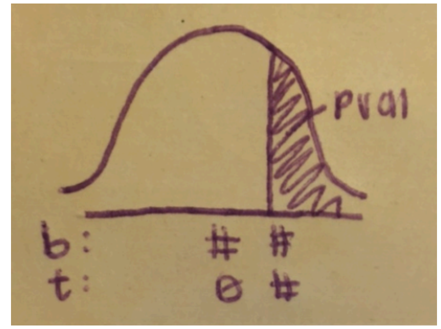
H_o = β = 0

H_a = β ? 0
- Replace ? with either ≠, >, or <

DEFINITIONS

β = Population slope (in context of regression)

r = population correlation coefficient



$t = \frac{b - \beta}{SE_b}$ $SE_b = S/[S_x\sqrt{(n-1)}]$

CALCULATOR

STAT ⇒ TESTS ⇒ G: LinRegTTest

H_o because the p-val, [value], is less/greater than α = [value]. There is sufficient/insufficient evidence to suggest [H_a in words].

	H_o True	H_o False
Reject H_o	Type I Error	Correct Conclusion
Fail to Reject H_o	Correct Conclusion	Type II Error

P(Type I Error) = α
P(Type II Error) = β
Power = 1 - β

If type I error is worse, statisticians set α = 0.01
If type II error is worse, statisticians set α = 0.1

If no α is stated, set α = 0.05

KEY CONCEPTS

- When α goes up, β goes down, and power goes up
- When α goes down, β goes up, and power goes down

PROBABILITY RULES

NAME	ABOUT/EQUATION	NAME	ABOUT/EQUATION
Probability of Event A	P(A) = (# of events corresponding to A) ÷ (Total # of events in the same space)	P(A|B)	Reads the probability that event A happens, given that event B happens • | symbol means “given”
Sample Space (s)	• P(S) = 1 • Every outcome that can occur for a given situation	Rule of addition	P(A ∪ B) = P(A) + P(B) - P(A ∩ B)

AP STATISTICS STUDY GUIDE

created by shayla huynh (*note: not everything from the course has been added yet*)

contact me if you have any questions

EMAIL FACEBOOK TWITTER INSTAGRAM

Union	∪ • Means "Or"	Rule of subtraction	$P(A^C) = 1 - P(A)$
Intersection	∩ • Means "And"	Rule of multiplication	• $P(A \cap B) = P(A) * P(B)$ ○ When A and B are independent • $P(A \cap B) = P(A) * P(B\|A)$ • $P(A \cap B) = 0$ ○ if event A and B are mutually exclusive
Complement	A^C • Means "Not A"		

RANDOM VARIABLES

IDENTIFYING

BINOMIAL RANDOM VARIABLE		**GEOMETRIC RANDOM VARIABLE**	
B	Binary	**B**	Binary
I	Independent	**I**	Independent
N	Fixed Number of Trials	**T**	Counting Trials Until First Success
S	Fixed Probability of Success	**S**	Fixed Probability of Success

MORE

BINOMIAL RANDOM VARIABLE		**GEOMETRIC RANDOM VARIABLE**	
EQNS	• $P(x=k) = (n \div k)\ p^k\ (1-p)^{n-k}$ • S.t. $(n \div k) = n^c k = [n! \div k!(n-k)!]$ • $\mu_x = np$ • $\sigma_x = \sqrt{[\ np(1-p)\ ]}$ Think of pascal triangle ! = Factorial • **EX:** 3! = 3 Factorial = 3 ∗ 2 ∗ 1 = 6 • **EX:** 5! = 5 Factorial = 5 ∗ 4 ∗ 3 ∗2 ∗ 1 = 120	**EQNS**	• $P(y=k) = (1-p)^{k-1} * p$ • $\mu_y = 1/p$ • $\sigma_y = \sqrt{[\ (1-p) \div p^2\ ]}$

AP STATISTICS STUDY GUIDE

created by shayla huynh (*note: not everything from the course has been added yet*)

contact me if you have any questions

EMAIL *FACEBOOK* *TWITTER* *INSTAGRAM*

ON THE CALCULATOR	DISTR (2ND VARS) • binompdf(n,p,k) = p(x = k) • binomcdf(n,p,k) = p(n ≤ k)	**ON THE CALCULATOR**	DISTR (2ND VARS) • geompdf(p,k) = p(µ = k) • geomcdf(p,k) = p(y ≤ k)

Important statistic formulas: http://stattrek.com/statistics/formulas.aspx
Tbh You don't rly need to memorize most of them. The important ones you should know are things like margin of error, expected value, etc.