Sentences to know:

**Unit 2 (Linear Regression)**

Interpreting slope:

For every additional 1 x-variable the y-variable increases/decreases by slope y-units on average.

Interpreting y-intercept:

When x-variable is equal to 0 x-units, the y-variable is equal to y-intercept y-units on average.

Interpreting r2:

r2% of the change in the y-variable is due to the change in the x-variable.

Is the linear model a good fit for the data?

Comment on these 3 things:

1- original plot (linear?)

2- correlation (high/strong?)

3- residual plot (scattered?)

**UNIT 5: Sentences to know:**

**Conclusion to a confidence interval**

We are \_\_% confident that the true parameter is between \_\_% and \_\_%.

**What does \_\_\_% confidence mean?**

\_\_\_% of all random samples of \_\_\_\_\_\_\_ will produce confidence intervals that catch the true parameter.

One sample: “percent of …”

Two sample: “difference in the percent of group 1 and group 2 …”

One sample: n things

Two sample: n1 things and n2 things

**Conclusion to a test of significance**

We reject/fail to reject Ho b/c the p-value of \_\_ is </> than α = \_\_.

We have sufficient/insufficient evidence that \_(Ha, in English words)\_.

**Interpreting the P-Value in context**

There is a p-value% chance of getting a sample where \_\_\_\_\_\_\_\_ or more extreme, if really the claim of Ho is true.

 

**Error & Power**

Type I = Conclude (Ha), when really it is not.

Type II = Conclude (NOT Ha), when really it is.

Power = Probability of concluding (Ha), when really it is.

FORMULAS:

1 sample Z Interval 1 sample Z test

$\hat{p}\pm Z^{\*}\sqrt{\frac{\hat{p}\hat{q}}{n}}=(a, b)$ $Z=\frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$

2 sample Z Interval 2 sample Z test

$\hat{p\_{1}}-\hat{p\_{2}}\pm Z^{\*}\sqrt{\frac{\hat{p\_{1}}\hat{q\_{1}}}{n\_{1}}+\frac{\hat{p\_{2}}\hat{q\_{2}}}{n\_{2}}}=(a, b)$ $Z= \frac{\hat{p\_{1}}-\hat{p\_{2}}}{\sqrt{\frac{\hat{p}\hat{q}}{n\_{1}}+\frac{\hat{p}\hat{q}}{n\_{2}}}}$ $\hat{p}\_{pooled= \frac{\hat{p\_{1}}+\hat{p\_{2}}}{n\_{1}+n\_{2}}}$

**UNIT 6:­ sentences and formulas to know**

**1 sample t test: on Calculator: T-Test**

Ho: µ = #

Ha: µ >, <, ≠ #

$t= \frac{\overbar{x}-μ}{^{s}/\_{\sqrt{n}}}$

P(t >, < \_\_\_\_\_) = tcdf(lower, upper, df) df = n-1

We reject/fail to reject Ho because the p-value of \_\_\_\_\_ is </> α = \_\_\_\_.

We have sufficient/insufficient evidence that the true average of \_\_\_\_\_\_ is \_\_(Ha)\_\_\_\_.

**1 sample t interval: on Calculator: T-Interval**

$\overbar{x}\pm (t^{\*})\left(^{s}/\_{\sqrt{n}}\right)$ = (a, b) df = n-1

We are \_\_\_% confident that the true average of \_\_\_\_\_ is between \_\_a\_\_ and \_\_b\_\_\_ units.

**2 sample t test: on Calculator: 2 Samp T Test**

Ho: µ1 = µ2

Ha: µ1 >, <, ≠ µ2

$t= \frac{\overbar{x}\_{1}-\overbar{x}\_{2}}{\sqrt{\frac{s\_{1}^{2}}{n\_{1}}+\frac{s\_{2}^{2}}{n\_{2}}}}$

P(t >, < \_\_\_\_\_) = tcdf(lower, upper, df)

df = given in 2 sample t-test on calculator OR use the smaller of n1 – 1 and n2 - 1

We reject/fail to reject Ho because the p-value of \_\_\_\_\_ is < / > α = \_\_\_\_.

We have sufficient/insufficient evidence that the true average of \_\_population 1\_\_\_\_ is \_\_(Ha symbol)\_\_\_\_ to \_\_\_population 2\_\_\_\_.

**2 sample t Interval: on Calculator: 2 Samp T-Interval**

$\left(\overbar{x}\_{1}-\overbar{x}\_{2}\right)\pm (t^{\*})\left(\sqrt{\frac{s\_{1}^{2}}{n\_{1}}+\frac{s\_{2}^{2}}{n\_{2}}}\right)$ = (a, b)

* get t\* from my program (INVT) or calculator invT() or formula sheet
* df is same as for 2 sample t test

We are \_\_\_% confident that the true difference between the average \_\_\_\_\_ of \_\_population 1\_\_\_ and \_\_\_population 2\_\_\_ is between \_\_a\_\_ and \_\_b\_\_ units.

OR

We are \_\_% confident that the average \_\_\_\_\_ of \_\_population 1\_\_ is between \_\_a\_\_ and \_\_b\_\_ units higher/lower than \_\_population 2\_\_\_.

**Paired t-test:**

L1 = first set of data L2 = second set of data L3 = L2 – L1 (the differences)

Then use the 1 sample t test on data in L3

Ho: µd = # ***define µd***

Ho: µd <, >, ≠ # \*note: often µd = 0

Conclusion: same as 1 sample t test, but make sure you answer the original question in context, or say “mean difference”

**Paired t interval:**

See paired t test for how to deal with the data…..

Use the 1 sample t interval on data in L3.

Conclusion: same as 1 sample t interval, but make sure you say “average difference between \_\_ and \_\_”

**SENTENCES:**

*Look at Ha to see what “more extreme” is*

**P-value:**

**1 sample:**

There is a p-value% chance of getting a sample with an average \_\_\_\_\_\_\_ of $\overbar{x} $ units or more extreme, if really the true average of \_\_\_\_\_\_ is (use # in Ho) units.

**2 sample:**

There is a p-value% chance of getting a sample with a difference averages of population 1 and population 2 things of ($\overbar{x}\_{1}-\overbar{x}\_{2})$ units or more, if really the two population averages are equal.

**Confidence:**

**1 sample:**

\_\_\_% of repeated random samples of n things will produce confidence intervals that catch the true average of population.

**2 sample:**

\_\_\_% of repeated random samples of n1 things and n2 things will produce confidence intervals that catch the true difference between the averages of population1 and population2.

**Errors/Power:**

Same as Unit 5 stuff

Type I = Conclude Ha, and its not

Type II = Conclude NOT Ha, and it is.

Power = Probability of concluding Ha, and it is.

**Unit 7:**

**Chi Square**

**GOF**

Hypotheses: Use the phrasing from the problem!

Ho: The observed distribution of \_\_\_\_\_\_ fits the expected distribution of \_\_\_\_\_.

Ha: The observed distribution of \_\_\_\_\_\_ DOES NOT fit the expected distribution of \_\_\_\_\_.

**Independence:**

Hypotheses: Use the phrasing from the problem!

Ho: There is no association between row variable and column variable. (or: no relationship, no affect)

Ha: There is an association between row variable and column variable. (or: relationship, affect)

Or

Ho: row variable and column variable are independent.

Ha: row variable and column variable are dependent.

**Conditions:**

1. Random
2. Categorical Data
3. All expected values > 5

**Formula:**

$χ^{2}=\sum\_{}^{}\frac{(observed-expected)^{2}}{expected}$

**Errors/Power:**

Same as previous units:

Type I = Conclude Ha, and its not

Type II = Conclude NOT Ha, and it is.

Power = Probability of concluding Ha, and it is.

**Unit 7:**

**Linear Regression t test**

**Hypotheses:**

Ho: β1 = 0

Ha: β1 >, <, ≠ 0

**Test Statistic:**

$t=\frac{b\_{1}-β\_{1}}{SE\_{b1}}$ = 🡪 Usually given computer output… see below:

Example:



**Conclusion:**

* Reject/Fail to reject …
* We have sufficient evidence that as x-variable increases, the y-variable increases/decreases/changes.

(or: there is a positive/negative/an association between x-variable and y-variable.)

**Conditions (not usually asked to do):**

1. Random
2. Linear Data
3. Independence (between the points)
4. Normal residuals
5. Equal Variance