Name______________________________

AP Statistics CED 9.1 Daily Video 1 (Skill 1.A)

Introducing Statistics – Do Those Points Align?

What Will We Learn?

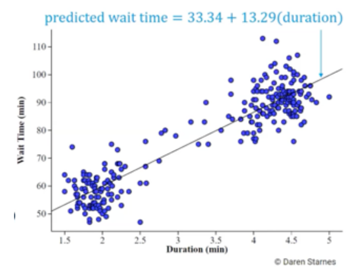
How can we determine if the slope of a sample regression line is consistent with random variation from a population regression model?

How Faithful is Old Faithful?

The Old Faithful geyser is the most popular attraction in Yellowstone National Park. People travel from all over the world to see this geyser erupt. The National Park Service helps visitors plan their time in the park by when Old Faithful will erupt next. The Starnes family took its first trip to Yellowstone National Park in July 1995. They only had six hours in the park, but were able to see Old Faithful erupt. Mr. and Mrs. Starnes returned to the park in July 2019. They wondered if the model used to predict eruptions of Old Faithful was still the same as in 1995.

Old Faithful: 1995

This scatterplot displays data on the duration (in minutes) and wait time until the next eruption (in minutes) for all 262 recorded eruptions of Old Faithful in July 1995.



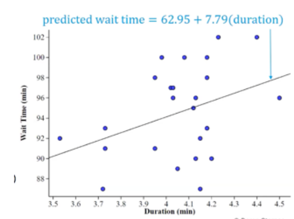
There is a ________________________, _______________, _______________ relationship between _______________ and ___________________ in this population of Old Faithful eruptions.

Its equation is:

predicted wait time = 33.34 + 13.29 (duration)

Old Faithful: 2019

The scatterplot displays data on the duration (in minutes) and wait time until the next eruption (in minutes) for a random sample of 25 Old Faithful eruptions in July 2019.



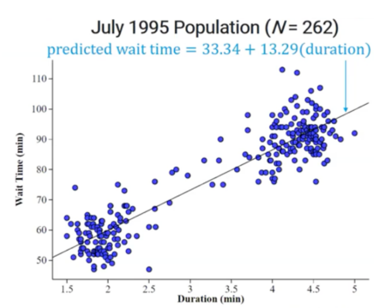
There is a ________________________, _______________, _______________ relationship between _______________ and ___________________ in this population of Old Faithful eruptions.

Its equation is

predicted wait time = 62.95 + 7.79 (duration)

Old Faithful: Then and Now

Is it believable that the population regression model from 1995 is still valid for predicting wait time from the duration of the previous Old Faithful eruption in 2019?



We need to ______________ the __________________ of getting a sample regression line with a ___________ as least as unusual as 7.79 in a random sample of n = 25 observations from the July 1995 population.

July 2019. Sample (n = 25) yielded a:

predicted wait time = 62.95 + 7.79(duration)

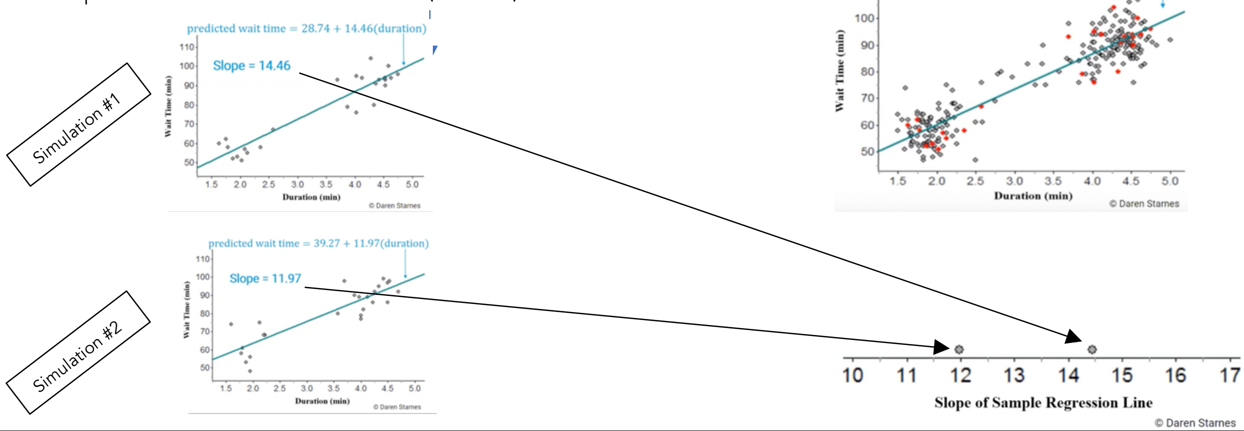
Name________________________________

Old Faithful: Simulation Trial 1

We will simulate random samples of *n* = _____ points from the July 1995 population to see how likely it is we will get a slope of __________.

Simulated random sample of *n*=25 points from the July 1995 population
predicted wait time = 28.74 + 14.46(duration)

July 1995 Population (*N* = 262)
predicted wait time = 33.34 + 13.29(duration)



Old Faithful: Simulation Results

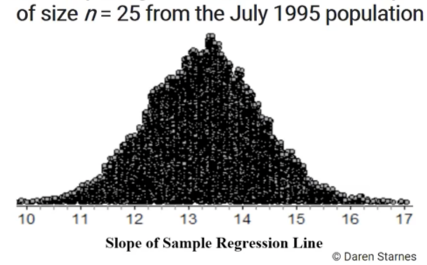
Is it believable that the population regression model from 1995 is still valid for predicting wait time from the duration of the previous Old Faithful eruption in 2019? After running many, many simulations we get the sampling distribution at the right.

July 2019 Sample (*n* = 25)
predicted wait time = 62.95 + 7.79(duration)

We need to ____________ the probability of getting a sample regression line with a __________ at least as unusual as __________ in a random sample of *n* = 25 observations from the July, 1995 population.

Looking at the sampling distribution, we see that we never got a slope of 7.79, so our estimated probability ≈______

July 1995 Population (*N* = 262)
predicted wait time = 33.34 + 13.29(duration)
Simulated sampling distribution of the slope of the sample regression line based on samples of size *n* = 25 from the July 1995 population



Old Faithful: Simulation Results

Is it believable that the population regression model from 1995 is still valid for predicting wait time from the duration of the previous Old Faithful eruption in 2019?

______. There is an _______________________ 0 probability of obtaining a ______________ regression line with a ______________ as least as surprising (in __________ directions) as 7.79 is the population regression model from __________ is still ____________.

What Should We Take Away?

- **Take _______ random samples of size ___ from the population.**
- **Calculate the ________ of each sample regression line.**
- **Build the _______________________________ distribution of the slope.**
- **See if the _______________ value of the sample slope can be explained by ______________ variation or not.**

Name________________________________

AP Statistics CED 9.2 Daily Video 1 (Skill 1.D)

Confidence Intervals for the Slope of a Regression Model

What Will We Learn?

What conditions must the population regression model meet to obtain valid confidence intervals and significance tests for the slope?

How can we determine the shape, center and variability of the sampling distribution of the slope of a sample regression line?

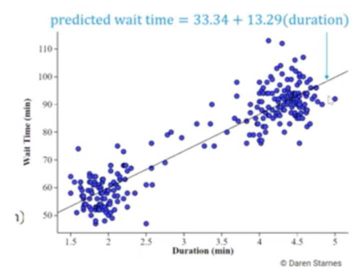
Is Old Faithful Still Faithful?

The Starnes family took its first trip to Yellowstone National Park in July 1995. They only had six hours in the park, but were able to see Old Faithful erupt. Mr. and Mrs. Starnes returned to the park in July 2019. They wondered if the model used to predict eruptions of Old Faithful was still the same as in 1995. Earlier, we used simulation to determine that the answer is "No."

Can we construct a confidence interval for the slope of the population regression line in 2019? To answer that question, we need to understand the connection between the population regression model and the sampling distribution of the slope.

Old Faithful: 1995

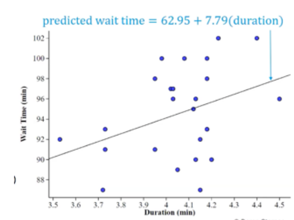
The scatterplot displays data on the duration (in minutes) and wait time until the next eruption (in minutes) for all 262 recorded eruptions of Old Faithful in July, 1995. We have added the population ______________________ line to scatterplot. Its equation is:



predicted wait time = 33.34 + 13.29(duration)

Old Faithful: 2019

The scatterplot displays data on the duration (in minutes) and wait time until the next eruption (in minutes) for a random sample of 25 Old Faithful eruptions in July 2019.



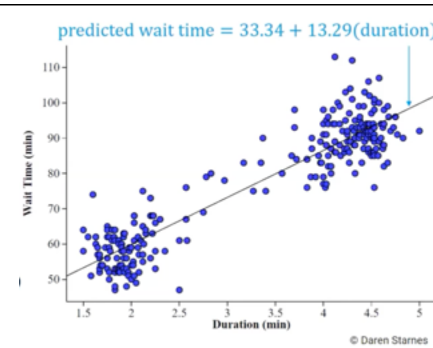
We have added the ___________________________________ line to the scatterplot. Its equation is:

predicted wait time = 62.95 + 7.79 (duration)

Can we construct a confidence interval for the slope of the population regression line in 2019?

Simulated Sampling Distribution

Suppose we take _______________ random sample of n = 25 observations from the __________________ of eruptions in July, 1995 and calculate the sample regression line $\hat{y} = a + bx$ for each one.

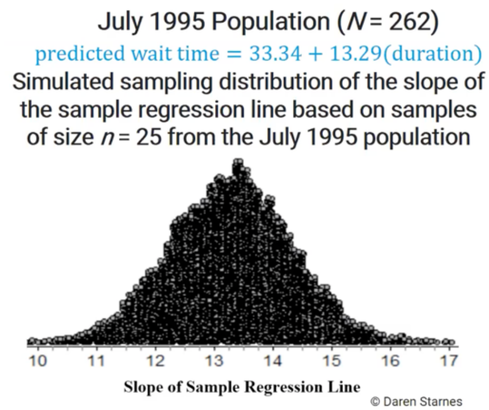


Can we determine the shape, center and variability of the sampling distribution of the slope b of the sample regression line from the population regression model?

Name________________________________

Simulated Sampling Distribution

Suppose we take repeated random samples of n = 25 observations from the population of eruptions in July 1995 and calculate the sample regression line $\hat{y} = a + bx$ for each one. Can we determine the shape, center and variability of the sampling distribution of the slope b of the sample regression line from the population model?



What are the shape, center and variability of the simulated sampling distribution of the slope of the sample regression line $\hat{y} = a + bx$?

Sampling Distribution of b (Use the distribution above.)

Shape: ______________________________

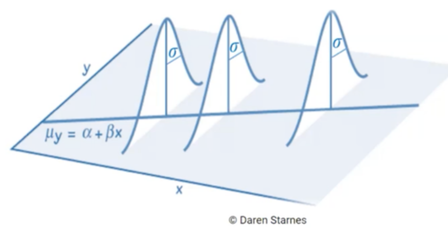
Center: _____________________________

Variability: __________________________

How are these characteristics related to the population regression model?

Population Regression Model

The population regression model is $\mu_y = \alpha + \beta x$ where μ_y is the mean value of the ______________ variable y for a given value of the __________________ variable x.



Confidence intervals and significance test for the slope β require that the population regression model meet these conditions:

- The ___________ relationship between x and y is __________________.
- The standard deviation of y, σ_y does not __________ with x.
- For a particular value of x, the responses (__________) are approximately normally distributed.

Population Regression Model

Confidence intervals and significance test for the slope β require that the population regression model meet these conditions:

- The true relationship between x and y is linear. To check this:
  Scatterplot: The scatterplot shows a ____ relationship between duration and wait time.
  Residual Plot: The residual plot show ____________ scatter about Residual = ____ line and no evidence of a _______________ curved pattern.

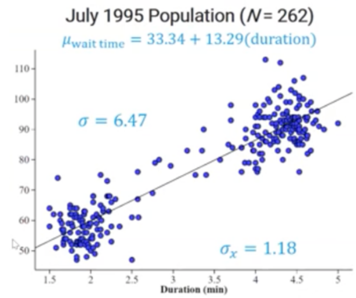
- The standard deviation of y, σ_y, does not vary with x.
  Scatterplot: The wait times __________ by a ___________ amount for the different eruption durations in the data set.
  Residual Plot: The residuals are _________ in ______ for the eruption durations in the data set.

- For a particular value of x, the responses (__________) are approximately normally distributed.
  Dotplot: A dotplot of residuals is roughly ______________________, ___________________, and somewhat ____________________________.

Name_______________________________

Sampling Distribution of Slope

When a population regression model $\mu_y = \alpha + \beta x$ meets the conditions, the sampling distribution of the slope b of the sample regression line has:



Shape: ______________________

Center: $\mu_b = \beta$ = ____________

Variability: $\sigma_b = \dfrac{\sigma}{\sigma_x\sqrt{n}}$

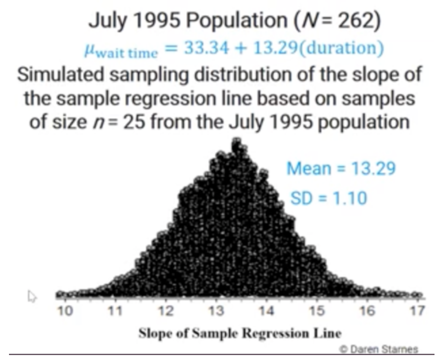
where: _____ = Standard deviation of __________ for the population regression line and

_____ = Standard deviation of __________ in population and

_____ = sample size.

Sampling Distribution of Slope

Suppose we take every possible random sample of _________ observations from the population and calculate the sample regression line $\hat{y} = a + bx$ for each one.



Sampling distribution of b

Shape: _________________

Center: $\mu_b = \beta$ = ____________

Variability: $\sigma_b = \dfrac{\sigma}{\sigma_x\sqrt{n}}$ = __________________________

When sampling without replacement check: (if $n \leq 10\%N$) ______________________

Inference for Slope

Can we construct a confidence interval for the slope of the population regression line in 2019?

Related questions:

How do we check the conditions about the __________________ regression model using only the ______________ data?

How can we _________________ the standard deviation of the ________________ distribution of the sample slope, $\sigma_b = \dfrac{\sigma}{\sigma_x\sqrt{n}}$, when we do not know the values of σ or σ_x?

What Should We Take Away?

What conditions must the population regression model meet to obtain valid confidence intervals and significance tests for the slope?

- **The _______ relationship between x and y is ___________.**
- **The standard deviation of y, σ_y, does ______________ with x.**
- **For a particular value of x, the responses __________ are approximately normally distributed.**

How can we determine the shape, center and variability of the sampling distribution of the slope of a sample regression line?

Shape: Approximately normal; Center: $\mu_b = \beta$ Variability: $\sigma_b = \dfrac{\sigma}{\sigma_x\sqrt{n}}$

Name______________________________

AP Statistics CED 9.2 Daily Video 2 (Skill 4.C)

Confidence Intervals for the Slope of a Regression Model

What Will We Learn?

How do we identify an appropriate confidence interval procedure for the slope of a population regression line?

How do we verify the conditions for calculating a confidence interval for the slope of a population regression line?

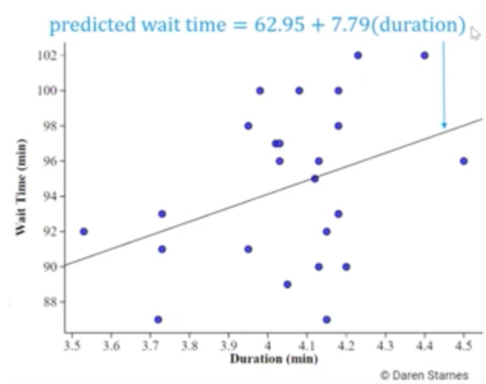
Predicting Old Faithful in 2019

The Starnes family took its first trip to Yellowstone National Park in July 1995. They only had six hours in the park, but were able to see Old Faithful erupt. Mr. and Mrs. Starnes returned to the park in July 2019. They wondered if the model used to predict eruptions of Old Faithful was still the same as in 1995. Earlier, we used simulation to determine that the answer is "No."

Can we construct a confidence interval for the slope of the population regression line in 2019?

Old Faithful: 2019

The scatterplot displays data on the duration (in minutes) and wait time until the next eruption (in minutes) for a random sample of 25 Old Faithful eruptions in July 2019, along with the sample regression line. Computer output from a least-squares regression analysis is shown below:



Predictor	Coef	SE Coef	T-Value	P-Value
Constant	62.95	16.4	3.85	0.001
Duration(min)	7.79	4.03	1.94	0.065
S=4.20970	R-sq=14.01%		R-sq(adj)=10.27%	

Can we construct a confidence interval for the slope of the population regression line in 2019?

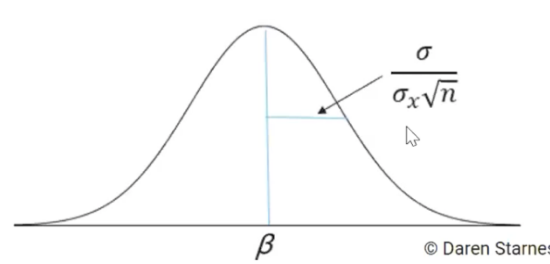
Population Regression Model

The population regression model is $\mu_y = \alpha + \beta x$ where μ_y is the mean value of the _______________ variable y for a given value of the ___________________ variable x. Confidence intervals and significance tests for the slope β require that the population regression model meet three conditions:

- The ___________ relationship between x and y is __________________.
- The standard deviation of y, σ_y does not __________ with x.
- For a particular value of x, the responses (__________) are approximately normally distributed.

Sampling Distribution of Slope

When the population regression model $\mu_y = \alpha + \beta x$ meets the conditions, the sampling distribution of the slope b of the sample regression line has:



Shape: __________________________;

Center: _____________________; and

Variability ____________________ (______________)

Where: ______ = standard deviation of residuals for the population regression line

______ = standard deviation of x-values in population and

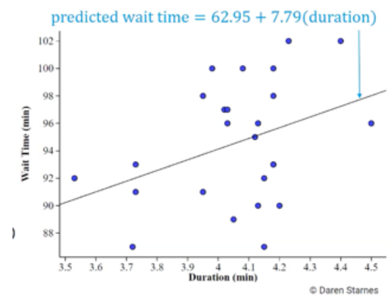
______ = sample size.

STATS MEDIC

Name________________________________

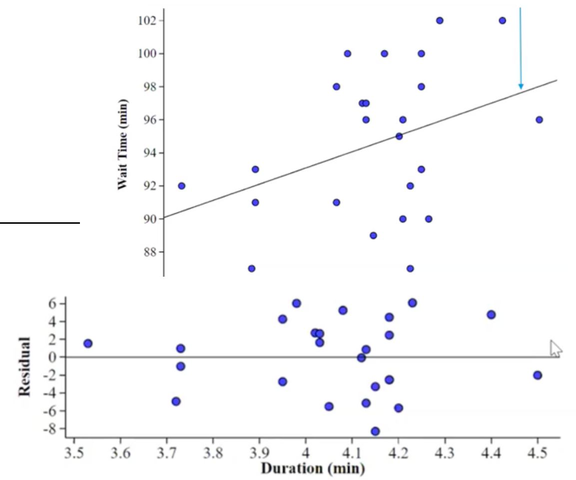
Inference for Slope

Can we construct a confidence interval for the slope of the population regression line in 2019?



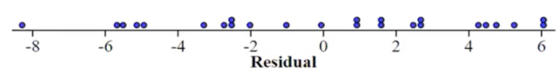
Related questions:
How do we check the conditions about the ________________ regression model using only the _____________ data?
How can we ________________ the standard deviation of the _______________ distribution of the sample slope, $\sigma_b = \frac{\sigma}{\sigma_x\sqrt{n}}$, when we do not know the values of σ or σ_x?

Checking Conditions



To calculate a confidence interval for the slope β of a population regression line we must check the following:

- **The true relationship between x and y is linear.**
  Scatterplot: The scatterplot shows a ____ relationship between duration and wait time.
  Residual Plot: The residual plot shows ___________ scatter about Residual = ____ line and no evidence of a _____________ curved pattern.
- **The standard deviation of y, σ_y, does not vary with x.**
  Scatterplot: The wait times _________ by a __________ amount for the different eruption durations in the data set.
  Residual Plot: The residuals are ________ in _____ for the eruption durations in the data set.
- **For a particular value of x, the responses (________) are approximately normally distributed.**



  Dotplot: A dotplot of residuals show no ______________________________________.

If the observed distribution is ___________ or shows other _____________________ from normality, n should be ______________________ or we cannot proceed.

- **There is independence in data collection.** Data are collected using a ____________ sample or a __________________ experiment.
  In this case, the data came from a _______________ sample of ______ Old Faithful eruptions in __________________.
  When sampling without replacement, check the $n \leq 10\%N$. In this case, ____ $\leq$ ____________

All of the conditions are _______________

Inference for Slope

Back to our question, "How can we estimate the standard deviation of the sampling distribution of the sample slope $\sigma_b = \frac{\sigma}{\sigma_x\sqrt{n}}$, when we do not know the values of σ or σ_x?"

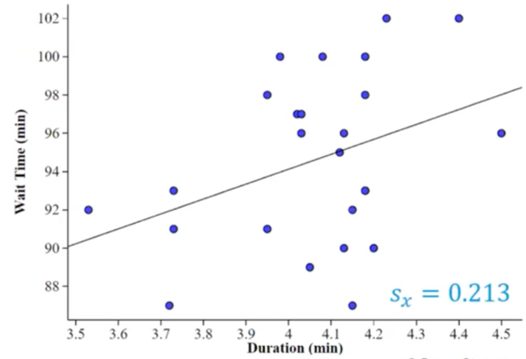
STATS MEDIC

Name________________________________

Estimating σ

We can use the sample regression _____________.

The ___________ standard deviation of the x-values is $s_x =$ _________. We can use s_x to estimate σ_x.



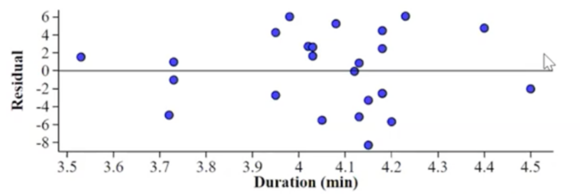
© Daren Starnes

Predictor	Coef	SE Coef	T-Value	P-Value
Constant	62.95	16.4	3.85	0.001
Duration(min)	7.79	4.03	1.94	0.065
$\boxed{\text{S=4.20970}}$	R-sq=14.01%		R-sq(adj)=10.27%	

The standard deviation of the residuals for the sample regression line:

$$s = \sqrt{\frac{\sum(y_i - \hat{y})^2}{n-2}} = \sqrt{\frac{\sum \text{residual}^2}{n-2}}$$

s estimates the size of the typical prediction error



© Daren Starnes

From the output, we can see that $s =$ _______.
We use this value to estimate σ.

Confidence Interval for Slope

Can we construct a confidence interval for the slope of the population regression line in 2019? YES!
The appropriate confidence interval is a ____.

What Should We Take Away?

How do we identify an appropriate confidence interval procedure for the slope of a population regression line?

The appropriate confidence interval is a ________________ because we are ____________ σ with *s*.

How do we verify the conditions for calculating a confidence interval for the slope of a population regression line?

*** The true relationship between ________ is linear. With _____________ and ____________________**

*** The standard deviation of y, σ_y, does vary with x. With _____________ and ___________________**

*** For a particular value of x, the *y*-values are approximately normally distributed. With _________ of the ______________. Need _______ if obvious _______________ or ____________.**

*** There is __________________ in data collection**

Data are collected using a _________ sample or a _______________ experiment.

When sampling ____________ replacement, check the ____________________.

STATS MEDIC

Name______________________________

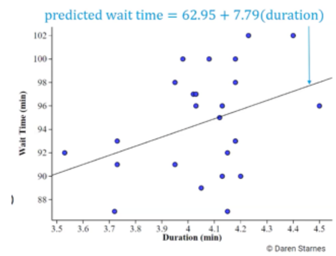
AP Statistics CED 9.2 Daily Video 3 (Skill 3.D)

Confidence Intervals for the Slope of a Regression Model

What Will We Learn?

How do we determine the margin of error when estimating the slope of a population regression line?
How do we calculate a confidence interval for the slope of a population regression line?

Predicting Old Faithful: 2019



The scatterplot displays data on the duration (in minutes) and wait time until the next eruption (in minutes) for a random sample of 25 Old Faithful eruptions in July 2019, along with the sample regression line.

Computer output from a least-square regression analysis is shown below.

Predictor	Coef	SE Coef	T-Value	P-Value
Constant	62.95	16.4	3.85	0.001
Duration(min)	7.79	4.03	1.94	0.065
S=4.20970	R-sq=14.01%		R-sq(adj)=10.27%	

Calculate and interpret a 95% confidence interval for the slope of the population regression line in 2019.

Calculating the Margin of Error

In AP Statistics, confidence intervals have the form:

CI = ________________ ± ________________

The margin of error describes how ________ a value of a ____________ statistic is ___________ to vary from the value of the ____________________ population __________________.

The margin of error is determined by _______ factors:

- How much the statistic ______________ varies from the ________________.
- How _______________ we want to be in our _____________________.

margin of error = (________________)(______________________)

Standard Error of Slope

The standard error of a statistic is an ______________ of the ______________________ of the sampling distribution of the __________________.

From topic 9.2, the standard deviation of the sampling distribution of b is:

$$\sigma_b = \frac{\sigma}{\sigma_x\sqrt{n}}$$

Because we don't know the value of σ or σ_x, we estimate them using the standard deviation of ________________ for the sample regression line, s, and the standard deviation of the x-values in the sample, s_x, to get the _____________________________:

$$SE_b = \frac{s}{s_x\sqrt{n-1}}$$

Calculating the Margin of Error

The critical value is a _______________ that makes the margin of error large enough to give a specific amount of confidence that the ______________ contains the value of the ________________.

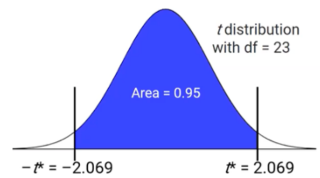
From confidence intervals for the ________ of a population regression line, the ___________________ represent the __________________ encompassing the middle _____ of the ____________________ with degrees of freedom _______________, where $C\%$ is the ____________________________.

Name________________________________

Calculating the Margin of Error

________________ = (critical value)(standard error of statistic)

In the Old Faithful example, we are asked to construct a 95% confidence interval based on a random sample of *n* = 25 eruptions. To find the critical value ______ for a 95% confidence interval, find the _______________ encompassing the middle ______ of the _______________ with _______________. This value can be found using Table B or using technology using InvT. Either way the critical value t^* = __________



Calculating the Confidence Interval

CI = ________________ ± _______________________

CI = ________________ ± (________________)(_______________________)

$$CI = b \pm t^* SE_b = b \pm t^* \frac{s}{s_x\sqrt{n-1}}$$

(Use this space to calculate the CI)

For our example about Old Faithful eruptions, b=_______ and s = ________. Unfortunately, s_x is not shown in the computer regression output. From the previous video, s_x=________.

Predictor	Coef	SE Coef	T-Value	P-Value
Constant	62.95	16.4	3.85	0.001
Duration(min)	7.79	4.03	1.94	0.065

S=4.20970 R-sq=14.01% R-sq(adj)=10.27%

SD of residuals standard error of slope

Calculating the Confidence Interval

All of the components needed to calculate a confidence interval can be found on the AP Statistics formula sheet. Make sure you can locate the information as you watch the video.

Factors that Affect Interval Width

Recall that confidence intervals in AP Statistics have the following structure:

CI = _________________ ± ___________________

The width of a confidence interval is ___________ the ________________________.

For a confidence interval about the slope of a population ______________________,

_____________ = ______________

We generally prefer _________ confidence intervals (________________), so we want the margin of error to be ____________. There are two common ways to decrease margin of error.

(1) ________ the ___________________ and (2) ___________ the _____________________________.

Calculating a Confidence Interval

Raoul performed an experiment using 16 windup rubber band single-propellor airplanes. He wound up the propeller a different number of times and recorded the amount of time (in seconds) that the airplane flew for each number of rotations that the propellor was wound. A regression analysis was performed and the partial computer output is given below. Assuming that the conditions for inference are satisfied and calculate a 95% CI.

Predictor	Coef	SE Coef	T	P
Constant	0.9241	0.6413	1.44	0.172
Rotation	0.04625	0.01565	2.96	0.010

S = 0.5426 R-Sq = 38.4% R-Sq(adj) = 34.0%

Name________________________________

Assuming that the conditions for inference are satisfied, which of the following is a 95 percent confidence interval for the slope of the regression line that relates to the number of rotations the rubber band is wound and plane's flight time?

(A) $0.04625 \pm (2.145)(0.01565)$
(B) $0.9241 \pm (2.145)(0.6413)$
(C) $0.04625 \pm (2.131)(0.01565)$
(D) $0.04625 \pm (2.131)\left(\frac{0.5426}{\sqrt{16}}\right)$
(E) $0.9241 \pm (2.131)(0.6413)$

$$\text{CI} = b \pm t^* \frac{s}{s_x\sqrt{n-1}} = b \pm t^* SE_b$$

What Should We Take Away?

How do we determine the margin of error when estimating the slope of a population regression line?

___________________ = (critical value)(standard error of statistic)

margin of error = ______________= ___________________

How do we calculate a confidence interval for the slope of a population regression line?

CI = _________________ ± ___________________

$$\text{CI} = b \pm t^* SE_b = b \pm t^* \frac{s}{s_x\sqrt{n-1}}$$

Name______________________________

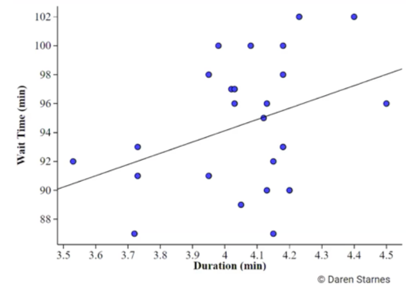
AP Statistics CED 9.3 Daily Video 1 (Skill 4.B)

What Will We Learn?

How do we interpret a confidence interval for the slope of a population regression line?
How do we justify a claim based on a confidence interval for slope?

Old Faithful: 2019

The scatterplot displays data on the duration (in minutes) and wait time until the next eruption (in minutes) for a random sample of 25 Old Faithful eruptions in July 2019, along with the sample regression line. Computer output from the least-squares regression analysis is shown below:




```
Predictor        Coef   SE Coef   T-Value   P-Value
Constant         62.95   16.4      3.85      0.001
Duration(min)    7.79    4.03      1.94      0.065
S=4.20970        R-sq=14.01%      R-sq(adj)=10.27%
```

Calculate and interpret a 95% confidence interval for the slope of the population regression line in 2019.

Interpreting the Confidence Interval

In general, here is how to interpret a confidence interval for the slope of a regression model:
"We are ________ confident that the _____________ from ____ to ____ ______________ the slope of the population regression line [_________________].
From Topic 9.2 Video 3, the 95% confidence interval is -0.55 to 16.13:
"We are _______ confident that the ___________ from _____ to _____ ______________ the slope of the ___________________ regression line for predicting ________________ until the next eruption (in minutes) from the _______________ of the previous eruption (in minutes) for _________ Old Faithful geyser eruptions in July 2019."

Interpreting Confidence Level

In ___________ random sampling with the _________ sample size, approximately *C*% of "*C*%" confidence intervals created _________________ the slope of the ________________ regression line.
If we take ________ random samples of size _____ from the population of Old Faithful eruptions in ______________________, and use each sample to ________________ a 95% confidence interval for the slope of the _____________________ regression line for predicting _________________ until the next eruption (in minutes) from the ___________ of the previous eruption (in minutes), about _______ of those ___________ would ______________ the __________ slope.

Justifying a Claim

Does the confidence interval (-0.55 to 16.13) provide convincing evidence that wait time until the next eruption of Old Faithful is linearly related to the duration of the previous eruption in July 2019?
Note that $\beta = 0$ would indicate a line with a _______________ is the model for ___________________ wait time from eruption duration. In other words, the regression model would predict the _________ wait time until the next Old Faithful eruption no matter the duration of the previous eruption.

Because the confidence interval (_____________________) contains ______ as a _________________ value of the slope of the ________________ regression line, there is __________________________ evidence that ________________ until the next eruption of Old Faithful is ________________ related to the _______________ of the previous eruption in July 2019.

STATS MEDIC

Name______________________________

Interpreting a Confidence Interval

Raoul performed an experiment using 16 windup rubber band single-propellor airplanes. He wound up the propeller a different number of times and recorded the amount of time (in seconds) that the airplane flew for each number of rotations that the propellor was wound. A regression analysis was performed, and the conditions for inference were verified. A 95% confidence interval for the slope of the regression line that relates the number of rotations the rubber band is would and the plane's flight time is given by (0.013, 0.080). Which of the following provides a correct interpretation of the confidence interval?

(A) There is a 0.95 probability that the slope of the population regression line that relates the number of rotations the rubber band is wound and the plane's flight time is between 0.013 and 0.080.
(B) If the data collection process were repeated many times, about 95% of the resulting sample regression lines would have slopes between 0.013 and 0.080.
(C) If the data collection process were repeated many times, about 95% of the resulting confidence intervals would contain the slope of the sample regression line.
(D) We are 95% confident that the slope of the sample regression line that relates the number of rotations the rubber band is wound and the plane's flight time is between 0.013 and 0.080.
(E) We are 95% confident that the slope of the population regression line that relates the number of rotations the rubber band is wound and the plane's flight time is between 0.013 and 0.080.

Justifying a Claim with a CI

(A) There is not convincing evidence of a linear relationship between the number of rotations of the rubber band and the flight times of these 16 windup airplanes because 0 is not included in the interval.
(B) There is not convincing evidence of a linear relationship between the number of rotations of the rubber band and the flight times of windup airplanes like these because 0 is not included in the interval.
(C) There is convincing evidence of a positive linear relationship between the number of rotations of the rubber band and the flight times of these 16 windup airplanes because all values in the interval are positive.
(D) There is convincing evidence of a positive linear relationship between the number of rotations of the rubber band and the flight times of windup airplanes like these because all values in the interval are positive.
(E) There is convincing evidence that there is not a linear relationship between the number of rotations of the rubber band and the flight times of windup airplanes like these because 0 is not included in the interval.

What Should We Take Away?

How do we interpret a confidence interval for the slope of a population regression line?

"We are ________ confident that the _____________ from ____ to ____ ______________ the slope of the population regression line [_________________].

How do we justify a claim based on a confidence interval for slope?

- **If all the values in the confidence interval are __________________ with the claim, there _____ convincing evidence for the claim.**
- **If one or more of the values in the confidence interval are __________________ with the claim, there is _______ convincing evidence for the claim.**

Name______________________________

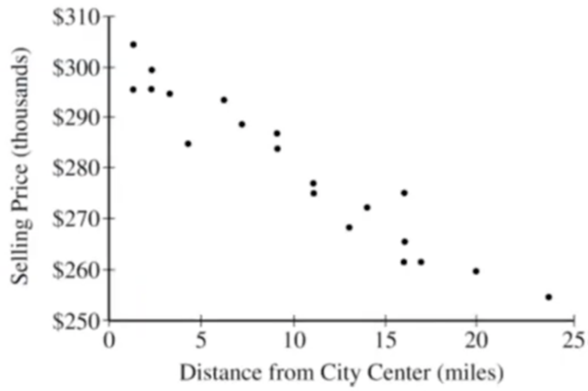
AP Statistics CED 9.3 Daily Video 2 (Skill 3.E)

What Will We Learn?

How do we construct and interpret a confidence interval for the slope of a population regression line?

2019 International Exam #2

A real estate agent working in a large city believes that, for three-bedroom houses, the selling price of the house decreases by approximately $2,000 for every mile increase in the distance of the house from the city center. To investigate the belief, the agent obtained a random sample of 20 three-bedroom houses that sold in the last year. The selling price, in thousands of dollars and the distance from the city center, in miles, for each of the 20 houses are shown in the scatterplot. The table shows computer output from a regression analysis of the data.



Predictor	Coef	SE Coef	T	P
Constant	301.7	1.80	167.17	0.000
Distance	−2.158	0.149	−14.45	0.000
	S = 4.4336		R-sq = 92.1%	

(a) Assume all conditions for inference are met. Construct and interpret a 95 percent confidence interval for the slope of the least-squares regression line.

(b) Does the confidence interval contradict the agent's belief about the relationship between selling price and distance from the city center? Justify your answer.

2019 International Exam #2(a) Calculate Interval

(a) Assume all conditions for inference are met. Construct and interpret a 95 percent confidence interval for the slope of the least-squares regression line.

We will construct a ______ confidence interval for β= _________ of the population regression line for _______________ selling price (in thousands) from distance from the city center (in miles) for all three-bedroom houses near this city.

We will use a ________________________________ and fortunately conditions are ___________.

(Use the output above to create the *t*-interval.)

b = __________ (Find on Output)   df = _______________ t^* = ______________ (use technology InvT)

Formula: CI = $b \pm t^* SE_b$

= ________________________________

= ________________________________

= ________________________________

2019 International Exam #2, Part (a) Interpret Interval   Given: 95% CI = -2.471 to -1.845

We are _______________ that the slope of the ________ regression line is between ______ and ______ thousands of ______________________. This implies that for ________ additional mile that a three-bedroom house is away from the _____________, the selling prince of the house is expected to _____________ between _____________ and _____________.

Name______________________________

2019 International Exam #2

A real estate agent working in a large city believes that, for three-bedroom houses, the selling price of the house decreases by approximately $2,000 for every mile increase in the distance of the house from the city center. To investigate the belief, the agent obtained a random sample of 20 three-bedroom houses that sold in the last year. The selling price, in thousands of dollars and the distance from the city center, in miles, for each of the 20 houses are shown in the scatterplot. The table shows computer output from a regression analysis of the data.

(b) Does the confidence interval contradict the agent's belief about the relationship between selling price and distance from the city center? Justify your answer.

Given: 95% CI = -2.471 to -1.845, Interpret the interval

Because the confidence interval contains ______, corresponding to a ___________ decrease, it is a _______________ value for the slope of the regression line. Consequently, the data _____________ contradict the agent's belief that the selling prices of three-bedroom houses ____________ about _____________ for every ____________ increase in distance of the house from the city center.

What Should We Take Away?

How do we construct and interpret a confidence interval for the slope of a population regression line?

Make sure to:

- **Define the ______________ you are trying to estimate.**
- **Identify the ______________ you are using.**
- **Verify that the _____________ for the procedure are _______.**
- **______________ the confidence interval.**
- **______________ the interval in ______________.**

Name________________________________

AP Statistics CED 9.4 Daily Video 1 (Skill 1.F)

Setting Up a Test for the Slope of a Regression Model

What Will We Learn?

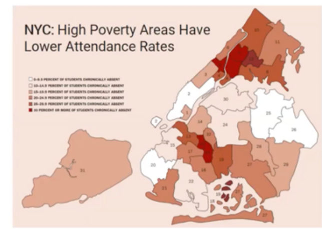
How do you state a null hypothesis in a test about the slope of a population regression line?

How do you state an alternative hypothesis in a test about the slope of a population regression line?

Schools and "Equal Opportunity"

Can education systems equalize opportunities for lower income students?

The Income Attendance Gap

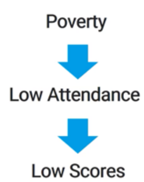


Nationally, higher income areas tend to have fewer chronically absent students. Possible reasons:

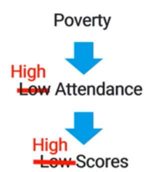
- Transportation access
- Work to support family

Is Attendance the Solution?

If



then does

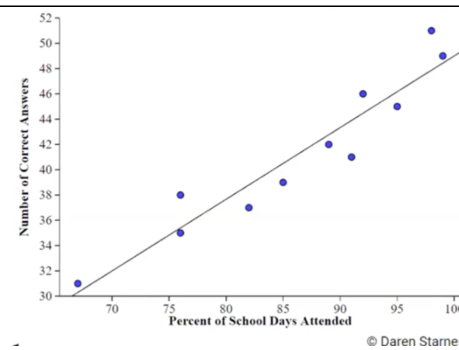


Some school systems have targeted attendance as the key to raising test scores for lower income students.

Let's Look at the Data

Researchers collected data on the percent of school days attended and the number of questions answered correctly on the state's end-of-course test for a random sample of 11 Texas Algebra 1 students. Here are the data, along with a scatterplot and computer output from a least-squares regression analysis.

Percent attendance	95	89	67	98	99	76	92	91	76	85	82
Questions correct	45	42	31	51	49	38	46	41	35	39	37



Predictor	Coef	SE Coef	T	P
Constant	−7.69	5.37	−1.43	0.186
Attendance	0.57	0.062	9.18	0.000

S = 1.99 R-Sq = 90.3% R-Sq (adj) = 89.3%

Do the data give convincing evidence at the $\alpha = 0.01$ significance level of a positive, linear relationship between state test score and percent attendance for Texas Algebra 1 students?

Null Hypothesis

In a statistical test, the _______ hypothesis is often a claim of "_________________", "______________", or "_________________________". In the attendance and test scores example, the null hypothesis is that there is ___________________________ between the percent of schools days attended and the number of questions answered correctly on the state test by Texas Algebra I students. In other words, the __________________ regression line would have a _______________. In symbols:

***H*$_0$: *β* = ____**; where β = _________ of the ________________ regression line for _________________ number of questions answered correctly on the state test from the percent of schools day attended for Texas Algebra 1 students.

Until we have ________________________ otherwise, we assume the __________ hypothesis is correct.

STATS MEDIC

Name______________________________

Alternative Hypothesis

In a statistical test, the _______________ hypothesis is the ________ that we hope to support with _____________ from the data collected. In the attendance and test scores example, the researchers wanted to know if there is a ____________, ____________ relationship between the ______________ of school days attended by Texas students and the _____________ of questions answered correctly on the state Algebra 1 test. So the alternative hypothesis is that the population regression line would have a ________________. In symbols:

***H_0*: $\boldsymbol{\beta}$ > ____**; where β = ___________ of the population regression line for predicting number of questions answered correctly on the state test from the percent of school days attended for Texas Algebra 1 students.

Stating Hypotheses: Summary

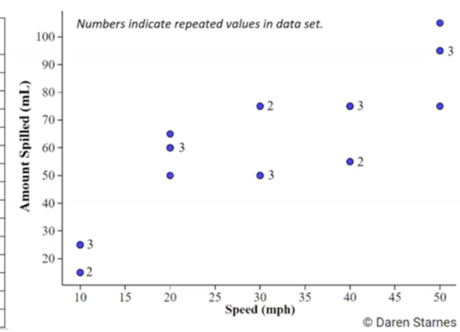
For hypotheses about the ________ of a population regression line:

- The null is a statement of ___________, typically ________________.
- The alternative always contains a strict ___________, typically ____________________________.
  When the inequality is _____________, the alternative is called _________________.
  When the inequality is _____________, the alternative is called _________________.
  The choice of alternative is determined by the ___________________ of interest and should be stated ____________ data collection begins.
- Never refer to ____________ (such as *b*) in the hypotheses!
- Remember to _____________ the _____________________.

Don't Spill my Drink!

Two AP Statistics students wondered if there is a linear relationship between speed and amount of drink spilled when driving on bumpy dirt roads. To find out, they filled a cup with 275 mL of water and placed it in the car's cup holder, then drove down a bumpy road at a specified speed and recorded how much water spilled out. The 25 trials were randomly assigned to be 5 trials each at 10, 20, 30, 40, or 50 miles per hour (mph). Here are the data. (Note that several data points have the same values for speed and amount spilled.) A scatterplot of the data, along with computer output from a least-squares regression analysis are shown. Do the data provide convincing evidence at the 0.05 significance level of a linear relationship between the car's speed on a bumpy dirt road and the amount of drink spilled? Assume the conditions for inference are met.

Speed (mph)	Spilled (mL)	Speed (mph)	Spilled (mL)
10	25	40	75
10	25	40	75
10	25	40	75
10	15	40	55
10	15	40	55
20	60	50	95
20	60	50	95
20	65	50	95
20	60	50	75
20	50	50	105
30	50		
30	75		
30	50		
30	75		
30	50		




```
Term         Coef   SE Coef
Constant     14.40   5.95
Speed (mph)  1.520   0.179
S=12.6800    R-sq = 75.75%
```

Stating the Hypotheses

***H_0*: _______________ and *H_a*: _______________; where____________________________________**

____

What Should We Take Away?

How do you state a null hypothesis in a test about the slope of a population regression line?

$H_0: \beta = \beta_0$ **For a null hypothesis of _____ linear relationship, ______________**

Note: remember to clearly define the ____________ $\boldsymbol{\beta}$.

How do you state an alternative hypothesis in a test about the slope of a population regression line?

$H_a: \beta > \beta_0$
$H_a: \beta < \beta_0$
$H_a: \beta \neq \beta_0$

For a test of a positive linear relationship, ______________. For the test of a negative relationship ______________. For the test of a linear relationship, ______________.

STATS MEDIC

Name________________________________

AP Statistics CED 9.4 Daily Video 2 (Skill 4.C)

Setting Up a Test for the Slope of a Regression Model

What Will We Learn?

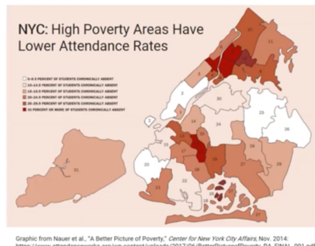
How do we identify an appropriate significance test procedure for the slope of a population regression line?

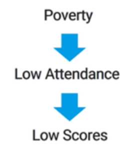
How do we verify the conditions for performing a test about the slope of a population regression line?

Is Attendance the Solution?

Nationally, higher income areas tend to have fewer chronically absent students. Possible reasons:

- Transportation access
- Work to support family





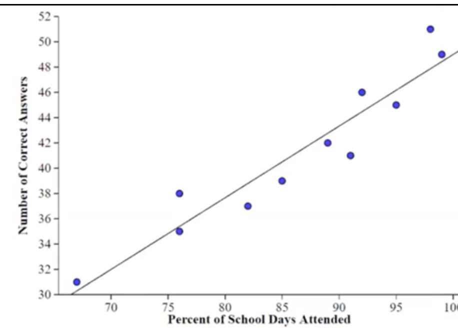
then does

Some school systems have targeted attendance as the key to raising test scores for lower income students.

Let's Look at the Data

Researchers collected data on the percent of school days attended and the number of questions answered correctly on the state's end-of-course test for a random sample of 11 Texas Algebra 1 students. Here are the data, along with a scatterplot and computer output from a least-squares regression analysis.

Percent attendance	95	89	67	98	99	76	92	91	76	85	82
Questions correct	45	42	31	51	49	38	46	41	35	39	37



© Daren Starnes

Predictor	Coef	SE Coef	T	P
Constant	−7.69	5.37	−1.43	0.186
Attendance	0.57	0.062	9.18	0.000
S = 1.99	R-Sq = 90.3%		R-Sq (adj) = 89.3%	

Do the data give convincing evidence at the $\alpha = 0.01$ significance level of a positive, linear relationship between state test score and percent attendance for Texas Algebra 1 students?

Test Scores and Attendance - Hypotheses

In a previous video, we stated the hypotheses: H_0: $\beta = 0$ and H_a: $\beta > 0$; where β = the slope of the population regression line for predicting number of questions answered correctly on the state test from the percent of schools day attended for Texas Algebra 1 students.

Identifying the Procedure

You have learned many different significance test procedures this year. Some involved one sample and others involve two samples. Some involved inference about categorical data: _______________ or _______________. Others involved inference about quantitative data: __________ or _____________.

When the goal is to test a claim about the __________ of a __________________ regression line, we use a __________________________.

Checking Conditions

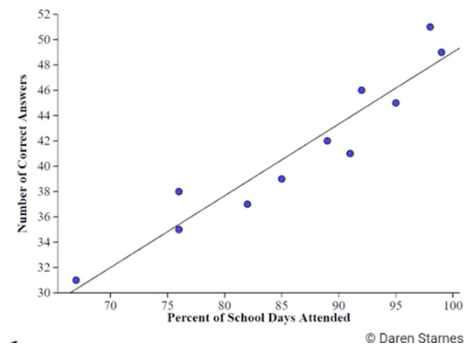
To perform a significance test about the slope of β of a population regression line, we must check the following:

- The true relationship between ______________ is linear.
- The standard deviation of _________ doe not vary with ______/
- For a particular value of x the ___________ (y – values) are approximately normally distributed.
- There is independence in data collection: Data are collected using __________ sample of a randomized experiment. When sampling without replacement check that ________________.

Name_______________________________

Checking Conditions (Be sure to ✓ your conditions!)

To perform a significance test about the slope β of a population regression line, we must check the following:

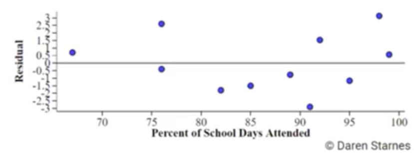


- The true relationship between x and y is linear.

Scatterplot: The scatterplot show a ________________________, _________________________ relationship between the percent of school days attended by a student and the number of correct answers on the state Algebra 1 test.

Residual Plot: The residual plots shows fairly _________________ scatter about ______________________ line and no obvious curved pattern.

- The standard deviation of y, σ_y does not vary with x.



Scatterplot: The number of correct answers vary by a ___________ amount for the different percents of school days attended in the data set.

Residual Plot: The residuals are ____________ in size for the different percents of school days attended in the data set.

- For a particular value of x, the responses (y – values) are approximately normally distributed.

Dotplot: A dotplot of residuals shows no obvious ______________ or _________________.

- There is independence in data collection.

Data are collected using a random sample or a randomized experiment. The data come from a _______________ sample of _______ Texas Algebra 1 students.

When sampling without replacement, check that $n \leq 10\%N$. ________(all Texas Algebra 1 students)

All of the conditions are _________.

What Should We Take Away?

How do we identify an appropriate significance test procedure for the slope of a population regression line?

When testing a claim about the slope of a population regression line, use a ___________________

How do we verify the conditions for performing a test about the slope of a population regression line?

*** The true relationship between ________ is linear. With ______________ and _____________________**

*** The standard deviation of y, σ_y, does vary with x. With ______________ and _____________________**

*** For a particular value of x, the y-values are approximately normally distributed. With _________ of the _______________. Need __________ if obvious ________________ or _____________.**

*** There is ___________________ in data collection**

Data are collected using a __________ sample or a ________________ experiment.

When sampling _____________ replacement, check the ______________________.

Name______________________________

AP Statistics CED 9.5 Daily Video 1 (Skill 3.E)

Carrying Out a Test for the Slope of a Regression Model

What Will We Learn?

How do we calculate an appropriate test statistic in a test about the slope of a population regression line?

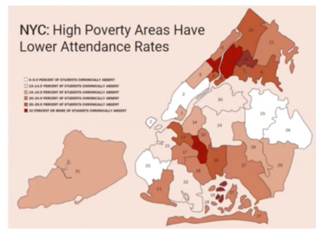
How do we calculate a *p*-value in a test about the slope of a population regression line?

Is Attendance the Solution?

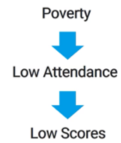
Nationally, higher income areas tend to have fewer chronically absent students.

Possible reasons:

- Transportation access
- Work to support family



Graphic from Nauer et al., "A Better Picture of Poverty," Center for New York City Affairs, Nov. 2014. https://www.attendanceworks.org/wp-content/uploads/2017/06/BetterPictureofPoverty_PA_FINAL_001.pdf



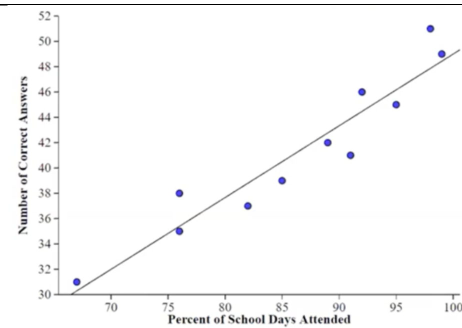
then does

Some school systems have targeted attendance as the key to raising test scores for lower income students.

Let's Look at the Data

Researchers collected data on the percent of school days attended and the number of questions answered correctly on the state's end-of-course test for a random sample of 11 Texas Algebra 1 students. Here are the data, along with a scatterplot and computer output from a least-squares regression analysis.

Percent attendance	95	89	67	98	99	76	92	91	76	85	82
Questions correct	45	42	31	51	49	38	46	41	35	39	37



Predictor	Coef	SE Coef	T	P
Constant	-7.69	5.37	-1.43	0.186
Attendance	0.57	0.062	9.18	0.000
S = 1.99	R-Sq = 90.3%		R-Sq (adj) = 89.3%	

Do the data give convincing evidence at the α = 0.01 significance level of a positive, linear relationship between state test score and percent attendance for Texas Algebra 1 students?

Test Scores and Attendance – Hypotheses

In a previous video, we stated the hypotheses: H_0: β = 0 and H_a: β > 0; where β = the slope of the population regression line for predicting number of questions answered correctly on the state test from the percent of schools day attended for Texas Algebra 1 students. We will use α = 0.01. We will use a *t*-test for slope and the conditions have all been met.

Calculating a Test Statistic

Predictor	Coef	SE Coef	T	P
Constant	-7.69	5.37	-1.43	0.186
Attendance	$\boxed{0.57}$	0.062	9.18	0.000
S = 1.99	R-Sq = 90.3%		R-Sq (adj) = 89.3%	

In the attendance and test score study, b = _____________.

This is evidence for H_a: β > ___ because b = ______ > __.

We want to know how ________________it is to get evidence for H_a this _______________________ by ________________ alone when H_0 is ___________.

After verifying the conditions are met, calculate the standardized test statistics:

$$standardized\ tests\ statistic = \frac{statistic - parameter}{standard\ error\ of\ the\ statistic}$$ from formula sheet!!

Calculating a Test Statistic

Predictor	Coef	SE Coef	T	P
Constant	-7.69	5.37	-1.43	0.186
Attendance	$\boxed{0.57}$	$\boxed{0.062}$	9.18	0.000
S = 1.99	R-Sq = 90.3%		R-Sq (adj) = 89.3%	

For a *t*-test for a slope, the standardized test statistic is: $t = \frac{b-\beta_0}{SE_b}$, where β_0 is the value of β specified by the null hypothesis. So, we would have $t = \frac{\_\_ - \_\_}{\_\_\_\_\_\_} =$ ____________

*** All components can be found on the AP Statistics Formula Sheet! Make sure you can locate them!**

STATS MEDIC

Name_______________________________

Calculating the *p*-value

Once we have calculated the standardized test statistic, use the _____________ with df = __________ to calculate the ___________. The *p*-value is the _________________ of values for the __________ distribution that are as ___________________________ than the observed value of the test statistic.

Predictor	Coef	SE Coef	T	P
Constant	−7.69	5.37	−1.43	0.186
Attendance	0.57	0.062	$\boxed{9.18}$	$\boxed{0.000}$
S = 1.99	R-Sq = 90.3%	R-Sq (adj) = 89.3%		

Because our alternative hypothesis is H_a: $\beta > 0$, we want to find $P(t \geq$ ________) in a *t* distribution with df = ___________.

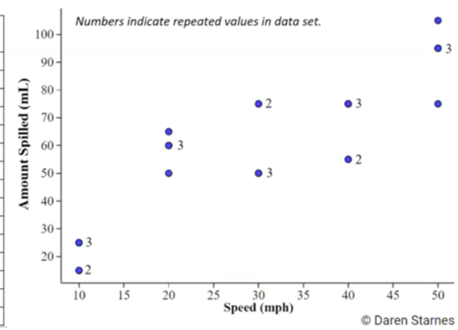
Note: divide by 2 for a one-tailed

You can use Table B or Technology to find the *p*-value = ______________ or use the output!

Don't Spill my Drink!

Two AP Statistics students wondered if there is a linear relationship between speed and amount of drink spilled when driving on bumpy dirt roads. To find out, they filled a cup with 275 mL of water and placed it in the car's cup holder, then drove down a bumpy dirt road at a specified speed and recorded how much water spilled out. The 25 trials were randomly assigned to be 5 trials each at 10, 20, 30, 40, or 50 miles per hour (mph). Here are the data. (Note that several data points have the same values for speed and amount spilled.) A scatterplot of the data, along with computer output from a least-squares regression analysis are shown. Do the data provide convincing evidence at the 0.05 significance level of a linear relationship between the car's speed on a bumpy dirt road and the amount of drink spilled? Assume the conditions for inference are met. Additionally, we have previously determined the hypotheses.

Speed (mph)	Spilled (mL)	Speed (mph)	Spilled (mL)
10	25	40	75
10	25	40	75
10	25	40	75
10	15	40	55
10	15	40	55
20	60	50	95
20	60	50	95
20	65	50	95
20	60	50	75
20	50	50	105
30	50		
30	75		
30	50		
30	75		
30	50		



Term	Coef	SE Coef
Constant	14.40	5.95
Speed (mph)	1.520	0.179
S=12.6800	R-sq = 75.75%	

Don't Spill My Drink!

Calculate the standardized test statistic and *p*-value. From the previous video we know that:

H_0: ____________ vs H_a: ____________; where β = __________________________ line for predicting the amount of drink spilled (in mL) from the car's speed (in mph) on a bumpy dirt road. We will use α = _____ because no significance level was stated. We will conduct a *t*-test for slope.

As a reminder: $t = \frac{b-\beta_0}{SE_b}$

Calculating a *p*-value (Remember to use Table B or technology!)

Because our alternative hypothesis is H_a: $\beta \neq 0$, we want to find $P(t \leq -8.49) + P(t \geq 8.49)$ in a *t* distribution with df = ____________ = _______.

p-value = ________________

What Should We Take Away?

How do we calculate an appropriate test statistic in a test about the slope of a population regression line? **The Formula is $t = \frac{b-\beta_0}{SE_b}$. But, often this information can be found on the computer output!**

How do we calculate a *p*-value in a test about the slope of a population regression line?

If ______, *p*-value = *P*($t \geq$ observed test statistics) If ______, *p*-value = *P*($t \geq$ observed test statistics)

If ______, *p*-value = 2✖*P*($t \geq$ |observed test statistics|) ; in a *t* distribution with ______ df.

Name______________________________

AP Statistics CED 9.5 Daily Video 2 (Skill 4.B)

Carrying Out a Test for the Slope of a Regression Model

What Will We Learn?

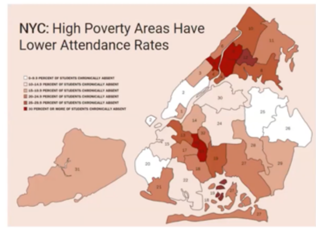
How do we interpret the *p*-value in a test about the slope of a population regression line?
How do we state a conclusion in a test about the slope of a population regression line?

Is Attendance the Solution?

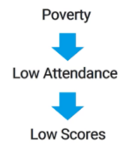
Can education systems equalize opportunities for lower income students?

Nationally, higher income areas tend to have fewer chronically absent students. Possible reasons:

- Transportation access
- Work to support family



Graphic from Nauer et al., "A Better Picture of Poverty," *Center for New York City Affairs*, Nov. 2014: https://www.attendanceworks.org/wp-content/uploads/2017/06/BetterPictureofPoverty_PA_FINAL_001.pdf



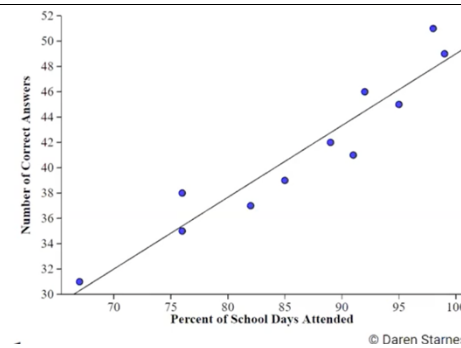
then does

Some school systems have targeted attendance as the key to raising test scores for lower income students.

Let's Look at the Data

Researchers collected data on the percent of school days attended and the number of questions answered correctly on the state's end-of-course test for a random sample of 11 Texas Algebra 1 students. Here are the data, along with a scatterplot and computer output from a least-squares regression analysis.

Percent attendance	95	89	67	98	99	76	92	91	76	85	82
Questions correct	45	42	31	51	49	38	46	41	35	39	37



Predictor	Coef	SE Coef	T	P
Constant	−7.69	5.37	−1.43	0.186
Attendance	0.57	0.062	9.18	0.000

S = 1.99 R-Sq = 90.3% R-Sq (adj) = 89.3%

Do the data give convincing evidence at the $\alpha = 0.01$ significance level of a positive, linear relationship between state test score and percent attendance for Texas Algebra 1 students?

Test Scores and Attendance - Hypotheses In a previous video, we stated the hypotheses: H_0: $\beta = 0$ and H_a: $\beta > 0$; where β = the slope of the population regression line for predicting number of questions answered correctly on the state test from the percent of schools day attended for Texas Algebra 1 students. Use $\alpha = 0.01$. We will use a *t*-test for slope and the conditions have all been met.

Interpreting *p*-value - From previous videos:

b = _________, t = __________ and *p*=value ≈ ____________

b = _________ is evidence ____________ H_0: β=____ and for H_a: β >____, because ________________.

The *p*-value measures how ____________ it is to get evidence for H_a as ________________________ than the observed evidence by ______________ alone when H_0 is ____________. Assuming there is _______________ relationship between state test score and percent attendance for Texas Algebra 1 students, there is an ___________________ probability of getting a sample regression line with slope ________ or ____________ by ___________ alone in a random sample of 11 Texas Algebra 1 students.

Making a Conclusion

Small *p*-values ⟶ test statistic ______________ to occur by ______________ chance alone.

Large *p*-values ⟶ test statistic _________________ to occur by ______________ chance alone.

- Because the *p*-value of ________ ≤ α = _____, we reject H_0.
  There is convincing ______________ evidence that [_________________].
- Because the *p*-value of ________ ≤ α = _____, we fail to reject H_0.
  There is not convincing ______________ evidence that [_________________].

STATS MEDIC

Name______________________________

Making a Conclusion

Do the data give convincing evidence at the α = 0.01 significance level of a positive, linear relationship between state test score and percent attendance for Texas Algebra 1 students?

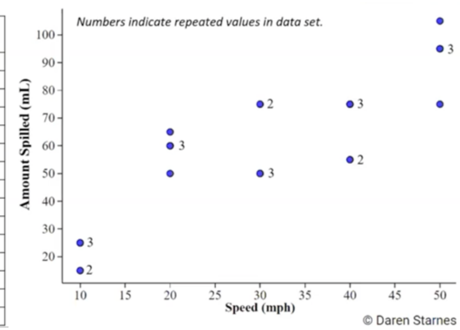
- H_0: β _____ vs. H_a: β _____, where β = the slope of the population regression line for predicting number of questions answered correctly on the state test from the percent of school days attended for Texas Algebra 1 students.
- Conditions are _______
- b = _________, t = __________ and p=value ≈ ____________

Because the p-value of ______________________ < ____________, we reject H_0. There _____ convincing statistical evidence of a ________________________________ between state test score and percent attendance for Texas Algebra 1 students.

Don't Spill my Drink!

Two AP Statistics students wondered if there is a linear relationship between speed and amount of drink spilled when driving on bumpy dirt roads. To find out, they filled a cup with 275 mL of water and placed it in the car's cup holder, then drove down a bumpy road at a specified speed and recorded how much water spilled out. The 25 trials were randomly assigned to be 5 trials each at 10, 20, 30, 40, or 50 miles per hour (mph). Here are the data. (Note that several data points have the same values for speed and amount spilled.) A scatterplot of the data, along with computer output from a least-squares regression analysis are shown. Do the data provide convincing evidence at the 0.05 significance level of a linear relationship between the car's speed on a bumpy dirt road and the amount of drink spilled? Assume the conditions for inference are met.

Speed (mph)	Spilled (mL)	Speed (mph)	Spilled (mL)
10	25	40	75
10	25	40	75
10	25	40	75
10	15	40	55
10	15	40	55
20	60	50	95
20	60	50	95
20	65	50	95
20	60	50	75
20	50	50	105
30	50		
30	75		
30	50		
30	75		
30	50		



Term	Coef	SE Coef
Constant	14.40	5.95
Speed (mph)	1.520	0.179
S=12.6800	R-sq = 75.75%	

Interpreting the p-value

From previous videos:

- H_0: β _____ vs. H_a: β _____, where β = the slope of the ________ regression line for predicting amount of _________ spilled (in mL) from the car's speed (in mph) on a bumpy dirt road.
- t-test for slope: Conditions are met.
- b = _________, t = __________ and p=value ≈ ____________

Interpret the p-value.

Assuming that there is ____________ relationship between the car's speed on a bumpy dirt road and the amount of drink spilled, there is a __________________ probability of getting a sample regression line with a slope as ________________________ than _________ in either direction by ________ alone.

Making a Conclusion

Do the data provide convincing evidence at the 0.05 significance level of a linear relationship between the car's speed on a bumpy dirt road and the amount of drink spilled? From previous video:

- H_0: β _____ vs. H_a: β _____, where β = the slope of the ________ regression line for predicting amount of _________ spilled (in mL) from the car's speed (in mph) on a bumpy dirt road.
- Conditions are met. And b = _________, t = __________ and p=value ≈ ____________

Because the p-value of _____________ < __________, we ____________. There is convincing statistical evidence of a _______________ relationship between the car's speed on a bumpy dirt road and the amount of drink spilled.

Name________________________________

What Should We Take Away?

How do we interpret the *p*-value in a test about the slope of a population regression line?

The *p*-value measures how ____________ it is to get evidence for H_a as ____________ or ___________ than the observed evidence by chance along with H_0 is ____________.

How do we state a conclusion in a test about the slope of a population regression line?

- **Because the *p*-value of ________ $\leq \alpha$ = _____, we reject H_0.**
  There is convincing ______________ evidence that [_________________].
- **Because the *p*-value of ________ $\leq \alpha$ = _____, we fail to reject H_0.**
  There is not convincing ______________ evidence that [_________________].

Name______________________________

AP Statistics CED 9.5 Daily Video 3 (Skill 4.B)

Carrying Out a Test for the Slope of a Regression Model

What Will We Learn?

How do we perform a complete significance test about the slope of a population regression line?

2001 Exam #6 (Modified)

The Statistics Department at a large university is trying to determine if it is possible to predict whether an applicant will successfully complete the Ph.D. program or will leave before completing the program. The department is considering whether GPA (grade point average) in undergraduate statistics and mathematics courses (a measure of performance) and mean number of credit hours per semester (a measure of workload) would be helpful measures. To gather data, a random sample of 20 entering students from the past 5 years is taken. The data are given below:

Successfully Completed Ph.D. Program

Student	A	B	C	D	E	F	G	H	I	J	K	L	M
GPA	3.8	3.5	4.0	3.9	2.9	3.5	3.5	4.0	3.9	3.0	3.4	3.7	3.6
Credit hours	12.7	13.1	12.5	13.0	15.0	14.7	14.5	12.0	13.1	15.3	14.6	12.5	14.0

Did Not Complete Ph.D. Program

Student	N	O	P	Q	R	S	T
GPA	3.6	2.9	3.1	3.5	3.9	3.6	3.3
Credit hours	11.1	14.5	14.0	10.9	11.5	12.1	12.0

The regression output below resulted from fitting a line to the data in the group of 7 students that did not complete the Ph.D. program. The residual plot (not shown) indicated no unusual patterns, and the assumptions necessary for inference were judged to be reasonable.

Did Not Complete Ph.D. Program

Predictor	Coef	StDev	T	P
Constant	24.200	3.474	6.97	0.001
GPA	−3.485	1.013	−3.44	0.018
S = 0.8408		R-Sq = 70.3%		

(a) For the students who did not complete the Ph.D. program, is there a significant relationship between GPA and mean number of credit hours per semester at the $\alpha = 0.01$ level?

H_0: $\beta = 0$ and H_a: $\beta \neq 0$; where β= the slope of the _____________ regression line for predicting ________ number of credit hours per semester from _______ for students who did ________ complete the statistics Ph.D. program at this large university. Use α = ___; t-test for slope; Conditions are met!

2001 Exam #6 (Modified)

(a) For the students who did not complete the Ph.D. program, is there a significant relationship between GPA and mean number of credit hours per semester at the $\alpha = 0.01$ level?

Did Not Complete Ph.D. Program

Predictor	Coef	StDev	T	P
Constant	24.200	3.474	6.97	0.001
GPA	−3.485	1.013	$\boxed{-3.44}$	0.018
S = 0.8408		R-Sq = 70.3%		

t = ___________; df = ___________; p-value =_________

Because the p-value of ____________________, we __________________. There is ______ convincing evidence of a ___________ relationship between GPA and mean number of credit hours in the population of students at this university who did not complete the statistics Ph.D. program.

2001 Exam #6 (Modified)

The regression output below resulted from fitting a line to the data in the group of 13 students that successfully completed the Ph.D. program. The residual plot (not shown) indicated no unusual patterns, and the assumptions necessary for inference were judged to be reasonable.

Successfully Completed Ph.D. Program

Predictor	Coef	StDev	T	P
Constant	23.514	1.684	13.95	0.000
GPA	−2.7555	0.4668	−5.90	0.000
S = 0.5658		R-Sq = 76.0%		

(b) For the students who successfully completed the Ph.D. program, is the evidence for a significant (linear) relationship between GPA and mean number of credit hours per semester stronger or weaker than for the students who did not complete the Ph.D. program? Justify your answer.

Name________________________________

2001 Exam #6 (Modified)

Hint: You do not need to run a full test to answer this question. You only need to use the computer output to the right.

Successfully Completed Ph.D. Program

Predictor	Coef	StDev	T	P
Constant	23.514	1.684	13.95	0.000
GPA	−2.7555	0.4668	−5.90	0.000
S = 0.5658	R-Sq = 76.0%			

_____________. The *p*-value for a test of H_0: $\beta = 0$ and H_a: $\beta \neq 0$; where β = the ___________ of the population regression line for ______________________ number of credit hours per semester _______ GPA for students who successfully completed the statistics Ph.D. program at this large university, is ______________________. This *p*-value is __________ than the __________ *p*-value for the test in part (a), and gives ______________ evidence of a _______________ relationship between GPA and the mean number of credit hours per semester. You could also compare the ____________ from the tests!

What Should We Take Away?

How do we perform a complete significance test about the slope of a population regression line?

Make sure to:

- **State the _________ and_____________ hypotheses, and define _________________.**
- **Give the ______________________________.**
- **Identify the __________________ you are using.**
- **Verify that the _________________ for the procedure are ________.**
- **Calculate the ________________ and the _____________.**
- **Make a _____________________ based on the _______________. (You do _______ need to interpret the *p*-value unless specifically asked.)**

Name________________________________

AP Statistics CED 9.6 Daily Video 1

Carrying Out a Test for the Slope of a Regression Model

What Will We Learn?

What inference methods did we learn about in Units 6 – 9?

How can we identify the appropriate inference procedure to use in a given setting?

Inference Recap

Two main goals of inference:

Estimating a parameter: ______________________

Testing a claim: ______________________

Appropriate inference method depends on type of data:

Unit 6 Inference for Categorical Data: _____________________

Unit 7 Inference for Quantitative Data: ____________________

Unit 8 Inference for Categorical Data: _____________________

Unit 9 Inference for Categorical Data: _____________________

Confidence Intervals

The formula on the formula sheet is the same for ___________ procedures.

CI = ____________ ± (_______________________)(_______________________________)

You should always do these steps:

- Define the _________________ you are trying to estimate.
- Identify the inference ______________________.
- Verify the ________________ for the procedure are _______.
- _________________ the confidence interval.
- _________________ the interval ____________________.

You do ________ need to interpret the confidence _________ unless specifically asked.

Significance Tests

$$\textit{standardized test statistic} = \frac{\textit{statistic} - \textit{parameter}}{\textit{standard error of the statistic}}$$

You should always do these steps:

- State the ________ and ____________ hypotheses. Be sure to ___________ parameter.
- Give the _____________________________.
- ______________ the inference procedure.
- Verify that the _______________ for the procedure are __________.
- Calculate the _________________ and the ____________.
- Make a conclusion based on the ________________.

You do _______ need to interpret the ___________ unless specifically asked.

STATS MEDIC

Name________________________

Unit 6

Inference for Categorical Data: Proportions

Goal	Method	Formula	Conditions
Estimate p	One-sample z interval for a proportion	$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	• Random sample/random assignment ○ $n \leq 10\%N$ (samp w/o replacement) • $n\hat{p}$ and $n(1-\hat{p}) \geq 10$
Test $H_0: p = p_0$	One-sample z test for a proportion	$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	• Random sample/random assignment ○ $n \leq 10\%N$ (samp w/o replacement) • $np_0 \geq 10$ and $n(1-p_0) \geq 10$
Estimate $p_1 - p_2$	Two-sample z interval for difference in proportions	$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	• Random samples/random assignment ○ $n_1 \leq 10\%N_1$ and $n_2 \leq 10\%N_2$ (samp w/o replacement) • $n_1\hat{p}_1, n_1(1-\hat{p}_1), n_2\hat{p}_2, n_2(1-\hat{p}_2) \geq 10$
Test $H_0: p_1 - p_2 = 0$	Two-sample z test for a difference in proportions	$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}_C(1-\hat{p}_C)}{n_1} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_2}}}$	• Random samples/random assignment ○ $n_1 \leq 10\%N_1$ and $n_2 \leq 10\%N_2$ (samp w/o replacement) • $n_1\hat{p}_C, n_1(1-\hat{p}_C), n_2\hat{p}_C, n_2(1-\hat{p}_C) \geq 10$

Unit 7

Inference for Quantitative Data: Means

Goal	Method	Formula	Conditions
Estimate μ	One-sample t interval for a mean	$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$	• Random sample/random assignment ○ $n \leq 10\%N$ (samp w/o replacement) • $n \geq 30$ or no strong skew/outliers
Test $H_0: \mu = \mu_0$	One-sample t test for a mean	$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$	• Random sample/random assignment ○ $n \leq 10\%N$ (samp w/o replacement) • $n \geq 30$ or no strong skew/outliers
Estimate $\mu_1 - \mu_2$	Two-sample t interval for difference in means	$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	• Random samples/random assignment ○ $n_1 \leq 10\%N_1$ and $n_2 \leq 10\%N_2$ (samp w/o replacement) • $n_1, n_2 \geq 30$ or no strong skew/outliers
Test $H_0: \mu_1 - \mu_2 = 0$	Two-sample t test for difference in means	$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	• Random samples/random assignment ○ $n_1 \leq 10\%N_1$ and $n_2 \leq 10\%N_2$ (samp w/o replacement) • $n_1, n_2 \geq 30$ or no strong skew/outliers

Unit 8

Inference for Categorical Data: Chi-Square

Goal	Method	Formula	Conditions
Test H_0: Categorical variable has specified distribution	Chi-square test for goodness of fit	$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$	• Random sample/random assignment ○ $n \leq 10\%N$ (samp w/o replacement) • All expected counts > 5.
Test H_0: Categorical variable has same distribution for each population or treatment	Chi-square test for homogeneity	$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$	• Random samples/random assignment ○ $n_1 \leq 10\%N_1, n_2 \leq 10\%N_2, \ldots$ (samp w/o replacement) • All expected counts > 5.
Test H_0: There is no association between two categorical variables in a population	Chi-square test for independence	$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$	• Random sample/random assignment ○ $n \leq 10\%N$ (samp w/o replacement) • All expected counts > 5.

Name______________________________

Inference for Quantitative Data: Slopes

Goal	Method	Formula	Conditions
Estimate β	t interval for a slope	$b \pm t^* SE_b = b \pm t^* \frac{s}{s_x\sqrt{n-1}}$	• The true relationship between x and y is linear. • The standard deviation of y doesn't vary with x. • For a particular value of x, the y-values are approximately normally distributed. • There is independence in data collection. ○ Random sample or randomized experiment. ○ $n \leq 10\%N$ (samp w/o replacement)
Test $H_0: \beta = \beta_0$	t test for a slope	$t = \frac{b - \beta_0}{SE_b}$	• The true relationship between x and y is linear. • The standard deviation of y doesn't vary with x. • For a particular value of x, the y-values are approximately normally distributed. • There is independence in data collection. ○ Random sample or randomized experiment. ○ $n \leq 10\%N$ (samp w/o replacement)

Selecting an Inference Procedure

Inference for ...

Categorical Data: Proportions	Quantitative Data: Means	Categorical Data: Chi-Square	Quantitative Data: Slopes
One-sample z interval for a proportion	One-sample t interval for a mean (Paired data)	Chi-square test for goodness of fit (Distribution of proportions for one categorical variable)	t interval for a slope
One-sample z test for a proportion	One-sample t test for a mean (Paired data)	Chi-square test for homogeneity (Distribution of a categorical variable for multiple populations or treatments)	t test for a slope
Two-sample z interval for a difference in proportions	Two-sample t interval for a difference in means	Chi-square test for independence (Relationship between two categorical variables)	
Two-sample z test for a difference in proportions	Two-sample t test for a difference in means		

1997 Exam #5

A company bakes computer chips in two ovens, oven A and oven B. The chips are randomly assigned to an oven and hundreds of chips are baked each hour. The percentage of defective chips coming from these ovens for each hour of production throughout a day is shown below.

Hour	Oven A	Oven B
1	45	36
2	32	37
3	34	33
4	31	34
5	35	33
6	37	32
7	31	33
8	30	30
9	27	24

The percentage of defective chips produced each hour by oven A has a mean of 33.56 and a standard deviation of 5.20. The percentage of defective chips produced each hour by oven B has a mean of 32.44 and a standard deviation of 3.78. The hourly differences in percentages for oven A minus oven B have a mean of 1.11 and a standard deviation of 4.28.

Does there appear to be a difference between oven A and oven B with respect to the mean percentages of defective chips produced? Give appropriate statistical evidence to support your answer.

Individual/case: ______________________________ Inference for: ________________________

Variable(s) of interest: _________________________ One sample, Paired data, or Two Samples?

Type of data: ___________ Estimate parameter or Test a Claim? Perform: _______________________

Name______________________________

Larry Green's Applet

Source: Larry Green's Comprehensive Review applet, http://www.ltcconline.net/greenl/java/Statistics/catStatProb/categorizingStatProblemsJavaScript.html

Is political affiliation related to the month that the person was born in? 3000 voters were studied.

Individual/case: ____________________________ Variable(s) of interest: _________________________
Type of Data: ______________________________ Inference for: ________________________________
1 Variable, 2 Variables (one sample) or 2 Variables (multiple samples or treatments)?
Inference Procedure: ____

- Confidence Interval for a Population Mean
- Confidence Interval for a Proportion
- Confidence Interval for the Diff. Between 2 Means (Independent Samples)
- Confidence Interval for Paired Data (Dependent Samples)
- Confidence Interval for the Difference Between 2 Proportions
- Prediction for a Single Value of *y* for a Fixed *x*
- Hypothesis Test for a Population Mean
- Hypothesis Test for a Population Proportion
- Hyp. Test for the Difference Between 2 Means (Independent Samples)
- Hyp. Test for Paired Data (Dependent Samples)
- Hyp. Test for the Difference Between 2 Proportions
- Chi-Square Goodness of Fit Test
- Chi-Square Test for Independence
- Chi-Square Test for Homogeneity

Is political affiliation related to the month that the person was born in? 3000 voters were studied.

Individual/case: ____________________________ Variable(s) of interest: _________________________
Type of Data: ______________________________ Inference for: ________________________________
1 Variable, 2 Variables (one sample) or 2 Variables (multiple samples or treatments)?
Inference Procedure: ____

What Should We Take Away?

What inference methods did we learn about in Units 6 – 9?

Unit 6 Inference for Categorical Data: ____________________
Unit 7 Inference for Quantitative Data: ___________________
Unit 8 Inference for Categorical Data: ____________________
Unit 9 Inference for Categorical Data: ____________________

How can we identify the appropriate inference procedure to use in a given setting?

Individual/case:
Variables of interest:
Type of Data:
Inference for: