Sample Regression Line:

Population (True) Regression Line:

STATE:

We want to estimate the slope \( \beta \) of the population regression line relating ____________ to ____________ with C\% confidence.

OR

We want to perform a test of

\( H_0: \beta = 0 \)

\( H_A: \beta < 0 \) (negative slope) or \( H_A: \beta > 0 \) (positive slope)

\( \beta = \)

where \( \beta \) is the slope of the population regression line relating ____________ to ____________. Our significance level is \( \alpha = \)__________

PLAN:

USE THE ACRONYMN LINER

- **Linear:** The actual relationship between \( x \) and \( y \) is linear. For any fixed value of \( x \), the mean response \( \mu_y \) falls on the population (true) regression line \( \mu_y = \alpha + \beta x \)
- **Independent:** Individual observations are independent of each other. When sampling without replacement, and also check the 10\% condition
- **Normal:** For any fixed value of \( x \), the response \( y \) varies according to a Normal distribution
- **Equal Standard Deviation:** The standard deviation of \( y \) (call it \( \sigma \)) is the same for all values of \( x \)
- **Random:** The data from from a well-designed random sample or randomized experiment

Because the conditions are met, we will use a \( t \)-interval for the slope of a regression line or we will use a \( t \)-test for the slope of a regression line.

DO:

Confidence Interval (\( t \)-interval for slope) – Plug values into calculator & include: calc inputs, df, \( t^* \), and the interval.

\[
b \pm t^* \frac{s}{s_x \sqrt{n-1}}
\]

with df = n-2 with \( SE_b = \frac{s}{s_x \sqrt{n-1}} \) → plug in values but then put in calculator

\[
\text{df} = \frac{s}{s_x \sqrt{n-1}}
\]

Confidence interval (________ , __________)

OR
**Significance Test (t-test for the slope)** – Plug values into calculator & include: calc inputs, df, test statistic, p-value, and graph.

\[ t = \frac{b - \beta_0}{SE_b} \] with \( df = n-2 \)

with \( SE_b = \frac{s}{s_x \sqrt{n-1}} \) don’t actually use, just plug right into calculator

\[ df = \]
\[ t = \]
\[ P\text{-value} = \]

**CONCLUDE:**

We are \%C confident that the interval between ___________ and ___________ captures the slope of the population regression line relating ___________ to ___________ in context.

OR

We reject or fail to reject the null because our P-value is _________ \( \alpha \). There is/isn’t convincing evidence of a positive linear relationship between ___________ and ___________ in context.

**Technology:**

Confidence Interval for slope of a line if given data otherwise do it by hand:

STAT → TESTS
G: LinRegTint

Significance Test for slope of a line if given data:

STAT → TESTS
F: LinRegTTest
Example 1:

Do Beavers benefit beetles? Researchers laid out 23 circular plots, each 4 meters in diameter, at random in an area where beavers were cutting down cottonwood trees. In each plot, they counted the number of stumps from trees cut by beavers and the number of clusters of beetle larvae. Ecologists think that the new sprouts from stumps are more tender than other cottonwood growth, so that beetles prefer them. If so, more stumps should produce more beetle larvae. Minitab output for a regression analysis on these data is shown blow. Construct and interpret a 99% confidence interval for the slope of the population regression line. Assume that the conditions for performing inference are met.

Regression Analysis: Beetle larvae versus Stumps

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.286</td>
<td>2.853</td>
<td>-0.45</td>
<td>0.657</td>
</tr>
<tr>
<td>Stumps</td>
<td>11.894</td>
<td>1.136</td>
<td>10.47</td>
<td>0.000</td>
</tr>
</tbody>
</table>

R-Sq=83.9%  R-Sq(adj)=83.1%
Example 2:

A researcher from the University of California, San Diego, collected data on average per capita wine consumption and heart disease death rate in a random sample of 19 countries for which data were available. The following table displays the data:

<table>
<thead>
<tr>
<th>Alcohol from Wine (liters/year)</th>
<th>Heart Disease Death rate (per 100,000)</th>
<th>Alcohol from Wine (liters/year)</th>
<th>Heart Disease Death rate (per 100,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>211</td>
<td>7.9</td>
<td>107</td>
</tr>
<tr>
<td>3.9</td>
<td>167</td>
<td>1.8</td>
<td>167</td>
</tr>
<tr>
<td>2.9</td>
<td>131</td>
<td>1.9</td>
<td>266</td>
</tr>
<tr>
<td>2.4</td>
<td>191</td>
<td>0.8</td>
<td>227</td>
</tr>
<tr>
<td>2.9</td>
<td>220</td>
<td>6.5</td>
<td>86</td>
</tr>
<tr>
<td>0.8</td>
<td>297</td>
<td>1.6</td>
<td>207</td>
</tr>
<tr>
<td>9.1</td>
<td>71</td>
<td>5.8</td>
<td>115</td>
</tr>
<tr>
<td>2.7</td>
<td>172</td>
<td>1.3</td>
<td>285</td>
</tr>
<tr>
<td>0.8</td>
<td>211</td>
<td>1.2</td>
<td>199</td>
</tr>
<tr>
<td>0.7</td>
<td>300</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Is there statistically significant evidence of a negative linear relationship between wine consumption and heart disease deaths in the population of countries? Carry out an appropriate significance test at the $\alpha = 0.05$ level.
Example 3:

A random sample of AP Statistics teachers was asked to report the age (in years) and mileage of their primary vehicles. A scatterplot of the data is shown. A computer output from a least-squares regression analysis of these data is shown below (df=19). Assume that the conditions for regression inference are met.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>7288.54</td>
<td>6591</td>
<td>1.11</td>
<td>0.2826</td>
</tr>
<tr>
<td>Car Age</td>
<td>11630.6</td>
<td>1249</td>
<td>9.31</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

S=19280     R-sq=82.0%    R-sq(Adj)=81.1%

**a)** Verify that the 95% confidence interval for the slope of the population regression line is 9016.4, 14,244.8).

**b)** A national automotive group claims that the typical driver puts 15,000 miles per year on his or her main vehicle. We want to test whether AP Statistics teachers are typical drivers. Explain why an appropriate pair of hypotheses for this test is $H_0: \beta = 15,000$ versus $H_1: \beta \neq 15,000$. 


c) Compute the test statistic and P-value for the test in part (b). What conclusion would you draw at the \( \alpha = 0.05 \) significance level?

d) Does the confidence interval in part (a) lead to the same conclusion as the test in part (c)? Explain.