Hypothesis testing - Steps

• Steps to do a *two-tailed test* of the hypothesis that $\beta_1 \neq 0$:

1. Set up the hypotheses:

   $H_0 : \beta_1 = 0$
   $H_a : \beta_1 \neq 0$.

2. Compute the test statistic:

   $$t = \frac{b_1 - 0}{\text{Std. error of } b_1} = \frac{b_1}{\hat{\sigma}_{b_1}}$$

3. Read critical value off the $t-$table: $t_{\alpha/2,n-2}$

4. Reach your conclusion: If $|t| > \text{critical value}$ conclude $H_a$. If $|t| \leq \text{critical value}$, conclude $H_0$ or *fail to reject* $H_0$ at level $\alpha$. 
One-tailed hypothesis tests

• We can also test one-sided or one-tailed hypothesis.

• A one-sided hypotheses would be:

\[ H_0 : \beta_1 = 0 \]
\[ H_a : \beta_1 > 0, \]

or \[ H_a : \beta_1 < 0. \]

• Steps are exactly the same, but now the critical value is \( t_{\alpha, n-2} \) (we do not divide \( \alpha \) into 2).

• If the test statistic \(|t|\) is larger than the critical value off the table and we conclude \( H_a \), we would be concluding that \( \beta_1 \) is larger (or smaller, depending on the test we set up) than zero.
Hypothesis testing (cont’d)

• If we wish to test a two-sided hypothesis about $\beta_1$ at level $\alpha$ we can also use the $100(1 - \alpha)\%$ confidence interval to do so.

• We will conclude $H_a$ whenever the CI does not include the hypothesized value for $\beta_1$ (the value in $H_0$).

• If the null hypothesis is $H_0 : \beta_1 = 0$, then we will reject the null at level $\alpha$ when the CI does not contain the value zero.

• Why?
Hypothesis testing (cont’d)

- Note that the test of hypothesis (two-tailed) says: fail to reject $H_0 : \beta_1 = 0$ if

$$\frac{b_1 - 0}{\hat{\sigma}_{b_1}} > t_{\frac{\alpha}{2},n-2} \text{ or } \frac{b_1 - 0}{\hat{\sigma}_{b_1}} < -t_{\frac{\alpha}{2},n-2}.$$ 

- Equivalently, by multiplying both sides of expressions above by $\hat{\sigma}_{b_1}$ and subtracting $b_1$ from both sides also, note that we fail to reject when

$$0 > b_1 - t_{\frac{\alpha}{2},n-2}\hat{\sigma}_{b_1} \text{ and } 0 < b_1 + t_{\frac{\alpha}{2},n-2}\hat{\sigma}_{b_1}$$

in other words, when 0 is inside the $100(1 - \alpha)\% \text{ CI}$ for $\beta_1$.

- This applies only to two-sided tests.
Hypothesis testing (cont’d)

- JMP and SAS produce a \( p \)-value.

- The \( p \)-value is the smallest \( \alpha \)-level that still leads to rejecting \( H_0 \).

- A \( p \)-value of 0.001 means that if I conclude \( H_a \) I only have a 1 in a thousand chance of reaching the wrong conclusion.

- A \( p \)-value of 0.3 means that if I conclude \( H_a \) I have a 1 in 3 chances of reaching the wrong conclusion.

- Typically we choose \( H_a \) only if the chance of making the wrong choice is small, say below 5%. Thus, \( p \)-values of 0.05 or smaller lead to \( H_a \).

- The \( p \)-values produced by computer programs the test \( H_0 : \beta_1 = 0 \) versus \( H_a : \beta_1 \neq 0 \).
Hypothesis testing (cont’d)

• We can test any hypothesis that might seem appropriate for the application at hand. For example, we might wish to test

\[ H_0 : \beta_1 = 250 \]
\[ H_a : \beta_1 > 250, \]

if that makes sense from a subject-matter point of view.

• The appropriate test statistic is

\[ t = \frac{b_1 - 250}{\hat{\sigma}_{b_1}} \]

and the rest is the same as in the one-tailed test described earlier.
Coefficient of correlation

• The coefficient of correlation $r$ measures the linear association between two variables.

• $r$ is defined to be between -1 and +1.

• Given $n$ measurements $x$ and $y$ on a sample of units:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$
Coefficient of correlation (cont’d)

• Since \( b_1 = \frac{SS_{xy}}{SS_{xx}} \), we can also compute \( r \) as

\[
r = b_1 \sqrt{\frac{SS_{xx}}{SS_{yy}}},
\]

so \( b_1 \) and \( r \) will always be of the same sign.

• A positive (negative) \( r \) means positive (negative) linear association between \( x \) and \( y \). An \( r \) close to zero means no linear association (but there can be nonlinear association!).

• Note that the correlation between \( x \) and itself is equal to 1 because:

\[
r_{xx} = \frac{SS_{xx}}{\sqrt{SS_{xx}SS_{xx}}} = \frac{SS_{xx}}{SS_{xx}} = 1.
\]
Correlation coefficient - Example

• Consider the five stores on which me measured advertising expenses and units of a product sold.

• From earlier example we had: $SS_{xx} = 10$, $SS_{xy} = 8$, $SS_{yy} = 8.8$ and $b_1 = 0.8$. Then

$$r = \frac{8}{\sqrt{10 \times 8.8}} = \frac{8}{9.38} = 0.85.$$ 

• Using the other formula:

$$r = 0.8 \times \sqrt{\frac{10}{8.8}} = 0.85.$$
Correlation coefficient - Test of hypothesis

• The sample correlation \( r \) estimates the population correlation \( \rho \).

• If \( r \) is close to zero, we tend to believe that the true population correlation is also zero.

• As in the case of \( \beta_1 \), we can test hypotheses about the true population correlation \( \rho \).

• Steps are the same as before:

  1. Set up hypothesis (one or two tailed) \( H_0 \) and \( H_a \) and choose \( \alpha \)
  2. Construct test statistic \( t \) and compare \( t \) to critical value from table.
  3. If statistic falls in critical region, reject \( H_0 \).
Correlation - Test of hypothesis (cont’d)

- For a two-tailed test, hypotheses are:

  \[ H_0 : \rho = 0 \text{ versus } H_a : \rho \neq 0. \]

- Test statistic is given by:

  \[ t = \frac{r \sqrt{n - 2}}{\sqrt{1 - r^2}} \]

- Test statistic is distributed as a \( t \) random variable with \( n - 2 \) degrees of freedom. Thus, for a test with level \( \alpha \), the critical value is \( t_{\alpha/2, n-2} \) from the \( t \) table.
Correlation - Test of hypothesis (cont’d)

- Decision is:
  - If $|t| > t_{\alpha/2,n-2}$: reject $H_0$.
  - If $|t| \leq t_{\alpha/2,n-2}$: fail to reject $H_0$.

- For a one-tailed test, the only change is in the formulation of $H_a$ and in the critical value, that now is equal to $t_{\alpha,n-2}$.

- A test of hypothesis for $\rho$ is equivalent to the test for $\beta_1$. If one test results in rejecting $H_0$, the other will too.

- Example in class.