

Instructions: Attempt to answer these questions by reading the textbook or with online resources before coming to class on the date above.

1. What is the mean estimate for \hat{y} under a linear regression model for a particular x^* ?

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x^*$$

2. What is the formula for the variance of \hat{y} ?

$$S_{\hat{y}}^2 = \left[S \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}} \right]^2$$

3. How can we construct a confidence interval for a particular value of \hat{y} ?

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2, n-2} S_{\hat{y}} = \hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2, n-2} S \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

4. How is the formula for a prediction interval different?

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2, n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

↑
here. since only a single value prediction

5. What kind of different assumptions are made for a prediction interval vs. a confidence interval?

confidence interval is for a sample of size n at that point (the mean of such samples is predicted). whereas, the prediction interval is for a single measurement.

6. What is the sample correlation coefficient?

$$r = \frac{S_{xy}}{\sqrt{S_{xx}} \sqrt{S_{yy}}}$$

7. What parameter does it estimate?

ρ

8. What properties does r have?

$$-1 \leq r \leq 1$$

9. What are the range of values for a weak, moderate or strong correlation?

according to our book

strong $|r| \geq .8$, moderate $.5 \leq |r| \leq .8$, weak $|r| \leq .5$

however other sources may use other ranges such as .7 or .4 for thresholds

10. What is the formula for a dependent joint probability distribution?

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho(x-\mu_1)(y-\mu_2)/\sigma_1\sigma_2 + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]/(2(1-\rho^2))}$$

11. What is the formula for the test statistic for testing if the correlation is zero or non-zero?

$$T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

12. What test from a previous section is this test equivalent to?

Lin Reg T Test test $\beta_1 = 0$ or $\rho = 0$

13. Explain the steps and provide the formulas necessary for creating a confidence interval for the correlation.

to create conf. interval on ρ , first transform into $V = \frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right)$

$$\omega / \mu_v = \frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right) \quad \sigma_v^2 = \frac{1}{n-3}$$

then CI on v is $v \pm \frac{z_{\alpha/2}}{\sqrt{n-3}}$

$$\text{let } v - \frac{z_{\alpha/2}}{\sqrt{n-3}} = c_1 \quad \& \quad v + \frac{z_{\alpha/2}}{\sqrt{n-3}} = c_2$$

then CI on ρ is

$$\left(\frac{e^{2c_1} - 1}{e^{2c_1} + 1}, \frac{e^{2c_2} - 1}{e^{2c_2} + 1} \right)$$

14. What is the formula for the standardized residuals?

$$e_i^* = \frac{y_i - \hat{y}_i}{\sqrt{1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}}}$$

15. What is a residual plot? What are the residuals plotted against?

it's a scatterplot of residuals plotted against one of the original variables; can be plotted against x or y

16. What are some problems that can be spotted with a model by looking at residual plots? Give some examples of what to look for.

non linear

outliers

unaccounted for variables

time observation

non-constant variance