Quiz 2
MSAN 601

Question 1 (6 pts)

Write out the hypothesis test which tests for the statistical significance of β_1 for an SLR model. Be sure to include the null and alternate hypothesis, the critical value including degrees of freedom (two-tailed test) for $\alpha = 0.05$ and an interpretation of both possible results.

Answer

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

$t^* = \frac{b_1}{s\{b_1\}}$, degrees of freedom = n-2

Interpretation of t^*:
If $|t^*| > t_{0.975;n-2}$, reject H_0. Thus, with 95% confidence we can say that β_1 is statistically significantly different from zero.
If $|t^*| \leq t_{0.975;n-2}$, do not reject H_0. Thus, with a 95% confidence, we fail to reject the null hypothesis and conclude that β_1 is not statistically significantly different from zero.

Question 2 (6 pts)

There is an SLR model with n = 17, $b_1 = 3.48$, $s\{b_1\} = 0.54$. Test the null hypothesis that β_1 is statistically significantly greater than 0, assuming $\alpha = 0.05$. Include null and alternate hypothesis, show all relevant calculations, and explicitly state your conclusion.

Answer

$H_0 : \beta_1 \leq 0$, $H_a : \beta_1 > 0$

$t^* = \frac{3.48}{0.54} = 6.44$

$t_\alpha = 1.753$ One-sided, $n - 2 = 15, \alpha = 0.05$

$t^* > t_\alpha$, thus, we reject the null hypothesis.

Thus, with a 95% confidence, we reject the null hypothesis and conclude that β_1 is statistically significantly greater than zero.

Question 3 (6 pts)

Fill in the blank values in the ANOVA table below

	Sum of Squares	df	Mean Square	F
Regression	**1494.465**	**1**	1494.465	**536.90**
Residual	**1664.54**	598	**2.78**	
Total	3159.009	**599**		

Question 4 (4 pts)

How can you use MSE and MSR to test whether $\beta_1 = 0$? Be complete in your answer including all relevant mathematical details.

Answer

$E[MSE] = \sigma^2$ and $E[MSR] = \sigma^2 + \beta_1^2 \sum(X_i - \bar{X})^2$

We can use the F-test to test whether $\beta_1 = 0$.

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

$F^* = \frac{MSR}{MSE}$, this is an upper-tail test as $MSE \geq MSR$. If $\beta_1 = 0$, F^* will be close to 1 and we fail to reject the null hypothesis. At higher values of F^*, however, we reject the null.

If $F^* \leq F(1-\alpha; 1; n-2)$, do not reject H_0 else reject H_0 with a confidence level of $1-\alpha$.

Question 5 (6)

Prove: $SSTO = SSE + SSR$

Answer

See slide 142 in SLR slides.

Question 6 (3)

What is coefficient of determination? What are its limitations?

Coefficient of determination is a measure of the goodness of fit of the regression line. It is equal to the fraction of the sample variation in Y that can be explained by the variation in X.

Answer

Limitations:

1. High R^2 does not imply that good predictions can be made.
2. High R^2 does not imply that the estimated regression line is a good fit.
3. $R^2 \sim 0$ does not imply that X and Y are not related.

Question 7 (3)

How is R^2 calculated? How is its value related to Pearson's coefficient of correlation? What is the range of values it can take?

Answer

$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$
$R^2 =$ square of Pearson's coefficient of correlation
$0 \leq R^2 \leq 1$

Question 8 (4)

What do you understand by a prediction interval? How does it differ from a confidence interval? (no math required)

Answer

A prediction interval is an estimate of an actual value of Y for given X. A confidence interval is an interval for the mean values of the Y for given X. Thus, a prediction interval is always wider than a confidence interval.

Question 9 (2)

The prediction intervals become wider as we predict Y for X values away from $\bar{X}$. True or false? Explain.

Answer

True. $s^2\{pred\} = MSE\left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})}{\sum(X_i - \bar{X})^2}\right]$, from the formula, as X_h goes further away from $\bar{X}$, the prediction interval widens. This happens because as the difference between X_h and $\bar{X}$ increases, we can make less precise predictions.