

For each problem on the test, you will be given a set of data, a set of intermediate sums, and a partial ANOVA Table. If you wish, you may use the data set with EXCEL to create an ANOVA Table including the statistics. Or you may use sums and/or the partial ANOVA Table to solve the problem. You may use a paper copy of a F Distribution Table or you may access the table via the Internet. You may use one page of your own course notes (NO SHARING). You may use your own calculator (NO SHARING) and MS Excel to help calculate your answers. CELL PHONES (even as calculators) are not permitted (PUT THEM AWAY). Making copies (pictures, notes, scans) of the tests or of your answers is NOT PERMITTED.

Understand that valid statistical conclusions are based on randomly selected, representative samples of the population. That is to say, if the sample is not representative of the population, then all bets are off.

Understand the caveats that statistical conclusions are based on the sample data and the level of significance. That is to say, if we were to use a different sample and/or a different level of significance, then we might arrive at a different conclusion regarding the hypotheses.

Be familiar with **Correlation, Regression, & Prediction** and the properties of the *correlation coefficient*  $r$ .

Understand that significant correlation is **never proof** of *Cause and Effect*.

Understand that the regression equation should **only** be used for *prediction* when there is *significant correlation*.

Apply the concepts of correlation, regression, and prediction, including;

calculating the correlation coefficient

hypothesis testing for significant correlation (both t-test and F-test)

determining the coefficients of the regression equation as appropriate; and if so

using the regression equation to predict values of the response variable  $Y$ , for values of the regressor variable  $X$ .

Course Notes:

Correlation & Regression

Homework Problems:

Homework #10a

Homework #10b

See page 2 & 3 for example questions, homework, and problems from previous exams.

Indicate whether or not the following statements regarding the correlation coefficient  $r$  are True or False.

- F / T An  $r$  value of  $+0.65$  offers stronger proof of cause and effect than does an  $r$  value of  $-0.50$ .
- F / T An  $r = 0.0$  implies there is absolutely no mathematical correlation between two variables.
- F / T An  $r$  value  $> 0.0$  implies that  $Y$  increases as  $X$  decreases.
- F / T An  $r$  value  $> 0.0$  always indicates significant correlation.
- F / T When comparing correlation between two variables, use averages rather than individual values.
- F / T If  $r > 0.5$ , it is permissible to extrapolate values of  $Y$  for values of  $X$  that are not in the data range of  $X$ .
- F / T Only positive values of  $r$  provide evidence of cause and effect.

When is strong correlation sufficient to prove cause and effect?

Under what circumstances is it acceptable to use the regression equation for predicting, even when there is statistically insufficient evidence to suggest correlation?

## Correlation & Regression

Reference:

5ed, Page 435, Problem 11-71

6ed, Page 462, Problem 11-81

$X$ 's = Statistics Scores

$Y$ 's = Ops Research Scores

$n = 20$

Sum of  $X = 1587$

Sum of  $Y = 1602$

Sum of  $X^2 = 127,731$

Sum of  $Y^2 = 129,818$

Sum of  $XY = 128,603$

$S_{xx} = 1803$

$S_{yy} = 1498$

$S_{xy} = 1484$

$SSR = 1222$

$SST = 1498$

Calculate the correlation coefficient  $r$ .

Calculate the Test Value  $t_0$  and Determine the Critical Value  $t_{\alpha/2, df}$ .

Calculate the Test Value  $F_0$  and Determine the Critical Value  $F$ .

Determine the Regression Equation (i.e., solve for  $B_0$  and  $B_1$ ).

Use the Regression Equation to predict an Ops Research Score given a Statistics Score = 96.

The table below is a list of randomly selected laptop computers and their respective performance rating and retail price. Use the data to determine whether or not there is any statistical correlation ( $\alpha = 5\%$ ) between a laptop's performance rating and its retail price. In addition to the tabular data, various intermediate computational values have also been included. Determine whether or not there is statistically significant correlation between Rating and Price. You must calculate the correlation coefficient  $r$ . However, you may choose to use either a  $t_{\text{test}}$  or an ANOVA Table  $F_{\text{test}}$  to test for statistically significant correlation. Write the Regression Equation, (i.e., calculate both  $\beta_0$  and  $\beta_1$ ).

Rating (X)	Price (Y)	
81	1050	$\Sigma X = 1883$
80	1050	
86	1080	$\Sigma Y = 24798$
87	1170	
85	1260	$\Sigma X^2 = 178463$
94	1125	
90	1215	$\Sigma Y^2 = 31618294$
96	1035	
96	1065	$\Sigma XY = 2351600$
88	1115	
109	1575	$S_{XX} = 1179$
104	1800	
103	1090	$S_{YY} = 871254$
103	1345	
98	1485	$S_{XY} = 16868$
98	1095	
95	1238	$SSR = 241432$
100	1080	
94	1485	$SST = 871254$
96	1440	

1. Correlation Coefficient  $r =$  \_\_\_\_\_

2a. Hypothesis Testing Using  $t_{\text{test}}$  Method

Null Hypothesis  $H_0:$  \_\_\_\_\_

$t_{\text{test}} =$  \_\_\_\_\_       $df =$  \_\_\_\_\_      Critical Value = \_\_\_\_\_

2b. Hypothesis Testing Using  $F_{\text{test}}$  Method

Null Hypothesis  $H_0:$  \_\_\_\_\_

$F_{\text{test}} =$  \_\_\_\_\_       $df's =$  \_\_\_\_\_ & \_\_\_\_\_      Critical Value = \_\_\_\_\_

3. Statistically Significant Correlation    Yes / No

4. Regression Equation \_\_\_\_\_

5. Given a rating of 103, predict the price. \_\_\_\_\_