

KEY

**Instructions:** This exam is in two parts: Part I is to be completed partly at home using the materials posted on Blackboard for Part I and you will answer questions about that work in class below; Part II is to be completed entirely in class. You may not use cell phones, and you may only access internet resources you are specifically directed to use. You may access your data file for Part I of the exam in Blackboard. You may access the data files posted to Blackboard for the Exam part II. Be sure you are using the data file that matches the exam version you are given.

#### Part I: At Home

This part was completed at home. You can upload the Excel file for Part I to the Part I folder in Blackboard for use during the Exam period. However, this submission will not be graded in this location, it must be submitted to the "to be graded folder" to receive credit.

#### Part II: In Class

1. Use the work done at home to answer the Part I questions.
2. Open the file from the in-class portion of the final posted on Blackboard that corresponds to the version of the exam you have. This is Exam C.
3. Answer the questions corresponding to the data file, and any additional calculation in Excel required.
4. When you have finished answering questions on the exam, and all your answers have been recorded on the paper test for grading, upload **both** the take home Excel file **and** the in-class Excel file to the same in-class Exam folder in Blackboard for grading. Only those files submitted to the correct folder will be graded. (If in doubt, put all work in one Excel file.)
5. Turn in your paper copy of the exam to your instructor.
6. Enjoy your break!

Part I:

1. Using the NoTip Table data, write the objective function you are using to maximize profit in their production process for the original problem data. (6 points)

$$200x_1 + 350x_2 = 16,500$$

2. How many of each type of cut should be made to produce the maximum profit? (6 points)

$$\text{Basic } (x_1) = 30$$

$$\text{Deluxe } (x_2) = 30$$

3. How did the result of the model change when we adjusted the profit for each type of table to \$120 and \$300? (6 points)

down from 16,500 to 12,910

(had to set integer constraints)

4. For the data on employee's year-end performance, which variable had the highest P-value (at any stage of the regression analysis)? State the name of the variable and the P-value. (6 points)

Test 2, P-value 0.524 in full model

5. Write the final regression equation you obtained, the  $R^2$  value, and explain your reasoning for choosing it. (9 points)

$$\text{Seniority} = x_1$$

$$R^2 = 0.989$$

$$\text{Test 1} = x_2$$

P-values all low

$$\text{Test 3} = x_3$$

$R^2$  high

$$y = 0.882x_1 + 0.38x_2 + 0.512x_3$$

6. Did any of the variables in the data set appear to be nonlinear? Why or why not? (6 points)

no, scatterplots all appear linear  
and residual plots random

7. State a 99% confidence interval for the coefficient for Test 1 in your final model. Interpret it in context. (6 points)

(0.121, 0.639)

We are 99% confident that the true value  
of the coeff for Test 1 is between 0.121 and 0.639.

8. Interpret the meaning of the slope for Test 3 in context. (6 points)

for each one unit increase in Test 3,  
performance score goes up by 0.51 points

9. Use your equation to predict the performance rating of an employee with 5 years of seniority, and scores of 72, 86, 99 on each of the three Tests. Construct a 95% prediction interval around that prediction. [Hint: Use your best model. If the model does not contain a particular variable, omit it as irrelevant.] (12 points)

$$\hat{y} = 82.45$$

(68.37, 96.46)

10. For the data on sea ice extent, describe the general trend of the data. (6 points)

general trend downward  
accelerating trend - nonlinear

11. Write the best regression equation you found and state the  $R^2$  value. (6 points)

$$Y = -0.0038x^2 + 15.236x - 15,103$$

$$R^2 = 0.8217$$

12. Interpret the meaning of the  $R^2$  value in the context of the problem. (8 points)

82% of the variability in sea ice extent is correlated w/ time

13. Are there any outliers in the data? Use the residuals and residual plots to determine which point is suspect. Redo your analysis without this point included and describe the results of this test. (10 points)

There are no obvious outliers

removing 1996 data improves fit slightly

removing 2012 data makes it worse

14. Using data on couples purchasing cars, determine if the two measurements are dependent or independent. Explain your reasoning. (8 points)

dependent

15. Conduct an appropriate  $t$ -test to determine if men are willing to pay more than their wives for a car. State the null and alternative hypotheses, test statistic, P-value and state the results in an English sentence understandable to a non-statistician. (12 points)

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_2 > \mu_1$$

$$T = 1.297 \dots$$

$$P\text{-value} = 0.10335 > .05$$

fail to reject null  
this is not strong evidence  
that men are willing to  
pay more

Calculations in Excel: (1) 30 points, (2) 35 points, (3) 20 points, (4) 20 points.

Part II:

16. A study shows that half of all Internet users in 2012 used Google for their preferred search engine. Before purchasing ads on Google, a company decides to conduct a survey to see if this is still true. They ask 1593 people and find the 791 respondents claim to use Google as their primary search engine. Conduct a hypothesis test of proportions to determine if this result has changed from previous results. State the hypotheses, test statistic, P-value and conclusion. Is this sufficient evidence to think use of Google has changed? (12 points)

$$H_0: p = 50\%$$

$$H_a: p \neq 50\%$$

$$Z\text{-test} = -0.2756$$

$$P\text{-value} = 0.782853$$

fail to reject null

the results are not different than before

17. Interpret a Type I and Type II error in the context of this problem. (6 points)

Type I: 50% of users do use Google but we conclude they do not

Type II: 50% of users do not use Google, but we conclude they do

18. Construct a 92% confidence interval for the true proportion of Internet users that prefer Google as their primary search engine. Interpret the interval in context. (8 points)

$$(47.46\%, 51.85\%)$$

we are 92% confident that the true population proportion of Google users is between 47.46% and 51.85%

19. Suppose that you wish to sample employees of a large company to determine factors that predict high inside sales commissions in order to prepare for a new training program. The company has 1300 employees in this position around the world. The company wants to select 8 of them for an initial study of best practices. Eligible employees are assigned numbers from 1 to 1300 based on their date of initial hire. Select a simple random sample and report the employee numbers you have selected below. (6 points)

*answers will vary*

1231

1046

213

268

121

897

352

779

1079

1111

**Standard errors:**  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$   $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$   $S_{pooled} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

$$S_{x_1-x_2} = S_{pooled} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

**Sample sizes:**  $n > \hat{p}(1-\hat{p}) \left(\frac{z_{\alpha/2}}{E}\right)^2$   $n > \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$   $m = n = \frac{4z_{\alpha/2}^2(\sigma_1^2 + \sigma_2^2)}{w^2}$

**Confidence intervals:**

One sample:  $\bar{x} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$   $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Two samples (independent):  $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2, n-1} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$   $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

**Test statistics:**

One sample:  $z$  or  $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$   $z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$

Two samples: dependent:  $z$  or  $t = \frac{\bar{d}_0 - \delta}{\frac{s_d}{\sqrt{n}}}$

Independent:  $z$  or  $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$   $z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$

Degrees of freedom (two samples, unpooled)  $\nu = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{\left(\frac{s_1^2}{m}\right)^2}{m-1} + \frac{\left(\frac{s_2^2}{n}\right)^2}{n-1}}$

$\chi^2$  Tests:  $\chi^2 = \sum_{\text{all cells}} \frac{(\text{obs} - \text{exp})^2}{\text{exp}}$

ANOVA:  $MSE = \frac{(\sum_{j=1}^J n_j (\bar{y}_j - \bar{y})^2)}{J-1}$   $MSS = \sum_{j=1}^J \frac{(n_j - 1) s_j^2}{n - J}$   $F = \frac{MSE}{MSS}$

Upload your completed Excel files to the Final Exam submission box in Blackboard, and submit your completed paper exam to your instructor. You may not modify anything once the exam is submitted.