Practice Final Exams #1 and #2

EXAM #1

1. A fast-food chain tests each day that the number of calories in their "Diet-Burger" is no more than 400. Due to imperfections in the cooking processes, the number of calories in their Diet-Burger is normally distributed with standard deviation 30 calories. The decision rule adopted by the fast-food chain is to reject the null hypothesis (that the mean calories is 400) if the sample mean number of calories is more than 410.

   a) If a random sample of size 40 burgers is selected, what is the probability of a Type I error, using this decision rule?

   b) If a random sample of size 10 burgers is selected and the same decision rule is applied, do you think the probability of a Type I error will be (check one):

      _____ lower than the one in part a).
      _____ the same as the one in part a).
      _____ higher than the one in part a).

   c) Suppose that the true mean number of calories is 422 (and the standard deviation is 30). If a random sample of 40 burgers is selected, what is the probability of a Type II error, using this decision rule?

2. A pizza business has sent out 5431 coupons good for $3 off their large pizza if redeemed within a month. From past experience, they expect on average about 6% of the coupons to be redeemed. If the actual sample count showed 315 redeemed coupons, would you consider this as being consistent with past experience? Explain your answer using a hypothesis test.

3. You are an insurance underwriter in the new field of space insurance, writing policies that cover the launching of commercial satellites. Space shuttle launches sometimes fail to put satellites in proper orbit, and the satellite is lost. The owner of a satellite, a Japanese company, wants to purchase insurance against such failure in the amount of 50 million dollars. Of the last 50 satellites that were launched, 12 were failures.

   At the 99% confidence level, find a confidence interval for the chances that a future satellite will fail to reach orbit.
4. A large consumer products company wants to measure the effect of different local advertising media on the sales of its products. Specifically, they considered TV and newspaper advertising, and also considered providing cents-off coupons in newspapers. Over a period of three months, these variables were measured in 22 cities of roughly equal population and demographics, and the results were analyzed using multiple regression. The variables were:

SALES = sales in 1,000,000 dollar units.

TVAD = TV ad budget, in 10,000 dollar units.

NEWSAD = Newspaper ad budget, in 1,000 dollar units.

COUPON = 1 if coupons were given out in local newspapers, and 0 otherwise.

A part of the output of the regression is given below:

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>R Squared (D)</td>
</tr>
<tr>
<td>Standard Error (C)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Df</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t-Stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.376</td>
<td>0.130</td>
<td></td>
</tr>
<tr>
<td>TVAD</td>
<td>0.127</td>
<td>0.017</td>
<td>(A)</td>
</tr>
<tr>
<td>NEWSAD</td>
<td>0.016</td>
<td>0.003</td>
<td>(B)</td>
</tr>
<tr>
<td>COUPON</td>
<td>0.100</td>
<td>0.075</td>
<td></td>
</tr>
</tbody>
</table>

a) Fill in the blank spaces above in the regression output:

(A) t-stat for TVAD = _______

(B) p-value for TVAD= _______

(C) R-squared= _______
(D) Standard Error of Estimate \( s_e = \) ________
(E) Degrees of Freedom, Error = __________

b) Interpret the coefficient for COUPON in words. Develop a 95% confidence interval for this number and interpret this confidence interval in words.

c) For Pittsburgh, a city typical of those studied, the proposed local advertising budgets were $47,000 for TV ads and $25,000 for newspaper ads. No coupons were distributed in this area. What is the predicted level of sales in Pittsburgh in dollars?

d) The total effective cost of distributing cents-off coupons in this city is $20,000. Of the three advertising-promotional media considered here, which is the most cost-effective way of increasing sales in Pittsburgh?

e) Which of the three advertising and promotional media (if any) may not be having a significant effect on sales?

f) Let \( \beta_2 \) denote the coefficient for NEWSAD in this regression. Test the hypothesis (using \( \alpha = 5\% \)):

\[ H_0: \beta_2 \leq 0.01 \text{ and } H_A: \beta_2 > 0.01. \]

Interpret your results in words suitable for a person who has little appreciation for statistics.

g) Is this regression likely to be useful? Explain.

5. The Circuit Systems Corporation was concerned about the amount of days of sick leave employees were taking during 1991. So, during January of 1992, a voluntary exercise program was implemented to try to improve the health of the employees. At the end of 1992, a study was commissioned to try to understand the effect of the exercise program. The number of days of sick leave in 1992 was recorded and a regression was run. The variables are:

- \( SL_{92} \) = sick leave days in 1992,
- \( PAY \) = pay rate per hour (in dollars),
- \( SL_{91} \) = sick leave days in 1991,
- \( EXERCISE \) = 1 if enrolled in exercise program, \( EXERCISE=0 \) if not.

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>R Square (B)</td>
</tr>
<tr>
<td>Standard Error (D)</td>
</tr>
<tr>
<td>Error Observations 24</td>
</tr>
</tbody>
</table>

ANOVA
Df | SS
---|---
Regression | 216.403
Residual (C) | 43.253
Total | 216.403

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t-Stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.098</td>
<td>5.051</td>
<td></td>
</tr>
<tr>
<td>PAY</td>
<td>0.0365</td>
<td>0.4454</td>
<td>(E)</td>
</tr>
<tr>
<td>SL91</td>
<td>0.63940</td>
<td>0.06688</td>
<td></td>
</tr>
<tr>
<td>EXERCISE</td>
<td>-2.4375</td>
<td>0.6448</td>
<td>(A)</td>
</tr>
</tbody>
</table>

a) Fill in the following blank parts of the output:

(A) $t$-Stat for EXERCISE = ______

(B) R-squared = ______

(C) degrees of freedom for error = ______

(D) $s_e$ = ______

(E) $p$-value for PAY = ______

b) At the 5% level, test the null hypothesis that the exercise class has no effect on sick leave in 1992.

c) The designer of the exercise program has claimed that, all else being equal, someone taking the exercise program will on average reduce the number of sick days by 3 days/year. At the 5% level, test this claim against the one-sided alternative that the program does not reduce the number of sick days by 3 days/year.

d) Determine a 90% confidence interval for the amount by which an extra dollar in hourly pay affects the number of sick leave days in 1992.

e) Test at the 1% level the claim that the number of sick leave days in 1991 has an effect on the number of sick leave days in 1992.
EXAM #2

1. Select the best answer in each case.

a) Consider a test of $H_0: \mu = 100$ vs. $H_\alpha: \mu \neq 100$, where the underlying population is normal and the population variance is unknown. With sample size $n_1$ you obtain a sample mean $\bar{X}$ that leads you to reject the null hypothesis at significance level $\alpha$. If you increase the sample size to $n_2$, with $n_2 > n_1$, and obtain the same sample mean as before you will still reject $H_0$. This is
   ______ always true.
   ______ sometimes true, but not always.
   ______ never true.

b) You observe one observation of value $X = 4.5$ from a normal distribution with mean $\mu$ and standard deviation $\sigma = 1$. Pick one of the following:
   ______ A 95% confidence interval for $\mu$ is given by: _______________
   ______ It is not possible to compute a 95% confidence interval for $\mu$ for the following reason: _______________________

c) True or False. In a hypothesis test,
   ______ The $p$-value is the probability that the null hypothesis is true.
   ______ The $p$-value is the probability of a Type I error.
   ______ A small $p$-value indicates strong evidence in favor of the alternative hypothesis.
   ______ A large $p$-value indicates strong evidence in favor of the null hypothesis.

2. A consulting firm has been retained by Alphatech, Inc., for an analysis of its salary structure. For its initial investigation, the consulting firm has information on salary and characteristics of 46 Alphatech employees. The information available on each of these 46 employees is as follows:
   $SALARY = \text{yearly salary in thousands of dollars}$
   $YRS EM = \text{years of employment at Alphatech}$
   $PRIOR YR = \text{years of prior employment}$
   $EDUC = \text{years of education, high school and up}$
   $\text{GENDER} = 0 \text{ if male, 1 if female}$

   A regression produces the following output:

   \begin{tabular}{ll}
   \textbf{Regression Statistics} \\
   \hline
   R Square & (A) \\
   Std. Error & 5.823 \\
   \hline
   \end{tabular}
a) Obtain the following information:

(A) R-squared = _______

(B) degrees of freedom for Regression = ________

(C) degrees of freedom for Residual = ________

(D) SSR = ________

b) Based on the results above, construct a 95% confidence interval for the difference in average pay between men and women with the same number of years in the company, same number of prior years, and same number of years of education. How significant is this difference?

c) Based on the regression above, how much is a year of higher education worth, as far as salaries at Alphatech are concerned? Give a 95% confidence interval for it.

d) It has always been assumed that every year of employment at Alphatech increases your salary by $950. Test this claim against a two-sided alternative at the 1% level. What is the p-value of this test?

e) Based on the regression above, what is the average salary among female employees who have been at Alphatech for 4 years, have no years of prior experience, and have 9 years of higher education. Give a 95% “error” interval for this estimate of the average salary (is it a confidence or prediction interval?).

3. A large university wants to determine the average income their students earn during the summer. A random sample of 45 first-year business students produced the following statistics measured in hundreds of dollars: $\bar{x} = 33.1$ and $s = 5.0$.

a) Estimate the mean summer employment income for all first-year business students, with 99% confidence.
b) A statistician provides a confidence interval that runs from 31.5 to 33.8. Assuming he/she used the same sample data, what is the probability content of this interval?

4. The following regression was performed on data of sales and advertising of Crest toothpaste over the years 1967 through 1980. Crest sales seem to be related to the amount of advertising expenditures, the ratio of Crest advertising to Colgate advertising (the major competitor of Crest), and U.S. disposable family income. The variables are:

\[
\begin{align*}
Y &= \text{Crest sales (in millions of $)} \\
X_1 &= \text{Crest advertising expenditure (in millions of $)} \\
X_2 &= \text{Advertising ratio (Crest advertising$/Colgate advertising $)} \\
X_3 &= \text{U.S. disposable family income (in billions of $)}
\end{align*}
\]

The output for the data was as follows:

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
</tr>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Adjusted R-sq</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Df</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>(B)</td>
</tr>
<tr>
<td>Residual</td>
<td>(D)</td>
</tr>
<tr>
<td>Total</td>
<td>29657.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ccoeff.</th>
<th>Standard Error</th>
<th>t-Stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>34.1046</td>
<td>17.6541</td>
<td></td>
</tr>
<tr>
<td>X1</td>
<td>3.7459</td>
<td>1.9760</td>
<td></td>
</tr>
<tr>
<td>X2</td>
<td>30.0463</td>
<td>22.8596</td>
<td></td>
</tr>
<tr>
<td>X3</td>
<td>0.0859</td>
<td>0.0379</td>
<td></td>
</tr>
</tbody>
</table>

a) Obtain the following information:

(A) Multiple R = _______

(B) degrees of freedom for Regression = _______

(C) SSR = _______

(D) degrees of freedom for Residual = _______

b) It has always been assumed that disposable family income has no effect on sales of Crest. Test this hypothesis against a two-sided alternative at the 5% level.

c) Give a 95% confidence interval for the additional Crest sales that the model predicts.
would result from an increase of $3 billion in disposable family income.

d) In 1981, advertising of Crest and Colgate were both $250 million and disposable income was $2 trillion (=$2,000 billion). What would you predict as actual sales of Crest in that year?

e) Give a 90% “error” interval for your estimate of d)? (Is it a confidence or prediction interval?)

f) In part d), if Crest had decided at the last minute to add $50 million to its advertising budget (and assuming Colgate did not change its budget), what would the model now predict as sales of Crest?

5. A manufacturer of light bulbs wants to estimate the mean length of life of a new type of bulb that is designed to be extremely durable. The firm's engineers test nine of these bulbs and find that the mean length of life is 5,200 hours with a standard deviation of 150 hours. Previous experience indicates that the lengths of life of individual bulbs of a particular type are normally distributed. Construct a 90% confidence interval for the average length of life of all bulbs of this new type.