COLUMBIA
BUSINESS
SCHOOL

Practice Final Exam – Solutions

Fall 2001   Professor Assaf Zeevi
B6014 Managerial Statistics   406 Uris Hall

1. [20 points] With the introduction of HDTV, TV stations will eventually have to start producing and broadcasting their shows in this new (digital) standard. NBC, trying to decide on their own strategy, conducted a poll in major metropolitan areas and found that 39 out of 200 households was likely to purchase an HDTV within the next 18 months.

(a) (5 points) Give an estimate and a 90% confidence interval for the true proportion of American households in major metropolitan areas that are likely to purchase an HDTV over the next 18 month period.

$\hat{p} = .195$. The 90% CI will be

$$.195 \pm 1.645\sqrt{\frac{.195 \times .805}{200}} \;=\; (.149, .242)$$

(b) (3 points) NBC's strategy is to proceed with a substantial investment in new technology if they are at least 90% confident that the true proportion estimated using their poll is above 17%. Based on this strategy, will NBC proceed with this investment?

Given the 90% CI given above we cannot guarantee with 90% confidence that the true proportion p is above 17%; what we can guarantee with 90% confidence is that p lies between .149 and .242.

(c) Based on this preliminary estimate they wanted to conduct a more extensive poll in order to increase the accuracy of their estimate.

(i) (2 points) What should their sample size be if they wanted to reduce their margin of error by a factor of two (always at a 90% confidence level)?
Four times larger, that is $n = 800$.

(ii) (4 points) What should their sample size be if they wanted to estimate the true proportion of American households in major metropolitan areas that are likely to purchase an HDTV over the next 18 months to within $\pm 1\%$ with 90% confidence? (Use your best guess for the true proportion they are trying to estimate.)
We need that $1.645\sqrt{\frac{.195 \times .805}{n}} = .01 \Rightarrow n = 4248$. (It was OK to use $\hat{p} = 1/2$ to get a worse case estimate provided that you stated why you were doing that.)

(d) (6 points) An independent poll conducted in Europe was based on a sample of 312 households. The proportion of households that own or are likely to purchase an HDTV over the next 18 month period was estimated to be 22%. Construct a 95% confidence interval for the difference between the US and Europe proportions and commenton the difference.

$\hat{p}_E = .22$. The 95% CI for $(p_{US} - p_E)$ is

$$(.195 - .22) \pm 1.96\sqrt{\frac{.195 \times .805}{200} + \frac{.22 \times .78}{312}} = (-.0966, .0466).$$

We cannot guarantee with 95% confidence that p_{US} is smaller than p_E.

2. [10 points] The chief of police wishes to estimate the mean number of car thefts reported per weekend at the Morningside Heights area. From a sample of 50 weekends, the chief of police found 2 weekends with 7 occurences, 18 weekends with 9, 16 weekends with 10, and the remainder with 11 reported occurences. The chief of police wants to check whether average number of car thefts per weekend has changed due to the new "weekend patrol program" he helped initiate last year. The average number of car thefts per weekend prior to the initiation of that program was 11.5.

(a) (6 points) Based on the information collected the chief of police held a press conference praising the success of this new patrol program. Do you agree with his reaction? (Answer this question using a one-sided test at the 5% significance level – you will need to calculate the sample standard deviation of your estimate.)

$\bar{X} = (2 \times 7 + 18 \times 9 + 16 \times 10 + 14 \times 11)/50 = .8$. Also, $s_{\bar{X}}^2 = (2 \times 2.8^2 + 18 \times .8^2 + 16 \times .2^2 + 14 \times 1.2^2)/49 = .98$. We will test for

$$H_0 : \mu \geq 11.5 \quad \text{vs.} \quad H_1 : \mu < 11.5.$$

The test statistic is $Z = \frac{9.8-11.5}{s/\sqrt{n}} = \frac{9.8-11.5}{.99/\sqrt{50}} = -12.23$, which needs to be compared with the critical value of 1.645. Clearly, we can reject the null and agree with the chief.

(b) (4 points) What is the p-value for your test?

The p-value of this test is equal to $P(Z \leq -12.23) = 0$.

3. [18 points] A study of the ski lodging industry was conducted to develop a model for predicting midweek occupancy levels. A linear regression model was proposed with the following variables:

- Y = Occupancy rate, expressed as a percentage of maximum capacity,

- $X1$ = Number of competitors within a 1-mile radius,

- $X2$ = Nightly room rate, in $\$$

- $X3$ = Advertising expense, ski coupons, etc. as a percentage of annual gross revenues.

The data expressed as a percentage means, e.g., 43% was entered as 43.0 and 6.5% was entered as 6.5. The regression produced the following output:

```
Multiple R        0.611
R-square          0.374
Adj R-square      0.296
Standard Error    8.250
Observations      58

Analysis of Variance
            df    Sum of Squares
Regression    3        975.4
Residual     54       1633.3
Total        57       2608.7

             Coeff.    Std. Err.
Constant     56.319     4.802
X1           -1.947     0.987
X2          -0.00801   0.01063
X3            1.8405    0.5582
```

(a) (6 points) Fill in the following:

- $p-$value for $X1$: 4.88%

- $p-$value for $X2$: 45.3%

- $p-$value for $X3$: .1%

(b) (3 points) Does the model suggest (at the 10% level) that the nightly room rate has a statistically significant effect on the occupancy level?

The p-value is 45.3%, which is too high. No, it appears that there is not a significant relation between the the nightly room rate and the accupancy level.

(c) (4 points) Give a 95% confidence interval for the effect on occupancy rate if two new competitors open a hotel within a 1-mile radius.

On the average the effect of 2 new competitors within a 1-mile radius will be $2 \times (-1.947) = -3.894$. The 95% CI is: $-3.894 \pm 2(1.96)(.987) = (-7.764, -.024)$.

(d) (5 points) Using a 95% prediction interval, predict the occupancy level of the Hotel Kirkwood. It has two competitors within a 1-mile radius, spends 10% of its annual gross revenues on advertising and charges $55 a night.

$\hat{Y} = 56.319 + 2(-1.947) + 10(1.8405) + 55(-.008) = 70.27\%$. The 95% CI is: $70.27 \pm 1.96(8.25) = (54.1, 86.44)$

4. [12 points] It has been found that time taken by people to navigate, find, and complete a purchase order at an online bookstore follows a normal distribution. (We shall refer to this process as a *connection*.) A random sample of 45 connections gives a sample mean of 15.3 minutes with a sample standard deviation of 6.7 minutes.

(a) (4 points) Give a 95% confidence interval for the true average time per connection.

$\bar{X} = 15.3$ and $s = 6.7$. The 95% CI is

$$15.3 \pm 1.96\frac{6.7}{\sqrt{45}} = (13.34, 17.26)$$

(b) (4 points) What is the probability that the estimate of 15.3 minutes differs from the true average by 2 minutes or more?

$$2P(\bar{X} - \mu \geq 2) = 2P(Z \geq \frac{2}{6.7/\sqrt{45}}) = 2P(Z \geq 2) = .0456.$$

(c) (4 points) With the current level of communication resources for this online bookstore and their projected growth over the next 6 months, the company will be able to provide satisfactory service if the average connection time per customer is no more than 13.5 minutes. Based on the random sample of 45 connections would you recommend that the company upgrades their communication resources? (Perform a one-sided test at a 5% significance level.)

The two competing hypotheses are

$$H_0 : \mu \leq 13.5 \quad \text{vs.} \quad H_1 : \mu > 13.5.$$

The test statistic is $Z = \frac{15.3-13.5}{6.7/\sqrt{45}} = 1.801$. The cutoff point for a 1-sided test at the 5% level is 1.645. Hence, we reject the null and recommend to upgrade the communication resources.

5. [20 points] The cellular phone company gave to each new subscriber a mail-in coupon to be redeemed within two months of initiating service with this provider. From past experience only 53% of these coupons get redeemed. This semester 226 out of a total of 385 coupons were redeemed. Would you consider this as being consistent with past experience? Using a hypothesis test, explain your answer.

$\hat{p} = 226/387 = .587$. We formulate the following competing hypotheses

$$H_0 : p \leq 53\% \quad \text{vs.} \quad H_1 : p > 53\%.$$

The test statistic is $Z = \frac{.587-.53}{\sqrt{(.587\times.413)/385}} = 2.27$. The cutoff is at 1.645. Therefore, we reject the null and declare that there is statistically significant evidence that more coupons are being redeemed this year.

You could have answewred this using a 2-sided test. In this case, the only change is the cutoff point, which is now 1.96; again you should have rejected the null.

6. [20 points] Snowboard sales in ski stores account for roughly 14% of a store's gross revenues. However, snowboard's contribution to revenue is not a function of sales alone, but is thought to be influenced by the available selling space. In a random sample of 15 ski stores, data on the following variables was recorded X_1, the average weekly sales volume (in number of snowboards), X_2, the square feet of selling area for snowboards, and Y, the percentage of a store's gross revenues. Again, a percentage of 1.25% was entered as 1.25 and not 0.0125. The regression is as follows:

```
Multiple R        0.9041
```

```
R-square          B
Adj R-square      0.7871
Standard Error    0.6020
Observations      15


Analysis of Variance
              df    Sum of Squares
Regression    C     A
Residual      D     4.344
Total         E     23.800


              Coeff.    Std. Err.
Constant  8.51127      0.67586
X1        0.07150      0.01120
X2        0.00150      0.00032
```

(a) (6 points) Fill in the blank parts of the regression output:

- (A) Regression sums of squares $= 19.456$

- (B) R-squared $= 81.75\%$

- (C) Regression degrees of freedom $= 2$

- (D) Residual degrees of freedom $= 12$

- (E) Total degrees of freedom $= 14$

(b) (4 points) Based on the results above, construct a 95% confidence interval for the increase in the share of gross revenues that would result from one thousand extra square feet of selling space for snowboards.

The 95% CI is: $1000(.00150) \pm 2.179(1000)(.00032) = (.803, 2.179)$.

(c) (5 points) With 90% confidence, predict the percentage of gross revenues from snowboard sales for a store with 2200 square feet of selling area allocated to snowboards and average weekly sales of 40 snowboards. Give an error interval for your estimate. What is the interpretation of this error interval?

$\hat{Y} = 8.51127 + 2200(.00150) + 40(.07150) = 14.67$ The 90% CI is: $14.67 \pm 1.783(.6020) = (13.60, 15.74)$.

(d) (5 points) In another study, the advertising budget (radio ads, coupons, discounts, newspaper ads) of each store was also included. The new variable is denoted by X_3 and is measured in tens of thousands of $ (i.e., $X_3 = 1.2$ implies that the advertising budget of that store was $12,000). Part of the new regression output was:

```
Standard Error    0.4020
Observations      15
```

```
          Coeff.    Std. Err.
Constant  6.01127   0.42345
X1        0.06242   0.00920
X2        0.00197   0.00041
X3        0.82357   0.14578
```

It has been assumed that the percentage of gross revenues due to snowboard sales increases by .6% for every \$10,000 spent in advertising and promotions. Test this claim at the 5% level against a one-sided alternative.

We are testing the following competing hypotheses

$$H_0 : \beta_{adv} \leq .6 \quad \text{vs.} \quad H_1 : \beta_{adv} > .6.$$

The T-statistic is $T = \frac{.82357-.6}{.14578} = 1.533$. The cutoff is the point $t_{n-k-1,.05} = t_{11,.05} = 1.796$. Hence, we cannot reject the null (since $1.533 < 1.796$); that is, the evidence is not strong enough (or statistically significant) to suggest that there is a greater effect than .6% for each \$10,000 spent in advertising and promotions.