MgtOp 215
Dr. Ahn

FINAL

Fall 2017

Name: ____________________

ID: ____________________

Section (Circle one): 7, 8, 9

Read the following **instructions** very carefully before you start the test.

- Make sure to write **your name** and **id** on the exam.
- This test is **closed** book and notes; three summary sheets are allowed.
- Show **all** your work **clearly** and **circle** your answers. **No** credit will be given to answers without proper work shown or without being circled.
- **Define** symbols properly, especially, symbols for random variables and parameters. Otherwise, **at least 2** points will be deducted.
- If you need more space, you may use the back of the test, and **indicate so**.
- Each question is worth **4** points; otherwise the value is specified.
- It is your responsibility to check your copy of the test consists of **9** pages including the cover page.
- You may find the MS Excel output on page 8 and the PHStat output on page 9 useful.
- There are questions on page 8.
- Make sure to put **your name** or initials on each and every page.

Good Luck
and
Have a Happy Holiday Season

Page 1: Cover page	Page 2:
Page 3:	Page 4:
Page 5:	Page 6:
Page 7:	Page 8: MS Excel output
Page 9: PHStat output	
Total:	out of **125**

Name: ______________

1. The following are the out-of-state tuition rates (in thousands of dollars) of randomly selected state universities.

   6.4, 12.0, 4.9, 6.4, 8.5, 11. 8, 7.0, 7.6

   a. Find the mean of the out-of-state tuition rates. (**3** pts.)

   b. Find the variance of the out-of-state tuition rates. (**3** pts.)

   c. Find the median of the out-of-state tuition rates. (**3** pts.)

   d. Find the 75^{th} percentile of the out-of-state tuition rates. (**3** pts.)

   e. Find the (estimated) standard error of the sample mean. (**3** pts.)

Name: ______________

2. According to a tire manufacturer, the tread life of a certain type of tire is a normal random variable with a mean of 45 thousand miles and a standard deviation of 5 thousand miles.

   a. Find the proportion of tires with the tread life between 35 thousand miles and 60 thousand miles.

   b. The manufacturer is to determine the warranty mileage for this tire. Any tire that wears out before this mileage will be replace free. What should be the warranty mileage if the manufacturer does not want to replace more than 1% of the tires free?

   c. Suppose 4 (randomly selected) tires are installed on a car. Find the probability that the sample mean of the tread life of these four tires is between 40 thousand and 50 thousand miles.

3. Ten percent of all shoppers who visit the Internet Web-site make a purchase. Suppose that twenty customers visit the Web-site. Find the probability that at least one shopper makes a purchase. Assume that these shoppers make independent shopping decisions.

4. The number of complaints about servers filed at a restaurant per day is a Poisson random variable with a mean of 3 complaints per day. Find the probability that there are three complaints about servers filed at a restaurant on a given day.

Name: ______________

5. A pharmaceutical company claims that its newly developed drug is more effective in that it can lower the level of harmful chemical in the body, on average, by more than 10 units when administered to patients suffering from certain illness. An investigator is about launch a statistical study.

   a. State the null and alternative hypothesis, **very specific to this study**. Justify your choice of the null and the alternative hypothesis

   b. Discuss when the investigator commits the Type I error **very specific to this study**. (3 pts.)

6. Refer to Problem 5. The investigator takes a random sample of 36 patients and administers the new drug. It is found that the average and the standard deviation of the levels of the chemical lowered are 12 and 8 units, respectively.

   a. Compute the test statistic appropriate for the hypotheses stated in Question 5-a. (**3** pts.)

   b. Compute the *p*-value of the test statistic in Question 6-a. (**3** pts.)

   c. What is the conclusion of the investigator's study? Be **very specific to this study** and justify using $\alpha = 0.05$. (**3** pts.)

   d. According the conclusion in Question 6-c, which type of error may be committed, Type I or Type II? Justify. (**3** pts.)

Name: ____________

7. In a poll of voters at a state in the US, 90 of 200 voters prefer free college education for everyone.

   a. Compute a 95% confidence interval for the proportion of all the voters this state who prefer free college education.

   b. Interpret the confidence interval obtained in Question 7-a.

   c. If one wants to estimate the proportion within 0.04 of the population proportion with 95% confidence, determine the **additional** sample size required based on your answer in Question 7-a.

8. Refer to Problem 7. It is claimed that the minority of the voters, that is, less than 50% of the voters, in this state favor free college education.

   a. State the null and alternative hypothesis to test the claim, **very specific to this problem**. Justify your choice of the null and the alternative hypothesis

   b. Based on the poll result in Problem 7, test the claim at the level of significance 0.05.

Name: ______________

9. A manufacturer of flashlight batteries took a sample of 8 batteries from a day's production and used them continuously until they were drained. The mean and the standard deviation of the number of hours they were used until failure were 470.4 and 95.4 hours, respectively.

   a. Assuming that the numbers of hours are approximately normally distributed, find the 95% confidence interval for the mean of the number of hours of all the batteries until they fail.

   b. If one tests whether the mean of the number of hours of all the batteries until they fail is 500 or not at $\alpha = 0.05$, would the null hypothesis be rejected? Justify your answer based on the confidence interval obtained in Question 9-a. (**3** pts.)

   c. Is the assumption made in Question 9-a appropriate? Justify your answer. (**3** pts.)

10. A magazine publishes restaurant ratings for various locations around the world. The magazine rates the restaurants for food, service, and the cost per person. In order to develop a regression model to predict the cost per person in dollars (Y) based on a rating variable (X) that represents the sum of the three ratings, a quant manager at the magazine sampled restaurants in New York City, used Excel, and produced the output on the last page of this exam.

    a. Find the correlation between the cost and the rating. (**3** pts.)

    b. How would you decide if a simple linear regression mode is appropriate for the data? Provide two ways to decide. (**3** pts.)

Name: ______________

c. Find the estimated regression line and interpret this line. (**6** pts.)

d. For an observation with $X = 75$ and $Y = 77$, find the residual for this observation. (**3** pts.)

e. Find the estimate of the variance of the error term. (**3** pts.)

f. Is there a strong evidence that there is a linear relationship between the rating (X) and the cost (Y)? Justify at α=0.05.

g. Find a 90% confidence interval for the slope parameter β_1.

h. For the data analyzed, test if the estimated β_1 is significantly smaller than 1.5. Use α=0.05. Make sure include the hypotheses and the conclusion. (**6** pts.)

Name: ______________

i. Refer to the PHStat output on p. 9. Interpret the 90 confidence interval for the mean of the costs of the restaurants with the rating of 70. (**4** pts.)

i. Find a 95% prediction interval for the cost with when the rating is 70 and interpret the interval. (**8** pts.)

Descriptive Statistics

Rating (X)		*Cost (Y)*	
Standard Deviation	6.629068219	Standard Deviation	13.42572
Minimum	49	Minimum	25
Maximum	78	Maximum	80
Sum	6257	Sum	4685
Count	100	Count	100

Excel output of covariance

	Rating (X)	*Cost (Y)*
Rating (X)	43.5051	
Cost (Y)	65.0955	178.4475

Regression output

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	9740.063	9740.063	117.775	1.727E-18
Residual	98				
Total	99	17844.75			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-46.772	8.675	-5.392	4.827E-07	-63.986	-29.557
Rating (X)	1.496	0.138	10.852	1.727E-18	1.223	1.770

Name: ______________

Confidence Interval Estimate

Data	
X Value	**70**
Confidence Level	**90%**

Intermediate Calculations	
Sample Size	100
Degrees of Freedom	98
t Value	1.660551
XBar, Sample Mean of X	62.57
Sum of Squared Differences from XBar	4350.51
Standard Error of the Estimate	9.094003
h Statistic	0.022689
Predicted Y (YHat)	57.96731

For Average Y	
Interval Half Width	2.2747
Confidence Interval Lower Limit	**55.6926**
Confidence Interval Upper Limit	**60.24197**

For Individual Response Y	
Interval Half Width	15.2714
Prediction Interval Lower Limit	**42.6959**
Prediction Interval Upper Limit	**73.23872**

Name: ____________

Solutions (Maybe subject to errors)

1.

a. mean: $\bar{x} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{6.4+12+4.9+6.4+8.5+11.8+7+7.6}{8} = \frac{64.6}{8} = 8.075$

b. Variance: $\bar{x} = \frac{\sum_{i=1}^{8}(x_i-\bar{x})^2}{8-1} = \frac{(6.4-8.075)^2+\cdots+(7.6-8.075)^2}{7} = \frac{2.806+\cdots+0.225625}{7} = \frac{46.535}{7} =$ 6.6479

c. Arrange in ascending order: 4.9, 6.4, 6.4, 7.0, 7.6, 8.5, 11. 8, 12.0.

Since there are even number of observations, the median is the simple average of the middle two numbers, which are the fourth and the fifth observation: $Median = \frac{7+7.6}{2} = 7.3$

d. The 75th percentile: $i = \frac{n \times i}{100} \rightarrow i = \frac{8 \times 75}{100} = 6$. Since this is an integer, the 75th percentile is the average of the 6th and 7th observations. 5^{th} percentile $= \frac{8.5+11.8}{2} = 10.15$

e. The estimated standard error: $\frac{s}{\sqrt{n}} = \frac{\sqrt{6.6479}}{\sqrt{8}} = \frac{2.5783}{\sqrt{8}} = 0.911582$

2. X: the tread life in thousand miles

a. $P(35 \leq X \leq 60) = P\left(\frac{35-45}{5} \leq Z \leq \frac{60-45}{5}\right) = P(-2 \leq Z \leq 3)$

$$= P(Z \leq 3) - P(Z \leq -2) = 0.9987 - 0.0228 = 0.9759$$

b. $P(X \leq x) = 0.01 \rightarrow P\left(\frac{X-45}{5} \leq \frac{x-45}{5}\right) = 0.01 \rightarrow P\left(Z \leq \frac{x-45}{5}\right) = 0.01$

We know from the standard normal table that: $P(Z \leq -2.33) \approx 0.01$. This suggests that

$$\frac{x-45}{5} = -2.33 \rightarrow x = 33.37$$

c.

$$(40 \leq \bar{X} \leq 50) = P\left(\frac{40-45}{5/\sqrt{4}} \leq Z \leq \frac{50-45}{5/\sqrt{4}}\right) = P\left(\frac{-5}{2.5} \leq Z \leq \frac{5}{2.5}\right)$$

$$= P(Z \leq 2) - P(Z \leq -2) = 0.9772 - 0.0228 = 0.9545$$

3. X: Number of shoppers who visits the internet website and purchase (Binomial with $n = 20, \pi = 0.1$)

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{20}{0}\ 0.1^0\ \ 0.9^{20} = 1 - 0.1216 = 0.8784$$

4. X: Number of complaints about servers filed per day (Poisson $\lambda = 3$)

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \to P(X = 3) = \frac{e^{-3}3^3}{3} = 0.2240$$

5.

a. μ: The mean reduction in the level of harmful chemicals in the body. The new claim (i.e., more than 10 units reduction on average in harmful chemical) is listed under the alternative hypothesis.

$$H_0: \mu \leq 10$$
$$H_1: \mu > 10$$

b. Type I error is committed when the investigator concludes that the new drug can reduce the average level of harmful chemical by more than 10 units (i.e., $\mu > 10$), although in reality, the newly developed drug is not more effective (i.e., $\mu \leq 10$).

6.

a.

$$\bar{x} = 12, s = 8, n = 36$$

$$Z_{stat} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \to Z_{stat} = \frac{12 - 10}{\frac{8}{\sqrt{36}}} = 1.5$$

b. $p - value = P(Z > 1.5) = 0.0668$

c. Since the $p - value$ is larger than $\alpha = 0.05$, the we fail to reject the null hypothesis. That means we don't have sufficient evidence that the new drug is more effective in lowering the level of harmful chemical by more than 10 units.

d. Since we failed to reject the null hypothesis, we might have committed type II error.

7.

a. $p = \frac{90}{200} = 0.45, \alpha = 0.05 \to Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$

$$p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \to 0.45 \pm 1.96\sqrt{\frac{0.45(1-0.45)}{200}} \to 0.45 \pm 0.0689$$

$$[0.3811, 0.5189]$$

b. With 95% confidence, the true proportion of voters in the US who prefer college education is between 38.11% and 51.89%.

c. $n \geq p\,(1-p)\left(\frac{z_{\frac{\alpha}{2}}}{e}\right)^2 \rightarrow n \geq 0.45\,(1-0.45)\left(\frac{1.96}{0.04}\right)^2 = 594.2257 \rightarrow n = 595$

The additional sample size is at least 595-200=395.

8.

a. π: the proportion of voters who favor free college education.

$$H_0: \pi \geq 0.5$$
$$H_1: \pi < 0.5$$

Need strong evidence that the minority of the voters, that is, less than 50% of the voters, in this state favor free college education

b. $p = \frac{90}{200} = 0.45, n = 200$

$$z_{stat} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0.45 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}} = -1.4142$$

Critical value: $-z_\alpha = -z_{0.05} = -1.645$

Since $z_{\text{stat}} \nleq -z_\alpha$, we don't reject the null hypothesis. That is, there is no strong evidence that the proportion of voters who favor free college education is at least 50%.

9.

a. $\bar{x} = 470.4, s = 95.4,\ n = 8, \alpha = 0.05$

$$t_{\frac{\alpha}{2},n-1} = t_{0.025,7} = 2.3646$$

$$\bar{x} \pm t_{\frac{\alpha}{2},n-1} \frac{s}{\sqrt{n}} \rightarrow 470.4 \pm 2.3646\,\frac{95.4}{\sqrt{8}} \rightarrow 470.4 \pm 79.7564$$

$$[390.64, 550.16]$$

b. Since 95% confidence interval contains 500, then $\mathrm{H_0}: \pi = 500$ cannot be rejected.

c. No. The battery life time cannot be negative. However, if the probability of negative life times is very small, (it is about 4.09×10^{-7}in this problem which is by $P(X < 0)$.), and if the histogram of the life time appears to be bell shaped, then the assumption would be appropriate.

10.

a.

$$r = \frac{s_{xy}}{s_x\ s_y} = \frac{65.0955}{\sqrt{43.5051} \times \sqrt{178.4475}} = 0.7388$$

b. Checking the scatter plot for an observable linear pattern

Checking the existing theory

c.

$$\hat{Y} = -46.772 + 1.496x$$

1.496 is the slope. So, if the rating is higher by 1, the cost per person will be higher by 1.496.

-46.772 is the intercept. So, if the rating is zero, the cost per person is -46.772 (Although this is not possible).

d.

$$\hat{Y} = -46.772 + 1.496 \times 75 = 65.428$$
$$e = 77 - 65.428 = 11.572$$

e.

$$\hat{\sigma}^2 = MSE = \frac{SSE}{d.f.} = \frac{SST - SSR}{d.f.} = \frac{17844.75 - 9740.063}{98} = 82.70089$$

or

$$\hat{\sigma}^2 = \frac{MSR}{F} = \frac{9740.063}{117.775} = 82.70089$$

f. Since the p-value for testing $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ is very small (1.727×10^{-7}), there is strong evidence $\beta_1 \neq 0$, that is strong evidence for a linear relationship.

g.

$$t_{\frac{\alpha}{2},n-2} \rightarrow t_{0.05,98=}1.6606$$
$$b_1 \pm t_{\frac{\alpha}{2},n-2}\, SE(b_1) = b_1 \pm t_{\frac{\alpha}{2},n-2}\, SE(b_1) = 1.496 \pm 0.2291$$
$$[1.2668,1.7252]$$

h.

$$H_0: \beta_1 \geq 1.5$$
$$H_1: \beta_1 < 1.5$$

$$t_{stat} = \frac{1.496 - 1.5}{0.138} = -0.02899$$

Critical value:$-t_{\frac{\alpha}{2},n-2} = -1.6606$

The null hypothesis is not rejected. There is no strong evidence that $\beta_1 b_1 < 1.5$. That is, b_1 is not significantly smaller than 1.5.

i. The average cost per person when the rating is 70 is between \$55.69 and \$60.24.

j.

$$t_{\frac{\alpha}{2},n-2} \to t_{0.025,98=}1.9845$$

$$\hat{Y}_j + t_{\frac{\alpha}{2},n-2}\hat{\sigma}\sqrt{1+h_j}$$

$$x_j = 70 \to \hat{Y}_j = -46.772 + 1.496 \times 70 = 57.948$$

$$57.948 \pm 1.9845 \times 9.094003\sqrt{1+0.022689} \to 57.948 \pm 18.25063$$

$$[39.6974, 76.1986]$$

With 95% confidence, the cost a person when the rating is 70 is between is between \$39.6974 and \$76.1986.

This solution is prepared by Mr. Sadegh Kazemi.