CROWDMARK

Please print in pen:

Waterloo Student ID Number:

WatIAM/Quest Login Userid:

UNIVERSITY OF WATERLOO

Examination
Quiz 3 (Version A)
Spring 2019
STAT 231

Times: Friday 2019-07-19 at 11:30 to 12:20

Duration: 50 minutes

Exam ID: 4165847

Sections: STAT 231 LEC 001-003

Instructors: James Adcock, Surya Banerjee

Special Materials

Candidates may bring only the listed aids.

- Calculator - Pink Tie (Blue goggle calculators are also permitted.)

Be sure to show all work where necessary.

Distribution tables / formulas provided separately.

Solutions.

By: Chi-kuang Yeh.

Date: Jul-26-2019.

	Q1 – Q8	8
	Q9	5
	Q10	13
	Q11	9
Total		35

Aids: Only question pages will be marked. Final numerical answers must be rounded to exactly 3 decimal places, unless the answers are exact to less than 3 decimal places. For intermediate steps, please be sure to keep all decimal places to avoid round off errors.

Only faculty-approved (pink-tie or blue goggle) calculators are allowed. Show all calculation based work.

Please initial:

*** This page was intentionally left blank. ***

CROWDMARK

Multiple Choice Questions – 1 mark for each question. Please clearly circle your answer.

1. Suppose you wanted to test H_o: $\mu = 5$ vs. H_a: $\mu \neq 5$. In order to do so, a random sample of 16 observations was taken from a Gaussian population with a known variance, $\sigma^2=4$. The mean of the sample was 4.35. Based on this information, determine the value of the test statistic / discrepancy measure to be used to test H_o.

A. 0.65   B. 0.8064   C. 1.30   D. 5

$$d = \frac{|\bar{y} - \mu_0|}{\sigma/\sqrt{n}} = 1.30$$

$\bar{y} = 4.35$, $\mu_0 = 5$

$n = 16$, $\sigma^2 = 4 \rightarrow \sigma = 2$.

2. A random sample of 20 observations drawn from a Gaussian population, with an unknown standard deviation. The 90% confidence interval for μ was given by [12.24, 14.60]. Would you reject H_o: $\mu = 15$ vs. H_a: $\mu \neq 15$ at the α = 10% level of significance?

A. Yes, since μ_o = 15 lies inside the 90% CI

B. Yes, since μ_o = 15 lies outside the 90% CI

C. No, since μ_o = 15 lies inside the 90% CI

D. No, since μ_o = 15 lies outside the 90% CI

$\mu_0 = 15 \in CI \rightarrow$ don't reject H_0

$\notin CI \rightarrow$ reject H_0.

3. Which of these best describes the least squares regression line – that is, the regression line obtained using the method of least squares?

A. It is the line which makes the sum of the squared distances perpendicular to the line as small as possible.

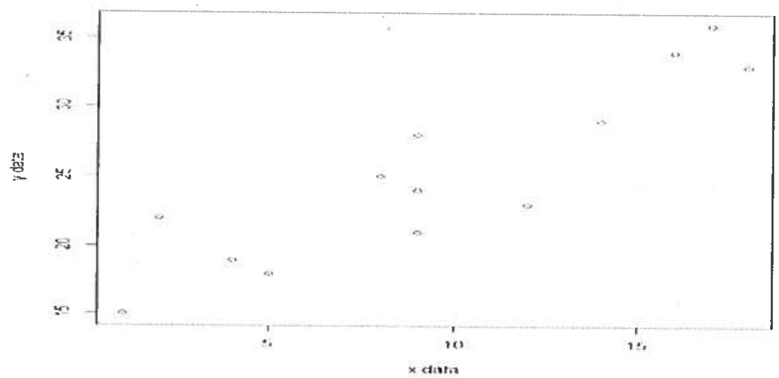
B. It is the line which best splits the data in half, with 50% of the data points lying above the regression line and 50% of the data points lying below the regression line.

C. It is the line which makes the sum of the squared horizontal differences between the observed values and predicted values as small as possible.

D. None of these describes the least squares regression line

Please initial:

Data were entered into R using labels x.data and y.data. Examine the scatterplot and numerical output below answer the following **TWO** questions:



Call:

lm(formula = y.data ~ x.data)

Residuals:

Min	1Q	Median	3Q	Max
-4.7172	-1.2621	-0.5931	2.1172	4.6966

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	15.2207	1.6492	9.229	1.64e-06 ***
x.data	1.0414	0.1505	6.922	2.51e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.93 on 11 degrees of freedom

$y = \alpha + \beta x.$

4. The p-value to test H_o: $\beta = 0$ vs. H_a: $\beta \neq 0$, suggests that we have ________________ against H_o.

A. weak evidence

B. evidence

C. strong evidence

D. very strong evidence

5. The predicted value of y when x = 5 is:

$\hat{y} = \hat{\mu}(5) = 15.2207 + 1.0414(5) = 20.4277.$

A. 20.4277

B. 15.2207

C. 5.207

D. 5

Please initial:

6. A random sample of 14 fibers was tested for their strength. Their strength had a sample mean of 266.5 and a sample standard deviation of 18.6. The strengths are approximately Normally distributed. Calculate the appropriate p-value in order to assess whether or not there is evidence to suggest that the average strength of fibres of this type is different from 260.

$\bar{y}$

$H_0: \mu = 260$
$H_a: \mu \neq 260$

2 sided.

$2(0.1) < p\text{-value} < 2(0.2)$

A. p-value = 0.0951

B. p-value = 1.31

C. 0.1 < p-value < 0.20

D. p-value > 0.20

$$d = \frac{|\bar{y} - \mu_0|}{s/\sqrt{n}} = \frac{|266.5 - 260|}{18.6/\sqrt{14}} = 1.307 \sim t_{(14-1)} \text{ under } H_0$$

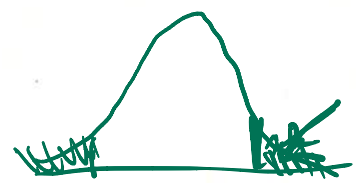
0.8   0.9

0.8702   1.3502

7. Suppose that a random sample of n observations is drawn from a G(μ, σ) population. Both μ and σ are unknown. We wish to test H_o: μ = 10 vs. H_a: μ ≠ 10. The sample data is to be stored in a vector, y, and the following R code has been run:

t — df = n-1 (1-parameter case)

```
ybar<-mean(y)
n<-length(y)
s2<-var(y)
s<-sqrt(s2)
mu.not<-10
d<-abs((ybar-mu.not))/(s/sqrt(n))
```



Which of the following R commands will calculate the p-value in order to test H_o?

A. 2*pt(d,n-1)-1   B. 2*(1-pnorm(d))   C. 2*(1-pt(d,n-1))   D. 2*(1-pt(d,n))   E. 2*pnorm(d)-1

8. A student was interested in modeling a relationship between final course grades (Y) and hours studied (x). A simple linear regression was performed, yielding the following equation: $\hat{y} = 33.4 + 3.25x$. How would you interpret $\hat{\beta}$ in this equation?

A. It represents the predicted change in (mean) course grade for every increase of 1 hour of study time.

B. It represents the predicted (mean) grade when a student does not study at all (x=0).   no change.

C. It represents the predicted change in (mean) study time for every increase of 1 mark in a course.

D. It represents the predicted (mean) study time that a student needs to study in order to get 100%.

E. More than one of these are correct.

9. True / False Questions (1 mark each). Please write True or False in each blank below.

a. Some data have been collected in order to learn about a model parameter, θ. A 95% confidence interval for θ is [0.016, 0.424] and a 99% confidence interval for θ is [-0.048, 0.488]. Consider a test of H_o: θ = 0 (versus a two-sided alternative). The resulting p-value would be less than 5% but greater than 1%. ___T___

$\theta \in CI_{99\%} \rightarrow$ p-value > 0.01

$\theta \notin CI_{95\%} \rightarrow$ p-value < 0.05

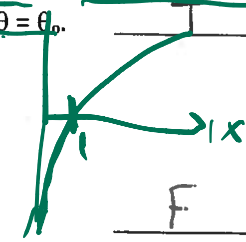
b. Suppose for data (x_i, y_i), i = 1, 2, ..., n we assume the model $Y_i \sim G(\alpha + \beta x_i, \sigma)$, i = 1, 2, ..., n; independent. Let $\tilde{\beta}$ be the estimator corresponding to the least squares estimate of β. Then $\tilde{\beta} \sim G(\hat{\beta}, \frac{\sigma}{\sqrt{S_{xx}}})$ ___F___

$\hat{\beta} \sim G(\beta, \sigma/\sqrt{S_{xx}})$

c. Let R(θ) be the relative likelihood function and Λ(θ) be the likelihood ratio test statistic. Smaller values of $R(\theta_o)$ and larger values of $\Lambda(\theta_o)$ are consistent with stronger evidence against H_o: $\theta = \theta_o$. ___T___

$0 < R(\theta_0) = \frac{L(\theta_0)}{L(\hat{\theta})} < 1 \quad \Lambda = -2\ln\left(\frac{L(\theta_0)}{L(\hat{\theta})}\right)$

↑ MLE



d. A p-value represents P(H_o is true, based on the data). ___F___

p-value = P(we obs. test statistics value as extreme or more extreme than the value observed | H_0 is true)

e. A likelihood ratio test statistic is to be used to test H_o: θ = 2, assuming that the data come from a Poisson model. A random sample of 20 observations yielded a sample mean of 3.3. The test statistic used to test H_o has an exact chi-squared distribution with 1 degree of freedom. ___F___

approxim. exact

$\Lambda(\theta_0) \sim \chi^2_{(1)}$?

Written Questions: Show all of your work in the space provided for full credit.

10. ***[13]*** A manufacturer of safety helmets for construction workers is concerned about the mean and the variation of the forces that its helmets transfer to those who wear the helmets, when subjected to an external force. The manufacturer has designed the helmets so that the mean force transferred by the helmets to the workers is 800 pounds or less, with a standard deviation of 40 pounds. Tests were run on a random sample of n = 20 helmets. The sample results are given below:

$$\bar{y} = 810 \text{ and } s = 45.3$$

Let Y_i represent a force transfer measurement. We will assume $Y_i \sim G(\mu, \sigma)$, i = 1,2, ..., 20; independent.

In order to see how well the helmets are working, a statistician was asked to test H_o: $\sigma = 40$ (vs. H_A: $\sigma \neq 40$).

a. Define the test statistic, U, to be used to test H_o. Be sure to state the distribution of U, assuming H_o is true. Do **NOT** give a value here, just give the test statistic, U, and the type of distribution that U follows. ***[2]***

$U = \frac{(n-1)S^2}{\sigma_0^2}$. Here, U has a χ^2 dist^{n} $\bar{w}$ $\overset{20}{n}-1=19$ d.f., assuming Ho is true. $\sigma_0 = 40$

p/22

b. Now, based on the sample information given, calculate the observed value of the test statistic. ***[2]***

$u_0 = \frac{(20-1)45.3^2}{40^2} = 24.369$ $\sim \chi^2(19)$ under Ho

c. Use the information in parts a. and b. to calculate the appropriate p-value to test H_o. **Note:** You may state a range of values for the p-value here, if necessary. ***[3]***

From the χ^2 table $\bar{w}$ d.f. = 19, we see:

0.8	0.9
23.9	27.204

$\uparrow u_0 = 24.369$

$\therefore$ 2-sided p-value $= 2\cdot(1 - P(U \cancel{\leq} 24.369))$ $\leq$

In our case, $2(0.1) < \text{p-value} < 2(0.2)$

$\therefore 0.2 < \text{p-value} < 0.4$

***** ***This question is continued on the next page.*** *****

d. Based on your p-value calculation in part c., assess the amount of evidence against H_o, using Table 5.1 guidelines. ***[2]***

According to Table 5.1 guidelines, $\overline{w}$ $0.1<$ $0.2 < p\text{-value} < 0.4$, we have no evidence against H_0.

e. At the $\alpha = 5\%$ level of significance, would you reject or not reject H_o? Justify your answer. ***[2]***

At $\alpha = 5\%$, we CANNOT REJECT / FAIL TO REJECT H_0 as the p-value $> \alpha = 0.05$.//

$\alpha = 0.05 <$ p-value

f. Now, draw a conclusion in the context of the question, at the 5% level of significance (i.e. do the helmets appear to be working properly, as designed?) ***[2]***

Conclusion: It appears as though the helmets are working properly, as designed.

11. ***[9]*** Suppose that a Binomial(n, θ) model has been proposed for the random variable Y. In this case, Y represents the number of successes in n independent Bernoulli trials and θ represents the probability of success. In this case, the maximum likelihood estimate of θ is $\hat{\theta} = \frac{y}{n}$. You do not have to show this though.

a. Given that the relative likelihood function for the Binomial model is $R(\theta) = \frac{L(\theta)}{L(\tilde{\theta})} = \left(\frac{\theta}{\tilde{\theta}}\right)^{Y}\left(\frac{1-\theta}{1-\tilde{\theta}}\right)^{n-Y}$, determine $\Lambda(\theta)$. ***[2]***

$$\Lambda(\theta) = -2\log\left[R(\theta)\right] = -2\log\left[\left(\frac{\theta}{\tilde{\theta}}\right)^{Y}\left(\frac{1-\theta}{1-\tilde{\theta}}\right)^{n-Y}\right]$$

$$\Rightarrow \Lambda(\theta) = -2\left[Y\log\left(\frac{\theta}{\tilde{\theta}}\right) + (n-Y)\log\left(\frac{1-\theta}{1-\tilde{\theta}}\right)\right]$$

You could substitute $\tilde{\theta} = \frac{Y}{n}$ also, but not necessary here for full credit.

b. We wish to test H_0: θ = 0.6 by performing a likelihood ratio test. From a random sample of n = 36 observations, the maximum likelihood estimate of θ was $\hat{\theta} = \frac{18}{36} = 0.5$. Using this information, determine the observed value of the likelihood ratio test statistic, $\lambda(\theta_0)$. ***[2]***

$$\lambda(0.6) = -2\left[18\log\left(\frac{0.6}{0.5}\right) + 18\log\left(\frac{0.4}{0.5}\right)\right]$$

$\lambda(0.6) = 1.470$

$\theta = 0.6 \ (H_0)$

$\hat{\theta} = 0.5 = y/n$

$= y/36$

$\rightarrow y = 18$

**** This question is continued on the next page. ****

Please initial

CROWDMARK

c. Use your result in part b. to calculate the p-value in order to test H_o. Be sure to state the approximate distribution of $\Lambda(\theta_o)$, assuming H_o is true, in order to justify your p-value calculation. *[3]*

$P(W > 1.4670) \approx$ p-value where $W \,\dot\sim\, \chi^2(1)$

p-value $\approx P(|Z| > \sqrt{1.4670})$

$= 2\cdot(1 - P(Z \leq \sqrt{1.4670}))$

$= 2\cdot(1 - P(Z \leq 1.21))$

$= 2\cdot(1 - 0.88686) = 0.226$

Aside: approx.

$\Lambda(\theta_0) \,\dot\sim\, \chi^2(1)$, assuming Ho is true

$Z \sim N(0,1)$

$Z^2 \sim \chi^2(1)$

d. Using Table 5.1 guidelines, assess the amount of evidence against H_o, based on the data. *[2]*

According to Table 5.1 guidelines, we have no evidence against Ho, based on the data, $\bar{w}$ a p-value $\approx 0.226 > 0.1$

Please initial: