Stats 105 - Midterm Spring 2014

NAME:

Answer Key

SECTION:

Academic Misconduct:

- **Any potential violation of UCLA's policy on academic integrity will be reported to the office of the Dean of Students. All work on this exam must be your own.**

Test Instructions

- Show all your work for the open ended question in order to get credit.
- Write clearly. Short answers are best.
- You may use a calculator, but you may NOT share with your neighbor.
- Cell phones, iPods or other Internet connected devices are NOT allowed to be used as calculators.

GOOD LUCK!

1. Patients in a hospital are classified as surgical or medical. A record is kept of the number of times patients require nursing service during the night and whether or not these patients are on Medicare. The data is presented in the table below.

	Surgical	Medical
Medicare	46	52
No Medicare	36	43

2 points

(a) Write the null and alternative hypothesis for a contingency table test to test whether type of medical patient is independent of whether patients are receiving Medicare.

H_0: The type of medicare the patient has is independent on type of patient they are

H_a: Type of medicare is dependent on type of patient.

3 points

(b) Compute the χ^2 test statistic for a contingency table test on these results. Show your work.

surgery $= \frac{82}{177} = 0.463$ Medical $\frac{95}{177} = .537$ | Medicare $\frac{98}{177} = .554$ No Medicare $\frac{79}{177} = .446$

$177(0.463)0.554 = 45.4$

$177(0.463)0.446 = 36.55$

$177(0.537)0.554 = 52.657$

$177(0.537)0.446 = 42.392$

	Surg.	Medical
Medi	45.4	52.657
No Medi	36.55	42.392

$$\chi^2 = \frac{(46-45.4)^2}{45.4} + \frac{(52-52.657)^2}{52.657} + \frac{(36-36.55)^2}{36.55} + \frac{(43-42.392)^2}{42.392}$$

$$\chi^2 = 0.033$$

2 points

(c) Given a critical $\chi^2 = 6.63$ for the rejection region on 1 degree of freedom what is the conclusion of your hypothesis test?

$\chi^2 = 0.033 < 6.63 = \chi^{2*}$ Fail to reject Ho

1 point

(d) What type of error might you have made with your conclusion? Type I or Type II?

Type II

2. Let X be exponentially distributed with parameter λ. The probability distribution function is

$$f(x, \lambda) = \lambda e^{-\lambda x}$$

(a) Find the maximum likelihood estimator of λ based on a random sample of size n. Show your work.

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

the log likelihood is $\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$

$$\frac{d \ln L(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

4 points

set this equal to zero to get $\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{X}}$

(b) Find the Fisher information $I(\lambda)$ of parameter λ. Show your work.

$$I(\lambda) = E\left[-\frac{d^2 \ln L(\lambda)}{d^2\lambda}\right]$$

3 points

so $\frac{d}{d\lambda}\left(\frac{n}{\lambda} - \sum x_i\right) = -\frac{n}{\lambda^2}$

the $E\left[-\left(-\frac{n}{\lambda^2}\right)\right] = \frac{n}{\lambda^2} = I(\lambda)$

w/ regards to x

A large sized coffee at Starbucks is called a *venti*, meaning *twenty* in Italian, because a a venti cup is supposed to hold 20 ounces of coffee. Jonathan believes that his venti always has less 20 ounces of coffee in it, so he wants to test Starbucks' claim. He randomly chooses 50 Starbucks locations and gets of cup of venti coffee from each one. He then measures the amount of coffee in each cup and finds that the mean amount of coffee is 19.08 ounces and the standard deviation is 0.42 ounces. Below are some possibly useful R-commands. Use this information to answer questions 1 - 3.

```
qt( 0.95, df = 49) = 1.676551

qt( 0.90, df = 49) = 1.299069

qnorm( 0.90) = 1.281552

qnorm( 0.95) = 1.644854
```

$$19.08 \pm 1.676551 \frac{0.42}{\sqrt{50}}$$

1. Construct a 90% confidence interval for the true mean amount of coffee in Starbucks venti cups.

   (a) (18.53, 19.63)
   (b) (18.98, 19.18)
   (c) (19.00, 19.16)
   (d) (18.38, 19.78)

2. Do you think Starbucks' claim that a venti cup has 20 ounces of coffee is reasonable based on the confidence interval you have just calculated?

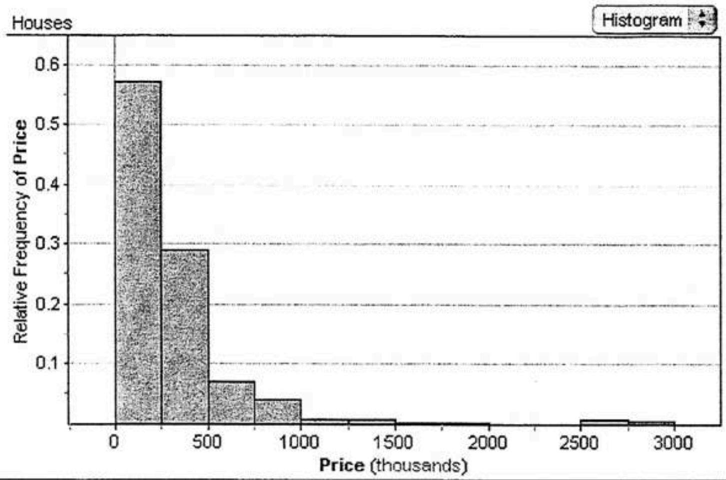
   (a) Based on the confidence interval, Jonathan should be relieved that Starbucks is giving him 20 ounces of coffee in his venti cup.
   (b) Based on the confidence interval, Jonathan should be angry at Starbucks and demand more coffee. Or he should go to Coffee Bean.

3. Which of the following actions could you take to make your confidence interval **more** narrow than the confidence interval you just constructed?

   I. Increase sample size. ✓
   II. Decrease sample size.
   III. Increase confidence level.
   IV. Decrease confidence level. ✓

   (a) I or II
   (b) I or III
   (c) I, II, or III
   (d) I or IV
   (e) I, II, III, or IV

4. The relative frequency histogram below shows the distribution of prices of all houses in a data set. Which of the following is the most appropriate measure of center and spread?



(a) Mean and standard deviation.

(b) Median and IQR.

(c) Mean and IQR.

(d) Median and standard deviation.

The nutrition label on a bag of potato chips says that a one ounce serving of potato chips has 130 calories and contains ten grams of fat, with three grams of saturated fat. A random sample of 55 bags yielded a sample mean of 138 calories with a standard deviation of 17 calories. Is there evidence that the nutrition label does not provide an accurate measure of calories in the bags of potato chips? Use $\alpha = 0.05$. Use this to answer questions 5 - 8.

5. Is this a one or two-sided hypothesis test?

(a) One

(b) Two

6. What is the test statistic you compute?

(a) t = 3.49

(b) t = - 3.49

(c) z = 3.49

(d) z = - 3.49

7. Given the following R code what is your p-value?

(a) `pt( 3.49, df = 54, lower.tail = FALSE )*2 = 0.0009700528`

(b) `pt( - 3.49, df = 54 )*2 = 0.0009700528`

(c) `pnorm(3.49, lower.tail = FALSE )*2 = 0.0004830205`

(d) `pnorm( - 3.49 )*2 = 0.0004830205`

8. What can you conclude about the measure of calories in the bags of potato chips?

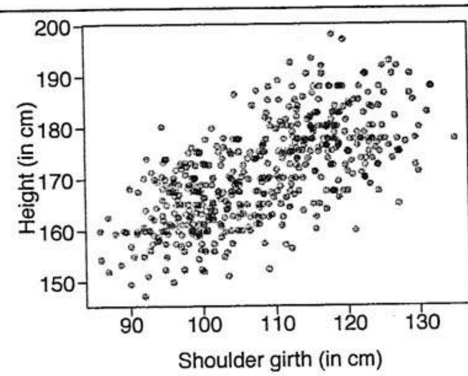
(a) Since our p-value is smaller than α we fail to reject the null hypothesis. There is no evidence that the nutrition label does not provide an accurate measure of calories.

(b) Since our p-value is smaller than α we reject the null hypothesis. There is evidence that the nutrition label does not provide an accurate measure of calories.

(c) Since our p-value is larger than α we fail to reject the null hypothesis. There is no evidence that the nutrition label does not provide an accurate measure of calories.

(d) Since our p-value is larger than α we reject to reject the null hypothesis. There is evidence that the nutrition label does not provide an accurate measure of calories.

Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender for 507 physically active individuals. The mean shoulder girth is 108.20 cm with a standard deviation of 10.37 cm. The mean height is 171.14 cm with a standard deviation of 9.41 cm. The correlation between height and shoulder girth is 0.666. Use this to answer questions 9 - 11.



9. Which of the below is the correct equation of the regression line for predicting height from shoulder girth?

   (a) $\widehat{shoulderGirth} = 105.79 + 0.604 * height$
   (b) $\widehat{height} = 91.72 + 0.734 * shoulderGirth$
   (c) $\widehat{shoulderGirth} = 91.72 + 0.734 * height$
   (d) $\widehat{height} = 0.604 + 105.79 * shoulderGirth$
   (e) $\widehat{height} = 105.79 + 0.604 * shoulderGirth$

10. What does the slope tell us in this context?

   (a) On average what we would expect height to be when shoulder girth is 0 centimeters.
   (b) On average what we would expect shoulder girth to be when height is 0 centimeters.
   (c) On average how many centimeters we would expect height to increase for each centimeter increase in shoulder girth.
   (d) When shoulder girth increases by one centimeter how many centimeters of an increase it causes in shoulder girth.
   (e) On average how many centimeters we would expect shoulder girth to increase for each centimeter decrease in height.

11. What is the residual for a randomly selected person with a shoulder girth of 100 cm and height of 180cm?

   (a) 10.00 cm
   (b) -10.00 cm
   (c) 13.81 cm
   (d) -13.81 cm

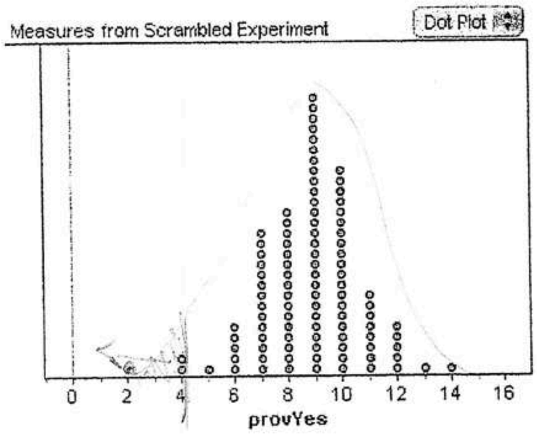
$\hat{y} = 166.19$

$y - \hat{y} = 180 - 166.19$

A "social experiment" conducted by a TV program questioned what people do when they see a very obviously bruised woman getting picked on by her boyfriend. On two different occasions at the same restaurant the same couple was depicted, where in one scenario the woman was dressed "provocatively" and in the second scenario the woman was dressed "conservatively". The contingency table below shows how many restaurant diners were present under each scenario, and whether or not they intervened. This is the observed data.

	Provocative	Conservative	Total
Intervened	4	16	20
Did Not Intervene	16	9	25
Total	20	25	45

We use a bootstrap simulation to test if dressing provocatively decreases the count of people who intervene. We write "yes" on 20 index cards and "no" on 25 index cards to indicate whether or not a diner (represented by each card) intervened. We next shuffle the cards and deal them into two groups of size 20 and 25, the provocative and conservative scenarios, respectively. Then we count and record many diners intervened under the provocative scenario (denoted as **provYes**). We repeat this 100 times. The dot plot below shows the distribution of **provYes** in these 100 simulations. We will use $\alpha = 0.05$ in this test. Use this information to answer questions 12 and 13.



$\frac{2}{100} = .02$

12. Based on the simulation results what is the estimated p-value for the observed number of diners intervening under the provocative scenario.

(a) 4

(b) 0.04

(c) 0.02

(d) Can't compute from this simulation.

13. Comparing the observed results to the simulation results which of the choices below best completes the following sentence?

The simulation results suggest that whether or not people intervene is ________ how the woman is dressed.

(a) dependent on

(b) independent of

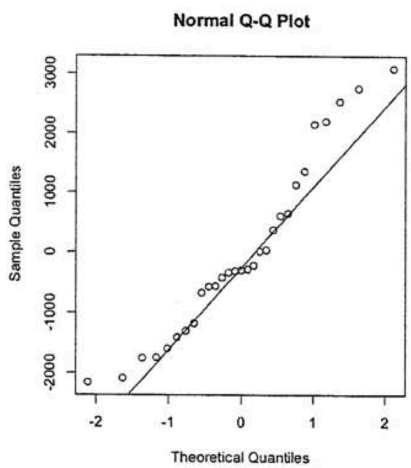
Below is the R regression output, Normal Q-Q plot, and residual plot for the relationship between the number of high school graduates the previous spring and the enrollment at the local university. Use this to answer questions 14 and 15.

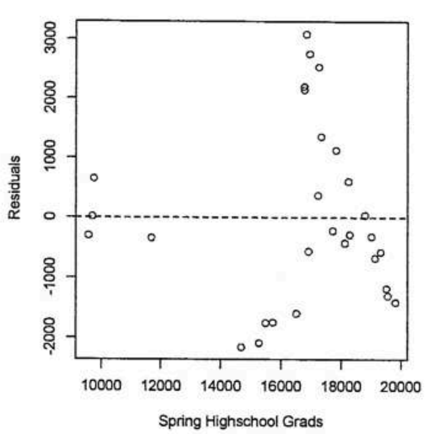
```
Call:
lm(formula = income$ROLL ~ income$HGRAD)

Residuals:
    Min      1Q  Median      3Q     Max
-2156.8 -1175.5  -301.0   649.8  3089.0

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.653e+03  1.635e+03  -2.235   0.0339 *
income$HGRAD  9.898e-01  9.743e-02  10.159 1.01e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1509 on 27 degrees of freedom
Multiple R-squared: 0.7926, Adjusted R-squared: 0.7849
F-statistic: 103.2 on 1 and 27 DF,  p-value: 1.012e-10
```





14. What is the Pearson correlation coefficient (r) for the relationship between the number of spring high school graduates and fall enrollment numbers?

(a) 0.7926
(b) 0.8903
(c) 0.7849
(d) 0.6282

15. Does a simple linear regression model adequately describe the relationship between the number of spring high school graduates and fall enrollment numbers?

(a) Yes, the regression results show a significant slope with p-value well below an α of 0.05
(b) Yes, the R^2 value of 89.03% shows the x-variable describing a large proportion of the variation seen in the y-variable.
(c) No, the residual plot of spring high school graduates versus residuals shows smaller residual value with smaller variability for smaller values of x.
(d) Cannot tell from the plots given. We must see the scatter plot of the x- and y-variables.