STAT303 Sec 507-510
Fall 2011
Exam #3
Form A

Instructor: Julie Hagen Carroll

Name: _____

1. Please PRINT your name in the blank above.
2. **Do not open this exam until you are told to do so.**
3. There are 20 multiple-choice questions on this exam, each worth the same. Please mark your answers **clearly** on a GRAY Scantron sheet. Multiple marks will be counted wrong.
4. You **must mark** your Scantron form with
   (a) Your NAME and UID.
   (b) Your correct SECTION (Thursday 11:10 is 507, 12:45 is 508, 2:20 is 509, and 3:55 is 510).
   (c) This test FORM (A, B, C, or D).
5. You will have 50 minutes to finish this exam.
6. You may use
   (a) Two $8\frac{1}{2} \times 11$ formula sheet (both sides) of your own making.
   (b) A copy of the standard normal table.
   (c) The categorical data handout and the bivariate data recap handout.
   (d) A stand-alone calculator, i.e., one that cannot communicate with the internet or anything outside itself.
7. You must put all possessions besides the materials listed under your table out of everyone else's way. This includes cell phones, which must be turned off.
8. If you have questions, please write out what you are thinking on the back of this test page so that we can discuss it after your results are returned to you.
9. If you are caught cheating or helping someone to cheat on this exam, you both will receive a grade of **zero** on the exam. You must work alone.
10. When you are finished please make sure you have marked your Section and Form and have an answer for every question. Place your scantron upside down on your exam and put your sheets inside the exam. It is a good idea to put your name on your sheets, too. Please sit in the same spot next week so that you can get this back.
11. I will not be back in town until Monday, but I hope to have grades posted before class.
12. Good luck!

1. The goal of a least squares regression line is

   A. to connect all the data points in a scatterplot.
   B. to connect all the residuals to the regression line.
   C. to minimize the sum of observed values of $\hat{y}$.
   D. to minimize the sum of squared residuals.
   E. none of the above.

2. Students more likely to 'drive American'. In one sample of 193 student drivers, 111 drove an American car. Is this sufficient evidence to support the claim? What are the test statistic value and p-value? Use at least 4 decimals in your calculations.

   A. $z = 2.09$ and p-value$= 0.0366$
   B. $z = 2.09$ and p-value$= 0.0183$
   C. $z = 2.11$ and p-value$= 0.0174$
   D. $z = 2.11$ and p-value$= 0.0348$
   E. $z = 0.58$ and p-value$= 0.2810$

3. Suppose the p-value in the previous problem was 0.46, how should we interpret this?

   A. We will get 46% of students with American cars when that actual proportion is more than 50%.
   B. 46% of the time we will get a student with an American car when only 50% of them actually have an American car.
   C. 46% of the time we will get at least this large of a proportion of students with American cars when the true proportion is not more than 50%.
   D. 46% of the time we will get less than 50% of the students with American cars when we should get more than 50%.
   E. We will get this proportion of students with American cars or more when we should get less than 46%.

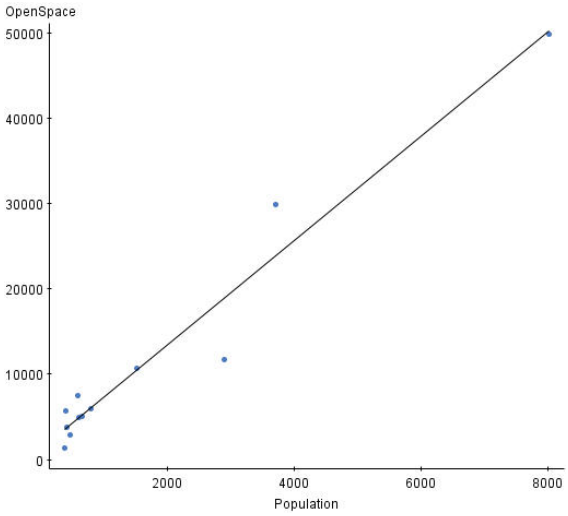
```
Hypothesis test results:
p1 : proportion with Hep-C where Tattoo=ComParlor
p2 : proportion Hep-C where Tattoo=Elsewhere
p1 - p2 : difference in proportions
H0 : p1 - p2 = 0
HA : p1 - p2 > 0
Ct1  Tot1  Ct2  Tot2  S Diff.   SE    Z   P-value
 17    52    8    61   0.196  0.078  2.5   0.0062
```

4. "Are tattoo parlors really risky?" Two samples of tattooed people were compared to see if Hep-C was more prevalent in patrons of tattoo parlors. Which of the following is the correct conclusion based on the output above?

   A. Tattoo parlors cause Hep-C.
   B. One is more likely to get Hep-C from a tattoo parlor.
   C. If one is getting a tattoo, not going to a tattoo parlor is a safer choice.
   D. The sample proportion isn't large enough for this to be a valid test.
   E. No one should get a tattoo.

5. Which of the following is correct for the set of hypotheses above?

   A. A Type I error would be claiming tattoo parlors are risky when in reality they're not so you should use an $\alpha = 0.01$
   B. A Type II error would be failing to prove tattoo parlors are risky when they really are so you should use an $\alpha = 0.01$.
   C. A Type I error would be claiming tattoo parlors are ok when in reality they're not so you should use an $\alpha = 0.01$
   D. A Type II error would be failing to prove tattoo parlors are risky when they really are so you should use an $\alpha = 0.10$.
   E. None of the above are correct.



6. If the point (8000, 50000), the top right point, was removed what would happen to the statistics?

   A. Nothing, the point is on the line.
   B. The correlation would be weaker.
   C. The equation of the line would be different.
   D. The correlation would be stronger.
   E. Two of the above are correct.

7. Which of the following is true about the scatterplot data in the previous problem?

   A. There is a strong linear relationship between *population* and *open space*.
   B. The points (8000, 50000), (3000, 12000), (3800, 30000) should all be removed because they are outliers.
   C. (3000, 12000) and (3800, 30000) are influential points.
   D. The Simple Linear Regression equation is not a good representation of this data.
   E. Two of the above are true.

	Staff	Student	Total
American	104	107	211
Asian	47	54	101
European	11	32	43
Total	162	193	355

8. The table above is comparing the country of origin of cars for staff and students. How likely is a student to have a European car?

   A. 32/43
   B. 32/193
   C. 32/355
   D. 193/355
   E. 43/355

9. How likely is an American car owning staffer?

   A. 104/162
   B. 104/211
   C. 104/355
   D. 162/355
   E. 211/355

10. If country of origin and driver are independent, what's the expected count for a student with an Asian car?

    A. 54/101
    B. 54/193
    C. 54/355
    D. 101/355 * 193/355
    E. 101*193/355

```
Chi-Square test:
 Statistic DF Value P-value
Chi-square    8.1386  0.0171
```

11. The χ^2 test for independence is given above. What is the correct conclusion?

    A. At the 1% level, we cannot say that the relationship between country of origin and driver is linear.
    B. At the 5% level, we can conclude that the relationship between country of origin and driver is linear.
    C. At the 1% level, we cannot say that country of origin and driver are independent.
    D. At the 1% level, we can say that country of origin and driver are independent.
    E. None of the above are correct conclusions.

12. A technical school claims that 90% of its graduates can pass a standardized test to identify simple engine problems. An automobile with one such problem was presented to 120 graduates. Only 90 of the 120 mechanics who examined the car correctly identified the problem. Does this support the school's claim? Which set of hypotheses will answer this question?

    A. $H_0 : \pi \leq 0.9$ vs. $H_A : \pi > 0.9$
    B. $H_0 : \pi = 0.9$ vs. $H_A : \pi \neq 0.9$
    C. $H_0 : \pi \leq 90/120$ vs. $H_A : \pi > 90/120$
    D. $H_0 : \pi = 90/120$ vs. $H_A : \pi \neq 90/120$
    E. We don't need to run a test because 90 students passed.

13. If I wanted to know if the true proportion of 'good' mechanics is 75% but all I was given was a 95% confidence interval, (0.321, 0.679), what should I conclude?

    A. At the 5% level, the true proportion of 'good' mechanics is 50%.
    B. At the 5% level, the true proportion of 'good' mechanics is not 75%.
    C. At the 1% level, the true proportion of 'good' mechanics is not 75%.
    D. All of the above are correct.
    E. Two of the above are correct.

14. A tally of 1939 winning and losing basketball scores produced the linear equation: $WinScore = 16.2 + 0.9 * LoseScore$. Which of the following is true?

    A. In general, the winning team was about 16 points ahead.
    B. For each point the loser scored, the winner also scored a point (rounding).
    C. If the loser didn't score at all, the winner should have about 16 points.
    D. All of the above are true.
    E. Seriously, 1939?

15. What is the residual for a 40/60 game using the previous equation (assuming it is within the range of the original scores)?

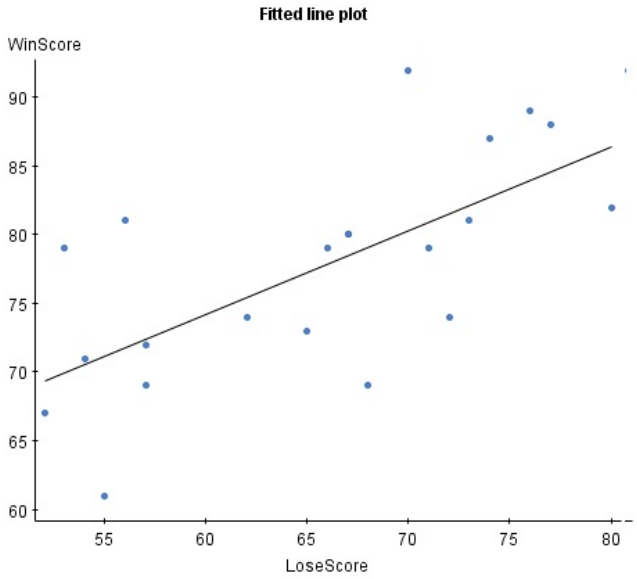
    A. 52.2
    B. 54
    C. 70.2
    D. -10.2
    E. 7.8

16. Define π as the true population proportion and p as the sample proportion from the population. A correct interpretation of a 95% confidence interval for π is

    A. "95 out of every 100 confidence intervals will contain π."
    B. Prior to collecting the data, the probability that the interval will contain π is 0.95.
    C. If many different random samples are taken, and a confidence interval for π is constructed from each sample, then about 95% of all these intervals will contain the point estimate p.
    D. $p \pm 0.95$ is very likely to contain π.
    E. The upper and lower endpoints of the interval are within 5% of π.

17. If $p_{50} \sim N(0.8, 0.057^2)$, how likely are we to get a proportion above 90%? (Note: $n = 50$.)

    A. 1.75
    B. 0.9599
    C. 0.0401
    D. 0.0802
    E. Since we don't know π, we don't know that $n\pi$ and $n(1-\pi)$ are greater than or equal to 10, so we can't use the normal approximation for p_{50}.



18. Describe the correlation between *Losescore* and *Winscore*.

    A. Strongly negative
    B. Moderately negative
    C. Weak
    D. Moderately positive
    E. Strongly positive

19. It was reported that SAT and family income have a correlation of $r = 0.89$. Which of the following conclusions are appropriate?

    A. Students from wealthier families have higher SAT scores than students from lower income families.
    B. There is a strong positive association between SAT and family income.
    C. The more money your family has the higher your SAT score will be.
    D. All of the above are correct.
    E. Only two of the above are correct.

20. Using the information below, what is the correct range of the p-value if I wanted to test $H_0 : \pi_1 = \pi_2$ vs. $H_A : \pi_1 \neq \pi_2$?

```
90% (0.286, 0.514)
95% (0.264, 0.536)
99% (0.222, 0.578)
```

    A. p-value > 0.10
    B. $0.10 > p$-value > 0.05
    C. $0.05 > p$-value > 0.01
    D. $0.01 > p$-value
    E. There is no hypothesized value.

1D,2B,3C,4D,5D,6B,7D,8B,9C,10E,11E,
12A,13B,14D,15E,16B,17C,18D,19E,20D