Exam #3 STA-209 Sections 03, 05

Name:_____

Directions

- Several questions have a *suggested* number of sentences for your answer. This is to help indicate the scope of solution I am looking for (i.e., you do not always need every single detail) and to discourage you from "information dumping"

- Information that is included with your answer that is not relevant to the problem will not help you but *will still be graded for correctness.* In other words, including more information than is asked for can generally only hurt you

- You **do not** need to write in complete sentences: bullet points are completely acceptable and even preferred

Question 1

Below I will describe a number of different study designs. You will be asked to

(1) Identify the null hypothesis
(2) Identify the correct statistical test for this hypothesis

Part A: For this study, I am interested in determining if a student's major (Humanities/STEM/Social Sciences) is associated with their final exam score in STA-209.

Part B: A two-day workshop for learning basic R has been created, where attendees are tested in their R skills both prior to the workshop and after the workshop has been completed. For each of these tests, a numeric score is given. We wish to determine whether or not the workshop has been effective in improving the R skill of the attendees.

Part C: Binge drinking is defined as a pattern of drinking that involves consuming 5 or more standard drinks within 2 hours. Respondents of a survey were asked for their sex and whether or not they have engaged in binge drinking more than twice in the previous week. We wish to determine whether or not there is a difference in binge drinking patterns between men and women.

Question 2

Cocaine addicts have been reported to have a significant depletion of stimulating neurotransmitters and thus continue to use cocaine to avoid feelings of depression and anxiety. A 3-year study with 72 chronic cocaine users compared an antidepressant drug called desipramine with lithium and a placebo (lithium is the standard treatment for cocaine addiction). One third of the subjects were randomly assigned to each treatment group with the following results:

	Relapse	No Relapse
Desipramine	10	14
Lithium	18	6
Placebo	20	4

Part A: What type of plot would you use to visually display these results

Part B: Describe the null hypothesis of this study and construct a table of expected counts under the assumption of the null hypothesis.

Part C: The χ^2 test statistic has a value of $\chi^2 = 10.5$ with a p-value of $p\text{-val} = 0.0052$. Based on this, what conclusion would you reach if testing at the $\alpha = 0.05$ level?

Question 3
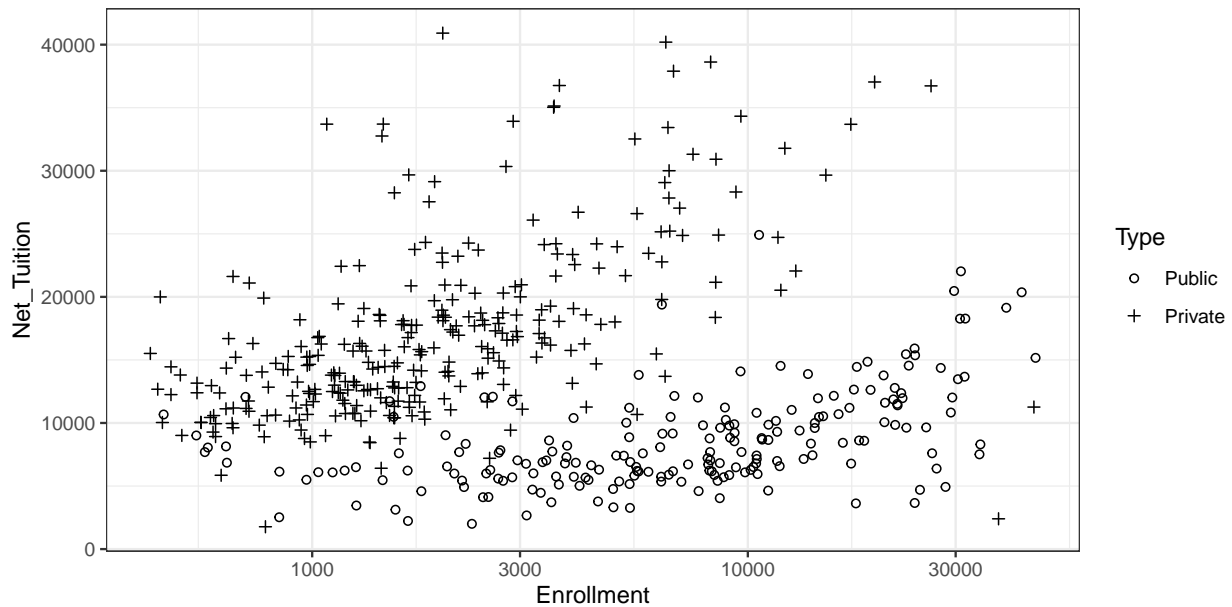
Model 1:

```
lm(formula = Net_Tuition ~ Enrollment, data = college)

Coefficients:
             Estimate Std. Error t value           Pr(>|t|)
(Intercept) 14225.3137   272.8034    52.1 <0.0000000000000002 ***
Enrollment     -0.0820     0.0265    -3.1               0.002 **

Residual standard error: 7180 on 1093 degrees of freedom
Multiple R-squared:  0.00869,   Adjusted R-squared:  0.00779
F-statistic: 9.58 on 1 and 1093 DF,  p-value: 0.00201
```

Model 2:

```
lm(formula = Net_Tuition ~ Enrollment + Type, data = college)

Coefficients:
             Estimate Std. Error t value           Pr(>|t|)
(Intercept)  5746.1019   377.2481    15.2 <0.0000000000000002 ***
Enrollment      0.2533     0.0239    10.6 <0.0000000000000002 ***
TypePrivate 10808.5970   398.6370    27.1 <0.0000000000000002 ***

Residual standard error: 5550 on 1092 degrees of freedom
Multiple R-squared:  0.408, Adjusted R-squared:  0.406
F-statistic:  376 on 2 and 1092 DF,  p-value: <0.0000000000000002
```

Part A: For this part, consider **Model 1** from above. What is the null hypothesis in linear regression? Based on the summary output, how would you describe the relationship between enrollment and tuition?

Part B: Now consider **Model 2**, which includes an indicator for whether or not a college is private. How would you interpret the intercept in this model? Is this a meaningful value in this model?

Part C: Compare the coefficient for Enrollment between **Model 1** and **Model 2**. What has changed? In other words, what impact has adding an indicator for Private had on this value, and why did it result in such a drastic change?