Non-Programming Stats Homework Questions

1. You have a data set from some population with two variables. One variable is your response and one is a group variable taking on either A, B, or C. You found the sample means for each group, say $\bar{y}_A$, $\bar{y}_B$, and $\bar{y}_C$. The sample means differ, can you conclude the population differ with respect to their mean? Why or why not?

2. What are the two major methods for conducting statistical inference? What are their goals?

3. Two six sided dice are rolled, and the sum of the face values is six. What is the probability that at least one of the dice came up a three? Show your work.

4. A drawer of socks contains seven black socks, eight blue socks, and nine green socks. Two socks are chosen in the dark.

   a. What is the probability that they match?

   b. What is the probability that a black pair is chosen

5. The first three digits of a university telephone exchange are 452. If all the sequences of the remaining four digits are equally likely, what is the probability that a randomly selected university phone number contains seven distinct digits?

6. We noted the for a Binomial random variable we could rewrite it as the sum of random variables taking on 1 and 0. Formally, the RV

   $$X = \begin{cases} 1 & \text{if success} \\ 0 & \text{if failure} \end{cases}$$

   is called a Bernoulli variable, $X \sim Ber(p)$, where $p = P(success)$. Show that the mean and variance of X are p and p(1-p), respectively. Hint this is very easy!

7. What are the differences between a PMF and a PDF?

8. Roughly explain what the central limit theorem tells us **and why it is useful**.

9. If the variance of a population is σ^2 and we take a random sample, the variance of the sample average is σ^2/n. Give an arguement as to why the variation in the average of n things less than that of the population.

10. Why don't we ever use a 100% confidence interval?

11. Is the p-value for a one-sided test bigger or smaller than that of the corresponding two-sided test? Why?

12. Why don't we accept the null hypothesis?

13. In the notes we said that for a random sample from an exponential distribution with rate parameter λ the MLE was $1/\bar{Y}$. Show this is the case using calculus (maximize the log likelihood given in the notes!).

14. Why do we care about doing multiple comparisons? Is there a requirement to adjust confidence intervals or only to adjust hypothesis tests?

15. Describe the Bonferroni multiple comparison adjustment (you'll need to look this up).

16. Which of the following are true statements about the p-value?

    a. The p-value is the probability the null hypothesis is correct.

    b. If the p-value is small it indicates there is a small probability that the null hypothesis is true.

    c. The p-value is the probability the alternative hypothesis is correct.

    d. If the p-value is small it indicates the data is unlikely assuming the null hypothesis is true.

17. Suppose a state congressman, collects data about support for a new tax and calculates a 95% confidence interval to be (0.48, 0.55). Which statements below are true?

    a. If we take many random samples, 95% of the resulting confidence intervals would contain the true proportion of voters who are in favor of the new tax.

    b. If we take many random samples, 95% of the resulting confidence intervals would contain the sample proportion of voters who are in favor of the new tax.

    c. We are 95% confident that the sample proportion of voters who are in favor of the new tax is between 48% and 55%.

    d. We are 95% confident that the true proportion of voters who are in favor of the new tax is between 48% and 55%.

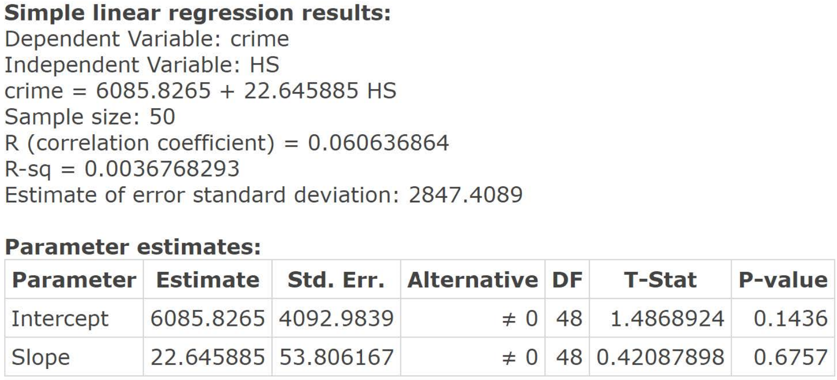
18. A manufacturing plant employs workers in 3 different shifts: Morning, Evening, and Night. In order to study the effect of shift time on work-related accidents, the manager of this plant wishes to test the hypothesis that the probability an accident occurs during the night shift is twice as high as for the morning or evening shifts. She uses a type of hypothesis test, called a Chi-square test, with the following null and alternative:

    $$H_0 : p_m = 0.25, p_b = 0.25, p_N = 0.5$$

    $$H_A : \text{At least one of the probabilities specified in the null is incorrect}$$

    She collects data on a random sample of 160 work-related accidents at the plant and calculates a test statistic of 6.53 and p-value of 0.038. What can we conclude based on the results of this test? (You may assume that the test run is valid.)

19. A criminologist studying the relationship between education and crime rate in medium-sized US cities collects data on a random sample of 50 counties.



Simple linear regression results:

```
Dependent Variable: crime
Independent Variable: HS
crime = 6085.8265 + 22.645885 HS
Sample size: 50
R (correlation coefficient) = 0.060636864
R-sq = 0.0036768293
Estimate of error standard deviation: 2847.4089
```

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-value
Intercept	6085.8265	4092.9839	≠ 0	48	1.4868924	0.1436
Slope	22.645885	53.806167	≠ 0	48	0.42087898	0.6757

He uses linear regression to study the relationship between the percentage of individuals in the county who have at least a high school diploma (X) to the crime rate per 100,000 residents (Y). The criminologist performed the linear regression analysis and found the results above.

a. What should the criminologist's null and alternative hypotheses be to assess the linear relationship between the percentage of individuals in the county who have at least a high school diploma and the crime rate?

b. The criminologist states that crime rate has a strong linear relationship with education level because the slope is large ($\hat{\beta}_1$=22.646). Is the criminologist's interpretation of the regression analysis correct? Why or why not?

20. A turkey egg incubator needs to be kept at a mean of 37.5 degrees Celsius to work create the correct conditions for hatching. A new incubator has been created and is considered as a replacement. Readings on 36 incubators are recorded and yield a mean temperature of 38.2 degrees Celsius. A 90% CI for the mean temperature comes out to be (37.4, 40.0). What is your conclusion about using this new incubator and why?