MAS 250 – FINAL

December 14, 2023 Time allowed: 180 minutes

Name: ______________________

Honor Code: I will be academically honest in all of my academic work and will not tolerate academic dishonesty of others.

Pledge: *I have neither given nor received any unauthorized aid during this exam.*

Instructions: Show all your work on your solutions. You will NOT receive credit if you do not justify your answers.

Good luck!

t-values:

- $t_{0.005,7} = 3.4995$, $t_{0.01,7} = 2.9980$
- $t_{0.005,8} = 3.3554$, $t_{0.01,8} = 2.8965$
- $t_{0.025,26} = 2.0555$, $t_{0.05,26} = 1.7056$
- $t_{0.025,27} = 2.0518$, $t_{0.05,27} = 1.7033$

χ^2-values:

- $\chi^2_{0.05,1} = 3.8415$, $\chi^2_{0.025,1} = 5.0239$
- $\chi^2_{0.05,2} = 5.9915$, $\chi^2_{0.025,2} = 7.3778$

1. (4 points) In order to determine how many bike racks a university needs to provide, they want to estimate the proportion of students that ride their bike to classes. They are reasonably certain that the percentage is around 25% but they would like to obtain a more scientifically sound estimate. How many students would you recommend they sample in order to be 98% confident the sample proportion is within 3% of the true population? Justify the reason for your recommendation.

2. (7 points) A researcher wants to test the average weight of air-freight packages, $H_0 : \mu = 80$ versus $H_1 : \mu > 80$. Assume that the weight follows a normal distribution with the standard deviation 3.2.

   (a) (4 points) Suppose that a sample of size 100 was collected. What is the type II error at $\mu = 81$ (alternative) and $\alpha = 0.05$?

   (b) (3 points) To achieve at least 95% of power at $\mu = 81$, how large does a random sample need to be? Use $\alpha = 0.05$.

3. (16 points) In the Spacelab Life Sciences 2 payload, 14 male rats were sent to space. Upon their return, the red blood cell mass (in milliliters) of the rats was determined. A control group of 14 male rats was held under the same conditions (except for space flight) as the space rats and their red blood cell mass was also determined when the space rats returned. Draw a random sample of 28 rats and randomly assign them to the flight group (x) or the control group (y). Calculate the sample means, $\bar{x} = 7.881$ and $\bar{y} = 8.430$, standard deviations, $s_x = 1.017$ and $s_y = 1.005$. Test the claim that the flight rats have a different mean red blood cell mass from the control rats at the $\alpha = 0.05$ significance level.

   (a) (3 points) State the null and alternative hypotheses for the problem.

   (b) (4 points) Test the hypotheses the null and alternative hypotheses in (a). Make sure to state your conclusion and interpretation in the context of the problem. (You can make necessary assumptions on the populations.)

   (c) (3 points) What assumptions did you make for answering (b)?

   (d) (3 points) Construct a 95% two-sided confidence interval for the difference of the two population means.

   (e) (3 points) What is the relationship between (b) and (d)?

4. (15 points) An experiment was conducted to examine the potential toxicity of red dye No. 40. Rats were randomly assigned to receive no Red 40 (the control group), a low dosage of Red 40, a medium dosage of Red 40, or a high dosage of Red 40. The time until death (in weeks) was measured. The goal of the experiment was to determine whether the mean times until death by consuming Red 40 were different. Data summaries are shown in the table. The overall sample mean $x_{..} = 75.55$.

	ctrl	low	med	high
Mean	91.36	69.89	71.50	65.25
Standard Deviation	11.01	11.86	23.77	28.06
Number Obs.	11	9	10	8

(a) (3 points) Write an appropriate statistical model. Specify the factor and response variable.

(b) (6 points) A partial ANOVA table is shown below. For parts (i)-(vi), fill in the missing entries.

Table 1: ANOVA table for Rat data.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Between	(i)	(iii)	(v)	(vi)	0.0246
Within	(ii)	(iv)	380.42		
Total	37	16984.72			

(c) (3 points) Based on the p-value shown in the ANOVA table what do you conclude? Use a significance level of $\alpha = 0.05$. Make sure to state the hypotheses and your conclusion in the context of the experiment.

(d) (3 points) 95% Tukey C.I.s for all possible pairwise comparisons of the group means are shown in the R output. Based on these intervals and/or the adjusted P-values, for which groups are the mean times until death significantly different at $\alpha = 0.05$?

```
Tukey multiple comparisons of means
95% family-wise confidence level

$group
     diff        lwr       upr     p adj
low-ctrl  -21.474747 -45.15437  2.204879 0.0868094
med-ctrl  -19.863636 -42.88286  3.155592 0.1109953
high-ctrl -26.113636 -50.59372 -1.633551 0.0328848
med-low     1.611111 -22.59544 25.817667 0.9978960
high-low   -4.638889 -30.23864 20.960858 0.9609302
high-med   -6.250000 -31.24014 18.740141 0.9056532
```

5. (6 points) To determine if chocolate milk was as effective as other carbohydrate replacement drinks, nine male cyclists performed an intense workout followed by a drink and a rest period. At the end of the rest period, each cyclist performed an endurance trial in which he exercised until exhausted and time to exhaustion was measured. Each cyclist completed the entire regimen on two different days. On one day the drink provided was chocolate milk and on the other day the drink provided was a carbohydrate replacement drink. The times (in minutes) to exhaustion appear in the table.

Cyclist	1	2	3	4	5	6	7	8	9
Chocolate Milk	24.85	50.09	38.30	26.11	36.54	26.14	36.13	47.35	35.08
Carbohydrate	10.02	29.96	37.40	15.52	9.11	21.58	31.23	22.04	17.02

(a) (3 points) Is there evidence that the mean time to exhaustion is greater after chocolate milk than after carbohydrate replacement drink? Interpret the result in the context at $\alpha = 0.01$. Let W =Chocolate Milk-Carbohydrate Replacement. You can use $\bar{w} = 14.0789$ (sample mean of W) and $s_W = 9.4746$ (sample standard deviation of W).

(b) (3 points) Construct a 99% two-sided confidence interval for the difference between two population means. What is your interpretation about the interval?

6. (10 points) The press release referenced in the previous exercise also included data from independent surveys of teenage drivers and parents of teenage drivers. In response to a question asking if they approved of laws banning the use of cell phones and testing while driving, 74% of the teens surveyed and 95% of the parents surveyed said they approved. Suppose that 600 teens and 400 parents of teens responded to the surveys and that it is reasonable to regard these sample as representative of the two populations. Do the data provide convincing evidence that the proportion of teens that approve of cell-phone and texting bans while driving is less than the proportion of parents of teens who approve?

(a) (3 points) State the null and alternative hypotheses.

(b) (4 points) Test the relevant hypotheses at $\alpha = 0.05$ by calculating the p-value and interpret the results.

(c) (3 points) Construct a two-sided 95% confidence interval for the different between the two proportions.

7. (24 points) Suppose that gas mileage (Y) and horsepower (x) are measured for 392 vehicles, and the following summary statistics are obtained.

$$\begin{array}{rll} n = 392 & \sum x = 409.52 & \sum y = 91.91 \\ S_{xx} = 57.93 & S_{xy} = -9.14 & S_{yy} = 2.38 \end{array}$$

You can use the normal approximation to answer the following questions.

(a) (3 points) Write the statistical model for a simple linear regression (using parameters not statistics) and describe the assumptions for the error term.

(b) (4 points) Calculate the sample correlation coefficient and interpret the result. Conduct a test for $H_0 : \rho = 0$ and $H_1 : \rho \neq 0$ at the significance level 0.05.

(c) (3 points) Find the least squares line.

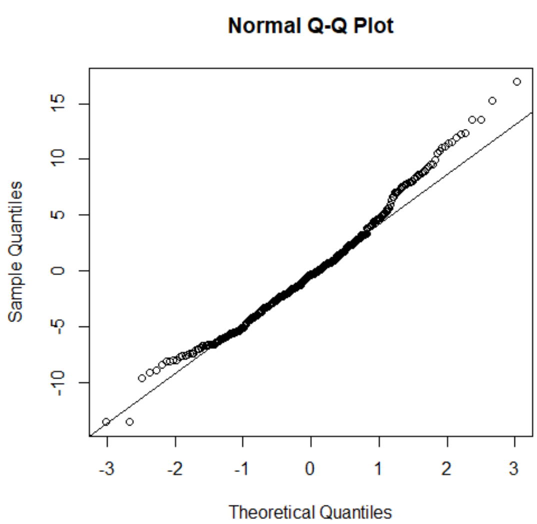
(d) (3 points) Estimate the standard deviation of the error term, $\hat{\sigma}$.

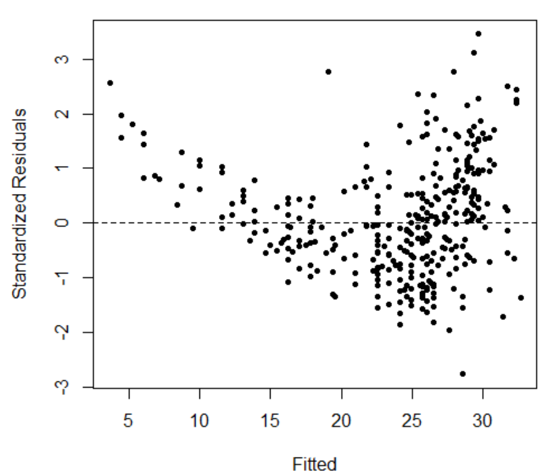
(e) (4 points) Is there sufficient evidence that mpg is linearly related to horsepower? Answer this question by testing the slope at two sides using the significance level 0.05. Specify the null and alternative hypotheses and draw your conclusion. What is the relationship with the test in (b)?

(f) (3 points) Calculate the coefficient of determination, and interpret the number. Relate the result with the sample correlation coefficient in (b).

(g) **(Extra 3 points)** Predict mpg for horsepower=1 and find a two-sided 95% confidence interval for $E(Y)$.

(h) (4 points) Interpret the two plots and check the assumptions of the regression model.





8. (10 points) An investigation was conducted to see whether exposure of women to atomic fallout affects the rate of birth defects. A sample of 500 children from mothers who were exposed to the atomic fallout from the atomic explosion at Hiroshima were assessed for birth defects. 400 children from a Japanese island far from Hiroshima were used as a control group. The count data below were obtained. Do these data give strong evidence of an effect of the mother's exposure on the frequency of birth defects?

	Birth defects present	Birth defects absent	Total
Mother exposed	84 (i)	416 (429.44)	500
Mother not exposed	43 (56.44)	357 (343.56)	400
Total	127	773	900

(a) (3 points) State the null and alternative hypotheses.

(b) (3 points) The values in the parentheses are the expected frequency under the null hypothesis. Calculate the expected frequency for (i).

(c) (4 points) Perform a proper test at $\alpha = 0.05$ and interpret the result.

9. (8 points) Let $X_1, \ldots, X_n$ denote independent and identically distributed random variables from the following distribution with parameter θ. Find the MLE for θ for each distribution.

(a) (4 points)

$$f(x|\theta) = \theta a^\theta x^{-(\theta+1)}, \quad x \geq a,$$

where a is a positive known constant.

(b) (4 points)

$$f(x|\theta) = e^{-(x-\theta)}, \quad x \geq \theta.$$

(c) **(Extra 3 points)** Calculate the bias of the MLE in (b).