STAT 305 D Final Exam

Show all your work.

1. (20 points) Find the following:

   a. (4 points) $P((Z-1, Z+1.5)$ contains $0)$, $Z \sim N(0,1)$
   b. (4 points) $P((X-5, X+5)$ contains $8)$, $X \sim N(5,4)$
   c. (4 points) $P((X-2, X+2)$ contains $0)$, $X \sim N(1,0.4)$
   d. (4 points) $P((X-2, X+2)$ contains $0)$, $X \sim N(0,4))$
   e. (4 points) $P((X-\sigma, X+\sigma)$ contains $\mu)$, $X \sim N(\mu, 4\sigma^2)$

2. (20 points) Every year, your eccentric neighbor carves an excessive number of pumpkins for Halloween and leaves them outside his front door through November. Being eccentric yourself, you weigh all the pumpkins and record their combined weight each year. Here are the data from the last six years:

Year	2007	2008	2009	2010	2011	2012
Weight (lb)	311.0	271.1	320.7	274.3	332.4	304.4

   a. (10 points) Conduct a hypothesis test at $\alpha = 0.05$ to find out if the true mean combined weight of your neighbor's pumpkins is different from 240 lb.
   b. (10 points) Attempt to convince your neighbor of his absurdity by constructing and interpreting a lower 95% confidence interval for the true mean pumpkin weight.

3. (20 points) Not to be outdone, you have been carving and displaying absurd levels of pumpkin matter this entire time. Here are *your* combined pumpkin weights for the last six years.

Year	2007	2008	2009	2010	2011	2012
Weight (lb)	517.7	568.0	541.8	567.1	518.1	592.1

   a. (10 points) In a typical year, do you exceed your neighbor in pumpkin weight? Conduct the appropriate hypothesis at $\alpha = 0.05$ to find out.
   b. (10 points) Construct and interpret a 2-sided 95% confidence interval for the true mean difference in pumpkin mass between the two of you.

4. (20 points) Now that you have bought and carved too many pumpkins, you have too many pumpkin seeds. You also have your neighbor's seeds because he is too eccentric for any kind of seed. You have thousands and thousands of seeds. Naturally, you decide to measure the length of each seed individually, painstakingly, carefully, painstakingly, and individually. Pain! Your 100 seeds (sample 1) have a mean length of 1.11 cm and a standard deviation of 1.00 cm. Your neighbor's 200 seeds (sample 2) have a mean length of 1.23 cm and a standard deviation of 1.00 cm.

a. (10 points) Use the appropriate hypothesis test at $\alpha = 0.01$ to test the claim that, on average, your seeds are a different length than your neighbor's.

b. (10 points) Compute and interpret a two-sided 99% confidence interval for the difference in true mean pumpkin seed lengths (yours - your neighbor's).

5. (20 points) Return to the New York rivers data:

	Name	X1	X2	X3	X4	Y
1	Olean	26	63	1.20	0.29	1.10
2	Cassadaga	29	57	0.70	0.09	1.01
3	Oatka	54	26	1.80	0.58	1.90
4	Neversink	2	84	1.90	1.98	1.00
5	Hackensack	3	27	29.40	3.11	1.99
6	Wappinger	19	61	3.40	0.56	1.42
7	Fishkill	16	60	5.60	1.11	2.04
8	Honeoye	40	43	1.30	0.24	1.65
9	Susquehanna	28	62	1.10	0.15	1.01
10	Chenango	26	60	0.90	0.23	1.21
11	Tioughnioga	26	53	0.90	0.18	1.33
12	West_Canada	15	75	0.70	0.16	0.75
13	East_Canada	6	84	0.50	0.12	0.73
14	Saranac	3	81	0.80	0.35	0.80
15	Ausable	2	89	0.70	0.35	0.76
16	Black	6	82	0.50	0.15	0.87
17	Schoharie	22	70	0.90	0.22	0.80
18	Raquette	4	75	0.40	0.18	0.87
19	Oswegatchie	21	56	0.50	0.13	0.66
20	Cohocton	40	49	1.10	0.13	1.25

Remember:

- Y is the mean nitrogen content (mg/liter) of the river based on samples based on regular intervals taken in the spring, summer, and fall months.
- $X1$ is the percentage of surrounding land used in agriculture.
- $X2$ is the % surrounding forested land.
- $X3$ is the % surrounding residential land.
- $X4$ is the % surrounding commercial/industrial land.

I fit the model:

$$Y_i = \beta_0 + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \varepsilon_i$$

And here is part of the JMP output:

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.1193202	0.282635	7.50	<.0001*
X2	-0.016004	0.004117	-3.89	0.0012*
X3	0.0162203	0.011483	1.41	0.1758

a. (4 points) State the assumptions on the ε_i's that we need in order to do inference.

b. (4 points) Does % forested land affect the nitrogen content in the rivers? Conduct the appropriate hypothesis test at $\alpha = 0.05$ to find out.

c. (4 points) Test $H_0 : \beta_3 = 0.01$ vs. $H_a : \beta_3 > 0.01$ at $\alpha = 0.01$.

d. (4 points) Construct and interpret a 2-sided 95% confidence interval for the intercept in the model.

e. (4 points) What does the model intercept represent? Is there anything problematic in your interpretation on a practical level?

6. EXTRA CREDIT (10 points). Suppose I fit a different model with all of X1 through X4:

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.7222135	1.234082	1.40	0.1832
X1	0.0058091	0.015034	0.39	0.7046
X2	-0.012968	0.013931	-0.93	0.3667
X3	-0.007227	0.03383	-0.21	0.8337
X4	0.3050278	0.163817	1.86	0.0823

Why is the coefficient on X2 no longer significantly different from zero?