Stat 185 - Final Review

Our final exam is this coming Friday, April 29 at 8:00 AM in our regular classroom. Here are some problems to help get you ready.

1. (Like problem 1 off of Exam I) In an effort to understand what factors impact GPA, we surveyed 5 UNCA students and asked about their GPA, number of hours they study at night, number of nights they go out, and their gender. The resulting data table is shown in table 1.

Table 1: GPA among UNCA students

GPA	**StudyPerWeek**	**SleepPerNight**	**OutPerWeek**	**Gender**
3.9	15	6	1	female
3.75	15	7	1	female
3.6	10	6	4	male
4.0	16	8	1	female
2.9	6	6	4	male

(a) What type of study is this - Observational or Experimental? (Circle one)

(b) What are the cases in the study? What feature in the table is associated with the cases?

(c) What are the variables in the study? What feature in the table is associated with the variables?

(d) Identify one variable that is definitely categorical.

(e) Identify one variable that is definitely numerical and continuous.

(f) Identify one variable that appears to be numerical and discrete, as represented in the table.

(g) (New material) Use the axes shown in figure 1 to draw a scatter plot relating GPA to StudyPerWeek, together with an approximate regression line.

(h) (New material) Suppose I tell you that an equation for the regression line is

$$y = 0.969x + 2.429.$$

What GPA is predicted for a student who studies 12 hours per week?

2. (An extension of problem 5 from Exam I to incorporate material involving the t-distribution.) In this problem, we'll consider the sample numeric data

$$\{1, 1, 2, 3, 5\}.$$

   (a) Write down a computation showing that the mean of the data is $\bar{x} = 12/5 = 2.4$.

   (b) Write down a computation showing that the sample standard deviation of the data is $s = \sqrt{14/5} \approx 1.673$.

   (c) Find the standard error of the data.

   (d) Use a t-distribution to write down a 95% confidence interval for the data.

3. (A fundamental problem) The weights of Lilliputians can be described via an L-distribution with mean 0.22 and standard deviation 0.1, as shown in figure 2 (a) Suppose we pick a simple random sample of 100 Lilliputians, compute the average weight of all those Lilliputians in the sample, and call that resulting random variable S.

   (a) What are the mean and standard deviation of S?

   (b) Which of the distributions shown in figure 2 might be the distribution of S?

   (c) Estimate $P(S < 0.21)$.

   (d) What properties of the random sample allow you to use a normal distribution in part (c)?

4. (A fundamental problem) Suppose we randomly select 500 college students and measure their heights in feet. We find that our data has an average of 5.8 with a standard deviation of 0.32. We wish to write down a 95% confidence interval for this data.

   (a) Find the standard error associated with this sample.

   (b) Use the rules of thumb for the normal distribution to find the z^* value that corresponds to a 95% confidence interval.

   (c) Write down a 95% confidence interval for the average height of college students based on this data.

5. (A fundamental problem) Supposedly, approximately 11% of the population is left handed, but we think it might be higher than that. Suppose that in a random sample of 100 people, we find 15 left handers. Let's use this data to explore the question of whether the 10% estimate is truly correct vs whether there might be more than 10% at the 95% level of confidence.

   (a) Write down the Null and Alternative Hypotheses for this problem.

   (b) Compute the standard error, test statistic, and p-value.

   (c) State the conclusion of the hypothesis test and your reasons why.

6. (Number 7 off of Exam II) A random sample of 66 bottlenose dolphins found their average length to be 9.2 feet with a standard deviation of 1.7 feet. Use that information to write down a 98% confidence interval for the length of bottlenose dolphins. *Note*: You'll need to use a normal table to find the z^*-multiplier.

7. (Number 8 off of Exam II) A recent survey of 2238 likely voters put President Biden's approval rating at 44%. Use that data to write down a 96% confidence interval for Biden's approval rating.
   Note: You'll need to use a normal table to find the z^*-multiplier.

8. (Number 4 off of Exam III) A random sample of 10 college texts compared their local bookstore price to their amazon.com price. The data indicate that, on average, the local price minus the amazon price was \$21.12 with standard deviation of \$31.8. We are curious if the amazon price is generally *less* than the book store price.
   (a) Write down the hypothesis test.
   (b) Compute the standard error for the problem and test statistic.
   (c) What is the conclusion of the test and what is your supporting evidence?

9. (Number 5 off of Exam III) In a sample of 100 Republicans, 34 were for issue 1 and 66 against. In a sample of 150 Democrats, 62 were for issue 1 and 88 against. Let's use this data to explore whether there is a genuine difference in the views of Democrats and Republicans on issue 1 to a 95% level of confidence.
   (a) Compute the observed proportions $\hat{p}_R$ and $\hat{p}_D$, as well as the difference
   $$\hat{p} = \hat{p}_D - \hat{p}_R.$$
   (b) Compute the standard error.
   (c) Compute the test statistic.
   (d) Use a normal table to compute the p-value.
   (e) State the conclusion of the test.
   (f) Why was it OK to use a normal table?

10. Match the scatter plots shown in figure 3 with their correlation from below:
    - -0.7
    - 0.9
    - 0.04
    - -0.9

11. (Regression) Figure 4 shows a scatter plot for a random sample of 1000 men who competed in the 2012 Boston Marathon. The x-coordinate of each point corresponds to the runner's age and the y-coordinate corresponds to the runners time in minutes. The regression line for the data is also shown and has formula $y = 1.5x + 186$.
    a. What time does this regression model predict for a 57 year old runner?
    b. Which of the following could be a reasonable value for the correlation between age and time: 0.8, 0.04, -0.3, or -0.8?
    c. Suppose I run a linear regression test on this data and a p-value of 0.00000015. Can I conclude at the 99% level of conficence that there is a linear relationship between age and speed?

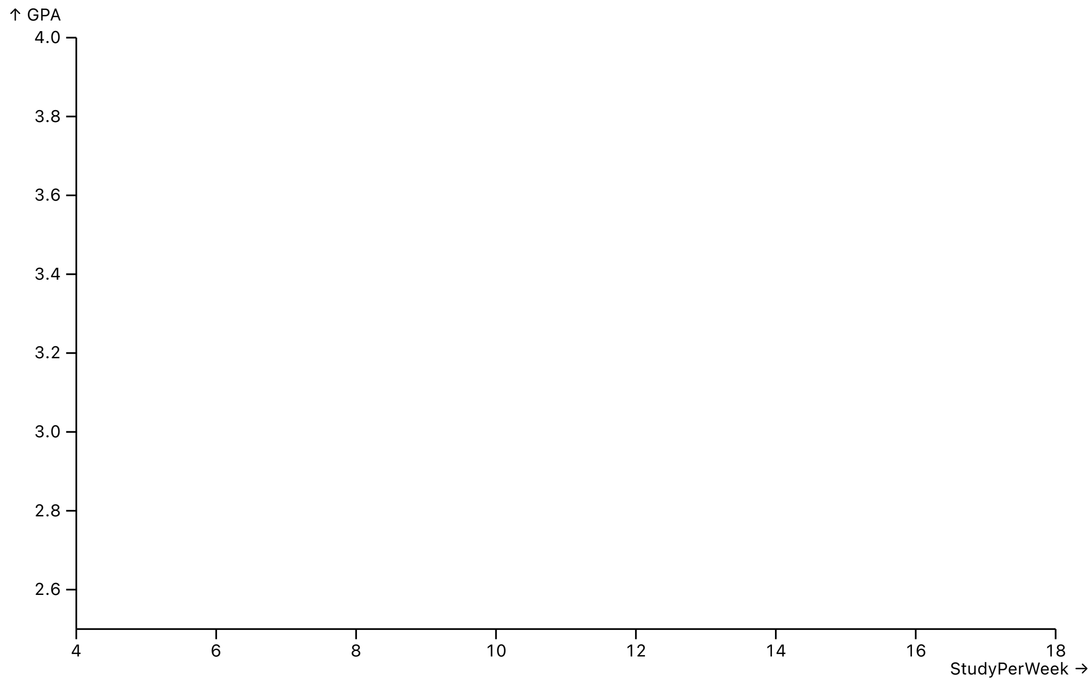

Figure 1: axes

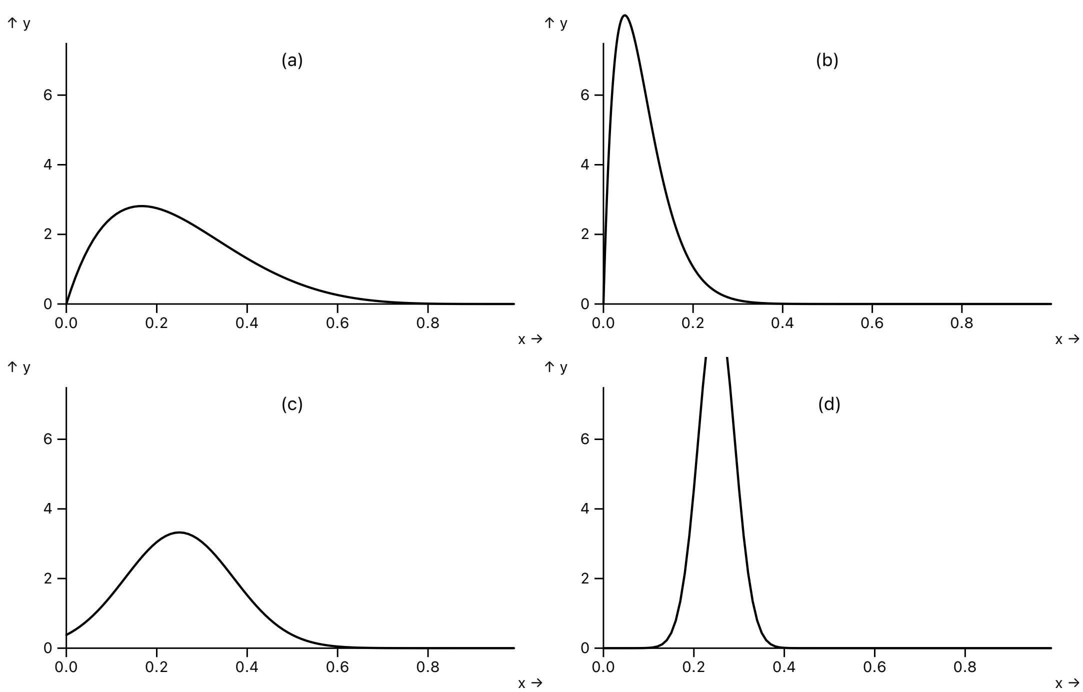

Figure 2: Potential distributions

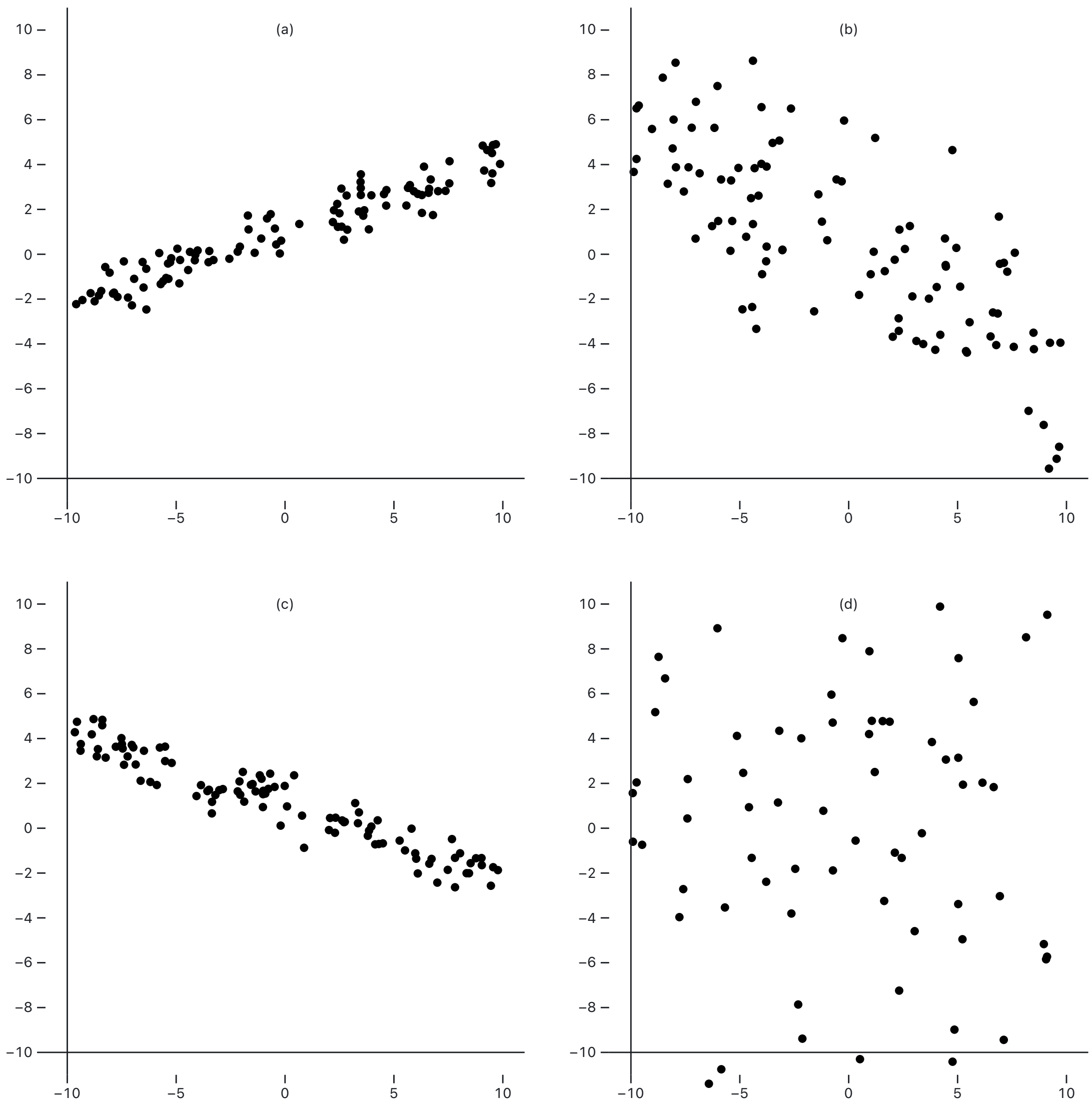

Figure 3: Several scatter plots

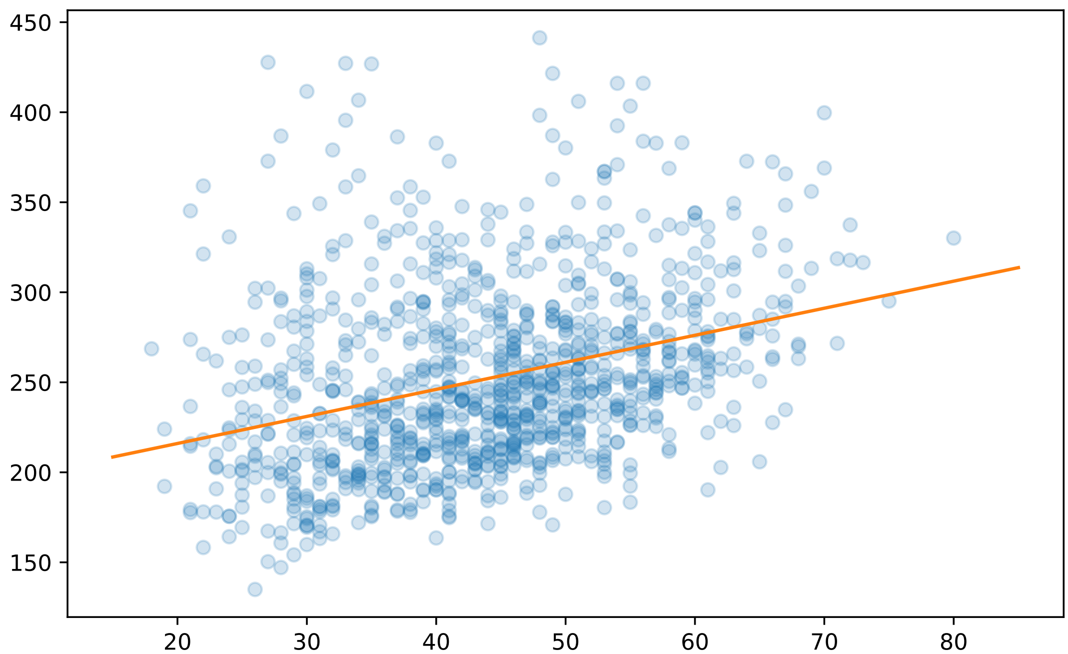

Figure 4: Linear regression for age and marathon time