Answer Key

Show your work wherever some is required. Do not assume I will know what you did.
You may use a graphing calculator where appropriate and 3"x5" card.

---

1. The following data reflect the number of citations certain medical facilities have received in the last two years.

0 0 1 1 1 1 1 2 2 2 2 3 3 3 3 3 3 4 4 4 4 5

(a) What are the mean and median of this dataset?

$$\overline{X} = 2.3\overline{6} \text{ citations} \qquad x_{.5} = 2.5 \text{ citations}$$

(b) What is the standard deviation for this dataset?

$$s = 1.398824491 \text{ citations}$$

(c) What is the range for this dataset?

$$\text{range } = 5 - 0 = 5 \text{ citations}$$

(d) Find $x_{.3}$ for this dataset?

$$.3 \cdot 22 = 6.6 \to 7 \Rightarrow x_{.3} = 1 \text{ citation}$$

(e) Find the deviation for the first observation (0).

$$0 - 2.3\overline{6} = -2.3\overline{6} \text{ citations}$$

(f) Find the variance.

$$s^2 = 1.956709957 \text{ citations}^2$$

2. In which of the main branches of statistics would we learn about sampling techniques and how to obtain useful data?

   Experimental Design

3. A psychologist has become concerned about post traumatic stress disorder (PTSD) in military veterans. While he has observed people from various occupations that have PTSD, he feels that it is much more common among veterans. To study this, he randomly selects 20 veterans from his patient records and records whether each of the patients has PTSD. Is this example an observational study or an experiment? Justify your choice.

   Observational study, since there is no control. He is simply observing his patients.

4. A banking oversight committee in Virginia has begun to hear complaints from minority populations that they are being discriminated against unfairly when applying for home mortgages. They claim that they are being turned down despite having higher credit ratings than whites who have been offered mortgages. The committee would like your help in deciding whether to launch a formal investigation into these claims. Discrimination is a serious charge, and would be bad publicity for falsely accused banks. However, allowing discrimination is ethically untenable. Assume investigations are not automatically launched, and can be costly and time consuming. Answer the following questions about this setting.

   (a) In this setting, what would be the null hypothesis?

   Should not launch an investigation

   (b) What would be the alternative hypothesis?

   Should launch an investigation

   (c) Describe in plain English what a Type I Error would entail here, and what the consequences would be.

   Decide that we should launch, but we shouldn't have $\longrightarrow$ False accusation, wasted money, but cleared up

   (d) Describe in plain English what a Type II Error would entail here, and the consequences.

   We decide that we should not launch an investigation, but we should $\longrightarrow$ any discrimination continues and we could have stopped it

   (e) In light of your work here, which choice would you consider to have the more serious risks? Briefly explain your reasoning.

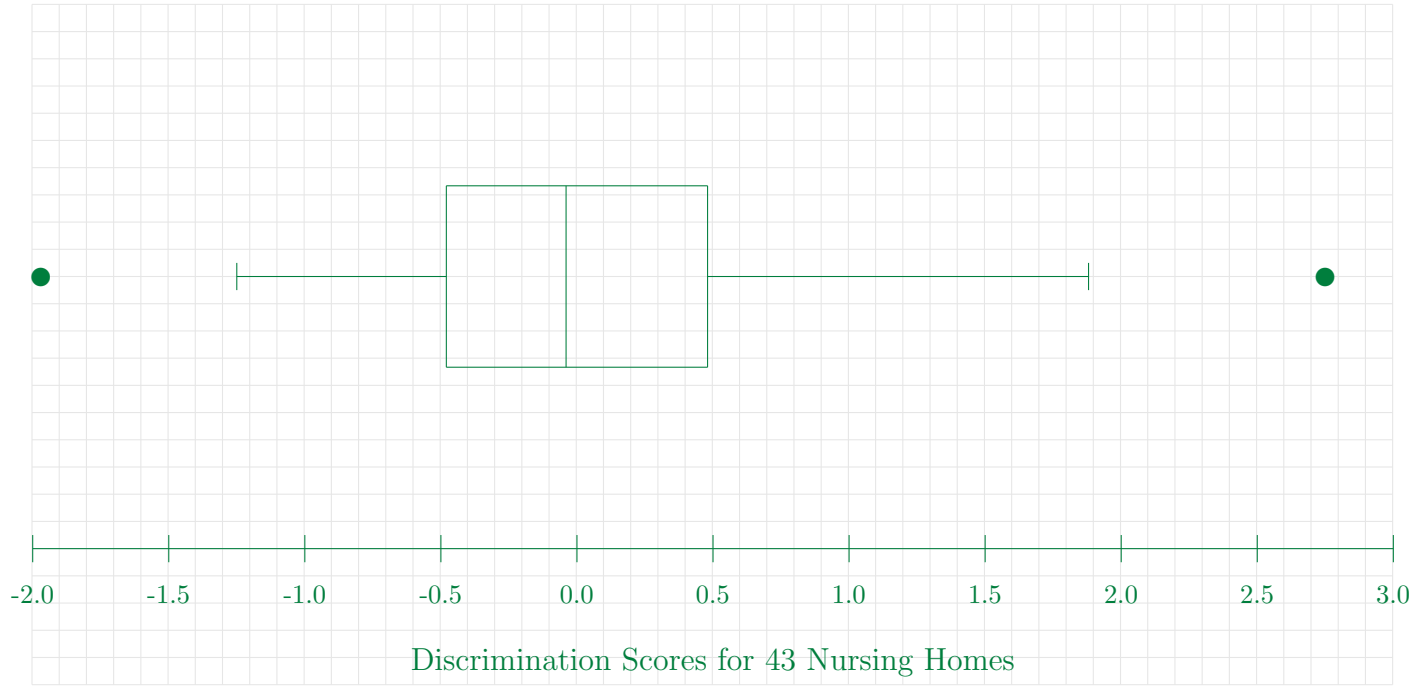
   Type II error is worse since money is not as important as the ethics of not ending discrimination if it exists

5. A nursing home in northeastern Pennsylvania is claiming that they are being unfairly targeted for fines due to discrimination. They commission a study to investigate the strength of their claim. A statistician constructs a model to predict the number of fines based on several measures of good care, and then gives each home of interest a "discrimination score" based on this model. A negative score indicates that the home was actually fined less often than the model predicted, while a positive score would indicate that a higher number of fines than predicted were received. The data that follow are the discrimination scores (sorted) for each of the 43 homes for which the model was able to make a prediction. Construct a Boxplot for this data, showing all work in the space provided. then use the graphpaper provided to make your plot. Finally, do you think your plot could be used by one or more of the homes to claim they were cited unfairly?

-1.97	-1.25	-1.03	-0.99	-0.82	-0.81	-0.79	-0.79	-0.58	-0.48	-0.48	-0.46	-0.45	-0.42	-0.31
-0.28	-0.24	-0.17	-0.17	-0.16	-0.13	-0.04	-0.03	0.13	0.15	0.15	0.18	0.19	0.25	0.29
0.45	0.46	0.48	0.52	0.56	0.66	0.97	1.24	1.28	1.37	1.65	1.88	2.75		

1. $x_{.5} = x_{(22)} = -0.04 \qquad x_{.25} = x_{(11)} = -0.48 \qquad x_{.75} = x_{(33)} = 0.48$
2. $iqr = 0.48 - (-0.48) = 0.96$
3. $L = 1.5 \cdot 0.96 = 1.44$
4. $F_1 = -0.48 - 1.44 = -1.92 \qquad F_3 = 0.48 + 1.44 = 1.92$
5. $a_1 = -1.25 \qquad a_3 = 1.88$

Based on the plot below, it looks like the home which is an outlier on the right side (score of 2.75) has a case for discrimination since they are an outlier and were cited more often than expected (hence a positive score). The outlier on the left side got "lucky" and was cited less than expected. Therefore this home should not complain.



Discrimination Scores for 43 Nursing Homes