## Chapter 1 Review Questions — Guth Statistics

1. A student is gathering information about students at her college for a statistics project. She asks students in each of her classes about how they feel about the Federal Government forcing insurers to provide birth control to all women that they insure. 85% of the students she asked feel that this is appropriate.
   *a.* In what population is she interested?
   > The population is students at the college.
   *b.* What is her sample?
   > The sample is students in her classes.
   *c.* Is the data she gathered quantitative or categorical?
   > Feelings about whether a program is appropriate are categorical.
   *d.* In what parameter is she interested? If she knows its value, what is it?
   > The proportion of all students at the college that feel the program is appropriate is the parameter, and this is unknown.
   *e.* In what statistic is she interested? If she knows its value, what is it?
   > The percent of students in her sample that feel the program is appropriate is a statistic. Its value is 85%.
   *f.* What method of sampling is she using?
   > She is using a convenience sample.
   *g.* Are her methods likely to introduce bias in her sample? If so, what biases?
   > There is a clear selection bias because she is only picking students in her classes. Also, the language used talks about the Federal Government *forcing* insurers to provide birth control. This is strong language that may evoke an emotional *response*, causing response bias.
   *h.* Is this study observational or experimental?
   > It is observational because there is no explanatory variable that is being manipulated.
   *i.* Would it be appropriate to use the results of this study to make an inference about the entire college?
   > The biased convenience sample cannot represent the entire population.
2. In a recent study researchers randomly gathered a collection volunteers and put them into two groups based on whether or not they drink diet soda. One group consisted of people who drink at least one can of diet soda each day. The other group consisted of people who do not drink diet soda. Each member was weighed, and then they were weighed again 1 month later. The non-diet soda drinkers had no change in weight. The diet soda drinkers gained 1 pound each, on average.
   *a.* Is this study experimental or observational?
   > It is observational, because the explanatory variable is not being manipulated.
   *b.* What is the explanatory variable? Is it quantitative or categorical?
   > The explanatory variable is whether or not people drink diet soda. This is not a quantity, so the values are categorical.
   *c.* What is the response variable? Is it quantitative or categorical?
   > The response variable is the increase in weight. This is quantitative.
   *d.* Are there confounding variables? If so, what might they be?
   > Because the groups in this study are not created to be similar – people are assigning themselves based on soda drinking preference. It may be that the diet soda drinkers are more inclined to gain weight, with or without the soda. This may be why they choose to drink diet soda. This inclination to gain weight is a confounding variable.
   *e.* Was a placebo implemented?
   > No.
   *f.* Which group is the control group?
   > This is not really an experiment, but the group which did not drink diet soda is would be a control group.

*g.* Which group is the treatment group?

> Again, this is not an experiment, but the group that is essentially a treatment group is the diet soda drinking group.

*h.* Explain whether it would be appropriate to conclude that diet soda causes members to gain weight? If not, what logical error has been made?.

> Because the groups are not created to be similar, no cause and effect conclusion can be made. The conclusion is erroneous because *correlation* (where one variable can be used to help predict another) does not imply *cause* (where we determine that a change in one variable causes a change in another).

*i.* Is the sampling method likely to create bias in the data? Explain.

> The sampling method was random, so there is *no selection bias*. There is *no reason* to think that there is *measurement* or *response bias*, since they are asking a simple question about diet soda, and then measuring weights. No information is given about non-response, so there is *no evidence of non-response bias*.

3. In the *Abecedarian* project, inner city children, 3 years of age, were randomly selected. Their parents were all from the same socio-economic background. Each child was assigned randomly to one of two groups. In one group, children were provided with high quality preschool education. In the other group, children received childcare without preschool learning. The children in both groups were provided with healthy food supplements. The students were followed for years, and many differences were noted. Among these, it was found that the children who had quality preschool were 4 times more likely to graduate from a four-year college (23% vs. 6%).

*a.* Is this study experimental or observational?

> It is an experiment, because the preschool variable (*yes* or *no*) was manipulated by the researchers.

*b.* What is the explanatory variable? Is it quantitative or categorical?

> The preschool variable (*yes* or *no*) is a categorical explanatory variable.

*c.* What is the response variable? Is it quantitative or categorical?

> The college variable (*yes* or *no*) is a categorical response variable.

*d.* Was blinding introduced into the study? How might this have affected the results?

> Blinding was not used, because the children knew that they were or were not receiving preschool education (in the sense that they lived through it). This lack of blinding could introduce a mindset that might influence the positive results for the preschool children – but note that *this was a desired response*.

*e.* Are there confounding variables? If so, what might they be?

> Students had similar backgrounds, and the nutritional supplements make their diets similar. Additionally, students were divided into groups through random assignment.

*f.* Which group is the control group?

> The control group included the children who did not receive preschool education.

*g.* Which group is the treatment group?

> The treatment group included the children who received preschool education.

*h.* Would it be appropriate to conclude that quality preschool causes children to be more likely to graduate from college? Explain.

> This sample can only represent the population of inner-city children from this socio-economic group. But among that population, a cause and effect relationship *can* be concluded.

## Chapter 2 Review Questions

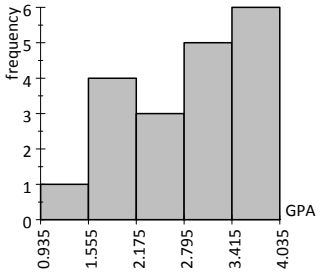
For questions #25 through #34, please refer to the sample data below. The following are GPA's of randomly selected students. The values are sorted from low to high.

{0.94, 1.69, 1.77, 1.91, 2.14, 2.26, 2.58, 2.64, 2.83, 2.83, 3.01, 3.13, 3.25, 3.42, 3.62, 3.74, 3.80, 3.81, 4.00}

1. Construct a frequency table and histogram, with the first lower class limit equal to 0.94. Use 5 classes.

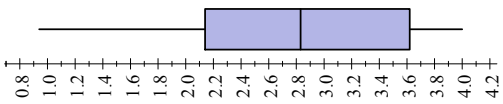
   The class width is (4.00 – 0.94)/5 = 0.612 ≈ 0.62 (this gets rounded *up*).
   The frequency table is below.

| Class | Tally | Frequency |
|---|---|---|
| 0.94 – 1.55 | I | 1 |
| 1.56 – 2.17 | IIII | 4 |
| 2.18 – 2.79 | III | 3 |
| 2.80 – 3.41 | ~~IIII~~ | 5 |
| 3.42 – 4.03 | ~~IIII~~ I | 6 |




2. Sketch a boxplot, noting any outliers. Show your work.

   The 5-point summary is: {*min*, $Q_1$, *median*, $Q_3$, *max*} = {0.94, 2.14, 2.83, 3.62, 4.00}. The lower fence for outliers is $Q_1 - 1.5 \cdot IQR = -0.08$. The upper fence is $Q_3 + 1.5 \cdot IQR = 5.84$. There are no outliers.




3. Without computations, describe the center, shape, and spread of the distribution. How does the shape affect the mean in relation to the median?

   The center of the distribution is around 2.8. The distribution's shape is skewed to the left. The spread could be measured by the range, max – min = 4.00 – 0.94 = 3.06. The left skew causes the mean to be less than the median.

4. Compute the sample mean (no need to show work).

   The sample mean is: $\bar{x} = 2.809$.

5. Compute the sample standard deviation and variance (don't show work).

   The sample standard deviation is: $s = 0.853$. The sample variance is: $s^2 = 0.728$.

6. Compute the usual range, and note any values that are *outside* the usual range (show your work)?

   The usual range is $\bar{x} - 2s = 1.103$ to $\bar{x} + 2s = 4.515$. Only 0.94 is outside of this range.

7. Compute the *z*-score of 0.94. Would this score be unusually low? Explain. What does the *sign* (+ or –) of the *z*-score tell us about the value (*x*) in relation to the mean?

   $z = \frac{x - \bar{x}}{s} = \frac{0.94 - 2.809}{0.853} \approx -2.19$. This is more than 2 standard deviations below the mean. It is unusual. The *z*-score is negative, indicating that the value is below the mean.

8. Compute the percentile of 3.80.

   The percentile is: $\frac{16}{19} \cdot 100 \approx 84^{th}$.

9. Compute the median.

   The median is: $\tilde{x} = 2.83$.

10. Compute the mode.

    The mode is 2.83. This value appears twice in the data set.

11. Compute the midrange.

    The midrange is: $\frac{0.94 + 4.00}{2} \approx 2.47$.

12. Compute the range.

    The range is 4.00 – 0.94 = 3.06.

13. Compute the lower and upper quartiles, and the inter-quartile range.

    The first quartile is: $Q_1 = 2.14$. The third quartile is $Q_3 = 3.62$. The IQR is $Q_3 - Q_1 = 3.62 - 2.14 = 1.48$.