1) (1 point) Estimate the mean GNP. Right answers in range 5500 – 7000. Use the diamonds from the boxplot to get the mean.

2) (1 point) The SD is closest to (circle one): 5000 (Actual SD = 8000. The typical deviation on the left is about 6000, and it is somewhat larger on the right, so that 5000 is the closest answer.)

3) (1 point) Estimate the percentage of countries with GNP greater than 10000. Right answers in range 15% – 24%. Use the fact that 25% of the data is to the right of about 8000 (the 75th percentile on the box plot).

4) (1 point) Estimate the percentage of countries with GNP less than the mean. Right answers in range 63% – 68%. Use the box plot and notice that the mean is more than half-way between the median and 75th percentile, but not too close to the 75th percentile.

5) (1 point) b) The average and standard deviation will decrease.

6) (2 points) Estimate the SD of Log(GNP). 2 points for 1.5 to 1.8, 1 point for 1.3 to 1.5 or 1.81 to 2.15. An answer of 1.8 implies that roughly 95% of the data fall between 4 and 11 (since these data are roughly normally distributed), which exceeds the range of the data. So, answers bigger than 1.8 are too high. The typical deviation from the average is a little more than 1.5 to the left of the average and around 1.7 or so to the right of the average. So, a good guess at the SD is around 1.65.

7) (1 point) Use a normal curve to estimate the 75th percentile of Log(GNP). Show work.

If you used your mean and SD from the previous two questions, and used a z-score of approximately 0.69 (you want the z-score that corresponds to 25% in each tail, which means 50% in the middle), you got full credit,

8) (1 point) Use a normal curve to estimate the percentage of countries with Log(GNP) less than 6.6. Show work.

If you used the mean and SD from the previous parts, and calculated correctly the area to the left of the z-score, you got full credit.

8) (1 point) Which region has most of the countries with high Log(GNP)s? Region 3

9) (2 points) Which region has the largest standard deviation of Log(GNP). Region 5

10) (1 points) Which region has the highest percentage of countries with Log(GNP)s above 7.5: Region 1

11) 1 point each for not checking the first two options.

(2 points) XXX South Asian countries and African countries for the most part have similar values of Log(GNP).

(2 points) XXX More than 50% of South Asian countries have Log(GNP)s less than the 25th percentile of Log(GNP) for South and Central American countries. (Note that the 25th percentile is the end of the box, not the end of the whiskers.)
12) (2 points) Here is an estimated regression line: \( \text{outcome} = -0.52 + 0.12 \text{ predictor} \). Which predictor variable corresponds to this regression line? Life expectancy. The slope is 0.12, and the only scatter plot showing a positive relationship is with life expectancy.

13) (3 points) Here is an estimated regression line: \( \text{outcome} = 9.49 - 0.18 \text{ predictor} \). Which predictor variable corresponds to this regression line? Death Rate. If you plug in some values of death rate, you can get predicted values of Log(GNP). The regression with predicted values close to the actual Log(GNP)s is the one that uses Death Rate.

14) (2 points) Estimate the correlation between Log(GNP) and the predictor you used in Problem #12. You can use information from other problems to arrive at your answer. Write your estimate, and explain how you got the estimate.

You have values for the SD of Log(GNP) from your answer on page 1, the SD of Life expectancy from the given information, and the slope of 0.12. Then, solve for \( r \) using the equation:

\[
0.12 = r \left( \frac{\text{SD LogGNP}}{\text{SD Life expectancy}} \right).
\]

If you estimated the correlation by eye, we accepted answers from 0.70 to 0.90. The actual correlation is 0.81.

15) (3 points) This is not valid. Although the calculations are fine, the researcher is trying to learn about cities from a regression that uses countries. The relationships are likely to be different for the cities, so the regression is an extrapolation and therefore not valid. This is related to ecological correlation (Chapter 9.4).

16 a) (3 points)

\[
\Pr(\text{makes at least one}) = 1 - \Pr(\text{makes none}) = 1 - \Pr(\text{misses first and misses second and misses third}).
\]

Since the shots are independent, \( \Pr(\text{misses first and misses second and misses third}) = .06 \times .06 \times .06 \)

Final answer: \( 1 - .06^3 \)

b) (3 points)

\[
\Pr(\text{pick at least one red in Box A}) = 1 - \Pr(\text{pick no reds in Box A}) = 1 - \Pr(\text{pick blue on first}) \times \Pr(\text{pick blue on second | blue on first}) \\
= 1 - (20/50) \times (19/49)
\]
\[
\Pr(\text{pick at least one red in Box B}) = 1 - \Pr(\text{pick no reds in Box B}) = 1 - \Pr(\text{pick blue on first}) \times \Pr(\text{pick blue on second | blue on first}) \\
= 1 - (2/5) \times (1/4)
\]

Box B gives the higher chance.

17) (4 points) No. This calculation assumes that men and their sons choose their jobs independently. This is not likely to be true: it is probably the case that men in manufacturing are more likely to have their sons in manufacturing.

Answers that said that they were independent and justified independence in terms of fathers’ and sons’ job choices got two points. Answers that made good points but missed the central idea of independence got some points. Answers that just said yes without justifying independence got zero points.

18) a) (5 points) From this study, can you conclude that owning a gun causes people to have higher likelihoods of experiencing a homicide than not owning a gun? Defend your answer. Be specific in your defense, referring to strengths or deficiencies of the study design.

You cannot make this conclusion. People who own guns are likely to differ from people who do not own guns in ways that affect homicide rates. For example, people who own guns may live in more dangerous areas or have riskier lifestyles. Because of these differences in background characteristics, we must conclude that the association found in the data does not prove that owning guns causes people to have higher likelihoods of experiencing a homicide. This answer earned five points.

Some answers suggested that the sample may not have been representative of the larger population. The researchers claim that it is, but it is good to be skeptical of that claim. This answer earned two points if the main issue of no balance was not mentioned.

Some answers suggested that the researchers need to look at all homicides, not just the ones where there is a gun in the home. Some possible reasons for looking only at homicides in homes include: 1) the researchers may not have the information on gun ownership for people shot outside their homes, and 2) the researchers want to focus on a particular type of homicide. Even if the researchers did use all the homicide data, they still would have to deal with possible lack of balance in the background characteristics. This answer, without any mention of potential imbalance in background characteristics, earned two points for pointing out that it would be helpful to look at all homicides.

b) (5 points)
1. Was the survey mailed to a random sample of scientists or not?
2. Was there any nonresponse bias?
3. How was the list of scientists used to pick the sample compiled?

Other answers that focused on the questionnaire wording did not count. It is important to consider this issue, but the problem specified not to worry about questionnaire wording. Answers with 1 of the correct answers earned two points. Answers with one right answer and almost correct answer earned extra points to get 3 or 4 points.