

Solutions

Part I. (20 points) Developmental psychology researchers suggest that some homework assignments can encourage the other members in the family to get involved (participate) in the homework solution process and hence help a child learn the school material better. The following table gives the mean family involvement score for two different types of homework assignments given to school children. One type of homework, called TIPS, is designed for increasing family involvement. The other one, called ATIPS, is an ordinary homework assignment.

The subjects are assigned to TIPS and ATIPS groups randomly (but they are not told which group). In the end, they are given a questionnaire to measure their family involvement score on a scale of 0 (never involved), 1,2,3 to 4 (always involved) for the homework assignments done. In each group, some families/children did the homework on a regular basis, and some others did not do most of the homework assignments.

Mean family involvement score

	TIPS	ATIPS
Did the HW regularly	2.55	1.43
Did not do most HW	1.08	1.01
Whole group	1.72	1.29

1. (3 points) Write down a variable included in this study and its values.

Family involvement score. Values: 0, 1, 2, 3, 4

2. (2 points) What are the treatment and control groups above?

\downarrow \downarrow
 TIPS ATIPS

3. (5 points) "This experiment started as a randomized controlled experiment, but then changed to an observational study". Do you agree? Explain in four sentences at most.

The subjects were randomly assigned to the treatment and control groups at first. However, some of them quit the study by not doing the HW. So, the remaining subjects are no more randomly assigned as in an observational study.

4. (1 point) Is this a "blind" study? How do you understand from above?

Yes, because the subjects do not know which type of HW they are assigned.

5. (4 points) Do you think TIPS assignments increase family involvement? Which numbers could you compare in the above table to decide?

Yes. Either the whole group numbers: 1.72 and 1.29 or the ones who did their HW's: 2.55 and 1.43.

6. (2 points) From the results given in the table, it looks like "doing the HW regularly increases family involvement". Which numbers are compared?

1.08 and 2.55 and 1.01 and 1.43.

7. (3 points) Do you think "family habits" could be confounding for the relationship stated in question 6? Explain with at most two sentences.

Yes, because the general habits of the family including the child (such as attention-tendency to interfere) can affect both "doing HW or not" and "family involvement". We do not expect a causal relationship as in 6.

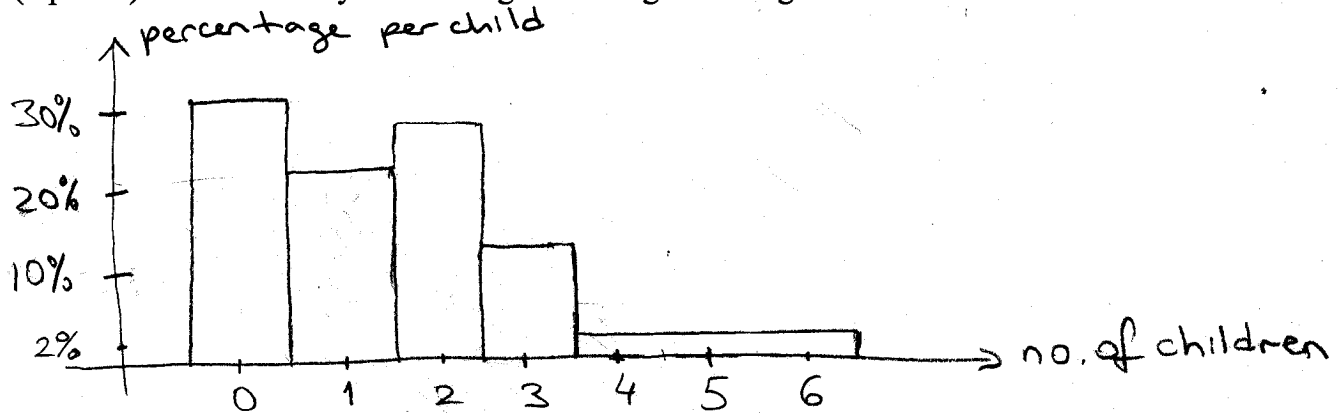
Part II. (20 points) Consider the distribution of women age 25-39 by educational level and their number of children given in the following table, obtained from the survey of a random sample.

Number of children	Women who are high-school graduates	Women with college degrees (university graduates)
0	30.3%	47.9%
1	21.7%	19.4%
2	28.4%	23.7%
3	13.6%	7.0%
4,5,6	6.0%	2.0%

1. (2 points) What is the variable under study?

The number of children

2. (6 points). Draw a density scale histogram for high-school graduates. Label the axes.



3. (4 points) Find the 90th percentile for each educational level.

3 for high-school graduates ($30.3 + 21.7 + 28.4 + 13.6 > 90$)
 2 for university graduates ($47.9 + 19.4 + 23.7 > 90$)

4. (5 points) Do an approximation as follows: The average number of children among those with 4-6 children can be taken as 5. That is, consider 5 instead of 4-6.

Based on this approximation, estimate the average (mean) number of children among high-school graduates.

$$0 \times 30.3\% + 1 \times 21.7\% + 2 \times 28.4\% + 3 \times 13.6\% + 5 \times 6\% \\ \approx 1.5$$

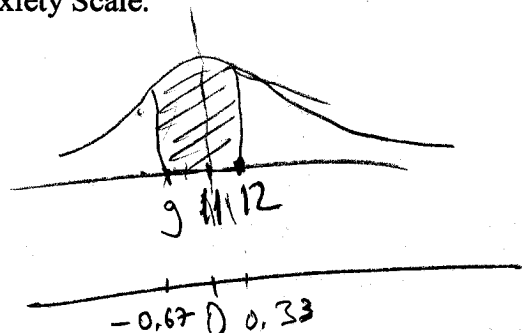
5. (3 points) Find the median for high-school graduates. Compare it with the mean: which one is larger? Explain in relation to the histogram of question 2 above (two sentences at most!).

Median is 1 because $30.3\% + 21.7\% > 50\%$

Mean is larger as we expect from the histogram which has a long-tail to the right.

Part III. (20 points) A person's anxiety at the dentists' office can be measured with "Dental Anxiety Scale" questionnaire. Scores on this scale, range from 0 (no anxiety) to 20 (extreme anxiety). Assume that the distribution of all scores is approximately normal with mean 11 and standard deviation 3.

1. (6 points) Find the percentage of people who would score between 9 and 12 on the Dental Anxiety Scale.

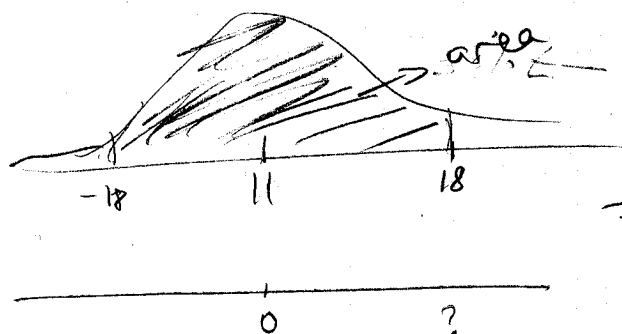


$$\frac{12-11}{3} \approx 0.33$$

$$\frac{9-11}{3} \approx -0.67$$

$$\text{Table} \Rightarrow \frac{48.43\% + 27.37\%}{2} = 37.9\%$$

2. (5 points) People with scores above 18 are considered to be "highly anxious". What percentile is 18?



$$\frac{18-11}{3} = \frac{7}{3} = 2.33$$

$$\Rightarrow \text{area} \approx 98\%$$

\Rightarrow blank areas, each 1%

\Rightarrow 99th percentile is 18.

3. (3 points) What is the expected value of the average score of a random sample of people?

The expected value is the population average, namely 11.

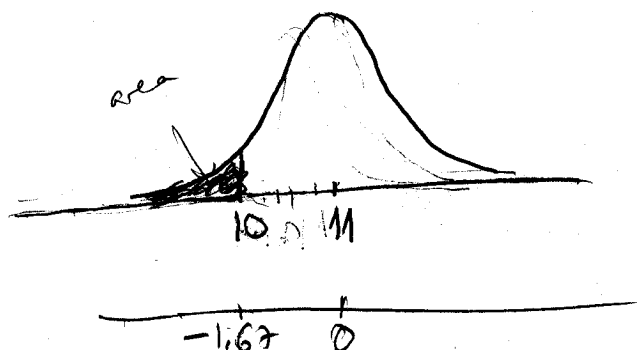
4. (6 points) What are the chances that a random sample of 25 subjects will have an average score less than 10 on the Dental Anxiety Scale?

SE for average score in the sample:

$$\frac{SD}{\sqrt{n}} = \frac{3}{\sqrt{25}} = 0.6$$

$$\Rightarrow \frac{10-11}{0.6} = -1.67$$

$$\text{table} \Rightarrow \text{area} = \frac{100-90}{2} = 5\%$$



Part IV. (20 points) In an archeological study at Efes, 180 pieces of ancient pottery have been found. Among these, 117 are painted pieces and the others have no paint on them.

1. (3 points) Estimate the percentage of painted pottery in ancient Efes (among all pottery used).

$$\frac{117}{180} = 65\%$$

2. (10 points) Construct a 90% confidence interval for the percentage of painted pottery in ancient Efes.

$$\hat{p} \pm z SE$$

$$90\% \Rightarrow z = 1.65$$

$$SE = \sqrt{\frac{(0.65)(0.35)}{180}} \approx 0.036 = 3.6\%$$

$$\Rightarrow 90\% \text{ CI for } p : 65\% \pm (1.65) 3.6\%$$

$$\Rightarrow (59.1\%, 70.9\%)$$

3. (3 points) Does the interval you constructed above contain the true percentage of painted pottery in ancient Efes? Explain in at most three sentences.

It may or may not. We are 90% confident that this interval contains p .

4. (4 points) Fill in the blanks with "sample" or "population":

The standard error measures the chance error in the sample percentage.

The confidence interval is for the population percentage.

Part V. (20 points) A random sample of 16 households are interviewed for the time (in hours) they spend watching television on Sundays. Here is a MINITAB output for the descriptive statistics of the data set:

Variable	N	Mean	SE	StDev
Time	16	5.6	0.508	?

1. (3 points) Note that the standard deviation in the sample is not written above. Find it from the given information.

$$SD = SE \sqrt{n} = (0.508) \sqrt{16} = 2.032$$

2. (8 points) Find a 99% confidence interval for the average time spent watching television on Sundays in the population.

For 99% \Rightarrow  $\Rightarrow z = 2.6$

$$\bar{x} \pm z SE$$

$$\Rightarrow 5.6 \pm (2.6)(0.508)$$

$$\Rightarrow (4.3, 6.9)$$

3. (4 points) In repeated sampling of 100 such samples, how many of the 99% confidence intervals (obtained from each of them) do you expect to cover the average time spent watching television on Sundays in the population? μ

Since each one is 99% CI, we expect

$$\frac{99}{100} (100 \text{ intervals}) = 99 \text{ intervals to contain } \mu$$

4. (5 points) In fact, the data set is 9, 3, 8, 5, 4, 5, 7, 8, 8, 5, 3, 6, 7, 4, 3, 4

Find the standard deviation from this sample (Hint: you can first sum all (entry-average)² values, then divide by n or $n-1$, and then take the square root. MINITAB output above uses " $n-1$ ")

$$\frac{(9-5.6)^2 + (3-5.6)^2 + \dots + (4-5.6)^2}{15} = 2.032$$