Homework 2 (Page 204 in 3rd Edition, exercises 4.3, 4.4, 4.7, 4.8)

4.3 College credits. A college counselor is interested in estimating how many credits a student typically enrolls in each semester. The counselor decides to randomly sample 100 students by using the registrar's database of students. The histogram below shows the distribution of the number of credits taken by these students. Sample statistics for this distribution are also provided.

Min	8
Q1	13
Median	14
Mean	13.65
SD	1.91
Q3	15
Max	18

(a) What is the point estimate for the average number of credits taken per semester by students at this college? What about the median?

Answer:

(1) The point estimate for the average number of credits taken per semester by students at this college is 13.65.

(2) The median is 14.

(b) What is the point estimate for the standard deviation of the number of credits taken per semester by students at this college? What about the IQR?

Answer:

(1) The point estimate for the standard deviation is 1.91.

(2) The IQR is Q3-Q1=15-13=2.

(c) Is a load of 16 credits unusually high for this college? What about 18 credits? Explain your reasoning. Hint: Observations farther than two standard deviations from the mean are usually considered to be unusual.

Answer:

The approximate 95% confidence interval is given by $(13.65 - 2 \times 1.91, 13.65 + 2 \times 1.91) = (9.83, 17.47)$. Therefore, 16 is not unusual but 18 is unusual.

(d) The college counselor takes another random sample of 100 students and this time finds a sample mean of 14.02 units. Should she be surprised that this sample statistic is slightly different than the one from the original sample? Explain your reasoning.

Answer: No, this is due to sampling error, as 14.02 is covered by the confidence interval in (c).

(e) The sample means given above are point estimates for the mean number of credits taken by all students at that college. What measures do we use to quantify the variability of this estimate (Hint: recall that $SD_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$)? Compute this quantity using the data from the original sample.

Answer:

(1) Standard error of mean is used to quantify the variability of this estimate.

(2) Standard error of mean is $SD_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.91}{10} = 0.191$.

4.4 Heights of adults. Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender, for 507 physically active individuals. The histogram below shows the sample distribution of heights in centimeters.

Min	147.2
Q1	163.8
Median	170.3
Mean	171.1
SD	9.4
Q3	177.8
Max	198.1

(a) What is the point estimate for the average height of active individuals? What about the median?
Answer:
The point estimate for the average height of active individuals is 171.1. The median is 170.3.

(b) What is the point estimate for the standard deviation of the heights of active individuals? What about the IQR?
Answer:
(1) the point estimate for the standard deviation of the heights of active individuals is 9.4.
(2) IQR = Q3-Q1= 177.8-163.8=14.

(c) Is a person who is 1m 80cm (180 cm) tall considered unusually tall? And is a person who is 1m 55cm (155cm) considered unusually short? Explain your reasoning.
Answer: The 95% CI is given by (162.3,189.9). So 180cm is not unusually tall, and 1.55 is unusually short.

(d) The researchers take another random sample of physically active individuals. Would you expect the mean and the standard deviation of this new sample to be the ones given above? Explain your reasoning.
Answer:
No the mean and the standard deviation would be different due to randomness in the sampling.

(e) The sample means obtained are point estimates for the mean height of all active individuals, if the sample of individuals is equivalent to a simple random sample. What measure do we use to quantify the variability of such an estimate (Hint: recall that $SD_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$)? Compute this quantity using the data from the original sample under the condition that the data are a simple random sample.
Answer:
(1) We use standard error to quantify the variability.
(2) Standard error of mean is $SD_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{9.4}{507} \approx 0.418$.

4.7 Chronic illness, Part I. In 2013, the Pew Research Foundation reported “45% of U.S. adults report that they live with one or more chronic conditions”. However, this value was based on a sample, so it may not be a perfect estimate for the population parameter of interest on its own. The study reported a standard error of about 1.2%, and a normal model may reasonably be used in this setting. Create a 95% confidence interval for the proportion of U.S. adults who live with one or more chronic conditions. Also interpret the confidence interval in the context of the study.

Answer:

(1) Mean is 45% and standard error is 1.2%, then the 95% confidence interval is (42.6%,47.4%)

(2) The confidence interval covers the true mean value with 95% probability.

4.8 Twitter users and news, Part I. A poll conducted in 2013 found that 52% of U.S. adult Twitter users get at least some news on Twitter. The standard error for this estimate was 2.4%, and a normal distribution may be used to model the sample proportion. Construct a 99% confidence interval for the fraction of U.S. adult Twitter users who get some news on Twitter, and interpret the confidence interval in context.

Answer:

(1) Mean is $\bar{X} = 0.52$ and standard error is $\frac{\sigma}{\sqrt{n}} = 0.024$, then the 99% confidence interval is (45.8%, 58.2%).

(2) The confidence interval covers the true mean with probability with 99% probability.