Chapter 4. (due Oct 11)

Graded: 4.4, 4.14, 4.24, 4.26, 4.34, 4.40, 4.48

**Q4.4**

(a) What is the point estimate for the average height of active individuals? What about the median?

Point estimate = mean = **171.1**

Median = **170.3**

(b) What is the point estimate for the standard deviation of the heights of active individuals? What about the IQR?

SD: **9.4**

IQR: 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.

Z(180cm) = 180 – 171.1/9.4 = **.95, which is less than 2 standard deviations, so not unusual**

Z(155cm) = 155 – 171.1/9.4 = **-1.7**, which is closer to 2 standard deviations, and more rare, but **not technically unusual**

(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.

**No, point estimates like the mean, and therefore stand deviation, only approximate population parameters, and vary from one sample to the other.**

(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? Compute this quantity using the data from the original sample under the condition that the data are a simple random sample.

The standard error describes the error associated with a sample estimate.

Se = 9.4/sqrt(507) = **.42**

**Q4.14**

(a) We are 95% confident that the average spending of these 436 American adults is between $80.31 and $89.11.

**False, we know their average spending, it is the confidence interval for all American adults.**

(b) This confidence interval is not valid since the distribution of spending in the sample is right skewed.

**False, it is a good estimation, especially considering the large enough sample size of 436.**

(c) 95% of random samples have a sample mean between $80.31 and $89.11.

**False, the 95% refers to the confidence of where the population’s point estimate will lie**

(d) We are 95% confident that the average spending of all American adults is between $80.31 and $89.11.

**True, that is what the confidence interval explains**

(e) A 90% confidence interval would be narrower than the 95% confidence interval since we don't need to be as sure about our estimate.

**True, a 90% confidence interval is narrower than a 95% confidence interval, and is for situations where less accuracy is needed.**

(f) In order to decrease the margin of error of a 95% confidence interval to a third of what it is now, we would need to use a sample 3 times larger.

**False, the sample does not have a direct multiplicative effect on the interval, since it is square-rooted in the denominator of the standard error. To have a 3x effect, we would need 9 times the sample.**

(g) The margin of error is 4.4.

**False, technically it is ±4.4.**

**Q4.24**

N = 36, min = 21, mean = 30.69, sd = 4.31, max = 39

(a) Are conditions for inference satisfied?

**Yes, the sample size is above 30, the distribution isn’t too heavily skewed, and the sample was randomly selected**

(b) Suppose you read online that children first count to 10 successfully when they are 32 months old, on average. Perform a hypothesis test to evaluate if these data provide convincing evidence that the average age at which gifted children fist count to 10 successfully is less than the general average of 32 months. Use a significance level of 0.10.

**Ho = the average gifted child counts to ten at the same age as other children, 32 months**

**Ho: μ = 32**

**Ha = gifted children count to ten sooner than other children, earlier than 32 months**

**Ha: μ < 32**

**Z = (30.69 – 32) / (4.31/6)**

**Z = -1.82**

**p-value = .0344, which is less than α = .10, so we reject the null hypothesis in favor of the alternative hypothesis, that gift children first count to 10 earlier than others.**

(c) Interpret the p-value in context of the hypothesis test and the data. P211

**The p-value of .0344 describes the probability of how unlikely the event is – that gifted children learn to read earlier. We set our significance level at .10, meaning we’ll reject the alternative hypothesis if the probability of observing the sample statistic is less than or equal to 90%. Because observing this sample statistic has a probability of over 96%, we rejected the null in favor of accepting it.**

(d) Calculate a 90% confidence interval for the average age at which gifted children first count to 10 successfully.

**30.69 ± 1.65 \* (4.31/6) = (29.5, 31.9)**

(e) Do your results from the hypothesis test and the confidence interval agree? Explain.

**They don’t agree, although the confidence interval is very close. The one tailed test hypothesis test had more leeway with .10 significance at one tail, while the confidence interval was only confident on each tail at only .05 significance.**

4.26 Gifted children, Part II. Exercise 4.24 describes a study on gifted children. In this study, along with variables on the children, the researchers also collected data on the mother's and father's IQ of the 36 randomly sampled gifted children. The histogram below shows the distribution of mother's IQ. Also provided are some sample statistics.

N = 36, min = 101, mean = 118.2, sd = 6.5, max = 131

(a) Perform a hypothesis test to evaluate if these data provide convincing evidence that the average IQ of mothers of gifted children is different than the average IQ for the population at large, which is 100. Use a significance level of 0.10.

Ho = Average IQ of mothers of gifted children is equal to IQ at large

**Ho: μ = 100**

Ha = Average IQ of mothers of gifted children is different to IQ at large

Ho: **μ ≠ 100**

**Z = (118.2-100) / (6.5/sqrt(36)) = 16.8, the probability is very close to 0, which is less than .10, so we reject the null in favor of the alternative**

(b) Calculate a 90% confidence interval for the average IQ of mothers of gifted children.

**115 ± 1.65 \* (6.5/6) = (113.2, 116.8)**

(c) Do your results from the hypothesis test and the confidence interval agree? Explain.

**Yes, they agree very strongly. Both the hypothesis test and confidence interval give very strong evidence that the IQ of mothers of gifted children is higher than average.**

**4.34 CLT.** Define the term “sampling distribution" of the mean, and describe how the shape,

center, and spread of the sampling distribution of the mean change as sample size increases.

**A sampling distribution for the mean is taking a collection of samples, calculating the mean of each, and analyzing their distribution, often graphically and with summary statistics.**

**The larger the size, the more closely it typically resembles the normal distribution. Smaller sample sizes are likely to be less normalized, and so the spread of the means and shape will be more haphazardly distributed. The center is more likely to be further away from the population mean with smaller samples.**

**4.40 CFLBs.** A manufacturer of compact fluorescent light bulbs advertises that the distribution of the lifespans of these light bulbs is nearly normal with a mean of 9,000 hours and a standard deviation of 1,000 hours.

(a) What is the probability that a randomly chosen light bulb lasts more than 10,500 hours?

Z = (10500 – 9000) / 1000 = 1.5

P-value = .0668, so **6.7%**

(b) Describe the distribution of the mean lifespan of 15 light bulbs.

**Because we need at least 30 events to start approximating the normal distribution, the distribution would be higher and less spread out due to a smaller standard error of 1000/sqrt(15) = 258.2, but we wouldn’t be confident in its accuracy.**

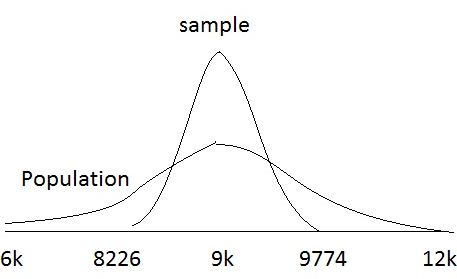
(c) What is the probability that the mean lifespan of 15 randomly chosen light bulbs is more than

10,500 hours?

Z = (10500 – 9000) / 258.2 = 5.8

**P-value approaches 0, so there’s a negligible probability that the mean is more than 10,500 hours.**

(d) Sketch the two distributions (population and sampling) on the same scale.



(e) Could you estimate the probabilities from parts (a) and (c) if the lifespans of light bulbs had

a skewed distribution?

**We can not estimate a if the distribution is not nearly normal, and c has an even smaller sample size, making it even more difficult.**

4.48 Same observation, different sample size. Suppose you conduct a hypothesis test based

on a sample where the sample size is n = 50, and arrive at a p-value of 0.08. You then refer back

to your notes and discover that you made a careless mistake, the sample size should have been

n = 500. Will your p-value increase, decrease, or stay the same? Explain.

**All things being equal, assuming I didn’t also miscalculate my standard deviation**, the standard error would decrease, pushing the Z value further to the tail(s), and therefore **decreasing the p-value**.

**If I had to recalculate the standard deviation as well**, the distribution would widen, even as the difference between the point estimate and the null value stayed the same, therefore the **p-value would increase** along with the difference between the tail(s) and the point estimate.