DATA 606 - Homework 4

Joshua Sturm

10/15/2017

Contents

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.

(a)

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

Solution

The point estimate for the average is the mean: 171.1. Similarly, the point estimate for the median would be 170.3.

(b)

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

Solution

The point estimate for the standard deviation is 9.4. The IQR is 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.

Solution

They are both unusual in that they are above the third quartile and below the first quartile, respectively.

```
n <- 507
mean <- 171.1
sd <- 9.4
shorter <- 155
taller <- 180
```

```r
# Calculate z-scores:
z_shorter <- (shorter - mean) / sd
z_taller <- (taller - mean) / sd
z_shorter
```

```
## [1] -1.712766
```

```r
z_taller
```

```
## [1] 0.9468085
```

Looking at the z-scores, though, the person with a height of 155cm is more unusual, since they are considerably farther from the mean.

(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.

Solution

I would expect the mean and sd to be similar but different from the first sample.

(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.

Solution

We can use the Standard Error: $SE = \frac{\sigma}{\sqrt{n}}$.

```r
SE <- sd / sqrt(n)
SE
```

```
## [1] 0.4174687
```

4.14 Thanksgiving spending, Part I.

The 2009 holiday retail season, which kicked off on November 27, 2009 (the day after Thanksgiving), had been marked by somewhat lower self-reported consumer spending than was seen during the comparable period in 2008. To get an estimate of consumer spending, 436 randomly sampled American adults were surveyed. Daily consumer spending for the six-day period after Thanksgiving, spanning the Black Friday weekend and Cyber Monday, averaged $84.71. A 95% confidence interval based on this sample is ($80.31, $89.11). Determine whether the following statements are true or false, and explain your reasoning.

(a)

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

Solution

False. The average spent in this sample is \$84.71. We are 95% confident that the average for the entire *population* is between \$80.31 and \$89.11.

(b)

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

Solution

False. So long as the sample is larger than 30, it's valid, and the skew should not matter.

(c)

95% of random samples have a sample mean between \$80.31 and \$89.11.

Solution

False. This confidence interval provides us with an idea of where the population mean is. A sample can differ from this, since the confidence interval is based on the mean and sd, which varies between samples.

(d)

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

Solution

True. This is the idea behind the confidence interval - to find the population mean.

(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.

Solution

True. If we're not measuring as precisely, the interval will be smaller, and less accurate.

(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.

Solution

False. The formula for the margin of error is: $MOE = z * \times SE$. If we want $\frac{1}{3} \times SE$, we get $\frac{\sigma}{\sqrt{(3^2 * n)}}$, so we'd have to use a sample 9 times larger.

(g)

The margin of error is 4.4.

Solution

True.

```
confl <- 80.31
confu <- 89.11
moe <- (confu - confl) / 2
moe
```

```
## [1] 4.4
```

4.24 Gifted children, Part I.

Researchers investigating characteristics of gifted children collected data from schools in a large city on a random sample of thirty-six children who were identified as gifted children soon after they reached the age of four. The following histogram shows the distribution of the ages (in months) at which these children first counted to 10 successfully. Also provided are some sample statistics.

(a)

Are conditions for inference satisfied?

Solution

I believe so. Since the sample is from a large pool (school children in a large city), we can assume they're independent. Furthermore, we have a sample size of 36, so $n = 36 \geq 30$. Lastly, there doesn't appear to be any strong skew in the histogram plot.

(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.

Solution

Null hypothesis $H_0 : \mu = 32$ months.
Alternate hypothesis $H_A : \mu < 32$ months.
$\alpha = 0.10$.
Since we're only interested if it's less than the average, it's a one-sided hypothesis test.

```
n <- 36
mean <- 30.69
sd <- 4.31
SE <- sd / sqrt(n)
z <- ((mean - 32)/SE)
```