705604096_stats101a_hw2

Jade Gregory

2023-04-12

Question 1

a)

```
n <- 30
mean <- 23606
sd <- 24757
lb <- mean - abs(qt(.025, 29) * sd / sqrt(n))
ub <- mean + abs(qt(.025, 29) * sd / sqrt(n))
lb
```

```
## [1] 14361.58
```

```
ub
```

```
## [1] 32850.42
```

The 95% confidence interval is (14361.58, 32850.42)

b) We have to assume the population distribution is Normal or our sample size is sufficiently large enough to provide us with a good approximation.

c) No, because a different sample will produce different sample statistics used to compute the confidence interval.

d)

```
a <- 25000
m <- 23606
sd <- 24757
n <- 30
z <- (m - a) / (sd / sqrt(n))
2 * pnorm(-abs(z))
```

```
## [1] 0.757772
```

Therefore, since our p-value is 0.757772 which is greater than our significance level of 0.05, we fail to reject the null hypothesis. So, the mean income is $25,000.

e) The smallest significance level we could have used so that we reject the null hypothesis would be 0.76.

Question 2

a)

```r
n <- 100
mean <- 23606
sd <- 24757
lb <- mean - abs(qt(.025, 99) * sd / sqrt(n))
ub <- mean + abs(qt(.025, 99) * sd / sqrt(n))
lb
```

```
## [1] 18693.67
```

```r
ub
```

```
## [1] 28518.33
```

The width of the interval decreases as the sample size increases.

b) The width of the interval decreases as our confidence level decreases.

Question 3

```r
cdc <- read.csv('cdc copy.csv')
t.test(cdc$weight ~ cdc$exerany, alternative = "two.sided", conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  cdc$weight by cdc$exerany
## t = 3.6842, df = 8024.9, p-value = 0.0002309
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  1.185482 3.881459
## sample estimates:
## mean in group 0 mean in group 1
##        171.5722        169.0387
```

a) Let

$$\bar{x}_e$$

represent the mean of desired weight for people who exercise and let

$$\bar{x}_n$$

represent the mean of desired weight for people who do not exercise.

$$H_0 : \bar{x}_e = \bar{x}_n H_a : \bar{x}_e \neq \bar{x}_n$$

b) The value of our test statistic is 3.6842.

c) The p-value of our test statistic is 0.0002309

d) With a significance level of 0.05, we reject our null hypothesis.

e) No, this is not the correct definition of the p-value. The p-value measures how likely it is that any observed difference between the null hypothesis and alternative hypothesis is due to chance.

f) The significance level is a measure of how small we want the probability of rejecting the null hypothesis when it is actually true to be. Using the p-value, it tests the probability of observing extreme values by chance.