PLCY 610 PS4

ENWONGO EKAH

4/6/2021

Step1: Clear all

```
rm(list = ls())
```

#Step 2: Load required packages

```
library (tidyverse)
```

#Step 3: Import datasets

```
library(readr)
ps4data <- read_csv("C:/Users/enwon/OneDrive/Desktop/PLCY610/PS4/ps4data.csv")

##
## -- Column specification --------------------------------------------------------
## cols(
##   abd = col_double(),
##   age = col_double(),
##   fthr_ed = col_double(),
##   educ = col_double(),
##   log.wage = col_double()
## )

View(ps4data)
```

ENWONGO EKAH

#Step4: Problem 1 (a)

Estimate the population average,μ, of the education variable by calculating the sample mean. Report the standard error and a 95% confidence interval for your estimate as well.

Solution:

Obs (n)	Sample Mean	Sample SD	Sample SE	Critical T Value	95% Confidence Interval
741	7.04	2.88	0.106	1.963	6.84, 7.25

Calculate the standard error:

$$SE_{\overline{X}} = \frac{s}{\sqrt{n}}$$

$$= \frac{2.88}{\sqrt{741}}$$

$$= 0.106$$

Confidence Interval:

$$\text{CI} = \bar{x} \pm t\,n - 1 \,*\, SE$$

$$= 7.04 \pm 1.963 * 0.106$$

$$= 7.04 \pm 0.208078$$

$$= 6.84, 7.25$$

Code:

```
#sample Mean
mean(ps4data$educ)

## [1] 7.044534

#sample sd
sample_sd<-sd(ps4data$educ)
sample_sd
```

```
## [1] 2.888357

n <- dim(ps4data)[1]
n

## [1] 741

#Sample SE
SE <- sample_sd/sqrt(n)
SE

## [1] 0.1061065

#Critical t value
qt(0.975, dim(ps4data)[1]-1)

## [1] 1.963175
```

#Subtract one for degrees of freedom of t

#Step 5 Problem 1b

Conduct a two sided t-test with the null hypothesis with $\mu_o = 5$ with $\alpha = .05$. Report the degrees of freedom, test statistic, and p- value of your test. What is the interpretation of this p-value? Can you reject the null hypothesis?

Solution:

Two-sided one sample t-test

$H_0 : \mu = 5$

$H_a : \mu \neq 5$

$\alpha = .05$

Sample Mean	Critical T-	Degrees of	T Test	95% Confidence	P Value

	Value	Freedom	Statistics	Interval	
7.044	1.963	740	19.27	6.84, 7.25	.000001

The T value is 1.963 and the t test statistics is 19.27 is in the rejection area. The p value of .000001 is statistically significant because it is less than the alpha, .05 and so close to 0 indicating that we would not see a sample mean value as extreme or more extreme than 7.04 if we conduct 100 studies with a sample size of 741 and the μ_o = 5. Therefore, we have sufficient evidence to reject the null hypothesis.

Code:

```
#two sided t-test
t.test(ps4data$educ, mu=5)

##
## One Sample t-test
##
## data:  ps4data$educ
## t = 19.269, df = 740, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 5
## 95 percent confidence interval:
##  6.836229 7.252840
## sample estimates:
## mean of x
##  7.044534
```

#Step 6 Probem 1 c

Repeat b) with the null hypothesis $\mu_o = 7.2$

Solution:

Two-sided one sample t-test

$H_0 : \mu = 7.2$

$H_a : \mu \neq 7.2$

$\alpha = .05$

Sample Mean	Critical T-Value	Degrees of Freedom	T Test Statistics	95% Confidence Interval	P Value
7.044	1.963	740	-1.47	6.84, 7.25	0.14

The T value is 1.963 and the t test statistics -1.47 is within the non -rejection area. The p value of 0.14 is statistically insignificant because it is greater than the alpha, .05 indicating that by random chance we would see a sample mean value as extreme or more extreme than 7.044 in 14 out of 100 studies with a sample size of 741 and the μ_o = 5. Therefore, we fail to reject the null hypothesis.

Code:

```
# 2 sided t-test with different null
t.test(ps4data$educ, mu=7.2)

##
## One Sample t-test
##
## data:  ps4data$educ
## t = -1.4652, df = 740, p-value = 0.1433
## alternative hypothesis: true mean is not equal to 7.2
## 95 percent confidence interval:
## 6.836229 7.252840
## sample estimates:
## mean of x
## 7.044534
```

#Step 7: Problem 1 d

Solution: Two Sample T test

$H_o : \mu Y_t = \mu Y_c$

$$H_a : \mu Y_t \neq \mu Y_c$$

$$\alpha = .05$$

Sample Mean of x	Sample Mean of y	Critical T-Value	Degrees of Freedom	T Test Statistics	95% Confidence Interval	P Value
6.820346	7.415771	1.963	551.58	-2.6798	-1.0318702, -0.1589784	0.007587

The test statistic is -2.6798 and the critical value is 1.963. We reject the null. The purpose of the T test was to see if the years of education is statistically different or not for those abducted and those not abducted. The alternative hypothesis mean difference is not equals to zero. The null is set in that the mean of education of those who were abducted minus the mean of education of those not abducted is actually zero. The P value goes ahead to certify this as it is 0.007587 and less than the alpha .05. Therefore, if we were to create a 95% confidence interval, we will not see zero in the confidence interval.

Code:

```
# 2 sided test for equality of treatment and control, var=abd
t.test(ps4data$educ[ps4data$abd==1],ps4data$educ[ps4data$abd==0])

##
##  Welch Two Sample t-test
##
## data:  ps4data$educ[ps4data$abd == 1] and ps4data$educ[ps4data$abd == 0]
## t = -2.6798, df = 551.58, p-value = 0.007587
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.0318702 -0.1589784
## sample estimates:
```

```
## mean of x mean of y
## 6.820346  7.415771
```

Step 8: Problem 1 e

Suppose your research assistant comes to you and says “ I think we can get stronger results if we ran a one- sided test instead”. Do you think this is a good idea? Why or why not? What assumptions would you be making?

Solution: A one sided test is not usually recommended and I don’t think it is a good idea as we have to make assumptions that the treatment effect is either negative(left) or positive (right). When using a two tailed test, regardless of the direction of the relationship you hypothesize, you are testing for the possibility of the relationship in two directions. Because one sided test provides more power to detect an effect, one may want to use it but will miss an effect in the other direction as in the case of this hypothesis. Using a one sided test means that we are then effectively forfeiting the right to the rest of the difference in the opposite direction.

Step 9: Problem 1 f

Repeat d) using a one-sided t test where the alternative hypothesis is $H_a: \mu Y_t < \mu Y_c$

Solution:

$H_o: \mu Y_t > \mu Y_c$

$$H_a: \mu Y_t < \mu Y_c$$

$$\alpha = .05$$

Sample Mean of x	Sample Mean of y	Critical T-Value	Degrees of Freedom	T Test Statistics	95% Confidence Interval	P Value
6.820346	7.415771	1.963	551.58	-2.6798	-Inf, -0.2293362	0.0038

The T value is 1.963 and the t test statistics for a one -sided test is -2.6798 is in the rejection area. The p value of 0.0038 is statistically significant because it is less than the alpha, .05. We reject the null because we are comparing the P value which is all on one side to the alpha of 0.5. The difference appears to be not zero and so we will not expect to see zero in the Confidence interval. Therefore, we have sufficient evidence to reject the null hypothesis.

Code:

```
t.test(ps4data$educ[ps4data$abd==1],ps4data$educ[ps4data$abd==0], alternative=
"less")

##
## Welch Two Sample t-test
##
## data:  ps4data$educ[ps4data$abd == 1] and ps4data$educ[ps4data$abd == 0]
## t = -2.6798, df = 551.58, p-value = 0.003794
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf -0.2293362
## sample estimates:

## mean of x mean of y
##  6.820346  7.415771
```

#Step 10 : Problem 1 g

Solution:

Two Sample T- test for equality of means

$H_o : \mu Y_t = \mu Y_c$

$$H_a : \mu Y_t \neq \mu Y_c$$

$$\alpha = .05$$

Sample Mean of x	Sample Mean of y	Critical T-Value	Degrees of Freedom	T Test Statistics	95% Confidence Interval	P Value
5.764069	6.068100	1.963	572.99	-1.1125	-0.8408032, 0.2327410	0.2664

The T value is 1.963 and the t test statistics is -1.1125 is in the non -rejection region. The p value of 0.266 is statistically insignificant. We fail to reject the null because so there is no significance difference in our sample. If our sample is as if random, it would mean that any significant difference would be attributed to the actual abduction itself. As if random means that although the design was not set up to randomly sample, the way the experiment worked out was as if the people were randomly allocated to each group. Yes, it makes sense because if they were randomly treated, we will not expect to see the difference in the fathers' education levels.

Code:

```
t.test(ps4data$fthr_ed[ps4data$abd==1],ps4data$fthr_ed[ps4data$abd==0])

##
##  Welch Two Sample t-test
##
## data:  ps4data$fthr_ed[ps4data$abd == 1] and ps4data$fthr_ed[ps4data$abd == 0]
## t = -1.1125, df = 572.99, p-value = 0.2664
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8408032  0.2327410
## sample estimates:
## mean of x mean of y
##  5.764069  6.068100
```

ENWONGO EKAH

#Step 11 : Problem 1 h

Solution: To minimize the probability of incorrectly rejecting the null hypothesis, we decrease the significance levels, alpha. This is a Type 1 error; the error of rejecting the null hypothesis when it is actually true. On the other hand, in reducing the probability of type I error, there is a trade-off. It increases the probability of committing Type II error, that is not rejecting the null even if it is false.

#Step 12 : Problem 1 i

Solution: We are evaluating within subject difference. Comparing the same person at two different times to see whether abductee income was higher after the program (time2) than before abductee time 1. Therefore, we are conducting a paired t – test.

Let's refer to our within sample difference as D =T_{A1} - T_{A2}

D =$T_{A,before}$ - $T_{A,after}$

H_O = μ_D = 0

H_O : μ income in Time 1 = μ Income in Time 2

Problem 2

I simulate 1000 sample size 10 from the above distribution. For each sample I perform

a t test H_O = μ_0 = 1 at the α = .05 level. Upon analyzing my results, I find that I reject the null hypothesis 9% of the time. Does this seem to make sense. What is going on?

Solution: Theoretically, if your alpha is .05, you should reject the null 5% of the time but here it is 9% and it is sensible because the sampling distribution is too small, and population is a non-normal distribution for the sample means. Hence, the Central limit theorem has not taken effect. CLT can only take effect if n is large no matter what the shape of the distribution is.

2b. I consult some SPP graduates who have already taken 610 and repeat part (a) with sample sizes of 100. Now I reject the null about 5% of the time. Why did my results change? What do my results suggest about t-tests with non-normal populations?

Solution: The results changed because the Central Limit Theorem has now taken effect as the sample size is now 100. This suggests that T- tests with non- normal populations may only work with larger sample sizes. t- test does not necessarily work when the sample size is less than 15 and the population is non-normal but when we have a sample size of over 40, the t-test is the best fit.