Abdulaziz Alqumayzi G200007615 Module6_CT2

Rudra S Bandhu

11/06/2021

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(caret)

## Loading required package: ggplot2

## Loading required package: lattice

library(ggplot2)
```

Problem 1

```
# Code for problem 1


cat("\n")

# Calculate the z-score
p1_zScore <- (52.5-50)/(20/sqrt(64))
# Calculate the p-value
p1_a <- 2*pnorm(-abs(p1_zScore))
paste("p = ", round(p1_a,digits = 5))

## [1] "p =  0.31731"

cat("\n")

#Calculate the z-score
p2_zScore <- (55-50)/(20/sqrt(64))
# Calculate the p-value
p1_b <- 2*pnorm(-abs(p2_zScore))
paste("p = ", round(p1_b,digits = 5))

## [1] "p =  0.0455"
```

```
cat("\n")

#Calculate the z-score
p3_zScore <- (57.5-50)/(20/sqrt(64))
# Calculate the p-value
p1_c <- 2*pnorm(-abs(p3_zScore))
paste("p = ", round(p1_c,digits = 5))

## [1] "p =  0.0027"
```

P-value of the test hypothesis:

a) P-value is 0.31731
b) P-value is 0.0455
c) P-value is 0.0027

Problem 2

The null hypothesis (H0) is the lead content of human hair removed from adult individuals that had died between 1880 and 1920 have no different from the lead content of present-day adults.

The alternative hypothesis (H1) is the lead content of human hair removed from adult individuals that had died between 1880 and 1920 have different from the lead content of present-day adults.

```
# Code for problem 2

# Calculate the pooled standard error
SE  <- sqrt(((30 - 1) * 14.5^2 + (100 - 1) * 12.3^2)/(30 + 100 - 2))
# Calculate the t-test
t   <- ((48.5 - 26.6) - 0) / (SE * sqrt(1/30 + 1/100))
# Calculate the p-value
P   <- pt(abs(t), df = 30 + 100 - 2, lower.tail = F)
paste("t(df = ", 30 + 100 - 2, ") = ", round(t,digits = 5), ", p = ", P)

## [1] "t(df =  128 ) =  8.19888 , p =  1.09512239977613e-13"
```

P-value as a real number is about 0.000000000000109512239977613

At the 1 percent level of significance (0.01) and a P-value that is near to zero, we have strong evidence to reject the null.

Problem 3

An environmental pollution scientist claims that two different solutions come from the same source.

```
# Code for problem 3

measurment_A <- c(6.24,6.31,6.28,6.3,6.25,6.26,6.24,6.29,6.22,6.28)
```

```
measurment_B <- c(6.27,6.25,6.33,6.27,6.24,6.31,6.28,6.29,6.34,6.27)

# Paired Samples t Test compares the means of two measurements taken from the
same individual, object, or related units.

# At a standard deviation equal to .05
# Calculate the t-test
t.test(measurment_A,measurment_B, alternative = "two.sided", paired = T,
var.equal=T, conf.level = 0.95)

##
##  Paired t-test
##
## data:  measurment_A and measurment_B
## t = -1.1173, df = 9, p-value = 0.2928
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.054445  0.018445
## sample estimates:
## mean of the differences
##                  -0.018

"P-value using our equation:"

## [1] "P-value using our equation:"

t <- ((mean(measurment_A)-mean(measurment_B))-
0)/sqrt((0.05^2)/10+(0.05^2)/10)
P     <- 2*pt(abs(t), df = 9, lower.tail = F)
paste("p = ", round(P,digits = 5))

## [1] "p =  0.44157"
```

We performed two tests, the first one using t.test() function of R. Second one is using the p-value equation.

Both results lead us to the same conclusion.

a

The data results show that the environmental pollution scientist claims was right. The data do not disproved the scientist.

At the 5 percent level of significance (0.05) and a P-value greater than 0.05, we have no evidence to reject the null. Therefore, we fail to reject the null.

b

P-value of t.test() function = 0.2928

P-value of our build equation = 0.4415

Problem 4

```r
# Code for problem 4a


"Point#2 results:"

## [1] "Point#2 results:"

t_stat1 <- (9.5 - 8)/(2/sqrt(5))
p1 <- 2*pnorm(t_stat1, lower.tail = F)
paste("t = ", round(t_stat1,digits = 5), ", p = ", round(p1,digits = 5))

## [1] "t =  1.67705 , p =  0.09353"

"Point#3 results:"

## [1] "Point#3 results:"

t_stat2 <- (8.5 - 8)/(2/sqrt(5))
p2 <- 2*pnorm(t_stat2, lower.tail = F)
paste("t = ", round(t_stat2,digits = 5), ", p = ", round(p2,digits = 5))

## [1] "t =  0.55902 , p =  0.57615"

"Point#4 results:"

## [1] "Point#4 results:"

t_stat3 <- (11.5 - 8)/(2/sqrt(5))
p3 <- 2*pnorm(t_stat3, lower.tail = F)
paste("t = ", round(t_stat3, digits = 5), ", p = ", round(p3,digits = 5))

## [1] "t =  3.91312 , p =  9e-05"
```

a

Point#1: The null hypothesis (H0) is that the signal value mean will be 8

Point#2: t = 1.67705 and p = 0.09353

At the 5 percent level of significance (0.05) and p-value is greater than 0.05, we fail to reject the null hypothesis.

Point#3: t = 0.55902 and p = 0.57615

At the 5 percent level of significance (0.05) and p-value is greater than 0.05, we fail to reject the null hypothesis.

Point#4: t = 4 and p = 9e-05

At the 5 percent level of significance (0.05) and p-value is less than 0.05, we can reject the null hypothesis.

```r
# Code for problem 4b

ct = qnorm(1-0.05) # calculate critical value
typeII <- pnorm(ct - (10-8)/(2/sqrt(5))) # calculate type II error
powerTest <- 1- typeII # calculate power of the test

"Point#2:"

## [1] "Point#2:"

paste("Type II error = ", round(typeII, digits = 5))

## [1] "Type II error =  0.27719"

"Point#3:"

## [1] "Point#3:"

paste("Power of the test = ", round(powerTest, digits = 5))

## [1] "Power of the test =  0.72281"
```

b

Point#1: The alternative hypothesis (H1) is equal (or no different from) 10.

Point#2: Probability of type-II error is 0.27719

Point#3: The power of the test is 0.72281

```r
# Code for problem 4c


{
  n = 1
  powerTest2 = 1-pnorm(ct - (9.2-8)/(2/sqrt(n))) #power of the test
  while(powerTest2 < 0.75)
  {
    n = n+1
    powerTest3 = 1-pnorm(ct - (9.2-8)/(2/sqrt(n))) #power of the test
    temp = n
    powerTest2 = powerTest3
   }
  cat("Number of signals sent to reach 75 percent:",n,"\n")
 }

## Number of signals sent to reach 75 percent: 15

cat("\n")

before_reach_perc <- 1-pnorm(ct - (9.2-8)/(2/sqrt(14)))
after_reach_perc <-  1-pnorm(ct - (9.2-8)/(2/sqrt(15)))
```

```
paste("The probability of sending 14 signals is
",round(before_reach_perc,digits = 5))

## [1] "The probability of sending 14 signals is  0.72579"

paste("The probability of sending 15 signals is
",round(after_reach_perc,digits = 5))

## [1] "The probability of sending 15 signals is  0.75141"
```

c

15 signals needed to reach a 75 percent probability of rejection when true mean μ = 9.2

The probability of sending 14 signals is 0.72579

The probability of sending 15 signals is 0.75141

References:

Field, A. P., Miles, J., & Field Zoë. (2017). Discovering statistics using R. W. Ross MacDonald School Resource Services Library.

Black, K. (2015). 10. calculating P values. Retrieved November 6, 2021, from https://www.cyclismo.org/tutorial/R/pValues.html.

Gimond, M. (2017). Comparing means: Z and T tests. Retrieved November 6, 2021, from https://mgimond.github.io/Stats-in-R/z_t_tests.html.

Grande, T. (2016). Calculating Type II Error (Beta) and Power using Excel. Retrieved November 6, 2021, from https://www.youtube.com/watch?v=gFlrJDh4N9E&t=287s&ab_channel=Dr.ToddGrande.