ST 411/511 Homework 3

Due on January 30

Vijay Vardhan Tadimeti

Winter 2019

Instructions

You should submit your assignment as either a PDF or Word document, which you can compile (should you choose – recommended) from the provided .Rmd (R Markdown) template. Please include your code.

Note: you may use the `t.test()` function for Question 4, but please do not use it for Questions 1-3.

Problems (25 points total)

Question 1: True or False, and explain your reasoning.

(a) (2 points) A p-value tells you the probability that the null hypothesis is true.

False. P values that are small usually indicate that there is an evidence of it being against the null hypothesis. Typically for a smaller p value we would reject the null hypothesis in favour of the alternate hypothesis.Usually a P value represents a for or against the hypothesis.

(b) (2 points) It is possible for a hypothesis test procedure to reject the null hypothesis even when the null hypothesis is true.

True. When we reject a null hypothesis when its true as well as possible leads to a type 1 error.

(c) (2 points) Consider the null hypothesis $H_0 : \mu = \mu_0$ versus a one-sided greater alternative $H_A : \mu > \mu_0$. For a fixed significance level α the critical value $z_{1-\alpha}$ will be greater than the critical value $t_{(4)1-\alpha}$ (i.e., the critical value for a t-distribution with 4 degrees of freedom).

It is false. Even if we assume alpha's value as 0.05, the critical value of its t distribution is more than the critical value of the normal distribution. The area of the tail under the t distribution is more when we compare it with the z of the normal distribution. It can be shown below.

```
qnorm(0.95)
```

```
## [1] 1.644854
```

```
qt(0.95, df = 4)
```

```
## [1] 2.131847
```

Question 2

A random sample of $n = 10$ OSU students is obtained, and the college GPA of each is recorded. The GPAs of the 10 students in the sample are provided in the vector `gpa`.

```
gpa <- c(3.1, 3.7, 4.0, 2.7, 2.5, 3.4, 3.5, 3.0, 1.9, 3.4)
```

(4 points) Test the null hypothesis $H_0 : \mu = 3.0$ versus the one-sided greater alternative $H_A : \mu > 3.0$ at significance level $\alpha = 0.05$. What do you conclude?

We notice that the P value is lesser than alpha, we hence reject the null hypothesis.

```
var <- var(gpa)
sd <- sd(gpa)
x  <- mean(gpa)
t  <- (x-3)/(sqrt(var/10))
p <- 1-pt(t,9)
sd
```

```
## [1] 0.6214678
```

```
x
```

```
## [1] 3.12
```

```
t
```

```
## [1] 0.6106082
```

```
p
```

```
## [1] 0.2782812
```

```
t.test(gpa,mu=3.0,alternative="greater")
```

```
##
##  One Sample t-test
##
## data:  gpa
## t = 0.61061, df = 9, p-value = 0.2783
## alternative hypothesis: true mean is greater than 3
## 95 percent confidence interval:
##  2.759747      Inf
## sample estimates:
## mean of x
##      3.12
```

Question 3

122 guinea pigs were randomly assigned to either a control group $(X_1, X_2, \ldots, X_m; m = 64)$ or to a treatment group that received a dose of *tubercle bacilli* $(Y_1, Y_2, \ldots, Y_n; n = 58)$. The lifetime, in days, for each guinea pig was recorded. The data are available as `ex0211` in the `Sleuth3` library.

```
data(ex0211)
```

(a) (2 points) Compute the sample mean and sample variance for each group.

```
meancontrol <- mean(ex0211$Lifetime[ex0211$Group=="Control"])
meanbasilli <- mean(ex0211$Lifetime[ex0211$Group=="Bacilli"])
varcontrol <- var(ex0211$Lifetime[ex0211$Group=="Control"])
varbasilli <- var(ex0211$Lifetime[ex0211$Group=="Bacilli"])
ncontrol <- length(which(ex0211$Group=="Control"))
nbasilli <- length(which(ex0211$Group=="Bacilli"))

meancontrol
```

```
## [1] 345.2344
```

```
meanbasilli
```

```
## [1] 242.5345
```

```
varcontrol
```

```
## [1] 49371.67
```

```
varbasilli
```

```
## [1] 13907.69
```

```
ncontrol
```

```
## [1] 64
```

```
nbasilli
```

```
## [1] 58
```

(b) (2 points) Compute the pooled variance estimate s_P^2.

```
sp2 <- ((ncontrol-1)*varcontrol + (nbasilli-1)*varbasilli)/(ncontrol+nbasilli-2)
sp2
```

```
## [1] 32526.28
```

(c) (2 points) Find the t-statistic for testing the null hypothesis that the difference in population mean survival time between these two treatments is zero ($H_0 : \mu_X - \mu_Y = 0$).

```
tstat <- (meancontrol-meanbasilli)/sqrt(sp2*(1/ncontrol + 1/nbasilli))
tstat
```

```
## [1] 3.141064
```

(d) (2 points) Find the critical value for a level $\alpha = 0.01$ one-sided test of the null hypothesis vs. the alternative that the difference in population mean survival time is greater than zero ($H_A : \mu_X - \mu_Y > 0$).

```
qt(1-0.01, df=ncontrol+nbasilli-2)
```

```
## [1] 2.357825
```

(e) (1 point) Find the one-sided greater p-value for the test in part (d).

```
p <- 1-pt(abs(tstat), df=ncontrol+nbasilli-2)
p
```

```
## [1] 0.00106009
```

(f) (2 points) Based on your answers to parts (d) and (e), would you reject the null hypothesis $H_0 : \mu_X - \mu_Y = 0$ vs. the alternative ($H_A : \mu_X - \mu_Y > 0$) at level $\alpha = 0.01$? State your conclusion in the original wording of the question.

Here the P value is less than alpha, we hence reject the null hypothesis.

```
control <- ex0211$Lifetime[ex0211$Group=="Control"]
basilli <- ex0211$Lifetime[ex0211$Group=="Bacilli"]
t.test(control, basilli, alternative="greater", var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  control and basilli
## t = 3.1411, df = 120, p-value = 0.00106
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  48.50152      Inf
## sample estimates:
## mean of x mean of y
##  345.2344  242.5345
```

Question 4

Researchers studied 15 pairs of identical twins where only one twin was schizophrenic ('Affected'). They measured the volume of the left hippocampus region of each twin's brain. This data is available as `case0202` in the `Sleuth3` library.

```
data(case0202)
View(case0202)
```

(a) (1 point) Is this paired data or two independent samples? Explain.

It is a paired test. Here we have 2 samples out of the total number of individuals, which are related to each other, but are also not independent. We must note that the both our mesurements are being performed on the same set of individuals.

(b) (3 points) Consider a hypothesis test to examine whether the difference in mean left hippocampus volume (Unaffected - Affected) is equal to zero, versus the two-sided alternative. Use the `t.test()` function in R to perform the appropriate t-test at significance level $\alpha = 0.01$. Report the p-value and what you conclude from the test

Here we have the P value as 0.006062 and our alpha value is 0.01. It is clear that the p value is lesser than the alpha. We hence reject the null hypothesis.

```r
unaffected <- case0202$Unaffected
affected <-case0202$Affected
t.test(case0202$Unaffected,case0202$Affected, data=case0202, alternative="two.sided", paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  case0202$Unaffected and case0202$Affected
## t = 3.2289, df = 14, p-value = 0.006062
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.0667041 0.3306292
## sample estimates:
## mean of the differences
##               0.1986667
```