

# ST 411/511 Homework 3

Due on January 29

Aashish Adhikari

Winter 2020

## Instructions

This assignment is due by 11:59 PM, January 29, 2020 on Canvas via Gradescope. **You should submit your assignment as a PDF which you can compile (should you choose – recommended) using the provide .Rmd (R Markdown) template.** If you opt to not use R Markdown, please format your solutions in a similar manner as provided in this document. Include your code in your solutions and indicate where the solutions for individual problems are located when uploading into Gradescope. You should also use complete, gramatically correct sentences for your solutions.

## 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. It gives us only the probability of observing a value as extreme as the critical value, not the probability that the null hypothesis is true. The test never tells us that we have enough evidence to support the null hypothesis. Thus, it does not tell us that we should accept the null hypothesis. We have to note that we are assuming the hypothesis' value to be true while calculating the  $p$  value.

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

True. We reject the null hypothesis depending upon the confidence level  $\alpha$  that we choose. However, there is always a non-zero probability of the true parameter being at any region of the numberline/  $x$ -axis.

**(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  $\alpha$  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).**

```
qnorm(1-0.01)
```

```
## [1] 2.326348
```

```
qt(1-0.01,4)
```

```
## [1] 3.746947
```

False. It is because the t-tests have higher-end tails than the z tests which results in farther critical values than the z test. It can also be seen in the example above.

## 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 are not given the population variance. Hence, we will use the t test.

```
samp_mean <- mean(gpa)
sv <- var(gpa)
t_st <- (samp_mean-3)/sqrt(sv/10)
p_val<-1-pt(t_st,9)
p_val
```

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

We fail to reject null hypothesis  $H_0 : \mu = 3.0$  at significance level alpha 0.05. Our p-value 0.27828 is greater than alpha(0.05) and we do not have strong evidence to prove the hypothesis  $H_A : \mu > 3.0$ .

## Question 3

122 guinea pigs were randomly assigned to either a control group ( $X_1, X_2, \dots, X_m; m = 64$ ) or to a treatment group that received a dose of *tubercle bacilli* ( $Y_1, Y_2, \dots, Y_n; n = 58$ ). 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.

```
control_group <- ex0211[ex0211$Group=="Control",]
bacilli_group <- ex0211[ex0211$Group=="Bacilli",]

mean_c <-mean(control_group$Lifetime)
mean_c
```

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

```
mean_b <-mean(bacilli_group$Lifetime)
mean_b
```

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

```
var_c <-var(control_group$Lifetime)
var_c
```

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

```
var_b <-var(bacilli_group$Lifetime)
var_b
```

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

The mean of control group is 345.2344 The mean of the bacilli group is 242.5345 The sample variance of the control group is 49371.67 The sample variance of the bacilli group is 13907.69

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

```
((64-1) * 49371.67) + ((58-1) * 13907.69)/(64+58-2)
```

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

The pooled variance is 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 ( $H_0 : \mu_X - \mu_Y = 0$ ).

```
((345.2344 - 242.5345) - 0) / (sqrt ( (32526.28) * ((1/64) + (1/58))))
```

```
## [1] 3.141065
```

The  $t$  statistic is 3.141065

(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(0.01,120)
```

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

```
qt(1-0.01,120)
```

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

The question did not specify the lesser or greater side, so I have included both. The lower critical value is -2.3578 and the upper critical value is 2.3578.

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

```
1-pt(3.141,120)
```

```
## [1] 0.001060305
```

0.00106 is the required  $p$  value.

(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.

I would reject the null hypothesis  $H_0 : \mu_X - \mu_Y = 0$  vs. the alternative ( $H_A : \mu_X - \mu_Y > 0$ ) at level  $\alpha = 0.01$  because the p value 0.0010 is less than the significance level 0.01. The p value gives the probability of observing the value as extreme as the critical value. Here, however, our critical value is beyond the t statistic even with a more lax significance level of 0.01.

## 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)
df <- case0202
names(df)

## [1] "Unaffected" "Affected"

df$Unaffected

## [1] 1.94 1.44 1.56 1.58 2.06 1.66 1.75 1.77 1.78 1.92 1.25 1.93 2.04 1.62 2.08

df$Affected

## [1] 1.27 1.63 1.47 1.39 1.93 1.26 1.71 1.67 1.28 1.85 1.02 1.34 2.02 1.59 1.97
```

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

The data is paired. It is because we are comparing two variables on the same group, not the measurements on the same variable for two different groups.

(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.

```
case0202$Unaffected

## [1] 1.94 1.44 1.56 1.58 2.06 1.66 1.75 1.77 1.78 1.92 1.25 1.93 2.04 1.62 2.08

case0202$Affected

## [1] 1.27 1.63 1.47 1.39 1.93 1.26 1.71 1.67 1.28 1.85 1.02 1.34 2.02 1.59 1.97

t.test(case0202$Unaffected, case0202$Affected, 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
```

The p value returned is 0.006062. From the test, we observe that p is less than the significance level. Thus, we reject the null hypothesis that the difference between means of two groups is 0.