HW1: Simple Linear Regression Overview

Amy Fan

DUE: 2/2/2024 11:59pm

(0) Instructions for installing tinytex for PDF rendering: **https://yihui.org/tinytex/**

```r
install.packages('tinytex')
tinytex::install_tinytex()
```

II. Applications [30 pts]

(1) [8 pts] (DAE 2.9) A Computer Program has produced the following output for a hypothesis testing problem:

Difference in sample means: 2.35

Degrees of freedom: 18

Standard error of the difference in sample means: ?

Test statistic: $t_0 = 2.01$

P-value: 0.0298

(a) [2 pts] What is the missing value for the standard error?

```r
d_bar <- 2.35
t_0 <- 2.01
df <- 18
n <- df + 1
s_d <- d_bar/(t_0/sqrt(n)); s_d
```

```
## [1] 5.096225
```

(b) [2 pts] Is this a two-sided or a one-sided test?

```r
pt(t_0,18,lower.tail = F) ### One-sided upper test
```

```
## [1] 0.02983103
```

(c) [2 pts] If $\alpha = 0.05$, what are your conclusions? **Reject the null hypothesis**

(d) [2 pts] Find a 90% two-sided CI on the difference in means. [-1.73, 1.73]

```
qt(.05,18);qt(.95,18)
```

```
## [1] -1.734064
```

```
## [1] 1.734064
```

(2) [8 pts] (DAE 2.26) The following are the burning times (in minutes) of chemical flares of two different formulations. The design engineers are interested in both the mean and veariance of the burning times.

Type 1	Type 1	Type 2	Type 2
65	82	64	56
81	67	71	69
57	59	83	74
66	75	59	82
82	70	65	79

(a) [3 pts] Test the hypothesis that the two variances are equal. Use $\alpha = 0.05$. **We fail to reject the null hypotehsis with a p value of 0.51. The variances can therefore be assumed equivalent.**

```
flares <- data.frame(type1 = c(65, 82, 81, 67, 57, 59, 66, 75, 82, 70),
                     type2 = c(64, 56, 71, 69, 83, 74, 59, 82, 65, 79)
                     )
Fstat <- var(flares$type1)/var(flares$type2)
1 - pf(Fstat,9,9)
```

```
## [1] 0.5128168
```

(b) [3 pts] Using the results of (a), test the hypothesis that the mean burning times are equal. Use $\alpha = 0.05$. What is the P-value for this test? **The p-value is 0.96 which is above the threshold. So, we conclude the burning times equivalent.**

```
t.test(flares$type1, flares$type2, var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  flares$type1 and flares$type2
## t = 0.048008, df = 18, p-value = 0.9622
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.552441  8.952441
## sample estimates:
## mean of x mean of y
##      70.4      70.2
```

(c) [2 pts] Discuss the role of the normality assumption in this problem. Check the assumption of normality for both types of flares.

The normality conditions need to be met for the t test which assumes that the data is distributed normally. The test statistic has a t distributed reference distribution for hypothesis testing which relies on the normally distributed data. The data appears to be normally distributed based on the qq plots.

```
qqnorm(flares$type1);qqline(flares$type1)
```

Normal Q–Q Plot
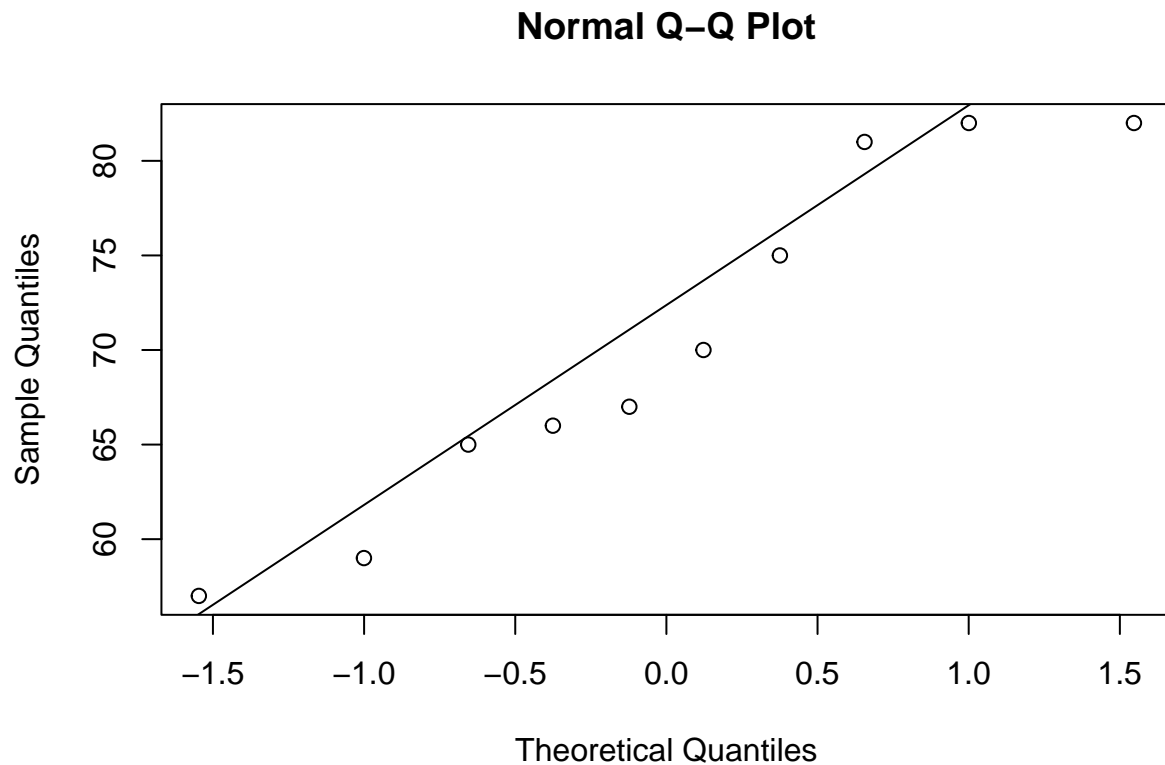

```
qqnorm(flares$type2);qqline(flares$type2)
```

Normal Q–Q Plot
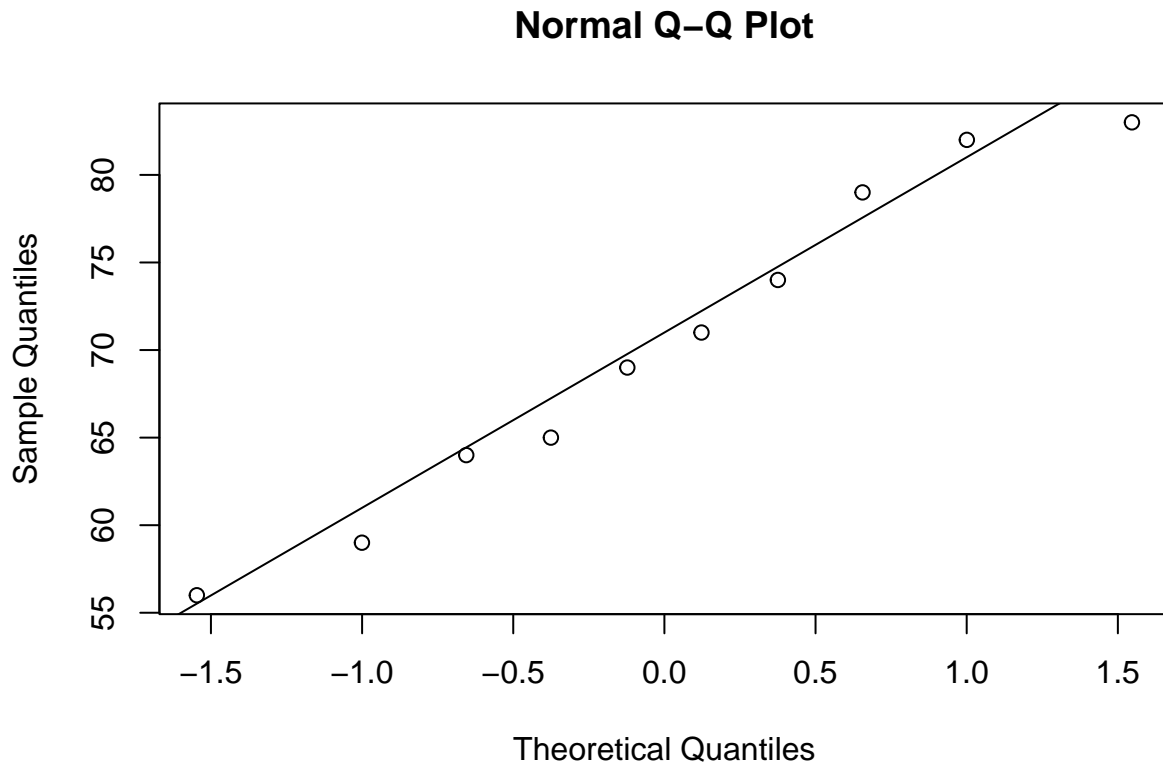
Verify you have entered the data correctly by checking that the sample means and variances for each type of flare are:

Type	Mean	Variance
1	70.4	85.82222
2	70.2	87.73333

```
mean(flares$type1); mean(flares$type2)
```

```
## [1] 70.4
```

```
## [1] 70.2
```

```
var(flares$type1); var(flares$type2)
```

```
## [1] 85.82222
```

```
## [1] 87.73333
```

(3) [8 pts] (DAE 2.32) *See text.* A .CSV file with the recorded observations is available on D2L.

(a) [2 pts] Is there a significant difference between the means of the population of measurements from which the two samples were selected? Use $/alpha = 0.05$. **There isn't a significant different since the pvalue is 0.6893.**

```
calipers <- read.csv("~/571_course/STAT571B/Datasets/DAE2_32.csv")
Fstat <- var(calipers$C1)/var(calipers$C2)
1 - pf(Fstat,11,11) # equal var
```

```
## [1] 0.8817792
```

```
t.test(calipers$C1, calipers$C2, var.equal = T)
```

```
##
##   Two Sample t-test
##
## data:  calipers$C1 and calipers$C2
## t = 0.40519, df = 22, p-value = 0.6893
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -0.001029568  0.001529568
## sample estimates:
## mean of x mean of y
##    0.26625    0.26600
```

(b) [2 pts] Find p value for test in a) **The p value is 0.6893.**
(c) [4 pts] Construct 95% confidence interval on the difference in mean diameter measurements for the two calipers. **The 95% confidence interval (-0.001029568, 0.001529568) includes zero which agrees with our previous conclusion to accept the null hypothesis.**

(4) [6 pts] (DAE 2.45) Construct a data set for which the paired t-test statistic is very large, but for which the usual two-sample or pooled t-test statistic is small. *Report both the data values and the t-statistics.* In general, describe how you created the data. Does this give you any insight regarding how the paired t-test works?

Paired t tests emphasize the ordering of the data, so the statistic is calculated using the differences between two columns in an ordered fashion. However, the independent t test calculates the mean and standard deviation for each sample independently and calculates the independent t statistic. In this way, we can construct a dataset by taking x1 from N(0,1). x2 is constructed by adding a small value to x1 with an even smaller standard deviation. This allows a significant difference to appear between paired groups, but a small difference between the pooled samples.

```
n <- 10
mu_1 <- 0
sd_1 <- 1
mu_2 <- 0.01
sd_2 <- .001

x1 <- rnorm(n, mu_1, sd_1)
x2 <- x1 + rnorm(n, mu_2, sd_2)
constructed_df <- data.frame(x1=x1, x2=x2)

t_paired <- t.test(x1,x2,paired = T)$statistic
t_ind <- t.test(x1,x2,paired = F)$statistic
paste('t_paired: ', t_paired,
               't_ind: ', t_ind)
```

```
## [1] "t_paired:  -32.7083556963302 t_ind:  -0.0223097380776043"
```

constructed_df

```
##             x1         x2
## 1  -0.7437853 -0.7328371
## 2  -0.2897028 -0.2793851
## 3  -2.0899493 -2.0806710
## 4   0.4609532  0.4714006
## 5   0.2460804  0.2577090
## 6   0.4722856  0.4822660
## 7  -1.0032391 -0.9943935
## 8   1.5541352  1.5645251
## 9   0.2002093  0.2104773
## 10 -0.7302585 -0.7218680
```