Computer Lab 3
Statistics for Business (STAT:1030, Bognar)

Homework: Inference for $\mu_1 - \mu_2$ using R

Suppose we have 2 brands of gasoline: *Brand 1* and *Brand 2*. Assume that the octane level of *Brand 1* gasoline has a $X_1 \sim N(\mu_1, \sigma_1)$ distribution, while the octane level of *Brand 2* gasoline has a $X_2 \sim N(\mu_2, \sigma_2)$ distribution. An inspector obtained a random sample of gasoline from 13 gas stations that sell *Brand 1*, and 12 gas stations that sell *Brand 2*. The measured octane levels are listed below.

Brand 1: 87.8, 88.1, 87.0, 87.3, 87.8, 87.5, 87.8, 87.2, 86.2, 87.1, 86.9, 86.8, 87.1
Brand 2: 86.2, 87.4, 87.7, 87.9, 88.2, 86.9, 87.4, 87.8, 89.3, 88.7, 88.5, 87.6

We will use R to find a 95% confidence interval for $\mu_1 - \mu_2$, and test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$ at the $\alpha = 0.05$ significance level. Initially, we will do inference assuming $\sigma_1 \neq \sigma_2$; we will then repeat the analysis assuming $\sigma_1 = \sigma_2$. The R commands can be found below. Copy, paste, and print your R commands, output, and graphics. Answer the following:

1. Based upon the histograms, does the normal assumption seem (roughly) reasonable? Why?
2. Is the normal assumption needed to do inference for $\mu_1 - \mu_2$? Why?
3. Clearly mark $n_1, \bar{x}_1, s_1, n_2, \bar{x}_2, s_2$ on the output.
4. Analysis assuming $\sigma_1 \neq \sigma_2$:
   (a) Clearly mark Satterthwaite's degrees of freedom, ν, on the output.
   (b) Clearly mark the 95% CI for $\mu_1 - \mu_2$ on the output. Based on the CI, is there a significant difference in the mean octane levels? Why?
   (c) Clearly mark the test statistic t^* and $p-$value for the test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$. Based upon the $p-$value, is there a significant difference in the mean octane levels at the $\alpha = 0.05$ significance level? Why?
   (d) Based upon the $p-$value, is there a significant difference in the mean octane levels at the $\alpha = 0.10$ significance level? Why?
5. Analysis assuming $\sigma_1 = \sigma_2$:
   (a) Clearly mark the degrees of freedom, $n_1 + n_2 - 2$, on the output.
   (b) Clearly mark the 95% CI for $\mu_1 - \mu_2$ on the output. Based on the CI, is there a significant difference in the mean octane levels? Why?
   (c) Clearly mark the test statistic t^* and $p-$value for the test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$. Based upon the $p-$value, is there a significant difference in the mean octane levels at the $\alpha = 0.05$ significance level? Why?
   (d) Based upon the $p-$value, is there a significant difference in the mean octane levels at the $\alpha = 0.01$ significance level? Why?
6. Which analysis seems more appropriate: assuming $\sigma_1 = \sigma_2$ or assuming $\sigma_1 \neq \sigma_2$? Why?

R Commands

Input the data

Open *R* (see `R-lab1.pdf` on our website if you need to refresh your memory). Load the Brand 1 data into an object called `x1`, and Brand 2 data into an object called `x2`.

```
x1 <- c(87.8, 88.1, 87.0, 87.3, 87.8, 87.5, 87.8, 87.2, 86.2, 87.1, 86.9, 86.8, 87.1)
x2 <- c(86.2, 87.4, 87.7, 87.9, 88.2, 86.9, 87.4, 87.8, 89.3, 88.7, 88.5, 87.6)
```

You can see the data inside of `x1` and `x2` by typing their names.

```
x1
x2
```

Plot the data

Make a histogram of each dataset:

```
hist(x1)
hist(x2)
```

Summary statistics

Brand 1 — Find n_1, $\bar{x}_1$, and s_1:

```
length(x1)
mean(x1)
sd(x1)
```

Brand 2 — Find n_2, $\bar{x}_2$, and s_2:
Fill in these commands yourself.

Inference for $\mu_1 - \mu_2$ assuming $\sigma_1 \neq \sigma_2$

Test $H_0 : \mu_1 = \mu_2$ vs $H_a : \mu_1 \neq \mu_2$ at the $\alpha = 0.05$ significance level (and obtain a 95% CI for $\mu_1 - \mu_2$) assuming $\sigma_1 \neq \sigma_2$:

```
t.test(x1, x2, var.equal=FALSE)
```

Inference for $\mu_1 - \mu_2$ assuming $\sigma_1 = \sigma_2$

Test $H_0 : \mu_1 = \mu_2$ vs $H_a : \mu_1 \neq \mu_2$ at the $\alpha = 0.05$ significance level (and obtain a 95% CI for $\mu_1 - \mu_2$) assuming $\sigma_1 = \sigma_2$:

```
t.test(x1, x2, var.equal=TRUE)
```

Quit *R*

To quit *R*:

```
q()
```