

Chapter 6 Exercise

Sherry WeMott

2020-05-12

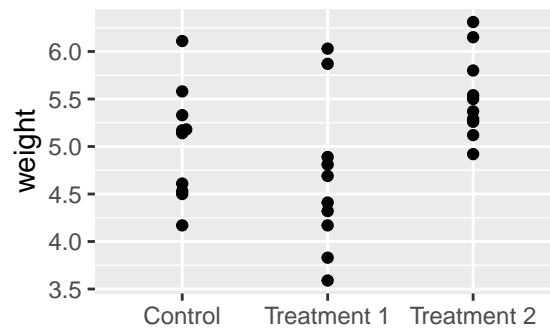
##Section 6.5 and Exercise 6.4

```
data("PlantGrowth")
PlantGrowth
```

```
##   weight group
## 1    4.17  ctrl
## 2    5.58  ctrl
## 3    5.18  ctrl
## 4    6.11  ctrl
## 5    4.50  ctrl
## 6    4.61  ctrl
## 7    5.17  ctrl
## 8    4.53  ctrl
## 9    5.33  ctrl
## 10   5.14  ctrl
## 11   4.81 trt1
## 12   4.17 trt1
## 13   4.41 trt1
## 14   3.59 trt1
## 15   5.87 trt1
## 16   3.83 trt1
## 17   6.03 trt1
## 18   4.89 trt1
## 19   4.32 trt1
## 20   4.69 trt1
## 21   6.31 trt2
## 22   5.12 trt2
## 23   5.54 trt2
## 24   5.50 trt2
## 25   5.37 trt2
## 26   5.29 trt2
## 27   4.92 trt2
## 28   6.15 trt2
## 29   5.80 trt2
## 30   5.26 trt2
```

```
PlantGrowth %>%
  mutate(group = fct_recode(group,
                             Control = "ctrl",
                             `Treatment 1` = "trt1",
                             `Treatment 2` = "trt2")) %>%
  ggplot(aes(x = group, y = weight)) +
  geom_beeswarm() +
```

```
labs(x = "")
```



```
PlantGrowth %>%
  filter(group %in% c("ctrl", "trt2")) %>%
  t.test(weight ~ group, data = .)

##
## Welch Two Sample t-test
##
## data: weight by group
## t = -2.134, df = 16.786, p-value = 0.0479
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.98287213 -0.00512787
## sample estimates:
## mean in group ctrl mean in group trt2
## 5.032 5.526
```

```
library("broom")
PlantGrowth %>%
  filter(group %in% c("ctrl", "trt2")) %>%
  t.test(weight ~ group, data = .) %>%
  tidy()
```

```
## # A tibble: 1 x 10
##   estimate estimate1 estimate2 statistic p.value parameter conf.low conf.high
##   <dbl>     <dbl>     <dbl>     <dbl>   <dbl>     <dbl>     <dbl>     <dbl>
## 1  -0.494      5.03      5.53     -2.13  0.0479     16.8    -0.983    -0.00513
## # ... with 2 more variables: method <chr>, alternative <chr>
```

When data is duplicated we get much smaller p-value due to increased sample size:

```
PlantGrowth %>%
  bind_rows(PlantGrowth) %>% # Add the duplicate of the dataset
  filter(group %in% c("ctrl", "trt2")) %>%
  t.test(weight ~ group, data = .) %>%
  tidy()
```

```
## # A tibble: 1 x 10
##   estimate estimate1 estimate2 statistic p.value parameter conf.low conf.high
##   <dbl>     <dbl>     <dbl>     <dbl>   <dbl>     <dbl>     <dbl>     <dbl>
## 1  -0.494      5.03      5.53     -3.10  0.00377     35.4    -0.817    -0.171
## # ... with 2 more variables: method <chr>, alternative <chr>
```

Resample only half the data:

```
PlantGrowth %>%
  sample_frac(size = 0.5) %>%
  bind_rows(., .) %>%
  filter(group %in% c("ctrl", "trt2")) %>%
  t.test(weight ~ group, data = .) %>%
  tidy()
```

```
## # A tibble: 1 x 10
##   estimate estimate1 estimate2 statistic p.value parameter conf.low conf.high
##   <dbl>      <dbl>      <dbl>      <dbl>   <dbl>      <dbl>      <dbl>      <dbl>
## 1  0.0382      5.21      5.17      0.246   0.809      17.4     -0.289      0.366
## # ... with 2 more variables: method <chr>, alternative <chr>
```

Resample half the data and add random noise:

```
pg1 <- PlantGrowth %>%
  sample_frac(size = 0.5)
pg2 <- pg1 %>%
  mutate(noise = rnorm(15, mean = 0, sd = 0.2),
         weight = weight + noise) %>%
  select(-noise)

pg1 %>%
  bind_rows(pg2) %>%
  filter(group %in% c("ctrl", "trt2")) %>%
  t.test(weight ~ group, data = .) %>%
  tidy()
```

```
## # A tibble: 1 x 10
##   estimate estimate1 estimate2 statistic p.value parameter conf.low conf.high
##   <dbl>      <dbl>      <dbl>      <dbl>   <dbl>      <dbl>      <dbl>      <dbl>
## 1 -0.404      4.95      5.35     -1.61   0.129      14.1     -0.940      0.132
## # ... with 2 more variables: method <chr>, alternative <chr>
```