MIE237

Neil Montgomery

2016-01-15

Inference with two independent numerical samples (9.8 and 10.5) continued…

"Key fact" equal variance version

Key fact:

$$\frac{\left(\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot}\right) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

9.40 continues

```
library(rio)
nitro <- import("Ex09.40.txt")
t.test(nitro$NoNitrogen, nitro$Nitrogen, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  nitro$NoNitrogen and nitro$Nitrogen
## t = -2.6191, df = 18, p-value = 0.01739
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.29915788 -0.03284212
## sample estimates:
## mean of x mean of y
##     0.399     0.565
```

9.40 with data in "real" form

```
library(tidyr)
nitro_tidy <- gather(nitro, treatment, weight)
```

Side-by-side boxplots (only n=10 each!)

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
nitro_tidy %>%
  ggplot(aes(y=weight, x=treatment)) + geom_boxplot()
```

Summaries

```
nitro_tidy %>%
  group_by(treatment) %>%
  summarize(n(), mean(weight), sd(weight))
```

```
## Source: local data frame [2 x 4]
##
##    treatment   n() mean(weight) sd(weight)
##       (fctr) (int)        (dbl)      (dbl)
## 1 NoNitrogen    10        0.399 0.07279347
## 2   Nitrogen    10        0.565 0.18674106
```

`t.test` again

```
nitro_tidy %>%
  t.test(weight ~ treatment, data=., var.equal = TRUE) -> nitro_t
nitro_t
```

```
##
##  Two Sample t-test
##
## data:  weight by treatment
## t = -2.6191, df = 18, p-value = 0.01739
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.29915788 -0.03284212
## sample estimates:
## mean in group NoNitrogen   mean in group Nitrogen
##                    0.399                    0.565
```

Note: `weight ~ treatment` is called a in R.

Details for hand calculation

```
## Source: local data frame [2 x 4]
##
##    treatment     n  mean         sd
##       (fctr) (int) (dbl)      (dbl)
## 1 NoNitrogen    10 0.399 0.07279347
## 2   Nitrogen    10 0.565 0.18674106
```

$s_p = 0.1503205$

$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 0.4472136$

$t_{18,0.025} = -2.100922$

$P(t_{18} < -2.6190945) = 0.0086932$

"Key Fact" no equal variance assumption

Key fact(actually an approximation):

$$\frac{\left(\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot}\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_\nu$$

with:

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

9.40 revisited

```
nitro_tidy %>%
  t.test(weight ~ treatment, data = .)
```

```
##
##  Welch Two Sample t-test
##
## data:  weight by treatment
## t = -2.6191, df = 11.673, p-value = 0.02286
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.30452438 -0.02747562
## sample estimates:
## mean in group NoNitrogen   mean in group Nitrogen
##                    0.399                    0.565
```