More fun with t-tests in R

Avalon C.S. Owens, Eric R. Scott

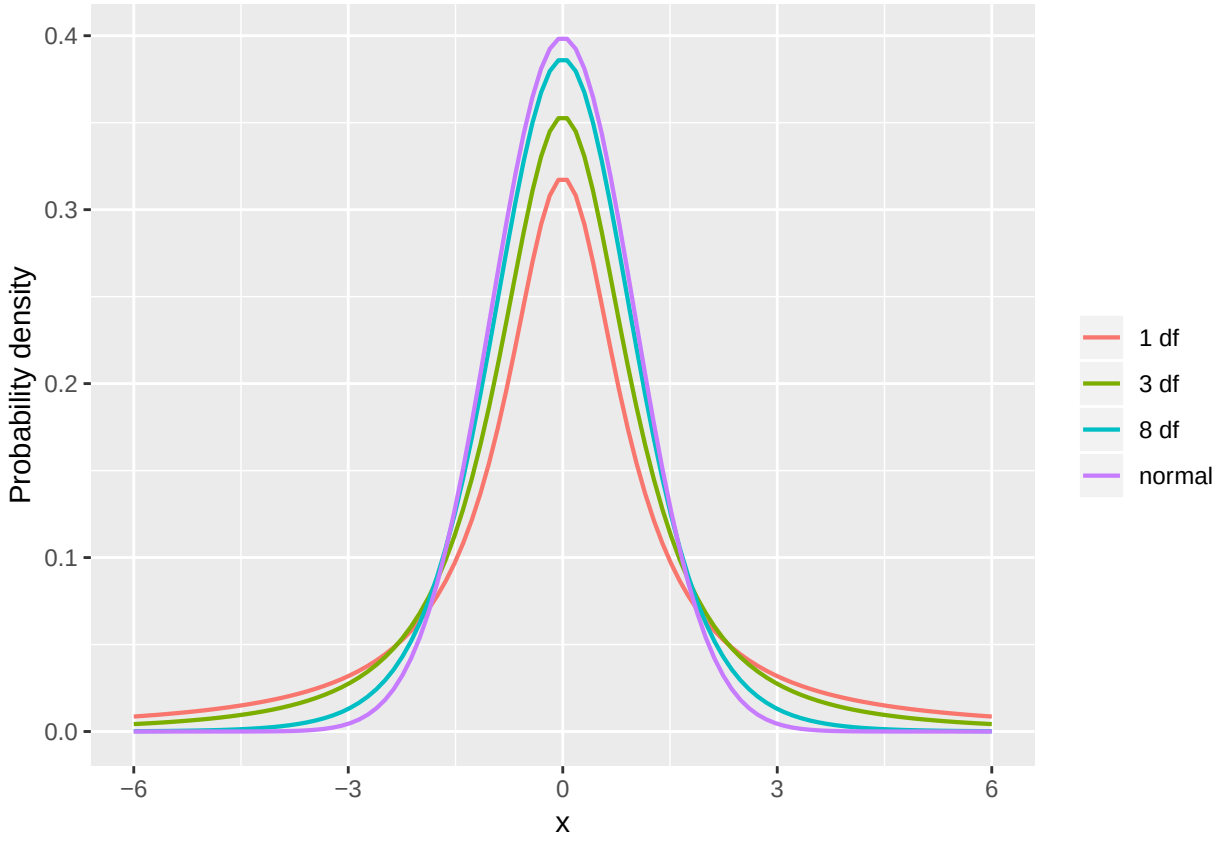
10/12/2018

Plan for today

- Performing paired t-tests in R
- Plotting paired data with `ggplot2`

Doing paired t-tests in R

The t-distribution



Data for today

Data for today

```
head(beaver1, 4)
```

```
##   day time  temp activ
## 1 346  840 36.33     0
```

```
## 2 346  850 36.34     0
## 3 346  900 36.35     0
## 4 346  910 36.42     0
```

```
head(beaver2, 4)
```

```
##   day time  temp activ
## 1 307  930 36.58     0
## 2 307  940 36.73     0
## 3 307  950 36.93     0
## 4 307 1000 37.15     0
```

One-sample t-test

$$t = \frac{\bar{Y} - \mu_0}{SE_{\bar{Y}}}$$

μ_0 = population mean (under H_0)

$\bar{Y}$ = sample mean

$SE_{\bar{Y}}$ = sample SE

One-sample t-test

H_0: $\mu = 37$

```
t.test(beaver1$temp, mu = 37)
```

```
##
##  One Sample t-test
##
## data:  beaver1$temp
## t = -7.6071, df = 113, p-value = 9.038e-12
## alternative hypothesis: true mean is not equal to 37
## 95 percent confidence interval:
##  36.82630 36.89808
## sample estimates:
## mean of x
##  36.86219
```

Two-tailed one-sample t-test

H_0: $\mu = 37$

```
t.test(beaver1$temp, mu = 37, alternative = "two.sided") #the default
```

```
##
##  One Sample t-test
##
## data:  beaver1$temp
## t = -7.6071, df = 113, p-value = 9.038e-12
## alternative hypothesis: true mean is not equal to 37
## 95 percent confidence interval:
##  36.82630 36.89808
```

```
## sample estimates:
## mean of x
##  36.86219
```

One-tailed one-sample t-test

H_0: $\mu \geq 37$

```
t.test(beaver1$temp, mu = 37, alternative = "less")
```

```
##
##  One Sample t-test
##
## data:  beaver1$temp
## t = -7.6071, df = 113, p-value = 4.519e-12
## alternative hypothesis: true mean is less than 37
## 95 percent confidence interval:
##      -Inf 36.89224
## sample estimates:
## mean of x
##  36.86219
```

One-tailed one-sample t-test

H_0: $\mu \leq 37$

```
t.test(beaver1$temp, mu = 37, alternative = "greater")
```

```
##
##  One Sample t-test
##
## data:  beaver1$temp
## t = -7.6071, df = 113, p-value = 1
## alternative hypothesis: true mean is greater than 37
## 95 percent confidence interval:
##  36.83215      Inf
## sample estimates:
## mean of x
##  36.86219
```

Independent two-sample t-test

$$t = \frac{(\bar{Y_1} - \bar{Y_2})}{SE_{\bar{Y_1} - \bar{Y_2}}}$$

Independent two-sample t-test

H_0: $\mu_1 = \mu_2$

```
t.test(beaver1$temp, beaver2$temp)
```

```
##
##  Welch Two Sample t-test
##
## data:  beaver1$temp and beaver2$temp
## t = -15.235, df = 131.12, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8298806 -0.6391334
## sample estimates:
## mean of x mean of y
##  36.86219  37.59670
```

Independent two-sample t-test

H_0: $\mu_1 = \mu_2$

```
t.test(beaver1$temp, beaver2$temp, var.equal = FALSE) #the default
```

```
##
##  Welch Two Sample t-test
##
## data:  beaver1$temp and beaver2$temp
## t = -15.235, df = 131.12, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8298806 -0.6391334
## sample estimates:
## mean of x mean of y
##  36.86219  37.59670
```

Independent two-sample t-test

H_0: $\mu_1 = \mu_2$

```
t.test(beaver1$temp, beaver2$temp, var.equal = TRUE) #equal variance
```

```
##
##  Two Sample t-test
##
## data:  beaver1$temp and beaver2$temp
## t = -15.937, df = 212, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8253593 -0.6436548
## sample estimates:
## mean of x mean of y
##  36.86219  37.59670
```

Independent two-sample t-test

H_0: $\mu_1 = \mu_2$

```r
t.test(temp ~ activ, data = beaver1) #formula interface
```

```
##
##  Welch Two Sample t-test
##
## data:  temp by activ
## t = -5.4346, df = 5.6263, p-value = 0.001978
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.5556401 -0.2067673
## sample estimates:
## mean in group 0 mean in group 1
##        36.84213        37.22333
```

REVIEW PART 1: TIDY DATA

REVIEW PART 1: TIDY DATA

```r
beaver_one <- beaver1 %>%
  add_column("beaver_ID" = 1)

beaver_two <- beaver2 %>%
  add_column("beaver_ID" = 2)

beavers <- bind_rows(beaver_one, beaver_two)

str(beavers)
```

```
## 'data.frame':    214 obs. of  5 variables:
##  $ day      : num  346 346 346 346 346 346 346 346 346 346 ...
##  $ time     : num  840 850 900 910 920 930 940 950 1000 1010 ...
##  $ temp     : num  36.3 36.3 36.4 36.4 36.5 ...
##  $ activ    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ beaver_ID: num  1 1 1 1 1 1 1 1 1 1 ...
```

Independent two-sample t-test

H_0: $\mu_1 = \mu_2$

```r
t.test(temp ~ beaver_ID, data = beavers) #formula interface
```

```
##
##  Welch Two Sample t-test
##
## data:  temp by beaver_ID
## t = -15.235, df = 131.12, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8298806 -0.6391334
## sample estimates:
## mean in group 1 mean in group 2
##        36.86219        37.59670
```

Independent two-sample t-test

H_0: $\mu_1 = \mu_2$

```
t.test(beaver1$temp, beaver2$temp) #x, y interface
```

```
##
##  Welch Two Sample t-test
##
## data:  beaver1$temp and beaver2$temp
## t = -15.235, df = 131.12, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8298806 -0.6391334
## sample estimates:
## mean of x mean of y
##  36.86219  37.59670
```

Paired-sample t-test

$$t = \frac{\bar{d} - \mu_{d0}}{SE_{\bar{d}}}$$

Paired-sample t-test

```
head(beaver1, 4)
```

```
##   day time  temp activ
## 1 346  840 36.33     0
## 2 346  850 36.34     0
## 3 346  900 36.35     0
## 4 346  910 36.42     0
```

```
head(beaver2, 4)
```

```
##   day time  temp activ
## 1 307  930 36.58     0
## 2 307  940 36.73     0
## 3 307  950 36.93     0
## 4 307 1000 37.15     0
```

Are these data paired?

Paired-sample t-test

```
t.test(beaver1$temp, beaver2$temp, paired = TRUE)
```

```
## Error in complete.cases(x, y): not all arguments have the same length
```

Paired-sample t-test

```
head(beaver1, 4)
```

```
##   day time  temp activ
## 1 346  840 36.33     0
## 2 346  850 36.34     0
## 3 346  900 36.35     0
## 4 346  910 36.42     0
```

```
tail(beaver1, 4)
```

```
##     day time  temp activ
## 111 347  310 36.88     0
## 112 347  320 36.93     0
## 113 347  330 36.97     0
## 114 347  340 37.15     1
```

Are these data paired?

Paired-sample t-test

```
t.test(beaver1$temp ~ beaver1$activ, paired = TRUE)
```

```
## Error in complete.cases(x, y): not all arguments have the same length
```

ALSO: The formula interface is ONLY FOR 2-sample t-tests

More data for today

More data for today

```
library(PairedData)
data(HorseBeginners)
HorseBeginners
```

```
##   Subject Actual Imaginary
## 1      S1  69.64     66.58
## 2      S2  62.26     25.59
## 3      S3  78.63     24.01
## 4      S4  76.00     38.35
## 5      S5  60.10     12.19
## 6      S6  68.51     34.25
## 7      S7  69.57      5.68
## 8      S8  74.48     15.02
```

Are these data paired?

Paired sample t-test

```
t.test(HorseBeginners$Actual, HorseBeginners$Imaginary, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  HorseBeginners$Actual and HorseBeginners$Imaginary
## t = 6.19, df = 7, p-value = 0.0004496
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  26.07319 58.30681
## sample estimates:
## mean of the differences
##                   42.19
```

Back to the help file

Back to the help file

```
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

Plotting paired data with ggplot2

Plotting paired data with ggplot2

```
HorseBeginners
```

```
##   Subject Actual Imaginary
## 1      S1  69.64     66.58
## 2      S2  62.26     25.59
## 3      S3  78.63     24.01
## 4      S4  76.00     38.35
## 5      S5  60.10     12.19
## 6      S6  68.51     34.25
## 7      S7  69.57      5.68
## 8      S8  74.48     15.02
```

Are these data tidy?

REVIEW PART 2: TIDY DATA

REVIEW PART 2: TIDY DATA

```
horse_tidy <- gather(HorseBeginners,
                     key = "Status", value = "Time")
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

REVIEW PART 2: TIDY DATA

```
head(horse_tidy, 14)
```

```
##     Status  Time
## 1  Subject    S1
## 2  Subject    S2
## 3  Subject    S3
## 4  Subject    S4
## 5  Subject    S5
## 6  Subject    S6
## 7  Subject    S7
## 8  Subject    S8
## 9   Actual 69.64
## 10  Actual 62.26
## 11  Actual 78.63
## 12  Actual    76
## 13  Actual  60.1
## 14  Actual 68.51
```

REVIEW PART 2: TIDY DATA

```
horse_tidy <- gather(HorseBeginners, -Subject,
                     key = "Status", value = "Time")
head(horse_tidy, 12)
```

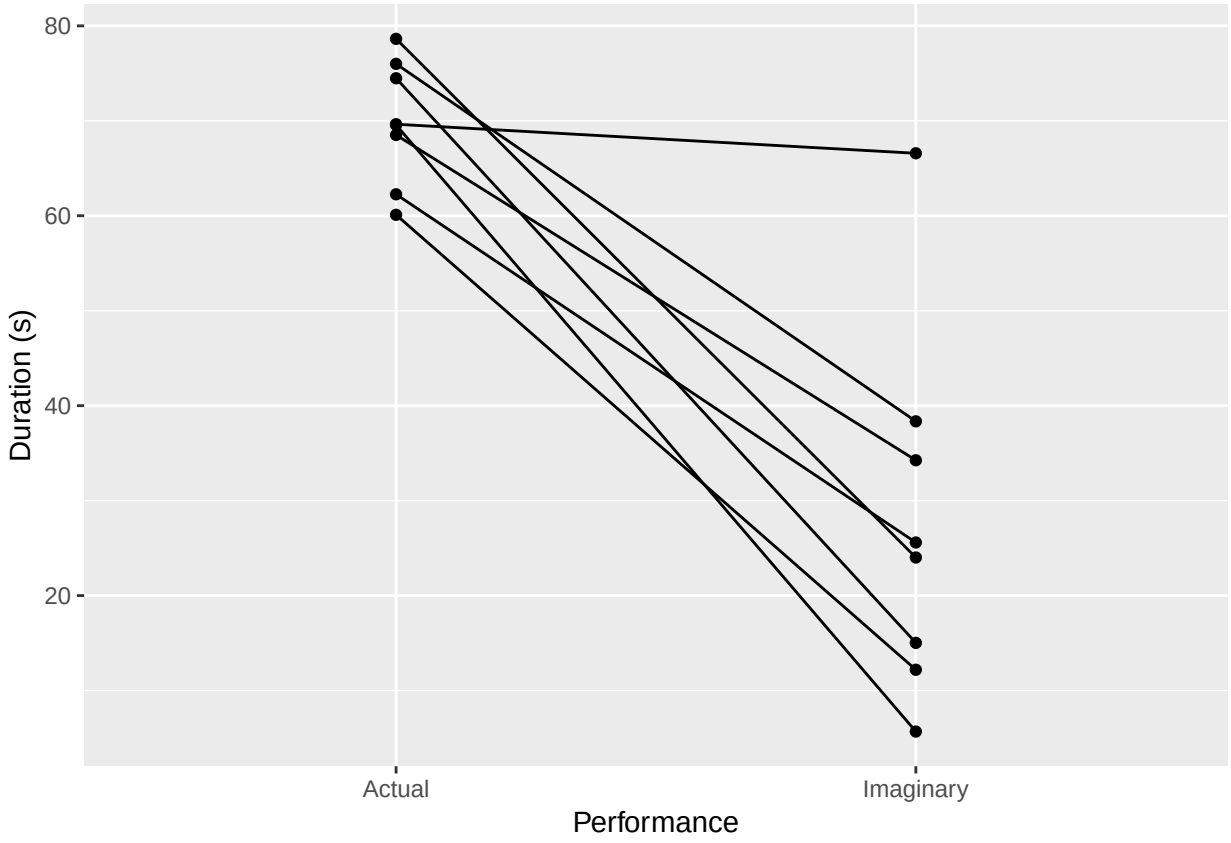
```
##    Subject    Status  Time
## 1       S1    Actual 69.64
## 2       S2    Actual 62.26
## 3       S3    Actual 78.63
## 4       S4    Actual 76.00
## 5       S5    Actual 60.10
## 6       S6    Actual 68.51
## 7       S7    Actual 69.57
## 8       S8    Actual 74.48
## 9       S1 Imaginary 66.58
## 10      S2 Imaginary 25.59
## 11      S3 Imaginary 24.01
## 12      S4 Imaginary 38.35
```

Plotting paired data with ggplot2

```
horse_plot <- ggplot(horse_tidy, aes(x = Status,
                                     y = Time,
                                     group = Subject)) +
  geom_point() + geom_line() +
  labs(x = "Performance", y = "Duration (s)")
```

Plotting paired data with ggplot2

```
horse_plot
```



Plotting paired data with ggplot2

```
head(horse_tidy, 3)
```

```
##   Subject Status  Time
## 1      S1 Actual 69.64
## 2      S2 Actual 62.26
## 3      S3 Actual 78.63
```

```
horse_tidy$Status <- factor(horse_tidy$Status,
                            levels = c("Imaginary", "Actual"))
head(horse_tidy, 3)
```

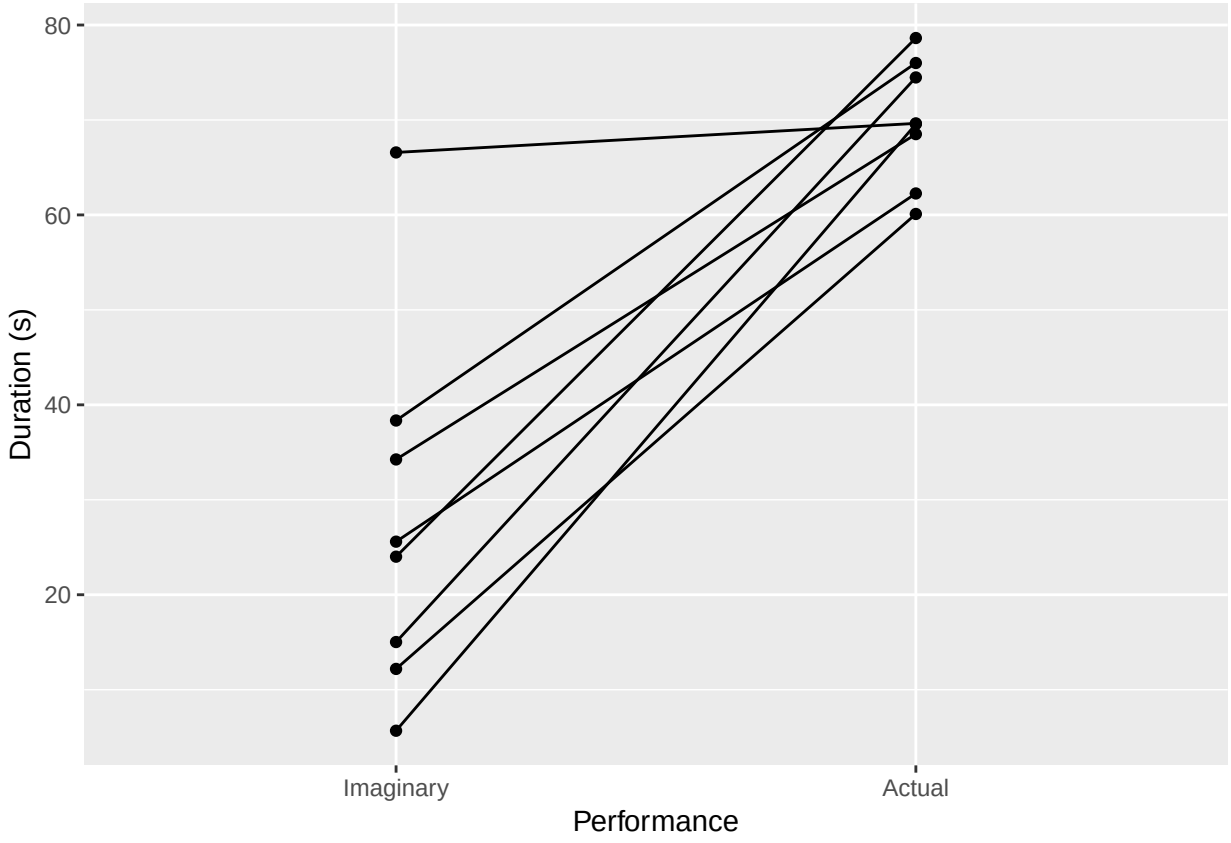
```
##   Subject Status  Time
## 1      S1 Actual 69.64
## 2      S2 Actual 62.26
## 3      S3 Actual 78.63
```

Plotting paired data with ggplot2

```
horse_plot <- ggplot(horse_tidy, aes(x = Status,
                                     y = Time,
                                     group = Subject)) +
  geom_point() + geom_line() +
  labs(x = "Performance", y = "Duration (s)")
```

Plotting paired data with ggplot2

```
horse_plot
```



Errors vs. warnings

Piazza

Homework time!