t_test

Jiao Lai

4/4/2021

This file documents how to conduct three different kinds of t test using R: paired t test, unpaired t test with equal variance, and unpaired t test with unequal variance. And assumptions, how to check assumptions for these tests.

1. Unpaired t test with equal variance. Assumptions: (1) The two samples are independent. (2) The two groups of samples (A and B), being compared, are normally distributed. Can be checked using Shapiro-Wilk test. (3) The variance of the two groups are equivalent, and hence pooled sample variance is used to calculate t statistic. This can be checked using F-test. R code: t.test(x1, x2 , alternative = 'two.sided', var.equal = TRUE)

Data simulation:

```
set.seed(123)
A <- rnorm(20)
B <- rnorm(20)
dat <- data.frame(group = rep(c('A', 'B'), each = 20),
                  x = c(A, B))
```

Compute summary statistics by groups.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

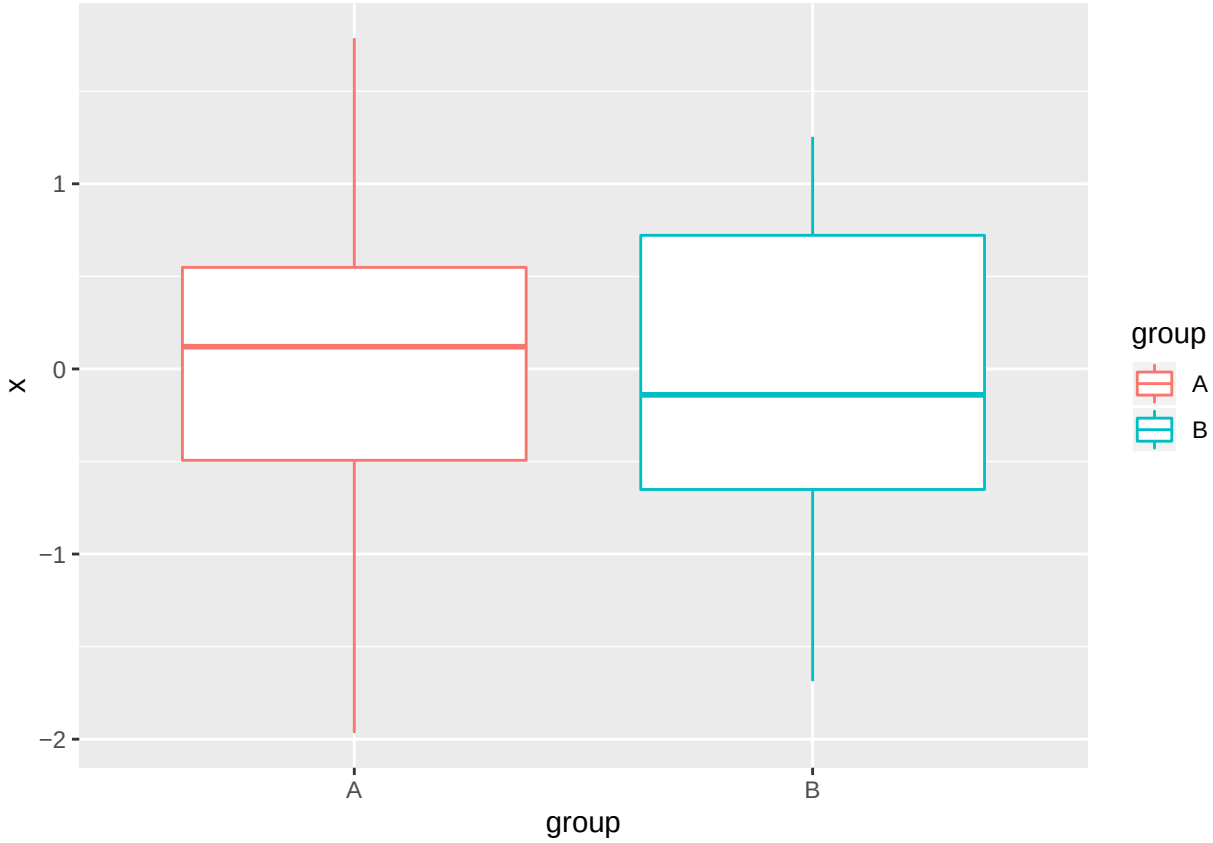
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
dat %>% group_by(group) %>%
  summarize(count = n(),
            mean = mean(x),
            sd = sd(x)
            )
```

```
## # A tibble: 2 x 4
##   group count    mean    sd
##   <fct> <int>   <dbl> <dbl>
## 1 A        20  0.142  0.973
## 2 B        20 -0.0513 0.830
```

Visualize summary data.

```
library(ggplot2)
ggplot(data = dat, aes(x = group, y = x, color = group))+
  geom_boxplot()
```



Test for normality.

```
with(dat, shapiro.test(x[group == 'A']))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x[group == "A"]
## W = 0.96858, p-value = 0.7247
```

```
with(dat, shapiro.test(x[group == 'B']))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x[group == "B"]
## W = 0.95594, p-value = 0.4663
```

Test for homogeneity in variances.

```
var.test(x ~ group, data = dat)
```

```
##
##  F test to compare two variances
##
## data:  x by group
```

```
## F = 1.3735, num df = 19, denom df = 19, p-value = 0.4957
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.5436557 3.4701298
## sample estimates:
## ratio of variances
##            1.37352
```

Independent sample t test.

```
# method 1
unpair_1 = t.test(x ~ group, data = dat, var.equal = TRUE)
# method 2
unpair_2 = t.test(A, B, var.equal = TRUE)
```

Save and extract test result.

```
library(broom)
unpair_tidy = tidy(unpair_1)
unpair_tidy$p.value
```

```
## [1] 0.5039985
```

2. Unpaired t test with un-equal variance. Assumptions: (1) The two samples are independent. (2) The two groups of samples (A and B), being compared, are normally distributed. Can be checked using Shapiro-Wilk test. (3) The variance of the two groups are inequivalent. This can be checked using F-test. R code: t.test(x1, x2 , alternative = 'two.sided', var.equal = FALSE)

```
t.test(x ~ group, data = dat, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  x by group
## t = 0.67462, df = 37.082, p-value = 0.5041
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3863821  0.7721440
## sample estimates:
## mean in group A mean in group B
##      0.14162380     -0.05125716
```

3. Paired t test. Assumptions: (1) The two samples are correlated. For example, the same group of subjects are observed at different time. (2) The two groups of samples (A and B), being compared, are normally distributed. Can be checked using Shapiro-Wilk test. R code: t.test(x1, x2 , alternative = 'two.sided', paired = TRUE) Cannot use formula for paired t test, instead, we have to specify the variables we are comparing. Make sure the data are paired correctly, and the sample sizes has to be the same.

```
t.test(A, B, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  A and B
## t = 0.646, df = 19, p-value = 0.526
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.4320476  0.8178095
## sample estimates:
## mean of the differences
##                0.192881
```