R Stats Club

t-test, ANOVA and linear regression

Damien DUPRE

27/03/2019

Statistical tests

test	DV	IV	baseR
t-test	1 continuous	1 categorical	t.test()
ANOVA	1 continuous	1+ categorical	aov()
linear regression	1 continuous	1+ continuous	lm()

Note: t-test and anova are special cases of linear models so they can be performed using the `lm()` function as well (see following slides)

Create a dataframe

```r
gender <- c(rep("male",50), rep("female",50))
# create an object gender which combines:
## the word male repeated 50 times
## and the word female repeated 50 times

education <- c(rep("second",33), rep("undergrad",33), rep("master",34))
# create an object education which combines:
## the word second repeated 33 times,
## the word undergrad repeated 33 times
## and the word master repeated 34 times

age <- sample(15:50, 100, replace = TRUE)
# create an object age which select a number between 15 and 50, 100 times

math_results <- rnorm(n = 100, mean = 50, sd = 25)
# create an object math_results which creates an array of

data_raw <- data.frame(gender, education, age, math_results)
# create a dataframe object with 4 columns: gender, education, age, math_results

str(data_raw) # check the structure of the dataframe
```

```
## 'data.frame':    100 obs. of  4 variables:
##  $ gender      : Factor w/ 2 levels "female","male": 2 2 2 2 2 2 2 2 2 2 ...
##  $ education   : Factor w/ 3 levels "master","second",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ age         : int  25 43 29 46 48 16 34 47 34 31 ...
##  $ math_results: num  56.3 49.3 48.9 84.2 44.4 ...
```

t-test with base R

t-test

```
res_ttest <- t.test(math_results ~ gender, data = data_raw)

res_ttest
```

```
##
##     Welch Two Sample t-test
##
## data:  math_results by gender
## t = -2.1106, df = 97.241, p-value = 0.03737
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -19.4181729  -0.5972622
## sample estimates:
## mean in group female   mean in group male
##             43.65249             53.66021
```

t-test with lm()

```
res_lm_ttest <- lm(math_results ~ gender, data = data_raw)

summary(res_lm_ttest)
```

```
##
## Call:
## lm(formula = math_results ~ gender, data = data_raw)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -61.389 -16.007  -0.386  13.046  58.850
##
## Coefficients:
##             Estimate Std. Error t value            Pr(>|t|)
## (Intercept)   48.656      2.371  20.523 <0.0000000000000002 ***
## gender1       -5.004      2.371  -2.111              0.0374 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.71 on 98 degrees of freedom
## Multiple R-squared:  0.04348,    Adjusted R-squared:  0.03372
## F-statistic: 4.455 on 1 and 98 DF,  p-value: 0.03735
```

ANOVA with base R

ANOVA

```
res_anova <- aov(math_results ~ gender + education, data = data_raw)

summary(res_anova)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## gender       1   2504    2504   4.479 0.0369 *
## education    2   1418     709   1.268 0.2860
## Residuals   96  53665     559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA with lm()

```
res_lm_anova <- lm(math_results ~ gender + education, data = data_raw)

# summary(res_lm_anova)

drop1(res_lm_anova, test = "F")
```

```
## Single term deletions
##
## Model:
## math_results ~ gender + education
##           Df Sum of Sq   RSS    AIC F value  Pr(>F)
## <none>                 53665 636.53
## gender     1    2811.5 56476 639.64  5.0295 0.02722 *
## education  2    1417.9 55083 635.14  1.2683 0.28599
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

linear regression with base R

linear regression

```
res_lm <- lm(math_results ~ age, data = data_raw)

summary(res_lm)
```

```
##
## Call:
## lm(formula = math_results ~ age, data = data_raw)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -55.995 -15.386  -0.476  14.891  55.464
##
## Coefficients:
##             Estimate Std. Error t value      Pr(>|t|)
## (Intercept)  50.5898     8.0486   6.286 0.00000000908 ***
## age          -0.0596     0.2366  -0.252         0.802
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.23 on 98 degrees of freedom
## Multiple R-squared:  0.0006471,    Adjusted R-squared:  -0.00955
## F-statistic: 0.06346 on 1 and 98 DF,  p-value: 0.8016
```

linear regression (main effects)

```
res_lm <- lm(math_results ~ age + gender, data = data_raw)

# summary(res_lm)

drop1(res_lm, test = "F")
```

```
## Single term deletions
##
## Model:
## math_results ~ age + gender
##        Df Sum of Sq   RSS    AIC F value  Pr(>F)
## <none>              54988 636.97
## age     1     94.97 55083 635.14  0.1675 0.68322
## gender  1   2561.56 57549 639.52  4.5187 0.03607 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

linear regression (interaction effect)

```
res_lm <- lm(math_results ~ age : gender, data = data_raw)

# summary(res_lm)

drop1(res_lm, test = "F")
```

```
## Single term deletions
##
## Model:
## math_results ~ age:gender
##            Df Sum of Sq   RSS    AIC F value Pr(>F)
## <none>                  55480 637.86
## age:gender  2    2106.4 57586 637.59  1.8414 0.1641
```

linear regression (main and interaction effects)

```
res_lm <- lm(math_results ~ age * gender, data = data_raw)

summary(res_lm)
```

```
##
## Call:
## lm(formula = math_results ~ age * gender, data = data_raw)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -60.34 -15.33  -0.51  13.96  58.85
##
## Coefficients:
##             Estimate Std. Error t value      Pr(>|t|)
## (Intercept) 51.64498    7.96649   6.483 0.00000000388 ***
## age         -0.09006    0.23450  -0.384         0.702
## gender1     -8.00625    7.96649  -1.005         0.317
## age:gender1  0.09050    0.23450   0.386         0.700
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.91 on 96 degrees of freedom
## Multiple R-squared:  0.04661,    Adjusted R-squared:  0.01681
## F-statistic: 1.564 on 3 and 96 DF,  p-value: 0.2031
```

statistics with tadaatoolbox

```
# install.packages("tadaatoolbox") # only once

library(tadaatoolbox)
```

t-test with tadaatoolbox

```
tadaa_t.test(data = data_raw, response = math_results, group = gender, print = "markdown")
```

Table 3: **Two Sample t-test** with alternative hypothesis: $\mu_1 \neq \mu_2$

Diff	μ_1 **female**	μ_2 **male**	**t**	**SE**	**df**	$CI_{95\%}$	**p**	**Cohen\'s d**	**Power**
-10.01	43.65	53.66	-2.11	4.74	98	(-19.42 - -0.6)	< .05	-0.42	0.55

ANOVA with tadaatoolbox

```
tadaa_aov(data = data_raw, math_results ~ gender + education, print = "markdown")
```

Table 4: **Two-Way ANOVA**: Using Type III Sum of Squares

Term	df	SS	MS	F	p	η^2_{part}	Cohen's f	Power
education	2	1417.94	708.97	1.27	.286	0.03	0.16	0.28
gender	1	2811.53	2811.53	5.03	< .05	0.05	0.23	0.61
Residuals	96	53664.68	559.01					
Total	99	57894.15	4079.51					

statistics with jmv

```
# install.packages("jmv") # only once

library(jmv)
```

t-test with jmv

```
ttestIS(formula = math_results ~ gender, data = data_raw)
```

```
##
##  INDEPENDENT SAMPLES T-TEST
##
##  Independent Samples T-Test
##  ---------------------------------------------------------------
##                                  statistic    df      p
##  ---------------------------------------------------------------
##    math_results    Student's t        -2.11    98.0    0.037
##  ---------------------------------------------------------------
```

ANOVA with jmv

```
ANOVA(formula = math_results ~ gender + education, data = data_raw)
```

```
##
##  ANOVA
##
##  ANOVA
##  ------------------------------------------------------------------------
##                  Sum of Squares    df    Mean Square    F       p
##  ------------------------------------------------------------------------
##    gender                  2812     1           2812    5.03    0.027
##    education               1418     2            709    1.27    0.286
##    Residuals              53665    96            559
##  ------------------------------------------------------------------------
```

Thank you for your attention

don't hesitate to send me an email: damien.dupre@dcu.ie