Paired and Independent t-test

Klinkenberg

26 sep 2017

Inhoud

- paired samples t test
  - calculate d
  - calculate t value
  - test for significance
  - effect size
  - confidence interval
- independent samples t test
  - calculate means
  - calculate variance
  - calculate pooled variance
  - calculate pooled se
  - calculate t value 2

Paired 2 samples

Paired-samples t-test

Paired-samples t-test

In the Paired samples t-test the deviation (D) for each pair is calculated and the mean of these deviations ($\bar{D}$) is tested against the null hypothesis where $\mu = 0$.

$$t_{n-1} = \frac{\bar{D} - \mu}{SE_D}$$

Where n (the number of cases) minus 1, are the degrees of freedom $df = n - 1$ and SE_D is the standard error of D, defined as $s_D/\sqrt{n}$.

Hypothesis

$$H_0 : \bar{D} = \mu_D$$
$$H_A : \bar{D} \neq \mu_D$$
$$H_A : \bar{D} > \mu_D$$
$$H_A : \bar{D} < \mu_D$$

Data structure

index	k1	k2
1	d	d
2	d	d
3	d	d
4	d	d

Where k is the level of the categorical predictor variabla and d is the value of the outcome/dependent variable.

Data example

We are going to use the IQ estimates we collected last week. You had to gues the IQ of the one sitting next to you and your own IQ.

Let's take a look at the data.

IQ estimates

	IQ.van.je.buur	Eigen.IQ
301	130	125
302	120	120
303	124	125
304	124	134
305	120	115
306	130	130
307	119	110
308	118	118
309	124	116

Calculate D

```
D = IQ.next.to.you - IQ.you
```

	IQ.van.je.buur	Eigen.IQ	D
301	130	125	5
302	120	120	0
303	124	125	-1
304	124	134	-10
305	120	115	5
306	130	130	0
307	119	110	9
308	118	118	0
309	124	116	8
310	125	125	0

Calculate $\bar{D}$

```
D      = na.omit(D) # get rid of all missing values
D.mean = mean(D)
D.mean
```

```
## [1] 2.416667
```

And we also need n.

```
n = length(D)
n
```

```
## [1] 48
```

Calculate t-value

$$t_{n-1} = \frac{\bar{D} - \mu}{SE_D}$$

```
mu = 0                  # Define mu

D.sd = sd(D)            # Calculate standard deviation
D.se = D.sd / sqrt(n) # Calculate standard error

df   = n - 1            # Calculate degrees of freedom

# Calculate t
t = ( D.mean - mu ) / D.se
t
```

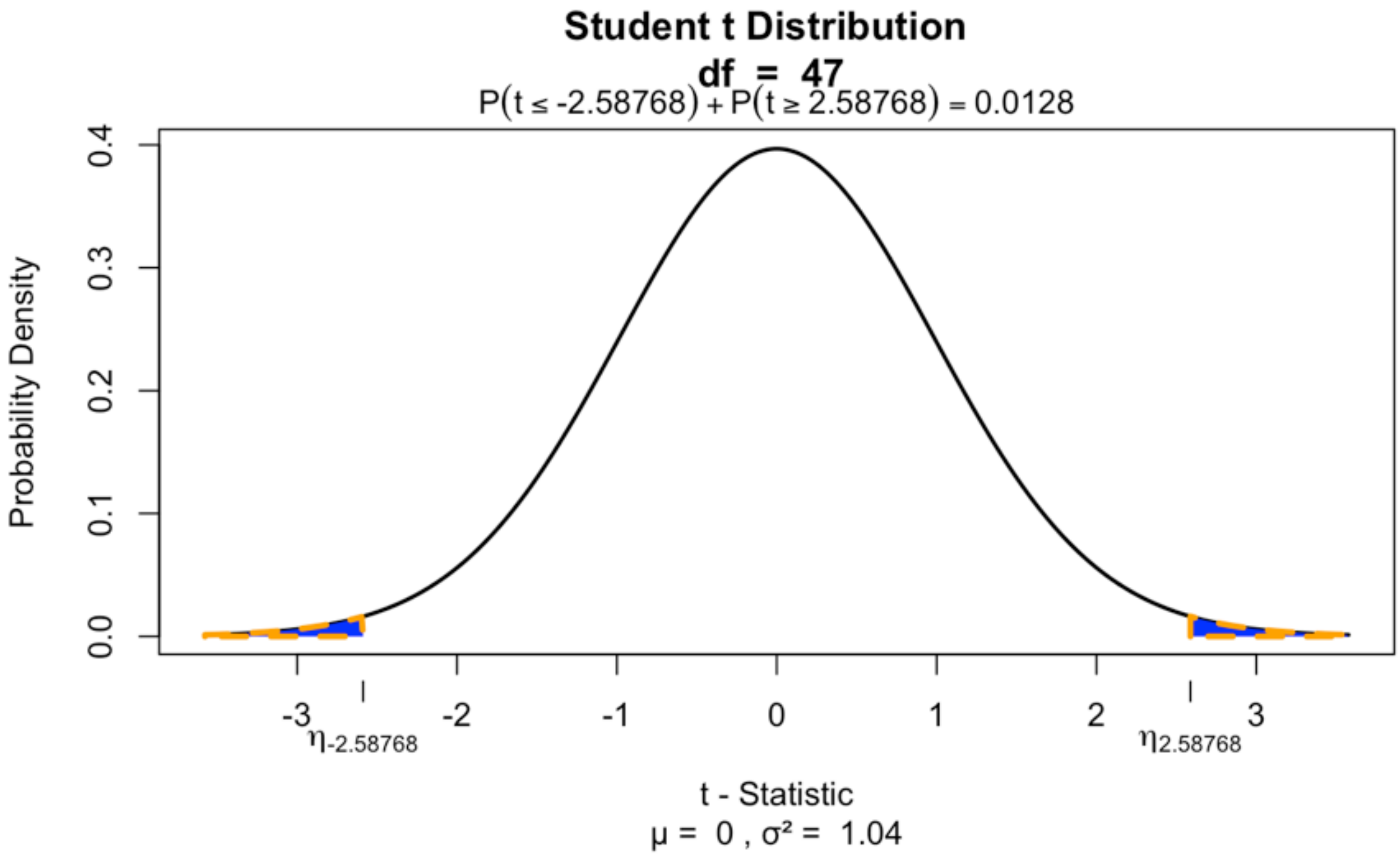
```
## [1] 2.58768
```

Test for significance

Two tailed

```r
if(!"visualize" %in% installed.packages()) { install.packages("visualize") }
library("visualize")

visualize.t(c(-t,t), df, section="tails")
```


Student t Distribution
df = 47
P(t ≤ -2.58768) + P(t ≥ 2.58768) = 0.0128
Probability Density
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
η-2.58768
η2.58768
t - Statistic
μ = 0 , σ² = 1.04

Effect-size

$$r = \sqrt{\frac{t^2}{t^2 + \mathrm{df}}}$$

```
r = sqrt(t^2/(t^2 + df))
r
```

```
## [1] 0.3531337
```

Confidence interval

To display correct conficance intervals in SPSS we need to correct the original scores for whithin subject variation.

```
** SPSS SYNTAX

COMPUTE personal_mean = MEAN(IQ.next.to.you, IQ.you).
EXECUTE.

AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES
  /BREAK=
  /total_mean = MEAN(personal_mean).

COMPUTE adjustment = total_mean - personal_mean.
EXECUTE.

COMPUTE IQ.next.to.you.adj = IQ.next.to.you + adjustment.
COMPUTE IQ.you             = IQ.you         + adjustment.
EXECUTE.
```

Compare 2 independent samples

Independent-samples t-test

Independent-samples t-test

In the independent-samples t-test the mean of both independent samples is calculated and the difference of these $(\bar{X}_1 - \bar{X}_2)$ means is tested against the null hypothesis where $\mu = 0$.

$$t_{n_1+n_2-2} = \frac{(\bar{X}_1 - \bar{X}_2) - \mu}{SE_p}$$

Where n_1 and n_2 are the number of cases in each group and SE_p is the pooled standard error.

Hypothesis

$$H_0 : t = 0 = \mu_t$$
$$H_A : t \neq 0$$
$$H_A : t > 0$$
$$H_A : t < 0$$

Data structure

index	k	outcome
1	1	d
2	1	d
3	2	d
4	2	d

Where k is the level of the categorical predictor variabla and d is the value of the outcome/dependent variable.

Additional assumption

Specificly for independent sample t-test.

- Equality of variance
  - H_0 : Variance = equal (p > .05)
  - H_A : Variance ≠ equal (p < .05)

Example

We are going to use the IQ estimates we collected last week again. You had to gues the IQ of the one sitting next to you and your own IQ. But we are going to add gender to the data set. We did not register this so we are going to simulate some genders.

```r
gender = sample(c("male", "female"), dim(data)[1], replace = TRUE)
```

The data

Calculate means

```
IQ.you.male   = subset(data, gender == "male",   select = IQ.you)$IQ.you
IQ.you.female = subset(data, gender == "female", select = IQ.you)$IQ.you

IQ.you.male.mean   = mean(IQ.you.male,   na.rm = T)
IQ.you.female.mean = mean(IQ.you.female, na.rm = T)

rbind(IQ.you.male.mean, IQ.you.female.mean)
```

```
##                        [,1]
## IQ.you.male.mean   118.9821
## IQ.you.female.mean 123.6735
```

Calculate variance

```r
IQ.you.male.var   = var(IQ.you.male,   na.rm = T)
IQ.you.female.var = var(IQ.you.female, na.rm = T)
rbind(IQ.you.male.var, IQ.you.female.var)
```

```
##                       [,1]
## IQ.you.male.var   262.2360
## IQ.you.female.var 112.8912
```

```r
IQ.you.male.n   = length(IQ.you.male)   - 1
IQ.you.female.n = length(IQ.you.female) - 1
rbind(IQ.you.male.n, IQ.you.female.n)
```

```
##                 [,1]
## IQ.you.male.n     55
## IQ.you.female.n   49
```

Calculate t-value

$$t_{n_1+n_2-2} = \frac{(\bar{X}_1 - \bar{X}_2) - \mu}{SE_p}$$

Where SE_p is the pooled standard error.

$$SE_p = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

And S_p^2 is the pooled variance.

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Where s^2 is the variance and n the sample size.

Calculate pooled variance

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

```
df = IQ.you.male.n + IQ.you.female.n - 2

s2.p = ( (IQ.you.male.n-1)*IQ.you.male.var + (IQ.you.female.n-1)*IQ.you.female

df
```

```
## [1] 102
```

```
s2.p
```

Calculate pooled SE

$$SE_p = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

```
se.p    = sqrt( ((s2.p/IQ.you.male.n) + (s2.p/IQ.you.female.n)) )
se.p
```

```
## [1] 2.721687
```

Calculate t-value

$$t_{n_1+n_2-2} = \frac{(\bar{X}_1 - \bar{X}_2) - \mu}{SE_p}$$

```
t = ( IQ.you.male.mean - IQ.you.female.mean ) / se.p

t
```

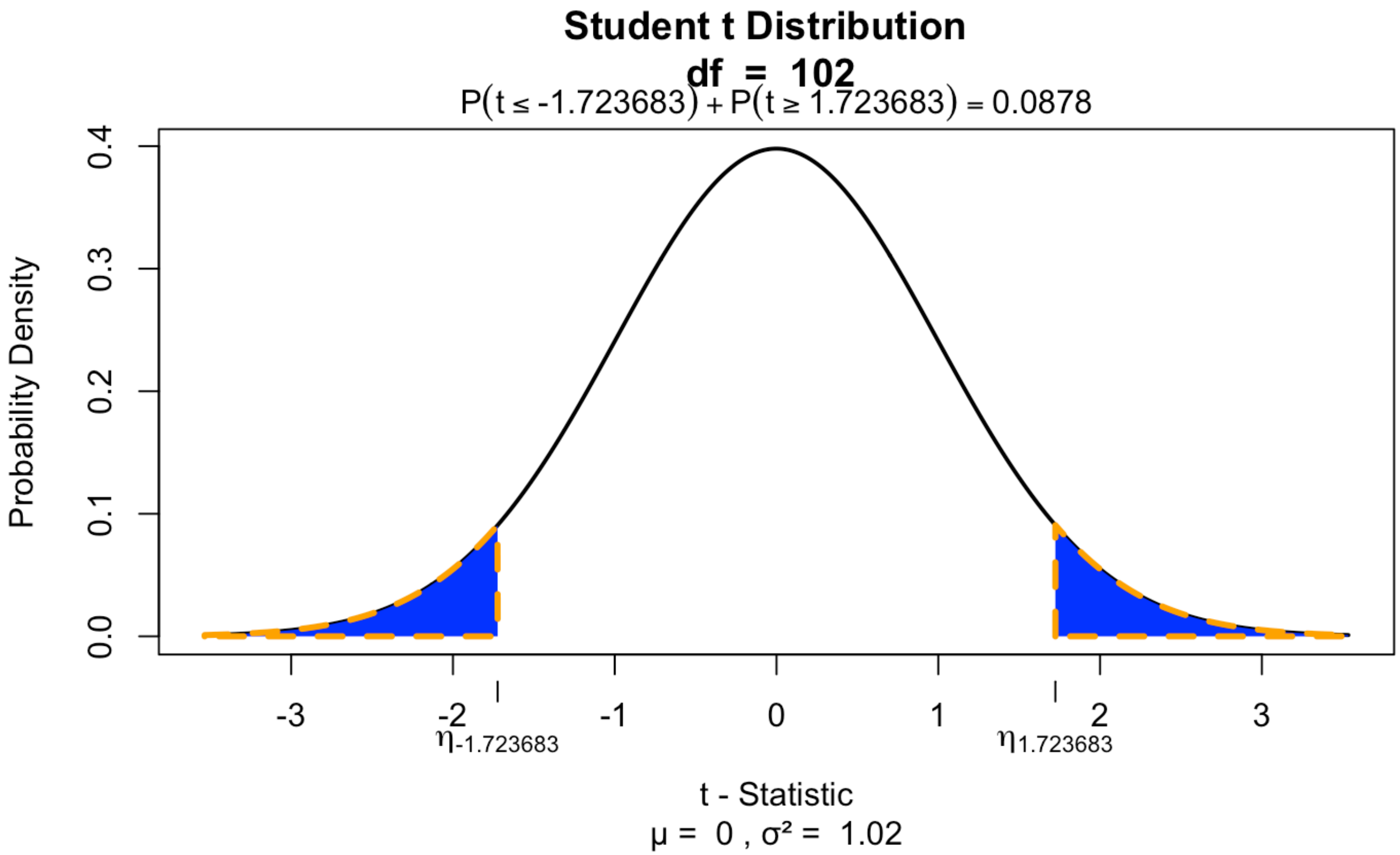
```
## [1] -1.723683
```

Test for significance

Two tailed

```
if(!"visualize" %in% installed.packages()) { install.packages("visualize") }
library("visualize")

visualize.t(c(-t,t), df, section="tails")
```



Student t Distribution
df = 102
P(t ≤ -1.723683) + P(t ≥ 1.723683) = 0.0878
Probability Density
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
η-1.723683
η1.723683
t - Statistic
μ = 0 , σ² = 1.02

Effect-size

$$r = \sqrt{\frac{t^2}{t^2 + \mathrm{df}}}$$

```
r = sqrt(t^2/(t^2 + df))

r
```

```
## [1] 0.1682374
```

- @shklinkenberg
- Klinkenberg
- S.Klinkenberg@UvA.nl
- ShKlinkenberg

END