Paired t-test

Data Analysis for Psychology in R 1

dapR1 Team

Department of Psychology
The University of Edinburgh

Learning Objectives

- Understand when to use an paired sample t-test
- Understand the null hypothesis for an paired sample t-test
- Understand how to calculate the test statistic
- Know how to conduct the test in R
- Know how to calculate Cohen's D for each form of t-test

Topics for today

- Recording 1: Conceptual background and introduction to our example
- Recording 2: Calculations and R-functions
- Recording 3: Assumptions and effect size

Purpose & Data

- The paired sample t-test is used when we want to test the difference in mean scores for a sample measured at two points in time.
    - Thus this is a first example of a repeated measures design.
- Data Requirements
    - A continuously measured variable.
    - A binary variable denoting time.

Example

- I want to assess whether a time-management course helps reduce exam stress in students.
- I ask 50 students to take a self-report stress measure during their winter exams.
- At the beginning of semester 2 they take a time management course.
- I then assess their self-report stress in the summer exam block.
    - Let's assume for the sake of this example that I have been able to control the volume and difficulty of the exams the students take in each block.

Data

```
## # A tibble: 6 × 3
##   ID    stress time
##   <chr>  <dbl> <fct>
## 1 ID1       14 t1
## 2 ID2        7 t1
## 3 ID3        8 t1
## 4 ID4        8 t1
## 5 ID5        7 t1
## 6 ID6        7 t1
```

Calculating difference

- In the paired t-test, we specifically calculate and analyse the difference in scores at time 1 and time 2 per participant.

$$d_i = x_{i1} - x_{i2}$$

Test statistic

- The resulting test statistic:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

- where:
  - $\bar{d}$ = mean of the individual difference scores (d_i)
  - s_d = standard deviation of the difference scores (d_i)
  - n = sample size

- The associated sampling distribution is a t-distributon with $n - 1$ degrees of freedom.
  - Note, this is just essentially a one sample test on the difference scores.

Hypotheses

- Two-tailed:

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d \neq 0$$

- One-tailed

$$H_1 : \mu_d < 0$$
$$H_1 : \mu_d > 0$$

Time for a break

Welcome Back!

Let's calculate a paried-sample t-test!

Our Example

- I elect to use a two-tailed test with alpha of .01
- I want to be quite sure the intervention has worked and stress levels have changed.
- So my hypotheses are:

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d \neq 0$$

Calculation

- Steps in my calculations:
  - Calculate the difference scores for individuals.
  - Calculate the mean of the difference scores.
  - Calculate the SD of the difference scores.
  - Check I know my N.
  - Calculate the standard error of the mean difference.
  - Use all this to calculate t
  - Calculate my degrees of freedom

Data organisation

- Our data is currently in what is referred to as long format.
  - All the scores are in one column, with two entries per participant.
- To calcuate the d_i values, we will convert this to wide format.
  - Where there are two columns representing the score at time 1 and time 2
  - And a single row per person

Data organisation

```
exam_wide <- exam %>%
  pivot_wider(id_cols =  ID,
              names_from = time,
              values_from = stress)
```

```
## # A tibble: 6 × 3
##   ID       t1    t2
##   <chr> <dbl> <dbl>
## 1 ID1      14     7
## 2 ID2       7     7
## 3 ID3       8     9
## 4 ID4       8    12
## 5 ID5       7    10
## 6 ID6       7     9
```

Calculation

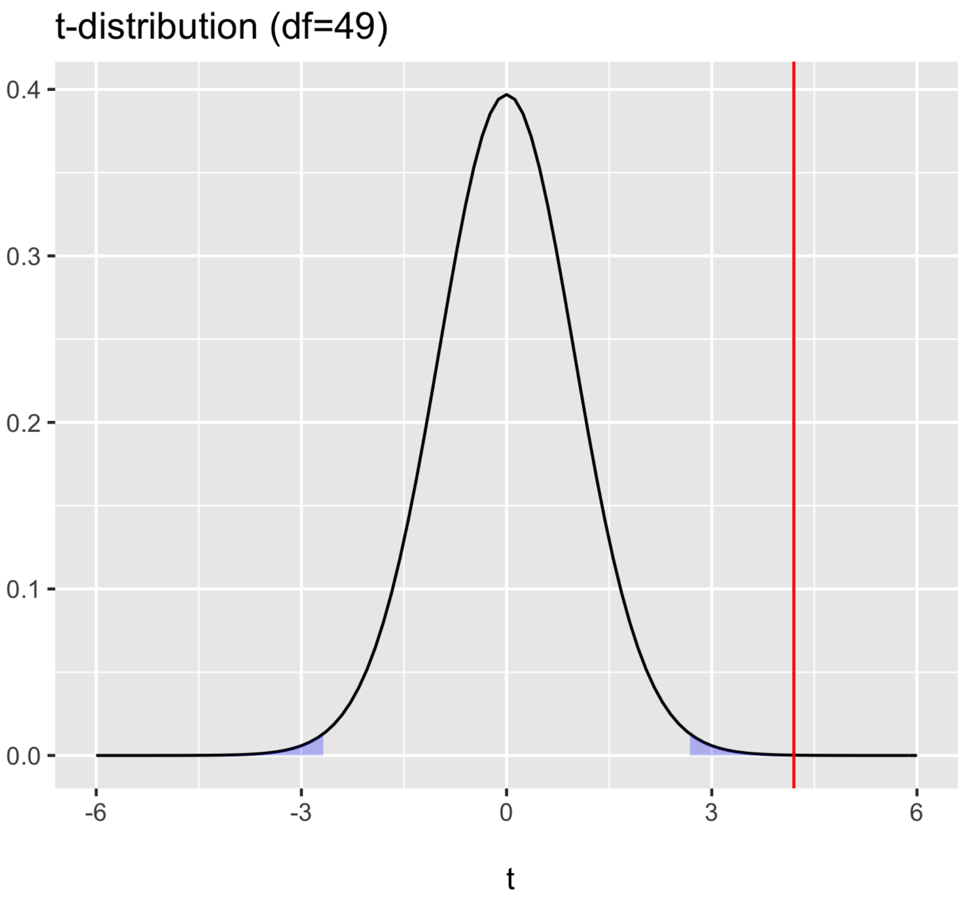
```
calc <- exam_wide %>%
  mutate(
    dif = t1 - t2) %>%

  summarise(
    D = mean(dif),
    SDd = round(sd(dif),2),
    N = n()) %>%

  mutate(
    SEd = round(SDd /sqrt(N),2),
    t = round(D/SEd,2)
  )
```

```
## # A tibble: 1 × 5
##       D   SDd     N   SEd     t
##   <dbl> <dbl> <int> <dbl> <dbl>
## 1   2.1  3.55    50   0.5   4.2
```

Is my test significant?

- So we have all the pieces we need:
  - t = 4.2
  - df = $n - 1$ = 49
  - Hypothesis to test (two-tailed)
  - $\alpha = 0.01$
- So now all we need is the critical value from the associated t-distribution in order to make our decision .

Is my test significant?




```
tibble(
  LowerCrit = round(qt(0.005, 49),2),
  UpperCrit = round(qt(0.995, 49),2),
  Exactp = round(2*(1-pt(calc[[5]], 49)),5)
)
```

```
## # A tibble: 1 × 3
##   LowerCrit UpperCrit  Exactp
##       <dbl>     <dbl>   <dbl>
## 1     -2.68      2.68 0.00011
```

In R

```
res <- t.test(exam_wide$t1, exam_wide$t2,
       paired = TRUE,
       alternative = "two.sided")
```

```
##
##      Paired t-test
##
## data:  exam_wide$t1 and exam_wide$t2
## t = 4.1864, df = 49, p-value = 0.0001174
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.091937 3.108063
## sample estimates:
## mean of the differences
##                     2.1
```

- Again, slight rounding differences.

Write-up

A paired-sample t-test was conducted in order to determine a if a statistically significant (α = .01) mean difference in self-report stress was present, pre- and post-time management intervention in a sample of 50 undergraduate students. The pre-intervention mean score was higher (Mean=9.72) than the post intervention score (Mean = 7.62). The difference was statistically significant (t (49)= 4.19, p < . 01, two-tailed). Thus, we reject the null hypothesis of no difference.

Time for a break

Welcome Back!

Now to check assumptions for a paired t-test and calculate Cohen's D

Assumption checks summary

	Description	One-Sample t-test	Independent Sample t-test	Paired Sample t-test
Normality	Continuous variable (and difference) is normally distributed.	Yes (Population)	Yes (Both groups/ Difference)	Yes (Both groups/ Difference)
Tests:	Descriptive Statistics; Shapiro-Wilks Test; QQ-plot			
Independence	Observations are sampled independently.	Yes	Yes (within and across groups)	Yes (within groups)
Tests:	None. Design issue.			
Homogeneity of variance	Population level standard deviation is the same in both groups.	NA	Yes	Yes
Tests:	F-test			
Matched Pairs in data	For paired sample, each observation must have matched pair.	NA	NA	Yes
Tests:	None. Data structure issue.			

Assumptions

1. Normality of the difference scores (d_i)
2. Independence of observations **within** group/time
3. Data are matched pairs (design)

- We will briefly show the normality assumptions again.
  - Hopefully these are becoming familiar.

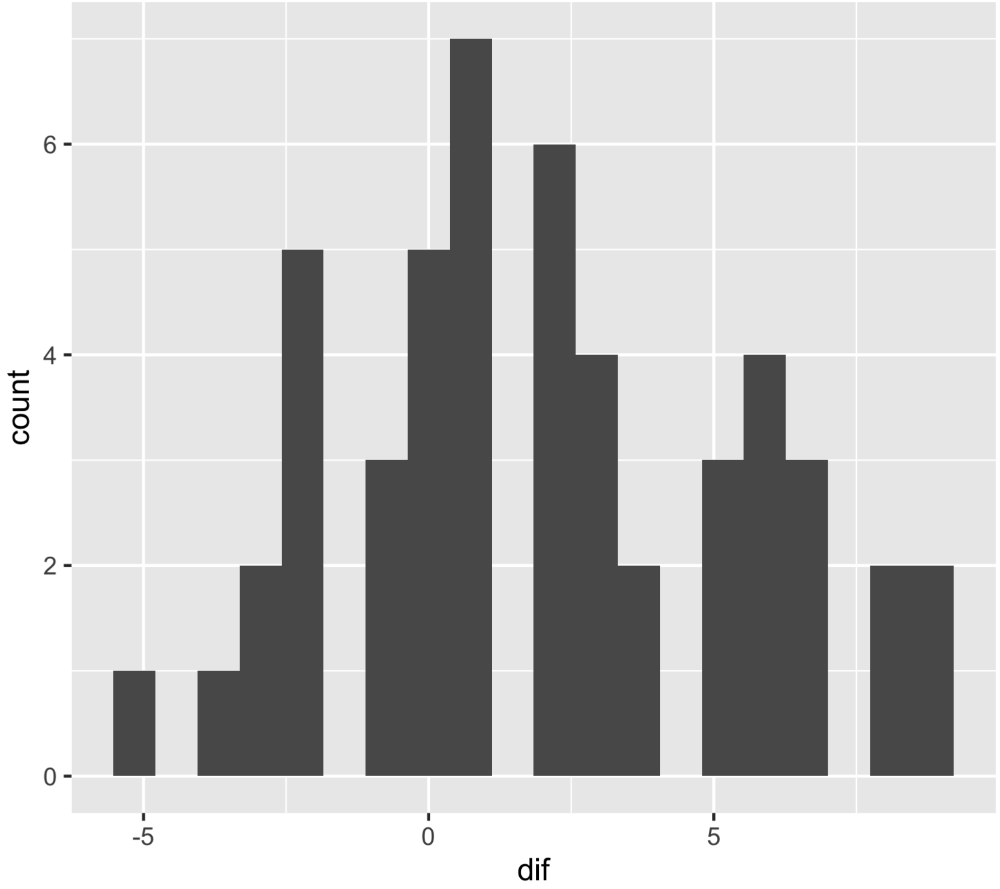
Adding the difference scores

- Our assumptions concern the difference scores.
- We showed these earlier in our calculations.
- Here we will add them to `exam_wide` for ease.

```
exam_wide <- exam_wide %>%
  mutate(
    dif = t1 - t2)
```

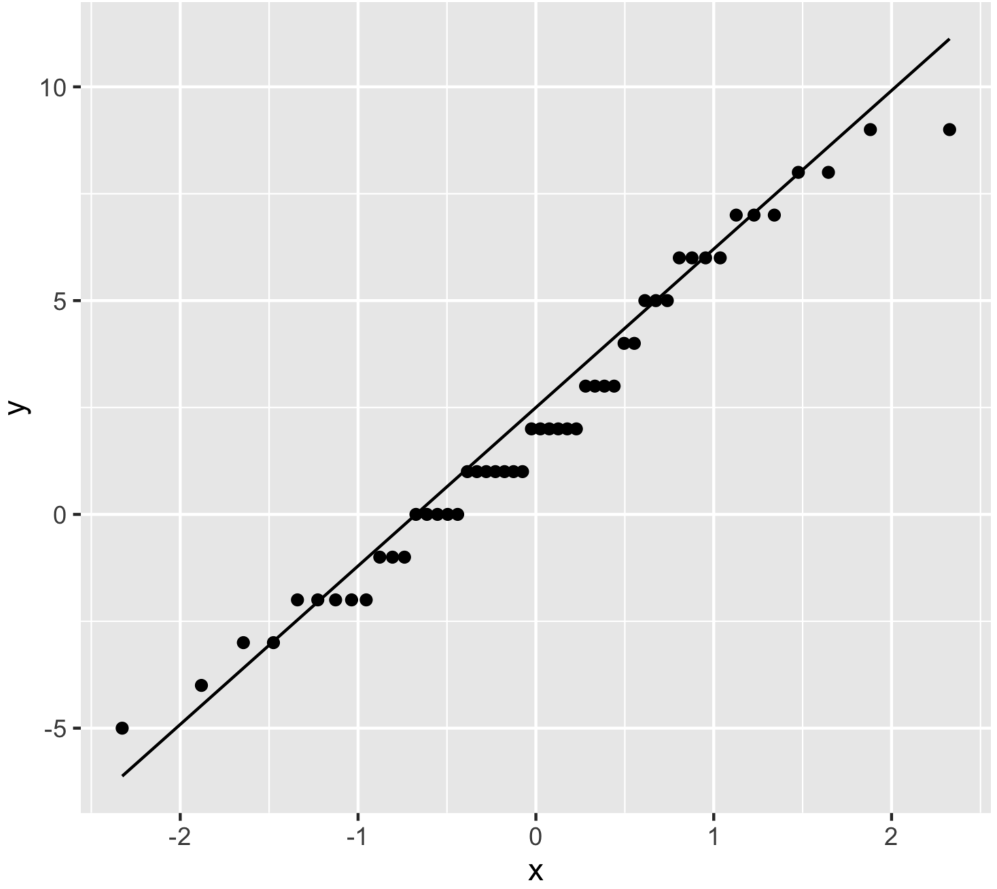
Histograms

```
exam_wide %>%
  ggplot(., aes(x=dif)) +
  geom_histogram(bins = 20)
```



QQ-plots

```
exam_wide %>%
  ggplot(., aes(sample = dif)) +
  stat_qq() +
  stat_qq_line()
```



Shapiro-Wilks R

```
shapiro.test(exam_wide$dif)
```

```
##
##      Shapiro-Wilk normality test
##
## data:  exam_wide$dif
## W = 0.97142, p-value = 0.264
```

- Fail to reject the null, $p > .05$
- Normality of the differences is met.

Cohen's D: Paired t

- Paired-sample t-test:

$$D = \frac{\bar{d} - 0}{s_d}$$

- $\bar{d}$ = mean of the difference scores (d_i)
- s_d = standard deviation of the difference scores (d_i)

Cohen's D in R

```r
library(effsize)
cohen.d(exam$stress, exam$time,
        subject = exam$ID,
        paired = TRUE,
        conf.level = .95)
```

```
##
## Cohen's d
##
## d estimate: 0.8822584 (large)
## 95 percent confidence interval:
##     lower     upper
## 0.3893289 1.3751880
```

Summary: Three different t-tests

	One-sample	Independent Sample	Paired (Dependent) Sample
Outcome	Continuous Variable	Continuous Variable	Continuous Variable
Predictor	Single group vs population	Categorical: two groups	Categorical: two time points
Sample	One sample vs population value	Two independent groups	One group sampled at two time points
Measure of difference	Observed - known population value	Group 1 - Group 2	Time 1 - Time 2
Measure of Variability	Standard error of the mean	Pooled standard error of difference in means	Standard error of the difference in means