Experimental Design and Data Analysis, Lecture 3

Eduard Belitser

VU Amsterdam

permutation test for two paired samples
0000000

Dependence in two paired samples
00000000

two independent samples
00000000000

Lecture overview

1. two paired samples (normal and not normal)

   - permutation test
   - dependence in two paired samples

     - Pearson's correlation test
     - Spearman's rank correlation test

2. two independent samples (normal and not normal)

   - two samples *t*-test
   - Mann-Whitney test
   - Kolmogorov-Smirnov test

permutation tests for two paired samples

Reminder: setting and design for two paired samples

Setting:

- An experiment with a numerical outcome measured according to two conditions per experimental unit;
- Interest is in a possible difference between the two outcomes per unit.

similar to paired setting & bootstr.

EXAMPLE Difference in average course grade for mathematical courses and informatics courses for BA-students at the VU.

EXAMPLE Difference in pain relief by an active drug and a placebo for patients.

is drug effective?

Design (the standard paired samples design):

- Take a random sample of experimental units from the relevant population.
- Measure the two outcomes on each unit.

Idea of permutation technique

- Data $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ must be two paired samples.

- In a permutation test we do not assume normality.

- We use any test statistic $T = T(X_1, Y_1, \ldots, X_n, Y_n)$ to test H_0: no difference between the distributions of X_j's and that of Y_j's. The choice of test statistics should express somehow the difference conjectured.

  needs to be a good indicator

- Like in a bootstrap test, we simulate the distribution of T under H_0, using B surrogate T^*-values. Repeat B times (for $i = 1, \ldots, B$):

  - generate each (X_j^*, Y_j^*), $j = 1, \ldots, n$, by applying a random permutation of the original (X_j, Y_j), i.e., choose between (X_j, Y_j) and (Y_j, X_j) with equal probability; $\longrightarrow$ *prob = 1/2 (flip a coin)*
  - next, compute $T_i^* = T(X_1^*, Y_1^*, \ldots, X_n^*, Y_n^*)$.

- Under H_0 of no difference between the distributions of X and Y within pairs permuting the labels should not change the distribution of T.

if it's in big prob. reg. = they come from same dist.

if it's in a small prob reg. = rej. H_0

we create many T^*s and look at dist. of them, then we look at where T-dist is.

dist. of T^* should be same as orig. T

Permutation test in R: data input and graphics

Recall dataset `ashina.txt` (headache after drug or placebo for 16 subjects).

```
> ashina=read.table("ashina.txt",header=TRUE)
> boxplot(ashina[,1],ashina[,2],names=c("active","placebo"))
> plot(ashina[,1],ashina[,2])
```


it seems like treat. decreases pain level
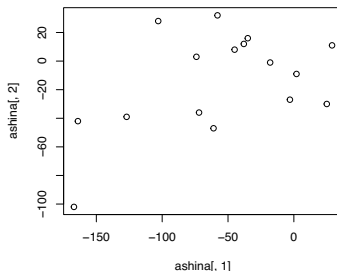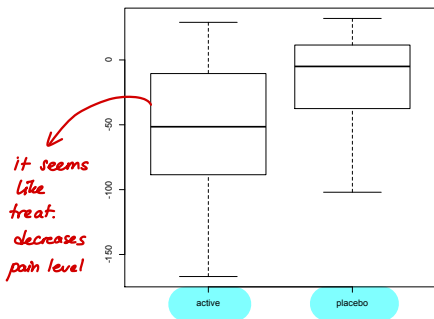
Based on the boxplots, we expect the active medicine to yield better pain relief.

Permutation test in R: testing (1)

choice of stat.s = $\overline{X} - \overline{Y}$

```
> mystat=function(x,y) {mean(x-y)}
> B=1000; tstar=numeric(B)
> for (i in 1:B) {
+ ashinastar=t(apply(cbind(ashina[,1],ashina[,2]),1,sample))
+    tstar[i]=mystat(ashinastar[,1],ashinastar[,2]) }
> myt=mystat(ashina[,1],ashina[,2])
```

*first col.*   *second col.*   *flip or not w/ prob = 1/2*

orig. T T^{*}

Instead of computing all $2^{16} = 65536$ possible permutations, we generate 1000 randomly chosen permutations to estimate the distribution of our test statistic under H_0. The function `apply` applies a function to either all rows or all columns in a matrix (parameter 1 indicates rows), `t(matrix)` means transposition of `matrix`.

Permutation test in R: testing (2)

→ T (based on orig. sample)

```
> myt
[1] -42.875
> hist(tstar)      → T*
> lines(rep(myt,2),c(0,20),
 + col="red",lwd=2)
> pl=sum(tstar<myt)/B
> pr=sum(tstar>myt)/B
> p=2*min(pl,pr); p
[1] 0.008
```

reject H_0

Histogram of tstar

in small region

myt

Frequency

tstar
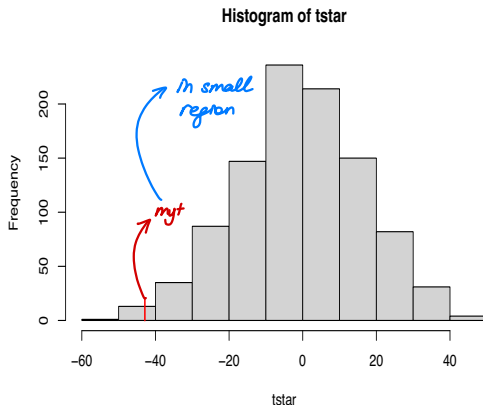
Conclusion: there is indeed a significant difference between the active drug and the placebo.

permutation test for two paired samples
Dependence in two paired samples
two independent samples

Permutation test: discussion

- A permutation test for two paired samples can be performed with any test statistic that expresses difference between the X and Y within pairs. (The mean of differences $Z_i = X_i - Y_i$ is most common to consider, but one may as well consider the median of the Z_i's.)

- Alternatives to the permutation test for two paired samples are the sign test and the Wilcoxon signed rank test applied to the differences.

permutation test for two paired samples
0000000

Dependence in two paired samples
●0000000

two independent samples
000000000000

Dependence in two paired samples

Dependence between two paired samples

* correlation = normalised covar.

Setting:

An experiment with two numerical outcomes (say X and Y) per experimental unit. Interest is in a possible dependence between the two outcomes per unit.

> if true dist. b.w. X and Y would have been known, I would compute covar. b.w. these 2 = covar = 0 → correlation = 0 ⤵ they are unrelat. (linearly) ⤵ does not mean indep. (linearly)

EXAMPLE Relation between shoe size and body mass index of a person.

EXAMPLE Relation between average course grade and number of students taking the course for courses at the VU.

EXAMPLE Relation between amount of precipitation and sun hours for different cities in Europe.

Design:

- Take a random sample of experimental units from the relevant population.
- Measure the two quantities on each unit. (The two outcomes are in principal related, because measured on the same experimental unit.)
- However, we possibly have measured unrelated quantities of the units and we want to test whether these quantities are correlated.

Pearson's correlation test *→ assumes norm.*

- Data $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$.

- The Pearson correlation test assumes normality of the both X_i's and Y_i's.
  (Rather, the asympt. normality of the sample correlation $\hat{\rho}$.)

- The test is based on the sample correlation coefficient (which estimates the "true" correlation $\rho = Cor(X, Y)$):

$$\hat{\rho} = \hat{\rho}_{X,Y} = \frac{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_N)(Y_i - \bar{Y}_n)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2}}.$$

sample corr. = good est.

- We test the null hypothesis $H_0 : \rho = \rho_0 = 0$ that the correlation between the two populations is $\rho_0 = 0$. The test statistic is given by

$$T_\rho = \frac{\hat{\rho} - \rho_0}{\sqrt{(1 - \hat{\rho}^2)/(n-2)}} = \frac{\hat{\rho}}{\sqrt{(1 - \hat{\rho}^2)/(n-2)}},$$

which has under $H_0 : \rho = 0$ a t-distribution with $n - 2$ degrees of freedom.

Spearman's rank correlation test

based on ranks

→ dependence on ranks (use in case you don't assume norm.)

- Data $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$.

- Spearman's rank correlation test does not assume normality. The test considers the ranks $R(X_i)$ and $R(Y_i)$ in the two samples, and compares the ordering of the ranks in the X_i and the Y_i.

- If the data are rank correlated, these sequences of ranks will run (approximately) in parallel or in opposite order.

- The test statistic is the sample correlation $\tilde{\rho}$ between the the rank vectors.

- We test the null hypothesis $H_0 : \rho_s = 0$. (Correlation of the rank variables.)

If all n ranks are distinct integers, the test statistic can be computed as $\tilde{\rho} = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)}$ where $d_i = R(X_i) - R(Y_i)$ is the difference between the two ranks of observations X_i and Y_i.

This test is useful in testing whether variable Y is a monotone transformation of variable X (or vice versa) in which case the true rank correlation is $\rho_s = 1$.

permutation test for two paired samples
○○○○○○○

Dependence in two paired samples
○○○○○●○○○

two independent samples
○○○○○○○○○○○○

Correlation tests in R: example

Consider the data frame `peruvians.txt`, where the rows correspond to 39 Peruvian men that moved from a native culture to a modern society. Amongst others, years since migration, systolic and diastolic blood pressure, heart rate (column `wrist`), weight, length were measured.

```
> peruvians=read.table("peruvians.txt",header=TRUE); peruvians
   age migration weight length chin  arm  calf wrist systolic diastolic
1   21         1   71.0   1629  8.0  7.0  12.7    88      170        76
2   22         6   56.5   1569  3.3  5.0   8.0    64      120        60
       [ some output deleted ]
39  54        40   87.0   1542 11.3 11.7  11.3    92      152        88
```

→ looks dependent

```
> attach(peruvians)
> plot(systolic~weight)
```

Based on this picture, we expect

dependence between `systolic` and `weight`.
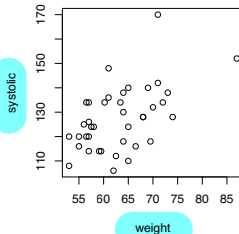
Pearson's test in R: example

```
> cor.test(systolic,weight)

 Pearson's product-moment correlation

data:  systolic and weight
t = 3.7164, df = 37, p-value = 0.0006654
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2463759 0.7186619
sample estimates:
      cor
0.5213643
```
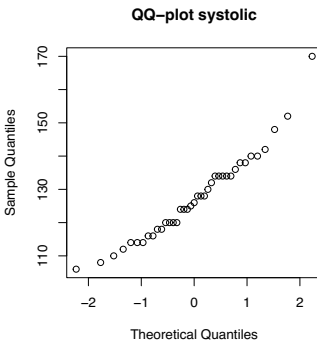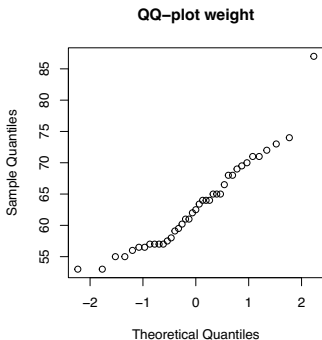
→ reject H_0 ⟶ they are corr.

Conclusion: there is significant correlation, if normality is assumed.

The default for `cor.test` is Pearson's correlation test, based on normality.

Correlation tests: diagnostics

Check the normality assumption on the two samples:

```
> par(mfrow=c(1,2)); qqnorm(peruvians$weight,main="QQ-plot weight")
> qqnorm(peruvians$systolic,main="QQ-plot systolic")
```


QQ-plots show that normality is doubtful for the weight sample. Hence, we use the rank correlation test of Spearman (and not Pearson's correlation test).

Spearman's test in R: example

```
> cor.test(systolic,weight,method="spearman")
```
→ *def. = Pearson*

```
        Spearman's rank correlation rho

data:  systolic and weight
S = 5322.352, p-value = 0.003119
```
→ *p. val is small → reject H_0 (same conclusion)*

```
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.4613004


Warning message:
In cor.test.default(systolic, weight, method = "spearman") :
  Cannot compute exact p-values with ties
```

Conclusion: there is indeed significant rank correlation.

There is a warning about ties, which means that some values occur multiple times in `weight` and/or `systolic`. Therefore R uses an approximation for the p-value.

two independent samples

Two independent samples: setting and design

Setting: an experiment with

- one numerical outcome per experimental unit,
- two independent groups of experimental units.

Interest is in a possible difference between the two populations. medskip → not paired

> EXAMPLE Comparing the weight of newborn children in two countries.

> EXAMPLE Total yield from an agricultural plot for two different fertilizers.
>
> not paired

Design:

- Take a random sample of experimental units of size M from the first population and a random sample of size N from the second population;
- Measure the outcome on each unit.

The numbers M and N need not be the same. → In paired: $M = N$

permutation test for two paired samples
○○○○○○○

Dependence in two paired samples
○○○○○○○○

two independent samples
○○●○○○○○○○○○○○

t-test for two independent samples

- Data $(X_1, \ldots, X_M)$ and $(Y_1, \ldots, Y_N)$. *size M* *size N*

- The <u>two samples t-test</u> assumes that both samples $X_1, \ldots, X_M$ and $Y_1, \ldots, Y_N$ come from independent normal populations. Denote the mean of the first population by μ and the mean of the second by ν.

- We test about the relation between the population means μ and ν:

$$H_0 : \mu \left\{ \begin{array}{c} = \\ \leq \\ \geq \end{array} \right\} \nu \quad \text{versus} \quad H_1 : \mu \left\{ \begin{array}{c} \neq \\ > \\ < \end{array} \right\} \nu.$$

- The test statistic is

$$T = \frac{\overline{X}_M - \overline{Y}_N}{S_{N,M}},$$ which has the t_{N+M-2}-distribution under H_0.

norm. fact. = makes sure that result. stat. has
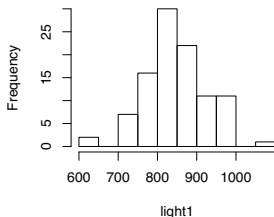*T-dist. w/ N+M−2 d.o.f. under H_0*

t-test in R: data input and graphics

→ we think they are indep. b/c they are meas. under diff. cond.s
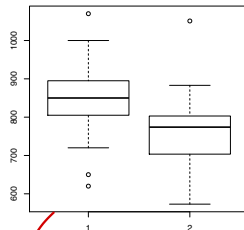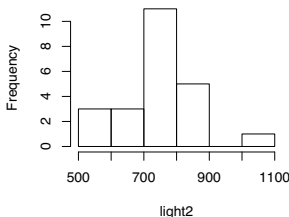
Consider two data sets of measurements of the speed of light (minus 299000) by Michelson in 1879 and in 1882.

```
>light1=scan("light1.txt"); light2=scan("light2.txt")
> hist(light1); hist(light2); boxplot(light1,light2)
```

→ looks norm. ←

sample sizes are not so big. (OK)

t-test in R: estimation and testing

The two samples *t*-test:

> *def = 2 indep. sample test*

```
> t.test(light1,light2)
```

        Welch Two Sample t-test

> H_0 is rejected ⟶ *2 samples have diff. means*

```
data:  light1 and light2
t = 4.0598, df = 27.754, p-value = 0.0003625
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  47.63387 144.73135
sample estimates:
mean of x mean of y
 852.4000   756.2174
```
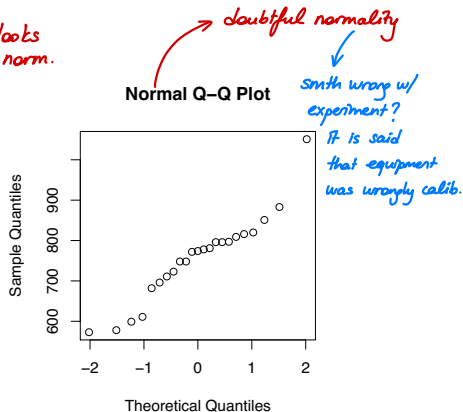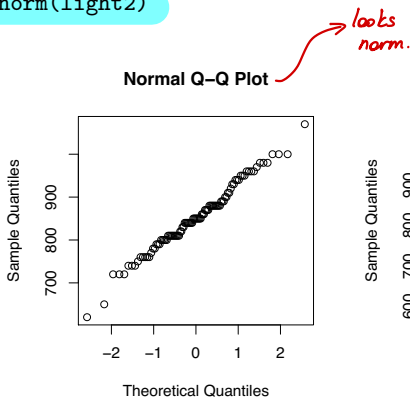
Conclusion: H_0 of equal means is rejected.

By default `t.test` with two arguments performs the two samples *t*-test for independent samples.

t-test in R: diagnostics

> `qqnorm(light1)`
> `qqnorm(light2)`

→ looks norm.

→ doubtful normality

↓ smth wrong w/ experiment? It is said that equipment was wrongly calib.

Normality of the second sample is actually doubtful.

permutation test for two paired samples
○○○○○○○○

Dependence in two paired samples
○○○○○○○○

two independent samples
○○○○○○●○○○○○

Mann-Whitney test → *based purely on ranks (does not rely on norm.)*

- Data: two independent samples $(X_1, \ldots, X_M)$ and $(Y_1, \ldots, Y_N)$.

- The Mann-Whitney test assumes that the sample $X_1, \ldots, X_M$ stems from population F and sample $Y_1, \ldots, Y_N$ stems from population G.

- We test the null hypothesis $H_0 : F = G$ (the distributions are the same).

- The Mann-Whitney test is again based on ranks. It considers the M ranks $R_1, \ldots, R_M$ of $X_1, \ldots, X_M$ in the combined sample $(X_1, \ldots, X_M, Y_1, \ldots, Y_N)$ of length $M + N$. If $F = G$ these M rank numbers should lie randomly between 1 and $M + N$. The test statistic is

$$T = \sum_{i=1}^{M} R_i,$$

the distribution of T under H_0 is (approximately) known.

- Large values of T indicate that F is shifted towards the right from G, i.e. that X-values are bigger than Y-values.

If responses are continuous, a significant result of Mann-Whitney test shows a difference in medians, actually this test is only consistent against the alternative $H_1 : P(X > Y) \neq P(Y > X)$.

Mann-Whitney test in R: testing

```
> wilcox.test(light1,light2)
```

```
        Wilcoxon rank sum test with continuity correction
```

```
data:  light1 and light2
W = 1829, p-value = 1.056e-05          → H₀ is reject.
alternative hypothesis: true location shift is not equal to 0
```

Conclusion: H_0 of equal medians is rejected. The underlying distribution of light1 is shifted to the right from that of light2.

When given two arguments wilcox.test will perform the Mann-Whitney test for two samples. The Mann-Whitney test is especially suited for detecting shift differences — differences in location — between two populations.

One-sided alternatives are also possible to test by the Mann-Whitney test. For example, to test whether the distribution of light1 is on the bigger values than the distribution of light2, we use wilcox.test(light1,light2,alt="g").

It will only see diff.s of cert. kind, not all small diff.s

fails to see if 2 dist. are truly diff.

Kolmogorov-Smirnov test *A: parametric*

→ can be used to test norm.

- Data: two independent samples $(X_1, \ldots, X_M)$ and $(Y_1, \ldots, Y_N)$.
- The Kolmogorov-Smirnov test assumes that the sample $X_1, \ldots, X_M$ stems from distribution F_X and sample $Y_1, \ldots, Y_N$ stems from distribution F_Y.
- We test the null hypothesis $H_0 : F_X = F_Y$ (the distributions are the same).
- The Kolmogorov-Smirnov test is based on the maximal difference of the two empirical distribution functions for two samples.
- The test statistic computes the maximal vertical difference in empirical distribution functions (summed histograms). Its distribution under H_0 is known (e.g., in R).

powerful test = it will always see diff. if you have suff. big M and N.
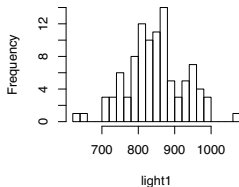
est. of dist. func.

The empirical distribution function for a sample $Z_1, \ldots, Z_n$ is defined as $\hat{F}_n(x) = \frac{\#(i : Z_i \le x)}{n}$ for all $x \in \mathbb{R}$. This is a non-decreasing from 0 to 1 step function making jumps of size $\frac{1}{n}$ in points $Z_{(1)}, \ldots, Z_{(n)}$.
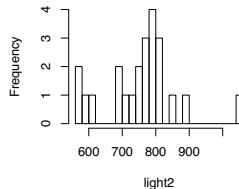
Kolmogorov-Smirnov test in R: graphics

permutation test for two paired samples
0000000

Dependence in two paired samples
00000000

two independent samples
0000000000●0
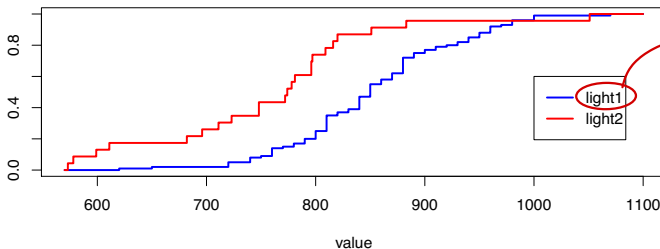
Testing in R by the Kolmogorov-Smirnov test

```
> ks.test(light1,light2)
```
def = 2-sided testing

```
        Two-sample Kolmogorov-Smirnov test

data:  light1 and light2
D = 0.5391, p-value = 3.803e-05
alternative hypothesis: two-sided
```
rej. H_0 = there is a diff. b.w. light1 and light2 → light1 is larger.

```
Warning message:
In ks.test(light1, light2) : cannot compute exact p-values with ties
```

A warning about ties again: R uses an approximation for computing the p-value.

Conclusion: $H_0 : F_X = F_Y$ is rejected (in fact, light1 is larger than light2).

One-sided alternatives are also possible to test by the Kolmogorov-Smirnov test. But take into account the counterintuitive interpretation: ks.test(x,y,alt="g") tests the alternative $F_X(x) \geq F_Y(x)$ which means that F_Y is shifted towards the right from F_X, i.e., that Y-values are bigger than X-values. For example, we suspect that light1 is larger than light2. To test this, ks.test(light1,light2,alternative="less") (or ks.test(light2,light1,alternative="greated").

To finish

Today we discussed: two samples tests (including permutation test); for paired and independent samples, for normal and not normal cases.

Next time: k samples, one way ANOVA.