Topic 6

Classical Statistical Tests for One- and Two-Sample Cases

Sergey Mastitsky ©
Klaipeda, 28-30 September 2011
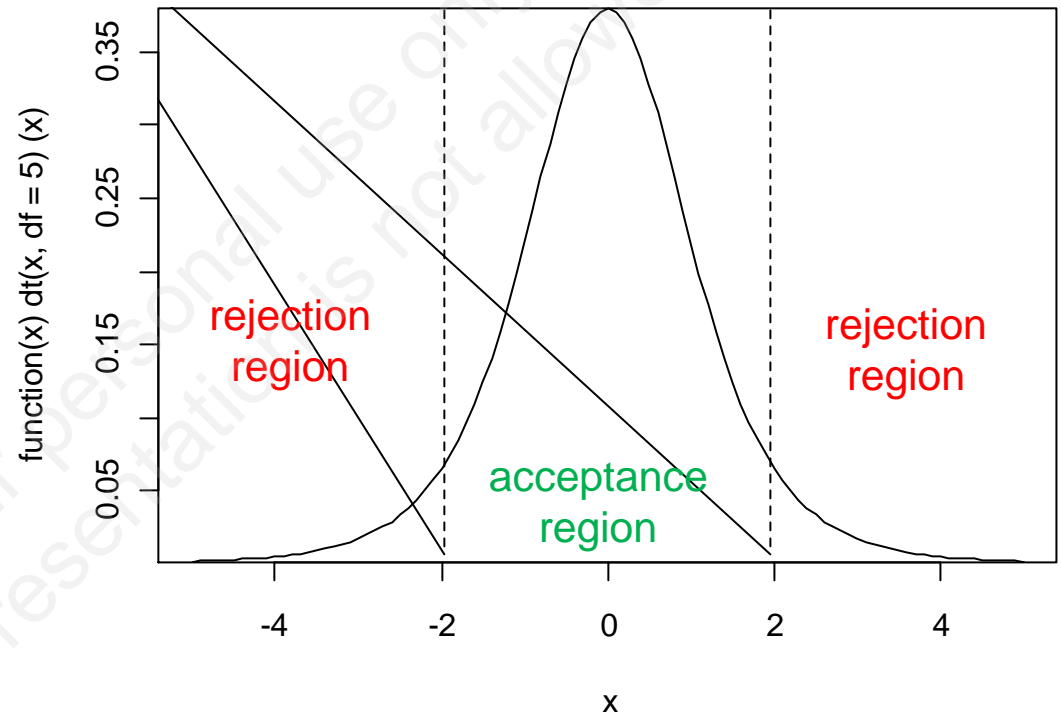
6. Classical Statistical Tests for One- and Two-Sample Cases

6.1. One-sample t-test

H_0: the sample of interest comes from a normally distributed population with mean μ_0

i.e. $H_0 : \mu = \mu_0$

Student *t* for one-sample case

$$t = \frac{\bar{x} - \mu_0}{S_{\bar{x}}}$$

- If *t* falls outside the acceptance region, reject H_0
- Alternatively, calculate the probability of obtaining a *t*-value as large or larger than the observed *t*-value. If *P* < 0.05, reject H_0

An example...

```
# Daily energy intake (kJ/day) in 11 women:
> d.intake <- c(5260, 5470, 5640,
        6180, 6390, 6515,
        6805, 7515, 7515,
        8230, 8770)
> mean(d.intake)
[1] 6753.6
> sd(d.intake)
[1] 1142.1
```

Does the sample mean differ significantly from a norm of 7725 kJ/day?

One-sample t-test:

```
> t.test(d.intake, mu = 7725)
```

```
	One Sample t-test

data:  d.intake
t = -2.8208, df = 10, p-value = 0.01814
alternative hypothesis: true mean is not equal to 7725
95 percent confidence interval:
 5986.348 7520.925
sample estimates:
mean of x
 6753.636
```

6.2. Two-sample t-test

Null hypothesis of a two-sample t-test

H_0: two samples come from normal distributions with the same mean

$$\text{i.e. } H_0 : \mu_1 = \mu_2$$

Student t for a two-sample case

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}}, \quad where \quad S_{\bar{x}_1 - \bar{x}_2} = \sqrt{S_{\bar{x}_1} - S_{\bar{x}_2}}$$

Only if group variances don't differ.
If they do, the Welch approximate test is used

- Again, if t falls outside the acceptance region, reject H_0
- Alternatively, calculate the probability of obtaining a t-value as large or larger than the observed t-value. If $P < 0.05$, reject H_0

An example...

\# Datasets that accompany the book of Peter Dalgaard (2008):

```
> library(ISwR)
```

\# Data on daily energy expenditure in lean and obese women:

```
> data(energy)
> attach(energy)
> energy
```

Does energy expenditure differ between lean and obese women?

```
> t.test(expend ~ stature)
```

```
	Welch Two Sample t-test

data:  expend by stature
t = -3.8555, df = 15.919, p-value = 0.001411
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.459167 -1.004081
sample estimates:
 mean in group lean mean in group obese
           8.066154            10.297778
```

The same test, but group variances are assumed to be equal

```
> t.test(expend ~ stature,
         var.equal = TRUE)
```

```
 Two Sample t-test

data:  expend by stature
t = -3.9456, df = 20, p-value = 0.000799
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.411451 -1.051796
sample estimates:
 mean in group lean mean in group obese
         8.066154            10.297778
```

Author: Sergey Mastitsky

5. Classical Statistical Tests for One- and Two-Sample Cases

5.3. Comparison of two variances

F-test to compare two variances

- Although R assumes that the group variances differ and thus automatically performs Welch test, we still can formally check this assumption:

```
> var.test(expend ~ stature)
```

```
 F test to compare two variances

data:  expend by stature
F = 0.7844, num df = 12, denom df = 8, p-value = 0.6797
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1867876 2.7547991
sample estimates:
ratio of variances
        0.784446
```

Important limitations of F-test

- Not robust against departures from normality

- Not applicable to paired data

Author: Sergey Mastitsky

6. Classical Statistical Tests

6.4. Formal testing for normality
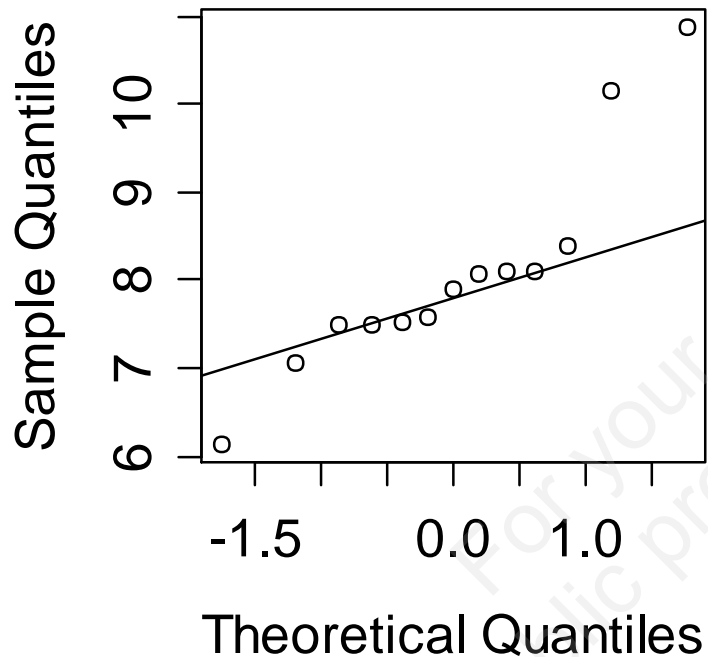
How to check if the data are normally distributed within each group?

```
> e.lean <- energy$expend[
    energy$stature == "lean"]

> e.obese <- energy$expend[
    energy$stature == "obese"]

> qqnorm(e.lean); qqline(e.lean)
> qqnorm(e.obese); qqline(e.obese)
```
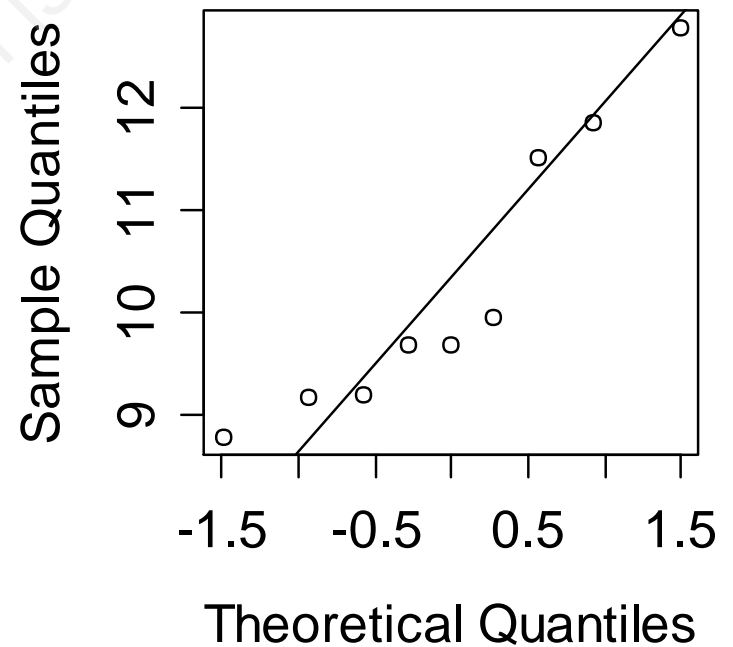
Normal probability plots

Shapiro-Wilk normality test

```
> shapiro.test(e.lean)
```

Shapiro-Wilk normality test

data: e.obese
W = 0.876, p-value = 0.1426

No deviation from normality

```
> shapiro.test(e.obese)
```

Shapiro-Wilk normality test

data: e.lean
W = 0.8673, p-value = 0.04818

Deviation from normality (marginal)

Author: Sergey Mastitsky

6.5. Wilcoxon test

Wilcoxon test

- t-tests are rather robust to deviations from normality, especially when samples are large

- However, in many cases (especially in Biology) it may be better to use a distribution-free method, e.g. Wilcoxon signed rank test

One-sample Wilcoxon test: theory

- H_0: the sample of interest comes from a symmetric distribution with the center at μ_0
- Subtract theoretical μ_0 from each observation
- Rank the resultant values, ignoring the sign
- Sum up the positive ranks to get the V-statistic
- Compare $V_{obs.}$ with $V_{crit.}$, or equivalently
- Calculate the probability P of obtaining $V \geq V_{obs.}$

An example...

Daily energy intake in 11 women compared against a recommended value:

```
> wilcox.test(d.intake, mu = 7725)
```

```
	Wilcoxon signed rank test with continuity correction

data:  d.intake
V = 8, p-value = 0.0293
alternative hypothesis: true location is not equal to 7725

Warning message:
In wilcox.test.default(d.intake, mu = 7725) :
  cannot compute exact p-value with ties
```

Two-sample Wilcoxon (= Mann Whitney) test: theory

- H_0: the two samples are coming from two distributions with different centers
- Rank all the values, ignoring the group and sign
- Sum up all the ranks in the first group to get the W-statistic
- Compare W_{obs} with its theoretically expected minimal value, or equivalently
- Or calculate the probability P of obtaining $W \geq W_{obs}$.

Author: Sergey Mastitsky

An example…

Daily energy expenditure in lean and obese women

```
> wilcox.test(expend ~ stature)
```

```
Wilcoxon rank sum test with continuity correction

data:  expend by stature
W = 12, p-value = 0.002122
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(x = c(7.53, 7.48, 8.08, 8.09, 10.15, 8.4,
  cannot compute exact p-value with ties
```

Limitation of the Wilcoxon tests

- Unless the sample size is 6 or above, the test simply cannot become significant at the 5% level

- The Wilcoxon tests are also susceptible to *ties* (i.e. observations sharing the same value). If ties are present, no exact *P*-value can be computed (not a problem for large samples)
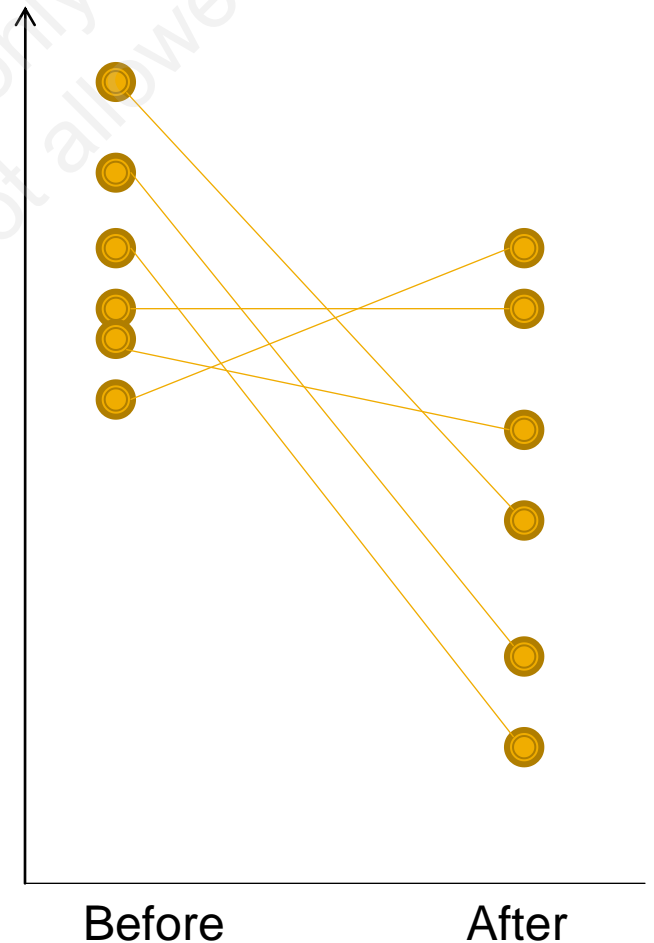
Author: Sergey Mastitsky

6.6. Comparison of paired samples

Paired tests

- Paired tests are used when there are two measurements on the same experimental unit
- Individual differences are calculated and then averaged, reducing the problem to a one-sample test

An example…

```
> library(ISwR)
```
Pre- and postmenstrual energy intake in a
  group of women
```
> data(intake)
> intake
   pre   post
1 5260  3910
2 5470  4220
3 5640  3885
. . .
```

The same 11 women were measured, so it makes sense to look at individual differences:

```
> attach(intake)

> post - pre
 [1] -1350 -1250 -1755 -1020 -745
 [6] -1835 -1540 -1540 -725 -1330
[11] -1435

> mean(post-pre)
[1] -1320.455
```

Testing Ho that the mean individual difference equals to 0:

```
> t.test(pre, post, paired = T)
```

```
Paired t-test

data:  pre and post
t = 11.9414, df = 10, p-value = 3.059e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1074.072 1566.838
sample estimates:
mean of the differences
              1320.455
```

Author: Sergey Mastitsky

The matched-pairs Wilcoxon test

```
> wilcox.test(pre, post, paired = T)
```

```
	Wilcoxon signed rank test with continuity correction

data:  pre and post
V = 66, p-value = 0.00384
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(pre, post, paired = T) :
  cannot compute exact p-value with ties
```

6. Classical Statistical Tests for One- and Two-Sample Cases

6.7. Testing proportions

H_0: the sample of interest comes from a population with the mean p_0

$$\text{i.e. } H_0 : p = p_0$$

Author: Sergey Mastitsky

Testing a single proportion: the theory

- The test is based on the binomial distribution
- If p is the probability (proportion) of interest and N is the sample size, then

$$sample\ mean = Np$$

$$sample\ variance = Np(1-p)$$

- If x is the number of binomial "successes", then the hypothesis of $p = p_0$ can be based on

$$u = \frac{x - Np_0}{\sqrt{Np_0(1 - p_0)}}$$

u^2 has an approximate χ^2 distribution with 1 d.f.

Author: Sergey Mastitsky

An example…

- Suppose, 40 out of 250 (16%) randomly chosen snails were found to be infected with a parasite
- We want to test the hypothesis that the probability of a randomly selected mollusc being infected is the expected 0.20 (20%)

Implementation of the single proportion test in R

```
> prop.test(40, 250, 0.20)
```

```
    1-sample proportions test with continuity correction

data:  40 out of 250, null probability 0.2
X-squared = 2.2563, df = 1, p-value = 0.1331
alternative hypothesis: true p is not equal to 0.2
95 percent confidence interval:
 0.1180086 0.2127286
sample estimates:
   p
0.16
```

Comparison of two independent proportions

- Suppose, we examined two populations of snails
- In population 1, 24 out of 158 (15.2%) snails were infected
- In population 2, 15 out of 165 (9.1%) snails were infected
- Are the two proportions significantly different between populations?

Data need to be presented in two vectors:

one with numbers of "successes"

```
> infected <- c(24, 15)
```

and one with total numbers of observations:

```
> examined <- c(158, 165)
```

Comparing two proportions in R

```
> prop.test(infected, examined)
```

```
 2-sample test for equality of proportions with continuity
 correction

data:  infected out of examined
X-squared = 2.2826, df = 1, p-value = 0.1308
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.01631243  0.13829171
sample estimates:
    prop 1     prop 2
0.15189873 0.09090909
```

The chi-squared test

- Two (or more proportions) can also be compared with the function of the chi-squared test: `chisq.test()`

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

- The data have to be presented as a matrix corresponding to a 2x2 contingency table

Comparing two proportions with the function `chisq.test()`

```
> inf.data <- matrix(
 c(infected, examined-infected),
 nrow = 2,
 dimnames = list(c("Pop1", "Pop2"),
 c("Infected", "Noninfected")))

> inf.data
        Infected    Noninfected
Pop1    24          134
Pop2    15          150
```

Comparing two proportions with the function `chisq.test()`

```
> chisq.test(inf.data)
```

```
	Pearson's Chi-squared test with Yates' continuity correction

data:  inf.data
X-squared = 2.2826, df = 1, p-value = 0.1308
```

Fisher's exact test

- When the expected values are ≤5, the chi-squared test produces unreliable *P*-values
- The Fisher's exact test has to be performed in such cases
- Suppose, the data on infection in snails were as follows:

```
> infected <- c(1, 9)
> examined <- c(9, 14)
```

Checking the expected values from a chi-squared test

```r
> inf.data <- matrix(
 c(infected, examined-infected),
 nrow = 2,
 dimnames = list(c("Pop1", "Pop2"),
 c("Infected", "Noninfected")))
> chisq.test(inf.data)$expected
```

```
     Infected Noninfected
Pop1 3.913043    5.086957
Pop2 6.086957    7.913043
Warning message:
In chisq.test(inf.data) : Chi-squared approximation may be incorrect
```

```
> fisher.test(inf.data)
    Fisher's Exact Test for Count Data

data:   inf.data
p-value = 0.02881
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
  0.001419354 0.867426585
sample estimates:
odds ratio
0.07882877
```

$$odds\ ratio = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

Compare with the chi-squared test's P-value:

```
> chisq.test(inf.data)$p.value
[1] 0.0376
```

Author: Sergey Mastitsky

6.8. Extension of the chi-squared and Fisher's exact tests to *r* x *c* contingency tables

r x *c* contingency tables

- In many research setups, there can be >2 classes on both sides of a contingency table
- Suppose, we examined 3 populations of snails
- In each population, we recorded the shell color ("light", "darker", "dark")
- Do the proportions of snails with various shell coloration differ among the three populations?

A 3×3 contingency table

```
> light <- c(12, 40, 45)
> darker <- c(87, 34, 75)
> dark <- c(3, 8, 2)
> color.data <- matrix(
 c(light, darker, dark), nrow = 3,
 dimnames = list(c(
"Pop1", "Pop2", "Pop3"),
 c("Light", "Darker", "Dark")))
```

A *3* × *3* contingency table

```
> color.data
        Light   Darker Dark
Pop1    12      87       3
Pop2    40      34       8
Pop3    45      75       2
```

```
> chisq.test(color.data)
```

```
 Pearson's Chi-squared test

data:  color.data
X-squared = 43.4337, df = 4, p-value = 8.411e-09

Warning message:
In chisq.test(color.data) : Chi-squared approximation may be incorrect
```

Finding where the differences lie (i.e. influential cells)

```
> E <- chisq.test(
        color.data)$expected
> O <- chisq.test(
        color.data)$observed
> (O - E)^2/E
```

```
          Light    Darker        Dark
Pop1 12.786942  7.1853741  0.4102564
Pop2  7.547399  6.5323333  5.8551423
Pop3  1.035041  0.1264774  1.9547594
```

Fisher's exact test for a 3 x 3 table

```
> fisher.test(color.data)
```

```
        Fisher's Exact Test for Count Data

data:  color.data
p-value = 1.191e-09
alternative hypothesis: two.sided
```

Author: Sergey Mastitsky