Statistical Computing
Week 12: 10th January 2020
Hypothesis Tests 2

Hypothesis Tests 2: Contents

- Recap of Lecture 11
- One sided hypothesis tests
- Other types of hypothesis tests:
  - Regression
  - Test of independence
- Summary

Reminder from last year

The **general principle** of a hypothesis test

A sample of 20 Berlin residents was obtained and the height of each person was measured.

The population mean is assumed to be 170 cm.
The sample mean is 169.0 cm

We set up two hypotheses to see if our sample mean fits the assumed population mean.

- The *null hypothesis* H_0: The population mean is 170 cm
- and the *alternative hypothesis* H_1: The population mean is **not** 170 cm

We compare our data with what we would expect, if the *null hypothesis* were true. If our data are very “unexpected” we don’t believe the null hypothesis, we reject it and “choose” the alternative hypothesis.

If we don’t reject the null hypothesis, either *the null hypothesis is true* or *it is untrue but we have not collected enough evidence yet to reject it*.

One sample t-test

```
> t.test(Berlin,mu=170)

        One Sample t-test
data:  Berlin
t = -0.33816, df = 19, p-value = 0.739
alternative hypothesis: true mean is not equal to 170
95 percent confidence interval:
 163.0398 175.0240
sample estimates:
mean of x
 169.0319
```

The p-Value is much greater than 0.05.
We do not have enough evidence to reject the null hypothesis.

Under the null hypothesis assumptions the following random variable follows a t-distribution:

Test statistic: $$t_{stat} = \frac{\overline{X} - \mu_0}{S_X}\sqrt{n} \sim t_{n-1}.$$

t_{n-1} is the t-distribution with $n-1$ degrees of freedom.
The test statistic t_{stat} is compared it to a **critical value** to see if the data are unlikely if the null hypothesis is true.

Critical region

Assuming H_0 to be true, the probability that t_{stat} lies in the central region is $1-\alpha = 95\%$. The probability that t_{stat} is extreme and lies in side regions is $\alpha = 5\%$ (the critical region).

The critical value t_{cr} is by definition the 0.975-quantile of the t_{n-1} distribution, and is found using the R command `qt(0.975,n-1)`

This leads to the decision rule:

If $|t_{stat}| \leqslant t_{cr}$, we „accept” the null hypothesis

If $|t_{stat}| > t_{cr}$, the null hypothesis is rejected, we do not believe that $\mu = \mu_0$.

On the last slide, we assumed that $\alpha = 5\%$. This is called the significance level and $\alpha = 5\%$ is the usual value by convention. Other common significance levels are 10% and 1%.

Instead of a critical value, statistical software outputs a *p*-value.

The exact definition of the *p*-value is: the smallest significance level required to reject the null hypothesis?

The *p*-value based decision rule is:

If *p*-value $> \alpha$, we „accept” the null hypothesis

If *p*-value $\leqslant \alpha$, the null hypothesis is rejected, we do not believe that $\mu = \mu_0$.

In our example the *p*-value is much greater than 0.05 we “accept” the null hypothesis.

One sided Tests

So far you have only seen tests where the null hypothesis is μ is equal to a null value,

$$H_0 : \mu = \mu_0 \qquad \text{vs} \qquad H_1 : \mu \neq \mu_0$$

The null hypothesis is rejected, if $\overline{x}$ is **much larger or much smaller** than the null value μ_0.

Often we are only "interested" in the result, if $\overline{x}$ is much smaller than the null value μ_0. This is a one sided test and has hypotheses in the form:

$$H_0 : \mu \geqslant \mu_0 \qquad \text{vs} \qquad H_1 : \mu < \mu_0$$

N.B. The null value always belongs to the null hypothesis.
H_0 and H_1 are mutually exclusive.
H_1 the alternative lies on the side of "interest".

When we are only "interested" in the result, if $\overline{x}$ is much **larger** than the null value μ_0, we have a one sided test of the form:

$$H_0 : \mu \leqslant \mu_0 \qquad \text{vs} \qquad H_1 : \mu > \mu_0$$

The test statistic is calculated in the same way as before:

$$t_{stat} = \frac{\overline{X} - \mu}{S_X}\sqrt{n}$$

As t_{stat} is the same it also has a t_{n-1} distribution, if the null hypothesis is true.

The critical region lies on one side only, which is why it is called a one sided test. The tests in the last lecture had a critical region to each side and is called a *two sided test*.

In a one sided test the entire 5% (α) type I error lies in one tail, the upper or lower tail depending on the alternative hypothesis test.

Example: at risk babies

In a North European population the expected weight of new born babies is 3500 grammes.

Midwives claim that complications during pregnancy lead to a reduces birth weight. We will call such infants “at risk babies”.

A sample of 20 at risk babies were weighed and compared to the expected mean of 3500 g.

The null and alternative hypotheses are

- The expected birth weight of at risk babies is at least 3500 g.
- The expected birth weight of at risk babies is less than 3500 g.

Our “interest” lies in the case that at risk babies have reduced birth weight, so the appropriate one sided hypotheses are:

$$H_0 : \mu \geqslant 3500 \qquad \text{vs} \qquad H_1 : \mu < 3500$$

The sampled birth weight data gave the following results $\overline{x} = 3280\,\mathrm{g}$ and $s_x = 490\,\mathrm{g}$. The null value is $\mu_0 = 3500\,\mathrm{g}$, the significance level is 5% and the sample size $n = 20$.

The test statistic is calculated using the same formula as in a two sided test

$$t_{stat} = \frac{\overline{x} - \mu_0}{s_x/\sqrt{n}} = \frac{3280 - 3500}{490/\sqrt{20}} = -2.0079.$$

The critical value t_{cr} is found using the 5% (α) quantile of the t_{19} distribution.

```
> qt(0.05,19)
[1] -1.729133
```

N.B. when the critical region lies in the upper tail use the $(1 - \alpha)$ quantile

```
> qt(0.95,19)
[1] 1.729133
```


$$t_{stat} = -2.01 < t_{cr}$$

The test statistic is in the critical region
$\Rightarrow$ the null hypothesis is rejected
$\Rightarrow$ we conclude that at risk pregnancies lead to a reduction in expected birth weight.

***p*-value:**
The *p*-value is 0.0295.
Because the *p*-value is less than 0.05, the null hypothesis is rejected.

It is important to decide in advance whether the test is one sided or two sided, and if it is unclear then the test should be a *two sided test*.

You should not use the approach:
„The sample mean is less than 3500 g so I will use a one sided t-Test with the alternative hypothesis $\mu < 3500$."
That is cheating and leads to false conclusions!

Two sample t-test

Also revision from last year.

To compare the population mean between two groups (e.g. test and control groups) we use a Two sample t-test

The null hypothesis is the two population means are the same.
The alternative hypothesis is the two population means are different.

The two sided test for unpaired data has the hypotheses:

$$H_0 : \mu_x = \mu_y \qquad \text{vs} \qquad H_1 : \mu_x \neq \mu_y$$

Test statistic is:

$$t_{\text{stat}} = \frac{\overline{x} - \overline{y}}{\sqrt{s_x^2/n_x + s_y^2/n_y}}$$

where n_x and n_y are the sample sizes, $\overline{x}$ and $\overline{y}$ the sample means, and s_x^2 and s_y^2 are the sample variances of the two groups.

If equal variances are assumed, then the formula reduces to:

$$t_{\text{stat}} = \frac{\overline{x} - \overline{y}}{s_p\sqrt{1/n_x + 2/n_y}},$$

where

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}.$$

S_p is called the *pooled standard deviation* and is a weighted mean of the sample standard deviation of the two groups.

The critical value is the same as in the one sample t-test, but the degrees of freedom is $n_x + n_y - 2$.

The decision rule is the same as in the one sample t-test:
If $|t_{stat}| \leqslant t_{cr}$, we “accept” the null hypothesis $\mu_x = \mu_y$.
If $|t_{stat}| > t_{cr}$, the null hypothesis is rejected, we do not believe that $\mu_x = \mu_y$.

One sided tests are also possible for comparing two samples. In which case the hypotheses are:

$$H_0 : \mu_x \leqslant \mu_y \qquad \text{vs} \qquad H_1 : \mu_x > \mu_y$$

or

$$H_0 : \mu_x \geqslant \mu_y \qquad \text{vs} \qquad H_1 : \mu_x < \mu_y$$

There is an example of a one sided two sample test in the worksheet.

Other types of hypothesis tests

The tests covered so far are all tests of locations One sample and two sample tests involving the population mean.

There are many different types of hypothesis tests. Every time you hear the term "is (not) statistically significant" then a Hypothesis test is implied.

We will now look a couple of other common scenarios where a hypothesis test is appropriate.

Regression

The criket chirp regression from lecture 6

```
> lm.obj<-lm(formula = chirp ~ temp, data = Grasshoppers)
> summary(lm.obj)
Call: lm(formula = chirp ~ temp, data = Grasshoppers)
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.882  0.908      13.084 7.36e-09 ***
temp         0.382  0.069       5.466 0.000108 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Temperature has a *p*-value of 0.0001, which means this is the result of a hypothesis test, but what are the hypotheses being tested?

We assume that the true relationship between the chirp rate y and temperature x is

$$y = a + bx.$$

If $b = 0$ then temperature has no influence on chirp rate.

If we could measure infinitely many crickets at infinitely many temperatures we might be able to calculate a and b, but in practice we estimate them using the formulae in Lecture 6, giving $\widehat{a}$ and $\widehat{b}$.

We use $\widehat{b}$ to test whether we believe temperature has an influence on chirp rate, i.e.

$$H_0 : b = 0 \qquad \text{vs} \qquad H_1 : b \neq 0$$

If $\widehat{b}$ is a long way from zero, taking into account the variation in the residuals, then the p-value will be small, we reject the null hypothesis and conclude that x has an influence on y.

If $\widehat{b}$ is close to zero, then the p-value will be large, and we accept that the calculated $\widehat{b}$ is non-zero only because of random error. We conclude that x has no influence on y.

Test of independence

Used for data summarised using a contingency table.

Is the incidence of red-green colour blindness the same for males and females?

	Colour blind		
	no	yes	total
Female	248	2	250
Male	230	20	250
Total	478	22	500

H_0: red-green colour blindness and sex are *independent*
H_1: red-green colour blindness and sex are *dependent*

Approach for a test of independence

(a) Find the ideal frequencies that we would expect if they are independent

$$e_{ij} = \frac{h_{i}.h_{.j}}{n},$$

where $h_i.$ and $h_{.j}$ are the row and column frequencies respectively.

(b) Calculate the test statistic $\chi^2_{stat} = \sum_{i,j} \dfrac{(h_{ij} - e_{ij})^2}{e_{ij}}$ which measures how far from this expected ideal the observed values are.

(c) If H_0 is true then the test statistic χ^2_{stat} is “chi-squared” χ^2 distributed random variable.

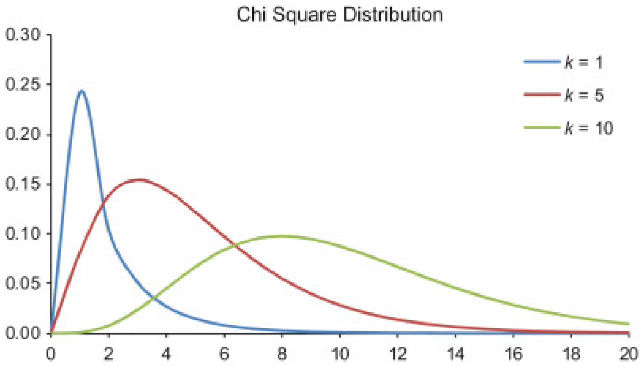
(d) Compare this test statistic to a critical value or obtain the p-value based

(e) Interpret the results.

χ^2 distribution

This distribution is positive and highly positively skewed.

It has one parameter k called degrees of freedom.



For this type of test the number of degrees of freedom is k=(nrows-1)(ncolumns-1).
The test is always one sided, only reject if the test statistic is large, so the critical value is the $(1-\alpha)$-quantile of the χ^2_k distribution.

In our example $k = (2-1)(2-1) = 1$ and the critical value at the 5% significance level is

```
> qchisq(0.95,1)
[1] 3.841459
```

The expected frequencies are:

	Colour blind		
	no	yes	total
Female	239	11	250
Male	239	11	250
Total	478	22	500

We calculate expected frequencies using row-total×column-total/sample size.

E.g. $250 \times 478/500 = 239$ (note this is not alway a whole number!)

- $\chi^2_{Pg} = \frac{9^2}{239} + \frac{9^2}{11} + \frac{9^2}{239} + \frac{9^2}{11}$

All the numerators are the same value, which is a result of there being 1 degree of freedom k=1

- Test statistic is greater than the critical value $\chi^2_{Pg} = 15.4 > 3.84$
- We reject H_0.
- We conclude that colour blindness and sex are dependent.

To obtain the p-value, find the probability that a χ^2_1 distributed random variable is less than the test statistic value 15.4

```
> 1-pchisq(15.4,1)
[1] 8.698829e-05
```


Hypothesis tests: summary

There are many different forms of hypothesis test, which are a fundamental part of statistical data analysis.

Although the types of test are different they all have similarities:

- Decide what parameter or property is to be tested (e.g. population mean, regression coefficient, independence)
- Is the test a comparison against a fixed value or between two groups (one sample or two)?
- Is a one sided test appropriate?
- Specify the null and alternative hypotheses.
- Calculate the test statistic.
- Choose the significance level α usually 0.05 in advance.
- continued ...

- Find the critical value or the p-value, and decide if the test statistic lies in the critical region.
    - If it does then we reject the null hypothesis
    - otherwise we accept the null hypothesis, although this may be due to a lack of evidence, e.g. a small sample size.
- Interpret the test result in the context of the area of application.