Statistics for Computing MA4413

Lecture 16

Hypothesis Testing Continued

Kevin Burke

kevin.burke@ul.ie

Hypothesis Testing

We have seen hypothesis testing in the cases of μ and p.

In both cases the test statistic is:

$$z = \frac{\text{statistic} \; - \; \text{hypothesised value}}{\text{standard error}}$$

which must then be compared to a critical value - either a z value (if n is large) or a t value (if n is small and data is approximately normal).

Mean and Proportion

Parameter	Statistic	Standard Error	Samples	Small $\Rightarrow t_\nu$
μ	$\bar{x}$	$\dfrac{s}{\sqrt{n}}$	large / small	$\nu = n - 1$
p	$\hat{p}$	$\sqrt{\dfrac{p_0\,(1 - p_0)}{n}}$	large	n/a

Note: the standard error for $\widehat{P}$ is different to that used for confidence intervals, i.e., use p_0 for hypothesis tests, **not** $\hat{p}$.

Hypothesis Testing in General

Of course, we also have the parameters $\mu_1 - \mu_2$ (difference in means) and $p_1 - p_2$ (difference in proportions).

In any case, the basic steps for hypothesis testing are the same.

1. State the null and alternative hypotheses:

$$\boxed{\begin{array}{llccccccl} H_0: & \text{parameter} & = & \textbf{or} & \leq & \textbf{or} & \geq & \text{hypothesised value} \\ H_a: & \text{parameter} & \neq & \textbf{or} & > & \textbf{or} & < & \text{hypothesised value} \end{array}}$$

where the parameter will be one of: μ, p, $\mu_1 - \mu_2$, $p_1 - p_2$.

Hypothesis Testing in General

2. Set the α level: commonly 0.1, 0.05 or 0.01.

3. Calculate the test statistic:

$$\boxed{z = \frac{\text{statistic} - \text{hypothesised value}}{\text{standard error}}}$$

4. Compare this to the appropriate critical value(s) from tables.
   - Use z for large samples and t for small samples.
   - Do **not** divide α by two if H_a has "$<$" or "$>$" sign (one-tailed test).

5. Provide your conclusion using both statistical *and* non-statistical language.

Difference Between Two Means: $\mu_1 - \mu_2$

Statistic	Standard Error	Samples	Small $\Rightarrow t_\nu$
$\bar{x}_1 - \bar{x}_2$	$\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$	large / small	$\nu = \dfrac{(a+b)^2}{\frac{a^2}{n_1-1} + \frac{b^2}{n_2-1}}$ $a = \dfrac{s_1^2}{n_1}, \quad b = \dfrac{s_2^2}{n_2}$
	$\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}$ where $s_p^2 = \dfrac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$	small assuming $\sigma_1^2 = \sigma_2^2$	$\nu = n_1 + n_2 - 2$

Example: Gender and Spending

We want to know if there is a difference in the amount that males and females spend in a particular shop.

The null and alternative hypotheses are:

$$H_0: \quad \mu_1 - \mu_2 = 0$$

$$H_a: \quad \mu_1 - \mu_2 \neq 0$$

If we test at the $\alpha = 0.05$ level, the critical values are $\pm z_{0.025} = \pm 1.96$ for large samples (replace with $\pm t_{\nu,\,0.025}$ for small samples).

Example: Gender and Spending

A sample of individuals were randomly selected and the results are:

	Male	Female
sample size	35	40
mean	54 €	56 €
variance	20 €^{2}	15 €^{2}

These are large samples so we will use z, the test statistic is:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_{01} - \mu_{02})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(54 - 56) - (0)}{\sqrt{\frac{20}{35} + \frac{15}{40}}} = \frac{-2}{0.9728} = -2.06.$$

Since $z = -2.06$ is outside of ± 1.96 (albeit only just), we reject H_0 at the 5% level of significance.

Example: Gender and Spending

We have rejected H_0 in favour of $H_a: \ \mu_1 - \mu_2 \neq 0$.

Conclusion: males and females spend different amounts in this shop.

Note that we can also calculate a p-value. Since this is a two-tailed test we have:

$$\text{p-value} = 2 \times \Pr(Z > |-2.06|) = 2 \times \Pr(Z > 2.06) = 2\,(0.01970)$$
$$= 0.0394.$$

The p-value is small (less than 0.05) which means that this data is unlikely assuming H_0 to be true $\Rightarrow H_0$ is unlikely.

Question 1

Consider the following samples:

Sample 1	Sample 2
$n_1 = 41$	$n_2 = 33$
$\bar{x}_1 = 20.5$	$\bar{x}_2 = 19.3$
$s_1^2 = 12.1$	$s_2^2 = 14.3$

Test the hypothesis that there is no difference between the true means, i.e., $\mu_1 - \mu_2 = 0$.

a) State the null and alternative hypotheses.

b) Calculate the test statistic and, hence, the p-value.

c) State your conclusion.

Small Samples

Note, for two small samples, we can either use the unequal variance approach or the equal variance approach.

- Unequal variance: the standard error is the same as for large samples, i.e., $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.
- Equal variance: the standard error is $\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$ and we must calculate the pooled variance, s_p^2.

Small samples $\Rightarrow$ use t-value. There is a formula for calculating the degrees of freedom, ν, in each case.

(note: if we wish to assume equal variances, then we **must** carry out the F-test first)

Example: CPU Temperature

A manufacturer of CPUs wishes to know if a new design runs cooler than the current model. Thus, a small sample of prototype CPUs will be produced and compared to a small sample of the current model.

By default, the company want to assume the current model is superior, i.e., that the temperature of the old model is lower: $\mu_{current} - \mu_{new} \leq 0$.

The null and alternative hypotheses are therefore:

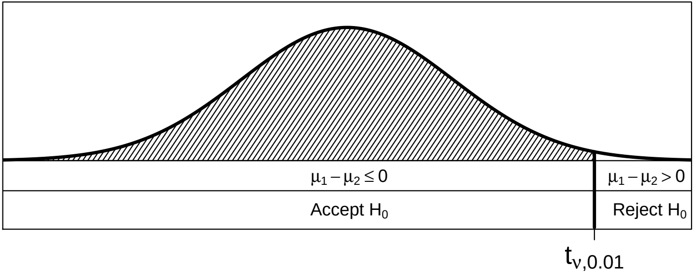
$$\begin{aligned} H_0 &: \quad \mu_1 - \mu_2 \leq 0 \\ H_a &: \quad \mu_1 - \mu_2 > 0 \end{aligned}$$

Note that we are using index 1 for the current model and index 2 for the new model.

Example: CPU Temperature

Assume that the company want to carry out the test at the 1% level of significance $\Rightarrow \alpha = 0.01$. (note: one-tailed test $\Rightarrow$ do not divide α by 2)

Since the alternative hypothesis points towards the rejection region, the critical value is $t_{\nu,0.01}$. The value of ν will depend on n_1, n_2 and the approach we use (equal or unequal variance).



Example: CPU Temperature

The data was collected and the results are:

	Current	New
sample size	6	4
mean	32.0 °C	29.2 °C
standard deviation	2.4 °C	1.3 °C

Note that we have $s_1 = 2.4$ and $s_2 = 1.3$.

For our calculations we will need:

$$s_1^2 = (2.4)^2 = 5.76 \qquad s_2^2 = (1.3)^2 = 1.69$$

Example: CPU Temperature (Equal Variances)

If we wish to assume **equal variances** then we must first carry out the F-test to see that this assumption is reasonable.

The null and alternative hypotheses for the F-test are:

$$H_0: \quad \sigma_1^2 = \sigma_2^2$$

$$H_a: \quad \sigma_1^2 \neq \sigma_2^2$$

The test statistic is $F = \frac{s^2_{\text{larger}}}{s^2_{\text{smaller}}} = \frac{5.76}{1.69} = 3.41$.

$$\left.\begin{aligned} \nu_1 &= n_{\text{top}} - 1 = 6 - 1 = 5 \\ \nu_2 &= n_{\text{bottom}} - 1 = 4 - 1 = 3 \end{aligned}\right\} \Rightarrow F_{5,3} = 14.9.$$

(Note: $F_{5,3} = 14.9$ is the value in brackets from the F tables which corresponds to the 5% level. For simplicity we will *always* use the 5% level for the F test.)

Example: CPU Temperature (Equal Variances)

For the F-test, the rejection region is that *above* the critical value.

Since $F = 3.41$ is below 14.9, we accept $H_0 : \sigma_1^2 = \sigma_2^2$. Hence, the equal variance assumption is reasonable.

We will need the pooled variance:

$$s_p^2 = \frac{(n_1 - 1)\,s_1^2 + (n_2 - 1)\,s_2^2}{n_1 + n_2 - 2} = \frac{(6-1)\,(5.76) + (4-1)\,(1.69)}{6+4-2}$$
$$= \frac{33.87}{8} = 4.234.$$

Also note that $\nu = n_1 + n_2 - 2 = 6 + 4 - 2 = 8$.

Example: CPU Temperature (Equal Variances)

Recall that

$$H_0: \quad \mu_1 - \mu_2 \leq 0$$

$$H_a: \quad \mu_1 - \mu_2 > 0$$

and we are testing at the $\alpha = 0.01$ level $\Rightarrow t_{8,\,0.01} = 2.896$ is the critical value (above which the rejection region lies).

The test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_{01} - \mu_{02})}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{(32.0 - 29.2) - (0)}{\sqrt{\frac{4.234}{6} + \frac{4.234}{4}}} = \frac{2.8}{1.3282} = 2.11.$$

Since this is below 2.896, we cannot reject H_0 at the 1% level.

Conclusion: we will continue to assume that the current CPU is superior to the new model (i.e., the current model runs cooler).

Example: CPU Temperature (Unequal Variances)

For the **unequal variance** approach we need to calculate:

$$a = \frac{s_1^2}{n_1} = \frac{5.76}{6} = 0.96, \qquad b = \frac{s_2^2}{n_2} = \frac{1.69}{4} = 0.4225.$$

$$\Rightarrow \nu = \frac{(a+b)^2}{\frac{a^2}{n_1-1} + \frac{b^2}{n_2-1}} = \frac{(0.96+0.4225)^2}{\frac{0.96^2}{6-1} + \frac{0.4225^2}{4-1}} = \frac{1.911}{0.2438} = 7.84.$$

Since only whole number ν values appear in the tables, we round this to $\nu = 8 \Rightarrow$ the critical value is $t_{8,0.01} = 2.896$.

Example: CPU Temperature (Unequal Variances)

The test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_{01} - \mu_{02})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(32.0 - 29.2) - (0)}{\sqrt{\frac{5.76}{6} + \frac{1.69}{4}}} = \frac{2.8}{1.1758} = 2.38.$$

Since this is below 2.896, we cannot reject H_0 at the 1% level, i.e., the conclusion is the same as the equal variance approach here.

Difference Between Two Proportions

When comparing two proportions via hypothesis testing, the standard error for the difference is:

$$\boxed{s(\widehat{P}_1 - \widehat{P}_2) = \sqrt{\frac{\hat{p}_c\,(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c\,(1-\hat{p}_c)}{n_2}}},$$

where $\hat{p}_c$ is the **combined proportion** over both groups:

$$\boxed{\hat{p}_c = \frac{x_1 + x_2}{n_1 + n_2}}.$$

Thus, as was the case with one proportion, the standard error is different to that used in the case of a confidence interval.

Difference Between Two Proportions

For testing the difference between two proportions we have:

Parameter	Statistic	Standard Error	Samples
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}$ where $\hat{p}_c = \frac{x_1 + x_2}{n_1 + n_2}$	large (hypothesis test)

Example: Concept Design

A car manufacturer has developed a concept design. The company wish to test the hypothesis that there is no difference between the proportions of younger and older people who like the concept design using the 10% level of significance.

The null and alternative hypotheses are

$$H_0: \quad p_1 - p_2 = 0$$
$$H_a: \quad p_1 - p_2 \neq 0$$

With $\alpha = 0.1$, and this being a two tailed test, the critical values are $\pm z_{0.05} = \pm 1.64$.

Example: Concept Design

The following data was collected:

	Younger	Older
Sample size	158	91
Like concept	122	60

Therefore

$$\hat{p}_1 = \frac{122}{158} = 0.7722, \qquad \hat{p}_2 = \frac{60}{91} = 0.6593,$$

and

$$\hat{p}_c = \frac{122 + 60}{158 + 91} = \frac{182}{249} = 0.7309.$$

Example: Concept Design

The test statistic is:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_{01} - p_{02})}{\sqrt{\frac{\hat{p}_c\,(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c\,(1-\hat{p}_c)}{n_2}}} = \frac{(0.7722 - 0.6593) - 0}{\sqrt{\frac{0.7309\,(0.2691)}{158} + \frac{0.7309\,(0.2691)}{91}}}$$

$$= \frac{0.1129}{0.0584} = 1.933.$$

As this is outside of ± 1.64, we reject the null hypothesis at the 10% level $\Rightarrow$ we accept that the true proportions are different.

Conclusion: The concept design is more appealing to younger people.

Example: Concept Design

We can also calculate a p-value. Being a two-tailed test, this is:

$$\begin{aligned} \text{p-value} &= 2 \times \Pr(Z > |1.933|) = 2 \times \Pr(Z > 1.933) \\ &= 2(0.0268) = 0.0536. \end{aligned}$$

The p-value is small and, hence, there is evidence against H_0.

Clearly we can reject H_0 at the 10% level of significance but it is worth noting that the evidence against H_0 is almost at the 5% level.

R Code: `t.test`

The code for producing a confidence interval using `t.test` was given in Lecture14. You will notice that the output of the `t.test` function contains more than just a confidence interval.

The function also displays the test statistic and corresponding p-value for a hypothesis test.

The direction of the test is set using the `alternative` option which must be one of `"two.sided"` (the default), `"less"` or `"greater"`.

The hypothesised value is set using the `mu` option (default is `mu=0`).

(note: `mu` represents the hypothesised mean in the case of one group and the hypothesised *difference* in means in the case of two groups)

R Code: One Mean

Consider the data from slide 14 of Lecture15 where we tested the hypothesis that the average CPU speed is 2.5Ghz. Thus:

$$H_0: \quad \mu = 2.5$$

$$H_a: \quad \mu \neq 2.5$$

We can test this hypothesis as follows:

```
cpu = c(2.53, 2.55, 2.54, 2.24)
t.test(cpu, mu=2.5)
```

Note: we do not have to specify `alternative` here since, by default this is already `alternative="two.sided"`.

R Code: One Mean

Output of `t.test(cpu, mu=2.5)`:

```
One Sample t-test

data:  cpu
t = -0.466, df = 3, p-value = 0.673
alternative hypothesis:  true mean is not equal to 2.5
95 percent confidence interval:
 2.225963 2.704037
sample estimates:
mean of x
 2.465
```

- `t = -0.466` is the test statistic.
- `df = 3` is the degrees of freedom.
- `p-value = 0.673` $\Rightarrow$ no evidence to reject $H_0 : \mu = 2.5$.
- Also note that the 95% CI contains the value 2.5Ghz.

R Code: One Mean

If we wanted to assume that the average speed was 2.5Ghz or more then we would have:

$$H_0: \quad \mu \geq 2.5$$
$$H_a: \quad \mu < 2.5$$

We can test this hypothesis as follows:

```
t.test(cpu, alternative="less", mu=2.5)
```

R Code: One Mean

Output of `t.test(cpu, alternative="less", mu=2.5)`:

```
One Sample t-test

data:  cpu
t = -0.466, df = 3, p-value = 0.3365
alternative hypothesis:  true mean is less than 2.5
95 percent confidence interval:
 -Inf 2.641764
```

- Note that the test statistic and degrees of freedom are the same but the p-value has changed since this is a one-tailed test.
- `p-value = 0.3365` $\Rightarrow$ no evidence to reject $H_0 : \mu \geq 2.5$.
- R also produces a one-tailed confidence interval - **ignore this**.

R Code: Difference Between Means

Consider the example on slide 18 of Lecture14. We wish to test the hypothesis that graduates from two universities have the same salary:

$$H_0: \quad \mu_1 - \mu_2 = 0$$

$$H_a: \quad \mu_1 - \mu_2 \neq 0$$

We can test this hypothesis as follows:

```
uni1 = c(32.1, 32.4, 33.2, 33.3, 33.6)
uni2 = c(35.7, 36.3, 39.4, 40.5)
t.test(uni1, uni2)
```

We do not need to specify `alternative="two.sided"` or `mu=0` since these are the defaults.

R Code: Difference Between Means

Output of `t.test(uni1, uni2)`:

```
 Welch Two Sample t-test

data:  uni1 and uni2
t = -4.2023, df = 3.359, p-value = 0.01961
alternative hypothesis:
true difference in means is not equal to 0
95 percent confidence interval:
 -8.661686 -1.448314
sample estimates:
mean of x mean of y
 32.920 37.975
```

- `p-val = 0.01961` $\Rightarrow$ strong evidence against $H_0 : \mu_1 - \mu_2 = 0$.
- We are not assuming equal variances here. In order to do so, use the option `var.equal=TRUE`.

R Code: Difference Between Means

Consider the example from slide 12 of this lecture where we compared the temperatures of two CPU designs. In this case, it was assumed that the current CPU runs cooler and hence:

$$H_0: \quad \mu_1 - \mu_2 \leq 0$$

$$H_a: \quad \mu_1 - \mu_2 > 0$$

We can test this hypothesis as follows:

```
cpu1 = c(33, 31, 32, 35, 33, 28)
cpu2 = c(29, 28, 29, 31)
t.test(cpu1, cpu2, alternative="greater")
```

If we wish to assume equal variances then use:

```
t.test(cpu1, cpu2, alternative="greater",
                      var.equal=TRUE)
```

R Code: Difference Between Means

Not assuming equal variances (compare with slide 19 of this lecture):

```
  Welch Two Sample t-test

data:  cpu1 and cpu2
t = 2.3853, df = 7.802, p-value = 0.02247
alternative hypothesis:
true difference in means is greater than 0
```

Assuming equal variances (compare with slide 17 of this lecture):

```
  Two Sample t-test

data:  cpu1 and cpu2
t = 2.1056, df = 8, p-value = 0.03417
alternative hypothesis:
true difference in means is greater than 0
```

R Code: Ratio of Variances

The F-test investigates the ratio of two variances. The null and alternative hypotheses for this test are:

$$H_0: \quad \sigma_1^2 = \sigma_2^2$$

$$H_a: \quad \sigma_1^2 \neq \sigma_2^2$$

For the CPU data we can carry out this test as follows:

```
var.test(cpu1, cpu2)
```

R Code: Ratio of Variances

Output of `var.test(cpu1, cpu2)`:

```
 F test to compare two variances

data:  cpu1 and cpu2
F = 3.5368, num df = 5,denom df = 3, p-value = 0.3274
alternative hypothesis:
true ratio of variances is not equal to 1

95 percent confidence interval:
 0.237614 27.458590
```

- `p-val = 0.3274` $\Rightarrow$ no evidence to reject $H_0 : \sigma_1^2 = \sigma_2^2$.
- The confidence interval is for the *ratio* of the true variances. Note that the above interval supports a true ratio of 1, i.e., no difference in variances.

R Code: One Proportion

For proportions, we use `prop.test`.

Consider the example from slide 34 of Lecture15 where a company wished to test the hypothesis that 70% of its users are teenagers:

$$H_0: \quad p = 0.7$$

$$H_a: \quad p \neq 0.7$$

The hypothesis test is carried out as follows:

```
x = 40
n = 75

prop.test(x, n, p=0.7)
```

R Code: One Proportion

Output of `prop.test(x, n, p=0.7)`:

```
data:  x out of n, null probability 0.7
X-squared = 9.1429, df = 1, p-value = 0.002497
alternative hypothesis:  true p is not equal to 0.7
95 percent confidence interval:
 0.4151494 0.6480820
sample estimates:
 p
0.5333333
```

- `p-val = 0.002` $\Rightarrow$ strong evidence against $H_0 : p = 0.7$.
- Note: R uses a different method for proportion tests to what we use in this course - so the results may differ slightly to what we do by hand.

R Code: One Proportion

Consider the example from slide 36 of Lecture15 concerning the opinion of a new operating system. Recall we had:

$$H_0: \quad p \leq 0.5$$

$$H_a: \quad p > 0.5$$

The hypothesis test is carried out as follows:

```
x = 38
n = 65

prop.test(x, n, alternative="greater", p=0.5)
```

Again the result will differ slightly from what we have done.

R Code: Two Proportions

For the concept car example on slide 22 of this lecture we had

$$H_0 : \quad p_1 - p_2 = 0$$

$$H_a : \quad p_1 - p_2 \neq 0$$

```
x = c(122,60)
n = c(158,91)

prop.test(x, n)
```

The result will differ slightly from what we have done.