

MTH5129 Probability & Statistics II  
Coursework 9

1. 14 male rats were sent into space. On their return, the red blood cell mass (in millilitres) of the rats were determined. A control group of 14 rats was held in the same conditions (except for space flight) and their red blood cell mass was determined. The sample mean and standard deviation of the flight animals were 7.881 and 1.017 respectively and of the control animals were 8.430 and 1.005. Test the claim that the flight animals have a different red blood cell mass to the control animals with a significance level of  $\alpha = 0.05$ . State clearly any assumptions and comment on them.

**Solution:** The null hypothesis is  $H_0 : \mu_1 = \mu_2$  and the alternative  $H_1 : \mu_1 \neq \mu_2$ .

The pooled estimate of variance is  $S_0^2 = 0.5S_1^2 + 0.5S_2^2$  since  $n_1 = n_2$ . Thus the observed value is

$$\begin{aligned} s_0^2 &= 0.5(1.017)^2 + 0.5(1.005)^2 \\ &= 1.022 \\ s_0 &= 1.011 \end{aligned}$$

The test statistic is

$$T = \frac{\bar{X} - \bar{Y}}{S_0 \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$T \sim t_{26}$  if  $H_0$  is true since  $n_1 + n_2 - 2 = 26$ .

Observed value of the test statistic is

$$t = \frac{7.881 - 8.430}{1.011 \sqrt{\left(\frac{1}{14} + \frac{1}{14}\right)}} = -1.44$$

```
> qt(0.975, 26)
[1] 2.055529
```

The rejection region is  $\{t : |t| > 2.056\}$  so we cannot reject  $H_0$  at the 5% significance level.

Assumptions are that the samples are independent and normally distributed with a common variance  $\sigma^2$ . The two samples would appear to be independent (although if the rats had been selected by choosing two from each of 14 litters they would not be). We cannot comment on normality as we have only the mean and standard deviation of the samples. The two standard deviations, and hence the sample variances, are close so the assumption of a common variance seems justified.

2. At the beginning of an academic year, students studying mathematics in their final year at a certain school were split into two tutorial groups, Group A (7 students) and Group B (11 students), with different teachers for the two groups. They all took the same mathematics examination at the end of the year and achieved the percentage marks tabulated below, where the marks for each group have been sorted into increasing order. The question to be investigated is whether there is significant evidence of a difference in the overall levels of performance between the two groups.

Group A	12	31	34	35	55	57	61				
Group B	26	29	36	38	48	56	59	60	67	74	75

- Test the hypothesis that there is no difference between the means of the two groups using a 1% significance level.
- Find the p value of the test.
- State the assumptions necessary for the test.
- Find a 95% confidence interval for the mean difference between the two groups.

**Solution:**

- a) We have  $\bar{x} = 40.7, \bar{y} = 51.6, s_1 = 17.7, s_2 = 17.4, n_1 = 7$  and  $n_2 = 11$ . So

$$s_0^2 = \frac{6 \times (17.7)^2 + 10 \times (17.4)^2}{16} \quad s_0 = 17.53.$$

The null hypothesis is  $H_0 : \mu_1 = \mu_2$  and the alternative  $H_1 : \mu_1 \neq \mu_2$ .

The test statistic is

$$T = \frac{\bar{X} - \bar{Y}}{S_0 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$T \sim t_{16}$  if  $H_0$  is true.

Observed value of the test statistic is

$$t = \frac{40.7 - 51.6}{17.53 \sqrt{\frac{1}{7} + \frac{1}{11}}} = -1.29$$

```
> qt(0.995, 16)
```

```
[1] 2.920782
```

The rejection region is  $\{t : |t| > 2.921\}$  so we cannot reject  $H_0$  at the 1% significance level.

- b) We can use either of the following R commands.

```
> 2*(pt(-1.29, 16))
```

```
[1] 0.2153845
```

```
> 2*(1-pt(1.29, 16))
```

```
[1] 0.2153845
```

- c) Assumptions are that the samples are independent and normally distributed with a common variance  $\sigma^2$ .

- d) The form of the 95% confidence interval is

$$\bar{x} - \bar{y} \pm t_{16}(0.975) s_0 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

```
> qt(0.975, 16)
[1] 2.119905
```

Substituting the values we find

$$\begin{aligned} 40.7 - 51.6 \pm 2.120 \times 17.53 \sqrt{\frac{1}{7} + \frac{1}{11}} \\ = -10.9 \pm 17.97 = (-28.87, 7.07). \end{aligned}$$

3. In a study of infants treated for pulmonary hypertension, the head circumference (in cm) of neonates treated for this condition (cases) was compared with others who were not treated (controls). Results were as follows

Cases	33.9	35.1	33.4	34.5	37.9
	31.3	32.5	32.9	36.3	34.2
Controls	35.2	35.6	36.7	35.1	36.0
	33.4	31.3	33.5	35.8	36.3
	34.3	33.1	32.4	35.1	33.6
	31.8	34.1	35.2	34.8	34.5
	31.6	31.9	31.9	32.8	34.0

Test the hypothesis that the population variance of head circumference for cases and controls are equal, use  $\alpha = 0.05$ .

**Solution:** For cases  $s_1^2 = 3.6366$  with  $n_1 = 10$ .

For controls  $s_2^2 = 2.5408$  with  $n_2 = 25$ .

The null hypothesis is  $H_0 : \sigma_1^2 = \sigma_2^2$  and the alternative  $H_1 : \sigma_1^2 \neq \sigma_2^2$ .

The test statistic is

$$F = \frac{S_1^2}{S_2^2}$$

$F \sim F_{24}^9$  if  $H_0$  is true.

Observed value of the test statistic is

$$F_{obs} = \frac{3.6366}{2.5408} = 1.431$$

We find the rejection region for  $\alpha = 0.05$  using R.

```
> qf(0.025, df1=9, df2=24)
[1] 0.2766867
> qf(0.975, df1=9, df2=24)
[1] 2.702711
```

So we reject if  $F_{obs} > 2.703$  or  $F_{obs} < 0.2767$ .

Thus we cannot reject  $H_0$  at the 5% significance level.

4. Two random samples were independently drawn from two normal populations. The first sample of size 8 had mean 14.1 and variance 107.3 and the second of size 11 had mean 17.5 and variance 20.7. Test the hypothesis that the populations have the same variance at the 5% significance level.

Test the hypothesis that the populations have the same mean at the 5% significance level.

**Solution:** The null hypothesis is  $H_0 : \sigma_1^2 = \sigma_2^2$  and the alternative  $H_1 : \sigma_1^2 \neq \sigma_2^2$ .

The test statistic is

$$F = \frac{S_1^2}{S_2^2}$$

$F \sim F_{10}^7$  if  $H_0$  is true.

Observed value of the test statistic is

$$F_{obs} = \frac{107.3}{20.7} = 5.18$$

We find the rejection region for  $\alpha = 0.05$  using R.

```

> qf(0.025, df1=7, df2=10)
[1] 0.2100348
> qf(0.975, df1=7, df2=10)
[1] 3.949824

```

So we reject if  $F_{obs} > 3.95$  or  $F_{obs} < 0.210$ .

Thus we can reject  $H_0$  at the 5% significance level.

This means we must use the approximate 2 sample test for the means.

The null hypothesis is  $H_0 : \mu_1 = \mu_2$  and the alternative  $H_1 : \mu_1 \neq \mu_2$ .

The test statistic is

$$T^* = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}}$$

$T^* \sim t_{\nu^*}$  if  $H_0$  is true.

$$\nu^* = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4/n_1^2}{n_1-1} + \frac{s_2^4/n_2^2}{n_2-1}}$$

So

$$\nu^* = \frac{(107.3/8 + 20.7/11)^2}{\frac{107.3^2/64}{7} + \frac{20.5^2/121}{10}} = 8.98.$$

Observed value of the test statistic is

$$t^* = \frac{14.1 - 17.5}{\sqrt{\frac{107.3}{8} + \frac{20.7}{11}}} = \frac{-3.4}{3.92} = -0.867$$

From R we have

```

> qt(.975, 8.98)
[1] 2.262925

```

The rejection region is  $\{t^* : |t^*| > 2.263\}$  so we cannot reject  $H_0$  at the 5% significance level.

You are given the following R output.

```
> qt(0.975,26)
[1] 2.055529
> qt(0.995,16)
[1] 2.920782
> 2*(pt(-1.29, 16))
[1] 0.2153845
> 2*(1-pt(1.29, 16))
[1] 0.2153845
> qt(0.975,16)
[1] 2.119905
> qf(0.025, df1=9, df2=24)
[1] 0.2766867
> qf(0.975, df1=9, df2=24)
[1] 2.702711
> qf(0.025, df1=7, df2=10)
[1] 0.2100348
> qf(0.975, df1=7, df2=10)
[1] 3.949824
> qt(.975, 8.98)
[1] 2.262925
```