

Question 1

Here $n = 18$, $\bar{x} = 40000$ and $s = 3125$.

- a) 80% confidence $\Rightarrow \alpha = 0.2$ remaining $\Rightarrow \alpha/2 = 0.1$ in each tail.

Small sample \Rightarrow use t value with degrees of freedom $\nu = n - 1 = 18 - 1 = 17$.

$$\begin{aligned} \bar{x} \pm t_{17, 0.1} \frac{s}{\sqrt{n}} \\ 40000 \pm 1.333 \frac{3125}{\sqrt{18}} \\ 40000 \pm 1.333(736.57) \\ 40000 \pm 981.85 \\ [39018.15, 40981.85] \end{aligned}$$

We are 80% confident that the true average salary of a software engineer, μ , is in the above interval.

- b) $\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$.

$$\begin{aligned} \bar{x} \pm t_{17, 0.025} \frac{s}{\sqrt{n}} \\ 40000 \pm 2.110(736.57) \\ 40000 \pm 1554.16 \\ [38445.84, 41554.16] \end{aligned}$$

- c) $\alpha = 0.01 \Rightarrow \alpha/2 = 0.005$.

$$\begin{aligned} \bar{x} \pm t_{17, 0.005} \frac{s}{\sqrt{n}} \\ 40000 \pm 2.898(736.57) \\ 40000 \pm 2134.58 \\ [37865.42, 42134.58] \end{aligned}$$

Note: to increase our confidence in capturing the true μ , the interval width increases.

Question 2

Variety 1

	Σ					
x_1	10	8	7	8	6	39
x_1^2	100	64	49	64	36	313

$$\bar{x}_1 = \frac{\sum x_1}{n_1} = \frac{39}{5} = 7.8.$$

$$s_1^2 = \frac{\sum x_1^2 - n_1 \bar{x}_1^2}{n_1 - 1} = \frac{313 - 5(7.8^2)}{4} = 2.2.$$

$$s = \sqrt{2.2} = 1.483.$$

Variety 2

	Σ						
x_2	5	6	8	6	7	7	39
x_2^2	25	36	64	36	49	49	259

$$\bar{x}_2 = \frac{\sum x_2}{n_2} = \frac{39}{6} = 6.5.$$

$$s_2^2 = \frac{\sum x_2^2 - n_2 \bar{x}_2^2}{n_2 - 1} = \frac{259 - 6(6.5^2)}{5} = 1.1.$$

$$s = \sqrt{1.1} = 1.049.$$

Thus we have:

Variety 1	Variety 2
$n_1 = 5$	$n_2 = 6$
$\bar{x}_1 = 7.8$	$\bar{x}_2 = 6.5$
$s_1^2 = 2.2$	$s_2^2 = 1.1$
$s_1 = 1.483$	$s_2 = 1.049$

- a) If we wish to assume that the true variances are equal (i.e., $\sigma_1^2 = \sigma_2^2$) we must carry out the F test first.

- b) The null and alternative hypotheses are:

$$\begin{aligned} H_0 : \sigma_1^2 &= \sigma_2^2 \\ H_a : \sigma_1^2 &\neq \sigma_2^2 \end{aligned}$$

The test statistic is

$$F = \frac{s_{\text{larger}}^2}{s_{\text{smaller}}^2} = \frac{2.2}{1.1} = 2.$$

The ν value that goes with s_{larger}^2 is $5 - 1 = 4$.
The ν value that goes with s_{smaller}^2 is $6 - 1 = 5$.

\Rightarrow The critical value is $F_{4,5} = 7.39$.

Question 2 continued

- b) Since $F = 2$ is within the acceptance region (i.e., it is below the critical value 7.39), we accept the null hypothesis that $\sigma_1^2 = \sigma_2^2$ at the 5% level of significance. Thus, it is reasonable to assume that the variances are equal.

(Note that the critical value, $F_{4,5} = 7.39$, was the one in brackets in the F tables; this corresponds to $\alpha/2 = 0.025 \Rightarrow \alpha = 0.05$. For simplicity, we will *always* use the 5% level of significance for F tests.)

- c) Since we are assuming equal variances, we will need to calculate the pooled variance:

$$\begin{aligned} s_p^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \\ &= \frac{(5 - 1)(2.2) + (6 - 1)(1.1)}{5 + 6 - 2} \\ &= \frac{4(2.2) + 5(1.1)}{9} \\ &= \frac{14.3}{9} = 1.5889. \end{aligned}$$

95% confidence $\Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$.

Degrees of freedom: $\nu = n_1 + n_2 - 2 = 9$.

Therefore, the t-value is: $t_{9,0.025} = 2.262$.

The 95% confidence interval for $\mu_1 - \mu_2$ is

$$\begin{aligned} &(\bar{x}_1 - \bar{x}_2) \pm t_{9,0.025} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \\ &(7.8 - 6.5) \pm 2.262 \sqrt{\frac{1.5889}{5} + \frac{1.5889}{6}} \\ &1.3 \pm 2.262 \sqrt{0.5826} \\ &1.3 \pm 2.262(0.7633) \\ &1.3 \pm 1.727 \\ &[-0.427, 3.027] \end{aligned}$$

We are 95% confident that the true difference, $\mu_1 - \mu_2$, lies in the interval $[-0.427, 3.027]$. Since this interval supports the possibility of no difference, $\mu_1 - \mu_2 = 0$, we cannot conclude that either variety is superior in quality.

Conclusion: The evidence suggests that there is no difference in the quality of stout produced using either variety of barley.

- d) The advantage of the unequal variance approach is it does not rely on the *assumption* of equal variances. Although we have carried out the F test and found the hypothesis $\sigma_1^2 = \sigma_2^2$ to be reasonable, it does not mean that it is true. For the unequal variance approach, the confidence interval is:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\nu,0.025} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Note that this is the same as the formula used for two large samples except that we have a t value rather than a z value.

The degrees of freedom must be calculated using a formula which first requires

$$\begin{aligned} a &= \frac{s_1^2}{n_1} = \frac{2.2}{5} = 0.44 \\ b &= \frac{s_2^2}{n_2} = \frac{1.1}{6} = 0.1833. \end{aligned}$$

$$\begin{aligned} \Rightarrow \nu &= \frac{(a + b)^2}{\frac{a^2}{n_1 - 1} + \frac{b^2}{n_2 - 1}} \\ &= \frac{(0.44 + 0.1833)^2}{\frac{0.44^2}{5 - 1} + \frac{0.1833^2}{6 - 1}} \\ &= \frac{0.3885}{0.0484 + 0.00672} = 7.048. \end{aligned}$$

There are only whole number ν values in the t-tables and, therefore, we use $\nu = 7$.

The 95% confidence interval for $\mu_1 - \mu_2$ is

$$\begin{aligned} &(\bar{x}_1 - \bar{x}_2) \pm t_{7,0.025} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\ &(7.8 - 6.5) \pm 2.365 \sqrt{\frac{2.2}{5} + \frac{1.1}{6}} \\ &1.3 \pm 2.365(0.7895) \\ &1.3 \pm 1.867 \\ &[-0.567, 3.167] \end{aligned}$$

Here, the confidence interval is much the same as in the equal variance approach. Hence, the conclusion is the same: there is no difference in quality using either variety of barley.

Question 3

This data is *paired* data (i.e., dependent) since the same individuals are being measured on two occasions. Thus, we need to work out the difference in times. Once we have done this, the procedure is the same as for *one* mean.

a)

Individual	1	2	3	4	5	6	7	
No Warm Up	13.6	12.8	12.3	11.7	12.0	13.3	10.5	Σ
Warm Up	13.9	12.4	12.2	11.6	11.9	12.7	10.4	
Difference: x	-0.3	0.4	0.1	0.1	0.1	0.6	0.1	1.1
x^2	0.09	0.16	0.01	0.01	0.01	0.36	0.01	0.65

$$\bar{x} = \frac{\sum x}{n} = \frac{1.1}{7} = 0.157142.$$

$$s^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \frac{0.65 - 7(0.157142^2)}{6} = 0.0795.$$

$$s = \sqrt{0.0795} = 0.282.$$

95% confidence $\Rightarrow \alpha = 0.05$ remaining $\Rightarrow \alpha/2 = 0.025$ in each tail.

Since the sample size is small, we use a t value with $\nu = n - 1 = 7 - 1 = 6$.

$$\bar{x} \pm t_{6,0.025} \frac{s}{\sqrt{n}}$$

$$0.157 \pm 2.447 \frac{0.282}{\sqrt{7}}$$

$$0.157 \pm 2.447 (0.1066)$$

$$0.157 \pm 0.2608$$

$$[-0.1038, 0.4178]$$

We are 95% confident that the true mean *difference*, μ , lies in the above interval. Note that the interval includes $\mu = 0$, i.e., the effect of warming up does not seem to change the 100m sprint time.