Estimation, confidence intervals and tests using a normal distribution 5C

1 $H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 > \mu_2$ 5% c.v. is $z = 1.6449$

t.s. is $z = \dfrac{(23.8 - 21.5) - 0}{\sqrt{\frac{5^2}{15} + \frac{4.8^2}{20}}} = 1.3699\ldots$

$1.3699 < 1.6449$ so result is not significant, accept H_0.

2 $H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2$ 5% c.v. is $z = \pm 1.96$

t.s. is $z = \dfrac{(51.7 - 49.6) - 0}{\sqrt{\frac{4.2^2}{30} + \frac{3.6^2}{25}}}$ Choose $\bar{x}_2 - \bar{x}_1$ to get $z > 0$

t.s. is $z = 1.996\ldots \quad > 1.96$ so result is significant, reject H_0.

3 $H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 < \mu_2$ 1% c.v. is $z = -2.3263$

t.s. is $z = \dfrac{(3.62 - 4.11) - 0}{\sqrt{\frac{0.81^2}{25} + \frac{0.75^2}{36}}} = -2.3946\ldots$

t.s. is $-2.3946\ldots < -2.3263$ so result is significant, reject H_0.

4
$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2$ 1% c.v. is $z = \pm 2.5758$

t.s. is $z = \dfrac{(112.0 - 108.1) - 0}{\sqrt{\frac{8.2^2}{85} + \frac{11.3^2}{100}}} = 2.712\ldots > 2.5758$

Significant result so reject H_0.

Central Limit Theorem applies since n_1, n_2 are large and enables you to assume $\bar{X}_1$ and $\bar{X}_2$ are both normally distributed.

5 $H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 > \mu_2$ 5% c.v. is $z = 1.96$

t.s. is $z = \dfrac{(72.6 - 69.5) - 0}{\sqrt{\frac{18.3^2}{100} + \frac{15.4^2}{150}}} = 1.396\ldots < 1.96$

Result is not significant so accept H_0.

Central Limit Theorem applies since n_1, n_2 are both large and enables you to assume $\bar{X}_1$ and $\bar{X}_2$ are normally distributed.

6 $H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 < \mu_2$ 1% c.v. is $z = -2.3263$

t.s. is $z = \dfrac{(0.863 - 0.868) - 0}{\sqrt{\frac{0.013^2}{120} + \frac{0.015^2}{90}}} = -2.5291\ldots < -2.3263$

Result is significant so reject H_0.

Central Limit Theorem is used to assume $\bar{X}_1$ and $\bar{X}_2$ are normally distributed since both samples are large.

7 $\sigma_1 = 0.011 \quad n_1 = 10 \quad \bar{x}_1 = 6.531$

$\sigma_2 = 0.015 \quad n_2 = 15 \quad \bar{x}_2 = 6.524$

$\text{H}_0 : \mu_1 = \mu_2 \;\; \text{H}_1 : \mu_1 \neq \mu_2, \quad$ 5% c.v. is $z = \pm 1.96$

t.s. is $z = \dfrac{(6.531 - 6.524) - 0}{\sqrt{\frac{0.011^2}{10} + \frac{0.015^2}{15}}} = 1.34466\ldots < 1.96$

Not significant.

Accept H_0.

There is insufficient evidence to suggest that the machines are producing pipes of different lengths.

8 a Let μ_{old} denote the mean yield of the old seed (in tonnes) and μ_{new} denote the mean yield of the new seed (also in tonnes). The null hypothesis is that the difference between the means is 1. The alternative hypothesis is that the difference is greater than 1.

$H_0 : \mu_{new} - \mu_{old} = 1$

$H_1 : \mu_{new} - \mu_{old} > 1.$

We have standard deviations and sample sizes of

$\sigma_{new} = \sqrt{0.8} = 0.8944,$

$n_{new} = 80,$

$\sigma_{old} = \sqrt{0.6} = 0.7746,$

$n_{old} = 70.$

$$z = \frac{\bar{x}_{new} - \bar{x}_{old} - (\mu_{new} - \mu_{old})}{\sqrt{\frac{\sigma_{new}{}^2}{n_{new}} + \frac{\sigma_{old}{}^2}{n_{old}}}}$$

The value of the test statistic is $= \dfrac{6.5 - 5 - (1)}{\sqrt{\frac{0.8}{80} + \frac{0.6}{70}}}$

$= 3.669$ (3 d.p.)

(Note that we used the null hypothesis of $H_0 : \mu_{new} - \mu_{old} = 1$ in this calculation).

The 5% (one-tailed) critical value for z is $z = 1.6449$ so our test statistic value is significant. So we reject H_0 and conclude that the mean yield of the new seed is more than 1 tonne greater than the mean yield of the old seed.

b The central limit theorem is relevant since the mean yield is normally distributed (sample size is large).

9 a Let μ_x denote the mean fat content per litre of milk for grain only fed cows and μ_y denote the mean fat content per litre of milk for grain/grass fed cows. The null hypothesis is that the difference between the means is 0. The alternative hypothesis is that the difference is greater than 0.

$H_0 : \mu_{grain} - \mu_{grain/grass} = 0$

$H_1 : \mu_{grain} - \mu_{grain/grass} \neq 0.$

We have standard deviations and sample sizes of

$\sigma_{grain} = \sqrt{0.8} = 0.8944,$

$n_{grain} = 60,$

$\sigma_{grain/grass} = \sqrt{0.75} = 0.8660,$

$n_{grain/grass} = 50.$

The value of the test statistic is:

$$z = \frac{\overline{x}_{grain} - \overline{x}_{grain/grass} - \left(\mu_{grain} - \mu_{grain/grass}\right)}{\sqrt{\frac{\sigma_{grain}^2}{n_{grain}} + \frac{\sigma_{grain/grass}^2}{n_{grain/grass}}}}$$

$$= \frac{4.1 - 3.7 - (0)}{\sqrt{\frac{0.8}{60} + \frac{0.75}{50}}}$$

$$= 2.376 \text{ (3 d.p.)}$$

(Note that we used the null hypothesis of $H_0 : \mu_{grain} - \mu_{grain/grass} = 0$ in this calculation).

The 5% (two-tailed) critical value for z is $z = 1.96$ so our test statistic value is significant. So we reject H_0 and conclude that the mean fat content per litre of milk from cows with a grain only diet is different from the mean fat content per litre of milk from cows with a grain/grass diet.

b We have assumed that mean fat content is normally distributed.

Challenge

a We find $\hat{\mu}$ the same way we find any other mean.
Sum all samples and divide by the number of samples.

This is written $\hat{\mu} = \dfrac{\sum_1^{n_x} x_i + \sum_1^{n_y} y_i}{\text{(number of samples)}}$.

We know that there were n_x samples of x, and n_y samples of y.
We also know that the sum of all samples of x is going to be equal to the sample mean of x, multiplied by the number of samples n_x.

Similarly for y, we find that $\sum_1^{n_y} y_i = n_y\bar{y}$.

Thus, $\hat{\mu} = \dfrac{\sum_1^{n_x} x_i + \sum_1^{n_y} y_i}{\text{(number of samples)}} = \dfrac{n_x\bar{x} + n_y\bar{y}}{n_x + n_y}$.

b In order to find a 99% confidence interval, we note that the variance is equal to

$$\sigma^2 = \frac{n_x{}^2\sigma_x{}^2 + n_y{}^2\sigma_y{}^2}{(n_x + n_y)^2} = \frac{100^2 \times 16 + 120^2 \times 24}{(100+120)^2} = \frac{1264}{121}$$

and so the standard deviation is $\sigma = \sqrt{\dfrac{1264}{121}} = 3.232$.

We also know from part **a** that

$$\hat{\mu} = \frac{n_x\bar{x} + n_y\bar{y}}{n_x + n_y}$$
$$= \frac{100 \times 46 + 120 \times 47}{100 + 120}$$
$$= \frac{512}{11}.$$

Using the statistics tables, we find that the value that corresponds with a 99% confidence interval is 2.58. This means our 99% confidence interval is found by substituting the values we have found into the equation:

$$\hat{\mu} \pm 2.58 \times \frac{\sigma}{\sqrt{n_x + n_y}} = \frac{512}{11} \pm 2.58 \times \frac{3.232}{\sqrt{100+120}}$$
$$= 46.545 \pm 0.562$$
$$= (45.983,\ 47.107)$$