Confidence intervals

Starter

1. **(Review of last lesson)**

A fish species in Wales is normally distributed with mean body length of 88.2 mm and variance 9 mm. Researchers believe the same species in Scotland is larger. A random sample of 40 taken from Scotland gave the values $\sum x_i = 3569.4$ and $\sum x_i^2 = 318935.9$. Test at the $5\,\%$ level, using the variance value calculated from the data of the random sample, whether the fish off the coast of Scotland are longer.

Working:

$$\bar{x} = \frac{3569.4}{40} = 89.235$$

$$s^2 = \frac{n}{n-1}\left(\frac{\sum x_i^2}{n} - \bar{x}^2\right)$$
$$= \frac{40}{39}\left(\frac{318935.9}{40} - 89.235^2\right)$$
$$= 10.78$$

By the central limit theorem: $\bar{X}_{40} \sim \mathrm{N}\left(83.7, \frac{10.78}{40}\right)$

$H_0 : \mu = 88.2$ (the fish off the coast of Scotland are not longer)
$H_1 : \mu > 88.2$ (the fish off the coast of Scotland are longer)

$p-$value method

$P(\bar{X} > 89.235) = 0.0231 \equiv 2.31\,\%$

Since $2.31\,\% < 5\,\%$, $\bar{x} = 89.235$ ***does lie*** in the critical region.

Therefore, we ***reject*** H_0 and conclude that there is evidence to suggest the fish off the coast of Scotland are longer.

Critical value method

Let x_{cv} be the critical value such that $P(\bar{X} > x_{cv}) = 0.05 \Rightarrow$
$x_{cv} = 89.1$

Since $\bar{x} = 89.235 > 89.1 = x_{cv}$, we reject H_0 and conclude that there is evidence to suggest to suggest the fish off the coast of Scotland are longer.

E.g. 1 (a) Given that $P(\bar{x} - a < \bar{X} < \bar{x} + a) = 0.95$, find the $Z-$ value that corresponds to a i.e $P(-z < Z < z) = 0.95$. Give your answer to 3 s.f..

(b) Using the formula $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ and the value found in (a), rearrange to find an inequality of the form $-k < \mu < k$. This is the $95\,\%$ confidence interval.

(c) Find similar inequalities of the form $-k < \mu < k$ for a:

(i) $90\,\%$ confidence interval

(ii) $98\,\%$ confidence interval

(d) State the width of the $98\,\%$ confidence interval.

Working: (a) $P(-z < Z < z) = 0.95 \quad \Rightarrow \quad P(Z < z) = 0.975$

$z = 1.96$ (3 s.f.)

(b) $P(-1.96 < Z < 1.96) = 0.95$

Let $z = 1.96$

Using $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$: $\quad 1.96 = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$$1.96 \times \frac{\sigma}{\sqrt{n}} = \bar{x} - \mu$$

$$\mu = \bar{x} - 1.96 \times \frac{\sigma}{\sqrt{n}}$$

When $z = -1.96$, $\quad \mu = \bar{x} + 1.96 \times \dfrac{\sigma}{\sqrt{n}}$

95% CI: $\quad \bar{x} - 1.96 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \times \dfrac{\sigma}{\sqrt{n}}$

(c) (i) 90% confidence interval $\Rightarrow P(-z < Z < z) = 0.90$

$\Rightarrow \quad P(Z < z) = 0.95$ so $z = 1.64$ (3 s.f.)

90% CI: $\quad \bar{x} - 1.64 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.64 \times \dfrac{\sigma}{\sqrt{n}}$

(ii) 98% confidence interval $\Rightarrow P(-z < Z < z) = 0.98$

$\Rightarrow \quad P(Z < z) = 0.99$ so $z = 2.33$ (3 s.f.)

98% CI: $\quad \bar{x} - 2.33 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.33 \times \dfrac{\sigma}{\sqrt{n}}$

(d) The width of the 98% CI is $2 \times 2.33 \times \dfrac{\sigma}{\sqrt{n}}$.

E.g. 2 The number of hours for which a certain light bulb stays lit is normally distributed with a standard deviation 45 hours. A random sample of 60 lightbulbs is tested and the sample mean is 734 hours. Calculate these confidence intervals for μ.

(a) 90% (b) 98%

Working: (a) $\bar{x} = 734, \sigma = 45$

90% CI: $\quad \bar{x} - 1.64 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.64 \times \dfrac{\sigma}{\sqrt{n}}$

$$734 - 1.64 \times \frac{45}{\sqrt{60}} < \mu < 734 + 1.64 \times \frac{45}{\sqrt{60}}$$

The 90% confidence interval for this sample is $724.5 < \mu < 743.5$.

(b) 98% CI: $\quad \bar{x} - 2.33 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.33 \times \dfrac{\sigma}{\sqrt{n}}$

$$734 - 2.33 \times \frac{45}{\sqrt{60}} < \mu < 734 + 2.33 \times \frac{45}{\sqrt{60}}$$

The 98% confidence interval for this sample is $720.5 < \mu < 747.5$.

E.g. 3 The $95\,\%$ confidence interval for a mean of μ is 85.3 ± 2.35. Find the following confidence intervals for μ.

(a) $90\,\%$ (b) $99\,\%$

Working: (a) $\bar{x} = 85.3$

$95\,\%$ CI: $\bar{x} - 1.96 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \times \dfrac{\sigma}{\sqrt{n}}$

So $1.96\dfrac{\sigma}{\sqrt{n}} = 2.35$

$$\frac{\sigma}{\sqrt{n}} = \frac{1.96}{2.35} \approx 0.834$$

$90\,\%$ CI: $\bar{x} - 1.64 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.64 \times \dfrac{\sigma}{\sqrt{n}}$

$$85.3 - 1.64 \times \frac{196}{235} < \mu < 85.3 + 1.64 \times \frac{196}{235}$$

The $90\,\%$ confidence interval is $83.33 < \mu < 82.27$.

(b) $99\,\%$ confidence interval $\Rightarrow$ $P(-z < Z < z) = 0.99$

$\Rightarrow$ $P(Z < z) = 0.995$ so $z = 2.58$ (3 s.f.)

$99\,\%$ CI: $\bar{x} - 2.58 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.58 \times \dfrac{\sigma}{\sqrt{n}}$

$$85.3 - 2.58 \times \frac{196}{235} < \mu < 85.3 + 2.58 \times \frac{196}{235}$$

The $99\,\%$ confidence interval is $82.21 < \mu < 88.39$.

E.g. 4 A normal distribution has standard deviation 8. Estimate the smallest sample size required if these confidence levels should have a width of less than 1.5.

(a) $95\,\%$ (b) $90\,\%$

Working: (a) $95\,\%$ CI: $\bar{x} - 1.96 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \times \dfrac{\sigma}{\sqrt{n}}$

Width of less than 1.5: $2 \times 1.96 \times \dfrac{8}{\sqrt{n}} < 1.5$

$$\sqrt{n} > 2 \times 1.96 \times \frac{8}{1.5}$$

$$n > 437.1$$

The smallest sample size is 438.

(b) $95\,\%$ CI: $\bar{x} - 1.64 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.64 \times \dfrac{\sigma}{\sqrt{n}}$

Width of less than 1.5: $2 \times 1.64 \times \dfrac{8}{\sqrt{n}} < 1.5$

$$\sqrt{n} > 2 \times 1.64 \times \frac{8}{1.5}$$

$$n > 306.02$$

The smallest sample size is 307.

E.g. 5 An exam board knows that each year the standard deviation of the marks in a certain subject is 13.5 but the mean mark, μ, will fluctuate. An examiner wishes to estimate the mean mark of all candidates nationally before all results are in and takes a random sample of 250 results. The sample mean is 68.4.

(a) Calculate a $95\ \%$ confidence interval for μ.

(b) Once all results were is, the actual value of μ was 65.4. What conclusions might the examiner draw about the sample?

Working: (a) Since the distribution of the population is not stated, we can use the central limit theorem to state that the distribution of the sample mean is $\bar{X}_{250} \sim \mathrm{N}\left(\mu, \dfrac{13.5^2}{250}\right)$

$\bar{x} = 68.4$

$95\ \%$ CI: $\quad \bar{x} - 1.96 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \times \dfrac{\sigma}{\sqrt{n}}$

$$68.4 - 1.96 \times \frac{13.5}{\sqrt{250}} < \mu < 68.4 + 1.96 \times \frac{13.5}{\sqrt{250}}$$

The $95\ \%$ confidence interval for this sample is $66.7 < \mu < 70.1$.

(b) Since the actual mean lies outside the confidence interval, it suggests that the sample was not representative of the population.

E.g. 6 A random sample of 80 chickens is weighed and it is found that $\sum x_i = 136.85$ kg and $\sum x_i^2 = 264.97$ kg², where x_i are the weights of the chickens in the sample. Find the $95\ \%$ confidence interval for μ.

Working: $\bar{x} = \dfrac{136.85}{80} = 1.710625$ kg

Since the variance of the population is not given, it must be calculated from the sample.

$s^2 = \dfrac{n}{n-1}\left(\dfrac{\sum x_i^2}{n} - \bar{x}^2\right)$: $\quad s^2 = \dfrac{n}{n-1}\left(\dfrac{\sum x_i^2}{n} - \bar{x}^2\right)$

$$= \frac{80}{79}\left(\frac{264.97}{80} - \left(\frac{136.85}{80}\right)^2\right)$$

$$\approx 0.391$$

Using the central limit theorem: $\quad \bar{X}_{80} \sim \mathrm{N}\left(\mu, \dfrac{0.391}{80}\right)$

$95\ \%$ CI: $\quad \bar{x} - 1.96 \times \dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \times \dfrac{\sigma}{\sqrt{n}}$

$$\frac{136.85}{80} - 1.96 \times \frac{0.391}{\sqrt{80}} < \mu < \frac{136.85}{80} + 1.96 \times \frac{0.391}{\sqrt{80}}$$

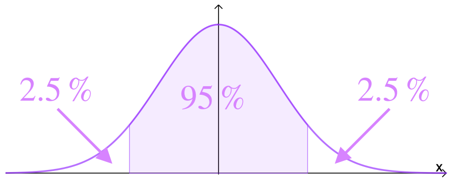
The $95\ \%$ confidence interval for this sample is $1.413 < \mu < 1.848$.

E.g. 7 State the percentage of the required confidence level for these hypothesis tests.

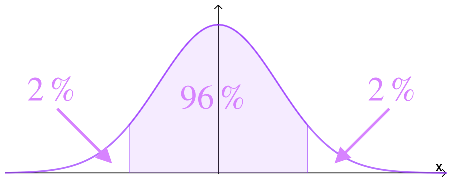
(a) A two-tailed test at the 5% significance level.
(b) A one-tailed test at the 2% significance level.
(c) A two-tailed test at the $\alpha\%$ significance level.
(d) A one-tailed test at the $\alpha\%$ significance level.

Working:

(a) Two-tailed tests split the significant level percentage equally either side of the mean (see diagram). So a 95% confidence interval would be needed



(b) One-tailed tests have the significant level percentage all on one side (see diagram). So a 96% confidence interval would be needed



(c) $(100\% - \alpha\%)$ confidence interval

(d) $(100\% - 2\alpha\%)$ confidence interval

E.g. 8 A tobacco company claims that the tar content of its cigarettes is 18.72 mg. Anti-smoking campaigners employ a research company as they think the real value is much higher. The tar content of 200 cigarettes is measure and the results, $\sum t_i = 3928.649$ mg and $\sum t_i^2 = 82415.74$ mg², where t_i is the tar content of the cigarettes. By finding a suitable confidence interval, test the company's claim at the 2% level.

Working: $\bar{x} = \dfrac{3928.649}{200} = 19.64$ mg (4 s.f.)

Since the variance of the population is not given, it must be calculated from the sample.

$$s^2 = \frac{n}{n-1}\left(\frac{\sum x_i^2}{n} - \bar{x}^2\right): \quad s^2 = \frac{200}{199}\left(\frac{82415.74}{200} - \left(\frac{3928.649}{200}\right)^2\right)$$
$$\approx 0.391$$

Using the central limit theorem: $\bar{X}_{200} \sim \mathrm{N}\left(18.72, \dfrac{26.35339}{200}\right)$

$H_0 : \mu = 18.72$ (the tobacco company's claim about tar content is correct)
$H_1 : \mu > 18{,}72$ (the tobacco company's claim about tar content is too low)
An 2% significance level requires the 96% confidence interval
$\Rightarrow \quad P(Z < z) = 0.98$ so $z = 2.05$ (3 s.f.)

96% CI: $\quad \bar{x} - 2.05\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.05\dfrac{\sigma}{\sqrt{n}}$

$$\frac{3928.649}{200} - 2.05 \times \frac{\sqrt{26.35}}{\sqrt{200}} < \mu < \frac{3928.649}{200} + 2.05 \times \frac{\sqrt{26.35}}{\sqrt{200}}$$

The 95% confidence interval for this sample is $18.99 < \mu < 20.39$.
Since the value given by the tobacco company lies outside the confidence interval for the sample, the null hypothesis is rejected. There is evidence to suggest that the company's claim about the tar content of the cigarettes is too low.

E.g. 9 The managing director of a firm commissioned a survey to estimate the mean expenditure on electrical appliances. A random sample of 100 people were questioned and the research team presented the managing director with a $95\ \%$ confidence interval of $(£128.14,\ £141.86)$. The director says this interval is too wide and wants a confidence interval of total width $£10$.

(a) Using the same value of $\bar{x}$, find the confident limits in this case.

(b) Find the level of confidence for the interval in part (a)

After reflection, the director is still not happy and now wishes to know how large a sample would be required to obtain a $95\ \%$ confidence interval of total width no more than $£10$.

(c) Find the smallest size of sample that will satisfy this request.

Working: (a) $$\bar{x} = \frac{128.14 + 141.86}{2} = 135$$

The new confidence interval would be $(£130,\ £140)$.

(b) $95\ \%$ CI: $$\bar{x} - 1.96 \times \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

$n = 100, \sigma = ?$: $$1.96 \times \frac{\sigma}{\sqrt{100}} = 141.86 - 135$$
$$\sigma = 35$$

Width is 10: $$z \times \frac{35}{\sqrt{100}} = 5$$
$$z = \frac{10}{7}$$

$P(Z > \frac{10}{7}) = 0.07656$ i.e. $7.656\ \%$

The level of confidence is $100\% - 2 \times 7.656\ \% = 84.7\ \%$

(c) $95\ \%$ CI, $\sigma = 35, n = ?$: $$1.96 \times \frac{35}{\sqrt{n}} \le 5$$
$$n \ge 188.2$$

The smallest size of sample is 189.

Video: Understanding confidence intervals
Video: How to calculate a confidence interval
Video: Confidence intervals example

Solutions to Starter and E.g.s

Exercise

p176 9B Qu 1, 2i, 3i, 4-9, (10-12 red)