Confidence intervals

Paula Moraga

Confidence interval

A $(c \times 100)\%$ confidence interval for a population parameter based on sample observations is an interval (A, B) such that we are $(c \times 100)\%$ confident that the population parameter is within the interval.

- When we use a different sample, we obtain a different confidence interval
- If we take many samples from the population and calculate 95% CIs for each sample, then 95% of these CIs will contain the true population parameter.
- Confidence level $c = 1 - \alpha$. Significance level $\alpha = 1 - c$
- 95% CI. Confidence level $c = 0.95$. Significance level $\alpha = 0.05$
- 90% CI. Confidence level $c = 0.90$. Significance level $\alpha = 0.10$
- 99% CI. Confidence level $c = 0.99$. Significance level $\alpha = 0.01$

Construct a confidence interval

A $(c \times 100)\%$ confidence interval for a population parameter is

$$\text{sample statistic} \pm \text{critical value} \times \text{SE}$$

- **Sample statistic** used to estimate the population parameter.
  Sample mean $(\bar{X})$ to estimate the population mean (μ)
  Sample proportion $(\hat{P})$ to estimate the population proportion (p)
- **Standard error SE** is the standard deviation of the sampling distribution of the sample statistic
- **Critical value** $(z^*,\ t^*(n-1))$.
  Measures the number of SE to be added and subtracted from the sample statistic to achieve the desired confidence level

Confidence interval for a proportion

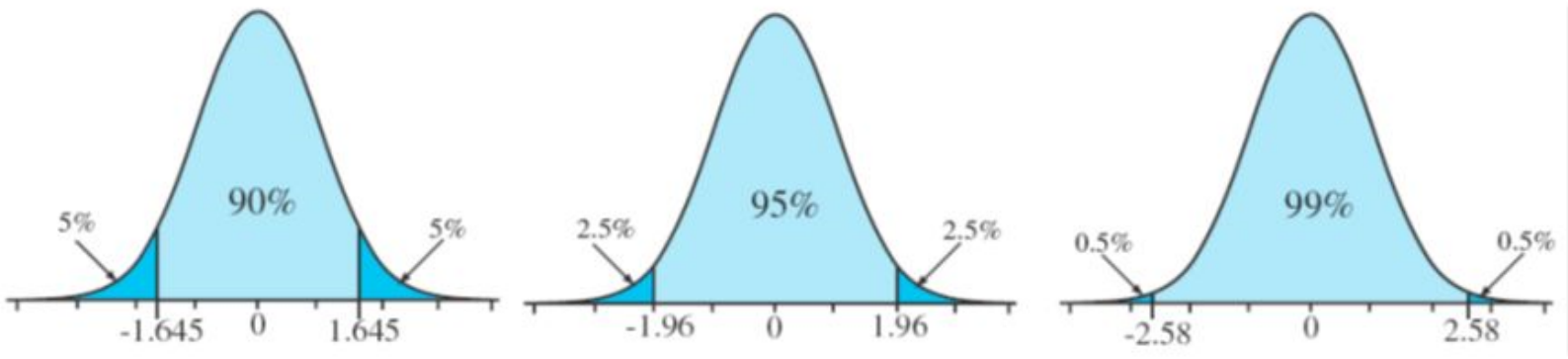
$(1-\alpha)\times 100\%$ confidence interval for the population proportion p

$$\hat{P} \pm z^* \times SE = \hat{P} \pm z^* \times \sqrt{\hat{P}(1-\hat{P})/n}$$

Assumptions:

1. Sample observations are independent (random sample)
2. In the sample, there are at least 10 successes and 10 failures

z^*: value such that $c = 1-\alpha$ of the probability in the N(0,1) falls between $-z^*$ and z^*.
Value such that $\alpha/2$ of the probability in the $N(0,1)$ is below z^*



Example. Confidence interval for a proportion

In a random sample of 40 students, 24 said they bought a copy of the textbook. Estimate the proportion of all students who bought the book and give a 95% confidence interval.

Solution

$$\hat{p} \pm z^* \times SE = \hat{p} \pm z^* \times \sqrt{\hat{p}(1-\hat{p})/n}$$

Assumptions:

- Sample observations independent (random sample)
- At least 10 successes ($24 \geq 10$) and 10 failures ($40 - 24 = 16 \geq 10$)

Sample statistic: $\hat{p} = 24/40 = 0.6$

Critical value: $z^* = -1.96$
($\alpha/2$ of the probability in $N(0, 1)$ is below z^*)

```
qnorm(0.025)
```

```
## [1] -1.959964
```

$\hat{p} \pm z^* \times SE = \hat{p} \pm z^* \times \sqrt{\hat{p}(1-\hat{p})/n} =$
$0.6 \pm 1.96 \times \sqrt{0.6(1-0.6)/40} = (0.448, 0.752)$

An estimate of the proportion of students who bought the book is 60%.

We are 95% confident that the proportion of students is between 44.8% and 75.2%.

Example. Confidence interval for a proportion

A random sample of 826 people living in UK was surveyed to better understand their political preferences. 70% of the responses supported the political party A. Calculate a 95% confidence interval for the proportion of people that support the political party A.

Solution

$$\hat{p} \pm z^* \times SE = \hat{p} \pm z^* \times \sqrt{\hat{p}(1-\hat{p})/n}$$

Assumptions:

- Observations are independent (they are from a random sample)
- In the sample at least 10 support
  ($n\hat{p} = 826 \times 0.70 = 578 \geq 10$ and at least 10 do not support
  ($n(1-\hat{p}) = 826 \times (1-0.70) = 248 \geq 10$)

Sample statistic: $\hat{p} = 24/40 = 0.70$

Critical value: $z^* = -1.96$
($\alpha/2 = 0.05/2$ of the probability in $N(0, 1)$ is below z^*)

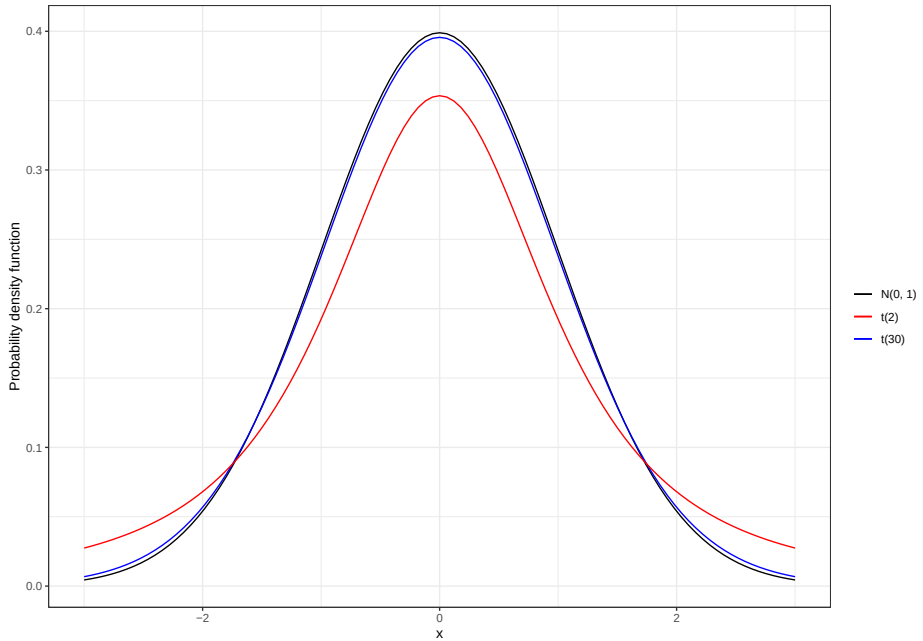
```
qnorm(0.025)
```

```
## [1] -1.959964
```

$\hat{p} \pm z^* \times SE = \hat{p} \pm z^* \times \sqrt{\hat{p}(1-\hat{p})/n} = 0.70 \pm 1.96 \times 0.016 =$
$(0.669, 0.731)$

We are 95% confident that the true proportion of people who support political party A is between 0.669 and 0.731.

t distribution



- Symmetric and bell-shaped like the normal distribution.
- Mean 0, Heavier tails than the normal distribution. Observations are more likely to fall beyond two standard deviations from the mean than under the normal distribution.

The t distribution has a parameter called degrees of freedom (df). The df describes the form of the t-distribution. When the df increases, the t-distribution approaches the standard normal N(0, 1)

Confidence interval for a mean

$(1-\alpha)\times 100\%$ confidence interval for the population mean μ

$$\bar{X} \pm t^*_{n-1} \times SE = \bar{X} \pm t^*_{n-1} \times \frac{S}{\sqrt{n}}$$

t^*_{n-1} value such that $\alpha/2$ of the probability in $t(n-1)$ is below t^*_{n-1}

Assumptions:

1. Sample observations are independent.
2. Sample observations from a normally distributed population.

When the population standard deviation is known:

$$\bar{X} \pm z^* \times SE = \bar{X} \pm z^* \times \frac{\sigma}{\sqrt{n}}$$

z^* value such that $\alpha/2$ of the probability in the N(0,1) is below z^*

Example. Confidence interval for a mean

In a class survey, students are asked how many hours they sleep per night. In a sample of 52 students, the mean was 5.77 hours with a standard deviation of 1.572 hours. Construct a 95% confidence interval for the mean number of hours slept per night in the population from which this sample was drawn.

Solution

σ is unknown, we use s and a t distribution with $n - 1 = 52 - 1$ df

$$\bar{x} \pm t^*_{51} \times SE = \bar{x} - t^*_{51} \times \frac{s}{\sqrt{n}}$$

Assumptions:

- We assume data are independent
- We assume data are normal

Sample statistic: $\bar{x} = 5.77$

Critical value: $t^*_{51} = -2.01$
($\alpha/2 = 0.05/2$ of the probability in t(51) is below t^*_{51})

```
qt(0.025, 51)
```

```
## [1] -2.007584
```

$\bar{x} \pm t^*_{51} \times SE = \bar{x} \pm t^*_{51} \times \frac{s}{\sqrt{n}} = 5.77 \pm 2.01 \times \frac{1.572}{\sqrt{52}} = (5.33, 6.21).$

We are 95% confident the population mean is between 5.33 and 6.21 hours.

Example. Confidence interval for a mean

Elevated mercury concentrations are an important problem for both dolphins and other animals who occasionally eat them. Calculate a 95% CI for the average mercury content in dolphins using a sample of 19 dolphins in Japan. In the sample, $n = 19$, $\bar{x} = 4.4$, $s = 2.3$, minimum $= 1.7$ and maximum $= 9.2$ μg/wet gram (micrograms of mercury per wet gram of muscle).

Solution

σ is unknown, we use s and a t distribution with $n - 1 = 19 - 1$ df

$$\bar{x} \pm t^*_{18} \times SE = \bar{x} \pm t^*_{18} \times \frac{s}{\sqrt{n}}$$

Assumptions:

- Independence seems reasonable because observations come from a random sample.
- Normality seems reasonable because there are not clear outliers (all observations are within 2.5 standard deviations of the mean).

Sample statistic: $\bar{x} = 4.4$

Critical value: $t^*_{18} = -2.10$
($\alpha/2 = 0.05/2$ of the probability in the t(18) distribution is below t^*_{18})

```
qt(0.025, 18)
```

```
## [1] -2.100922
```

$\bar{x} \pm t^*_{18} \times SE = \bar{x} \pm t^*_{18} \times \frac{s}{\sqrt{n}} = 4.4 \pm 2.10 \times \frac{2.3}{\sqrt{19}} = (3.29, 5.51)$

We are 95% confident the average mercury content in dolphins is between 3.29 and 5.51 μg/wet gram.