University of Glasgow

Statistical Inference, Lecture 11

Dr Benn Macdonald

Room 225 Maths and Stats Building,
Benn.Macdonald@glasgow.ac.uk

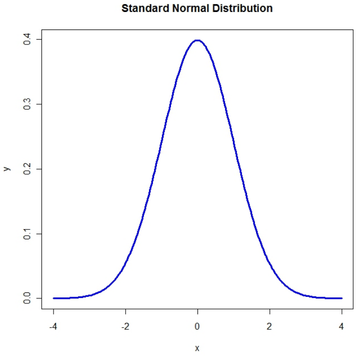
February 2021

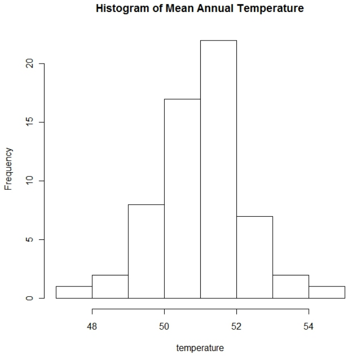
- Likelihood Intervals
- **One and two-sample t-intervals and tests (exact coverage)**
- **Wilks and Wald Intervals (approximate coverage)**
- Multiparameter models
  - Wilks and Wald confidence regions

Example 9 - Mean Annual Temperature in New Haven

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3069.6}{60} = 51.16, s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 1.602$$

3.6.4 One-sample t-interval and t-test





3.6.4 One-sample t-interval and t-test

You can also test hypotheses about the population mean (**one-sample t-test**).

The confidence interval and t-test both require **the assumption of normality** to be true.

If the assumption of normality is not true, probably use a Wilcoxon Signed Ranks test.

3.6.4 One-sample t-interval and t-test

(**one-sample t-test**)

Suppose that we wish to investigate the following hypotheses about the population mean μ_T:

$\mathrm{H}_0 : \mu_T = 51.2$

$\mathrm{H}_1 : \mu_T \neq 51.2$

3.6.4 One-sample t-interval and t-test

```
t.test(temp, mu=51.2)

One Sample t-test

data:  temp
t = -0.2448, df = 59, p-value = 0.8074
alternative hypothesis: true mean is not equal to 51.2
95 percent confidence interval:
 50.83306 51.48694
sample estimates:
mean of x
    51.16
```

3.6.4 One-sample t-interval and t-test

A **two-sample t-interval and t-test** to compare two population means is the parametric **(assuming normality)** equivalent of a Mann-Whitney U Test.

In general, for ‘large’ sample sizes use one-sample/two-sample t-tests. For ‘small’ sample sizes use nonparametric tests.

See details of ‘the central limit theorem’ in Probability (Level M).

3.6.4-3.6.5 One-sample/two-sample t-test

Read the notes on pages 16 to 19 and try the tutorial examples (tutorial sheet 4).

One and two-sample t-tests were covered in the practical.

The practical questions/solutions are on MOODLE.

- Likelihood Intervals
- One and two-sample t-intervals and tests (exact coverage)
- **Wilks and Wald Intervals (approximate coverage)**
- Multiparameter models
    - Wilks and Wald confidence regions

3.7 Intervals based on Approx. Confidence

One and two-sample t-intervals have exact coverage properties within normal models.

However, we can use results based on the **large sample properties of Maximum Likelihood Estimators** to provide confidence intervals for a range of statistical models.

We will introduce and use these results without proof here.

3.7.1 Wilks Intervals

When θ is set to its true value θ_T then, approximately,

$$2[\ell(\hat{\theta}_{MLE}) - \ell(\theta_T)] \sim \chi^2_1$$

This means that the quantity $2[\ell(\hat{\theta}_{MLE}) - \ell(\theta_T)]$ is an approximate **pivotal quantity** for θ.

The large sample distribution of $2[\ell(\hat{\theta}_{MLE}) - \ell(\theta_T)]$ is approximately χ^2.

3.7.1 Wilks Intervals

A confidence interval stems from the result that

$$P\{2[\ell(\hat{\theta}_{MLE}) - \ell(\theta_T)] \leq \chi_1^2(c)\} \approx c,$$

where $\chi_1^2(c)$ denotes the c^{th} quantile of the χ_1^2 distribution.

i.e. a χ_1^2 random variable will be less than $\chi_1^2(c)$ with probability c.

3.7.1 Wilks Intervals

An approximate 100c% confidence (Wilks) interval for θ:

$$\{\theta : 2[\ell(\hat{\theta}_{MLE}) - \ell(\theta)] \leq \chi_1^2(c)\}.$$

In terms of the relative log-likelihood function:

$$\{\theta : -2r(\theta) \leq \chi_1^2(c)\}$$

$$\{\theta : r(\theta) \geq -\frac{1}{2}\chi_1^2(c)\}$$

The approximation is valid provided the sample size (i.e. the number of independent observations) is large.

3.7.1 Wilks Intervals

For 95% confidence this result is:

$$\{\theta : -2r(\theta) \leq \chi_1^2(0.95)\}$$

$$\{\theta : -2r(\theta) \leq 3.84\}$$

$$\{\theta : r(\theta) \geq -1.92\}$$

** see the connection here to the result at the bottom of page 10.

Example 7 continued: Air Conditioning Failures:

Assuming $X_1, X_2, \ldots, X_n \sim (\text{Expo}(\theta)), \theta > 0$:

The **log relative likelihood** (Page 3) is given by:

$$r(\theta) = \ell(\theta) - \ell(\hat{\theta}_{MLE})$$

$$r(\theta) = n\log_e(\theta) - \theta\sum_{i=1}^{n} x_i - n\log_e(\hat{\theta}_{MLE}) + \hat{\theta}_{MLE}\sum_{i=1}^{n} x_i$$

$$= 24\log_e(\theta) - 1539\theta + 123.9$$

Air Conditioning Example 7 continued:

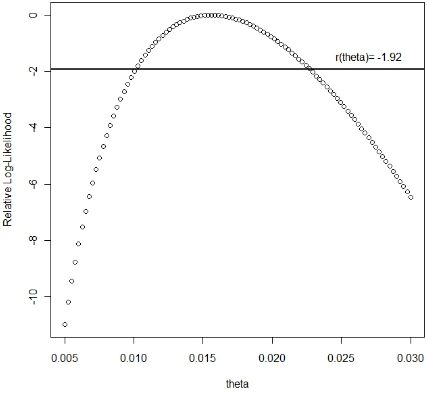


Figure: Relative log likelihood for the air conditioning data

Air Conditioning Example 7 continued:

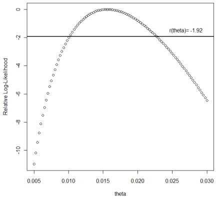


Figure: Relative log likelihood for the air conditioning data

It can be seen from the Figure that **the Wilks interval is approximately (0.010, 0.024)**.

Air Conditioning Example 7 continued:

To obtain the Wilks interval with approximate confidence 0.95 the **Newton-Raphson** algorithm can be used.

This is very similar to the routine described for calculating $100p\%$ likelihood intervals in Section 3.5.

Starting values can be estimated from the plot for the lower and upper bounds of the interval i.e. 0.010 and 0.024 respectively.

Air Conditioning Example 7 continued:

Newton-Raphson algorithm:

$$\theta_B^{(j+1)} = \theta_B^{(j)} - \frac{g(\theta_B^{(j)})}{g'(\theta_B^{(j)})}$$

with

$$\{\theta : r(\theta) \geq -1.92\}$$

and hence

$$g(\theta_B) = r(\theta) + 1.92$$

3.7.1 Wilks Intervals

Air Conditioning Example 7 continued:

The **Newton-Raphson** algorithm converged in two-steps to the Wilks interval with approximate confidence 0.95 of:

$$(0.010, 0.023)$$

3.7.1 Wilks Intervals

Invariance properties:

Wilks intervals have attractive invariance properties.

Let θ and β be parameters, $\beta = g(\theta)$ and $g()$ is monotonic.

If (θ_a, θ_b) is a Wilks interval for θ then,

$(g(\theta_a), g(\theta_b))$ would be a Wilks interval for β.

3.7.2 Wald Intervals

This approach is based on another distributional approximation.

When θ_T is the true value of θ then, approximately,

$$\hat{\theta}_{MLE} \sim N(\theta_T, 1/k(\mathbf{x})).$$

The **large sample distribution** of a maximum likelihood estimate $\hat{\theta}$ is **approximately normal** with mean θ_T and variance $1/k(\mathbf{x})$.

This result holds as the sample size tends to infinity.

3.7.2 Wald Intervals

Sample information = $k(\mathbf{x}) = -l''(\hat{\theta}_{MLE})$.

This means that, approximately,

$$\frac{\hat{\theta}_{MLE} - \theta_T}{\sqrt{1/k(\mathbf{x})}} \sim N(0,1)$$

and so the quantity $(\hat{\theta}_{MLE} - \theta_T)/\sqrt{1/k(\underline{x})}$ is an **approximate pivotal quantity** for θ.

3.7.2 Wald Intervals

An approximate $100c\%$ confidence (Wald) interval for $\boldsymbol{\theta}$:

$$(\hat{\theta}_{MLE} - z\,\sqrt{1/k(\mathbf{x})},\ \hat{\theta}_{MLE} + z\,\sqrt{1/k(\mathbf{x})})$$

$z = \phi^{-1}(1 - \frac{1-c}{2})$.

Typically, $c = 0.95$ and hence $z = \phi^{-1}(0.975) = 1.96$.

Note:

$$\text{var}\{\hat{\theta}_{MLE}\} \approx \frac{1}{k(\mathbf{x})}$$

3.7.2 Wald Intervals

Wald intervals are usually very easy to obtain relative to Wilks intervals, but their properties are less satisfactory for finite sample sizes.

The intervals are **not invariant** .

The intervals are **always symmetric** .

Example 7 continued: Air Conditioning Failures:

Assuming an exponential distribution for the data with parameter θ $(\mathrm{Expo}(\theta)), \theta > 0$:

Likelihood: (Page 2)

$$L(\theta) \propto \prod_{i=1}^{n} \theta\, e^{-\theta x_i} = \theta^n e^{-\theta \sum x_i}$$

3.7.2 Wald Intervals

Log-likelihood:

$$\ell(\theta) = n \log_e \theta - \theta \sum_{i=1}^{n} x_i$$

$$\ell'(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i$$

$$\hat{\theta}_{MLE} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$$

$$\ell''(\theta) = -\frac{n}{\theta^2}$$

and this is < 0 for all $\theta > 0$.

Air Conditioning Example 7 continued...

So, we have

$$k(\mathbf{x}) = -\ell''(\hat{\theta}_{MLE}) = \frac{n}{\hat{\theta}^2}$$

Since $\hat{\theta}_{MLE} = 0.016$ and $n = 24$, the **standard error** of $\hat{\theta}_{MLE}$ is

$$\sqrt{1/k(\mathbf{x})} = \sqrt{\frac{\hat{\theta}^2}{n}} = \sqrt{\frac{0.016^2}{24}} = 0.00327$$

Air Conditioning Example 7 continued...

The 0.975 quantile of the standard normal distribution is 1.96.

An **approximate 95% confidence interval for θ** is then

$$(\hat{\theta} - 1.96 \times 0.00327, \hat{\theta} + 1.96 \times 0.00327)$$

$$(0.0096, 0.0224)$$

$$(0.010, 0.022)$$

It is therefore highly likely that θ lies in the range 0.010 to 0.022.