Tutorial 4

YANG YANG

The Australian National University

Week 5, 2017

Overview

Confidence Intervals for $\hat{\beta}_0$ and $\hat{\beta}_1$

A 95% confidence interval for $\hat{\beta}_0 \longrightarrow \hat{\beta}_0 \pm t(0.975) \times se(\hat{\beta}_0)$

A 95% confidence interval for $\hat{\beta}_1 \longrightarrow \hat{\beta}_1 \pm t(0.975) \times se(\hat{\beta}_1)$

Confidence Interval for parameters

- It is an observed interval (i.e., it is calculated from the observations), that potentially includes the unobservable **true parameter of interest**.
- How frequently the observed interval contains the true parameter **if the experiment is repeated** is called the confidence level.
- population mean μ and population standard deviation σ
- "We are 95% confident that the **true value of the parameter** is in our confidence interval."

Prediction Interval for future observations

- It is an estimated interval in which **future observations** will fall, with a certain probability, given what has already been observed.
- Prediction intervals predict the distribution of **observable** future points.
- observations or sample point X_{n+1}
- "On repeated applications of this computation, the future observation X_{n+1} will fall in the predicted interval 95% of the time."

Confidence Intervals for Prediction

We often interested in predicting the response value for future observations at particular values of predictor (e.g., x_0).

$$\hat{Y}(x_0) = b_0 + b_1 x_0$$

Associated standard error of the estimate

$$\text{Var}\{\hat{Y}(x_0)\} = \sigma^2 \left\{ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right\}$$

Substitute our estimator, s, for the scale parameter, σ, we then have an estimate of the standard error of prediction

$$s\{\hat{Y}(x_0)\} = s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

Confidence intervals for prediction

Assuming normal errors, a 100(1-α)% interval for the expected response $E(Y|x_0)$ is

$$\hat{Y}(x_0) \pm t_{n-2}(1-\alpha/2)s\{\hat{Y}(x_0)\}$$

This interval is indeed a confidence interval, since in gives a plausible range for a **fixed, population quantity.**

Prediction intervals for a future response value

We also make prediction for the value of a single future response value, Y_0, at a particular value of the predictor value, x_0. It is called a prediction interval since we want to find a likely range for a random quantity.

A future value Y_0 is independent of the observed data used to estimate b_0 and b_1. So,

$$\text{Var}\{Y_0 - \hat{Y}(x_0)\} = \text{Var}(Y_0) + \text{Var}\{\hat{Y}(x_0)\} = \sigma^2 + \sigma^2 \left\{ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right\}$$

Prediction intervals for a future response value

Since $E\{Y_0 - \hat{Y}(x_0)\} = 0$, the standard normal theory assumptions show that

$$\frac{Y_0 - \hat{Y}(x_0)}{s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}} \sim t_{n-2}.$$

(see Page 19 of Lecture Notes)

CI and PI

A 100(1-α)% confidence interval for a given value x_0:

$$\hat{Y}(x_0) \pm t_{\alpha/2,n-2} \times s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}^2}}$$

A 100(1-α)% prediction interval for a given value x_0:

$$\hat{Y}(x_0) \pm t_{\alpha/2,n-2} \times s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}^2}}$$

- The further away from the $\bar{x}$ we are predicting, the wider our prediction interval will be.
- PI is wider than CI for the same value of x_0

Hypothesis test of the $\rho_{x,y}$

Step One: $H_0 : \rho_{x,y} = 0$ v.s. $H_A : \rho_{x,y} \neq 0$

Step Two: Test Statistic $= \frac{r-0}{se(r)} = \frac{r\sqrt{n-2}}{1-r^2}$

Step Three: Refers to the t distribution table with $n-2$ degrees of freedom and find the critical values.

Step Four: Compare the calculated test statistics with the critical values and make a decision.

Step Five: Conclusion

Question 1 (d)-(g)

In R, we use "predict()" command to compute 95% confidence intervals and prediction intervals

- CI: predict(object,newdata=as.data.frame(cbind(col.name=new.x)), interval="**confidence**")
- PI: predict(object,newdata=as.data.frame(cbind(col.name=new.x)), interval="**prediction**")