The prediction interval formulas for the next observation from a normal distribution when σ is unknown

December 13, 2005

1 Introduction

In this lecture we will derive the formulas for the symmetric two-sided prediction interval for the $n+1$-st observation and the upper-tailed prediction interval for the $n+1$-st observation from a normal distribution *when the variance σ^2 is unknown.* We will need the following theorem from probability theory that gives the distribution of the statistic $\overline{X} - Xn + 1$.

Suppose that $X_1, X_2, \ldots, X_n, X_{n+1}$ is a random sample from a normal distribution with mean μ and variance σ^2.

Theorem 1. *The random variable $T = (\overline{X} - X_{n+1})/(\sqrt{\frac{n+1}{n}}S)$ has t distribution with $n-1$ degrees of freedom.*

2 The two-sided prediction interval formula

Now we can prove the theorem from statistics giving the required prediction interval for the next observation x_{n+1} in terms of n observations $x_1, x_2, \cdots, x_n$. Note that it is symmetric around $\overline{X}$. This is one of the basic theorems that you have to learn how to prove. There are also asymmetric two-sided prediction intervals.

Theorem 2. *The random interval $(\overline{X} - t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}S \ , \ \overline{X} + t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}S)$ is a $100(1-\alpha)\%$-prediction interval for x_{n+1}.*

Proof. We are required to prove

$$P(X_{n+1} \in (\overline{X} - t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}S \ , \ \overline{X} + t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}S)) = 1 - \alpha.$$

We have

$$\begin{aligned}
LHS &= P(\overline{X} - t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}S < X_{n+1}, X_{n+1} < \overline{X} + t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}S) = P(\overline{X} - X_{n+1} < t_{\alpha/2,n-} \\
&= P(\overline{X} - X_{n+1} < t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}S \ , \ \overline{X} - X_{n+1} > -t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}S) \\
&= P((\overline{X} - X_{n+1})/\sqrt{\frac{n+1}{n}}S < t_{\alpha/2,n-1} \ , \ (\overline{X} - X_{n+1})/\sqrt{\frac{n+1}{n}}S > -t_{\alpha/2,n-1}) \\
&= P(T < t_{\alpha/2,n-1} \ , \ T > -t_{\alpha/2,n-1}) = P(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}) = 1 - \alpha
\end{aligned}$$

To prove the last equality draw a picture. □

Once we have an actual sample $x_1, x_2, \ldots, x_n$ we obtain the the observed value $(\overline{x} - t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}s, \overline{x} + t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}s)$ for the prediction (random) interval $(\overline{X} - t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}S \ , \ \overline{X} + t_{\alpha/2,n-1}\sqrt{\frac{n+1}{n}}S)$ The observed value of the prediction (random) interval is also called the two-sided $100(1-\alpha)\%$ prediction interval for x_{n+1}.

3 The upper-tailed prediction interval

In this section we will give the formula for the upper-tailed prediction interval for the next observation x_{n+1}.

Theorem 3. *The random interval* $(\overline{X} - t_{\alpha,n-1}\sqrt{\frac{n+1}{n}}S \ , \ \infty)$ *is a* $100(1-\alpha)\%$*-prediction interval for the next observation* x_{n+1}.

Proof. We are required to prove

$$P(X_{n+1} \in (\overline{X} - t_{\alpha,n-1}\sqrt{\frac{n+1}{n}}S \ , \ \infty)) = 1 - \alpha.$$

We have

$$
\begin{aligned}
LHS &= P(\overline{X} - t_{\alpha,n-1}\sqrt{\frac{n+1}{n}}S < X_{n+1}) \\
&= P(\overline{X} - X_{n+1} < t_{\alpha,n-1}\sqrt{\frac{n+1}{n}}S) \\
&= P((\overline{X} - X_{n+1})/\sqrt{\frac{n+1}{n}}S < t_{\alpha,n-1}) \\
&= P(T < t_{\alpha,n-1}) \\
&= 1 - \alpha
\end{aligned}
$$

To prove the last equality draw a picture - I want *you* to draw the picture on tests and the final. $\square$

Once we have an actual sample $x_1, x_2, \ldots, x_n$ we obtain the observed value $(\overline{x} - t_{\alpha,n-1}\sqrt{\frac{n+1}{n}}s, \infty)$ of the upper-tailed prediction (random) interval $(\overline{X} - t_{\alpha,n-1}\sqrt{\frac{n+1}{n}}S, \infty)$ The observed value of the upper-tailed prediction (random) interval is also called the upper-tailed $100(1-\alpha)\%$ prediction interval for x_{n+1}.

The number random variable $\overline{X} - t_{\alpha,n-1}\sqrt{\frac{n+1}{n}}S$ or its observed value $\overline{x} - t_{\alpha,n-1}\sqrt{\frac{n+1}{n}}s$ is often called a prediction *lower bound* for x_{n+1} because

$$P(\overline{X} - t_{\alpha,n-1}\sqrt{\frac{n+1}{n}}S < X_{n+1}) = 1 - \alpha.$$