# Lecture 31
# The prediction interval formulas for the next observation from a normal distribution when $\sigma$ is known

## 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* $\sigma^2$ *is known*. 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 $\mu$ and variance $\sigma^2$. We assume $\mu$ is unknown but $\sigma^2$ is known.

**Theorem 1.** *The random variable* $\overline{X} - X_{n+1}$ *has normal distribution with mean zero and variance* $\frac{n+1}{n}\sigma^2$*. Hence we find that the random variable* $Z = (\overline{X} - X_{n+1})/(\sqrt{\frac{n+1}{n}}\sigma)$ *has standard normal distribution.*

## 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* $(\overline{X} - z_{\alpha/2}\sqrt{\frac{n+1}{n}}\sigma \ , \ \overline{X} + z_{\alpha/2}\sqrt{\frac{n+1}{n}}\sigma)$ *is a* $100(1-\alpha)\%$*-prediction interval for* $x_{n+1}$*.*

*Proof.* We are required to prove

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

We have

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

To prove the last equality draw a picture. □

Once we have an actual sample $x_1, x_2, \ldots, x_n$ we obtain the observed value $(\overline{x} - z_{\alpha/2}\sqrt{\frac{n+1}{n}}\sigma\ ,\ \overline{x} + z_{\alpha/2}\sqrt{\frac{n+1}{n}}\sigma)$ for the prediction (random) interval $(\overline{X} - z_{\alpha/2}\sqrt{\frac{n+1}{n}}\sigma\ ,\ \overline{X} + z_{\alpha/2}\sqrt{\frac{n+1}{n}}\sigma)$ 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* $(\overline{X} - z_{\alpha}\sqrt{\frac{n+1}{n}}\sigma\ ,\ \infty)$ *is a* $100(1-\alpha)\%$*-prediction interval for the next observation* $x_{n+1}$*.*

*Proof.* We are required to prove

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

We have

$$\begin{aligned} LHS &= P(\overline{X} - z_\alpha\sqrt{\frac{n+1}{n}}\sigma < X_{n+1}) \\ &= P(\overline{X} - X_{n+1} < z_\alpha\sqrt{\frac{n+1}{n}}\sigma) \\ &= P((\overline{X} - X_{n+1})/\sqrt{\frac{n+1}{n}}\sigma < z_\alpha) \\ &= P(Z < z_\alpha) \\ &= 1 - \alpha \end{aligned}$$

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

Once we have an actual sample $x_1, x_2, \ldots, x_n$ we obtain the observed value $(\overline{x} - z_\alpha\sqrt{\frac{n+1}{n}}\sigma, \infty)$ of the upper-tailed prediction (random) interval $(\overline{X} - z_\alpha\sqrt{\frac{n+1}{n}}\sigma, \infty)$ 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 $\overline{X} - z_\alpha\sqrt{\frac{n+1}{n}}\sigma$ or its observed value $\overline{x} - z_\alpha\sqrt{\frac{n+1}{n}}\sigma$ is often called a prediction *lower bound* for $x_{n+1}$ because

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