Tolerance And Prediction Intervals

Prediction Interval for a Future Observation

In some problem situations, we may be interested in predicting a future observation of a variable. This is a different problem than estimating the mean of that variable, so a confidence is not appropriate.

Suppose that $X_1, X_2, \ldots, X_3$ is a random sample from a normal population. We wish to predict the value X_{n+1}, a single future observation. A point prediction of X_{n+1} is $\bar{X}$, the sample mean.

- **The prediction error is $X_{n+1} - \bar{X}$**
- **The expected value of the prediction error is**
$$E(X_{n+1} - \bar{X}) = \mu - \mu = 0$$
- **The variance of the prediction error is**
$$V(X_{n+1} - \bar{X}) = \sigma^2 + \frac{\sigma^2}{n} = \sigma^2\left(1 + \frac{1}{n}\right)$$
- **Because of the future observation X_{n+1} is independent of the mean of the current sample $\bar{X}$. The prediction error $X_{n+1} - \bar{X}$ is normally distributed. Therefore**
$$Z = \frac{X_{n+1} - \bar{X}}{\sigma\sqrt{\left(1 + \frac{1}{n}\right)}}$$
has a standard normal distribution.
- **Replacing σ with S results in**
$$T = \frac{X_{n+1} - \bar{X}}{S\sqrt{\left(1 + \frac{1}{n}\right)}}$$
which has a t distribution with n-1 degrees of freedom.

Prediction Interval

Manipulating T as we have done previously in the development of a CI leads to a prediction interval on the future observation X_{n+1}.

A 100(1-α)% prediction interval on a single future observation from a normal distribution is given by

$$\bar{X} - t_{\alpha/2,n-1} S\sqrt{1+\frac{1}{n}} \le X_{n+1} \le \bar{X} + t_{\alpha/2,n-1} S\sqrt{1+\frac{1}{n}}$$

The prediction interval for X_{n+1} will always be longer than the confidence interval for μ because there is more variability associated with the prediction error than with the error of estimation.

Example:

```
MTB > random 20 c1;
SUBC> normal 15 3.
MTB > print c1
```

Data Display

```
C1
    9.0680   18.6516   14.0500   20.3649   11.8317   17.9297   13.7635
   11.8171   15.0659   16.3931   12.0878   10.6404   12.9046   16.0074
    8.2335   17.9648   12.5913   14.0011   12.6949   11.9322
```

Find a 95% CI on the mean.

Find a 95% prediction interval for the next observation.

Descriptive Statistics: C1

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
C1	20	0	13.900	0.717	3.205	8.233	11.857	13.334	16.297

Variable	Maximum
C1	20.365

```
MTB > tint 95 c1
```

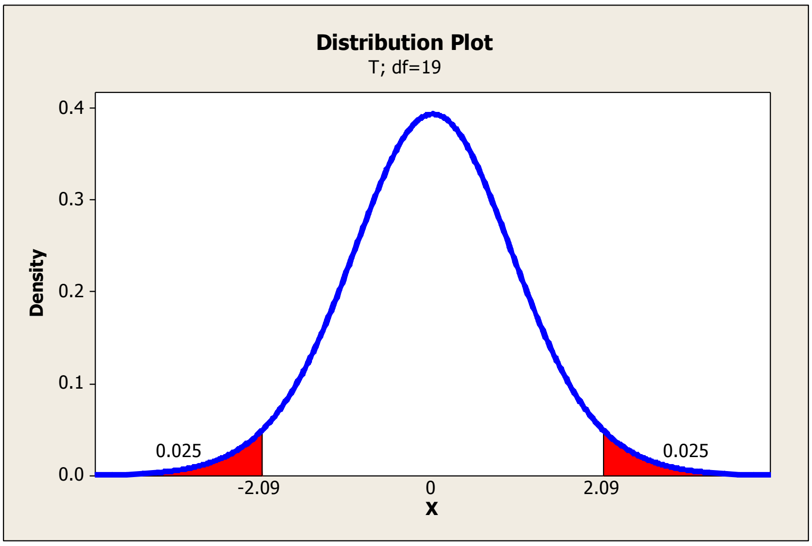
One-Sample T: C1

Variable	N	Mean	StDev	SE Mean	95% CI
C1	20	13.900	3.205	0.717	(12.400; 15.400)

$$12.40 \leq \mu \leq 15.4$$

Prediction interval

$$\bar{X} - t_{\alpha/2,n-1} S\sqrt{1+\frac{1}{n}} \le X_{n+1} \le \bar{X} + t_{\alpha/2,n-1} S\sqrt{1+\frac{1}{n}}$$



$$13.9 - 2.09(3.205)\sqrt{1+\frac{1}{20}} \le X_{21} \le 13.9 + 2.09(3.205)\sqrt{1+\frac{1}{20}}$$

$$7.036 \le X_{21} \le 20.763$$

Compare with the following

$$12.40 \le \mu \le 15.4$$

```
MTB > random 1 c2;
SUBC> normaL 15 3.
MTB > print c2
```

Data Display

```
C2
   12.4073
```