Confidence vs Prediction Intervals

Confidence Intervals

We've all constructed confidence intervals to estimate the mean μ of a random variable. The general form of a two-sided interval estimator of parameter θ based on the distribution of $Z \sim N(0,1)$ or t_{df} is either

- $\hat{\theta} \pm z_{\alpha/2} \times \sigma_{\hat{\theta}}$, where $\hat{\theta}$ is an unbiased estimator of θ and $\sigma_{\hat{\theta}}$ is the *standard error* of $\hat{\theta}$, or
- $\hat{\theta} \pm t_{df,\alpha/2} \times \hat{\sigma}_{\hat{\theta}}$, where $\hat{\sigma}_{\hat{\theta}}$ is the *estimated* standard error of $\hat{\theta}$ derived from the same sample used to compute $\hat{\theta}$, and the degrees of freedom *df* is the sample size *n* minus the number of parameters *m* estimated prior to estimating $\sigma_{\hat{\theta}}$.

Both types of confidence interval assume we are sampling from a normally distributed random variable, but this assumption becomes less critical for large *n*. We'll also assume observations are independent.

In Stat 50 we were given key facts about the Estimator $\hat{\theta}$, such as its expected value, standard error, and distribution, but in Stat 103 we must derive these prior to arriving at a confidence interval for the parameter θ. If you are surprised to learn that this involves finding the mean and variance of a linear combination, then you haven't been paying attention!

Confidence vs Prediction Intervals

In addition to confidence intervals, we will also be interested in prediction intervals in regression. We'll begin with a quick discussion of the difference between them, then do a simple exercise to develop some intuition, and finally proceed to deriving forms for the confidence and prediction intervals used in Simple Linear Regression.

A (1-α)100% confidence interval is constructed to contain the true (but unknown, hence requiring estimation) value of some parameter θ with predetermined probability (1-α). The key to confidence intervals is that they are constructed to estimate parameters (constants of the random variable such as its mean) and that we can never know for certain whether the parameter lies withing the final constructed interval. Thus, we are restricted to talking about the confidence we have that the parameter lies within the final interval. (Probability belongs to the future. Once we've drawn a sample, the interval either contains the parameter or it doesn't and we won't know which is the case.)

A (1-α)100% prediction interval *predicts* that the *next* value of a random variable lies within the interval with probability (1-α). The key to a prediction interval is that it is constructed to contain the next value of a random variable, and we will know if it succeeded when we make the next observation on the variable.

As an example, suppose I am a realtor using a regression model that uses house size to predict price. A house with 2150 ft^2 comes on the market. I may wish to estimate the mean price of 2150 ft^2 houses in the target population (houses located in the same area, for instance) by constructing a confidence interval for the mean price of all houses with 2150 ft^2, but I can't be certain that the interval contains the mean price of such houses. Instead, I may wish to predict the price of the particular house that just came to market. I'm predicting the price of the *next* 2150 ft^2 house drawn from the target population. Unlike my confidence interval for the mean price of such houses, I'll know whether the house's price falls within my prediction interval when the house sells! This illustrates the difference between a (1-α)100% confidence interval for a parameter and a (1-α)100% prediction interval for the next value of a random variable.

Motivation: A Simple Prediction Interval

Example: Let $\boldsymbol{Y} \sim N\left(\mu, \sigma^2\right)$, find the probability that $\boldsymbol{Y}$ lies within the interval $\left(\mu - z_{\alpha/2}\sigma, \mu + z_{\alpha/2}\sigma\right)$.

Answer: $(1 - \alpha)$ **Note:** The answer did not require knowledge of either μ or σ.

Example: Now, construct an interval that has probability $(1 - \alpha)$ of containing the next value drawn from random variable $\boldsymbol{Y}$.

Answer: $\boxed{\left(\mu - z_{\alpha/2}\sigma, \mu + z_{\alpha/2}\sigma\right)}$ will do the trick. Congratulations, you've just constructed your first prediction interval! **Note:** Of course, constructing this interval requires knowledge of both μ and σ.

A Small Complication

The interval constructed in the previous example, $\left(\mu - z_{\alpha/2}\sigma, \mu + z_{\alpha/2}\sigma\right)$, assumed that we knew both μ and σ. (Throughout we will also assume that all variables are normally distributed.) Suppose instead that $\boldsymbol{Y} \sim N\left(?, \sigma^2\right)$. We will have to estimate the unknown mean μ prior to predicting the next value of $\boldsymbol{Y}$.

Example: Construct an interval, based on a simple random sample of n observations on $\boldsymbol{Y}$, that has probability $(1 - \alpha)$ of containing the mean value μ of random variable $\boldsymbol{Y}$.

Answer: $\boxed{\bar{\boldsymbol{Y}} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$. This is a $(1 - \alpha)100\%$ confidence interval for the mean μ of random variable $\boldsymbol{Y}$.

Discussion: You've constructed so many confidence intervals like the one above that you may have forgotten the logic behind them, but have no fear, I'm here to remind you of it.

- ○ $\bar{\boldsymbol{Y}} = \frac{\boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_n}{n}$. Have I mentioned the importance of linear combinations before?
- ○ $\mu_{\bar{\boldsymbol{Y}}} = \mu$, i.e., the sample mean $\bar{\boldsymbol{Y}}$ is an unbiased estimator of the mean μ of $\boldsymbol{Y}$.
- ○ $\sigma^2_{\bar{\boldsymbol{Y}}} = \frac{\sigma^2}{n}$. The independence of the $\boldsymbol{Y}_i$ in the simple random sample is important to deriving this.
- ○ $\bar{\boldsymbol{Y}}$ is normally distributed because linear combinations of normal random variables are normal.

Example: For $\boldsymbol{Y} \sim N\left(?, \sigma^2\right)$, construct an interval that has probability $(1 - \alpha)$ of containing the next observation on $\boldsymbol{Y}$. The solution involves a two-step process: we'll first draw n values from $\boldsymbol{Y}$ to estimate μ, then construct a prediction interval for *next* $\boldsymbol{Y}$.

Answer: $\boxed{\bar{\boldsymbol{Y}} \pm z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}}}$. **Note:** The standard error is now $\sigma\sqrt{1 + \frac{1}{n}}$.

Discussion: The meaning of the interval above is $P\left(\bar{\boldsymbol{Y}} - z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}} \le \boldsymbol{Y}_{next} \le \bar{\boldsymbol{Y}} + z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}}\right) = 1 - \alpha$.

Now, $P\left(\bar{\boldsymbol{Y}} - z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}} \le \boldsymbol{Y}_{next} \le \bar{\boldsymbol{Y}} + z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}}\right) = P\left(-z_{\alpha/2} \le \frac{\boldsymbol{Y}_{next} - \bar{\boldsymbol{Y}}}{\sigma\sqrt{1 + \frac{1}{n}}} \le z_{\alpha/2}\right)$. The claim is that

$\dfrac{\boldsymbol{Y}_{next}-\bar{\boldsymbol{Y}}}{\sigma\sqrt{1+\frac{1}{n}}} \sim N(0,1)$, from which the result follows, but why is this quotient the standard normal variate? Have I mentioned the importance of linear combinations?

- $\boldsymbol{Y}_{next}-\bar{\boldsymbol{Y}}$ is normally distributed because it is a linear combination of normal random variables.
- $\mu_{\boldsymbol{Y}_{next}-\bar{\boldsymbol{Y}}} = \mu_{\boldsymbol{Y}_{next}} - \mu_{\bar{\boldsymbol{Y}}} = \mu - \mu = 0$
- $\sigma^2_{\boldsymbol{Y}_{next}-\bar{\boldsymbol{Y}}} = \sigma^2_{\boldsymbol{Y}_{next}} + \sigma^2_{\bar{\boldsymbol{Y}}} = \sigma^2 + \dfrac{\sigma^2}{n} = \sigma^2\left(1+\dfrac{1}{n}\right)$ because $\boldsymbol{Y}_{next} \sim N\left(\mu,\sigma^2\right)$ is independent of $\bar{\boldsymbol{Y}}$.

Therefore, $\dfrac{\boldsymbol{Y}_{next}-\bar{\boldsymbol{Y}}}{\sigma\sqrt{1+\frac{1}{n}}} \sim N(0,1)$. **Note:** The added variability of the prediction interval relative to the confidence interval stems from the two-step nature of estimating μ prior to predicting the next value of $\boldsymbol{Y}$.

Who Needs the Variance?

In practice, we rarely know the variance σ^2 of $\boldsymbol{Y}$. Making the usual assumptions about $\boldsymbol{Y}$ and the sample, the only adjustments made to confidence and prediction intervals involves replacing σ by $\boldsymbol{S}$ and $z_{\alpha/2}$ by $t_{n-1,\alpha/2}$.

$$\boxed{\begin{array}{ll} \bar{\boldsymbol{Y}} \pm t_{n-1,\alpha/2}\dfrac{\boldsymbol{S}}{\sqrt{n}} & \text{is a CI for } \mu \\[2ex] \bar{\boldsymbol{Y}} \pm t_{n-1,\alpha/2}\boldsymbol{S}\sqrt{1+\dfrac{1}{n}} & \text{is a PI for } \boldsymbol{Y}_{next} \end{array}}$$

Confidence Interval for the Mean of *Y* at *X* = *x* in Simple Linear Regression

The goal is to construct an interval for $\mu_{\boldsymbol{Y}|X=x}$. In the notes on Simple Linear Regression (SLR), it is shown that $\hat{\beta}_1 \sim N\left(\beta_1, \dfrac{\sigma^2}{\sum\left(X_i-\bar{X}\right)^2}\right)$ and that $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of the true intercept and slope. Then

- $\mu_{\hat{\boldsymbol{Y}}|X=x} = \mu_{\hat{\beta}_0+\hat{\beta}_1 x} = \mu_{\hat{\beta}_0} + x\mu_{\hat{\beta}_1} = \beta_0+\beta_1 x = \mu_{\boldsymbol{Y}|X=x}$, i.e., the fitted values on the regression line are unbiased estimates of the conditional means $\mu_{\boldsymbol{Y}|X=x}$.
- Using the alternate form for the fitted value at $X = x$, $\hat{\boldsymbol{Y}} = \bar{\boldsymbol{Y}} + \hat{\beta}_1\left(x-\bar{X}\right)$, the variance at $X = x$ is

$$\sigma^2_{\hat{\boldsymbol{Y}}|X=x} = \sigma^2_{\bar{\boldsymbol{Y}}+\hat{\beta}_1\left(x-\bar{X}\right)} = \sigma^2_{\bar{\boldsymbol{Y}}} + \left(x-\bar{X}\right)^2\sigma^2_{\hat{\beta}_1} = \frac{\sigma^2}{n} + \left(x-\bar{X}\right)^2\frac{\sigma^2}{\sum\left(X_i-\bar{X}\right)^2} = \sigma^2\left(\frac{1}{n}+\frac{\left(x-\bar{X}\right)^2}{\sum\left(X_i-\bar{X}\right)^2}\right)$$

If $\sigma^2_{\boldsymbol{\varepsilon}}$ is unknown, the interval $\boxed{\left(\hat{\beta}_0+\hat{\beta}_1 x\right) \pm t_{n-2,\alpha/2}\boldsymbol{S}_{\boldsymbol{\varepsilon}}\sqrt{\dfrac{1}{n}+\dfrac{\left(x-\bar{X}\right)^2}{\sum\left(X_i-\bar{X}\right)^2}}}$ has probability $1-\alpha$ of covering $\mu_{\boldsymbol{Y}|X=x}$.

Prediction Interval for the Mean of $\boldsymbol{Y}$ at $X = x$ in Simple Linear Regression

A $(1 - \alpha)$ prediction interval for next $\boldsymbol{Y}$ at $X = x$ is $\boxed{\left(\hat{\beta}_0 + \hat{\beta}_1 x\right) \pm t_{n-2,\alpha/2} \boldsymbol{S}_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{\left(x - \bar{X}\right)^2}{\sum \left(X_i - \bar{X}\right)^2}}}$

Discussion: $\boldsymbol{S}_\varepsilon$ is the estimate of the standard deviation σ of the error variable. $\boldsymbol{S}_\varepsilon = \sqrt{MSE}$ and has $n - 2$ degrees of freedom because the two parameters β_0 and β_1 are estimated prior to computing the residuals upon which the mean square error *MSE* is based. The variance $\sigma^2 \left(1 + \frac{1}{n} + \frac{\left(x - \bar{X}\right)^2}{\sum \left(X_i - \bar{X}\right)^2}\right)$ for the *next* observation on $\boldsymbol{Y}$ at $X = x$ comes from adding the variance σ^2 of $\boldsymbol{Y}$ about its true mean $\mu_{\boldsymbol{Y}|X=x}$ at $X = x$ to the variance $\sigma^2 \left(\frac{1}{n} + \frac{\left(x - \bar{X}\right)^2}{\sum \left(X_i - \bar{X}\right)^2}\right)$ from estimating $\mu_{\boldsymbol{Y}|X=x}$.

Finally, The variance calculation for $\hat{\boldsymbol{Y}}$ at $X = x$ relied, as usual, on the assumption that $\bar{\boldsymbol{Y}}$ and $\hat{\beta}_1$ are independent. This is hardly an obvious proposition since both random variables are linear combinations of the *same n* observations, $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n$. While showing that $\bar{\boldsymbol{Y}}$ and $\hat{\beta}_1$ are independent is straightforward, the proof hinges on evaluating the covariance of the two linear combinations of the $\boldsymbol{Y}_i$ plus a theorem that jointly normally distributed random variables are independent *if and only if* they are uncorrelated.

Postscript

The term $\frac{\left(x - \bar{X}\right)^2}{\sum \left(X_i - \bar{X}\right)^2}$ that appears in the variance of the confidence interval for $\mu_{\hat{\boldsymbol{Y}}|X=x}$ and the prediction interval for $\boldsymbol{Y}$ at $X = x$ adjusts the variance based on the distance $\left|x - \bar{X}\right|$ between the *x*-coordinate of the confidence/prediction interval and the mean value of all *x*-coordinates in the sample. Result: the farther x is from $\bar{X}$ the wider the intervals are. Heuristically, the true regression line $y = \beta_0 + \beta_1 x$ and the estimated line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ tend to be closest near $\bar{X}$, but small errors in estimating the true slope β_1 by $\hat{\beta}_1$ lead to greater divergence of the lines the further removed x is from $\bar{X}$. The additional likely error is incorporated into the error variance at x through the $\frac{\left(x - \bar{X}\right)^2}{\sum \left(X_i - \bar{X}\right)^2}$ term.