What is the difference between Confidence Interval and Prediction Interval?

Authored by

stats writer

May 6, 2024

Confidence Interval and Prediction Interval are both statistical tools used to estimate the true value of a population based on a sample. However, there are fundamental differences between the two.

A Confidence Interval is a range of values calculated from a sample that is likely to contain the true population parameter with a certain level of confidence. It is used to estimate the unknown population parameter, such as the mean or proportion, and is often used in hypothesis testing. The confidence level indicates the probability that the true value falls within the calculated interval.

On the other hand, a Prediction Interval is a range of values that is likely to contain a future observation from the same population with a specified level of confidence. It takes into account the variability of the data and is used to make predictions about individual data points rather than population parameters.

In summary, the main difference between Confidence Interval and Prediction Interval is that the former is used to estimate population parameters while the latter is used to make predictions about individual data points. Additionally, Confidence Interval is based on sample data while Prediction Interval takes into account both sample and population variability.

Confidence Interval vs. Prediction Interval: What's the Difference?

Two types of intervals that are often used in regression analysis are confidence intervals and prediction intervals.

Here's the difference between the two intervals:

Confidence intervals represent a range of values that are likely to contain the true mean value of some response variable based on specific values of one or more predictor variables.

Prediction intervals represent a range of values that are likely to contain the true value of some response variable for a *single new observation* based on specific values of one or more predictor variables.

For example, suppose we fit a that uses the number of bedrooms to predict the selling price of a house:

Price = β0 + β1(number of bedrooms)

If we'd like to estimate the mean selling price of houses with three bedrooms, we would use a confidence interval.

However, if we'd like to estimate the selling price of a specific new home that just came on the market with three bedrooms, we would use a prediction interval.

Note: Since prediction intervals attempt to create an interval for a specific new observation, there's more uncertainty in our estimate and thus prediction intervals are always wider than confidence intervals.

Confidence Interval vs. Prediction Interval: Difference in Formulas

We use the following formula to calculate a confidence

interval:

?0 +/- tα/2,n-2 * Syx√((x0 - x?)2/SSx + 1/n)

We use the following formula to calculate a prediction interval:

?0 +/- tα/2,n-2 * Syx√((x0 - x?)2/SSx + 1/n + 1)

where:

?0: Estimated mean value of response variabletα/2,n-2: t-critical value with n-2 degrees of freedomSyx: Standard error of response variablex0: specific value of predictor variable x?: mean value of predictor variableSSx: Sum of squares for predictor variablen: Total sample size

Notice that the formula for a prediction interval contains an extra one in the square root portion, which means the standard error will always be larger than a confidence interval.

Example: Interpreting Confidence Intervals vs. Prediction Intervals

Suppose we have the following dataset that shows the number of bedrooms and the selling price for 20 houses

in a particular neighborhood:

Bedrooms	Price (thousands)
1	120
1	133
1	139
2	185
2	148
2	160
2	192
3	205
3	244
3	213
3	236
3	280
3	275
3	273
4	312
4	311
4	304
5	415
5	396
6	488

Now suppose we fit a simple linear regression model to this dataset in R:

```
#define data
df <- data.frame(beds=c(1, 1, 1, 2, 2, 2, 2, 3, 3, 3,
3, 3, 3, 3, 4, 4, 4, 5, 5, 6),
price=c(120, 133, 139, 185, 148, 160, 192, 205, 244, 213,
```

236, 280, 275, 273, 312, 311, 304, 415, 396, 488))

#fit simple linear regression model
model <- lm(price~beds, data=df)

#view model fit
summary(model)

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 39.450 13.248 2.978 0.00807 **
beds 70.667 4.031 17.529 9.26e-13 ***

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Residual standard error: 24.19 on 18 degrees of freedom
Multiple R-squared: 0.9447, Adjusted R-squared: 0.9416
F-statistic: 307.3 on 1 and 18 DF, p-value: 9.257e-13

The fitted regression model turns out to be:

Selling price (thousands) = 39.450 + 70.667(number of bedrooms)

We can use the following code to calculate a confidence

interval for the mean selling price of houses that have three bedrooms:

#define new house
new <- data.frame(beds=c(3))

#confidence interval for mean selling price of house with 3 bedrooms
predict(model, newdata = new, interval = "confidence")

fit lwr upr
1 251.45 240.087 262.813

The 95% confidence interval for the mean selling price of a house with three bedrooms is .

We can then use the following code to calculate a prediction interval for the selling price of a new house that just came on the market that has three bedrooms:

#define new house
new <- data.frame(beds=c(3))

#confidence interval for mean selling price of house with 3 bedrooms

predict(model, newdata = new, interval = "prediction")

fit lwr upr
1 251.45 199.3783 303.5217

The 95% prediction interval for the selling price of a new house with three bedrooms is .

Notice that the prediction interval is much wider than the confidence interval because there is more uncertainty around the selling price of a single new house as opposed to the mean selling price of all houses with three bedrooms.

The following tutorials offer additional information about confidence intervals:

The following tutorials offer additional information about prediction intervals: