Intervals

To find the confidence interval in R, create a new `data.frame` with the desired value to predict. The prediction is made with the `predict()` funct... The `interval` argument is set to 'confidence' to output the mean inter...


```
new.dat <- data.frame(speed=30)
predict(cars.lm, newdata = new.dat, interval = 'confidence')
```

```
##        fit      lwr      upr
## 1 100.3932 87.43543 113.3509
```

From the output, the fitted stopping distance at a speed of 30 mph is just above 100 feet. The confidence interval of (87.44, 113.35) signifies the range in which the true population parameter lies at a 95% level of confidence.

Prediction Intervals

The prediction interval is rather similar to the confidence interval in calculation, but as mentioned earlier, there are significant differences. The prediction interval equation is defined as:

$$\hat{y}_h \pm t_{\alpha/2,n-2}\sqrt{MSE\left(1+\frac{1}{n}+\frac{(x_k-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right)}$$

Where $\hat{y}_h$ is the fitted reponse at predictor value x_k and the critical t-value is $t_{\alpha/2,n-2}$ with $n-2$ degrees of freedom. $\sqrt{MSE\left(1+\frac{1}{n}+\frac{(x_k-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right)}$ is the standard error of prediction.

To find the prediction interval in R, the `predict()` function is utilized once again, but this time, the interval argument is given 'prediction.'

```
predict(cars.lm, newdata = new.dat, interval = 'prediction')
```

```
##        fit      lwr     upr
## 1 100.3932 66.86529 133.921
```

Notice the fitted value is the same as before, but the interval is wider. This is due to the additional term in the standard error of prediction. It should be noted prediction and confidence intervals are similar in that they are both predicting a response, however, they differ in what is being represented and interpreted. The best predictor of a random variable (assuming the variable is normally distributed) is the mean μ. The best predictor for an observation from a sample of x data points, $x_1, x_2, \cdots, x_n$ and error σ is $\bar{x}$. Since the prediction interval must take into account the variability of the estimators for μ and σ, the interval will be wider.