# More on linear regression, chapter 4.7

Grethe Hystad

November 12, 2012

$$\hat{Y} = \hat{\alpha} + \hat{\beta}(x - \overline{x})$$

is the prediction of the value of $Y$ for some given $x$. Question: How close is $\hat{Y}$ to the mean of $Y$, $E(Y)$, or to Y itself? In this section we will find

- Confidence interval for $E(Y) = \alpha + \beta(x - \overline{x})$
- Prediction interval for $Y$, given a particular value of x.

# Confidence interval for $E(Y)$

The endpoints for a $100(1-\gamma)\%$ confidence interval for $E(Y) = \alpha + \beta(x - \overline{x})$ are

$$\hat{\alpha} + \hat{\beta}(x - \overline{x}) \pm c t_{\gamma/2}(n-2),$$

where

$$c = \sqrt{\frac{n\hat{\sigma}^2}{n-2}} \sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}.$$

Details are given in class.

The points $(x_1, y_1)$, $(x_2, y_2)$,....,$(x_n, y_n)$ were used to estimate $\alpha$ and $\beta$.
Suppose that we are given a value of $x$, say $x_{n+1}$. A point estimate of the corresponding value of $Y$ is

$$\hat{y}_{n+1} = \hat{\alpha} + \hat{\beta}(x_{n+1} - \overline{x}).$$

However, $\hat{y}_{n+1}$ is just one possible value of the random variable

$$Y_{n+1} = \alpha + \beta(x_{n+1} - \overline{x}) + \epsilon_{n+1}.$$

We will obtain a **prediction interval** for $Y_{n+1}$, where $x = x_{n+1}$.

# Prediction interval for $\hat{Y}_{n+1}$

The endpoints for a $100(1-\gamma)\%$ prediction interval for $\hat{Y}_{n+1}$ are

$$\hat{\alpha} + \hat{\beta}(x_{n+1} - \overline{x}) \pm d t_{\gamma/2}(n-2),$$

where

$$d = \sqrt{\frac{n\hat{\sigma}^2}{n-2}} \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}.$$

Details are given in class.

# Prediction interval

Consider the example from chapter 4.6, where we measured average January minimum temperature in degrees Fahrenheit with the latitude of 56 U.S. cities from 1931-1960.versus. We show here the 95% prediction interval for the average Jan minimum temperature for the first 8 values given in the order below.

```
> Latitude
 [1] 31.2 32.9 33.6 35.4 34.3 38.4 40.7 41.7 40.5 39.7 31.0 25.0 26.3 33.9 43.7
[16] 42.3 39.8 41.8 38.1 39.0 30.8 44.2 39.7 42.7 43.1 45.9 39.3 47.1 41.9 43.5
[31] 39.8 35.1 42.6 40.8 35.9 36.4 47.1 39.2 42.3 35.9 45.6 40.9 40.9 33.3 36.7
[46] 35.6 29.4 30.1 41.1 45.0 37.0 48.1 48.1 43.4 43.3 41.2
> Temp
 [1] 44 38 35 31 47 42 15 22 26 30 45 65 58 37 22 19 21 11 22 27 45 12 25 23 21
[26]  2 24  8 13 11 27 24 14 27 34 31  0 26 21 28 33 24 24 38 31 24 49 44 18  7
[51] 32 33 19  9 13 14
> newdata = data.frame(Latitude)
> predict(lm(Temp~Latitude), newdata, interval="predict")
        fit        lwr      upr
1  42.908601 28.1672274 57.64998
2  39.322302 24.6845201 53.96008
3  37.845590 23.2432345 52.44795
4  34.048332 19.5174243 48.57924
5  36.368879 21.7976843 50.94007
6  27.719569 13.2440411 42.19510
7  22.867517  8.3800653 37.35497
8  20.757929  6.2505759 35.26528
```

# Prediction interval

For example the 95% prediction interval of the average Jan minimum temperature for the latitude of 31.2 degrees north of equator is between 28.1672274 F and 57.64998 F.

# Confidence interval

We show here the 95% confidence interval for the mean average Jan minimum temperature for the first 13 values given in the order below.

```
> predict(lm(Temp~Latitude), newdata, interval="confidence")
         fit       lwr      upr
1  42.908601 39.519708 46.29750
2  39.322302 36.416936 42.22767
3  37.845590 35.124323 40.56686
4  34.048332 31.741000 36.35566
5  36.368879 33.820113 38.91765
6  27.719569 25.791514 29.64762
7  22.867517 20.851892 24.88314
8  20.757929 18.603912 22.91195
9  23.289434 21.294846 25.28402
10 24.977105 23.042053 26.91216
11 43.330519 39.882068 46.77897
12 55.988046 50.610283 61.36581
13 53.245582 48.301855 58.18931
```

For example the 95% confidence interval of the mean average Jan minimum temperature for the latitude of 31.2 degrees north of equator is between 39.519708 F and 46.29750 F.