11.2 Linear Regression

The least squares regression line can be computed for any collection of data points, regardless of how they are obtained. But in many cases, we may want to view the points as the values of random variables. In particular, we might suppose that X and Y are random variables. Often, X will represent a variable we have control over, have easier access to observe, or believe to determine the values of other variables, while the variable Y may one about which we want to learn, or whose values are determined in by other variables. We call the variable X the **explanatory** (or **predictor**) variable, and the variable Y the **response** variable.

Using conditional expectation, we can decompose Y as

$$Y = E[Y|X] + (Y - E[Y|X])$$

We call the function $f(X) = E[Y|X]$ the **regression function** of Y on X, and call $\epsilon = Y - E[Y|X]$ the **residual**. Using properties of conditional expectation, the residual is *uncorrelated* with X.

The simplest regression model is linear, which assumes that

$$E[Y|X] = \beta_0 + \beta_1 X$$

The coefficients of this linear function β_0, β_1 are called the **regression coefficients**.

In general, we treat these as unknown parameters of a model, and seek to estimate them. To do so, we obtain a random sample of size n from the distribution of (X, Y). We can record the sample as a vector $\mathbf{y} = (y_1, \dots, y_n)$ and a vector $\mathbf{x} = (x_1, \dots, x_n)$.

One candidate for estimators for the regression coefficients β_0, β_1 are the coefficients for least squares line based on the data $\mathbf{x}, \mathbf{y}$ discussed previously. We'll now consider properties of the distribution of these estimators.

Simple Linear Regression

Assume that for all value $X = x$, the random variable Y can be represented as

$$Y = \beta_0 + \beta_1 x + \epsilon$$

We make the following assumptions about this model:

1. The values $x_1, \dots, x_n$ of the predictor X are fixed.
2. The conditional distribution of Y_i given $(x_1, \dots, x_n)$ is normal.
3. There are parameters β_0, β_1 so that
$$E[Y_i|x_1, \dots, x_n] = \beta_0 + \beta_1 x_i$$
4. There is a parameter σ^2 so that
$$\mathrm{Var}(Y_i|x_1, \dots, x_n) = \sigma^2$$
5. The variables $Y_1, \dots, Y_n$ are conditionally independent, given $x_1, \dots, x_n$.

Together, these assumptions imply that the joint conditional pdf of $\mathbf{y}$ given the other parameters is

$$f(\mathbf{y}|\mathbf{x}, \beta_0, \beta_1, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum (y_i - (\beta_0 + \beta_1 x_i))^2\right]$$

Thm: Under the assumptions above, the MLEs of β_0, β_1 are the least-squares estimates and the MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n} \sum (y_i - (\hat{\beta}_0 - \hat{\beta}_1 x_i))^2$$

Proof. Consider the log-likelihood function:

$$\log f = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum(y_i - (\beta_0 + \beta_1 x_i))^2$$

For fixed value of σ^2, differentiating with respect to β_0 and β_1 produces the same system of equations we solved to find the least squares estimates.

To find the MLE of σ^2, replace β_0, β_1 with their least squares estimates and differentiate with respect to σ^2. (Exercise) $\square$

Consider now the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum(Y_i - \bar{Y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{x}$$

Let

$$s_x = \sqrt{\sum(x_i - \bar{x})^2}$$

Note that these estimators are functions of fixed data $\mathbf{x}$ and random data $\mathbf{Y}$.

<u>Thm:</u> Under the assumptions above, $\hat{\beta}_0$ and $\hat{\beta}_1$ are Multivariate Normally. Moreover, the distribution of $\hat{\beta}_1$ is Normal with mean β_1 and variance $\frac{\sigma^2}{s_x^2}$ and the distribution of $\hat{\beta}_0$ is Normal with mean β_0 and variance

$$\sigma^2\left(\frac{1}{n} + \frac{\bar{x}}{s_x^2}\right).$$

The covariance of $\hat{\beta}_0, \hat{\beta}_1$ is

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\bar{x}\sigma^2}{s_x^2}$$

Proof. By definition,

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{s_x^2} \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{x}$$

Note that $\bar{Y}$ is a linear combination of the Y_i, and so $\hat{\beta}_1$ is a linear combination of the Y_i, and thus, $\hat{\beta}_0$ is also a linear combination of the Y_i. Hence, every linear combination of $\hat{\beta}_0$ and $\hat{\beta}_1$ is a linear combination of the Y_i. Since the Y_i are conditionally independent and Normally distributed, then every linear combination of $\hat{\beta}_0$ and $\hat{\beta}_1$ are Normally distributed, which shows that $\hat{\beta}_0, \hat{\beta}_1$ are Multivariate Normal. In particular, this also implies that $\hat{\beta}_0$ and $\hat{\beta}_1$ are themselves Normally distributed.

Now, to find the mean of $\hat{\beta}_1$, we use linearity of expectation:

$$\begin{aligned}
E[\hat{\beta}_1] =& \frac{1}{s_x^2}\sum(x_i - \bar{x})E[Y_i - \bar{Y}] \\
=& \frac{1}{s_x^2}\sum(x_i - \bar{x})\left(E[Y_i] - E[\bar{Y}]\right) \\
=& \frac{1}{s_x^2}\sum(x_i - \bar{x})\left(\beta_0 + \beta_1 x_i - (\beta_0 + \beta_1\bar{x})\right) \\
=& \frac{\beta_1}{s_x^2}\sum(x_i - \bar{x})(x_i - \bar{x}) = \beta_1
\end{aligned}$$

To find the variance of β_1, note again it is a linear combination of independent variables, so

$$\text{Var}(\hat{\beta}_1) = \frac{\sum(x_i - \bar{x})^2\text{Var}(Y_i)}{s_x^4} = \frac{\sigma^2}{s_x^2}$$

To find the mean of $\hat{\beta}_0$,

$$E[\hat{\beta}_0] = E[\bar{Y}] - E[\hat{\beta}_1]\bar{x} = \frac{1}{n}\sum(\beta_0 + \beta_1 x_i) - \beta_1\bar{x} = \beta_0$$

Computing the variance of $\hat{\beta}_0$ and the covariance between $\hat{\beta}_0$ and $\hat{\beta}_1$ is left as an exercise. $\square$

As a corollary of the preceding theorem, $\hat{\beta}_0, \hat{\beta}_1$ are unbiased estimators of the corresponding parameters. In fact, it turns out that these estimators are the **minimum variance linear unbiased estimators** for these parameters.

Consider the random observations $(x_1, Y_1), \ldots, (x_n, Y_n)$, and suppose we wish to predict the value of Y for a specific value of x. By assumption, Y is Normal with mean $\beta_0 + \beta_1 x$ and variance σ^2, and so we may use $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$ as an estimate for Y.

Thm: The MSE of this estimate is

$$E[(\hat{Y} - Y)^2] = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{s_x^2}\right]$$

Proof. Observe that $E[\hat{Y}] = E[Y] = \beta_0 + \beta_1 x = \mu$. Then

$$E[(\hat{Y} - Y)^2] = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(Y) - 2\mathrm{Cov}(\hat{Y}, Y) = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(Y)$$

Now, as $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$, then

$$\begin{aligned}
\mathrm{Var}(\hat{Y}) + \mathrm{Var}(Y) =& \mathrm{Var}(\hat{\beta}_0) + x^2 \mathrm{Var}(\hat{\beta}_1) + 2x\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) + \sigma^2 \\
=& \frac{\sigma^2}{s_x^2} + x^2\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}\sigma^2}{s_x^2}\right) - \frac{2x\bar{x}\sigma^2}{s_x^2} + \sigma^2 \\
=& \sigma^2 \left(1 + \frac{1}{n} + \frac{(\bar{x} - x)^2}{s_x^2}\right)
\end{aligned}$$

□

Note that MSE is smallest when $x = \bar{x}$ and increases as x gets further from $\bar{x}$.