richardwu.ca

STAT 331 Course Notes

Applied Linear Models

Peter Balka • Winter 2018 • University of Waterloo

Last Revision: January 9, 2018

Table of Contents

Abstract

These notes are intended as a resource for myself; past, present, or future students of this course, and anyone interested in the material. The goal is to provide an end-to-end resource that covers all material discussed in the course displayed in an organized manner. If you spot any errors or would like to contribute, please contact me directly.

1 January 4, 2018

1.1 Simple linear regression review

In SLRM, there is a single explanatory variate and a response variate.
A good graphical summary for SLRM are **scatterplots**.
A good numerical summary for SLRM is the **correlation coefficient** defined as

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

where $-1 \leq r \leq 1$. If $|r| \approx 1$ then the explanatory/response variates have a strong linear relationship.

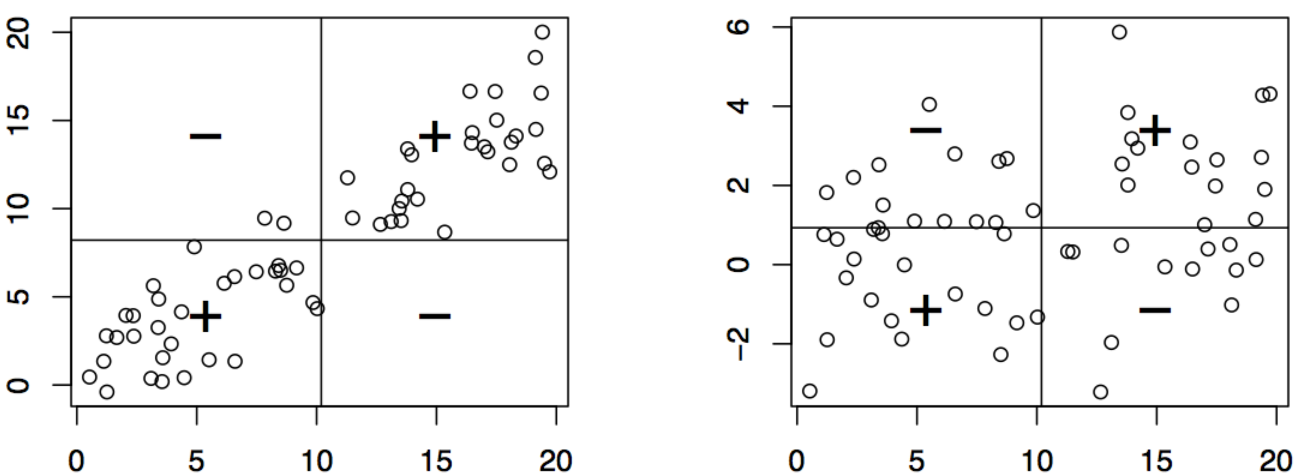
2 January 9, 2018

2.1 Correlation coefficient and covariance

Note: the measure r is also the covariance divided by the standard deviations or

$$r = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Note that the covariance $E[(X - E[X])(Y - E[Y])]$ can be graphically separated by the means $\bar{X}$ and $\bar{Y}$.



One can see that the covariance signage is determined by the sum of the magnitudes in the positive and negative quadrants.

2.2 Simple linear regression (SLR) model

An SLR model can be thought of as a line with covariates x and y where

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, \ldots, n$$

where ϵ_i is some error term for each i.

Example 2.1. From the dataset

Overhead	Office Size
218955	1589
224513	1912
$\vdots$	$\vdots$

Thus we have the SLR model

$$\begin{aligned} 218955 &= \beta_0 + \beta_1(1589) + \epsilon_1 \\ 224513 &= \beta_0 + \beta_1(1912) + \epsilon_2 \end{aligned}$$

2.3 Methods of least squares

Find (estimate) the value of β_0, β_1 (denoted by $\hat{\beta}_0, \hat{\beta}_1$, respectively) that minimizes the sum of squares of the errors $\sum_{i=1}^n \epsilon_i^2$. That is: we find values of β_0, β_1 that minimizes the function

$$S(\beta_0, \beta_1) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2$$

We take the partial derivatives and set to 0 to find the minimum (assuming convexity)

$$\begin{aligned} \frac{\partial S}{\partial \beta_0} &= -2 \sum_{i=1}^n y_i - (\beta_0 + \beta_1 x_i) = 0 \\ \frac{\partial S}{\partial \beta_1} &= -2 \sum_{i=1}^n x_i [y_i - (\beta_0 + \beta_1 x_i)] = 0 \end{aligned}$$

which yields (the notation changes to estimates of β assuming we can calculate those)

$$\begin{aligned} \sum_{n=1}^n y_i &= n\hat{\beta}_0 + \sum_{n=1}^n x_i \hat{\beta}_1 \\ \sum_{n=1}^n x_i y_i &= \sum_{i=1}^n x_i \hat{\beta}_0 + \sum_{n=1}^n x_i^2 \hat{\beta}_1 \end{aligned}$$

which gives us the estimates

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

The corresponding fitted line is

$$\hat{\mu}_{y|X=x} = \hat{\mu} + \hat{\beta}_0 + \hat{\beta}_1 x$$

For the example with overhead above, we'd have

$$\hat{\mu} = -27877.06 + 126.33x$$

2.4 (Fitted) residuals

These are the difference between the actual values and our fitted value (distinct from the error terms previously)

$$e_i = (y_i - \hat{\mu}_i) = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

Some **key points** regarding this model

- By estimating two parameters (β_0, β_1), we have imposed two constraints on our residuals (from our partial derivatives)

  $$\sum e_i = 0$$
  $$\sum x_i e_i = 0$$

  These reduces our number of n independent measures by 2 since we can compute the remaining two residuals from $n - 2$ observations. Thus we have $n - 2$ **degrees of freedom** (or in general, $n - k$ dfs where k is the number of estimated parameters>).

2.5 Interpretation of estimated parameters $\hat{\beta}_i$

β_1

$$\begin{aligned} \hat{\mu} &= \hat{\beta}_0 + \hat{\beta}_1 x \\ \hat{\mu_{x+1}} &= \hat{\beta}_0 + \hat{\beta}_1(x+1) \\ &= \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_1 \\ &= \hat{\mu} + \hat{\beta}_1 \end{aligned}$$

thus $\hat{\beta}_1$ can be interpreted as the estimated mean change in the response (y) associated with one unit change of x.

β_0 For $x = 0$, $\hat{\mu} = \hat{\beta}_0$.

However, in the example with overhead, it's evident that when $x = 0$ overhead is negative (-27877.06) which is nonsensical.

Never extrapolate results outside the range of the values of the explanatory variate(s).