MIE237

Neil Montgomery
2016-01-29

Regression

Linear models

- Model: Output = Input + Noise
- $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad x_i \in \mathbb{R}$
- $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, \quad x_{ji} \in \mathbb{R}$
- $Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i, \quad x_{ji} \in \mathbb{R}$

In each case $\varepsilon_i \sim N(0, \sigma^2)$.

Linear models in matrix form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Linear models in matrix form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

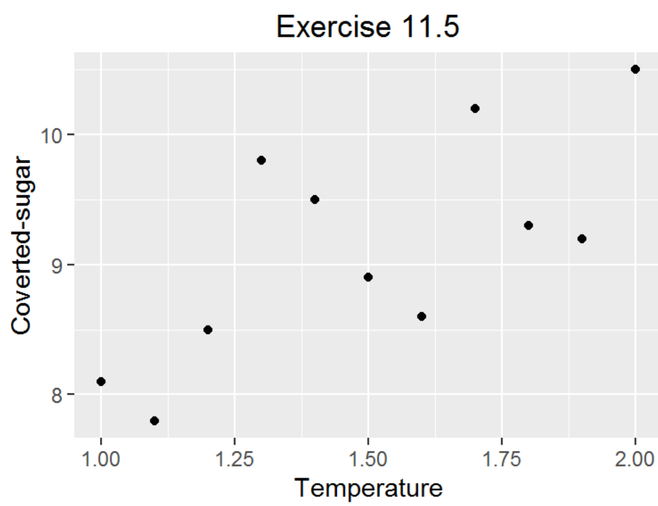
$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}$$

"Simple" linear regression

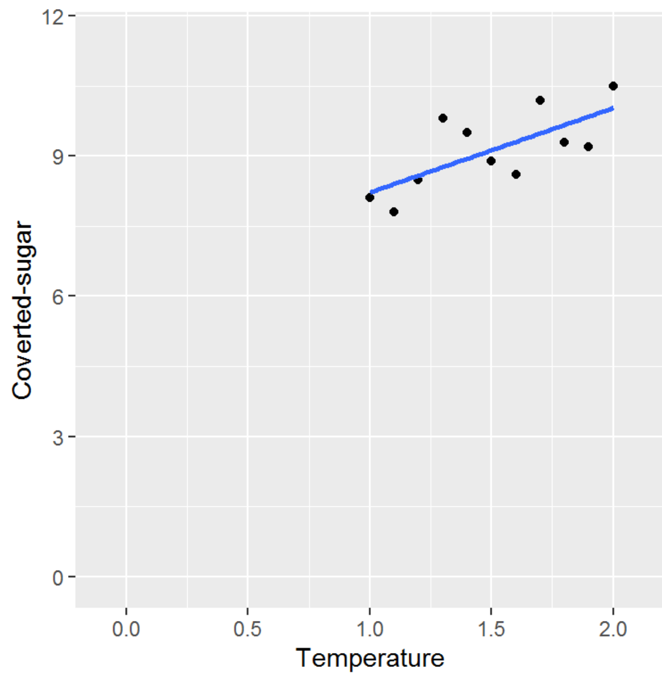
$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$\varepsilon \sim N(0, \sigma^2)$$

β_1 is the "slope" parameter and β_1 is the "intercept" parameter. So there are three unknown parameters that need to be estimated using data.



Parameter interpretation



The slope is the change in y when x changes by 1 unit.

The intercept rarely matters in and of itself.

Caution: plots usually present regression data with axes rescaled to be pleasing to the eye.

Example regression results using R (ugly version)

```
##
## Call:
## lm(formula = `Coverted-sugar` ~ Temperature, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7082 -0.4868 -0.1227  0.5109  1.0346
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.4136     0.9246   6.936 6.79e-05 ***
## Temperature   1.8091     0.6032   2.999    0.015 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6326 on 9 degrees of freedom
## Multiple R-squared:  0.4999, Adjusted R-squared:  0.4443
## F-statistic: 8.996 on 1 and 9 DF,  p-value: 0.01497
```

Pretty version

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.4136	0.9246	6.94	0.0001
Temperature	1.8091	0.6032	3.00	0.0150

Preview of "correlation"

Recall $\mathrm{Cov}(X,Y) = E(X - E(X))(Y - E(Y)) = E(XY) - E(X)E(Y)$.

Measures linear relationship between X and Y.

A related measure that divides by the two standard deviations:

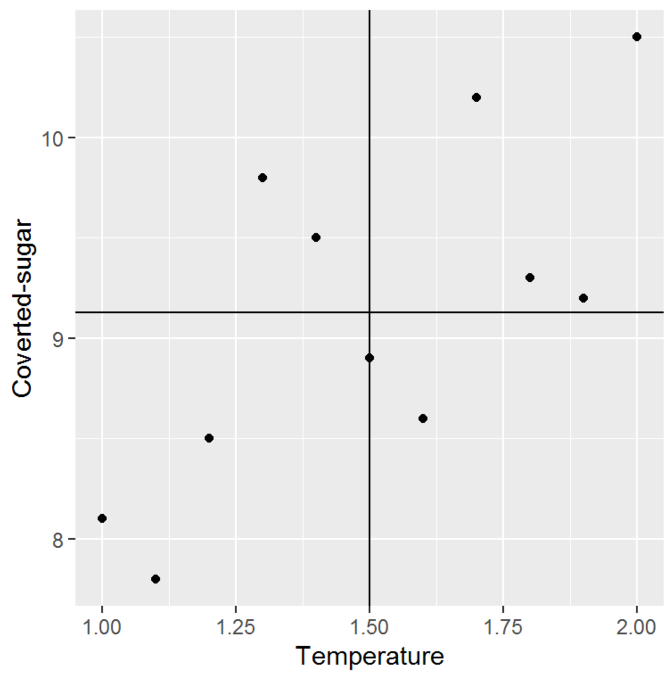
$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

"Correlation coefficient"

Preview of correlation

ρ is a measure of the linear relationshipbetween the random variables X and Y.

We'd like an analogous measure for :



We will base it on:

$S_{xy} = \sum_{i=1}^{n} \left(x_i - \overline{x}\right) \left(y_i - \overline{y}\right)$

"Least squares" estimation for model parameters

Model: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ (Or: $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$)

Estimation method is to minimize (for β_0 and β_1) the "squared error":

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2$$

More generally, minimize:

$$\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} = (\boldsymbol{Y} - \boldsymbol{X\beta})' (\boldsymbol{Y} - \boldsymbol{X\beta})$$

Results

$$\hat{\beta}_1 = \frac{S_{xY}}{S_{xx}}$$

where $S_{xx} = \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2$ (scales out the variation in the x direction)

and

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{x}$$

More generally:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}' \boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{y}$$

where the key thing to notice is that it involves a matrix inversion.

More on $\hat{\beta}_1$

Estimation based on the model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.

Other model assumptions: ε_i are i.i.d. $N(0, \sigma^2)$. So the Y_i are also normal, with mean $\beta_0 + \beta_1 x_i$ and variance σ^2 .

Important to remember that $\hat{\beta}_1$ is a with a distribution, mean, and variance of its own. (β_0 too.)

In fact $E\left(\hat{\beta}_1\right) = \beta_1$ and its variance is \frac{\sigma^2}{S_{xx}.