# Regression

## An upgraded review

Jan Rovny

December 19, 2019

# Regression Logic

- Regression equation:

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- $e$ captures all the reasons why the prediction $\hat{y}_i$ deviates from the observed value $y_i$. This includes:
  - Measurement error
  - Omitted predictors
  - Idiosyncratic sources of behavior
- Error can thus be written as:

$$\begin{aligned} e_i &= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \\ &= y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \\ &= y_i - \hat{y}_i \end{aligned}$$

# Regression Estimation

- To estimate regression coefficients we minimize:

$$
\begin{aligned}
S &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \\
&= \sum_{i=1}^{n} e_i^2
\end{aligned}
$$

- We minimize the squared residuals
- This means we find a regression line that is closest to the observed values on $y_i$

# Regression Assumption #1

- **I. The most important assumption**
- *1. Model is correctly specified*
- Formally: *Mean Independence:* $E(\epsilon_i) = 0$, which means that the mean value of $\epsilon$ does not depend on any of the predictors.
- Model includes all relevant predictors in the correct functional form (squares, interactions etc.).
- If this does not hold, there is omitted variable bias, the OLS estimator is biased and inconsistent = WRONG
- Specification error is a central problem for which there is no statistical solution.
- **We must turn to theory!**

# Assumptions about Errors

- *2. Linearity*: $y$ is a linear function of the $x$s.
  - Violation of 1. and 2. causes point estimate bias!
- *3. Normality*: $\epsilon_i \sim N(0, \sigma^2)$ We assume that the error is normally distributed (around the regression line).
  - 3. is important for inference, allows us to use t-tests.
- *4. Homoscedasticity*: $Var(\epsilon_i) = \sigma^2$: variance of errors is constant.
- *5. Nonautocorrelation*: $Cov(\epsilon_i, \epsilon_j) = 0$ $(i \neq j)$ , errors are independent. (Problem in time-series data.)
  - 4. and 5. do not effect point estimates, only determine the standard errors.

# Interpretations

- We are interested in the effects of our predictors $x$
- These are summarized by $\beta_x$ coefficients, which have a clear interpretation:
  - for every 1 unit change in $x$, $y$ responds by $\beta_x$ units
- We want to make sure that the effect of $x$ is assessed independently of some other predictor $z$
- Consequently we specify a multiple regression equation:

$$y_i = \beta_0 + \beta_1 x1_i + \beta_2 x2_i + ... + \beta_k xk_i + e_i$$

- Here, the individual $\beta$s become partial regression coefficients, independent of all effects of the other $x$s in the model

# Predicted values

- To assess the specific effects of some $x$, a useful method is the calculation and visualization of *predicted values* $\hat{y}$.
- To make this meaningful, we specify the values of the other $x$s, effectively holding them constant, while varying the $x$ of interest, and calculating $\hat{y}$ for every iteration.
- Example:
  - Predict left-right placement as a function of: gender, age, education, income, religious affiliation, climate views
    ```
    m<-lm(lr~female+age+as.factor(educ)+inc+mus+clim_deny,
                   data=D)
    ```
  - Assess the effect of climate views as other variables are held constant.

# Predicted values exercise

- To asses predicted values
  - Enter values of $\beta$s
  - Set values of predictors ($x$s), keeping all but one constant.
  - Solve for $\hat{lr}$

$$\hat{lr} = \beta_0 + \beta_2 * female + \beta_3 * age + \beta_4 * educ1 + \beta_5 * educ2 + \beta_6 * educ3 + \beta_7 * inc + \beta_8 * mus + \beta_9 * clim\_deny$$

- $\hat{\beta}$s estimated in R:

| (Intercept) | female | age | educ1 | educ2 | educ3 | inc | mus | clim_deny |
|---|---|---|---|---|---|---|---|---|
| 4.37 | 0.09 | 0.005 | -0.08 | -0.119 | -0.45 | 0.06 | -0.93 | 0.43 |

- $x$ values set by us

[See demonstration in R]

# Matrix Algebra

# Vectors and Matrixes

- We have so far written our equations in normal algebraic notation
- But really, our dependent variable $y_i$ is a vector of length $n$
- Our predictors $x_1, x_2, x_3 ... x_k$ form a matrix of $n$ rows and $k$ collumns
- Our estimates $\beta_0, \beta_1 ... \beta_k$ are a vector of length $k + 1$ ($k$ predictors + intercept)
- We can thus note them, and operate with them accordingly
- To do this, we use *matrix algebra* notation and operations

# Matrix notation

- Note that vectors and matrixes are in **bold**, while individual values are not
- Vectors are smaller case letters ($\boldsymbol{y}$, $\boldsymbol{e}$), while matrixes are capitalized ($\boldsymbol{X}$)

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix} \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$$

$$\mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{pmatrix}$$

# Matrix operations - addition, subtraction

- Note two matrixes **A** and **B**:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \mathbf{B} = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix}$$

- We can add and subtract:

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} 3 & 4 & 5 \\ 6 & 7 & 8 \\ 9 & 10 & 11 \end{pmatrix} \qquad \mathbf{A} - \mathbf{B} = \begin{pmatrix} -1 & 0 & 1 \\ 2 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix}$$

# Matrix operations - transpose

- We can transpose a matrix:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$
$$\mathbf{A}' = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$$

- Here **A**' (read "A prime") is a transpose of **A**
- Transpose is in a sense a 'two-dimensional' rotation of the matrix

# Matrix operations - multiplication

- *Scalar multiplication*:

$$4 * \mathbf{A} = 4 \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 4 & 8 & 12 \\ 16 & 20 & 24 \\ 28 & 32 & 36 \end{pmatrix}$$

- *Matrix multiplication* is not straight forward
- It combines both multiplying and adding

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} * \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}$$

# Matrix operations - identity matrix

- *Identity matrix* is a special kind of matrix which is the size of *nxn*
- contains 1s in the diagonal positions
- 0s in the off-diagonal positions

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

- Multiplying by an identity matrix leads to no change $\mathbf{A} * \mathbf{I} = \mathbf{A}$

# Matrix operations - inversion

- *Inversion* of a matrix $\mathbf{C}$ is an operation by which we look for an inverse matrix $\mathbf{C}^{-1}$, so that: $\mathbf{C}\mathbf{C}^{-1} = \mathbf{I}$
- Not all matrixes are inversible. We call these singular.
- While $\mathbf{A}$ and $\mathbf{B}$ are singular, $\mathbf{C}$ is inversible:

$$\mathbf{C} = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 6 \\ 4 & 6 & 8 \end{pmatrix} \mathbf{C}^{-1} = \begin{pmatrix} 1 & -2 & 1 \\ 2 & 2 & -0.5 \\ 1 & -0.5 & 0 \end{pmatrix}$$

# OLS in Matrix Notation

# OLS estimation in matrix notation

- From the basic regression equation, we know that:

$$e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

- In matrix notation, we simplify this to:

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\hat{\beta}$$

- Here, $\boldsymbol{X}$ and $\boldsymbol{y}$ are known, and $\hat{\beta}$ is to be estimated so as to minimize $\boldsymbol{e}$

# OLS estimation in matrix notation

- In effect, we minimize the sum of squared residuals (RSS), which is $\boldsymbol{e'e}$

$$(e_1, e_2, \ldots, e_n)_{1xn} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}_{nx1} = (e_1 * e_1 + e_2 * e_2 + \ldots + e_n + e_n)_{1x1}$$

- $\boldsymbol{e'e}$ can be written using the equation above as:

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

# OLS estimation in matrix notation

- Now, we need to minimize $\boldsymbol{e'e}$:

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

- This yields so-called ‘normal equations’

$$\boldsymbol{X}'\boldsymbol{X}\hat{\beta} = \boldsymbol{X}'\boldsymbol{y}$$

- Or, if $\boldsymbol{X}$ is inversible (non-singular):

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

- In R: `beta<-solve(t(X) %*% X) %*% t(X) %*% y`

# OLS estimation in matrix notation

- Under what conditions can $\boldsymbol{X}'\boldsymbol{X}$ be inverted?
  1. $N \geq k + 1$, this is usually met easily
  2. the columns of $\boldsymbol{X}$ must be linearly independent (no multicollinearity)