STAT1006 Week 9 Cheat Sheet

Lisa Luff

10/17/2020

Contents

Matrix Approach to Linear Regression

- Still exactly the same as linear regression, just a different representation
- Notation:
  - Vectors are lowercase bold letters: $\boldsymbol{y}, \hat{\boldsymbol{\varepsilon}}, \boldsymbol{\beta}, \hat{\boldsymbol{\beta}}$
  - Matrices are uppercase bold letters: $\boldsymbol{X}, \boldsymbol{I}$
  - Scalars represented by italiscised letters: y_i, x_i

Matrix Revision

Matrix Position Labels

$$\boldsymbol{X} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \rightarrow a_{ij}$$

* This pattern is represented as a_{ij}, where * i represents the i-th row * j represents the j-th column * Vectors are just n x 1 matrices, and we will treat them as matrices * They are represented as a single column of data with n rows

Matrix Transpose:

- The transpose of an r x c matric, is a c x r matrix
  - Written as $\boldsymbol{X'}$
  - The elements of $\boldsymbol{X}$ are x_{ij}
  - The elements of $boldsymbol X'$ are x_{ji}

$$\boldsymbol{X} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \rightarrow \boldsymbol{X'} = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$$

Special Matrices

- Symmetrical matrices:
  - $\boldsymbol{A} = \boldsymbol{A'}$
  - This will be n x n matrices
  - This means that when transposed, the top left, to bottom right diagonal is the same as the original

$$\boldsymbol{A_{2x2}} = \begin{pmatrix} \boxed{1} & 2 \\ 3 & \boxed{4} \end{pmatrix} \rightarrow \boldsymbol{A'_{2x2}} = \begin{pmatrix} \boxed{1} & 3 \\ 2 & \boxed{4} \end{pmatrix}$$

- Diagonal matrices:
  - Where the matrix is made up of zeros, except for the top left, to bottom right diagonal
  - Will be a n x n matrix

$$\boldsymbol{A_{3x3}} = \begin{pmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{2} & 0 \\ 0 & 0 & \boxed{3} \end{pmatrix}$$

- Identity matrices:
  - The same as a diagonal matrix, except the diagonal is filled with ones

$$\boldsymbol{I_{3x3}} = \begin{pmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \end{pmatrix}$$

- Unity Matrices:
  - All ones
  - Shape/size is irrelevant

$$\mathbf{1_{2x3}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$$

- Zero matrices:
  - Same as unity matrix, except all zeroes

$$\mathbf{1_{2x3}} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Matrix Inversion

- You can use the identity matrix to find the inverse
- The inverse is the matrix you multiply the original matrix by to end up with an identity matrix

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

- We want to find unknowns a, b, c, d

$$\boldsymbol{A^{-1}} = \begin{pmatrix} \text{a} & \text{b} \\ \text{c} & \text{d} \end{pmatrix} \rightarrow$$

- We know that

$$\boldsymbol{AA^{-1}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \rightarrow$$

- And we have the equations

$$\boldsymbol{AA^{-1}} = \begin{pmatrix} \text{1a + c} & \text{1b + d} \\ \text{4c} & \text{4d} \end{pmatrix}$$

- We are then able to then show that

$$\boldsymbol{A^{-1}} = \begin{pmatrix} 1 & \text{-} \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix} \rightarrow \text{ and so },a = 1;\, b = -\frac{1}{4};\, c = 0;\, d = \frac{1}{4}$$

Matrix Addition/Subtraction

- Just add the corresponding elements
- Can only add/subtract matrices with matching dimensions

Matrix Multiplication

- For each element of matrix $\boldsymbol{A}$, multiply x_i with x_j from matrix $\boldsymbol{B}$, where $i = j$
- Therefore Matrix $\boldsymbol{A}$ must have the same number of rows, as matrix $\boldsymbol{B}$ has columns

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} \text{a} & \text{b} \\ \text{c} & \text{d} \\ \text{e} & \text{f} \end{pmatrix} = \begin{pmatrix} \text{1a} & \text{2c} & \text{3e} \\ \text{4b} & \text{5d} & \text{6f} \end{pmatrix}$$

SLR in Matrix Form

- For $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$; $Var(\varepsilon_i) = \sigma^2$
- It can be written as follows:

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}; \boldsymbol{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ 1 & X_n \end{pmatrix}; \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{pmatrix}; \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

  - Where **Y** is n x 1, **X** is n x 2, $\boldsymbol{\varepsilon}$ is n x 1, and $\boldsymbol{\beta}$ is 2 x 1
- X has a row of 1's because you need to have as many rows as there are unknown parameters (not including σ). The 1 allows us to find β_0
- So;

$$\boldsymbol{X\beta} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \beta_0 & \beta_1 X_1 \\ \beta_0 & \beta_1 X_2 \\ . & . \\ . & . \\ \beta_0 & \beta_1 X_n \end{pmatrix}$$

- So the simple linear model in matrix notation is:
  - $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$
- Due to transposing, the sum of squares terms are given by
  - $\sum \varepsilon_i^2$
  - For any vector a, $a'a$ represents the sum of the squares of the elements of a

SLR Matrix Means, Variance and Covariance

- Mean:
  - We know that $E(Y_n|X) = \beta_0 + \beta_1 X_n$
  - This is equivalent to $E(\boldsymbol{Y}|\boldsymbol{X}) = \boldsymbol{X\beta}$
    * Because $E(\boldsymbol{\varepsilon}) = 0$
- For any set of observations Y, p is the number of parameters estimated (number of explanatory variables plus one for β_0)
- Variance:
  - we know that $Var(Y) \approx \frac{1}{n-c} \sum (Y_i - \hat{Y}_i)^2$
    * And that the linear assumptions means $Var(\varepsilon_i) = \sigma^2$
- Covariance
  - We know that $Cov(Y, X) \approx \frac{1}{n} \sum (Y_i - \overline{Y}_i)(X_i - \overline{X}_i) = \frac{1}{n}(\sum X_i Y_i - n\overline{XY})$
    * And the linear assumptions means that $Cov(\varepsilon_i, \varepsilon_j) = 0$
- In matrix form $Var(\varepsilon) = \sigma^2 I_n$
  - Where I_n is an identity matrix of size n
  - For a vector the term Var refers to a matrix containing the variance of each element of the vector, plus all the possible covariances between elements in the vector

$$E(\boldsymbol{\varepsilon}) = \begin{pmatrix} \mathrm{E}(\varepsilon_1) = 0 \\ \mathrm{E}(\varepsilon_2) = 0 \\ . \\ . \\ \mathrm{E}(\varepsilon_n) = 0 \end{pmatrix}; Var(\boldsymbol{\varepsilon}) = \begin{pmatrix} Var(\varepsilon_1) = \sigma^2 & Cov(\varepsilon_1, \varepsilon_2) = 0 \\ Cov(\varepsilon_2, \varepsilon_2) = 0 & Var(\varepsilon_2) = \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

- so for any vector **Y**:
  - Var(**Y**) = E[(**Y** - E(**Y**))(**Y** - E(**Y**))']
  - Var(a**Y**) = a^2Var(**Y**)
  - Var(**AY**) = **A**Var(**Y**)**A'**
  - E(**AY**) = **A**E(**Y**)
  - Cov(**X**, **Y**) = E[(**X** - E(**X**))'(**Y** - E(**Y**))]

Least Squares Estimation

- To estimate $\boldsymbol{\beta}$, we must minimise SSE
  - $SSE = \boldsymbol{\varepsilon'\varepsilon} = (\boldsymbol{Y} - \boldsymbol{X\beta})$
- To do this we solve the Normal equations
  - $\sum Y_i = n\beta_0 + \beta_1 \sum X_i$
  - And $\sum X_i Y_i = \beta_0 \sum X_i + \beta_1 \sum X_i^2$
- In matrix form
  - $\boldsymbol{X'Y} = \boldsymbol{X'X\beta}$
  - $(\boldsymbol{X'X})^{-1}\boldsymbol{X'Y} = \hat{\boldsymbol{\beta}}$

Estimates, Predicted Values, Sample Variance and Covariance

- Unbiased estimate
  - $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$
  - The LSR estimates are unbiased
- Predicted values and residuals
  - $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$
  - $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$
  - And $\hat{\boldsymbol{Y}} = \boldsymbol{HY}$, or the hat matrix
    - It transforms the observations $\mathbf{Y}$ into the predicted values
    - $\mathbf{HH} = \mathbf{H}$
- So, $\hat{\boldsymbol{\varepsilon}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$
- And so the variances of the parameter estimates
  - $Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X'X})^{-1}$
  - This matrix will be the same as before, except it will contain $Cov(\hat{\beta}_0, \hat{\beta}_1)$

Confidence interval for the Regression Line

- We know that $Var(\hat{Y}|X_0) = \sigma^2\left(\frac{1}{n} + \frac{(X_0-\overline{X})^2}{\sum(X_i-\overline{X})^2}\right)$
  - So, $Var(\hat{(Y)}|X_0) = \sigma^2\boldsymbol{X_0'}(\boldsymbol{X'X})^{-1}X_0$
- For a given point $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_0 = (1X_0)\hat{\boldsymbol{\beta}} = \boldsymbol{X_0'}\hat{\boldsymbol{\beta}}$
  - So for a given point $Var(Y|X_0) = \sigma^2[1 + \boldsymbol{X_0'}(\boldsymbol{X'X})^{-1}X_0]$

Thinking in Terms of Matrices

1. Write down the Normal Equations for regression
2. Write these as a matrix equation
3. Consider how each element of the matrix and vector that don't involve $\boldsymbol{\beta}$ can be written as an inner product
4. What vectors are involved in these inner products?
5. Define a matrix $\mathbf{X}$ and a vector $\mathbf{Y}$, so that you can write the Normal Equations as $\boldsymbol{X'X\beta} = \boldsymbol{X'Y}$
6. Use the same $\mathbf{X}$ and $\mathbf{Y}$ to write out the original equations relating responses to covariates

Multiple Regression

- When you want to look at multiple explanatory variables for one response variable
- One Y, multiple X's
- Population multiple regression equation:
  - $\mu_y = \beta_0 + \beta_1 x_i + \beta_2 x_2 + ... + \beta_p x_p$
    - Where p is the number of explanatory variables

Multiple linear regression model:

- $y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \varepsilon_i$
  - For i = 1, 2, . . . , n
  - This is a linear function of the explanatory variables
    - It is no longer a straight line, but it is a multidimentional straight(linear) plane
    - Can use ggplot to plot 3 dimentions in R (max dimensions)
- Matrix form:
  - $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix} ; \boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & ... & X_{1(p+1)} \\ 1 & X_{21} & X_{22} & ... & X_{2(p+1)} \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & X_{n1} & X_{n2} & ... & X_{n(p+1)} \end{pmatrix} ; \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{pmatrix} ; \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_{p+1} \end{pmatrix}$$

  - Where $X = X_{ij}$, where i = 1, 2, . . . , n, and j = 1, 2, . . . , p, p + 1
  - And **Y** is n x 1, **X** is n x (p + 1), $\boldsymbol{\varepsilon}$ is n x 1, and $\boldsymbol{\beta}$ is (p + 1) x 1
- You can write **X** as $\boldsymbol{X} = [1, X_1, X_2, ..., X_p]'$
- The coefficient β_i represents the average change in the response when the variable x_i increases by one unit and **all other variables are held constant**
- Assumptions:
  - The deviations ε_i are independent and Normally distributed $N(0, \sigma)$
    - That σ^2 is unknown
  - Check these by ensuring there is no pattern in the residual plot
  - That n > p
    - if n < p, possibly infinitely many solutions
    - if n = p, a solution, if it exists, is unique
    - if n > p, no solution
      - So we estimate using least squares

Parameter Estimation

- We know that $E(Y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$
  - Don't need to include variables that are not significant
- And that $\sigma^2 \approx s^2 = \frac{\sum e_i^2}{n-p-1} = \frac{\sum (y_i - \hat{y}_i)^2}{n-p-1}$
- For the least squares estimate:
  - $y = \boldsymbol{X\beta} + \varepsilon, \varepsilon \sim N(0, \boldsymbol{I}\sigma^2)$
  - And so the least squares estimate is
    - $\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$
    - So We can find the fitted values and hat matrix the same as for SLR
- Covariance matrix:
  - $Var(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X'X})^{-1}\sigma^2$
  - $s^2 = \frac{SSE}{n-p-1} = \frac{1}{n-p-1}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$
    - The only difference is the denominator

Confidence Intervals

- Confidence intervals are per β_j
  - A level C confidence interval for β_j is
    - $\beta_j \pm t^* SE_{\beta_j}$

Significance Test

- Also per β_j
- Testing the hypothesis $H_0 : \beta_j = 0$
  - Calculate the t statistic
    - $t = \frac{\beta_j}{SE_{\beta_j}} \sim t(n-p-1)$
    - When the null is true
  - Software usually gives a two-sided p-value, so halve for one sided
- If we test $H_0 : \beta_j = 0$ for each j, and find that NONE of the p tests are significant, this means we fail to reject the null, and we remove at least one variable and re-test
  - If all are significant, reject the null hypothesis
- So for testing $H_0 : \beta_i = 0$ using a t-test
  - $\frac{\hat{\beta}_i - 0}{se(\hat{\beta}_i)} \sim t_{n-p-1}$
  - And $se(\hat{\beta}_i)$ is the square root of the ith diagonal element of $Var(\hat{\boldsymbol{\beta}})$
- Useful to look at a scatterplot matrix
  - Need to make sure you're looking at the plots where the explanatory variable is on the x-axis.
  - Can also look at the relationship between the explantory variables themselves
  - They do not provide enough information to understand the joint relationship between response and the predictors
    - Need to take into account the relationship between the predictors
- Focus on univariate summaries (one explanatory at a time), beware multivariate summaries

Multivariate Linear Regression in R

- *variable*.lm = lm(*response* ~ *x1* + *x2* + ... + *xp*, data = *df*)
  - If your using all other columns, can simplify to:
  - *variable*.lm = lm(*response* ~ ., data = *df*)
- Need to use Adjusted R-Squared for R^2
- For **X** matrix (separated from **Y** matrix)
  - $(\boldsymbol{X'X})$ is t(**X**) %*% **X**
  - $(\boldsymbol{X'X})^{-1}$ is solve(t(**X**) %*% **X**)
  - $(\boldsymbol{X'Y})$ is t(**X**) %*% **Y**
  - $\hat{\beta}$ is solve(t(**X**) %*% **X**) %*% t(**X**) %*% **Y**
  - $Var(\hat{\boldsymbol{\beta}})$ & $Cov(\hat{\boldsymbol{\beta}})$ is print(solve(t(**X**) %*% **X**) %*% **X**, digits = *n*)

$$\begin{array}{cccc} & & x1 & x2 \\ & \boxed{Var} & \text{Cov} & \text{Cov} \\ x1 & \text{Cov} & \boxed{Var} & \text{Cov} \\ x2 & \text{Cov} & \text{Cov} & \boxed{Var} \end{array}$$

  - $\sigma \sim s$ is in the summary table, or you can call it directly with summary(*variable*.lm)$sigma
  - Then use that for Var with s * sqrt(diag(solve(t(**X**) %*% **X**) %*% **X**))