HW 11 due on wed
HW 12 due next wed
OH today 11:30-12:30
wed 2:30-4

16: CORRELATION AND MATRIX APPROACH

Larsen & Marx 11.4; Rice 14.3; 14.4

Prof Amanda Luby

1 Covariance and Correlation

When we started linear regression, we began with the simplest scenario from a statistical standpoint – the case where each (x_i, y_i) are just constants with no probabilistic structure. When we moved into inference for this setting, we treated x_i as constant and Y_i as a random variable. We'll now move into the next layer of complexity: assuming both X_i and Y_i are random variables.

Covariance

Let X and Y be two random variables. The *covariance* of X and Y is given by:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

Let X and Y be two random variables with finite variances. Then,

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

$X_1, X_2, \dots, X_n \overset{iid}{\sim} F_X$ "independent"

If X and Y are independent:
$f_{X,Y} = f_X f_Y$
$E(X \cdot Y) = E(X)E(Y)$
$Cov(X,Y) = 0$

all of these are if and only if statements → can show independence w/ any property

The covariance of two random variables gives us a sense of how/what direction they are "related", but it also depends on the scale of the mean/variance for each RV. The *correlation coefficient* gives us a similar measure that is comparable across all RV's:

Correlation coefficient

Let X and Y be two random variables. The correlation coefficient of X and Y is given by:

"rho"

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \text{Cov}(X^*, Y^*)$$

ρ = population parameter

where $X^* = \frac{X - \mu_X}{\sigma_X}$ $\quad Y^* = \frac{Y - \mu_Y}{\sigma_Y}$

Note: $|\rho(X, Y)| \leq 1$:

$$0 \le Var(X^* \pm Y^*) = 1 + 1 \pm 2Cov(X^*, Y^*)$$
$$= 2 \pm 2 \cdot \rho(X,Y)$$
$$= 2[1 \pm \rho(X,Y)]$$
$$0 \le 1 \pm \rho(X,Y)$$
$$\pm \rho(X,Y) \le 1$$

Example: Suppose the correlation coefficient between X and Y is unknown, but we have observed n measurements $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. How could we use this data to estimate ρ?

$$\rho(X,Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$$

function of theoretical moments of $X + Y$

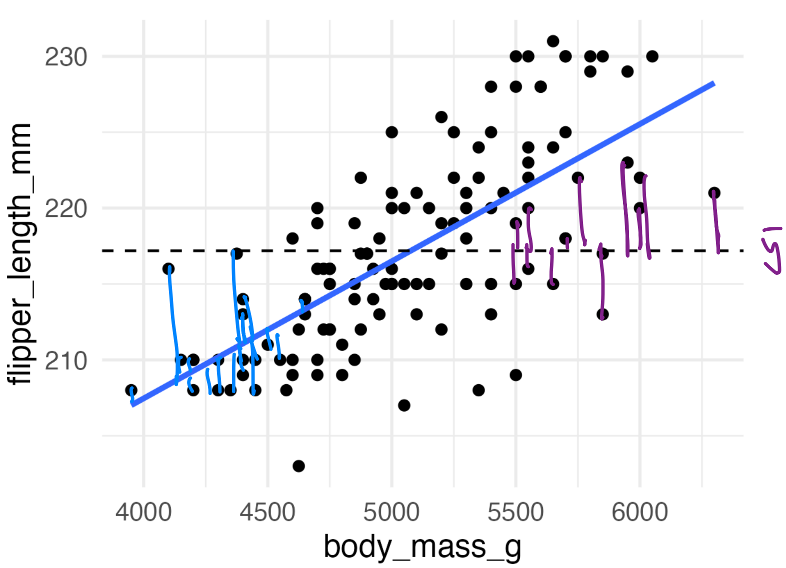
idea: plug in the sample moments

$$r = \frac{\frac{1}{n}\sum X_i Y_i - \frac{1}{n}\sum X_i \frac{1}{n}\sum Y_i}{\sqrt{\frac{1}{n}\sum(X_i - \bar{X})^2}\sqrt{\frac{1}{n}\sum(Y_i - \bar{Y})^2}} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{n\sum X_i^2 - (\sum X_i)^2}\sqrt{n\sum Y_i^2 - (\sum Y_i)^2}}$$

* nice relationship between r + $\hat{\beta}_1$ (HW, yay!)

If we square the (estimated) correlation coefficient, we can simplify to: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

$$R^2 = r^2 = \frac{\sum(y_i - \bar{y})^2 - \sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$



$\sum(y_i - \bar{y})^2$: Total variability in Y_i's (SS_{total})

$\sum(y_i - \hat{y}_i)^2$: Total variability "left over" after fitting regression model (SS_{resid})

Interpretation of R^2:

$$R^2 = \frac{\text{variability in } Y_i\text{'s that are explained by regression}}{\text{total variability in the } Y\text{'s}}$$

proportion of total variability in the Y_i's that is explained by the linear regression on X.

$r = .6 \rightarrow R^2 = .36 \rightarrow$ 36% of the variability in Y is explained by the regression on X (and therefore 64% is due to other factors) correlation

```r
cor(gentoo$body_mass_g, gentoo$flipper_length_mm, use = "complete.obs")
```

Deals w/ missing data

r =
```
[1] 0.7026665
```

$y \sim x$

```r
gentoo_lm = lm(flipper_length_mm ~ body_mass_g, data = gentoo)
summary(gentoo_lm)
```

```
Call:
lm(formula = flipper_length_mm ~ body_mass_g, data = gentoo)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.0194  -2.7401   0.1781   2.9859   8.9806 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.713e+02  4.244e+00   40.36   <2e-16 ***
body_mass_g 9.039e-03  8.321e-04   10.86   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.633 on 121 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.4937,	Adjusted R-squared:  0.4896 
F-statistic:   118 on 1 and 121 DF,  p-value: < 2.2e-16
```

Not important until we have >1 x variable

12/6
- HW 11 due today
- OH 2:30-3:45
- Final project comments up

2 Matrix Approach to Least Squares

2.1 Deriving the least squares solutions for 1 variable case

Define:

$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ $n \times 2$ "design matrix"

$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$ $n \times 1$

$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ 2×1

$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$

$\hat{\mathbf{Y}} = \mathbf{X}\beta$ $= \begin{bmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_1 x_2 \\ \vdots \\ \beta_0 + \beta_1 x_n \end{bmatrix}$

$n \times 2 \cdot 2 \times 1$

$n \times 1$

$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & x_4 & \dots & x_n \end{bmatrix}$

Note: $\|\vec{u}\|^2 = \sum u_i^2$

The least squares problem is to find $\hat{\beta}$ to minimize $L = \sum(y_i - (\beta_0 + \beta_1 x_i))^2$.

$$= \|Y - X\beta\|^2$$

$$= \|Y - \hat{Y}\|^2$$

$$Y - \hat{Y} = \begin{bmatrix} Y_1 - \hat{Y}_1 \\ Y_2 - \hat{Y}_2 \\ \vdots \\ Y_n - \hat{Y}_n \end{bmatrix}$$

In Notes14, we should that the least squares estimates satisfy:

$$\left.\begin{aligned} \frac{\partial L}{\partial \beta_0} &= \sum(y_i - (\beta_0 + \beta_1 x_i)) = 0 \\ \frac{\partial L}{\partial \beta_1} &= \sum(y_i - (\beta_0 + \beta_1 x_i))x_i = 0 \end{aligned}\right\} \quad \begin{aligned} \sum y_i &= \sum(\beta_0 + \beta_1 x_i) \\ &= n\beta_0 + \beta_1 \sum x_i \\ \sum x_i y_i &= \beta_0 \sum x_i + \beta_1 \sum x_i^2 \end{aligned}$$

In matrix form, these equations are equivalent to:

equivalent

$$X^T X\hat{\beta} = X^T Y$$

$$\begin{bmatrix} n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i \\ \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} y_1 + y_2 + \dots + y_n \\ x_1 y_1 + x_2 y_2 + \dots + x_n y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 + 1 + \dots + 1 & x_1 + x_2 + \dots + x_n \\ x_1 + x_2 + \dots + x_n & x_1^2 + x_2^2 + \dots + x_n^2 \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$X^T X\hat{\beta} = \begin{bmatrix} n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i \\ \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

$$X^T X\hat{\beta} = X^T Y$$

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

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

$$X^TX = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix}$$

Aside: $x_1, x_2, \ldots, x_{p-1}$

$X = n \times p$, $Y = n \times 1$, $\hat{\beta} = p \times 1$

(1) p partial derivatives + find joint solution

(2) Find $p \times 1$ matrix $(X^TX)^{-1}X^Ty$

Which means that the least squares solution is (assuming (X^TX) invertible)

$$\hat{\beta} = (X^TX)^{-1}X^TY$$

$$(X^TX)^{-1} = \frac{1}{n\Sigma x_i^2 - (\Sigma x_i)^2} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix}$$

$$(X^TX)^{-1}X^Ty = \frac{1}{n\Sigma x_i^2 - (\Sigma x_i)^2} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix} \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}$$

$$\hat{\beta} = \frac{1}{n\Sigma x_i^2 - (\Sigma x_i)^2} \begin{bmatrix} \Sigma y_i \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i \\ -\Sigma x_i \Sigma y_i + n\Sigma x_i y_i \end{bmatrix}$$

$$\hat{\beta}_0 = \frac{\Sigma y_i \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} \leftarrow \bar{Y} - \hat{\beta}_1 \bar{X} \ \ldots$$ algebra :)

$$\hat{\beta}_1 = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} \leftarrow$$ matches form of $\hat{\beta}_1$ from Notes 14

2.2 Mean and Covariance of Vector-Valued RV's

Let **Y** be a random vector where $\underbrace{E(Y_i) = \mu_i}_{\text{Not identical}}$ and $\underbrace{Cov(Y_i, Y_j) = \sigma_{ij}}_{\text{if } \sigma_{ij} \neq 0 \text{, not independent}}$

$Y_1, Y_2, \ldots, Y_n \overset{\text{independent identically distributed}}{\sim} N(\mu, \sigma^2)$

$\downarrow$ extend to multivariate case

$\vec{Y} \sim MVN(\vec{\mu}, \Sigma_Y)$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$E(\vec{Y}) = \vec{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \qquad \Sigma_Y = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

$\uparrow$ each (i,j) element is covariance

Linear functions of random variables

Let $\mathbf{Z} = \mathbf{c} + \mathbf{A}\mathbf{Y}$. Then
$E(\mathbf{Z}) = \mathbf{c} + \mathbf{A}E(\mathbf{Y})$ and $\Sigma_Z = \mathbf{A}\Sigma_Y\mathbf{A}^T$

c = vector of constants $(n \times 1)$
A = matrix of constants $(n \times n)$

$\hookrightarrow$ kind of equivalent to $V(a + bX) = b^2 V(X)$

2.3 Mean and Covariance of Least Squares Estimates

Let $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, where: $\epsilon_i \sim N(0, \sigma^2)$

$Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$

$E(\epsilon_i) = 0$

$V(\epsilon_i) = \sigma^2$

$Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$

$$\Sigma_\epsilon = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \cdot I$$

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Matrix form: $\epsilon_i \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \sigma^2 I\right)$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \sim MVN(X\beta, \sigma^2 I)$$

Mean and covariance of LS estimates (Matrix Form)

$E(\hat{\beta}) = \beta$

$\Sigma_{\hat{\beta}} = \sigma^2 (X^T X)^{-1}$

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T y \\ &= (X^T X)^{-1} X^T (X\beta + \epsilon) \\ &= (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T \epsilon \\ &= \beta + (X^T X)^{-1} X^T \epsilon \end{aligned}$$

$$\begin{aligned} E(\hat{\beta}) &= \beta + (X^T X)^{-1} X^T E(\epsilon) \\ &= \beta \end{aligned}$$

$\hat{\beta}_0$ and $\hat{\beta}_1$ (and $\hat{\beta}_2, \hat{\beta}_3, \ldots, \hat{\beta}_p$) are all unbiased for $\beta_0, \beta_1, \ldots, \beta_p$

true even if ϵ_i's are dependent / not constant variance!

(even if we have weird residual plots, $\hat{\beta}$ still unbiased as long as $E(\epsilon) = \vec{0}$)