16: MULTIPLE REGRESSION

Rice 14.4

Prof Amanda Luby

1 The Hat Matrix

$$\epsilon = \mathbf{Y} - \hat{\mathbf{Y}}$$

Note: The "hat matrix" has some nice properties: $H = H^T = H^2$ and $(I - H) = (I - H)^T = (I - H)^2$.

2 Estimation of σ^2

In Notes 15, two of the properties that we worked with were:

(1) $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$
(2) $S^2 = \frac{n}{n-2}\hat{\sigma}^2$

In matrix notation, we can write:

$$\sum (Y_i - \hat{Y}_i)^2 = ||Y - HY||^2$$

Then, using some nice properties for finding means of matrices (see Rice 14.4), we can show that $E(||Y - \hat{Y}||^2) = (n-p)\sigma^2$. This leads to the unbiased estimate for σ^2 for the multiple regression case:

Errors vs Residuals:

Covariance matrix of the residuals:

Cross-covariance matrix

Let X be a random vector of length n with covariance matrix Σ_X. If $Y = AX$ and $Z = BX$, where $A = p \times n$ and $B = m \times n$, then the cross-covariance matrix of Y and Z is given by:

If the errors have covariance matrix $\sigma^2 I$, the residuals are uncorrelated with the predicted values

Proof:

3 CI's for β

Sampling distribution for $\hat{\beta}$

4 CI's and PI's for predictions

Let $x^T = (1, x_1, ..., x_p)$ be a vector of predictors for a new observation Y.

5 Multiple R^2

In the simple regression case, recall that

$R^2 =$

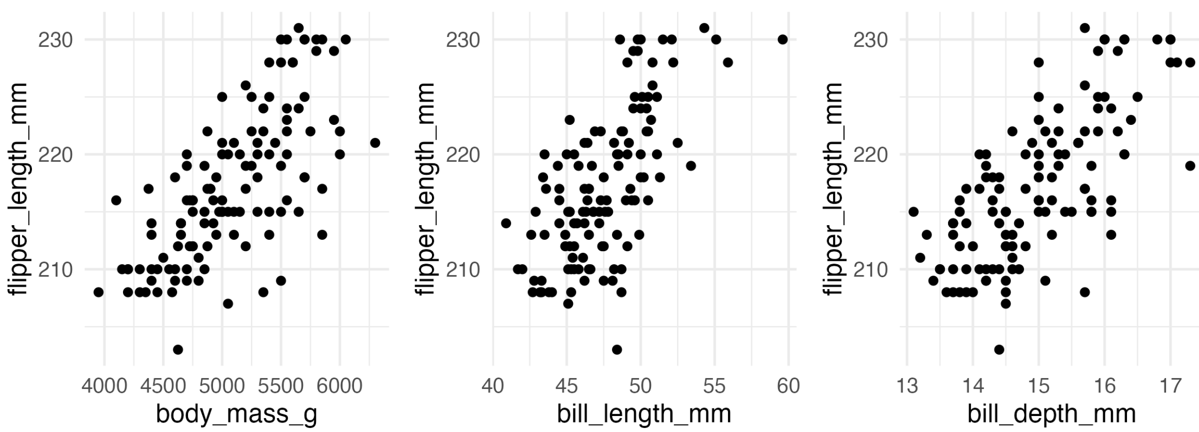
In simple linear regression, $R^2 = r^2$, where r is the sample correlation between X and Y. In the multiple regression case, we define $R = \text{Cor}(\hat{y}, y)$.

In multiple regression, whenever we add another predictor variable, R^2 *never gets worse*. The *Adjusted* R^2 is more often used in practice:

Adjusted $R^2 =$

as the number of predictors increase, what happens to the adjusted R^2?

6 Interpretation of β_i in Multiple Regression




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

Coefficients:
(Intercept)  body_mass_g
  1.713e+02    9.039e-03


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

Coefficients:
   (Intercept)  bill_length_mm
       151.096           1.391
```

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

Coefficients:
  (Intercept)  bill_depth_mm
       147.22           4.67


Call:
lm(formula = flipper_length_mm ~ body_mass_g + bill_length_mm +
    bill_depth_mm, data = gentoo)

Coefficients:
   (Intercept)     body_mass_g  bill_length_mm   bill_depth_mm
     139.99254         0.00382         0.52150         2.20463


Call:
lm(formula = flipper_length_mm ~ body_mass_g + bill_length_mm +
    bill_depth_mm, data = gentoo)

Residuals:
    Min      1Q  Median      3Q     Max
-12.440  -2.492   0.023   2.829   8.322

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.400e+02  6.527e+00  21.448  < 2e-16 ***
body_mass_g    3.820e-03  1.153e-03   3.314 0.001217 **
bill_length_mm 5.215e-01  1.711e-01   3.047 0.002846 **
bill_depth_mm  2.205e+00  5.748e-01   3.836 0.000202 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.11 on 119 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.6082,    Adjusted R-squared:  0.5983
F-statistic: 61.58 on 3 and 119 DF,  p-value: < 2.2e-16
```