For modeling freshman GPA to depend linearly on high school GPA, a standard linear regression model is:

\[ Y_i = \beta_0 + \beta_1 x_i + e_i, \ i = 1, 2, \ldots, n. \]

Suppose that different intercepts were to be allowed for females and males, and write the model as

\[ Y_i = I_F(i)\beta_F + I_M(i)\beta_M + \beta_1 x_i + e_i, \ i = 1, 2, \ldots, n. \]

where \( I_F(i) \) and \( I_M(i) \) are indicator variables taking on values of 0 and 1 according to whether the gender of the \( i \)th person is female or male.

The design matrix for such a model will be

\[
X = \begin{bmatrix}
I_F(1) & I_M(1) & x_1 \\
I_F(2) & I_M(2) & x_2 \\
\vdots & \vdots & \vdots \\
I_F(n) & I_M(n) & x_n
\end{bmatrix}
\]

Note that \( X^T X = \begin{bmatrix} n_F & 0 & \sum_{Female} i x_i \\
0 & n_M & \sum_{Male} i x_i \\
\sum_{Female} i x_i & \sum_{Male} i x_i & \sum_i x_i^2 \end{bmatrix} \)

where \( n_F \) and \( n_M \) are the number of females and males, respectively.

The regression model is setup as

\[
Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = X \begin{bmatrix} \beta_F \\ \beta_M \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}
\]
3. 14.9.6. The 4 weighings correspond to 4 outcomes of the dependent variable \( y \)

\[
Y = \begin{bmatrix} 3 \\ 3 \\ 1 \\ 7 \end{bmatrix}
\]

For the regression parameter vector

\[
\beta = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}
\]

the design matrix is

\[
X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}
\]

The regression model is

\[
Y = X\beta + e
\]

(b). The least squares estimates of \( w_1 \) and \( w_2 \) are given by

\[
\hat{\beta} = \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \end{bmatrix} = (X^T X)^{-1} X^T Y
\]

Note that

\[
(X^T X) = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}
\]

so

\[
(X^T X)^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/3 \end{bmatrix}
\]

and

\[
(X^T Y) = \begin{bmatrix} 11 \\ 9 \end{bmatrix},
\]

so

\[
\hat{\beta} = \begin{bmatrix} 1/3 \\ 0 \\ 0 \\ 1/3 \end{bmatrix} \times \begin{bmatrix} 11 \\ 9 \end{bmatrix} = \begin{bmatrix} 11/3 \\ 3 \end{bmatrix}
\]

(c). The estimate of \( \sigma^2 \) is given by the sum of squared residuals divided by \( n - 2 \), where 2 is the number of columns of \( X \) which equals the number of regression parameters estimated.

The vector of least squares residuals is:

\[
\hat{e} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ y_3 - \hat{y}_3 \\ y_4 - \hat{y}_4 \end{bmatrix} = \begin{bmatrix} 3 - 11/3 \\ 3 - 3 \\ 1 - 2/3 \\ 7 - 20/3 \end{bmatrix} = \begin{bmatrix} -2/3 \\ 0 \\ 1/3 \\ 1/3 \end{bmatrix}
\]

From this we can compute
\[
\hat{\sigma}^2 = \frac{(-2/3)^2 + 0^2 + (1/3)^2 + (1/3)^2}{4} = \frac{6/9}{2} = 1/3
\]

(d). The estimated standard errors of the least squares estimates of part (b) are the square roots of the diagonal entries of
\[
\hat{\text{Cov}}(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1}
\]
which are both equal to \(\sqrt{\sigma^2 \times 1/3} = 1/3\).

(e). The estimate of \(w_1 - w_2\) is given by \(\hat{\beta}_1 - \hat{\beta}_2 = 11/3 - 3 = 2/3\)
The standard error of the estimate is the estimate of its standard deviation which is the square root of the estimate of its variance.

Now
\[
\text{Var}(\hat{\beta}_1 - \hat{\beta}_2) = \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2) - 2\text{Cov}(\hat{\beta}_1, \hat{\beta}_2).
\]
This is the sum of the diagonal elements of \(\hat{\text{Cov}}(\hat{\beta})\) minus the sum of off-diagonal entries (which are 0).
So
\[
\text{Var}(\hat{\beta}_1 - \hat{\beta}_2) = 1/9 + 1/9 - 2 \times 0 = 2/9.
\]
and the standard error is \(\sqrt{2/9}\)

(f). To test \(H_0 : w_1 = w_2\), we can compute the t-statistic
\[
t = \frac{\hat{w}_1 - \hat{w}_2}{\text{stErr}(\hat{w}_1 - \hat{w}_2)} = \frac{2/3}{\sqrt{2/9}} = \sqrt{2}
\]
Under \(H_0\) this has a t distribution with \(n - 2 = 2\) degrees of freedom.
Using R we can compute the P-value as
\[
P\text{-Value} = 2 \times (1 - \text{pt}(\sqrt{2}, \text{df}=2)) = 0.2928932
\]
For normal significance levels (.05 or .01) the null hypothesis is not rejected because the P-Value is higher.


Suppose that
\[
Y_i = \beta_0 + \beta_1 x_i + e_i, \ i = 1, \ldots, n
\]
where the \(e_i\) are i.i.d. \(N(0, \sigma^2)\). Find the mle’s of \(\beta_0\) and \(\beta_1\) and verify that they are the least squares estimates. This follows immediately from the lecture notes: Regression Analysis II in the section on maximum likelihood. The likelihood function is a monotonic function of the least-squares criterion \(Q(\beta) = \sum_1^n (Y_i - \hat{Y}_i)^2\). Therefore the least-squares estimates of \(\beta\) and the mle’s are identical.
