Homework 5

1. The file `bodytemp.csv` contains normal body temperature readings (in degrees Fahrenheit) and heart rates (beats per minute) of 65 males (coded by 1) and 65 females (coded by 2).

   (a) For both males and females make scatter plots of heart rate versus body temperature. Comment on the relationship or lack thereof.

   (b) Does the relationship for males appear to be the same as that for females? Examine this question graphically, by making a scatter plot showing both females and males and identifying females and males by different plotting symbols.

   (c) Construct an F test to test whether the two regression lines are identical.

   The F-statistic is 0.38 on 2,126 df. The p-value equals 0.68.

   (d) Devise and apply a permuation test of this hypothesis.

   A permutation test reaches the same conclusion: the variable gender is permuted many times, and the value of the F statistic is recorded each time. The p-value is the proportion of the permutation statistics that are larger than the observed F statistic.

   (e) Test whether the slopes are the same but the intercepts are not equal.

   $F = .769$ with 1,126 df. The p-value = 0.38, so there is little evidence that the intercepts are unequal

2. Consider simple linear regression, $Y_i = \beta_0 + \beta_1 x_i + e_i$ where $e \sim MVN(0, \sigma^2 I)$. Assume for simplicity that $n = 25$ and that $\sum_{i=1}^{n} x_i = 0$.

   (a) What is a 95% confidence interval for β_0?

   From direct calculation

$$\begin{aligned}
\hat{\beta}_0 &= \bar{Y} \\
s_{\hat{\beta}_0} &= \frac{s}{\sqrt{n}} \\
\hat{\beta}_1 &= \frac{\sum x_i Y_i}{\sum x_i^2} \\
s_{\hat{\beta}_1} &= \frac{s}{\sqrt{\sum x_i^2}}
\end{aligned}$$

where $s = \sqrt{\frac{RSS}{n-2}}$ A 95% CI for β_0 is thus

$\hat{\beta}_0 \pm t_{23}(.025)s_{\hat{\beta}_0} = \hat{\beta}_0 \pm 2.07 s_{\hat{\beta}_0}$

(b) What is a 95% confidence interval for β_1?

$\hat{\beta}_1 \pm 2.07 s_{\hat{\beta}_1}$

(c) What is a 95% confidence region for (β_0, β_1)?

Confidence region is $\{\beta \ : \ \parallel X\beta - X\hat{\beta} \parallel^2 \leq 2s^2 F_{2,23}(.05)\}$

$$\parallel X\beta - X\hat{\beta} \parallel^2 = n(\beta_0 - \hat{\beta}_0)^2 + (\beta_1 - \hat{\beta}_1)^2 \sum x_i^2$$

and $2F_{2,23}(.05) = 2 \times 3.42 = 6.84$.

The confidence region is thus an ellipse with axes parallel to the β_0 and β_1 axes. The ellipse cuts the β_0 axis at $n(\beta_0 - \hat{\beta}_0)^2 = 6.84s^2$, that is at $\hat{\beta}_0 \pm s\sqrt{6.84/n}$ and cuts the β_1 axis at $\hat{\beta}_1 \pm s\sqrt{6.84 \sum x_i^2}$.

(d) Compare graphically the regions found in (a) and (b) to the region found in (c).

Since $\sqrt{6.84} = 2.61 > 2.07$, the confidence intervals in (a) and (b) are contained in the intervals of (c).

(e) How can the hypothesis that $\beta_0 = 0$ and $\beta_1 = 1$ be tested?

The test rejects if the point (0,1) is not in the confidence ellipse.

3. Verify explictly that the least squares estimates for the α_i in the balanced two way layout are the same whether or not the β_j are set equal to zero.

Minimize

$$\sum_{jk} (Y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2$$

and also the same form without the β_j and use the constraints to verify the claim.