Mickey Warner

1a. The six groups of students are made of the possible combinations of major and background. Major=1 and BG=1 are the baselines so the estimated effects are added to the intercept for each student group. The scores for each group are given in the following table:

Major 1, BG 1	$\hat{\mu}$	0.8893
Major 1, BG 2	$\hat{\mu} + \hat{\eta}_{BG2}$	2.1457
Major 2, BG 1	$\hat{\mu} + \hat{\alpha}_{Major2}$	2.8753
Major 2, BG 2	$\hat{\mu} + \hat{\alpha}_{Major2} + \hat{\eta}_{BG2} + \hat{\gamma}_{Major2,BG2}$	2.4686
Major 3, BG 1	$\hat{\mu} + \hat{\alpha}_{Major3}$	2.0782
Major 3, BG 2	$\hat{\mu} + \hat{\alpha}_{Major3} + \hat{\eta}_{BG2} + \hat{\gamma}_{Major3,BG2}$	1.7216

Thus, the group with the lowest score is Major 1 and BG 1 (Economics and Rural) at 0.8893, and the group with the highest score is Major 2 and BG 1 (Anthropology and Rural) at 2.8753.

1b. i. The null hypothesis is that the reduced model (the intercept-only model) is sufficient. That is, $H_0 : E(y_{ij}) = \mu$ for some μ. The alternative is that the correct model has main effects and interactions in addition to an intercept, $H_1 : E(y_{ij}) = \mu + \alpha_i + \eta_j + \gamma_{ij}$.

ii. The F-statistic yields a p-value of 0.04945. At the $\alpha = 0.05$ level, we would reject the null hypothesis. This is to say that an intercept-only model does not adequately explain the variation in mathematics ineptitude scores. The model that gives each major and high school background its own effect, together with an interaction between the two, is preferable. However, this is not to say this full model is ideal.

1c. The two p-values for BG do not contradict each other because they are testing two different hypotheses. In the `lm` output, the test is whether $\eta_{BG2} = 0$ in the full model and this is accomplished using a t-test. In the `anova` output, the test is whether the high school background provides enough of an improvement to be included in the model *after* including major in the model.

Even though the `anova` output gives a large p-value for high school background, there appears to be some evidence that there is an interaction effect

1

(at the 0.10 level). If there is an interaction (between major and background), then we would keep background in the model, regardless of its main effect contribution. I would be fine with keeping the interaction effect in the model, and this is to say that student background is relevant in predicting the score. Others may think the p-values are too high (all are above 0.05) for the effects to be considered real.

2a. We estimate the parameters by $\hat{\beta} = (X^\top X)^{-1} X^\top y$ and $\hat{\sigma}^2 = (y - X\hat{\beta})^\top (y - X\hat{\beta})/(n - r)$ where $n = 8$ and $r = rank(X) = 2$. We obtain $\hat{\beta}_1 = 2.64$, $\hat{\beta}_2 = 3.73$, $\hat{\sigma}^2 = 4.70$.

2b. For estimable $\lambda^\top \beta$, a 95% confidence interval is given by

$$\lambda^\top \hat{\beta} \pm t_{0.975}(n - r)\sqrt{\hat{\sigma}^2 \lambda^\top (X^\top X)^{-1} \lambda}$$

β_1 and $\beta_1 + \beta_2$ are both estimable and the corresponding λ are $\lambda_1 = (1, 0)^\top$ and $\lambda_2 = (1, 1)^\top$. The two 95% confidence intervals are $(1.12, 4.17)$ and $(5.93, 6.83)$ for β_1 and $\beta_1 + \beta_2$, respectively.

2c. To test $H_0 : \beta_2 = 3$, we use $\lambda = (0, 1)^\top$ and compute the test statistic

$$t = \frac{\lambda^\top \hat{\beta} - 3}{\sqrt{\hat{\sigma}^2 \lambda^\top (X^\top X)^{-1} \lambda}} = 1.64$$

This statistic yields a p-value of 0.15 based on the two tails of a t distribution with $n - r = 6$ degrees of freedom. At the $\alpha = 0.01$ level, we fail to reject the hypothesis that $\beta_2 = 3$.

2d. Testing $H_0 : \beta_1 - \beta_2 = 0$ is done similarly as in part **2c**, but now $\lambda = (1, -1)^\top$. We compute the following test statistic

$$t = \frac{\lambda^\top \hat{\beta} - 0}{\sqrt{\hat{\sigma}^2 \lambda^\top (X^\top X)^{-1} \lambda}} = -1.01$$

which gives a p-value of 0.34. This is again calculated as the probability of exceeding $|t|$ from a t distribution with $n - r$ degrees of freedom.

3. We use $\mathbf{1}$ to denote the $n \times 1$ vector of ones, $\mathbf{y} = (y_1, \ldots, y_n)^\top$, $\hat{\mathbf{y}} = (\hat{y}_1, \ldots, \hat{y}_2)^\top$, and $\bar{y} = 1/n \sum_{i=1}^n y_i$. Note that the mean of the predicted values is equal to the mean of the data, i.e. $\bar{\hat{y}} = \bar{y}$. The correlation between $\mathbf{y}$ and $\hat{\mathbf{y}}$ is given by

$$r(\mathbf{y}, \hat{\mathbf{y}}) = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{y})^2}}$$

$$= \frac{(\mathbf{y} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})}{\sqrt{[(\mathbf{y} - \bar{y}\mathbf{1})^\top (\mathbf{y} - \bar{y}\mathbf{1})] [(\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})]}}$$

$$= \frac{(\mathbf{y} - \hat{\mathbf{y}} + \hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})}{\sqrt{[(\mathbf{y} - \bar{y}\mathbf{1})^\top (\mathbf{y} - \bar{y}\mathbf{1})] [(\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})]}}$$

$$= \frac{(\mathbf{y} - \hat{\mathbf{y}})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1}) + (\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})}{\sqrt{[(\mathbf{y} - \bar{y}\mathbf{1})^\top (\mathbf{y} - \bar{y}\mathbf{1})] [(\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})]}}$$

$$= \frac{\mathbf{y}^\top \hat{\mathbf{y}} - \bar{y}\mathbf{y}^\top \mathbf{1} - \hat{\mathbf{y}}^\top \hat{\mathbf{y}} + \bar{y}\hat{\mathbf{y}}^\top \mathbf{1} + (\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})}{\sqrt{[(\mathbf{y} - \bar{y}\mathbf{1})^\top (\mathbf{y} - \bar{y}\mathbf{1})] [(\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})]}}$$

$$= \frac{\mathbf{y}^\top P_x \mathbf{y} - n\bar{y}^2 - \hat{\mathbf{y}}^\top \hat{\mathbf{y}} + n\bar{y}^2 + (\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})}{\sqrt{[(\mathbf{y} - \bar{y}\mathbf{1})^\top (\mathbf{y} - \bar{y}\mathbf{1})] [(\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})]}}$$

$$= \frac{\mathbf{y}^\top P_x P_x \mathbf{y} - \hat{\mathbf{y}}^\top \hat{\mathbf{y}} + (\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})}{\sqrt{[(\mathbf{y} - \bar{y}\mathbf{1})^\top (\mathbf{y} - \bar{y}\mathbf{1})] [(\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})]}}$$

$$= \frac{\hat{\mathbf{y}}^\top \hat{\mathbf{y}} - \hat{\mathbf{y}}^\top \hat{\mathbf{y}} + (\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})}{\sqrt{[(\mathbf{y} - \bar{y}\mathbf{1})^\top (\mathbf{y} - \bar{y}\mathbf{1})] [(\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})]}}$$

$$= \frac{(\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})}{\sqrt{[(\mathbf{y} - \bar{y}\mathbf{1})^\top (\mathbf{y} - \bar{y}\mathbf{1})] [(\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})]}}$$

$$= \sqrt{\frac{(\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})}{(\mathbf{y} - \bar{y}\mathbf{1})^\top (\mathbf{y} - \bar{y}\mathbf{1})}}$$

Squaring both sides and expanding the terms yield

$$(r(\mathbf{y}, \hat{\mathbf{y}}))^2 = \frac{(\hat{\mathbf{y}} - \bar{y}\mathbf{1})^\top (\hat{\mathbf{y}} - \bar{y}\mathbf{1})}{(\mathbf{y} - \bar{y}\mathbf{1})^\top (\mathbf{y} - \bar{y}\mathbf{1})}$$

$$= \frac{\hat{\mathbf{y}}^\top \hat{\mathbf{y}} - \bar{y}\hat{\mathbf{y}}^\top \mathbf{1} - \bar{y}\mathbf{1}^\top \hat{\mathbf{y}} + \bar{y}^2 \mathbf{1}^\top \mathbf{1}}{\mathbf{y}^\top \mathbf{y} - \bar{y}\mathbf{y}^\top \mathbf{1} - \bar{y}\mathbf{1}^\top \mathbf{y} + \bar{y}^2 \mathbf{1}^\top \mathbf{1}}$$

$$= \frac{\hat{\mathbf{y}}^\top \hat{\mathbf{y}} - n\bar{y}^2}{\mathbf{y}^\top \mathbf{y} - n\bar{y}^2}$$

$$= \frac{\mathbf{y}^\top P_x P_x \mathbf{y} - n(\frac{1}{n}\mathbf{y}^\top \mathbf{1})(\frac{1}{n}\mathbf{1}^\top \mathbf{y})}{\mathbf{y}^\top \mathbf{y} - n(\frac{1}{n}\mathbf{y}^\top \mathbf{1})(\frac{1}{n}\mathbf{1}^\top \mathbf{y})}$$

$$= \frac{\mathbf{y}^\top P_x P_x \mathbf{y} - \mathbf{y}^\top (\frac{1}{n}\mathbf{1}\mathbf{1}^\top)\mathbf{y}}{\mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top (\frac{1}{n}\mathbf{1}\mathbf{1}^\top)\mathbf{y}}$$

$$= \frac{\mathbf{y}^\top P_x \mathbf{y} - \mathbf{y}^\top P_1 \mathbf{y}}{\mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top P_1 \mathbf{y}}$$

$$= \frac{\mathbf{y}^\top (P_x - P_1)\mathbf{y}}{\mathbf{y}^\top (I - P_1)\mathbf{y}}$$

$$= R^2$$

$$y_t = \text{logit}(p_t)$$