1. (30 points) Problem 1.

   (a) (10 points) Let X and Y both be ordinal. The polychoric correlation ρ is the correlation between two *latent* continuous Gaussian variables (Z_1, Z_2) underlying (X, Y). That is,
$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$
   For two randomly selected individuals with responses (X_1, Y_1) and (X_2, Y_2), their responses are concordant if $X_1 < X_2$ and $Y_1 < Y_2$ or $X_1 > X_2$ and $Y_1 > Y_2$. Let Π_c be the probability of this event and Π_d be the probability of discordance, defined similarly (but *not* $1 - \Pi_d$ – need to consider ties). Then $\gamma = (\Pi_c - \Pi_d)/(\Pi_c + \Pi_d)$ is a measure of concordance/discordance and lies between -1 and 1.

   (b) (10 points)
$$\frac{\left[\frac{P(Y=1|X=1,Z=1)}{P(Y=2|X=1,Z=1)}\right]}{\left[\frac{P(Y=1|X=2,Z=1)}{P(Y=2|X=2,Z=1)}\right]} = \frac{e^{\alpha+\beta+\tau}}{e^{\alpha+\tau}} = e^{\beta} \text{ and } \frac{\left[\frac{P(Y=1|X=1,Z=2)}{P(Y=2|X=1,Z=2)}\right]}{\left[\frac{P(Y=1|X=2,Z=2)}{P(Y=2|X=2,Z=2)}\right]} = \frac{e^{\alpha+\beta}}{e^{\alpha}} = e^{\beta},$$
   both independent of Z.

   (c) (5 points) $\hat{\theta} = \dfrac{n_{11}n_{22}}{n_{12}n_{21}}$.

   (d) (5 points). (1) the random component, here $Y_i \sim \text{Pois}(\mu_i)$, (2) the systematic component, here the linear predictor $\eta_i = \alpha + \beta x_i$, and (3) the link relating $\mu_i = E(Y_i)$ to η_i, here the log-link: $\log(\mu_i) = \eta_i = \alpha + \beta x_i$.

   (e) Not graded – too ambiguous.

2. (25 points) Problem 2.

   (a) (5 points) Let X be college of enrollment and Y be political affiliation. We reject $H_0 : X \perp Y$ at the 5% level based on either the Pearson or LRT p-values, 0.0129 and 0.0118. The expected cell counts are all well over 5, so we can be confident in the Pearson p-value. see the top of page 80 in your text.

   (b) (5 points) $r_5 = -3.24$ corresponds to those Engineering students who are Democrats. The direction of residuals switches from Engineering to Letters, and three of these six residuals are larger than 2 in magnitude. So it would seem that Engineering and Letters are particularly unlike each other in terms of political affiliation.

   (c) (5 points) $G_1^2 + G_2^2 = 10.8539 + 5.5361 = 16.39 = G^2$ and $df_1 + df_2 = 2 + 4 = 6 = df$ as advertised.

   (d) (10 points) The table where non-Engineering students are combined into one category shows strong evidence of dependence ($p \approx 0.005$), whereas the table with Engineering omitted shows no evidence of dependence ($p \approx 0.22$). So it would seem that college of enrollment and political affiliation are dependent largely due to Engineering students' differences compared to the other three schools. In particular, Engineering has fewer Democrats and more Republicans than we expect under independence.

3. (10 points) Problem 3. Let π_i be the probability of having one or more satellites for the colors $i = 1, 2, 3, 4$. Let $Y = 1$ if a crab has one or more satellites. The model is

$$P(Y = 1) = \beta_0 + \beta_1 I\{C = 1\} + \beta_2 I\{C = 2\} + \beta_3 I\{C = 3\},$$

and estimated to be

$$P(Y = 1) = 0.3182 + 0.4318 I\{C = 1\} + 0.4081 I\{C = 2\} + 0.2727 I\{C = 3\}.$$

Note that $\pi_4 = \beta_0$, $\pi_1 = \beta_0 + \beta_1$, $\pi_2 = \beta_0 + \beta_2$ and $\pi_3 = \beta_0 + \beta_3$.

   (a) (5 points)
   $$\frac{\hat{\pi}_1}{\hat{\pi}_4} = \frac{0.3182 + 0.4318}{0.3182} \approx 2.36.$$
   Light-medium female horseshoe crabs are estimated to be 2.36 times as likely to have one or more satellites relative to those with dark shells.

   (b) (5 points) We estimate $\pi_1 - \pi_4$ by $\hat{\pi}_1 - \hat{\pi}_4 = 0.4318$ with a 95% CI of $(0.1189, 0.7447)$. These are both in the `Analysis Of Parameter Estimates` table. Testing $H_0 : \pi_4 - \pi_1 = 0$ is equivalent to testing $H_0 : \beta_1 = 0$, with Wald p-value 0.0068. You could also note that the 95% CI does not include zero. Either way the hypothesis is rejected at the 5% significance level.

4. (35 points) Problem 4. The `par=ref` gives zero-one dummy variables. The fitted model is therefore

$$\begin{aligned} \text{logit } P(\text{high aspirations}) = & -2.0327 - 0.6002\ I\{\text{female}\} + 0.2004\ I\{\text{large urban}\} \\ & -0.0522\ I\{\text{rural}\} + 1.7720\ I\{\text{high IQ}\} \\ & +1.7092\ I\{\text{high ses}\} + 0.4112\ I\{\text{female}\} I\{\text{high ses}\} \\ & -0.0785\ I\{\text{large urban}\} I\{\text{high ses}\} - 0.5823\ I\{\text{rural}\} I\{\text{high ses}\}. \end{aligned}$$

   (a) (5 points) This is given under `Odds Ratio Estimates`. High IQ increases the odds of having high occupational aspirations by a factor of 5.9 with 95% CI $(5.1, 6.8)$.

   (b) (10 points) The odds of high occupational aspirations is lower for females; this difference is more pronounced among those with low socioeconomic status:
   $$\frac{\left[\frac{P(\text{high aspirations}|\text{female,high ses})}{P(\text{low aspirations}|\text{female,high ses})}\right]}{\left[\frac{P(\text{high aspirations}|\text{male,high ses})}{P(\text{low aspirations}|\text{male,high ses})}\right]} = e^{-0.6002+0.4112} \approx 0.83.$$
   $$\frac{\left[\frac{P(\text{high aspirations}|\text{female,low ses})}{P(\text{low aspirations}|\text{female,low ses})}\right]}{\left[\frac{P(\text{high aspirations}|\text{male,low ses})}{P(\text{low aspirations}|\text{male,low ses})}\right]} = e^{-0.6002} \approx 0.55.$$

(c) (10 points) For high socioeconomic status, those living in large cities have slightly greater odds of high occupational aspirations than those in smaller towns, who have much greater odds than rural dwellers.

$$\frac{\left[\frac{P(\text{high aspirations}|\text{large urban,high ses})}{P(\text{low aspirations}|\text{large urban,high ses})}\right]}{\left[\frac{P(\text{high aspirations}|\text{small urban,high ses})}{P(\text{low aspirations}|\text{small urban,high ses})}\right]} = e^{0.2004-0.0785} \approx 1.13.$$

$$\frac{\left[\frac{P(\text{high aspirations}|\text{rural,high ses})}{P(\text{low aspirations}|\text{rural,high ses})}\right]}{\left[\frac{P(\text{high aspirations}|\text{small urban,high ses})}{P(\text{low aspirations}|\text{small urban,high ses})}\right]} = e^{-0.0522-0.5823} \approx 0.53.$$

For low socioeconomic status, those living in large cities have slightly greater odds of high occupational aspirations than those in smaller towns (the difference is a bit larger than for high socioeconomic status), who have slightly greater odds than rural dwellers.

$$\frac{\left[\frac{P(\text{high aspirations}|\text{large urban,low ses})}{P(\text{low aspirations}|\text{large urban,low ses})}\right]}{\left[\frac{P(\text{high aspirations}|\text{small urban,low ses})}{P(\text{low aspirations}|\text{small urban,low ses})}\right]} = e^{0.2004} \approx 1.22.$$

$$\frac{\left[\frac{P(\text{high aspirations}|\text{rural,low ses})}{P(\text{low aspirations}|\text{rural,low ses})}\right]}{\left[\frac{P(\text{high aspirations}|\text{small urban,low ses})}{P(\text{low aspirations}|\text{small urban,low ses})}\right]} = e^{-0.0522} \approx 0.95.$$

(d) (5 points) Yes, there is good replication in the data set at each set of covariate values; we can trust the p-values. There is no evidence of gross lack-of-fit.

(e) (5 points) The plot looks fine. There is no obvious pattern and all standardized Pearson residuals are < 3 in magnitude.