STAT 659 — Solution to Exam 2 Spring 2015

1. (a) Do not reject $H_o : \beta_{S\times G} = 0$ since $G^2 = 0.7713$ with a $p-$value$= 0.3798$. There is insufficient evidence to include the interaction term in the model. Hence, there is insufficient evidence to indicate that the ORs differ for men and women.

   (b) Using the homogeneous association model, reject $H_0 : \beta_S = 0$ since the Wald $X^2 = 29.0$ with a $p-$value< 0.0001. There is strong evidence of partial association between smoking and depression, controlling for gender.

   (c) The log odds ratio for depression comparing females to males is
   $$\text{logit}(\hat{\pi}(F, yes)) - \text{logit}(\hat{\pi}(M, yes)).$$

      i. Homogeneous association model:
      $$\text{logit}(\hat{\pi}(F, yes)) - \text{logit}(\hat{\pi}(M, yes)) = 0.9368.$$
      Thus, $\widehat{OR} = e^{0.9368} = 2.552.$

      ii. Model with interaction:
      $$\text{logit}(\hat{\pi}(F, yes)) - \text{logit}(\hat{\pi}(M, yes)) = 0.6474 + 0.3648 = 1.0122.$$
      Thus, $\widehat{OR} = e^{1.0122} = 2.752.$

2. (a) Since the response categories are ordered, we can consider using a POM. The score statistic for the POM assumption is $X^2 = 1.6130$ with a $p-$value$= 0.4464$. Thus, we have no reason to doubt the adequacy of the POM.

   (b)
   $$P(Y = 1) = \frac{1}{1 + e^{-3.12+0+2.50} + e^{-3.43+0+2.32}} = 0.535.$$

   (c) $\widehat{\text{logit}}(P(Y \le 1)) = 2.475 - 0 - 2.159 = 0.316$, so $\hat{P}(Y = 1) = e^{0.316}/(1 + e^{0.316}) = 0.578$, which is similar to the value in (b).

   (d) Since the coefficents of task and ventilation are negative, choosing task=insulation and ventilation=ordinary will make $\widehat{logit}(P[Y \le j])$ smaller. Thus, there will be a smaller probability for small exposure. Thus, the combination of (insulation,ordinary) leads to a high probability for higher exposure (larger values of Y). A similar argument (changing the signs) will say that the combination of (tile,negative pressure) will lead to the lowest exposure.

3. (a) We fail to reject $H_0 : \beta_{\texttt{race}} = \beta_{\texttt{ses1}} = beta_{\texttt{ses2}} = 0$ since $G^2 = 354.36 - 353.63 = 0.73 < 7.81 = \chi^2_{3,0.05}$. Thus, there is insufficient evidence that these two variables are needed in the model.

   (b) Keeping the other variables constant, the difference in logits is

   $\text{logit}(\hat{\pi}(\texttt{crowding} = 1, \texttt{passive} = 1, ...)) - \text{logit}(\hat{\pi}(\texttt{crowding} = 0, \texttt{passive} = 0, ...)) = 0.2493 + 0.7131 = 0.9624.$

   Thus, the odds ratio equals $e^{0.9624} = 2.618$ indicating that the odds of having at least one lower respiratory infection when in a crowded setting exposed to passive smoke are 2.618 times the odds of having at least one lower respiratory infection when in a noncrowded setting not exposed to passive smoke.

   (c)
   - Estimated sensitivity: $\hat{P}(\hat{y} = 1|y = 1) = 62/114 = 0.544$
   - Estimated specificity: $\hat{P}(\hat{y} = 0|y = 0) = 104/170 = 0.612$
   - Estimated probability of correct classification: $(62 + 104)/284 = 0.585$

   (d) The model with lowest AIC_C is model 4 with `crowding,agegroup,risk`. To see if `passive` is needed as well, we do not reject $H_0 : \beta_{\texttt{passive}} = 0$ since $G^2 = 355.5 - 354.4 = 1.1 < 3.84 = \chi^2_{1,0.05}$. Thus, model 3 does not improve upon model 4. We compare the model 5 to see if `agegroup` is useful using $G^2 = 366.8 - 355.5 = 11.3 > 5.99 == \chi^2_{2,0.05}$. We need to include `agegroup` in the model and conclude that Model 4 is the most appropriate model.

   (e) The Hosmer-Lemeshow test has $X^2 = 11.7599$ with a $p-$value$= 0.1622$. Thus, there is no indication of lack of fit for the preliminary main effects model.