PubH 7430 – Fall 2007

EXAM SOLUTIONS

For some questions, the answers below give MORE details than were required to get full credit. Questions were worth 5 points each; only 10 of the 12 questions were to be answered.

APPLICATION

1. (a) From the summary statistics on p. 6, we see that women had higher average satisfaction scores than men (6.87 vs. 7.55).

   (b) There was a positive and approximately linear association of health status score with satisfaction that appears to differ by gender: the slope for men is steeper than the slope for women.

2. (a) Σ_i is 8 by 8, because there are 8 survey respondents per nursing home.

   (b) Unstructured is not a reasonable structure to consider because there is no logical ordering of respondents within home.

   (c) Compound symmmetry is a reasonable structure to consider because it assumes equicorrelation within home: the reponses for all persons within a home are equally correlated. This is appropriate because there is no logical ordering of respondents within home.

3. Compound symmetry with a group effect for nursing home location is the most appropriate structure:

A	Ind vs. CS (no group)	$T^* = 781.6 - 770.7 = 10.9$	$df = 2 - 1 = 1$	$p < 0.05$ (p=0.001)
B	Ind vs. CS (with group)	$T^* = 779.0 - 761.3 = 17.7$	$df = 4 - 2 - 2$	$p < 0.05$ (p=0.0001)
C	CS with vs. no group	$T^* = 770.7 - 761.3 = 9.4$	$df = 4 - 2 = 2$	$p < 0.05$ (p=0.009)

   Extra details: Test A indicates that CS is significantly better than independent when there is no group. Test B indicates that CS is significantly better than independent when there is a group. Therefore, we no longer consider the independence models. Test C indicates that CS with group is significantly better than CS without group.

4. The association of gender with satisfaction did differ significantly by health status score (interaction test t=-3.56, p=0.0004). At a health score of 10, women reported satisfaction that was lower on average by 0.559 points [0.511-(0.107)(10)=-0.559], while at a health score of 14, women reported satisfaction that was lower on average by 0.987 points [0.511-(0.107)(14)=-0.987]; thus the gender effect was larger for larger health status scores.

This is not quite answering the question of interest, so only partial credit was given for it: The association of gender with satisfaction did differ significantly by health status score (interaction test t=-3.56, p=0.0004). Higher health scores among men were associated with higher satisfaction scores (0.23 point increase in satisfaction for each 1 point increase in health status), similarly increasing but to a lesser degree among women (0.12 point increase in satisfaction for each 1 point increase in health status).

5. (a) From the plot on p. 20, variability in the fitted values appears to be somewhat smaller than variability in the satisfaction scores.
   (b) From the plot on p. 7, we see that the slope for both men and women is approximately linear. The residual plot on the bottom of p. 21 also shows a very random scatter above and below the horizontal line at zero, thus giving no indication of lack of linearity.
   (c) The histogram of residuals is unimodal and reasonably bell-shaped, thus indicating a good approximation to normality.

CONCEPTS

6. (a) That usually decreases our power to test the predictors of interest.
   (b) We would observe an age effect if we did a cross-sectional study and saw that 60 year olds have on average more cardiovascular disease than 40 year olds. We would see an aging effect if we enrolled 40 year olds, followed them for 20 years, and saw that on average they had more cardiovascular disease as they got older.

7. (a) The likelihood ratio test statistic would be the difference in -2 log likelihood between the model with the interaction and the model without the interaction. Degrees of freedom would equal 1, the degrees of freedom for that interaction term that is being thrown out to get the reduced model.
   (b) We are testing whether the regression coefficient for that interaction is zero ($H_0 : \alpha_{gender*hlthscore} = 0$).
   (c) Maximum likelihood.

8. The research question indicates that our outcome should be average satisfaction score per nursing home (so, 56 observations, not 8*56) and our predictor of interest should be proportion of women in each nursing home. Extra details: This is a derived variable analysis.

9. (a) Σ_i is now 4 by 4 (for the four time points) and i is indexing person.
   (b) Toeplitz would not be a reasonable structure because the time points are not equally spaced.
   (c) Make the time points equally spaced, e.g., month 1, month 2, month 3, month 4.

10. (a) Gender, time, and gender by time interaction.
    (b) Repeated measures ANOVA fits a compound symmetry structure.

11. Since Σ is assumed known here:

    (a)

$$\begin{aligned} E[\hat{\alpha}] &= E[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y] \\ &= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}E[Y] \\ &= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}X\alpha \\ &= \alpha \end{aligned}$$

    (b)

$$\begin{aligned} Var[\hat{\alpha}] &= Var[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y] \\ &= \left((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\right) Var[Y] \left((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\right)' \\ &= \left((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\right) \Sigma \left((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\right)' \\ &= \left((X'\Sigma^{-1}X)^{-1}X'\right) \left((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\right)' \\ &= \left((X'\Sigma^{-1}X)^{-1}X'\right) \left(\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}\right) \\ &= \left((X'\Sigma^{-1}X)^{-1}\right) \left(X'\Sigma^{-1}X\right) \left((X'\Sigma^{-1}X)^{-1}\right) \\ &= (X'\Sigma^{-1}X)^{-1} \end{aligned}$$

12. Two possible answers: (a) It accounts for the extra variability in $Var[\hat{\alpha}]$ that is introduced by using an estimated Σ_i instead of a known Σ_i. (b) It allows the data to directly impact the structure of Σ_i by including $\hat{\varepsilon}_i\,\hat{\varepsilon}_i'$, the empirical covariance matrix obtained from the model residuals (just like we use OLS residuals during EDA to show us what structure the data say that Σ_i should take).

EXTRA CREDIT

1. (5 points)

    (a) $Var[Y_{ij}|urban] = 0.3013 + 0.0275 = 0.3288$

    (b) $Corr[Y_{ij}, Y_{ij'}|urban] = \frac{0.0275}{0.3288} = 0.0836$