STAT502 Lab #9

1. This is adapted from the example on the Apex Enterprises candidate data in Chapter 25. Five personnel officers were first sampled from a population of many, and for each personnel officer a sample of four candidates was then collected. The response was a rating of hiring potential. Letting Y_{ij} denote the rating for the jth candidate from the i personnel officer, the one-way ANOVA model for Y_{ij} is

$$Y_{ij} = \mu_i + \varepsilon_{ij},$$

where $\mu_i \sim N(\mu_., \sigma_\mu^2)$, and $\varepsilon_{ij} \sim N(0, \sigma^2)$, with all μ_i and ε_{ij} independent.

   (a) Find the following covariances, and explain what these represent in terms of the personnel officers and candidates.

      i. $Cov(Y_{11}, Y_{11})$
      ii. $Cov(Y_{11}, Y_{21})$
      iii. $Cov(Y_{11}, Y_{12})$
      iv. $Cov(Y_{11}, Y_{22})$

   (b) State the hypotheses for testing whether there is significant variation due to personnel officers.

   (c) Letting Factor A represent the various personnel officers, use the fact that $E(MSA) = \sigma^2 + n\sigma_\mu^2$ and $E(MSE) = \sigma^2$ to argue that MSA/MSE is an appropriate test statistic for the hypotheses in part (b).

   (d) Use the following commands to read in the data and fit the model for this situation. Summarize the data with an ANOVA table.

```
data = read.table('apex.txt',header=T,sep='\t')
rating = data[,1]
A = as.factor(data[,2])
fit = lm(rating~A)
```

   (e) With $\alpha = .05$, what is the conclusion for the test in part (b)?

   (f) Another primary objective for this study is estimating the overall average rating among all candidates (and all personnel officers):

$$E(Y_{ij}) = E(\mu_i + \varepsilon_{ij}) = \mu.$$

   With equal sample sizes, a suitable estimator for $\mu.$ is $\overline{Y}_{..} = \frac{1}{4(5)} \sum \sum Y_{ij}$. Find $\overline{Y}_{..}$ for this sample.

   (g) To find the standard error of $Y_{..}$, first note that

$$\overline{Y}_{..} = \frac{1}{4(5)} \sum_{i=1}^{5} \sum_{j=1}^{4} (\mu_i + \varepsilon_{ij}) = \overline{\mu}_. + \overline{\varepsilon}_{..}$$

   so that by the independence of the μ_i and ε_{ij},

$$Var(\overline{Y}_{..}) = Var(\overline{\mu}_. + \overline{\varepsilon}_{..}) = \frac{\sigma_\mu^2}{5} + \frac{\sigma^2}{4(5)} = \frac{4\sigma_\mu^2 + \sigma^2}{4(5)} = \frac{E(MSA)}{4(5)}$$

   Thus, the standard error for $\overline{Y}_{..}$ is $\sqrt{MSA/20}$. Find this value for the data here.

(h) Use the fact that

$$\frac{\overline{Y}_{..} - \mu_{.}}{\sqrt{MSA/20}} \sim t_{df}$$

to find a 95% confidence interval for $\mu_{..}$. Can you guess what the degrees of freedom are?

(i) We can also estimate the variance components of this model. For σ^2, note that

$$\frac{SSE}{\sigma^2} \sim \chi^2_{df}$$

where χ^2_{df} represents a chi-squared distribution with $df = 5(4-1) = 15$ (the degrees of freedom for error). From a chi-square table (or software), we can obtain critical values such that

$$P\left(\chi^2(\alpha/2) < \frac{SSE}{\sigma^2} < \chi^2(1-\alpha/2)\right) = 1 - \alpha \tag{1}$$

where $1-\alpha$ is a desired confidence level. The R function `qchisq(p,df)` gives the chi-squared critical value with cumulative probability p. Use it to find the critical values above for $\alpha = .10$ (i.e., 90% confidence).

(j) Starting with result (1) above and your critical values from R, find confidence limits L and U such that

$$P\left(L < \sigma^2 < U\right) = 1 - \alpha$$

Hint: you may use the fact that $SSE = 5(4-1)MSE$

(k) To obtain a point estimate for σ^2_μ, recall that

$$E(MSE) = \sigma^2 \quad \text{and} \quad E(MSA) = \sigma^2 + 4\sigma^2_\mu$$

Use these results to find an unbiased estimator $\hat{\sigma}^2_\mu$ such that $E(\hat{\sigma}^2_\mu) = \sigma^2_\mu$. What is the numeric value of this estimator for this sample?