SOLUTION

# STAT120C Midterm: Version A (Pluta 2019)

**Instructions:**

1. There are total 3 questions (all with multiple parts).
2. There are 100 points total. The point total for each question is given at the begining of the question.
3. This is a close book exam. You are allowed to bring a double sided 8.5"X11" cheat sheet.
4. Calculators and other electronic devices are not allowed.

1. (5 + 5 + 5 + 10 + 10 + 5 + 10 = 50 points) Consider the balanced one-way ANOVA model

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad \text{with } \varepsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2),$$

for $i = 1, \ldots, I$ and $j = 1, \ldots, J$.

(a) What 3 assumptions of the ANOVA model are implied by the distribution of $\varepsilon_{ij}$?

1. Independence of observations.
2. Normally distributed errors.
3. Errors have constant variance.

(b) What is the overall mean $\mu$ of the population in terms of the parameters in the model?

$$\mu = \frac{1}{I}\sum_{i=1}^{I} \mu_i$$

(c) Assuming $\mu$ is the overall population mean, what is the interpretation of $\mu_i - \mu$?

$\mu_i - \mu$ is the difference between the mean response for group $i$ with the overall mean response. This is the effect size of the $i^{th}$ level of the treatment.

(d) Compute the MLE of the $\mu_i$, denoted $\hat{\mu}_i$. Use the notation from class, e.g. $\bar{Y}_{..}$ for the overall sample mean.

$$\mathcal{L}(\mu_1, \mu_2, \ldots, \mu_I | Y) = \prod_{i=1}^{I}\prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(Y_{ij} - \mu_i)^2\right\}$$

$$= (2\pi\sigma^2)^{-\frac{IJ}{2}} \exp\left\{\frac{-1}{2\sigma^2}\sum_i\sum_j (Y_{ij} - \mu_i)^2\right\}.$$

$$\ell(\mu_1, \ldots, \mu_I) = -\frac{IJ}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i\sum_j (Y_{ij} - \mu_i)^2$$

$$\frac{\partial \ell}{\partial \mu_i} = \frac{1}{\sigma^2}\sum_j (Y_{ij} - \mu_i)$$

$$0 = \sum_j (Y_{ij} - \hat{\mu}_i) \Rightarrow \boxed{\hat{\mu}_i = \bar{Y}_{i.}}$$

(e) Show that $SSTotal = SSW + SSB$. That is, verify the equality

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{..})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{i.})^2 + J\sum_{i=1}^{I}(\bar{Y}_{i.}-\bar{Y}_{..})^2.$$

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{..})^2 = \sum_i\sum_j(Y_{ij}-\bar{Y}_{i.}+\bar{Y}_{i.}-\bar{Y}_{..})^2$$

$$= \sum_i\sum_j(Y_{ij}-\bar{Y}_{i.})^2 + 2\sum_i\sum_j\overbrace{\cancel{(Y_{ij}-\bar{Y}_{i.})}}^{0}(\bar{Y}_{i.}-\bar{Y}_{..}) + \sum_i\sum_j(\bar{Y}_{i.}-\bar{Y}_{..})^2$$

$$= \sum_i\sum_j(Y_{ij}-\bar{Y}_{i.})^2 + \sum_i\sum_j(\bar{Y}_{i.}-\bar{Y}_{..})^2$$

$$= \sum\sum(Y_{ij}+\bar{Y}_{i.})^2 + J\sum_i(\bar{Y}_{i.}-\bar{Y}_{..})^2, \text{ since } \sum_j(Y_{ij}-\bar{Y}_{i.}) = \sum_j Y_{ij} - J\bar{Y}_{i.} = 0, \ \forall i.$$

(f) The $F$-statistic for testing $H_0$ : All $\mu_1 = \mu_2 = \cdots = \mu_I = \mu$ against $H_1$ : At least one $\mu_i \neq \mu$ is

$$F = \frac{SSB/(I-1)}{SSW/[I(J-1)]}.$$

Give a brief explanation for why this statistic intuitively makes sense as a test of $H_0$.

F is the ratio of the variance explained by assuming each group has its own mean $\mu_i$ to the residual variance. When $H_0$ is true, the between group variance will be small relative to SSW, so we can reject $H_0$ when F is large.

(g) Show that the likelihood ratio test for $H_0$ from the previous part is equivalent to the ANOVA $F$-test. (Hint: show that the likelihood ratio test rejects for large values of $SSB/SSW$.)

Likelihood ratio test statistic is: $\Lambda(Y) = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right)^{-IJ/2}$,

where $\hat{\sigma}_0^2 = \frac{1}{IJ}\sum_i\sum_j(Y_{ij}-\hat{\mu})^2 = \frac{1}{IJ}\sum_i\sum_j(Y_{ij}-\bar{Y}_{..})^2$

$\hat{\sigma}^2 = \frac{1}{IJ}\sum_i\sum_j(Y_{ij}-\hat{\mu}_i) = \frac{1}{IJ}\sum_i\sum_j(Y_{ij}-\bar{Y}_{i.})$

$\rightarrow$ Reject $H_0$ when $\Lambda(Y)$ is small $\Rightarrow$ Reject when $\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}$ is large

$$\Rightarrow \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = \frac{SSTot}{SSW} = \frac{SSW+SSB}{SSW}, \text{ by (e)}$$

$$= 1 + \frac{SSB}{SSW}$$

$\Rightarrow$ Reject when $\frac{SSB}{SSW}$ is large

$\Rightarrow$ Reject $H_0$ when $F = \frac{SSB/(I-1)}{SSW/[I(J-1)]}$ is large.

2. (10 + 5 + 5 = 20 points) A manufacturer of computer hard drives is interested in the effect of the coating used to insulate the hard drives on overall drive access speed. In particular, the manufacturer is interested in comparing the average drive speed across seven different types of coating. In order to determine if there is a difference, a balanced experiment was carried out on a total of 28 different hard drives.

(a) Based on the above information, complete the partial ANOVA table

| Source | $SS$ | $df$ | $MS$ | $F$ |
|---|---|---|---|---|
| Treatment | 240 | 6 | 40 | 4 |
| Error | 210 | 21 | 10 | |
| Total | 450 | 27 | | |

(b) Identify the 95th-percentile of the null distribution for the ANOVA $F$-test from (a).

i. $F_{0.95}(1, 6) = 5.98$
ii. $F_{0.95}(1, 21) = 4.32$
iii. $F_{0.95}(6, 21) = 2.57$
iv. $F_{0.95}(1, 27) = 4.21$
v. $F_{0.95}(7, 27) = 2.37$
vi. $F_{0.95}(7, 21) = 2.49$
vii. $F_{0.95}(6, 28) = 2.44$
viii. $F_{0.95}(6, 27) = 2.46$

(c) Using the completed table in (a) and $F$ percentile identified in (b), carry out the hypothesis test to determine if the average drive speed differs among the coating types at the 0.05 significance level. Be sure to clearly state your hypotheses, test statistic, and conclusions in context of the data.

Hypotheses $\begin{cases} H_0: \text{Mean drive speed is the same across coating types} \\ H_1: \text{At least one coating type produces a different mean speed} \end{cases}$

$F = 4 \overset{H_0}{\sim} F(6,21)$

Reject $H_0$ when $F > F_{.95}(6,21) = 2.57$

$\Rightarrow$ There is evidence to conclude that at least one coating type produces a different mean drive speed, at the $\alpha = 0.05$ level of significance.

3. (10 + 5 + 5 + 5 + 5 = 30 points) A laboratory is conducting a small study to determine the effects of a new drug combination on the reduction of tumor growth in lab rats. A balanced two-way design is proposed, with drug A administered at 2 levels: 0 mg (control), 200mg; and drug B administered at 4 levels: 0mg, 10mg, 50mg, 300mg. The measured outcome is the change in tumor size 2 weeks after treatment. It is belived that the drugs have individual effects on tumor growth, but it is unkown how the drugs work in combination.

(a) Write an appropriate model for this study, including all distributional assumptions, and parameter constraints.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}, \quad i = 1,2; \; j = 1,2,3,4; \; k = 1, \dots, K$$

$$\sum_i \alpha_i = 0, \quad \sum_i \delta_{ij} = \sum_j \delta_{ij} = 0, \quad \varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$$

$$\sum_j \beta_j = 0$$

(b) What hypothesis should be tested to answer the primary question of the study?

Test for Interactions
$$\begin{cases} H_0: \text{All } \delta_{ij} = 0 \\ H_1: \text{At least one } \delta_{ij} \neq 0 \end{cases}$$

(c) Write the test statistic and reference distribution to test this hypothesis. The statistic should be expressed in terms of the observed data $Y_{ij}$ and the appropriate sample means, using our notation from class.

$$F^* = \frac{SSAB/[(I-1)(J-1)]}{SSE/[IJ(K-1)]} \overset{H_0}{\sim} F_{(I-1)(J-1),\, IJ(K-1)}$$

where $SSAB = K \sum_i \sum_j (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$

$SSE = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2$

(d) Due to budgetary restrictions, each treatment combination will only be administered to two rats (i.e. $K = 2$ in our notation). Briefly explain what impact this small sample size may have on the hypothesis test.

This study would likely be underpowered to detect significant interactions since each $\delta_{ij}$ would be estimated from only 2 obs, possibly resulting in large variance for $\hat{\delta}_{ij}$.

Note that the type I error will not be affected, only the power.

(e) Suppose it is determined that the 200mg dose of drug A decreases the mean outcome relative to the control for drug A, and that the mean outcome decreases as the dosage of drug B is increased. Further suppose no significant interaction between the drugs is present. Sketch a possible interaction plot for these data, with clearly labeled axes and legend.