STAT120C Homework 4

Assigned Thursday April 24, 2019

Due Thursday May 2, 2019 by 5pm in the Dropbox in DBH

1. Assume that we have I random samples. For $i = 1, \cdots, I$, we assume that the ith random sample $(Y_{i1}, Y_{i2}, \cdots, Y_{iJ})$ came from the normal distribution with mean μ_i and variance σ^2. These assumptions can be written as follows

$$Y_{ij} = \mu_i + \epsilon_{ij} \ , \ i = 1, \cdots, I \ ; \ j = 1, \cdots, J$$

where $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$.

   (a) Obtain the liklihood function and show that maximizing the likelihood is equivalent to minimizing $Q = \sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{ij} - \mu_i)^2$.

   (b) Show that for a fixed i, the MLE of μ_i is $\hat{\mu}_i = \bar{Y}_{i\cdot} = \frac{1}{J} \sum_{j=1}^{J} Y_{ij}$.

   (c) Derive and simplify the likelihood ratio test statistic for $H_0 : \mu_1 = \cdots = \mu_I = \mu$.

   (d) The sum of squares of total is defined as $SSTO = \sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2$. Prove that it can be partitioned into SSB and SSW, where $SSB = J \sum_{i=1}^{I} (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$, $SSW = \sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{ij} - \bar{Y}_{i\cdot})^2$, and $\bar{Y}_{\cdot\cdot} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij}}{IJ}$.

   (e) What are the distributions of $\bar{Y}_{1\cdot}, \cdots, \bar{Y}_{I\cdot}$ when $\mu_1 = \cdots = \mu_I$? Justify your answer.

   (f) Prove that SSB/σ^2 follows a chi-squared distribution when $\mu_1 = \cdots = \mu_I$. Specify the degrees of freedom.

   (g) What distribution does SSW/σ^2 follow? Justify your answer. Be sure to specify the degrees of freedom.

   (h) Are the results in (e) and (f) sufficient to construct an F statistic? If not, what other results(s) would you need? You don't have to prove the result(s).

   (i) Give the F statistic in terms of SSB and SSW. Which distribution does F follow when $\mu_1 = \cdots = \mu_I$?

2. Samples of each of three types of stopwatches were tested. 21 watches of type I, 16 watches of tyep II, and 11 watches of type III were randomly selected. The cycles (on-off-restart) survied until some part of the mechanism failed were measured. Below shows part of the R output from analyzing these data.

	Df	Sum Sq	Mean Sq	F value
type	____	754.4	_______	_______
Residuals	____	______	56.7	

   (a) Complete the ANOVA table.

   (b) Write the formula for estimating σ^2 (using our usual one-way ANOVA model notation), and compute the estimate from the table.

   (c) Calculate the total sum of squares.

   (d) Test whether there is significant difference among the types. Conduct your test at the significance level $\alpha = 0.05$.

3. The staff of a service center for electronic equipment includes three technicians who specialize in repairing three widely used types of disk drives for desktop computers. A study was desired to study the effects of technician and type of disk drives on the service time. In the study, each technician was

randomly assigned to six jobs on each type of disk drive. The minutes required to complte each job were recorded. We can study the data using the following two-way ANOVA table:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ij} \;, i = 1, \cdots, I \;;\; j = 1, \cdots, J \;;\; k = 1, \cdots, K$$

where $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{i=1}^{I} \delta_{ij} = \sum_{j=1}^{J} \delta_{ij} = 0$.

(a) Specify I, J, K in the study.

(b) What assumption(s) do we place on ϵ_{ijk}'s?

(c) Interpre the parameters $\mu, \alpha_i, \beta_j, \delta_{ij}$.

(d) Write (do not prove) the sum of squares of total, which is $SSTO = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2$, as a sum of four sums of squares. The sums of squares need to be given in terms of Y_{ijk}'s.

(e) Complete the following partial ANOVA table.

Source	*df*	*SS*	*MS*	*F*
Drive Type		200		
Technician		10		
Interaction		90		
Error				–
Total		500	–	–

(f) What is an unbiased estimate of σ^2? You can use the information in the above table.

(g) Does the type have an effect on service time? Based on the table you completed, carry out a hypothesis test at level 0.05. Be sure to clearly state your hypothesis, test statistic, rejection region, and conclusion.

(h) Based on the table you completed, carry out a hypothesis test (at level $\alpha = 0.05$) to determine whether the effects of type depend on technician. Be sure to clearly state your hypothesis, test statistic, rejection region, and conclusion.

4. Answer the following short questions.

(a) Suppose that $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ where μ is unknown. Construct a χ^2 random variable from this sample.

(b) Consider two normal samples $Y_{ij} \overset{iid}{\sim} N(\mu_i, \sigma^2)$, $i = 1, 2, j = 1, \ldots, J_i$, with sample means $\bar{Y}_{1\cdot}$, $\bar{Y}_{2\cdot}$, and sample variances s_1^2, s_2^2. Show that the pooled sample variance s_p^2 is an unbiased estimator of σ^2, where

$$s_p^2 = \frac{(J_1 - 1)s_1^2 + (J_2)s_2^2}{J_1 + J_2 - 2}.$$

(c) Suppose $Y_{ij} \overset{iid}{\sim} N(\mu_i, \sigma_i^2)$, $i = 1, 2, j = 1, \ldots, J_i$, and assume $\sigma_1^2 \neq \sigma_2^2$.

i. Which assumption of the one-way ANOVA model is violated?

ii. What are the possible effects on the performance of the ANOVA F-test when this assumption is violated?

(d) State the Bonferonni correction for testing K hypotheses simultaneously, and explain why such a correction is necessary.

Table: Percentiles of the F-distribution

df_1	df_2	0.975	0.950	0.900
1	6	8.813	5.987	3.776
1	24	5.717	4.260	2.927
1	45	5.377	4.057	2.820
1	54	5.316	4.020	2.801
2	6	7.260	5.143	3.463
2	24	4.319	3.403	2.538
2	45	4.009	3.204	2.425
2	54	3.953	3.168	2.404
3	6	6.599	4.757	3.289
3	24	3.721	3.009	2.327
3	45	3.422	2.812	2.210
3	54	3.369	2.776	2.188
4	6	6.227	4.534	3.181
4	24	3.379	2.776	2.195
4	45	3.086	2.579	2.074
4	54	3.034	2.543	2.052
9	6	5.523	4.099	2.958
9	24	2.703	2.300	1.906
9	45	2.412	2.096	1.774
9	54	2.360	2.059	1.750