

# STAT3600A Linear Statistical Analysis

## Assignment 3

(Hand in the answers on or before 11:59PM, November 30, 2020)

### Notes:

- Tutors only accept PDF files, other format will NOT be graded.
- Please write all necessary intermediate steps, not just the final answer.

### Question 1

In an experiment to investigate the effect of colour of paper (blue, green, orange) on response rates for questionnaires distributed by the “windshield method” in supermarket parking lots, 18 representative supermarket parking lots were chosen in a metropolitan area, so that each colour was assigned at random to six lots. The response rates (in %) follow. Assume that a one-way classification model is appropriate.

Blue	28	27	30	27	34	33
Green	33	29	25	30	28	31
Orange	33	27	26	29	29	28

- Specify the one-way classification model in a formal way. And obtain the fitted values for the model parameters.
- Obtain the residuals.
- Compile the ANOVA table.
- Conduct a size 0.1 test to determine whether or not the mean response rates for the three colours differ.
- When informed of the findings, an executive said: “See? I was right all along. We might as well print the questionnaires on plain white paper, which is cheaper.”  
Does this conclusion follow from the findings of the study? Discuss.

## Question 2

In their report to the Senate Public Works Committee, *Texaco Inc.* presented a dataset as evidence that the Octel filter, developed by Associated Octel Company for reducing air pollution, was at least as good as a standard silencer in controlling vehicle noise levels. The findings were based on a 2-way factorial experiment with 6 replications. The two factors are:

- TYPE – type of filter (1 = standard silencer, 2 = Octel filter),
- SIZE – vehicle size (1 = small, 2 = medium, 3 = large).

The response variable (NOISE) is the noise level reading in decibels. The complete dataset is detailed below:

SIZE	TYPE	NOISE					
1	1	810	820	820	835	835	835
2	1	840	840	845	845	855	850
3	1	785	790	785	760	760	770
1	2	820	820	820	825	825	825
2	2	820	820	825	815	825	825
3	2	775	775	775	770	760	765

- (a) Write down an expression of a two-way classification model relevant to the above study. Specify the model assumptions carefully.
- (b) Compile an ANOVA table for testing whether there are interaction effects between the two factors at the 5% significance level.
- (c) Based on the outcome of the test in (b), it is recommended that the comparison between standard silencer and Octel filter be made separately for the three vehicle sizes. Explain.
- (d) Denote by  $\sigma^2$  the variance of an individual noise level reading.
  - (i) Calculate the six estimated treatment means obtained by fitting a 2-way classification model to the data.
  - (ii) Show that the difference between any two estimated treatment means has variance  $\sigma^2/3$ .
  - (iii) Extract from the ANOVA table in (b) an estimate of  $\sigma^2$ .
  - (iv) Calculate a 95% confidence interval for the difference in the expected noise level reading between standard silencer and Octel filter for small vehicles.
  - (v) Calculate a 95% confidence interval for the difference in the expected noise level reading between standard silencer and Octel filter for medium vehicles.

- (vi) Calculate a 95% confidence interval for the difference in the expected noise level reading between standard silencer and Octel filter for large vehicles.
- (e) With reference to the above findings, do you think that the claim of *Texaco Inc.* is supported by the observed data? Explain.

### Question 3

**(Assignment 2 Question 3 revisit)**

In a small-scale experimental study of the relation between degree of brand liking ( $Y$ ) and moisture content ( $X_1$ ) and sweetness ( $X_2$ ) of the product, the following results were obtained from the experiment based on a completely randomized design [**dataset is attached in the Moodle site**]

**Note:** You are allowed to directly use the result you found in assignment 2.

$i$	1	2	3	...	14	15	16
$x_{i1}$	4	4	4	...	10	10	10
$x_{i2}$	2	4	2	...	4	2	4
$Y_i$	64	73	61	...	95	94	100

- (a) Obtain the studentized deleted residuals.
- (b) Obtain the cook's distance  $D_i$  for each case and prepare an index plot.
- (c) Which case has the largest absolute studentized deleted residual? For this case, please obtain the *DFFITs* and the cook's distance values for this case to assess the influence of this case. What do you conclude?

### Question 4

A dataset consists of  $n = 565$  swim times (in sec) for male and female swimmers competing in 500-yard freestyle at three biennial swimming competitions (held in 2009, 2011 and 2013 respectively). The response variable  $Y$  is taken to be  $\ln(\text{swim time})$ . A breakdown of sample sizes into the six gender/year categories is given below:

Gender	Year	2009	2011	2013	Total
Male		89	76	83	248
Female		110	113	94	317
Total		199	189	177	565

A two-way ANCOVA model is set up to study the effects of the factors "gender" (male/female) and "year" (2009/2011/2013) on  $Y$ , using "age" ( $X$ ) as a concomitant variable. The model has the following expression:

$$Y = \mu + \alpha_1 A_1 + \beta_1 B_1 + \beta_2 B_2 + (\alpha\beta)_{11} A_1 B_1 + (\alpha\beta)_{12} A_1 B_2 + \gamma X + \epsilon,$$

where  $\epsilon \sim N(0, \sigma^2)$  and the dummy variables  $A_1, B_1, B_2$  are defined by

$$A_1 = \begin{cases} 1, & \text{female,} \\ 0, & \text{male;} \end{cases} B_1 = \begin{cases} 1, & \text{year 2011,} \\ 0, & \text{otherwise;} \end{cases} B_2 = \begin{cases} 1, & \text{year 2013,} \\ 0, & \text{otherwise.} \end{cases}$$

Fitting the above model to the dataset yields the least squares estimates

$$\hat{\mu} = 4.8178, \hat{\alpha}_1 = 0.2513, \hat{\beta}_1 = 0.04155, \hat{\beta}_2 = -0.01442, \\ \widehat{(\alpha\beta)}_{11} = -0.09820, \widehat{(\alpha\beta)}_{12} = -0.04873, \hat{\gamma} = 0.02014$$

for the parameters  $\mu, \alpha_1, \beta_1, \beta_2, (\alpha\beta)_{11}, (\alpha\beta)_{12}, \gamma$ , respectively.

- (a) Calculate the six fitted treatment means for the six treatments defined by the two factors "gender" and "year", assuming a constant age  $X = 0$ .
- (b) To test for any significant interactions between "gender" and "year", the following two-way ANCOVA table is compiled:

Source	s.s.	d.f.	m.s.	F-ratio
Age & Main effects of "gender" & "year"	24.7208	?		
Interactions between "gender" & "year"	0.2276	?	?	?
Error	17.7429	?	?	
Total	42.6914	564		

Calculate the missing entries marked with "?" in the above ANCOVA table.

- (c) Test whether there is significant interactions between "gender" and "year" at 5% level.