COSC6323 - Homework 4

April 2021

Instructions

Please compile your report as a pdf. Submit both: FirstName_LastName.rmd (if used) and FirstName_LastName.pdf files to the black board.

1 Task 1

In a study for determining the effect of weaning conditions on the weight of 9-week-old pigs, data on weaning (WWT) and 9-week (FWT) weights were recorded for pigs from three litters. One of these litters was weaned at approximately 21 days (EARLY), the second at about 28 days (MEDIUM), and the third at about 35 days (LATE). The data are given in Table 1. Perform an analysis of covariance using FWT as the response, weaning time as the factor, and WWT as the covariate. Comment on the results. Is there a problem with assumptions?

EARLY		MEDIUM		LATE	
WWT	FWT	WWT	FWT	WWT	FWT
9	37	16	48	18	45
9	28	16	45	17	38
12	40	15	47	16	35
11	45	14	46	15	38
15	44	14	40	14	34
15	50	12	36	14	37
14	45	10	33	13	37

Table 1. Data for exercise 1

2 Task 2

Skidding is a major contributor to highway accidents. The following experiment was conducted to estimate the effect of pavement and tire tread depth

on spinout speed, which is the speed (in mph) at which the rear wheels lose friction when negotiating a specific curve. There are two asphalt (ASPHALT1 and ASPHALT2) pavements and one concrete pavement and three tire tread depths (1-, 2-, and six-sixteenths of an inch). This is a factorial experiment, but the number of observations per cell is not the same. The data are given in Table2.
(a) Perform the analysis of variance using both the dummy variable and "standard" approaches. Note that the results are not the same although the differences are not very large.
(b) The tread depth is really a measured variable. Perform any additional or alternative analysis to account for this situation.
(c) It is also known that the pavement types can be characterized by their coefficient of friction at 40 mph as follows:
ASPHALT1: 0.35
ASPHALT2: 0.24
CONCRETE: 0.48
Again, perform an alternative analysis suggested by this information. Which of the three analyses is most useful?

OBS	PAVE	TREAD	SPEED	OBS	PAVE	TREAD	SPEED
1	ASPHALT1	1	36.5	14	CONCRETE	2	45.0
2	ASPHALT1	1	34.9	15	CONCRETE	6	47.1
3	ASPHALT1	2	40.2	16	CONCRETE	6	51.2
4	ASPHALT1	2	38.2	17	OCNCRETE	6	51.2
5	ASPHALT1	2	38.2	18	ASPHALT2	1	33.4
6	ASPHALT1	6	43.7	19	ASPHALT2	1	38.2
7	ASPHALT1	6	43.0	20	ASPHALT2	1	34.9
8	CONCRETE	1	40.2	21	ASPHALT2	2	36.8
9	CONCRETE	1	41.6	22	ASPHALT2	2	35.4
10	CONCRETE	1	42.6	23	ASPHALT2	2	35.4
11	CONCRETE	1	41.6	24	ASPHALT2	6	40.2
12	CONCRETE	2	40.9	25	ASPHALT2	6	40.9
13	CONCRETE	2	42.3	26	ASPHALT2	6	43.0

Table 2. Spinout Speeds for Exercise 2

3 Task 3

Cochran and Chamlin (2006) used data from the National Opinion Research Council - General Social Survey (NORC-GSS) to compare whites' and blacks' opinions of the death penalty. The data consisted of responses from 32,937 participants collected between 1972 and 1996. (The question was not asked every year.) The outcome variable was whether the respondent did or did not support

the death penalty. Their hypotheses concerned both the possible difference between blacks and whites, and the possible change in that difference over time. The authors provided a table of the percentage of whites and blacks each year that supported the death penalty, shown in Table 3.
(a) Convert the percentages given in Table 3 to the Ln(odds) within each race and year, and plot the ln(odds) versus year. Comment on any patterns you see. If there is a trend in time, does it appear linear or quadratic?
(b) Use logistic regression to model the probability a person will support the death penalty, as a function of race and year. Is there significant evidence that a quadratic term in year improves the model? Assume that in each year's sample there were 1100 whites and 400 blacks.
(c) Attempt to improve your model by adding interactions of race with the linear and quadratic variables in time. Do the interactions significantly improve the model?
(d) The authors of the study refer to the gap between white and black support as "enduring". Are your results in part (c) consistent with this?

Year	White%	Black%	Year	White%	Black%
1972	57.4	28.8	1985	79.0	49.7
1973	63.6	35.8	1986	75.3	42.7
1974	66.3	36.3	1987	73.7	42.9
1975	63.2	31.9	1988	76.0	42.5
1976	67.5	41.1	1989	76.5	56.1
1977	70.0	41.6	1990	77.7	52.3
1978	69.4	43.0	1991	71.4	42.7
1980	70.3	39.1	1993	75.4	51.5
1982	76.9	48.4	1994	78.3	50.7
1983	76.2	45.0	1996	75.5	50.3
1984	74.5	43.5			

Table 3. Data on Death Penalty Opinion for Exercise 3

4 Task 4

Popkin (1991) presented the data shown in Table 4 for number of auto crashes and number of alcohol-related (A/R) auto crashes for young drivers in North Carolina. You are interested in whether the probability a crash will be A/R is related to age and gender.

	Age < 18		18 <= Age <= 20		21 <= Age <= 24		25 <= Age	
	Male	Female	Male	Female	Male	Female	Male	Female
Total	14589	8612	21708	10941	25664	13709	41304	25183
A/R	553	117	2147	470	3250	540	4652	794

Table 4. A/R Crashes for exercise 4

(a) Construct a profile plot (similar to those for the two-way ANOVA) for the Ln(odds) that a crash will be alcohol-related, using the age category on the horizontal axis and separate symbols for gender. Discuss the apparent effects. Is there a graphical suggestion of an interaction?
(b) Construct a profile plot in the same way as for part (a), using the empirical probability that a crash will be alcohol-related. Is there a graphical suggestion of an interaction?
(c) Construct a dummy variable system for the age category and gender, and fit a logistic regression that only includes main effects. Interpret the main effects, using the profile plot from part (a).
(d) Fit a logistic regression that includes main effects and interactions.
(e) Construct a likelihood ratio test for the null hypothesis that none of the interactions are significant. Interpret the results.

5 Task 5

Van den Bos et. al. (2006) analyzed Y = Outcome Satisfaction for 138 participants in an experiment with two factors: Cognitive Busyness (low to high) and Outcome (equal to others, better than others, worse than others) The mean values for Y within each cell are given below.

		Outcome		
		Equal to others	Better than others	Worse than others
Cognitive	Low	6.5(a)	3.0(c)	1.6 (d)
Busyness	High	6.3(a)	4.0(b)	2.0 (d)

(a) Construct a profile plot that will allow you to inspect the apparent effects in the data.
(b) The authors cite the following test statistics from the two-way ANOVA:
Main effect for Outcome: $F(2,132) = 236.56$, $p < 0.001$
Main effect for Busyness: $F(1,132) = 4.36$, $p < 0.04$
Interaction: $F(2,132) = 3.38$, $p < 0.04$
Use this information, together with your profile plot, to write a short paragraph explaining the effects of these factors on Outcome Satisfaction.
(c) The authors carried out "the least significant difference test for means ($p < 0.05$) with the six cells of our design serving as the independent variable." how many independent samples t-tests are implied by this statements?
(d) In the table above, cell means with the same letter in parentheses were not significantly different using the method described in part (c). The authors state "there were no efforts of Cognitive Busyness within the equal-to-other and

worse-than-other conditions.' Is this consistent with you profile plot and the formal test statistics given in part (b)?