

MATH 4044 – Statistics for Data Science

Assessable Practical Exercise 3 (SP5 2022)

Due 9 Oct 2022 by 11:59pm

Instructions:

- This exercise is worth 2.5% of your final mark and it is due no later than **11:59pm on Sunday 9 Oct** in Week 9.
- The exercise will be marked out of 20.
- You will need to submit your **individual** work via Learnonline as a **single file**, in either a Microsoft Word (doc or docx) or pdf file format. Your submission should consist of the SAS output you have generated (40%), plus the requested interpretation (50%). Please include only the most relevant SAS output (10%).
- You are welcome to discuss the exercise, give advice and share tips with other students, but there should be no sharing of files or output.

Assessment task:

A company studied the effects of three different types of promotions (*treatments*) on the sales of a specific brand of crackers (*response*):



Promotion 1 – The crackers were on their regular shelf, but free samples were given out in the store;

Promotion 2 – The crackers were on their regular shelf, but were given additional shelf space;

Promotion 3 – The crackers were given special display shelves at the end of the aisle in addition to their regular shelf space.

The company selected 15 stores to participate in the study. Each store was randomly assigned one of the 3 promotion types, with 5 stores assigned to each promotion. Data was collected on the number of boxes of crackers sold during the promotion period, as well as the number sold during the preceding time period.

(a) Use SAS to perform a one-way ANOVA test to determine whether there is a statistically significant difference in the mean number of boxes sold during the promotion period, by type of promotion:

- Check the necessary conditions;
- Examine and comment on residuals;
- If appropriate, perform post-hoc tests;
- Report and briefly discuss your results.

(b) Use SAS to perform a one-way ANCOVA with the number of the cases sold in the preceding period (variable `last`) as a covariate:

- Confirm that there is a linear relationship between the response variable and the covariate (a scatterplot and a correlation coefficient plus a comment will suffice);
- Check the two additional ANCOVA assumptions (report and comment only on the parts of the output most directly relevant to condition checking for this exercise):
 - Independence of the covariate and the treatment effect (perform a one-way ANOVA test; there should be no statistically significant difference);
 - Equality of slopes (add and check significance of the interaction term);
- Decide what your final ANCOVA model should be (with or without the interaction term) and perform post-hoc analysis for this model;
- Examine and comment on residuals;
- Report and briefly discuss your results.

Note: You should obtain and examine Type III Sum of Squares (`ss3`). Also obtain estimates of 'least squares means' (`lsmeans`) which are means adjusted for the covariate.

(c) Compare results from parts (a) and (b). Did including the covariate reduce the error variance and thus produce better estimates of mean sales levels by the type of promotion? Which model is a better fit to the data? Which type of promotion appears to be the most effective? Explain briefly.

Data file for this exercise:

The data is stored in a SAS data file called `crackers.sas7bdat` located in `mydata` library on the SAS OnDemand server. Variables in that file are as follows:

Variable	Description
<i>Obs</i>	Observation number
<i>cases</i>	The number of boxes of crackers sold during the promotion period
<i>last</i>	The number of boxes of crackers sold during the preceding period
<i>treat</i>	Treatment number: 1 = crackers on regular shelves plus free samples given out in store 2 = crackers on regular shelves with additional shelf space 3 = crackers on regular shelves plus special display shelves at the end of the aisle
<i>store</i>	Store number, 1 to 5