Business Analytics (110-1)

Assignment 3

Due: 9:00 am, Tue 23-Nov-2021

1.

A farmer wants to conduct an experiment to compare eight varieties of oats. The farmer knows that the growing area is heterogeneous so he decided to group the area into five blocks. He randomly plants each variety of oats in each block and records the yields accordingly. Dataset oatvar is the experiment result.

(a) What kind of experiment is this farmer using?

(b) Is there any interaction effect between the variety of oats and the growing area block?

(c) Conduct a hypothesis test to determine whether the yield of oats is affected by different varieties at 5% significance level.

(d) Check the diagnostics. Is there any unusual findings?

(e) Is it necessary to perform multiple comparisons? If yes, carry out the comparison following the structure and procedure presented in the lecture.

2.

Detergent manufacturer frequently makes claims about the effectiveness of their products. A consumer-protection service decided to test the five best selling brands of detergent, where each manufacturer claims that its product produces the "whitest whites" in all water temperatures. The experiment was conducted in the following way. One hundred fifty white sheets were equally soiled. Thirty sheets were washed in each brand – 10 with cold water, 10 with warm water, and 10 with hot water. After washing, the "whiteness" scores for each sheet were measured with laser equipment. Dataset detergent is the experimental result.

Column 1: water temperature code
Column 2: scores for detergent 1 (first 10 rows = cold water, middle 10 rows = warm, last 10 = hot)
Column 3: scores for detergent 2 (same format as column 2)
Column 4: scores for detergent 3 (same format as column 2)
Column 5: scores for detergent 4 (same format as column 2)
Column 6: scores for detergent 5 (same format as column 2)

Is there sufficient statistical evidence to infer that there are differences in whiteness scores between the five detergents, differences in whiteness scores between the three water temperatures, or interaction between detergent and temperatures?

3.

When car dealers lease a car, how do they decide what to charge? One answer, if you've got a lot of unpopular cars to move, is to charge whatever it takes to get the cars off the lot. A different answer considers the so-called "*residual price*" of the car at the end of the lease. The residual price of a leased car is the value of this car in the used-car market.

How should we estimate the residual price of a car? The residual price depends on how much the car was worth originally, such as the manufacturer's list price. Let's take this off the table by limiting our attention to a particular type of car. Let's also assume that we are looking at cars that have not been damaged in an accident.

What else matters? Certainly, the age of the car affects its residual price. Customers expect to pay less for older cars, on average. Older cars have smaller residual price. The term of the lease, say 2 or 3 years, has to cover the cost of the ageing of the car. Another factor that affects the residual price is the type of use. An older car that is in the great condition might be worth more than a newer car that has been heavily driven. It seems as though the cost of a lease ought to take both duration and use into account.

Dataset `used_bmw` lists 218 BMW's popular 3-series.

(a) Check scatterplots of the variables. Do the relationships appear straight enough to permit using multiple linear regression with these variables?
(b) Fit the appropriate multiple linear regression model. Does this model meet the assumptions for multiple linear models?
(c) Build confidence intervals for the partial effects of age and mileage.
(d) Summarize the results of your model. Recommend terms for leases that cover the costs of ageing and mileage.
(e) Do you have any caveats that should be pointed out with your recommended terms? For example, are there any evident lurking variables?