COSC 6323 – Statistical Methods in Research
Homework 3
February 21, 2020

COMPUTATIONAL PHYSIOLOGY LAB

Problems

1. A manufacturer of air conditioning ducts is concerned about the variability of the tensile strength of the sheet metal among the many suppliers of this material. Four samples of sheet metal from four randomly chosen suppliers are tested for tensile strength. The data are given in **AirCondition.txt**

(a) Perform the appropriate analysis to ascertain whether there is excessive variation among suppliers.
(b) Estimate the appropriate variance components.

2. For laboratory studies of an organism, it is important to provide a medium in which the organism flourishes. The data for this exercise shown below are from a completely randomized design with four samples for each of seven media. The response is the diameters of the colonies of fungus.

Medium	*Fungus Colony Diameters*			
WA	***4.5***	***4.1***	***4.4***	***4.0***
RDA	***7.1***	***6.8***	***7.2***	***6.9***
PDA	***7.8***	***7.9***	***7.6***	***7.6***
CMA	***6.5***	***6.2***	***6***	***6.4***
TWA	***5.1***	***5***	***5.4***	***5.2***
PCA	***6.1***	***6.2***	***6.2***	***6***
NA	***7.0***	***6.8***	***6.6***	***6.8***

(a) Perform an analysis of variance to determine whether there are different growth rates among the media.
(b) Is this exercise appropriate for preplanned or post hoc comparisons? Perform the appropriate method and make recommendations.

3. A research article states that y = satisfaction with police (measured using a survey questionnaire) is related to x = neighborhood social disorder (vandalism, traffic, decayed buildings and streets, etc.). They summarize their findings as

> ***Satisfaction with police declines swiftly as perceptions of neighborhood social disorder increase (n =178, beta1 $\widehat{\beta 1}$= -0.62, r^2 = 0.18).***

(a) Calculate the ***F*** test for a linear relationship between the two variables. Are the authors justified in claiming the two are related?
(b) Give a 95% confidence interval for the true value off beta1. What allows the authors to use the word "declines" in their statement?

4. Given the values in below table for the independent variable ***x*** and dependent variable ***y***:

x	y	x	y
-1	7	1	5
-1	3	1	8
-1	6	1	12
-1	6	1	8
-1	7	1	6
-1	4	1	8
-1	2	1	9

(a) Perform the linear regression of ***y*** on ***x***. Test $H_0: \beta_1 = 0$.
(b) Note that half of the observations have ***x*** = -1 and the rest have ***x*** = 1. Does this suggest an alternative analysis? If so, perform such an analysis and compare results with those of part(a).

5. Data were collected to discern environmental factors affecting health standards. For 21 small regions we have data on the following variables:

POP: population (in thousands),
VALUE: value of all residential housing, in millions of dollars; this is the
proxy for economic conditions,
DOCT: the number of doctors,
NURSE: the number of nurses,
VN: the number of vocational nurses, and
DEATHS: number of deaths due to health-related causes (i.e., not accidents); this is the proxy for health standards.

The data are given in file: **Health.txt**

(a) Perform a regression relating DEATHS to the other variables, excluding POP. Compute the variance-inflation factors; interpret all results.
(b) Obviously multicollinearity is a problem for these data. What is the cause of this phenomenon? It has been suggested that all variables should be converted to a per capita basis. Why should this solve the multicollinearity problem?
(c) Perform the regression using per capita variables. Compare results with those of part (a). Is it useful to compare *R*^2 values? Why or why not?

Submission

- Zip your submission folder (contains code, documents, pdf, and plots) and submit to blackboard. Name your zip file "**FirstName_LastName_HW3.zip**". One submission per individual.
- The homework should be submitted on BlackBoard by **3pm on March 20**, **2019**.
- The homework report should be a pdf generated in Latex with figures from R only. All the figures and plots should have appropriate labels, titles and annotations.
- The code should also be provided.

Grading

Total points **25 pts** (5 pts each)