CHAPTER 2 REVIEW QUESTIONS WITH ANSWERS

CHAPTER 2:

For G. David Garson, *Data Analytics for Social Science: Applications in R*

1. Which R package was used for the Levene's test in ANOVA and for hypothesis testing?

    A. library(car)

    B. library(caret)

    C. library(cluster)

    D. library(corrplot)

    ANSWER: The correct answer is A. Others were:

    library(caret)   # Used for logistic regression

    library(cluster)   # Used for cluster analysis

    library(corrplot)   # Used for visualizing correlation matrices

2. Which R package is NOT used in conjunction with crosstabulation?

    A. library(expss)

    B. library(gmodels)

    C. library(plyr)

    D. library(vcd)

    ANSWER: The correct answer is C. The "plyr" package which is used, for example, to apply a

    statistical function to each subset of a data frame. The "expss" and "gmodels" packages

    contain crosstabulation commands (the `cro()` and `CrossTable()` commands

    respectively). The "vcd" package supports measures of association for crosstabulation

through its `assocstats()` command. There are also other crosstabulation programs, such as `xtabs()` from the "stats" package in R's System Library.

3. Given that a data frame called "survey" is loaded and has a column labeled "marital", what is wrong with this command?: `table(marital)`

A. The `table()` command requires two variables, not just one.

B. Should be `TABLE(survey$marital)`

C. Should be `table::survey$marital`

D. Should be `table(survey$marital)`

ANSWER: D is correct. A is incorrect because `table()` does support 1-way tables, which are frequency distributions. B is incorrect because R is case-sensitive in most contexts and will not recognize the `table()` command if capitalized. C is incorrect because commands like table() require the argument in parentheses, whereas the double-colon is used in relation to packages.

4. What command in R is used most frequently to implement ordinary least squares (OLS) regression? Fill In: _____

ANSWER: The `lm()` command is for linear models. This is not the only R command which can implement OLS regression, but it is the most common.

5. Give a short answer: What is the "base::" part in the command below and what does it do?
`base::table(survey$marital,survey$sex)`

ANSWER: The "base::" part is the command prefix for the `table()` command, assuring that the command is the one in the "base" package and not a `table()` command from some other package.

6. Cronbach's alpha, a leading measure of reliability, is produced by which R package?

A. reliability

B. psych

C. alpha

D. alpha()

ANSWER: B. The psych package is an important statistical package which supports the `alpha()` command and many other statistical procedures. A is wrong because there is no package "reliability". There is a package "Reliability", but it is not used in the text and does not compute Cronbach's alpha b but rather serves specialized purposes in software reliability. C is incorrect because there is no "alpha" package. D is wrong because `alpha()` is a command, not a package.

7. Which was NOT used for cluster analysis in the text?

A. `clusplot()` from the "cluster" package

B. `cluster.stats()` from the "fpc" package

C. `NbClust()` from the "NbClust" package.

D. All of the above were used

ANSWER: D. All were used. Also used was `pmm()` from the "cluster" package.

8. The ANOVA family has four main variants: ANOVA, ANCOVA, MANOVA, and MANCOVA. In the text, which command(s) was/were used to implement them?

A. anova()

B. anova() and manova()

C. aov() and manova()

D. glm()

ANSWER: C. The `aov()` command from the "stats" package in R's System Library was used for ANOVA and ANCOVA and the `manova()` command, also from the "stats" package, was used for MANOVA and MANCOVA. Answer A is incorrect because there is no "anova" package (there is an "Anova" package but it was only mentioned, not used, in the text). Answer B is incorrect for the same reason. Answer D is in error also: the `glm()` command was used for logistic regression.

9. In logistic regression, what is a synonym for a "confusion table", created by the `confusionMatrix()` command?

A. a regression table

B. a logistic regression table

C. a table of logistic coefficients

D. a classification table

ANSWER: D is the correct answer. A "hit table" is another synonym. For binary logistic regression, this is a 2-b-y-2 table of predicted and observed values on the binary dependent variable. Percent accuracy is a metric based on a confusion table. Answer A is vague and is not a synonym. Answer B is also vague but could refer to the same thing as Answer C, which is not a synonym and is incorrect. The `confusionMatrix()` command is from the "caret" package.

10. The "processR" package implements which type of statistical analysis?

   A. mediation and moderation analysis, as per the work of Andrew Hayes

   B. correlation, as per the work of Francis Galton

   C. correlation, as per the work of Karl Pearson

   D. processR is not a statistical package

   ANSWER: A is correct. B and C refer to correlation, which is not the focus of processR. Francis Galton is credited with being the inventor of correlation, but the most widely used variant, Pearsonian correlation, was invented by Karl Pearson. Answer D is simply wrong.