PH1700 Homework 9

**Instructions**.

You may work in groups, but you must submit your own write up of the homework. Create a Word or PDF document with text as if you were writing up a statistical analysis report. Embed the figures and tables in your document. You can copy tables by selecting the table, right-click, and copy as a picture – then paste into Word. Similarly with graphs, click on the window with the graph, Ctrl-C (Command-C on Mac) to copy, then Ctrl-V (Command-V on Mac) to paste into Word. Important: it is unacceptable to just answer with a graph or a number. You have to write a sentence or two describing what you observe and your interpretation. Please include units (mm, inches, pounds, mm/L, etc.) when reporting any numbers. Be professional in your homework report. For questions involving hypothesis testing or confidence intervals, assume the 0.05 significance level, unless otherwise specified.

**For all homework problems, when performing any statistical test, please state the hypotheses, and report the test statistic, and report the p-value (or critical value, if requested) as part of performing and interpreting test.**

**Problems**

**12.44 from ROSNER:** Additionally, Assess the distribution of Maxfwt to verify that non-parametric methods are needed. (Discuss why or why not) regardless of your answer, use non-parametric methods. In practice, we would only perform testing for pairwise differences if the overall test is significant.

Environmental Health, Pediatrics Refer to Data Set LEAD.DAT at www.cengagebrain.com.

12.44 Use nonparametric methods to compare MAXFWT among the three exposure groups defined by the variable LEAD\_GRP.

**Required additional problems**

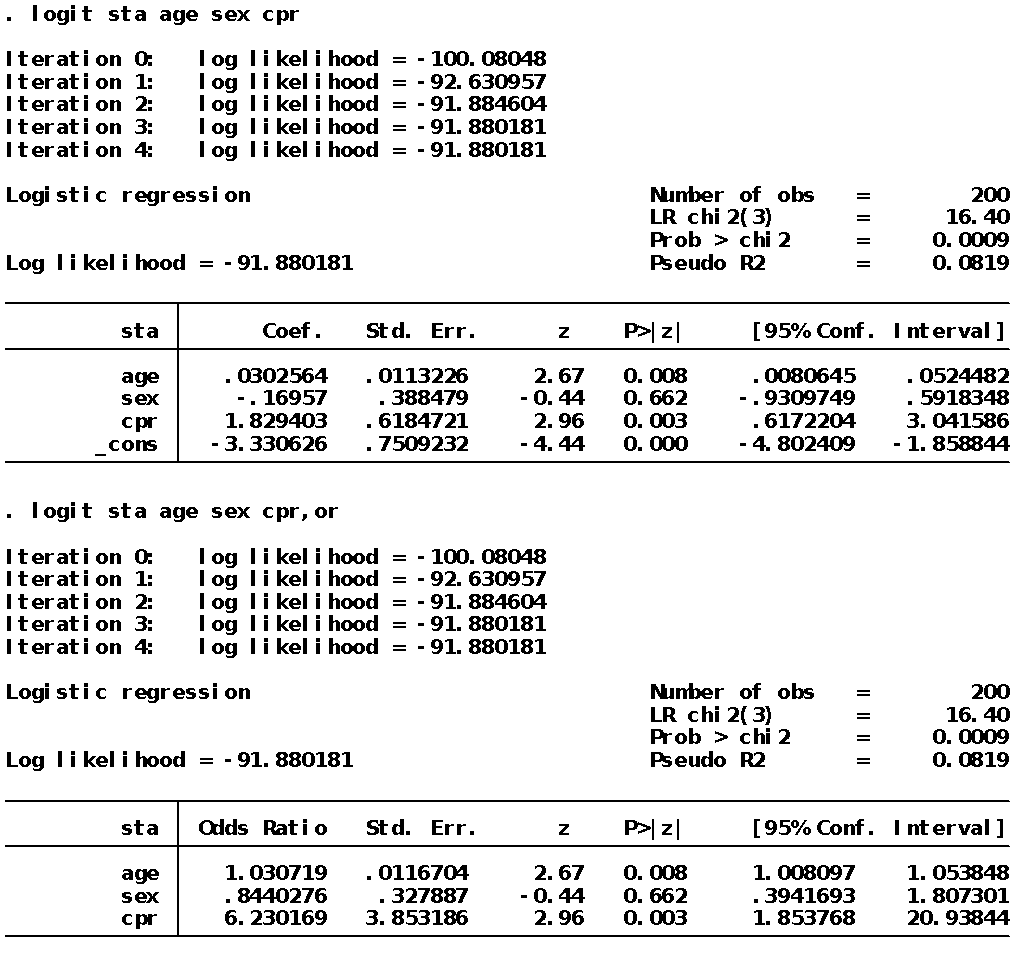
**1.** Using LEAD.DAT.dta,

1. Test whether there are any differences in full IQ score across the 3 exposure groups defined by lead\_grp (no exposure, past exposure and current exposure). For homework purposes, do so both parametrically (hint: refer to ANOVA from 1690 course to perform the parametric test or Rosner chapter 12; you can also refer to the ANOVA section from the STATA lab) and non-parametrically. State the hypothesis, report the test statistics and p value
2. Assess the distribution of full IQ scores and discuss which test is more valid and justify why.
3. Given your results to the overall test, is there sufficient evidence to perform further pairwise testing between the means (yes or no, and explain why or why not, in 1 or 2 sentences.) For each test performed, state the null and alternative hypotheses, report the test statistic, and p-value, and interpret your tests.

**2.** This problem uses the ICU data from Hosmer and Lemeshow, Applied logistic regression, 2nd edition, described on page 24-25 “The ICU study data set consists of a sample of 200 subjects who were part of a much larger study on survival of patient following admission to an adult intensive care unit (ICU). The major goal of this study was to develop a logistic regression model to predict the probability of survival to hospital discharge of these patients.”

We wish to assess the relationship between Surviving the ICU and age, sex, and CPR prior to entry into the ICU. Therefore, we fit a logistic regression model using vital status (STA) 0 = lived and 1 = died; age in years, sex (0=Male, 1= Female) and CPR prior to ICU entry (CPR, 0=no, 1=yes) The full data can be found in icu.csv and its data dictionary in icu.txt.

Use the following stata output to answer the following questions.



2.a) Write the appropriate logistic model for this analysis

2.b) Is the model significant? Report the hypothesis, appropriate test statistics and p-value.

2.c) Is age a significant risk factor for dying in the ICU? Report the hypothesis, appropriate test statistics and p-value. How would we interpret the coefficient for age? What could we have done to interpret the coefficient for age as an odds ratio?

2.d) Is sex a significant risk factor for dying in the ICU? Report the hypothesis, appropriate test statistics and p-value. Interpret its odds ratio appropriately, based on the significance.

2.e) Is CPR a significant risk factor for dying in the ICU? Report the hypothesis, appropriate test statistics and p-value. Interpret its odds ratio appropriately.