

# Biostat 510

## Homework 3

### Due Thursday, Feb 12, 2009

This homework uses the dataset, AllCases.sas7bdat, which was created for homework 2. You can use the file from my web page: [http://www.umich.edu/~kwelch510/2009/homework\\_2009.html](http://www.umich.edu/~kwelch510/2009/homework_2009.html), or you can use your own dataset. Please remember to report the test statistic, df, and p-value for any statistical tests. State the alpha level that you are using for statistical tests at the beginning of your homework write-up. If you use a different alpha level for a particular test, please indicate that in your write-up. Also, please report the results of your statistical tests in plain English, don't just say that there is a significant difference or not, explain the results. You should create all new variables in a data step, and you can do that in a single data step, at the start of your commands. Just go back and add new variables and then rerun the data step to re-create the new dataset, containing new variables.

1. Create a format to display the values of RAN, so that 1 = Ran and 2 = Not Ran. Use the format where appropriate in your SAS commands, so the values are displayed properly.
2. Carry out an **independent-samples t-test** to compare the mean of HR1 and HR2 for those who ran vs. those who did not run.
  - a) Write out the null and alternative hypotheses.
  - b) What do you conclude for HR1, for HR2?
  - c) Which t-test (equal or unequal variances) is appropriate for HR1 and for HR2 and why?
  - d) Include the output from these t-tests in your write-up.
3. **Paired-samples t-test**
  - a) Use a **paired-samples t-test** to compare the mean of HR1 and HR2 for all students. What do you conclude?
  - b) Carry out a **paired-samples t-test** to compare the mean of HR1 and HR2 separately for students who ran and those who did not run. What do you conclude for students who ran? For students who did not run? (You will have to sort your data first by RAN, and then carry out the paired-samples t-test by RAN).

```
proc sort data=b510.Allcases;
  by ran;
run;
proc ttest data=b510.Allcases;
  by ran;
  /*more SAS code*/
```

- c) Include the output from all of these paired-samples t-tests in your write-up.
4. Create a new variable called HRDIFF by subtracting HR1 from HR2.
    - a) Carry out an **independent-samples t-test** to compare the mean of HRDIFF for those who ran to the mean of HRDIFF for those who did not run.

- b) What do you conclude about the change of heart rate for each of these groups of students?
  - c) Include the output from this independent-samples t-test in your write-up.
5. Carry out a **one-sample t-test** to decide if the resting heartrate (HR1) of students is equal to 72 beats per minute.
- a) What do you conclude?
  - b) Include the output from this one-sample t-test in your write-up.
6. Create a **scatter plot**
- a) Make a **scatter plot** with HR1 as the X-variable and HR2 as the Y-variable, and include a loess fit in this scatter plot.
  - b) What is the nature of the relationship between HR1 and HR2?
  - c) Create another **scatter plot**, where you plot HR1 as X and HR2 as Y, but this time, panelby RAN (remember to use the formatted values for this variable). Include a loess fit in these plots also.
  - d) What does the relationship between these two variables look like for those who ran, vs. those who did not run?
  - e) Include all of these scatter plots in your write-up.
7. Fit a **linear regression model**, with HR2 as Y and HR1 as X.
- a) Write out this linear regression model.
  - b) How many observations are included in this regression model?
  - c) What is the R-square?
  - d) What is the parameter estimate for HR1? Please interpret it.
  - e) Check the residuals for this model, include a plot of studentized deleted residuals vs. predicted values.
    - Does there appear to be homoskedasticity for this model?
    - Are there any apparent outliers?
  - f) Create a histogram and q-qplot of the studentized deleted residuals.
    - Do the residuals appear to be normally distributed?
    - Comment on any departures from normality that you observe.
  - g) Include the output from the regression model, and the graphs in your write-up.
8. Fit a separate **linear regression model**, with HR2 as Y and HR1 as X for those who ran and those who did not run. Use a By statement for this analysis.
- a) How many observations are included in each model?
  - b) What is the R-square for each model?
  - c) What is the relationship between HR1 and HR2 for those who ran? For those who did not run? (Interpret the parameter estimate for HR1 for those who ran and those who did not run.)
  - d) Include the output from these regression models in your write-up.
9. Fit a **dummy variable regression model**, with X=dummy variable for RAN, and Y=HR2.
- a) Create a dummy variable for RAN, where 1=Ran and 0=didn't run. (You don't need to create any formats for this variable).
  - b) Include the dummy variable as a predictor in a model where Y=HR2.
  - c) How many observations are included in this model?
  - d) What is the model R-square?

- e) Interpret the t-test for your dummy variable that is in the Parameter Estimates table.
  - f) Compare the results for this dummy variable regression model with the results from the independent-sample t-test for HR2 from Question 2 above.
  - g) Include the output from this regression model in your write-up.
10. Save your SAS commands as homework3.sas. Re-run all of the commands in this command file and make sure there are no errors in your log.