

Biostat 510

Homework 7

Due Thursday, April 2, 2009

State the alpha level that you are use for statistical tests in this homework. Include the test statistic, df, and p-value for the results of any statistical tests that you report.

This homework uses the Afifi data (**afifi.dat**), which you can download from my web page. This dataset has 2 rows of data for each patient. The first row contains information collected upon admission to the hospital; the second row contains the same information collected just before the patient died or was discharged.

1. Use SPSS commands to create the Afifi dataset. Variable descriptions are shown below.
 - a) Create SHOCKDUM, with a value of 1 for patients in shock and 0 if not.
 - b) Create value labels for SEX and SHOKTYPE.

Variables	Name	Columns	Format	Description
1,22	ID	1-4	4.0	Id number
2,23	AGE	5-8	4.0	Age (years)
3,24	HEIGHT	9-12	4.0	Height (cm)
4,25	SEX	13-15	3.0	Sex (1=male, 2=female)
5,26	SURVIVE	16	1.0	Survival (1=lived, 3=died)
6,27	SHOKTYPE	17-20	4.0	Type of Shock (2=non-shock, 3=hypovolemic shock, 4=cardiogenic shock, 5=bacterial shock, 6=neurogenic shock, 7=other)
7,28	SBP1,SBP2	21-24	4.0	Systolic Blood Pressure (mm Hg)
8,29	MAP1, MAP2	25-28	4.0	Mean Arterial Pressure (mm Hg)
9,30	HR1,HR2	29-32	4.0	Heartrate (beats per minute)
10,31	DBP1,DBP2	33-36	4.0	Diastolic blood pressure (mm Hg)
11,32	CVP1,CVP2	37-40	4.1	Mean central venous BP (mm Hg)
12,33	BSA1,BSA2	41-44	4.2	Body surface area (m sq)
13,34	CI1,CI2	45-48	4.2	Cardiac index (l/min/min squared)
14,35	APP1,APP2	49-52	4.1	Appearance time (sec)
15,36	CT1,CT2	53-56	4.1	Mean circulation time (sec)
16,37	UR1, UR2	57-60	4.0	Urinary Output (ml/hr)
17,38	PL1,PL2	61-64	4.1	Plasma volume index (ml/kg)
18,39	RC1,RC2	65-68	4.1	Red cell index (ml/kg)
19,40	HGB1,HGB2	69-72	4.1	Hemoglobin (gm)
20,41	HCT1,HCT2	73-76	4.1	Hematocrit (%)
21,42	CARD1, CARD2	80	1.0	Card (1=initial, 2=final)

Sample SPSS commands to read in selected variables are shown below.

You should include *all* variables when you read in this dataset, not just those shown in the commands below. Modify these commands to read all variables into the afifi data set.

Note: the records=2 option is used to indicate that the data are present on *two* rows for each subject. The / in front of IDNUM and SBP2 indicates that SPSS is to go to a new line to begin to read those variables. These options are not used to read data with only one row per subject.

The number after the column location for a variable indicates the number of decimal places to use for that variable (e.g. HGB1 69-72 (1) means that Hemoglobin at time 1 is in columns 69-72, with 1 place after the decimal). Be sure to include the correct number of decimals for your variables. Check the data description to get the correct decimal values for each variable.

DATA LIST

```
FILE="c:\documents and settings\kwelch\desktop\b510\afifi.dat" RECORDS=2
/ IDNUM 1-4 AGE 5-8 SEX 13-15 SURVIVE 16 SHOKTYPE 17-20 SBP1 21-24 MAP1 25-28
  HGB1 69-72 (1)
/ SBP2 21-24 MAP2 25-28 HGB2 69-72 (1).
```

Execute.

2. Get a scatterplot with SBP2 as Y and SBP1 as X. Include a loess fit in the scatterplot.
 - a) Describe the relationship between SBP1 and SBP2. Does this relationship between SBP1 and SBP2 appear to be fairly linear?
 - b) Paste your SPSS commands for this and all other questions into a syntax window.
 - c) Include this scatterplot in your writeup.

3. Fit a linear regression model with SBP2 as the dependent variable and SBP1 as the independent variable.
 - a) What is the model R-square? How many observations are in your model?
 - b) Is there a significant linear relationship between SBP1 and SBP2? Interpret the coefficient for SBP1 in the fitted model.
 - c) Using the Plots... button in the Regression window, get a plot with the studentized-deleted residuals (SDRESID) as Y and the standardized predicted values (ZPRED) as X.
 - d) Using the Save... button, save the Unstandardized Residuals, Unstandardized Predicted Values, and the Studentized Deleted Residuals in your dataset.
 - i. After you run this model, go to the data window and check out the new variables that were created.
 - ii. What are the saved variables called?
 - iii. Check out the variable labels for these variables in the variable view window.
 - e) Use Descriptives...Explore to get information about the distribution of the studentized deleted residuals that were saved from the regression.
 - i. Include the test for normality, a histogram and a normal q-q plot for the studentized deleted residuals in your write-up (Please don't include all the output and the other plots that weren't requested in your write-up).

- ii. Discuss the distribution of the residuals. Do they appear to be relatively normal? Why or Why not?
 - f) Include the output for this linear regression model fit and only the requested plots in your write-up.
- 4. Create a correlation matrix for the variables SBP2, SBP1, AGE, DBP1, MAP1, CVP1, and BSA1. Use the Options...button and Exclude cases Listwise to get only those cases that are complete for all variables.
 - a) How many observations are included in this correlation matrix?
 - b) Include this correlation matrix in your write-up.
- 5. Create a scatterplot matrix with these same variables.
 - a) Include this scatterplot matrix in your write-up.
- 6. Fit a multiple regression model with SBP2 as the dependent variable and ALL of the predictors used in the correlation matrix above as independent variables. Request the collinearity diagnostics (from the Statistics...button). Don't save the residuals and predicted values for this model (you will need to go to the Save...window and unclick those options).
 - a) What is the sample size for this model? What is the R-square for this model?
 - b) What is the overall significance of this model? (Report the F-test for the model).
 - c) Which of the predictors are significant in this model?
 - d) What is the parameter estimate for SBP1 in this model? Is it significant?
 - e) Which predictors appear to be collinear based on the VIF? Explain.
 - f) What is the condition index for this model? Which variables appear to be collinear, based on the collinearity diagnostics? Explain.
 - g) **Include the tables for the Model Summary, ANOVA, Coefficients, and the Collinearity Diagnostics only from this regression model in your write-up.**
- 7. Fit a new multiple regression model, but this time exclude DBP1 and MAP1 as predictors. Check for collinearity in this new model. Don't save the residuals and predicted values for this model.
 - a) What is the R-square for this model? Is this model significant overall? Report the F-test for this model.
 - b) What is the coefficient for SPB1 in this model? Is it significant?
 - c) Does there appear to be any collinearity in this model? Why or why not?
 - d) **Include the tables for the Model Summary, ANOVA, Coefficients, and the Collinearity Diagnostics only from this regression model in your write-up.**
- 8. Create a boxplot for SBP2, using SHOKTYPE as the categories.
 - a) Which category of SHOKTYPE has the highest level of SBP2?
 - b) Include the boxplot in your write-up.
- 9. Fit a linear regression model with SBP2 as the outcome (dependent) variable using dummy variables for each level of SHOKTYPE as the independent variables.