SPSS Lab 2

Roc Curves, Distributions and More Descriptive Statistics

Demo 2

Do a ROC curve analysis for the following data:

testvar	case	count
1	1	3
1	0	33
2	1	2
2	0	6
3	1	2
3	0	6
4	1	11
4	0	11
5	1	33
5	0	2

- Go to menu data\weight cases\weight cases by \Frequency variable: count
- Go to Analyze\ROC Curve\test variable:testvar and state variable:case and value of state variable =1 (because 1 is for disease present)
    - Check all checkboxes
- Read the result tables and write down your observations. Write these observations in the body of the email that you will send me.
- Double Click on the ROC curve graph to open it in the chart editor and change the color of the curve to red.
    - Double click on the curve (several times until it is highlighted), then right click and select "properties window." On tab "Lines" select the red color and click apply.
    - Also change the background of the curve area from grey to white: right click on the grey area and select properties. On the fill and border tab choose the white color, click apply.
    - Close Chart Editor window
- Double click on the table that says "Area under the Curve" to open the Pivot table editor
    - Right click in the field that contains the AUC value and select cell properties.
    - Go to the "Shading" tab and select a yellow background color.

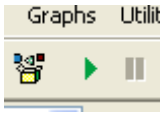
Simulations

Do SPPS calculations of Binomial random variable simulation for problems in book: 4.95

1. Problem 4.98 asks to draw 200 random samples from a binomial distribution each based on 100 trials with probability of success =0.01 on each trial.
   a. Draw on a piece of paper your guess of how this data will be distributed.
   b. Then, obtain a frequency distribution over the 200 random samples and plot the distribution
   c. Use a script called "Create empty data with n cases.SBS" to populate a spreadsheet with 200 cases numbered from 1 to 200:
      i. Go to File\New \Data

      ii. Goto to File\open\script. Select the file "Create empty data file with n cases.SBS"

      iii. Click the "run" button, the green sideways triangle:



      iv. You will be asked how many cases you want, enter 200.

   d. Now create a field with random samples from a binomial distribution. Go to transform\compute variable. In Target Variable, write probBinom. In the Function group list select "Random Numbers" In "Function and Special Variables" Select "Rv.Binom" In the first question mark"?" Type in the number of trials as 100, replace the second question mark"?" with the probability 0.01. Click OK.

   e. Run Analyze\Descriptive Statistics\Frequencies\ to obtain statistics and a histogram.
   f. How do the results compare to your guess? Redo the simulation for p=0.99. Describe your results!

Descriptive Statistics

1. Open file hospital_ch2.sav.

2. Create a new field called stay_cat (to summarize stay categories) that has the values 1: for 5 or less days in the hospital, 2: for 5<days≤10, 3: 10<days≤15, 4: 15<days. *Example for storing value 1 in stay_cat, when dur_stay<=5:*

   a. Go to Transform\Compute Variables. In Target variable type in the new field name "stay_cat"

   b. In Numeric Expression type in 1

   c. Click the button "If". Select "Include if case satisfies condition." In the white space type in dur_stay<=5. Instead of just typing you may click on the dur_stay field on the left and then click on the "<=" button in the center section of the window. Click Continue. Click Ok.

3. Let's compute all the descriptive measures for field temperature (temp)

    a. Click on Analyze\Descriptive Statistics\Frequencies…

    b. Select temp as the variable.

    c. Select button " Statistics..." Select all checkboxes, except "Values are group midpoints". In Percentiles type in 95, then click Add. Click Continue. Be aware that Skewness is a measure of symmetry (A distribution, or data set, is symmetric if it looks the same to the left and right of the center point). Skewness of the normal distribution is 0. Positive skewness indicates a long right tail. Negative Skewness indicates a long left tail. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. Kurtosis of the normal distribution is 3. Positive Excess Kurtosis indicates flatness. Negative Excess Kurtosis indicates peakedness.

    d. Click the "Charts…" button. Click Histogram with Normal curve. Click Continue, then click Ok.

Lab Assignment 2

This time, you will have to answer some questions and write the answers in the body of the email that you will send me. Send me your files (data and output file) as an attachment to your email.

Descriptive Statistics

After executing the above described steps in SPSS, continue with the following steps:

    e. Briefly describe the results. Write your observations in the body of the email.

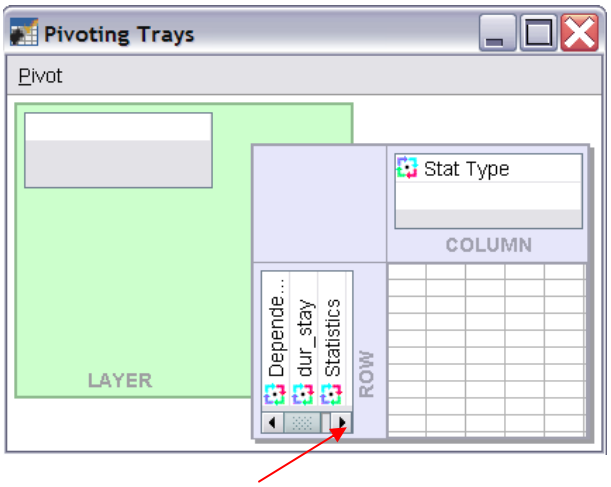
4. Let's compute summary statistics for groups of cases. For example let's compute the temperature dependant on the number of days a patient stayed in the hospital

    a. Click on Analyze\Descriptive Statistics\Explore. Add "temp" to the Dependant List, add Dur_stay to the Factor List.

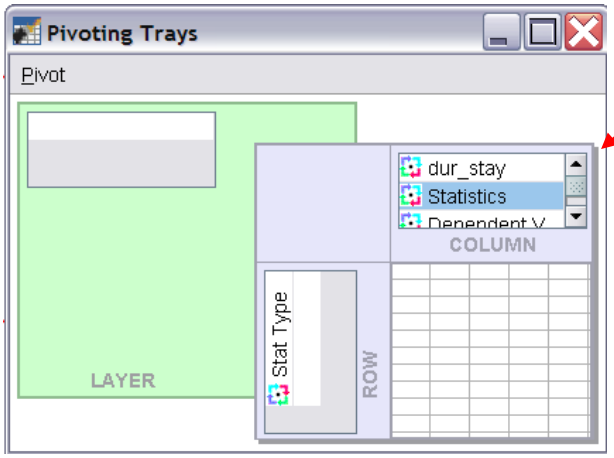
    b. Click the button "Statistics…" and select all checkboxes. Select the button "Plots" and select all checkboxes.

    c. You will get some warnings, read them to understand, but you can ignore them. They say that under certain duration stays the temperature was the same (constant).

    d. Click on the table called "Descriptives" in the Pivot. The Pivot table editor will open up. Go to menu Pivot\Pivoting Trays, you will see a window like this:



e. Drag the Row "statistics, Dependant, and dur_stay" to the top Column table and "Stat Type" to the Row table. After you dragged them it will look like this:



f. Close that window.

g. Study the descriptive table and write down your observations in the body of the email you will send me.

h. Scroll down and look at the percentiles table

i. Scroll down to the boxplots and write down your observations.

Descriptive Statistics 2

1. Do the same as in the above "Descriptive Statistics 1" section but use the field you created stay_cat instead of dur_stay.

2. Did you get the warning again? Why not? Write answer in body of email

3. Compare these boxplots with the boxplots you analyzed before. What trend do you see in the temperatures for patients who stayed for a short time vs. those who stayed longer? How can you explain this? Write your answers in the body of the email.

Simulations

1. Do Problem 4.99 from the book: Draw 200 random samples from a Poisson distribution each with mean=1. Obtain a frequency distribution of the number of success over 200 random samples and plot the distribution.

    a. Run the script to generate 200 cases as described in the Demo section

    b. Now create a field with random samples from a Poisson distribution. Go to transform\compute variable. In Target Variable, write probPoisson. In the Function group list select “Random Numbers” In “Function and Special Variables” Select “Rv.Poisson” In the first question mark”?” Type in the mean. (The Poisson distribution depends only on one parameter: the mean). Click Ok.

    c. Run a descriptive analysis like described before.

2. Problem 4.100: Compare your results from Problem 4.98 and 4.99. Do you feel that the Poisson approximation to the binomial is adequate in this case? Write your answer in the email that you will send me.