SAS PROGRAMMING FOR BUSINESS ANALYTICS

ASSIGNMENT 2

SUBMITTED BY:

TANAY BHALERAO

U47707491

FEBRUARY 20, 2015

UNIVERSITY OF SOUTH FLORIDA

Management Information Systems

Homework 2

1. Refer to DOGS1 data. The investigator may want to ensure that the dogs allocated to each treatment group were of similar compositions with respect to gender and hair coat. Use PROC FREQ to conduct Fisher's exact test to see if the concentration of the drug received was statistically independent of the gender of the dog. Likewise, see if the length of the coat and the drug treatment were statistically independent with Fisher's exact test. Write your interpretation of the results of these tests.

```
DATA infile_dogs1;
      infile
"\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\dogs1.txt" LRECL=
200 firstobs=2;
      input
      dog $ 1-8 concent 16 sex $ 17 age 31-32 haircoat $ 33-37 weight 45-48;
RUN;

PROC FREQ DATA=infile_dogs1;
      TITLE "FISHER'S EXACT TEST";
      TABLES (haircoat sex)*concent /EXACT;
      RUN;
```

FISHER'S EXACT TEST

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of haircoat by concent			
	concent			
haircoat	0	1	2	Total
Med	2 8.00 20.00 20.00	3 12.00 30.00 42.86	5 20.00 50.00 62.50	10 40.00
Short	8 32.00 53.33 80.00	4 16.00 26.67 57.14	3 12.00 20.00 37.50	15 60.00
Total	10 40.00	7 28.00	8 32.00	25 100.00

Statistics for Table of haircoat by concent

Statistic	DF	Value	Prob
Chi-Square	2	3.3780	0.1847
Likelihood Ratio Chi-Square	2	3.4968	0.1741
Mantel-Haenszel Chi-Square	1	3.2377	0.0720
Phi Coefficient		0.3676	
Contingency Coefficient		0.3450	
Cramer's V		0.3676	

WARNING: 83% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test	
Table Probability (P)	0.0270
Pr <= P	0.2361

Sample Size = 25

Frequency Percent Row Pct Col Pct	Table of sex by concent			
	concent			
sex	0	1	2	Total
F	7 28.00 46.67 70.00	3 12.00 20.00 42.86	5 20.00 33.33 62.50	15 60.00
M	3 12.00 30.00 30.00	4 16.00 40.00 57.14	3 12.00 30.00 37.50	10 40.00
Total	10 40.00	7 28.00	8 32.00	25 100.00

Statistics for Table of sex by concent

Statistic	DF	Value	Prob
Chi-Square	2	1.2946	0.5234
Likelihood Ratio Chi-Square	2	1.2876	0.5253
Mantel-Haenszel Chi-Square	1	0.1435	0.7048
Phi Coefficient		0.2276	
Contingency Coefficient		0.2219	
Cramer's V		0.2276	

WARNING: 83% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test	
Table Probability (P)	0.0720
Pr <= P	0.5235

Sample Size = 25

Interpretation:

The size of the dataset is small. Using Fisher's Test for independence we are figuring out if the proportion of one nominal variable is different depending on the value of the other variable.

After carrying out the Fisher's Exact Test we can infer the following:

- The length of the coat and the drug treatment were found to be statistically **dependent** since P-value is less than 0.05 **(P-Value: 0.0270).** This is less than 5% and hence the null hypothesis can be rejected and the test turned out to be significant.
- Also the concentration of the drug received was found to be statistically **independent** from the sex of the dog since P-value in this case is more than 0.05**(P-Value: 0.0720)** i.e. more than 5%. In such a situation we do not reject the null hypothesis since the test is insignificant.

2. Refer to the following data, obtained from Michael Radelet, in the book Categorical Data Analysis by Alan Agresti. The data describe the circumstances of 326 homicide cases in Florida from 1976-1977.

Defendant's race	Victim's race	Death penalty	Count
Black	White	Yes	11
Black	White	No	52
Black	Black	Yes	6
Black	Black	No	97
White	White	Yes	19
White	White	No	132
White	Black	Yes	0
White	Black	No	9

Use PROC FREQ to create appropriate tables to answer the following questions. Write down your answers to these questions, and mark (circle or highlight) those same numbers on your SAS output. (In other words, make sure that your tables explicitly show the requested percentages.)

a. When the defendant was white and the victim was black, in what percentage of cases was the death penalty verdict given?
b. When the defendant was black and the victim was white, in what percentage of cases was the death penalty verdict given?
c. When the races of the victim and the defendant were the same, in what percentage of cases was the death penalty given?

```
DATA Homicide;
      infile
"\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\Homicide.txt"
LRECL= 200 DLM="09"x firstobs=2;
      input
      Def_race $ Vic_race $ Death_Penalty $ Count;
RUN;

PROC FREQ DATA=Homicide;
      TITLE "HOMICIDE-DEATH PENALTY PERCENTAGES";
      TABLES Def_race*Vic_race*Death_Penalty;
      TABLES Death_Penalty /NOCUM;
      WEIGHT Count;
      RUN;
```

a) In what percentage of cases was the death penalty verdict given?

11.04%

Death_Penalty	Frequency	Percent
No	290	88.96
Yes	36	11.04

b) When the defendant was white and the victim was black, in what percentage of cases was the death penalty verdict given?

0.0%

Frequency
Percent
Row Pct
Col Pct

Table 2 of Vic_race by Death_Penalty

Controlling for Def_race=White

Vic_race	Death_Penalty: No	Death_Penalty: Yes	Total
Black	9 5.63 100.00 6.38	0 0.00 0.00 0.00	9 5.63
White	132 82.50 87.42 93.62	19 11.88 12.58 100.00	151 94.38
Total	141 88.13	19 11.88	160 100.00

c) When the defendant was black and the victim was white, in what percentage of cases was the death penalty verdict given?

17.46%

HOMICIDE-DEATH PENALTY PERCENTAGES

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table 1 of Vic_race by Death_Penalty Controlling for Def_race=Black		
	Death_Penalty		
Vic_race	No	Yes	Total
Black	97 58.43 94.17 65.10	6 3.61 5.83 35.29	103 62.05
White	52 31.33 82.54 34.90	11 6.63 17.46 64.71	63 37.95
Total	149 89.76	17 10.24	166 100.00

d) When the races of the victim and the defendant were the same, in what percentage of cases was the death penalty given?

9.84%

```
PROC FREQ DATA=Homicide;
      TITLE "HOMICIDE-DEATH PENALTY PERCENTAGES";
      TABLES Death_Penalty/NOCUM;
      WHERE Def_race=Vic_race;
      WEIGHT Count;
         RUN;
```

HOMICIDE-DEATH PENALTY PERCENTAGES

The FREQ Procedure

Death_Penalty	Frequency	Percent
No	229	90.16
Yes	25	9.84

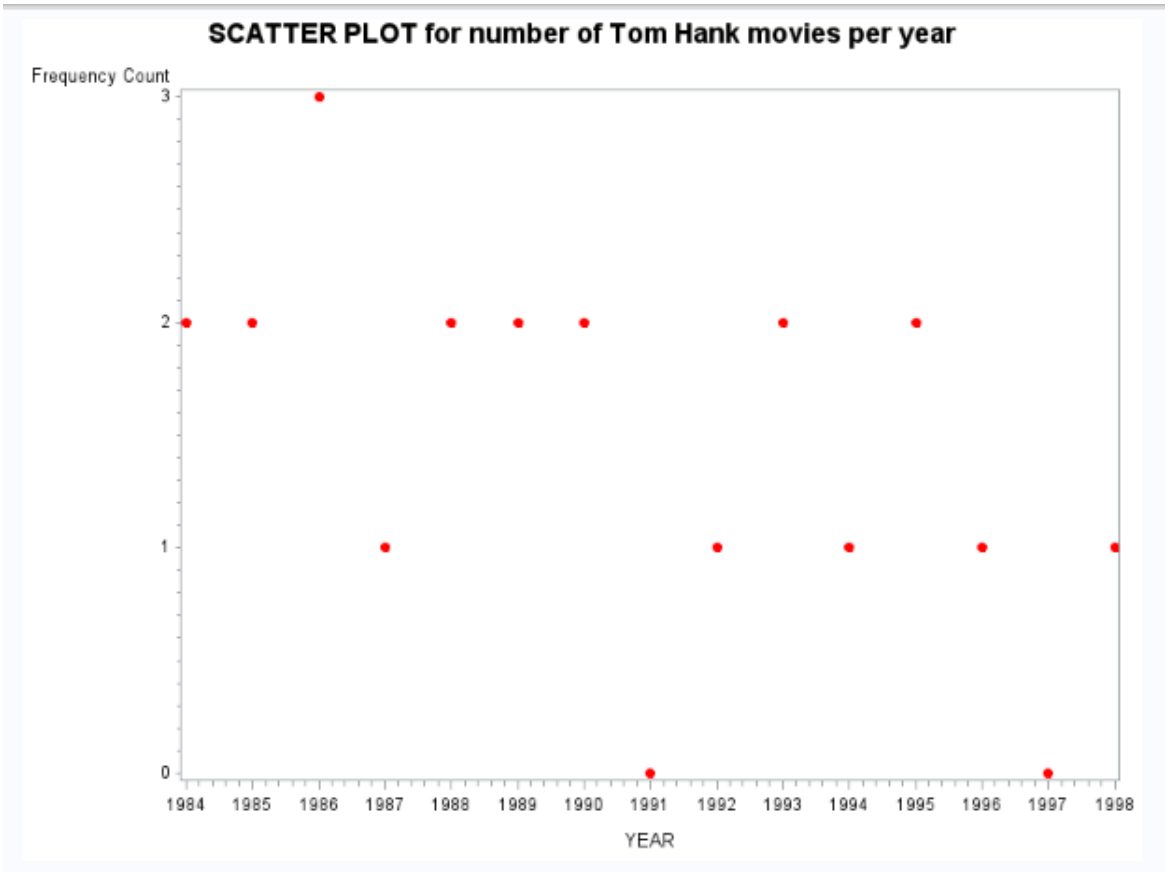
3. Refer to the HANKS data. Your task is to use a SAS program to count the number of movies in which Tom Hanks appeared in each of the years from 1984-1998, then make a scatterplot with the number of movies made in each year on the vertical axis versus the year on the horizontal axis. In some years, such as 1997, he did not appear in any movies. On the scatterplot, indicate those years by plotting a point at zero. To do this you could create another dataset with all of the years from 1984-1998 by using a Do loop, then MERGE that dataset with the dataset containing the movie counts. Then, you will need to replace missing values for movie counts with zeroes.

```
DATA Hanks;
      infile
"\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\Hanks.txt" LRECL=
200 firstobs = 2;
      Input TITLE $ 1-25 YEAR 26-30 LENGTH 34-37 MPAA $ 42-47 ACTION 50-52
DRAMA 58-60 HUMOR 66-68 SEX 74 VIOLENCE 81-83 SUSPENSE 90 OFFBEAT 98;
RUN;

PROC FREQ DATA=Hanks;
      TABLES YEAR /NOCUM NOPERCENT OUT=Hanks_dataset1;
Run;
DATA Hanks_dataset2;
      DO YEAR = 1984 to 1998;
      OUTPUT;
      END;
RUN;

DATA Hanks_total_data;
      MERGE Hanks_dataset1 Hanks_dataset2;
      BY YEAR;
      IF Count=. THEN count=0;
RUN;

SYMBOL VALUE=dot COLOR=red;
PROC GPLOT DATA=Hanks_total_data;
      PLOT count*YEAR;
      TITLE "SCATTER PLOT for number of Tom Hank movies per year";
RUN;
```



4. Book page 117 Questions 3.10, 3.16, 3.18

3.10)

```
DATA Carotid_Sten;
      INPUT Invasive_M1 $ Non_invasive_M2 $ Count;
DATALINES;
Occuled Occuled 15
Occuled Non-Occuled 8
Non-Occuled Occuled 10
Non-Occuled Non-Occuled 67
;
RUN;

PROC FREQ DATA = Carotid_Sten;
      TITLE "Kappa Coefficient - Carotid Stenosis";
      TABLES Invasive_M1*Non_invasive_M2 /AGREE;
      WEIGHT Count;
RUN;
```

Statistics for Table of Invasive_M1 by Non_invasive_M2

McNemar's Test	
Statistic (S)	0.2222
DF	1
Pr > S	0.6374

Simple Kappa Coefficient	
Kappa	0.5068
ASE	0.1008
95% Lower Conf Limit	0.3092
95% Upper Conf Limit	0.7045

Sample Size = 100

3.16)

```
DATA Clinical_trial;
   INPUT TABLET_GR $ OUTCOME $ COUNT_HEART_ATTACK COUNT_STROKE;
DATALINES;
ASPIRIN 1-YES 80 65
ASPIRIN 2-NO 920 935
PLACEBO 1-YES 240 165
PLACEBO 2-NO 1760 1835
;
PROC FREQ DATA=Clinical_trial;
   TITLE "Problem Clinical_trial Data-MI";
```

```
   TABLES TABLET_GR*OUTCOME /CMH;
   WEIGHT COUNT_HEART_ATTACK;
RUN;
PROC FREQ DATA=Clinical_trial;
   TITLE "Problem Clinical_trial Data-Stroke";
   TABLES TABLET_GR*OUTCOME /CMH;
   WEIGHT COUNT_STROKE;
RUN;
```

Problem Clinical_trial Data-MI

Common Odds Ratio and Relative Risks				
Statistic	**Method**	**Value**	**95% Confidence Limits**	
Odds Ratio	Mantel-Haenszel	0.6377	0.4891	0.8314
	Logit	0.6377	0.4891	0.8314
Relative Risk (Column 1)	Mantel-Haenszel	0.6667	0.5237	0.8487
	Logit	0.6667	0.5237	0.8487
Relative Risk (Column 2)	Mantel-Haenszel	1.0455	1.0202	1.0713
	Logit	1.0455	1.0202	1.0713

Total Sample Size = 3000

Problem Clinical_trial Data-Stroke

Common Odds Ratio and Relative Risks				
Statistic	**Method**	**Value**	**95% Confidence Limits**	
Odds Ratio	Mantel-Haenszel	0.7731	0.5741	1.0411
	Logit	0.7731	0.5741	1.0411
Relative Risk (Column 1)	Mantel-Haenszel	0.7879	0.5974	1.0391
	Logit	0.7879	0.5974	1.0391
Relative Risk (Column 2)	Mantel-Haenszel	1.0191	0.9979	1.0407
	Logit	1.0191	0.9979	1.0407

Total Sample Size = 3000

The Relative Risk of Aspirin with respect to heart attack is 0.667 while with respect to stroke is 0.7879. Both of which are statistically insignificant.

3.18)

```
Data MAG_SUR;
      INPUT STUDY $ OUTPUT $ DRUG_TYPE$ COUNT;
Datalines;
one survived mgso 20
one survived placebo 25
one died mgso 100
one died placebo 155
two survived mgso 25
two survived placebo 21
two died mgso 150
two died placebo 150
three survived mgso 30
three survived placebo 28
three died mgso 200
three died placebo 240
;
PROC FREQ DATA = MAG_SUR;
      TITLE "Conducting meta-analysis to determine effect of MGSO4";
      TABLES STUDY*DRUG_TYPE*OUTPUT /ALL;
      WEIGHT COUNT;
RUN;
```

Common Odds Ratio and Relative Risks

Statistic	Method	Value	95% Confidence Limits	
Odds Ratio	Mantel-Haenszel	0.8050	0.5695	1.1379
	Logit	0.8050	0.5695	1.1378
Relative Risk (Column 1)	Mantel-Haenszel	0.9721	0.9288	1.0174
	Logit	0.9722	0.9291	1.0172
Relative Risk (Column 2)	Mantel-Haenszel	1.2070	0.8942	1.6294
	Logit	1.2068	0.8940	1.6289

Breslow-Day Test for Homogeneity of the Odds Ratios	
Chi-Square	0.0331
DF	2
Pr > ChiSq	0.9836

Total Sample Size = 1144

The Overall Relative Risk is 0.9721 with lower 95% confidence Interval: 0.9288 and higher 95% confidence Interval: 1.01174. Also the Breslow Day Statistic for Homogeneity of the odds is 0.9836