Execution Environment

Code: Homework 6.sas

```sas
* Programmed by Charles Hwang  *
* Coded in SAS OnDemand        *
* Wednesday, November 11, 2020 *
* Course: STAT 403             *
* Title: Homework 6            *;

/* 1 */ Data Drug;
Input ID$ 1-3 GENDER$ 4 SES$ 5 DRUG$ 6 AGE 7-8; * It appears holding the line does not work here ;
Datalines;
0011LB15
0022MZ35
0032HF76
0041Lc21
0052H.58
0062LG47
0072LD23
0081ME51
0091HA32
0101Hg19
0111.b21
0122L?38
0131MF26
0141Hd36
0152Ma25
0161LD32
0172HA54
0181LG78
0192MG54
0202LH47
0212HC49
0221HB38
0232MB56
0241HA29
0252LD77
0261MG27
0272LH36
0281HB19
0292MA28
0301HB54
;
/* 1a */ Proc Format;
Value $Gender '1'='Male' '2'='Female' Other='Missing';
Value $SES 'L'='Low' 'M'='Medium' 'H'='High' Other='Missing';
Value Age 0-20='0-20' 21-40='21-40' 41-99='41+' Other='Missing';
Run;

/* 1b */ Data Drug;
Infile "/home/chwang10/drugdata.txt" firstobs=2; * Skipping header row, starting on row 2 ;
Input ID$ 1-3 GENDER$ 4 SES$ 5 DRUG$ 6 AGE 7-8;
/* 1b(i) */ Format GENDER $Gender. SES $SES.;
/* 1b(ii) */ Label SES="Socioeconomic Status" DRUG="Drug Group" AGE="Age of subject";
/* 1b(iii) */ If UPCASE(DRUG)='A' or UPCASE(DRUG)='B' or UPCASE(DRUG)='C' or UPCASE(DRUG)='F' then COST='Generic';
else if UPCASE(DRUG)='D' or UPCASE(DRUG)='E' or UPCASE(DRUG)='G' or UPCASE(DRUG)='H' then COST='Premium';
else COST='Miscode'; * The DRUG variable seems more likely to be miscoded than missing ;
Run;
```

```sas
/* 1c */ Proc Print data=Drug;
Title "1c. Drug Data Set with Labels";
Label SES="Socioeconomic Status" DRUG="Drug Group" AGE="Age of subject"; * not working? ;
Run;

/* 1d */ Proc Freq data=Drug;
Title "1d. Frequencies of Socioeconomic Status and Cost, Grouped by Age";
Format AGE Age.;
Table AGE*SES*COST;
Run;

/* 1e */ Proc Gchart data=Drug;
Title "1e. Bar Chart of Cost, Grouped by Socioeconomic Status";
Vbar COST /group=SES discrete;
Run; * There is a clear distinction between those with high and low socioeconomic statuses. Those with
high socioeconomic statuses tend to be associated with the generic drug, while those with low
socioeconomic statuses tend to be associated with the premium drug. ;

/* 1f */ * H0: There is no relationship between cost and socioeconomic status.
HA: There is a relationship between cost and socioeconomic status. ;
Proc Freq data=Drug;
Title "1f. Chi-Squared Test of Independence on Cost and Socioeconomic Status";
Table COST*SES /chisq;
Run; * We fail to reject H0 at the α = .05 level. There is insufficient
evidence (χ = 6.5164, p = 0.1638) that there is a relationship between cost and socioeconomic status.
These results do not support the conclusion drawn from the bar chart because of the small sample sizes
in all but one of the cells (n < 5), and we should exercise caution with using this test as the sample
size assumption of the chi-squared test is violated. ;

/* 2 */ Data CrimeDrinker;
Length Alcoholic$ 10; * Extending character length to properly display "Nondrinker" ;
Input Crime$ Alcoholic$ N @@;
Datalines;
Arson Drinker 50 Rape Drinker 88 Violence Drinker 155 Stealing Drinker 379 Coining Drinker 18 Fraud Drinker 63 Arson Nondrinke
;
/* 2(i) */ Proc Freq data=CrimeDrinker;
Title "2(i). Chi-Squared Test of Independence on Alcoholism and Crime";
Table Alcoholic*Crime /chisq deviation expected norow nocol nopercent;
Weight N;
Run;

/* 2(ii) */ * H0: There is no relationship between alcoholism and crime.
HA: There is a relationship between alcoholism and crime. ;

/* 2(iii) */ * We reject H0 at the α = .05 level. There is sufficient
evidence (χ = 49.7306, p < 0.0001) that there is a relationship between alcoholism and crime. ;

/* 2(iv) */ * It appears that drinkers are generally more likely to commit crimes, as the data show
they commit approximately 11.887073 percent more than nondrinkers. However, fraud is disproportionately
committed by nondrinkers, with only 30.4347826 percent of fraud is committed by drinkers—the only crime
with such a proportion below 53 percent. If fraud were removed from the data, drinkers
would (coincidentially) commit approximately 30.4347826 percent more crimes than nondrinkers. ;

/* 2(v) */ * See Problem 2(i) ;

/* 3a */ Data HA;
Length Race$ 9; * Extending character length to properly display "Caucasian" ;
/* 3a(i) */ Input Race$ Survival$ Percent @@;
Datalines;
Black Died 99 Black Survived 1 Caucasian Died 97.5 Caucasian Survived 2.5
;
/* 3a(ii) */ Data HAC;
Set HA;
Count1=400*Percent;
Count2=1000*Percent;
Run;

/* 3b */ %Macro CS(dataname=, r=, c=, options=, weight=);
Proc Freq data=&dataname;
Title "3c. Chi-Squared Test of Independence on Race and Survival Rate of Heart Attack";
Table &r*&c /&options;
Weight &weight;
Run;
%Mend CS;

/* 3c(i) */ * H0: There is no relationship between race and the survival rate of a heart attack.
HA: There is a relationship between race and the survival rate of a heart attack. ;
```

```sas
/* 3c(ii) */ %CS(dataname=HAC, r=Race, c=Survival, options=chisq deviation expected norow nocol, weight=Count1);
/* 3c(iii) */ * We reject H0 at the α = .05 level. There is sufficient evidence (p < 0.0001) that there
is a relationship between race and the survival rate of a heart attack. ;
/* 3c(iv) */ * See Problem 3c(ii) ;
/* 3c(v) */ %CS(dataname=HAC, r=Race, c=Survival, options=chisq relrisk, weight=Count2);
/* 3c(vi) */ * We reject H0 at the α = .05 level. There is sufficient evidence (p < 0.0001) that there
is a relationship between race and the survival rate of a heart attack. ;
/* 3c(vii) */ * Relative risk ratio: 1.0154 (95 percent confidence interval: (1.0142,1.0166)) ;
* African-Americans are approximately 1.54 percent more likely than Caucasians to die from a heart
attack, and we are 95 percent confident this percentage is between 1.42 and 1.66. Additionally, the
relative risk ratio is significant at the α = .05 level because its 95 percent confidence interval does
not include 1. ;
/* 3c(viii) */ * We get the same p-value (p < 0.0001) and conclusion for both sample sizes, although it
is worth noting that the test statistic was much greater for Count2 (χ = 654.3075 vs. χ = 261.7230). It
is not feasible to perform an accurate chi-squared test of independence using percentages, as not
having the correct sample sizes will make it unable to verify whether the sample size assumption is
violated. Also, because the test statistic varies depending on the sample size, inputting an arbitrary
sample size by multiplying the percentage by some constant may lead to the incorrect conclusion. ;
/* 3c(ix) */ * See Problem 3c(v) ;

/* 4a */ * H0: Test-driving a car does not change one's opinion of it.
HA: Test-driving a car does change one's opinion of it. ;

/* 4b */ Data Car;
Input Pre$ Post$ N @@;
Datalines;
Yes Yes 75 Yes No 14 No Yes 26 No No 83
;
Proc Freq data=Car;
Title "4c. McNemar's Test on Pre- and Post-Test Drive Opinion of Car";
Table Pre*Post /agree noprint;
Weight N;
Run; * We fail to reject H0 at the α = .05 level. There is insufficient evidence (χ = 3.60, p = 0.0578)
that test-driving a car changes one's opinion of it. ;

/* 4c */ * See Problem 4b :
```

Log: Homework 6.sas

Notes (40)

```
1           OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
70
71          * Programmed by Charles Hwang  *
72          * Coded in SAS OnDemand         *
73          * Wednesday, November 11, 2020 *
74          * Course: STAT 403              *
75          * Title: Homework 6            *;
76
77          /* 1 */
77      !          Data Drug;
78          Input ID$ 1-3 GENDER$ 4 SES$ 5 DRUG$ 6 AGE 7-8; * It appears holding the line does not work here ;
79          Datalines;

NOTE: The data set WORK.DRUG has 30 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time            0.00 seconds
      user cpu time        0.00 seconds
      system cpu time      0.00 seconds
      memory               670.03k
      OS Memory            30376.00k
      Timestamp            11/12/2020 09:40:16 AM
      Step Count                        50  Switch Count  2
      Page Faults                       0
      Page Reclaims                     158
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272


110         ;
111         /* 1a */
111     !          Proc Format;
112         Value $Gender '1'='Male' '2'='Female' Other='Missing';
NOTE: Format $GENDER has been output.
113         Value $SES 'L'='Low' 'M'='Medium' 'H'='High' Other='Missing';
NOTE: Format $SES has been output.
114         Value Age 0-20='0-20' 21-40='21-40' 41-99='41+' Other='Missing';
NOTE: Format AGE has been output.
115         Run;
```

```
NOTE: PROCEDURE FORMAT used (Total process time):
      real time             0.00 seconds
      user cpu time         0.00 seconds
      system cpu time       0.01 seconds
      memory                293.68k
      OS Memory             30116.00k
      Timestamp             11/12/2020 09:40:16 AM
      Step Count                        51  Switch Count  2
      Page Faults                       0
      Page Reclaims                     58
      Page Swaps                        0
      Voluntary Context Switches        13
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           56


116
117        /* 1b */
117      !          Data Drug;
118        Infile "/home/chwang10/drugdata.txt" firstobs=2; * Skipping header row, starting on row 2 ;
119        Input ID$ 1-3 GENDER$ 4 SES$ 5 DRUG$ 6 AGE 7-8;
120        /* 1b(i) */ Format GENDER $Gender. SES $SES.;
121        /* 1b(ii) */ Label SES="Socioeconomic Status" DRUG="Drug Group" AGE="Age of subject";
122        /* 1b(iii) */ If UPCASE(DRUG)='A' or UPCASE(DRUG)='B' or UPCASE(DRUG)='C' or UPCASE(DRUG)='F' then COST='Generic';
123        else if UPCASE(DRUG)='D' or UPCASE(DRUG)='E' or UPCASE(DRUG)='G' or UPCASE(DRUG)='H' then COST='Premium';
124        else COST='Miscode'; * The DRUG variable seems more likely to be miscoded than missing ;
125        Run;

NOTE: The infile "/home/chwang10/drugdata.txt" is:
      Filename=/home/chwang10/drugdata.txt,
      Owner Name=chwang10,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=11Nov2020:19:47:07,
      File Size (bytes)=351

NOTE: LOST CARD.
ID=  GENDER=Missing SES=Missing DRUG=  AGE=. COST=  _ERROR_=1 _N_=31
NOTE: 32 records were read from the infile "/home/chwang10/drugdata.txt".
      The minimum record length was 0.
      The maximum record length was 8.
NOTE: SAS went to a new line when INPUT statement reached past the end of a line.
NOTE: The data set WORK.DRUG has 30 observations and 6 variables.
NOTE: DATA statement used (Total process time):
      real time             0.07 seconds
      user cpu time         0.01 seconds
      system cpu time       0.00 seconds
      memory                795.56k
      OS Memory             30632.00k
      Timestamp             11/12/2020 09:40:16 AM
      Step Count                        52  Switch Count  3
      Page Faults                       0
      Page Reclaims                     138
      Page Swaps                        0
      Voluntary Context Switches        23
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


126
127        /* 1c */
127      !          Proc Print data=Drug;
128        Title "1c. Drug Data Set with Labels";
129        Label SES="Socioeconomic Status" DRUG="Drug Group" AGE="Age of subject"; * not working? ;
130        Run;

NOTE: There were 30 observations read from the data set WORK.DRUG.
NOTE: PROCEDURE PRINT used (Total process time):
      real time             0.06 seconds
      user cpu time         0.06 seconds
      system cpu time       0.00 seconds
      memory                2411.18k
      OS Memory             30888.00k
      Timestamp             11/12/2020 09:40:16 AM
      Step Count                        53  Switch Count  1
      Page Faults                       0
      Page Reclaims                     243
      Page Swaps                        0
      Voluntary Context Switches        6
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           16


131
132        /* 1d */
132      !          Proc Freq data=Drug;
133        Title "1d. Frequencies of Socioeconomic Status and Cost, Grouped by Age";
134        Format AGE Age.;
135        Table AGE*SES*COST;
136        Run;
```

```
NOTE: There were 30 observations read from the data set WORK.DRUG.
NOTE: PROCEDURE FREQ used (Total process time):
      real time            0.10 seconds
      user cpu time        0.11 seconds
      system cpu time      0.01 seconds
      memory               1828.06k
      OS Memory            31920.00k
      Timestamp            11/12/2020 09:40:16 AM
      Step Count                        54  Switch Count  5
      Page Faults                       0
      Page Reclaims                     282
      Page Swaps                        0
      Voluntary Context Switches        31
      Involuntary Context Switches      12
      Block Input Operations            0
      Block Output Operations           568


137
138          /* 1e */
138      !        Proc Gchart data=Drug;
139          Title "1e. Bar Chart of Cost, Grouped by Socioeconomic Status";
140          Vbar COST /group=SES discrete;
141          Run;

141      !      * There is a clear distinction between those with high and low socioeconomic statuses. Those with
142          high socioeconomic statuses tend to be associated with the generic drug, while those with low
143          socioeconomic statuses tend to be associated with the premium drug. ;
144
145          /* 1f */ * H0: There is no relationship between cost and socioeconomic status.
146          HA: There is a relationship between cost and socioeconomic status. ;

NOTE: There were 30 observations read from the data set WORK.DRUG.
NOTE: PROCEDURE GCHART used (Total process time):
      real time            0.17 seconds
      user cpu time        0.16 seconds
      system cpu time      0.01 seconds
      memory               7017.75k
      OS Memory            37656.00k
      Timestamp            11/12/2020 09:40:17 AM
      Step Count                        55  Switch Count  5
      Page Faults                       0
      Page Reclaims                     2112
      Page Swaps                        0
      Voluntary Context Switches        44
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           328


147          Proc Freq data=Drug;
148          Title "1f. Chi-Squared Test of Independence on Cost and Socioeconomic Status";
149          Table COST*SES /chisq;
150          Run;

NOTE: There were 30 observations read from the data set WORK.DRUG.
NOTE: PROCEDURE FREQ used (Total process time):
      real time            0.04 seconds
      user cpu time        0.05 seconds
      system cpu time      0.00 seconds
      memory               1298.43k
      OS Memory            35504.00k
      Timestamp            11/12/2020 09:40:17 AM
      Step Count                        56  Switch Count  5
      Page Faults                       0
      Page Reclaims                     212
      Page Swaps                        0
      Voluntary Context Switches        37
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           528


150      !       * We fail to reject H0 at the α = .05 level. There is insufficient
151          evidence (χ = 6.5164, p = 0.1638) that there is a relationship between cost and socioeconomic status.
152          These results do not support the conclusion drawn from the bar chart because of the small sample sizes
153          in all but one of the cells (n < 5), and we should exercise caution with using this test as the sample
154          size assumption of the chi-squared test is violated. ;
155
156          /* 2 */
156      !        Data CrimeDrinker;
157          Length Alcoholic$ 10; * Extending character length to properly display "Nondrinker" ;
158          Input Crime$ Alcoholic$ N @@;
159          Datalines;

NOTE: SAS went to a new line when INPUT statement reached past the end of a line.
NOTE: The data set WORK.CRIMEDRINKER has 12 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time            0.00 seconds
      user cpu time        0.00 seconds
```

```
      system cpu time        0.00 seconds
      memory                675.31k
      OS Memory              34984.00k
      Timestamp              11/12/2020 09:40:17 AM
      Step Count                        57  Switch Count  2
      Page Faults                        0
      Page Reclaims                     91
      Page Swaps                         0
      Voluntary Context Switches        15
      Involuntary Context Switches       0
      Block Input Operations             0
      Block Output Operations          264
```

```
161         ;
162         /* 2(i) */
162      !            Proc Freq data=CrimeDrinker;
163         Title "2(i). Chi-Squared Test of Independence on Alcoholism and Crime";
164         Table Alcoholic*Crime /chisq deviation expected norow nocol nopercent;
165         Weight N;
166         Run;
```

```
NOTE: There were 12 observations read from the data set WORK.CRIMEDRINKER.
NOTE: PROCEDURE FREQ used (Total process time):
      real time             0.04 seconds
      user cpu time         0.05 seconds
      system cpu time       0.01 seconds
      memory                1156.25k
      OS Memory              35504.00k
      Timestamp              11/12/2020 09:40:17 AM
      Step Count                        58  Switch Count  5
      Page Faults                        0
      Page Reclaims                    194
      Page Swaps                         0
      Voluntary Context Switches        36
      Involuntary Context Switches       0
      Block Input Operations             0
      Block Output Operations          552
```

```
167
168         /* 2(ii) */ * H0: There is no relationship between alcoholism and crime.
169         HA: There is a relationship between alcoholism and crime. ;
170
171         /* 2(iii) */ * We reject H0 at the α = .05 level. There is sufficient
172         evidence (χ = 49.7306, p < 0.0001) that there is a relationship between alcoholism and crime. ;
173
174         /* 2(iv) */ * It appears that drinkers are generally more likely to commit crimes, as the data show
175         they commit approximately 11.887073 percent more than nondrinkers. However, fraud is disproportionately
176         committed by nondrinkers, with only 30.4347826 percent of fraud is committed by drinkers—the only crime
177         with such a proportion below 53 percent. If fraud were removed from the data, drinkers
178         would (coincidentially) commit approximately 30.4347826 percent more crimes than nondrinkers. ;
179
180         /* 2(v) */ * See Problem 2(i) ;
181
182         /* 3a */
182      !            Data HA;
183         Length Race$ 9; * Extending character length to properly display "Caucasian" ;
184         /* 3a(i) */ Input Race$ Survival$ Percent @@;
185         Datalines;
```

```
NOTE: SAS went to a new line when INPUT statement reached past the end of a line.
NOTE: The data set WORK.HA has 4 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time             0.00 seconds
      user cpu time         0.00 seconds
      system cpu time       0.00 seconds
      memory                680.18k
      OS Memory              34984.00k
      Timestamp              11/12/2020 09:40:17 AM
      Step Count                        59  Switch Count  2
      Page Faults                        0
      Page Reclaims                     88
      Page Swaps                         0
      Voluntary Context Switches        15
      Involuntary Context Switches       0
      Block Input Operations             0
      Block Output Operations          272
```

```
187         ;
188         /* 3a(ii) */
188      !                Data HAC;
189         Set HA;
190         Count1=400*Percent;
191         Count2=1000*Percent;
192         Run;
```

```
NOTE: There were 4 observations read from the data set WORK.HA.
NOTE: The data set WORK.HAC has 4 observations and 5 variables.
NOTE: DATA statement used (Total process time):
```

```
      real time           0.00 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              948.40k
      OS Memory           35244.00k
      Timestamp           11/12/2020 09:40:17 AM
      Step Count                      60  Switch Count  2
      Page Faults                     0
      Page Reclaims                   134
      Page Swaps                      0
      Voluntary Context Switches      13
      Involuntary Context Switches    0
      Block Input Operations          0
      Block Output Operations         264


193
194         /* 3b */ %Macro CS(dataname=, r=, c=, options=, weight=);
195         Proc Freq data=&dataname;
196         Title "3c. Chi-Squared Test of Independence on Race and Survival Rate of Heart Attack";
197         Table &r*&c /&options;
198         Weight &weight;
199         Run;
200         %Mend CS;
201
202         /* 3c(i) */ * H0: There is no relationship between race and the survival rate of a heart attack.
203         HA: There is a relationship between race and the survival rate of a heart attack. ;
204         /* 3c(ii) */ %CS(dataname=HAC, r=Race, c=Survival, options=chisq deviation expected norow nocol, weight=Count1);

NOTE: There were 4 observations read from the data set WORK.HAC.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.04 seconds
      user cpu time       0.04 seconds
      system cpu time     0.00 seconds
      memory              1254.65k
      OS Memory           35760.00k
      Timestamp           11/12/2020 09:40:17 AM
      Step Count                      61  Switch Count  5
      Page Faults                     0
      Page Reclaims                   211
      Page Swaps                      0
      Voluntary Context Switches      33
      Involuntary Context Switches    0
      Block Input Operations          0
      Block Output Operations         552


205         /* 3c(iii) */ * We reject H0 at the α = .05 level. There is sufficient evidence (p < 0.0001) that there
206         is a relationship between race and the survival rate of a heart attack. ;
207         /* 3c(iv) */ * See Problem 3c(ii) ;
208         /* 3c(v) */ %CS(dataname=HAC, r=Race, c=Survival, options=chisq relrisk, weight=Count2);

NOTE: There were 4 observations read from the data set WORK.HAC.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.05 seconds
      user cpu time       0.05 seconds
      system cpu time     0.00 seconds
      memory              1199.78k
      OS Memory           35760.00k
      Timestamp           11/12/2020 09:40:17 AM
      Step Count                      62  Switch Count  5
      Page Faults                     0
      Page Reclaims                   201
      Page Swaps                      0
      Voluntary Context Switches      33
      Involuntary Context Switches    3
      Block Input Operations          0
      Block Output Operations         552


209         /* 3c(vi) */ * We reject H0 at the α = .05 level. There is sufficient evidence (p < 0.0001) that there
210         is a relationship between race and the survival rate of a heart attack. ;
211         /* 3c(vii) */ * Relative risk ratio: 1.0154 (95 percent confidence interval: (1.0142,1.0166)) ;
212         * African-Americans are approximately 1.54 percent more likely than Caucasians to die from a heart
213         attack, and we are 95 percent confident this percentage is between 1.42 and 1.66. Additionally, the
214         relative risk ratio is significant at the α = .05 level because its 95 percent confidence interval does
215         not include 1. ;
216         /* 3c(viii) */ * We get the same p-value (p < 0.0001) and conclusion for both sample sizes, although it
217         is worth noting that the test statistic was much greater for Count2 (χ = 654.3075 vs. χ = 261.7230). It
218         is not feasible to perform an accurate chi-squared test of independence using percentages, as not
219         having the correct sample sizes will make it unable to verify whether the sample size assumption is
220         violated. Also, because the test statistic varies depending on the sample size, inputting an arbitrary
221         sample size by multiplying the percentage by some constant may lead to the incorrect conclusion. ;
222         /* 3c(ix) */ * See Problem 3c(v) ;
223
224         /* 4a */ * H0: Test-driving a car does not change one's opinion of it.
225         HA: Test-driving a car does change one's opinion of it. ;
226
227         /* 4b */
227     !        Data Car;
228         Input Pre$ Post$ N @@;
```

```
229          Datalines;

NOTE: SAS went to a new line when INPUT statement reached past the end of a line.
NOTE: The data set WORK.CAR has 4 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time            0.00 seconds
      user cpu time        0.00 seconds
      system cpu time      0.00 seconds
      memory               676.87k
      OS Memory            35240.00k
      Timestamp            11/12/2020 09:40:17 AM
      Step Count                        63  Switch Count  2
      Page Faults                        0
      Page Reclaims                     90
      Page Swaps                         0
      Voluntary Context Switches        16
      Involuntary Context Switches       0
      Block Input Operations             0
      Block Output Operations          264


231          ;
232          Proc Freq data=Car;
233          Title "4c. McNemar's Test on Pre- and Post-Test Drive Opinion of Car";
234          Table Pre*Post /agree noprint;
235          Weight N;
236          Run;

NOTE: There were 4 observations read from the data set WORK.CAR.
NOTE: PROCEDURE FREQ used (Total process time):
      real time            0.32 seconds
      user cpu time        0.10 seconds
      system cpu time      0.02 seconds
      memory               14460.59k
      OS Memory            46512.00k
      Timestamp            11/12/2020 09:40:17 AM
      Step Count                        64  Switch Count  5
      Page Faults                        0
      Page Reclaims                   3525
      Page Swaps                         0
      Voluntary Context Switches       481
      Involuntary Context Switches       0
      Block Input Operations             0
      Block Output Operations         1352


236     !      * We fail to reject H0 at the α = .05 level. There is insufficient evidence (χ = 3.60, p = 0.0578)
237          that test-driving a car changes one's opinion of it. ;
238
239          /* 4c */ * See Problem 4b ;
240
241          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
252
```

Results: Homework 6.sas

1c. Drug Data Set with Labels

Obs	ID	GENDER	SES	DRUG	AGE	COST
1	001	Male	Low	B	15	Generic
2	002	Female	Medium	Z	35	Miscode
3	003	Female	High	F	76	Generic
4	004	Male	Low	c	21	Generic
5	005	Female	High		58	Miscode
6	006	Female	Low	G	47	Premium
7	007	Female	Low	D	23	Premium
8	008	Male	Medium	E	51	Premium
9	009	Male	High	A	32	Generic
10	010	Male	High	g	19	Premium
11	011	Male	Missing	b	21	Generic
12	012	Female	Low	?	38	Miscode
13	013	Male	Medium	F	26	Generic
14	014	Male	High	d	36	Premium
15	015	Female	Medium	a	25	Generic
16	016	Male	Low	D	32	Premium
17	017	Female	High	A	54	Generic
18	018	Male	Low	G	78	Premium
19	019	Female	Medium	G	54	Premium
20	020	Female	Low	H	47	Premium
21	021	Female	High	C	49	Generic
22	022	Male	High	B	38	Generic
23	023	Female	Medium	B	56	Generic
24	024	Male	High	A	29	Generic
25	025	Female	Low	D	77	Premium
26	026	Male	Medium	G	27	Premium
27	027	Female	Low	H	36	Premium
28	028	Male	High	B	19	Generic

Obs	ID	GENDER	SES	DRUG	AGE	COST
29	029	Female	Medium	A	28	Generic
30	030	Male	High	B	54	Generic

## 1d. Frequencies of Socioeconomic Status and Cost, Grouped by Age

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table 1 of SES by COST			
Controlling for AGE=0-20			
	COST		
SES(Socioeconomic Status)	Generic	Miscode	Premium	Total

SES(Socioeconomic Status)	Generic	Miscode	Premium	Total
High	1	0	1	2
	33.33	0.00	33.33	66.67
	50.00	0.00	50.00	
	50.00	.	100.00	
Low	1	0	0	1
	33.33	0.00	0.00	33.33
	100.00	0.00	0.00	
	50.00	.	0.00	
Medium	0	0	0	0
	0.00	0.00	0.00	0.00
	.	.	.	
	0.00	.	0.00	
Total	2	0	1	3
	66.67	0.00	33.33	100.00

Frequency
Percent
Row Pct
Col Pct

Table 2 of SES by COST			
Controlling for AGE=21-40			
	COST		
SES(Socioeconomic Status)	Generic	Miscode	Premium	Total

SES(Socioeconomic Status)	Generic	Miscode	Premium	Total
High	3	0	1	4
	21.43	0.00	7.14	28.57
	75.00	0.00	25.00	
	42.86	0.00	20.00	
Low	1	1	3	5
	7.14	7.14	21.43	35.71
	20.00	20.00	60.00	
	14.29	50.00	60.00	
Medium	3	1	1	5
	21.43	7.14	7.14	35.71
	60.00	20.00	20.00	
	42.86	50.00	20.00	
Total	7	2	5	14
	50.00	14.29	35.71	100.00
Frequency Missing = 1			

Frequency
Percent
Row Pct
Col Pct

Table 3 of SES by COST			
Controlling for AGE=41+			
	COST		
SES(Socioeconomic Status)	Generic	Miscode	Premium	Total

SES(Socioeconomic Status)	Generic	Miscode	Premium	Total
High	4	1	0	5
	33.33	8.33	0.00	41.67
	80.00	20.00	0.00	
	80.00	100.00	0.00	
Low	0	0	4	4
	0.00	0.00	33.33	33.33
	0.00	0.00	100.00	
	0.00	0.00	66.67	
Medium	1	0	2	3
	8.33	0.00	16.67	25.00
	33.33	0.00	66.67	
	20.00	0.00	33.33	
Total	5	1	6	12
	41.67	8.33	50.00	100.00

1e. Bar Chart of Cost, Grouped by Socioeconomic Status
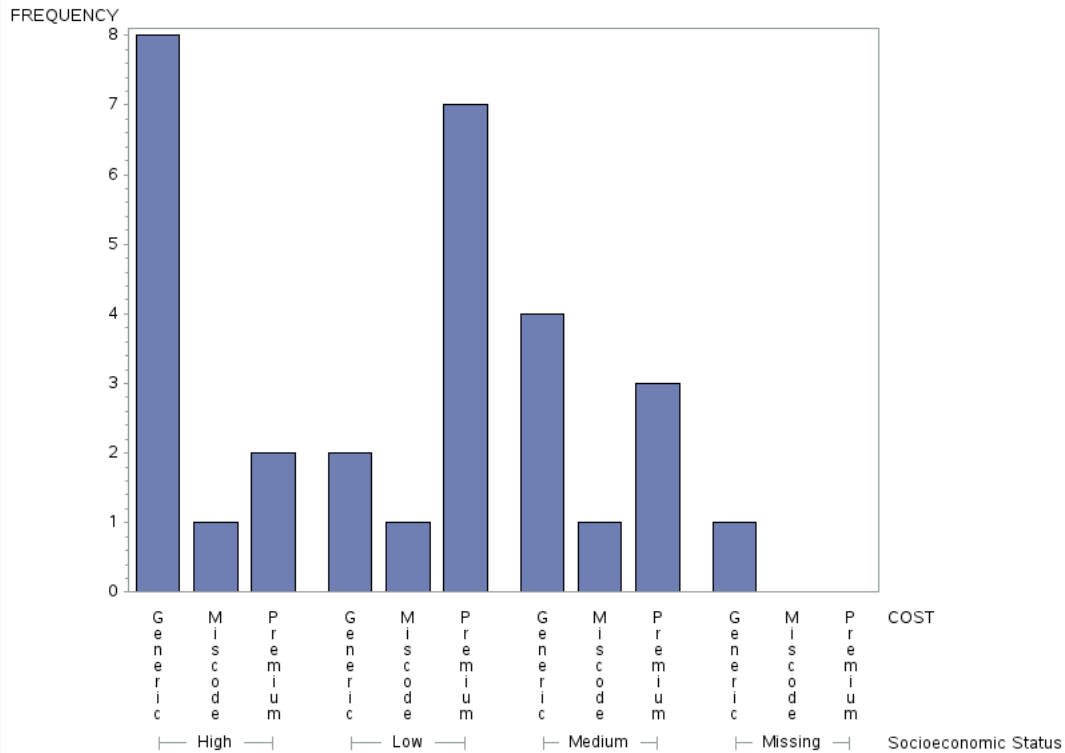
1f. Chi-Squared Test of Independence on Cost and Socioeconomic Status

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of COST by SES				
		SES(Socioeconomic Status)			
	COST	High	Low	Medium	Total
	Generic	8 27.59 57.14 72.73	2 6.90 14.29 20.00	4 13.79 28.57 50.00	14 48.28
	Miscode	1 3.45 33.33 9.09	1 3.45 33.33 10.00	1 3.45 33.33 12.50	3 10.34
	Premium	2 6.90 16.67 18.18	7 24.14 58.33 70.00	3 10.34 25.00 37.50	12 41.38
	Total	11 37.93	10 34.48	8 27.59	29 100.00
	Frequency Missing = 1				

Statistics for Table of COST by SES

Statistic	DF	Value	Prob
Chi-Square	4	6.5164	0.1638
Likelihood Ratio Chi-Square	4	6.8447	0.1443
Mantel-Haenszel Chi-Square	1	1.3308	0.2487
Phi Coefficient		0.4740	
Contingency Coefficient		0.4283	
Cramer's V		0.3352	
WARNING: 89% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 29
Frequency Missing = 1

2(i). Chi-Squared Test of Independence on Alcoholism and Crime

The FREQ Procedure

Frequency Expected Deviation	Table of Alcoholic by Crime						
		Crime					
	Alcoholic	Arson	Coining	Fraud	Rape	Stealing	Violence	Total
	Drinker	50 49.109 0.8913	18 16.898 1.1024	63 109.31 -46.31	88 79.208 8.7924	379 358.55 20.454	155 139.93 15.067	753
	Nondrinker	43 43.891 -0.891	14 15.102 -1.102	144 97.694 46.306	62 70.792 -8.792	300 320.45 -20.45	110 125.07 -15.07	673
	Total	93	32	207	150	679	265	1426

Statistic	DF	Value	Prob
Chi-Square	5	49.7306	<.0001
Likelihood Ratio Chi-Square	5	50.5173	<.0001
Mantel-Haenszel Chi-Square	1	13.0253	0.0003
Phi Coefficient		0.1867	
Contingency Coefficient		0.1836	
Cramer's V		0.1867	

Sample Size = 1426

3c. Chi-Squared Test of Independence on Race and Survival Rate of Heart Attack

The FREQ Procedure

Frequency Expected Deviation Percent	Table of Race by Survival		
		Survival	
Race	Died	Survived	Total
Black	39600 39300 300 49.50	400 700 -300 0.50	40000 50.00
Caucasian	39000 39300 -300 48.75	1000 700 300 1.25	40000 50.00
Total	78600 98.25	1400 1.75	80000 100.00

Statistics for Table of Race by Survival

Statistic	DF	Value	Prob
Chi-Square	1	261.7230	<.0001
Likelihood Ratio Chi-Square	1	270.2375	<.0001
Continuity Adj. Chi-Square	1	260.8513	<.0001
Mantel-Haenszel Chi-Square	1	261.7197	<.0001
Phi Coefficient		0.0572	
Contingency Coefficient		0.0571	
Cramer's V		0.0572	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	39600
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 80000

3c. Chi-Squared Test of Independence on Race and Survival Rate of Heart Attack

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Race by Survival		
		Survival	
Race	Died	Survived	Total
Black	99000 49.50 99.00 50.38	1000 0.50 1.00 28.57	100000 50.00
Caucasian	97500 48.75 97.50 49.62	2500 1.25 2.50 71.43	100000 50.00
Total	196500 98.25	3500 1.75	200000 100.00

Statistics for Table of Race by Survival

Statistic	DF	Value	Prob
Chi-Square	1	654.3075	<.0001
Likelihood Ratio Chi-Square	1	675.5936	<.0001
Continuity Adj. Chi-Square	1	653.4354	<.0001
Mantel-Haenszel Chi-Square	1	654.3043	<.0001
Phi Coefficient		0.0572	
Contingency Coefficient		0.0571	
Cramer's V		0.0572	

Fisher's Exact Test

Fisher's Exact Test	
Cell (1,1) Frequency (F)	99000
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Odds Ratio and Relative Risks		
Statistic	Value	95% Confidence Limits
Odds Ratio	2.5385	2.3577 2.7331
Relative Risk (Column 1)	1.0154	1.0142 1.0166
Relative Risk (Column 2)	0.4000	0.3719 0.4302

Sample Size = 200000

4c. McNemar's Test on Pre- and Post-Test Drive Opinion of Car

The FREQ Procedure

Statistics for Table of Pre by Post

McNemar's Test		
Chi-Square	DF	Pr > ChiSq
3.6000	1	0.0578

Simple Kappa Coefficient			
Estimate	Standard Error	95% Confidence Limits	
0.5968	0.0565	0.4860	0.7076

Sample Size = 198

Agreement of Pre and Post