B. Kedem, STAT 430
SAS Examples SAS3
==========
ssh abc@glue.umd.edu, tap sas82, sas <--Old
tap sas913, sas <--New Version

https://www.statlab.umd.edu/sasdoc/sashtml/onldoc.htm

CH3: CATEGORICAL DATA
==========

Speak of categorical data: Nominal, Ordinal.
Nominal: Hair color, Political affiliation, Religion,
Transportation type, Sleep stage, Sport activity.
Ordinal: Blood pressure, Age group, Degree of feeling, Discrete data,
Opinion.
Levels: Low,Normal,High, etc.

Basic operation: Counting!!!

Introduce basic categorical data analysis, pearson chi-sq test stat, likelihood ration test stat, Fisher exact test, Odds Ratio, Relative Risk, Mantel-Haenszel.

Content
---
1. Questionnaires
2. Associate categories with names using PROC FORMAT
3. Categorizing (Quantizing) AGE: Either by IF or PROC FORMAT
4. Two by Two Tables
5. Now Also Chi Square Test
6. Chi Square test from a given 2 by 2 table
7. Odds ratio and Relative risk
8. Mantel-Haenszel Meta Analysis
9. Chi Square for a 2 by 3 Table

```
1. Questionnaires
---------------------
Wish to design a questionnaire with the following items.

1. AGE in years
2. GENDER 1=Male, 2=Female
3. RACE 1=W, 2=B, 3=H, 4=Other
4. MARITAL STATUS 1=Single, 2=Married, 3=Widowed, 4=Divorced
5. EDUCATION LEVEL 1=Elementary 2=HS 3=College 4=Grdauate Degree
6. In the following statements
   1=Strongly Disagree, 2= Disagree, 3=Neutral 4=Agree 5=Strongly Agree

a. President is doing a good job.
b. Arms budget should increase.
c. There should be more federal aid fo big cities.

NOTE: Here we store categorical variables as characters but we give the
      categories names in terms of numbers for now. We'll fix this
      later. NUMBERS ARE CONVENIENT FOR RECORDING THE DATA.

OPTIONS PS=35 LS=70;

DATA QUEST;
INPUT ID 1-3 AGE 4-5 GENDER $ 6 RACE $ 7 MARITAL $ 8 EDUC $ 9
      PRES 10 ARMS 11 CITIES 12;

*Description of input using LABELS;

LABEL  MARITAL ='Marital Status'
       EDUC    ='Level of Education'
       PRES    ='President Doing a Good Job'
       ARMS    ='Arms Budget Should Increase'
       CITIES  ='More Federal Aid to Cities';

DATALINES;
001301133111
002282234555
003192112321
```

```
004271122135
005601441552
006592114334
007231222222
008452333335
009312123444
010372114115
;

PROC MEANS DATA=QUEST N MEAN STD;
TITLE 'Age Statistics';
VAR AGE;
RUN;
```

The MEANS Procedure

Analysis Variable : AGE

N	Mean	Std Dev
10	35.9000000	14.3406958

```
PROC FREQ DATA=QUEST;
TITLE 'Count Information for Categorical Variables';
TABLES GENDER RACE MARITAL EDUC PRES ARMS CITIES;
RUN;
```

Count Information for Categorical Variables

The FREQ Procedure

Cumulative Cumulative

GENDER	Frequency	Percent	Frequency	Percent
1	4	40.00	4	40.00
2	6	60.00	10	100.00

RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	6	60.00	6	60.00
2	2	20.00	8	80.00
3	1	10.00	9	90.00
4	1	10.00	10	100.00

Marital Status

MARITAL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	3	30.00	3	30.00
2	3	30.00	6	60.00
3	3	30.00	9	90.00
4	1	10.00	10	100.00

Level of Education

EDUC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1	10.00	1	10.00
2	3	30.00	4	40.00
3	3	30.00	7	70.00
4	3	30.00	10	100.00

President Doing a Good Job

PRES	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	3	30.00	3	30.00
2	1	10.00	4	40.00
3	3	30.00	7	70.00
4	1	10.00	8	80.00
5	2	20.00	10	100.00

Arms Budget Should Increase

ARMS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	2	20.00	2	20.00
2	2	20.00	4	40.00
3	3	30.00	7	70.00
4	1	10.00	8	80.00
5	2	20.00	10	100.00

More Federal Aid to Cities

CITIES	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	2	20.00	2	20.00
2	2	20.00	4	40.00
4	2	20.00	6	60.00
5	4	40.00	10	100.00

2. Associate categories with names using PROC FORMAT

Now we'll give names to the categories instead of the numerical names

1,2,3, etc we gave the categories before. All we have to do is recode the category names. The names will appear in the output.
This is done by FORMATS.

The complete program is given next.

```
OPTIONS PS=35 LS=70;

PROC FORMAT;
*Note: VALUE can be at most 8 long!!!
*For the character data need quotes on numbers and $ on VALUE;
VALUE $MF        '1'='MALE' '2'='FEMALE';
VALUE $RACE      '1'='W' '2'='B' '3'='H' '4'='OTHER';
VALUE $MSTATUS   '1'='SINGLE' '2'='MARRIED' '3'='WIDOWED'
                 '4'='DIVORCED';
VALUE $EDUCAT    '1'='ELEMENTARY' '2'='HS' '3'='COLLEGE'
                 '4'='GRAD ED';
*For the numerical data NO $ ON VALUE AND NO QUOTES ON NUMBERS;
VALUE AGREEYN     1='STR DISAGREE' 2='DISAGREE' 3='NEUTRAL'
                  4='AGREE' 5='STR AGREE';
RUN;

DATA QUEST;
INPUT ID 1-3 AGE 4-5 GENDER $ 6 RACE $ 7 MARITAL $ 8 EDUC $ 9
      PRES 10 ARMS 11 CITIES 12;

*Description of input using LABELS;

LABEL  MARITAL ='Marital Status'
       EDUC    ='Level of Education'
       PRES    ='President Doing a Good Job'
       ARMS    ='Arms Budget Should Increase'
       CITIES  ='More Federal Aid to Cities';

*Now use the coded categories. Note PRES ARMS CITIES;
*all coded by AGREEYN!!!;
FORMAT GENDER $MF. RACE $RACE. MARITAL $MSTATUS.
       EDUC $EDUCAT. PRES ARMS CITIES AGREEYN.;
```

```
*Before I used $AGREEYN in PRES ARMS CITIES AGREEYN. and it worked;
*just the same!!! Better to use PRES ARMS CITIES AGREEYN.;

DATALINES;
001301133111
002282234555
003192112321
004271122135
005601441552
006592114334
007231222222
008452333335
009312123444
010372114115
;


PROC FREQ DATA=QUEST;
TITLE 'Count Information for Categorical Variables';
TABLES GENDER RACE MARITAL EDUC PRES ARMS CITIES;
RUN;
```

Count Information for Categorical Variables

The FREQ Procedure

GENDER	Frequency	Percent	Frequency	Percent
MALE	4	40.00	4	40.00
FEMALE	6	60.00	10	100.00

	Cumulative	Cumulative

RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
W	6	60.00	6	60.00
B	2	20.00	8	80.00
H	1	10.00	9	90.00
OTHER	1	10.00	10	100.00

Marital Status

MARITAL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
SINGLE	3	30.00	3	30.00
MARRIED	3	30.00	6	60.00
WIDOWED	3	30.00	9	90.00
DIVORCED	1	10.00	10	100.00

Level of Education

EDUC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
ELEMENTARY	1	10.00	1	10.00
HS	3	30.00	4	40.00
COLLEGE	3	30.00	7	70.00
GRAD ED	3	30.00	10	100.00

President Doing a Good Job

PRES	Frequency	Percent	Cumulative Frequency	Cumulative Percent
STR DISAGREE	3	30.00	3	30.00
DISAGREE	1	10.00	4	40.00
NEUTRAL	3	30.00	7	70.00

AGREE	1	10.00	8	80.00
STR AGREE	2	20.00	10	100.00

Arms Budget Should Increase

ARMS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
STR DISAGREE	2	20.00	2	20.00
DISAGREE	2	20.00	4	40.00
NEUTRAL	3	30.00	7	70.00
AGREE	1	10.00	8	80.00
STR AGREE	2	20.00	10	100.00

More Federal Aid to Cities

CITIES	Frequency	Percent	Cumulative Frequency	Cumulative Percent
STR DISAGREE	2	20.00	2	20.00
DISAGREE	2	20.00	4	40.00
AGREE	2	20.00	6	60.00
STR AGREE	4	40.00	10	100.00

3. Categorizing (Quantizing) AGE: Either by IF or PROC FORMAT

```
*Giving names to categories obtained from CLIPPED age;
PROC FORMAT;
VALUE AGEGRP 0-20 ='0-20'
             21-40 ='21-40'
             41-60 ='41-60'
             61-80 ='61-80'
             OTHER ='Greater than 80';
```

```
RUN;

NOTE: Can associate the format in the DATA or PROC statements!!!


PROC FREQ DATA=QUEST;
*Give label describing the input using LABELS;
LABEL  AGE  = 'Age Frequency';
TABLES AGE;
*Associate AGE with AGEGRP;
FORMAT AGE AGEGRP.;
RUN;
```

The FREQ Procedure

Age Frequency

AGE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0-20	1	10.00	1	10.00
21-40	6	60.00	7	70.00
41-60	3	30.00	10	100.00

```
Now cross tabulate: AGE*PRES;
---------------------------------

PROC FREQ DATA=QUEST;
*Give label describing the input using LABELS;
LABEL  AGE  = 'Age Frequency';
TABLES AGE PRES AGE*PRES;
*Associate AGE with AGEGRP;
FORMAT AGE AGEGRP. PRES AGREEYN.;
RUN;

AGE*PRES Table has the form: (entries incorrect!!!)
```

Table of AGE by PRES

AGE(Age Frequency) PRES(President Doing a Good Job)

Frequency Percent Row Pct Col Pct	STR DISAGREE	DISAGREE	NEUTRAL	AGREE	Total
0-20	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 10.00 50.00 33.33	1 10.00 50.00 33.33	2 20.00
21-40	0 0.00 0.00 0.00	1 10.00 50.00 33.33	0 0.00 0.00 0.00	1 10.00 50.00 33.33	2 20.00
41-60	1 10.00 50.00 100.00	1 10.00 50.00 33.33	0 0.00 0.00 0.00	0 0.00 0.00 0.00	2 20.00
Total	1 10.00	3 30.00	3 30.00	3 30.00	10 100.00

(Continued)

4. Two by Two Tables:

```
DATA ELECT;
INPUT GENDER $ CANDID $;
```

```
DATALINES;
M DEWEY
M TRUMAN
F TRUMAN
F TRUMAN
M DEWEY
F DEWEY
F TRUMAN
F TRUMAN
F TRUMAN
M DEWEY
M DEWEY
M TRUMAN
F TRUMAN
F TRUMAN
M DEWEY
M DEWEY
F TRUMAN
M TRUMAN
M TRUMAN
F DEWEY
;
PROC FREQ DATA=ELECT;
*One and two way tables;
TABLES
GENDER CANDID CANDID*GENDER;
RUN;
```

The FREQ Procedure

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	10	50.00	10	50.00
M	10	50.00	20	100.00

CANDID	Frequency	Percent	Cumulative Frequency	Cumulative Percent
DEWEY	8	40.00	8	40.00
TRUMAN	12	60.00	20	100.00

Table of CANDID by GENDER

CANDID	GENDER		
Frequency / Percent / Row Pct / Col Pct	F	M	Total
DEWEY	2	6	8
	10.00	30.00	40.00
	25.00	75.00	
	20.00	60.00	
TRUMAN	8	4	12
	40.00	20.00	60.00
	66.67	33.33	
	80.00	40.00	
Total	10	10	20
	50.00	50.00	100.00

5. Now Also Chi Square Test:

```
PROC FREQ DATA=ELECT;
*One and two way tables;
TABLES
```

```
GENDER CANDID CANDID*GENDER/CHISQ;
RUN;
```

The FREQ Procedure

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	10	50.00	10	50.00
M	10	50.00	20	100.00

Chi-Square Test for Equal Proportions <------NOTE!!!

Chi-Square	0.0000
DF	1
Pr > ChiSq	1.0000 <--- Clear since 50/50

Sample Size = 20

CANDID	Frequency	Percent	Cumulative Frequency	Cumulative Percent
DEWEY	8	40.00	8	40.00
TRUMAN	12	60.00	20	100.00

Chi-Square Test for Equal Proportions

Chi-Square	0.8000
DF	1
Pr > ChiSq	0.3711

Sample Size = 20

Table of CANDID by GENDER

CANDID	GENDER		
Frequency Percent Row Pct Col Pct	F	M	Total
DEWEY	2 10.00 25.00 20.00	6 30.00 75.00 60.00	8 40.00
TRUMAN	8 40.00 66.67 80.00	4 20.00 33.33 40.00	12 60.00
Total	10 50.00	10 50.00	20 100.00

Statistics for Table of CANDID by GENDER

Statistic	DF	Value	Prob
Chi-Square	1	3.3333	0.0679
Likelihood Ratio Chi-Square	1	3.4522	0.0632
Continuity Adj. Chi-Square	1	1.8750	0.1709
Mantel-Haenszel Chi-Square	1	3.1667	0.0752
Phi Coefficient		-0.4082	
Contingency Coefficient		0.3780	
Cramer's V		-0.4082	

```
        WARNING: 50% of the cells have expected counts less
                 than 5. Chi-Square may not be a valid test.


                       Fisher's Exact Test
                 ---------------------------------
                 Cell (1,1) Frequency (F)        2
                 Left-sided Pr <= F         0.0849
                 Right-sided Pr >= F        0.9901

                 Table Probability (P)      0.0750
                 Two-sided Pr <= P          0.1698

                          Sample Size = 20
```

Discuss the example from the book on p. 77.

6. Chi Square test from a given 2 by 2 table

Suppose the data are already in the form of 2 by 2 table and we wish to perform a chi-sqtest, use WEIGHT statement to tell SAS the count in each cell.

```
                          OUTCOME
                 |   Dead        Alive
-----------------------------------------
       Control   |    20          80
                 |
GROUP            |
                 |
       Drug      |    10          90
                 |
```

```
OPTIONS PS=35 LS=70;

DATA T_BY_T;
INPUT GROUP $ OUTCOME $ COUNT;
DATALINES;
```

```
CONTROL DEAD 20
CONTROL ALIVE 80
DRUG DEAD 10
DRUG ALIVE 90
;
PROC FREQ DATA=T_BY_T;
TABLES GROUP*OUTCOME/CHISQ;
WEIGHT COUNT;
RUN;
```

Note: Table is reversed alphabetically!!! First Alive then Dead!!!

The FREQ Procedure

Table of GROUP by OUTCOME

GROUP	OUTCOME		
Frequency Percent Row Pct Col Pct	ALIVE	DEAD	Total
CONTROL	80 40.00 80.00 47.06	20 10.00 20.00 66.67	100 50.00
DRUG	90 45.00 90.00 52.94	10 5.00 10.00 33.33	100 50.00
Total	170 85.00	30 15.00	200 100.00

Statistics for Table of GROUP by OUTCOME

Statistic	DF	Value	Prob	
Chi-Square	1	3.9216	0.0477	<--???
Likelihood Ratio Chi-Square	1	3.9866	0.0459	<--???
Continuity Adj. Chi-Square	1	3.1765	0.0747	
Mantel-Haenszel Chi-Square	1	3.9020	0.0482	
Phi Coefficient		-0.1400		
Contingency Coefficient		0.1387		
Cramer's V		-0.1400		

Fisher's Exact Test

Cell (1,1) Frequency (F)	80	
Left-sided Pr <= F	0.0367	
Right-sided Pr >= F	0.9859	<-- Accept!!!
Table Probability (P)	0.0226	
Two-sided Pr <= P	0.0734	<-- Accept?

Sample Size = 200

7. Odds ratio and Relative risk

In our case above: OddsRatio=(80/90)/(20/10)=0.4444
RelatRisk=(80/100)/(90/100)=0.8889

To get the odds ratio and relative risk use option CMH:

```
PROC FREQ DATA=T_BY_T;
TABLES GROUP*OUTCOME/CHISQ CMH;
WEIGHT COUNT;
RUN;
```

The FREQ Procedure

Summary Statistics for GROUP by OUTCOME

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value
Case-Control	Mantel-Haenszel	0.4444
(Odds Ratio)	Logit	0.4444
Cohort	Mantel-Haenszel	0.8889
(Col1 Risk)	Logit	0.8889
Cohort	Mantel-Haenszel	2.0000
(Col2 Risk)	Logit	2.0000

Type of Study	Method	95% Confidence Limits	
Case-Control	Mantel-Haenszel	0.1964	1.0057
(Odds Ratio)	Logit	0.1964	1.0057
Cohort	Mantel-Haenszel	0.7901	1.0000
(Col1 Risk)	Logit	0.7901	1.0000
Cohort	Mantel-Haenszel	0.9866	4.0545
(Col2 Risk)	Logit	0.9866	4.0545

Total Sample Size = 200

8. Mantel-Haenszel Meta Analysis

Have two 2 by 2 tables. Wish to combine thje data and then test for independence!!! Will consider the combined odds ratio and the Cochran-Mantel-Hanszel test for independence given two tables.

	Boys test		Girls test	
	Fail	Pass	Fail	Pass
Low	20	100	30	100
High	15	150	25	200

Sleep

Note: The WEIGHT statement feeds SAS the cell counts!!!

```
OPTIONS PS=35 LS=70;

DATA ABILITY;
INPUT GENDER $ RESULTS $ SLEEP $  COUNT;
DATALINES;
BOYS FAIL 1-LOW 20
BOYS FAIL 2-HIGH 15
BOYS PASS 1-LOW 100
BOYS PASS 2-HIGH 150
GIRLS FAIL 1-LOW 30
GIRLS FAIL 2-HIGH 25
GIRLS PASS 1-LOW 100
GIRLS PASS 2-HIGH 200
;

PROC FREQ DATA=ABILITY;
TITLE 'Mantel Haenszel Tet of Independene';
TABLES GENDER*SLEEP*RESULTS/ALL;
WEIGHT COUNT;
RUN;
```

Some of the output for Boys only, Girls only, Combined Boys & Girls:

Statistics for Table 1 of SLEEP by RESULTS
Controlling for GENDER=BOYS

Statistic	DF	Value	Prob
Chi-Square	1	3.7013	0.0544
Likelihood Ratio Chi-Square	1	3.6494	0.0561
Fisher n11=20 (two sided)			0.0674

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	2.0000	0.9777	4.0911
Cohort (Col1 Risk)	1.8333	0.9795	3.4313
Cohort (Col2 Risk)	0.9167	0.8349	1.0064

Statistics for Table 2 of SLEEP by RESULTS
Controlling for GENDER=GIRLS

Statistic	DF	Value	Prob
Chi-Square	1	9.0106	0.0027
Likelihood Ratio Chi-Square	1	8.7000	0.0032
Fisher n11=30 (two sided)			0.0037

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	2.4000	1.3404	4.2973
Cohort (Col1 Risk)	2.0769	1.2789	3.3728
Cohort (Col2 Risk)	0.8654	0.7792	0.9611

Summary Statistics for SLEEP by RESULTS

Controlling for GENDER

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	12.4770	0.0004
2	Row Mean Scores Differ	1	12.4770	0.0004
3	General Association	1	12.4770	0.0004

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value
Case-Control (Odds Ratio)	Mantel-Haenszel	2.2289
	Logit	2.2318
Cohort (Col1 Risk)	Mantel-Haenszel	1.9775
	Logit	1.9822
Cohort (Col2 Risk)	Mantel-Haenszel	0.8891
	Logit	0.8936

Type of Study	Method	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	1.4185	3.5024
	Logit	1.4205	3.5064
Cohort (Col1 Risk)	Mantel-Haenszel	1.3474	2.9021
	Logit	1.3508	2.9087
Cohort (Col2 Risk)	Mantel-Haenszel	0.8283	0.9544
	Logit	0.8334	0.9582

Breslow-Day Test for

```
Homogeneity of the Odds Ratios
------------------------------
Chi-Square              0.1501
DF                           1
Pr > ChiSq              0.6985


   Total Sample Size = 640
```

9. Chi Square for a 2 by 3 Table

```
Run exactly as in 2 by 2 tables. We have:
Gender at 2 levels M,F, Candidate at 3 levels DEWEY,TRUMAN,WASH.

Note: In each cell the expected count must be at least 5 for the
      results to be valid.

Note: Here df=(2-1)*(3-1)=2

DATA ELECT;
INPUT GENDER $ CANDID $;
DATALINES;
M DEWEY
M TRUMAN
F TRUMAN
F TRUMAN
M DEWEY
F DEWEY
F WASH
F TRUMAN
F TRUMAN
M WASH
M DEWEY
M TRUMAN
F WASH
F TRUMAN
```

```
M DEWEY
M DEWEY
F WASH
M TRUMAN
M TRUMAN
F WASH
F TRUMAN
F TRUMAN
F TRUMAN
M TRUMAN
M DEWEY
M DEWEY
F DEWEY
F WASH
F WASH
M WASH
M WASH
M WASH
M WASH
F TRUMAN
;


PROC FREQ DATA=ELECT;
*One and two way tables;
TABLES
GENDER CANDID CANDID*GENDER/CHISQ;
RUN;
```

The FREQ Procedure

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	17	50.00	17	50.00
M	17	50.00	34	100.00

Chi-Square Test
for Equal Proportions

Chi-Square	0.0000
DF	1
Pr > ChiSq	1.0000

Sample Size = 34

CANDID	Frequency	Percent	Cumulative Frequency	Cumulative Percent
DEWEY	9	26.47	9	26.47
TRUMAN	14	41.18	23	67.65
WASH	11	32.35	34	100.00

Chi-Square Test
for Equal Proportions

Chi-Square	1.1176
DF	2
Pr > ChiSq	0.5719

Sample Size = 34

The FREQ Procedure
Table of CANDID by GENDER

CANDID	GENDER		
Frequency | Percent | Row Pct | Col Pct	F	M	Total
DEWEY	2	7	9
	5.88	20.59	26.47
	22.22	77.78	
	11.76	41.18	
TRUMAN	9	5	14
	26.47	14.71	41.18
	64.29	35.71	
	52.94	29.41	
WASH	6	5	11
	17.65	14.71	32.35
	54.55	45.45	
	35.29	29.41	
Total	17	17	34
	50.00	50.00	100.00

The FREQ Procedure
Statistics for Table of CANDID by GENDER

Statistic	DF	Value	Prob
Chi-Square	2	4.0115	0.1346
Likelihood Ratio Chi-Square	2	4.1919	0.1230
Mantel-Haenszel Chi-Square	1	1.7574	0.1849
Phi Coefficient		0.3435	
Contingency Coefficient		0.3249	
Cramer's V		0.3435	

WARNING: 33% of the cells have expected counts less than 5. Chi-Square may not be a valid test. <-- Note!

Sample Size = 34

Now, suppose we have the counts already. Then we input the data as follows:

```
DATA TT;
INPUT  CANDID $ GENDER $ COUNT;
DATALINES;
DEWEY  F 2
DEWEY  M 7
TRUMAN F 9
TRUMAN M 5
WASH   F 6
WASH   M 5
;

PROC FREQ DATA=TT;
TABLES CANDID*GENDER/CHISQ;
WEIGHT COUNT;
RUN;
```

Get the same results as above.