STAT 515 - Practice 3 Solutions

Pg 714: 13.15a by hand OR SAS. Also state the null and alternate hypothesis in terms of the parameters and the problem, and check the assumptions.

H_0: p_0=0.26, p_1=0.30, p_2=0.11, p_3=0.14, p_4=0.19
H_A: not H_0

The expected values in this case would be 85*.26=22.1, 85*.30=25.5, 85*.11=9.35, 85*.14=11.9, and 85*.19=16.15... as all of these are greater than five the test should be reliable.

By SAS:

```
DATA avonex;
INPUT exacerbations $  count;
CARDS;
0              32
1      26
2              15
3              6
4plus          6
;

PROC FREQ DATA=avonex ORDER=data;
TABLES exacerbations / TESTP=(.26,.30,.11,.14,.19);
WEIGHT count;
RUN;
```

```
                                    Chi-Square Test
                                 for Specified Proportions
                                 ffffffffffffffffffffffff
                                 Chi-Square        17.1631
                                 DF                      4
                                 Pr > ChiSq         0.0018
```

With a p-value of 0.0018 (much less than α=0.05) we reject the null hypothesis and conclude that there is a difference.

By hand:

$$\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp} = \frac{(32-22.1)^2}{22.1} + \frac{(26-25.5)^2}{25.5} + \frac{(15-9.35)^2}{9.35} + \frac{(6-11.9)^2}{11.9} + \frac{(6-16.15)^2}{16.15} = 17.16313$$

We compare this to 9.48773 (df=5-1=4, α=0.05) and reject the null hypothesis (so we conclude there is a difference).

Pg. 732: 13.35 b-d by hand and by SAS. Is the sample size large enough? Is this a test of homogeneity or of independence? Why?

```
DATA tumors;
INPUT tumors $ diet $ count;
CARDS;
Y      High_No 27
Y      High_Fib       20
Y      Low_No  19
Y      Low_Fib 14
N      High_No 3
N      High_Fib       10
N      Low_No  11
N      Low_Fib 16
;
PROC FREQ DATA=tumors ORDER=DATA;
  WEIGHT count;
  TABLES tumors*diet / chisq expected nopercent;
RUN;
```

Frequency,					
Expected,					
Row Pct ,					
Col Pct	High_No	High_Fib	Low_No	Low_Fib	Total
Y	27	20	19	14	80
	20	**20**	**20**	**20**	
	33.75	25.00	23.75	17.50	
	90.00	66.67	63.33	46.67	
N	3	10	11	16	40
	10	**10**	**10**	**10**	
	7.50	25.00	27.50	40.00	
	10.00	33.33	36.67	53.33	
Total	30	30	30	30	120

Statistics for Table of tumors by diet

Statistic	DF	Value	Prob
Chi-Square	3	**12.9000**	**0.0049**

By SAS:

b) As can be seen from the highlighted portions of the contingency table above the expected values are all 20 for the cancer groups and 10 for the non-cancer groups.

c) As can be seen from the statistics table the χ^2 value is 12.900.

d) With a p-value of 0.0049 (compared to α=0.05) we reject the null hypothesis and find that they are not independent.

By Hand:

b) **The table of expected values is the row total times the column total divided by the grand total**

	Diet			
Cancer	High/No	High/Fib	Low/No	Low/Fib
Yes	80*30/120=20	80*30/120=20	80*30/120=20	80*30/120=20
No	40*30/120=10	40*30/120=10	40*30/120=10	40*30/120=10

c) $$\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp} = \frac{(27-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(14-20)^2}{20} + \frac{(3-10)^2}{10} + \frac{(10-10)^2}{10}$$

$$+ \frac{(11-10)^2}{10} + \frac{(16-10)^2}{10} = 2.45 + 0 + 0.05 + 1.8 + 4.9 + 0 + 0.1 + 3.6 = 12.9$$

d) We compare 12.9 (the χ^2) to 7.18473 (df=(4-1)(2-1)=3, α=0.05) and reject the null hypothesis (so we conclude they are not independent).

It is a test of <u>homogeneity</u> because the <u>column</u> totals of 30 are fixed. (Note... in this case the row totals were not fixed, but if you turned the table sideways they would have been!)