Homework 03
Joseph Blubaugh
jblubau1@tamu.edu
STAT 659-700

2.16

a) Response: Have Smoked (Yes, No), Explanatory: Lung Cancer (Cases, Controls)
b) Case-Control Study
c) Yes because the sample size is large
d) The odds ratio of cancer patients is 5.8% higher than cancer patients that do not smoke. The 95% confidence iterval of the odds ratio crosses 1 so there is not enough evidence to support that smoking and lung cancer are related.
Odds Ratio:$((688/709) * (21/709))/((21/709) * (650/709)) = 1.058$
95% Confidence: $log(1.058) \pm 1.96\sqrt{\frac{1}{688} + \frac{1}{650} + \frac{1}{21} + \frac{1}{58}} = (-.053, .967) \rightarrow (e^{-.053}, e^{1.76}) = (.634, 1.76)$

Table 1: Smokers and Lung Cancer

	Cancer	No.Cancer	Total
Smoker Yes	688	650	1338
Smoker No	21	58	79
Total	709	709	1418

2.18

a) $(168/1362) * (290/1362) * 1362 = 35.8$

Table 2: Happiness and Income

	Not.too.Happy	Pretty.Happy	Very.Happy	Total
Above Average	21	159	110	290
Average	53	372	221	646
Below Average	94	249	83	426
Total	168	780	414	1362

b) DF: $(3 - 1) * (3 - 1) = 4$, pvalue: `1 - pchisq(73.4, 4) = 0`
c) 21: The standardized error is nearly 3 times less than the standard error of the expected probability
83: The standardized error is nearly 6 times less than the standard error of the expected probability

d) 110: the standardized error is approximately 3.1 times larger than the standard error of the expected probability
94: the standardized error is approximately 7.5 times larger than the standardized error of the expected probability

e)

```
proc freq order=data data=happy; weight count;
tables happiness*income/chisq expected norow nocol nopercent deviation cmh1;
run;
```

Table 3: Output from SAS

Statistic	DF	Value	Pvalue
Chi-Squared	4	73.3	<.001
Likelihood Ratio Chi-Sq	4	71.3	<.001
CMH	1	55.9	<.001

f)

2.19

a) $X^2 = 167.83$, pvalue = 0, conclusion is that race and political affiliation are related

Table 4: Race vs Party Affiliation

	Democrat	Independent	Republican	Total
White	871	444	873	2188
Black	302	80	43	425
Total	1173	524	916	2613

Table 5: Expected Values

	Democrat	Independent	Republican
White	982.2	438.8	767
Black	190.8	85.23	149

Table 6: Chi-squared Values

	Democrat	Independent	Republican
White	12.59	0.06169	14.65
Black	64.81	0.3209	75.41

b) There are much fewer white democrats and black republicans than what is expected under the null hypothesis. There are also many more black democrats and white republicans than what is expected. The portion of independents is close to what is expected under the null hypothesis.

Table 7: Standardized Residuals

	Democrat	Independent	Republican
White	-11.85	0.7564	4.554
Black	11.85	-0.6924	-8.943

c) Conclude that republican and democrat affiliation by race are not independent however independent affiliation is independent `1 - pchisq(.3825, 1) = .536`

Table 8: Partitioned Chi-squares

	Chi.squared	DF
Democrat and Republican	167.5	1
Independent	0.3825	1
Total	167.8	2

2.20

```
data cancer;
input stage $ living $ count;
cards;
advanced spouse 109
advanced alone 59
advanced other 53
local spouse 100
local alone 85
local other 36
;

proc freq order=data data=cancer; weight count;
tables stage*living/chisq expected norow nocol nopercent deviation cmh1;
run;
```

Table 9: Output from SAS

Statistic	DF	Value	Pvalue
Chi-Square	2	8.32	0.0155

2.21

a) Yes the chi-squared independence test is appropriate for this table because all expected values are greater than 5 and the responses are not correlated.

b) You can compare the responses to the 3 questions by gender by computing the Chi-Squared statistic based on the expected values of each cell. Overall the difference in responses are not significantly different. You can also isolate the response to question A by breaking out the chi-squared values and comparing it to the combined answers of B and C. That relationship is also not significant.

Table 10: Data

Gender	A	B	C	Total
Male	60	81	75	216
Female	75	87	86	248
Total	135	168	161	464

Table 11: Expected Values

Gender	A	B	C
Male	62.84	78.21	74.95
Female	72.16	89.79	86.05

Table 12: Chi-squared values

Gender	A	B	C
Male	0.1248	0.1003	0.0001335
Female	0.1166	0.08731	0

Table 13: Contingency Table

Gender	A	BC
Male	0.1248	0.1004
Female	0.1166	0.08731

Test for whether Response A is significantly different: 1 - pchisq(.2414, 1) = 0.6231968

2.22

a) Since the not all of the expected frequencies are > 5 we can use Fishers Exact Test. Because the pvalue is near 0 we conclude that Drug use and Diagnosis are related.

Diagnosis	Schizophrenia	Affective	Neurosis	Personality	Special	Total
Drugs	105	12	18	47	0	182
No Drugs	8	2	19	52	13	94
Total	113	14	37	99	13	276

Table 15: Expected Frequencies

Diagnosis	Schizophrenia	Affective	Neurosis	Personality	Special
Drugs	74.51	9.232	24.4	65.28	8.572
No Drugs	38.49	4.768	12.6	33.72	4.428

```
    Fisher's Exact Test for Count Data

data:  x
p-value < 2.2e-16
alternative hypothesis: two.sided
```

b) The value for Schizophrenia is much higher than expected and much lower than expected for No Drug use suggesting that Drug use and schizophrenia are related. The opposite is true for Neurosis, Personality, and Special. The large standard errors in the table suggest that as a whole drug use and diagnosis are related.

Table 16: Standard Errors

Diagnosis	Schizophrenia	Affective	Neurosis	Personality	Special
Drugs	7.879	1.624	-2.386	-4.846	-5.148
No Drugs	-7.877	-1.574	2.386	4.848	5.172

c)

Table 17: Chi-Squared Values

Diagnosis	Schizophrenia	Affective	Neurosis	Personality	Special
Drugs	12.49	0.8522	1.679	3.585	8.6
No Drugs	24.16	1.551	3.251	9.937	16.81

Table 18: Partitioned Table (i)

Diagnosis	Schizophrenia	Affective
Drugs	12.49	0.8522
No Drugs	24.16	1.55

Table 19: Partitioned Table (ii)

Diagnosis	Schizophrenia	Affective
Drugs	12.49	0.8522
No Drugs	24.16	1.55

Table 20: Partitioned Table (iii)

Diagnosis	Schizophrenia	Affective
Drugs	12.49	0.8522
No Drugs	24.16	1.55

```
## Chi-Squared (i)
1 - pchisq(sum(dta.prt[, 2:3]), 1)
```

```
[1] 4.126239e-10
```

```
## Chi-Squared (ii)
1 - pchisq(sum(dta.prt2[, 2:3]), 1)
```

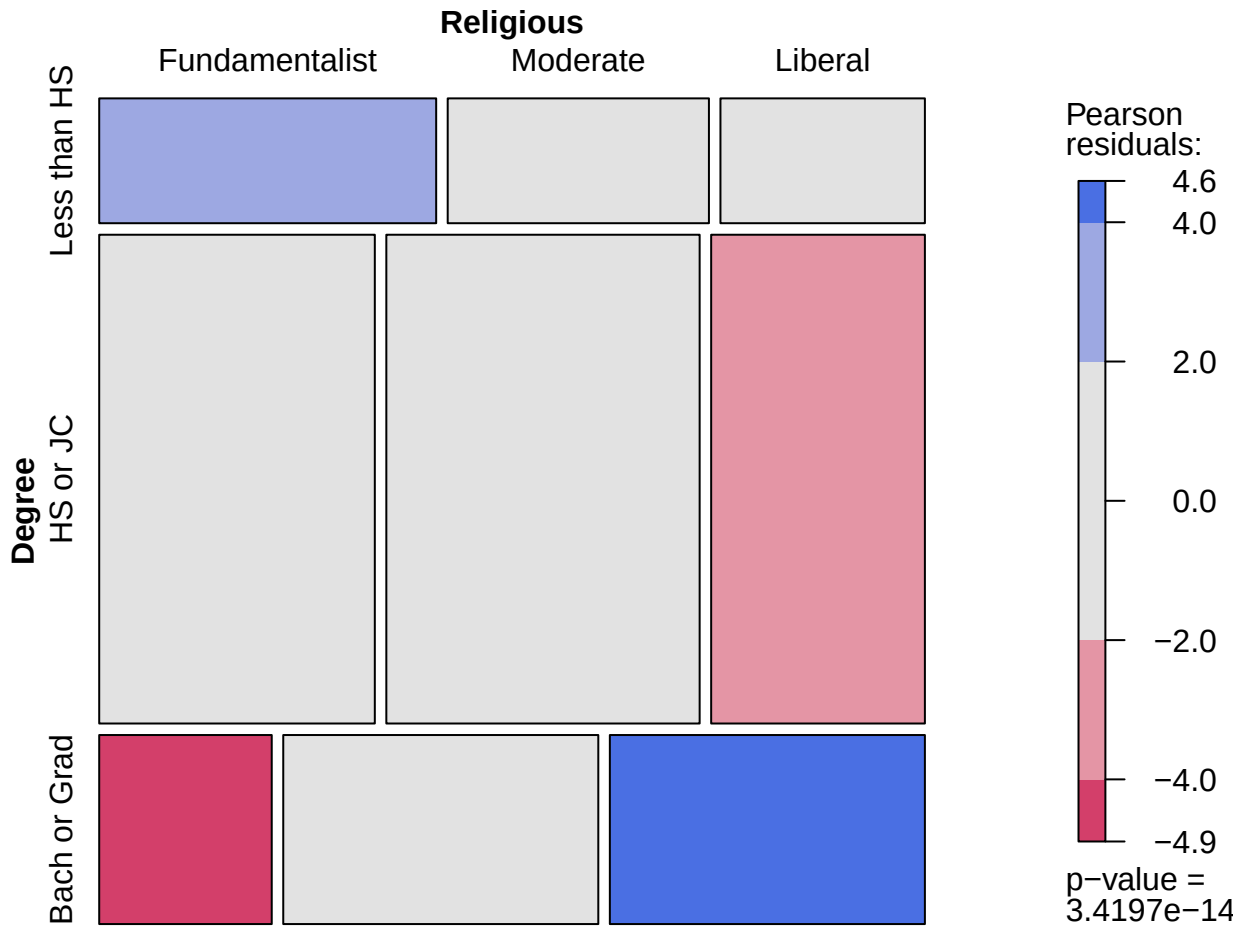
```
[1] 1.79058e-05
```

```
## Chi-Squared (iii)
1 - pchisq(sum(dta.prt3[, 2:4]), 2)
```

```
[1] 0
```

2.23

People with less than HS education tend to be heavily into the fundamental religious category. Those with HS or JC tend to be in the Fundamentalist or Moderate category. And those with a college degree tend to be liberal. It appears that the relationship between those with a bachelors degree or higher is significantly related to religious beliefs.



2.27

a)

Table 21: Apsiration and Income

.	Low	Medium	High	Total
< HS	9	11	9	29
HS	43	52	41	136
< Grad	13	23	12	48
Grad	10	22	27	59
Total	75	108	89	272

b)

Table 22: Standard Error

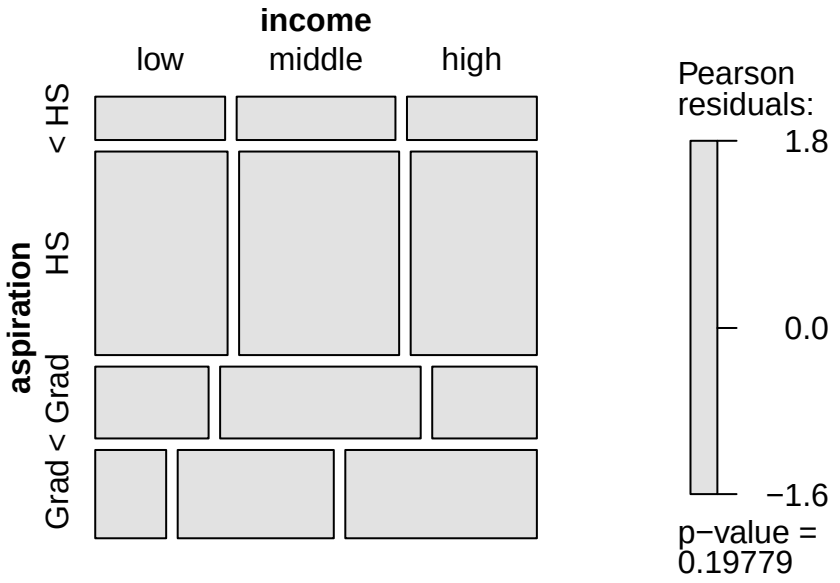
.	Low	Medium	High
< HS	0.4865	-0.2009	-0.2092
HS	1.494	-0.4957	-0.9046
< Grad	-0.07133	1.302	-1.255
Grad	-2.047	-0.4212	2.415

c)

Table 23: Exact Chi-Squared from SAS

Chi.Square	DF	Asymptotic.Pr...ChiSq	Exact.Pr....ChiSq
8.593	6	0.1978	0.199

d) The exact Chi-Squared test confirms that that relationship between aspiration and income is not significant and All of the standard errors of the expected values are reasonably close the actual values.



2.25

a)

Table 24: N

	n.1	n.2	Total
n1.	n11	n12	n1+
n2.	n21	n22	n2+
Total	n+1	n+2	N

Table 25: Mu

	mu.1	mu.2	Total
mu1.	n+1/N * n1+/N * N	n+2/N * n1+/N * N	n1+/N * N
mu2.	n+1/N * n2+/N * N	n+2/N * n2+/N * N	n2+/N * N
Total	n+1/N * N	n+2/N * N	N++/N * N

b)

$$\frac{(\frac{n_{+1}}{N}\frac{n_{1+}}{N}N)(\frac{n_{+2}}{N}\frac{n_{2+}}{N}N)}{(\frac{n_{+1}}{N}\frac{n_{2+}}{N}N)(\frac{n_{+2}}{N}\frac{n_{1+}}{N}N)} = 1$$

2.26

a) A Chi-squared distribution with DF degrees of freedom

b) The DF for the sum of two chi-squared variables is equal to the sum of the DF for each variable: $X_1^2 + X_2^2 = X_3^2$