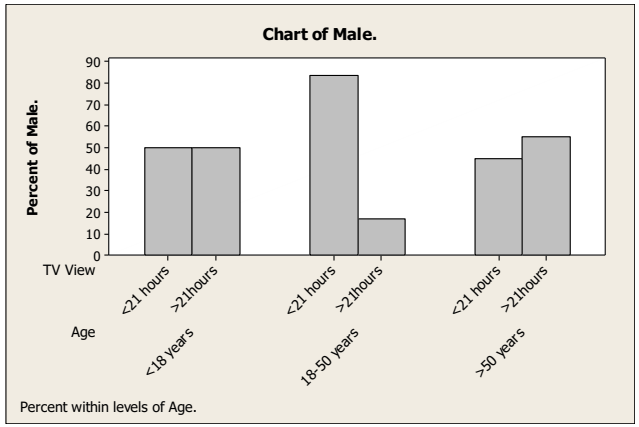
MATU9D2 : Practical Statistics Spring 2016

Solutions to Practical 4

Question 1 (i)

Males Only **Are TV Viewing and Age associated?**

Subjective Impression From the plot below of the percentages within each age group we can see wildly differing % of those watching less than 21 hours of TV a week. This suggests an association between Age and TV Viewing for Males



Chi-Square Test: <18 years, 18-50 years, >50 years

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	<18 years	18-50 years	>50 years	Total
1	10	20	9	39
	12.19	14.63	12.19	
	0.393	1.975	0.834	
2	10	4	11	25
	7.81	9.38	7.81	
	0.613	3.082	1.301	
Total	20	24	20	64

Chi-Sq = 8.196, DF = 2, P-Value = 0.017

Formal Test : χ^2 *Test* of Independence

H_o : TV Viewing and Age Independent

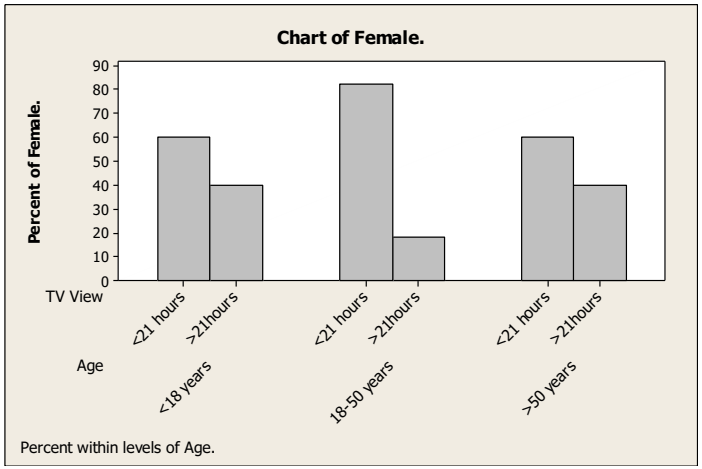
H_1 :TV Viewing and Age associated

Test Statistic $X^2 = 8.196$ p=0.017

Can reject Ho in favour of H1 at 5% level (p<0.05) so evidence of an association between Age and TV Viewing for Males

Females Only **Are TV Viewing and Age associated?**

Subjective Impression From the plot below of the percentages within each age group we can see more similar % for those watching less than 21 hours of TV a week. This suggests no association between Age and TV Viewing for Females



Chi-Square Test: <18 years., 18-50 years., >50 years.

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	<18 years.	18-50 years.	>50 years.	Total
1	18	18	12	48
	20.00	14.67	13.33	
	0.200	0.758	0.133	
2	12	4	8	24
	10.00	7.33	6.67	
	0.400	1.515	0.267	
Total	30	22	20	72

Chi-Sq = 3.273, DF = 2, P-Value = 0.195

Formal Test : χ^2 *Test* of Independence

H_o : TV Viewing and Age Independent

H_1 :TV Viewing and Age associated

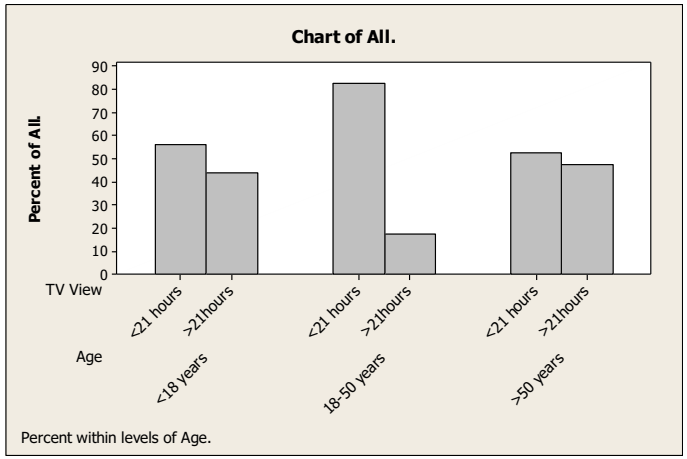
Test Statistic $X^2 = 3.273$ p=0.195

Cannot reject Ho in favour of H1 at 5% level (p>0.05) so insufficient evidence of an association between Age and TV Viewing for Females

(ii) Combined i.e. Males & Females Together

Are TV Viewing and Age associated?

Subjective Impression From the plot below of the percentages within each age group we can see wildly differing % of those watching less than 21 hours of TV a week. This suggests an association between Age and TV Viewing.



Chi-Square Test: < 18 years, 18 - 50 years, > 50 years

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

        < 18  18 - 50   > 50
       years    years  years  Total
1         28       38     21     87
       31.99    29.43  25.59
       0.497    2.498  0.823

2         22        8     19     49
       18.01    16.57  14.41
       0.882    4.435  1.461

Total     50       46     40    136

Chi-Sq = 10.595, DF = 2, P-Value = 0.005
```

Formal Test : χ^2 *Test* of Independence

H_o : TV Viewing and Age Independent

H_1 :TV Viewing and Age associated

Test Statistic $X^2 = 10.595$ p=0.005

Can reject Ho in favour of H1 at 1% level (p<0.01) so sufficient evidence of an association between Age and TV Viewing.

(iii) Part (ii) shows that if the data is a representative sample of the population then there is a statistically significant association between Age and amount of TV views for adults. However, part (i) shows that this association is only reflected in the results for adult males and that there is insufficient evidence of an association amongst the females.

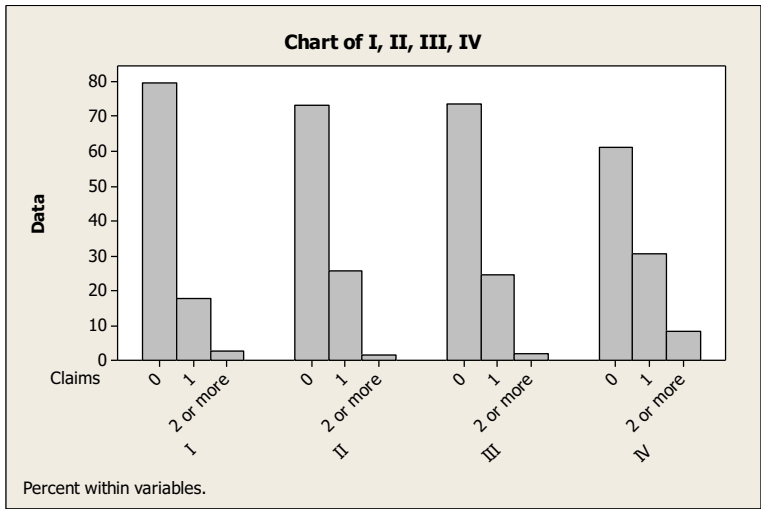
(iv) Question asked for a 95% confidence interval for the proportion of the population who watched more than 21 hours of TV per week. 49 males and females watched more than 21 hours out of a total of 136 people sampled. So from output below, 95% confident that the proportion of people who watch more than 21 hours of TV per week lies between 0.28 and 0.45.

Test and CI for One Proportion

```
Sample   X    N  Sample p          95% CI
1       49  136  0.360294  (0.279805, 0.447011)
```

Question 2 Are Number of Claims and Insurance Group associated?

Subjective Impression From the plot below of the percentages within each insurance group we can see differing % of those making 0, 1 or 2 or more claims. This suggests an association between Number of Claims and Insurance Group.



Chi-Square Test: I, II, III, IV

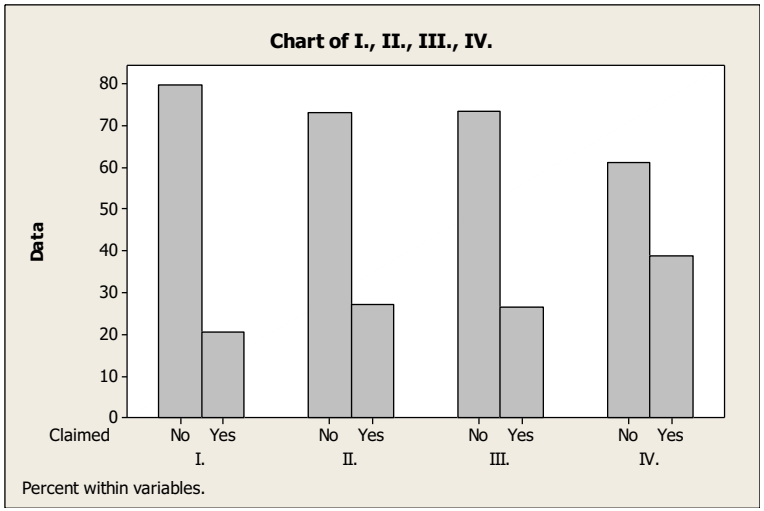
```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

             I       II      III       IV  Total
1          900     2000     1500       30   4430
         840.06  2036.95  1516.56    36.43
          4.277    0.670    0.181    1.134

2          200      700      500       15   1415
         268.33   650.63   484.41    11.64
         17.398    3.746    0.502    0.973

3           30       40       40        4    114
         21.62    52.42    39.03     0.94
         3.250    2.942    0.024   10.006

Total     1130     2740     2040       49   5959

Chi-Sq = 45.104, DF = 6
WARNING: 1 cells with expected counts less than 1. Chi-Square approximation
     probably invalid.
1 cells with expected counts less than 5.
```

Chisquared Test invalid because of low expected value so must combine rows and/or columns to increase expected values

Are Number of Claims and Insurance Group associated?

Combine rows/columns to increase expected values : chose to add those making 1 and 2 or more claims.

Subjective Impression From the plot below of the percentages within each insurance group we can see differing % of those making no claims. This suggests an association between whether you make a claim or not and Insurance Group.



Chi-Square Test: I., II., III., IV.

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

             I.      II.     III.      IV.  Total
1          900     2000     1500       30   4430
         840.06  2036.95  1516.56    36.43
          4.277    0.670    0.181    1.134

2          230      740      540       19   1529
         289.94   703.05   523.44    12.57
         12.393    1.942    0.524    3.286

Total     1130     2740     2040       49   5959

Chi-Sq = 24.407, DF = 3, P-Value = 0.000
```

Formal Test : χ^2 *Test* of Independence

H_o : Insurance Group and making a claim are independent

H_1 : Insurance Group and making a claim are associated

Test Statistic $X^2 = 24.407$ $p<0.001$

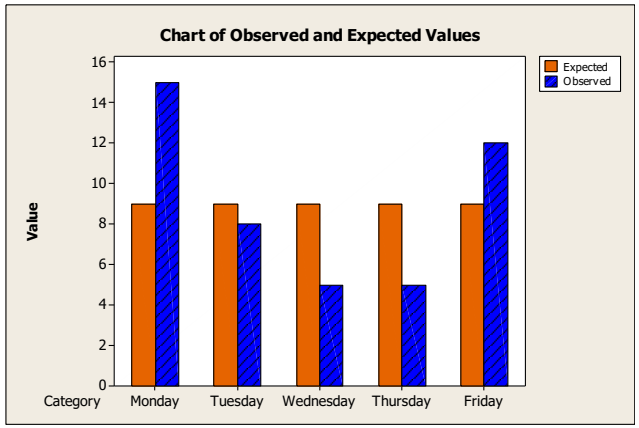
Can reject Ho in favour of H1 at 1% level ($p<0.01$) so sufficient evidence of an association between Insurance Group and making a claim

Question 3

Question asked whether the Number of Defective Items was due to chance.
This is more formally described as: Are the Number of Defective Items Uniformly distributed?

If Uniformly distributed we would expect the same number each day.

Subjective Impression The number (in blue) look fairly different each day. However, this may not be statistically significant.



Chi-Square Goodness-of-Fit Test for Observed Counts in Variable: Defectives

Using category names in Day

Category	Observed	Test Proportion	Expected	Contribution to Chi-Sq
Monday	15	0.2	9	4.00000
Tuesday	8	0.2	9	0.11111
Wednesday	5	0.2	9	1.77778
Thursday	5	0.2	9	1.77778
Friday	12	0.2	9	1.00000

N	DF	Chi-Sq	P-Value
45	4	8.66667	0.070

Formal Test : χ^2 *Test* : Goodness of Fit

H_o : Number of defectives follow a Uniform Distribution

H_1 : Number of defectives follow some other distribution

Test Statistic $X^2 = 8.667$ p=0.070

Cannot reject Ho in favour of H1 at 5% level (p>0.05) so insufficient evidence that number of defectives follow some other distribution.

So can assume Data follows a Uniform distribution