HW8

Problem 1.

CATEGORIES	1	2	3
1	0.13	0.19	0.28
2	0.07	0.11	0.22

<u>Part a)</u>

CATEGORIES	1	2	3	$N_{I.}$
1	13	19	28	60
2	7	11	22	40
$N_{.J}$	20	30	50	100

$$\widehat{e_{11}} = \frac{60 * 20}{100} = 12$$

We repeat this step for all the elements in the table.

$\widehat{e_{11}}$	1	2	3
1	12	18	30
2	8	12	20

Now we do:

$$\frac{(13 - 12)^2}{12} = 0.0833$$

We repeat this step for all the elements in the table

	1	2	3
1	0.0833	0.0556	0.1333
2	0.125	0.0833	0.2

If we add all the elements, we get that the chi-square is:

$$\chi^2 = 0.6806$$

The degrees of freedom are: df= (2-1) (3-1) =1*2=2

$$df = 2$$

With a significance level of 0.1 we can get the p-value:

$$p - value = 0.7116$$

Since

$$\chi^2_{\alpha,df} = \chi^2_{0.1,2} = 4.605$$

The rejection region is

$$\chi^2 \geq \chi^2_{\alpha,df}$$

And it doesn't hold:

$$0.6806 \geq 4.605$$

So, we don't reject H_o

Part b)

CATEGORIES	1	2	3	$N_{I.}$
1	130	190	280	600
2	70	110	220	400
$N_{.J}$	200	300	500	1000

$$\widehat{e_{11}} = \frac{600 * 200}{1000} = 120$$

We repeat this step for all the elements in the table.

$\widehat{e_{11}}$	1	2	3
1	120	180	300
2	80	120	200

Now we do:

$$\frac{(130 - 120)^2}{120} = 0.833$$

We repeat this step for all the elements in the table

	1	2	3
1	0.833	0.556	1.333
2	1.25	0.833	2

If we add all the elements, we get that the chi-square is:

$$\chi^2 = 6.806$$

The degrees of freedom are: df= (2-1) (3-1) =1*2=2

$$df = 2$$

With a significance level of 0.1 we can get the p-value:

$$p - value = 0.03328$$

Since

$$\chi^2_{\alpha,df} = \chi^2_{0.1,2} = 4.605$$

The rejection region is

$$\chi^2 \geq \chi^2_{\alpha,df}$$

And it holds:

$$6.806 \geq 4.605$$

So, we reject H_o

Part c)

For us to reject the independence, χ^2 needs to be 4.605 or more. This value comes from the next formula:

$$\chi^2 = \sum_{ij} \frac{\left(n_{ij} - \widehat{e_{ij}}\right)^2}{\widehat{e_{ij}}}$$

Since the estimated expected comes from this other formula:

$$\widehat{e_{ij}} = \frac{n_{i*} * n_{*j}}{n}$$

$$n_{i*} = \sum_{j=1}^{J} n_{ij} \qquad n_{*j} = \sum_{i=1}^{I} n_{ij}$$

If we substitute we get:

$$\chi^2 = \sum_{ij} \frac{\left(n_{ij} - \frac{\sum_{j=1}^{J} n_{ij} * \sum_{i=1}^{I} n_{ij}}{n}\right)^2}{\frac{\sum_{j=1}^{J} n_{ij} * \sum_{i=1}^{I} n_{ij}}{n}}$$

Since we have the proportions of each category, base on the first table

$$n_{ij} = n * p_{ij}$$

p_{ij}	1	2	3	p_{i*}
1	0.13	0.19	0.28	0.6
2	0.07	0.11	0.22	0.4
p_{*j}	0.2	0.3	0.5	

$$\chi^2 = \sum_{ij} \frac{\left(n * p_{ij} - \frac{\sum_{j=1}^{J} n * p_{ij} * \sum_{i=1}^{I} n * p_{ij}}{n}\right)^2}{\frac{\sum_{j=1}^{J} n * p_{ij} * \sum_{i=1}^{I} n * p_{ij}}{n}}$$

We want to solve for n, given that $\chi^2 = 4.605$

$$\sum_{ij} \frac{\left(n * p_{ij} - \frac{\sum_{j=1}^{J} n * p_{ij} * \sum_{i=1}^{I} n * p_{ij}}{n}\right)^2}{\frac{\sum_{j=1}^{J} n * p_{ij} * \sum_{i=1}^{I} n * p_{ij}}{n}} = \sum_{ij} \frac{\left(n * p_{ij} - \frac{n * \sum_{j=1}^{J} p_{ij} * n * \sum_{i=1}^{I} p_{ij}}{n}\right)^2}{\frac{n * \sum_{j=1}^{J} p_{ij} * n * \sum_{i=1}^{I} p_{ij}}{n}}$$

$$\sum_{ij} \frac{\left(n * p_{ij} - n * \sum_{j=1}^{J} p_{ij} * \sum_{i=1}^{I} p_{ij}\right)^2}{n * \sum_{j=1}^{J} p_{ij} * \sum_{i=1}^{I} p_{ij}} = \sum_{ij} \frac{\left(n * \left(p_{ij} - \sum_{j=1}^{J} p_{ij} * \sum_{i=1}^{I} p_{ij}\right)\right)^2}{n * \sum_{j=1}^{J} p_{ij} * \sum_{i=1}^{I} p_{ij}} =$$

$$\sum_{ij} \frac{n^2 * \left(p_{ij} - \sum_{j=1}^{J} p_{ij} * \sum_{i=1}^{I} p_{ij}\right)^2}{n * \sum_{j=1}^{J} p_{ij} * \sum_{i=1}^{I} p_{ij}} = \sum_{ij} \frac{n * \left(p_{ij} - \sum_{j=1}^{J} p_{ij} * \sum_{i=1}^{I} p_{ij}\right)^2}{\sum_{j=1}^{J} p_{ij} * \sum_{i=1}^{I} p_{ij}} =$$

$$n * \sum_{ij} \frac{\left(p_{ij} - \sum_{j=1}^{J} p_{ij} * \sum_{i=1}^{I} p_{ij}\right)^2}{\sum_{j=1}^{J} p_{ij} * \sum_{i=1}^{I} p_{ij}}$$

$$p_{i*} = \sum_{j=1}^{J} p_{ij} \qquad p_{*j} = \sum_{i=1}^{I} p_{ij}$$

$$n * \sum_{ij} \frac{\left(p_{ij} - p_{i*} * p_{*j}\right)^2}{p_{i*} * p_{*j}}$$

Now we have:

$$\chi^2 = n * \sum_{ij} \frac{\left(p_{ij} - p_{i*} * p_{*j}\right)^2}{p_{i*} * p_{*j}}$$

We clear for n:

$$n = \frac{\chi^2}{\sum_{ij} \frac{\left(p_{ij} - p_{i*} * p_{*j}\right)^2}{p_{i*} * p_{*j}}}$$

Now we solve the equation:

$$\frac{(p_{11} - p_{1*} * p_{*1})^2}{p_{1*} * p_{*1}} = \frac{(0.13 - 0.6 * 0.2)^2}{0.6 * 0.2} = 0.00083$$

We repeat this for all the elements in the table

	1	2	3
1	0.00083	0.00056	0.0013
2	0.00125	0.00083	0.002

We add them all together and get:0.00680556

$$\sum_{ij} \frac{\left(p_{ij} - p_{i*} * p_{*j}\right)^2}{p_{i*} * p_{*j}} = 0.00680556$$

$$n = \frac{\chi^2}{\sum_{ij} \frac{\left(p_{ij} - p_{i*} * p_{*j}\right)^2}{p_{i*} * p_{*j}}} = \frac{4.605}{0.00680556} = 676.653$$

Since n needs to be a Natural number, we need to make the sample size of 677 so that we reject the independence hypothesis.

(All the operations are done on the excel)

Problem 2.

	Males	Females	n_{i*}
<16	25	10	35
16-17	24	32	56
18-20	28	17	45
>21	19	34	53
N_{*j}	96	93	189

$$\widehat{e_{11}} = \frac{35 * 96}{198} = 17.78$$

We repeat this step for all the elements in the table.

e_{ij}	Males	Females
<16	17.78	17.22
16-17	28.44	27.56
18-20	22.85	22.14
>21	26.92	26.07

Now we do:

$$\frac{(25 - 17.78)^2}{17.78} = 2.93$$

We repeat this step for all the elements in the table

	Males	Females
<16	2.93	3.02
16-17	0.69	0.71
18-20	1.15	1.19
>21	2.33	2.4

If we add all the elements, we get that the chi-square is:

$$\chi^2 = 14.4616$$

The degrees of freedom are: df= (4-1) (2-1) =3*1=3

$$df = 3$$

With a significance level of 0.1 we can get the p-value:

$$p - value = 0.00234$$

Since

$$\chi^2_{\alpha,df} = \chi^2_{0.1,3} = 6.251$$

The rejection region is

$$\chi^2 \geq \chi^2_{\alpha,df}$$

And it holds:

$$14.4616 \geq 6.251$$

So, we reject H_o

(All the operations are done on the excel)

Problem 3.

Same Rating [(G, G), (B, B), (S, S)]: 69

Differs by one [(G, S), (S, G), (S, B), (B, S)]: 102

Differs by two [(G, B), (B, G)]: 45

n=69+102+45=216

Each pair have a 1/9 probability

$$p_1 = \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{3}{9}$$

$$p_2 = \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{4}{9}$$

$$p_3 = \frac{1}{9} + \frac{1}{9} = \frac{2}{9}$$

$$H_o: p_1 = \frac{3}{9}, \qquad p_2 = \frac{4}{9}, \qquad p_3 = \frac{2}{9}$$

$$H_a: otherwise$$

$$\chi^2 = \sum_{i=1}^{3} \frac{(n_i - n * p_i)^2}{n * p_i}$$

$$\chi^2 = \frac{\left(69 - 216 * \frac{3}{9}\right)^2}{216 * \frac{3}{9}} + \frac{\left(102 - 216 * \frac{4}{9}\right)^2}{216 * \frac{4}{9}} + \frac{\left(45 - 216 * \frac{2}{9}\right)^2}{216 * \frac{2}{9}}$$

$$\chi^2 = \frac{(-3)^2}{72} + \frac{(6)^2}{96} + \frac{(-3)^2}{48} = 0.125 + 0.375 + 0.1875 = 0.6875$$

$$\chi^2 = 0.6875$$

The df: 3-1=2

$$p - value = 0.7091062$$

The H_o holds and we cannot reject it.

Since

$$\chi^2_{\alpha,df} = \chi^2_{0.1,2} = 4.605$$

The rejection region is

$$\chi^2 \geq \chi^2_{\alpha,df}$$

And doesn't hold:

$$0.6875 \geq 4.605$$

Problem 4.

$$f(x) = \lambda * e^{-\lambda * x}$$

For the case that λ is 1 for both variables we get:

$$f(x) = e^{-x}$$

$$f(y) = e^{-y}$$

$$f(x, y) = e^{-x-y}$$

$$U = X, V = \frac{X}{Y}$$

$$X = U, Y = \frac{U}{V}$$

Let's do a Jacobian transformation

$$|J| = \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ \frac{1}{v} & -\frac{u}{v^2} \end{vmatrix} = -\frac{u}{v^2}$$

$$f_{uv}(u, v) = f_{xy}\left(u, \frac{u}{v}\right) * \left|-\frac{u}{v^2}\right| = e^{-u-\frac{u}{v}} * \left|-\frac{u}{v^2}\right|$$

$$f_v(v) = \int_0^{\infty} e^{-u-\frac{u}{v}} * \left|-\frac{u}{v^2}\right| du = \int_0^{\infty} e^{-u-\frac{u}{v}} * \frac{u}{v^2} du =$$

To do the integration I used an Integral Calculator

$$= \left[-\frac{((v+1)u + v)e^{-\frac{u}{v}-u}}{v^3 + 2v^2 + v}\right]_0^{\infty} = \frac{1}{\left(\frac{1}{v}+1\right)^2 v^2} = \frac{1}{(v+1)^2}$$

$$(v+1)^{-2} \sim F_{2,2}$$

For λ= 1 we get that it follows a F distribution with 2,2 degrees of freedom.

Now for other values of λ:

$$f(x) = \lambda_1 * e^{-\lambda_1 * x}$$

$$f(y) = \lambda_2 * e^{-\lambda_2 * y}$$

$$f_{uv}(u, v) = f_{xy}\left(u, \frac{u}{v}\right) * \left|-\frac{u}{v^2}\right| = \lambda_1 * e^{-\lambda_1 * u} * \lambda_2 * e^{-\lambda_2 * \frac{u}{v}} * \left|-\frac{u}{v^2}\right| =$$

$$\lambda_1 * \lambda_2 * e^{-\lambda_1 * u - \lambda_2 * \frac{u}{v}} * \left|-\frac{u}{v^2}\right|$$

$$f_v(v) = \int_0^\infty \lambda_1 * \lambda_2 * e^{-\lambda_1 * u - \lambda_2 * \frac{u}{v}} * \frac{u}{v^2} du = \int_0^\infty \lambda_1 * \lambda_2 * e^{-u\left(\lambda_1 + \lambda_2 * \frac{1}{v}\right)} * \frac{u}{v^2} du =$$

To do the integration I used an Integral Calculator

$$= \left[-\frac{\lambda_2 \lambda_1 * \left((\lambda_1 v + \lambda_2) u + v \right) e^{-\frac{\lambda_2 u}{v} - \lambda_1 u}}{v * \left({\lambda_1}^2 * v^2 + 2\lambda_2 \lambda_1 v + {\lambda_2}^2 \right)} \right]_0^\infty = \frac{\lambda_2 \lambda_1}{\left(\frac{\lambda_2}{v} + \lambda_1 \right)^2 v^2} = \frac{\lambda_2 \lambda_1}{(\lambda_1 v + \lambda_2)^2} =$$

$$\lambda_2 \lambda_1 * (\lambda_1 v + \lambda_2)^{-2}$$

Since

$$E[X] = \frac{1}{\lambda} = \bar{X}$$

$$\lambda_2 \lambda_1 * (\lambda_1 v + \lambda_2)^{-2} = \frac{1}{\bar{Y}} \frac{1}{\bar{X}} * \left(\frac{1}{\bar{X}} v + \frac{1}{\bar{Y}} \right)^{-2} = \frac{1}{\frac{\sum_{j=1}^n y_j}{n}} \frac{1}{\frac{\sum_{i=1}^m x_i}{m}} * \left(\frac{1}{\frac{\sum_{i=1}^m x_i}{m}} v + \frac{1}{\frac{\sum_{j=1}^n y_j}{n}} \right)^{-2} =$$

$$\frac{n}{\sum_{j=1}^n y_j} * \frac{m}{\sum_{i=1}^m x_i} * \left(\frac{m}{\sum_{i=1}^m x_i} v + \frac{n}{\sum_{j=1}^n y_j} \right)^{-2} \sim F_{2*n, 2*m}$$