Q1. The table below is based on records of accidents in 1988 compiled by the Department of Highway Safety and Motor Vehicles in Florida. The analyst would like to know if there is a relationship between Injury type and seatbelts
a. Would you consider this an experiment? Do you think that these data were collected as part of an experiment? Explain
b. Give the null and alternative for the test
c. Use the chi-square test to run the test at α = 0.05
d. Give the real-world answer
e. Determine the expected counts for each cell and comment on the validity of the Chi-Square test

Solution:
a)
This is not an experiment. The data was not collected as a part of an experiment. This falls under the category of retrospective observational study. This is a study based on the previously collected or historical data. The records were pulled out in 1988 and then classified according to the requirement of the department.

b)The null hypothesis is given as follows - H_0: $X$is independent of Y
The alternate hypothesis is given as follows - H_A: $X$is nor independent of Y

c) Given alpha value is 0.05

$$\chi^2_{stat} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

$$E_{i,j} = \frac{R_i C_j}{N}$$

To check is each count is greater than 5

$$E_{1,1} = \frac{164128 * 2111}{577006} = 600.468$$

$$E_{1,2} = \frac{164128 * 574895}{577006} = 163527.531$$

$$E_{2,1} = \frac{412878 * 2111}{577006} = 1510.531$$

$$E_{2,2} = \frac{412878 * 574895}{577006} = 411367.468$$

Hence we can deduce that every count from the above value is greater than 5.
Using the formula, we get

$$\chi^2_{stat} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

$\chi^2_{stat} = \frac{(1601 - 600.468)^2}{600.468} + \frac{(162527 - 163527.531)^2}{163527.531} + \frac{(510 - 1510.531)^2}{1510.531} + \frac{(412368 - 411367.468)^2}{411367.468}$

χ^2_{stat}= 1667.140 + 6.121 + 662.722 + 2.433 = 2338.416

$df = (i - 1)(j - 1)\ (2 - 1)(2 - 1) = 1\ p - value = P(\chi^2_{df} \geq \chi^2_{stat})$

$= 1 - P(\chi^2 \geq \chi^2_{stat})$ = 0

Therefore, we get that 0 < 0.05, hence we reject H_0 and we accept the alternative hypothesis H_A

d) From the above we can state there is a relationship between injury type and the seatbelts.

e) $E_{1,1} = \frac{164128 * 2111}{577006}$ = 600.468

$E_{1,2} = \frac{164128 * 574895}{577006}$ = 163527.531

$E_{2,1} = \frac{412878 * 2111}{577006}$ = 1510.531

$E_{2,2} = \frac{412878 * 574895}{577006}$ = 411367.468

Each of the expected counts displayed above are greater than 5. Hence we can check the validity for the Chi-Square test. We can validate the Chi-Square test for the above solution.

Q2. It is expected that the true blood types proportions for Americans is given in the table below. For a sample of 200 Americans, their blood types were tested and also are given in the table below.
a. Give the null and alternative hypotheses for this test
b. Test this hypothesis at a=0.05
c. Give the real-world answer
d. Check to see if the test that you used in part b is justified.

Solution:
a) The null hypothesis is as follows - $H_0: \pi_A = 0.41$

$\pi_B = 0.1$

$\pi_{AB} = 0.04$

$\pi_0 = 0.45$

The alternative hypothesis is as follows - $H_A = \pi_i \neq \pi_{i,0}$ for i

b) Given that $\alpha = 0.05$, testing hypothesis as:

$\chi^2_{stat} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ and $E_{i,j} = \frac{R_i C_j}{N}$

We calculate the expected count as follows,

$E_1 = 200 * 0.41 = 82$

$E_2 = 200 * 0.1 = 20$

$E_3 = 200 * 0.04 = 8$

$E_4 = 200 * 0.45 = 90$

Each expected count is greater than 5. Formulating Chi-Square test we get,

$\chi^2_{stat} = \frac{(89-82)^2}{82} + \frac{(18-20)^2}{20} + \frac{(12-8)^2}{8} + \frac{(81-90)^2}{90}$

χ^2_{stat} = 0.5975 + 0.2 + 2 + 0.9 = 3.6975

$df = (a - 1) = 4 - 1 = 3$

$p - value = P(\chi^2_{df} \geq \chi^2_{stat})$

$= 1 - P(\chi^2 \leq \chi^2_{stat})$ = 0.2957

p-value= 0.2957

From the above result we can conclude that 0.295 > 0.05, hence we fail to reject the null hypothesis H_0 and there is enough evidence to support alternative hypotheses.

c) The real world answer for the above solution is that the probabilities of the true blood types proportions as mentioned in the above are viable. From the above solution we have enough evidence to support the alternative hypothesis.

d) The expected count of each is shown as follows :

$E_1 = 200 * 0.41 = 82$

$E_2 = 200 * 0.1 = 20$

$E_3 = 200 * 0.04 = 8$

$E_4 = 200 * 0.45 = 90$

We can justify that each expected count is greater than 5 hence we can validate the Chi-Square Test.

Q3. Refer to the table below that compares party identification and race. Test to see if there is a relationship between party identification and race. Do all of the checks needed and give the real-world results at the end.

Solution:

The null H_0 hypothesis and alternative H_A hypothesis is given as follows:

a) H_0: X is independent of Y

H_A: $X$is not independent of Y

b) Given that $\alpha = 0.05$, N = 980

$\chi^2_{stat} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ and $E_{i,j} = \frac{R_i C_j}{N}$

$E_{1,1} = \frac{129 * 444}{980}$= 58.44

$E_{1,2} = \frac{129 * 120}{980}$= 15.795

$E_{1,3} = \frac{129 * 416}{980}$= 54.759

$E_{2,1} = \frac{851 * 444}{980}$= 385.55

$E_{2,2} = \frac{851 * 120}{980}$= 104.204

$E_{2,3} = \frac{851 * 416}{980}$= 361.240

Each of the expected counts is greater than 5, hence we can validate the Chi-Square Test.

$$\chi^2_{stat} = \frac{(103-58.44)^2}{58.44} + \frac{(15-15.795)^2}{15.795} + \frac{(11-54.759)^2}{54.759} + \frac{(341-385.55)^2}{385.55} + \frac{(105-104.204)^2}{104.20} + \frac{(405-361.240)^2}{361.240}$$

χ^2_{stat}= 33.968 + 0.040 + 34.968 + 5.148 + 0.006 + 5.3010 = 79.431

$df = (i - 1)(j - 1)\ (2 - 1)(3 - 1) = 2$

$p - value = P(\chi^2_{df} \geq \chi^2_{stat})$

$1 - P(\chi^2 \leq \chi^2_{stat}) = 0$

Since 0 < 0.05 we reject the null hypothesis H_0 and there is enough evidence to support alternative hypotheses.

Real world answer : According to the above calculation,we can conclude that there is a relationship between race and party identification.