1. Is gender independent of education level? A random sample of 395 people were surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in the following table:

	High School	Bachelors	Masters	Ph.d.	Total
Female	60	54	46	41	201
Male	40	44	53	57	194
Total	100	98	99	98	395

Are gender and education level dependent at 5% level of significance? In other words, given the data collected above, is there a relationship between the gender of an individual and the level of education that they have obtained?

Ans: The null hypothesis is that "gender and education are not related"

We know,

$$\text{Expected Value} = \frac{(\text{Row Total}) * (\text{Column Total})}{\text{Total Number of Observations}}$$

Table with Expected values

	High School	Bachelors	Masters	Ph.d	Total
Female	60 (50.886)	54 (49.8683)	46 (50.3772)	41 (49.8734)	201
Male	40 (49.1139)	44 (48.1316)	53 (48.6227)	57 (48.1316)	194
Total	100	98	99	98	395

Degree of freedom = $3 \times 1 = 3$

We know,

$$\chi^2 = \frac{\text{Observed value} - \text{Expected value})^2}{\text{Expected value}}$$

$$= \frac{(O-E)^2}{E}$$

Table with $(O-E)^2/E$

	High School	Bachelors	Masters	Ph.d	Total
Female	1.6823	0.3423	0.3803	1.5770	
Male	1.6912	0.3546	0.3940	1.6340	
~~Total~~	~~100~~				

$$\chi^2 = 1.6823 + 0.3423 + 0.3803 + 1.5770 + 1.6912 + 0.3546 + 0.3940 + 1.6340$$

$$= 8.0057$$

Now according to Chi square Table at degree of freedom 3 & at 5% significance the value is 7.81 which is less than our calculated Chi square value hence our Null hypothesis is rejected.

Hence, gender and education level are dependent at 5% level of significance.

2. The sales of a company (in million dollars) for each year are shown in the table below.

x (year)	2005	2006	2007	2008	2009
y (sales)	12	19	29	37	45

a) Find the least square regression line $y = ax + b$.

b) Use the least squares regression line as a model to estimate the sales of the company in 2012.

Sol

a) Let the year 2005 be represented by 0. We put $t = x - 2005$ for simple calculation.

The table becomes,

t	0	1	2	3	4
y	12	19	29	37	45

t	y	ty	t^2
0	12	0	0
1	19	19	1
2	29	58	4
3	37	111	9
4	45	180	16

$\Sigma t = 10 \quad \Sigma y = 142 \quad \Sigma ty = 368 \quad \Sigma x^2 = 30$

$$a = (n\Sigma ty - \Sigma t \Sigma y) / (n\Sigma t^2 - (\Sigma t)^2)$$
$$= (5 \times 368 - 10 \times 142) / (5 \times 30 - 10^2)$$
$$= 8.4$$

$$b = \left(\frac{1}{n}\right)(\Sigma y - a \Sigma t)$$
$$= \left(\frac{1}{5}\right)(142 - 8.4 \times 10) = 11.6$$

$y = 8.4t + 11.6$ or $y = 8.4(x - 2005) + 11.6$

b) In 2012, $t = 2012 - 2005 = 7$

The estimate sales in 2012 were:

$y = 8.4 \times 7 + 11.6$

$= 70.4$ million dollars