HW8 solution

Ch 9 - #16: Records pertaining to the monthly number of job-related injuries at an underground coal mine were being studied by a federal agency. The values for the past 100 months were as follows:

Injuries per month	frequency of occurence
0	35
1	40
2	13
3	6
4	4
5	1
6	1

(a) apply the chi-square test to these data to test the hypothesis that the underlying distribution is Poisson. Use a level of significance of α=.05

Find the mean value of the distribution:

$X_0 := 35$ $X_1 := 40$ $X_2 := 13$ $X_3 := 6$ $X_4 := 4$

$X_5 := 1$ $X_6 := 1$

$$m := \frac{\sum_{j=0}^{6} (j \cdot X_j)}{\sum_{j=0}^{6} X_j} \qquad m = 1.11$$

The Mathcad function dpois(x,y) returns the P(X=x) for Poisson distribution with mean y

$x := 0..6$ $p_x := \text{dpois}(x, m)$ $O_x := X_x$ $E_x := 100 \cdot p_x$ $T_x := \frac{(O_x - E_x)^2}{E_x}$

$$x = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{pmatrix} \quad p = \begin{pmatrix} 0.33 \\ 0.366 \\ 0.203 \\ 0.075 \\ 0.021 \\ 4.628 \times 10^{-3} \\ 8.561 \times 10^{-4} \end{pmatrix} \quad O = \begin{pmatrix} 35 \\ 40 \\ 13 \\ 6 \\ 4 \\ 1 \\ 1 \end{pmatrix} \quad E = \begin{pmatrix} 32.956 \\ 36.581 \\ 20.302 \\ 7.512 \\ 2.085 \\ 0.463 \\ 0.086 \end{pmatrix} \quad T = \begin{pmatrix} 0.127 \\ 0.32 \\ 2.627 \\ 0.304 \\ 1.76 \\ 0.624 \\ 9.766 \end{pmatrix}$$

Since the deviation for x=6 skews the entire calculation, lump x=3..6 as one cell

$x := 0..3$ $\quad p2_x := p_x$ $\quad O2_x := O_x$

$p2_3 := p_3 + p_4 + p_5 + p_6$ $\quad O2_3 := O_3 + O_4 + O_5 + O_6$ $\quad E2_x := 100 \cdot p2_x$

$$T2_x := \frac{\left(O2_x - E2_x\right)^2}{E2_x}$$

x =

x
0
1
2
3

$$p2 = \begin{pmatrix} 0.33 \\ 0.366 \\ 0.203 \\ 0.101 \end{pmatrix} \quad O2 = \begin{pmatrix} 35 \\ 40 \\ 13 \\ 12 \end{pmatrix} \quad E2 = \begin{pmatrix} 32.956 \\ 36.581 \\ 20.302 \\ 10.145 \end{pmatrix} \quad T2 = \begin{pmatrix} 0.127 \\ 0.32 \\ 2.627 \\ 0.339 \end{pmatrix}$$

$$\chi_{sq} := \sum_{x=0}^{3} T2_x \qquad \chi_{sq} = 3.412$$

For 2 degrees of freedom and α=.05, the critical value of χ.sq is 5.99, therefore, we cannot reject H0.

(b) Apply the chi-square test to these data to test the hypothesis that it distribution is Poisson with mean 1.0. Again let α=.05

$m := 1$

$x := 0..6$ $\quad p_x := \text{dpois}(x, m)$ $\quad O_x := X_x$ $\quad E_x := 100 \cdot p_x$ $\quad T_x := \dfrac{\left(O_x - E_x\right)^2}{E_x}$

x =

x
0
1
2
3
4
5
6

$$p = \begin{pmatrix} 0.368 \\ 0.368 \\ 0.184 \\ 0.061 \\ 0.015 \\ 3.066 \times 10^{-3} \\ 5.109 \times 10^{-4} \end{pmatrix} \quad O = \begin{pmatrix} 35 \\ 40 \\ 13 \\ 6 \\ 4 \\ 1 \\ 1 \end{pmatrix} \quad E = \begin{pmatrix} 36.788 \\ 36.788 \\ 18.394 \\ 6.131 \\ 1.533 \\ 0.307 \\ 0.051 \end{pmatrix} \quad T = \begin{pmatrix} 0.087 \\ 0.28 \\ 1.582 \\ 2.813 \times 10^{-3} \\ 3.971 \\ 1.569 \\ 17.623 \end{pmatrix}$$

Again, since the deviation for x=6 skews the entire calculation, lump x=3..6 as one cell

$x := 0..3$

$p2_x := p_x$

$p2_3 := p_3 + p_4 + p_5 + p_6$

$O2_x := O_x$

$O2_3 := O_3 + O_4 + O_5 + O_6$

$E2_x := 100 \cdot p2_x$

$$T2_x := \frac{(O2_x - E2_x)^2}{E2_x}$$

x =

0
1
2
3

$$p2 = \begin{pmatrix} 0.368 \\ 0.368 \\ 0.184 \\ 0.08 \end{pmatrix} \quad O2 = \begin{pmatrix} 35 \\ 40 \\ 13 \\ 12 \end{pmatrix} \quad E2 = \begin{pmatrix} 36.788 \\ 36.788 \\ 18.394 \\ 8.022 \end{pmatrix} \quad T2 = \begin{pmatrix} 0.087 \\ 0.28 \\ 1.582 \\ 1.973 \end{pmatrix}$$

$$\chi_{sq} := \sum_{x=0}^{3} T2_x \qquad \chi_{sq} = 3.922$$

For 3 degrees of freedom and α=.05, the critical value of χ.sq is 7.81, therefore, we cannot reject H0.

(c) what are the differences between parts (a) and (b) and when might each case arise?

In part (a), we are estimating the mean value of the distribution from the data and find a mean of 1.11. In part (b), the mean value is given to be 1.0. If there were some external information or other points in the distribution that we are not analyzing, we might be given the mean value. Having the mean value not only changes the values of the expected number of observations, it also changes the number of degrees of freedom, which changes the chi-squared critical value.

Ch 9 - #19: The crosstowner is a bus that cuts a diagonal path from northeast Atlanta to southwest Atlanta. The time required to complete the route is maintained by the bus operator. The bus runs Monday through Friday. The times of the last fifty 8:00 am runs in minutes, are as follows:

$$t := \begin{pmatrix} 92.3 & 92.8 & 106.8 & 108.9 & 106.6 \\ 115.2 & 94.8 & 106.4 & 110.0 & 90.9 \\ 104.6 & 72.0 & 86.0 & 102.4 & 99.8 \\ 87.5 & 111.4 & 105.9 & 90.7 & 99.2 \\ 97.8 & 88.3 & 97.5 & 97.4 & 93.7 \\ 99.7 & 122.7 & 100.2 & 106.5 & 105.5 \\ 80.7 & 107.9 & 103.2 & 116.4 & 101.7 \\ 84.8 & 101.9 & 99.1 & 102.2 & 102.5 \\ 111.7 & 101.5 & 95.1 & 92.8 & 88.5 \\ 74.4 & 98.9 & 111.9 & 96.5 & 95.9 \end{pmatrix}$$

How are these run times distributed? Develop and test a suitable model.

t is a matrix for ease of data entry and for display purposes only. Convert to a vector

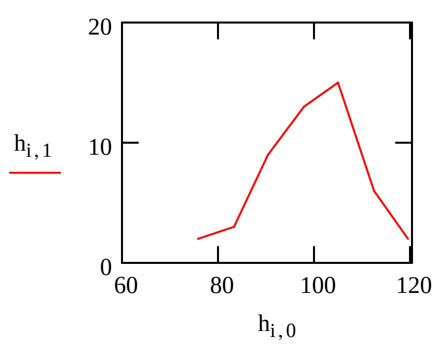
$i := 0..9 \qquad j := 0..4$

$T_{i+10\cdot j} := t_{i,j}$

Plot a histogram with roughly sqrt(N) cells

$N := 7$

$h := \text{histogram}(N, T) \qquad i := 0..N$



This might be gaussian

$m := \text{mean}(T) \qquad m = 99.222$

$s := \text{Stdev}(T) \qquad s = 10.169$

Create K classes with equal probability, based on a gaussian distribution with appropriate mean.

$\varepsilon := 1 \cdot 10^{-8}$ ε is needed to keep qnorm happy with P( ) from 0 - 1

$K := 10$

$k := 0..K$

$a_k := \text{qnorm}\left(\frac{k}{K+\varepsilon} + \frac{\varepsilon}{100}, m, s\right)$

$j := 0..K-1$

$E_j := \frac{50}{K}$ $O := \text{hist}(a, T)$

$\text{Term}_j := \frac{(O_j - E_j)^2}{E_j}$

a =

	0
0	34.532
1	86.19
2	90.663
3	93.889
4	96.646
5	99.222
6	101.798
7	104.555
8	107.781
9	112.254
10	160.389

O =

	0
0	5
1	3
2	6
3	4
4	6
5	5
6	5
7	7
8	6
9	3

E =

	0
0	5
1	5
2	5
3	5
4	5
5	5
6	5
7	5
8	5
9	5

Term =

	0
0	0
1	0.8
2	0.2
3	0.2
4	0.2
5	0
6	0
7	0.8
8	0.2
9	0.8

$\chi_{sq} := \sum_{j=0}^{K-1} \text{Term}_j$ $\chi_{sq} = 3.2$

Repeat for 8 classes

$a := 0$ clear the variable

$K := 8$

$k := 0 .. K$

$$a_k := \text{qnorm}\left(\frac{k}{K+\varepsilon} + \frac{\varepsilon}{100}, m, s\right)$$

$j := 0 .. K-1$

$E_j := \frac{50}{K}$ $O := \text{hist}(a, T)$

$$\text{Term}_j := \frac{\left(O_j - E_j\right)^2}{E_j}$$

$$a = \begin{pmatrix} 34.532 \\ 87.524 \\ 92.363 \\ 95.982 \\ 99.222 \\ 102.462 \\ 106.081 \\ 110.92 \\ 159.984 \end{pmatrix} \qquad O = \begin{pmatrix} 6 \\ 5 \\ 6 \\ 7 \\ 8 \\ 5 \\ 7 \\ 6 \end{pmatrix}$$

E =

	0
0	6.25
1	6.25
2	6.25
3	6.25
4	6.25
5	6.25
6	6.25
7	6.25
8	5
9	5

Term =

	0
0	0.01
1	0.25
2	0.01
3	0.09
4	0.49
5	0.25
6	0.09
7	0.01
8	0.2
9	0.8

$$\chi_{sq} := \sum_{j=0}^{K-1} \text{Term}_j \qquad \chi_{sq} = 1.2$$

Repeat for 5 classes

$a := 0$ clear the variable

$K := 5$

$k := 0..K$

$$a_k := \text{qnorm}\left(\frac{k}{K+\varepsilon} + \frac{\varepsilon}{100}, m, s\right)$$

$j := 0..K-1$

$E_j := \frac{50}{K}$ $O := \text{hist}(a, T)$

$$\text{Term}_j := \frac{(O_j - E_j)^2}{E_j}$$

$$a = \begin{pmatrix} 34.532 \\ 90.663 \\ 96.646 \\ 101.798 \\ 107.781 \\ 159.146 \end{pmatrix} \qquad O = \begin{pmatrix} 8 \\ 10 \\ 11 \\ 12 \\ 9 \end{pmatrix}$$

E =

	0
0	10
1	10
2	10
3	10
4	10
5	6.25
6	6.25
7	6.25
8	5
9	5

Term =

	0
0	0.4
1	0
2	0.1
3	0.4
4	0.1
5	0.25
6	0.09
7	0.01
8	0.2
9	0.8

$$\chi_{sq} := \sum_{j=0}^{K-1} \text{Term}_j \qquad \chi_{sq} = 1$$

Summarizing the results:

# classes	χ_{sq}	$\chi_{sq,\,.05,\,K-3}$
10	3.2	14.1
8	1.2	11.1
5	1.0	5.99

Conclusion: Do not reject H0 for any of these cases.