Chi-square tests

1. Chi-square test of Association

2× 2 Contingency Table

The outcome of a certain IQ test is tabularized as follows.

	Pass		**Fail**		**Total**
Male	**O=28**	**E=**	**O=12**	**E=**	**40**
Female	**O=34**	**E=**	**O=26**	**E=**	**60**
Total	**62**		**38**		

The two variables here are "Gender of Candidate" and "Results of the Candidate".

When you have two categorical variables from the same population, you may test whether there is a significant association between the two variables using the Chi-square Test.

Hypothesis

H_0 : $There\ is\ no\ relationship\ between\ gender\ and\ results$

H_1 : $There\ is\ a\ significant\ relationship\ between\ gender\ and\ results$

Test Statistic

$$\chi^2_{cal} = \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{df,\ \alpha\%} \quad ; df = (r-1)(c-1)$$

Under H_0

$$\text{Expected frequency} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

$\chi^2_{cal}=$

$\chi^2{}_{1,\ 5\%} = 3.84$

Decision

$h \times k$ Contingency Table

Example: A survey of 200 families, known to the regular television viewers was undertaken. They were asked which of the TV channels they watched most during a common week, and the observations are as follows.

TV channel watched most	Region			
	North	East	South	West
1	19	16	42	23
2	6	11	26	7
3	15	3	12	10

Test the hypothesis that there is no association between the TV channel watched most and the Region, using the Chi-square test.

2. Chi-square Goodness of Fit test

Example 1

From a list of 500 digits the occurrence of each distinct digit is observed. Test at 5% significance level, whether the sequence is a random sample from **the Uniform distribution.**

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	40	58	49	53	38	56	61	53	60	32

Example 2

The table below gives the number of heavy rainstorms reported by 330 weather stations over a one year period.

a) Find the expected frequencies of rainstorms given by the Poisson distribution having the same mean and the total as the observed distribution.

b) Use the Chi-square test to check the adequacy of the **Poisson distribution** to model these data.

# rainstorms	0	1	2	3	4	5	More than 5
# weather stations	102	114	74	28	10	2	0

Moments and Moment Generating Function

Moments

In Statistics, the **mathematical expectation** is called the **moments** of the distribution of a random variable.

Definition(r^{th} moment of ar.v.)

The r^{th} moment of a random variable X denoted by $\boldsymbol{\mu_r'}$ is the expected value of the random variable's r^{th} power; i.e.$\boldsymbol{E(X^r)}$.

$$For\ r = 1,2,3,\ldots$$

$$\mu_r' = E(X^r) = \sum x^r P(x) \text{ ; when } X \text{ is discrete}$$

$$\mu_r' = E(X^r) = \int_{-\infty}^{+\infty} x^r f(x)dx \text{ ; when } X \text{ is continuous}$$

***Special Application*: $\boldsymbol{\mu_1' = E(X^1) = E(X) = \mu}$ (mean of r.v. X)**

Definition (r^{th} moment about the mean of ar.v.)

The r^{th} moment about the mean of a random variable X denoted by $\boldsymbol{\mu_r}$ is the expected value of $(X-\mu)^r$; i.e.$\boldsymbol{E[(X-\mu)^r]}$.

$$For\ r = 1,2,3,\ldots$$

$$\mu_r = E[(X-\mu)^r] = \sum (x-\mu)^r P(x) \text{ ; when } X \text{ is discrete}$$

$$\mu_r = E[(X-\mu)^r] = \int_{-\infty}^{+\infty} (x-\mu)^r f(x)dx \text{ ; when } X \text{ is continuous}$$

***Special Application*: $\boldsymbol{\mu_2 = E[(X-\mu)^2] = V(X) = \sigma^2}$ (variance of r.v. X)**

Theorem:

$\boldsymbol{\sigma^2 = \mu_2' - \mu^2}$**; i.e.** $\boldsymbol{V(X) = E(X^2) - [E(X)]^2}$

Proof:

Moment Generating Function (MGF)

- The moments of most distributions can be determined directly by evaluating the respective integrals or sums.
- MGF is an alternative procedure, which sometimes provides considerable simplifications to find the moments.
- MGF can be used to find the expected value of ar.v. and its variance.

Definition

$M_X(t)$is the value which the function M_X assumes for the real variable (t).

The MGF of ar.v. X is given by;

$$M_X(t) = E(e^{tx}) = \sum e^{tx} P(x); \quad \text{when} X \text{ is discrete}$$

$$M_X(t) = E(e^{tx}) = \int_{-\propto}^{+\propto} e^{tx} f(x)dx; \quad \text{when} X \text{ is continuous}$$

$$where, M_X(t) = E(e^{tx}) = E[1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots + \frac{(tx)^r}{r!} + \cdots$$

$$M_X(t) = E(e^{tx}) = 1 + tE(x) + \frac{t^2E(x^2)}{2!} + \frac{t^3E(x^3)}{3!} + \cdots + \frac{t^rE(x^r)}{r!} + \cdots$$

$$and, \quad M'_X(t) = \frac{d\ M_X(t)}{dt}$$

$$= \frac{d\ E(e^{tx})}{dt} = 0 + E(x) + \frac{2t\ \ E(x^2)}{2!} + \frac{3t^2E(x^3)}{3!} + \cdots + \frac{rt^{r-1}E(x^r)}{r!} + \cdots$$

$$M''_X(t) = \frac{d^2\ M_X(t)}{dt^2}$$

$$= \frac{d^2\ E(e^{tx})}{dt^2} = 0 + 0 + \frac{2E(x^2)}{2!} + \frac{6t\ \ E(x^3)}{3!} + \cdots + \frac{r(r-1\_t^{r-2}E(x^r)}{r!} + \cdots$$

Properties of the MGF

1. $M'_X(t)|_{t=0} = E(X)= \mu$
2. $M''_X(t)|_{t=0} = E(X^2)$
3. $V(X) = M''_X(t)|_{t=0} - (M'_X(t)|_{t=0})^2$

Example 1

Find the MGF and E(X) and V(X) for the r.v.X whose *pdf* is given by

$$f(x) = \begin{Bmatrix} e^{-x}, & for\ x > 0 \\ 0, & otherwise \end{Bmatrix}$$

Example 2

Suppose Y~ Bin (n,p). Find E(X) and V(X) using its MGF.

Exercises

1. Let Y be a continuous r.v with *pdf* $f(y)=2e^{-3y}$; $y\geq 0$, Find the mean and the variance of Y.
2. Given that the probability distribution of a r.v. is $1/8\ {}^{3}_{r}C$ for r=1,2,3. Find the MGF, mean and variance for this random variable.