BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: R Programming

Course Code: 23ST05CC

Unit-V

Continuous Probability Distribution

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. I. Angel Agnes Mary
Guest Faculty
Department of Statistics

UNIT – V

Continuous Probability Distribution

Distribution	Mean	Variance	PDF
Normal	μ	σ^2	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\left(\frac{x-\mu}{2\sigma}\right)^2}$, $x=0,1,2,\ldots.$
Log Normal	$\exp\left(\mu+\frac{\sigma^2}{2}\right)$	[Exp(σ^2) – 1] exp(2 μ + σ^2)	$f_x(u)=\frac{1}{u\sigma\sqrt{2\pi}}e^{-\left(\frac{\log u-\mu}{2\sigma}\right)^2}$, $u>0$
Exponential	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$f(x)=\theta.e^{(-\theta x)}$, $x>0, \quad \theta>0$
Cauchy	undefined	undefined	$f(x)=\frac{1}{\pi}\left[\frac{\pi}{2}+tan^{-1}(x-\mu)\right]$
Gamma	kθ	Kθ^2	$f(x)=\frac{e^{-x}x^{\alpha-1}}{Gamma(\alpha)}$, $\alpha>0, \quad 0<x<\infty$
Beta	$\frac{\alpha}{(\alpha+\beta)}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$f(x)=\frac{(1-x)^{\beta-1}x^{\alpha-1}}{\frac{Gamma(\alpha).Gamma(\beta)}{Gamma(\alpha+\beta)}}$

16. Fitting of Normal Distribution

Obtain the equation of the normal distribution that may be fitted to the following data.

Class	60 – 65	65 – 70	70 – 75	75 – 80	80 – 85	85 – 90	90 – 95	95 – 100
Frequency	3	21	150	335	326	135	26	4

Also find expected frequency and goodness of fit for the data

Aim

Fit an appropriate Normal distribution for the above data and calculate theoretical frequency and testing the goodness of fit of a Normal distribution.

Procedure

- Define the Null Hypothesis (H_0) and Alternate Hypothesis (H_1) for given data.
- For a given Normal Distribution Calculate Mean and Standard Deviation.

$$Mean = \frac{\sum fm}{\sum f}$$

$$SD = \sqrt{\frac{\sum fm^2}{\sum f} - \left(\frac{\sum fm}{\sum f}\right)^2}$$

- Compute the respective probability using the recurrence relation of Normal Distribution.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{2\sigma}\right)^2}, \qquad x = 0,1,2,....$$

- Compute the Expected Frequency [E_i].
- Compute the test Statistic of Chi-Square Test,

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

- Taking decision about the hypothesis H_0 with respect to the level of significance and degree of freedom.

Calculation

Hypothesis

H_0: The data is good fit for the Normal Distribution.

H_1: The data is not good fit for the Normal Distribution.

Calculate Mean and Variance for the data.

X	Mid Value (m)	f	Fm	m^2	fm^2
60 – 65	62.5	3	187.5	3906.25	11718.75
65 – 70	67.5	21	1417.5	4556.25	95681.25
70 – 75	72.5	150	10875	5256.25	788437.5
75 – 80	77.5	335	25962.5	6006.25	2012093.75
80 – 85	82.5	326	26895	6806.25	2218837.5
85 – 90	87.5	135	11812.5	7656.25	1033593.75
90 – 95	92.5	26	2405	8556.25	222462.5
95 – 100	97.5	4	390	9506.25	380.25
Total		1000	79945		6420850

$$\mu = \frac{\sum_{i=1}^{n} fm}{\sum_{i=1}^{n} f} = \frac{79945}{1000} = 79.945$$

$$\sigma = \sqrt{\frac{\sum fm^2}{\sum f} - \left(\frac{\sum fm}{\sum f}\right)^2}$$

$$= \sqrt{\frac{6420850}{1000} - \left(\frac{79945}{1000}\right)^2}$$

$$= \sqrt{6420.85 - 6391.20}$$

$$= \sqrt{29.647}$$

= 5.4449

N = 1000, μ = 79.945, σ = 5.4449

Calculate the Probability and Expected Frequency,

X	Lower limit	$z = \left(\frac{x-\mu}{\sigma}\right)$	**Φ(z) z-table value**	**ΔΦ(z) = Φ(z+1) - Φ(z)**	**Expected Frequency N . ΔΦ(z)**
Below 60	-∞	-∞	0	0.00012	0
60 – 65	60	-3.6631	0.00012	0.0030	3
65 – 70	65	-2.7448	0.00307	0.0306	30.6 = 31
70 – 75	70	-1.8265	0.03362	0.1478	147.8 = 148
75 – 80	75	-0.9082	0.18141	0.3220	322.01 = 322
80 – 85	80	0.0101	0.50399	0.3191	319.11 = 319
85 – 90	85	0.9284	0.82381	0.1440	144
90 – 95	90	1.8467	0.96784	0.0295	29.50 = 30
95 - 100	95	2.7650	0.99720	0.0027	2.70 = 3
Above 100	100	3.683	0.99988	0	-

Compute the test Statistic of Chi-Square Test,

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

Frequency (O_i)	Expected Frequency (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
3	3	0	0	0
21	31	10	100	3.2258
150	148	2	4	0.0270
335	322	13	169	0.5248
326	319	7	49	0.1536
135	144	-9	81	0.5625
26	30	-4	16	0.5333
4	3	1	1	0.3333
			Total	5.3603

Degrees of Freedom

Calculate the Degrees of Freedom = n – p – 1

= 8 – 2 – 1

= 5

Table Value from Chi-Square Table in 5df = 11.070

Calculated Value = 5.3603

Conclusion

Since Calculated Value is less than the Table Value (i.e. 5.3603 < 11.070). So, we accept the null Hypothesis. Hence we conclude that the data is good fit for the Normal Distribution.

R Coding

```
#Input Data

k<-8

bins<-c("below 60","60-65","65-70","70-75","75-80","80-85","85-90","90-95",
"95-100","above100")

lowerlimit<-c(0,60,65,70,75,80,85,90,95,100)

mid<-c(62.5,67.5,72.5,77.5,82.5,87.5,92.5,97.5)

frequency<-c(3,21,150,335,326,135,26,4)

#obtaining theoretical probability

mu<- sum(mid*frequency)/sum(frequency)

x<-(mid-mu)

variance<-sum(frequency*(x*x))/sum(frequency)

sigma<-sqrt(variance)

cumprob<-round(c(pnorm(lowerlimit,mean=mu,sd=sigma,lower.tail=TRUE)),4)

#obtaining required columns in chi-square goodness of fit

prob<-seq(1,10)

for(i in 2:10)

{

prob[1]<-cumprob[1]

prob[i]<-round(cumprob[i+1]-cumprob[i],4)

}
```

```
exp_freq<-round(prob*(sum(frequency)))

out<-data.frame(cbind(bins,lowerlimit,cumprob,prob,exp_freq))

out


bins<-c("60-65","65-70","70-75","75-80","80-85","85-90","90-95","95-100")

frequency<-c(3,21,150,335,326,135,26,4)

exp_freq<-c(3,31,148,322,319,144,30,3)

oi_ei_sq<-(frequency-exp_freq)^2

oi_ei<-round(oi_ei_sq/exp_freq,4)
```

 # Output Table, Chi-square test statistics & P-Value

```
output<-data.frame(cbind(bins,frequency,exp_freq, oi_ei_sq, oi_ei))

output

chi_sq<-sum(oi_ei)

chi_sq

pvalue<-pchisq(chi_sq,8-2-1,lower.tail=FALSE)

pvalue
```

<u>R Output</u>

> mu

[1] 79.945

>sigma

[1] 5.444904

>Output

bins	**lowerlimit**	**cumprob**	**prob**	**exp_freq**
below 60	0	0	0	0
60-65	60	0.0001	0.0029	3
65-70	65	0.003	0.0309	31
70-75	70	0.0339	0.148	148
75-80	75	0.1819	0.3221	322
80-85	80	0.504	0.3194	319
85-90	85	0.8234	0.1442	144
90-95	90	0.9676	0.0296	30
95-100	95	0.9972	0.0027	3
above100	100	0.9999	<NA>	<NA>

> Output

bins	**frequency**	**exp_freq**	**oi_ei_sq**	**oi_ei**
60-65	3	3	0	0
65-70	21	31	100	3.2258065
70-75	150	148	4	0.027027
75-80	335	322	169	0.5248447
80-85	326	319	49	0.153605
85-90	135	144	81	0.5625
90-95	26	30	16	0.5333333
95-100	4	3	1	0.3333333

> chi_sq

[1] 5.36045

> pvalue

[1] 0.3734908

Interpretation

Here the p-value is greater than the specified level of significance (i.e.) **0.373 > 0.05**, so we accept the null hypothesis. Hence, we conclude that the data is good fit for the Normal Distribution.

Observation Problem

The following table baseball throws for a distance by 303 for first year wise school girls.

X	15 – 25	25 – 35	35 – 45	45 – 55	55 – 65	65 – 75
No. of girls	1	2	7	25	33	53

X	75 – 85	85 – 95	95 – 105	105 – 115	115 – 125	125 – 135
No. of girls	64	44	31	27	11	4

Fit a normal distribution and find the theoretical frequencies for the classes of the above frequency distribution and also test the goodness of fit.

17. Fitting of Log Normal Distribution

The following 50 values denote the random observation from the standard normal population construct an appropriate frequency population distribution for the observations. Fit the Log Normal Distribution and estimate the parameter by using the method of moments. The random observations are given below.

1.801, -0.175, -0.861, -0.577, -0.827, 0.459, -0.754, -1.406, 0.335, 0.802, 1.102, -0.134, 0.526, -0.820, -0.457, -0.072, 1.231, 0.239, 0.140, 0.560, -0.336, 1.483, 0.205, 0.333, 0.643, 0.942, 0.049, 2.021, 0.426, -0.729, -0.290, 0.550, 0.313, 0.209, -0.249, -0.706, 1.401, -0.240,-0.253, 0.338, 1.396, -1.142, -0.891, 0.323, -0.251, -0.466, 0.205, 1.223, 1.135, -1.804

Test whether Log Normal distribution is a good fit for the given data.

Aim

Fit an appropriate Log Normal distribution for the above data and calculate theoretical frequency and testing the goodness of fit of a Log Normal distribution.

Procedure

- Define the Null Hypothesis (H_0) and Alternate Hypothesis (H_1) for given data.
- Construct an appropriate frequency distribution for random observation which is drawn from standard normal population which can be expressed in Log Normal.

$$y = e^x$$

$$x = e^y$$

- For a given Log Normal Distribution estimate the parameter of Mean and Variance using method of moments.

$$Mean = \frac{\sum fx}{\sum f}$$

$$Variance = \frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2$$

$$E(X) = \mu = e^{\mu + \frac{1}{2}\sigma^2}$$

$$V(X) = \sigma^2 = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$$

$$SD(X) = \sigma = \sqrt{\sigma^2}$$

- Compute the respective probability density function is given by for a random variable x.

Let $y = log_e x \sim N(\mu, \sigma^2)$

$$f_x(u) = \frac{1}{u\sigma\sqrt{2\pi}} e^{-\left(\frac{logu - \mu}{2\sigma}\right)^2}, \quad u > 0$$

- Compute the Expected Frequency [E_i].
- Compute the test Statistic of Chi-Square Test,

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

- Taking decision about the hypothesis H_0 with respect to the level of significance and degree of freedom.

Calculation

Hypothesis

H_0: The data is good fit for the Log Normal Distribution.

H_1: The data is not good fit for the Log Normal Distribution.

Let the value of x is computed. Taking expectation,

6.0557	0.8395	0.4227	0.5616	0.4374	1.5825	0.4705	0.2451	1.3979	2.2300
3.0102	0.8746	1.6922	0.4404	0.6332	0.9305	3.4247	1.2700	1.1503	1.7507
0.7146	4.4061	0.8146	0.7168	1.9022	2.5651	1.0502	7.5459	1.5311	0.4824
0.7483	1.7333	1.3675	1.2324	0.7796	0.4936	4.0593	0.7866	0.7765	1.4021
4.0390	0.3192	0.4102	1.3813	0.7780	0.6275	1.2275	3.3974	3.1112	0.1646

To form a frequency distribution the data might be classified interms of class intervals.

$$i = \frac{Maximumvalue - Minimumvalue}{k}$$

Where,

k = 1 + (3.322 $\log_{10}$N)

= 1+ (3.322 $\log_{10}$50)

= 1+ (3.322 x 1.70)

k = 6.6439

$$i = \frac{7.5458 - 0.1646}{6.6439}$$

i = 1.11

Calculate Mean and Variance for the data.

x	Mid Value (x)	f	fx	d	fd	d^2	fd^2
0 – 1.11	0.555	25	13.88	-3	-75	9	225
1.11 – 2.22	1.665	14	23.31	-2	-28	4	56
2.22 – 3.33	2.775	4	11.1	-1	-4	1	4
3.33 – 4.44	3.885	5	19.43	0	0	0	0
4.44 – 5.55	4.995	0	0	1	0	1	0
5.55 – 6.66	6.105	1	6.11	2	2	4	4
6.66 – 7.77	7.215	1	7.22	3	3	9	9
Total		50	81.05		-102		298

$$\mu = \frac{\sum_{i=1}^{n} fx}{\sum_{i=1}^{n} f} = \frac{81.05}{50} = 1.6206$$

$$\sigma^2 = \frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2 \times i$$

$$= \left[\frac{298}{50} - \left(\frac{-102}{50}\right)^2\right] \times 1.11$$

$$= 1.996224$$

$$\sigma = \sqrt{1.996224}$$

$$= 1.412878$$

Estimating the parameter using method of moments,

To find Mean,

$$\mu = e^{\mu + \frac{1}{2}\sigma^2}$$

$$e^{\mu + \frac{1}{2}\sigma^2} = 1.6206 \qquad (1)$$

Taking log on both sides,

$$\mu + \frac{1}{2}\sigma^2 \log_e e = \log(1.6206)$$

$$\mu = 0.482797 - 0.345629$$

$$\mu = 0.137168$$

To find Variance,

$$\sigma^2 = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$$

$$e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2} = 1.996224$$

Taking log on both sides,

$$2\mu + 2\sigma^2 \log_e e - 2\mu - \sigma^2 \log_e e = \log(1.996224)$$

$$\sigma^2 = 0.69125$$

$$\sigma = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{0.69125}$$

$$\sigma = 0.83142$$

Calculate the Probability and Expected Frequency,

X	**upper limit**	$z = \frac{\log_e(x) - 0.1372}{0.8314}$	**Φ(z) z-table value**	**ΔΦ(z) = Φ(z) - Φ(z-1)**	**N . ΔΦ(z) (E_i)**
0 – 1.11	1.11	-0.039	0.4840	0.4840	24.21
1.11 – 2.22	2.22	0.7942	0.7852	0.3012	15.11
2.22 – 3.33	3.33	1.2819	0.8997	0.1145	5.68
3.33 – 4.44	4.44	1.6279	0.9484	0.0487	2.4
4.44 – 5.55	5.55	1.8963	0.9713	0.0229	1.14
5.55 – 6.66	6.66	2.1156	0.9830	0.0117	0.59
6.66 – 7.77	7.77	2.3010	0.9893	0.0063	0.32

Compute the test Statistic of Chi-Square Test,

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

Frequency (O_i)	Expected Frequency (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
25	24.21	0.79	0.6241	0.0258
14	15.11	-1.11	1.2321	0.0815
4	5.68	-1.68	2.8224	0.4969
5	2.4	2.6	6.76	2.8167
0	1.14	-1.14	1.2996	1.14
1	0.59	0.41	0.1681	0.2849
1	0.32	0.68	0.4624	1.445
			Total	6.2908

Degrees of Freedom

Calculate the Degrees of Freedom = n – p – 1

= 7 – 2 – 1

= 4

Table Value from Chi-Square Table in 4df = 9.488

Calculated Value = 6.2908

Conclusion

Since Calculated Value is less than the Table Value (i.e. 6.2908 < 9.488). So, we accept the null Hypothesis. Hence we conclude that the data is good fit for the Log Normal Distribution.

<u>R Coding</u>

#given data

```
scores<-c(1.801, -0.175, -0.861, -0.577, -0.827, 0.459, -0.754, -1.406, 0.335, 0.802, 1.102, -0.134, 0.526, -0.820, -0.457, -0.072, 1.231, 0.239, 0.140, 0.560, -0.336, 1.483, -0.205, -0.333, 0.643, 0.942, 0.049, 2.021, 0.426, -0.729, -0.290, 0.550, 0.313, 0.209, -0.249, -0.706, 1.401, -0.240, -0.253, 0.338, 1.396, -1.142, -0.891, 0.323, -0.251, -0.466, 0.205, 1.223, 1.135, -1.804)

N<-length(scores)

scoreslog<-exp(scores)
```

#determining width of the class interval for a fixed k values

```
k<-6.6439

width<-(max(scoreslog)-min(scoreslog))/k

c=1.11

bins<-c("0.0-1.11","1.11-2.22","2.22-3.33","3.33-4.44","4.44-5.55","5.55-6.66","6.66-7.77")

uplimit<-c(1.11,2.22,3.33,4.44,5.55,6.66,7.77)
```

#function to get frequency for each class interval

```
countobs<-function(data,l,u)
{
data2<-ifelse(data>=l & data<u,1,0)
count<-sum(data2)
}
```

#obtaining the frequency for each class

```
frequency<-c(countobs(scoreslog,0.0,1.11), countobs(scoreslog,1.11,2.22),
countobs(scoreslog,2.22,3.33), countobs(scoreslog,3.33,4.44), countobs(scoreslog,4.44,5.55),
countobs(scoreslog,5.55,6.66), countobs(scoreslog,6.66,7.77))
```

#obtaining theoretical probabilities

```
midx<-seq(0.555,7.215,1.11)
midx
mu<- sum(midx*frequency)/sum(frequency)
A<-3.885
d<-(midx-A)/c
fd<-frequency*d
fd_sq<- frequency*(d^2)
variance<-(((sum(fd_sq)/N)-(sum(fd)/N)^2))*c
sigma<-sqrt(variance)
mulog<-log(mu)-(log(variance)/2)
variancelog<-log(variance)
sdlog<-sqrt(variancelog)
cumprob<-round(c(plnorm(uplimit,meanlog=mulog,sdlog=sdlog,lower.tail=TRUE)),4)
cumprob
```

#obtaining frequency columns in chi-square goodness of fit

```
prob<-seq(1,7)

for(i in 2:7)

{

prob[1]<-cumprob[1]

prob[i]<-round(cumprob[i]-cumprob[i-1],4)

}

exp_freq<-round((prob*N),2)

oi_ei_sq<-(frequency-exp_freq)^2

oi_ei<-round(oi_ei_sq/exp_freq,4)
```

#output table, chi-square test statistic & p-value

```
out<-data.frame(cbind(bins,frequency,cumprob,prob,exp_freq,oi_ei_sq,oi_ei))

out

chi_sq<-sum(oi_ei)

chi_sq

pvalue<-pchisq(chi_sq,7-2-1,lower.tail=FALSE)

pvalue
```

R Output

>mu

[1] 1.6206

>variance

[1] 1.996224

>mulog

[1] 0.1371678

>variancelog

[1] 0.6912574

>sdlog

[1] 0.8314189

Output:

bins	frequency	cumprob	prob	exp_freq	oi_ei_sq	oi_ei
0.0-1.11	25	0.4843	0.4843	24.21	0.6241	0.0258
1.11-2.22	14	0.7865	0.3022	15.11	1.2321	0.0815
2.22-3.33	4	0.9001	0.1136	5.68	2.8224	0.4969
3.33-4.44	5	0.9482	0.0481	2.4	6.76	2.8167
4.44-5.55	0	0.971	0.0228	1.14	1.2996	1.14
5.55-6.66	1	0.9828	0.0118	0.59	0.1681	0.2849
6.66-7.77	1	0.9893	0.0065	0.32	0.4624	1.445

>chi_sq

[1] 6.2908

>pvalue

[1] 0.1784582

Interpretation

Here the p-value is greater than the specified level of significance (i.e.) **0.178 > 0.05**, so we accept the null hypothesis. Hence, we conclude that the data is good fit for the Log Normal Distribution.

Observation Problem

The height in inches of 1000 students of same age in a particular standard given the following frequency table. Assuming the height of the population follows Log Normal Distribution of the given frequency distribution and test for goodness of fit.

Height in inches	60 - 62.5	62.5 - 65	65 – 67.5	67.5 – 70	70 – 72.5	72.5 - 75	75 – 77.5
No. of Students	30	20	145	332	329	140	24

18. Fitting of Cauchy Distribution

Fit an appropriate Cauchy distribution for the table given below which shows the distribution of height of men in a college, test for goodness of fit.

Height	61 – 62	62 – 63	63 – 64	64 – 65	65 – 66	66 – 67	67 – 68
frequency	4	20	23	75	114	186	212

Height	68 – 69	69 – 70	70 – 71	71 – 72	72 – 73	73 – 74	74 – 75
frequency	252	218	175	149	46	18	8

Aim

Fit an appropriate Cauchy Distribution for the table given below which shows the distribution of heights of men in a college and testing the goodness of fit of a Cauchy Distribution.

Procedure

- Define the Null Hypothesis (H_0) and Alternate Hypothesis (H_1) for given data.
- For a given Cauchy Distribution Calculate respecting Mean.
- Compute the respective probability density function of Cauchy distribution with location parameter x, μ and scale parameter units is given by,

$$f(x) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{dx}{1+(x-\mu)^2}$$

$$f(x) = \frac{1}{\pi}\left[\frac{\pi}{2}+tan^{-1}(x-\mu)\right]$$

- Compute the Expected Frequency [E_i].
- Compute the test Statistic of Chi-Square Test,

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

- Taking decision about the hypothesis H_0 with respect to the level of significance and degree of freedom.

Calculation

Hypothesis

H_0: The data is good fit for the Cauchy Distribution.

H_1: The data is not good fit for the Cauchy Distribution.

Calculate Mean for appropriate formula of Continuous Distribution.

x	f	Mid x	fx
61 – 62	4	61.5	246
62 – 63	20	62.5	1250
63 – 64	23	63.5	1460.5
64 – 65	75	64.5	4837.5
65 – 66	114	65.5	7467
66 – 67	186	66.5	12369
67 – 68	212	67.5	14310
68 – 69	252	68.5	17262
69 – 70	218	69.5	15151
70 – 71	175	70.5	12337.5
71 – 72	149	71.5	10653.5
72 – 73	46	72.5	3335
73 – 74	18	73.5	1323
74 – 75	8	74.5	596
Total	1500		102598

$$\mu = \frac{\sum_{i=1}^{n} fx}{\sum_{i=1}^{n} f}$$

$$= \frac{102598}{1500} = 68.39867$$

For a Cauchy Distribution the probability density function and Expected Frequency is given by,

$$f(x) = \frac{1}{\pi}\left[\frac{\pi}{2} + tan^{-1}(x - \mu)\right]$$

X	Upper limit	x - μ	f(x) = (1/2)tan^{-1}(x - μ)	Δf(x) = f(x) – f(x-1)	Expected Frequency N . Δf(x)
61 – 62	62	-6.40	0.04875	0.049	74
62 – 63	63	-5.40	0.0575	0.0088	13
63 – 64	64	-4.40	0.07014	0.0126	19
64 – 65	65	-3.40	0.0895	0.198	30
65 – 66	66	-2.40	0.1231	0.0334	52
66 – 67	67	-1.40	0.1920	0.0682	108
67 – 68	68	-0.60	0.3653	0.1717	272
68 – 69	69	0.60	0.6601	0.2950	440
69 – 70	70	1.50	0.8177	0.1591	225
70 – 71	71	2.60	0.8812	0.0640	91
71 – 72	72	3.60	0.9128	0.00318	46
72 – 73	73	4.60	0.9313	0.0186	27
73 – 74	74	5.60	0.9434	0.0122	18
74 – 75	75	6.60	0.9521	0.0033	12

Compute the test Statistic of Chi-Square Test,

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

Frequency (O_i)	Expected Frequency (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
4	74	-70	4900	66.2162
20	13	7	49	3.7692
23	19	4	16	6.8421
75	30	45	2025	67.5
114	52	62	3844	73.9230
186	108	78	6084	56.3333
212	272	-60	3600	13.2353
252	440	-188	35344	80.3273
218	225	-7	49	0.2178
175	91	84	7056	77.5385
149	46	103	10609	230.6304
46	27	19	361	13.3704
18	18	0	0	0
8	12	-4	16	1.3333
			Total	685.2365

Degrees of Freedom

Calculate the Degrees of Freedom = n – p – 1

= 14 – 1 – 1

= 12

Table Value from Chi-Square Table in 12df = 21.026

Calculated Value = 685.2365

Conclusion

Since Calculated Value is greater than the Table Value (i.e. 685.2365 > 21.026). So, we reject the null Hypothesis. Hence we conclude that the data is not good fit for the Cauchy Distribution.

R Coding

#Input Data

```
bins<-c("61-62","62-63","63-64","64-65","65-66","66-67","67-68","68-69","69-70","70-71","71-72",
"72-73","73-74","74-75")

uplimit<-c(62,63,64,65,66,67,68,69,70,71,72,73,74,75)

midx<-seq(61.5,74.5,1)

frequency<-c(4,20,23,75,114,186,212,252,218,175,149,46,18,8)
```

#obtaining theoretical probability

```
mu<- sum(midx*frequency)/sum(frequency)

q<-( uplimit-mu)

cumprob<-round(c(pcauchy(q ,location=0,scale=1,lower.tail=TRUE,log.p=FALSE)),4)

cumprob
```

#obtaining required columns in chi-square goodness of fit

```
prob<-seq(1,14)

for(i in 2:14)

{

prob[1]<-cumprob[1]

prob[i]<-round(cumprob[i]-cumprob[i-1],13)

}

exp_freq<-round(prob*(sum(frequency)))

exp_freq

oi_ei_sq<-(frequency-exp_freq)^2

oi_ei_sq

oi_ei<-round(oi_ei_sq/exp_freq,4)

oi_ei
```

#output table, chi-square test statistic & p-value

```
out<-data.frame(cbind(bins,frequency,cumprob,prob,exp_freq,oi_ei_sq,oi_ei))

out

chi_sq<-sum(oi_ei)

chi_sq

pvalue<-pchisq(chi_sq,14-1-1,lower.tail=FALSE)

pvalue
```

<u>R Output</u>

mu

[1] 68.39867

> out

bins	frequency	cumprob	prob	exp_freq	oi_ei_sq	oi_ei
61-62	4	0.0493	0.0493	74	4900	66.2162
62-63	20	0.0583	0.009	13	49	3.7692
63-64	23	0.0712	0.0129	19	16	0.8421
64-65	75	0.0911	0.0199	30	2025	67.5
65-66	114	0.1257	0.0346	52	3844	73.9231
66-67	186	0.1976	0.0719	108	6084	6.3333
67-68	212	0.3792	0.1816	272	3600	13.2353
68-69	252	0.6723	0.2931	440	35344	80.3273
69-70	218	0.8223	0.15	225	49	0.2178
70-71	175	0.8832	0.0609	91	7056	77.5385
71-72	149	0.9138	0.0306	46	10609	230.6304
72-73	46	0.9319	0.0181	27	361	13.3704
73-74	18	0.9438	0.0119	18	0	0
74-75	8	0.9521	0.0083	12	16	1.3333

> chi_sq

[1] 685.2369

> pvalue

[1] 6.367015e-139

Interpretation

Here the p-value is less than the specified level of significance (i.e.) **0.000 < 0.05**, so we reject the null hypothesis. Hence, we conclude that the data is not good fit for the Cauchy Distribution.

Observation Problem

Obtain the equation of the Cauchy Distribution that may be fitted to the following data.

Class	60 – 65	65 – 70	70 – 75	75 – 80	80 – 85	85 – 90	90 – 95	95 – 100
Frequency	3	21	150	335	326	135	26	4

Also find expected frequency and goodness of fit for the data

19. Fitting of Exponential Distribution

200 electric light bulbs were tested and the average life time of the bulbs was found to be 25 hours. Using the summary given below test the hypothesis that the lifetime is exponentially distributed.

Life t in hours	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
No. of bulbs	104	56	24	12	4

You are given that an exponential distribution with parameter $\theta>0$ has the p.d.f.:

$$f(x) = \theta.e^{-\theta x}, x \geq 0$$

Also calculate the goodness of fit.

Aim

Fit an appropriate Exponential Distribution for the above data and calculate theoretical frequency. Testing the goodness of fit of a Exponential Distribution.

Procedure

- Define the Null Hypothesis (H_0) and Alternate Hypothesis (H_1) for given data.
- For a given Exponential Distribution Calculate respecting Mean.
- Compute the respective probability density function of Exponential distribution with location parameter θ is given by,

$$f(x) = \theta.e^{(-\theta x)}, \quad x > 0, \quad \theta > 0$$

- The Cumulative density function is given by

$$F(x) = 1 - e^{(-\theta x)}$$

- Compute the Expected Frequency [E_i].
- Compute the test Statistic of Chi-Square Test,

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

- Taking decision about the hypothesis H_0 with respect to the level of significance and degree of freedom.

Calculation

Hypothesis

H_0: The data is good fit for the Exponential Distribution.

H_1: The data is not good fit for the Exponential Distribution.

Calculate Mean for appropriate formula of Continuous Distribution.

x	f	Mid x	fx
0 – 20	104	10	1040
20 – 40	56	30	1680
40 – 60	24	50	1200
60 – 80	12	70	840
80 – 100	4	90	360
Total	200		5120

$$\mu = \frac{\sum_{i=1}^{n} fx}{\sum_{i=1}^{n} f} = \frac{5120}{200} = 25.6$$

By the method of moments, the parameter θ is calculated,

$$\mu = \frac{1}{\theta}$$

$$\theta = \frac{1}{\mu} = \frac{1}{25.6} = 0.0391$$

For a Exponential Distribution the Cumulant density function and Expected Frequency is given by,

$$F(x) = 1 - e^{(-\theta x)}$$

Bins	Upper limit (x)	f(x) = 1-$e^{(-\theta x)}$	Δf(x) = f(x) – f(x-1)	Expected Frequency N . Δf(x)
0 – 20	20	0.5422	0.5422	108
20 – 40	40	0.7907	0.2482	50
40 – 60	60	0.9042	0.1135	23
60 – 80	80	0.9562	0.0520	10
80 – 100	100	0.9800	0.0238	5

Compute the test Statistic of Chi-Square Test,

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

Frequency (O_i)	Expected Frequency (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
104	108	-4	16	0.1481
56	50	6	36	0.72
24	23	1	1	0.0435
12	10	2	4	0.4
4	5	-1	1	0.200
			Total	1.5116

Degrees of Freedom

Calculate the Degrees of Freedom = n – p – 1

= 5 – 1 – 1

= 3

Table Value from Chi-Square Table in 12df = 7.815

Calculated Value = 1.5116

Conclusion

Since Calculated Value is less than the Table Value (i.e. 1.5116 < 7.815). So, we accept the null Hypothesis. Hence we conclude that the data is good fit for the Exponential Distribution.

R Coding

```
#given data
bins<-c("0-20","20-40","40-60","60-80","80-100")
frequency<-c(104,56,24,12,4)
N<-sum(frequency)
uplimit<-seq(20,100,20)

#obtaining theoretical probabilities
midx<-seq(10,90,20)
mu<- sum(midx*frequency)/sum(frequency)
theta<-1/mu
cumprob<-round(c(pexp(uplimit,rate=theta,lower.tail=TRUE)),4)
cumprob

#obtaining frequency columns in chi-square goodness of fit
prob<-seq(1,5)
for(i in 2:5)
{
prob[1]<-cumprob[1]
prob[i]<-round(cumprob[i]-cumprob[i-1],4)
}
exp_freq<-round(prob*N)
oi_ei_sq<-(frequency-exp_freq)^2
oi_ei<-round(oi_ei_sq/exp_freq,4)
```

#output table, chi-square test statistic & p-value

```
out<-data.frame(cbind(bins,frequency,cumprob,prob,exp_freq,oi_ei_sq,oi_ei))
out
chi_sq<-sum(oi_ei)
chi_sq
pvalue<-pchisq(chi_sq,5-1-1,lower.tail=FALSE)
pvalue
```

<u>R Output</u>

> mu

[1] 25.6

> theta

[1] 0.0390625

Output:

bins	frequency	cumprob	prob	exp_freq	oi_ei_sq	oi_ei
0-20	104	0.5422	0.5422	108	16	0.1481
20-40	56	0.7904	0.2482	50	36	0.72
40-60	24	0.904	0.1136	23	1	0.0435
60-80	12	0.9561	0.0521	10	4	0.4
80-100	4	0.9799	0.0238	5	1	0.2

> chi_sq

[1] 1.5116

> pvalue

[1] 0.6795957

Interpretation

Here the p-value is Greater than the specified level of significance (i.e.) **0.6796 > 0.05**, so we accept the null hypothesis. Hence, we conclude that the data is good fit for the Exponential Distribution.

Observation Problem

The process completition time by 40 employees are given below.

19.74	2.28	10.43	10.8	20.42	20.71	63.01	12.81	19.09	14.7
11.93	34.88	17.35	63.51	97.79	47.89	17.02	22.28	106.49	15.46
10.89	80.58	47.72	16.18	43.18	13.52	10.83	19.13	14.42	19.5
32.01	19.32	111.85	130.93	35.26	16.51	51.09	62.28	45.65	10.42

Test whether exponential distribution is a good fit for the given data.

20. Fitting of Gamma Distribution

Fit a Gamma Distribution for the following data and test for the goodness of fit.

x	0 – 2	2 – 4	4 – 6	6 – 8	8 – 10
Frequency	10	56	24	12	4

Aim

Fit an appropriate Gamma distribution for the above data and calculate theoretical frequency and testing the goodness of fit of a Gamma distribution.

Procedure

- Define the Null Hypothesis (H_0) and Alternate Hypothesis (H_1) for given data.
- For a given Normal Distribution Calculate Mean and Variance.

$$Mean = \frac{\sum fm}{\sum f} \text{ and } Variance = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$

- By method of moments we calculate the shape parameter (α) and scale parameter (β). Scale parameter value should be 1.

$$\alpha = \frac{[E(X)]^2}{V(X)} \text{ and } \beta = \frac{V(`X)}{E(X)}$$

- Compute the respective probability using the recurrence relation of Gamma Distribution.

$$f(x) = \frac{e^{-x} x^{\alpha-1}}{Gamma(\alpha)}, \quad \alpha > 0, \quad 0 < x < \infty$$

- The cumulative function is given by,

$$f(x) = \frac{1}{Gamma(\alpha)} \int_0^x y^{\alpha-1} e^{-y} dy$$

- Compute the Expected Frequency [E_i].
- Compute the test Statistic of Chi-Square Test,

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

- Taking decision about the hypothesis H_0 with respect to the level of significance and degree of freedom.

Calculation

Hypothesis

H_0: The data is good fit for the Gamma Distribution.

H_1: The data is not good fit for the Gamma Distribution.

Calculate Mean and Variance for the data.

x	Mid Value (m)	f	fm	x = m - μ	fx	x^2	fx^2
0 – 2	1	10	10	-2.9434	-29.434	8.6636	86.6360
2 – 4	3	56	168	-0.9434	-52.831	0.8900	49.8400
4 – 6	5	24	120	1.0566	25.358	1.1164	26.7936
6 – 8	7	12	84	3.0566	36.679	9.3428	112.1136
8 – 10	9	4	36	5.0566	20.226	25.5692	102.2768
Total		106	418		-0.002		377.6600

$$\mu = \frac{\sum_{i=1}^{n} fm}{\sum_{i=1}^{n} f} = \frac{418}{106} = 3.9434$$

$$Variance = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f} \right)^2$$

$$\sigma^2 = \frac{377.6600}{106} - \left(\frac{-0.002}{106} \right)^2$$

$$= 3.5628 - 0$$

$$= 3.5628$$

To find shape and scale parameter,

$$\alpha = \frac{(Mean)^2}{Variance}$$

$$\alpha = \frac{(3.9434)^2}{3.5628}$$

$$= 4.3647 \approx 4$$

$$\beta = \frac{Variance}{Mean}$$

$$\beta = \frac{3.5628}{3.9434}$$

$$= 0.9035 \approx 1$$

Calculate the Cumulative Probability and Expected Frequency,

$$F(x) = \frac{1}{Gamma(\alpha)} \int_0^x y^{\alpha-1} e^{-y} dy$$

bins	Lower limit	F(x)	ΔF(x) = F(x -1) - F(x)	Expected Frequency
0 – 2	2	0.143	0.143	15
2 – 4	4	0.567	0.424	45
4 – 6	6	0.849	0.282	30
6 – 8	8	0.958	0.109	12
8 – 10	10	0.990	0.032	3

Compute the test Statistic of Chi-Square Test,

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

Frequency (O_i)	Expected Frequency (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
10	15	-5	25	1.6667
56	45	11	121	2.6889
24	30	-6	36	1.2000
12	12	0	0	0
4	3	1	1	0.3333
			Total	5.8889

Degrees of Freedom

Calculate the Degree of Freedom = n – p – 1

= 5 – 2 – 1

= 2

Table Value from Chi-Square Table in 5df = 5.991

Calculated Value = 5.8889

Conclusion

Since Calculated Value is less than the Table Value (i.e. 5.8889 < 5.991). So, we accept the null Hypothesis. Hence we conclude that the data is good fit for the Gamma Distribution.

<u>R Coding</u>

#given data

```
bins<-c("0-2","2-4","4-6","6-8","8-10")

frequency<-c(10,56,24,12,4)

N<-sum(frequency)

uplimit<-seq(2,10,2)

uplimit
```

#obtaining theoretical probabilities

```
midx<-seq(1,9,2)

mu<- sum(midx*frequency)/sum(frequency)

x<-(midx-mu)

variance<-(sum(frequency*(x^2))/sum(frequency))-( sum(frequency*x)/sum(frequency))^2

beta<-round(variance/mu)

alpha<-round(((mu)^2)/variance)

cumprob<-round(c(pgamma(uplimit,shape=alpha,scale=beta,lower.tail=TRUE)),3)

cumprob
```

#obtaining frequency columns in chi-square goodness of fit

```
prob<-seq(1,5)

for(i in 2:5)

{

prob[1]<-cumprob[1]

prob[i]<-round(cumprob[i]-cumprob[i-1],4)

}

exp_freq<-round(prob*N)

oi_ei_sq<-(frequency-exp_freq)^2

oi_ei<-round(oi_ei_sq/exp_freq,4)
```

#output table, chi-square test statistic & p-value

```
out<-data.frame(cbind(bins,frequency,cumprob,prob,exp_freq,oi_ei_sq,oi_ei))

out

chi_sq<-sum(oi_ei)

chi_sq
```

```
pvalue<-pchisq(chi_sq,5-2-1,lower.tail=FALSE)

pvalue
```

R Output

> mu

[1] 3.943396

> variance

[1] 3.562834

> alpha

[1] 4

> beta

[1] 1

Output

bins	frequency	cumprob	prob	exp_freq	oi_ei_sq	oi_ei
0-2	10	0.143	0.143	15	25	1.667
2-4	56	0.567	0.424	45	121	2.6889
4-6	24	0.849	0.282	30	36	1.2
6-8	12	0.958	0.109	12	0	0
8-10	4	0.990	0.032	3	1	0.333

> chi_sq

[1] 5.8889

> pvalue

[1] 0.052631

Interpretation

Here the p-value is greater than the specified level of significance (i.e.) **0.053 > 0.05**, so we accept the null hypothesis. Hence, we conclude that the data is good fit for the Gamma Distribution.

Observation Problem

Fit a Gamma Distribution for the following data and test for the goodness of fit.

X	0 – 2	2 – 4	4 – 6	6 – 8	8 – 10
Frequency	30	44	24	16	6

21. Fitting of Beta Distribution

Fit a beta distribution for the following data and test for goodness of fit.

X	0 – 0.2	0.2 – 0.4	0.4 – 0.6	0.6 – 0.8	0.8 – 1
Frequency	75	90	80	95	70

Aim

Fit an appropriate Beta distribution for the above data and calculate theoretical frequency and testing the goodness of fit of a Beta distribution.

Procedure

- Define the Null Hypothesis (H_0) and Alternate Hypothesis (H_1) for given data.
- For a given Normal Distribution Calculate Mean and Variance.

$$Mean = \frac{\sum fm}{\sum f}$$

$$Variance = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$

- By method of moments we calculate the shape parameter (α) and (β).

$$\alpha = E(X)\left[\frac{[E(X)\times(1-E(X)]}{V(X)}\right]-1$$

$$\beta = [1-E(X)]\left[\frac{[E(X)\times(1-E(X)]}{V(X)}\right]-1$$

- Compute the respective probability using the recurrence relation of beta Distribution.

$$f(x)=\frac{(1-x)^{\beta-1}x^{\alpha-1}}{B(\alpha,\beta)}$$

$$f(x)=\frac{(1-x)^{\beta-1}x^{\alpha-1}}{\frac{Gamma(\alpha).Gamma(\beta)}{Gamma(\alpha+\beta)}}$$

Where, B(α , β) is

$$B(\alpha,\beta)=\frac{Gamma(\alpha).Gamma(\beta)}{Gamma(\alpha+\beta)}$$

- The cumulative function is given by,

$$F(x)=\frac{Gamma(\alpha+\beta)}{Gamma(\alpha).Gamma(\beta)}\int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt$$

- Compute the Expected Frequency [E_i].
- Compute the test Statistic of Chi-Square Test,

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

- Taking decision about the hypothesis H_0 with respect to the level of significance and degree of freedom.

Calculation

Hypothesis

H_0: The data is good fit for the Beta Distribution.

H_1: The data is not good fit for the Beta Distribution.

Calculate Mean and Variance for the data.

X	Mid Value (m)	f	fm	x = m - μ	fx	x^2	fx^2
0 – 0.2	0.1	75	7.5	-0.398	-29.82	0.1581	11.8564
0.2 – 0.4	0.3	90	27	-0.198	-17.78	0.0390	3.5141
0.4 – 0.6	0.5	80	40	0.002	0.192	0.000	0
0.6 – 0.8	0.7	95	66.5	0.202	19.228	0.0410	3.8917
0.8 – 1	0.9	70	63	0.402	28.168	0.1619	11.3348
Total		410	204		0.016		30.5970

$$\mu = \frac{\sum_{i=1}^{n} fm}{\sum_{i=1}^{n} f} = \frac{204}{410} = 0.4976$$

$$Variance = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$

$$\sigma^2 = \frac{30.5970}{410} - \left(\frac{0.016}{410}\right)^2 = 0.0746$$

To find shape parameter α and β,

$$\alpha = E(X)\left[\frac{[E(X)\times(1-E(X)]}{V(X)}\right] - 1$$

$$\alpha = 0.4976 \times \left[\frac{[0.4976 \times (1-0.4976]}{0.0746}\right] - 1$$

$$\alpha = 0.4976 \times 2.3511$$

$$= 1.1699 \approx 1$$

$$\beta = [1 - E(X)]\left[\frac{[E(X)\times(1-E(X)]}{V(X)}\right] - 1$$

$$\beta = (1 - 0.4976) \times \left[\frac{[0.4976 \times (1-0.4976]}{0.0746}\right] - 1$$

$$= 0.5024 \times 2.3511$$

$$= 1.1812 \approx 1$$

Calculate the Cumulative Probability and Expected Frequency,

$$F(x) = \frac{Gamma(\alpha + \beta)}{Gamma(\alpha).Gamma(\beta)} \int_0^x t^{\alpha - 1}(1 - t)^{\beta - 1} dt$$

bins	Lower limit	F(x)	$\Delta F(x)$ = F(x -1) - F(x)	Expected Frequency N . $\Delta F(x)$
0 – 0.2	0.2	0.2	0.2	82
0.2 – 0.4	0.4	0.4	0.2	82
0.4 – 0.6	0.6	0.6	0.2	82
0.6 – 0.8	0.8	0.8	0.2	82
0.8 – 1	1	1	0.2	82

Compute the test Statistic of Chi-Square Test,

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

Frequency (O_i)	Expected Frequency (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
75	82	-7	49	0.5976
90	82	8	64	0.7805
80	82	-2	4	0.0488
95	82	13	169	2.0610
70	82	-12	144	1.7561
			Total	5.2440

Degrees of Freedom

Calculate the Degrees of Freedom = n – p – 1

= 5 – 2 – 1

= 2

Table Value from Chi-Square Table in 5df = 5.991

Calculated Value = 5.2440

Conclusion

Since Calculated Value is less than the Table Value (i.e. 5.2440 < 5.991). So, we accept the null Hypothesis. Hence we conclude that the data is good fit for the Beta Distribution.

R Coding

#given data

```
bins<-c("0-0.2","0.2-0.4","0.4-0.6","0.6-0.8","0.8-1")

frequency<-c(75,90,80,95,70)

N<-sum(frequency)

uplimit<-seq(0.2,1,0.2)
```

#obtaining theoretical probabilities

```
midx<-seq(0.1,0.9,0.2)

mu<- sum(midx*frequency)/sum(frequency)

x<-(midx-mu)

variance<-(sum(frequency*(x^2))/sum(frequency))-( sum(frequency*x)/sum(frequency))^2

M<-(((mu*(1-mu))/variance)-1)

beta<-round(M*(1-mu))

alpha<-round(M*mu)

cumprob<-round(c(pbeta(uplimit,shape1=alpha,shape2=beta,lower.tail=TRUE)),4)

cumprob
```

#obtaining frequency columns in chi-square goodness of fit

```
prob<-seq(1,5)

for(i in 2:5)

{

prob[1]<-cumprob[1]

prob[i]<-round(cumprob[i]-cumprob[i-1],4)

}
```

```
exp_freq<-round(prob*N)

oi_ei_sq<-(frequency-exp_freq)^2

oi_ei<-round(oi_ei_sq/exp_freq,4)
```

#output table, chi-square test statistic & p-value

```
out<-data.frame(cbind(bins,frequency,cumprob,prob,exp_freq,oi_ei_sq,oi_ei))

out

chi_sq<-sum(oi_ei)

chi_sq

pvalue<-pchisq(chi_sq,5-2-1,lower.tail=FALSE)

pvalue
```

<u>R Output</u>

> mu

[1] 0.497561

> variance

[1] 0.0746282

> alpha

[1] 1

> beta

[1] 1

Output

bins	frequency	cumprob	prob	exp_freq	oi_ei_sq	oi_ei
0-0.2	75	0.2	0.2	82	49	0.5979
0.2-0.4	90	0.4	0.2	82	64	0.7805
0.4-0.6	80	0.6	0.2	82	4	0.0488
0.6-0.8	95	0.8	0.2	82	169	2.061
0.8-1.0	70	1	0.2	82	144	1.7561

> chi_sq

[1] 5.244

> pvalue

[1] 0.0726574

Interpretation

Here the p-value is greater than the specified level of significance (i.e.) **0.073 > 0.05**, so we accept the null hypothesis. Hence, we conclude that the data is good fit for the Beta Distribution.

Observation Problem

Fit a beta distribution for the following data and test for goodness of fit.

Class	0 – 0.1	0.1 – 0.2	0.2 – 0.3	0.3 – 0.4	0.4 – 0.5
Frequency	14	12	18	18	12

Class	0.5 - 0.6	0.6 – 0.7	0.7 – 0.8	0.1 – 0.2	0.2 – 0.3
Frequency	15	13	20	27	20

22. One sample t-Distribution

A random sample of 10 boys had the following I.Q.'s: 70, 120, 110, 101, 88, 83, 95, 98, 107 and 100. Do these data support the assumption of a population mean I.Q. of 100? Find a reasonable range in which most of the mean I.Q. values of samples of 10 boys lie.

Aim

To find a reasonable range in which most of the mean I.Q. values of samples of 10 boys lie.

Procedure

- State the null hypothesis and alternative hypothesis
- State alpha, in other words, determines the significance level
- Compute the test statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

- To determine the critical value
- Compare the calculated test statistic to the critical value.

Calculation

Hypothesis

H_0: The data are consistent with the assumption of a mean I.Q. of 100 in the population. i.e., $\mu = 100$.

H_1: The data are not consistent with the assumption of a mean I.Q. of 100 in the population. i.e., $\mu \neq 100$.

Level of Significance

$\alpha = 0.05$

Test Statistic

Calculate mean and SD from the sample values of I.Q. is,

X	$\bar{x}$	**(x- $\bar{x}$)^2**
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
972		**1833.60**

Here, n = 10, $\bar{x}$ = 972/10 = 97.2 and S = 1833.60/9 = 14.27

Under the null hypothesis, the test statistic is

$$t = \frac{97.2 - 100}{14.27 / \sqrt{10}}$$

$$t = \frac{2.8}{4.514}$$

= 0.62

Degrees of freedom

The degree of freedom = n – 1

= 10 – 1 = 9df

Table Value = 2.262

Conclusion

Since the calculated value is less than the table value (0.62 < 2.262). We do not reject the null hypothesis. Hence, we conclude that the data are consistent with the assumption of a mean I.Q. of 100 in the population.

R Coding

 # one sample t distribution

```
IQ = c(70, 120, 110, 101, 88, 83, 95, 98, 107, 100)

t.test(IQ, mu=100,alternative="less")
```

R Output

One Sample t-distribution

data: IQ

t = -0.62034, df = 9, p-value = 0.2752

alternative hypothesis: true mean is less than 100

95 percent confidence interval:

  -Inf 105.4741

sample estimates:

mean of x

  97.2

Conclusion

Here the p-value is greater than the specified level of significance (0.275 > 0.05). We do not reject the null hypothesis. Hence, we conclude that the data are consistent with the assumption of a mean I.Q. of 100 in the population.

Observation Problem

The following data represents the number of hours that a rechargeable hedge trimmer operates before a recharge is required.

1.5	2.2	0.9	1.3	2.0	1.6	1.8	1.5	2.6	1.2	1.7

To test the hypothesis that this particular trimmer operates on the average of 1.8 minutes before requiring a recharge.

23. Two sample t-distribution

Below are given the gain in weights of pig fed on two diets A and B. gain in weight using:

Diet A: 25, 32, 30, 34, 24, 14, 32, 24, 30, 31, 35, 25

Diet B: 44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 22

Test if the two diets differ significantly regards their effect on increase in weight.

Aim

To test the two diets differ significantly regards their effect on increase in weight.

Procedure

- State the null hypothesis and alternative hypothesis
- State alpha,
- Compute the test statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

- To determine the critical value
- Compare the calculated test statistic to the critical value.

Calculation

Hypothesis

H_0: There is no significant difference between the mean increase in weight diet A and diet B. i.e., $\mu_x = \mu_y$.

H_1: There is significant difference between the mean increase in weight diet A and diet B. i.e., $\mu_x \neq \mu_y$.

Level of Significance

$\alpha = 0.05$

Test Statistic

Under the null hypothesis (H_o), the test statistic is,

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

Here, $n_1 = 12$, and $n_2 = 15$

Calculate mean and SD from the sample values of diet A and diet B,

x	**x-$\bar{x}$**	**(x- $\bar{x}$)^2**	**y**	$y - \bar{y}$	$(y - \bar{y})^2$
25	-3	9	44	14	196
32	4	16	34	4	16
30	2	4	22	-8	64
34	6	36	10	-20	400
24	-4	16	47	17	289
14	-14	196	31	1	1
32	4	16	40	10	100
24	-4	16	30	0	0
30	2	4	32	2	4
31	3	9	35	5	25
35	7	49	18	-12	144
25	3	9	21	-9	81
			35	5	25
			29	-1	1
			22	-8	64
336	**0**	**380**	**450**	**0**	**1410**

$$\bar{x} = \frac{336}{12} = 28 \text{ and } \bar{y} = \frac{450}{15} = 30$$

$$s^2 = \frac{\left[\sum(x-\bar{x})^2 + \sum(y-\bar{y})^2\right]}{n_1 + n_2 - 1}$$

$$s^2$$

$$t = \frac{28 - 30}{\sqrt{71.6\left(\frac{1}{12} + \frac{1}{15}\right)}} = \frac{-2}{\sqrt{10.74}}$$

$$t = -0.609 = |-0.609| = 0.609$$

Confidence Interval

A 95% confidence interval for the difference between sample Mean:

$$LL = [(\bar{x} - \bar{y})] - t_{0.05}\left(\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$

$$LL = [(28 - 30)] - 2.06\left(\sqrt{71.6\left(\frac{1}{12} + \frac{1}{15}\right)}\right)$$

$$LL = -2 - 2.06\sqrt{10.74}$$

$$LL = -2 - 6.40$$

$$LL = -8.40$$

$$UL = [(\bar{x} - \bar{y})] + t_{0.05}\left(\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$

$$LL = [(28 - 30)] + 2.06\left(\sqrt{71.6\left(\frac{1}{12} + \frac{1}{15}\right)}\right)$$

$$LL = -2 + 2.06\sqrt{10.74}$$

$$LL = -2 + 6.40$$

$$LL = 4.40$$

The 95% confidence interval is -8.40 to 4.40.

Degrees of freedom

The degrees of freedom = $n_1 + n_2 - 2$

= 12 + 15 – 2 = 25 d.f.

Table Value = 2.06

Conclusion

Since the calculated value is less than the table value (0.609 < 2.06). We do not reject the null hypothesis. Hence, we conclude that there is no significant difference between the mean increase in weight diet A and diet B. The 95% confidence interval is -8.40 to 4.40.

R Coding

 # Two sample t distribution

```
Weight = c(25, 32, 30, 34, 24, 14, 32, 24, 30, 31, 35, 25, 44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 22)

Diet = c(rep(1,12),rep(2,15))

t.test(Weight ~ Diet)
```

R Output

Two Sample t-distribution

data: Weight by Diet

t = -0.64573, df = 23.16, p-value = 0.5248

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-8.404739 4.404739

sample estimates:

mean in group 1 mean in group 2

28 30

Conclusion

Here the p-value is greater than the specified level of significance (0.525 > 0.05). We do not reject the null hypothesis. Hence, we conclude that there is no significant difference between the mean increase in weight diet A and diet B. The 95% confidence interval is -8.40 to 4.40.

Observation Problem

Measurement of the fat content of two kinds of ice cream, Brand A and Brand B yielded from ten ice cream types are given below:

Brand A	13.5	14	13.6	12.9	13	12.4	13.8	13.5	12.7	12.9
Brand B	12.9	13	12.4	13.5	12.7	12.8	12.9	12.3	13.4	12.6

Do the brands differ with respect to fat content?

APPENDIX

NORMAL DISTRIBUTION TABLE

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0394	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0,0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0722	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0,4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

CHI-SQUARE TABLE

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169