African Journal of Mathematics and Computer Science Research Vol. 5(8), pp. 148-152, May 2012
Available online at http://www.academicjournals.org/AJMCSR
DOI: 10.5897/AJMCSR12.028
ISSN 2006-9731©2012 Academic Journals

Full Length Research Paper

Statistical modeling of wastage using the beta distribution

Ebuh G. U.[1]*, Nwoke C.[1] and Ebuh A. C.[2]

[1]Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.
[2]Department of Computer Science, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.

Accepted 11 May, 2012

This work attempts to fit a set of observed industrial wastage to the beta distribution using the Chi-square goodness of fit test and it was found to fit the beta distribution at 1% significance level. The mean and the variance of the wastage proportion were found to be 0.429 and 0.0395, respectively. The parameters of the beta distribution were α = 2.25 and β = 2.97. The skewness of the beta distribution is 0.286 confirming the claim that the distribution is skewed to the right whenever β > α and the distribution is a unimodal distribution since α and β is greater than 1.

Key words: Parameter, goodness of fit, chi-square, probability, random variable, distribution.

INTRODUCTION

It is believed that in management of a firm, there must be proportionate wastage out of the general output or yield. These wastages can be seen as loss of money, time and manpower. The beta distribution has two positive parameters denoted as α and β. It is a continuous distribution defined on the interval (0, 1). In Bayesian statistics, It can be seen as the posterior distribution of the parameter P of the binomial distribution after observing α-1 independent events with probability P and β-1 with probability 1-P, if the prior distribution of P was uniform (Oyeka, 2009). The beta distribution can be of two kinds, the beta distribution of the first kind and the second kind. According to Feller (1968), a random variable X is said to have a beta distribution of the first kind if and only if its probability density is

$$F(x)=\frac{1}{\beta(\alpha,\beta)}\ X^{\alpha-1}1-X^{\beta-1},0 elsewhere 0<X<1\alpha,\beta>0 \quad (1)$$

The mean of the distribution is $\frac{\alpha}{\alpha+\beta}$ (2)

Its variance is $\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)}$ (3)

The beta distribution can take on different shapes depending on the values of the two parameters of α and β. When α = 1 and β = 1, the distribution is uniform over (0, 1) when α>1 and β>1, the distribution is unimodal when α = β, the beta distribution is symmetric about ½ whenever β > α, the distribution is skewed to the right while if α > β, the distribution is skewed to the left.

METHOD OF ANALYSIS

The major statistical tool that was used for this paper is the Chi-square goodness of fit test.

Hypothesis

H_o: The data fits the Beta distribution
H_1: The data does not fit the Beta distribution

The test statistics is given by

$$\chi^2_{cal}=\sum_{i=1}^{k}\left[\frac{(O_i-e_i)^2}{e_i}\right] \quad (4)$$

*Corresponding author. E-mail: ablegod007@yahoo.com.

where o_i is the observed frequency, e_i is the expected frequency and $e_i = nP_x$ where P_x is the expected relative frequency.

Decision rule

From the test statistics, if the calculated χ^2 is greater than the tabulated χ^2, the null hypothesis is rejected, that is rejected H_o if $\chi^2_{cal} > \chi^2_{1-\alpha,\ k-m-1}$ at significance level of α, otherwise the null hypothesis is rejected when k-m-1 is the degree of freedom.

K is the number of categories or class interval for which the observed and the expected frequency is available and m is the number of parameters estimated for the hypothesized distribution.

Relationship between the beta and other distribution using cumulating moment generating function (CMGF)

According to Oyeka et al. (2008), the CMGF is the weighted sum or cumulating of a set of its constituent moments of the distribution of the random variable of interest about zero. It is denoted by $g_n(c)$. The CMGF of the beta distribution is

$$g_n(c) = \sum_{t=0}^{n} \binom{n}{t} \frac{\lambda^{n-t}\Gamma(\alpha+\beta)\Gamma(ct+\alpha)}{\Gamma(\alpha)\Gamma(ct+\alpha)} \quad (5)$$

where c is the power of the distribution, α and β are the parameters of the distribution, n and t are the moments of the distribution. To get the CMGF of the gamma distribution (Equation 7) from the CMGF of the beta distribution, we put Equation (6) into (5)

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(ct+\alpha)} = \beta^{ct} \quad (6)$$

$$g_n(c) = \sum_{t=0}^{n} \binom{n}{t} \frac{\lambda^{n-t}\Gamma(\alpha+\beta)\Gamma(ct+\alpha)}{\Gamma(\alpha)\Gamma(ct+\alpha)} \quad (7)$$

Similarly to get the CMGF of an exponential distribution (Equation 8) from the CMGF of a beta distribution, we put α = 1 and $\frac{\Gamma(\alpha+\beta)}{\Gamma(ct+\alpha)} = \beta^{ct}$ into Equation (7) to obtain:

$$g_n(c) = \sum_{t=0}^{n} \binom{n}{t} \lambda^{n-t} \beta^{ct} \Gamma(ct+1) \quad (8)$$

To also get a chi-square distribution from the beta

$$g_n(c) = \sum_{t=0}^{n} \binom{n}{t} \frac{\lambda^{n-t} 2^{ct} \Gamma\left(ct+\frac{k}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \quad (9)$$

ANALYSIS

Data presentation on the total number of items produced and the number of defective products for thirty-one randomly selected days is shown in Table 1. Table 2 shows the group frequency table for Table 1.

Parameter estimation

Mean, $\bar{x} = \frac{\sum fx}{\sum f} = \frac{13.295}{31} = 0.429$

Variance, $s^2 = \frac{\sum f(x-\bar{x})^2}{N-1} = \frac{1.186352}{31-1} = \frac{1.186352}{30}$

$= 0.0395$

Where the relationship among the mean, alpha and beta distributions is given by the

$$\bar{x} = \frac{\alpha}{\alpha+\beta},$$

and the relationship among the variance, alpha and beta distributions is given by the

$$s^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Thus

$$\bar{x} = \frac{\alpha}{\alpha+\beta} = 0.429 \quad (10)$$

$$s^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.0395 \quad (11)$$

From Equation 10, $\alpha = 0.429(\alpha+\beta)$

$\therefore\ \alpha = 0.751\beta$ (12)

Substituting Equation 12 into (11) gives

Table 1. Data presentation on the total number of items produced and the number of defective products for thirty-one randomly selected days.

S/N	Total number produced (X)	Total number of defective products	Proportion of defective products
1	5900	2320	0.393
2	5628	1830	0.325
3	5954	1480	0.249
4	6000	2400	0.400
5	5488	2090	0.381
6	4800	900	0.188
7	5004	1760	0.352
8	4570	2080	0.455
9	6400	3200	0.500
10	4870	2980	0.612
11	6030	1680	0.279
12	6200	1920	0.310
13	6077	2420	0.398
14	5320	1978	0.372
15	5328	2368	0.444
16	4600	2840	0.609
17	5368	3404	0.634
18	5523	1800	0.326
19	4830	2000	0.414
20	6380	3276	0.513
21	5418	2800	0.517
22	5540	1540	0.298
23	4836	2380	0.492
24	5930	2168	0.366
25	4634	1480	0.319
26	4800	1200	0.250
27	5892	2000	0.475
28	5085	4082	0.803
29	5251	1238	0.236
30	5450	3418	0.627
31	5974	3348	0.560
	163240	71060	

$$s^2 = \frac{0.751\beta^2}{(0.751\beta+\beta)^2(0..751\beta+\beta+1)} = 0.0395$$

β = 2.97

From Equation 12, α = 0.751 × 3 = 2.25. Therefore, α = 2.25, β = 2.97.

COMPUTATION OF THE OBSERVED VALUES OF WASTAGE PROPORTION INTO THE BETA DISTRIBUTION

From Equation 1

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } 0<x<1 \ \alpha,\beta>0 \\ 0 & \text{elsewhere} \end{cases}$$

Substitute the values of α = 2 and β = 3 in the above equation

$$= \frac{\Gamma 5}{\Gamma 2\,\Gamma 3} = \frac{4!}{1!2!} = 12$$

$$\therefore f(x) = \begin{cases} 12\,x\,(1-x)^2 & 0<x<1 \ \alpha,\beta>0 \\ 0 & \text{Elsewhere} \end{cases} \qquad (13)$$

Calculation of the expected frequencies

We first calculate the probability corresponding to the

Table 2. A group frequency table for Table 1.

Class interval	Tally	Frequency (f)	Class mark (x)	Fx	x- $\bar{x}$	(x- $\bar{x}$)^2	f(x- $\bar{x}$)^2	Class boundaries
0.10 – 0.19	I	1	0.145	0.145	-0.284	0.080655	0.080655	0.095 – 0.195
0.20 – 0.29	IIII	4	0.245	0.980	-0.184	0.033856	0.135424	0.195 – 0.295
0.30 – 0.39	~~IIII~~ ~~IIII~~	10	0.345	3.450	-0.084	0.007056	0.07056	0.295 – 0.395
0.40 – 0.49	~~IIII~~ II	7	0.445	3.115	0.016	0.000256	.001792	0.395 – 0.495
0.50 – 0.59	IIII	4	0.545	2.180	0.116	0.013456	0.53824	0.495 – 0.595
0.60 – 0.69	IIII	4	0.645	2.580	0.216	0.046656	0.186624	0.595 – 0.695
0.70 – 0.79		0	0.745	0	0.316	0.099856	0	0.695 – 0.795
0.80 – 0.89	1	1	0.845	0.845	0.416	0.173056	0.173056	0.795 – 0.895
		31		13.295			1.186352	

Table 3. Calculation of expected frequencies and $\frac{(O_i - e_i)^2}{e_i}$.

Class interval	Class boundaries	Observed frequency (O_i)	Expected relative frequency P(x)	Expected frequency e_i = nPx	$(O_i - e_i)^2$	$\frac{(O_i - e_i)^2}{e_i}$
0.10 – 0.29	0.095 – 0.195	5	0.291955688	9.050626328	16.40757365	1.8128661
0.30 – 0.39	0.295 – 0.395	10	0.17665135	5.47619185	20.46484018	3.737056834
0.40 – 0.49	0.395 – 0.495	7	0.163820369	5.078431129	3.6924269	0.72708024
0.50 – 0.59	0.495 – 0.595	4	0.135029352	4.185909912	0.034562495	0.008256866
0.60 – 0.89	0.595 – 0.895	5	0.180742044	5.603003364	0.363613057	0.064896098
						6.350156

lower upper class boundaries of each interval, and then calculate their expected frequency for each interval.

The expected frequency for the first class interval 0.10 to 0.29 is

$$P(0.095<x<0.295) = 12\int_{0.095}^{0.295} x - 2x^2 + x^3 \partial x =$$

$$12\left[\frac{x^2}{2} - \frac{2x^3}{3} + \frac{x^4}{4}\right]_{0.095}^{0.295}$$

= 12(0.02829092 – 0.0039612793) = 0.291955688.

The expected frequency for this class interval is calculated as

np(x) = 31(0.291955688) = 9.050626328

The expected frequencies for the other class interval are computed as shown in Table 3.

CHI-SQUARE (χ^2) GOODNESS OF FIT TEST AND HYPOTHESIS

Chi-Square would be used to see if this data fits the beta distribution at 1% level of significance. H_o: the data fits the beta distribution, H_1: the data does not fit the beta distribution

$$\chi^2 = \sum\left(\frac{(O_i - e_i)^2}{e_i}\right) = 6.350156$$

Chi-square tabulated = $\chi^2_{\alpha,k-m-1}$ where $\chi^2_{0.01,\ 5-2-1}$ = 9.210

Decision and conclusion

Since 6.350156 <9.210, we accept that the null hypothesis, hence implying that the data fits the Beta distribution with its parameter α = 2.25 and β = 2.97 at 1% level of significance.

Calculation of skewness

The skewness of the Beta distribution is given as

$$\frac{2(\beta-\alpha)\sqrt{(\alpha+\beta+1)}}{(\alpha+\beta+2)\sqrt{\alpha\beta}} = \frac{2(3-2)\sqrt{(2+3+1)}}{(2+3+2)\sqrt{(2)(3)}} = \frac{2\sqrt{6}}{7\sqrt{6}} = 0.286$$

Conclusion

From the results obtained, the mean and variance of the wastage proportion are 0.429 and 0.0395, respectively, the parameters of the distribution are α = 2.25 and β = 2.97, the skewness of the distribution is 0.286, the observed proportion of wastage (defective) fits the Beta distribution with a chi-square calculated value of 6.350156 which is less than the tabulated chi-square value of 9.210 at 1% significance level.

It also shows that the values of the wastage proportion fit the beta distribution at 1% significance level and that the density of the distribution is skewed to the right with α = 2.25 and β = 2.97. The mean and the wastage proportion are 0.429 and 0.0395, respectively, which is high and therefore need to be minimized.

REFERENCES

Feller W (1968). An Introduction to Probability Theory and its Applications. Wiley, New York. pp. 123-124.

Oyeka ICA (2009). An Introduction to Applied Statistics Methods, (Eight Edition) Nobern Avocation Publishing Company, Enugu. pp. 56-141.

Oyeka ICA, Ebuh GU, Nwosu CR, Utazi EC, Ikpebu PA, Obiora-Ilouno HO (2008). "The n^{th} moment of x^{c} " ISSN 0331-9504. J. Niger. Stat. Assoc., 20: 24-33.