

THE F- DISTRIBUTION WITH APPLICATIONS TO NUMBER OF FEMALE PATIENTS WITH CANCER.

ABSTRACT

The F-distribution is a particular parameterization of the beta prime distribution, which is also called the beta distribution of the second kind. This considered a new four-parameter generalized version of the Fisher Snedecor distribution called Beta- F distribution is introduced. The comprehensive account of the statistical properties of the new distributions was considered. Formal expressions for the cumulative density function, moments, and moment generating function and maximum likelihood estimation as well as its Fisher information were obtained.

This study employed a descriptive survey research design. The target population was female patients with cancer in Lagos State University teaching Hospital for complete for 36 days. The data collected was analyzed using both descriptive statistics and presented in tables, charts and figures. Some well-known distributions were fit, some existing single distributions and the proposed F distribution to these data using the method of maximum likelihood was applied using Micro-Soft Excel 2007.

The study established that the minimum AIC and minimum statistic on kolmogorov Smirnov of the F-distribution shows that female patients with cancer data better than the other life time distributions considered. The study concluded that Fisher-distribution is better off than existing single distribution with reference to Akaike Information Criterion (AIC) been used. It is instructive to note that the F-distribution preserves the originality of the data without transformation.

Key Words: *Fisher-Snedecor distribution, Beta-F distribution, Outlier, Maximum likelihood method.*

1.0 INTRODUCTION

F-distribution is a continuous probability distribution (also known as Snedecor's F distribution or the Fisher-Snedecor distribution) named in after R.A. Fisher and George W. Snedecor. It arises frequently as the null distribution of a test statistic (hypothesis testing) and used to develop confidence interval and in the Analysis of Variance (ANOVA) for comparison of several populations means. The F-distribution arises frequently as the null distribution of a test statistics most notably in ANOVA where for example the null hypothesis that two independent normal variances are equal and the observed sums of some appropriately selected squares are then examined to see whether their ratio is significantly incompatible with this null hypothesis. This

paper focuses on developing a generalized f-distribution that is capable of handling data that are normal distributed.

2.0 LITERATURE REVIEW

The practical use of distributions can be traced back to the use of Green functions in the 1830s to solve ordinary differential equations, but was not formalized until much later. According to Kolmogorov & Fomin (1957), generalized functions originated in the work of Sergei Sobolev (1936) on second-order hyperbolic partial differential equations, and the ideas were developed in somewhat extended form by Laurent Schwartz in the late 1940s. According to his autobiography, Schwartz introduced the term "distribution" by analogy with a distribution of electrical charge, possibly including not only point charges but also dipoles and so on. Gårding (1997) comments that although the ideas in the transformative book by Schwartz (1951) were not entirely new, it was Schwartz's broad attack and conviction that distributions would be useful almost everywhere in analysis that made the difference.

3.0 THE F DISTRIBUTION

3.1 DERIVATION OF F-DISTRIBUTION

Proof:

3.2 Let $F=h(V)$ and $V=h^{-1}(F)$ and obtain $f_F(f/w)$ by method of transformations:

$$\text{Fix } W=w \quad f(v|w)=f(v)=\frac{1}{\Gamma\left(\frac{k_1}{2}\right)2^{k_1/2}}v^{\frac{k_1}{2}-1}e^{-v/2} \quad v>0$$

$$F=h(V)=\frac{V/k_1}{w/k_2}=cV \quad c=\frac{k_2}{wk_1} \Rightarrow V=h^{-1}(F)=\frac{F}{c}$$

$$\Rightarrow f_F(f|w)=f_V(h^{-1}(f)|w)\left|\frac{dV}{dF}\right|=\frac{1}{\Gamma\left(\frac{k_1}{2}\right)2^{k_1/2}}\left(\frac{f}{c}\right)^{\frac{k_1}{2}-1}e^{-(f/c)/2}\left|\frac{1}{c}\right|$$

$$=\frac{1}{\Gamma\left(\frac{k_1}{2}\right)2^{k_1/2}}\left(\frac{fwk_1}{k_2}\right)^{\frac{k_1}{2}-1}e^{-\frac{fwk_1}{2k_2}}\left(\frac{wk_1}{k_2}\right)=\frac{1}{\Gamma\left(\frac{k_1}{2}\right)}\left(\frac{k_1}{2k_2}\right)^{\frac{k_1}{2}}f^{\frac{k_1}{2}-1}w^{\frac{k_1}{2}}e^{-\frac{fwk_1}{2k_2}} \dots \dots \dots i)$$

3.3 CONDITIONAL DISTRIBUTION OF F|W=W

$$f_W(w)=\frac{1}{\Gamma\left(\frac{k_2}{2}\right)2^{k_2/2}}w^{\frac{k_2}{2}-1}e^{-w/2} \quad w>0$$

$$\Rightarrow f_{F,W}(f,w)=f_F(f|w)f_W(w)=\frac{1}{\Gamma\left(\frac{k_1}{2}\right)}\left(\frac{k_1}{2k_2}\right)^{\frac{k_1}{2}}f^{\frac{k_1}{2}-1}w^{\frac{k_1}{2}}e^{-\frac{fwk_1}{2k_2}}\frac{1}{\Gamma\left(\frac{k_2}{2}\right)2^{k_2/2}}w^{\frac{k_2}{2}-1}e^{-w/2}$$

$$=\frac{1}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)2^{\frac{k_1+k_2}{2}}}\left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}}f^{\frac{k_1}{2}-1}w^{\left(\frac{k_1}{2}+\frac{k_2}{2}\right)-1}\exp\left\{-\frac{w}{\left[\frac{fk_1}{2k_2}+1\right]}\right\} \quad f,w>0 \dots \dots \dots i)$$

3.4 Marginal Distribution of W, Joint Distribution of F,W

$$\begin{aligned}
\Rightarrow f_F(f) &= \int_0^{\infty} f_{F,W}(f,w)dw = \frac{1}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)2^{\frac{k_1+k_2}{2}}}\left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}}f^{\frac{k_1}{2}-1}\int_0^{\infty}w^{\left(\frac{k_1+k_2}{2}\right)-1}\exp\left\{-\frac{w}{\left[\frac{fk_1}{2k_2}+\frac{1}{2}\right]^H}\right\}dw \\
&= \frac{1}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)2^{\frac{k_1+k_2}{2}}}\left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}}f^{\frac{k_1}{2}-1}\Gamma\left(\frac{k_1+k_2}{2}\right)\left\{\left[\frac{fk_1}{2k_2}+\frac{1}{2}\right]^H\right\}^{\frac{k_1+k_2}{2}} \\
&= \frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}\left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}}f^{\frac{k_1}{2}-1}\left(\frac{1}{2^{\frac{k_1+k_2}{2}}}\right)\left(\frac{1}{2}\right)^{\frac{k_1+k_2}{2}}\left[1+\frac{fk_1}{k_2}\right]^{\frac{k_1+k_2}{2}} \\
&= \frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}\left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}}f^{\frac{k_1}{2}-1}\left[1+\frac{fk_1}{k_2}\right]^{\frac{k_1+k_2}{2}} \quad f > 0 \dots\dots\dots (i)
\end{aligned}$$

4.0 APPLICATION

The data represent the cases of female patients with cancer in Lagos state university teaching hospital for complete for 36 days. We fit some well-known distributions, some existing single distributions and the proposed F distribution to these data using the method of maximum likelihood.

The (maximum likelihood estimation) MLEs of the parameters, the Akaike Information Criterion (AIC) and test of goodness of fit using Kolmogorov Smirnov for the fitted models are listed in Table 2 and 3

TABLE 1: DESCRIPTIVE STATISTICS ON FEMALE PATIENTS WITH CANCER

Min	Q ₁	Q ₂	Mean	Q ₃	Max	Var	Skewness	Kurtosis	Std error	Range
20	27	34	35.139	41	56	71.494	0.29457	-0.27126	1.4092	36

TABLE 2: NORMALITY TEST OF SOME SINGLE DISTRIBUTIONS

#	MODEL	KOLMOGOROV SMIRNOV	
		Statistic	Rank
1	Beta	0.18195	5
2	Exponential	0.43935	7
3	Exponential (2P)	0.23867	6
4	Gamma	0.11098	3
5	Gamma (3P)	0.11238	4
6	Normal	0.10989	2
7	F-distribution	0.09909	1

TABLE 3: MLE AND STATISTICS FOR DIFFERENT SINGLE DISTRIBUTIONS

	MODEL	MLE	AIC	RANK
1	Beta	$\alpha_1=1.4372$ $\alpha_2=1.9802$ $a=16.001$ $b=56.002$	830.7356	5
2	Exponential	$\lambda=0.02846$	832.1738	7
3	Exponential (2P)	$\lambda=0.06606$ $\mu=20.0$	830.8829	6
4	Gamma	$\alpha=17.27$ $\beta=2.0346$	830.1552	3
5	Gamma (3P)	$\alpha=15.553$ $\beta=2.1355$ $\gamma=1.9253$	830.6838	4
6	Normal	$\mu=8.4554$ $\sigma=35.139$	829.7247	2
7	F-distribution	$a=20.494$ $b=49.784$	828.1691	1

5. CONCLUSION

We studied the use of Fisher distribution in the obtaining some single distributions. We noted some of the results which were obtained in doing this. Following the pattern of these distributions, we have derive the Fisher Snedecor distribution and studied some of its properties

which include the moments of the beta- Fisher Snedecor distribution from which its mean, variance, skewness and kurtosis could be derived. The Fisher distribution is said to be used in virtually all fields where interest is on testing the equality of means across populations. Aside this function, its capabilities to model life time data was uncovered by comparing it with other existing single distributions and beta compounded distributions using cancer remission data previously studied by Lee and Wang (2003) , Lemonte and Cordeiro (2011) and Luz M. Zea (2012). Based on the minimum AIC and minimum statistic on kolmogorov Smirnov the F distribution therefore described female patients with cancer data better than the other life time distributions considered in this paper.

REFERENCES:

- Adepoju, K.A**, Chukwu A.U and Min Wang (2014)” The Beta Power exponential distribution. Journal of Statistical Science and Application 2(2014) 37-46
- Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. Commun. Stat. Theory, 31, 497-512.
- Famoye, F., Lee, C. and Olumolade, O. (2005). The Beta-Weibull Distribution. J. Stat. Theory Appl., 4, 121-136.
- Lee, E. T., & Wang, J. W. (2003). *Statistical Methods for Survival Data Analysis* (3rd ed.). New York: Wiley.
- Lemonte, A. J., & Cordeiro, G. M. (2011). An extended Lomax distribution. *Statistics*, 1-17.
<http://dx.doi.org/10.1080/02331888.2011.568119>

Nadarajah, S., Gupta, A.K., (2004). The beta Frechet distribution. *Far East Journal of Theoretical Statistics*, 14, 15-24.

Nadarajah, S., Kotz, S., (2004). The beta Gumbel distribution. *Mathematical Problems in Engineering*, 10, 323-332.

Nadarajah, S., Kotz, S.,(2006). The beta exponential distribution. *Reliability Engineering and System Safety*, 91, 689-697.

Shittu, O.I, Kazeem A. Adepoju **(2013)**.”On the Beta-Nakagami Distribution” (Canada journal of progress in applied Mathematics, vol 5, No1, 2013)