Available online at www.pelagiaresearchlibrary.com

Pelagia Research Library

Advances in Applied Science Research, 2014, 5(2):220-228

ISSN: 0976-8610
CODEN (USA): AASRFC

Determination of maturity date barley crops using lognormal parametric survival function

Nwaosu S. C. and Onuche P.

Department of Mathematics/Statistics/Computer Sciences, University of Agriculture, Makurdi, Benue State

ABSTRACT

Determination of maturity date and group classification of forty nine (49) varieties of Barley Crop were classified out using the experimental data collected from lake Chad, Research Institute in Maiduguri after the 2011cropping season. The purpose was to group the varieties into early, normal and late maturity varieties so as to assist plant breeders carry out further work on varieties evaluations. Data were collected on 50% heading date, yield and plant height as three major characteristics in Crop varieties needed in breeding work. Lognormal density function was used in the parametric Survival analysis since it has the least Anderson - Darling (adj) value after the goodness of fit test. One way analysis variance (ANOVA) was employed to test for the group homogeneity as the groups were determined using the first, second and third quartile values. The results show that the median heading date was 63 ± 0.96 days with an average yield of 0.77kg/plot. The first quartile (early maturity date) have an average 58 ±0.93 days with the highest average yield of 1.077kg/plot, and the third quartile (late maturity date) have an average1.075 days of 66 ± with an average yield of 0.58kg/plot. The authors recommend the selection of varieties from the first group that exhibited short maturity date highest average yield for further breeding work and an extension of this work using germination date, panicle count and weight of 1000 grains.

INTRODUCTION

Barley is a common name for any of a genus of cereal grass (Hordeum). It originated from Asia and Ethiopia, and is one of the most ancient of cultivated plants (Encarta, 2012). It is now the fourth largest grain crop after Wheat, Rice, and Corn.

In United States, Canada, as well as in the greater portion of Europe, Barley is sown in the spring. Along the Mediterranean Sea and in parts of California and Arizona, it is sown in the fall. Drought resistant and hardy barley can be grown on marginal Crop land. Salt resistant strains are being developed to increase its usefulness in coastal regions. Barley germinates at about the same temperature as wheat. Barley grain, hay, strain and several by product are used as feed. The grain is used for malt beverages and in cooking. Like other Cereals, Barley contains a large proportion of Carbohydrate (65%) and Protein (12.8%).

In Nigeria, Barley is grown in the Northern part of the Country between latitude 10^0-14^0N and altitude of 240-340m^2 during the cold season under irrigation which involve huge cost of production. As in wheat, it during heading, the temperature exceeds35^0C, the yield is reduced, (Nwaosu, 2010, Nwaosu *et al* 2010). In the recent times experiments carried out in Lake Chad Research Institute is focused on rainfall conditions on Mambilla Plateau in Taraba State of Nigeria.

Forty nine (49) varieties obtained from *Cymmit* during the 2012 were evaluated.

Since the crop thrives in cold environment within which certain temperature requires that does not last long, varieties that are short maturity duration, with average plant height and high yield are most desirable. Since no variety meets these characteristics, cross breeding becomes imperative. Breeding improves, stabilize crop varieties and increase productivity which culminates into food availability (Spencer, 1988).

It is therefore necessary to develop Statistical methods that incorporate these characteristics of Barley to characterize the varieties into early, normal and late maturing varieties. The breeders can then advance their activities with these varieties. This will also greatly reduce the cost of repeating these trials in order to look out for these characteristics.

Fifty percent (50%) heading is the date in which 50% of the varieties have headed, and then is converted into days from the date the varieties germinated. These data were used for this exercise in the absence of maturity date, since like wheat; from heading date to maturity date takes a maximum of two (2) weeks. In Agricultural production, plant height and growth duration are among the variables that can be used in grouping measurement (Kwanchai *et al* 1984).

This work is divided into four (4) sections. Section one (1) is introduction, section two(2) is the methodology, section three (3) is the results while section four (4) is the general discussion, conclusion and recommendations.

MATERIALS AND METHODS

Forty nine (49) exotic varieties of Barley were acquired and tested at Lake Chad Research Institute Experimental Farm in 2011-2012 farming season. The data for this paper were lifted from the experimental log book of the Institute. The authors verified that the experimental procedures were strictly adhered to. The data used for the paper were the 50% heading date. Also data were collected for plan height and yield as these were the parameters that are needed for the breeding work.

To determine the appropriate density function for the survival analysis, four probability density functions, Exponential, Weibull, Lognormal base e and Gamma (Seefig.1). Lognormal distribution was used because it stood out of other distributions. Percentiles were used to segregate the varieties into groups. Analysis of Variance (ANOVA) and Duncan multiple range test (DMRT) were employed to test for homogeneity between groups.

2.1 Survival Analysis

Survival Analysis also known as time-to-event analysis or time –to-failure analysis is a branch of Statistics which deals with the death of Biological process or organisms. The function denoted by $S(t)$, describes the probability that an individual survives longer time t. (Consuelo *et al*, 2005, Allison, 1995, Cox *et al*, 1999, Terry *et al*, 2000 and Ismor, 2008) and is given as follows;

$$S(t) = P[T > t] = 1 - P[T \leq t] = 1 - F(t)$$
$$= 1 - \int_{-\infty}^{t} f(t)dx$$

where t is a realization of the time random variable T.

$$f(t)dx = \frac{d}{dt}(1 - S(t))$$

2.2 Hazard function

Let us consider what happens to the conditional probability of time t as the time interval shrinks towards 0.

$$h(t) = f(t|T \geq t) = F'(t|T \geq t)$$
$$= \lim_{\Delta t \to 0} \frac{f(t + \Delta t|T \geq t) - F(t|T \geq t)}{\Delta t}$$
$$= \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t|T \geq t)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{P[t \leq T \leq t + \Delta t, T \geq t]}{\Delta t P[T \geq t]}$$
$$= \lim_{\Delta t \to 0} \frac{P[t \leq T \leq t + \Delta t]}{\Delta t[1 - F(t)]}$$
$$\frac{P[maturity\ between\ t\ and\ t + \Delta t]}{P[not\ matured\ at\ time\ t]}$$
$$= \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$$

This is known as the hazard function or the failure rate function (maturity rate function). It is also known as an instantaneous failure rate or the conditional density of failure rate at time t, given that the variety has survived until time t (Engelhart, 1995).

An increasing $h(t)$ at time t indicates that the variety is more likely to mature in the next increment of time $(t, t + \Delta t)$. That is the variety is maturing with time. In this study, we made use ofnon-censored time since there is no fixed time for the varieties of Barley to mature or head and the experiment terminated when the varieties completely matured and were harvested.

2.3 Lognormal Distribution

Suppose we have a random variable T that has a Lognormal distribution then $X = log\ T$ is normally distributed with mean $\mu_z = \mu$ and variance ${\sigma_z}^2 = \sigma^2$ then the probability density function (pdf) of T by change of variable technique gives (http://www.math.uah.edu/stat)

$$f_T(t) = \left|\frac{dlog(t)}{dt}\right| * f_X(logt)$$
$$= \frac{1}{t^2\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(logt-\mu)^2}$$

Hence,

$$F(t) = \emptyset(\frac{ln\, t - \mu}{\sigma}), \quad t \in (0, \infty)$$

and the quartile function is given by

$$F^{-1}(P) = Exp[\mu + \sigma\emptyset^{-1}(P)], \qquad 0<P<1$$

It then follows that

$$P[T \leq t_p] = P$$

And

$$Z_p = \frac{lnt_p - \mu_{lnt}}{\sigma_{lnt}} \quad \rightarrow t_p = e^{Z_p * \sigma_{lnt} + \mu_{lnt}}$$

RESULTS

The data analyses were carried out using Minitab Statistical Software Version 13 (see Appendix).Figure 1 shows the plot of four probability density functions with the adjusted Anderson- Darling test of goodness of fit. The Lognormal base e showed the least value of the goodness of fit test value of 0.858 while Exponential shows the lack of fit. This is not unconnected to the fact that the result of the hazard function is not a constant. Figure 2 shows the probability plot of the Lognormal density function with 95% confidence interval. It can be seen that the entire data lie within the region, strengthening the fact that Lognormal is the best fit. Table 1 shows the result of the maximum likelihood estimation of Lognormal distribution with log-likelihood value of -160.034 which is significant. The location parameter μ_{logx}= 4.128 with the scale parameter valueσ^2_{logx} = 0.102. The Anderson- Darling (Adj) for Lognormal is 1.064. The estimate of the median is62.39 ± 0.913 days with the first quartile giving 57.93$\pm$0.94 days and the third quartile 66.49 ± 1.07 days and the interquartile range of 8.56 ± 0.86 days.

Figure 3 shows the parametric survival plot with 95% confidence interval. One can easily read the probabilities of varieties failing and any given day while Figure 4 shows the parametric hazard plot. From the graph, we can read off the instantaneous heading date for the varieties. From the graph, the instantaneous heading date of 63 is 0.1 in other words 10 varieties are expected to mature daily. This result shows that by day 63rd 50% of the varieties are expected to have headed with the expected maturing date of 70 days. Also 25% of the varieties are expected to have headed by day 58 with expected maturing date of 63rd day. Also 75% of the varieties were expected to have headed by 66th day with expected maturing day of 74.These values were used in classifying the varieties into early, normal and late heading dates. The result of the segregation showed that 18 varieties are in early heading group, 18 varieties are also in normal heading group while 13 varieties are in late heading group. The one way analysis of variance (ANOVA) shows that there is significant differences ($p<0.001$) between the average heading dates for the groups. The early maturing group has an average maturing date of 56days followed by the normal maturing group with 63 days, then the late group with 71 days. The groups were also compared in their yield and height parameters. The early maturing group shows a significant highest mean yield of 1.077kg/plot followed by normal group with the mean yield of 0.73kg/plot while the late group gave an average yield of 0.58kg/plot. From the table also there is no significant difference between the groups with respect to the plant height.

CONCLUSION

From the ongoing, it was observed that there are differences in maturity days and in yield parameter between the three groups. The 95% confidence limits for the mean values shows that there are some varieties in the normal and late groups whose average yield are greater than some varieties in the early group. The authors suggest that these varieties should be selected by the plant breeders for further work. The authors also suggest that some other parameters like the panicle count, tiller count and days to germination to be included in the further work since they also important parameters in breeding work.

REFERENCES

[1] "Barley" Microsoft (CR) Encarta ® 2012 (DVD). Redmond, W. A.; Microsoft Corporation **2008**.

[2] Bain, L. J., and Engelhart, M. (**1993**). Introduction To Probability and Mathematical Statistics, 476-496.

[3] Consuello, G., Dorain, S. Chris, S., and Angela, W. (**2005**). Estimation of The Cumulative Hazard function.

[4] Http://www.Math.Uah.edu/Stat. downloaded on the 24-07-**2013**

[5] Ismor Fischer (**2008**). Survival Analysis (Pdf) Lecture notes STAT 541/8-4

[6] Kwanchai, A. G., and Arturo, A. G. (**1984**) Statistical Procedure for Agricultural Research. An International Rice Research Institute Book. 2nd Edition. A wiley-Inter-Science Publication. John Wiley and sons, New York 75

[7] Nwaosu, S. C. (**2009**). Statistical Analysis on the effect of seed rates, row spacing and blocking on the yield of Bread Wheat varieties at Ngala Vertisols of Borno State. MSc thesis submitted to the Department of Statistics, University of Ibadan.

[8] Nwaosu, S. C., Bakari, H. C., and Chukwu, A. U (**2010**). *Journal of Mathematical Sciences*. 21(1):67-75.

[9] Olabanji, O. G., Ikwelle, M. C., Nwaosu, S. C, and Onyenwe, C. N. (**1990**). Effect of Sowing date on Growth and Grain Yield of Wheat Grown In the Highland of mambilla (Gembu). In Wheat In Nigeria: Production, Processing And Utilization. (eds. Rayar, A. J., Kaigama, B. K., Olukosi, J. © LCRI, IAR and UNIMAID)

[10] Spencer, D. S. C(**1988**). Improving Farming System In West Africa and Central Africa. "The role of IITA on Improved Agricultural Technologies For Small Scale Nigerian Farmers" (eds. Abalu, G. O and Kalu, B. A)

APPENDIX

Distribution Analysis: Data (time in days)

Step	Log-Likelihood
0	-160.037
1	-160.034
2	-160.034
3	-160.034

Variable: time

Censoring Information	Count
Uncensored value	49

Estimation Method: Maximum Likelihood
Distribution : Lognormal base e

Parameter Estimates

		Standard	95.0% Normal CI	
Parameter	Estimate	Error	Lower	Upper
Location	4.12815	0.01460	4.09954	4.15675
Scale	0.10217	0.01032	0.08382	0.12454

Log-Likelihood = -160.034

Goodness-of-Fit

Anderson-Darling (adjusted) = 1.0639

Characteristics of Distribution

		Standard	95.0% Normal CI	
	Estimate	Error	Lower	Upper
Mean(MTTF)	62.3875	0.9130	60.6235	64.2028
Standard Deviation	6.3910	0.6623	5.2162	7.8303
Median	62.0627	0.9059	60.3124	63.8638
First Quartile(Q1)	57.9298	0.9368	56.1225	59.7953
Third Quartile(Q3)	66.4906	1.0752	64.4162	68.6317
Interquartile Range(IQR)	8.5608	0.8751	7.0065	10.4599

Table of Percentiles

		Standard	95.0% Normal CI	
Percent	Percentile	Error	Lower	Upper
1	48.9333	1.3750	46.3112	51.7037
2	50.3153	1.2949	47.8403	52.9184
3	51.2123	1.2438	48.8317	53.7091
4	51.8977	1.2055	49.5880	54.3149
5	52.4619	1.1745	50.2096	54.8152
6	52.9470	1.1485	50.7431	55.2466
7	53.3760	1.1260	51.2140	55.6292
8	53.7630	1.1062	51.6381	55.9754
9	54.1175	1.0885	52.0256	56.2935
10	54.4458	1.0725	52.3839	56.5889
20	56.9489	0.9673	55.0843	58.8767
30	58.8250	0.9157	57.0573	60.6475
40	60.4768	0.8968	58.7445	62.2603
50	62.0627	0.9059	60.3124	63.8638
51	62.2219	0.9083	60.4668	64.0279
52	62.3816	0.9111	60.6212	64.1931
53	62.5419	0.9142	60.7756	64.3595
55	62.8647	0.9212	61.0849	64.6964
56	63.0275	0.9252	61.2400	64.8671
57	63.1913	0.9295	61.3955	65.0396
58	63.3563	0.9341	61.5516	65.2138
60	63.6902	0.9444	61.8658	65.5684
61	63.8594	0.9501	62.0242	65.7490
62	64.0304	0.9561	62.1835	65.9321
63	64.2031	0.9626	62.3440	66.1177
64	64.3779	0.9694	62.5057	66.3062
65	64.5548	0.9766	62.6688	66.4976
66	64.7341	0.9842	62.8335	66.6922
69	65.2880	1.0098	63.3385	67.2975
70	65.4787	1.0193	63.5110	67.5073
71	65.6729	1.0293	63.6861	67.7217
73	66.0729	1.0511	64.0446	68.1653
74	66.2793	1.0628	64.2286	68.3955
80	67.6357	1.1488	65.4212	69.9252
90	70.7452	1.3935	68.0660	73.5298
91	71.1744	1.4315	68.4233	74.0362
92	71.6437	1.4741	68.8120	74.5919
93	72.1632	1.5224	69.2403	75.2096
94	72.7479	1.5780	69.7198	75.9076
95	73.4206	1.6438	70.2685	76.7141
96	74.2188	1.7239	70.9157	77.6757
97	75.2120	1.8267	71.7157	78.8788
98	76.5529	1.9702	72.7872	80.5134
99	78.7150	2.2118	74.4972	83.1716

Table of Survival Probabilities
95.0% Normal CI

Time	Probability	Lower	Upper
51.0000	0.9727	0.9263	0.9917
52.0000	0.9583	0.9013	0.9851
52.0000	0.9583	0.9013	0.9851
53.0000	0.9388	0.8708	0.9750
55.0000	0.8815	0.7932	0.9391
55.0000	0.8815	0.7932	0.9391
55.0000	0.8815	0.7932	0.9391
56.0000	0.8428	0.7462	0.9114
56.0000	0.8428	0.7462	0.9114
56.0000	0.8428	0.7462	0.9114
56.0000	0.8428	0.7462	0.9114
56.0000	0.8428	0.7462	0.9114
57.0000	0.7975	0.6942	0.8765
57.0000	0.7975	0.6942	0.8765
58.0000	0.7462	0.6381	0.8344
58.0000	0.7462	0.6381	0.8344
58.0000	0.7462	0.6381	0.8344
58.0000	0.7462	0.6381	0.8344
61.0000	0.5671	0.4550	0.6740
62.0000	0.5039	0.3935	0.6141
62.0000	0.5039	0.3935	0.6141
63.0000	0.4417	0.3343	0.5536
63.0000	0.4417	0.3343	0.5536
63.0000	0.4417	0.3343	0.5536
63.0000	0.4417	0.3343	0.5536
63.0000	0.4417	0.3343	0.5536
63.0000	0.4417	0.3343	0.5536
63.0000	0.4417	0.3343	0.5536
63.0000	0.4417	0.3343	0.5536
64.0000	0.3818	0.2786	0.4942
64.0000	0.3818	0.2786	0.4942
64.0000	0.3818	0.2786	0.4942
64.0000	0.3818	0.2786	0.4942
64.0000	0.3818	0.2786	0.4942
64.0000	0.3818	0.2786	0.4942
64.0000	0.3818	0.2786	0.4942
65.0000	0.3254	0.2277	0.4370
66.0000	0.2736	0.1824	0.3830
69.0000	0.1498	0.0831	0.2451
69.0000	0.1498	0.0831	0.2451
70.0000	0.1194	0.0615	0.2080
70.0000	0.1194	0.0615	0.2080
71.0000	0.0940	0.0446	0.1751
73.0000	0.0561	0.0222	0.1215
73.0000	0.0561	0.0222	0.1215
73.0000	0.0561	0.0222	0.1215
74.0000	0.0426	0.0153	0.1002
74.0000	0.0426	0.0153	0.1002
74.0000	0.0426	0.0153	0.1002

Goodness of Fit

Distribution	Anderson-Darling (adj)
Weibull	1.286
Lognormal base e	0.858
Exponential	18.159
Logistic	0.938

Table of Percentiles

Distribution	Percent	Percentile	Standard Error	95% Normal CI Lower	Upper
Weibull	1	42.0374	2.27783	37.8018	46.7476
Lognormal base e	1	48.9333	1.37496	46.3112	51.7037
Exponential	1	0.6270	0.08957	0.4739	0.8296
Logistic	1	44.8989	2.22756	40.5330	49.2649
Weibull	5	49.1433	1.87251	45.6069	52.9538
Lognormal base e	5	52.4619	1.17454	50.2096	54.8152
Exponential	5	3.2001	0.45715	2.4186	4.2341
Logistic	5	51.1149	1.59632	47.9861	54.2436
Weibull	10	52.6525	1.64620	49.5228	55.9799
Lognormal base e	10	54.4458	1.07245	52.3839	56.5889
Exponential	10	6.5732	0.93903	4.9679	8.6972
Logistic	10	53.9286	1.34433	51.2938	56.5634
Weibull	50	63.0685	1.01999	61.1007	65.0996
Lognormal base e	50	62.0627	0.90587	60.3124	63.8638
Exponential	50	43.2439	6.17770	32.6832	57.2170
Logistic	50	62.2026	0.94764	60.3453	64.0600

Table of MTTF

Distribution	Mean	Standard Error	95% Normal CI Lower	Upper
Weibull	62.2558	1.02913	60.2711	64.3059
Lognormal base e	62.3875	0.91299	60.6235	64.2028
Exponential	62.3878	8.91254	47.1519	82.5467
Logistic	62.2026	0.94764	60.3453	64.0600

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
time	Between Groups	1794.940	2	897.470	201.686	.000
	Within Groups	204.692	46	4.450		
	Total	1999.633	48			
height	Between Groups	156.164	2	78.082	1.367	.265
	Within Groups	2626.530	46	57.098		
	Total	2782.694	48			
yield	Between Groups	2090858.556	2	1045429.278	3.860	.028
	Within Groups	12458940.704	46	270846.537		
	Total	14549799.259	48			

Time

Duncan[a,b]

rep	N	Subset for alpha = 0.05 1	2	3
early	18	55.50		
normal	18		63.17	
late	13			70.85
Sig.		1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 15.955.
b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Yield			
Duncan[a,b]			
rep	N	Subset for alpha = 0.05	
		1	2
late	13	576.908	
normal	18	734.106	734.106
early	18		1077.117
Sig.		.398	.069

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 15.955.
b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Fig.1