A
PROJECT REPORT ENTITLED
ON

"Innovative data analysis of ANOVA models: An Empirical study".

A PROJECT REPORT SUBMITTED TO

DEPARTMENT OF STATISTICS
SHIVAJI UNIVERSITY
KOLHAPUR.

FOR THE PARTIAL FULFILLMENT OF THE DEGREE

M.Sc. –II (Statistics)
2021-22

PRESENTEDBY,
Mr. Kate Vijay Bhagawat
Mr. Basarwadkar Suraj Sudhakar
Mr. Kamble Vaibhav Bhairu

UNDER THE GUIDANCE OF,
Prof. Mrs. H.V. Kulkarni
Department of Statistics,
Shivaji University, Kolhapur.

CERTIFICATE

This is to certify that the project entitled, " **Innovative data analysis of ANOVA models: An Empirical study**". as partial fulfilment for the award of the degree of M.Sc. in Statistics of Shivaji University, Kolhapur, is a record of bonafide work carried out by them under my supervision and guidance. To the best of my knowledge, the matter presented in the project has not been submitted earlier.

This project is submitted by:

Mr. Kate Vijay Bhagawat

Mr. Basarwadkar Suraj Sudhakar

Mr. Kamble Vaibhav Bhairu

Place: **Kolhapur**
Date:

Prof. Dr. H. V. Kulkarni
Project Guide,
Department of Statistics,
Shivaji University, Kolhapur.

Prof. Dr. S. B. Mahadik
Head of Department,
Department of Statistics,
Shivaji University, Kolhapur.

ACKNOWLEDGEMENT

We wish to thank the Department of Statistics (Shivaji University, Kolhapur) for allowing us to do the project. This project report has been prepared under the guidance of Prof. Mrs. H.V. Kulkarni and Mr. S.M. Patil. We would like to express our profound gratitude towards them for their guidance and timely support throughout the completion of this project.

Also, we would like to thank Prof. Dr. S. B. Mahadik (HOD, Dept of Statistics), Dr. S. D. Pawar, Mr. S. V. Rajguru, and Mr. S. K. Ganjave Sir for their support, suggestions, and guidance for this project.

We would like to thank all my friends, M.Sc. part-1 students, research students, Non-teaching staff of our department for their co-operation and help which we received from them throughout the work.

INDEX

INTRODUCTION

Parametric data analysis is crucial in any investigation where sample sizes are small to moderate. The choice of analytical technique for the inference problem under consideration should take into account the best fitted probability distribution for the underlying data. Although data obtained from most disciplines like health, education, engineering, chemical industries and social sciences research are often not normally distributed, there are very few studies dealing with parametric analysis under specific probability distributions for the particular statistical tests or tools that need to be applied for particular inferential problems to be addressed.

Problem:

Many of the statistical procedures including correlation, regression, t-tests, and ***analysis of variance***, under parametric set-up are based on the assumption that the data follows a normal distribution. However, ANOVA is an extensively applicable technique used in almost every discipline where data sets are quite likely to come from non-normal distributions. Furthermore, often the sample sizes are too small to use the asymptotic-theory-based non-parametric methods, emphasizing the need of parametric ANOVA procedures based on popular non-normal distributions arising in practice.

The classical ANOVA test for testing the equality of mean effects uses the F-test that is based on the assumption of normality and equality of variances of underlying populations. However as mentioned before, there can be ample real situations where these assumptions are not met. Consequently, the inferences based on classical ANOVA applied to these data sets may often be misleading. Kulkarni and Patil (2020,2021,2022) have developed refined parametric ANOVA test procedures based on the novel ILRT test (Integrated Likelihood Ratio Test) for several parametric

distributional set-ups. This test does not consider the assumption of equality of variances. These tests together with the use of recent model adequacy checking procedures can offer most reliable inference in a majority of non-normal situations. In this project we have used ILRT test procedure under several parametric distributions and model adequacy checks on a very large number of real data sets collected from various books and presented some selected insightful analyses of real data sets where the inference based on the classical F-test based ANOVA was quite misleading.

What is ILRT test?

Let L ($\boldsymbol{\psi}$**,** λ*)* be a likelihood function under consideration $\boldsymbol{\psi}$ being the vector of parameter of interest and $\lambda \in \Lambda$ the vector of nuisance parameter. An IL is of the form

$$\bar{L}(\psi) = \int_{\Lambda} L(\psi, \lambda)\ \Pi(\lambda/\psi)\ d\lambda,$$

Where, Π *is a* non-negative weight function on Λ making the above integral convergent for every fixed ψ.

$\bar{L}$ depends on the data and ψ, and can be used as standard likelihood function for all likelihood-based inference procedures.

The Integrated Likelihood Ratio Test (ILRT) eliminates the nuisance parameter by integrating them out.

Objectives:

1) To unfold the impact of the blind use of traditional F-Test on real life data sets.

2) In-depth parametric analysis of real-life data sets under factorial setup in various disciplines like engineering, textile, pharmaceutical, chemical department etc. under specific best fitted distributions.

3) Model adequacy checks under non-normal parametric analysis to support the validity of inferences based on the parametric distributions used for analysis.

Data Collection:

We collected 100 real life datasets where the underlying inference problem needs use of design of experiments (ANOVA).

1. From the 100 collected datasets we got 36 one-way ANOVA datasets
2. The data sets were checked for variance homogeneity and appropriateness of classical F-Test through various model adequacy procedures
3. Out of 36 datasets 9 data sets did not satisfy the assumption of normality or equal variance property.
4. Remaining 64 data sets are two-way ANOVA type data sets. Out of them 10 data sets did not satisfy the assumption of normality or equal variance property.

Statistical Software used:

i) MATLAB
ii) Minitab
iii) Microsoft-Excel

General Procedure:

1) First Employ classical F-Test on the data set under consideration and obtain residuals of F-Test.

2) Employ Bartlett Test of homoscedasticity and Anderson Darling Test of normality on residuals to check variance homogeneity and normality respectively.

3) When both assumptions are satisfied then F-test gives valid results.

4) Those data sets which do not satisfy homoscedasticity assumption, but satisfy normality assumption are analyzed through normal based ILRT Test, which takes care of the heteroscedasticity in the data.

5) Those which do not satisfy normality assumption are analyzed through ILRT based on i) Gamma distribution ii) Weibull distribution iii) Log normal distribution iv) Inverse Gaussian distribution v) Birnbaum Saunders (BS) distribution based ILRT Tests.

6) A best parametric model-based analysis is chosen, based on various model adequacy checks like quantile residuals, interval plots etc. appropriate for the specific parametric distribution under consideration and the AIC, BIC and MSE criteria.

Statistical Analysis:

A) One-way setup:

Dataset-1: Discipline: Civil Engineering

This is a real-life dataset in which experiment is performed to check the effect of rodding level on compressive strength of concrete. Compressive strength of concrete depends on many factors such as water cement ratio, cement strength, quality of concrete material. Test for compressive strength is carried out either on cube or cylinder. Experimenter wants to know whether the rodding level significantly affects compressive strength of concrete.

An article in the *ACI Materials Journal* (Vol. 84, 1987, pp. 213–216) describes several experiments investigating the rodding of concrete to remove entrapped air. A 3-inch × 6-inch cylinder was used, and the number of times this rod was used is the design variable. The resulting compressive strength of the concrete specimen is the response. The data are shown in the following table:

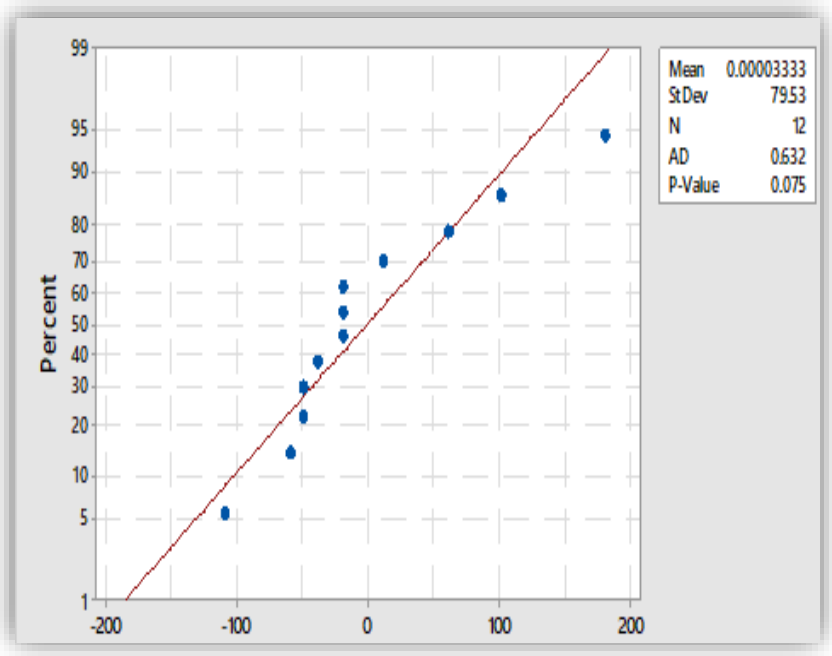
Rodding Level	Compressive Strength		
10	1530	1530	1440
15	1610	1650	1500
20	1560	1730	1530
25	1500	1490	1510

Application of classical F-Test:

i) P-value for normality of F Test based residuals = **0.0755** **(marginally indicates normality of underlying data)**

ii) P-value for equality of variance Test= **0.115**(**marginally indicates homogeneity of underlying data)**

iii) Normal Probability Plot of Residuals:



Residuals under F-test based on normal

Result based on classical F-test:

P-value for equality of Rodding level performances = **0.0002**

There is significant effect of rodding level on compressive strength of concrete.

But from P-value of normality of residuals .07 is almost on the boundary of the critical region so the result given by F-Test is likely to be invalid. we employed the aforementioned Step 5 which yielded the following results:

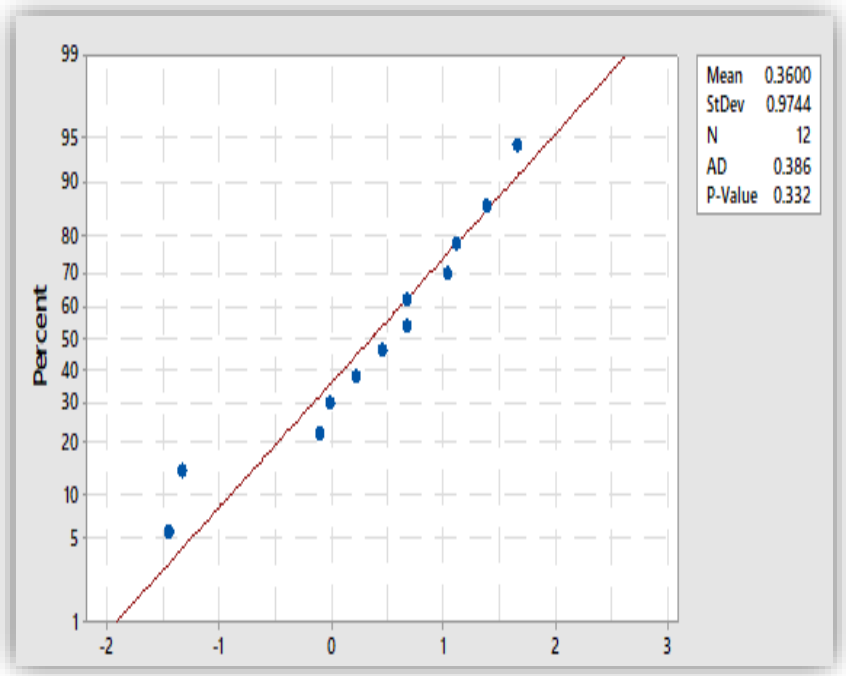
Distribution/Test	P-value for testing effect	P-value for Normality	AIC	BIC	MSE
F TEST	0.0002	0.0755	142	143	5797.2
ILRT NORMAL	0.2862	0.0755	146.183	148.6075	5797.2
ILRT BS	0.2926	0.3451	130.9184	133.3429	3411.1
ILRT GAMMA	0.2951	0.0949	153.1443	155.5688	7748.3
ILRT WEIBUL	0.3702	0	322.395	324.8195	8875.8
ILRT LOGNORMAL	0.2777	0.1431	378.4735	379.4433	Inf
ILRT IG	0.299	0.3621	137.2367	139.6613	8042.10

Clearly, Based on all criteria, the BS distribution gives the best results.

Other details of Application of Birnbaum-Saunders distribution based ILRT:

i) P-value for normality on Residuals obtained from BS based ILRT is **0.3451** which is sufficiently greater than 0.05.

ii) Quantile Residual plot:



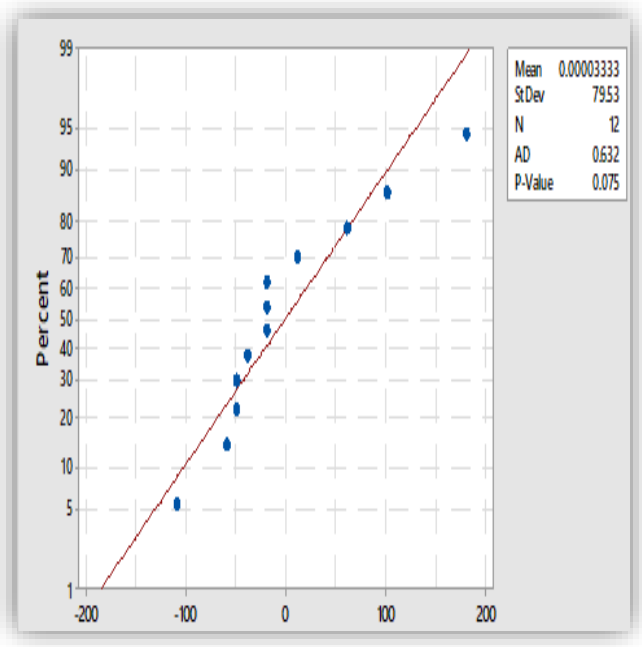
Quantile residuals under ILRT based on BS

iii) Inference based on BS-ANOVA:

P-value for testing the Rodding level effect is **0.2926** So BS-ILRT **strongly accept the null hypothesis i.e there is no significant effect of rodding level on compressive strength of concrete.**

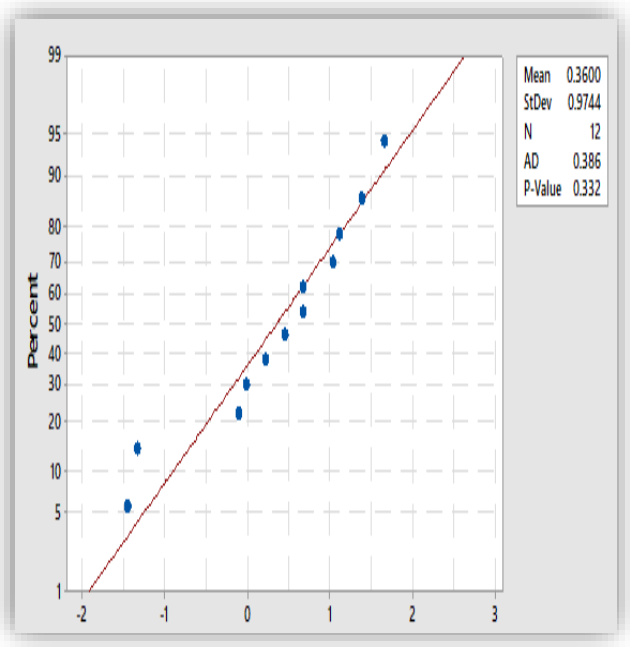
Note that this conclusion is exactly and strongly opposite to the one based on mis-used F-test ANOVA

Normal Probability Plots:



Residuals under F-test based on normal

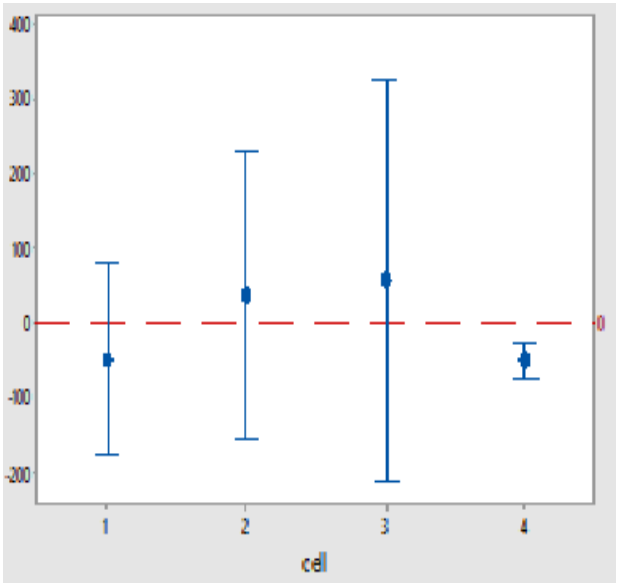
p-value=0.075



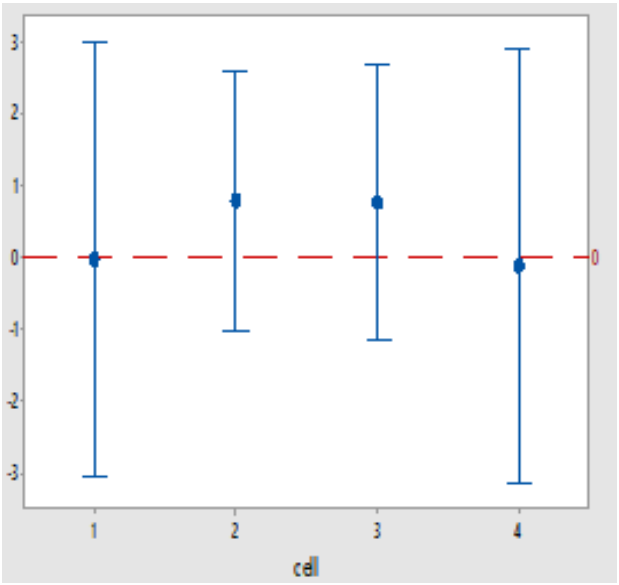
Quantile residuals under ILRT based on BS

p-value=0.332

95% confidence interval plots of the mean of residuals for each treatment:



Residuals under F-test based on normal



Quantile residuals under ILRT based on BS

Each individual interval covers zero under the BS model goes in the favor of the decision based on BS-ILRT. Note that the similar interval plot for F-based residuals indicates heterogeneous variances for the residuals and the last cell does not cover zero indicating that its mean is different from zero, these two things going against the decision given by classical F-test.

Conclusion:

So from model adequacy measures BS distribution is best fit for given dataset therefore result given by BS-ILRT is more valid where decision based on F-Test not valid. This is totally and strongly accepted in favor of BS based ILRT. That is **there is no significant effect of rodding level on compressive strength of concrete, conclusion exactly opposite to that based the classical F-test.**

Data set 2: Discipline: Electrical Engineering

This is real life dataset of experiment in which experimenter wants investigate effect of four bleaching chemicals on pulp brightness (diffusive reflectance of blue light from pad of pulp sheets)

An article in the *Journal of Quality Technology* (Vol. 13, No. 2, 1981, pp. 111–114) describes an experiment that investigates the effects of four bleaching chemicals on pulp brightness. These four chemicals were selected at random from a large population of potential bleaching agents. The data are as follows:

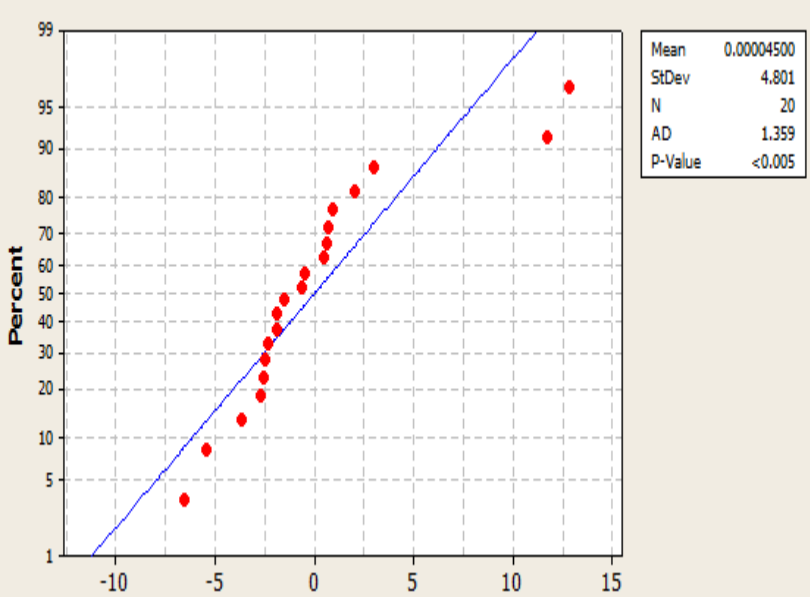
Chemical	Pulp Brightness				
1	77.199	74.466	92.746	76.208	82.876
2	80.522	79.306	81.914	80.346	73.385
3	79.417	78.017	91.596	80.802	80.626
4	78.001	78.358	77.544	77.364	77.386

Application of classical F-Test:

i) P-value for normality of F-test based residuals **=0.0011** **(strongly indicates non-normality of underlying data)**

ii) P-value for equality of variance Test= **0.001**(**strongly indicates heteroscedasticity of underlying data)**

iii) Normal Probability Plot of Residuals:



Residuals under F-test based on normal

Result based on classical F-test:

P-value for equality of effect of four bleaching chemicals = 0.5383 There is no significant effect of four bleaching chemicals on pulp brightness.

But from P-value of normality of residuals 0.0011 and P-value for equality of variances 0.001 is not satisfied so the result given by F-Test is likely to be invalid. We employed the aforementioned Step 5 which yielded the following results:

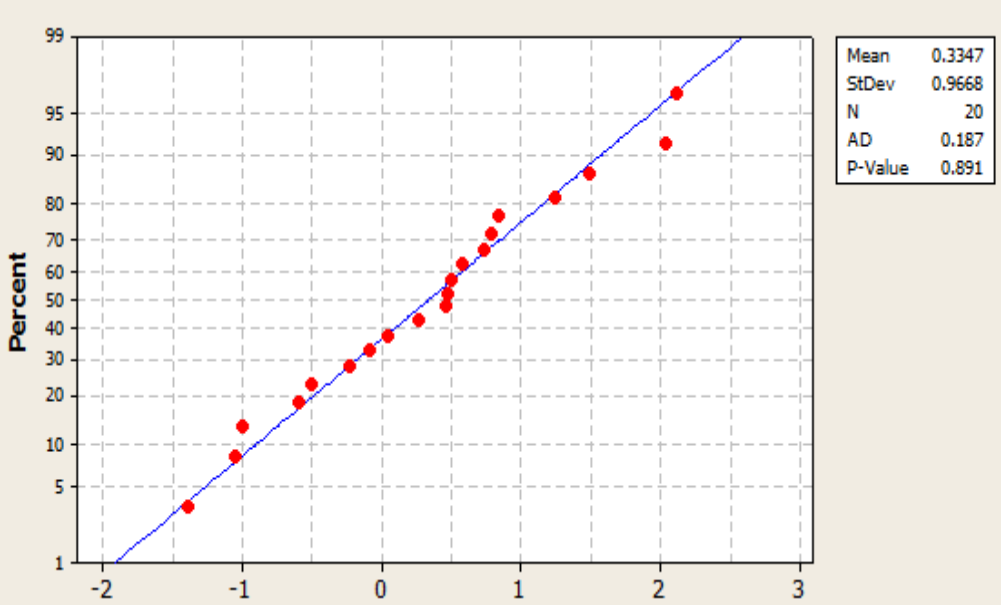
Distribution/Test	P-value for testing effect	P-value for Normality	AIC	BIC	MSE
F TEST	0.5383	0.0011	122.4860	124.4775	21.8986
ILRT NORMAL	0.2958	0.0011	121.6747	126.6533	21.8986
GAMMA ILRT	0.3119	0.0023	132.3243	137.3030	25.7366
WEIBULL ILRT	0.3075	0.0005	163.9161	168.8948	29.7684
BS ILRT	0.3098	0.9042	102.4057	107.3843	19.1995
LOGNORMAL ILRT	0.2965	0.0042	394.9599	399.9385	Inf
IG ILRT	0.2620	0.2115	107.6892	112.6678	26.4192

Clearly, Based on all criteria, the only BS distribution gives the best results.

Other details of Application of Birnbaum-Saunders distribution based ILRT:

i) P-value for normality on Residuals obtained from BS based ILRT is **0.9042** which is sufficiently greater than 0.05.

ii) Quantile Residual Plot:

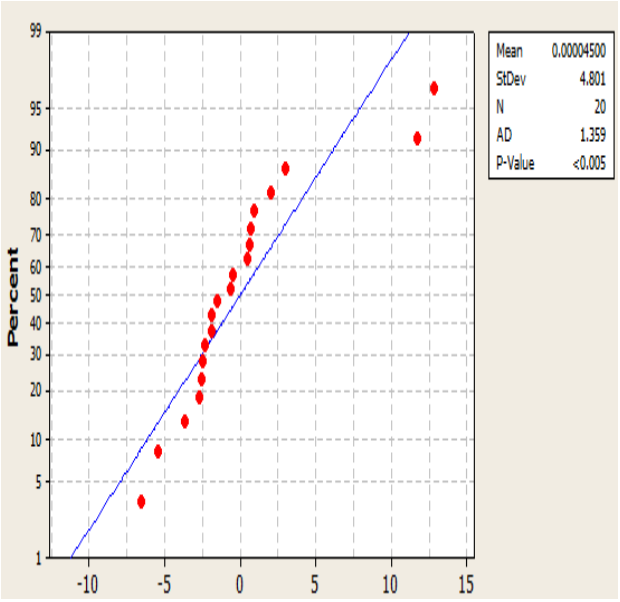


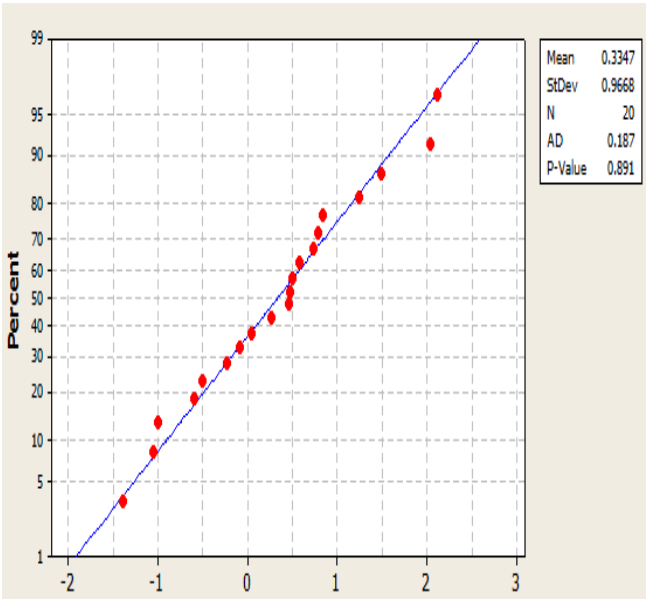
Quantile residuals under ILRT based on BS

iii) **Inference based on BS-ANOVA:**

P-value for testing effect of four chemicals is 0.3098 So BS-ILRT **strongly accept the null hypothesis i.e. there is no significant effect of four bleaching chemicals on pulp brightness.**

Normal Probability Plots:





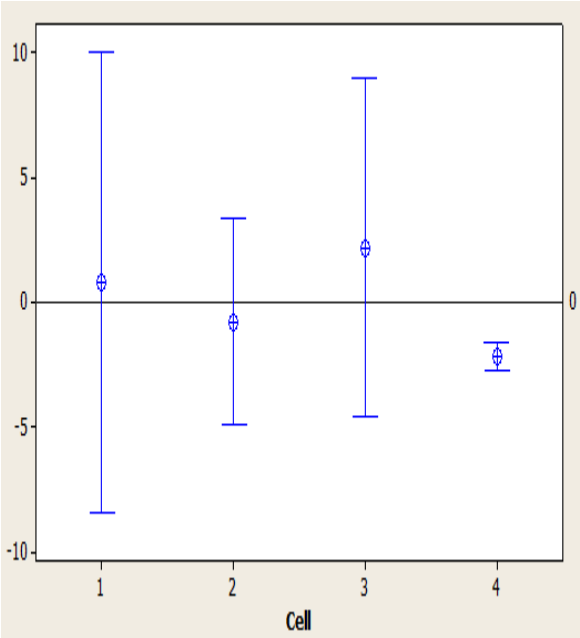
Residuals under F-test based on normal

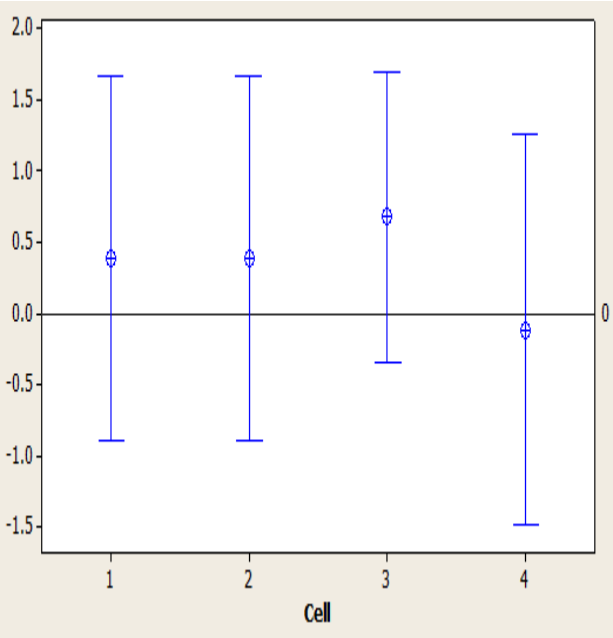
P-value=0.0000

Quantile residuals under ILRT based on BS

P-value=0.891

95% confidence interval plots of the mean of residuals for each treatment:





Residuals under F-test based on normal

Quantile residuals under ILRT based on BS

Each individual interval covers zero under the BS model goes in the favor of the decision based on BS-ILRT. Note that the similar interval plot for F-based residuals indicates heterogeneous variances for the residuals and the last cell does not cover zero indicating that its mean is different from zero, these two things going against the decision given by classical F-test.

Conclusion:

From model adequacy measures BS distribution is best fit for given dataset therefore result given by BS-ILRT is more valid where decision based on F-Test not valid. This is totally and strongly accepted in favor of BS based ILRT. That is **there is no significant effect of four bleaching chemicals on pulp brightness.**

B) Two Way Setup:

Data set 1): discipline: Mechanical Engineering

This is real life data set in which experimenter wants to know that effect of cyclic loading (Factor A) and environmental conditions (Factor B) on fatigue crack growth at a constant 22 MPa stress for a particular material. Fatigue is a failure type when material is subjected to repeated cycles of stress or strain and occurs even at lower stress than the material yield stress.

An article in the *Journal of Testing and Evaluation* (Vol. 16, no. 2, pp. 508–515) investigated the effects of cyclic loading and environmental conditions on fatigue crack growth at a constant 22 MPa stress for a particular material. The data from this experiment are shown below (the response is crack growth rate):

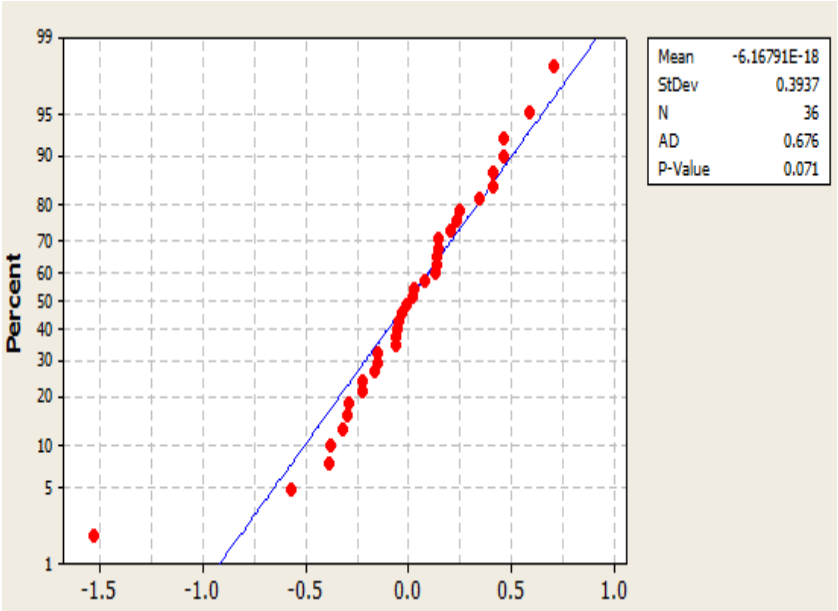
	Environment		
Frequency	Air	H_2O	Salt H_2O
10	2.29	2.06	1.90
	2.47	2.05	1.93
	2.48	2.23	1.75
	2.12	2.03	2.06
1	2.65	3.20	3.10
	2.68	3.18	3.24
	2.06	3.96	3.98
	2.38	3.64	3.24
0.1	2.24	11.00	9.96
	2.71	11.00	10.01
	2.81	9.06	9.36
	2.08	11.30	10.40

Application of classical F-Test:

1) P-value for normality of F-Test based residuals = **0.0707**(**marginally indicates normality of underlying data)**

2) P-value for equality of variance Test= **0.006** **(strongly indicates heteroscedasticity of underlying data)**

3) Normal Probability plot of residuals:



Residuals under F-test based on normal

Result:

Tests	P-value for test of effects			P-value for normality of residuals	AIC	BIC	MSE
	A	B	AB				
Conventional F-Test	0.0000	0.0000	0.0000	0.0707	68.0277	94.9476	0.1507

Cyclic loading, environment conditions and simultaneously both have significant effect on fatigue crack growth of particular material.

But from P-value of normality of residuals 0.07 is almost on the boundary of critical region and P-value for equality of variances does not satisfy so the result given by F-Test is likely to be invalid. we employed the aforementioned Step 5 which yielded the following results:

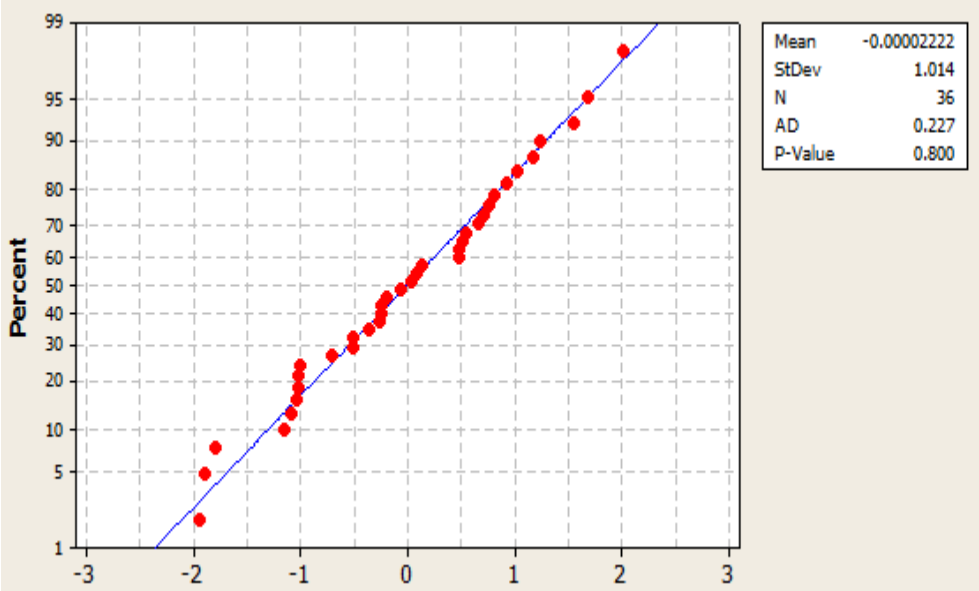
Tests	P-value for test of effects			P-value for normality of residuals	AIC	BIC	MSE
	A	B	AB				
Conventional F-Test	0.0000	0.0000	0.0000	0.0707	68.0277	94.9476	0.1507
Normal ILRT	0.3628	0.0493	0.0000	0.3062	52.1824	91.7704	0.1507
BS ILRT	0.0000	0.0508	0.0266	0.0446	67.8049	107.3928	19.9450
Gamma ILRT	0.3854	0.0556	0.0000	0.8054	46.5624	73.4823	0.1507
Weibull ILRT	0.4583	0.0409	0.0108	0.0005	202.1463	229.0661	6.2755
Lognormal ILRT	0.3883	0.0677	0.0000	0.3109	52.3131	91.9010	0.1507
IG ILRT	0.0000	0.0399	0.0533	0.1693	52.2869	91.8749	0.1507

Clearly, Based on all criteria, the Gamma distribution gives the best results.

Other details of Application of Gamma distribution based ILRT:

i) P-value for normality based on residuals obtained from gamma based ILRT is **0.8054** which is sufficiently greater than 0.05.

Quantile Residual Plot:



Quantile residuals under ILRT based on Gamma distribution.

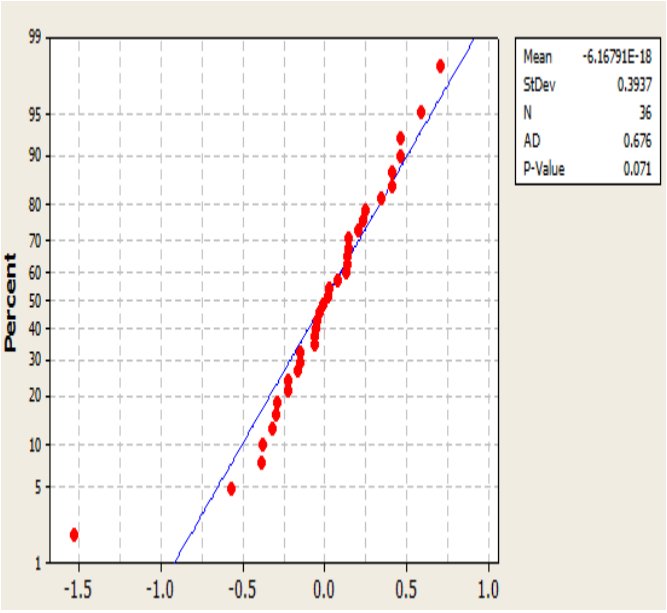
ii)Inference based on Gamma ANOVA:

i) P-value for testing effect of factor A (cyclic loading frequency) is **0.3854**, for factor B (Environmental conditions) is **0.0556** and for interaction effect of cyclic loading frequency and Environmental conditions (AB) is 0.

ii) Gamma ILRT strongly indicating that there is no significant effect of factor A (cyclic loading frequency) and factor B (Environmental conditions) independently.

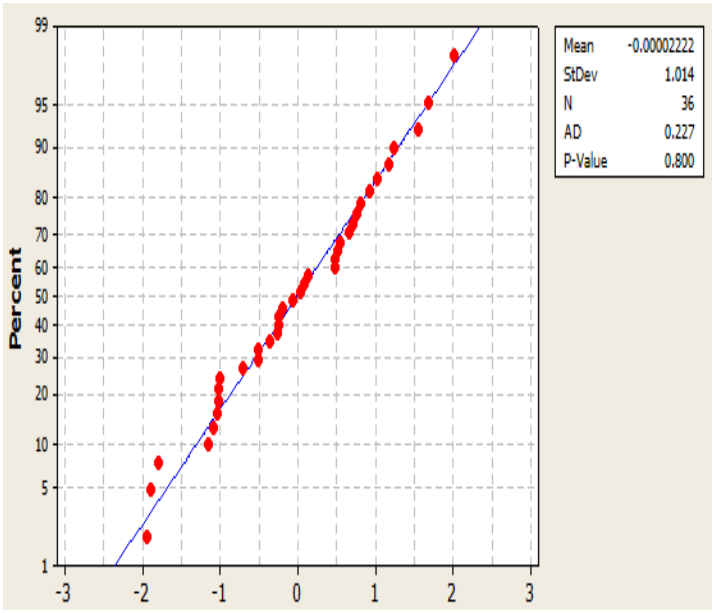
Note that this conclusion is exactly and strongly opposite to the one based on mis-used F-test ANOVA

Normal Probability plots:



Residuals under F-test based on normal

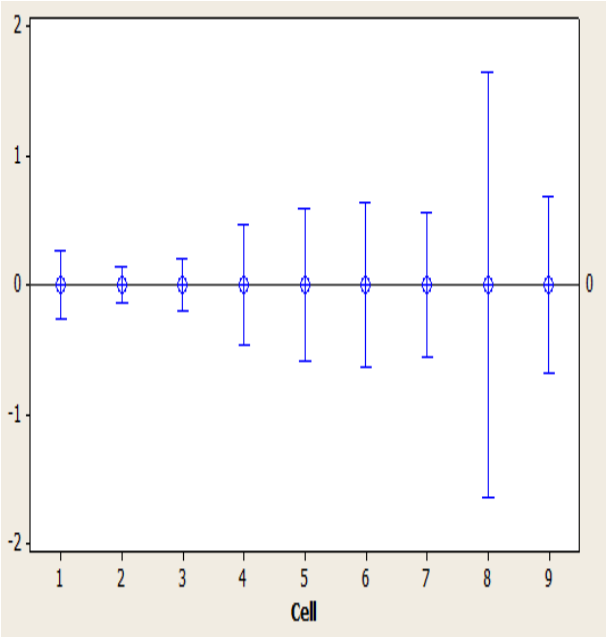
P-value=0.071



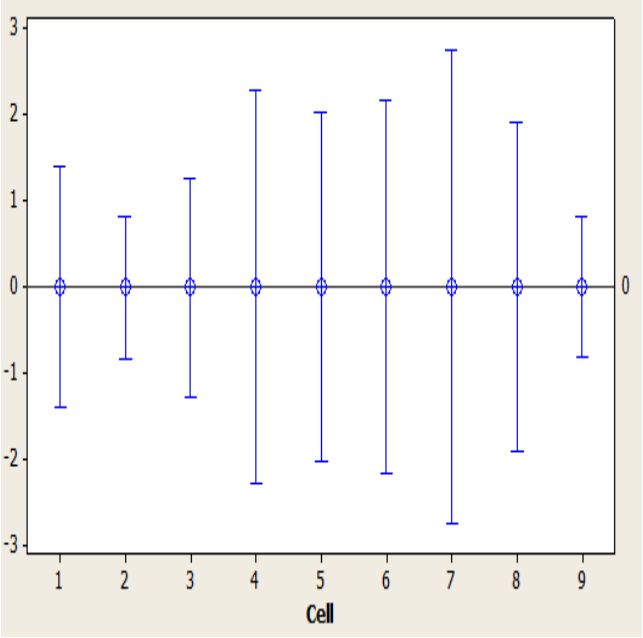
Quantile residuals under ILRT based on Gamma

P-value=0.800

95% confidence interval plots of the mean of residuals for each treatment combination:



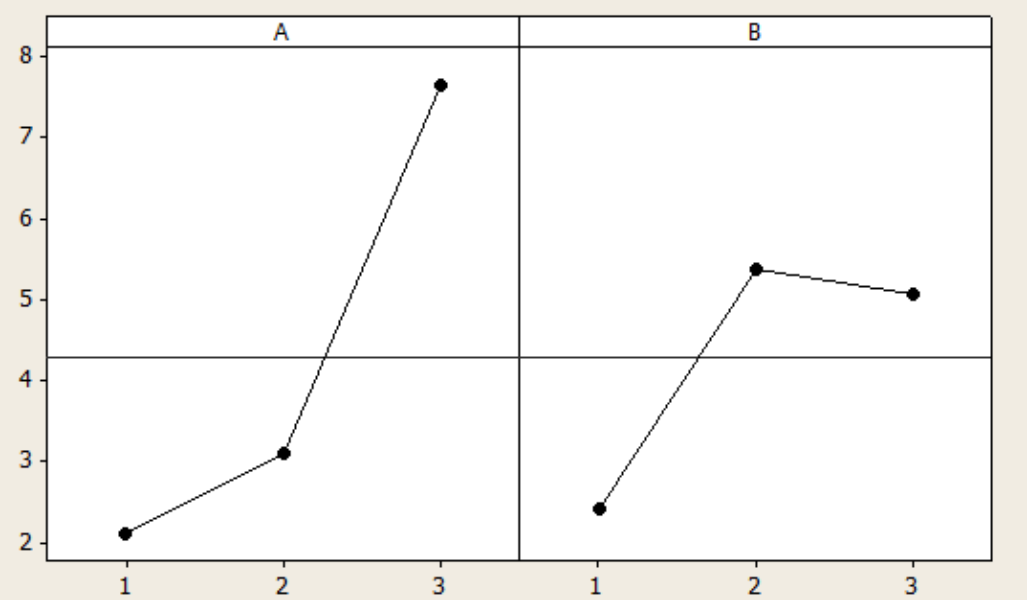
Residuals under F-Test based on normal



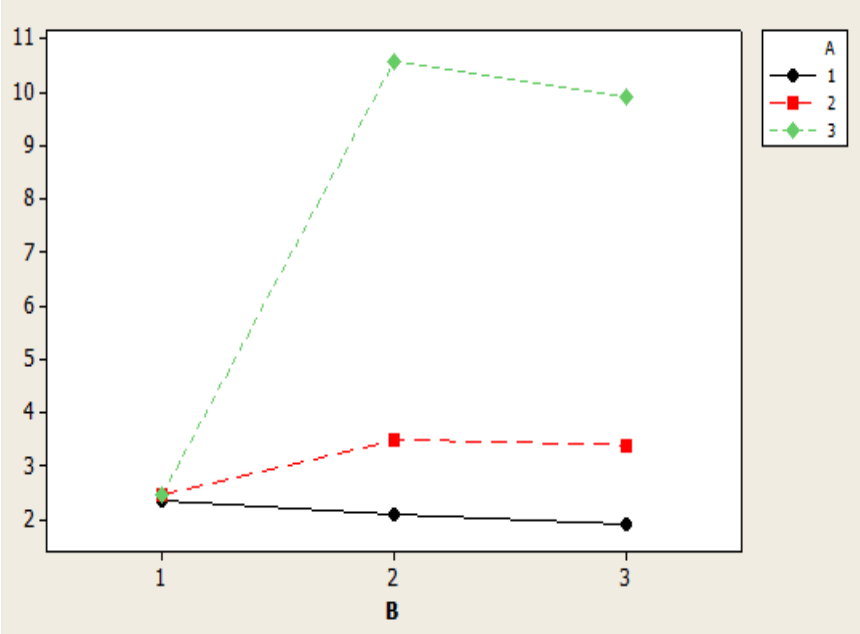
Quantile residuals under ILRT based on Gamma

Form Interval plot of Quantile residuals under ILRT based on Gamma, for each individual treatment combination have almost uniform length goes under favor of the decision based on Gamma-ILRT. Note that similar interval plot for F-based indicating heterogenous variances for the residuals, this thing going against the decision given by classical F-test.

Main effect plots:



Interaction effect plot:



Conclusion:

Based on all model adequacy measures Gamma distribution is best fit for given dataset therefore result given by Gamma-ILRT is more valid where decision based on F-Test not valid. This is totally and strongly accepted in favor of Gamma based ILRT. that there is no significant effect of factor A (cyclic loading frequency) and factor B (Environmental conditions) independently, however there is strong interaction between the two factors.

Data set 2): Discipline: Clinical Trails

This is real life data set in which experimenter wants to know that effect of two products placebo and allergenic product and two positions ear and back on skin in a skin sensitivity study.

6 A skin sensitivity study was conducted using two products, one a placebo (control) and the other a potentially allergenic product (treatment). Three animals were exposed to the control product and three to the treatment product. Each animal was administered the products on the ear and on the back. Duplicate tests were run. The data are tabulated below. Analyze these data by constructing an ANOVA table, including the expected mean squares. Perform mean comparisons and estimate variance components where appropriate.

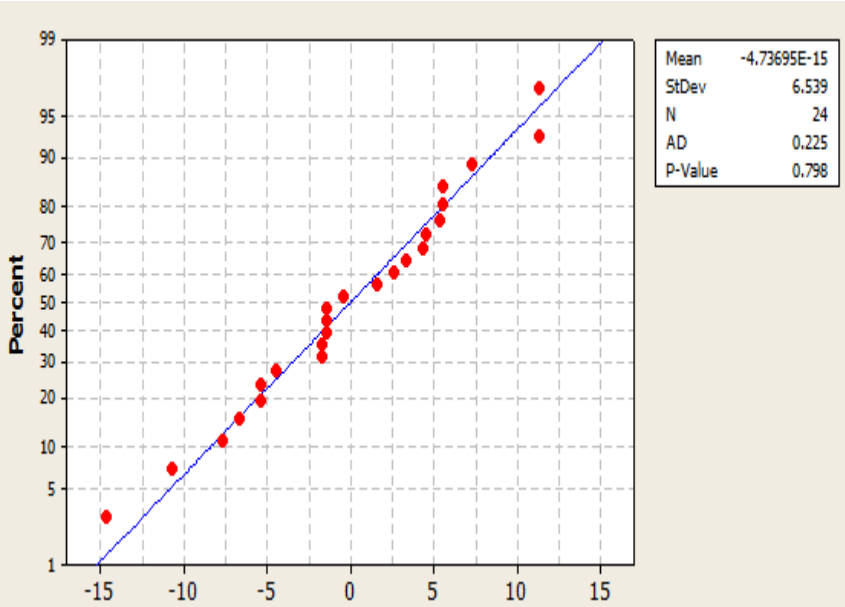
		Measurements	
Group	**Animal**	**Ear**	**Back**
Control	1	73, 70	90, 92
	2	66, 66	83, 83
	3	71, 70	77, 74
Treatment	4	77, 74	89, 88
	5	67, 70	70, 78
	6	76, 77	96, 96

Application of classical F-Test:

1) P-value for normality of F-Test based residuals = **0.8073** **(Strongly indicates normality of underlying data)**

2) P-value for equality of variance Test= **0.042** **(marginally indicates homoscedasticity of underlying data)**

3) Normal Probability Plot of Residuals:



Residuals under F-test based on normal

Result from classical F-test:

Tests	P-value for test of effects			P-value for normality of residuals	AIC	BIC	MSE
	A	B	AB				
Conventional F-Test	0.2070	0.0000	0.8340	0.8073	165.2246	169.9368	40.9826

The position of body of animal (factor B) significantly contributes in the skin sensitivity study

But from P-value of homogeneity of variances = **0.042** is almost on the boundary of the critical region the result given by F-Test is likely to be invalid. we employed the aforementioned Step 5 which yielded the following results.

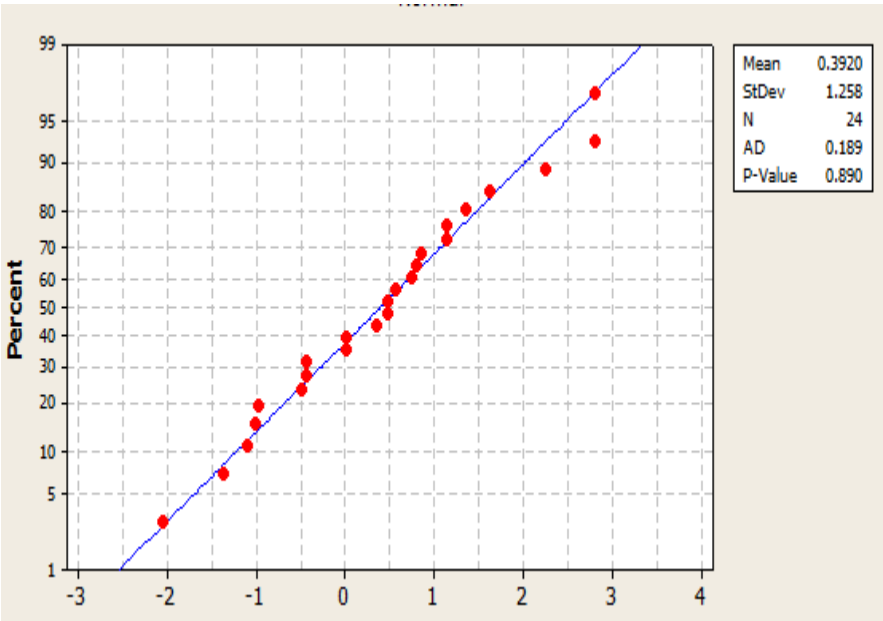
Tests	**P-value for test of effects**			**P-value for normality of residuals**	**AIC**	**BIC**	**MSE**
	A	**B**	**AB**				
Conventional F-Test	**0.2070**	**0.0000**	**0.8340**	**0.8073**	**165.2246**	**169.9368**	**40.9826**
N-ILRT	**0.0615**	**0.0003**	**0.8319**	**0.0699**	**163.0079**	**171.2543**	**41.7502**
BS-ILRT	**0.0853**	**0.0006**	**0.7994**	**0.0481**	**163.4910**	**171.7373**	**41.2630**
G-ILRT	**0.0679**	**0.0003**	**0.7374**	**0.6648**	**163.7650**	**168.4772**	**41.6339**
W-ILRT	**0.0477**	**0.0011**	**1.0000**	**0.9007**	**168.5166**	**175.5849**	**40.3612**
Lognormal ILRT	**0.0704**	**0.0003**	**0.7073**	**0.0480**	**163.5080**	**171.7544**	**41.2600**
IG ILRT	**0.0640**	**0.0002**	**0.8438**	**0.0788**	**163.7298**	**171.9762**	**41.5685**

Clearly, based on p-value of normality and MSE, the Weibull distribution gives the best results.

Other details of Application of Weibull distribution based ILRT:

i) P-value for normality on residuals obtained from Weibull-based ILRT is 0.9007 which is sufficiently greater than 0.05

Quantile Residual Plot:



Quantile residuals under ILRT based on Weibull distribution

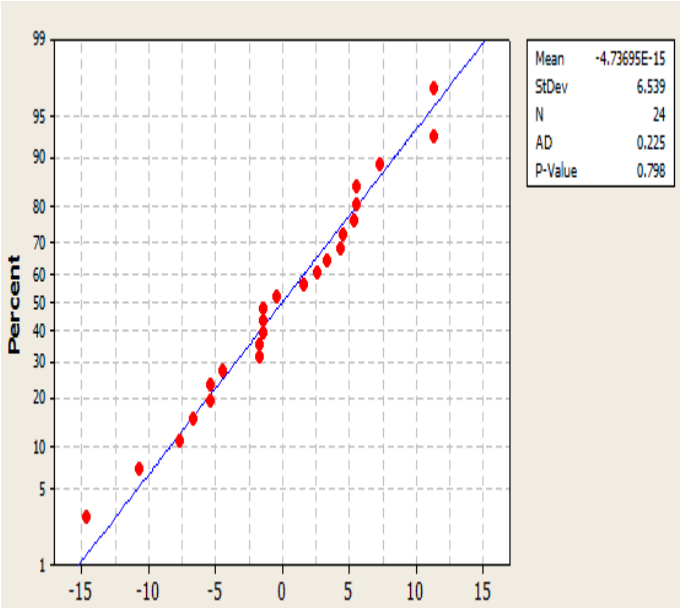
ii)Inference Based on Weibull ANOVA:

i) P-value for testing effect of factor A (treatment groups placebo and allergenic product) is **0.04**, for factor B (position of body) is **0.001** and interaction effect AB is **1**.

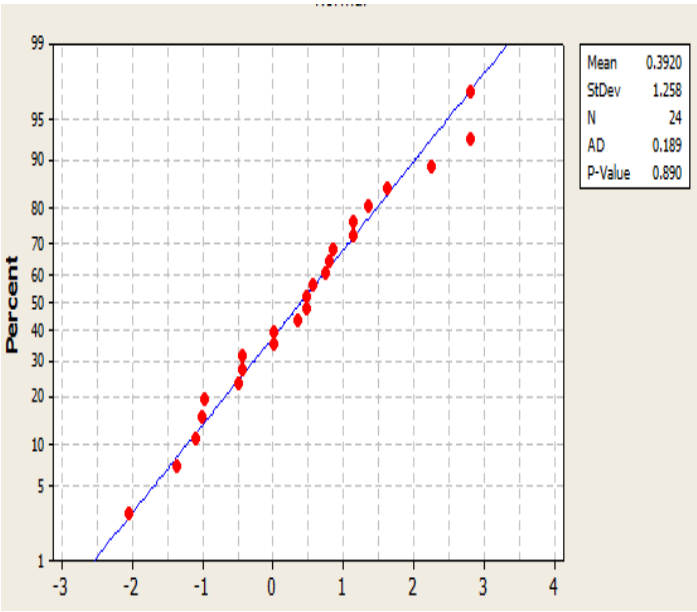
ii) Weibull ILRT saying that there is significant effect of factor A (treatment groups placebo and allergenic product) and factor B (position of body Ear and Back) independently on skin.

Note that this conclusion is exactly and strongly opposite to the one based on mis-used F-test ANOVA

Normal Probability plots:

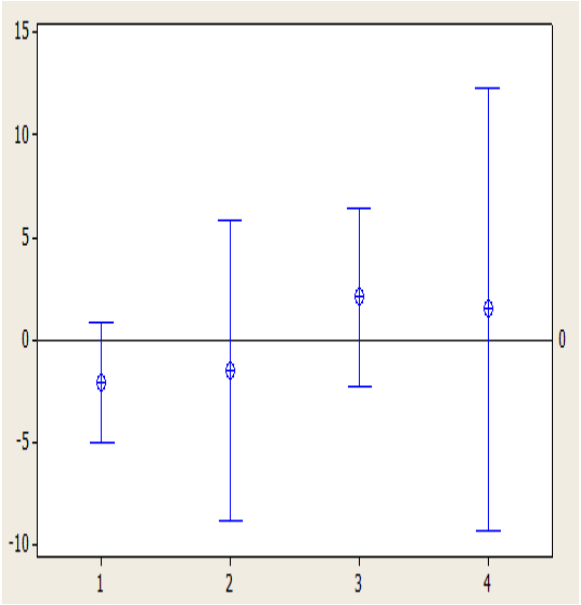


Residuals under F-test based on normal
P-value=0.789

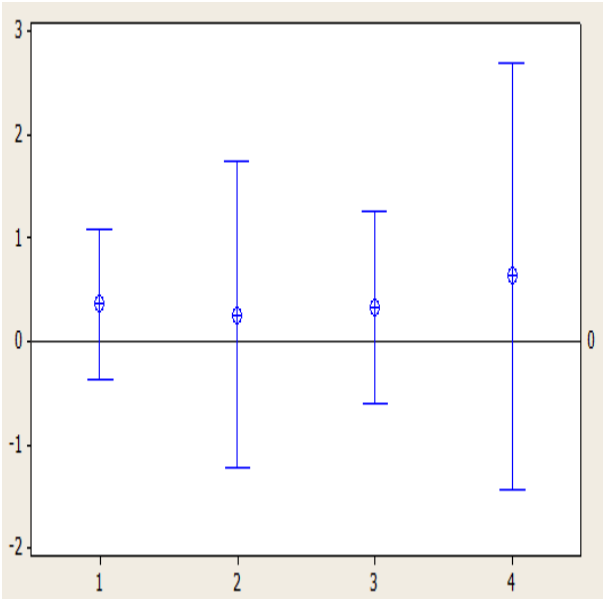


Quantile residuals under ILRT based on Gamma
P-value=0.890

95% confidence interval plots of the mean of residuals for each treatment combination:



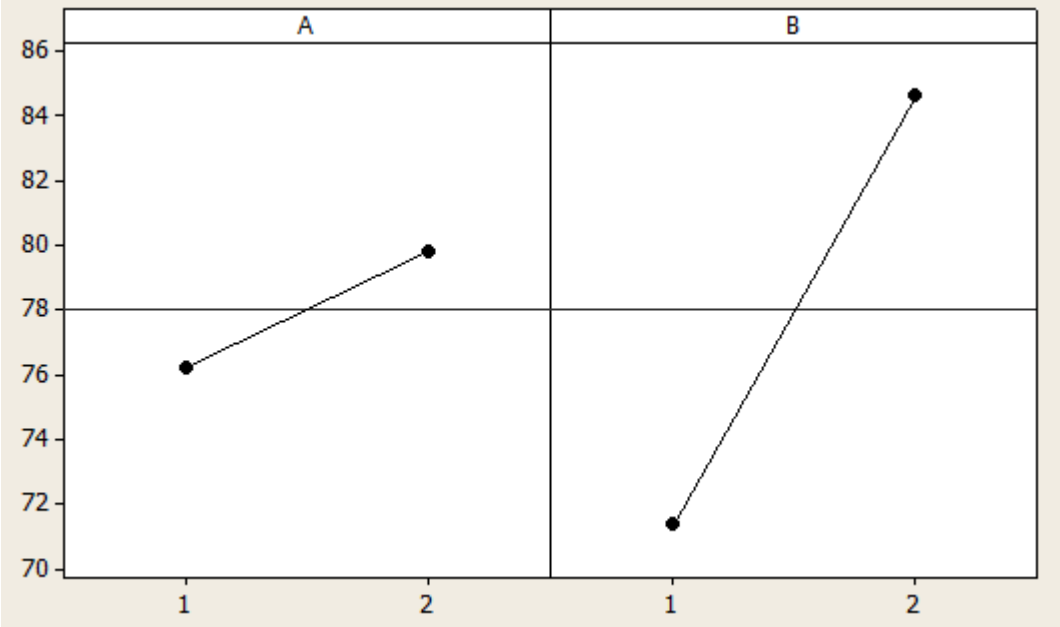
Residuals under F-Test based on normal



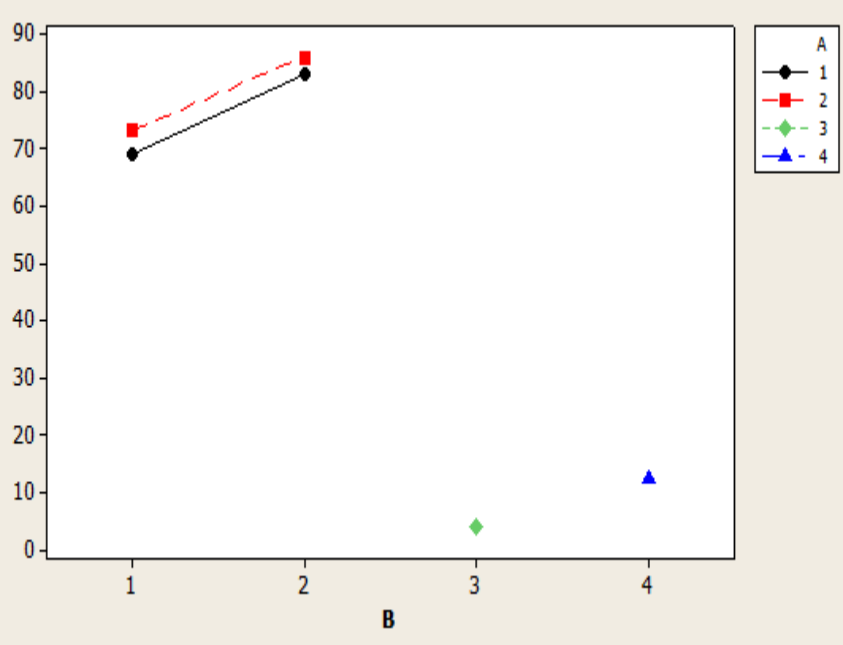
Quantile residuals under ILRT based on Weibull

Form Interval plot of Quantile residuals under ILRT based on Weibull, for each individual treatment combination have almost uniform length goes under favor of decision based on Weibull ILRT. Note that similar interval plot for F-based residuals indicating heterogeneity in variances for the residuals, this thing going against the decision given by classical F-test.

Main effect plots of factor A (treatment groups placebo and allergenic product) and factor B (position of body):



Interaction plot:



Conclusion:

From model adequacy measures Weibull distribution is best fit for given dataset therefore result given by Weibull-ILRT is more valid where decision based on F-Test not valid. This is totally and strongly accepted in favor of Weibull based ILRT. **that there is significant effect of factor A (treatment groups placebo and allergenic product) and factor B (position of body Ear and Back) independently.**

Dataset 3): Discipline: Aerospace Engineering.

This is real life data set in which experimenter wants to know that two batches of ammonium perchlorate and two types of catalysts significantly affects the propellant burning rate.

Problem 1.44
In a pilot-plant for producing composite rocket propellants 12 batches were produced, by which seven experimental rocket motors were cast and cured for each batch. By static firing of these motors at 25 °C burning-rate laws as a function of pressure were obtained. From such a law a rate value at 70bar pressure was obtained. The propellant batches were mixed with two batches of ammonium perchlorate and two types of catalysts. Besides, each batch was repeated three times. The burning-rate data at 25 °C and P=70bar for all mixing conditions of the propellants are given in the table below. By applying the two-way analysis of variance with replication, determine with 95% confidence level whether the catalyst type and preparation of 10 μm of ammonium perchlorate essentially affect the propellant burning rate at 25 °C and 70bar?

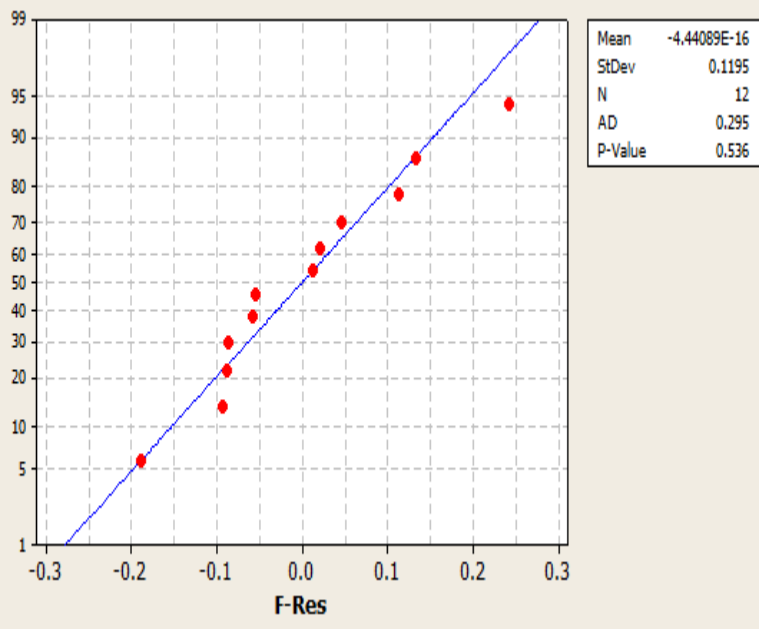
	Catalyst A			Catalyst B		
Batch AP (A)	14.199	14.197	14.193	15.716	15.612	15.682
Batch AP (B)	14.398	14.418	14.307	15.616	15.912	15.482

Application of classical F-Test:

i) P-value for normality of F-Test based residuals = **0.5724** **(Strongly indicates normality of underlying data)**

ii) P-value for equality of variance Test= **0.002** **(Strongly indicates heteroscedasticity of underlying data)**

iii) Normal Probability Plot of Residuals:



Residuals under F-test based on normal

Result from classical F-test:

Tests	P-value for test of effects			P-value for normality of residuals	AIC	BIC	MSE
	A	B	AB				
Conventional F-Test	0.2240	0.0000	0.2240	0.5724	-9.9830	-8.0434	0.0131

The catalysts type (factor B) significantly affects the propellant burning rate.

But from P-value of homogeneity of variances = **0.002** which strongly says that there is heteroscedasticity in data so the result given by F-Test is likely to be invalid. we employed the aforementioned Step 5 which yielded the following results.

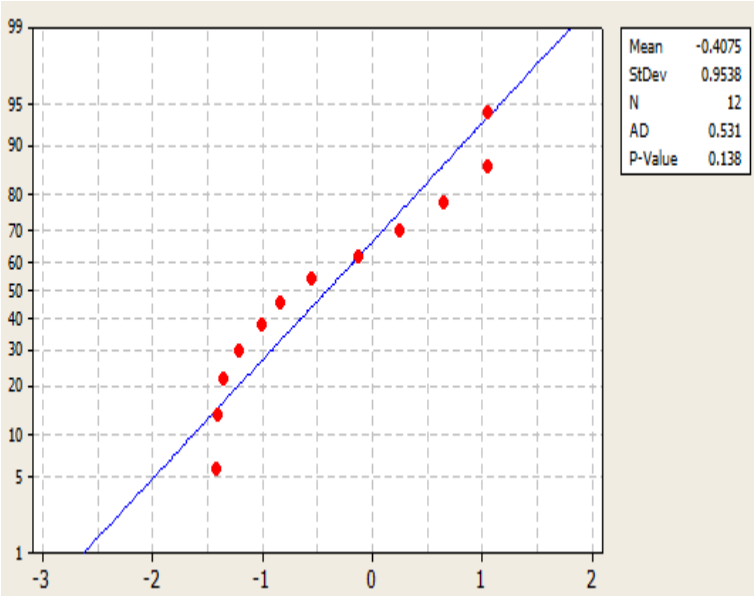
Tests	**P-value for test of effects**			**P-value for normality of residuals**	**AIC**	**BIC**	**MSE**
	A	**B**	**AB**				
Conventional F-Test	**0.2240**	**0.0000**	**0.2240**	**0.5724**	**-9.9830**	**-8.0434**	**0.0131**
ILRT Normal	**0.0493**	**0.0000**	**0.2538**	**0.1429**	**-21.3138**	**-16.949**	**0.0321**
BS-ILRT	**0.0638**	**0.0000**	**1.0000**	**0.0652**	**-27.1676**	**-23.7733**	**0.0169**
GAMMA -ILRT	**0.0533**	**0.0000**	**0.2284**	**0.1883**	**-2.8689**	**-0.9293**	**0.0170**
WEIBUL	**1.0000**	**1.0000**	**1.0000**	**0.0708**	**16.8581**	**21.7072**	**0.0092**
LN-ILRT	**0.0547**	**1.66E-06**	**0.2204**	**0.0642**	**-27.1674**	**-23.7731**	**0.0169**
IG	**1**	**1**	**0.9629**	**0.0005**	**9.6707**	**10.1556**	**0.9685**

Clearly, based on p-value of normality, AIC and BIC the ILRT Normal distribution based test gives the best results.

Other details of Application of ILRT Normal distribution based ILRT:

i) P-value for normality on residuals obtained from Normal-based ILRT is **0.1429** which is sufficiently greater than 0.05

Quantile Residual Plot:



Quantile residuals under ILRT based on Normal distribution

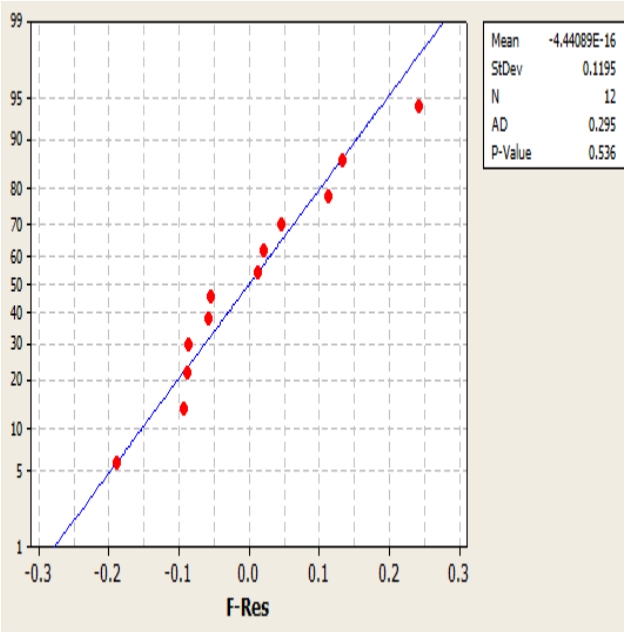
ii)Inference Based on Normal ILRT based ANOVA:

ii) P-value for testing effect of factor A (ammonium perchlorate) is **0.0493**, for factor B (catalyst type) is **0.000** and interaction effect AB is **0.2538**.

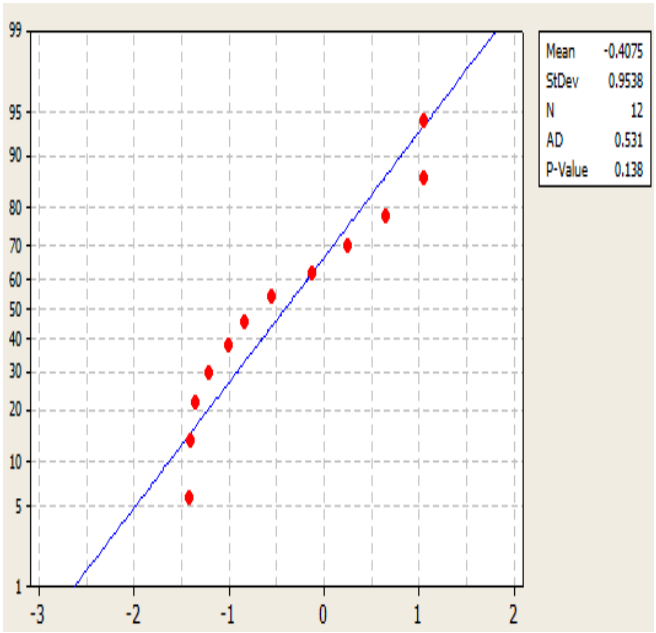
iii) Normal ILRT saying that there is significant effect of factor A (batches of ammonium perchlorate) and factor B (catalyst type) independently on propellant burning at constant temperature and pressure.

Note that this conclusion is exactly and opposite to the one based on mis-used F-test ANOVA

Normal Probability plots:

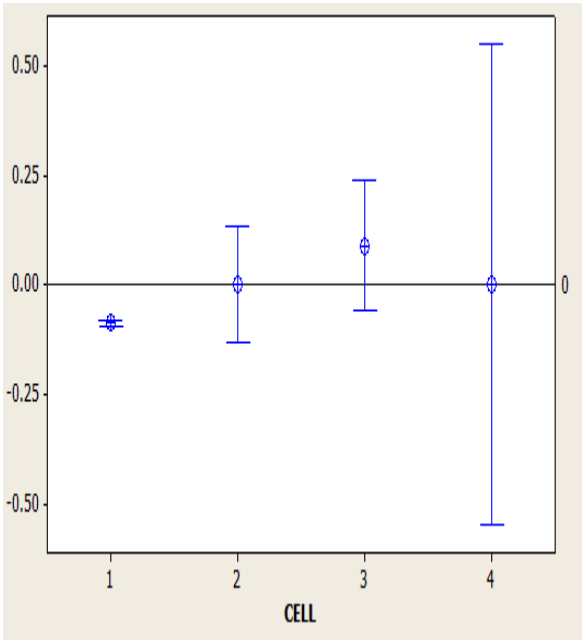


Residuals under F-test based on normal
P-value=0.536

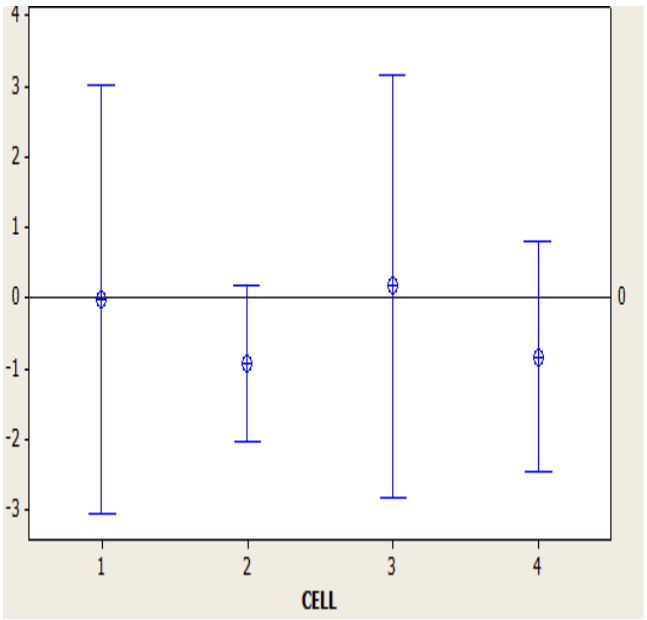


Quantile residuals under ILRT based on Gamma
P-value=0.138

95% confidence interval plots of the mean of residuals for each treatment combination:



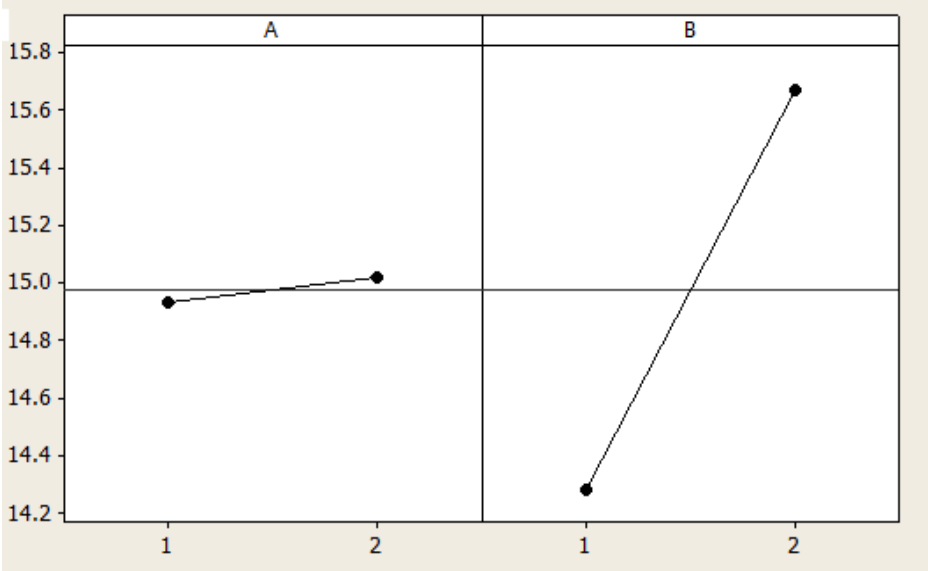
Residuals under F-Test based on Normal



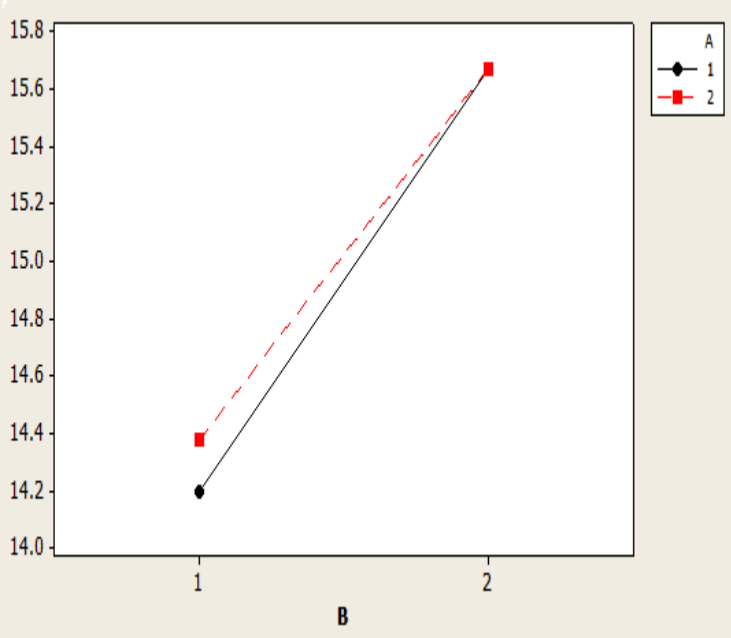
Quantile residuals under ILRT based on Normal

Each individual interval covers zero under the ILRT normal model goes in the favor of the decision based on Normal-ILRT. Note that the similar interval plot for F-based residuals indicates heterogeneous variances for the residuals and the first cell does not cover zero indicating that its mean is different from zero, these two things going against the decision given by classical F-test.

Main effect plots:



Interaction plot:



Conclusion:

From model adequacy measures Normal distribution is best fit for given dataset therefore result given by Normal-ILRT is more valid where decision based on F-Test not valid. This is totally and accepted in favor of Normal based ILRT. **that there is significant effect of factor A (batches of ammonium perchlorate) and factor B (Catalyst type) independently.**

Major Findings:

1) From collected data sets we observed that **almost 25% data sets** were did not satisfied normality and homogeneity of variances of residuals.
2) In some situations (data sets mentioned above), ILRT completely changed the result given by classical F-test.
3) Another non-normal distribution (BS, Gamma, Weibull, IG, Lognormal) based ILRT gives satisfactory results.
4) There is much need of ILRT in industry for statistical analysis to draw the inference about the particular problem.

References:

1) Analysis of medians under two-way model with and without interaction for Birnbaum–Saunders distributed response
SM Patil, HV Kulkarni - Journal of Applied Statistics, 2022 - Taylor & Francis

2) Douglas C. Montgomery - Design and Analysis of Experiments (2001, John Wiley)

3)Wikipedia