

Intro to Data Science

PROJECT

CODE LINK : [R CODE](#)
DATASHEET : [G-SHEET](#)

Group Members

Raghav Vijayvergia (B21ES017)
Aditya Anand (B21ES003)
Mayank Srivastava (B21BB019)
Praveen Kumar (B21CI032)
Mamta Kumari (B21CI025)

Objective: Analyzing various factors and their impact on LAPTOP Battery life .

Tools Used:

- RStudio
- Battery Analytics Application
- Stopwatch
- Hp Laptop

Factors Considered:

A = Refresh Rate
B = Brightness
C = RESOLUTION
D = SPEAKER VOL.

A .Data Collection :

- **Dataset with different Parameters:**

	A-Refresh Rate		B-Brightness		C-RESOLUTION		D-SPEAKER.		RATE
	60Hz(Low)	144Hz(High)	0 (Low)	100 (High)	900x600p	1920x1080p	NO(MUTE)	YES(Full)	(%DISCHARGE/Hrs.)
1									31.08
a									39.84
b									37.68

ab					59.68
c					31.32
ac					37.32
bc					35.4
abc					56.72
d					34.2
ad					43.08
bd					38.52
abd					61.36
cd					32.76
acd					39
bcd					37.08
abcd					60.04

INDICATORS :

LOW LEVEL	HIGH LEVEL
-----------	------------

B .Data Preprocessing & Transformation :

1 .OUR COLLECTED DATA / WITH LEVEL

```
> data.rate
      A  B  C  D  Rate
(1) -1 -1 -1 -1 31.08
a      1 -1 -1 -1 39.84
b     -1  1 -1 -1 37.68
ab      1  1 -1 -1 59.68
c     -1 -1  1 -1 31.32
ac      1 -1  1 -1 37.32
bc     -1  1  1 -1 35.40
abc      1  1  1 -1 56.72
d     -1 -1 -1  1 34.20
ad      1 -1 -1  1 43.08
bd     -1  1 -1  1 38.52
abd      1  1 -1  1 61.36
cd     -1 -1  1  1 32.76
acd      1 -1  1  1 39.00
bcd     -1  1  1  1 37.08
abcd      1  1  1  1 60.04
```

2. DESIGN MATRIX GENERATED

> Design.matrix

	I	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD	Rate
(1)	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	31.08
a	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1	1	1	-1	-1	39.84
b	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1	37.68
ab	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	59.68
c	1	-1	-1	1	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	31.32
ac	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1	37.32
bc	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	35.40
abc	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	56.72
d	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	34.20
ad	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	43.08
bd	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	38.52
abd	1	1	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	61.36
cd	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	32.76
acd	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	39.00
bcd	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	37.08
abcd	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	60.04

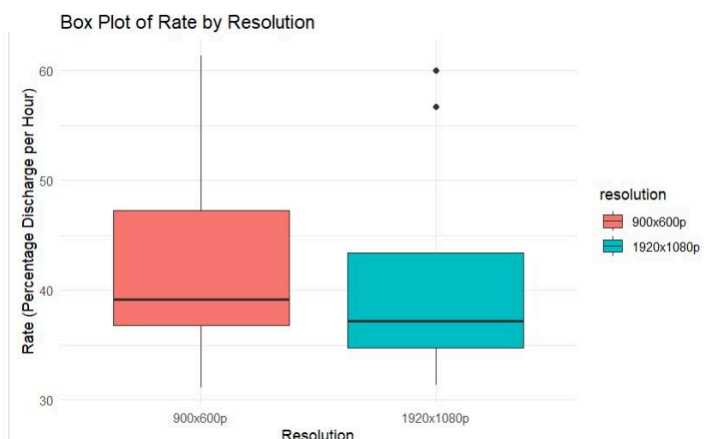
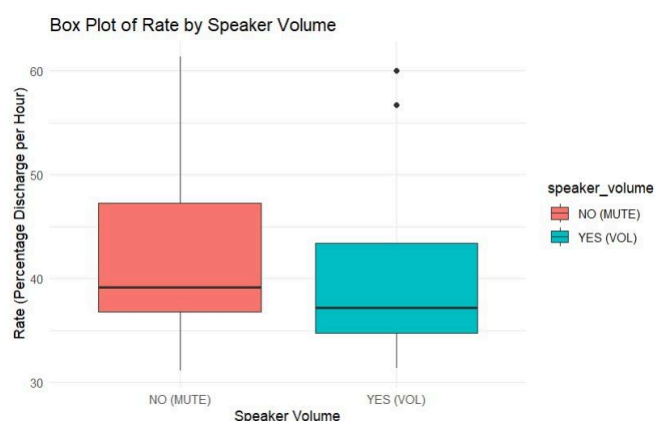
3. FACTOR ESTIMATES

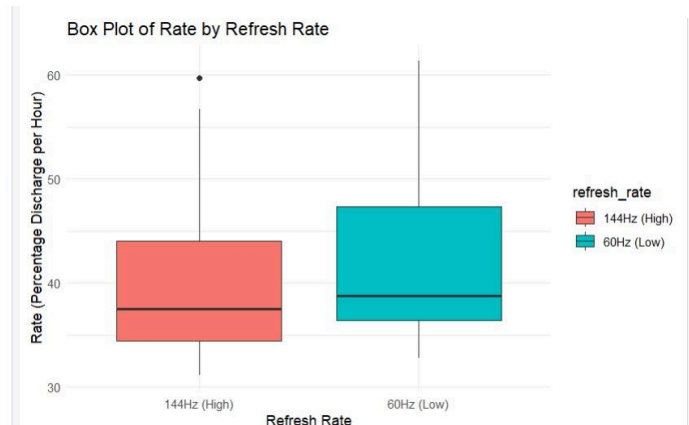
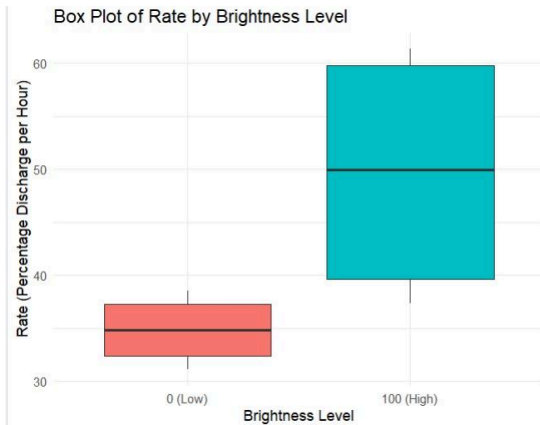
	I	A	B	AB	C	AC	BC		
Rate	42.1925	14.875	12.235	7.405	-1.975	-0.745	-0.025		
	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
	0.605	2.125	0.355	-0.245	0.265	-0.095	0.115	0.715	0.085

>> FROM ABOVE RESULT WE CAN SEE that : WE HAVE BOTH +VE AND -VE EFFECT OF FACTORS ON RATE OF DISCHARGE

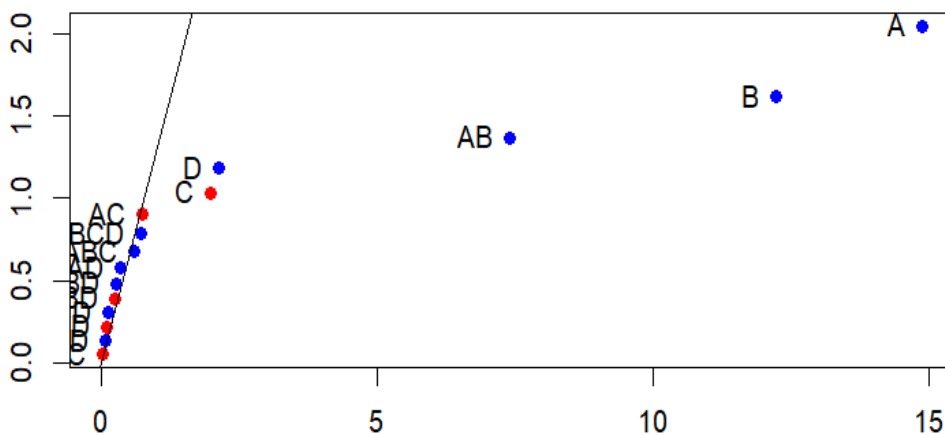
C .Data Exploration and Analysis :

1.Box plots





2. HALF NORMAL PLOTS



>>PLOT ANALYSIS

- The **red points** in the plot show that the factors have **Negative effects**.
 - The **blue points** in the plot show that the factors have **Positive effects**.
 - The factors far apart from the line are more significant.
 - Factor **C(Resolution)** has a Negative **effect Surprising**. Means better resolution is good for battery life. May be due to inbuilt internal optimization
 - From the obtained plot, we get that the main factor effects **A and B** are highly **significant**. Others C & D are less significant
 - Also, The Interaction Effect AB is also far from the line to be considered as significant Here .
-

D. REGRESSION MODEL :

```
> lm.rate=lm(Factors ~ A*B, data=data.rate)
> summary(lm.rate)
```

Call:

```
lm.default(formula = Factors ~ A * B, data = data.rate)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.73	-1.08	0.13	0.78	3.27

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	42.1925	0.4595	91.822	< 2e-16	***
A	7.4375	0.4595	16.186	1.62e-09	***
B	6.1175	0.4595	13.313	1.51e-08	***
A:B	3.7025	0.4595	8.058	3.49e-06	***

— — —

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.838 on 12 degrees of freedom

Multiple R-squared: 0.9768, Adjusted R-squared: 0.9709

F-statistic: 168.1 on 3 and 12 DF, p-value: 4.585e-10

>>Result ANALYSIS : Estimated discharge rate :

As : The regression model is given by

$$y_p = \beta_0 + \beta_1 x_1 + \beta_2 x_3 + \beta_3 x_4 + \beta_4 x_1 x_3 + \beta_5 x_1 x_4$$

In this case

$$y_p = 42.1925 + 7.4375 x_1 + 6.1175 x_2 + 3.7025 x_1 x_3$$

Here x1, x2 are levels of A and B respectively for each exp.

>Residuals are the differences between the observed values and the predicted values from the regression model that represent the errors or unexplained variance

>>e(x) = yactual - ypredicted

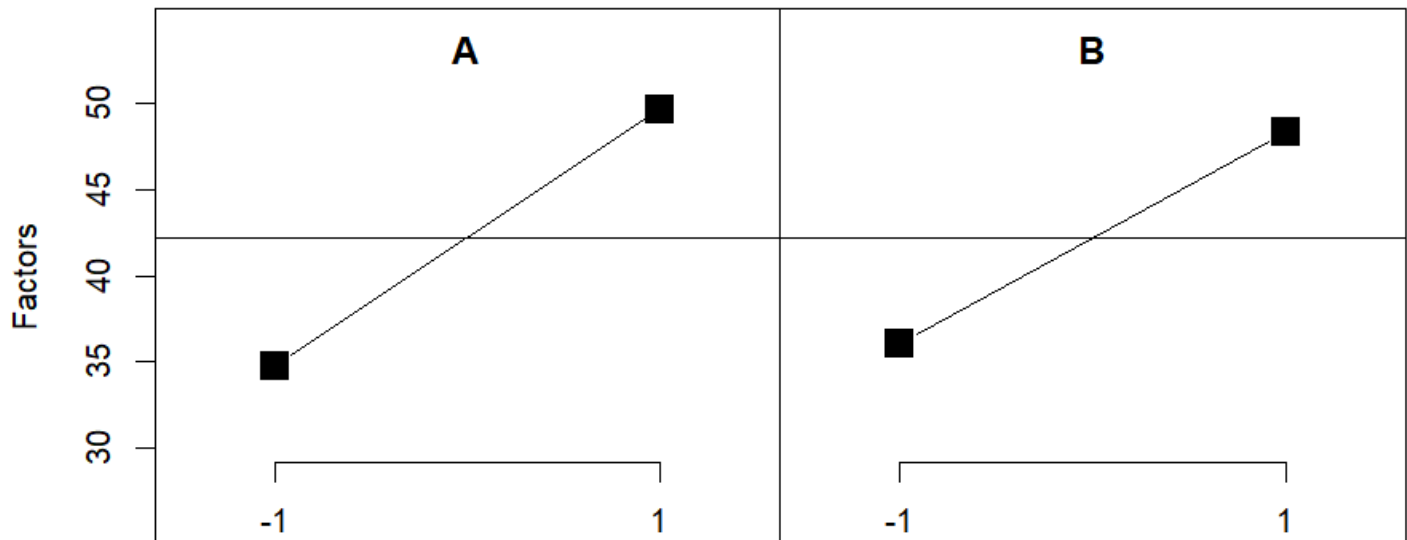
```
> residuals=mod$res
```

```
> residuals
```

[illegible]

2. MAIN EFFECTS PLOTS

Main effects plot for Factors

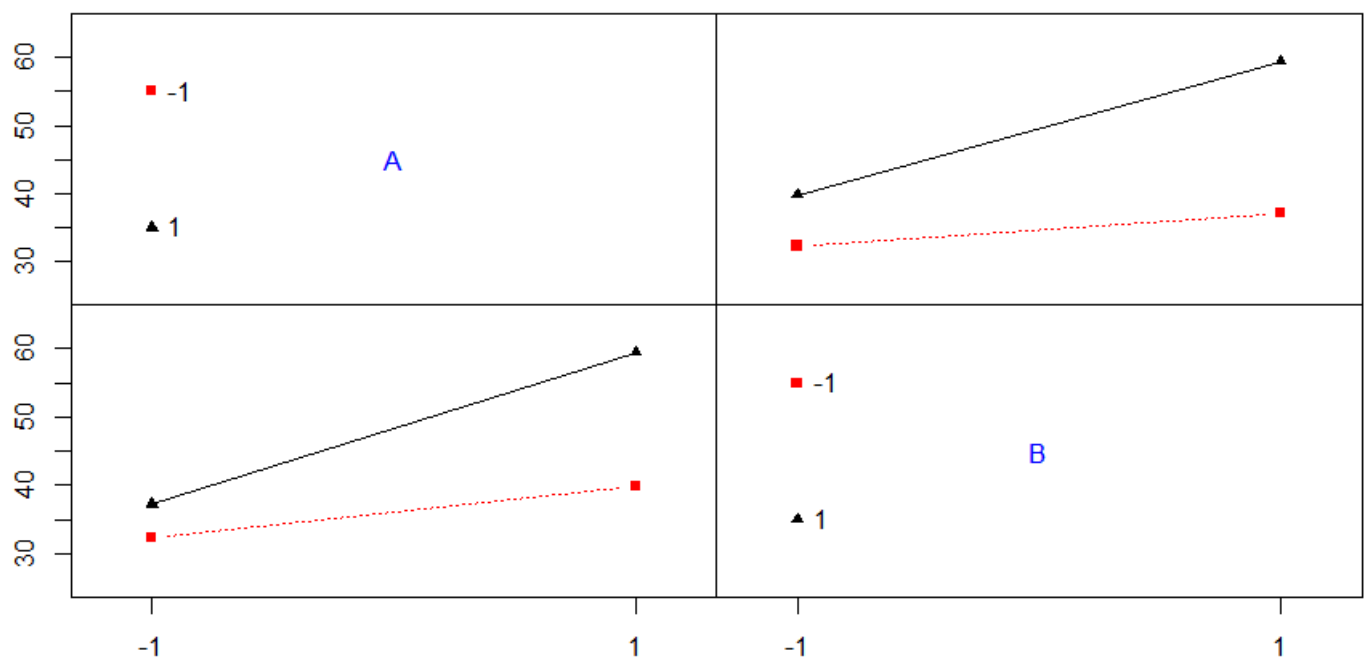


>>PLOT ANALYSIS

- The plot signifies that for the low level of Refresh Rate (A), the rate of discharging is less, and high for the high level.
 - Also , at the low level of Brightness (B), the rate of discharging is less, and high for the high level.
-

3. INTERACTION EFFECT PLOTS

Interaction plot matrix for Factors



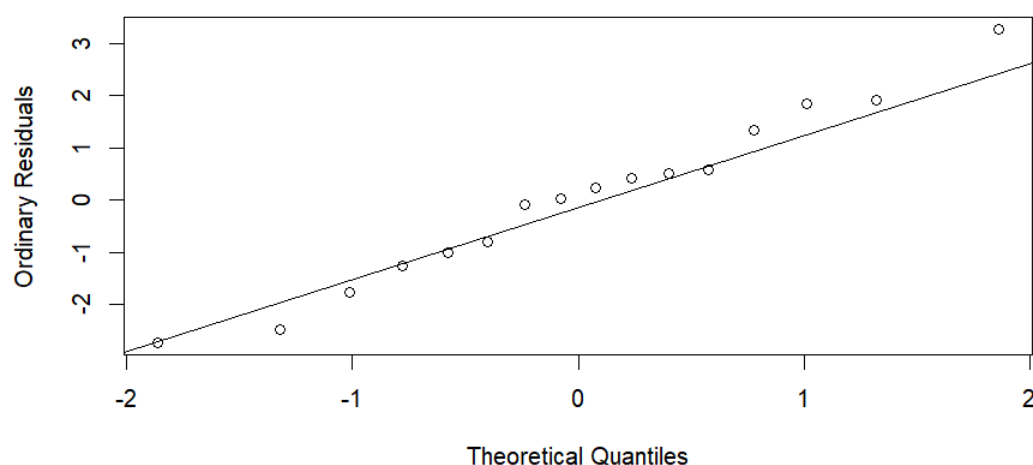
>>PLOT ANALYSIS

- In GRAPHS, the **black** line signifies the **high level** of factor A and B respectively . The **red** line signifies the **low level** of factor A and B respectively .
 - OBSERVATION says as in each block, as the lines are **not parallel to each other**, hence, the interactions between these factors A and B are significant.
-

4. Normal QQ Plot

Below , The residuals Exactly follow the normal distribution.

Normal Q-Q Plot



D. Projection ANOVA(by omitting factors C and D)

```
> Total <- apply(values, 1, sum)
> Total
      (1)      a      b      ab
129.36 159.24 148.68 237.80
> A <- rep(c(-1, +1), 2)
> B <- rep(c(-1, -1, +1, +1), 1)
> data.mat <- data.frame(A, B, as.vector(values))
> model <- aov(as.vector(values) ~ A*B, data = data.mat)
> summary(model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
A	1	885.1	885.1	261.99	1.62e-09	***
B	1	598.8	598.8	177.24	1.51e-08	***
A:B	1	219.3	219.3	64.92	3.49e-06	***
Residuals	12	40.5	3.4			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Above Analysis shows factors A and B are highly significant.

USING MORE MULTIPLE LEVEL FOR FACTOR A and B

HIGH, MID, LOW

Multiple levels of significant factors			
A-Refresh Rate	B-Brightness		
RATE	L1- 0	L2 -50	L3 -100
144	42.44	50.11	63.2
90	36.3	42.1	49.8
60	34.1	35.7	37.89

>>STRUCTURE

```
> str(data.rate)
'data.frame':  9 obs. of  3 variables:
 $ refresh_rate      : chr  "144Hz" "90Hz" "60Hz" "144Hz" ...
 $ brightness_level  : chr  "L1" "L1" "L1" "L2" ...
 $ rate              : num  42.4 36.3 34.1 50.1 42.1 ...

> data.rate
  refresh_rate brightness_level  rate
1      144Hz             L1 42.44
2       90Hz             L1 36.30
3       60Hz             L1 34.10
4      144Hz             L2 50.11
5       90Hz             L2 42.10
6       60Hz             L2 35.70
7      144Hz             L3 63.20
8       90Hz             L3 49.80
9       60Hz             L3 37.89
```

>>ANOVA

```
> rate.aov <- aov(rate ~ refresh_rate + brightness_level, data = data.rate)
> summary(rate.aov)
              Df Sum Sq Mean Sq F value Pr(>F)
refresh_rate    2  387.7   193.86  10.398 0.0260 *
brightness_level 2  244.8   122.39   6.565 0.0545 .
Residuals       4    74.6    18.64
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

FROM Above results we find that compared to Brightness (B) << Refresh Rate (A) is more significant.

E. Hypothesis Testing

1. LSD TEST

```
$groups
      rate groups
63.2  63.20      a
50.11 50.11     ab
49.8   49.80     ab
42.44 42.44      b
42.1   42.10      b
37.89 37.89      b
36.3   36.30      b
35.7   35.70      b
34.1   34.10      b
```

Obtained the rates into groups
are as follows into a,b & ab

Where alpha = 0.05

2. Tukey's HSD TEST

```
$refresh_rate
      diff      lwr      upr      p adj
90Hz-60Hz  6.836667 -5.728199 19.40153 0.2424679
144Hz-60Hz 16.020000  3.455134 28.58487 0.0226806
144Hz-90Hz  9.183333 -3.381532 21.74820 0.1224995

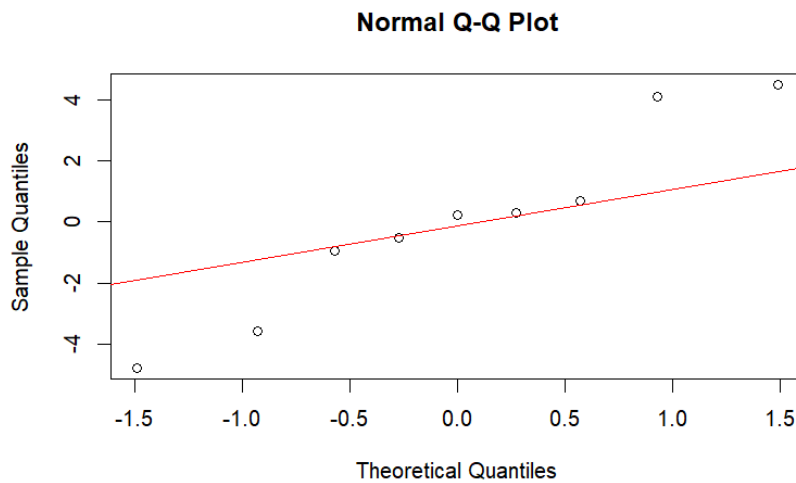
$brightness_level
      diff      lwr      upr      p adj
L2-L1  5.023333 -7.5415323 17.58820 0.4122775
L3-L1 12.683333  0.1184677 25.24820 0.0485634
L3-L2  7.660000 -4.9048657 20.22487 0.1901441
```

>>Result ANALYSIS

- Above, shows factors A & B have impact on each level where conf. Level = 0.95
- For Factor **A-Refresh rate** : there more difference between mid to high than low to mid shows there is higher variation on high Refresh rate .
- For Factor **B-Brightness level** : there more difference between mid to high than low to mid shows there is higher variation on high Brightness .
>>**NOT LINEARLY INCREASING**
- Rate of Discharging:
144Hz > 90Hz > 60Hz
L3>L2>L1

ABOVE RESULT implies that using 144Hz or 60Hz is the a crucial factor.

3. QQ-PLOT



Above , The residuals Exactly follow the normal distribution.

F. Conclusion

- Effects of Refresh Rate & Brightness are highly significant.
 - Others RESOLUTION & SOUND are less significant .
 - Factor C(Resolution) has a Negative effect.
 - Means better resolution is good for battery life.
 - Compared to Brightness (B) Refresh Rate (A) is more significant.
 - Rate of Discharging:
 - 144Hz > 90Hz > 60Hz
 - L3>L2>L1 (but NOT LINEARLY INCREASING)
 - **IDEAL CONDITION TO SET FACTORS :**
 - REFRESH RATE -60Hz
 - RESOLUTION - HIGH (RECOMMENDED)
 - BRIGHTNESS - ON LOWER SIDE
 - SPEAKER VOL - NOT MATTER MUCH
-

THANKYOU