Data Quest – Unlocking the Power of Data

(Case Study Challenge '23)

Track	***From fork to fitness**: Relationship between Eating Patterns and Obesity*
College/University	*Sardar Patel University , V.V Nagar Anand Gujarat India*
Software Used	*Excel , R – Programming, Python*

Student Name	Email ID	PG Course	Current semester
Kunal Chatur Salunkhe	kunal31301@gmail.com	M.Sc. Applied Statistics	III
Rupali Shivaji Kharat	kharatr002@gmail.com	M.Sc. Statistics	III
Khushi Vishnubhai Patel	khushi.4932@gmail.com	M.Sc. Statistics	III

Instructions:

(a) The below slides are mandatory. Addition of slides are allowed. The final PPT should not exceed more than 8 slides.

(b) Use bullet pointers/graphs to make your presentation more concise and effective.

What is your data speaking?

- Source of data is coming from clinical trial.
- Our data has been clean. There is no null values.
- In the data class of age are between 12 to 18.
- In our data most number of visit from “China” i.e 436 visit.
- Our data says both treatment are equally effective for reducing weight.
- Data tells about maximum reduce weight between Baseline visit & End of Treatment visit.
- In our data there is no particular relationship between BMI- Body Mass Index & UCS - Uncontrolled Eating Score, ECS - Emotional Eating Score, RCS - Unrestrained Control Score.
- Our data is not independent.
- Data tells the average median income of visitors is 29975.093040.
- Data tells the real story of magic Pill – *Testa*

Summary/Analysis/Visualization

Task 1 ***: Which of the two treatments is your magic Pill – Testa in reducing body weight..?***

- ## ***H0 : Treatment A & Treatment B are equal reducing effect body weight.***
- ## ***H1 : Treatment A is better than Treatment B in reducing effect body weight.***
- *# We check the summary Treatment A & Treatment B*

```
> summary(Tr_A)
```

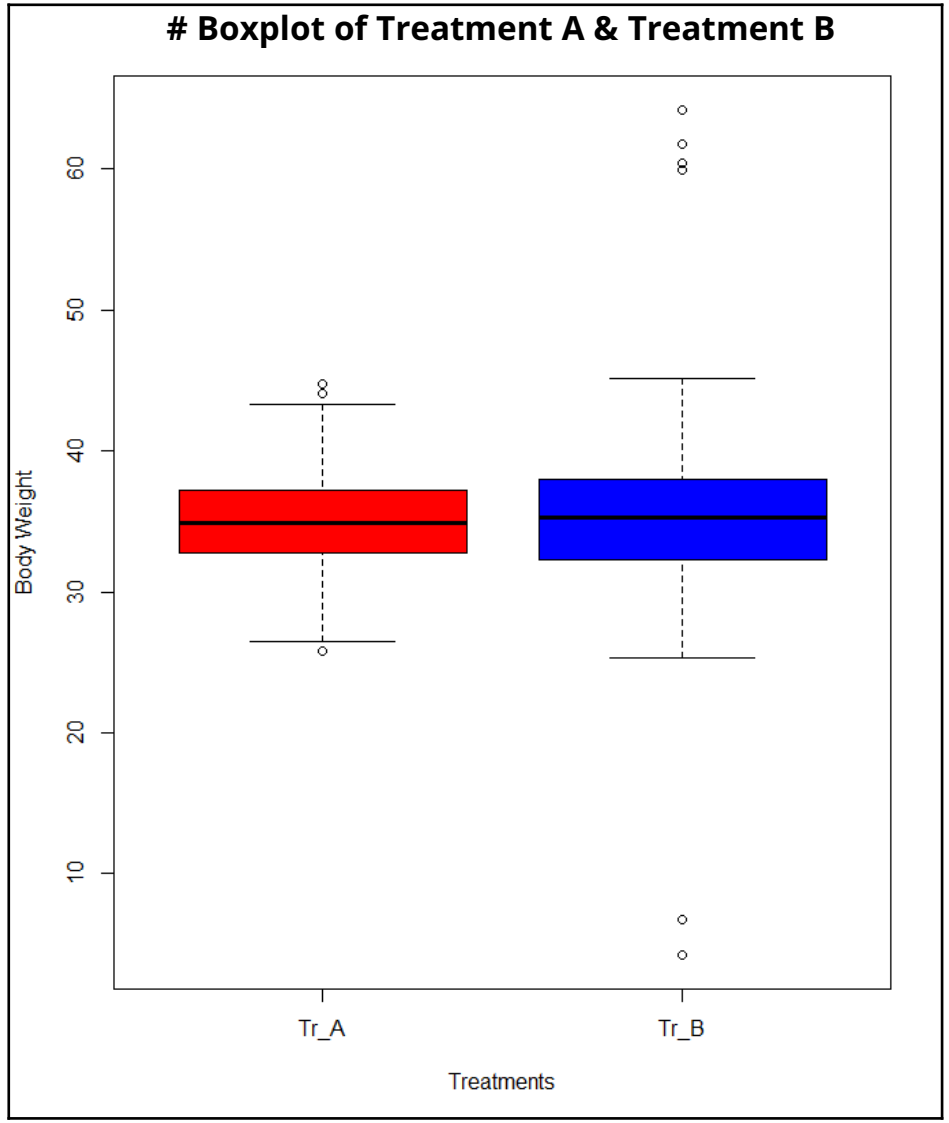
Min.	*1st Qu.*	*Median*	Mean	*std.*	*3rd Qu.*	*Max.*
25.79	*32.76*	34.84	34.86	3.7773	*37.16*	*44.73*

```
> summary(Tr_B)
```

Min.	*1st Qu.*	Median	Mean	*Std.*	*3rd Qu.*	*Max.*
4.263	32.292	35.242	35.242	7.0989	37.958	64.168

- **Here variance of "Treatment A" & variance of "Treatment B" are not same.**

So we use ***"FISHER BEHREN test"*** is used because variance are not same.



Summary/Analysis/Visualization

Task 2 *: How TestQ scores better explain eating behavior patterns..?*

- **1) For the Treatment A: Import the data in R**
  - `> tr_A=read.csv(file.choose(),header = TRUE)`
  - `> y=tr_B$Avg.BMI, x1=AVG.UCS , x2=AVG.ECS , x3=AVG.RCS ["AVG = Average"]`

2) Checking assumption to perform a regression analysis

i) Data is random or not | ii) Data is normally distributed | iii) Error are normally distributed

- **3) To checking data is independent (Random) or not, we perform runs test.**

H_0: DATA IS RANDOM & H_1: DATA IS NOT RANDOM

```
Runs Test - Two sided

data: y | Standardized Runs Statistic = -1.7057, p-value = 0.08806
```

- So here p-value is greater than 0.05 so we do not reject our H0 and data is random.
- **4) To checking data is normally distributed , we perform "shapiro-wilks normality test" & Draw boxplot, histogram, Quantile- Quantile plot | H0: DATA IS NORMALLY DISTRIBUTED**
- `> shapiro.test(y)`

```
Shapiro-Wilk normality test | data:  y | W = 0.98843, p-value = 0.3747
```

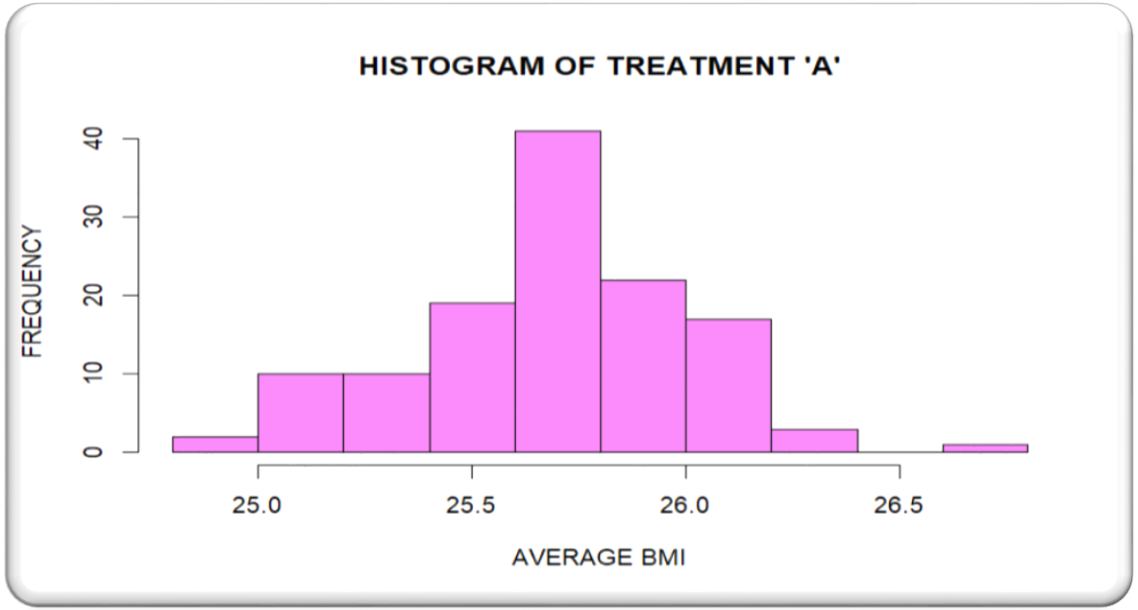
We accept our Ho & data is normally distributed

- `> MODEL=lm(y~x1+x2+x3)`

```
Call: lm(formula = y ~ x1 + x2 + x3)

Coefficients:

(Intercept)          x1          x2          x3

  25.021405    0.008257   -0.025032    0.052320
```

```
> summary(MODEL)

Residuals:
 Min       1Q   Median       3Q      Max -
   0.88891 -0.17914  0.01405  0.18114  0.96529
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  25.021405   1.235519  20.252   <2e-16 ***
x1            0.008257   0.046376   0.178    0.859
x2           -0.025032   0.071881  -0.348    0.728
x3            0.052320   0.033905   1.543    0.125
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.3094 on 121 degrees of freedom
Multiple R-squared:  0.02069,       Adjusted R-squared:  -
0.003591
 F-statistic: 0.8521 on 3 and 121 DF,  p-value: 0.4681
```

From above P-value is greater than 0.05 so we accept H_0 that is there is no significant relationship between independent and dependent variable.



Summary/Analysis/Visualization

Novo Nordisk®

Task 2 *: How TestQ scores better explain eating behavior patterns..?*

- **1) For the Treatment B : Import the data in R**
- `> tr_B=read.csv(file.choose(),header = TRUE)`
- `> y=tr_B$Avg.BMI, x1=AVG.UCS, x2=AVG.ECS, x3=AVG.RCS ["AVG = Average"]`
- **2) Checking assumption to perform a regression analysis**

  i) Data is random or not | ii) Data is normally distributed | iii) Error are normally distributed
- **3) To checking data is independent (Random) or not, we perform runs test.**

  `H0: DATA IS RANDOM & H1: DATA IS NOT RANDOM`

  Runs Test - Two sided

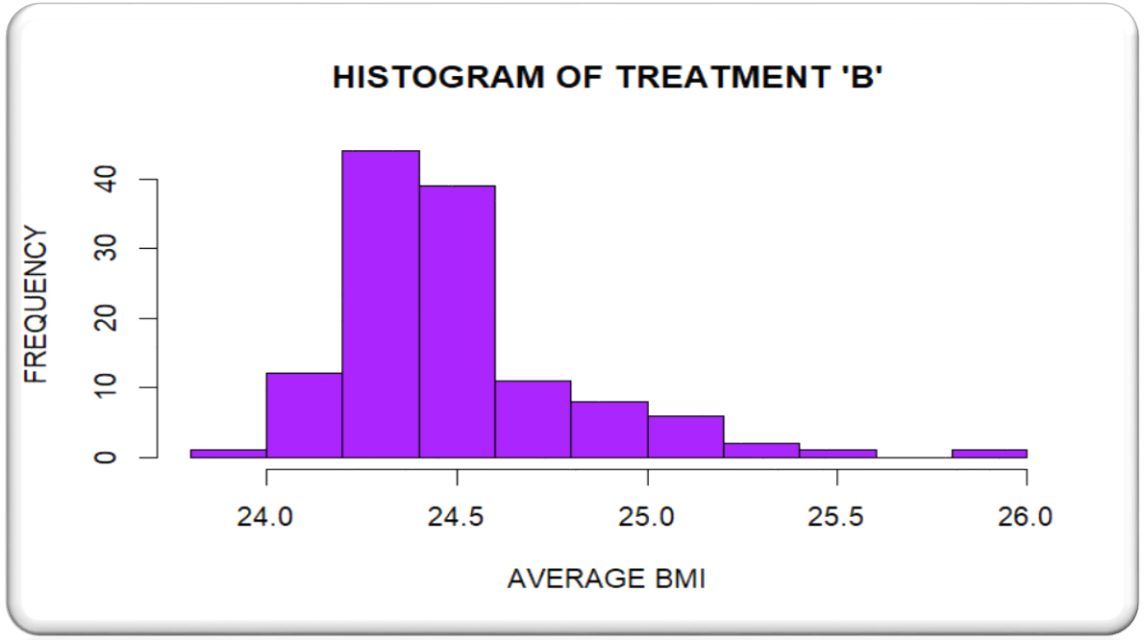
  data: y | Standardized Runs Statistic = 0.09053, p-value = 0.9279
- So here p-value is greater than 0.05 so we do not reject our H0 and data is random.
- **4) To checking data is normally distributed , we perform "shapiro-wilks normality test" & Draw boxplot, histogram, Quantile - Quantile plot**
- `H0: DATA IS NORMALLY DISTRIBUTED & H1: DATA IS NON NORMALLY DISTRIBUTED`
- `> shapiro.test(y)`
- Shapiro-Wilk normality test | data: y | W = 0.87883, p-value =1.119e-08

  We reject H0 data does not follows normal distribution
- `> Mmodel = rlm(tr_B$Avg.BMI ~ tr_B$Avg.UCS + tr_B$AVG.ECS + tr_B$Avg.RCS ,tr_B = tr_B )`

```
Call:rlm(formula = y ~ x1 + x2 + x3, tr_B = tr_B)
Converged in 6 iterations
Coefficients:
(Intercept)          x1          x2          x3
25.00696003  0.00970891 -0.09181338  0.01074824
Degrees of freedom: 125 total; 121 Residual, Scale estimate: 0.195
```

```
> summary(model)

Call: rlm(formula = y ~ x1 + x2 + x3, tr_B = tr_B)
Residuals:
    Min      1Q  Median      3Q     Max
-0.48565 -0.12901 -0.03773  0.13149  1.37787
Coefficients:
            Value   Std. Error t value
(Intercept) 25.0070  1.2605    19.8393
x1           0.0097  0.0397     0.2444
x2          -0.0918  0.1218    -0.7539
x3           0.0107  0.0091     1.1760
Residual standard error: 0.1949 on 121 degrees of freedom
```

From above t statistics is less than t critical value so we accept H_0 that is there is no significant relationship between independent and dependent variable.



Method used and Justification

Task 1 *: Which of the two treatments is your magic Pill – Testa in reducing body weight..?*

- ***Method* : "FISHER BEHREN t-test" using Excel & R- Programming**
- ***Justification* : We have taken difference to make "independent observations" weight of "Treatment_A & Treatment_B" through the 1st visit (Visit 0) & last visit (Visit 3).**

  We take summary & check the mean & variance | [Using Boxplot visualization]

  Here variance of "Treatment A" & variance of "Treatment B" are not same.

  So we used ***"FISHER BEHREN t-test"*** because variance are not same.

Task 2 : *How TestQ scores better explain eating behavior patterns?*

***Method* :**

- **i) "Multiple linear regression" for treatment A**

  ii) "Robust linear regression" for treatment B
- ***Justification :***

  i) Use "Multiple linear regression" for treatment A because it follows normality & Independent in data.

  ii) Use "Robust linear regression" for treatment B because it follows independency & non – normality in data.

Explanation/Conclusion/Interpretation

Task 1 : ***Which of the two treatments is your magic Pill – Testa in reducing body weight..?***

- ***Explanation*** **: We have taken difference to make "independent observations" weight of "Treatment_A & Treatment_B" through the "Baseline visit & End of treatment visit".**

  We take summary & check the mean & variance | [Using Boxplot visualization]

  Here variance of "Treatment A" & variance of "Treatment B" are not same.

  So we used ***"FISHER BEHREN t-test"*** because variance are not same. [**Using R & Python**]

  We have apply both '**ONE TAIL TESTING" & "TWO TAIL TESTING"**

- ***Decision*** **:**

  P-value = 0.2799 > 0.05 [Here P-value is greater than 0.05]

  t-critical value = 1.65296 > test statistics t |-1.0837|,

  Hence we fail to reject H0 hence both treatment A & B has equal effect on reducing weight.

- ***Conclusion:*** **So , We accept the null hypothesis.**

- ***Interpretation*** **: We can say that Treatment A & Treatment B are our Magic Pill-Testa which is helpful in reducing body weight of young teenagers.**

Explanation/Conclusion/Interpretation

Task 2 *: How TestQ scores better explain eating behavior patterns..?*

- ***Explanation*** **: For the treatment A, We have mention in summary. After "shapiro-wilks normality test" , we have done through the scatter plot & correlation to find the relationship between independent variables (ECS, UCS, RCS) & dependent (BMI). There is no relationship between BMI & independent variables through scatter plot. Here data is normal so we perform regression analysis.**
- **HYPOTHESIS :**
- ***Null hypothesis (H0):*** **The regression coefficient (BETA CAP) for a specific independent variable is equal to zero, indicating no significant relationship between that independent variable and the dependent variable.**
- ***Alternative hypothesis (H1):*** **The regression coefficient (BETA CAP) for a specific independent variable is not equal to zero, indicating a significant relationship between that independent variable and the dependent variable.**
- ***Through the model Conclusion*** **:**
- **Beta 0 cap = The estimated value of beta 0 is 25.021405 which** Interprets That when UCS ,RCS & ECS are equal to zero then **BMI is 25.021405**
- **Beta 1 cap =** 0.008257 interprets that BMI is increased by 0.008257 times when change in **1 unit of average uncontrolled eating score** and average Uncontrolled Eating Score, average Unrestrained Control Score remains fixed.
- **Beta 2 cap = -0.025032** interprets that BMI is decreased by 0.025032 times when change in **1 unit of average Emotional Eating Score and** average uncontrolled eating score and Unrestrained Control Score remains fixed.
- **Beta 3 cap = 0.052320** interprets that BMI is increased by 0.052320 times when change in **1 unit of average Unrestrained Control Score** and average Uncontrolled Eating Score and Emotional Eating Score remains fixed.

- ***Explanation*** **: For the treatment B, We have mention in summary. After "shapiro-wilks normality test" , we have done through the scatter plot & correlation to find the relationship between independent variables (ECS, UCS, RCS) & dependent (BMI). Here *data is non normal* so we could not perform regression analysis. Here assumptions of traditional regression are violated so we perform "Robust linear regression."**
- **HYPOTHESIS :**
- ***Null hypothesis (H0):*** **The regression coefficient (BETA CAP) for a specific independent variable is equal to zero, indicating no significant relationship between that independent variable and the dependent variable.**
- ***Alternative hypothesis (H1):*** **The regression coefficient (BETA CAP) for a specific independent variable is not equal to zero, indicating a significant relationship between that independent variable and the dependent variable.**
- ***Through the model Conclusion*** **:**
- **Beta 0 cap = The estimated value of beta 0 is 25.00696003 which** Interprets That when UCS ,RCS & ECS are equal to zero then **BMI is 25.00696003.**
- **Beta 1 cap = 0.00970891** interprets that BMI is increased by 0.0097089times when change in **1 unit of average uncontrolled eating score** and average Uncontrolled Eating Score, average Unrestrained Control Score remains fixed.
- **Beta 2 cap** = **-0.09181338** interprets that BMI is decreased by 0.091813 times when change in **1 unit of average Emotional Eating Score** and average uncontrolled eating score and Unrestrained Control Score remains fixed.
- **Beta 3 cap** = **0.01074824** interprets that BMI is increased by 0.01074824 times when change in **1 unit of average Unrestrained Control Score** and average Uncontrolled Eating Score and Emotional Eating Score remains fixed.

Interpretation: TESTQ score does not explain eating behavior pattern as it has no relationship with BMI. As it is stated that higher the TESTQ score higher the obesity and obesity is directly proportional to BMI . But here BMI and TESTQ score has no relation , So TESTQ score does not explain eating behavior pattern.

Recommendations/Scope

- By using odds ratio method we can find association between BMI and median income. Through this we can identify which median income group should be focus for marketing.
- We can find association between treatment effect and age group by correlation . So that further precise clinical trial can be done.
- By dividing BMI group of baseline visit in to class where 250 young teens are divide into three part as per their BMI (below, above and healthy BMI associated with obesity) and then finding the mean of reduce weight for each class interval and then conclusion can be deducted that which class interval is more effective with treatment. This can be done for further clinical trial.
- We can also find which country people are responding to testa pill by annova taking dependent variable country and independent variable reduce weight or reduce BMI.
- We can also find comparison of treatment by its effect on gender by testing of hypothesis methodology.

Thank You..!! ☺