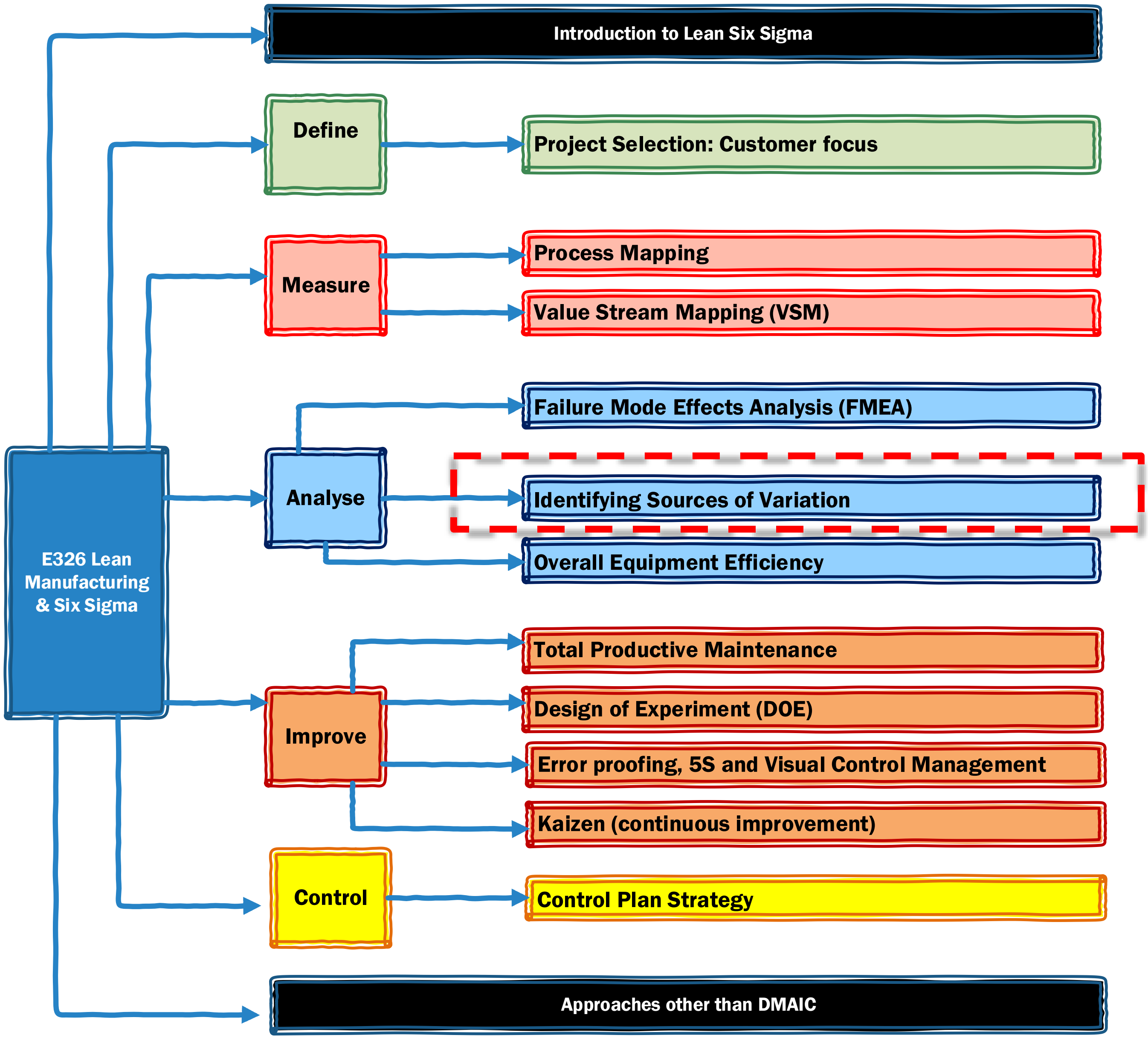
Problem 06
Sources of Variation

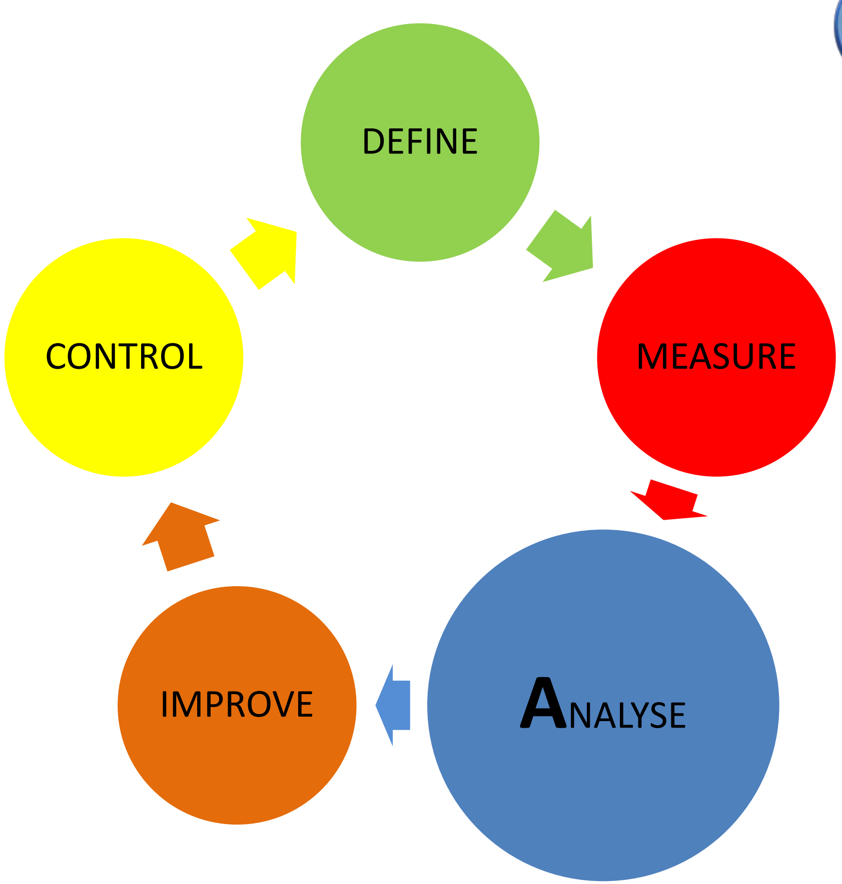
E326 – Lean Manufacturing & Six Sigma

SCHOOL OF ENGINEERING

E326 Lean Manufacturing and Six Sigma Topic Tree



DMAIC



- Verify the influence or Significant Factor (X's)
- Data analysis with proper tools
- Analyze sources of variations
- Prioritize opportunities to improve

Objectives of Analyse phase:

- To stratify and analyse the opportunity to identify a specific problem and define an easily understood problem statement
- To identify and validate the root causes that assure the elimination of "real" root causes and thus the problem the team is focused on
- To determine true sources of variation and potential failure modes that lead to customer dissatisfaction

Analyze Phase in Lean Six Sigma

- The Analyze phase aims to identify critical factors of a ‘good’ product or service, and the root causes of ‘defects’.
- It has less of a logical flow, but functions more as a toolbox of tools and techniques. The following questions are answered in Analyze phase in sequence:

  1) Q: How does the process actually work?

  A: Analyze the process;

  2) Q: What does the existing process knowledge say?

  A: Develop theories and ideas (potential root causes)

  3) Q: What does the data say?

  A: Analyze the data

  4) Q: How does the root cause affect the process output?

  A:Verify root causes and understand cause and effect.

Sources of Variation (SOV)

- Sources of variation are categorized into families of related causes and quantified to reveal the largest causes.
- Typical categories are -
  1) *Positional variation*: often called within-part variation and refers to variation of a characteristic on the same product;
  2) *Cyclical variation*: covers part-to-part variation;
  3) *Temporal variation*: occurs as change over time

Example: ***Study of the variation of metal hardness :***

Positional variation: *Collection of hardness data at different locations on the part.*

Cyclical variation: *Average hardness on consecutive parts to detect any pattern*

Temporal variation: *Average of hardness of parts selected from production on several days*

Tools to Analyze

- Graphical Tools
  - Fishbone Diagram
  - Marginal Plot
  - Scatter Plot
  - Histogram
  - Dot Plot
  - Box Plot
  - Multi-Vari Charts
  - ...
- Regression Analysis
- ANOVA (One-way ANOVA & Balanced ANOVA)
  - Main Effects Plots (ANOVA)
  - Interaction Plots (ANOVA)

Note: Minitab helps us with various plots and analysis for problem solving.

Analysis of Variance (ANOVA)

- Analysis of variance (ANOVA) is a hypothesis testing method used to evaluate how two or more levels of one or more factors affect the mean of a response. It does this by comparing the variances "within" groups and variances "between" groups.
  - One-way ANOVA performs a comparison of the means of a number of replications of the experiments performed where *a single input factor is varied at different settings or levels*.
  - Balanced ANOVA compares the response means with ***multiple*** key-process-input-variables (KPIVS) and ***interactions***.
- ANOVA Assumptions:
  1) *Each sample is an independent, random sample -*
     - *Independent. The selection of any sample is not dependent on any other sample being selected or not selected*
     - *Random. All members of the population have an equal chance of being selected.*
  2) *The measurement within each group are normally distributed and have equal variances -*
     - *This only applies to the within group variation, not between group variation.*
     - *The variances for each group are equal (or the differences between the variances are not significant)*

ANOVA – Test Procedures

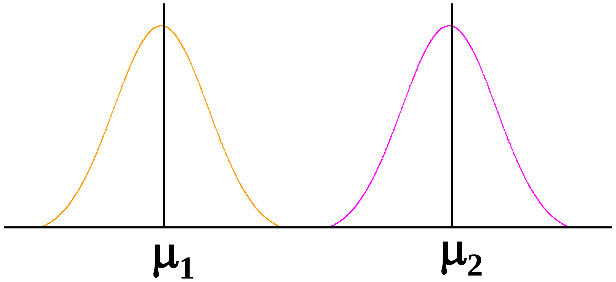
1. State the problem on-hand
2. State the Null hypothesis (as in Hypothesis Testing)
3. Do the model assumptions hold? (Not the focus of today's problem)
   - Normality Test
   - Test for Equal Variances
   - Independent sample?
   - random sample?
4. Construct the ANOVA Table
   - Manual calculation (Not the focus of today's problem)
   - Statistical program like Minitab
5. Interpret the output of analysis
6. Translate the conclusion into practical terms

t-test or 1-way ANOVA?

- Testing for two means
  - 2-sample t-test
  - Paired t-test

H_0: $\mu_1 = \mu_2$

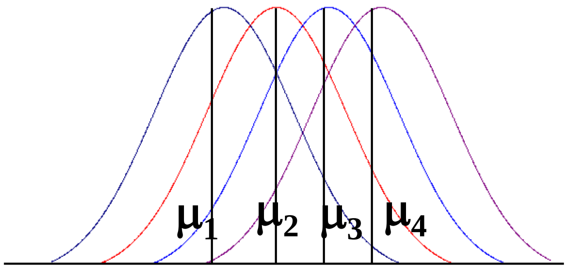
H_a: $\mu_1 \neq \mu_2$



- Testing for three or more means
  - 1-way ANOVA
  - Balanced ANOVA

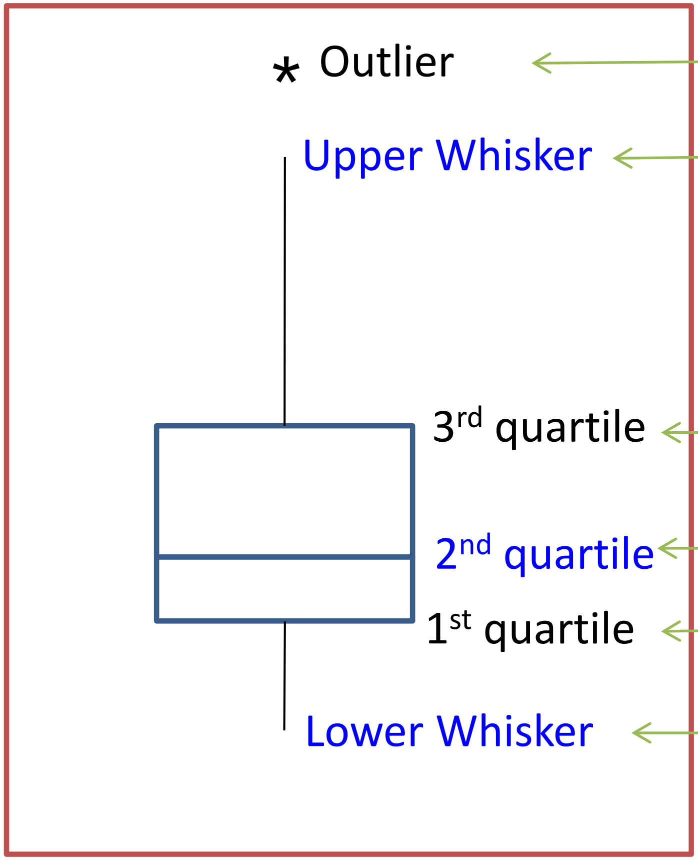
H_0: $\mu_1 = \mu_2 = \ldots = \mu_i$

H_a: at least one is different from the others



Recall - Boxplot

- Boxplots, or box-and-whisker diagrams, give a quick look at the distribution of a set of data.
- They allows easy comparison of multiple data sets.



More than 1.5x of upper quartile

Extends to the maximum data point within 1.5 box heights from the top of the box

75% of the data are less than or equal to this value

50% of the data are less than or equal to this value

25% of the data are less than or equal to this value.

Extends to the minimum data point within 1.5 box heights from the bottom of the box

Recall – Boxplot - example

Outlier will be any points above 15.65

Q3 + (1.5 * IQR)
= 14.9 + (1.5 * 0.5) = 15.65

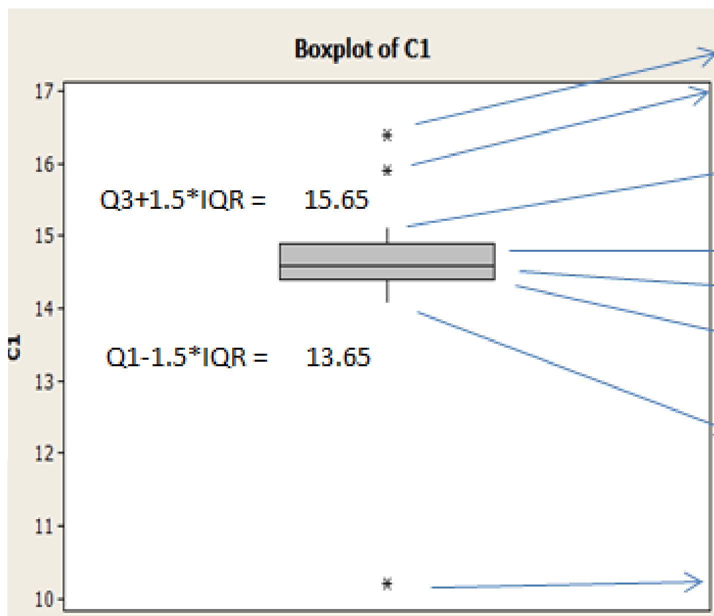
	16.4
	15.9
	15.1
Q3	14.9
	14.7
	14.7
	14.7
Median	14.6
	14.5
	14.5
	14.4
Q1	14.4
	14.4
	14.1
	10.2

IQR (Interquartile range)
= Q3-Q1 == 14.9 - 14.4
= 0.5

Q1 - (1.5 * IQR)
= 14.4 - (1.5 * 0.5) = 13.65

Outlier will be any points below 13.65

Minitab : Graph > Box plot



16.4 (Outlier -> Point above 15.65)

15.9 (Outlier -> Point above 15.65)

15.1 (Max data within 1.5xIQR or max data exclude outlier)

14.9 (Q3)
14.6 (Median)
14.4 (Q1)
IQR = 0.5

14.1 (Min data within 1.5xIQR or min data exclude outlier)

10.2 (Outlier -> Point below 13.65)

Multi-Vari Charts

- The Multi-Vari chart is used to display ANOVA data for up to 4 factors. It is commonly used as a preliminary analysis of the data to assess the means for each factor level. Each factor should have at least 2 distinct levels.
- View the data to see if there are any visible trends or interactions by creating a Multi-Vari chart.
- To provide directions and inputs for improvement activities
- A way of presenting analysis of variance (ANOVA) data in a graphical form, providing *a visual alternative to ANOVA.*
- It is a useful graphical tool for identifying sources of variations.

P06 Suggested Solution

Today's Problem

BEST ELECTRICITY
A PRIME GROUP COMPANY

Diamond Electric

ESPower
Positively Different

Geneco™
Powered by Seraya Energy

iswitch®
LICENSED ELECTRICITY RETAILER

Keppel Electric

Ohm.
Fuss-free energy

PacificLight

sembcorp
Energy that makes a difference

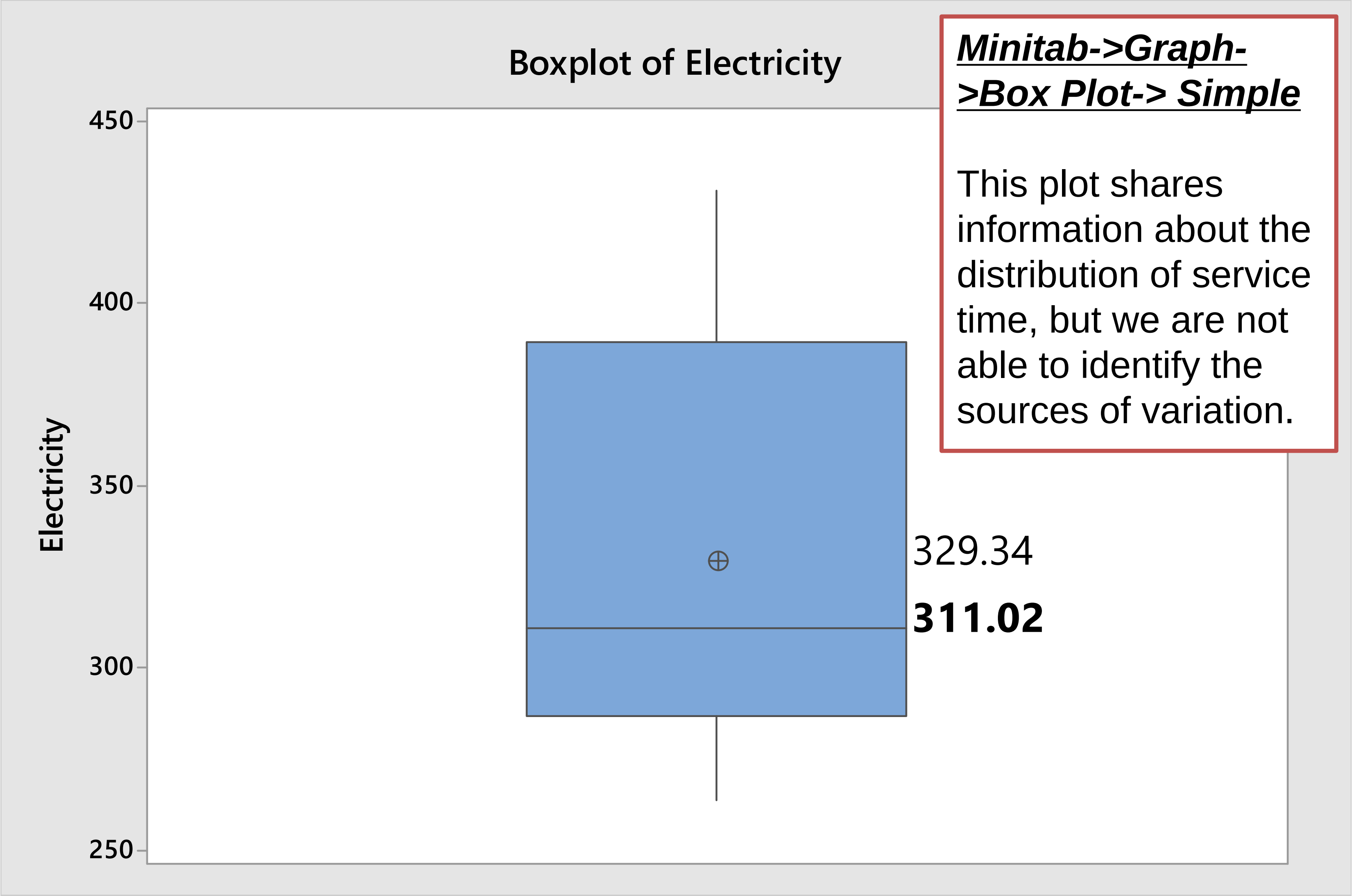
Senoko
energy for life

SUNSEAP

TUAS POWER SUPPLY

union POWER

Simple Box Plot

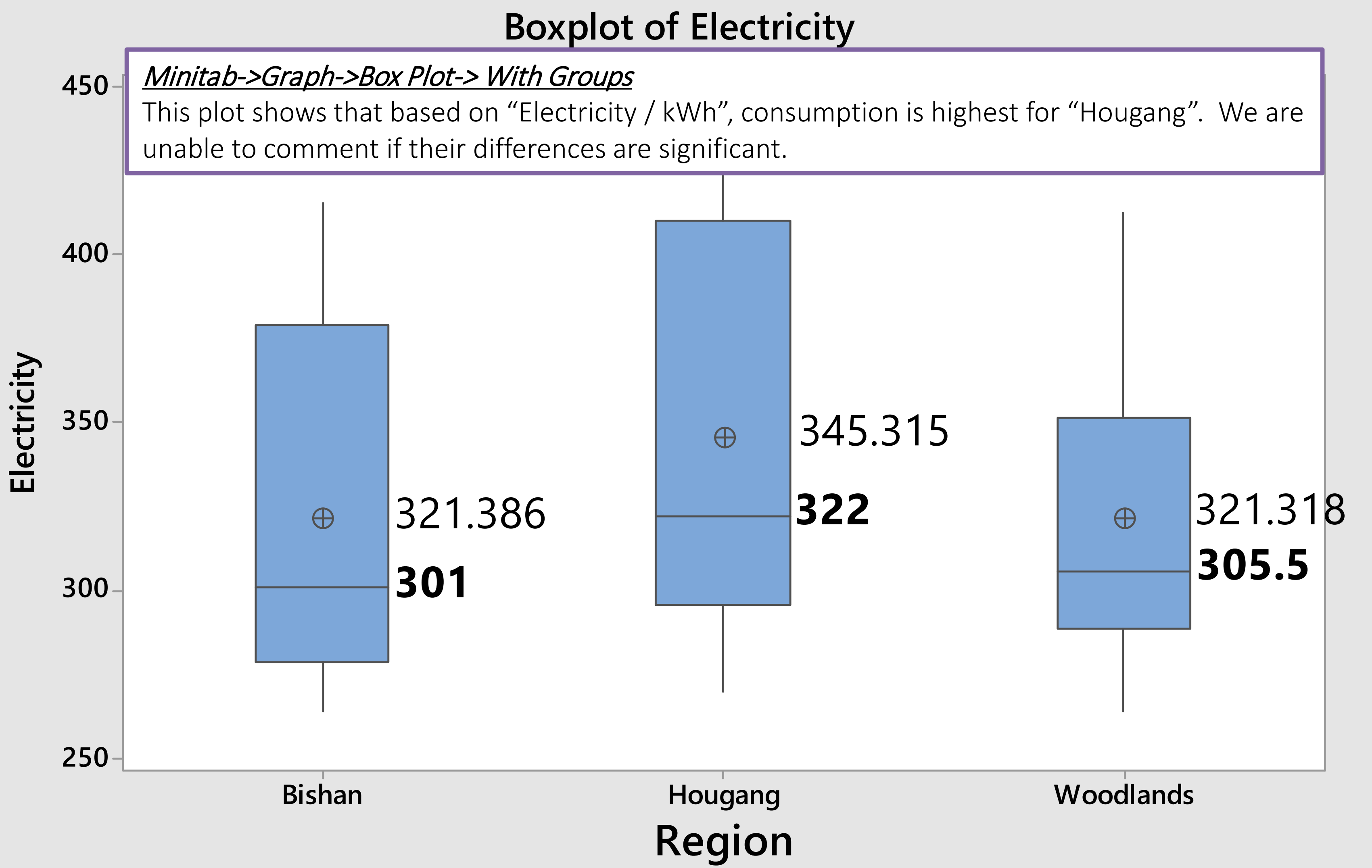


<u>Minitab->Graph->Box Plot-> Simple</u>

This plot shares information about the distribution of service time, but we are not able to identify the sources of variation.

Box Plot With Groups – “Region”

Minitab->Graph->Box Plot-> With Groups

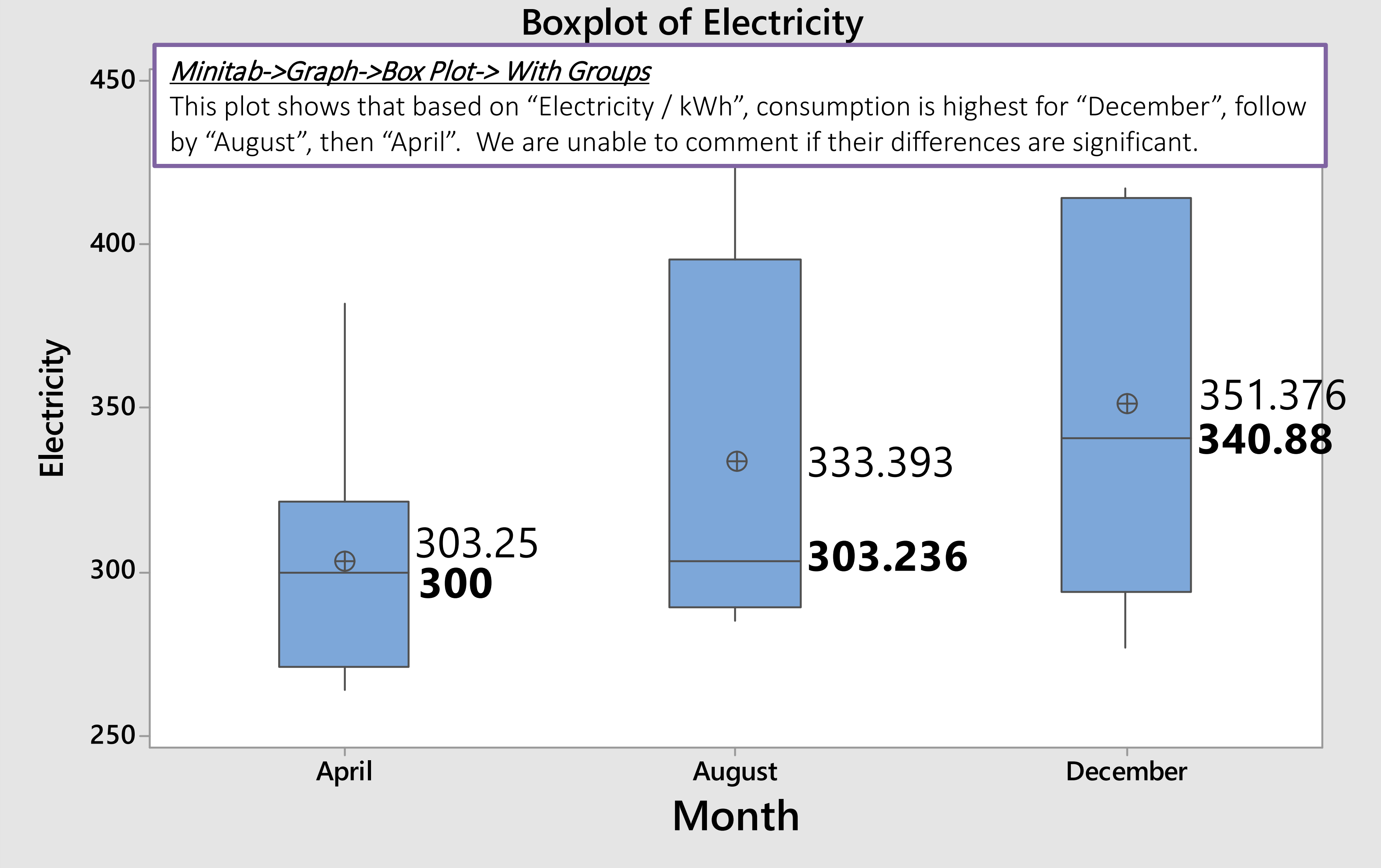


Minitab->Graph->Box Plot-> With Groups

This plot shows that based on “Electricity / kWh”, consumption is highest for “Hougang”. We are unable to comment if their differences are significant.

Box Plot With Groups – "Month"

Minitab->Graph->Box Plot-> With Groups

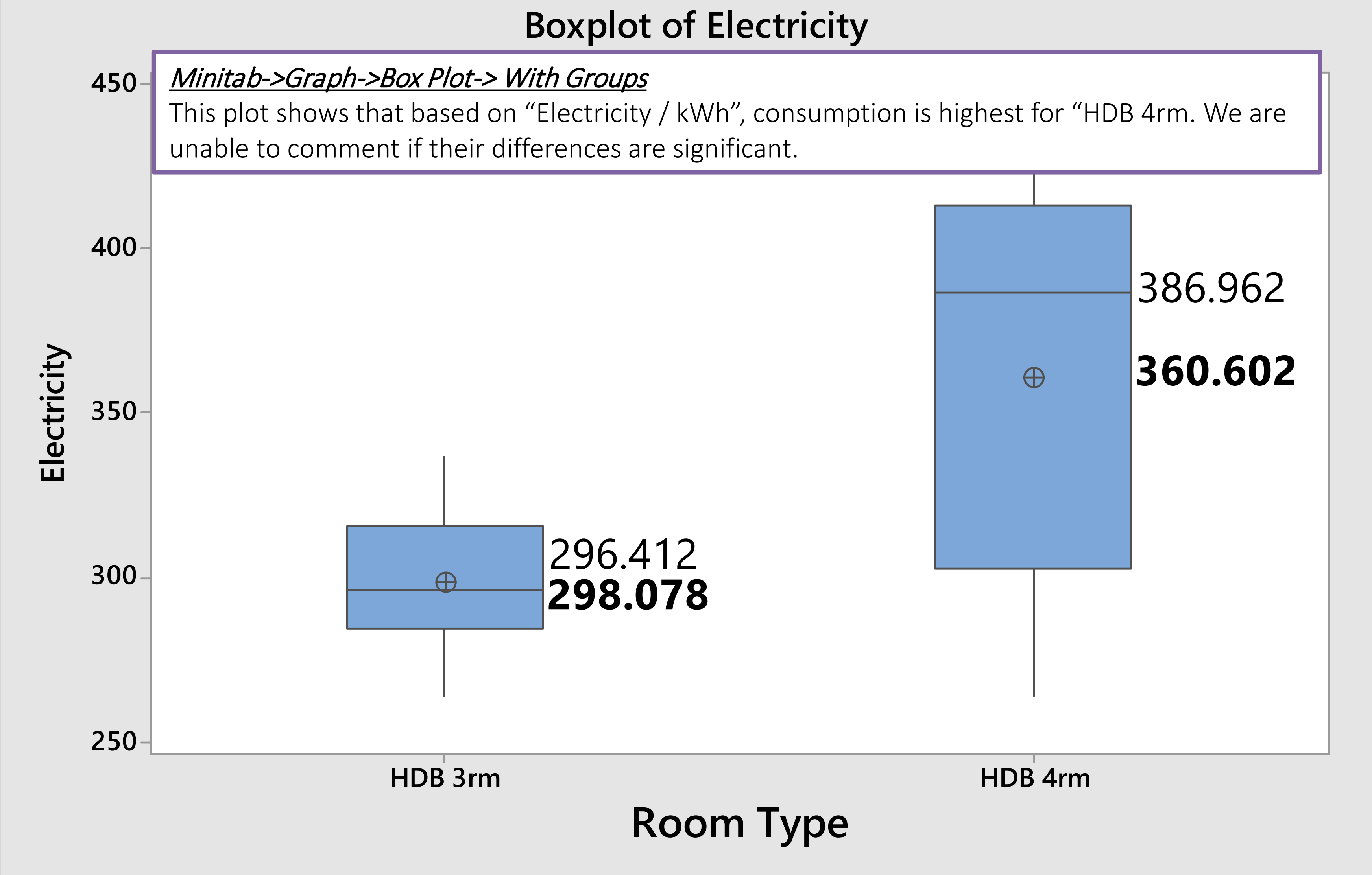


Minitab->Graph->Box Plot-> With Groups

This plot shows that based on "Electricity / kWh", consumption is highest for "December", follow by "August", then "April". We are unable to comment if their differences are significant.

Box Plot With Groups – “Room Type”

Minitab->Graph->Box Plot-> With Groups



Multi-Vari Chart

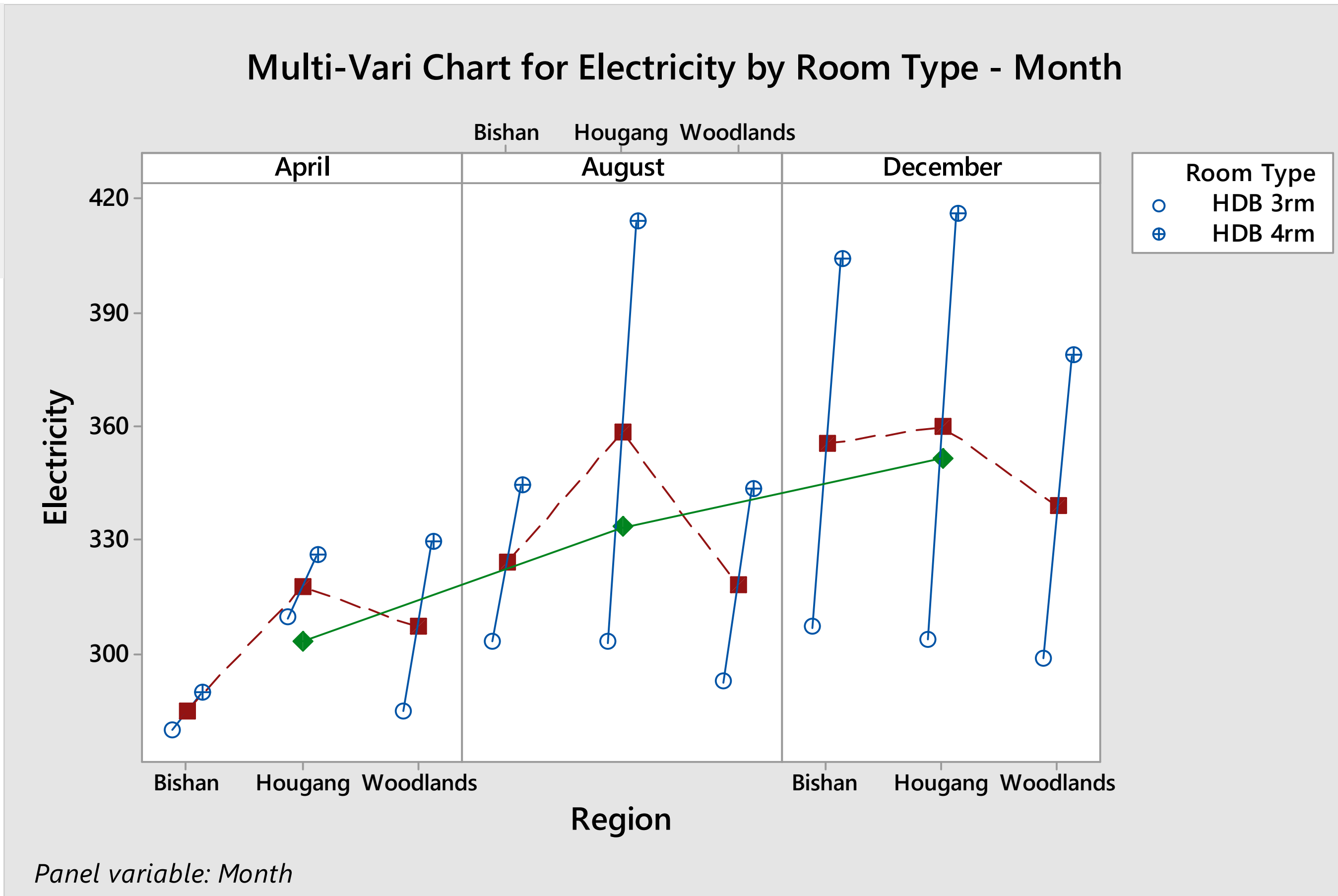
Minitab->Stat>Quality Tools -> Multi-Vari Chart

Response: Electricity

Factor 1: 'Room Type'

Factor 2: Region

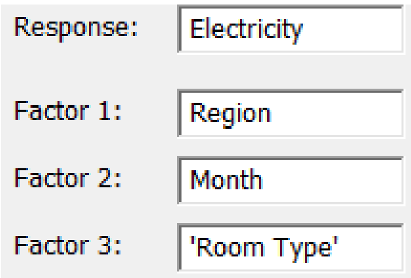
Factor 3: Month

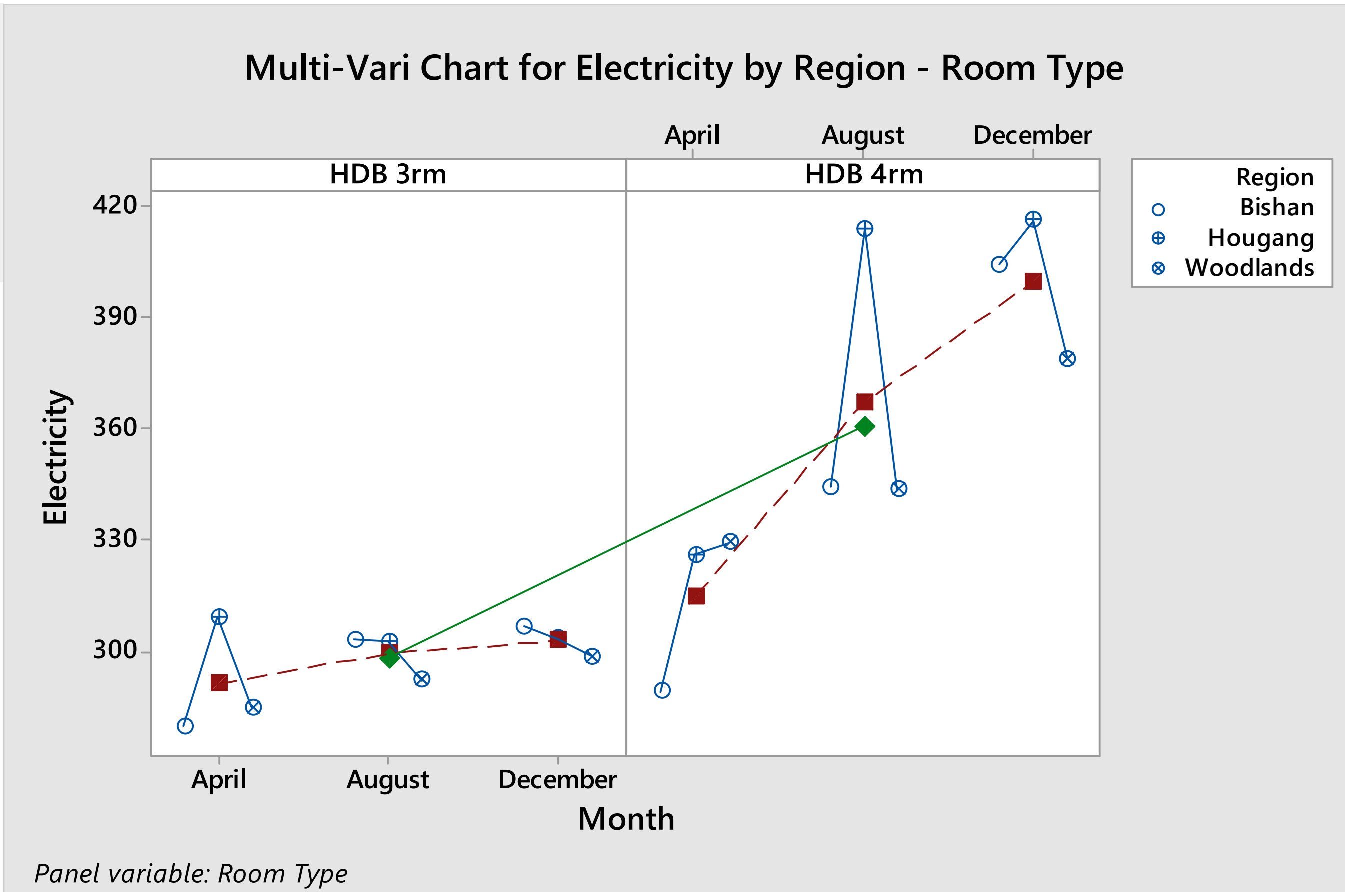


- Consumption is highest for "December", follow by "August", then "April". (Green line)
- Consumption is highest for "Hougang". (Red lines)
- Consumption is highest for "HDB 4rm". (Blue lines)

Multi-Vari Chart

Minitab->Stat>Quality Tools -> Multi-Vari Chart





- Consumption is highest for “December”, follow by “August”, then “April”. (Green lines)
- No observable trend for “Region”. (Blue lines)
- Consumption is highest for “HDB 4rm”. (Green line)

Multi-Vari Chart

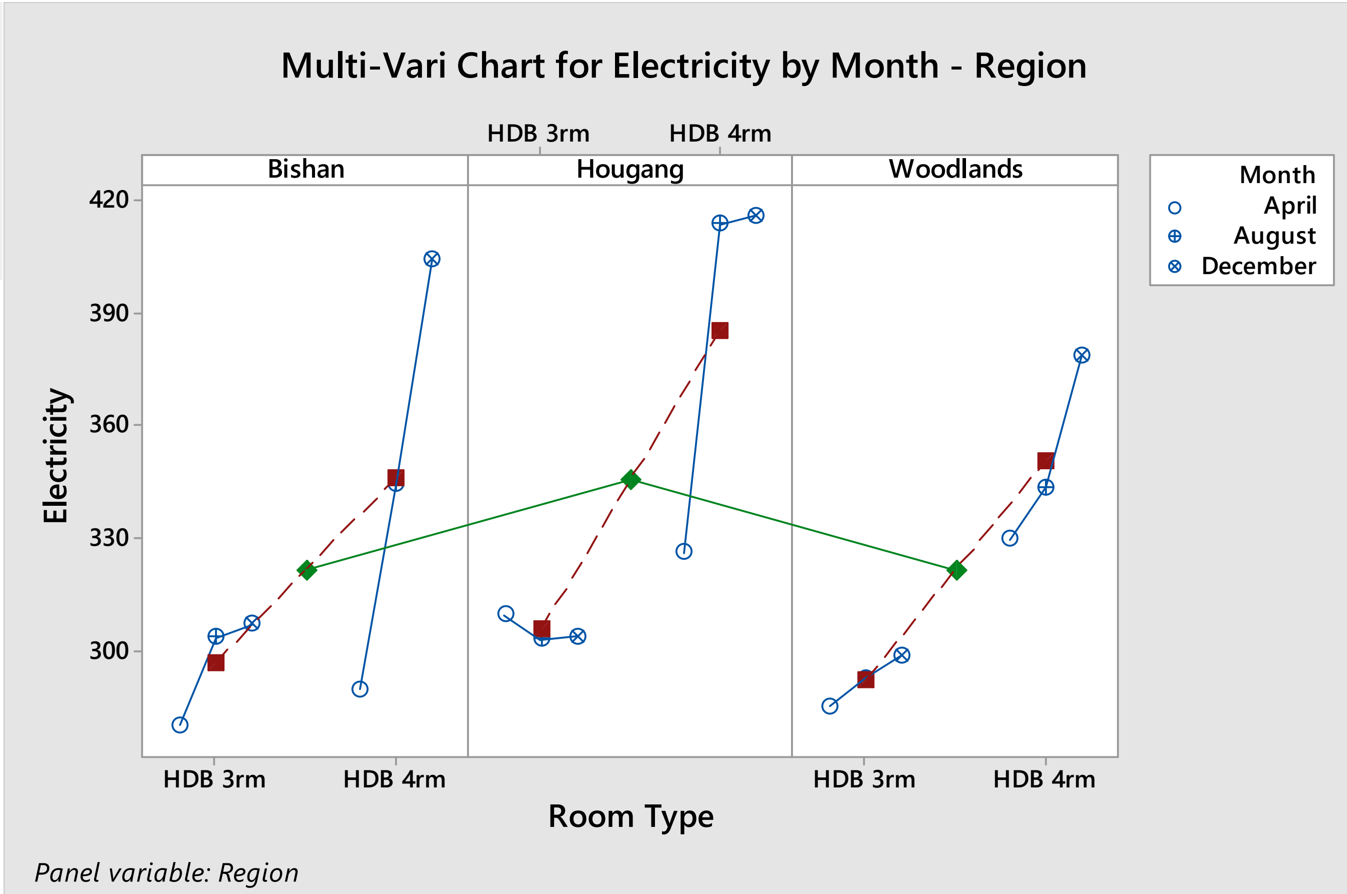
Minitab->Stat>Quality Tools -> Multi-Vari Chart

Response: Electricity

Factor 1: Month

Factor 2: 'Room Type'

Factor 3: Region



- No observable trend for “Month”. (Blue lines)
- Consumption is highest for “Hougang”. (Green line)
- Consumption is highest for “HDB 4rm”. (Red lines)

Setting up Hypothesis testing for ANOVA

- ANOVA allows us to compare the means of k independent populations in a single test, to determine if there are any significant differences among the means.
- Null hypothesis,

  $\mathrm{H_0}$: Population means are all equal

  Alternative hypothesis,

  $\mathrm{H_1}$: At least one mean is different or Not all population means are equal

- ANOVA uses the F test statistic:

  $F = \frac{MS(Tr)}{MSE}$ which follows F distribution with

  $(k - 1)$ numerator degrees of freedom, and

  $(n - k)$ denominator degrees of freedom

One-way ANOVA – “Region”

One-way ANOVA: Electricity versus Region

```
Method

Null hypothesis         All means are equal
Alternative hypothesis  At least one mean is different
Significance level      α = 0.05

Equal variances were assumed for the analysis.


Factor Information

Factor  Levels  Values
Region       3  Bishan, Hougang, Woodlands


Analysis of Variance

Source  DF  Adj SS  Adj MS  F-Value  P-Value
Region   2    4594    2297     0.82    0.448
Error   33   92158    2793
Total   35   96752


Model Summary

      S   R-sq  R-sq(adj)  R-sq(pred)
52.8456  4.75%      0.00%       0.00%


Means

Region      N    Mean  StDev       95% CI
Bishan     12   321.4   53.9  (290.3, 352.4)
Hougang    12   345.3   58.5  (314.3, 376.4)
Woodlands  12   321.3   45.3  (290.3, 352.4)

Pooled StDev = 52.8456
```

Minitab-> Stat-> ANOVA ->One Way

H_0: All means are equal

H_1: At least one mean is different

P-value = 0.448>0.05. **Accept** Null hypothesis

Thus, “Region” **IS NOT** a significant factor contributing to the Electricity Consumption

Numeric interpretation that shows differences in consumption

One-way ANOVA – "Month"

One-way ANOVA: Electricity versus Month

Minitab-> Stat-> ANOVA ->One Way

Method

Null hypothesis	All means are equal
Alternative hypothesis	At least one mean is different
Significance level	α = 0.05

Equal variances were assumed for the analysis.

H_0: All means are equal

H_1: At least one mean is different

Factor Information

Factor	Levels	Values
Month	3	April, August, December

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Month	2	14192	7096	2.84	0.073
Error	33	82559	2502		
Total	35	96752			

P-value = 0.073>0.05. **Accept** Null hypothesis

Thus, "Month" **IS NOT** a significant factor contributing to the Electricity Consumption

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
50.0180	14.67%	9.50%	0.00%

Means

Month	N	Mean	StDev	95% CI
April	12	303.3	36.6	(273.9, 332.6)
August	12	333.4	54.8	(304.0, 362.8)
December	12	351.4	56.3	(322.0, 380.8)

Pooled StDev = 50.0180

Numeric interpretation that shows differences in consumption

One-way ANOVA – “Room Type”

One-way ANOVA: Electricity versus Room Type

Minitab-> Stat-> ANOVA ->One Way

```
Method

Null hypothesis         All means are equal
Alternative hypothesis  At least one mean is different
Significance level      α = 0.05

Equal variances were assumed for the analysis.


Factor Information

Factor     Levels  Values
Room Type       2  HDB 3rm, HDB 4rm


Analysis of Variance

Source     DF  Adj SS  Adj MS  F-Value  P-Value
Room Type   1   35183   35183    19.43    0.000
Error      34   61569    1811
Total      35   96752


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
42.5541  36.36%     34.49%      28.66%


Means

Room Type   N    Mean  StDev       95% CI
HDB 3rm    18  298.08  19.65  (277.69, 318.46)
HDB 4rm    18   360.6   56.9  ( 340.2,  381.0)

Pooled StDev = 42.5541
```

H_0: All means are equal

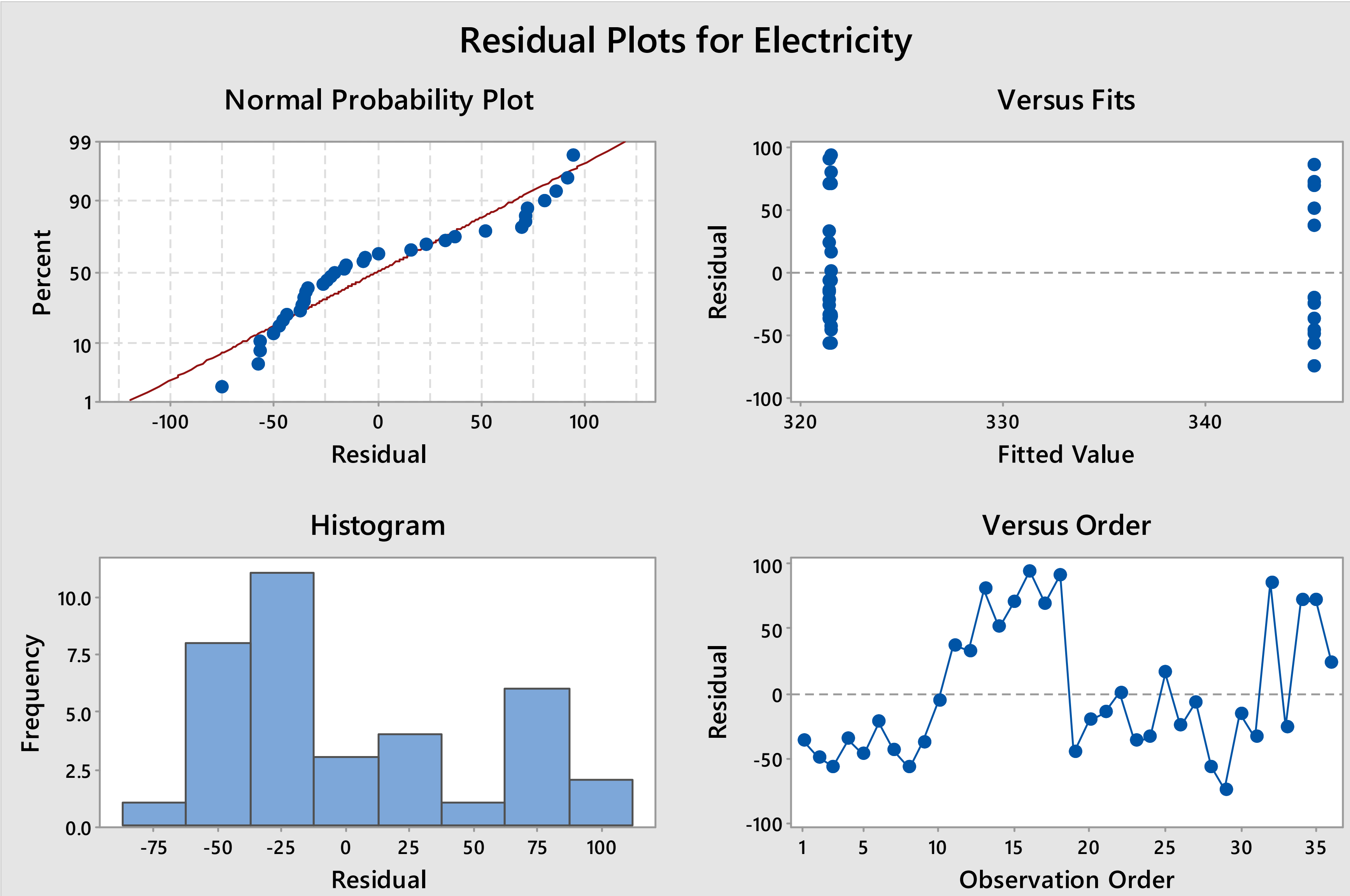
H_1: At least one mean is different

P-value = 0.000<0.05. **Reject** Null hypothesis

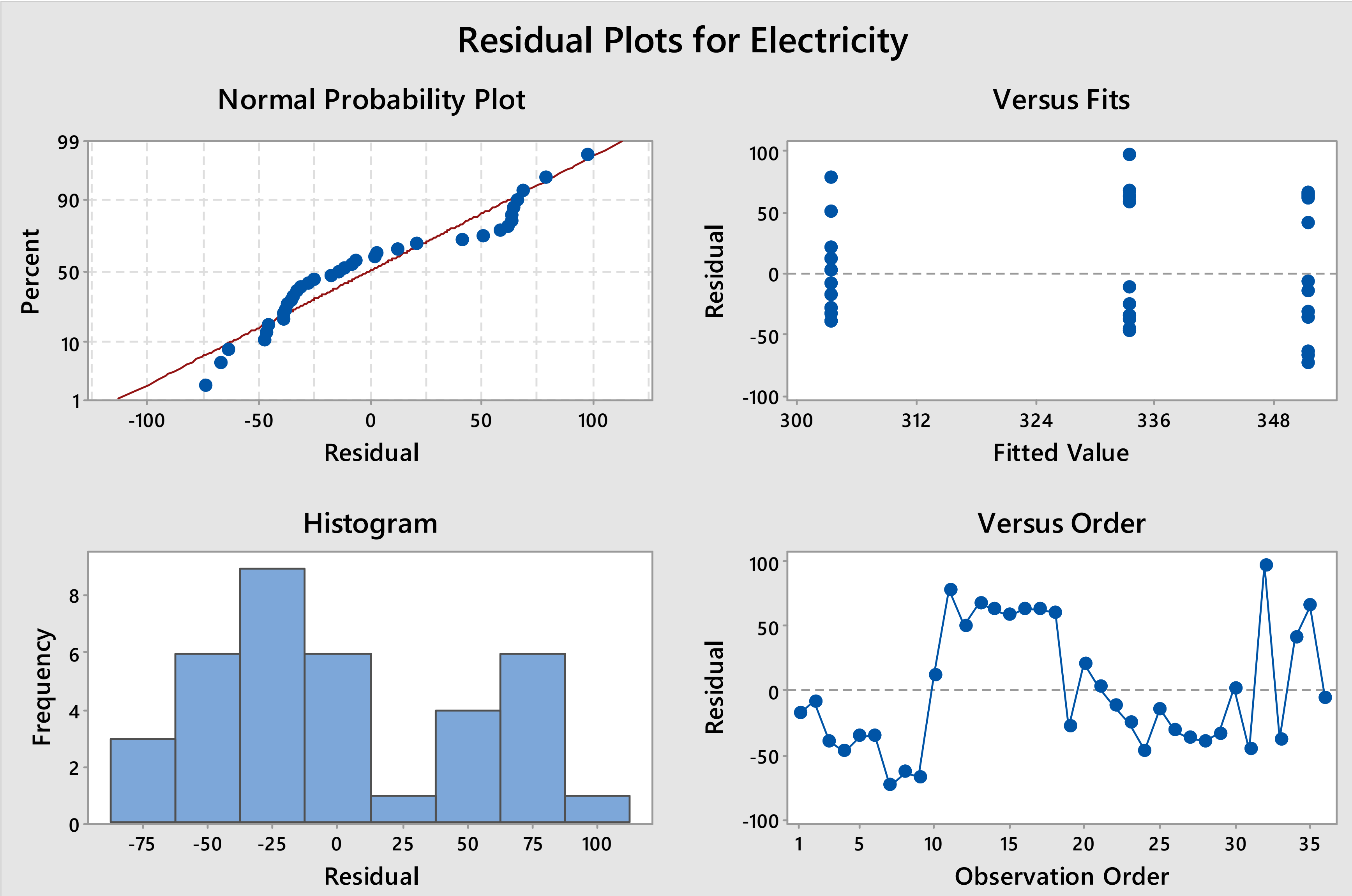
Thus, “Room Type” **IS** a significant factor contributing to the Electricity Consumption

Numeric interpretation that shows differences in consumption

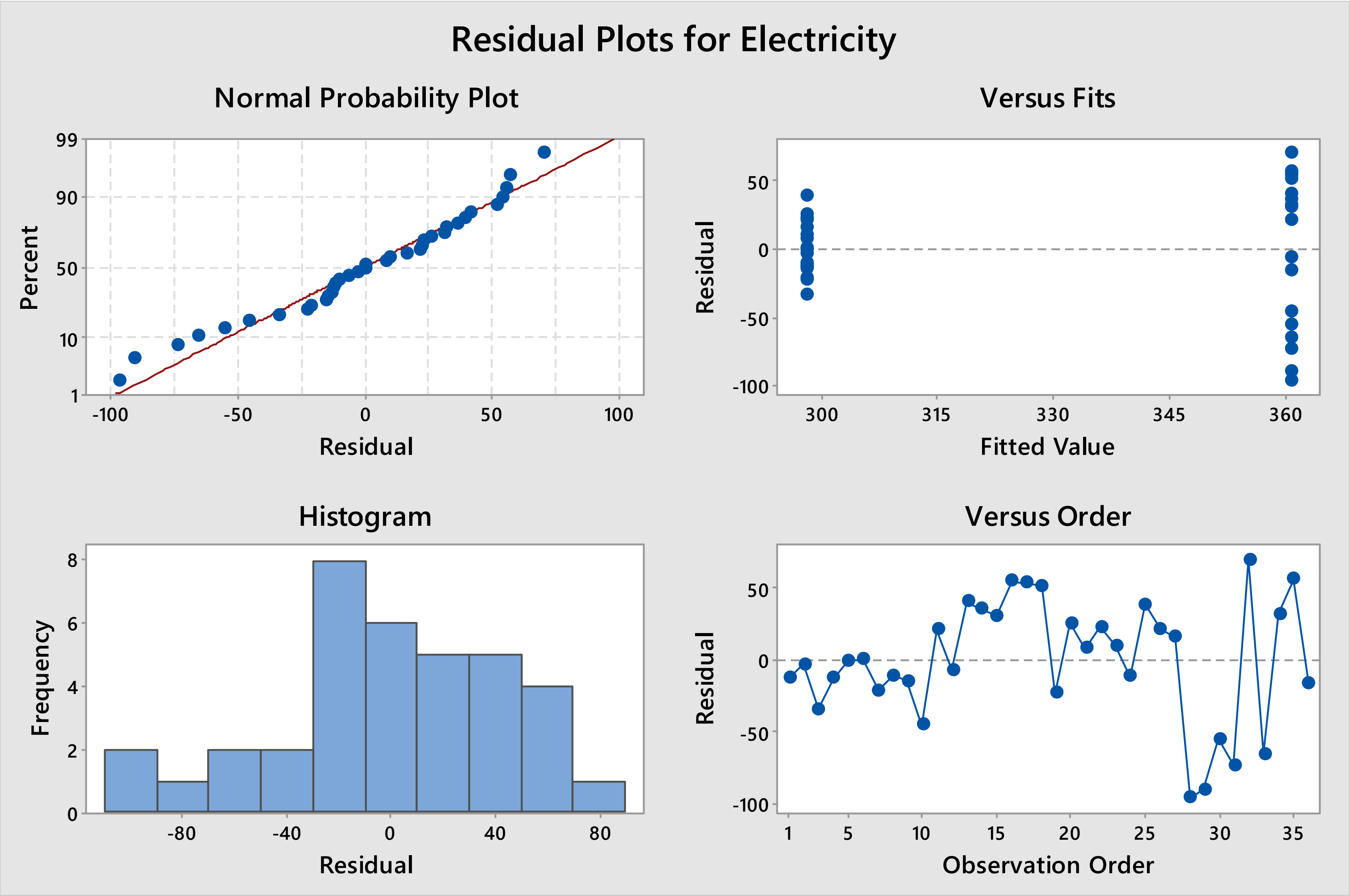
Residual Plot – “Region”



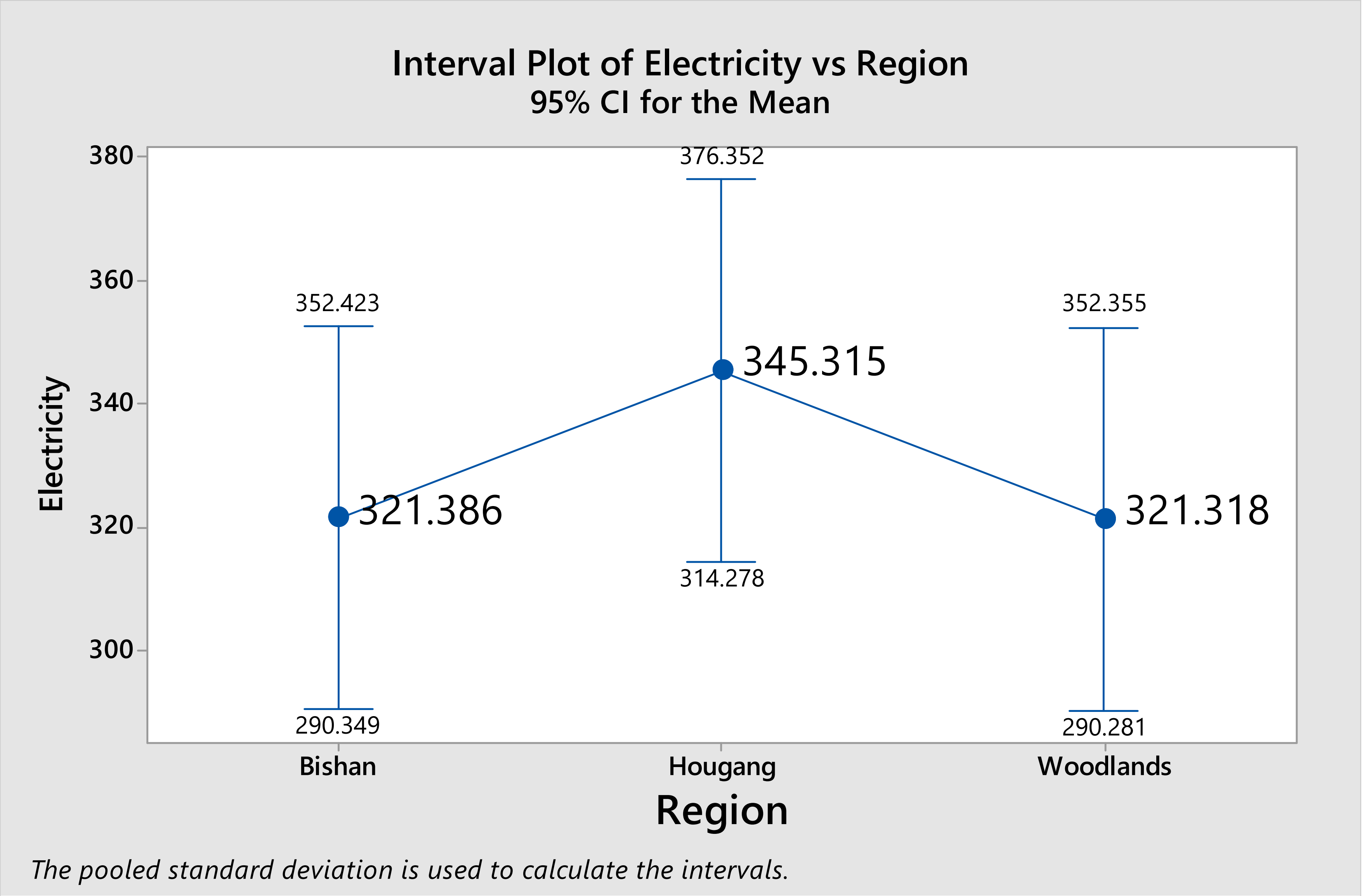
Residual Plot – "Month"



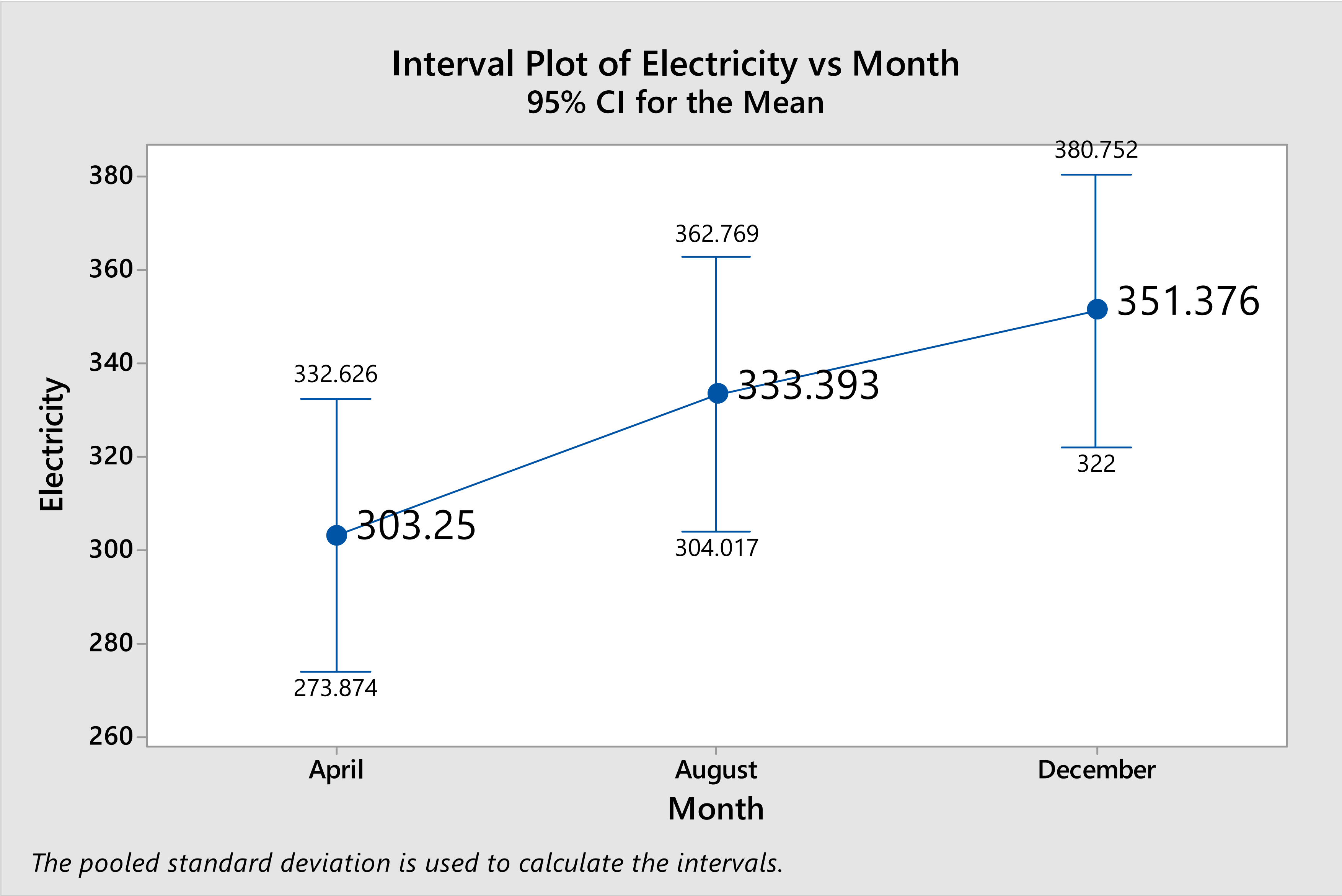
Residual Plot – “Room Type”



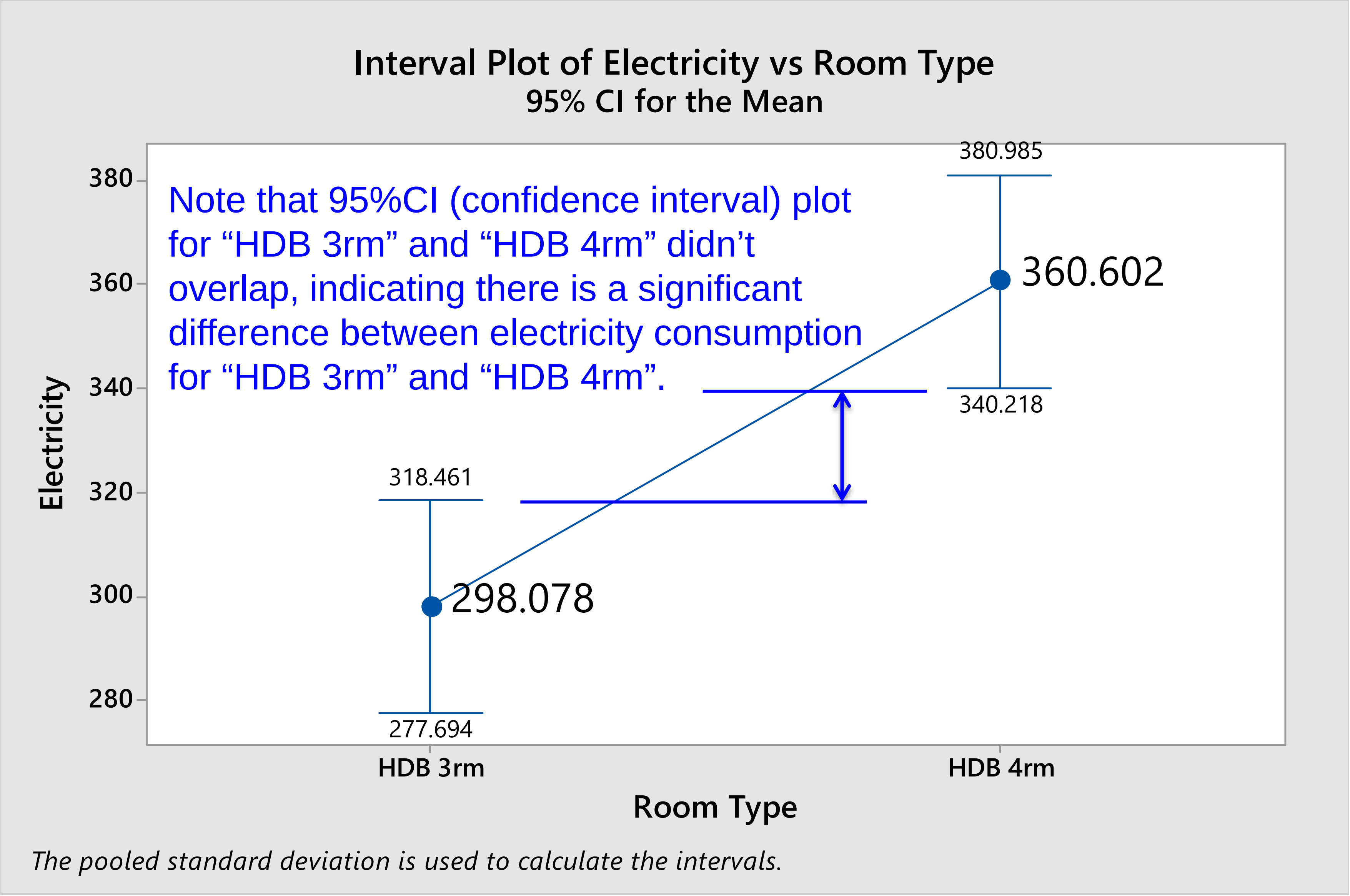
Interval Plot – “Region”



Interval Plot – “Month”



Interval Plot – “Room Type”



Balanced ANOVA

Minitab-> Stat-> ANOVA ->Balanced ANOVA

ANOVA: Electricity versus Region, Month, Room Type

Factor	Type	Levels	Values
Region	fixed	3	Bishan, Hougang, Woodlands
Month	fixed	3	April, August, December
Room Type	fixed	2	HDB 3rm, HDB 4rm

Analysis of Variance for Electricity

Source	DF	SS	MS	F	P
Region	2	4594	2297	1.47	0.256
Month	2	14192	7096	4.54	0.025
Room Type	1	35183	35183	22.51	0.000
Region*Month	4	2483	621	0.40	0.808
Region*Room Type	2	1481	741	0.47	0.630
Month*Room Type	2	8105	4053	2.59	0.102
Region*Month*Room Type	4	2583	646	0.41	0.797
Error	18	28129	1563		
Total	35	96752			

S = 39.5314   R-Sq = 70.93%   R-Sq(adj) = 43.47%

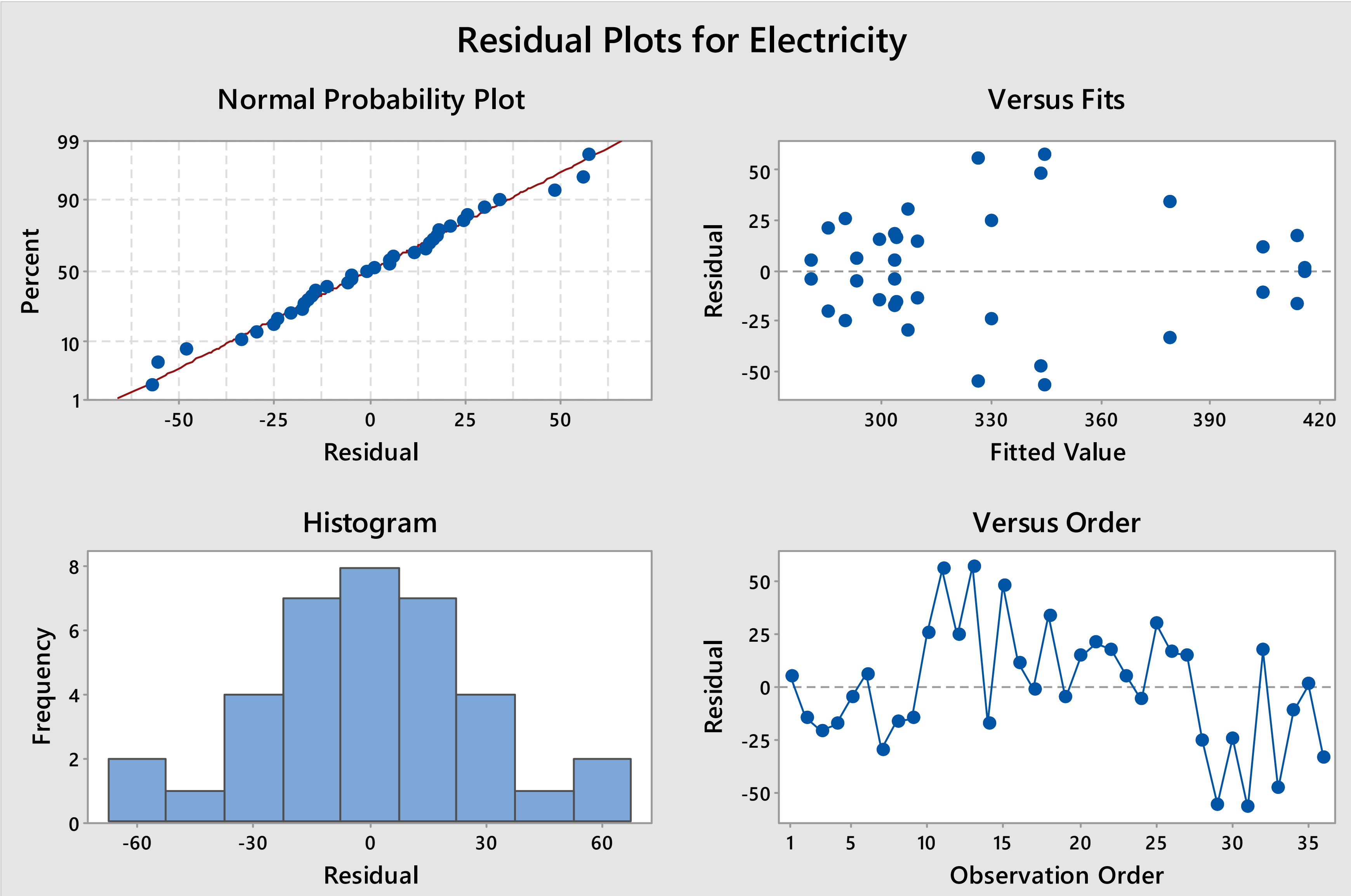
- Balanced ANOVA indicates that 2 main effects: “Month“ and “Room Type“ are significant.
- If P>0.05 for any of the factors, it indicates that the factor is not significant.

- P value for all interaction effects > 0.05, hence insignificant.

Balanced ANOVA

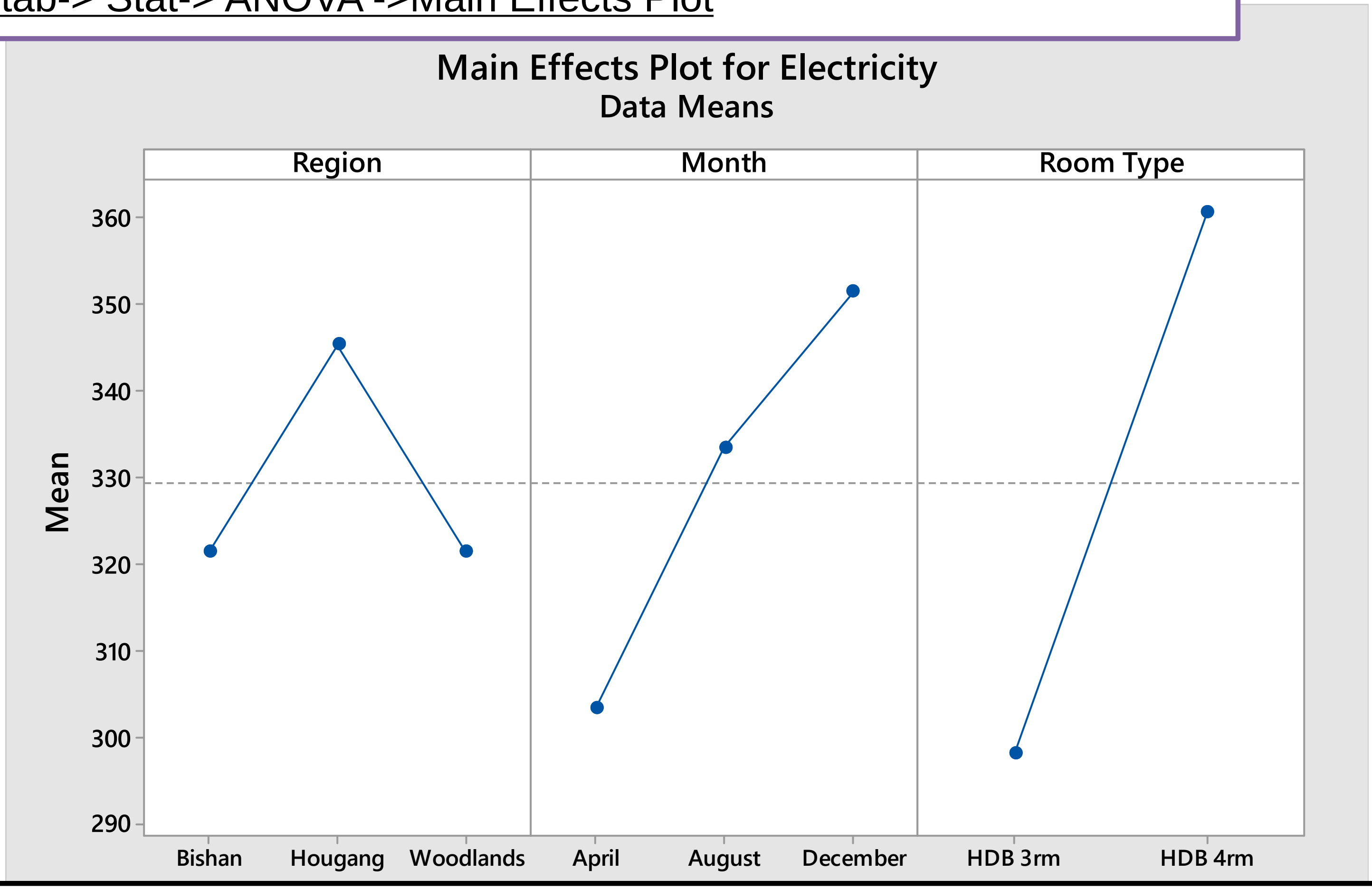
- Analyses all 3 factors *at the same time*
- Includes the interactions among the factors
- Takes into account the blocking effect (blocking out the effects of different factors) into the analysis
- Is more sensitive in identifying the significant factors compared with One-way ANOVA.

Residual Plot – Balanced ANOVA



Main Effects Plot (ANOVA)

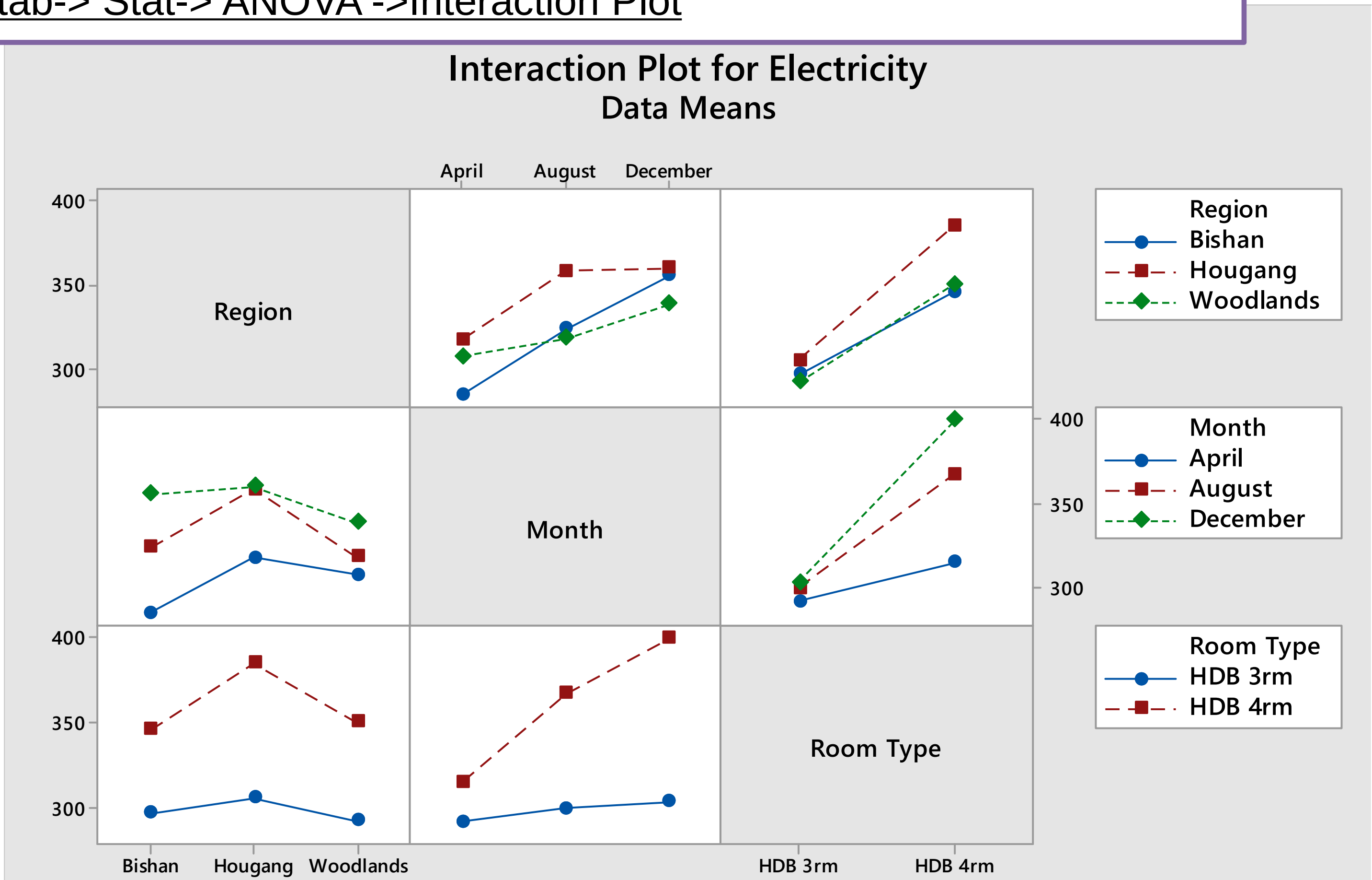
Minitab-> Stat-> ANOVA ->Main Effects Plot



- The differences between the response mean of each factor level (individual points) and the grand mean (horizontal line) indicate the effects.
- Main effect “Room Type” has the steepest gradient, hence is the most significant effect.

Interaction Plot (ANOVA)

Minitab-> Stat-> ANOVA ->Interaction Plot



- Interaction Plot creates a single interaction plot for two factors, or a matrix of interaction plots for three to nine factors.
- Among all Interaction effects, "Month*Room Type" exhibited the most non-parallel lines. To ascertain whether this interaction effect is a significant one, we have to refer to P-value in Balance ANOVA (p-value < 0.05, then significant).

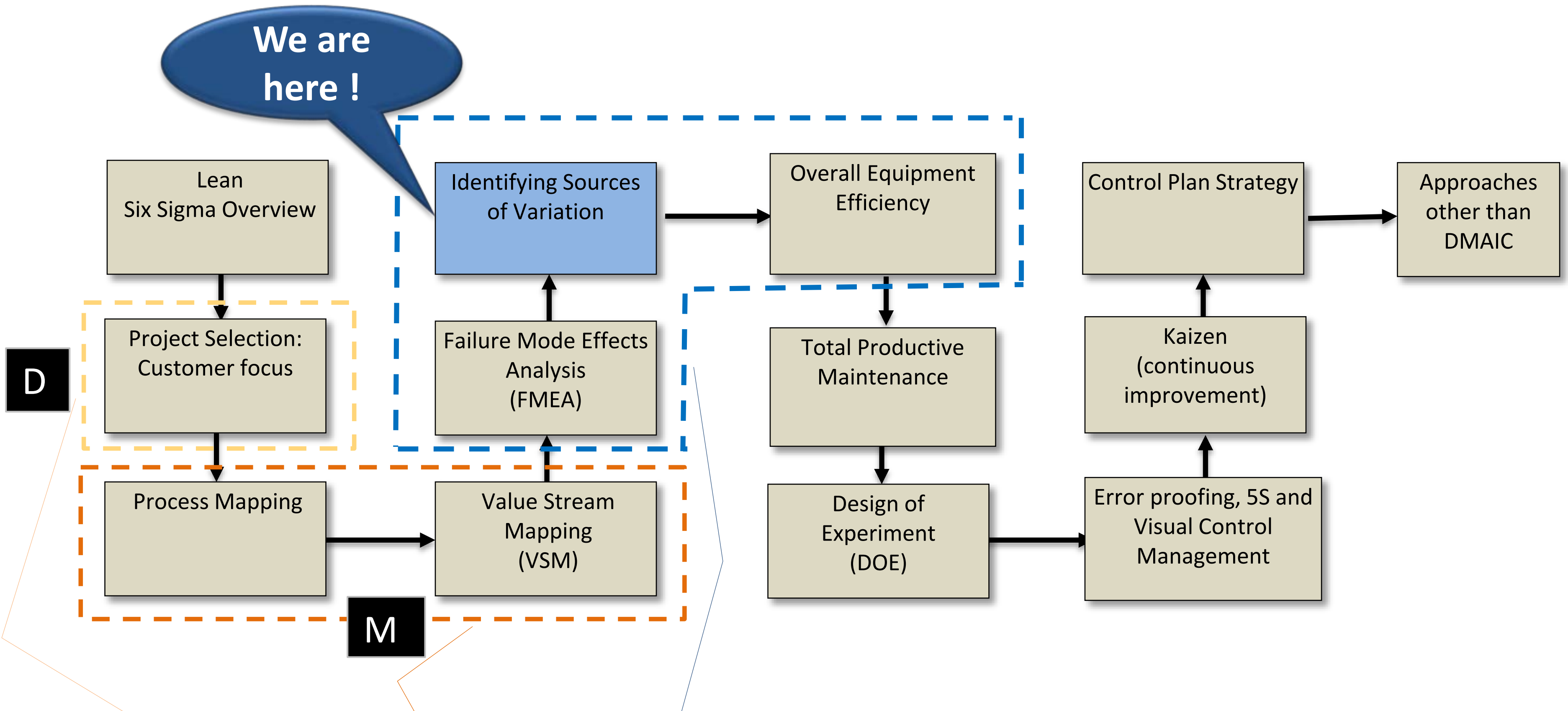
Summary of Findings

- Various plots and One-way ANOVA show that the “Room Type” is the sources of variations, i.e., factors significantly affecting the “Electricity Consumption / kWh” in the Open Electricity Market (OEM).
- Based on the Balanced ANOVA, 2 factors, “Month” and “Room Type”, significantly affecting the “Electricity Consumption / kWh” in the Open Electricity Market (OEM).
- You can proceed to the next Phase of the Six Sigma project to take actions to “Improve” on the 3 factors identified in today’s “Analyze” phase to improve the “Electricity Consumption / kWh” in the Open Electricity Market (OEM).
- Compared with One-way ANOVA, Balanced ANOVA analyses all factors and their interactions at the same time. It is more sensitive in identifying the significant factors compared with One-way ANOVA.

Learning Objectives

- Recognize the Lean Six Sigma Analyze Phase
- Apply tools used in Analyze Phase
  - Multi-Vari Charts
  - One-way ANOVA, Balanced-ANOVA
- Identify Possible Sources of Variation

Overview of E326 Lean Manufacturing and Six Sigma



Identify improvement opportunity
Define-M-A-I-C

Establish project baseline and measurable
D-**Measure**-A-I-C

Extract useful information and identify variation in data
D-M-**Analyze**-I-C