Part B Question

11. Briefly explain Response Surface Methodology.

The Response Surface Methodology, sometimes known as RSM, is a statistical approach that can be utilized in the process of designing and optimizing systems or procedures. In this process, mathematical models are constructed to reflect the relationship between the variables that are fed into the system and the response that is of interest to the researcher. Through a series of trials, Response Surface Methodology (RSM) can assist in determining the ideal values of the input variables that will produce the intended response. This technique utilizes statistical methods such as regression analysis and analysis of variance (ANOVA) to estimate the model parameters and evaluate the relevance of the input variables in order to determine an appropriate course of action. RSM is an effective strategy for the optimization and enhancement of processes since it helps maximize the intended outcomes while simultaneously reducing the number of trials that need to be conducted.

12. Explain various approaches that can be taken when designing response surfaces?

When it comes to the development of response surfaces using the Response Surface Methodology (RSM), there are a few options that can be pursued. Here are three methods that are frequently used:

CCD stands for central composite design and is an approach that is frequently utilized in RSM. Choosing a group of experimental points, such as factorial points, axial points, and center points, is a necessary step in the process. The factorial points are responsible for capturing the primary effects that are caused by the input variables, the axial points are responsible for estimating the effects of curvature, and the center points are responsible for assisting in the evaluation of the model's inadequacy. The use of CCD is efficient in terms of the number of experimental runs that are needed because it enables the fitting of second-order models.

The Box-Behnken Design is a method to design that is appropriate when the factors of interest have a moderate to high number of levels. This design makes use of a three-level factorial arrangement with additional center points. When compared to CCD, the Box-Behnken Design has the benefit of requiring fewer experimental runs while still enabling the fitting of second-order models. This is an advantage of the Box-Behnken Design.

The use of a factorial design is possible in circumstances in which one is primarily interested in the primary effects that the input variables have on the outcome of the experiment. The selection of a collection of experimental points that encompass the many possible combinations of level settings for each factor is required for this method. Although factorial designs are relatively straightforward and effective, they might not take into account the effects of curvature or the interactions between components.

13. How is a central composite design different from other response surface designs?

A central composite design (CCD) is different from other response surface designs in several ways:

Inclusion of Axial Points: CCD adds additional points termed axial points, which are situated at a distance of from the center of the design space along each factor axis. Axial points are described as being "axial" since they are located along each factor axis. These axial points make it possible to estimate the curvature or quadratic effects of the variables that are fed into the model. When opposed to designs that only take into consideration linear effects, CCD provides a more realistic depiction of the response surface since it takes into account both linear and quadratic effects thanks to the inclusion of axial points.

Evaluation of the model's unsuitability: CCD consists of center points that are reproduced a number of times in order to conduct an evaluation of the model's unsuitability. The term "lack of fit" refers to the disparity that exists between the model that was fitted and the data that was seen at the center locations. CCD gives a measure of how well the response surface model fits the data by measuring the lack of fit. This helps evaluate whether higher-order terms or interactions should be added in the model. CCD also provides a measure of how well the model matches the data.

Flexibility in selecting: The value of in CCD can be modified based on the desired emphasis on curvature effects. This allows for greater flexibility in the design process. Researchers are able to investigate a broader range of curvature and adjust the balance between the complexity of the model and its accuracy by picking alternative values for the parameter.

Utilization of experimental runs in an effective manner: CCD was developed to strike a balance between the number of runs that are conducted in an experiment and the capacity to fit second-order models. In comparison to full factorial designs, it calls for a smaller number of experimental runs, but it nevertheless generates estimates of the model parameters that may be relied upon.

14. What is ANOVA? What are the assumptions of ANOVA?

The abbreviation for "Analysis of Variance" is "ANOVA." It is a method of statistical analysis that compares the means of two or more groups or treatments in order to determine the degree of difference between them. The analysis of variance (ANOVA) breaks down the total variation in the data into its component parts, or sources, so that the relative importance of each of these sources may be determined.

- Independence: It is presumed that the observations made within each group or treatment did not influence or affect one another in any way. This indicates that the significance of one observation does not have an effect on the significance of another observation.
- Normality: The data are supposed to follow a normal distribution within each group or therapy, which is what is meant by "normality." The analysis of variance (ANOVA) relies on the normalcy assumption in order to perform hypothesis tests and create confidence intervals, hence this assumption is very significant.

- Homogeneity: The assumption that the variance of the response variable is the same across all groups or treatments is known as homogeneity of variances, which is often referred to as homoscedasticity. To put it another way, the distribution of the data is the same throughout all of the distinct categories.
- Equality: The analysis of variance (ANOVA) presupposes that the sample sizes for each group or treatment are equivalent. This is referred to as a "balanced design." When there is a disparity in the sizes of the samples being analyzed, it is possible that the ANOVA model will require some adjustments.

15. What is the F-test in ANOVA?

The F-test is a statistical test that is used in Analysis of Variance (ANOVA) to assess if there are significant differences between the means of two or more groups or treatments. The test can be used to compare up to four different sets of data. It examines the degree to which the variability between the group means and the variability within each group are comparable.

The ratio of the mean square between groups to the mean square within groups is what the F-test uses to determine significance. The mean square can be calculated by dividing the sum of squares (SS) by the degrees of freedom (df) to get the mean square.

16. What does MANOVA stand for, and what is its primary purpose?

The abbreviation for "Multivariate Analysis of Variance" is "MANOV." The major objective of this method is to compare the means of numerous dependent variables concurrently across two or more groups or treatments in order to determine whether or not there is a significant difference between them. To put it another way, MANOVA makes it possible to conduct an analysis of several response variables or outcome measures at the same time.

When compared to the practice of doing separate basic ANOVAs for each dependent variable, MANOVA has a number of advantages. It is possible to conduct an exhaustive investigation of the multivariate data structure, it takes into account the correlations that exist between the dependent variables, it boosts the statistical power by making use of the information that is common to several variables, and it accounts for the interdependencies that exist between the variables.

Part C Question

17. Explain briefly about the Box-Behnken design for three factor Experiment.

Box-Behnken designs (BBD): Box-Behnken designs usually have fewer design points than central composite designs, thus, they are less expensive to run with the same number of factors. They can efficiently estimate the first- and second-order coefficients; however, they can't include runs from a factorial experiment. Box-Behnken designs always have 3 levels per factor, unlike central composite designs which can have up to 5. Also unlike central composite designs, Box-Behnken designs never include runs where all factors are at their extreme setting, such as all of the low settings.

This section will show how to develop the smallest design possible taking into account three different factors. This gives us a well-balanced incomplete block design with three different treatments and three different blocks.

Block	*Treatments*		
	1	2	3
1	X	X	
2	X		X
3		X	X

The three treatments are considered as three factors A, B, C in the experiment. The two crosses (Xs) in each block are replaced by the two columns of the 22 factorial design and a column of zeros are inserted in the places where the cross do not appear. This procedure is repeated for the next two blocks and some centre points are added resulting in the Box-Behnken design for k = 3 which is given in Table below. Similarly, designs for k = 4 and k = 5 can be obtained. The advantage of this design is that each factor requires only three levels.

TABLE Box–Behnken design for $k = 3$

Run	*A*	*B*	*C*
1	–1	–1	0
2	–1	1	0
3	1	–1	0
4	1	1	0
5	–1	0	–1
6	–1	0	1
7	1	0	–1
8	1	0	1
9	0	–1	–1
10	0	–1	1
11	0	1	–1
12	0	1	1
13	0	0	0
14	0	0	0
15	0	0	0

18. The table below shows the height, x, in inches and the pulse rate, y, per minute, for 6 people. Find the correlation coefficient and interpret your result.

x	5	7	4	15	12	9
y	8	9	12	26	16	13

Solution:

Calculate by yourself

$\Sigma X_i = 52$ $\Sigma X^2 = 540$ $\Sigma XY = 850$

$\Sigma Y_i = 84$ $\Sigma Y^2 = 1390$

$$n\sum(xy) - \left(\sum x\right)\left(\sum y\right)$$

$= 732$

$$\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y\right)^2}$$

≈ 829.46

divide to get $r \approx 732 / 829.46$

$\approx .88$

There is a strong positive correlation since r is very close to 1 (positive)

19. Suppose the National Transportation Safety Board (NTSB) wants to examine the safety of compact cars, midsize cars, and full-size cars. It collects a sample of three for each of the treatments (car types). Using the hypothetical data provided below; perform an ANOVA test to determine whether the mean pressure applied to the driver's head during a crash is equal for each type of cars. Use a=5%.

Compact Cars	MidSize cars	Full size cars
7	4	6
5	8	4
6	6	7

Calculate the overall mean

Compact Cars: (7 + 5 + 6) / 3 = 6; MidSize Cars: (4 + 8 + 6) / 3 = 6; Full-Size Cars: (6 + 4 + 7) / 3 = 5.67

Grand Mean: (6 + 6 + 5.67) / 3 = 5.89

Calculate the sum of squares between treatments (SSB):

= (3 * (6 - 5.89)^2) + (3 * (6 - 5.89)^2) + (3 * (5.67 - 5.89)^2)

= 0.171 + 0.171 + 0.209

= 0.551

Calculate the degrees of freedom between treatments (dfB) (check formula from PPT)

= 3 – 1 = 2

mean squares between treatments (MSB): (check formula from PPT)

= 0.551 / 2 = 0.276

squares within treatments (SSW): (check formula from PPT) = 12.187

degrees of freedom within treatments (dfW) (check formula from PPT) = 6

mean squares within treatments (MSW) (check formula from PPT) = 12.187 / 6 = 2.031

F-statistic: F = 0.276 / 2.031 = 0.136

critical value for F at $\alpha = 0.05$ and (dfB, dfW): From the F- table and $\alpha = 0.05$ critical value is 5.14 (approximately)

Compare the F-statistic with the critical value 0.136 < 5.14, reject the null hypothesis

There is not enough evidence to draw the conclusion, based on the results of the ANOVA test, that the mean pressure that is applied to the driver's head during a collision is substantially different amongst the three different types of cars (compact cars, midsize cars, and full-size cars) at the 5% significance level.

20. Researchers want to test a new anti-anxiety medication. They split participants into three conditions (0mg, 50mg, and 100mg), then ask them to rate their anxiety level on a scale of 1-10. Are there any differences between the three conditions using alpha = 0.05 and F table= 5.1433

0 mg	50 mg	100 mg
9	7	4
8	6	3
7	6	2

Procedural steps for One-Way ANOVA

1. Define Null and Alternative Hypotheses
2. State Alpha
3. Calculate Degrees of Freedom
4. State Decision Rule
5. Calculate Test Statistic
6. State Results
7. State Conclusion

1. Define Null and Alternative Hypotheses

$$H_0; \mu_{0mg} = \mu_{50mg} = \mu_{100mg}$$

$$H_1; not\ all\ \mu's\ are\ equal$$

Steps for One-Way ANOVA

2. State Alpha

Alpha = 0.05

3. Calculate Degrees of Freedom

Now we calculate the degrees of freedom using N = 21, n = 7, and a = 3.

You should already recognize N and n. "a" refers to the number of groups ("levels") you're dealing with:

$$df_{Between} = a - 1 = 3 - 1 = \mathbf{2}$$

$$df_{Within} = N - a = 21 - 3 = \mathbf{18}$$

$$df_{Total} = N - 1 = 21 - 1 = \mathbf{20}$$

Steps for One-Way ANOVA

4. State Decision Rule

To look up the critical value, we need to use two different degrees of freedom.

$$df_{Between} = a - 1 = 3 - 1 = \mathbf{2}$$

$$df_{Within} = N - a = 21 - 3 = \mathbf{18}$$

We now head to the F-table and look up the critical value using (2, 18) and alpha = 0.05. This results in a critical value of 3.5546, so our decision rule is:

If F is greater than 3.5546, reject the null hypothesis.

Check F table colum 2 and row 16

Steps for One-Way ANOVA

5. Calculate Test Statistic

$SS_{between}$ SS_{within} SS_{total}

$$SS_{between} = \frac{57^2 + 47^2 + 21^2}{7} - \frac{125^2}{21} = 98.67$$

$$\text{0mg Group: } 9 + 8 + 7 + 8 + 8 + 9 + 8 = 57$$
$$\text{50mg Group: } 7 + 6 + 6 + 7 + 8 + 7 + 6 = 47$$
$$\text{100mg Group: } 4 + 3 + 2 + 3 + 4 + 3 + 2 = 21$$

Steps for One-Way ANOVA

5. Calculate Test Statistic

$$SS_{within} = 853 - \frac{57^2 + 47^2 + 21^2}{7} = 10.29$$

$$\sum Y^2 = 9^2 + 8^2 + 7^2 + 8^2 + 8^2 + 9^2 + 8^2 + 7^2 + 6^2 + 6^2 + 7^2 + 8^2 + 7^2 + 6^2 + 4^2 + 3^2 + 2^2 + 3^2 + 4^2 + 3^2 + 2^2 = \mathbf{853}$$

Figure 7.

$$SS_{total} = 853 - \frac{125^2}{21} = 108.95$$

	SS	df	MS	F
Between	98.67	2		
Within	10.29	18		
Total	108.95	20		

$$MS_{between} = \frac{98.67}{2} = 49.34 \quad \Bigg| \quad MS_{within} = \frac{10.29}{18} = 0.57$$

Figure 10.

And finally, we can calculate our F:

$$F = \frac{MS_{between}}{MS_{within}} = \frac{49.34}{0.57} = 86.56$$

	SS	df	MS	F
Between	98.67	2	49.34	86.56
Within	10.29	18	0.57	
Total	108.95	20		

6. State Results
F = 86.56
Result: Reject the null hypothesis.
7. State Conclusion
The three conditions differed significantly on anxiety level, $F(2, 18) = 86.56$, $p < 0.05$.