Marketing Research using Analysis of Variance (ANOVA)

Sep 02, 2023.

One-Way ANOVA

1. One-way ANOVA (Analysis of Variance) is a statistical method used to test the equality of means between two or more groups.
2. It is used to determine if there is a significant difference in the mean of the dependent variable (response variable) between different levels (categories) of an independent variable (explanatory variable).
3. One-way ANOVA is used to compare the means of multiple groups to see if any group is significantly different from the others, and it is a type of hypothesis testing.

Marketing Applications of One-Way ANOVA

One-way ANOVA has several applications in marketing research, including:

- Market segmentation: One-way ANOVA can be used to segment customers based on their demographics, behavior, or purchase history.
- Customer satisfaction: One-way ANOVA can be used to compare the satisfaction levels of customers in different regions or with different demographics.
- Product comparison: One-way ANOVA can be used to compare the sales of different product lines to see if there are significant differences in their means.
- Advertising effectiveness: One-way ANOVA can be used to compare the effectiveness of different advertising campaigns.
- Brand preference: One-way ANOVA can be used to compare the brand preferences of different customer segments.
- Price sensitivity: One-way ANOVA can be used to determine if there is a significant difference in the price sensitivity of customers in different regions or with different demographics.

Understanding one-way ANOVA using mtcars data

The `mtcars` dataset provides data extracted from the 1974 Motor Trend US magazine, presenting specifications and performance tests for various 1973-74 car models.

Overview of the mtcars dataset:

1. **Number of Rows**: 32 (representing 32 car models)
2. **Number of Columns**: 11 (each representing a specific attribute or measurement)

Variables in the dataset: 1. `mpg`: Miles/(US) gallon (fuel efficiency) 2. `cyl`: Number of cylinders 3. `disp`: Displacement (cu.in.) 4. `hp`: Gross horsepower 5. `drat`: Rear axle ratio 6. `wt`: Weight (in 1000 lbs.) 7. `qsec`: 1/4 mile time (in seconds) 8. `vs`: Engine (0 = V-shaped, 1 = straight) 9. `am`: Transmission (0 = automatic, 1 = manual) 10. `gear`: Number of forward gears 11. `carb`: Number of carburetors

Sample cars in the dataset: Mazda RX4, Hornet 4 Drive, Duster 360, Merc 240D, and so on.

Motor Trend Magazine. (1974). Motor Trend 1974 Car Road Tests. Los Angeles: Petersen Publishing Co.

Understanding one-way ANOVA using mtcars data

1. Recall that One-way ANOVA is a statistical technique used to ascertain if there are notable differences in the means of three or more independent groups. In essence, it lets you compare the averages of these groups to see if they're really different or if the observed differences could just be due to random chance.

2. The `mtcars` dataset comprises various specifications and performance metrics for 32 car models from the 1970s. For this illustration, let's take a hypothetical scenario where we wish to analyze whether cars' fuel efficiency (measured in miles per gallon or mpg) varies based on the number of gears a car possesses. The dataset categorizes cars as having 3, 4, or 5 gears.

One-Way ANOVA Process

1. **Setting the Hypotheses**:

   - **Null Hypothesis (H0)**: The average fuel efficiency (mpg) remains constant across cars, irrespective of whether they have 3, 4, or 5 gears.
   - **Alternative Hypothesis (HA)**: The average fuel efficiency varies for at least one of the gear categories.

2. **Understanding Variance**:

    - **Between-group Variance**: This reflects how much each group's average (mean) deviates from the overall average of all groups. If the between-group variance is substantially larger than the within-group variance, it could indicate that the gear types impact fuel efficiency.
    - **Within-group Variance**: Represents the variability of individual observations from their respective group's average.

3. **Computing the F-Statistic**:

    - The F-statistic is a crucial value in ANOVA. It's calculated by taking the ratio of between-group variance to the within-group variance. A high F-statistic suggests that the groups (in this case, gear types) are distinct from each other in terms of their averages more than one would expect by random chance.

Once you compute the F-statistic from your data, you refer to the F-distribution (a theoretical probability distribution) to determine the p-value associated with that F-statistic, given the degrees of freedom of the test. The degrees of freedom depend on the number of groups and the number of observations in the groups.

4. **p-value**:

The p-value is a probability that provides context for the F-statistic. Specifically, it answers the question: "If there were no real differences between the groups (i.e., the null hypothesis is true), how likely would I observe an F-statistic as extreme as, or more extreme than, the one I calculated from my data?"

Low p-value (typically 0.05): This suggests that the observed data (and the resultant F-statistic) is unlikely under the assumption that the null hypothesis is true. In other words, if you get a low p-value, it's evidence against the null hypothesis, and you might decide to reject the null hypothesis in favor of the alternative hypothesis, which posits that there's a difference between the groups.

High p-value: This suggests that the observed data (and the F-statistic) is plausible under the null hypothesis. You would not reject the null hypothesis, implying you haven't found strong enough evidence to say that the group means are different.

5. **Assessing the Result**:

    - To determine if the observed differences (or lack thereof) are statistically significant, the F-statistic is compared to a threshold value from the F-distribution. If the F-statistic is sufficiently high, we reject the null hypothesis, implying there's a significant difference in mpg across the gear types.

Performing a one-way ANOVA on the `mtcars` dataset might reveal that the number of gears in cars does (or does not) significantly influence their average fuel efficiency.

If differences are found, further analyses would be needed to pinpoint where these differences lie.

6. **Post-Hoc Analysis**:

    - If the ANOVA reveals a significant result, subsequent post-hoc tests can pinpoint which specific gear types exhibit different mpg averages.

The goal of ANOVA is to gauge if the groups being compared have different averages. It doesn't specify which groups are distinct, which is why follow-up analyses are essential when differences are identified.

Running One-Way ANOVA in SPSS

1. You can run a one-way ANOVA (analysis of variance) in SPSS by following these steps:

- Open your data file in SPSS.
- Select "Analyze" from the top menu, then select "Compare Means," and then select "One-Way ANOVA."
- In the "One-Way ANOVA" window, select the dependent variable that you want to analyze by dragging it to the "Dependent List" box on the right-hand side of the window.
- Select the independent variable that you want to use by dragging it to the "Factor" box on the left-hand side of the window.
- If you want to run post-hoc tests to compare specific pairs of means, click on the "Post Hoc" button and select the tests you want to run (e.g., Bonferroni, Tukey, or LSD).
- If you want to test for violations of the assumptions of normality or homogeneity of variance, click on the "Options" button and select the tests you want to run (e.g., Shapiro-Wilk, Levene's test).
- Click "OK" to run the analysis.

2. SPSS will generate a table with the results of the ANOVA, including the F-statistic, the degrees of freedom, the p-value, and the effect size (partial eta-squared).
3. If you ran post-hoc tests, SPSS will also generate a table with the results of those tests. If you ran tests for normality or homogeneity of variance, SPSS will also provide those results in the output window.
4. You can also generate graphical output, such as a boxplot or a mean plot, by selecting the appropriate options in the "One-Way ANOVA" window or in the "Chart Editor" window.

Two-Way ANOVA

1. Two-way ANOVA (Analysis of Variance) is a statistical method used to analyze the effects of two independent variables on a dependent variable.
2. It is used to determine if there is a significant interaction between the two independent variables on the dependent variable.
3. Two-way ANOVA is an extension of one-way ANOVA and allows for the examination of the effects of two independent variables on the dependent variable.

Marketing Applications of Two-Way ANOVA

Two-way ANOVA has several applications in marketing research, including:

- Customer segmentation: Two-way ANOVA can be used to segment customers based on their demographics and behavior to understand their preferences and buying habits.
- Product design: Two-way ANOVA can be used to test the effects of different product features on customer satisfaction to inform product design decisions.
- Advertising effectiveness: Two-way ANOVA can be used to test the interaction between different advertising channels (e.g. TV, social media) and target demographics to determine the most effective advertising strategies.
- Price sensitivity: Two-way ANOVA can be used to test the interaction between price and product features to understand the most price-sensitive customers.
- Promotion effectiveness: Two-way ANOVA can be used to test the interaction between different promotions and target demographics to determine the most effective promotions for different customer segments.

Two-Way ANOVA and Interactions

1. In a two-way ANOVA, the experimenter manipulates two independent variables, each with multiple levels, to see if there is an interaction between the variables.
2. The dependent variable is then measured and the means of the groups formed by the combinations of the two independent variables are compared.

3. This allows the experimenter to determine if there is a significant interaction between the two independent variables and if the effect of one independent variable on the dependent variable depends on the level of the other independent variable.
4. The main difference between two-way ANOVA with and without interactions is the way that the effects of the independent variables are analyzed.
5. Two-way ANOVA without interactions:

- In this type of two-way ANOVA, the main effects of each independent variable are analyzed independently, without considering the possible interaction between the two independent variables.
- This is also known as a main effects model.
- The main purpose of this type of analysis is to determine if there is a significant effect of each independent variable on the dependent variable.

6. Two-way ANOVA with interactions:

- In this type of two-way ANOVA, the possible interaction between the two independent variables is taken into account.
- An interaction between independent variables means that the effect of one independent variable on the dependent variable depends on the level of the other independent variable.
- The purpose of this type of analysis is to determine if there is a significant interaction between the independent variables and if the effect of one independent variable on the dependent variable depends on the level of the other independent variable.

Overall, two-way ANOVA is a useful tool for marketers to understand the relationships between multiple variables and make data-driven decisions about their target audience and marketing strategies.

Two-Way ANOVA in SPSS

1. You can run a two-way ANOVA (analysis of variance) in SPSS by following these steps:

- Open your data file in SPSS.
- Select "Analyze" from the top menu, then select "General Linear Model," and then select "Univariate."
- In the "Univariate" window, select the dependent variable that you want to analyze by dragging it to the "Dependent Variable" box on the right-hand side of the window.
- Select the two independent variables that you want to use by dragging them to the "Fixed Factor(s)" box on the left-hand side of the window.
- If you want to include interactions between the two independent variables, click on the "Model" button and select the interaction term (e.g., "Variable 1 * Variable 2").

- If you want to test for violations of the assumptions of normality or homogeneity of variance, click on the “Options” button and select the tests you want to run (e.g., Shapiro-Wilk, Levene’s test).
- Click “OK” to run the analysis.

2. SPSS will generate a table with the results of the ANOVA, including the F-statistic, the degrees of freedom, the p-value, and the effect size (partial eta-squared) for each main effect and interaction effect.
3. Additionally, SPSS will generate tables with the means and standard deviations for each level of the independent variables, as well as any post-hoc tests you may have specified.
4. You can also generate graphical output, such as a mean plot or an interaction plot, by selecting the appropriate options in the “Univariate” window or in the “Chart Editor” window.