Assignment 6 - Means comparisons and experimentation

INSTRUCTIONAL DETAILS

ANOVA (Analysis of Variance) is a statistical test used to compare the mean of different groups or samples. It is used to determine whether there is a significant difference between the means of two or more groups.

To perform ANOVA in Python, you can use the f_oneway() function from the scipy.stats module. This function takes in the data for each group as separate arguments, and returns the F-statistic and p-value for the test.

Here is an example of how to perform ANOVA in Python:

```
from scipy.stats import f_oneway

group1 = [1, 2, 3, 4, 5]
group2 = [2, 3, 4, 5, 6]
group3 = [3, 4, 5, 6, 7]

statistic, pvalue = f_oneway(group1, group2, group3)

print('F-statistic:', statistic)
print('p-value:', pvalue)
```

If the p-value is less than the significance level (typically 0.05), then we can conclude that there is a significant difference between the means of the groups. Otherwise, we cannot reject the null hypothesis that there is no difference between the means.

The output of an ANOVA test typically includes several statistics, including the F-value, the p-value, and the degrees of freedom. Here is a brief overview of what these statistics mean and how you can interpret them:

F-value: The F-value is a measure of the ratio of the variance between the groups to the variance within the groups. A larger F-value indicates that there is a greater difference between the means of the groups, and therefore a greater likelihood that the differences are significant.

p-value: The p-value is a measure of the probability that the observed differences between the groups could have occurred by chance. A small p-value (typically less than 0.05) indicates that it is unlikely that the differences occurred by chance and that there is a significant difference between the means of the groups.

Degrees of freedom: The degrees of freedom are used to calculate the F-value and are based on the number of samples and the number of variables being tested.

To interpret the output of an ANOVA test in Python, you can use the f_oneway() function from the scipy.stats module. This function takes in the data for each group and returns the F-value, p-value, and degrees of freedom. You can then use these statistics to determine whether there is a significant difference between the means of the groups.

For example, suppose you have data for three groups and you want to test whether there is a significant difference between the means of the groups. You could use the following code to perform an ANOVA test:

```
from scipy.stats import f_oneway

# data for each group
```

```
group1 = [1, 2, 3, 4]
group2 = [2, 3, 4, 5]
group3 = [3, 4, 5, 6]

# perform ANOVA test
f_value, p_value, df = f_oneway(group1, group2, group3)

# interpret results
if p_value < 0.05:
    print("There is a significant difference between the means of the groups.")
else:
    print("There is not a significant difference between the means of the groups.")
```

This code would perform an ANOVA test on the data and print the results indicating whether there is a significant difference between the means of the groups.

BACKGROUND

ANOVA (analysis of variance) is a statistical technique used to compare the means of different groups or treatments. It is a useful tool for businesses because it allows them to understand whether there are significant differences between the means of different groups and identify the factors that may be contributing to these differences.

For example, a business might use ANOVA to compare the means of different customer segments in order to understand whether there are significant differences in their purchasing behavior. This might include factors such as the types of products they purchase, the frequency of their purchases, and the amount they spend. Understanding these differences can inform marketing and sales strategies and help the business target its efforts more effectively.

ANOVA can also be used to compare the means of different treatments or interventions in order to understand their effectiveness. For example, a business might use ANOVA to compare the means of sales before and after implementing a new marketing campaign in order to understand the impact of the campaign on sales.

Overall, ANOVA is an important tool for businesses because it helps them understand whether there are significant differences between the means of different groups and identify the factors that may be contributing to these differences. This can inform decision-making and help businesses achieve their goals.

RESEARCH QUESTION

The Federal Communications Commission's (FCC) E-Rate program is a program that provides discounts on telecommunications services to schools and libraries. Data analysis can be a valuable tool in customer segmentation for the E-Rate program, as it can help the FCC identify trends and patterns in the data and better understand the needs of different customer groups. Here are a few ways in which data analysis can be used to segment customers in the E-Rate program:

Demographic data: Data analysis can be used to segment customers based on demographic data, such as the location of the school or library, the type of institution (e.g. public or private), and the size of the institution.

Service needs: Data analysis can be used to segment customers based on their service needs, such as the type of telecommunications services they require, the amount of bandwidth they need, and their budget for those services.

Past usage data: Data analysis can be used to segment customers based on their past usage of telecommunications services, such as the amount of data they have used in the past and the number of devices they have connected to the network.

Future needs: Data analysis can be used to segment customers based on their future needs, such as their plans for expanding their telecommunications services or adopting new technologies.

Are there statistically significant or experimentally important differences based on the “status” of an application for the following dependent variables: enrollment, nslp_count, nslp_percent, and request_amount?

Same question, but change the independent variable to locale.

REQUIREMENTS FOR SUBMISSION

See “Assignment 1 - Descriptives” for a detailed list of submission requirements.

FORMATTING

See “Assignment 1 - Descriptives” for a detailed list of assignment formatting guidelines. Also, assignment formatting guidelines can be found in the course document cache.

DATASET DETAILS (all data sets can be found here or here)

segmentation.csv

DATASET FIELDS

application_number
funding_year
state
form_version
window_status
status
category
application_name
address1
city
zipcode
entity_number
fcc_number
applicant_type
enrollment
nslp_count
nslp_percent
locale
request_amount
congress_district