Mini Project 4 - One Way Anova

Rishi Laddha

Section 2

Question 1 - State the assumptions of one-way ANOVA. Comment if these assumptions seem reasonable in the context of the PDR data provided here.

Answer -

The following assumptions hold for one-way ANOVA -

1. *Samples with each group are taken from a normally distributed population.*
2. *Samples are taken independently of each other.*
3. *Homogeneity of variances must be observed.*

Question 2 - Implementation

Code:

1. **Function** -

```
function []=hypothesis_checker(hmatrix);

a = hmatrix;
a2=a.*a;
[row, column] = size(a2);

SSW = sum(sum(a2,2))-sum(sum(a,2).^2)/column;
dof1 = (row-1);


SSB = sum((sum(a,2).^2)/column)-sum((sum(sum(a)).^2)/(row*column));
dof2 = (row*(column-1));

MSB=SSB/dof1;
MSW=SSW/dof2;

Fcal = MSB/MSW;
Ftab = finv(0.99,dof1,dof2);

if Ftab>Fcal
    if sum(sum(a))/(column*row)>5.0;
        disp("Severe Issue and universally relevant")
    else
        disp("Failed to reject hypothesis.")
    end
 else Ftab<Fcal
    disp("Hypothesis rejected.")
end

end
```

2. **Function Call** -

```
clc
clear


c=[3 2 9 8 9 9;5 9 10 5 8 9;6 7 10 5 7 8;8 9 9 8 2 8;3 8 7 10 10 4;2 7 9 10 6 7];
n=[3 2 9 8 9 7;5 4 4 5 3 2;5 2 4 5 1 2;3 1 1 2 6 2;3 8 7 10 10 4;3 1 9 8 6 7];
o=[9 9 10 8 7 8;5 4 4 5 3 2;4 6 7 2 8 9;3 1 5 2 6 2;7 4 5 1 2 3;3 2 9 8 6 7];
p=[6 9 5 5 7 6;6 4 6 5 7 8;4 6 7 2 8 9;4 6 6 1 8 9;10 7 8 1 5 6;3 2 9 8 6 7];

hypothesis_checker(p)
```

3. **Outputs** -

a. Community Participation -

```
dof1 =

     5


dof2 =

    30


TSS =

  213.5556


MSB =

    0.7778


MSW =

    6.9889


Fcal =

    0.1113


Ftab =

    3.6990

Severe Issue and universally relevant
```

b. Funding -

```
dof1 =

     5


dof2 =

    30


TSS =

  278.7500


MSB =
```

```
   20.1833


MSW =

    5.9278


Fcal =

    3.4049


Ftab =

    3.6990

Failed to reject hypothesis.
```

c. Land Ownership -

```
dof1 =

     5


dof2 =

    30


TSS =

  249.0000


MSB =

   24.3333


MSW =

    4.2444


Fcal =

    5.7330


Ftab =

    3.6990


ans =

  logical

   1

Hypothesis rejected.
```

d. Shortage of technical staff -

```
dof1 =

     5
```

```
dof2 =

    30


TSS =

   184


MSB =

    0.3333


MSW =

    6.0778


Fcal =

    0.0548


Ftab =

    3.6990

Severe Issue and universally relevant
```

Question 3 - F-distribution -

1. Explain how we calculate F_t for one way ANOVA.
   Answer - Calculation of F_t requires degrees of freedom of the dependent variables (observations within each group), degrees of freedom of the independent variable (groups themselves) and the level of significance, α. Then, the matlab function *finv()* or a F-table can be used to find out the F_t value for the given degrees of freedom and α value.
2. For the given problem, calculate F_t using either MATLAB pr from the F0distribution table corresponding to a level of significance of $\alpha = 0.01$.
   Answer -

   MATLAB code -

```
dof1 = 30; %(for SSW)
dof2 = 5;  %(for SSB)

Ftab = finv(0.99,dof2,dof1) %% 0.99 is the confidence level i.e. 99 percent confiedence
```

Output -

```
Ftab =

    3.6990
```

Question 5 - Hypothesis Testing -

1. Explain the concept of hypothesis testing. What are the possible outcomes of the ANOVA test?
   Answer - Hypothesis testing is used to use data from a random sample to make inferences about the population. This process rely on the use of various test statistics and distributions like the standard normal distribution, the X^2 distribution, the t distribution and the F distribution.

   Hypothesis testing involves a hypothesis about a parameter and a measure of reliability of that statement in terms of probability.

   The following steps are undertaken to perform hypothesis testing -

[Quote begins]

> 1. Specify H_0 and H_α and an acceptable level of significance α.
> 2. Define a sample based test statistic (eg. mean, standard deviation, etc.) and a rejection (or critical) region for H_0 that is most suitable for the experiment.
> 3. Collect the sample data and calculate the test statistic.
> 4. Make a decision to either reject or fail to reject H_0.
> 5. Interpret the result in the language of the problem at hand and provide an estimate of the error in the decision.
>
> (p.82, Chapter 6 - Statistical Experiments, Amrik Sen)Qu

[Quote ends]

In the ANOVA test, the following hypothesis are set -
$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_t$

H_α : one of the above equality is not satisfied

We reject H_0 when the calculated F is **greater** than the critical F. We fail the reject H_0 when the calculated F is **less** than the critical F.

Based on the values of calculated F and the critical F, we may reject or fail to reject H_0.

2. What is the significance of $\alpha = 0.01$? If $\alpha = 0.01$, does that mean that H_0 is true with probability 0.99?
   Answer - $\alpha = 0.01$ signifies that when H_0 **is <u>true</u>,** the probability of the decision given by the test being correct is $1 - \alpha$ i.e. the probability of the test being correct is 0.99 (when H_0 is not rejected). Similarly, when H_0 is true, the probability of the decision given by the test being incorrect (Type 1 error) is α i.e the probability of the test being incorrect is 0.01 (when H_0 is rejected)
3. If there are some issues which could be universally relevant and need immediate redressal across all groups, select and mention them.
   Answer - From output of the MATLAB code mentioned in the response to Question 2, **<u>Community Participation</u>** and **<u>Shortage of Technical Staff</u>** are universally relevant and need immediate redressal across all groups (with grand means of 7.1111 and 6 respectively
   (> 5)).