**Assignment 3**

1. Suppose that a random sample of n = 5 was selected from the vineyard properties for sale in Sonoma County, California, in each of three years. The following data are consistent with summary information on price per acre for disease-resistant grape vineyards in Sonoma County. Carry out an ANOVA to determine whether there is sufficient evidence to support the claim that the mean price per acre for vineyard land in Sonoma County was not the same for each of the three years considered. Test at the 0.05 level.

1996: 30000 34000 36000 38000 40000

1997: 30000 35000 37000 38000 40000

1998: 40000 41000 43000 44000 50000 Step 0:

Model : Yij = μ + Ai + eij , i = 1, 2, 3, …, j = 1, 2, 3, …

Ho : μ1996 = μ1997 = μ1998 H1 : Ho is not true.

Step 1:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Years | Observations | | | | | Mean  i | Âi | μ = Mean  (Total) |
| 1996 | 30000 | 34000 | 36000 | 38000 | 40000 | 3560  0 | -2800 | 38400 |
| 1997 | 30000 | 35000 | 37000 | 38000 | 40000 | 3600  0 | -2400 |  |
| 1998 | 40000 | 41000 | 43000 | 44000 | 50000 | 4360  0 | 5200 |  |

Step 2:

DF (Treatment) = 3 – 1 = 2

DF (Total) = 15 – 1 = 14

DF (Error) = 14 – 2 = 12

SS (Treatment) = Σ (Âi \* no. of observations) = 203200000

SS (Residual) = Σ (Yij – Meani)2 = 178400000

SS (Total) = SS (Years) + SS (Residual)

# = Σ (Yij – μ)2 = 381600000

Mean SS (Treatment) = SS(Treatment)/DF( Treatment) = 101600000

Mean SS (Residual / Error) = SS(Residual)/DF( Residual) = 14866666.7

F statistic = Mean SS (Treatment) / Mean SS (Residual / Error) = 6.83408072 F(0.05, 2,12) = 3.89

Conclusion:

Since Test Statistic, F > tabulated value of F

At α = 0.05 level, we do reject the null hypothesis and say that the mean price per acre was not the same in each year.

1. The following data on calcium content of wheat are consistent with summary quantities that appeared in the article “Mineral Contents of Cereal Grains as Affected by Storage and Insect Infestation” (Journal of Stored Products Research [1992]). Four different storage times were considered. Is there sufficient evidence to conclude that the mean calcium content is not the same for the four different storage times? Test the appropriate hypotheses at the 0.05 level.

|  |  |
| --- | --- |
| **Storage Time** | **Observations** |
| 0 months | 58.75 57.94 58.91 56.85 55.21 57.30 |
| 1 month | 58.87 56.43 56.51 57.67 59.75 58.48 |
| 2 months | 59.13 60.38 58.01 59.95 59.51 60.34 |
| 3 months | 62.32 58.76 60.03 59.36 59.61 61.95 |

Step 0:

Model : Yij = μ + Ai + eij , i = 1, 2, 3, …, j = 1, 2, 3, …

Ho : μ0 = μ1 = μ2 = μ3 H1 : Ho is not true.

Step 1:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Storage Time | Observations | |  |  |  |  | Meani | Âi |
| 0 month | 58.75 | 57.94 | 58.91 | 56.85 | 55.21 | 57.30 | 57.49333 | -1.3408 |
| 1 month | 58.87 | 56.43 | 56.51 | 57.67 | 59.75 | 58.48 | 57.9567 | -0.8825 |
| 2 months | 59.13 | 60.38 | 58.01 | 59.95 | 59.51 | 60.34 | 59.55333 | 0.7191  7 |
| 3 months | 62.32 | 58.76 | 60.03 | 59.36 | 59.61 | 61.95 | 60.33833 | 1.5041  7 |

Mean, μ = 58.83417

Step 2:

DF (Treatment) = 4 – 1 = 3

DF (Total) = 24 – 1 = 23

DF (Error) = 23 – 3 = 20

SS (Treatment) = Σ (Âi \* no. of observations) = 32.13815

SS (Residual) = Σ (Yij – Meani)2 = 32.901033

SS (Total) = Sum of Squares (Years) + Sum of Squares (Residual)

= Σ (Yij – μ)2 = 65.0391833

Mean SS (Treatment) = SS(Treatment)/DF(Treatment) = 10.7127166

Mean SS (Residual) = SS(Residual)/DF(Residual) = 1.64505

F statistic = Mean SS (Treatment)/Mean SS (Residual) = 6.51

F(0.05, 3,20) = 3.10

Conclusion:

Since, Test Statistic, F > Tabulated value of F, we reject the null hypothesis. Thus, we conclude that the mean calcium content is not the same for the four different storage times.

1. Use the data below, showing a summary of highway gas mileage for several observations, to decide if the average highway gas mileage is the same for midsize cars,

SUV’s, and pickup trucks. Test the appropriate hypotheses at the α = 0.05 level.

Hint: Var(X) =  ̅ , i = 1, 2, 3, …, n

𝑛

(Compare Residual Sum of Square with it).

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***n*** | ***Mean*** | ***Std. Dev.*** |
| ***Midsize*** | 31 | 25.8 | 2.56 |
| ***SUV’s*** | 31 | 22.68 | 3.67 |
| ***Pickups*** | 14 | 21.29 | 2.76 |

Step 0:

Model : Yij = µ + Ai + eij , i = 1, 2, 3, …, j = 1, 2, 3, … Ho : µM = µS = µP H1 : Ho is not true.

Step 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | n | Mean ( yഥi) | Standard Deviation ( Si ) | Ti |
| Midsize | 31 | 25.8 | 2.56 | 799.8 |
| SUV | 31 | 22.68 | 3.67 | 703.08 |
| Pickup | 14 | 21.29 | 2.76 | 298.06 |

DF (Treatment) = 3 – 1 = 2

DF (Total) = 76 – 1 = 75

DF (Error) = 75 – 2 = 73

G = ∑i ∑j yij = ∑i ni yiഥ = ∑i Ti

Cumulative Frequency, CF = G2/N

∑ yij2 = ni (Si2 + yiഥ2) Si2 = 1/ ni  ( ∑ yij2 - yiഥ2 )

G = (31\*25.8) + (31\*22.68) + (14\*21.29)

= 1800.94

CF = (1800.94)2/76

= 42676.116

∑ yij2 = 31\*(2.562 + 25.82) + 31\*(3.672+ 22.682) + 14\*(21.292 + 2.762) = 43653.7357

Total SS = ∑i ∑j ∑ yij2 – G2/N

Treatment SS = ∑i (Ti2 / ni) – CF

Error SS = Total SS – Treatment SS

Total SS = 977.6197

Treatment SS = 42926.3918 - 42676.116 = 250.2758

Error SS = 727.3439

Mean SS (Treatment) = Total SS / DF (treatment)

= 125.1379 Mean SS (Residual) = Error SS / DF (Error)

# = 9.996361

F statistic = Mean SS(Treatment) / Mean SS(Error) = 12.51834

F(0.05, 2, 73) = 3.122

Conclusion:

Since, Test Statistic, F > Tabulated value of F.

We reject the null hypothesis.

So, we conclude that the mean highway gas mileage is not the same for the three types of vehicles.

4. To examine the effects of pets and friends in stressful situations, researchers recruited 45 people to participate in an experiment. Fifteen of the subjects were randomly assigned to each of three groups to perform a stressful task alone (control group), with a good friend present, or with their dog present. Each subject’s mean heart rate during the task was recorded. Test the appropriate hypotheses at the α = 0.05 level to decide if the mean heart rate differs between the groups.

|  |  |  |  |
| --- | --- | --- | --- |
|  | n | Mean | Std. Dev. |
| Control | 15 | 82.52 | 9.24 |
| Pets | 15 | 73.48 | 9.97 |
| Friends | 15 | 91.325 | 8.34 |

Step 0:

Model : Yij = µ + Ai + eij , i = 1, 2, 3, …, j = 1, 2, 3, … Ho : µC = µP = µF H1 : Ho is not true.

Step 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | n | Mean ( yഥi) | Standard Deviation ( Si ) | Ti |
| Control | 15 | 82.52 | 9.24 | 1237.8 |
| Pets | 15 | 73.48 | 9.97 | 1102.2 |
| Friends | 15 | 91.325 | 8.34 | 1369.875 |

DF (Treatment) = 3 – 1 = 2

DF (Total) = 45 – 1 = 44

DF (Error) = 44 – 2 = 42

G = ∑i ∑j yij = ∑i ni yiഥ = ∑i Ti

Cumulative Frequency, CF = G2/N

∑ yij2 = ni (Si2 + yiഥ2) Si2 = 1/ ni  ( ∑ yij2 - yiഥ2 )

G = (15\*82.52) + (15\*73.48) + (15\*91.325)

= 3709.875

CF = (3709.875)2/45

= 305848.2781

∑ yij2 = 15\*(9.242 + 82.522) + 15\*(9.972+ 73.482) + 15\*(8.342 + 91.3252) = 312051.7579

Total SS = ∑i ∑j ∑ yij2 – G2/N

Treatment SS = ∑i (Ti2 / ni) – CF

Error SS = Total SS – Treatment SS

Total SS = 6203.4798

Treatment SS = 308236 - 305848.2781 = 2388.468275

Error SS = 3815.011525

Mean SS (Treatment) = Treatment SS / DF (treatment)

= 1194.234138 Mean SS (Residual) = Error SS / DF (Error)

# = 90.83360

F statistic = Mean SS(Treatment) / Mean SS(Residual) = 13.14749

F(0.05, 2, 42) = 3.220

Conclusion:

Since, Test Statistic, F > Tabulated value of F, we reject the null hypothesis. Thus, we conclude that the mean heart rate is not the same for each of the three groups.

5. High productivity and carbohydrate storage ability of the Jerusalem artichoke make it a promising agricultural crop. Consider the following data on chlorophyll concentration (in grams per square meter) for four varieties of Jerusalem artichoke:

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***n*** | ***Mean*** | ***Std. Dev.*** |
| ***Variety 1*** | 5 | 0.3 | 0.12 |
| ***Variety 2*** | 5 | 0.24 | 0.089 |
| ***Variety 3*** | 4 | 0.41 | 0.1 |
| ***Variety 4*** | 6 | 0.33 | 0.054 |

Do the data suggest that true average chlorophyll concentration depends on the variety? State and test the appropriate hypotheses at a level of 0.05. If it is significant, find the variety where difference is more.

Step 0:

Model: Yij = µ + Ai + eij , i = 1, 2, 3, …, j = 1, 2, 3, … Ho: µ1 = µ2 = µ3 = µ4 H1: Ho is not true.

Step 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | n | Mean ( yഥi) | Standard Deviation ( Si ) | Ti |
| Variety 1 | 5 | 0.3 | 0.12 | 1.5 |
| Variety 2 | 5 | 0.24 | 0.089 | 1.21 |
| Variety 3 | 4 | 0.41 | 0.1 | 1.64 |
| Variety 4 | 6 | 0.33 | 0.054 | 1.98 |

DF (Treatment) = 4 – 1 = 3

DF (Total) = 20 – 1 = 19

DF (Error) = 19 – 3 = 16 G = ∑i ∑j yij = ∑i ni yiഥ = ∑i Ti

Cumulative Frequency, CF = G2/N

∑ yij2 = ni (Si2 + yiഥ2) Si2 = 1/ ni  ( ∑ yij2 - yiഥ2 )

G = (5\*0.3) + (5\*0.24) + (4\*0.41) + (6\*0.33) = 6.33s

CF = (6.33)2/20 = 2.003445

∑ yij2 = 5\*(0.122 + 0.32) + 5\*(0.0892+ 0.242) + 4\*(0.12 + 0.412) + 6\*(0.0542 + 0.332) = 2.2329

Total SS = ∑i ∑j ∑ yij2 – G2/N = 0.229455

Treatment SS = ∑i (Ti2 / ni) – CF = 2.06862 – 2.003445 = 0.065175

Error SS = Total SS – Treatment SS = 0.16428

Mean SS (Treatment) = Treatment SS / DF (treatment)

= 0.021725

Mean SS (Residual) = Error SS / DF (Error)

# = 0.0102675

F statistic = MSST / MSSE = 2.11589

F(0.05, 3, 16) = 3.24

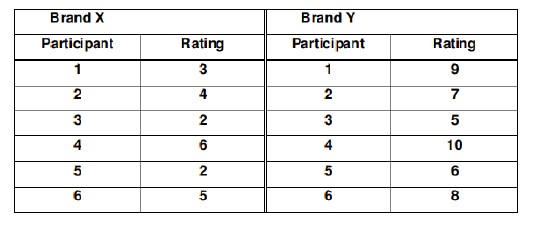
Conclusion:

Since, Test Statistic, F < Tabulated value of F, we accept the null hypothesis.

There is not enough evidence to reject the null hypothesis.

Thus, it is not significant and so, we cannot conclude that the mean chlorophyll concentration depends on the variety.

6. The effectiveness of advertising for two rival products (Brand X and Brand Y) was compared. Market research at a local shopping centre was carried out, with the participants being shown adverts for two rival brands of coffee, which they then rated on the overall likelihood of them buying the product (out of 10, with 10 being "definitely going to buy the product"). Half of the participants gave ratings for one of the products, the other half gave ratings for the other product.



Perform appropriate non-parametric test to determine whether the ratings for both the brands are same. Test to be done at 5% level of significance.

Step 0:

Ho : µ1 = µ2

H1 : Ho is not true.

Step 1:

Rank all scores together, ignoring which group they belong to.

If there is any tie in the ranking, take the average of those ranks.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Brand X |  |  | Brand Y |  |
| Participant | Rating | Rank | Participant | Rating | Rank |
| 1 | 3 | 3 | 1 | 9 | 11 |
| 2 | 4 | 4 | 2 | 7 | 9 |
| 3 | 2 | 1.5 | 3 | 5 | 5.5 |
| 4 | 6 | 7.5 | 4 | 10 | 12 |
| 5 | 2 | 1.5 | 5 | 6 | 7.5 |
| 6 | 5 | 5.5 | 6 | 8 | 10 |

Step 2:

∑R1 = Sum of ranks of Brand X

∑R1 = 3 + 4 + 1.5 + 7.5 + 1.5 + 5.5 = 23

∑R2 = Sum of ranks of Brand Y

∑R2 = 11 + 9 + 5.5 + 12 + 7.5 + 10 = 55

Number of participants in each group = n1 = n2 = 6

Step 3:

U1 = n1n2 + (n1(n1+1)) / 2 - ∑R1

U1 = 6\*6 + (6(6+1)) / 2 - 23

U1 = 34

U2 = n1n2 + (n2(n2+1)) / 2 - ∑R2

U2 = 6\*6 + (6(6+1)) / 2 - 55

U2 = 2

U2 < U1

Thus, U = U2 = 2

Step 4:

Critical value of U = 5 for n1 = n2 = 6.

Conclusion:

Since Test Statistic value, U < U critical value, we reject Ho.

Hence, we conclude that the ratings for both the brands are not same.

7. The study assessed the effectiveness of a new drug designed to reduce repetitive behaviours in children affected with autism. A total of 8 children with autism are enrolled in the study and the amount of time that each child is engaged in repetitive behaviour during three hour observation periods are measured both before treatment and then again after taking the new medication for a period of 1 week. The data are shown below.

|  |  |  |
| --- | --- | --- |
| Child | Before Treatment | After 1 Week of Treatment |
| 1 | 85 | 75 |
| 2 | 70 | 50 |
| 3 | 40 | 50 |
| 4 | 65 | 40 |
| 5 | 80 | 20 |
| 6 | 75 | 65 |
| 7 | 55 | 40 |
| 8 | 20 | 25 |

Perform appropriate non-parametric test to see whether there is any effect of the new drug.

Step 0:

Ho : µB = µA

H1 : Ho is not true.

Step 1:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Child | Before | After 1Week | Difference | Ranks for magnitude | Signed Rank |
| 1 | 85 | 75 | 10 | 3 | 3 |
| 2 | 70 | 50 | 20 | 6 | 6 |
| 3 | 40 | 50 | -10 | 3 | -3 |
| 4 | 65 | 40 | 25 | 7 | 7 |
| 5 | 80 | 20 | 60 | 8 | 8 |
| 6 | 75 | 65 | 10 | 3 | 3 |
| 7 | 55 | 40 | 15 | 5 | 5 |
| 8 | 20 | 25 | -5 | 1 | -1 |

Sum of Positive Ranks = 32

Sum of Negative Ranks = -4

Mininum = 4

Calculated Rank = 4

Critical Value of Wilcoxon Signed Rank Test = 4 when n = 8 and α = 0.05.

Conclusion:

Since calculated rank = Wilcoxon Signed Rank.

Hence, we accept the null hypothesis.

We conclude that there is effect of new drug at 0.05 level of significance.

8. In the same data given in the previous problem, use appropriate parametric test to check whether there is any effect.

Step 0:

Ho : µ = µd = 0 H1 : Ho is not true.

Step 1:

Level of significance α = 0.05.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Child | Before Treatment | After 1 Week of Treatment | Difference (d) | d2 |
| 1 | 85 | 75 | 10 | 100 |
| 2 | 70 | 50 | 20 | 400 |
| 3 | 40 | 50 | -10 | 100 |
| 4 | 65 | 40 | 25 | 625 |
| 5 | 80 | 20 | 60 | 3600 |
| 6 | 75 | 65 | 10 | 100 |
| 7 | 55 | 40 | 15 | 225 |
| 8 | 20 | 25 | -5 | 25 |
|  | Total | | 125 | 5175 |

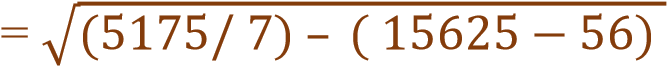
Step 2:

Student’s T distribution

Test Statistic, t = ത(ത𝑑ത − µd)/ (Sd / √n) ≈ t(α, n-1)

𝑑ҧ = ∑di / n = 125 / 8 = 15.625

Sd = ඥ[ ∑(di)2/(n − 1)] – [(∑di)2 / n(n − 1) ]

Sd 

Sd 

= 21.4538 t = 15.625 / [21.4538 / √8 ]

= 2.0599

Step 3:

Critical value of t, for α= 0.05 and df = n – 1 = 7

t(0.025, 7) = 4.785

Conclusion:

Since value of Test statistic, t < t critical value, Ho is accepted. We conclude that there is effect of the new drug.

1. Ratings are given in depression scale to different set of people (those who perform no exercise, those who perform jogging for 20 minutes and those who perform jogging for 40 minutes)

Perform appropriate non-parametric test to see whether ratings are same for 3 groups of people.

|  |  |  |  |
| --- | --- | --- | --- |
|  | No exercise | Jogging for 20 mins | Jogging for 60 mins |
| 1 | 23 | 22 | 59 |
| 2 | 26 | 27 | 66 |
| 3 | 51 | 39 | 38 |
| 4 | 49 | 29 | 49 |
| 5 | 58 | 46 | 56 |
| 6 | 37 | 48 | 60 |
| 7 | 29 | 49 | 56 |
| 8 | 44 | 65 | 62 |
| Mean rating | 39.63 | 40.63 | 55.75 |
| Std Deviation | 12.85 | 14.23 | 8.73 |

Step 0:

Ho : µ1 = µ2 = µ3 H1 : Ho is not true.

Step 1:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | No exercise | | Jogging for 20 mins | | Jogging for 60 mins | |
| Scale | Rank | Scale | Rank | Scale | Rank |
| 1 | 23 | 2 | 22 | 1 | 59 | 20 |
| 2 | 26 | 3 | 27 | 4 | 66 | 24 |
| 3 | 51 | 16 | 39 | 9 | 38 | 8 |
| 4 | 49 | 14 | 29 | 5.5 | 49 | 14 |
| 5 | 58 | 19 | 46 | 11 | 56 | 17.5 |
| 6 | 37 | 7 | 48 | 12 | 60 | 21 |
| 7 | 29 | 5.5 | 49 | 14 | 56 | 17.5 |
| 8 | 44 | 10 | 65 | 23 | 62 | 22 |
| Rank (R) |  | 76.5 |  | 79.5 |  | 144 |
| Mean Rating | 39.63 | | 40.63 | | 55.75 | |
| Std Deviation | 12.85 | | 14.23 | | 8.73 | |

n = total no of observations = 8 + 8+ 8 = 24

R12 = 5852.25

R22 = 6320.25

R32 = 20736

H = (12 / n(n+1)) ∑ [ ( R2i / ni ) -3(n+1) ]

H = (12 / 24(25)) ((5852.25/8 – 3\*25) \* (6320.25/8 – 3\*25) \* (20736/8 – 3\*25))

H = 7.27125

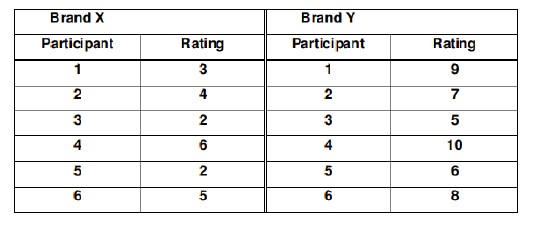
H has a χ2 distribution df = 2 and α = 0.05 is 5.991.

Conclusion:

Since H calculated value > H critical value, we reject Ho.

Thus, we can conclude that there is a significant effect of exercise on depression.

1. In the same data given in the previous problem, use appropriate parametric test to determine whether the ratings for both the brands are same. Test to be done at 5% level of significance.



Step 0:

Ho : µ1 = µ2

H1 : Ho is not true.

Step 1:

|  |  |  |  |
| --- | --- | --- | --- |
| Brand X | | Brand Y | |
| Participant | Rating | Participant | Rating |
| 1 | 3 | 1 | 9 |
| 2 | 4 | 2 | 7 |
| 3 | 2 | 3 | 5 |
| 4 | 6 | 4 | 10 |
| 5 | 2 | 5 | 6 |
| 6 | 5 | 6 | 8 |
| Total | 22 | Total | 45 |

ത𝑥ത1ത = 22 / 6 = 3.67 ത𝑥ത2ത = 45 / 6 = 7.5 n1 = n2 =6

S12 = ∑ [ ∑ (xi − xത)2 / (n-1) ] = 2.67

S22 = ∑ [ ∑ (yi − yത)2 / (n-1) ] = 3.5

Sp2 = [ (n1-1)S12 + (n2-1)S22 ] / (n1 + n2 -2)

= [ (6-1)\*2.67 + (6-1)\*3.5] / ( 6 + 6 – 2) = 3.08

t = ( xത1 − xത2 ) / √ [Sp2 ( ( 1/n1) + (1/ n2) ) ]

= (7.5 - 3.67) / √ [3.08 (( 1/6) + (1/6))]

= -3.78

t(0.05,7) = 2.228

Conclusion:

Since tcalc < tcritic , we accept H0 at α = 0.05 and df = 10.

We accept the null hypothesis. i.e, we conclude that ratings for both the brands are same.