Statistisk Dataanalyse 2023

Instruktør: Vivek Misra

Pointgivende Aktiviteter

- **I har 2 Pointgivende Aktiviteter i faget Statistisk Dataanalyse.**
  - Hvert Pointgivende Aktivitet tæller 5% og bidrager samlet til 10% i den endelige karakter, da der er 2 tests.
  - **FORDELE:**
    - Hjælper den enkelte studerende med at forhøje karakteren.
    - Hjælper i ”værste tilfælde” med at bestå, hvis studerende ligger på overkanten til at dumpe.
    - Hjælper med at give mavefornemmelse over Eksamensspørgsmål som kommer til faget.
  - **ULEMPE:**
    - Studerende har mindre chance for at forhøje karakteren og skal gøre ekstra indsats til endelig eksamen.
    - Studerende har mindre chance for at blive reddet til den endelig eksamen, hvis dumpet.
  - **DER ER INGEN RE-PRØVER I POINTGIVENDE AKTIVITET**
    - Deltager man ikke testen, afholdes testen IKKE igen da det afholdes engang i specifikke tider.

Snyd og Plagiat!

- **DET ER STRENGT FORBUDT, AT SNYDE TIL POINTGIVENDE AKTIVITETER, TESTS, EKSAMEN OG RAPPORT ETC.**
  - **BRUGEN AF CHATGPT OG ANDRE AI ER STRENGT FORBUDT FRA SDU'S SIDE TIL EKSAMENS AKTIVITETER! (WARRANT).**
  - I MÅ HELLER **IKKE** BRUGE VIMIS22.GITHUB.IO TIL POINTGIVENDE AKTIVITETER OG EKSAMEN.
    - JEG DEAKTIVERER SIDEN, SÅLEDES AT I IKKE KAN KOMME IND. DET ER JER, DER ER ANSVARLIG FOR BRUG AF JERES EGEN HJÆLPEMIDLER!
  - **NOTE: Fejler man Pointgivende Aktivitet eller klarer sig dårligt. Vær ikke bange!**
    - **Se det som en mulighed for forbedring, og kom ENDELIG gerne til Øvelsestimerne og få spurgt om tingene.**
  - **RESULTAT PÅ POINTGIVENDE AKTIVITET, KOMMER STRAKS EFTER AFSLUTNING AF PRØVEN.**
    - **AFHOLDELSE: PÅ ITSLEARNING.COM VED FORELÆSNINGSLOKALET**
  - **KOMMUNIKATION IKKE TILLADT PÅ TVÆRS AF ELEVER, HVERKEN VERBALT ELLER GENNEM INTERNETTET. DETTE SES OGSÅ SOM SNYD UNDER POINTGIVENDE AKTIVITET & EKSAMEN.**
  - **ANBEFALING TIL PA: Er du tvivl om et spørgsmål, gå med første mavefornemmelse.**
    - **VEND TILBAGE TIL DEN TVIVLENDE SPØRGSMÅL, NÅR DU HAR LØST DE ANDRE OPGAVER!**
  - **PS: Hvis du er Ordblind, så kontakt venligst lektoren eller sekretæren for ekstra tid!**

Exercise Class NR6

Solutions to the Tasks

Task 1 - Description

Fertilizer 1	Fertilizer 2	Fertilizer 3
10	6	5
12	8	9
9	3	12
15	0	8
13	2	4

- We want to investigate if there is an effect of the type of fertilizer applied to apple trees and the production of apples. We randomly select 15 trees and randomly assign them to one of three groups (5 trees per group). We perform a test in which we apply one type of fertilizer (fertilizer 1, 2 or 3) to each group. The data are shown. At a $\alpha=0.05$, can it be concluded that there is a significant difference in the production of apples depending on which fertilizer is used? Which fertilizer/fertilizers causes a higher/lower production than the other/others?

Task 1 - Solution

- Solution: We need to compare the means of three sample groups.
- First we need to find the mean of the three sample groups.

Fertilizer Mean 1	Fertilizer Mean 2	Fertilizer Mean 3
11,8	3,8	7,6

$$Mean\ of\ Total\ Fertilizer = \frac{11{,}8 + 3{,}8 + 7{,}6}{3} = \frac{23{,}2}{3} = 7{,}73$$

Task 1 - Solution

- Now we need to find the sum of square totals and betweens.

Sum of Square Total	$(10-7{,}73)^2+(12-7{,}73)^2+(9-7{,}73)^2+(15-7{,}73)^2$ $+(13-7{,}73)^2+(6-7{,}73)^2+(8-7{,}73)^2+(3-7{,}73)^2$ $+(0-7{,}73)^2+(2-7{,}73)^2+(5-7{,}73)^2+(9-7{,}73)^2$ $+(12-7{,}73)^2+(8-7{,}73)^2+(4-7{,}73)^2 = 264{,}9335$
Sum of Square Between	$5\cdot(11{,}8-7{,}73)^2 = 82{,}8245$
Sum of Square Between	$5\cdot(3{,}8-7{,}73)^2 \approx 77{,}2245$
Sum of Square Between	$5\cdot(7{,}6-7{,}73)^2 \approx 0{,}0845$
Sum of Total Square Between	$82{,}8245+77{,}2245+0{,}0845 = 160{,}1335$

- Now the Sum Square Total and Sum Square Between will be subtracted.

SS-Total – SS-Between	$\mathbf{264,9335-160,1335=104,8}$

Task 1 - Solution

- Now we will create the table shown in the presentation.

X	Sum of Squares	d.f.	Mean	F-Ratio
Factor	SS-Between = 160,1335	Antal gruppe - 1 => 3-1 = 2	Dividere Total Square Between med 2 $\frac{160{,}1335}{2} = 80{,}06675$	Dividere Mean Faktor med Mean Residual $\frac{80{,}06675}{8{,}733333} \approx 9{,}167949$
Residual	$\mathbf{264{,}9335 - 160{,}1335 \approx 104{,}8}$	$14 - 2 = 12$	Dividere SS-Between Residual med df-Residual $\frac{104{,}8}{12} \approx 8{,}733333$	
Total	SS-Total = 264,9335	Antal dataværdier => 15-1 = 14		

- Now we will find the degrees of freedom.

k-1, N-k	(3,89)

Task 1 - Solution

- We can see, that the F-ratio is bigger than the critical value and therefore we can say as followed: F-Ratio>F-Table.
- Because F-Ratio is bigger than F-table, then we can reject the null-hypothesis which means that there is a difference between the means of the samples.
- Now the LSD-Test needs to be implemented, to see a clear difference among the sample groups.

Task 1 - Solution

- Here we have created an table, where the LSD-Test i conducted.

F-Ratio > F-Table			LSD-Intervals
Fertilizers	Mean	Lower Interval	Upper Interval
Type 1	11,8	9,76	13,83
Type 2	3,8	1,76	5,83
Type 3	7,6	5,56	9,63
(95%,12)			

- We can see, that there is a significant different among the 3 different fertilizers. The LSD-Test shows that Type 1 is the most different among the 3, because it produces more apples than Type 2 and 3 according to its interval.

Task 2 - Description

- We want to evaluate three different methods to lower the blood pressure of individuals that have been diagnosed with high blood pressure. Eighteen subjects are randomly assigned to three groups (6 per group): the first group takes medication, the second group exercises, and the third one follows a specific diet. After four weeks, the reduction in each person's blood pressure is recorded. Is there a significant difference among the reduction obtained from each of the three methods? If yes, which method was more effective?

Medication	Exercise	Diet
12	5	6
8	9	10
11	2	5
17	0	9
16	1	8
15	3	6

Task 2 - Solution

- Solution: We need to compare the means of three sample groups.
- First we need to find the mean of the three sample groups.

Medication Mean	Exercise Mean	Diet Mean
13,16	3,33	6,33

$$Mean\ of\ Total\ Methods = \frac{13{,}16 + 3{,}33 + 6{,}33}{3} = \frac{22{,}82}{3} = 7{,}606$$

Task 2 - Solution

- Now we need to find the sum of square totals and betweens.

Sum of Square Total	$(12-7{,}61)^2+(8-7{,}61)^2+(11-7{,}61)^2+(17-7{,}61)^2+(16-7{,}61)^2+(15-7{,}61)^2+(5-7{,}61)^2+(9-7{,}61)^2+(2-7{,}61)^2+(0-7{,}61)^2+(1-7{,}61)^2+(3-7{,}61)^2+(6-7{,}61)^2+(10-7{,}61)^2+(5-7{,}61)^2+(9-7{,}61)^2+(8-7{,}61)^2+(6-7{,}61)^2=426{,}9578$
Sum of Square Between	$6\cdot(13{,}16-7{,}61)^2=184{,}815$
Sum of Square Between	$6\cdot(3{,}33-7{,}61)^2\approx 109{,}9104$
Sum of Square Between	$6\cdot(7{,}33-7{,}61)^2\approx 0{,}4704$
Sum of Total Square Between	$184{,}815+109{,}9104+0{,}4704\approx 295{,}1958$

- Now the Sum Square Total and Sum Square Between will be subtracted.

SS-Total – SS-Between	$\mathbf{426,9578-295,1958=131,762}$

Task 2 - Solution

- Now we will create the table shown in the presentation.

X	Sum of Squares	d.f.	Mean	F-Ratio
Factor	SS-Between = 295,1958	Antal gruppe - 1 => 3-1 = 2	Dividere Total Square Between med 2 $\frac{295,1958}{2} = 147,5979$	Dividere Mean Faktor med Mean Residual $\frac{147,5979}{8,235} \approx 17,92324$
Residual	$\mathbf{426,9578 - 295,1958 = 131,762}$	$18 - 2 = 16$	Dividere SS-Between Residual med df-Residual $\frac{131,76}{16} = 8,235$	
Total	SS-Total = 426,9589	Antal dataværdier => 18-1 = 17		

- Now we will find the degrees of freedom.

k-1, N-k	$(3,68)$

Task 2 - Solution

- We can see, that the F-ratio is bigger than the critical value and therefore we can say as followed: F-Ratio>F-Table.
- Because F-Ratio is bigger than F-table, then we can reject the null-hypothesis which means that there is a difference between the means of the samples.
- Now the LSD-Test needs to be implemented, to see a clear difference among the sample groups.

Task 2 - Solution

- Here we have created an table, where the LSD-Test i conducted.

F-Ratio > F-Table			LSD-Intervals
Treatments	Mean	Lower Interval	Upper Interval
Medication	13,16	11,00	15,31
Exercise	3,33	1,17	5,48
Diet	6,33	4,17	8,48
(95%,12)			

- We can see, that there is a significant different among the three treatments in reducing the bloodpresure. Medication in this case is the most effective, and has the highest mean among the three methods. We can see, that there is no effect in using treatment-methods such as Exercise and Diet, which actually have the same effect. Therefore Medication is different!

Task 3 - Description

Region Hovedstaden	Region Sjælland	Region Syddanmark	Region Midtjylland	Region Nordjylland
68	78	89	62	57
75	79	87	74	65
95	65	75	71	78
85	67	65	70	88
84	60	84	72	67
88	79	92	72	77
85	57	84	64	72
75	74	72	75	69

- In the table below, there are randomly selected scores for eight amateur basketball teams in each of five Danish regions, for a particular weekend. Is there sufficient evidence to support that there is a difference in mean scores by region? If yes, which region/s got the highest scores and which one the lowest?

Task 3 - Solution

- Solution: We need to compare the means of three sample groups.
- First we need to find the mean of the three sample groups.

Capital Region	Region of Sjælland	Region of Syddanmark	Region of Midtjylland	Region of Nordjylland
81,875	69,875	81	70	71,625

$$Mean\ of\ Total\ Fertilizer = \frac{81{,}875 + 69{,}875 + 81 + 70 + 71{,}625}{3} = \frac{374{,}375}{3} = 74{,}875$$

Task 3 - Solution

- Now we need to find the sum of square totals and betweens.

Sum of Square Total	$\mathbf{3618,375}$
Sum of Square Between	$8 \cdot (81{,}875 - 74{,}875)^2 = 392$
Sum of Square Between	$8 \cdot (74{,}875 - 69{,}875)^2 = 200$
Sum of Square Between	$8 \cdot (81 - 74{,}875)^2 = 300{,}125$
Sum of Square Between	$8 \cdot (70 - 74{,}875)^2 = 190{,}125$
Sum of Square Between	$8 \cdot (71{,}625 - 74{,}875)^2 = 84{,}5$
Sum of Total Square Between	$392 + 200 + 300{,}125 + 190{,}125 + 84{,}5 = 1166{,}75$

- Now the Sum Square Total and Sum Square Between will be subtracted.

SS-Total – SS-Between	$\mathbf{3618,375 - 1166,75 = 2451,625}$

Task 3 - Solution

- Now we will create the table shown in the presentation.

X	Sum of Squares	d.f.	Mean	F-Ratio
Factor	SS-Between = 1166,75	Antal gruppe - 1 => 5-1 = 4	Dividere Total Square Between med 2 $\frac{1166,75}{4} = 291,6875$	Dividere Mean Faktor med Mean Residual $\frac{291,68}{70} \approx 4,166857$
Residual	$\mathbf{3618,375} - \mathbf{1166,75} = \mathbf{2451,625}$	$39 - 4 = 35$	Dividere SS-Between Residual med df-Residual $\frac{2451,625}{35} \approx 70,04643$	
Total	SS-Total = 3618,375	Antal dataværdier => 40-1 = 39		

- Now we will find the degrees of freedom.

k-1, N-k	$(\mathbf{2}, \mathbf{63})$

Task 3 - Solution

- We can see, that the F-ratio is bigger than the critical value and therefore we can say as followed: F-Ratio>F-Table.
- Because F-Ratio is bigger than F-table, then we can reject the null-hypothesis which means that there is a difference between the means of the samples.
- Now the LSD-Test needs to be implemented, to see a clear difference among the sample groups.

Task 3 - Solution

- Here we have created an table, where the LSD-Test i conducted.

F-Ratio > F-Table			LSD-Intervals
Fertilizers	Mean	Lower Interval	Upper Interval
Capital Region	81,875	81,01	82,73
Region of Sjælland	69,875	69,01	82,73
Region of Syddanmark	81	80,13	81,86
Region of Midtjylland	70	69,13	70,86
Region of Nordjylland	71,625	70,76	72,48
(95%,12)			

- The Capital Region, and the Region of Southern Denmark has the highest interval in terms of scores. Whereas the Region of Sjælland is also above, but the Region of Midtjylland and Nordjylland have the lowest scores and therefore we can say that there is a significant difference among the regions, which is the acceptance of the alternative hypothesis.

Tak for i dag!

Instruktør: Vivek Misra