

Indhold

General information	2
Valg af undersøgelse	2
Fortolkning af P-værdi	3
ANOVA Tabel	3
Notation - Sandsynlighed:	3
One-way ANOVA	5
Konfidensinterval for gennemsnit (Bonferroni)	7
Test for om 2 eller flere σ^2 er ens - Test for varianshomogenitet.	8
Two-way ANOVA	9
Two-Factor ANOVA	11
χ^2-test - Antalstabel / Contingency table / Uafhængighedstest	14
χ^2-test - Goodness of fit	17
Simpel lineær regression	19
Forventet værdi, Prædiktions- og konfidensinterval	22
Multipel regression	23
Forventet værdi, Prædiktions- og konfidensinterval	27
Multipel regression - Unrestricted vs. Restricted	28
Logistisk Regression	31
Tidserier	34
- Moving Average.....	34
- Seasonal Indeks	34
- Trendmodel - Simple lineær regression m. en tidsvariable	34
- Timeseries - Autoregressiv model	38

One-way ANOVA

One-way Anova vælges hvis den afhængige variable (y) er intervallskaleret og den uafhængige variable er nominal eller ordinalt skaleret, gerne med flere indbyrdes grupper.

1. Opstil den/de stokastiske variable

- Hvad er det vi undersøger?

2. Opstil hypoteserne

$$H_0: \mu_1 = \mu_2 = \mu_3 \dots = \mu_k$$

H_1 : Mindst to gennemsnit er forskellige

3. Model

$$Y_i = \mu + \alpha_i + \varepsilon$$

4. Opstil signifikansniveau samt alfa-stjerne (hvis ikke andet fremgår af opgaven er $\alpha = 0.05$)

$$\alpha = 0.05$$

$$\alpha^* = \frac{\alpha}{k * (k - 1) / 2}$$

5. Valg af observator

$$\frac{MST}{MSE} \sim F_{k-1; n-k}$$

6. Forudsætninger

- H_0 -hypotesen er sand
- $X_j \sim ND$ - Hver stikprøve der indgår i undersøgelsen, skal være (tilnærmelsesvis) normalfordelt (UCGS).
 - Fit Y by X \rightarrow y i "y" og x'erne i "by"
- Stikprøverne skal være simpelt tilfældigt udvalgt
- Uafhængighed
 - Indenfor grupper
 - Hvis STU opfyldt så er disse ofte ikke et problem
 - Afrapportering
 - Imellem grupper
- Troværdighed - kan man stole på at oplysningerne er korrekte?
- Der skal være varianshomogenitet:

7. Undersøg for varianshomogenitet

- Opstil hypoteser til test for varianshomogenitet

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \dots = \sigma_k^2$$

H_1 : Mindst to varianser er forskellige

- Valg af signifikansniveau

$$\alpha = 0.05$$

$$f_{obs} = \frac{S_{max}^2}{S_{min}^2}$$

e. Kritiske grænser (brug Excel ark)

$$f_{n_{max}-1; n_{min}-1; \frac{\alpha^*}{2}}$$

i. Husk at dividere alfa-stjerne med 2!!

f. Konklusion

i. Fasthold / Forkast

ii. Varianshomogenitet eller heterogenitet?

6) Beregning af observatorværdien

- Indsæt "Analysis of variance" og "effect tests" fra jmp

$$f_{obs} = \frac{MST}{MSE}$$

7) Kritiske grænse (brug Excel)

$$f_{k-1; n-k; \alpha}$$

Hvor k er antal grupper (stikprøver)

8) Konklusion på varianshomogenitetstesten

9) Udregning af observator værdier og p-værdi

a. Indsæt Analysis of variance og effect tests fra JMP output

$$\frac{MS(A)}{MSE} \sim F_{a-1; n-ab}$$

$$\frac{MS(B)}{MSE} \sim F_{b-1; n-ab}$$

10) Konklusion

a. Modellen som helhed (analysis of variance → P-værdi og observatorværdi)

i. Hvis P-værdien er mindre end signifikansniveauet er modellen signifikant

b. Effekterne (effect test → kommenter på parametrene og p-værdierne)

i. Hvis P-værdien er mindre end signifikansniveauet er modellen signifikant

ii. Hvis kun én af hovedeffekterne er signifikante reduceres modellen til en oneway anova. Herefter testes der om denne ene effekt forsat er signifikant - Husk at opskriv den nye model.

c. Hvad betyder resultatet for opgavens problemstilling?

11) Fortolkning - Hvis begge hovedeffekter er signifikante:

a. Bonferroni konfidensintervaller

b. LSMeans differences student's t -> husk at korrigerer alpha til α^* .

i. Kig på forskelle for variableerne.

- i. Er observationerne uafhængige af hinanden, eller har man haft mulighed for at påvirke hinandens svar?
 - 1. Hvis observationerne er simpelt tilfældigt udvalgt er der ikke noget problem her.
- Troværdighed
 - i. Er svarene blevet påvirket udefra? - Har vedkommende et incitament til at lyve?
 - ii. Kan det sandsynliggøres at informanterne kender til den adspurgte information nøjagtigt?
 - iii. Afrapporteringsproblemer
- STU

6) Udregn observatorværdi

- Indsæt JMP-output der fremkommer ved at indsætte to nominalt skalerede variabler i Fit Y by X - den ene i Y response og den anden i X.
 - i. Analyze -> Fit Y by X -> markér de to faktorer -> Indsæt i X & Y -> under rød pil (contingency table) skal "Count", "Expected", "Deviation" og "Cell Chi^2" krydses af
- Indsæt tabellen "Tests" fra jmp-outputtet og aflæs Pearson-værdien
 - i. Hvis P-værdien er mindre end signifikansniveauet forkastes H₀ og modellen er signifikant (de to variabler er afhængige)

$$\sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \sim \chi^2_{(r-1) \cdot (c-1)}$$

- Indsæt værdier fra outputtene i formlen ovenfor.
 - i. f_{ij} = den aktuelle kolonne og rækkes faktiske værdi
 - ii. e_{ij} = den aktuelle kolonne og rækkes forventede værdi

7) Kritiske grænser (Brug Excel filen)

$$\chi^2_{(r-1) \cdot (c-1); \alpha}$$

$r = \text{antal rækker}, c = \text{antal kolonner}$

8) P-værdi

$$P(\chi^2_{(r-1) \cdot (c-1)} > \chi^2_{obs}) =$$

9) Konklusion

- På foreliggende grundlag forkastes/fastholdes H₀ + argument for hvorfor
 - o Kritiske grænse >< observatorværdien
- Beskrivelse af p-værdi → Sikkerheden i testen
- Forudsætningsproblemer
- Tekst der kobler det sammen med opgaveteksten
- Tendensen i de observerede værdier ift. de forventede værdier