HD Dataanalyse, efterår 2018 Opgaveark 8



Baggrund og databeskrivelse (opgave 1-2)

Denne opgave ser nærmere på forskellige regnskabsnøgletal for et udsnit af små, mellem-store og store danske virksomheder. Opgaven tager udgangspunkt i filen Experian2008 (håndværk).jmp, hvis indhold er den del af datasættet beskrevet på opgaveark 5, der om-handler virksomheder indenfor de otte forskellige håndværk: mekaniker, el-installatør, ba-ger, maler, murer, tømrer, gulvlægger og blikkenslager.

Opgave 1 (egenkapitalens forrentning)

a). Gør rede for om der er statistisk belæg for en nulhypotese om, at den forventede egenkapitalforrentning er den samme blandt de betragtede brancher (Primær branche) (brug = 5% som signifikansniveau). Angiv i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.

JMP-vink: “Analyze” -> “Fit Y by X”. Placér Egenkapitalens forrentning (i %) under “Y, Response”,

Primær branche under “X, Factor”. Vælg “Means/Anova” via den røde trekant.

|  |  |
| --- | --- |
| ANOVA/ANAVA | |
| Forudsætninger | Populationerne er normalfordelte.  Populationerne har samme varians. |
| Hypotese | Alle egenkapitalforrentninger er ens for de 8 brancher.  Ikke alle egenkapitalforrentninger er ens for de 8 brancher.  Eller    Ikke alle egenkapitalforrentninger er ens for de 8 brancher. |
| P-værdi/signifikanssandsynlighed | 38.97% |
| Signifikansniveau | 5% |
| Teknisk konklusion | Da p-værdien 38.97% er større end signifikansniveauet på 5%, kan vi ikke forkaste nulhypotesen. |
| Ikke-teknisk konklusion | Alle egenkapitalforrentninger er ens for de 8 brancher |
|  | |

b). Gør rede for om antagelsen om normalfordelte data ser ud til at være opfyldt ved at se på histogrammer af fordelingen af egenkapitalforrentningen indenfor hver af de betragtede brancher.

JMP-vink: “Analyze” -> “Fit Y by X”. Vælg “Display Options” -> “Histograms” via den røde trekant.



Fordelingerne har ikke den typiske normalfordelings klokkeform, fordelingerne synes mere spidse med lange haler pga. outliers. Dette giver os problemer med normalfordelingsantagelsen

Gør rede for om antagelsen om normalfordelte data ser ud til at være opfyldt ved at se på middelværdi og standardafvigelse af egenkapitalforrentningen indenfor hver branche og sammenligne histogrammerne ovenfor med den empiriske regel.

JMP-vink: “Analyze” -> “Fit Y by X”. Vælg “Means and Std Dev” via den røde trekant.



Vi kan se at standardafvigelserne er meget høje i forhold til niveauet for middelværdierne i stikprøven. Fx har autorep værksteder en typisk afvigelse fra gennemsnittet på 135 procent point. Den empiriske regel siger så at ca. 32% af observationerne skal ligge udenfor spændet på 270 procent point, dette synes ikke opfyldt, da kun få virksomheder har en egenkapital forrentning på mindre end -117.43% og større end 152.91%.

Gør rede for hvordan eventuelle outliers i datamaterialet kan påvirke konklusionen i delspg. a).

Outliers kan påvirke middelværdierne hvilket kan føre til en forkert konklusion.



1/7

HD Dataanalyse, efterår 2018 Opgaveark 8



I den resterende del af opgave 1 begrænser vi datamaterialet til udelukkende at omhandle virksomheder med en egenkapitalforrentning på mellem -100% og 100%.

JMP-vink: “Rows” -> “Data Filter”

c). Gør rede for om der er statistisk belæg for en nulhypotese om, at den forventede egenkapitalforrentning er den samme blandt de betragtede brancher (brug = 5% som signifikansniveau). Angiv i den forbindelse antagelser, hypoteser, signifikans-sandsynlighed samt teknisk og let forståelig konklusion.

JMP-vink: “Analyze” -> “Fit Y by X”. Placér Egenkapitalens forrentning (i %) under “Y, Response”,

Primær branche under “X, Factor”. Vælg “Means/Anova” via den røde trekant.

|  |  |
| --- | --- |
| ANOVA/ANAVA | |
| Forudsætninger | Populationerne er normalfordelte.  Populationerne har samme varians. |
| Hypotese | Alle egenkapitalforrentninger er ens for de 8 brancher.  Ikke alle egenkapitalforrentninger er ens for de 8 brancher.  Eller    Ikke alle egenkapitalforrentninger er ens for de 8 brancher. |
| Teststørrelse | 3.7660 |
| P-værdi/signifikanssandsynlighed | 0.05% |
| Signifikansniveau | 5% |
| Teknisk konklusion | Da p-værdien 0.05% er mindre end signifikansniveauet på 5%, kan vi forkaste nulhypotesen. |
| Ikke-teknisk konklusion | Alle egenkapitalforrentninger er ikke ens for de 8 brancher |
|  | |

d). Tegn box plots af fordelingen af egenkapitalforrentningen indenfor hver af de betragtede brancher.

JMP-vink: “Analyze” -> “Fit Y by X”. Vælg “Display Options” -> “Box Plots” via den røde trekant. Fjern

samtidig “Points”, “Mean Diamonds” og “X Axis Proportional”.



Gør rede for at egenkapitalforrentningen blandt bagereog blikkenslagere ser ud til i et vist omfang at skille sig ud fra forrentningen blandt de øvrige brancher.

Ja bagere synes at have en lavere egenkapitalforrentning. Blikkenslagere synes at have en højere egenkapitalforrentning.

e). Gør rede for om konklusionen i delspg. c) ændres, såfremt vi ser bort fra bagerne.



Konklusionen i delspg. c) ændres ikke, når vi ser bort fra bagerne, vi forkaster fortsat nulhypotesen.

Gør rede for om konklusionen i delspg. c) ændres, såfremt vi ser bort fra blikkenslagerne.



Konklusionen i delspg. c) ændres, når vi ser bort fra blikkenslagerne, vi forkaster nu ikke nulhypotesen.

Gør rede for hvad der kan være årsagen til, at der er forskel på betydningen af bagere og blikkenslagere for analysens resultater.

JMP-vink: Se på antallet af observationer i hver af de to brancher ved at vælge “Means and Std Dev” via den røde trekant.

Der er 19 bagere og 148 blikkenslagere, de 19 bagere er for lille en stikprøve til at have en effekt på den samlede konklusion.

2/7

HD Dataanalyse, efterår 2018 Opgaveark 8



Opgave 2 (primært resultat)

Opgave 2 omhandler udelukkende virksomheder med et primært resultat (Primært resultat (i 1.000 kr.)) mellem -3 mio. kr. og 3 mio. kr.

JMP-vink: “Rows” -> “Data Filter”

a). Gør rede for om der er statistisk belæg for en nulhypotese om, at det forventede pri-mære resultat er det samme blandt de betragtede brancher (brug = 5% som signifi-kansniveau). Angiv i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.

JMP-vink: “Analyze” -> “Fit Y by X”. Placér Primært resultat under “Y, Response”, Primær branche

under “X, Factor”. Vælg “Means/Anova” via den røde trekant.

|  |  |
| --- | --- |
| ANOVA/ANAVA | |
| Forudsætninger | Populationerne er normalfordelte.  Populationerne har samme varians. |
| Hypotese | Alle de primære resultater er ens for de 8 brancher.  Ikke alle de primære resultater er ens for de 8 brancher.  Eller    Ikke alle de primære resultater er ens for de 8 brancher. |
| F-Teststørrelse | 3.1222 |
| P-værdi/signifikanssandsynlighed | 0.29% |
| Signifikansniveau | 5% |
| Teknisk konklusion | Da p-værdien 0.29% er mindre end signifikansniveauet på 5%, kan vi forkaste nulhypotesen. |
| Ikke-teknisk konklusion | Alle de primære resultater er ikke ens for de 8 brancher |
|  | |

b). Tegn box plots af fordelingen af det primære resultat indenfor hver af de betragtede brancher.

JMP-vink: “Analyze” -> “Fit Y by X”. Vælg “Display Options” -> “Box Plots” via den røde trekant. Fjern

samtidig “Points”, “Mean Diamonds” og “X Axis Proportional”.



Gør rede for at det primære resultat blandt el-installatører og blikkenslagere ser ud til i et vist omfang at skille sig ud fra resultatet blandt de øvrige brancher. JMP-vink: “Zoom ind” ved at begrænse værdierne på 2. aksen til intervallet -1,5 mio. til 1,5 mio. ved at højreklikke på 2. aksen, vælge “Axis Settings...” og indtaste de relevante værdier i felterne “Minimum” og “Maximum”.

El-installatører og blikkenslagere ser ud til at variere meget og have højere max end de øvrige brancher.



c). Gør rede for om konklusionen i delspg. a) ændres, såfremt der ses bort fra el-installatører og blikkenslagere i analysen.

JMP-vink: “Rows” -> “Data Filter” og “Analyze” -> “Fit Y by X”.

|  |  |
| --- | --- |
| ANOVA/ANAVA | |
| Forudsætninger | Populationerne er normalfordelte.  Populationerne har samme varians. |
| Hypotese | Alle de primære resultater er ens for de 6 brancher.  Ikke alle de primære resultater er ens for de 6 brancher.  Eller    Ikke alle de primære resultater er ens for de 6 brancher. |
| Teststørrelse | 1.5828 |
| P-værdi/signifikanssandsynlighed | 16.23% |
| Signifikansniveau | 5% |
| Teknisk konklusion | Da p-værdien 16.23% er større end signifikansniveauet på 5%, kan vi ikke forkaste nulhypotesen. |
| Ikke-teknisk konklusion | Alle de primære resultater er ens for de 6 brancher |
|  | |



3/7

HD Dataanalyse, efterår 2018 Opgaveark 8



Baggrund og databeskrivelse (opgave 3)

Denne opgave beskæftiger sig med hvilke faktorer, der er afgørende for, om ansatte er tilfredse med deres job. Opgaven tager udgangspunkt i filen Jobtilfredshed.jmp, hvis indhold er beskrevet på opgaveark 4.

Opgave 3

a). Gør rede for om der er statistisk belæg for en nulhypotese om, at jobtilfredsheden er den samme i hele landet (brug = 5% som signifikansniveau). Angiv i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklu-sion.

JMP-vink: “Analyze” -> “Fit Y by X”. Placér Jobtilfredshed under “Y, Response”, Region under “X, Fa-

ctor”.

|  |  |
| --- | --- |
| Chi i anden test | |
| Forudsætninger | De forventede celleværdier skal være større end 5, for at sikre kvaliteten af testen. Dette er ikke helt opfyldt her. |
| Hypotese | Der er uafhængighed dvs. ingen sammenhæng mellem Jobtilfredshed og region  Der er afhængighed dvs. sammenhæng mellem mellem Jobtilfredshed og region |
| P-værdi/signifikanssandsynlighed | 41.25% |
| Signifikansniveau | 5% |
| Teknisk konklusion | Da p-værdien 41.25% er større end signifikansniveauet på 5%, kan vi ikke forkaste nulhypotesen. |
| Ikke-teknisk konklusion | Der er uafhængighed dvs. ingen sammenhæng mellem mellem Jobtilfredshed og region. |
|  | |

b). Gør rede for om der er statistisk belæg for en nulhypotese om, at der ikke er nogen sammenhæng mellem jobtilfredshed og følelsen af at være motiveret af sit arbejde (Motivation) (brug = 5% som signifikansniveau). Angiv i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.

JMP-vink: “Analyze” -> “Fit Y by X”. Placér Jobtilfredshed under “Y, Response”, Motivation under “X,

Factor”.

|  |  |
| --- | --- |
| Chi i anden test | |
| Forudsætninger | De forventede celleværdier skal være større end 5, for at sikre kvaliteten af testen. |
| Hypotese | Der er uafhængighed dvs. ingen sammenhæng mellem Jobtilfredshed og motivation  Der er afhængighed dvs. sammenhæng mellem mellem Jobtilfredshed og motivation |
| Teststørrelse | 208.330 |
| P-værdi/signifikanssandsynlighed | <0.01% |
| Signifikansniveau | 5% |
| Teknisk konklusion | Da p-værdien <0.01% er mindre end signifikansniveauet på 5%, kan vi forkaste nulhypotesen. |
| Ikke-teknisk konklusion | Der er afhængighed dvs. sammenhæng mellem mellem Jobtilfredshed og motivation. |
|  | |

c). Gør rede for om der er statistisk belæg for en nulhypotese om, at der ikke er nogen sammenhæng mellem jobtilfredshed og følelsen af at være stresset på sit arbejde (Stress) (brug = 5% som signifikansniveau). Angiv i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.

JMP-vink: “Analyze” -> “Fit Y by X”. Placér Jobtilfredshed under “Y, Response”, Stress under “X, Fa-

ctor”.



|  |  |
| --- | --- |
| Chi i anden test | |
| Forudsætninger | De forventede celleværdier skal være større end 5, for at sikre kvaliteten af testen. |
| Hypotese | Der er uafhængighed dvs. ingen sammenhæng mellem Jobtilfredshed og stress  Der er afhængighed dvs. sammenhæng mellem mellem Jobtilfredshed og stress |
| Teststørrelse | 18.246 |
| P-værdi/signifikanssandsynlighed | 0.01% |
| Signifikansniveau | 5% |
| Teknisk konklusion | Da p-værdien 0.01% er mindre end signifikansniveauet på 5%, kan vi forkaste nulhypotesen. |
| Ikke-teknisk konklusion | Der er afhængighed dvs. sammenhæng mellem mellem Jobtilfredshed og stress. |
|  | |

HD Dataanalyse, efterår 2018 Opgaveark 8



d). Gør rede for om der er statistisk belæg for en nulhypotese om, at der ikke er nogen sammenhæng mellem jobtilfredshed og hvor stresset man føler sig på sit arbejde, hvis vi kun ser på dem, der føler sig stressede2 (brug = 5% som signifikansniveau). Angiv i den forbindelse antagelser, hypoteser, signifikanssandsynlighed samt teknisk og let forståelig konklusion.

JMP-vink: “Rows” -> “Data Filter” og “Analyze” -> “Fit Y by X”.

|  |  |
| --- | --- |
| Chi i anden test | |
| Forudsætninger | De forventede celleværdier skal være større end 5, for at sikre kvaliteten af testen. |
| Hypotese | Der er uafhængighed dvs. ingen sammenhæng mellem Jobtilfredshed og stress, når vi kun ser på på dem, der føler sig stressede  Der er afhængighed dvs. sammenhæng mellem mellem Jobtilfredshed og stress, når vi kun ser på på dem, der føler sig stressede |
| Teststørrelse | 0.559 |
| P-værdi/signifikanssandsynlighed | 45.48% |
| Signifikansniveau | 5% |
| Teknisk konklusion | Da p-værdien 45.48% er større end signifikansniveauet på 5%, kan vi ikke forkaste nulhypotesen. |
| Ikke-teknisk konklusion | Der er uafhængighed dvs. ingen sammenhæng mellem mellem Jobtilfredshed og stress, når vi kun ser på på dem, der føler sig stressede. |
|  | |

Gør rede for hvordan ovenstående analyse også kan laves som en sammenligning af to andele og foretag denne analyse.

JMP-vink: “Rows” -> “Data Filter” og “Analyze” -> “Fit Y by X”.

|  |  |
| --- | --- |
| Test af 2 andele, tosidet alternativ hypotese, vi benytter unpooled estimat her. | |
| Forudsætninger | skal være større end 9  skal være større end 9 |
| Hypotese |  |
| Teststørrelse | 0.6693 |
| P-værdi/signifikanssandsynlighed | 50.33% |
| Signifikansniveau | 5% |
| Teknisk konklusion | Da p-værdien/ signifikanssandsynligheden 50.33% er større end 5% signifikansniveauet forkaster vi ikke nulhypotesen. |
| Ikke-teknisk konklusion | Andelen af tilfredse bland i nogen grad stressede er den samme som andelen af tilfredse bland i høj grad stressede. |
| Vi kan se vi får samme p-værdi som i Chi i anden testet, når vi bruger pooled test: | |

* Dvs. at vi i dette delspørgsmål ser bort fra svarkategorien “I mindre grad eller slet ikke” af variablen Stress.



5/7

HD Dataanalyse, efterår 2018 Opgaveark 8



Kortfattede talløsninger

Opgave 1

a): *H*0 : *Mek* = *El* = *Bager* = *Maler* = *Murer* = *T* ø*mrer* = *Gulv* = *Blik*

*Ha* :Mindst to af middelværdierne *Mek; El ; Bager ; Maler ; Murer ; T* ø*mrer ; Gulv; Blik* er ikke ens

Signifikansssh. = 38*;*97%

c): *H*0 : *Mek* = *El* = *Bager* = *Maler* = *Murer* = *T* ø*mrer* = *Gulv* = *Blik*

*Ha* :Mindst to af middelværdierne *Mek; El ; Bager ; Maler ; Murer ; T* ø*mrer ; Gulv; Blik* er ikke ens

Signifikansssh. = 0*;*05%

d): *H*0 : *Mek* = *El* = *Maler* = *Murer* = *T* ø*mrer* = *Gulv* = *Blik*

*Ha* :Mindst to af middelværdierne *Mek; El ; Maler ; Murer ; T* ø*mrer ; Gulv; Blik* er ikke ensSignifikansssh. = 0*;*05%

*H*0 : *Mek* = *El* = *Bager* = *Maler* = *Murer* = *T* ø*mrer* = *Gulv*

*Ha* :Mindst to af middelværdierne *Mek; El ; Bager ; Maler ; Murer ; T* ø*mrer ; Gulv* er ikke ensSignifikansssh. = 11*;*30%

Opgave 2

a): *H*0 : *Mek* = *El* = *Bager* = *Maler* = *Murer* = *T* ø*mrer* = *Gulv* = *Blik*

*Ha* :Mindst to af middelværdierne *Mek; El ; Bager ; Maler ; Murer ; T* ø*mrer ; Gulv; Blik* er ikke ens

Signifikansssh. = 0*;*29%

c): *H*0 : *Mek* = *Bager* = *Maler* = *Murer* = *T* ø*mrer* = *Gulv*

*Ha* :Mindst to af middelværdierne *Mek; Bager ; Maler ; Murer ; T* ø*mrer ; Gulv* er ikke ensSignifikansssh. = 16*;*23%

Opgave 3

a): *H*0 : Der er uafhængighed mellem Jobtilfredshed og Region *Ha* :Der er afhængighed mellemJobtilfredshedogRegionSignifikansssh. = 41*;*25%

b): *H*0 : Der er uafhængighed mellem Jobtilfredshed og Motivation *Ha* :Der er afhængighed mellemJobtilfredshedogMotivationSignifikansssh. = Mindre end 0*;*01%



c): *H*0 : Der er uafhængighed mellem Jobtilfredshed og Stress *Ha* :Der er afhængighed mellemJobtilfredshedogStressSignifikansssh. = 0*;*01%

d): *H*0 : Der er uafhængighed mellem Jobtilfredshed og Stress *Ha* :Der er afhængighed mellemJobtilfredshedogStressSignifikansssh. = 45*;*48%

*H*0 : *p*I høj grad *p*I nogen grad= 0 *Ha* : *p*I høj grad *p*I nogen grad,0;

Signifikansssh. = 36*;*39%

7/7