

Labbrapport i Statistik

OPPOSITION

Kurskod: 732G53

Duy Thai Pham
Viet Tien Trinh



Avdelningen för Statistik och maskininlärning
Institutionen för datavetenskap
Linköpings universitet
2024-09-13

Innehåll

| | |
|------------------------|----------|
| Introduktion | 1 |
| Uppgifter | 2 |
| Uppgifter A) | 2 |
| i) | 2 |
| ii) | 2 |
| iii) | 2 |
| Uppgifter B) | 2 |
| i) | 2 |
| ii) | 2 |
| iii) | 2 |
| iv) | 3 |
| Uppgifter C) | 3 |
| i) | 3 |
| ii) | 3 |
| Uppgifter D) | 3 |
| Sammanfattning | 4 |

Introduktion

I denna genomgång kommer vi att presentera vad vi anser vara brister eller felaktigheter i grupp 16 rapport.

Uppgifter

Uppgifter A)

i)

Tabellen från grupp 16 är inte en riktigt tabell utan output från R kod och de numeriska variabler är inte tydligt särskiljanda som diskret eller Kontinuerlig. Samtidigt var skala av andra, tredje och sista variabler fel.

ii)

I figur 3 och 4 används fel diagram, istället för spridningsdiagram som skulle passa till “numerisk” variabeltyp, så valde grupp 16 fioldiagrammen för att visualisera samband mellan 2 numeriska variabler. Samtidigt saknas det tolkning för alla diagram.

iii)

På grund av att det saknas tolkning, så finns det inte heller användbara information för att motivera varför variablerna bör inkluderas in i modellen. Motiveringen är därmed inte tydligt presenterad och det är bara en slutsats att inkludera samtliga variabler i regressionsmodellen som dras utan anledning.

Uppgifter B)

i)

Alla skattningsvärden är korrekta men grupp 16 skapade inte en riktigt tabell utan output från R och skattningsparametrarna tolkas inte. Det finns inte heller medelfel, t-värde eller p-värde i deras svar.

ii)

Intervallskattning av den första och sista variabler är korrekt men ingen tolkning finns. Beräkningar visas inte heller.

iii)

Rapporten visar inte en förväntad ANOVA-tabell utan bara en detaljerad utskrift som skapas med hjälp av summary()-funktionen i R. Inga hypoteser presenteras. Kritiska värden som fås fram är felaktigt. Dessutom finns det ingen testvariabel och p-värdet som fås fram är också felaktigt.

I slutändan används funktionen drop1, vilket verkar visa rätt väg men återigen finns det ingen tolkning eller lämpliga slutsatser som dras.

iv)

Enligt tidigare uppgifter har grupp 16 visualiserat 4 fioldiagram för 4 variabler, därav 2 kvantitativa variabler och 2 kvalitativa variabler. Dock, på denna del har de valt rätt att testa ta bort `property_type` och `economic_conditions`, vilket är 2 kvalitativa variablerna som de har i sitt data. Inga hypoteser presenteras men testet som genomförs i R är korrekt med relevant tolkning. Dock skulle lite mer detaljerat tolkning behövas för att tydliggöra. En observation är att en kommentar från R inte syns i rapporten, vilket kan vara fel från programmet. Ingen signifikansnivå väljs men om det är som det står, så kanske borde en högre signifikansnivå väljas.

I denna del av uppgiften finns det en funktion som skulle passa till tidigare del, vilket är `reviderad_modell <- step(linreg, direction = backward)`. Dock, i slutändan kan grupp 16 redovisa vilka variabler som bör inkluderas men det krävs tydligare resonemang och tolkning för beslutet.

Uppgifter C)

i)

Med denna uppgift måste vi bedöma varje regressionsantagande med hjälp av lämpliga visualiseringar och grupp 16 inte visar det tydligt. Residualerna bedöms inte är normalfördelade genom histogrammet. Residualerna verkar ha lika varians och inte finns någon samband linjärt kan man se det genom spridningsdiagrammet men fortfarande sakna två parallella linjer som ett lock och en botten. Samtidigt finns inte tolkning under diagrammet.

ii)

3 extremvärden identifierades inte, samtidigt identifierade grupp 16 inte vilken/vilka förklarande eller respon-svariabel observationen är ett extremvärde.

Uppgifter D)

Saknade beräkna utvärderingsmått R^2 för att motivera linjär modellen är bra och lämpligt eller inte.

Sammanfattning

Sammantaget är denna rapport inte välgjord. Informationen är inte tydligt förklarad, kommandona från R döljas inte och beskrivs inte hur de fungera så att läsare enkelt kan följa dem. Använder några onödvisande diagrammet i uppgiften C och rapporten finns inte referenser/källa som grupp 16 hänvisade på.