

# Met4 Hjemmeeksamen: Kredittkort og mislighold

Insitutt for Foretaksøkonomi  
NHH

Fra 25.04.2018 kl 09:00  
Til 27.04.2018 kl 14:00

## Introduksjon

I dette caset skal dere jobbe med data for kredittkortbruk, og bruke enkle modeller til å vurdere sannsynligheten for mislighold av kredittkortgjeld. Datasettet er hentet fra University of California, Irvine sin database med maskinlæringsdatasett<sup>1</sup>, og inneholder 30 000 observasjoner av kredittkortbrukere i Taiwan i 2005. Datasettet har en binær variabel (dummyvariabel) som viser hvorvidt et individ har misligholdt kredittkortgjelden eller ikke, samt alder, kjønn, sivilstatus, utdanning og betalingshistorikk. Se nettsiden i fotnoten, eller Yeh and Lien (2009) for en detaljert oversikt over variablene. Merk at det øvrige innholdet i denne artikkelen ikke er pensum i Met4.

Forklaringene til enkelte variabler i artikkelen og på nettsiden kan være mangelfull. Gjør en selvstendig vurdering på hvordan dere håndterer dette.

## Referanser

**Yeh, I-Cheng, and Che-hui Lien.** 2009. "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients." *Expert Systems with Applications*, 36(2): 2473–2480.

---

<sup>1</sup><https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients>

## Spørsmål

- (1) Presenter deskriptiv statistikk av datasettet. Fokuser på de egenskapene ved datasettet som er relevant for resten av eksamenen.
- (2) Estimer en regresjonsmodell (dvs. OLS, Logit og/eller Probit) med “default payment next month” som *responsvariabel*. Bruk kjønn, oppnådd “graduate”-utdanning, og status som gift som *forklaringsvariabler*. Drøft på hvilken måte forklaringsvariablene forklarer sannsynligheten for mislighold.
- (3) Anta at du er rådgiver for en bedrift på Taiwan som selger kredittkort. Bedriften planlegger å lansere en stor salgskampanje, der gateselgere henvender seg til forbipasserende for å tilby dem kredittkort. Ta høyde for følgende:
1. Bedriften ønsker ikke å tilby kredittkort til individer med misligholdssannsynlighet høyere enn 25%.
  2. Det finnes ingen register hvor bedriften kan observere nye kunders betalingshistorikk.
  3. Lavere misligholdssannsynlighet gir høyere forventet profit.
- Bedriften ønsker dine råd om *hvem* av de forbipasserende selgerne skal henvende seg til. Analysen dere utfører skal være så presis som mulig. Bruk det utdelte datasettet. Estimer en regresjonsmodell etter eget valg og med de forklaringsvariablene dere mener er relevant. Gjør de forutsetningene dere finner nødvendig, men som støtter hensikten med analysen. Forklar og begrunn valgene dere gjør.
- Konklusjonene fra analysen må oppsummeres slik at de lett kan kommuniseres til et stort salgskorps uten utdanning i hverken økonomi eller statistikk, men som er gode på raske møter med mennesker.
- (4) Skriv et kort notat til direktøren i selskapet. Vurder bedriftens salgsstrategi som du presenterte fra oppgave (3) opp mot målet om profittmaksimering. Bør bedriften ha tillit til konklusjonene? Kom med anbefaling om hvordan selskapet bør gå frem. Bruk datasettet for å understøtte poengene deres om nødvendig.

## A R-syntax

Dersom vi estimerer en logit/probit-modell, kan vi predikere sannsynligheten for suksess ved bruk av f.eks. kommandoen `predict.glm(reg, newdata=newdat, type="response")`. Denne kommandoen gir oss ikke konfidensintervall til den predikerte sannsynligheten. Dersom vi ønsker dette, kan vi gjøre følgende:

```
1 # Installerer devtools-pakken, som gjør at vi kan laste ned
2 # innhold fra GitHub
3 install.packages("devtools")
4
5 # Aktiverer devtools:
6 library(devtools)
7
8 # Bruker devtools til å installere en pakke som inneholder
9 # funksjonen vi vil ha:
10 install_github("jgabry/QMSS-package")
11
12 # Når denne er installert kan vi laste den inn som vanlig:
13 library(QMSS)
14
15 # Antar at vi har en dataframe med navn df i minnet, med
16 # variablene Y og X1 og X2. Kjører en logit (merk bruken
17 # av I(X1^2) for å få med kvadratledd, samt X1:X2 som gir
18 # interaksjonen av disse to variablene):
19 reg <- glm(Y ~ X1 + X2 I(X1^2) + X1:X2,
20           family=binomial(link = 'logit'), data = df)
21
22 # Lager en ny dataframe. Her ser vi på alle kombinasjonene
23 # av X1 og X2 fra 1 til 4:
24 newdat <- expand.grid(X1 = 1:4, X2 = 1:4)
25
26 # Bruker funksjonen predProb. Denne trenger først
27 # regresjonsmodellen (kan være logit eller probit), deretter
28 # en dataframe med forklaringsvariabler ved de verdiene
29 # vi er interessert i å predikere. De to siste argumentene
30 # avgjør om vi vil ha konfidensintervall – og i såfall hvilket
31 # signifikansnivå.
32 pred <- predProb(reg, newdat, ci = TRUE, level = .95)
33
34 # Kikker på prediksjonene med konfidensintervall:
35 head(pred)
```

## Administrative bestemmelser

- Hjemmeeksamen i Met4 må leveres i grupper på 2, 3, eller 4 studenter.
- Se § 9 i FORSKRIFT OM EKSAMEN VED NHH (FULLTIDSSTUDIENE), og del 2 i UTFYLLENDE BESTEMMELSER TIL EKSAMENSFORSKRIFTEN for regelverk.
- Det er ikke tillatt å diskutere eksamen med studenter utenfor din gruppe etter at oppgavesettet er frigitt.
- Besvarelsene vil bli rettet iht rubrikk postet på Canvas.
- Du kan besvare eksamen på norsk eller engelsk.
- Send en mail til *både* Ole-Petter Moe Hansen (s9705@nhh.no) og Håkon Otneim (s12203@nhh.no) ved spørsmål til oppgaven. Tilleggsinformasjon av betydning vil bli lagt ut på kursets hjemmeside på Canvas.
- Rapporten må ikke være lengre inn 10 sider. Tabeller, figurer og referanser er inkludert i de 10 sidene. Dersom rapporten har en forside uten noen form for svar på oppgavene kan forsiden komme i tillegg til de 10 sidene. Innholdsfortegnelse er ikke nødvendig. Prioriter hva dere tar med i rapporten!
- Rapporten skal skrives med fonten Times New Roman, størrelse 12 og linjeavstand 1.15. Tekst i figurer og tabeller kan ha font ned til størrelse 9.
- Eksamen administreres i Wiseflow. Besvarelsen må leveres som en enkelt pdf-fil. Andre format (f.eks. .doc, .docx eller .R) er ikke akseptert.