
Arbeidskrav nr. 2: MET 3590 Metode og økonometri

Innlevering: Via It's Learning

Datsett: *kyllingdata.xls* (tilgjengelig i It's Learning)

Tekniske problemer? Hvis du like før fristens utløp har tekniske problemer (f.eks. med at.bi eller It's Learning), så kan du sende svarene i en mail til met3590.arbeidskrav@gmail.com. VIKTIG: En evt. slik mail må sendes *innen* fristens utløp!

- For å kunne avgi eksamen i MET 3590 Metode og økonometri må begge arbeidskravene bestås
- Hvert arbeidskrav regnes som bestått hvis 30% eller flere av spørsmålene er korrekt besvart
- Et korrekt svar gir 1 poeng
- Poeng trekkes IKKE for gale svar (det vil det derimot bli gjort på eksamen)
- Svarene må avgis innen to uker etter publikasjon av arbeidskravet. I løpet av de to ukene så kan kandidaten maksimalt gjennomføre arbeidskravet 8 ganger
- Fasiten til arbeidskravene publiseres i It's Learning etter fristen for innlevering

For å kunne besvare oppgavene 1 til 10, så må først analysene i caset under gjennomføres med statistisk programvare (f.eks. STATA). For å kunne besvare oppgavene 11 til 20, så er det ikke nødvendig å løse caset, eller å bruke statistisk programvare.

Case: Virkningen av ulike fórtyper på vekten til kyllinger

Et internasjonalt selskap (“NyeEnzymer”) som driver med forskning og utvikling av enzymer (biologiske katalysatorer eller “turboer”) har utviklet fire nye fórtyper for kyllinger. For å studere virkningen av de fire fórtypene på vekten til kyllinger, så veies vekten til 50 kyllinger ved fødsel og etter en periode med fóring. For de fleste kyllingene er perioden med fóring 21 dager lang, men for noen er den betydelig mindre. I perioden med fóring ble hver enkelt kylling fóret med kun én fórtipe.

Datsettet fra studiet er inneholdt i *kyllingdata.xls*:

- $vekt_i$: Vekten (målt i gram) til kylling i etter fóeringsperioden
- tid_i : Antall dager med fóring siden fødsel for kylling i
- $fortype_i$: Fórtipe gitt til kylling i . Verdien 1 betyr at kyllingen ble fóret med type 1, verdien 2 betyr at kyllingen ble fóret med type 2, verdien 3 betyr at kyllingen ble fóret med type 3 og verdien 4 betyr at kyllingen ble fóret med type 4
- $vektfdsl_i$: Vekten (målt i gram) til kylling i ved fødsel

Du skal nå gjøre følgende:

A. Lag fire dummy variabler $d1_i$, $d2_i$, $d3_i$ og $d4_i$ for hver variant av fórtypen: $d1_i$ for verdien 1, $d2_i$ for verdien 2, $d3_i$ for verdien 3 og $d4_i$ for verdien 4. Beregn også korrelasjonsmatrisen til dummyvariablene, tid_i , $vektfdsl_i$ og $vekt_i$ (dvs. beregn utvalgskorrelasjonene mellom disse variablene)

B. Beregn modellene:

$$vekt_i = B_1 + B_2 tid_i + u_i \quad (1)$$

$$vekt_i = B_1 + B_2 vektfdsl_i + u_i \quad (2)$$

C. Beregn en modell med $vekt$ på venstresiden og *kun* dummy variablene på høyresiden (dvs. utelat konstantleddet fra høyresiden). [Hint til STATA: Husk å velge “Suppress constant term”.]

D. Beregn en modell med $vekt_i$ som venstresidevariabel og følgende på høyresiden: Konstantleddet, $d2_i$, $d3_i$, $d4_i$, tid_i og $vektfdsl_i$. [Hint til STATA: Husk å *ikke* velge “Suppress constant term”.]

E. Beregn en modell *uten* de ikke-signifikante variablene i forrige oppgave (men behold konstantleddet selv om det er ikke-signifikant). Bruk en tosidig t -test for hver variabel med et signifikansnivå lik 5% for å avgjøre om en variabel er signifikant eller ikke.

F. Beregn en modell med den naturlige logaritmen til $vekt_i$ som venstresidevariabel og følgende variabler på høyresiden: Konstantleddet, $d2_i$, $d3_i$, $d4_i$, tid_i og $\ln(vektfdsl_i)$

G. Beregn følgende modell:

$$\ln(vekt_i) = B_1 + B_2(d2_i + 4 \cdot d3_i + 3 \cdot d4_i) + B_5 tid_i + B_6 \ln(vektfdsl_i) + u_i$$

H. Beregn en modell med gjennomsnittlig daglig vektøkning siden fødsel (det vil si $(vekt_i - vektfdsl_i)/tid_i$) som venstresidevariabel og *kun* dummy variablene som høyresidevariabler [Hint til STATA: Husk å velge “Suppress constant term”.]

Oppgavene

1. Hvilket av følgende utsagn om variablenes målenivå er *galt*:

- (a) Variabelen $vekt$ er på forholdstallsnivå
- (b) Variabelen tid er på forholdstallsnivå
- (c) Variabelen $vektfdsl$ er på forholdstallsnivå
- (d) Variabelen $fortype$ er på forholdstallsnivå

2. Ta utgangspunkt i Case oppgave F, og beregn et tosidig 99% konfidensintervall for parameteren til $\ln(vektfdsl)$:

- (a) Ca. $[-7,61 ; 1,99]$
- (b) Ca. $[-2,81 ; 2,81]$

- (c) Ca. $[-2, 69 ; 2, 69]$
- (d) Ca. $[-7, 40 ; 1, 78]$
3. Hvilken av følgende beregnede modeller har størst forklaringskraft hvis man tar hensyn til antall parametre som er beregnet?:
- (a) $\widehat{vekt}_i = b_1 + b_2 tid_i$
- (b) $\widehat{vekt}_i = b_1 + b_2 vektfdsl_i$
- (c) $\widehat{vekt}_i = b_1 + b_3 d3_i + b_4 d4_i + b_5 tid_i$
- (d) $\widehat{vekt}_i = b_1 + b_2 d2_i + b_3 d3_i + b_4 d4_i + b_5 tid_i + b_6 vektfdsl_i$
4. Med utgangspunkt i resultatene fra Case oppgave D, hvilket av følgende utsagn er korrekt:
- (a) Ingen av variablene er signifikante på 5%
- (b) Variablene $d3_i, d4_i$ og tid_i er signifikante på 5%
- (c) Variablene $d3_i, d4_i$ og tid_i er signifikante på 1%
- (d) Variablene $d3_i, d4_i, tid$ og $vektfdsl_i$ er signifikante på 10%
5. Beregningene i Case oppgave H tyder på at gjennomsnittlig daglig vektøkning som følge av fôring med type 4 er:
- (a) Ca. 5 gram
- (b) Ca. 8 gram
- (c) Ca. 11 gram
- (d) Ca. 9 gram
6. Hvilke restriksjoner må pålegges modellen beskrevet i Case oppgave F for å få følgende modell?
- $$\ln(vekt_i) = B_1 + B_2(d2_i + 4 \cdot d3_i + 3 \cdot d4_i) + B_5 tid_i + B_6 \ln(vektfdsl) + u_i$$
- (a) $B_2 = 1, B_3 = 4$ og $B_4 = 3$
- (b) $B_3 = 4B_2$ og $B_4 = 3B_2$
- (c) $B_2 = 1 + 4 + 3$
- (d) $B_2 > 0, B_3 > 0$ og $B_4 > 0$
7. Modellene fra Case oppgave F og G kan brukes til å teste restriksjonene fra forrige oppgave. Hva er resultatet av denne testen?
- (a) Hypotesen om at restriksjonene holder beholdes på alle de vanlige signifikansnivåene (dvs. 1%, 5% og 10%)
- (b) Hypotesen om at restriksjonene holder beholdes på et signifikansnivå lik eller lavere enn 1%, men ikke på 5% og 10%
- (c) Hypotesen om at restriksjonene holder beholdes på signifikansnivå lik eller lavere enn 5%, men ikke på 10%

- (d) Hypotesen om at restriksjonene holder forkastes på alle de vanlige signifikansnivåene (dvs. 1%, 5% og 10%)
8. Ta utgangspunkt i den beregnede modellen fra Case oppgave D. Hva er den anslåtte vekten til kyllinger som føres med type 3 gitt at tid_i settes lik 21 og at $vekt\ f\ d\ s\ l_i$ settes lik sitt utvalgsgjennomsnitt 41,06?
- (a) Ca. 46 gram
 (b) Ca. 87 gram
 (c) Ca. 267 gram
 (d) Ca. 613 gram
9. Ta utgangspunkt i den beregnede modellen fra Case oppgave E. Hva er den anslåtte vekten til kyllinger som føres med type 3 gitt at tid_i settes lik 21?
- (a) Ca. 49 gram
 (b) Ca. 90 gram
 (c) Ca. 270 gram
 (d) Ca. 616 gram
10. Du har nå gjennomført en del analyser omkring sammenhengen mellom fórtype og vekten til kyllinger. Hvilken av følgende konklusjoner er riktig?
- (a) Fórtype 1 er den som har størst effekt på vekt
 (b) Fórtype 2 har dobbel så stor effekt på vekt som fórtype 3
 (c) Fórtype 3 har størst effekt på vekt
 (d) Fórtype 4 har en tredel så stor effekt som fórtype 3

11. Ta utgangspunkt i den beregnede modellen

$$\hat{Y} = 1,08 + 0,19X_2 + 33,64 \ln X_3 + 68,11 \frac{1}{X_4}.$$

Hvilket av følgende utsagn om sammenhengen mellom X_2 og Y er riktig?

- (a) Hvis X_2 øker med én enhet, så øker Y i gjennomsnitt med 19% under den forutsetning at de andre X -ene ikke endrer seg
 (b) Hvis X_2 øker med 1%, så øker Y i gjennomsnitt med 19% under den forutsetning at de andre X -ene ikke endrer seg
 (c) Hvis X_2 øker med 1%, så øker Y i gjennomsnitt med 0,19 under den forutsetning at de andre X -ene ikke endrer seg
 (d) Hvis X_2 øker med én enhet, så øker Y i gjennomsnitt med 0,19 under den forutsetning at de andre X -ene ikke endrer seg
12. Ta utgangspunkt i modellen fra forrige oppgave. Hvilket av følgende utsagn om sammenhengen mellom X_3 og Y er riktig?

- (a) Hvis X_3 øker med én enhet, så øker Y i gjennomsnitt med ca. 0,3364 under den forutsetning at de andre X -ene ikke endrer seg
 - (b) Hvis X_3 øker med 1%, så øker Y i gjennomsnitt med ca. 0,3364
 - (c) Hvis X_3 øker med 1%, så øker Y i gjennomsnitt med ca. 0,3364 under den forutsetning at de andre X -ene ikke endrer seg
 - (d) Hvis X_3 øker med 1%, så øker Y i gjennomsnitt med ca. 33,64 under den forutsetning at de andre X -ene ikke endrer seg
13. Ta utgangspunkt i modellen fra oppgave 11. For et nivå på X_4 lik 10, hvilket av følgende utsagn er riktig hvis X_4 øker med 1 enhet?
- (a) Y faller i gjennomsnitt et sted mellom 0,6 og 0,7 enheter under den forutsetning at de andre X -ene ikke endrer seg
 - (b) Y øker i gjennomsnitt et sted mellom 60 og 70 enheter under den forutsetning at de andre X -ene ikke endrer seg
 - (c) Y øker i gjennomsnitt et sted mellom 6 og 7 enheter under den forutsetning at de andre X -ene ikke endrer seg
 - (d) Y øker i gjennomsnitt et sted mellom 0,6 og 0,7 enheter under den forutsetning at de andre X -ene ikke endrer seg

14. Betrakt den beregnede modellen

$$\hat{Q} = b_1 + b_2P + b_3X + b_4R, \quad n = 44,$$

og ta utgangspunkt i testen

$$H_0 : B_2 = 0, B_3 = 0 \text{ og } B_4 = 0,$$

$$H_A : \text{Én eller flere av påstandene i } H_0 \text{ er gale.}$$

En versjon av testuttrykket til denne testen inneholder R^2 til både modellen med og uten restriksjoner. Hva er R^2 til modellen med restriksjoner?

- (a) Mindre enn 0
 - (b) Nøyaktig lik 0
 - (c) Større enn 0, men mindre enn R^2 til modellen uten restriksjoner
 - (d) Nøyaktig lik R^2 til modellen uten restriksjoner
15. Ta utgangspunkt i testen fra forrige oppgave. Hva er sannsynligheten for at testverdien (dvs. verdien til testuttrykket eller testobservator) er større enn 2,226 hvis H_0 er riktig?
- (a) Ca. 1%
 - (b) Ca. 2,5%
 - (c) Ca. 5%
 - (d) Ca. 10%

16. La testverdien i forrige oppgave være 2,839. Hva er p -verdien?

- (a) Ca. 10%
- (b) Ca. 5%
- (c) Ca. 2,5%
- (d) Ca. 1%

17. Betrakt følgende beregnede modell:

$$\hat{Y} = 0,9815 + 0,5714X - 0,0037X^2, \quad n = 29,$$

(0,3926)
(0,1270)
(0,0044)

hvor tallene i parentes er standardfeilene til de beregnede verdiene. Du ønsker å studere innvirkningen av X på Y ved hjelp av t -tester. Hvilket av følgende utsagn er korrekt på de vanlige signifikansnivåene?

- (a) Ifølge t -testene så er det ingen sammenheng mellom X og Y
 - (b) Ifølge t -testene så er det en invers (dvs. resiprokk) sammenheng mellom X og Y
 - (c) Ifølge t -testene så er det en lineær sammenheng mellom X og Y
 - (d) Ifølge t -testene så er det en ikke-lineær sammenheng mellom X og Y
18. En bedrift ønsker å sammenligne effekten av tre ulike internettreklamer, og bestemmer seg for å gjøre dette ved å studere antall klikk på reklamen i løpet av dagen. La Y_i være antall klikk i løpet av dag i , og la $D1_i$, $D2_i$ og $D3$ være dummy variabler. $D1_i$ er lik 1 for de dagene reklame nr. 1 ble brukt og 0 ellers, $D2_i$ er lik 1 for de dagene reklame nr. 2 ble brukt og 0 ellers, mens $D3_i$ er lik 1 for de dagene reklame nr. 3 ble brukt og 0 ellers. Kun én av reklamene ble brukt per dag.

Hvilken av følgende modeller kan *ikke* beregnes?

- (a) $Y_i = B_1 + B_3D2_i + B_4D3_i + u_i$
- (b) $Y_i = B_1 + B_2D1_i + B_4D3_i + u_i$
- (c) $Y_i = B_2D1_i + B_3D2_i + B_4D3_i + u_i$
- (d) $Y_i = B_1 + B_2D1_i + B_3D2_i + B_4D3_i + u_i$

19. Ta utgangspunkt i den beregnede modellen

$$\hat{Y}_i = 58,73 - 9,21D2_i + 38,44D3_i, \quad R^2 = 0,08, \quad n = 68,$$

hvor Y_i , $D2_i$ og $D3_i$ er definert som i forrige oppgave. Hva er anslått antall klikk i løpet av en dag for reklame nr. 1?

- (a) Ca. 59 klikk
 - (b) Ca. 38 klikk
 - (c) Ca. 97 klikk
 - (d) Ca. 50 klikk
20. Ta utgangspunkt i den beregnede modellen i forrige oppgave, og test om gjennomsnittlig antall klikk er ulik for de tre reklamene. Hva er resultatet av testen?

- (a) Gjennomsnittlig antall klikk er ulik for de tre reklamene, hvis vi bruker et signifikansnivå på 1%
- (b) Gjennomsnittlig antall klikk er ulik for de tre reklamene, hvis vi bruker et signifikansnivå på 5%
- (c) Gjennomsnittlig antall klikk er ulik for de tre reklamene, hvis vi bruker et signifikansnivå på 10%
- (d) Gjennomsnittlig antall klikk er lik for de tre reklamene ifølge de tre vanligste signifikansnivåene (dvs. 1%, 5% og 10%)