

NORGES HANDELSHØYSKOLE

Eksamen høst 2015

Kurskode: INT010

Tittel: Anvendt metode

Dato: 10.11.2015 Kl. 09.00-12.00

Faglærer går ikke rundt i eksamenslokalene, men kan kontaktes av eksamensvakten på
tlf. 59612/996 99 305

Hjelpemidler til eksamen:

Hjelpemidler tillatt: JA, alle trykte/egenskrevne

Kalkulator: JA [i tråd med retningslinjer for bruk av kalkulator, jf. utfyllende bestemmelser til Forskrift om eksamen ved Norges Handelshøyskole (fulltidsstudiene).]

***Alle delspørsmål i oppgavesettet teller likt.
Husk at dere må skrive slik at sensor kan lese det som er skrevet.***

Eksamensoppgaven består av 6 sider, inkludert denne.

Oppgave 1

Ordet regresjon stammer fra Sir Francis Galtons studie av sammenhengen mellom foreldres høyde og deres barns høyde som voksne. Analysen ble publisert i 1886 med tittelen *Regression towards Mediocrity in Hereditary Staure*. Hovedfunnet var at barn forventes å få en høyde som er nærmere gjennomsnittet for hele populasjonen enn høyden til foreldrene. Dette er et statistisk fenomen som gjør seg gjeldende i mange sammenhenger og er i dag kjent som «regression towards the mean».

I 1886 var multippel regresjon ikke utviklet. Galton normaliserte derfor kvinners høyde til menns høyde ved å multiplisere kvinners høyde med en faktor på 1,08. Deretter brukte han en grafisk teknikk til å trekke en «regresjonslinje» med barnas høyde som avhengig variabel og gjennomsnittet av foreldrenes høyde som forklaringsvariabel. Han hadde samlet inn data for 928 barn fra 205 familier. Denne oppgaven er basert på et datasett med 197 av disse familiene som er rekonstruert fra Galtons notater og gjort tilgjengelig av James A. Hanley ved MacGill University.

Datasettet inneholder følgende variabler:

| | |
|------------|---|
| Height | Barnets høyde (som voksen) i tommer |
| Male | Dummyvariabel som er én for menn og null for kvinner |
| Height_adj | $\text{Height} * \text{Male} + \text{Height} * 1,08 * (1 - \text{Male})$ |
| Father | Fars høyde i tommer |
| Mother | Mors høyde i tommer |
| Midparent | Foreldrenes middelhøyde definert som $(\text{Father} + 1,08 * \text{Mother}) / 2$ |
| Kids | Antall barn i familien |

Det er 898 barn inkludert i datasettet. Merk at 1 tomme=1 inch=2,54 cm.

| | Father | Mother | Son's order of height | Daughter's order of height |
|----|--------|-----------|------------------------|------------------------------|
| 1 | 18.5 | 7.0 | 13.2 | 9.2, 9.0, 9.0 |
| 2 | 18.5 | 6.5 | 13.5, 12.5 | 8.5, 8.5 |
| 3 | 18.0 | about 4.0 | 11.0 | 8.0 |
| 4 | 18.0 | 4.0 | 10.5, 8.5 | 7.0, 4.5, 3.0 |
| 5 | 18.0 | 1.5 | 12.0, 9.0, 2.0 | 6.5, 2.5, 2.5 |
| 6 | 14.0 | 8.0 | | 9.5 |
| 7 | 14.0 | 2.0 | 16.5, 14.0, 13.0, 13.0 | 10.5, 4.0 |
| 8 | 14.0 | 6.5 | | 10.5, 8.0, 6.0 |
| 9 | 14.5 | 6.0 | | 6.0 |
| 10 | 14.0 | 8.5 | | 5.5 |
| 11 | 14.0 | 2.0 | 14.0, 10.0 | 8.0, 7.0, 7.0, 6.0, 3.5, 3.0 |
| 12 | 14.0 | 1.0 | | 5.0 |

Figur 1. Første side i Galtons notatbok. Familiene er sortert etter fars høyde.

Tabell 1. Sammenhengen mellom barns høyde og foreldres høyde

| | Modell 1 | Modell 2 | Modell 3 | Modell 4 | Modell 5 | Modell 6 |
|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|--------------------|
| Utvalg | Begge kjønn | Begge kjønn | Sønner | Døtre | Begge kjønn | Begge kjønn |
| Constant | 18,8** (2,84) | 15,3** (2,75) | 19,4** (4,13) | 16,6** (3,62) | 16,2** (2,79) | 64,1*** (0,101) |
| Midparent | 0,729** (0,0410) | | | | | |
| Father | | 0,406** (0,0292) | 0,412** (0,0467) | 0,401** (0,0363) | 0,398** (0,0296) | |
| Mother | | 0,322** (0,0313) | 0,334** (0,0460) | 0,308** (0,0421) | 0,321** (0,0313) | |
| Male | | 5,23** (0,144) | | | 5,21** (0,144) | 5,19** (0,148) |
| Kids | | | | | -0,0438 (0,0272) | |
| Antall obs. | 898 | 898 | 465 | 433 | 898 | 898 |
| Adj. R ² | 0,261 | 0,639 | 0,236 | 0,294 | 0,639 | 0,767 |
| FE | Nei | Nei | Nei | Nei | Nei | Ja |

I modell 1 er *Height_adj* brukt som avhengig variabel. I modell 2-6 er *Height* avhengig variabel. Modell 1-5 er estimert med minste kvadraters metode. Modell 6 inneholder en paneldataregresjon med en familiespesifikk fast effekt (FE). Standardfeilen til koeffisientene er oppgitt i parentes.

* signifikant på 10 %-nivå. ** signifikant på 5 %-nivå. *** signifikant på 1 %-nivå

- Gi en kortfattet gjennomgang av hva vi lærer fra modell 5.
- Motivasjonen for å ta med antall barn som en forklaringsvariabel er at barnerike familier ofte er fattige, og at fattigdom kan påvirke ernæring og vekst. Vis hvordan vi gjennomfører en t-test med 5 % signifikansnivå av nullhypotesen $\beta^{Kids}=0$ i modell 5. Du må selv ta stilling til om du vil bruke ensidig eller tosidig test.
- Modell 1, 3 og 4 har vesentlig lavere forklaringskraft (justert R²) enn modell 2, 5 og 6. Hva skyldes dette?
- Faren i den første familien i Galtons notatbok er 78,5 tommer høy og moren er 67 tommer. Prediker sønnens høyde ved å bruke modell 1 og modell 2.
- Modell 1 innebærer at $Height_adj = \alpha + \beta (0,5 * Father + 0,5 * 1,08 * Mother) + \epsilon$. Anta at dette er en riktig spesifisering og at sann β er lik estimatet 0,729. Da er forventet verdi på koeffisienten til variabelen *Father* i modell 2 lik $0,5 * 0,729 = 0,365$. Test om den estimerte koeffisienten til *Father* i modell 2 er signifikant forskjellig fra 0,365.
- I modell 3 og 4 har vi åpnet opp for at betydningen av mors og fars høyde ikke er lik for sønner og døtre. Men vi kan ikke uten videre se om forskjellen mellom de estimerte koeffisientene er signifikant. Foreslå en regresjonsmodell der vi enkelt kan teste om betydningen av mors og fars høyde er forskjellig for sønner og døtre.
- Hvorfor er det kun kjønnsdummiene (*Male*) som lar seg estimere i modell 6?

Oppgave 2

Under finner du deskriptiv statistikk for høyde fra datasettet som er brukt i oppgave 1.

| | Gj.snitt | Median | Varians | Antall obs. |
|--------|----------|--------|---------|-------------|
| Fedre | 69,349 | 69,5 | 6,875 | 197 |
| Mødre | 63,984 | 64,0 | 5,549 | 197 |
| Sønner | 69,229 | 69,2 | 6,925 | 465 |
| Døtre | 64,110 | 64,0 | 5,618 | 433 |

- Ernæringssituasjonen blir typisk bedre over tid. Test om vi kan forkaste en hypotese om at døtrene er like høye som sine mødre.
- Anta at variansen til fedres og mødres høyde er den samme. Kall denne σ^2 . Anta at variansen til variabelen *Midparent* er 3,30. Finn et uttrykk for variansen til *Midparent* som en funksjon av σ^2 og beregn deretter σ^2 .
- Kvinner er i gjennomsnitt lavere enn menn. Test om variansen til mødrenes høyde likevel er lik variansen til fedrenes høyde slik vi antok ovenfor.
- Korrelasjonen mellom ektefellers høyde er 0,101. Test om denne er signifikant forskjellig fra null. Kommenter kort konklusjonen.
- Anta at vi kjører en regresjon $fars_h\ddot{o}yde = \alpha + \beta * mors_h\ddot{o}yde + \epsilon$. Hva blir den estimerte regresjonskoeffisienten, β , i denne regresjonen?

Oppgave 3

Alle moderne trådløse routere har såkalt dual-band hvilket innebærer en mulighet for å kjøre trådløst nettverk på 2,4- og 5 GHz-båndet samtidig. Ifølge nettstedet *klikk.no* har de to båndene litt ulike egenskaper. Vi kan grovt sett si at 2,4 GHz-båndet har dårligere hastighet, men større rekkevidde, mens 5 GHz har bedre hastighet, men dårligere rekkevidde.

Professor Jarle Møen er litt usikker på hva dette innebærer i praksis og hvilket bånd han bør bruke i ulike rom. Siden han likevel må lage en eksamensoppgave til sine studenter bestemmer han seg for å foreta et eksperiment.

I fire forskjellige rom gjennomfører han i rask rekkefølge gjentatte målinger av nedlastingshastigheten i det trådløse nettet ved å bruke nettsiden *speedtest.net*. I hvert rom gjennomfører han fem målinger med bruk av 2,4 GHz båndet og fem målinger med bruk av 5 GHz båndet. Rådatene og to analyser fra Excel er gjengitt under.

Rom 1 er det rommet hvor routeren står. Rom 2 er et naborom i samme etasje som routeren, men det er en brannmur mellom rommene. Rom 3 er et rom i etasjen under. I luftlinje er avstanden til routeren omtrent den samme i rom 2 og 3. Rom 4 ligger i en murkjeller to etasjer under routeren.

RÅDATA: Nedlastingshastighet målt i Mbps.

| Bånd | Rom1 | Rom2 | Rom3 | Rom4 |
|---------|-------|-------|-------|-------|
| 2,4 GHz | 42,34 | 32,02 | 48,15 | 36,39 |
| 2,4 GHz | 40,88 | 37,93 | 44,22 | 29,22 |
| 2,4 GHz | 37,99 | 41,14 | 44,76 | 31,98 |
| 2,4 GHz | 37,40 | 35,12 | 31,87 | 20,88 |
| 2,4 GHz | 39,01 | 42,63 | 39,22 | 19,64 |
| 5 GHz | 43,24 | 40,04 | 41,24 | 4,05 |
| 5 GHz | 46,73 | 40,70 | 43,01 | 9,12 |
| 5 GHz | 43,64 | 45,57 | 40,75 | 10,19 |
| 5 GHz | 41,67 | 40,39 | 40,87 | 11,82 |
| 5 GHz | 41,68 | 43,41 | 33,40 | 17,45 |

ANALYSE 1: Anova: Two-Factor With Replication

| SUMMARY | Rom1 | Rom2 | Rom3 | Rom4 | Total |
|----------------|--------|--------|--------|--------|--------|
| <i>2,4 GHz</i> | | | | | |
| Count | 5 | 5 | 5 | 5 | 20 |
| Sum | 197,62 | 188,84 | 208,22 | 138,11 | 732,79 |
| Average | 39,52 | 37,77 | 41,64 | 27,62 | 36,64 |
| Variance | 4,22 | 18,77 | 40,02 | 51,90 | 54,71 |
| <i>5 GHz</i> | | | | | |
| Count | 5 | 5 | 5 | 5 | 20 |
| Sum | 216,96 | 210,11 | 199,27 | 52,63 | 678,97 |
| Average | 43,39 | 42,02 | 39,85 | 10,53 | 33,95 |
| Variance | 4,28 | 5,71 | 13,84 | 23,41 | 204,12 |
| <i>Total</i> | | | | | |
| Count | 10 | 10 | 10 | 10 | |
| Sum | 414,58 | 398,95 | 407,49 | 190,74 | |
| Average | 41,46 | 39,90 | 40,75 | 19,07 | |
| Variance | 7,94 | 15,91 | 24,83 | 114,66 | |

ANOVA

| Source of Variation | SS | df | MS | F | P-value | F crit |
|---------------------|---------|-------|---------|-------|---------|--------|
| Sample | 72,41 | 1,00 | 72,41 | 3,57 | 0,07 | 4,15 |
| Columns | 3520,10 | 3,00 | 1173,37 | 57,89 | 0,00 | 2,90 |
| Interaction | 748,92 | 3,00 | 249,64 | 12,32 | 0,00 | 2,90 |
| Within | 648,65 | 32,00 | 20,27 | | | |
| Total | 4990,08 | 39,00 | | | | |

ANALYSE 2: Anova: Two-Factor With Replication

| SUMMARY | Rom1 | Rom2 | Rom3 | Total |
|----------------|--------|--------|--------|--------|
| <i>2,4 GHz</i> | | | | |
| Count | 5 | 5 | 5 | 15 |
| Sum | 197,62 | 188,84 | 208,22 | 594,68 |
| Average | 39,52 | 37,77 | 41,64 | 39,65 |
| Variance | 4,22 | 18,77 | 40,02 | 20,70 |

| | | | | |
|--------------|--------|--------|--------|--------|
| <i>5 GHz</i> | | | | |
| Count | 5 | 5 | 5 | 15 |
| Sum | 216,96 | 210,11 | 199,27 | 626,34 |
| Average | 43,39 | 42,02 | 39,85 | 41,76 |
| Variance | 4,28 | 5,71 | 13,84 | 9,08 |

| | | | | |
|--------------|--------|--------|--------|--|
| <i>Total</i> | | | | |
| Count | 10 | 10 | 10 | |
| Sum | 414,58 | 398,95 | 407,49 | |
| Average | 41,46 | 39,90 | 40,75 | |
| Variance | 7,94 | 15,91 | 24,83 | |

ANOVA

| Source of Variation | SS | df | MS | F | P-value | F crit |
|---------------------|--------|-------|-------|------|---------|--------|
| Sample | 33,41 | 1,00 | 33,41 | 2,31 | 0,14 | 4,26 |
| Columns | 12,25 | 2,00 | 6,12 | 0,42 | 0,66 | 3,40 |
| Interaction | 57,24 | 2,00 | 28,62 | 1,98 | 0,16 | 3,40 |
| Within | 347,41 | 24,00 | 14,48 | | | |
| Total | 450,32 | 29,00 | | | | |

- a) Hva lærer vi av analyse 1.
- b) Hva lærer vi i tillegg av analyse 2.
- c) Drøft kort om forutsetningene for ANOVA er oppfylt.