



# SKOLEEKSAMEN MET4

Høst, 2017

**Dato:** 28.11.2017

**Tidsrom:** 09:00 – 12:00

**Antall timer:** 3

**Foreleser/emneansvarlig kan kontaktes av eksamensvakt på telefon: 55959902**

TILLATTE HJELPEMIDLER:

Alle trykte/egenskrevne, kalkulator

Ordbok: èn tospråklig ordbok tillatt

**Antall sider, inkludert forside:5, inkludert denne**

Instruksjoner: Skriv kortfattet. Hvert problem presenterer en oppgave og flere spørsmål. Merk: Engelsk oppgavetekst følger etter den norske oppgaveteksten. Du skal kun svare på hver oppgave en gang.

1. En lærer sjekker individuelle hjemmeeksamener i kurs AAA-123 og finner slående likhet mellom studentens svar. Hun mistenker fusk ved plagiering, der elevene kopierer svar fra hverandre. For å identifisere om fusk bidrar til å oppnå høyere score, bestemmer hun seg for å lage et datasett av studentrespons og gjennomføre noen tester. Læreren setter signifikansnivået til 1 prosent og ønsker å bruke tosidige tester. Datasettet inneholder følgende variabler:

**Student\_id** – unikt nummer som identifiserer hver student

**RQ1** - Svar på spørsmål 1. Dette er svaret på beregningen forespurt i spørsmål 1 på eksamen. Det riktige svaret er 20.

**Rest** – Andel korrekte svar på de resterende spørsmålene.

**Total** – Totalt antall poeng.

**Group** – En indikatorvariabel for gruppen studenter som er mistenkt for juks. Den mistenkellige gruppen er betegnet som gruppe A. De øvrige studentene er i gruppe B.

**Gender** – En indikatorvariabel for kvinner.

Tabell 1 viser deskriptiv statistikk for variablene. Tabell 2 viser resultatene fra en variansanalyse (ANOVA). Tabell 3 viser et tilfeldig utvalg av studenter.

Er det en signifikant forskjell i svarene mellom de to studentgruppene? Svar på følgende spørsmål for å avgjøre dette:

- a. Variansen i svarene kan være en indikasjon på fusk. Dette kan ses med **RQ1**, hvor standardavviket for gruppe A er lavere enn standardavviket for gruppe B. Bruk formelen for varians og forklar hvorfor plagiat kan føre til et slikt utfall.
- b. Utfør en test for å sjekke om variansen av **RQ1** er den samme for studenter i gruppe A og studenter i gruppe B. Viser resultatet av testen at studentene i gruppe A kanskje har plagiert?
- c. Utfør en test for å sjekke om det er forskjell mellom de gjennomsnittlige svarene på **RQ1** for gruppe A og B. Viser resultatet av testen at studentene i gruppe A svarte annerledes enn studenter i gruppe B?
- d. Utfør en test for å sjekke om det er forskjell mellom de gjennomsnittlige svarene på **Rest** for gruppe A og B. Viser resultatet av testen at studentene i gruppe A har andre resultater enn studentene i gruppe B?
- e. Endres resultatene i noen av testene som ble utført i de foregående 3 punktene (b, c og d) hvis vi benytter en ensidig test? Hvorfor/Hvorfor ikke?
- f. Se resultatene i tabell 2 fra en variansanalyse (ANOVA). Responsvariabelen er **Total**. Hvor mange faktorer er det? Hva er faktornivåene for hver faktor? Kan du skrive ned null- og alternativhypotesene? Hvilken forutsetning for ANOVA-testen synes å være brutt?
- g. Tolk resultatene av variansanalysen. (bruk maks 10 setninger)
- h. Tabell 3 viser variabelen **Totalt** for et tilfeldig utvalg på 7 studenter hver fra gruppe A og B. Bruk en tosidig Wilcoxon Rank Sum-test med 5 prosent signifikansnivå. Skriv opp nullhypotesen og alternativhypotesen. Gjennomfør testen. Tolk resultatene dine.

- i. Basert på svarene dine og tilleggsberegninger (om nødvendig), hva kan du si om fordelingen av variabelen **Total** for de to studentgruppene? (Hint: Bruk ordene gjennomsnitt, varians og lokasjon (location på engelsk)) (maks 6 setninger)
2. Du skal analysere effekten på kriminalitet av «Medical Marijuana Law» (MML) i New Mexico, som ble satt i kraft i 2007. MML tillater reseptbelagt cannabis til pasienter som lider av en rekke helseproblemer, med symptomer som alvorlig eller svekkende smerte. For å finne effekten velger du nabostaten Texas som en kontrollgruppe. Du samler paneldata for fylkene (counties på engelsk) i de to statene i 10 år.

Datasettet inneholder følgende variabler:

**MML** – Indikatorvariabel, som tar en verdi 1 når MML er aktiv. Variabelen begynner å ta en verdi på 1 i 2007 for New Mexico

**Violent Crime Rate** – målt som antall voldelige forbrytelser per 100 000 innbyggere

**Fraction of Hispanics** – Andelen av befolkningen med latinamerikansk opprinnelse.

**Median Income** – Et mål på medianen av den rapporterte skattepliktige inntekt for familier som bor i fylket.

**Year** – observasjonsåret

**County** – unik kode tildelt hvert fylke

Tabell 4 viser deskriptiv statistikk for datasettet. Tabell 5 viser resultater for estimerte modeller. Angi tydelig hvilken kolonne i tabell 5 du refererer til når du svarer. I oppgaven nedenfor er "modell" et synonym med "kolonne".

- a. Hva er forskjellen mellom paneldata og et tverrsnitt? Basert på antall observasjoner, hvor mange fylker er i dataene? Hvor mange fylker er det i New Mexico?
- b. Koeffisienten på MML i kolonne (1) i tabell 5 er 117,5 med en standardfavvik på 16,76. Finn en omtrentlig p-verdi for denne koeffisienten. Tolk koeffisienten.
- c. Hva skjer med koeffisienten til MML når vi legger til faste effekter på tid- og fylkesnivå? Hva er effekten på forklaringsgraden? Hvorfor øker de faste effektene forklaringsgraden? Baser svaret ditt på kolonnene (1), (3) og (4) i tabell 5.
- d. Hva er din foretrukne modell? Hvorfor? Oppsummer MMLs virkning på kriminalitet ved å skrive et notat til en politiker som er kjent med reformen (maks 7 setninger).
- e. Basert på den økonomiske modellen av kriminalitet, tror vi at en økning i lovlige inntektsmuligheter (her er dette representert av median lønn) burde føre til reduksjon av kriminalitet. Kommenter dette ved å tolke fortegnene og betydningen av estimatene for median lønn i kolonnene (2) og (5) i tabell 5
- f. President Trump har lovet å bygge en mur mot Mexico og stoppe innstrømmingen av innvandrere. Kommenter dette ved å bruke kolonne (5) i Tabell 5 til å predikere forskjellen i **Violent Crime Rate** i to fylker som er identiske, med unntak av at de har henholdsvis den minimale og maksimale verdien av variabelen **Fraction of Hispanics** (hint: bruk informasjonen i tabell 4).

**Tabell 1. Deskriptiv statistikk for oppgave 1.**

	Hele utvalget		Gruppe A		Gruppe B	
	Gjennom- snitt	Standard- avvik	Gjennom- snitt	Standard- avvik	Gjennom- snitt	Standard- avvik
RQ1	19.45	7.16	20.10	1.16	18.80	10.06
Rest	0.78	0.22	0.98	0.01	0.58	0.11
Total	64.27	17.96	80.92	1.66	47.63	9.14
Gender	0.60	0.49	0.64	0.48	0.56	0.50
Antall observasjoner	100		50		50	

**Tabell 2. ANOVA av Total**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	27709	27709	639.226	0.000
Gender	1	23	23	0.524	0.471
Group:Gender	1	42	42	0.961	0.329
Residuals	96	4161	43		

**Tabell 3. Total score for 14 av studentene**

Gruppe A	Gruppe B
133	81
127	69
131	107
111	92
116	84
134	126
131	67

**Tabell 4. Deskriptiv statistikk for kriminalitetsdata til oppgave 2.**

	Gjennom- snitt	Standard- avvik	Min	Max
Violent Crime Rate	274.41	210.76	0.00	1909.11
MML	0.06	0.23	0.00	1.00
Fraction of Hispanics	0.32	0.23	0.02	0.97
Median Income	10.51	0.24	9.75	11.61
Year	2006.5	2.87	2002	2011
Number of Observations				<b>2870</b>

Tabellnoter: Min angir minimumsverdien i utvalget. Max angir maksimumsverdien av variabelen.

**Tabell 5. Resultater for effekten av MML på kriminalitet**

	(1)	(2)	(3)	(4)	(5)
	Violent Crime	Violent Crime	Violent Crime	Violent Crime	Violent Crime
MML	117.5 (16.76)	99.17 (16.77)	-51.00 (14.04)	-46.72 (14.87)	-51.05 (15.14)
Median Income		-5.365 (17.20)			82.85 (63.61)
Fraction of Hispanics		132.0 (18.34)			-670.2 (232.2)
Constant	267.7 (4.019)	282.3 (183.1)			
Observations	2,870	2,870	2,870	2,870	2,870
R-squared	0.017	0.037	0.670	0.674	0.675
Adj. R-squared	0.0165	0.0364	0.634	0.637	0.638
F-statistic	49.18	37.15	13.19	9.873	6.504
Fixed Effects	no	no	County	County and Year	County and Year

Tabellnoter: Standardavvik i parantes. "Fixed-effects" angir hvilken type faste effekter som er inkludert i modellen. Den øverste kolonnen viser den avhengige variabelen. Adj. R-Squared står for Adjusted R-Square.