BOKMÅL

NHH

SKOLEEKSAMEN MET4

Vår 2019

Dato: 22. mai 2019

Tidsrom: 09:00 - 12:00

Antall timer: 3

Foreleser/emneansvarlig kan kontaktes av eksamensvakt på telefon: 97599593

TILLATTE HJELPEMIDLER:
Alle trykte/egenskrevne hjelpemidler, kalkulator.

Ordbok: én tospråklig ordbok tillatt

Antall sider, inkludert forside: 9

Oppgave 1

Et kontroversielt spørsmål de senere år har vært om bruk av voldelige dataspill fører til mer voldelig oppførsel. Flere studier antyder at det er en slik sammenheng, men en særlig profilert artikkel som påviste en sammenheng mellom bruk av skytespill og ferdigheter med ekte våpen ble nylig trukket tilbake på grunn av talltriksing.[1] En annen innvendig mot forskning av dette fenomenet er at personer som spiller skytespill kan vise høyere aggresjonsnivå like gjerne på grunn av spillets *vanskelighetsnivå* eller *stressnivå*, som på grunn av skyting og vold som sådan.

I en helt ny studie[2] kan vi lese om et eksperiment som ble utført for å undersøke dette nærmere, der 275 mannlige studenter ble tilfeldig trukket ut til å spille et dataspill i en av to varianter:

- I den første varianten var oppdraget å *skyte* og *drepe* et romvesen ved å sikte med et våpen og trekke av (voldelig).
- I den andre varianten var oppdraget å *redde* et romvesen fra fare (ikke-voldelig).

I den ikke-voldelige varianten var ordene *sikte* og *skyte* erstattet med *finne* og *redde*, men utover det var spillene utformet på eksakt samme måte. Etter at studentene hadde spilt en stund, ble deres aggresjonsnivå målt på en skala fra 1 til 9 ved hjelp av en standard psykologisk test, der 9 er mest aggressiv. Under følger en tabell med deskriptiv statistikk for aggresjonsnivået i de to gruppene:

Spillversjon	Min.	Median	Gj.snitt	Max.	St.avvik	N
Voldelig	1	6	6.22	9	2.35	136
Ikke-voldelig	1	6	5.91	9	2.52	139

(a) Test om forventet aggresjonsnivå er det samme i de to gruppene. Begrunn eventuelle valg du gjør underveis.

Svar: Vi skal teste

$$H_0 : \mu_1 = \mu_2 \text{ mot } H_1 : \mu_1 \neq \mu_2.$$

Standardavvikene i de to populasjonene er nesten like, så vi antar at de to populasjonene har lik varians (kan også teste det formelt). Regner ut en felles varians:

$$S_P^2 = \frac{(136-1)\cdot 2.35^2 + (139-1)\cdot 2.52^2}{136+139-2} = 5.94.$$

[1] Whitaker og Bushman: *«Boom, Headshot!»: Effect of Video Game Play and Controller Type on Firing Aim and Accuracy*, opprinnelig publisert i 2012 i tidsskriftet *Communications Research*, senere trukket tilbake.

[2] Hilgard, Engelhardt, Rouder, Segert og Bartolow: *Null Effects of Game Violence, Game Difficulty, and 2D:4D Digit Ratio on Aggressive Behavior* (2019), *Psycological Science*

Testobservator for t-test med lik varians:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_P^2\left(\frac{1}{136} + \frac{1}{139}\right)}} = 1.05.$$

Kritisk verdi for tosidig t-test med $136 + 139 - 2 = 273$ frihetsgrader på 5% nivå er 1.96. Siden $1.05 < 1.96$ kan vi **ikke** forkaste H_0 om likt aggresjonsnivå.

Til sensor: Trekk 3p for manglende forklaringer og begunnelser. Formell test for varians ikke nødvendig, men det bør være en setning om hvorfor man velger hva. Trekk 5p eller mer for grove regnefeil, 2-3p for mindre regnefeil. Det er også lov å konkludere ut fra testobservatoren at det ikke blir noe forkastning uansett hvilke frihetsgrader man velger.

Studentene ble også bedt om å selv vurdere graden av voldelighet i spillet, og svarte på en skala fra 1 til 9. For enkelhets skyld sorterer vi svarene i to grupper: Studentene som vurderer spillet til å være *lite voldelig* (1–4) og studentene som vuderer spillet til å være *voldelig* (5–9). Vi lager følgende krysstabell mot spillets *faktiske* voldelighet (vi mangler svar fra 26 studenter):

		Faktisk voldelig		
		Ja	Nei	Sum
Opplevd voldelig	Ja	93	9	102
	Nei	33	114	147
	Sum	126	123	249

(b) Test om faktisk og opplevd voldelighet er uavhengige kjennetegn. På hvilken måte er dette et relevant spørsmål i denne konteksten?

Svar: Vi skal teste

$$H_0 : \text{Faktisk og opplved voldelighet er uavhengige kjennetegn ,mot}$$
$$H_1 : \text{Faktisk og opplved voldelighet er ikke uavhengige kjennetegn.}$$

I oppgaven har vi fått oppgitt de observerte frekvensene (f_{ij}). Vi regner ut de forventede frekvensene under H_0 ved hjelp av formelen $e_{ij} = f_{i\bullet} f_{j\bullet}/n$:

$$e_{11} = 102 \cdot 126/249 \approx 51.6, \qquad e_{12} = 123 \cdot 102/249 \approx 50.4$$
$$e_{21} = 126 \cdot 147/249 \approx 74.4 \qquad e_{22} = 123 \cdot 147/249 \approx 72.6$$

Testobservatoren for en χ^2-test for uavhengighet er gitt ved

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \frac{(93-51.6)^2}{51.6}+\frac{(9-50.4)^2}{50.4}+\frac{(33-74.4)^2}{74.4}+\frac{(114-72.6)^2}{72.6}$$
$$= 113.9 \text{ (litt avhengig av hvor man gjør avrunding).}$$

Kritisk verdi for en ensidig χ^2-test med $(2-1)\cdot(2-1) = 1$ frihetsgrad på 5% nivå er 3.84, så vi har en veldig klar forkastning.

Dette er et viktig resultat for eksperimentets gyldighet. De observerte frekvensene viser at deltakerne i stor grad selv opplever spillets voldelighet i samsvar med det som faktisk er tilfelle. Resultatet vårt viser at denne sammenhengen er statistisk signifikant.

Til sensor: De to delene av spørsmålet vekter med 7 poeng for testen, og 3 poeng for å forstå konteksten. Minst 5p trekk for grove regnefeil, 2-3p for mindre regnefeil.

Studentene har i eksperimentet spilt med enten *høyt* eller *lavt* vanskelighetsnivå (`difficulty_treatment`), og vi vil undersøke om dette kan være med på å forklare variasjon i aggresjonsnivået i tillegg til spillets voldsnivå (`violent_treatment`). Vi gjennomfører derfor en *toveis variansanalyse med samspill* i en statistisk programvarepakke og får følgende utskrift:

```
                                      Df Sum Sq Mean Sq F value Pr(>F)
violent_treatment                      1    6.8   6.783   1.147  0.285
difficulty_treatment                   1    6.0   5.955   1.007  0.316
violent_treatment:difficulty_treatment 1   15.3  15.322   2.592  0.109
Residuals                            271 1601.9   5.911
```

(c) Hvilke tester er gjennomført her? Hva er resultatet? Hva er fortolkningen?

Svar: Modellen som ligger til grunn for toveis variansanalyse med samspill er

$$\mathrm{E}(X_{ijr}) = \mu + \alpha_i + \beta_j + \phi_{ij},$$

der X_{ijr} er observert aggressivitet til person nummer r som har spilt med voldelighetsnivå i og vanskelighetsgrad j, μ er en felles forventningsverdi, α_i er en tilleggseffekt for voldsnivå i, β_j er en tilleggseffekt for vanskelighetsgrad j, og ϕ_{ij} en en samspilleseefekt som er spesifikk for kombinasjonen av voldelighetsnivå i og vanskelighetsgrad j. Vi tester tre nullhypoteser i dette tilfellet:

$$H_0: \alpha_1 = \alpha_2 = 0$$
$$H_0: \beta_1 = \beta_2 = 0$$
$$H_0: \phi_{11} = \phi_{12} = \phi_{21} = \phi_{22} = 0,$$

og i alle tilfeller alternativhypotesen at minst en parameter er forskjellig fra null. Fra utskriften ser vi at vi ikke forkaster noen av disse nullhypotesene. Med andre ord: det ser ikke ut til å være en statistisk signifikant sammenheng mellom voldelighetsnivå og aggressivitet, selv når vi kontrollerer for vanskelighetsgrad og samspillseffekter.

Til sensor: Spørsmålets tre deler vektes likt, men del 2 og 3 kan gjerne besvares med en og samme setning som i løsningsforslaget over.

Oppgave 2

Høsten 2016 innførte Regjeringen den såkalte *fraværsgrensen* i den videregående skole som medfører at elever med mer enn 10% udokumentert fravær i et fag ikke vil få karakter i faget. Det er samlet inn fraværsdata fra alle landets studiespesialiserende utdanningsprogram for skoleårene 2015–16 og 2016–17. [3] Hver skole har oppgitt median timefravær og median dagsfravær[4]. I tillegg har vi en rekke andre variabler ved den aktuelle videregående skole fra skoleåret 2016–17.[5]

Under følger et sammendrag av variablene:

	Gj.snitt	St.avvik	Antall	Beskrivelse
`time_1516`	13.0	7.2	362	Median timefravær i skoleåret 2015–16
`dag_1516`	5.8	2.5	362	Median dagsfravær i skoleåret 2015–16
`time_1617`	9.4	6.7	374	Median timefravær i skoleåret 2016–17
`dag_1617`	3.5	1.8	374	Median dagsfravær i skoleåret 2016–17
`antall_elever`	243.5	203.7	345	Antall elever ved skolen som går studiespesialiserende retning
`privat`	0.2	0.4	345	Dummyvariabel som tar verdien 1 dersom skolen er privat
`arbeidsro`	4.0	0.4	304	Elevenes gj.snittsvurdering på om de har tilstrekkelig arbeidsro, skala: 1–9
`trivsel`	4.3	0.2	305	Elevenes gj.snittsvurdering av egen trivsel på skolen, skala: 1–9
`karakterbidrag`	-0.1	1.6	344	Et mål på hvor mye den enkelte skole bidrar til elevenes standpunktkarakterer når man han kontollert for elevenes utgangspunkt

(a) **Forklar hvorfor det er fornuftig for skolene å rapportere *median*fraværet i stedet for *gjennomsnitts*fraværet.**

Svar: Det er naturlig å tenke seg at noen få enkeltelever av ulike grunner vil ha svært mye fravær gjennom et skoleår, noe som kan trekke gjennomsnittsfraværet mye opp selv om slike elever ikke er representative for det store flertallet. Medianen er et mer robust sentermål, som ikke blir påvirket av store uteliggere.

[3] `https://www.udir.no/tall-og-forskning/statistikk/statistikk-videregaende-skole/fravar-vgs/`
[4] Fravær i enkelttimer og hele dager blir registrert hver for seg
[5] `skoleporten.udir.no`

Vi ser at medianeleven i gjennomsnitt har redusert sitt timefravær fra 13 til 9.4 timer etter at fraværsgrensen ble innført. Vi ønsker å finne ut om denne forskjellen er statistisk signifikant, og gjennomfører en test for det i programpakken `R`. Utskriften er som følger:

```
    Paired t-test

data:  vgs$time_1516 and vgs$time_1617
t = 19.291, df = 349, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.199610 3.926104
sample estimates:
mean of the differences
               3.562857
```

(b) Hvilken test er gjennomført her? Skriv opp denne testen og bruk informasjon fra utskriften til å gjennomføre den og konkludere. Hva vil det si at testen er «paired»?

Svar: Dette er en parret t-test der vi først regner ut differansene før og etter endringen for hver skole og regner ut den gjennomsnittlige differansen $\overline{X}_D$ og det standardavvik $s_D/\sqrt{n}$. Testobservatoren T er forholdet mellom disse to tallene og fra utskriften ser vi at verdien er 19.3. Kritisk verdi er 1.96 (tosidig, evt 1.645 for ensidig test), så vi forkaster klart nullhypotesen om ingen forskjell i fravær før og etter endringen. Det kan vi også lese ut fra p-verdien, som er nesten null.

I en parret test ser vi *gjennomsnitt av differanser* i stedet for *differanser av gjennomsnitt* som i dette tilfellet betyr at vi eliminerer variasjon mellom skolene fra testen.

Til sensor: For full pott skal studenten vise tydelig at vedkommende har forstått hva en parret t-test er. Denne forklaringen har lik vekt som første del av spørsmålet.

Da fraværstallene som vi bruker i denne oppgaven ble publisert høsten 2018, møttes statssekretær Atle Simonsen i Kunnskapsdepartementet (som argumenterer *for* fraværsgrensen) og leder i Elevorganisasjonen Agathe Waage (som argumenterer *mot* fraværsgrensen) til debatt i *Dagsnytt Atten* (NRK, 28.10.2018). Ved å forkorte og parafrasere debatten sitter vi igjen med følgene inntrykk:

Atle Simonsen: Fraværsgrensen har virket, fordi medianeleven har redusert sitt fravær med 40%.

Agathe Waage: Medianeleven har redusert sitt dagsfravær fra 5 til 3 dager. Utgangspunktet var allerede lavt, så endringen har ingen praktisk betydning.

Agathe Waage: Over halvparten av elevene som ikke får karakter, får ikke karakter nettopp på grunn av fraværsgrensen, som igjen fører til *mer* frafall i videregående skole.

Atle Simonsen: Det *totale* antall elever som ikke får karakter har gått ned, og frafallet er også gått ned, ergo fører fraværsgrensen til *mindre* frafall.

(c) **Bruk minst et begrep/konsept som du har lært i MET4 til å skrive en kort kommentar (max. en halv side med normal håndskrift) til denne debatten. Hvordan kan det samme datamaterialet tas til inntekt for to stikk motsatte syn?**

Svar: Dette er en typisk situasjon i samfunnsdebatten der vi kan mistenke at politiske aktører alltid klarer å presentere tilsynelatende objektive tall til fordel for sitt ideologiske syn. Ofte viser det seg at det er først når vi aksepterer at det finnes nyanser og legitime argumenter både for og mot vårt syn at vi faktisk kan *lære* noe av å gjøre empiriske analyser.

Disse to utvekslingene peker på to momenter vi har diskutert i MET4. Den første handler om at den ene aktøren peker på en forskjell som nok er *statistisk* signifikant, mens den andre aktøren påpeker at forskjellen neppe kan sies å være viktig *i praksis*.

Den andre utvekslingen minner oss om hvor vanskelig det kan være å skille mellom kausalitet og korrelasjon. Begge aktørene blander selektivt inn nye variable (årsak til karaktermangel, karaktermangel totalt, totalt frafall) som hver for seg kan gi inntrykk av to motsatte kausale effekter. Her kan det hende at den fulle rapporten kan hjelpe oss til å faktisk få svar på om fraværsgrensen fungerer etter hensikten.

Til sensor: Vurder etter beste skjønn, men legg vekt på at studenten klarer å knytte debatten til et punkt som er (eller kunne vært) pensum i kurset. Presise svar belønnes, og det kan fint holde med 2-3 setninger. Vage svar som ikke tar stilling til noe skal straffes, svar som flyter over en halv side skal også straffes.

Et viktig element i argumentasjonen *for* fraværsgrensen er at elever lærer mer når de er til stede i undervisningen. Vi skal undersøke dette nærmere på skolenivå ved å analysere i hvilken grad fraværsnivået ved hver enkelt skole bidrar til elevenes læring gjennom variabelen `karakterbidrag`. Enkelt sagt betyr et karakterbidrag på 1 at skolen bidrar med en hel karakter til elevenes eksamenskarakterer utover det man ellers kunne forvente ut fra elevenes utgangspunkt. På samme måte betyr negative verdier at elevene går ut med dårligere karakterer enn det man ellers kunne forvente.

Tabell 1 presenterer to regresjonsutskifter med `karakterbidrag` som responsvariabel.

(d) **Forklar kort hva vi lærer om sammenhengen mellom karakterbidrag og variablene som måler fravær.**

Svar: I modell 1 har vi brukt de fire forklaringsvariablene som *ikke* måler fravær og ser at flere av koeffisientene er signifikant forskjellige fra null. I modell 2 legger vi til de to fraværsvariablene, men vi ser at ingen av dem har koeffisienter som er signifikant forskjellige fra null, og justert R^2 går *ned* fra modell 1 til modell 2. Vi kan i det aktuelle skoleåret ikke her se noen spesiell sammenheng mellom fravær og skolens karakterbidrag ved hjelp av lineær regresjon.

Her bør vi dog kikke nærmere på datamaterialet før vi trekker bastante konklusjoner. En naturlig tanke er at det er lite selvstendig variasjon i i fravær som ikke er drevet av trivsel, og at fravær i liten grad er en skolevariabel, men mer knyttet til elevenes utgangspunkt, og dermed ikke er en del av skolenes bidrag til karakterene.

Til sensor: For full pott *må* svaret være kort, som spesifisert i spørsmålet. Legg kun vekt på det som faktisk svarer på spørsmålet, og ikke på strøtanker og spekulasjoner. Svar som er ekvivalent med første del av løsningsforslaget over kan gi full pott, men diverse advarsler mot å fortolke resultatet for mye á la andre del, skal telle positivt, selv om det kanskje ikke er vanntett. Forsøk på kausale fortolkninger skal straffes hardt.

(e) Regn ut 95% konfidensintervaller for koeffisienten til `privat` i de to modellene. Hvordan kan du se ut fra intervallene at koeffisienten måles som signifikant forskjellig fra null i den første modellen, men ikke i den andre?

Svar: Konfidensintervall for regresjonskoeffisienter er gitt som $\widehat{\beta} \pm t_{n-2,\alpha/2} s_\beta$, der s_β er koeffisisntens standardavvik som vi finner i parantes i regresjonstabellen. I begge tilfeller er $n > 200$, så vi setter $t_{n-2,\alpha/2} = 1.96$ for 95% konfidensintervaller. I modell 1 får vi:

$$\widehat{\beta} \pm t_{\alpha/2} s_\beta = 0.533 \pm 1.96 \cdot 0.227 = [0.088, 0.978].$$

I modell 2 får vi

$$\widehat{\beta} \pm t_{\alpha/2} s_\beta = 0.196 \pm 1.96 \cdot 0.239 = [-0.272, 0.664].$$

Det andre intervallet inneholder 0, noe som betyr at koeffisienten ikke er signifikant forskjellig fra null i denne modellen.

Til sensor: Hver intervall gir 3p hver, 4p for andre del av spørsmålet.

En av dine medstudenter advarer mot å gå direkte fra modell 1 til modell 2 fordi vi risikerer å få problemer med multikolinearitet.

(f) Forklar hvorfor det er en gyldig innvending i denne spesifikke situasjonen og hvordan multikolinearitet eventuelt kan komme til uttrykk her. Hva kan vi gjøre for å undersøke om vi faktisk har problemer med multikolinearitet i modell 2?

Svar: Ved å legge til to variable som potensielt er sterkt korrelerte fordi de måler det samme fenomenet («fravær» i dette tilfellet) kan det oppstå multikolinearitetseffekter. I praksis betyr det at standardavvikene til regresjonskoeffisientene blåses opp, og dermed kanskje også at ingen av koeffisientene estimeres til å være signifikant forskjellige fra null, selv om det faktisk skulle være en viktig sammenheng mellom fravær og karakterbidrag.

I vårt tilfelle *er* ingen av fraværskoeffisientene er signifikant forskjellige fra null. Er det fordi det ikke er noen sammenheng, eller kan det skyldes nettopp multikolinearitet?

En rask sjekk er å bruke de to fraværsvariablene hver for seg. Vi kan også sjekke om koeffisientenes VIF (*variance inflation factor*) er store (større enn 5 f.eks.), siden dette nettopp er mål på hvor mye koeffisientenes varians øker som følge av multikolinearitet.

(Vi kan avsløre at de to fraværsvariablene ikke er signifikante hver for seg, at ingen VIF-verdier i modell 2 er større enn 2, og at korrelasjonen mellom de to fraværsvariablene er 0.40. Altså kan vi langt på vei puste lette ut og slå fast at multikolinearitet neppe er et problem i modell 2.)

Til sensor: Spørsmålet har tre deler som teller likt: 1. Hvorfor er det en gyldig innvending i denne spesifikke situasjonen? 2. Hvordan vil det komme til uttrykk i denne spesifikke situasjonen? 3. Hva kan vi gjøre for å sjekke det?

(g) Diagnostiser kort regresjonsmodell (2) ved hjelp av plottene i Figur 1.

Svar: Residualplottet avslører ingen spesielle mønster eller tegn på heteroskedastisitet. Histogrammet og QQ-plottet antyder at fordelingen til residualene ikke er helt symmetrisk, men det er ikke et dramatisk avvik fra normalitet. Vi har 277 observasjoner, som under normale omstendigheter er mer enn nok til å stole på at den statistiske inferensen er temmelig presis pga sentralgrenseteoremet. Plottet over Cooks avtand viser at det er et titalls skoler som har mye større innvirkning på regresjonsanalysen enn andre. Et naturlig neste steg vil være å undersøke om det er noen systematikk i dette. Hvilke skoler kan det være snakk om? Hva er det som gjør at de har så stor innvirkning på resultatet?

Til sensor: Igjen, belønn presise svar, selv om man kanskje ikke har fått med alle detaljer. Straff vage svar som ikke klarer å komme med konkrete innsikter.

Den 1. januar 2020 slås de tre fylkene Buskerud, Akershus og Østfold sammen til nye Viken Fylkeskommune. En ansatt i utdanningsetaten i den nye fylkeskommunen ønsker å finne ut om det er signifikante forskjeller i karakterbidrag mellom de tre fylkene i dag, og samler et datasett over alle videregående skoler i de tre fylkene med fire variabler: en variabel for skolenes karakterbidrag, og tre dumyvariabler, en for hvert fylke, som indikerer om skolen ligger i Buskerud, Akershus eller Østfold.

Hun prøver først å forklare skolenes karakterbidrag ved hjelp av de tre fylkesdummyene ved å tilpasse følgende regresjonsmodell:

$$\texttt{karakterbidrag} = \beta_0 + \beta_1\texttt{fylkeBuskerud} + \beta_2\texttt{fylkeAkershus} + \beta_3\texttt{fylke}\text{Ø}\texttt{stfold} + \epsilon.$$

(h) Hun får bare feilmeldinger i programvaren når hun forsøker å estimere koeffisientene i regresjonsmodellen over. Hvorfor gjør hun det, og hva må hun gjøre for å løse problemet?

Svar: Når vi bruker dummyvariable må vi alltid passe på at en kategori utelates fra regresjonsmodellen og brukes som referansekategori. Hvis ikke får vi perfekt multikolinearitet ($\texttt{fylkeBuskerud}_i + \texttt{fylkeAkerhus}_i + \texttt{fylke}\text{Ø}\texttt{stfold}_i = 1$ for alle i), som igjen fører til feilmeldinger.

Altså: dropp en fylkeskategori.

Hun finner ut av feilen, og kjører en ny regresjon som får følgende output i programvaren:

```
Call:
lm(formula = karakterbidrag ~ fylke, data = viken)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5800 -0.7076 -0.0576  1.1205  2.9424

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.05758    0.25331   0.227    0.821
fylkeBuskerud  0.12137    0.41906   0.290    0.773
fylkeØstfold  -0.67758    0.45313  -1.495    0.140

Residual standard error: 1.455 on 64 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:  0.04405,   Adjusted R-squared:  0.01418
F-statistic: 1.475 on 2 and 64 DF,  p-value: 0.2365
```

(i) Skriv opp og fortolk regresjonsmodellen som er estimert over. Kan vi ved hjelp av denne utskiften konkludere om det var signifikante forskjeller i `karakterbidrag` mellom de tre fylkene i skoleåret 2016–17?

Svar: Vi estimerer følgende modell:

$$\texttt{karakterbidrag} = 0.058 + 0.121 \cdot \texttt{fylkeBuskerud} - 0.678 \cdot \texttt{fylkeØstfold},$$

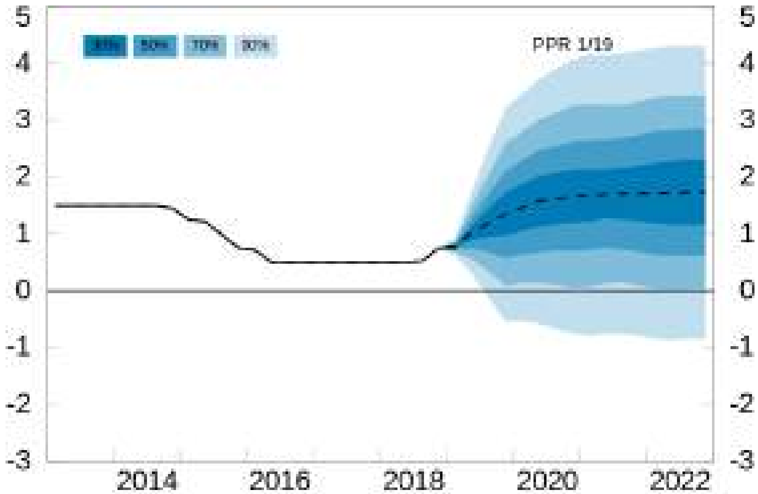
så referansekategorien er Akershus, som vi estimerer til å ha et forventet karakterbidrag lik 0.058. På samme måte estimerer vi skolene i Buskerud til å ha et forventet karakterbidrag på $0.058 + 0.121 = 0.179$ og skolene i Østfold til å ha et forventet karakterbidrag på $0.058 - 0.678 = -0.62$. Regresjonsmodellen har dog svært lav forklaringskraft, og regresjonens F-test gir ikke grunn til å forkaste nullhypotesen om at de tre fylkene har likt forventet karakterbidrag.

Dette er ekvivalent med å gjøre en enveis variansanalyse.

Til sensor: Besvarelser som bare ser at ingen av variablene er signifikante, derfor ingen signifikante forskjeller mellom fylkene, får maksimal halv uttelling. For full pott *må* det tas høyde for multippel testing og refereres til F-testen. Det er ikke et krav for full pott å se sammenhengen med variansanalyse.

Oppgave 3

I figuren under ser vi en graf over den norske styringsrenten siden 2013 og anslag over hvilken bane renten skal følge de neste tre årene med fire usikkerhetsintervaller.



(a) Bruk figuren til å anslå sannsynligheten for negativ styringsrente ved utgangen av 2022.

Svar: Det nest ytterste usikkerhetsintervallet inneholder 70% av sannsynligheten, og siden intervallene ser symmetriske ut kan vi konkludere med at 15 % sannsynlighet ligger under det nest ytterste intervallet. Ved utgangen av 2022 ser det ut til at den horisontale linjen som representerer 0% styringsrente sammenfaller ganske nøyaktig med den nederste grensen for det nest ytterste intervallet, så vi estimerer sannsynligheten for negativ styringsrente til å være nettopp 15% på dette tidspunktet.

Til sensor: Det vi tester her er evnen til å lese informasjon ut fra en graf. Gi poeng etter beste skjønn.

Oppgave 4

Anta at sammenhengen mellom to stokastiske variabler X og Y er gitt ved regresjonsligningen

$$Y = X + \epsilon, \qquad (*)$$

der $\mathrm{E}(X) = 0$, $\mathrm{Var}(X) = 1$, og der ϵ er en tredje stokastisk variabel som er uavhengig av X, og som tilfredsstiller $\mathrm{E}(\epsilon) = 0$ og $\mathrm{Var}(\epsilon) = 1$.

(a) Regn ut $\mathbf{E}(\boldsymbol{Y})$, $\mathbf{Var}(\boldsymbol{Y})$ og $\mathbf{Cov}(\boldsymbol{X}, \boldsymbol{Y})$.

Svar: Bruker regneregler for forventning, varians og kovarians og får:

$$\mathrm{E}(Y) = \mathrm{E}(X + \epsilon) = \mathrm{E}(X) + \mathrm{E}(\epsilon) = 0 + 0 = 0,$$

$$\mathrm{Var}(Y) = \mathrm{Var}(X + \epsilon) = \mathrm{Var}(X) + \mathrm{Var}(\epsilon) = 1 + 1 = 2,$$

der vi har brukt uavhengigheten mellom X og ϵ. Videre,

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{Cov}(X, X + \epsilon) \\
&= \mathrm{Cov}(X, X) + \mathrm{Cov}(X, \epsilon) \\
&= \mathrm{Var}(X) + 0 \\
&= 1 + 0 = 1.
\end{aligned}$$

Til sensor: Tre deler som teller likt.

Anta at du har tilgang til et sett observasjoner (X_i, Y_i), $i = 1, \ldots, n$ fra modell $(*)$. Ved å estimere en enkel lineær regresjonsmodell med Y som responsvariabel og X som forklaringsvariabel vil du naturlig nok få minste kvadraters estimater av β_0 (konstantleddet) og β_1 (stigningstallet) i nærheten av 0 og 1 henholdsvis, der presisjonen avhenger av utvalgsstørrelsen n.

Anta at du heller snur på regresjonen, og estimerer de to koeffisientene i følgende modell ved hjelp av minste kvadraters metode:

$$X = \beta_0^* + \beta_1^* Y + \epsilon^*.$$

(b) Vis at vi da vil få ut estimater i nærheten av $\widehat{\beta}_0^* = 0$ og $\widehat{\beta}_1^* = 1/2$.

Svar: Vi kan sette tallene fra forrige deloppgave rett inn i formel for miste kvadraters estimater, og få ut følgende estimat for stigningstallet:

$$\widehat{\beta}_1^* = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} \approx \frac{1}{2},$$

der vi bruker "tilnærmet lik"-tegnet på grunn av estimeringsusikkerhet i kovarians og varians. (De som bruker hintet uten å ha klart spm. (a) får samme svar: $\widehat{\beta}_1^* = 0.5/1 = 1/2$). Videre bruker vi at $\mathrm{E}(Y) = 0$, så $\overline{X} \approx \overline{Y} \approx 0$, og ved å bruke formel for estimat av konstantleddet har vi

$$\widehat{\beta}_0^* = \overline{X} - \widehat{\beta}_1^* \overline{Y} \approx 0.$$

Til sensor: Belønn gode forsøk, selv om man ikke kommer helt i mål.

Hint: Du trenger resultatene fra spørsmål (a). Hvis du ikke fikk til (a) kan du i denne oppgaven sette $\mathrm{E}(Y) = 0$, $\mathrm{Var}(Y) = 1$ og $\mathrm{Cov}(X, Y) = 0.5$ uten å miste uttelling på dette spørsmålet, men det er ikke nødvendigvis riktige svar i (a).

Tabell 1: Regresjonsutskrift (standardavvik i parantes)

	Dependent variable:	
	karakterbidrag	
	(1)	(2)
privat	0.533**	0.196
	(0.227)	(0.239)
log(antall_elever)	0.250**	0.245**
	(0.099)	(0.113)
trivsel	1.419***	1.878***
	(0.505)	(0.522)
arbeidsro	0.095	-0.068
	(0.213)	(0.241)
time_1617		-0.003
		(0.019)
dag_1617		0.073
		(0.079)
Constant	-7.985***	-9.461***
	(2.062)	(2.373)
Observations	286	277
R2	0.077	0.076
Adjusted R2	0.063	0.055
Residual Std. Error	1.327 (df = 281)	1.267 (df = 270)
F Statistic	5.825*** (df = 4; 281)	3.677*** (df = 6; 270)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figur 1: Diagnoseplott til regresjonsmodell (2) i Tabell 1