Løsningsforslag

Til sensor: Hvert delspørsmål gir i utgangspunktet 10 poeng. Det er til hvert spørsmål oppgitt trekk for en del standardfeil. Andre feil skal etter beste evne trekkes på samme måte.

Oppgave 1

(a) Vi antar uavhengige observasjoner fra to normalfordelte populasjoner, og tester om de to variansene er like:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_A : \sigma_1^2 \neq \sigma_2^2.$$

Testobservator:

$$F = \frac{S_1^2}{S_2^2} = \frac{9.97^2}{3.72^2} = 7.18.$$

Kritisk verdi for en tosidig F-test med (14,14) frihetsgrader på 5% nivå er 2.98, så **nullhypotesen om lik varians forkastes**.

Feil	Trekk
Riktig test, men regnefeil	-5
Setter brøken opp-ned, uten å finne riktig kritisk verdi	-5
Riktig regning, men gal konklusjon	-7
Riktig regning, men ingen konklusjon	-5

(b) fra spørsmål (a) har vi at variansene i de to normalfordelte populasjonene er ulike, så vi benytter t-test for populasjoner med ulik varians. Vi velger en tosidig test, fordi det ikke er noe som utelukker avvik til den ene eller andre siden, og begge sider er beslutningsrelevante:

$$H_0 : \mu_1 = \mu_2$$
$$H_A : \mu_1 \neq \mu_2.$$

Testobservator:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_1^2}{n_2}}} = \frac{56.47 - 54.60}{\sqrt{\frac{9.97^2}{15} + \frac{3.72^2}{15}}} = 0.68.$$

Antall frihetsgrader for t-test får vi fra formelen

$$\nu = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}} = \frac{(9.97^2/15 + 3.72^2/15)^2}{\frac{(9.97^2/15)^2}{14} + \frac{(3.72^2/15)^2}{14}} \approx 18.$$

Kritisk verdi for en tosidig t-test med 18 frihetsgrader på 5% nivå er 2.101, så vi **kan ikke forkaste nullhypotesen om lik reisetid**.

Feil	Trekk
Riktig test, men uten å bruke formel for frihetsgrader	-3
Mangler begrunnelse for tosidig vs. ensidig test	-2
Riktig test, men regnefeil	-5
Riktig utført test for like varianser	-5
Riktig regning, men gal konklusjon	-7
Riktig regning, men ingen konklusjon	-5

(c) Vi tester om alle tre reiserutene har lik reisetid:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
$$H_A : \text{Minst en rute har forskjellig reisetid.}$$

Fra utskriften har vi at p-verdien til testen er 0.003, så vi har en **klar forkastning på 5% nivå**. For at vi skal kunne gjøre variansanalyse må vi ha **uavhengige, normalfordelte observasjoner** fra populasjoner med **lik varians**.

Siden observasjonene er gjort etter hverandre i tid, er de neppe uavhengige. I tillegg har vi i spørsmål (a) slått fast at minst to av variansene er ulike, så **forutsetningene er ikke oppfylt**.

Å sette opp hypotesen, liste opp forutsetninger og vurdere om de er oppfylt gir 5 poeng. Å skrive rett konklusjon ved hjelp av p-verdien gir de resterende 5 poengene.

(d) La p_1 og p_2 være andelen som snek i 2015 og 2016 hhv. Vi ønsker å teste

$$H_0 : p_1 = p_2$$
$$H_A : p_1 \neq p_2.$$

Vi har følgende estimater:

$$\widehat{p}_1 = \frac{13\,064}{426\,382} = 0.0306, \quad \widehat{p}_2 = \frac{12\,904}{342\,085} = 0.0377, \quad \widehat{p} = \frac{13\,064 + 12\,904}{426\,382 + 342\,085} = 0.0338$$

Testobservator:

$$Z = \frac{\widehat{p}_2 - \widehat{p}_1}{\sqrt{\left(\frac{1}{n_2} + \frac{1}{n_2}\right)\widehat{p}(1-\widehat{p})}} = \frac{0.0377 - 0.0306}{\sqrt{\left(\frac{1}{426\,382} + \frac{1}{342\,085}\right) \cdot 0.0338 \cdot (1 - 0.0338)}} = 17.11$$

Kritisk verdi for en tosidig z-test på 5% nivå er 1.96 (evt. 1.64 for ensidig test, som også kan forsvares her), så vi **forkaster nullhypotesen om lik andel** med klar margin. Med så mange observasjoner vil vi forkaste nullhypotesen selv ved veldig små avvik, om forskjellen fra 2015 til 2016 har noen praktisk betydning, er ikke åpenbart.

Feil	Trekk
Riktig test, men regnefeil	-5
Riktig test, men bruker feil tall	-7
Riktig regning, men gal konklusjon	-7
Riktig regning, men ingen konklusjon	-5

(e) Hvis vi antar at hver reise skjer uavhengig av hverandre og at det er en konstant sannsynlighet p for billettkontroll, er antall ganger postdoktoren tas for sniking i løpet av en måned, X, binomisk fordelt med $n = 2 \cdot 20 = 40$ og «suksess»-sannsynlighet p. Vi estimerer p som

$$\widehat{p} = \frac{342\ 085}{56.5 \text{ mill}} = 0.006.$$

Da estimerer vi $\mathrm{E}(X)$ som $n\widehat{p} \approx 0.242$, og **forventet månentlig utgift som kr $\mathbf{1150 \cdot 0.242 \approx 279}$**.

Postdoktoren vil da i det lange løp spare $960 - 279 = 671$ kroner i måneden på å snike i stedet for å kjøpe månedskort, men at forutsetningene om uavhengighet og konstant kontollsannsynlighet er oppfylt, er derimot heller tvilsomt (se neste spørsmål).

Full pott dersom studenten gir et fullverdig argument via den binomiske fordelingen eller tilsvarende. Ellers:

Feil	Trekk
Kommer frem til riktig svar ved hjelp av uformelle argumenter	-2
Bruker feil tall, men ellers riktig argument	-5
Bruker 20 i stedet for 40 (altså t/r) i utregningen	-2

(f) I spørsmål (d) betyr den ekstra informasjonen av det ikke er kontollert tilfeldige utvalg, men at kontollørene bevisst går etter avganger der de mistenker en høy snikeandel, og at de er blitt flinkere til dette med tiden. **Vi kan derfor ikke tolke forkastningen som at snikeandelen har økt** (men heller ikke som at kontrollørene er blitt flinkere, for det *kan* jo hende at snikeandelen faktisk har økt!).

I spørsmål (e) har den nye informasjonen absolutt noe å si, fordi sannsynligheten for å bli tatt, og dermed også forventet utgift, vil være avhengig av om de aktuelle avgangene som postdoktoren benytter er blant de som kontrollørene sikter seg inn på.

To deler, hver del teller like mye. Presise svar med gode begrunnelser kan belønnes med full pott, selv om de ikke er akkurat som løsningsforslaget. Lange, upresise og vage svar trekkes det for. Forsøk på helgardering gir null poeng.

(g) Vi har tre observasjoner:

$$y_1 = \frac{11\ 709}{574\ 005} = 0.0204, \;\; y_2 = 0.0306, \;\text{og}\; y_3 = 0.0377.$$

Den eksponensielle glattingen blir da

$$\begin{aligned} S_1 &= y_1 = 0.0204 \\ S_2 &= 0.5 \cdot y_2 + 0.5 \cdot S_1 = 0.0255 \\ S_3 &= 0.5 \cdot y_3 + 0.5 \cdot S_2 = 0.0316 \end{aligned}$$

Prediksjonen for 2017 er gitt ved $\boldsymbol{S_3 = 0.0316}$.

OBS! Følgende formel verserer i diverse gamle kompendier, og skal ikke gi trekk. Den markerte $\boldsymbol{w}$'en skal ikke være der:

$$S_t = wy_t + w(1-w)y_{t-1} + w(1-w)^2 y_{t-2} + \cdots + \boldsymbol{w}(1-w)^{t-1}y_1.$$

Feil	Trekk
Riktige formler, feil tall	-5
Riktig, men kommer ikke helt i mål med prediksjonen	-3

Oppgave 2

(a) I analyse 1 er det fem forklaringsvariable, hvorav to er signifikant forskjellige fra null. Dersom driftsutgiftene til en kommune øker med 1000 kroner, går forventet score på leseprøven *ned* med 0.02 poeng, alt annet likt. Videre ser vi at kommuner som har nynorsk som målform scorer 0.78 poeng dårligere enn kommuner med nøytral målform. Bokmålskommuner, derimot, estimeres til å score 0.21 poeng bedre enn kommuner med nøytral målform, men denne effekten er ikke signifikant på 5% nivå. Heller ikke koeffisientene til mobbevariabelen eller Nord-Norge-dummyen er signifikant forskjellige fra null.

Konstantleddet er klart signifikant, men har ingen fornuftig tolkning i dette tilfellet.

Vi legger merke til at det bare er 288 av 425 kommuner som er med i denne analysen, og at det fra informasjonen vi har kan tyde på at det er de minste kommunene som er utelatt. Siden begge variablene med signifikante koeffisienter (`driftsutgifter` og `nynorsk`) sannsynligvis samvarierer med kommunestørrelsen, er det grunn til å tro at det ligger viktig informasjon i de utelatte observasjonene, og det er noe som bør undersøkes nærmere.

Forklaringsgraden på 0.07 er forholdsvis lav, så mesteparten av variasjonen i testscore lar seg ikke forklare med disse variablene.

Full score gis til svar som er relativt korte og konsise, og der studenten viser fortsåelse for størrelsene som estimeres, og ikke bare lister opp koeffisientene og sier om de er signifikante eller ikke, og som skriver mer om forklaringsgraden enn at den er «bra» eller «dårlig». Ellers:

Feil	Trekk
Ingen referanse til at manglende observasjoner kan være problematisk her	-2
Mekansik opplisting, uten å vise forståelse	-5 eller mer
«R^2 er dårlig» e.l., uten noe mer	-2
Langt, usammenhengende svar	-5

(b) Setter inn i formelen for regresjonslinjen, og får

$$\widehat{y} = -0.02 \cdot 110 + 0.04 \cdot 5 + 0.21 - 0.60 + 50.81 = \mathbf{48.42}$$

Feil	Trekk
Bruker feil benevning (110000 i stedet for 110, desimal i stedet for prosenttall)	-7
Feilen over, men viser at en skjønner at svaret er feil	-5
Bruker bare koeffisientene som er signifikant forskjellige fra null	-7

(c) Vi kaller denne koeffisienten β, og tester

$$H_0 : \beta = 0$$
$$H_A : \beta \neq 0.$$

Testobservatoren er

$$T = \frac{\widehat{\beta}}{S(\widehat{\beta})} = \frac{-0.02}{0.01} = -2.$$

Antall frihetsgrader i t-testen er $n - k - 1 = 288 - 5 - 1 = 282$, men så langt går ikke tabellen vår, så vi bruker i stedet kritisk verdi fra normalfordelingen, som for en tosidig test på 5% nivå er 1.96. **Vi forkaster følgelig nullhypotesen.**

Feil	Trekk
Riktig test, men regnefeil	-5
Riktig regning, men gal konklusjon	-7
Riktig regning, men ingen konklusjon	-5

(d) Bortsett fra tre avvikende observasjoner i den øvre halen ligger punktene i QQ-plottet pent på en linje. I tillegg er histogrammet forholdsvis symmetrisk og klokkeformet, så residualene ser ut til å være temmelig normalfordelte. Vi ser ingen tegn til heteroskedastisitet dersom vi plotter residualene mot de predikerte verdiene, men dersom vi plotter dem mot logaritmen til antallet 5.-klassinger i kommunen, er det tydelig heteroskedastisitet: resultatene fra små kommuner varierer mer enn fra større kommuner. (Det er for øvrig ikke overraskende, se oppgave (f)). Heteroskedastisitet er ikke nødvendigvis et alvorlig problem i regresjonsanalyse, men inferensen blir ikke eksakt.

Feil	Trekk
Langt, usammenhengende svar	-5

(e) Ved å gå fra analyse 3 til analyse 4 bytter vi ut `log(antall)` som forklaringsvariabel med `log(folketall)`. Resten av koeffisientene er tilnærmet uforandret, og koeffisientene til `log(antall)` og `log(folketall)` er tilnærmet like. Den logaritmiske funksjonsformen fører til at vi kan tolke koeffisientene som prosent: 1% økning i elevtall (eller folketall) fører til en økning i testresultat på 0.0042 poeng (eller 0.0044). Dette er veldig naturlig, siden folketall og antall elever henger nøye sammen, og er sterkt korrelerte.

I analyse 5 bruker vi både `log(antall)` og `log(folketall)` som forklaringsvariable, men da mister de begge sin signifikans. Det er et klassisk tilfelle av multikolinearitet, siden de to variablene forklarer den samme tingen (kommunestørrelse), og vi kan ikke skille effektene fra hverandre. En F-test vil sannsynligvis avsløre at begge disse koeffisientene ikke begge kan være null samtidig, men det har vi ikke gjort her.

Vi legger merke til at `nynorsk` forblir signifikant negativ i alle disse tre analysene, noe som er en interessant observasjon siden det ser ut til at det nynorskkommuner gjør det dårligere enn andre kommuner, *selv når vi har kontrollert for kommunestørrelsen.*

I analyse 6 benytter vi bare `log(folketall)`, og den koeffisienten blir klart signifikant. Det er denne modellen som har størst justert R^2, noe som forteller oss at resten av variablene ikke har noen særlig forklaringskraft.

Et mulig neste modell kunne være å bruke `log(folketall)` sammen med `nynorsk`, for å se om forklaringskraften kan øke enda litt mer.

Her er det noen elementer som bør være med: multikolinearitet, og at justert R^2 er størst i modellen med færrest forklaringsvariable. Bonus for å fange opp poenget med at nynorsk ser ut til å ha en selvstendig effekt. Som før straffes vissvass hardt, og vi er ikke ute etter fire opplistinger a la spørsmål (a). Det er spesifisert i spørsmålet at vi er ute etter en overordnet vurdering. Gi poeng etter beste skjønn.

(f) La n_j være antall elever i kommunen, da er gjennomsnittsscoren gitt som $Y_j = (1/n_j)\sum_{i=1}^{n_j} Z_{ij}$. Ved å ta gjennomsnitt over alle elever i kommunen får vi følgende regresjonsmodell:

$$\begin{aligned} Y_j &= \frac{1}{n_j}\sum_{i=1}^{n_j} Z_{ij} = \frac{1}{n_j}\sum_{i=1}^{n_j}(\alpha + \beta X_j + \epsilon_{ij}) \\ &= \alpha + \beta X_j + \bar{\epsilon}_j. \end{aligned}$$

Så alt er likt, bortsett fra feilleddet, som nå er gitt ved $\bar{\epsilon}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\epsilon_{ij}$. Vi får da at $\mathrm{E}(\bar{\epsilon}_j) = 0$ som før, men at

$$\mathrm{Var}(\bar{\epsilon}_j) = \mathrm{Var}\left(\frac{1}{n_j}\sum_{i=1}^{n_j}\epsilon_{ij}\right) = \frac{1}{n_j^2}\cdot n_j \cdot \mathrm{Var}(\epsilon_{ij}) = \frac{\sigma}{n_j}.$$

Feilleddet er dermed heteroskedastisk, fordi variansen til feilleddet blir mindre for større kommuner. Det stemmer overens med bildet i diagnoseplott 3 i Figur 1, selv om det er en mer komplisert situasjon fordi det er flere forklaringsvariable.

Dette skal være det vanskelige spørsmålet. For å få full pott kreves det et formelt argument som utleder variansen til feilleddet. Ellers:

Feil	Trekk
Klarer å sette opp modellen korrekt, men klarer ikke å vise heteroskedastisitet	-8
Klarer seg greit og viser forståelse, men med et mer uformelt argument	-5

(g) **Svaret er nei**. Tolkningen av regresjonskoeffisienten til `fritak` i analyse 2 er at for hvert prosentpoeng økning i fritaksgraden, øker kommunens testscore med 0.10. Effekten er signifikant forskjellig fra null, men den sier ikke noe om det er en kausal sammenheng mellom høyere fritaksgrad og bedre testscore, og vi kan *i alle fall* ikke si noe om det ligger en ond hensikt fra kommunens side bak en slik effekt.

For å få full pott bør det være både en generell betraktning om at vi ikke kan tolke regresjonskoeffisienter kausalt, og at vi uansett ikke kan si noe om den ene eller andre hensikten ligger bak. Utover det skal svar som viser forståelse for denne problemstillingen belønnes. Forsøk på helgardering gir null poeng.

(h) Fortolkningen av regresjonskoeffisienten til `fritak` i dette eksempelet at at et prosentpoeng høyere fritaksgrad fører til 1.10 poeng høyere testscore. Denne koeffisienten er signifikant forskjellig fra null, med en p-verdi på 0.0208.

Man må være oppmerksom på at det ser ut til at en skole er meget innflytelsesrik. Det er heller tvilsomt at vi ville fått et stigningstall signifikant forskjellig fra null dersom vi hadde tatt ut skolen med høyest fritaksgrad.

Journalisten bør først bekrefte at resultatet fra denne skolen er korrekt, og deretter samle inn mer data for å fosøke å kontrollere for andre variable, som geografisk plassering, andel minoritetsspråklige elever osv, som kan være med å forklare variasjon i testresultatet.

Dersom fritaksgraden fortsatt har selvstendig forklaringskraft, kan det være betimelig med noen spørsmål til de ansvarlige myndigheter om hvordan ordningen med fritak fungerer, og om det er mulig for skolene å utnytte den for å oppnå høyere testresultat.

Tre elementer for full score, som teller like mye: 1. Riktig tolkning av koeffisient. 2. Identifisere problemet med en særlig innflytelsesrik observasjon. 3. Komme med et godt svar om hvordan journalisten kan håndtere det. På siste punktet er det mange frihetsgrader, der forståelse og klarhet belønnes, mens langtekkelighet og vaghet trekkes for.