

Oppgave 1 Forskningsmidler

Anta at 1635 menn og 1188 kvinner har sendt inn søknader om prosjektmidler. Resultatene viser at 290 av mennene og 177 av kvinnene fikk et positivt svar.

- a) Kan man konkludere at tildeling av prosjektmidler er avhengig av kjønn? Skriv ned nullhypotesen og den alternative hypotesen. Sett opp kontingenstabellen og beregn forventede frekvenser under nullhypotesen.
Bruk en χ^2 -test med signifikansnivå $\alpha = 0.05$. Hva er konklusjonen?
- b) Dataanalysen kan også gjennomføres ved et to-utvalgs oppsett hvor man sammenligner den sanne andel av menn som får prosjektmidler med den sanne andel av kvinner som får prosjektmidler.
Skriv ned nullhypotesen og den alternative hypotesen, og utfør en z-test for å teste hypotesen. Bruk signifikansnivå $\alpha = 0.05$
- c) Hvordan kan man sammenligne testene i deloppgavene a) og b)? Vis at de to testene er ekvivalente. Er det grunner til å foretrekke én av dem over den andre?

Oppgave 2 Integralberegning

a) Beregn følgende integral ved å bruke en Student t-fordeling:

$$\int_0^{\infty} \frac{1}{1+t^2} dt$$

Oppgave 3 Enkel lineær modell

Anvendte statistikere finner det svært nyttig å se på residualer for å vurdere tilpasningen av en rett linje gjennom et gitt sett med n punkter $(x_1, Y_1), \dots, (x_n, Y_n)$.

- a) Definer hva et residual er. Bruk en liten skisse i forklaringen.

For en enkel lineær modell kan man plotte residualer mot x_i for alle i i et såkalt residualplott for å kontrollere modellantakelsene. Skissér to mulige residualplott der hvert av dem viser at én av modellantakelsene for en lineær modell er brutt.

Forklar for hvert av dem hvilken antakelse som er brutt og hvordan dette vises i residualplottet.

- b) Her antar vi at regresjonslinjen skal gå gjennom $(0,0)$, slik at

$$Y_i = \beta x_i + \epsilon_i$$

for $i = 1, \dots, n$, der β er en ukjent parameter og ϵ_i er normalfordelt med forventning 0 og ukjent varians σ^2 . Videre antas at $\epsilon_1, \dots, \epsilon_n$ er uavhengige.

Beregn sannsynlighetsmaksimeringsestimatorene $\hat{\beta}$ og $\hat{\sigma}^2$ for henholdsvis β og σ^2 .

Finn fordelingen for $\hat{\beta}$ (inkludert alle parametre).

Oppgave 4 Pris på fjellhytter

Firmaet Fjellmegleren i Oppdal har trukket ut 12 hytter i Oppdalsområdet som er solgt det siste året. Anta at prisen på ei hytte, Y , i tusen kroner, er en kontinuerlig fordelt stokastisk variabel. Anta videre at prisene på forskjellige hytter er uavhengige stokastiske variabler. Observasjonene er gitt nedenfor.

Salgspriser i Oppdal (i 1000 kr, ordnet i stigende rekkefølge):

1990 2390 2450 2580 2590 2610 2675 2770 3190 3440 4400 5500

- a) Hva menes med medianen, $\tilde{\mu}$, for en kontinuerlig fordelt stokastisk variabel Y ?

Fjellmegleren vil teste om medianen i fordelingen for hytteprisene i Oppdal er over 2.5 millioner kroner (dvs. over 2500 tusen kroner). Formuler dette spørsmålet som et testingsproblem om $\tilde{\mu}$. Sett opp nullhypotesen H_0 og den alternative hypotesen H_1 .

Forklar hvordan testingen kan gjøres ved hjelp av tegntesten. Utfør testen ved å regne ut p-verdien. Hva blir konklusjonen dersom signifikansnivået settes til 5%?

- b) Hvilken tilleggsantagelse må gjøres dersom man ønsker å teste H_0 mot H_1 i forrige delpunkt ved hjelp av Wilcoxons ett-utvalgstest (*Wilcoxon signed rank test*)?

Utfør også denne testen med de gitte dataene. Hva blir konklusjonen dersom signifikansnivået settes til 5%?

Fjellmegleren vil videre teste om det er forskjeller i prisnivået på hytter i Oppdal og to andre populære fjellområder, Hafjell og Trysil. Firmaet har derfor også trukket ut 8 hytter fra Hafjell og 8 hytter fra Trysil som er solgt de siste månedene. Vi antar at sannsynlighetstettheten for hyttepriser i Hafjell og i Trysil har samme form som sannsynlighetstettheten for hyttepriser i Oppdal, og at prisene på forskjellige hytter er uavhengige variabler. Observasjonene er gitt nedenfor, med tallene fra Oppdal gjentatt her for ordens skyld.

Salgspriser for hytter (i 1000 kr, ordnet i stigende rekkefølge innenfor hvert område):

Oppdal:	1990	2390	2450	2580	2590	2610	2675	2770	3190	3440	4400	5500
Hafjell:	2660	2810	2880	2900	2950	3290	4830	5320				
Trysil:	2070	2730	3080	3150	3230	3370	3620	4125				

- c) Tyder salgprisene ovenfor på at det er prisforskjeller på hytter i de tre fjellområdene? Formuler dette spørsmålet som et testingsproblem. Sett opp nullhypotesen H_0 og den alternative hypotesen H_1 .

Gjennomfør testingen ved hjelp av en Kruskal-Wallis test. Bruk signifikansnivå 5% og angi konklusjonen med ord.

- d) Hvilken annen ikke-parametrisk test enn den som ble brukt i punkt c), kunne man ha brukt dersom salgspriser for hytter i bare to fjellområder skulle sammenlignes? Forklar kort hvordan du da ville gjennomføre testingen.

Hvilken test kan man bruke dersom man i punkt c) antar at salgprisene innenfor hvert område er normalfordelte med samme varians σ^2 ? Forklar kort hvordan du da ville gjennomføre testingen.