

Oppgave 1 Forskningsmiddel

Anta at 1635 menn og 1188 kvinner har sendt inn søknader om prosjektmiddel. Resultata viser at 290 av mennene og 177 av kvinnene fekk eit positivt svar.

- a) Kan ein konkludere at tildeling av prosjektmiddel er avhengig av kjønn? Skriv ned nullhypotesen og den alternative hypotesen. Set opp kontingenstabellen og berekn forventa frekvensar under nullhypotesen.

Bruk ein χ^2 -test med signifikansnivå $\alpha = 0.05$. Kva er konklusjonen?

- b) Dataanalysen kan òg gjennomførast ved eit to-utvals oppsett der ein samanliknar den sanne andelen av menn som får prosjektmiddel med den sanne andelen av kvinner som får prosjektmiddel.

Skriv ned nullhypotesen og den alternative hypotesen, og utfør ein z-test for å teste hypotesen. Bruk signifikansnivå $\alpha = 0.05$

- c) Korleis kan ein samanlikne testane i deloppgåvene a) og b)? Vis at dei to testane er ekvivalente. Er det grunnar til å føretrekke ein av dei over den andre?

Oppgave 2 Integralberegning

a) Berekn følgende integral ved å bruke ei Student-t fordeling:

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

Oppgave 3 Enkel lineær modell

Anvende statistikker finn det svært nyttig å sjå på residual for å vurdere tilpassinga av ei rett linje gjennom eit gitt sett med n punkt $(x_1, Y_1), \dots, (x_n, Y_n)$.

- a) Definer kva eit residual er. Bruk ei lita skisse i forklaringa.

For ein enkel lineær modell kan ein plote residual mot x_i for alle i i eit såkalla residualplott for å kontrollere modellføresetnadene. Skissér to moglege residualplott der kvart av dei viser at ein av modellføresetnadene for ein lineær modell er broten. Forklar for kvart av dei kva antakelse som er broten og korleis dette blir vist i residualplottet.

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

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

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

Berekn sannsynsmaksimeringsestimatorane $\hat{\beta}$ og $\hat{\sigma}^2$ for høvesvis β og σ^2 .

Finn fordelinga for $\hat{\beta}$ (inkludert alle parametrar).

Oppg ve 4 Pris p  fjellhytter

Firmaet Fjellmegleren i Oppdal har trekt ut 12 hytter i Oppdalsområdet som er selt det siste  ret. Anta at prisen p  ei hytte, Y , i tusen kroner, er ein kontinuerleg fordelt stokastisk variabel. Anta vidare at prisane p  forskjellige hytter er uavhengige stokastiske variablar. Observasjonane er gitt nedanfor.

Salsprisar i Oppdal (i 1000 kr, ordna i stigande rekkef lgje):

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

- a) Kva meiner ein med medianen, $\tilde{\mu}$, for ein kontinuerleg fordelt stokastisk variabel Y ?

Fjellmegleren vil teste om medianen i fordelinga for hytteprisane i Oppdal er over 2.5 millionar kroner (dvs. over 2500 tusen kroner). Formuler dette sp rsmålet som eit testingsproblem om $\tilde{\mu}$. Set opp nullhypotesen H_0 og den alternative hypotesen H_1 .

Forklar korleis testinga kan gjerast ved hjelp av teikntesten. Utf r testen ved   rekne ut p-verdien. Kva blir konklusjonen dersom signifikansniv et blir sett til 5%?

- b) Kva tilleggsf resetnad m  gjerast dersom ein  nsker   teste H_0 mot H_1 i f rre delpunkt ved hjelp av Wilcoxons ett-utvalgstest (*Wilcoxon signed rank test*)?

Utf r  g denne testen med dei gitte dataa. Kva blir konklusjonen dersom signifikansniv et blir sett til 5%?

Fjellmegleren vil vidare teste om det er forskjellar i prisniv et p  hytter i Oppdal og to andre populære fjellomr de, Hafjell og Trysil. Firmaet har derfor ogs  trekt ut 8 hytter fr  Hafjell og 8 hytter fr  Trysil som er selt dei siste m nadene. Vi antar at sannsynstettleiken for hytteprisar i Hafjell og i Trysil har same form som sannsynstettleiken for hytteprisar i Oppdal, og at prisane p  forskjellige hytter er uavhengige variablar. Observasjonane er gitt nedanfor, med tala fr  Oppdal gjentatt her for ordens skyld.

Salsprisar for hytter (i 1000 kr, ordna i stigande rekkef lgje innanfor kvart 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 salsprisane ovanfor på at det er prisforskjellar på hytter i dei tre fjellområda? Formuler dette spørsmålet som eit testingsproblem. Set opp nullhypotesen H_0 og den alternative hypotesen H_1 .

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

- d) Kva annan ikkje-parametrisk test enn den som vart brukt i punkt c), kunne ein ha brukt dersom salsprisar for hytter i berre to fjellområde skulle samanliknast? Forklar kort korleis du då ville gjennomføre testinga.

Kva test kan ein bruke dersom ein i punkt c) antar at salsprisane innanfor kvart område er normalfordelte med same varians σ^2 ? Forklar kort korleis du då ville gjennomføre testinga.