

Plenumsregning – Uke 12

Geir-Arne Fuglstad, IMF

Vår 2024 – Oppgave 6

Lakseprodusenten SalFisk AS ønsker å undersøke om en avlusingsbehandling påvirker velferd hos laks. SalFisk AS vil gjennomføre et forsøk der 15 laksemerder behandles med samme metode og under like omstendigheter. En egnet numerisk skadeindikator skal måles i hver av de 15 merdene før og etter behandling. En *økning* i denne indikatoren kan indikere dårligere velferd som følge av behandling, men målingene er forbundet med usikkerhet.

La de uavhengige stokastiske variablene $D_1, D_2, \dots, D_{15}$ representere differansen i skadeindikator fra før til etter behandling.

- a) Anta $D_1, \dots, D_{15}$ er uavhengige og normalfordelte med forventning μ og varians σ^2 . Vi skal teste $H_0: \mu = 0$ mot $H_1: \mu > 0$ gjennom testobservator $T = \bar{D} / \sqrt{S^2/15}$, der S^2 er empirisk varians.
1. Skisser fordeling til testobservator under H_0 , og marker forkastningsområde ved signifikansnivå $\alpha = 0.05$ i skissen.
 2. Vi observerer $\sum_{i=1}^{15} d_i = 8.55$ og $\sum_{i=1}^{15} (d_i - \bar{d})^2 = 17.64$. Hva blir konklusjonen av testen?
 3. **Ekstra:** Finn en tilnærming til minste n som kreves for at styrke er minst 0.99 for $\mu = 0.2$ ved å anta kjent $\sigma^2 = s^2$ fra verdi regnet ut i 2..

Vår 2024 – Oppgave 6

Lakseprodusenten SalFisk AS ønsker å undersøke om en avlusingsbehandling påvirker velferd hos laks. SalFisk AS vil gjennomføre et forsøk der 15 laksemerder behandles med samme metode og under like omstendigheter. En egnet numerisk skadeindikator skal måles i hver av de 15 merdene før og etter behandling. En *økning* i denne indikatoren kan indikere dårligere velferd som følge av behandling, men målingene er forbundet med usikkerhet.

La de uavhengige stokastiske variablene $D_1, D_2, \dots, D_{15}$ representere differansen i skadeindikator fra før til etter behandling.

b) SalFisk AS er usikre på normalantagelsen, og vil i stedet gjennomføre en fortegnstest for å undersøke hypotesene

$$H_0 : \tilde{\mu} = 0 \text{ mot } H_1 : \tilde{\mu} > 0,$$

der $\tilde{\mu}$ er medianen til hver av de 15 målingene $D_1, D_2, \dots, D_{15}$. De ønsker å gjennomføre testen ved signifikansnivå $\alpha = 0.05$.

- (i) Skriv ned en egnet testobservator og beslutningsregel for fortegnstesten.
- (ii) Hva blir det faktiske signifikansnivået til testen, og hvorfor kan vi ikke oppnå et signifikansnivå nøyaktig lik 0.05 i slike tester?
- (iii) SalFisk AS observerte at skadeindikatoren økte i 10 av 15 merder. Hva blir resultatet av testen?

Ekstra 1: To-utvalg permutasjonstest

NB: Det er foreløpig ingen tidligere eksamensoppgaver om permutasjonstester.

Vi har samlet 9 målepunkter, $x_1, \dots, x_9$, fra Maskin A og 7 målepunkter, $y_1, \dots, y_7$, fra Maskin B. Slik som visualisert til høyre. Disse maskinene er i utgangspunktet identifiserte og vi er ikke villig til å gjøre noen parametriske antagelser.

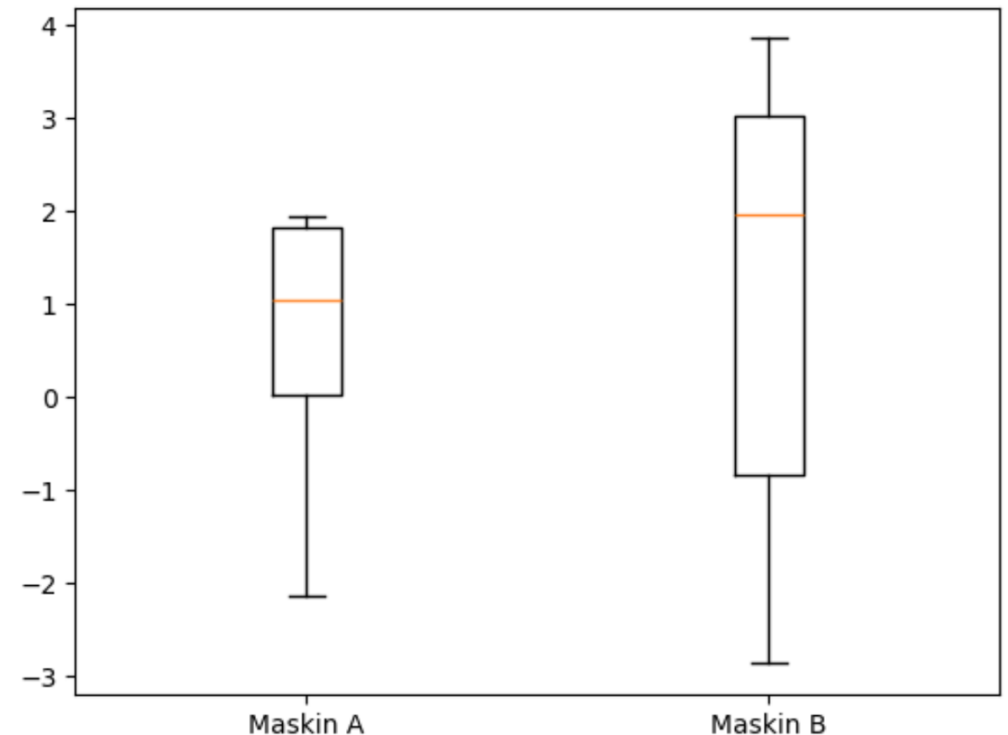
- a) Forklar hvorfor vi kan bruke en to-utvalg permutasjonstest for å undersøke om spredningen er større for Maskin B enn Maskin A.
- b) Formuler en nullhypotese og alternativ hypotese.
- c) Forklar hvorfor en mulig testobservator er

$$T = \frac{1}{7} \sum_{i=1}^7 (Y_i - \bar{Y})^2 - \frac{1}{9} \sum_{i=1}^9 (X_i - \bar{X})^2$$

der $X_1, \dots, X_9, Y_1, \dots, Y_7$ er stokastiske under tilfeldige permutasjoner av målepunktene.

- d) Estimer fordelingen til T under nullhypotesen i et histogram, og marker observert verdi.
- e) Beskriv hvordan man kan estimere p-verdi, og estimer tallverdi (i Python). Hva blir konklusjonen?

Visualisering av forskjeller



Ekstra 2

Anta $X_1, \dots, X_n \sim_{iid} N(\mu, \sigma^2)$ der $\mu = 10$ er **kjent** og σ^2 er ukjent. Vi vurderer to ulike testobservatorer:

1. $T_1 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ og

2. $T_2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

for å teste $H_0: \sigma^2 = 4$ mot $H_1: \sigma^2 > 4$.

a) Bestem beslutningsregelen under signifikansnivå $\alpha = 0.05$ for hver testobservator.

b) Anta $n = 10$. Bestem styrken til den resulterende hypotesetesten når $\sigma^2 = 9$ for hver av testobservatorene. (Dere trenger Python for å finne tallverdier til styrkene)

Sommer 2019 – Oppgave 3

Anta at tidene $X_1, X_2, \dots, X_n$ det tar før pasientene til selskap A sovner er et tilfeldig utvalg med ukjent forventningsverdi μ og kjent standardavvik $\sigma = 10$. Videre kan du anta at de tilsvarende tidene $Y_1, Y_2, \dots, Y_n$ for selskap B er uavhengige og normalfordelt med ukjent forventningsverdi θ og kjent standardavvik $\tau = 12$. Du kan anta at de to utvalgene er uavhengige av hverandre.

Selskap B ønsker å utføre hypotesetesten

$$H_0: \mu = \theta \quad \text{mot} \quad H_1: \mu > \theta. \quad (2)$$

med et signifikansnivå $\alpha = 0.1$ basert på de tilfeldige utvalgene $X_1, X_2, \dots, X_n$ og $Y_1, Y_2, \dots, Y_n$.

- b) Anta at den sanne differansen mellom effekten til de to selskapene er $\mu - \theta = 5$ minutt.

Utløst et uttrykk for det minste antallet pasienter n begge selskapene må gi sovemedisinen sin til dersom de krever at testens styrke skal være minst 95 % dersom den sanne differansen er $\mu - \theta = 5$ minutt.