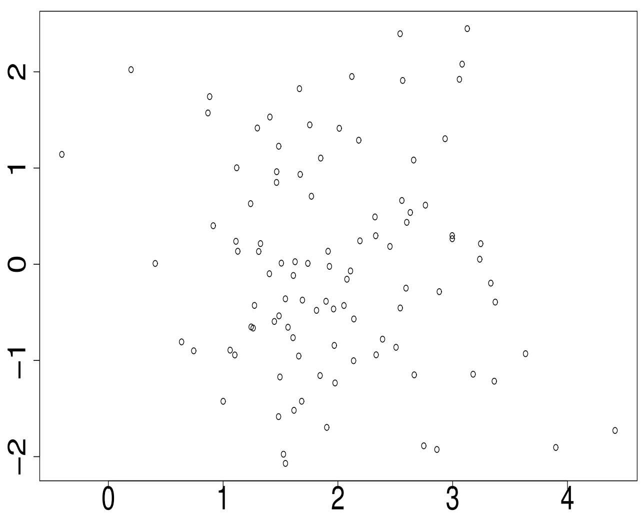
TMA4245 Statistikk
Eksamen juni 2015

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

**Oppgave 1**




Figur 1: Parvise data fra en simultanfordeling.

Figur 1 viser parvise utfall av to stokastiske variabler $X$ (horisontalt) og $Y$ (vertikalt).

**a)** Hva vil du anslå at korrelasjonen mellom $X$ og $Y$ er? Gi en kort begrunnelse.

Forventningsverdien og standardavviket for hver av de to variablene har heltallsverdier. Anslå disse verdiene ved å studere Figur 1 visuelt.

**Oppgave 2**

Anta at $A$ og $C$ er uavhengige hendelser, og at $B$ og $C$ også er uavhengige hendelser.

**a)** Kan $A \cap B$ og $C$ være avhengige? Kan $A \cup B$ og $C$ være avhengige? (Hint: Se f.eks. på situasjonen $A \cap B \cap C = \emptyset$.)

Kan $A \cup B$ og $C$ være avhengige dersom $A$ og $B$ er disjunkte?

**Oppgave 3**

En farmasøytisk fabrikk produserer en medisin i flytende form som selges på flasker med gitt innhold. Denne medisinen inneholder en viktig komponent som krever at det foretas en fortløpende kvalitetskontroll. Dette gjøres ved at det for hver produksjonsserie tas ut et tilfeldig utvalg flasker hvor innholdet analyseres. Hver gang en serie prøver analyseres, analyseres også en kontrolløsning med kjent konsentrasjon 0.10 mg/l. Dette gjøres for å sikre at analysemetoden er riktig kalibrert. Siden analysemetoden ikke er helt presis, vil målt konsentrasjon variere. Utfallet av analysen av kontrolløsningen kan regnes som en normalfordelt stokastisk variabel $X$ med forventningsverdi $\mu$ og varians $\sigma^2$, der $\sigma^2$ er variansen i målefeilen ved analysemetoden og $\mu$ under normale omstendigheter er lik 0.10 mg/l.

En *alarmhendelse* $A$ for analysemetoden er definert ved at målt verdi $X$ i kontrolløsningen avviker mer enn ett standardavvik fra konsentrasjonen 0.10 mg/l, altså $|X - 0.1| > \sigma$ (dvs. $X < 0.1 - \sigma$ eller $X > 0.1 + \sigma$). En *aksjonshendelse* $B$ er definert ved at målt verdi $X$ i kontrolløsningen avviker mer enn 2 standardavvik fra 0.10 mg/l, dvs. $|X - 0.1| > 2\sigma$.

**a)** Anta (bare i dette punktet) at $\sigma = 0.01$ mg/l.

Beregn $P(B)$ og $P(B \mid A)$ når $\mu = 0.10$ mg/l.

Anta så at en urenhet har sneket seg inn i prøven, slik at $\mu = 0.11$ mg/l, og beregn nå $P(B)$.

Anta i resten av oppgaven at $\mu = 0.10$ mg/l. For å estimere $\sigma^2$ skal vi benytte måleresultatene $x_1$, $x_2$, $\ldots$, $x_n$, fra $n$ uavhengige analyser av kontrolløsningen. Vi kan dermed betrakte $x_1, x_2, \ldots, x_n$ som utfall fra $n$ uavhengige stokastiske variabler $X_1, X_2, \ldots, X_n$, hvor hver $X_i$, $i = 1, \ldots, n$, har samme fordeling som den stokastiske variabelen $X$.

**b)** Vis at sannsynlighetsmaksimeringsestimatoren (SME) for $\sigma^2$ er

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

Selv om $\mu$ er kjent, kan en også bruke estimatoren

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

hvor $\overline{X} = \sum_{i=1}^{n} X_i / n$.

I de neste punktene kan du bruke uten bevis at

$$\sum_{i=1}^{n} (X_i - \mu)^2 / \sigma^2 \ \text{ og } \ \sum_{i=1}^{n} (X_i - \overline{X})^2 / \sigma^2$$

er $\chi^2$-fordelte(kjikvadratfordelte) med henholdsvis $n$ og $n-1$ frihetsgrader.

**c)** Vis at både $\widehat{\sigma}^2$ og $S^2$ er forventningsrette.

Hvilken av de to estimatorene vil du foretrekke? Begrunn svaret.

Resultatene målt i mg/l av 20 analyser av kontrolløsningen har gitt $\sum_{i=1}^{20} x_i = 1.9240$ og $\sum_{i=1}^{20} x_i^2 = 0.1866$.

**d**) Utled et 90 %-konfidensintervall for $\sigma^2$ ved å benytte den estimatoren du anbefalte i (c).

**Oppgave 4**

La $X_1, ..., X_n$ angi et tilfeldig utvalg (uavhengige, identisk fordelte stokastiske variabler) fra en eksponensialfordeling med ukjent parameter $\mu$, dvs. fra en fordeling med sannsynlighetstetthet

$$f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}} ;\ x \geq 0,\ \mu > 0.$$

Nullhypotesen $H_0 : \mu = 1$ skal testes mot alternativet $H_1 : \mu < 1$. I fÃ¸rste omgang benyttes testobservatoren

$$T = \min\{X_1, ..., X_n\},$$

**a**) Vis at sannsynlighetstettheten til $T$ er

$$f_T(t) = \frac{n}{\mu} e^{-\frac{nt}{\mu}} ;\ t \geq 0.$$

Bestem forventningsverdien og variansen til $T$.

Testen som baserer seg på bruk av $T$, forkaster $H_0$ for små verdier av $T$. Dette kan formuleres slik:

Test1: Hvis $T < c_1$, forkastes $H_0$.

Signifikansnivået for testene i denne oppgaven er $\alpha$.

**b**) Vis at

$$c_1 = \frac{1}{n} \ln \frac{1}{1 - \alpha}.$$

SME for $\mu$ er utvalgsgjennomsnittet $\overline{X} = \Sigma_{i=1}^{n} X_i / n$.

**c**) Formuler en Test2 basert på $\overline{X}$. Bestem en tilnærmet sannsynlighetsfordeling til $\overline{X}$, og beregn (tilnærmet) kritisk område for Test2 dersom du kan anta at $n \geq 30$.

**d**) Finn et uttrykk for teststyrken for hver av de to testene ($n \geq 30$). (Teststyrke er sannsynlighet for å forkaste $H_0$ når $H_0$ ikke er riktig; styrken er avhengig av et skjerpet alternativ.)

Hvilken test er best? Hvorfor? Du får lov til å basere din konklusjon på resultater funnet for $\alpha = 0.05$, $n = 30$, og skjerpet alternativ hypotese $H_1' : \mu = 0.8$.

Hvorfor er Test1 en dårlig test?

# Fasit

**2**. **a**) Ja, Ja, Nei

**3**. **a**) 0.0456, 0.1437, 0.16 **c**) Foretrekker $\widehat{\sigma}^2$ **d**) $[0.000057, 0.000166]$

**4**. **a**) $\mathrm{E}[T] = \mu/n$, $\mathrm{Var}[T] = \mu^2/n^2$ **c**) Forkast $H_0$ dersom $\overline{X} \leq 1 - \frac{z_\alpha}{n^{1/2}}$.