**Norges teknisk–**
**naturvitenskapelige universitet**
**Institutt for matematiske fag**



# LØSNINGSFORSLAG TIL EKSAMEN I FAG TMA4240/TMA4245 STATISTIKK

Lørdag 19.august 2006

## Oppgave 1 Kvalitetskontroll av poteter

**a)** Vekten $X$ er normalfordelt med forventning $\mu = 110$ og varians $\sigma^2 = 100$.

Sannsynligheten for at en tilfeldig valgt potet veier mer enn 120 $g$:

$$\begin{aligned} P(X > 120) &= 1 - P(X \le 120) = 1 - P(\frac{X-110}{10} \le \frac{120-110}{10}) \\ &= 1 - \Phi(1) = 1 - 0.8413 = \underline{0.1587}. \end{aligned}$$

Sannsynligheten for at en tilfeldig valgt potet er innenfor spesifikasjonen:

$$\begin{aligned} P(|X-110| \le 20) &= = P(X \le 130) - P(X \le 90) \\ &= P(\frac{X-110}{10} \le \frac{130-110}{10}) - P(\frac{X-110}{10} \le \frac{90-110}{10}) \\ &= \Phi(2) - \Phi(-2) = \Phi(2) - (1 - \Phi(2)) \\ &= 2\Phi(2) - 1 = 2 \cdot 0.9772 - 1 = \underline{0.9544}. \end{aligned}$$

Sannsynligheten for at en pose med 14 tilfeldig valgte poteter veier mindre enn $1500g$:

La $X_1, \ldots, X_{14}$ være vekten til de 14 potetene, og la $Z$ være vekten på posen med 14 poteter. Da er $Z = \sum_{i=1}^{14} X_i$. Siden $Z$ er en lineærkombinasjon av uavhengige og normalfordelte variabler, er $Z$ også normalfordelt, med forventning og varians lik

$$\begin{aligned} \mathrm{E}(Z) &= \mathrm{E}(\sum_{i=1}^{14} X_i) = \sum_{i=1}^{14} \mathrm{E}(X_i) = \sum_{i=1}^{14} \mu = 14\mu = 14 \cdot 110\, g = 1540\, g \\ \mathrm{Var}(Z) &= \mathrm{Var}(\sum_{i=1}^{14} X_i) \overset{\text{uavh}}{=} \sum_{i=1}^{14} \mathrm{Var}(X_i) = \sum_{i=1}^{14} \sigma^2 = 14\sigma^2 = 14 \cdot 100\, g^2 = 1400\, g^2 \end{aligned}$$



Vi får da

$$\begin{aligned} P(Z < 1500) &= P(Z \le 1500) = P(\frac{Z-1540}{\sqrt{1400}} \le \frac{1500-1540}{\sqrt{1400}}) \\ &= \Phi(-1.07) = \underline{0.1423}. \end{aligned}$$

**b)** $Y$ er antall poteter blant de $n$ i stikkprøven som *ikke* tilfredstiller spesifikasjonen.

Vi sammenligner forsøket beskrevet i oppgaven med kriteriene for en Bernoulliforsøksrekke (et binomisk forsøk):

- Vi gjør $n$ forsøk, der hvert forsøk er å undersøke om vekten på en tilfeldig valgt potet avviker mindre enn 20 $g$ fra forventet vekt eller ikke.
- Det er to mulige utfall av hvert forsøk. Enten følger poteten *ikke* vektspesifikasjonen ("suksess"), eller den oppfyller kravet ("fiasko").
- Sannsynligheten $p$ for at poteten ikke følger spesifikasjonen er $p = P(|X - 110| \le 20)$, og denne er lik i alle forsøk.
- Siden potetene i stikkprøven er trukket tilfeldig, kan vi anta at forsøkene er uavhengige.

Dermed oppfyller forsøket kravene til en Bernoulliforsøksrekke (et binomisk forsøk), og vi antar at $Y$ er binomisk fordelt med parametre $n$ og $p = P(|X - 110| \le 20)$.

Forventningsverdi for estimatoren $\hat{p} = \dfrac{Y}{n}$ for $p$:

$$\begin{aligned} \mathrm{E}(\hat{p}) &= \mathrm{E}(\frac{Y}{n}) = \frac{1}{n}\mathrm{E}(Y) \\ &= \frac{1}{n}np = p. \end{aligned}$$

Siden $\mathrm{E}(\hat{p}) = p$, er $\hat{p}$ en forventningsrett estimator for $p$.

Variansen til $\hat{p}$:

$$\begin{aligned} \mathrm{Var}(\hat{p}) &= \mathrm{Var}(\frac{Y}{n}) = \frac{1}{n^2}\mathrm{Var}(Y) \\ &= \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}. \end{aligned}$$

Vi skal finne et estimat for variansen $\sigma^2$ basert på estimatet $\hat{p}$ for $p = P(|X - 110| \le 20)$.

Vi finner først et uttrykk for sammenhengen mellom $p$ og $\sigma^2$:

$$\begin{aligned}
p &= 1 - P(|X-110| \le 20) \;=\; 1 - [P(X-110 \le 20) - P(X-110 \le -20)] \\
&= 1 - \left[P(\frac{X-110}{\sigma} \le \frac{20}{\sigma}) - P(\frac{X-110}{\sigma} \le \frac{-20}{\sigma})\right] \\
&= 1 - \left[\Phi(\frac{20}{\sigma}) - \Phi(\frac{-20}{\sigma})\right] \;=\; 1 - \left[\Phi(\frac{20}{\sigma}) - (1 - \Phi(\frac{20}{\sigma}))\right] \\
&= 2 \cdot (1 - \Phi(\frac{20}{\sigma}))
\end{aligned}$$

Dette gir at

$$\begin{aligned}
2\Phi(\frac{20}{\sigma}) &= 2 - p \\
\Phi(\frac{20}{\sigma}) &= 1 - \frac{p}{2}
\end{aligned}$$

For en gitt verdi av $p$, kan vi finne $\frac{20}{\sigma}$ og dermed $\sigma$ fra $\frac{p}{2}$-kvantilen i standardnormalfordelingen. Men siden $p$ er ukjent, erstatter vi $p$ med $\hat{p}$, og finner et estimat for $\sigma$ basert på den tilsvarende kvantilen. Estimatet for $p$ er $\hat{p} = \frac{y}{n} = \frac{2}{19}$, og vi får at

$$\Phi(\frac{20}{\sigma}) \approx 1 - \frac{\hat{p}}{2} = 1 - \frac{2/19}{2} = 0.9474.$$

Fra tabell over kumulativ fordeling i normalfordelingen, brukt baklengs, får vi

$$\frac{20}{\hat{\sigma}} = 1.62 \Rightarrow \hat{\sigma} = \frac{20}{1.62} = 12.35\ g.$$

Et estimat for variansen blir da $\widehat{\sigma^2} = (\frac{20}{1.62})^2 = \underline{152.4\ g^2}$.

**Oppgave 2 Internettkafe**

a) Vi har $P(T < 15) = F(15) = 1 - \exp(-15 \cdot 0.1) = 0.78$.

Videre er $P(T < 45 | T > 30) = \frac{P(30<T<45)}{P(T>30)} = \frac{F(45)-F(30)}{1-F(30)} = \frac{\exp(-30\lambda)-\exp(-45\lambda))}{\exp(-30\lambda)} = 0.78$.

Siden de to uttrykkene er like, sier man at fordelingen er glemsom. Sannsynligheten for å være pålogget mindre enn 15 minutter til, er ikke påvirket av hvor lenge du har vært pålogget.

b)

Derivering av den kumulative sannsynlighetsfunksjonen gir: $f(t) = \frac{dF(t)}{dt} = \lambda \exp(-\lambda t)$.



Rimelighetsfunksjonen er $L(\lambda) = \prod_{i=1}^{10} \lambda \exp(-\lambda t_i)$. Logaritmen av rimelighetsfunksjonen er $l(\lambda) = 10 \ln(\lambda) - \lambda \sum_{i=1}^{10} t_i$. Derivering gir

$$\frac{dl}{d\lambda} = \frac{10}{\lambda} - \sum_{i=1}^{10} t_i$$

Ved å sette $\frac{dl}{d\lambda} = 0$ får vi SME $\hat{\lambda} = \frac{10}{\sum_{i=1}^{10} t_i}$. Estimatet er $\frac{10}{125.1} = 0.08$.

c)

Vi har at $T = \min(T_1, T_2, T_3)$. Videre er $P(T < 15) = 1 - P(T > 15) = 1 - P(T_i > 15)^3$. Dette gir $P(T < 15) = 1 - \exp(-\lambda \cdot 3 \cdot 15)$.

Siden $P(T > t) = \exp(-3\lambda t)$, er ventetiden til første ledige maskin, $T$, eksponensialfordelt med parameter $3 \cdot \lambda$, for $t > 0$. La $U$ være den totale ventetiden til den andre kunden får en ledig maskin. La $V = U - t$ være tiden fra første maskin ble ledig til andre maskin blir ledig. Vi betinger her på den første tiden. Ved glemsomhetsegenskapen er ventetiden fra første maskin ble ledig til andre maskin blir ledig eksponensialfordelt. Her er da $V|T = t$ også eksponensialfordelt med parameter $3 \cdot \lambda$, for $v > 0$. Marginal tetthet for $U$ er

$$f(u) = \int_0^u f(u|t) f(t) dt = \int_0^u 3\lambda \exp[-3\lambda(u-t)] 3\lambda exp(-3\lambda t) dt = (3\lambda)^2 u \exp(-3\lambda u).$$

Dette er en gamma fordeling med parameter 2 og $3\lambda$. Her er forventningen $E[U] = \frac{2}{3\lambda}$.

Dette gjelder generelt utfra at ventetiden til to hendelser i en Poisson prosess er Gammafordelt. Forventningen kan dermed også ses direkte ved at $T$ og $V$ er identisk fordelt og uavhengige, og at total tid er summen av de to. Vi har $E(T) = \frac{1}{3\lambda}$ og $E(V) = \frac{1}{3\lambda}$.

Innsetting av $\lambda = \hat{\lambda} = 0.08$ gir hhv: $P(T < 15) = 1 - \exp(-0.08 \cdot 3 \cdot 15) = 0.97$ og $E[U] = \frac{2}{3 \cdot 0.08} = 8.33$.

**Oppgave 3 Kneoperasjon** Vi jobber med to uavhengige normalfordelte utvalg med samme varians.

Vi har konvensjonell kirurgi: $X_1, X_2, ..., X_n$ u.i.f. normal med $\mathrm{E}(X_i) = \mu$ og $\mathrm{Var}(X_i) = \sigma^2$, og computer-assistert kirurgi: $Y_1, Y_2, ..., Y_m$ u.i.f. normal med $\mathrm{E}(Y_j) = \eta$ og $\mathrm{Var}(Y_j) = \sigma^2$.

Observasjoner:
Konvensjonell kirurgi, $n = 25$ uavhengige operasjoner $x_1, ..., x_{25}$. $\bar{x} = \frac{1}{25} \sum_{i=1}^{25} x_i = 2.3$, $\sum_{i=1}^{25} (x_i - \bar{x})^2 = 294$.

Computer-assistert kirurgi, $m = 27$ uavhengige operasjoner $y_1, ..., y_{27}$. $\bar{y} = \frac{1}{27}\sum_{j=1}^{27} y_j = 1.5$, $\sum_{j=1}^{27}(y_j - \bar{y})^2 = 115$.

**a)** $S_p^2$ er en god estimator for varians, når vi baserer oss på to utvalg der variansen antas den samme i begge utvalgene.

$$\begin{aligned}
\bar{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
\bar{Y} &= \frac{1}{m}\sum_{j=1}^{m} Y_j \\
S_p^2 &= = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{j=1}^{m}(Y_j - \bar{Y})^2}{n+m-2}
\end{aligned}$$

Numerisk verdi: $s_p^2 = \frac{294+115}{25+27-2} = 8.18$.

Forberedelse til forventningsverdi: Siden $X_i \sim N(\mu, \sigma^2)$ så vet vi at $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}$ er kji-kvadratfordelt med $n-1$ frihetsgrader. Tilsvarende, siden $Y_j \sim N(\eta, \sigma^2)$ vet vi at $\frac{\sum_{j=1}^{m}(Y_j-\bar{Y})^2}{\sigma^2}$ er kji-kvadratfordelt med $m-1$ frihetsgrader. Siden $X_i$ og $Y_j$ er uavhengige så vil summen $V = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{j=1}^{m}(Y_j-\bar{Y})^2}{\sigma^2}$ være kji-kvadratfordelt med $n+m-2$ frihetsgrader.

Sammenhengen mellom $S_p^2$ og $V$: $S_p^2 = \frac{V\sigma^2}{n+m-2}$.

Forventningsverdi:

$$\mathrm{E}(S_p^2) = \mathrm{E}(\frac{V\sigma^2}{n+m-2}) = \frac{\sigma^2}{n+m-2}\cdot \mathrm{E}(V) = \frac{\sigma^2}{n+m-2}\cdot(n+m-2) = \sigma^2$$

Konfidensintervall: start med en kji-kvadratfordelt størrelse mellom to kvantiler i kji-kvadratfordeling.

$$\begin{aligned}
P(\chi^2_{1-\alpha/2,n+m-2} \le V \le \chi^2_{\alpha/2,n+m-2}) &= 1-\alpha \\
P(\chi^2_{1-\alpha/2,n+m-2} \le \frac{S_p^2\cdot(n+m-2)}{\sigma^2} \le \chi^2_{\alpha/2,n+m-2}) &= 1-\alpha \\
P(\frac{S_p^2\cdot(n+m-2)}{\chi^2_{\alpha/2,n+m-2}} \le \sigma^2 \le \frac{S_p^2\cdot(n+m-2)}{\chi^2_{1-\alpha/2,n+m-2}}) &= 1-\alpha
\end{aligned}$$

Numerisk intervall ($\alpha = 0.05$):
$[\frac{s_p^2\cdot(n+m-2)}{\chi^2_{\alpha/2,n+m-2}}, \frac{s_p^2\cdot(n+m-2)}{\chi^2_{1-\alpha/2,n+m-2}}] = [\frac{8.18\cdot(25+27-2)}{32.357}, \frac{8.18\cdot(25+27-2)}{71.420}] = \underline{\underline{[5.73, 12.64]}}$.



**b)** Vi ønsker å undersøke om forventet vinkelutslag mellom leggbeinet og lårbeinet er forskjellig for de to operasjonsmetodene.

Null- og alternativ hypotese:

$$\begin{aligned}
H_0 : \mu = \eta \qquad & H_1 : \mu \ne \eta \\
H_0 : \mu - \eta = 0 \qquad & H_1 : \mu - \eta \ne 0
\end{aligned}$$

De ukjente parameterene er $\mu$, $\eta$ og $\sigma^2$, og vi setter opp følgende estimatorer:

$$\begin{aligned}
\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \\
\hat{\eta} &= \frac{1}{m}\sum_{j=1}^{n} Y_j = \overline{Y} \\
S_p^2 &= \frac{1}{n+m-2}[\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{m}(Y_j - \overline{Y})^2]
\end{aligned}$$

Vi vet at under $H_0$ så er

$$T_0 = \frac{(\overline{X} - \overline{Y}) - 0}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \quad t\text{-fordelt med } (n+m-2) \text{ frihetsgrader.}$$

Vi vil forkaste $H_0$ når $|T_0| \ge k$, der konstanten $k$ finnes slik at Type-I feilen er kontrollert på nivå $\alpha$.

$$\begin{aligned}
P(|T_0| \ge k | H_0 \text{ sann}) &= \alpha \\
k &= t_{\frac{\alpha}{2},(n+m-2)}
\end{aligned}$$

der $t_{\frac{\alpha}{2},(n+m-2)}$ er $\frac{\alpha}{2}$-kvantilen i en $t$-fordeling med $n+m-2$ frihetsgrader.

Forkastningsområde: Forkast $H_0$ når $|T_0| \ge t_{\frac{\alpha}{2},(n+m-2)}$.

Når $\alpha = 0.05$ og $n = 25$, $m = 27$, er $t_{0.025,50} = 2.009$. Innsatt data fra oppgaven har vi:

$$t_0 = \frac{\overline{x} - \overline{y} - 0}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{2.3 - 1.5 - 0}{\sqrt{8.18}\cdot\sqrt{\frac{1}{25}+\frac{1}{27}}} = 1.01$$

Siden $|t_0| = 1.01 < t_{0.025,50} = 2.009$ så forkaster vi *ikke* $H_0$ på nivå $\alpha = 0.05$, og finner ikke grunn til å tro at det er forskjell i forventet vinkelutslag på de to operasjonsmetodene.

Når vi forkastet $H_0$ på nivå 0.05 så betyr det at $p$-verdien må være større enn 0.05. $P$-verdien er gitt som

$$P(|T_0| > |t_0| | H_0 \text{ sann}) = 2\cdot P(T_0 > 1.01 | \mu - \eta = 0) = 2\cdot(1 - P(T_0 \le 1.01 | \mu - \eta = 0))$$

Fra tabell 2 i oppgaven så slår vi opp på $P(T \leq t)$ med $t = 1.01$ og $\nu = 50$, og finner 0.851, som gir $p$-verdi $2 \cdot (1 - 0.841) = \underline{\underline{0.32}}$.

## Oppgave 4 Simultanfordeling

Simultanfordelingen, $f(x, y)$, til de to diskrete stokastiske variablene $X$ og $Y$ er gitt i følgende tabell:

| | y=0 | y=1 | y=2 | f(x) |
|---|---|---|---|---|
| x=-1 | $\frac{1}{6}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{3}$ |
| x=0 | $\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{1}{12}$ | $\frac{1}{3}$ |
| x=1 | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{1}{3}$ |
| f(y) | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | 1 |

Marginalfordelingen til $X$ og til $Y$ sees i tabellen.

$$
\begin{aligned}
f(x) &= \begin{cases} \frac{1}{3} & \text{for } x = -1, 0, 1 \\ 0 & \textit{ellers.} \end{cases} \\
f(y) &= \begin{cases} \frac{1}{3} & \text{for } y = 0, 1, 2 \\ 0 & \textit{ellers.} \end{cases}
\end{aligned}
$$

Forvening og varians til $X$:

$$
\begin{aligned}
\mathrm{E}(X) &= \frac{1}{3} \cdot (-1) + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 = \underline{\underline{0}} \\
\mathrm{Var}(X) &= \frac{1}{3} \cdot (-1-0)^2 + \frac{1}{3} \cdot (0-0)^2 + \frac{1}{3} \cdot (1-0)^2 = \underline{\underline{\frac{2}{3}}}
\end{aligned}
$$



Forvening og varians til $Y$:

$$
\begin{aligned}
\mathrm{E}(Y) &= \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 2 = \underline{\underline{1}} \\
\mathrm{Var}(Y) &= \frac{1}{3} \cdot (0-1)^2 + \frac{1}{3} \cdot (1-1)^2 + \frac{1}{3} \cdot (2-1)^2 = \underline{\underline{\frac{2}{3}}}
\end{aligned}
$$

Kovarians:

$$
\begin{aligned}
\mathrm{E}(X \cdot Y) &= \frac{1}{6} \cdot (-1) \cdot 0 + \frac{1}{12} \cdot (-1) \cdot 1 + \frac{1}{12} \cdot (-1) \cdot 2 \\
&+ \frac{1}{12} \cdot 0 \cdot 0 + \frac{1}{6} \cdot 0 \cdot 1 + \frac{1}{12} \cdot 0 \cdot 2 \\
&+ \frac{1}{12} \cdot 1 \cdot 0 + \frac{1}{12} \cdot 1 \cdot 1 + \frac{1}{6} \cdot 1 \cdot 2 \\
&= \frac{1}{6} \\
\mathrm{Cov}(X, Y) &= \frac{1}{6} - 0 \cdot 1 = \underline{\underline{\frac{1}{6}}}
\end{aligned}
$$

Siden kovariansen mellom $X$ og $Y$ ikke er null så kan ikke $X$ og $Y$ være uavhengige. Vi ser også at simultanfordelingen til $X$ og $Y$ ikke er lik produktet av de to marginalfordelingene, noe som ville vært tilfellet hvis $X$ og $Y$ hadde vært uavhengige.