E6 opgavesæt 4

Jogvan M. Poulsen

18. december 2000

Opgave 1 (sommer 93, opgave 4)

For at undersøge virkningen af tre gødningstyper dyrkedes tomatplanter i 30 lige store potter. I hver potte var nedsat 5 planter. 10 af potterne fik gødning A, 10 gødning B og 10 gødning C. Gødnngen blev tilført i lige store mængder og dyrkningen foregik i øvrigt under ens betingelser. Efter 4 uger høstedes planterne og tørvægten af de grønne dele fra hver potte bestemtes. Vægtene er givet i nedenstående tabel, hvor observationerne kan antages at være normalfordelte med ens varianser.

Tørvægt for tomatplanter ved brug at tre gødningstyper

	A	B	C
	90	117	126
	120	98	141
	109	132	93
	98	114	123
	124	123	89
	74	134	112
	118	136	122
	97	141	125
	98	144	83
	114	151	119
n	10	10	10
$\overline{x}$	104,2	129,0	113,3
s^2	243,73	255,78	354,46

a Undersøg, om de tre gødningstyper giver samme udbytte Vi antager at obsevationerne $x_{ij}, i = 1,2,3; j = 1 \cdots 10$ er værdier af indbyrdes uafhænige stokastiske variable X_{ij} (Her svarer $I = 1,2,3$ til hhv. A, B, C i skemaet). Vi skal undersøge middelværdierne og varianserne for de tre gødningstyper, som vi kalder hhv. μ_1, μ_2, μ_3 og σ^2. Vores model bliver således:

$$(\mathbf{R}^{30}, (N_{(\mu_1,\mu_2,\mu_3,\sigma^2)})_{(\mu_1,\mu_2,\mu_3,\sigma^2)\in\mathbf{R}^3\times[0,\infty]})$$

hvor tætheden $N_{(\mu_1,\mu_2,\mu_3,\sigma^2)}$ er givet ved

$$\Phi_{(\mu_1,\mu_2,\mu_3,\sigma^2)}(x) = \prod_{r=1}^{3}\prod_{s=1}^{10} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{\sigma^2}(x_{rs}-\mu_r)^2}$$

Likelihoodfunktionen bliver

$$\begin{aligned} L \quad : \quad & \mathbf{R}^{10} \times \mathbf{R}^3 \times]0,\infty[\to [0,\infty[\\ L(x,(\mu_1,\mu_2,\mu_3,\sigma^2)) = & \frac{1}{(\sqrt{2\pi\sigma^2})^{30}} e^{-\frac{1}{\sigma^2}\sum_{r=1}^{3}\sum_{s=1}^{10}(x_{rs}-\mu_r)^2} \end{aligned}$$

Fra sætning 7.3.1IH får vi, at maksimaliseringsestimatorerne for μ_1, μ_2, μ_3 er givet ved

$$(\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3) = (\overline{x}_1, \overline{x}_2, \overline{x}_3) = (104,2; 129,0; 113,3)$$

Vores nullhypotese bliver her:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu \in \mathbf{R}$$

Fra sætning 7.5.1IH er givet et test-sandsynlighed ved

$$\epsilon_0(x) = P\left(V_0 \geq \frac{s_1^2}{s_0^2}\right)$$

hvor V_0 er F-fordelt med frihedsgrader $(k-1, n-k) = (2, 27)$. Størrelserne s_1^2 og s_0^2 er givet ved

$$\begin{aligned} s_0^2 &= \frac{1}{n-k}\sum_{r=1}^{k}(n_r - 1)s_r^2 = \frac{9}{27}(243,73 + 255,78 + 354,46) = 284,66 \\ s_1^2 &= \frac{1}{k-1}\sum_{r=1}^{k} n_r(\overline{x}_r - \overline{x})s_r^2 = \frac{10}{2}(127,69 + 182,25 + 4,84) = 1573,9 \end{aligned}$$

hvor $\overline{x} = 115,5$. Nu kan vi indsætte i testsandsynligheden

$$\epsilon_0(x) = P\left(V_0 \geq \frac{1573,9}{284,66}\right) = P(V_0 \geq 6,14)$$

Når vi ser i F-fordelingstabellen i TT, ser vi, at ved $F(2,27)$ er sandsynligheden at få $V_0 > 4,26$ er 5%, hvilket betyder at sandsynligheden for at få $V_0 > 6,14$ er minder end 5%. Derfor kan vi konkludere, at vi ikke kan godtage hypotesen. Gødningsudbyttet er altså ikke ens for de tre gødningstyper.

b Under forudsætning af, at gødning C var en blanding af A og B i forholdet 1:1, undersøg da ved en T-test, om $\mu_C = {}^{(\mu_A+\mu_B)}/_2$, hvor μ_A, μ_B og μ_C er niveauet for tørvægten ved brug af hhv. gødning A, B og C. Vi skal altså se om

$$H_1 : \mu_c = \frac{\mu_a + \mu_b}{2}$$

hvor $\hat{\mu}_c = 113,3$ og ${}^{(\hat{\mu}_A+\hat{\mu}_B)}/_2 = 116,6$ T-testet er givet ved (ifølge s.157IH)

$$\epsilon_1(x) = 2P\left(T \geq \frac{|\frac{\overline{x}_1+\overline{x}_2}{2} - \overline{x}_3|}{s_0\sqrt{\frac{1}{n_1+n_2} + \frac{1}{n_3}}}\right)$$

hvor $k = 2$ og $n_1 = n_2 = n_3 = 10$. Vi kan nu indsætte i testes

$$\epsilon_1(x) = 2P\left(T \geq \frac{|116,6 - 113,3|}{16,87\sqrt{\frac{1}{20} + \frac{1}{10}}}\right) = 2P(T \geq 0,51)$$

Her kan man sige udfra T-tabellen, at $2P(T \geq 0,51)$ med $b - k = 27$ frihedsgrader, giver og en værdi der er meget større end 5%, hvilket må føre til konklusionen om, at vi ikke kan afvise at hypotesen var rigtig.

c Find maksimallingsestimatorerne for middelværdierne i (b). Det skal med andre ord forstås sådan, at vi skal finde μ_1 og μ_2 sådan, at Likelihoodfunktionen bliver maksimal. Vi opstiller først L-funktionen.

$$\begin{aligned} L \quad &: \quad \mathbf{R}^{10} \times \mathbf{R}^3 \times]0,\infty[\to [0,\infty] \\ &L(x, (\mu_1, \mu_2, (\frac{\mu_1+\mu_2}{2}), \sigma^2)) = \frac{1}{(\sqrt{2\pi\sigma^2})^{30}} e^{-\frac{1}{\sigma^2}\sum_{r=1}^{3}\sum_{s=1}^{10}(x_{rs}-\mu_r)^2} \end{aligned}$$

For at den bliver maksimal, skal vi minimere $\sum_{r=1}^{3}\sum_{s=1}^{10}(x_{rs}-\mu_r)^2$. Udfra lemma 6.2.1IH får vi at

$$\sum_{r=1}^{3}(\sum_{s=1}^{10}(x_{rs}-\mu_r)^2)=\sum_{r=1}^{3}(\sum_{s=1}^{10}(x_{rs}-\overline{x}_r)^2+n(\overline{x}_r-\mu_r)^2)$$

Vi har nu reduceret problemet til at vi kun skal minimere $(\overline{x}_r-\mu_r)^2$, og det gør vi så

$$\sum_{r=1}^{3}(\overline{x}_r-\mu_r)^2=(\overline{x}_1-\mu_1)^2+(\overline{x}_2-\mu_2)^2+(\overline{x}_3-\frac{\mu_1+\mu_2}{2})^2$$

Vi kan definere en funktion Y ved

$$Y(\mu_1,\mu_2)=(\overline{x}_1-\mu_1)^2+(\overline{x}_2-\mu_2)^2+(\overline{x}_3-\frac{\mu_1+\mu_2}{2})^2$$

For at finde de værdier til μ_1 og μ_2 der minimerer funktionen, kan vi udregne ${}^{\partial Y}/_{\partial\mu_1}=0$ og ${}^{\partial Y}/_{\partial\mu_2}=0$ og derved finde værdierne. Det kræver en masse regne regne, så det vil jeg spare jer for, ved at springe 'overflødige' mellemregninger over. Vi fortsætter

$$\frac{\partial Y}{\partial\mu_2}=\frac{5}{2}\mu_1-2\overline{x}_1+\frac{1}{2}\mu_2-\overline{x}_3=0 \quad \text{og} \quad \frac{\partial Y}{\partial\mu_1}=\frac{5}{2}\mu_2-2\overline{x}_2+\frac{1}{2}\mu_1-\overline{x}_3=0$$

$$\Rightarrow \mu_1=\frac{4}{5}\overline{x}_1+\frac{2}{5}\overline{x}_3-\frac{1}{5}\mu_2 \quad \text{og} \quad \mu_1=\frac{4}{5}\overline{x}_1+\frac{2}{5}\overline{x}_3-\frac{1}{5}\mu_2$$

Ved at isolere og indsætte værdierne fra skemaet, får vi at

$$\hat{\mu}_1=103,10 \qquad \hat{\mu}_2=127,9 \qquad \hat{\mu}_3=\frac{\mu_1+\mu_2}{2}=115,5$$

Disse værdier maksimerer Likelihoodfunktionen

Opgave 2 (vinter 93-94, opgave 2)

Lad $f:[0,1]\rightarrow\mathbf{R}$ være en kontinuert, ikke-negativ funktion med $\int_0^1 f(x)dx=1$. Definer $p:\mathbf{R}^2\rightarrow\mathbf{R}$ ved

$$p(x,y)=\begin{cases}1 \text{ for } 0\leq x\leq 1 \text{ og } o\leq y\leq f(x)\\ \\ 0 \text{ ellers}\end{cases}$$

a Gør rede for, at p er en sandsynlighedstæthed. Siden $p(x,y)\geq 0$ for $x,y\in\mathbf{R}$ og at funktionen $p:\mathbf{R}^2\rightarrow\mathbf{R}$ opfylder

$$\int p(x,y)dxdy=1$$

kan $p(x,y)$ antage en sandsynlighedstæthed. Eller sagt på en anden måde, at når kun er defineret på indetvallet $0,1$, hvor $x=1$ og hvor 'arealet

b Lad (X,Y) være en todimentional stokastisk variabel med tæthed p. Vis da, at den marginale fordeling af X er givet ved tætheden f. Den marginale fordeling af x får ved at integrere y ud

$$p(x) = \int_{-\infty}^{\infty} p(x,y)dy = \int_0^{f(x)} 1dy = [y]_0^{f(x)} = f(x)$$

c Vis, at

$$\mathrm{E}(Y) = \frac{1}{2}\int_0^1 f(x)^2 dx$$

Middelværdien for y finder vi ved at gange y på $p(x,y)$ og integrere.

$$\mathrm{E}(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)ydydx = \int_0^1\int_0^{f(x)} 1\cdot ydydx$$

Da $p(x,y) = 1$ i intervallet. Vi fortsætter

$$\mathrm{E}(Y) = \int_0^1 [{}^1\!/\!_2 y^2]_0^{f(x)} dx = \frac{1}{2}\int_0^1 f(x)^2 dx$$

Og det var det vi skulle vise

Opgave 3 (IH 6.13)

Lad $X_{11}, X_{12}, X_{21}, X_{22}$ være stokastisk uafhængige, normalfordelte variable med samme ukendte varians σ^2. Antag at

$$\begin{aligned} \mathrm{E}(X_{11}) &= \mathrm{E}(X_{12}) = \alpha + \beta \\ \mathrm{E}(X_{21}) &= \mathrm{E}(X_{22}) = \alpha - \beta \end{aligned}$$

a Find maksimalliseringsestimatorerne $\hat{\alpha}$ og $\hat{\beta}$. Vi begynder med at kalde middelværdierne for hhv $\alpha + \beta = \mu_1$ og $\alpha - \beta = \mu_2$. Vi har fire variable, og vores model bliver ifølge sætning 6.2.1IH

$$\text{Model}: (\mathbf{R}^4, (N_{(\mu_1,\mu_2,\sigma^2)})_{(\mu_1,\mu_2,\sigma^2)\in\mathbf{R}^2\times]0,\infty[})$$

hvor tætheden for N er givet ved

$$\Phi_(\mu_1,\mu_2,\sigma^2)(x) = \prod_{r=1}^{2}\prod_{s=1}^{2}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{\sigma^2}(x_{r,s}-\mu_r)^2}$$

Likelihoodfunktionen bliver

$$\begin{aligned} L \quad : \quad & \mathbf{R}^4 \times \mathbf{R}^2 \times]0,\infty[\to [0,\infty[\\ & L(x,(\mu_1,\mu_2,\sigma^2)) = \frac{1}{(\sqrt{2\pi\sigma^2})^4}e^{-\frac{1}{\sigma^2}\sum_{r=1}^2\sum_{s=1}^2(x_{rs}-\mu_r)^2} \end{aligned}$$

For at maksimere L, skal vi minimere $\sum_{r=1}^2\sum_{s=1}^2(x_{rs}-\mu_r)^2$ (tilsvarende som i opgave 1c). Lemma 6.2.1IH giver os igen, at

$$\sum_{r=1}^{2}(\sum_{s=1}^{2}(x_{rs}-\mu_r)^2) = \sum_{r=1}^{2}(\sum_{s=1}^{2}(x_{rs}-\overline{x}_r)^2 + n(\overline{x}_r - \mu_r)^2)$$

Vi definerer nu funktionen Y som vi skal minimere

$$\begin{aligned} Y(\mu_1, \mu_2) &= (\overline{x}_1 - \mu_1)^2 + (\overline{x}_2 - \mu_2)^2 \\ \Rightarrow Y(\alpha, \beta) &= (\overline{x}_1 - \alpha - \beta)^2 + (\overline{x}_2 - \alpha + \beta)^2 \end{aligned}$$

Nu kan vi finde $\hat{\alpha}$ ved at differentiere med hensyn til α og sætte lig nul

$$\begin{aligned} \frac{\partial Y}{\partial \alpha} &= -2(\overline{x}_1 - \alpha - \beta) - 2(\overline{x}_2 - \alpha + \beta) \\ \frac{\partial Y}{\partial \beta} &= -2(\overline{x}_1 - \alpha - \beta) + 2(\overline{x}_2 - \alpha + \beta) \end{aligned}$$

Ved at isolere får vi at

$$\begin{aligned} &4\alpha = 2\overline{x}_1 + 2\overline{x}_2 \Rightarrow \alpha = \frac{1}{2}\overline{x}_1 + \frac{1}{2}\overline{x}_2 \Rightarrow \hat{\alpha} = \frac{1}{4}(x_{11} + x_{12} + x_{21} + x_{22}) \\ &4\beta = 2\overline{x}_1 - 2\overline{x}_2 \Rightarrow \beta = \frac{1}{2}\overline{x}_1 - \frac{1}{2}\overline{x}_2 \Rightarrow \hat{\alpha} = \frac{1}{4}(x_{11} + x_{12} - x_{21} - x_{22}) \end{aligned}$$

b Find fordelingen af $\hat{\alpha}$ og $\hat{\beta}$. Da X_{rs} er normalfordelt med parametrene (μ, σ^2), og at summen af normalfordelinger er normalfordelt, får vi fra sætning 6.4.1IH at vide, at $\hat{\mu}$ er normalfordelt med middelværdi μ og varians $1/n\sigma^2$ - altså normalfordelt med parametrene $(\mu, 1/n\sigma^2)$. Det betyder at $\hat{\mu}_1 + \hat{\mu}_2 = 2\hat{\alpha}$ er normalfordelte. med parametrene $((\mu_1 + \mu_2), (1/2\sigma^2 + 1/2\sigma^2))$ som også kan skrives som $(2\alpha, \sigma^2)$. På side 130IH får vi at vide, at hvis X er normalfordelt (μ, σ^2) og $a \in \mathbf{R}$ og $b \neq 0$, er konstanter, så er $a + bX$ normalfordelt $(a + b\mu, b^2\sigma^2)$. Det bruger vi nu. Vi kan omskrive til

$$\frac{1}{2}(2\alpha, \sigma^2) = (\frac{1}{2} \cdot 2\alpha, \left(\frac{1}{2}\right)^2 \sigma^2) = (\alpha, \frac{\sigma^2}{4})$$

Resultater bliver, at $\hat{\alpha}$ er normalfordelt $(\alpha, \frac{1}{4}\sigma^2)$. Tilsvarende bliver det for $\hat{\beta}$. Da $\hat{\mu}_2$ er normalfordelt $(\mu_2, 1/2\sigma^2)$ er $-\hat{\mu}_2$ normalfordelt $(-\mu_2, 1/2(-1)^2\sigma^2)$. Summen $\hat{\mu}_1 + (-\hat{\mu}_2)$ er normalfordelt $(\mu_1 - \mu_2, 1/2\sigma^2 + 1/2\sigma^2)$ som også er $(2\beta, \sigma^2)$. $\hat{\beta}$ er dermed normalfordelt $(\beta, \frac{1}{4}\sigma^2)$.

c Vis, at $\hat{\alpha}$ og $\hat{\beta}$ er ukorrelerede Korrelationen mellem $\hat{\alpha}$ og $\hat{\beta}$ er i følge s.66TT givet ved:

$$\mathrm{corr}(\hat{\alpha}, \hat{\beta}) = \frac{\mathrm{cov}(\hat{\alpha}, \hat{\beta})}{\sqrt{\mathrm{var}(\hat{\alpha})\mathrm{var}(\hat{\beta})}}$$

Så hvis cov $= 0$ er de ukorrelerede. Covariansen er givet ved

$$2\mathrm{cov}(\hat{\alpha}, \hat{\beta}) = \mathrm{var}(\hat{\alpha} + \hat{\beta}) - \mathrm{var}(\hat{\alpha}) - \mathrm{var}(\hat{\beta})$$

Nu kan vi indsætte værdierne for hhv. $\hat{\alpha}$ og $\hat{\beta}$, hvor vi får

$$\begin{aligned} 2\mathrm{cov}(\hat{\alpha}, \hat{\beta}) &= \mathrm{var}(\hat{\alpha} + \hat{\beta}) - \mathrm{var}(\hat{\alpha}) - \mathrm{var}(\hat{\beta}) \\ &= \mathrm{var}(\frac{1}{2}(x_{11} + x_{12})) - \mathrm{var}(\frac{1}{4}(x_{11} + x_{12} + x_{21} + x_{22})) \\ &\quad -\mathrm{var}(\frac{1}{4}(x_{11} + x_{12} - x_{21} - x_{22})) \end{aligned}$$

Men da X_{rs} er stokastisk uafhængige med samme varisns σ^2, gælder der, at variansen af en sum er lig summen af varianserne. Vi kan nu indsætte

$$\begin{aligned}
2\mathrm{cov}(\hat{\alpha},\hat{\beta}) &= (\frac{\sigma^2}{4}+\frac{\sigma^2}{4}) - (\frac{\sigma^2}{16}+\frac{\sigma^2}{16}+\frac{\sigma^2}{16}+\frac{\sigma^2}{16}) \\
&\quad -(\frac{\sigma^2}{16}+\frac{\sigma^2}{16}+\frac{\sigma^2}{16}+\frac{\sigma^2}{16}) \\
&= \frac{\sigma^2}{2}-\frac{\sigma^2}{4}-\frac{\sigma^2}{4} \\
&= 0
\end{aligned}$$

Dermed er $\hat{\alpha}$ og $\hat{\beta}$ er ukorrelerede