Eksamen i Stat 100 - Statistikk.
Tirsdag 30. mai 2006 9.00 - 12.30.
Forslag til løsning.

Oppgave 1 Vannverk

a) F_1 og F_2 er ikke uavhengige, fordi

$$P(F_1) = 0{,}8 \neq 0{,}9 = P(F_1 \mid F_2)$$

F_1 og F_2 er ikke disjunkte, fordi

$$P(F_1 \cap F_2) = P(F_1 \mid F_2) \cdot P(F_2) = 0{,}9 \cdot 0{,}8 = 0.72 \neq 0.$$

b) $P(F_1 \cup F_2) = P(F_1) + P(F_2) - P(F_1 \cap F_2) =$

$$0{,}8 + 0{,}8 - 0.72 = \underline{\underline{0.88}}$$

c) "i det lange løp i 997 av 1000 dager" betyr mer presist at sannsynligheten er 0,997 (=997/1000).

Vannverkets krav er: $P(F_1 \cup F_2 \cup F_3) = 0{,}997.$

Vi har: $P(F_1 \cup F_2 \cup F_3) = P(F_1 \cup F_2) + P(F_3)$
$- P[(F_1 \cup F_2) \cap F_3] = \langle$ fordi F_3 er uavhengig av $F_1 \cup F_2 \rangle =$
$P(F_1 \cup F_2) + P(F_3) - P(F_3) \cdot P(F_1 \cup F_2) =$
$P(F_1 \cup F_2) + P(F_3)[1 - P(F_1 \cup F_2)]$.

Med andre ord:

$$P(F_3) = \frac{P(F_1 \cup F_2 \cup F_3) - P(F_1 \cup F_2)}{1 - P(F_1 \cup F_2)}$$

For å tilfredstille vannverkets krav må vi da ha:

$$\underline{\underline{P(F_3)}} = \frac{0{,}997 - 0{,}88}{1 - 0{,}88} = \underline{\underline{0.975}}$$

Oppgave 2 Maratonløp

a) $P(X < 165) = P(X \leq 165) = P\left(\frac{X-230}{\sqrt{625}} \leq \frac{165-230}{\sqrt{625}}\right) =$

$P\left(Z \leq \frac{-65}{25}\right) = P(Z \leq -2{,}6) = G(-2.6) =$

$\underline{\underline{0.0047}}$, der $Z \stackrel{\text{def}}{=} \frac{X-230}{\sqrt{625}} \sim N(0,1)$

$P(X > 270) = 1 - P(X \leq 270) =$

$1 - P\left(\frac{X-230}{\sqrt{625}} \leq \frac{270-230}{\sqrt{625}}\right) = 1 - P(Z \leq 1.6) =$

$1 - G(1.6) = 1 - 0{,}9452 = \underline{\underline{0{,}0548}}$

b) Samlet tid for laget blir $S = X_1 + X_2 + \cdots + X_5$.

Vi har: $E(S) = E(X_1) + E(X_2) + \cdots + E(X_5) = 5 \cdot \mu =$
$5 \cdot 230 = 1150$.

$Var(S) = Var(X_1) + Var(X_2) + \cdots + Var(X_5) =$
$5 \cdot \sigma^2 = 5 \cdot 625 = 3125$

Herav: $S \sim N(1150, 3125)$

$$\underline{\underline{P(S > 1100)}} = 1 - P(S \leq 1100) =$$

$$1 - P\left(\frac{S-1150}{\sqrt{3125}} \leq \frac{1100-1150}{\sqrt{3125}}\right) = 1 - G(-0.8944)$$

$$= 1 - 0.1856 = \underline{\underline{0.8144}}$$

$\underline{\underline{P(\text{alle fem bruker mindre enn 230 minutter})}} =$

$$P(X_1 < 230 \text{ og } X_2 < 230 \text{ og } \cdots\cdots \text{ og } X_5 < 230) \underset{\text{uavhengige}}{=}$$

$$P(X_1 < 230) \cdot P(X_2 < 230) \cdots\cdots P(X_5 < 230) =$$

$$\left[G\left(\frac{230-230}{\sqrt{625}}\right)\right]^5 = \left[G(0)\right]^5 = \left(\tfrac{1}{2}\right)^5 = \underline{\underline{0.03125}}$$

$\underline{\underline{P(\text{minst en av de fem bruker mer enn 300 minutter})}} =$

$$1 - P(\text{alle fem bruker mindre enn 300 minutter}) =$$

$$1 - P(X_1 < 300 \text{ og } X_2 < 300 \text{ og } \cdots\cdots \text{ og } X_5 < 300) \underset{\text{uavhengige}}{=}$$

$$1 - P(X_1 < 300) \cdot P(X_2 < 300) \cdots\cdots P(X_5 < 300) =$$

$$1 - \left[G\left(\frac{300-230}{\sqrt{625}}\right)\right]^5 = 1 - \left[G(2.8)\right]^5 = 1 - (0.9974)^5 =$$

$$1 - 0.9871 = \underline{\underline{0.0129}}$$

c) $H_0 : \mu = 230$ mot $H_1 : \mu < 230$

Vi bruker en t-test som betyr å forkaste H_0 hvis

$$T = \frac{\bar{X} - 230}{\sqrt{\frac{S^2}{n}}} < -t_{\alpha}^{(n-1)}$$

Vi har:

$$T = \frac{210 - 230}{\sqrt{\frac{2550}{25}}} = \frac{-20}{10.1} = -1.98$$

$$-t_{\alpha}^{(n-1)} = -t_{0.05}^{(24)} = -1.711$$

Fordi $-1.98 < -1.711$ må H_0 forkastes.

Konklusjon: Vi påstår at forventet sluttid er mindre enn 230 minutter og maratonløperen synes å ha rett.

d) Et 99 % konfidensintervall for σ^2 er gitt ved

$$\left[\frac{(n-1)S^2}{\chi^{2\,(n-1)}_{\frac{0.01}{2}}} , \frac{(n-1)S^2}{\chi^{2\,(n-1)}_{1-\frac{0.01}{2}}} \right] =$$

$$\left[\frac{24 \cdot 2550}{\chi^{2\,(24)}_{0.005}} , \frac{24 \cdot 2550}{\chi^{2\,(24)}_{0.995}} \right] =$$

$$\left[\frac{61200}{45.56} , \frac{61200}{9.89} \right] =$$

$$\underline{\underline{\left[1343{,}3 \; , \; 6188{,}1 \right]}}$$

Fordi $625 \notin [1343{,}3 \, , \, 6188{,}1]$ må $\underline{\underline{H_0 \text{ forkastes}}}$

Oppgave 3 Akselerasjon på biler

a) Regresjonsmodell:

$$Y_i = \alpha + \beta x_i + e_i \quad , i=1,2,\dots,n \ (n=14 \text{ her})$$

med de vanlige antagelsene, blant annet at
$e_1, e_2, \dots, e_n$ uavhengige med $e_i \sim N(0,\sigma^2)$

α, β og σ^2 er parametre med ukjente verdier.

Estimater for parameterne (fra Minitab utskrift):

$$\hat{\alpha} = \text{"Constant"} = 11,8346$$

$$\hat{\beta} = \text{" Hk"} = -0.015285$$

$$\hat{\sigma}^2 = \text{"}MS_E\text{"} = 0.428$$

$$R^2 = \frac{SS_R}{SS_T} = \frac{12,656}{17,794} = 0.711$$

Det betyr at ca. 71,1% av den observerte variasjonen i akselerasjonstidene forklares ved denne modellen.

b) Et 95% prediksjonsintervall er gitt ved:

$$\hat{\alpha} + \hat{\beta} \cdot x_0 \pm t_{0.025}^{(n-2)} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \left(\frac{x_0 - \overline{x}}{\frac{s}{SE(\hat{\beta})}} \right)^2}$$

der $s = \sqrt{MS_E}$ og $SE(\hat{\beta}) = \sqrt{\hat{Var}(\hat{\beta})}$.

Vi har $\overline{x} = \frac{1}{14}(160 + 155 + \cdots + 230) = 182.6$

Fra Minitab utskriften har vi de fleste andre nødvendige størrelsene. Da har vi følgende prediksjonsintervall:

$$11.8346 - 0.015285 \cdot 190 \pm 2.179 \cdot 0.654 \cdot \sqrt{1 + \frac{1}{14} + \left(\frac{190 - 182.6}{\frac{0.654}{0.002812}} \right)^2} =$$

$$8.93 \pm 2.179 \cdot 0.677 = \underline{\underline{[7.45, 10.41]}}$$

c) Den "nye" tilpassede regresjonsmodellen er

$$Y = \alpha^* + \beta^* \cdot Z$$

der

$$Z = \text{"effekt i kW"} = 0.7355 \cdot \text{"effekt i Hk"} = 0.7355 \cdot x$$

Dette gir $\bar{Z} = 0.7355 \cdot \bar{x}$.

For enkelhets skyld definerer vi $k = 0.7355$.

Da har vi:

$$\beta^* = \frac{\sum_{i=1}^{n}(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Z_i - \bar{Z})^2} = \frac{\sum_{i=1}^{n}(kx_i - k\bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(kx_i - k\bar{x})^2} =$$

$$\frac{k\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{k^2 \sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{1}{k} \cdot \hat{\beta}$$

$$\alpha^* = \bar{Y} - \beta^* \bar{Z} = \bar{Y} - \frac{1}{k}\hat{\beta} \cdot k\bar{x} = \bar{Y} - \hat{\beta}\bar{x} = \hat{\alpha}$$

Den "nye" tilpassede regresjonsmodellen blir da:

$$\underline{\underline{Y}} = \alpha^* + \beta^* Z = \hat{\alpha} + \frac{1}{k}\hat{\beta}\, Z$$

$$11.8346 + \frac{1}{0.7355}(-0.015285)\, Z =$$

$$\underline{\underline{11.8346 - 0.02078\, Z}}$$