Løsningsforslag eksamen i Stat100 29/5 - 07.

Oppgave 1

X_i = c-vitamininnholdet i appelsinjuice kartong nr. i

Y_j = c-vitamininnholdet i grapefruktjuice kartong nr. j

Antar: $X_i \sim N(\mu_1, \sigma)$, $Y_j \sim N(\mu_2, \sigma)$
alle obs. antas å være uavhengige.

Tester $H_0: \mu_1 \geq \mu_2$ mot $H_1: \mu_1 < \mu_2$

$$t = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -1.95$$

$$t_{0.05, 22} = 1.717$$

$-1.95 < -1.717$ og vi forkaster H_0

Vi påstår at c-vitamininnholdet i appelsinjuicen er mindre enn i grapefruktjuicen.

Oppgave 2

a) Y_i = vannforbruk pr person i hushold. i

X_i = antall personer i hushold nr. i

Modell: $Y_i = \alpha + \beta x_i + e_i$ der feilleddene $e_1 \ldots e_{10}$ antas uavhengige $N(0, \sigma)$.

$\hat{\alpha} = 136.7$

$\hat{\beta} = -12.4$

$\hat{\sigma} = s = 10.41$

Tolkning av $\hat{\beta}$: Hvis antall personer i husholdningen øker med 1, estimeres vannforbruket pr person å minke med 12.4 l/døgn.

b) $R^2 = 78\%$ som er høy, og som tyder på at modellen er god. Men plot av residualer mot fits viser at det er systematikk i feilleddene. Derfor er ikke modellen så god likevel.

c) $x = 3 \Rightarrow \hat{Y} = 136.7 - 12.4 \cdot 3 = 99.5$ $\quad \bar{x} = 3$

Nedre grense $= \hat{Y} - t_{0.025,8} S\sqrt{1 + \frac{1}{n} + \frac{(3-\bar{x})^2}{\Sigma(x_i - \bar{x})^2}}$

$= 99.5 - 2.306 \cdot 10.4121\sqrt{1 + \frac{1}{10}} = 99.5 - 25.18 = 74.32$

Øvre grense $= \hat{Y} + t_{0.025,8} S\sqrt{1 + \frac{1}{n} + \frac{(3-\bar{x})^2}{\Sigma(x_i - \bar{x})^2}}$

$= 99.5 + 25.18 = 124.68$

[74.32, 124.68] er et 95% prediksjons-intervall for vannforbruket pr. person i en husholdning på 3 personer.
Tolking: Det er 95% sannsynlig at vannforbruket pr person vil ligge i intervallet hvis husholdningen har 3 personer.
Hvis $x = 12$: $\hat{Y} = 136.7 - 12.4 \cdot 12 = -12.1$
Estimert vannforbruk ~~pr~~ person blir negativt. Modellen er tilpasset for husholdninger med 1-5 personer, og er ikke brukbar for husholdninger med mange flere personer.

Oppgave 3

X = fart til bil som kjører på veien

$X \sim N(51, 10)$

a) $P(X > 50) = 1 - P(X \leq 50) = 1 - G\left(\frac{50-51}{10}\right)$

$= 1 - G(-0.1) = 1 - 0.4602 = \underline{\underline{0.5398}}$

La Y = antall biler som kjører fortere enn 50 km/t. blant 40 biler.

Y er bin, $n = 40$, $p = 0.5398$

$P(Y \geq 20) = 1 - P(Y \leq 19) \approx 1 - G\left(\frac{19.5 - 40 \cdot 0.5398}{\sqrt{40 \cdot 0.5398 \cdot 0.4602}}\right)$

$= 1 - G(-0.66) = 1 - 0.2546 = \underline{\underline{0.7454}}$

(Med TI 84: $P(Y \geq 20) = 1 - \text{binomcdf}(40, 0.5398, 19)$
$= \underline{\underline{0.7471}}$)

b) Tester $H_0: \sigma = 10$ mot $H_1: \sigma \neq 10$

$\bar{x} = 58$, $S = 4.359$

Finner et 95% konfidensintervall for σ:

Nedre grense $= \sqrt{\frac{(n-1)\, S^2}{\chi_{0.025,8}}} = \sqrt{\frac{8}{17.53}} \cdot 4.359 = 2.94$

Øvre grense $= \sqrt{\frac{(n-1)\, S^2}{\chi_{0.975,8}}} = \sqrt{\frac{8}{2.18}} \cdot 4.359 = 8.35$

[2.94, 8.35] er et 95% konfidensintervall for σ. Vi forkaster H_0 fordi 10 ligger utenfor intervallet. Påstår at standardavviket har forandret seg.

c) $H_0: \mu \geq 60$ mot $H_1: \mu < 60 = \mu_0$

$$t = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{58 - 60}{\frac{4.359}{\sqrt{9}}} = -1.38$$

$t_{0.05,8} = 1.86$

Vi kan ikke forkaste H_0 fordi $-1.86 < -1.38$.

Kan ikke si at forventet fart er mindre enn fartsgrensen.

Oppgave 4

La D_1 = drektig på første forsøk

D_2 = — " — andre — " —

D_3 = — " — tredje — " —

$P(D_1) = 0.7$ $P(D_2 / \bar{D_1}) = 0.6$

$P(D_3 / \bar{D_1} \cap \bar{D_2}) = 0.55$.

a) $P(X=1) = P(D_1) = \underline{0.7}$

$P(X=2) = P(\bar{D_1} \cap D_2) = P(D_2 / \bar{D_1}) P(\bar{D_1})$

$= 0.6 \cdot (1-0.7) = 0.6 \cdot 0.3 = \underline{0.18}$

$P(X=3) = 1 - P(Y=1) - P(Y=2) = 1 - 0.7 - 0.18 = \underline{\underline{0.12}}$

b) $\mu = E(X) = \sum_{x=1}^{3} x P(X=x) = 1 \cdot 0.7 + 2 \cdot 0.18 + 3 \cdot 0.12$

$= \underline{\underline{1.42}}$

Vi forventer at en kalv på denne gården trenger mellom 1 og 2 forsøk før å bli drektig.

$\sigma^2 = \sum_{x=1}^{3} x^2 P(X=x) - \mu^2 = 1^2 \cdot 0.7 + 2^2 \cdot 0.18 + 3^2 \cdot 0.12$

$- 1.42^2 = 0.4836$, $\sigma = 0.695 \approx \underline{\underline{0.70}}$

Standardavviket til antall forsøk for en kalv er: 0,70

c) $P(\text{slaktet}) = P(\bar{D}_1 \cap \bar{D}_2 \cap \bar{D}_3) =$

$P(\bar{D}_3 / \bar{D}_1 \cap \bar{D}_2)\, P(\bar{D}_1 \cap \bar{D}_2) = (1-0{,}55) \cdot P(\bar{D}_2 | \bar{D}_1) P(\bar{D}_1)$

$= 0.45 \cdot (1-0.6) \cdot (1-0.7) = \underline{\underline{0{,}054}}$

Y = antall kalver som må slaktes pga fertilitetsproblemer blant 37 kunnkalver.

Y er bin, $n = 37$, $p = 0{,}054$

$E(Y) = np = 37 \cdot 0{,}054 = 1{,}998 \approx 2$

Han må regne med å slakte 2 kunnkalver pga fertilitetsproblemer.