TMA4245 Statistikk
Eksamen 10. august 2013

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsskisse

Oppgave 1

a)

$$P(X > 17.5) = P(\frac{(X - 16.14}{0.63} > \frac{17.5 - 16.14}{0.63}) = P(Z > 2.16)$$
$$= 1 - P(Z < 2.16) = 1 - 0.9846 = 0.0154.$$

Let X_i be sample $i = 1, \ldots, 5$, and $p =$ probability that at least one sample larger than 17.5. The complement of this event is that all samples are below the limit.

$$p = 1 - P(X_1 < 17.5 \cap \ldots X_5 < 17.5) = 1 - \prod_{i=1}^{5} P(X_i < 17.5) = 1 - 0.986^5 = 0.075$$

The product assumption relies on independent samples.

b) $H_0 : \sigma^2 \geq 0.63^2$, $H_1 : \sigma^2 < 0.63^2$.

Reject if the sample variance $s^2 = \frac{1}{14}\sum_{i=1}^{15}(x_i - \bar{x})^2$ is significantly small. We know that

$$\frac{14s^2}{0.63^2} \sim \chi^2_{14}$$

under H_0.

Thus we reject H_0 if $\frac{\sum_{i=1}^{15}(x_i-\bar{x})^2}{0.63^2} < \chi^2_{14,0.05} = 6.57$.

We observe $\frac{\sum_{i=1}^{15}(x_i-\bar{x})^2}{0.63^2} = \frac{2.2}{0.63^2} = 5.54$.

This means we reject H_0 at 0.05 significance level.

Oppgave 2

a)

$$P(Y > 10) = 1 - P(Y \leq 10) = 1 - \sum_{y=0}^{10} \frac{5^y}{y!} e^{-5} = 1 - 0.986 = 0.014$$

$$P(Y < 5) = P(Y \leq 4) = 0.44$$

$$P(Y = 0 | Y < 5) = \frac{P(Y = 0 \cap Y < 5)}{P(Y < 5)} = \frac{P(Y = 0)}{P(Y < 5)} = \frac{e^{-5}}{0.44} = 0.015$$

b) The likelihood function is the probability of getting $Y = 261$ with parameter $4 \cdot 52\lambda = 208\lambda$. This likelihood is a function of the parameter λ. The log-likelihood becomes

$$l(\lambda) = \log P(Y = 261; \lambda) = 261 \log(208\lambda) - \log(261!) - 208\lambda$$

The maximum is found by differentiation.

$$l'(\lambda) = \frac{261}{\lambda} - 208 = 0 \Rightarrow \lambda = \frac{261}{208}.$$

This gives $\hat{\lambda} = 261/208 = 1.25$.

c) The moment generating function of a sum of two independent variables is the product of the moment generating functions.

The moment generating function of a Poisson distribution is

$$M_X(t) = \sum_{y=0}^{\infty} e^{ty} \frac{\lambda^y}{y!} e^{-\lambda} = \sum_{y=0}^{\infty} \frac{(e^t\lambda)^y}{y!} e^{-\lambda} = e^{\lambda(e^t-1)} \sum_{y=0}^{\infty} \frac{(e^t\lambda)^y}{y!} e^{-e^t\lambda} = e^{\lambda(e^t-1)}$$

where the last sum equal 1 because it is the sum over a Poisson variable with parameter $e^t\lambda$. Then

$$M_Z(t) = M_X(t) M_Y(t) = e^{\lambda(e^t-1)} e^{\nu(e^t-1)} = e^{(\lambda+\nu)(e^t-1)}$$

We recognize the functional form of the moment generating function. This is the moment generating function of a Poisson distribution with parameter $\lambda + \nu$.

Oppgave 3

a)

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\mathrm{Var}(\sum_{i=1}^n Y_i(x_i - \bar{x}))}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} = \frac{\sum_{i=1}^n \mathrm{Var}((x_i - \bar{x})Y_i)}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \mathrm{Var}(Y_i)}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2}$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \tau^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} = \frac{\tau^2 \sum_{i=1}^n (x_i - \bar{x})^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} == \frac{\tau^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Residuals are defined by $Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$, i.e. the distance from the data to the fitted line. The expression uses the sum of these residuals. The usual average would take n in the denominator, but we correct for the number of estimated parameters in the line (which is here 2) to get an unbiased estimator for τ^2.

b) For both kinds of ice cream: $\hat{\beta}_1^a \sim N(\beta_1^a, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2})$. Thus, by defining $\delta = \beta_1^f - \beta_1^s$ we have

$$\hat{\delta} = \hat{\beta}_1^f - \hat{\beta}_1^s \sim N\left(\delta, \frac{2\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right).$$

We estimate σ^2 from the residuals as follows:

$$s^2 = \frac{1}{2 \cdot 14 - 2 \cdot 2} \left(\sum_{i=1}^{14} (Y_i^f - \hat{\beta}_0^f - \hat{\beta}_1^f x_i)^2 + \sum_{i=1}^{14} (Y_i^s - \hat{\beta}_0^s - \hat{\beta}_1^s x_i)^2\right)$$

$$= (7300^2 + 7046^2)/24 = 2071^2.$$

This gives that

$$\frac{\widehat{\delta} - \delta}{\sqrt{\frac{2s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t_{n-4}.$$

We get $n-4$ degrees of freedom because we have estimated a total of four parameters in the two regression lines. In turn this gives

$$P\left(t_{24,0.05} < \frac{\hat{\delta} - \delta}{\sqrt{\frac{2s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} < t_{24,0.95} \right) = 0.90$$

Solving each of the two inequalities seperately with respect to δ we get

$$P\left(\hat{\delta} + t_{24,0.05}\sqrt{\frac{2s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} < \delta < \hat{\delta} + t_{24,0.95}\sqrt{\frac{2s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right) = 0.90$$

A 90% confidence interval for δ is thereby given by

$$\left(\hat{\beta}_1^f - \hat{\beta}_1^s + t_{24,0.05}\sqrt{\frac{2s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \hat{\beta}_1^f - \hat{\beta}_1^s + t_{24,0.95}\sqrt{\frac{2s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right).$$

We have $t_{24,0.95} = 1.71$, and by symmetri av the t-distibution $t_{24,0.05} = -1.71$. Inserting numbers we get a 90 percent confidence interval: $(278 \pm 1.71 \cdot \sqrt{2/260.4} \cdot 2071) = (278 \pm 310) = (-32, 588)$.

c) The figure shows that residuals are very similar for common days. Both are positive day 1, both negative day 2, positive day 3 and 4, and so on. This indicates that cream and limonade sales are not independent. If sales are very high on one type a given day, it is likely higher for the other one that day as well. The model with independence does not account for this. Maybe a regression model with some common model parameters would be a better option here.

Oppgave 4

$P(\textit{direkte seier}) = P(7) + P(11) = 6/36 + 2/36 = 8/36$.

La B_i = 4 kastet i runde i, $P(B_i) = 3/36$. La C_i = Hverken 4 eller 7 kastet i runde i, $P(C_i) = 27/36$. Da er $P(A_4) = P(B_1 \cap B_2) + P(B_1 \cap C_2 \cap B_3) + P(B_1 \cap C_2 \cap C_3 \cap B_4) + \ldots$, dvs $P(A_4) = P(B_1)P(B_2) + P(B_1)P(C_2)P(B_3) + P(B_1)P(C_2)P(C_3)P(B_4) + \ldots$, $P(A_4) = (\frac{3}{36})^2 \sum_{k=0}^{\infty} (\frac{27}{36})^k = (\frac{3}{36})^2 \frac{1}{1-27/36} = 1/36$.

$P(\textit{vinne}) = P(\textit{direkte seier}) + P(A_4) + P(A_5) + \ldots + P(A_{10})$.

Her er $P(A_4) = P(A_{10}) = 1/36$, $P(A_5) = P(A_9) = 16/360$, $P(A_6) = P(A_8) = 25/396$. Da blir $P(\textit{vinne}) = 0.493$.