MATH7501 Examination 2011: Solutions and Marking Scheme

1. (a) E and F are independent if $P(E \cup F) = P(E)P(F)$.

   (b) E and F are disjoint if $E \cap F = \emptyset$.

   (c) $P(A_1 \cap A_2 \cap A_3) = P(A_3|A_1 \cap A_2)P(A_1 \cap A_2)$, where
   $P(A_1 \cap A_2) = P(A_2|A_1)P(A_1)$.
   Hence $P(A_1 \cap A_2 \cap A_3) = P(A_3|A_1 \cap A_2)P(A_2|A_1)P(A_1)$.

   (d) $P(A|B) > P(A) \Rightarrow P(A \cap B)/P(B) > P(A) \Rightarrow P(A \cap B)/P(A) > P(B)$, i.e. $P(B|A) > P(B)$.

   (e) Here it is important to bear in mind that X and Y represent the number of heads in the *first* two tosses and in the *last* two tosses, respectively. Hence,

      i. $P(X = 0) = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$ and $P(Y = 1) = 2\left(\frac{1}{2}\right)^2 = \frac{1}{2}$.
      $P(X = 0, Y = 1) = P(TTH) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$, and $P(X = 0, Y = 1) = \frac{1}{8} = \frac{1}{4} \times \frac{1}{2} = P(X = 0)P(Y = 1)$.
      Hence $\{X = 0\}$ and $\{Y = 1\}$ are independent events.

      ii. $P(Y = 2) = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$.
      $P(X = 0, Y = 2) = 0 \neq P(X = 0)P(Y = 2)$.
      Hence $\{X = 0\}$ and $\{Y = 2\}$ are dependent events.

      iii. X and Y are not independent random variables.
      This would require

$$P(X = i, Y = j) = P(X = i)P(Y = j) \ \forall i, j.$$

2. (a) i. E.g., $\{B_1, B_2, \dots, B_n\}$ is a partition if $B_i \cap B_j = \emptyset \; \forall i \neq j$ and $\cup_{i=1}^{n} B_i = \Omega$.

ii. $\forall A \in \Omega$, $A = A \cap (\cup_{i=1}^{n} B_i) = \cup_{i=1}^{n} (B_i \cap A)$
Note that $(B_i \cap A) \cap (B_j \cap A) = \emptyset \; \forall i \neq j$.
Hence $P\left(\cup_{i=1}^{n} (B_i \cap A)\right) = \sum_{i=1}^{n} P(B_i \cap A)$.
It follows that

$$P(A) = \sum_{i=1}^{n} P(B_i \cap A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

iii.

$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}, \; \forall j.$$

(b) $A = \{\text{hamster trodden on}\}$.
$B_i = \{\text{brother i plays with hamster}\}$.
$P(B_1) = \frac{2}{10}$, $P(B_2) = \frac{5}{10}$, $P(B_3) = \frac{3}{10}$.
$P(A|B_1) = \frac{3}{10}$, $P(A|B_2) = \frac{4}{10}$, $P(A|B_3) = \frac{3}{10}$.
The required probabilities are given by

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{P(A)},$$

where

$$P(A) = P(B_1)P(A|B_1) + P(S_2)P(A|B_2) + P(S_3)P(A|B_3).$$

Hence,
$P(B_1|A) = \frac{\frac{3}{10}\frac{2}{10}}{\frac{3}{10}\frac{2}{10} + \frac{4}{10}\frac{5}{10} + \frac{3}{10}\frac{3}{10}} = \frac{6}{35}$,
$P(B_2|A) = \frac{\frac{4}{10}\frac{5}{10}}{\frac{35}{100}} = \frac{20}{35}$,
$P(B_3|A) = \frac{9}{35}$.
Brother no. 2 is the most likely.

3. (a) $\Pi_X(s) = \sum_{k=1}^{\infty} s^k p(1-p)^{k-1} = sp\sum_{k=1}^{\infty}[s(1-p)]^{k-1}$.

Now by setting $n = k-1$, it follows that

$$sp\sum_{n=0}^{\infty}[s(1-p)]^n = \frac{sp}{1-(1-p)s}.$$

(b) Recall that $E(X) = \Pi'_X(s)|_{s=1}$. In this case

$\Pi'_X(s) = \frac{p}{1-(1-p)s} + \frac{(1-p)sp}{[1-(1-p)s]^2} = \frac{[1-(1-p)s]p+sp(1-p)}{[1-(1-p)s]^2} = \frac{p}{[1-(1-p)s]^2}$, hence

$E(X) = \frac{p}{p^2} = \frac{1}{p}$.

As for the variance, recall that

$Var(X) = E(X^2) - E(X)^2 = E[X(X-1)] + E(X) - E(X)^2$

and that $E[X(X-1)] = \Pi''_X(s)|_{s=1}$.

Here

$\Pi''_X(s) = \frac{2p(1-p)}{[1-(1-p)s]^3}$ and $E[X(X-1)] = \frac{2p(1-p)}{p^3} = \frac{2(1-p)}{p^2}$.

Hence

$Var(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2(1-p)+(p-1)}{p^2} = \frac{1-p}{p^2}$.

(c) For $n = 1, 2, 3, \ldots$,

$P(X-k+1 \geq n | X \geq k) =$

$$\frac{P(X \geq n+k-1, X \geq k)}{P(X \geq k)} = \frac{P(X \geq n+k-1)}{(1-p)^{k-1}} = \frac{p(1-p)^{n+k-2}}{(1-p)^{k-1}} = p(1-p)^n.$$

(The denominator indicate "first $k-1$ trials must fail")

The result implies that $X-k+1$ conditioned on $X \geq k$ is again geometrically distributed.

(d) Since $X-k+1 \geq n | X \geq k$ follows a Geometric distribution with parameter p, we have that

$E[(X-k+1)^2 | X \geq k] = E(X^2) = E[X(X-1)] + E(X)$,

which is equal to

$\frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2(1-p)+p}{p^2} = \frac{2-p}{p^2}$.

4. (a) $b(T,\theta) = E(T) - \theta$, $MSE(T) = E\big[(T-\theta)^2\big]$.

(b)

$$\begin{aligned} MSE(T) &= E\big[\big(T - E(T) + E(T) - \theta\big)^2\big] \\ &= E\big[\big(T - E(T)\big)^2 + \big(E(T) - \theta\big)^2 + 2\big(T - E(T)\big)\big(E(T) - \theta\big)\big] \\ &= E[\big(T - E(T)\big)^2 + \big(E(T) - \theta\big)^2. \end{aligned}$$

Note that $E\big[\big(T - E(T)\big)\big(E(T) - \theta\big)\big] = \big[E(T) - \theta\big]E\big[T - E(T)\big] = 0$. Hence,

$$MSE(T) = Var(T) + b^2(T,\theta).$$

(c)

$$L(\theta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\theta}\exp\left(-\frac{x_i}{\theta}\right) = \frac{1}{\theta^n}\exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right) = \frac{1}{\theta^n}\exp\left(-\frac{n\overline{x}}{\theta}\right)$$

and

$$l(\theta) = -n\log(\theta) - \frac{n\overline{x}}{\theta}.$$

Differentiating $l(\theta)$ w.r.t. θ yields $l'(\theta) = -\frac{n}{\theta} + \frac{n\overline{x}}{\theta^2}$. Solving $l'(\theta) = 0$ gives $\widehat{\theta} = \overline{x}$ (we may readily check that $l''(\overline{x}) = -\frac{n}{\overline{x}^2}$, which implies that $\overline{x}$ is a maximum). Therefore, the estimator is

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(d) $Var(X_i) = E(X_i^2) - E(X_i)^2 = 2\theta^2 - \theta^2 = \theta^2$. Now,

$$E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \theta \quad \text{so that} \quad b(T,\theta) = E(\overline{X}) - \theta = 0,$$

which shows that $\overline{X}$ is an unbiased estimator of θ. As a result,

$$MSE(\overline{X}) = Var(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\theta^2}{n}.$$

5. (a) If $Z \sim N(0,1)$ and $U \sim \chi_n^2$, and Z and U are independent, then the distribution of the ratio $T = Z/\sqrt{U/n}$ is called student's t distribution with n degrees of freedom.

(b) s_X^2 and s_Y^2 are independent with $\frac{(n-1)s_X^2}{\sigma^2} \sim \chi_{n-1}^2$ and $\frac{(n-1)s_Y^2}{\sigma^2} \sim \chi_{m-1}^2$. As the sum of two independent χ^2-distributions is also a χ^2-distribution with degrees of freedom equal to the sum of the two degrees of freedom we have that

$$\frac{(n-1)\,s_X^2}{\sigma^2} + \frac{(n-1)\,s_Y^2}{\sigma^2} = \frac{(n+m-2)\,s_P^2}{\sigma^2} \sim \chi_{n+m-2}^2.$$

Noting that

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0,1),$$

and that Z and s_P^2 are independent, then

$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \Bigg/ \sqrt{\frac{s_P^2}{\sigma^2}} = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{s_P\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}.$$

(c) i. We assume that $X_1, \ldots, X_{15}$ are *iid* $N(\mu_X, \sigma^2)$, $Y_1, \ldots, Y_{15}$ are *iid* $N(\mu_Y, \sigma^2)$ and that the X_i are independent of the Y_j. We test the hypothesis $H_0 : \mu_X = \mu_Y$ *vs* $H_1 : \mu_X \neq \mu_Y$. If H_0 is true then $\widehat{t} = \frac{\overline{x}-\overline{y}}{s_P\sqrt{\frac{1}{15}+\frac{1}{15}}}$ is a realization from a t-distribution with 28 degrees of freedom. The rejection region for a t-test at the 5% significance level is $C = \{(\mathbf{x}, \mathbf{y} : t \leq -2.05, t \geq 2.05)$ as $t_{28,0.025} = 2.05$. Using the data summaries,

$$\overline{x} = \frac{1195}{15} = 79.6, \quad s_X^2 = \frac{1}{14}\left\{97997 - 15\left(\frac{1195}{15}\right)^2\right\} = \frac{2795.3}{14} = 199.6,$$

$$\overline{y} = \frac{1302}{15} = 86.8, \quad s_Y^2 = \frac{1}{14}\left\{114344 - 15\left(\frac{1302}{15}\right)^2\right\} = \frac{1330.4}{14} = 95.028,$$

$$s_P^2 = \frac{14\left(\frac{2795.3}{14}\right) + 14\left(\frac{1330.4}{14}\right)}{28} = \frac{4125.73}{28} = 147.347.$$

Hence,

$$\widehat{t} = \frac{\left(\frac{1195}{15}\right) - \left(\frac{1302}{15}\right)}{\sqrt{147.347\left(\frac{1}{15} + \frac{1}{15}\right)}} = \frac{-7.13}{4.4324} = -1.61.$$

We can not reject H_0 at the 5% level as $-1.61 \notin C$.

ii. Let $\delta = \mu_X - \mu_Y$ and use the fact that $\frac{\overline{X}-\overline{Y}-\delta}{s_P\sqrt{\frac{1}{15}+\frac{1}{15}}} \sim t_{28}$.

It follos that

$P(-2.05 < t_{28} < 2.05) = P\left(-2.05 < \frac{\overline{X}-\overline{Y}-\delta}{s_P\sqrt{\frac{1}{15}+\frac{1}{15}}} < 2.05\right) =$

$P\left(\overline{X}-\overline{Y}-2.05s_P\sqrt{\frac{1}{15}+\frac{1}{15}} < \delta < \overline{X}-\overline{Y}+2.05s_P\sqrt{\frac{1}{15}+\frac{1}{15}}\right) = 0.95.$

Hence,

$\left(\overline{X}-\overline{Y}-2.05s_P\sqrt{\frac{1}{15}+\frac{1}{15}};\overline{X}-\overline{Y}+2.05s_P\sqrt{\frac{1}{15}+\frac{1}{15}}\right)$ is a random interval which contains δ with probability 0.95.

A realisation of this, $\left(\overline{x}-\overline{y}-2.05s_P\sqrt{\frac{1}{15}+\frac{1}{15}};\overline{x}-\overline{y}+2.05s_P\sqrt{\frac{1}{15}+\frac{1}{15}}\right)$, is a 95% confidence interval for δ.

In this case we have,

$\left(\frac{-107}{15}-2.05\sqrt{147.347\left(\frac{1}{15}+\frac{1}{15}\right)};\frac{-107}{15}+2.05\sqrt{147.347\left(\frac{1}{15}+\frac{1}{15}\right)}\right) =$

$(-7.13-2.05(4.4324);-7.13+2.05(4.4324)) = (-16.22, 1.95)$.

There is a duality between the 95% confidence interval and the 5% significance test. Specifically, the confidence intervals contain all the values of the parameter for which we accept H_0. We note, in particular, that it contains 0.

6. (a) The normal equations for the estimated coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ are

$$\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\text{and} \quad \sum_{i=1}^{n} x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0\,.$$

The predicted value for the i^{th} case is $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i$, hence $r_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$. Substituting into the normal equations gives the required results directly.

(b) The two constraints on the residuals arise as a result of estimating the two parameters β_0 and β_1, therefore the estimate of σ^2 has $n-2$ degrees of fredom in a regression problem.

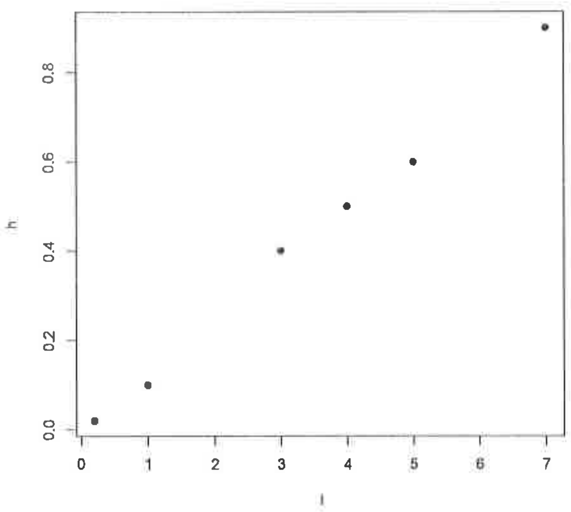
(c)

$$\frac{\partial}{\partial \beta}\sum_{i=1}^{n}(y_i - \beta)^2 = -2\sum_{i=1}^{n}(y_i - \beta) = 0 \Rightarrow \hat{\beta} = \overline{y}.$$

Given the assumed model it is obvious that $\hat{\beta} = \overline{y}$.

(d) None of these!

(e) The scatter plot suggests that a linear model with *zero* intercept can represent the pattern in the data. Hence, a suitable model to predict h_i is



$$h_i = \beta l_i + \epsilon_i.$$

The least squares estimate of β is

$$\hat{\beta} = \frac{\sum_{i=1}^{6} h_i l_i}{\sum_{i=1}^{6} l_i^2} = \frac{0.004 + 0.1 + 1.2 + 2.0 + 3.0 + 6.3}{0.04 + 1 + 9 + 16 + 25 + 49} = \frac{12.604}{100.04} = 0.126,$$

which implies a speed of $1/\hat{\beta} = 7.93$ kilometers per hour.

(f) i. The regression line is of the form $y = \hat{\beta}_0 + \hat{\beta}_1 x$, where $\hat{\beta}_1 = C_{xy}/C_{xx} = 5.42$ and $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = 0.6538$.

ii. The estimated error variance is $S(\hat{\beta}_0, \hat{\beta}_1)/(n-2)$, where $S(\hat{\beta}_0, \hat{\beta}_1) = C_{yy} - C_{xy}^2/C_{xx} = 4.68 \Rightarrow \hat{\sigma}^2 = 0.26$.

The estimated standard errors for $\hat{\beta}_0$ and $\hat{\beta}_1$ are

$$\hat{\text{s.e.}}(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 \left[\frac{1}{20} + \frac{\overline{x}^2}{C_{xx}}\right]} = 0.2256 \quad \text{and} \quad \hat{\text{s.e.}}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{C_{xx}}} = 1.7698.$$

The upper 2.5% point of t_{18} is 2.1009, so confidence intervals are:

for slope: $5.42 \pm (2.1009 \times 1.7698) = (1.701,\ 9.138)$.

for intercept: $0.6538 \pm (2.1009 \times 0.2256) = (0.179, 1.127)$.

The CI for the intercept does not include zero, so we reject $H_0 : \beta_0 = 0$ using a 2-tailed test at the 5% level.

iii. When $x = 0.9$, $\hat{y} = 0.6538 + (5.42 \times 0.9) = 5.5318$ units.

iv. The coefficient of determination here is $r_{xy}^2 = C_{xy}^2/C_{xx}C_{yy} = 0.34$. Hence 34% of the variability in fat content is explained by the relationship, which is not extremely useful.