Royal Statistical Society
GRADUATE DIPLOMA 2000
Statistical Theory and Methods I

1 (a) Let the events $E_1, E_2, \cdots$ in a sample space S be such that $(i) E_1 \bigcup E_2 \bigcup E_3 \cdots = S$; $(ii) P(E_i \bigcap E_j) = 0, i \neq j$; $(iii) P(E_i) > 0$ ($\{E_i\}$ are a set of mutually exclusive, exhaustive events).
Let A be any event in S such that $P(A) > 0$. Then

$$P(E_j|A) = \frac{P(A|E_j)P(E_j)}{\sum_i P(A|E_i)P(E_i)}$$

(b) Define the events $E_1 \equiv$Patient is resistant, $E_2 \equiv$Patient is not resistant, $S \equiv$Patient is successfully treated. With A, $P(S|E_1) = 0.92$ *and* $P(S|E_2) = 0.87$. With B, $P(S|E_1) = 0.75$ *and* $P(S|E_2) = 0.95$ Also $P(E_1) = \theta,\ P(E_2) = (1-\theta)$.

$$P(E_1|not\ S) = \frac{P(not\ S|E_1)P(E_1)}{P(not\ S|E_1)P(ES_1) + P(not\ S|E_2)P(E_2)}$$

and for B this is $= 0.25\theta/\{0.25\theta + 0.05(1-\theta)\}$ For A, $P(S) = P(S|E_1)P(E_1) + P(S|E_2)P(E_2) = 0.92\theta + 0.87(1-\theta)$ and for B, $P(S) = 0.75\theta + 0.95(1-\theta)$
$0.75\theta + 0.95(1-\theta) > 0.92\theta + 0.87(1-\theta) \Leftrightarrow 0.08(1-\theta) > 0.17\theta \Leftrightarrow 0.08 > 0.25\theta$ *or* $0.32 > \theta$
When $\theta = 0.25,\ P(S) = 0.75 \times 0.25 + 0.95 \times 0.75 = 1.20 \times 0.75 = 0.90$. for treatment B, Let $X = number\ of\ successfully\ treated\ patients\ out\ of\ 20, 40$ that X is Binomial (20,0.90)

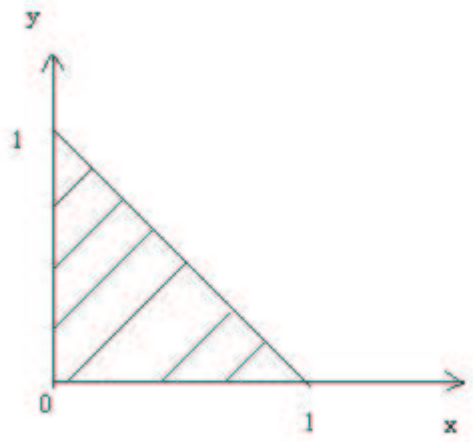
$$P(X \geq 18) = 0.9^{20} + 20 \times 0.9^{19} \times 0.1 + \frac{20 \times 19}{2} \times 0.9^{18} \times 0.1^2$$

$$0.9^{18}(0.9^2 + 2 \times 0.9 + 190 \times 0.01) = 0.9^{18} \times 4.51 = 0.6769$$

2 (i) $E[X^r Y^s] = \int_{x=0}^{1} \int_{y=0}^{1-x} X^r Y^s X^{\alpha-1} Y^{\beta-1} (1-X-Y)^{r-1} \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} dY dX$ evaluating over the region.

The inner integral in Y is $\int_0^{1-x} Y^{s+\beta-1}(1-x-y)^{\gamma-1} dY$; substitute $u = \frac{Y}{1-X}$ to give $\int_0^1 u^{s+\beta-1}(1-x)^{s+\beta-1}\{1-x-u(1-x)\}^{\gamma-1}(1-x)du = (1-x)^{s+\beta+\gamma-1} \int_0^1 u^{s+\beta-1}(1-u)^{\gamma-1} du = (1-x)^{s+\beta+\gamma-1} B(\beta+s, \gamma)$
and the whole integral is

$$\int_0^1 X^{\alpha+r-1}(1-X)^{\beta+\gamma+s-1} \cdot \frac{\Gamma(\alpha+\beta+\gamma)B(\beta+s,\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} dX$$



$$= \frac{B(\alpha + r, \beta + \gamma + s)\Gamma(\alpha + \beta + \gamma)B(\beta + s, \gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}$$

So that

$$E[X^r Y^s] = \frac{\Gamma(\alpha + r)\Gamma(\beta + \gamma + s)\Gamma(\alpha + \beta + \gamma)\Gamma(\beta + s)\Gamma(\gamma)}{\Gamma(\alpha + \beta + \gamma + r + s)\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)\Gamma(\beta + \gamma + s)}$$

$$= \frac{\Gamma(\alpha + r)\Gamma(\beta + s)\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + \gamma + r + s)}$$

(ii) *For* $r = 1,\ s = 0$ $E[X] = \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)} \cdot 1 \cdot \frac{\Gamma(\alpha+\beta+\gamma)}{(\alpha+\beta+\gamma)\Gamma(\alpha+\beta+\gamma)} = \frac{\alpha}{\alpha+\beta+\gamma}$
For $r = 2,\ s = 0$ $E[X^2] = \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)} \cdot 1 \cdot \frac{\Gamma(\alpha+\beta+\gamma)}{(\alpha+\beta+\gamma+1)(\alpha+\beta+\gamma)\Gamma(\alpha+\beta+\gamma)} = \frac{(\alpha+1)\alpha}{(\alpha+\beta+\gamma+1)(\alpha+\beta+\gamma)}$

$$V[X] = E[X^2] - (E[X])^2 = \frac{(\alpha + 1)\alpha}{(\alpha + \beta + \gamma + 1)(\alpha + \beta + \gamma)} - \frac{\alpha^2}{(\alpha + \beta + \gamma)^2}$$

$$= \frac{\alpha}{\alpha + \beta + \gamma}\{\frac{\alpha + 1}{\alpha + \beta + \gamma + 1} - \frac{\alpha^2}{\alpha + \beta + \gamma}\}$$

$$= \frac{\alpha(\beta + \gamma)}{(\alpha + \beta + \gamma + 1)(\alpha + \beta + \gamma)^2}$$

(iii)

$$P(X, Y) = \frac{cov(X, Y)}{\sqrt{V(X)V(Y)}} \text{ and by symmetry } E(Y)\frac{\beta}{\alpha + \beta + \gamma}$$

For

$$r = s = 1,\ E(XY) = \frac{\alpha\beta}{(\alpha+\beta+\gamma+1)(\alpha+\beta+\gamma)} \ \textit{which gives}$$

$$\begin{aligned} cov(X,Y) &= \frac{\alpha\beta}{(\alpha+\beta+\gamma+1)(\alpha+\beta+\gamma)} - \frac{\alpha\beta}{(\alpha+\beta+\gamma)^2} \\ &= \frac{-\alpha\beta}{(\alpha+\beta+\gamma+1)(\alpha+\beta+\gamma)^2} \end{aligned}$$

and

$$\begin{aligned} P(X,Y) &= \frac{-\alpha\beta}{(\alpha+\beta+\gamma+1)(\alpha+\beta+\gamma)^2} \Big/ \sqrt{\frac{\alpha(\beta+\gamma)\beta(\alpha+\gamma)}{(\alpha+\beta+\gamma+1)^2(\alpha+\beta+\gamma)^4}} \\ &= \frac{-\alpha\beta}{\sqrt{\alpha\beta(\alpha+\gamma)(\beta+\gamma)}} \ \textit{or} \ -\sqrt{\frac{\alpha\beta}{(\alpha+\gamma)(\beta+\gamma)}} \end{aligned}$$

3 (i)

$$x\begin{pmatrix} M \\ x \end{pmatrix} = \frac{xM!}{x!(M-x)!} = \frac{M(M-1)!}{(x-1)!(M-x)!} = M\begin{pmatrix} M-1 \\ x-1 \end{pmatrix}$$

for any $x = 1, 2, \cdots n$.

$$\begin{aligned} E[X] &= \sum_{x=0}^{n} \frac{x\begin{pmatrix} M \\ x \end{pmatrix}\begin{pmatrix} N-M \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} = \sum_{x=1}^{n} \frac{M\begin{pmatrix} M-1 \\ x-1 \end{pmatrix}\begin{pmatrix} N-M \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \\ &= \sum_{x=1}^{n} \frac{M\begin{pmatrix} M-1 \\ x-1 \end{pmatrix}\begin{pmatrix} (N-1)-(M-1) \\ (n-1)-(x-1) \end{pmatrix}}{\frac{N}{n}\begin{pmatrix} N-1 \\ n-1 \end{pmatrix}} \\ &= \frac{nM}{N}\sum_{n=0}^{n-1} \frac{\begin{pmatrix} M-1 \\ u \end{pmatrix}\begin{pmatrix} (N-1)-(M-1) \\ (n-1)-u \end{pmatrix}}{\begin{pmatrix} N-1 \\ n-1 \end{pmatrix}} = \frac{nM}{N} \end{aligned}$$

Using the result given,

$$V[X] = E[X(X-1)] + E[X] - \{E[X]\}^2 = \frac{n(n-1)M(M-1)}{N(N-1)} + \frac{nM}{N} - \frac{n^2M^2}{N^2}$$

$$= \frac{nM}{N^2(N-1)} \begin{pmatrix} N(n-1)(M-1)+N(N-1) \\ -nM(N-1) \end{pmatrix}$$

$$= \frac{nM(N-M)(N-n)}{N^2(N-1)} = n \cdot \frac{M}{N} \cdot (1-\frac{M}{N}) \cdot (\frac{N-n}{N-1})$$

(ii)

$$P(X=x \bigcap Y=y) = \begin{pmatrix} 60 \\ x \end{pmatrix} \begin{pmatrix} 20 \\ y \end{pmatrix} \begin{pmatrix} 20 \\ 5-x-y \end{pmatrix} / \begin{pmatrix} 100 \\ 5 \end{pmatrix}$$

$$y = 0, 1, \cdots 5 \quad x = 0, 1, \cdots 5-y$$

$$P(X=x|Y=y) = P(X=x \bigcap Y=y)/P(Y=y) \; and \; P(Y=y)$$

$$= \begin{pmatrix} 20 \\ y \end{pmatrix} \begin{pmatrix} 80 \\ 5-y \end{pmatrix} / \begin{pmatrix} 100 \\ 5 \end{pmatrix}$$

therefore the conditional probability is

$$\begin{pmatrix} 60 \\ x \end{pmatrix} \begin{pmatrix} 20 \\ (5-y)-x \end{pmatrix} / \begin{pmatrix} 80 \\ 5-y \end{pmatrix}$$

Using results(a),with $N = 80, M = 60, n = 5-y$

$$E[X|Y=y] = \frac{60}{80}(5-y) = \frac{3}{4}(5-y), \quad y = 0, 1, \cdots 5$$

and

$$V[X|Y=y] = \frac{3}{16}(5-y)(\frac{75+y}{79}), \quad y = 0, 1, \cdots 5$$

.

4 (i)

$$M_x(t) = \sum_{x=1}^{\infty} e^{xt}(1-\theta)^{x-1}\theta = \theta e^t \sum_{x=1}^{\infty} \{(1-\theta)e^t\}^{x-1} = \theta e^t/(1-(1-\theta)e^t)$$

$$E[X^r] = M_x^{(r)}(0), \; M_x^1(t) = \frac{\theta e^t(1-(1-\theta)e^t) + \theta e^t(1-\theta)e^t}{\{1-(1-\theta)e^t\}^2} = \frac{\theta e^t}{\{1-(1-\theta)e^t\}^2}$$

When $t = 0, \; we \; find \; E[X] = 1/\theta$

$$M_x''(t) = \frac{\{1-(1-\theta)e^t\}^2\theta e^t + \theta e^t \cdot 2\{1-(1-\theta)e^t\}\{(1-\theta)e^t\}}{\{1-(1-\theta)e^t\}^4} = \frac{\theta e^t(1+(1-\theta)e^t)}{\{1-(1-\theta)e^t\}^3}$$

$$When t = 0, E[X^2] = \frac{2-\theta}{\theta^2}, \; and \; V[X] = \frac{2-\theta}{\theta^2} - \frac{1}{\theta^2} = \frac{1-\theta}{\theta^2}$$

(ii) Let X_i be the number of trials after the $(i-1)^m$until the i^m, then $Y = X_1+X_2+\cdots+X_n$, each having the geometric distribution of part(a); and by the independence property of Bernouilli trials the $\{X_i\}$ are independent.
Thus $E[Y] = n/\theta$ *and* $V[Y] = \frac{n(1-\theta)}{\theta^2}$; and by the Central Limit Theorem, as

$$n \to \infty, \frac{Y - n/\theta}{\sqrt{\frac{n(1-\theta)}{\theta^2}}} \to N(0,1)$$

(Y has a negative binomial distribution in general.)

(iii) The required number n=400 is as in(b) with $\theta = 0.8$, Hence

$$P(Y \geq 520) = P\{\frac{0.8Y - 400}{\sqrt{400 \times 0.2}} \geq \frac{0.8 \times 520 - 400}{\sqrt{400 \times 0.2}}\}$$

$$= P\{\frac{0.8Y - 400}{\sqrt{80}} \geq \frac{16}{\sqrt{80}}\} = P(Z \geq 1.7889)$$

Where Z is N(0,1): this is $1 - \phi(1.7889) = 1 - 0.9632 = 0.0368$
(Applying a continuity correction, to find $P(Y \geq 519.5)$ leads to $P(Z > 1.744) = 0.0408$)

5 (i) The joint pdf is that of X and Y, which is by independence

$$\frac{x^{\frac{1}{2}r-1}e^{-\frac{1}{2}x}}{2^{\frac{1}{2}r}\Gamma(\frac{1}{2}r)} \cdot \frac{y^{\frac{1}{2}s-1}e^{-\frac{1}{2}y}}{2^{\frac{1}{2}s}\Gamma(\frac{1}{2}s)} \; for \; x > 0, y > 0.$$

$U = \frac{X}{r}/\frac{Y}{s}$ and $V = Y/s$, so that $Y = sV$ and $X = rUV$.The Jacobian

$$\begin{vmatrix} \frac{\partial X}{\partial U} & \frac{\partial X}{\partial V} \\ \frac{\partial Y}{\partial U} & \frac{\partial Y}{\partial V} \end{vmatrix} = \begin{vmatrix} rV & rU \\ 0 & s \end{vmatrix} = rsV$$

The pdf of U,V is

$$f(U,V) = \frac{(rUV)^{\frac{1}{2}r-1}(sV)^{\frac{1}{2}s-1}e^{-\frac{1}{2}(rUV+sV)}}{2^{\frac{1}{2}(r+s)}\Gamma(\frac{1}{2}r)\Gamma(\frac{1}{2}s)} \cdot rsV$$

$$= \frac{U^{(r-2)/2}V^{\frac{r+s}{2}-1}e^{-\frac{(rU+s)V}{2}}r^{\frac{1}{2}}s^{\frac{1}{2}s}}{2^{\frac{1}{2}(r+s)}\Gamma(\frac{1}{2}r)\Gamma(\frac{1}{2}s)} \; (U > 0, V > 0)$$

(ii) Integrate out V:

$$f_v(u) = \frac{r^{\frac{r}{2}}s^{\frac{s}{2}}u^{\frac{1}{2}r-1}}{2^{\frac{1}{2}(r+s)}\Gamma(\frac{1}{2}r)\Gamma(\frac{1}{2}s)} \int_0^\infty v^{\frac{r+s}{2}-1}e^{-\frac{1}{2}v(ru+s)}dv$$

and by writing $z = \frac{1}{2}v(ru+s)$, the integral is

$$\frac{2^{\frac{r+s}{2}}\Gamma(\frac{r+s}{2})}{(ru+s)^{\frac{1}{2}(r+s)}}$$

So that the marginal pdf is

$$\frac{\Gamma((r+s)/2)r^{r/2}s^{s/2}u^{r/2-1}}{\Gamma(r/2)\Gamma(s/2)(ru+s)^{\frac{1}{2}(r+s)}} \; or \; \frac{r^{\frac{1}{2}r}s^{\frac{1}{2}s}u^{\frac{1}{2}r-1}}{B(\frac{1}{2}r,\frac{1}{2}s)(ru+s)^{(r+s)/2}} \; which \; is \; F(r,s)$$

(iii) The exponential distribution with mean (expected value)2 is $\chi^2_{(2)}$. If X,Y are independent exponentials with mean θ^{-1}, then $2\theta x$ and $2\theta y$ are independent exponentials with mean 2, i.e. are $\chi^2_{(2)}$ U becomes

$$\frac{2\theta x/2}{2\theta y/2} \sim F_{(2,2)} \; i.e. \; \frac{X}{Y} \sim F_{(2,2)}$$

6 (i) For $\bigcup(-\theta,\theta),\ f(x)=\frac{1}{2\theta},\ (-\theta<\theta)\ and\ F(x)=\frac{x}{2\theta},\ (-\theta<\theta)\ Hence:$

$$F(u_1,u_n)=P(U_1\le u_1\ and\ U_n\le u_n)=P(U_n\le u_n)-P(U_1>u_1\ and\ U_n\le u_n)$$

$$=P(all\ data\le u_n)-P(all\ data\ between\ u_1\ and\ u_n)=\{F(u_n)\}^n-\{F(u_n)-F(u_1)\}^n$$

$$=(\frac{u_n}{2\theta})^n-(\frac{u_n-u_1}{2\theta})^n \qquad (-\theta<u_1<u_n<\theta)$$

$$Therefore\ f(u_1,u_n)=\frac{\partial^2}{\partial u_1\partial u_n}F(u_1,u_n)=\frac{n(n-1)(u_n-u_1)^{n-2}}{(2\theta)^n}$$

(ii) Change the variables to $R=U_n-U_1,\ T=U_1,\ i.e.\ U_1=T,\ U_n=R+T$

$$The\ Jacobian \begin{vmatrix} \frac{\partial U_1}{\partial R} & \frac{\partial U_1}{\partial T} \\ \frac{\partial U_2}{\partial R} & \frac{\partial U_2}{\partial T} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} \quad (in\ modulus)$$

giving $f(r,t)=n(n-1)r^{n-2}/(2\theta)^n\ (for\ -\theta<t<\theta;\ 0<r<\theta-t)$
Integrating out T from $-\theta$ to $\theta-r$ we have

$$f(r)=\frac{n(n-1)}{(2\theta)^n}\int_{-\theta}^{\theta-r}r^{n-2}dt=\frac{n(n-1)r^{n-2}}{(2\theta)^n}[t]_{-\theta}^{\theta-r}=\frac{n(n-1)r^{n-2}(2\theta-r)}{(2\theta)^n}\ (0<r<2\theta)$$

(iii)

$$E[R]=\frac{n(n-1)}{(2\theta)^n}\int_0^{2\theta}r^{n-2}(2\theta-r)rdr=\frac{n(n-1)}{(2\theta)^n}\int_0^{2\theta}(2\theta r^{n-1}-r^n)dr$$

$$=\frac{n(n-1)}{(2\theta)^n}(\frac{(2\theta)^{n+1}}{n}-\frac{(2\theta)^{n+1}}{n+1})=\frac{2\theta(n-1)}{n+1}$$

Thus $E[\frac{1}{2}R]=\theta(1-\frac{2}{n+1})$ and $\frac{1}{2}R$ is biased for θ, though asymptotically unbiased.

7

(a)

$$Y = \theta X^{\alpha},\ so\ X = (\frac{Y}{\theta})^{1/\alpha}\ and\ dX = (\frac{1}{\alpha})(\frac{Y}{\theta})^{\frac{1}{\alpha}-1}\frac{dY}{\theta}.$$

Y is a monotonic function (strictly increasing)of X, and so $f(Y)dY = f(X)dX$, which when written in terms of Y gives

$$\alpha\theta(\frac{Y}{\theta})^{\frac{\alpha-1}{\alpha}}exp(-Y)\frac{1}{\alpha\theta}(\frac{Y}{\theta})^{\frac{1}{\alpha}-1}dY,\ i.e.\ f(Y)dY = e^{-y}dy.\ \ (y>0)$$

This is the exponential distribution with mean 1.

(b) (i)

$$\left(\begin{array}{c}10\\3\end{array}\right) = \frac{10!}{3!7!} = 120.\ \ P(X=0) = \frac{1}{120}\cdot 1\cdot\left(\begin{array}{c}5\\3\end{array}\right) = \frac{1}{12} = P(X=3)$$

$$and\ P(X=1) = \frac{1}{120}\cdot 5\cdot\left(\begin{array}{c}5\\2\end{array}\right) = \frac{5}{12} = P(X=2)$$

$Hence:$	$x=0$	1	2	3
$P(x)$	0.0833	0.4167	0.4167	0.0833
$F(x)$	0.0833	0.5000	0.9167	1.000

For random numbers (to 3 d.p.) 001-083, take x=0; $084-500 \Rightarrow x=1$; $501-917 \Rightarrow x=2$; $918-000 \Rightarrow x=3$. Hence we obtain x=1, 1, 3, 2.

(ii) $F(x) = 1-e^{-x}$ $(x>0)$, so use $x = -ln(1-a)$ where u are the given random numbers.
We obtain 0.507, 0.089, 4.269, 0.772. The method used is the "inverse c.d.f."method.

(iii) From part (a), this is a Weibull with $\theta = 2$ and $\alpha = 1/2$, Hence $Y = 2\sqrt{x}$ is Exponential(1), i.e. $X = (\frac{1}{2}Y)^2$. Taking the results of(ii), $X = (\frac{1}{2}\times 0.507)^2 = 0.0643$, and the other values are 0.0020, 4.5560, 0.1490.

8 Transition matrix is

$$P = \left(\begin{array}{cccccc}
1-(1-\phi_1)\beta & (1-\phi_1)\beta & 0 & \cdots & 0 & 0\\
\phi_2 & (1-\phi_2)(1-\beta) & (1-\phi_2)\beta & \cdots & 0 & 0\\
0 & \phi_3 & (1-\phi_3)(1-\beta) & \cdots & 0 & 0\\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\
0 & 0 & 0 & \cdots & (1-\phi_{N-1})(1-\beta) & (1-\phi_{N-1})\beta\\
0 & 0 & 0 & \cdots & \phi_N & 1-\phi_N
\end{array}\right)$$

For the stationary distribution $\Pi = P\Pi$ and so

$$
\begin{aligned}
&\Pi_1 = (1-\beta+\phi_1\beta)\Pi_1 + \phi_2\Pi_2 \\
&\Pi_2 = (1-\phi_1)\beta\Pi_1 + (1-\phi_2)(1-\beta)\Pi_2 + \phi_3\Pi_3 \\
&\vdots \\
&\Pi_i = (1-\phi_{i-1})\beta\Pi_{i-1} + (1-\phi_i)(1-\beta)\Pi_i + \phi_{i+1}\Pi_{i+1} \\
&\vdots \\
&\Pi_N = (1-\phi_{N-1})\beta\Pi_{N-1} + (1-\phi_N)\Pi_N
\end{aligned}
$$

The first equation gives $\Pi_2 = \beta(1-\phi_1)\Pi_1/\phi_2$
Then the second equation becomes

$$\Pi_2 = \phi_2\Pi_2 + (1-\phi_2)(1-\beta)\Pi_2 + \phi_3\Pi_3, \ or \ \Pi_3 = \beta(1-\phi_2)\Pi_2/\phi_3$$

and this result generalizes to the remaining equations so that

$$\Pi_{i+1} = \beta(1-\phi_i)\Pi_i/\phi_{i+1} \ for \ i = 0, 1, \cdots N-1.$$

If

$$\Pi_i < \frac{\beta}{1+\beta},$$

then

$$\frac{\Pi_i}{\Pi_{i-1}} = \beta\frac{(1-\phi_{i-1})}{\phi_i} \geq \frac{\beta(1-\phi_i)}{\phi_i}$$

since it is given that $\phi_i \geq \phi_{i-1}$; so

$$\frac{\Pi_i}{\Pi_{i-1}} \geq \beta\frac{1/(1+\beta)}{\beta/(1+\beta)}$$

because $\Phi_i < \frac{\beta}{1+\beta}$,Thus $\frac{\Pi_i}{\Pi_{i-1}} \geq 1$ Conversely, if

$$\phi_i > \frac{\beta}{1+\beta},$$

then

$$\frac{\Pi_{i+1}}{\Pi_i} = \beta\frac{(1-\phi_i)}{\phi_{i+1}} \leq \beta\frac{(1-\phi_i)}{\phi_i}$$

since

$$\phi_i \leq \phi_{i+1};$$

and now

$$\frac{\Pi_{i+1}}{\Pi_i} < 1$$

The mode of the stationary distribution is therefore found at(approximately)$\phi = \frac{\beta}{1+\beta}, or \ \beta = \frac{\phi}{1-\phi}$
Therefore choose $\beta = \frac{0-1}{1-0.1} = 1/9$.

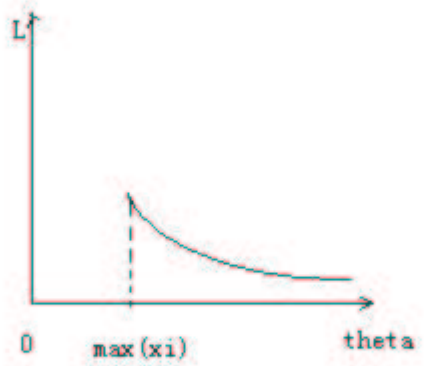
Statistical Theory and Methods II

1 (a) For a random sample of n observations drawn from a distribution with probability density (or mass)function $f(X, \theta)$, the likelihood function of the observations $X = (X_1, X_2, \cdots, X_n)^T$ is the joint probability function $L(\theta) = \prod_{i=1}^{n} f(X_i, \theta)$, considered as a function of the parameter θ in the distribution.

(b) (i) When $X \sim U(0, \theta), then\ E[X] = \frac{1}{2}\theta,\ and\ V[X] = \frac{1}{12}\theta^2$ Thus $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ also has expectation $\frac{1}{2}\theta$ and the first moment estimator gives $\overline{X} = \frac{1}{2}\theta$ *or* $\tilde{\theta}_1 = 2\overline{X}$

$= \frac{\theta^2}{3n}$

(ii)



The function decreases steadily as θ increases. It does not have a zero derivative for any value of $\theta \geq max(x_i)$. The value at $max(x_i)$ is therefore the maximum likelihood in the range $max(x_i) < \theta < \infty$ and this shows than $max(x_i)$ is the m.l.e. $\hat{\theta}_2$.

(iii)

$Y = max(x_i),\ P(Y \leq y) = P(\{x_1, x_2, \cdots, x_n\} \leq y) = \Pi P(x_i \leq y) by\ independence,$

$$i.e.\ F(Y) = (\frac{y}{\theta})^n \quad for\ 0 < y < \theta$$

So

$$f(y) = \frac{ny^{n-1}}{\theta^n} \quad for\ 0 < y < \theta$$

$$E[y] = \int_0^\theta yf(y)dy = \frac{n}{\theta^n}\int_0^\theta y^n dy = \frac{n\theta^{n+1}}{\theta^n(n+1)} = \frac{n\theta}{n+1}$$

(iv) Clearly $\tilde{\theta} = \frac{(n+1)Y}{n}$ is an unbiased estimator of θ
Its variance is $(\frac{n+1}{n})^2 V[Y]$ Now

$$E[Y^2] = \frac{n}{\theta^n}\int_0^\theta y^{n+1}dy = \frac{n\theta^2}{n+2}$$

so that

$$V[Y] = \frac{n\theta^2}{n+2} - \frac{(n\theta)^2}{(n+1)^2} = \frac{\theta^2(n(n+1)^2 - n^2(n+2))}{(n+2)(n+1)^2} = \frac{n\theta^2}{(n+2)(n+1)^2}$$

and

$$V[\tilde{\theta}] = (\frac{n+1}{n})^2 \cdot \frac{n\theta^2}{(n+2)(n+1)^2} = \frac{\theta^2}{n(n+2)}$$

(V(Y) was given in this question-but is easy enough to find!)
$V(\tilde{\theta_1})/V(\tilde{\theta}) = \frac{n+2}{3}$ This is the relative efficiency.

2 Suppose that the null hypothesis $H_0 : \theta = \theta_0$ is to be tested against the alternative $H_1 : \theta = \theta_1$, and that a random set of observations $(x_1, x_2, x_3, \cdots, x_n)$ is available, The likelihood functions on the two hypotheses are $L_n(\theta_0), L_n(\theta_1)$ and the ratio $\lambda_n = L_n(\theta_0)/L_n(\theta_1)$

The sequential probability ratio test continues sampling more observations so long as $A < \lambda_n < B$, for suitably chosen constants A, B with $A < B$. It stops and accepts H_0 if $\lambda_n \geq B$, and stops and accepts H_1 if $\lambda_n \leq A$. If $\alpha,\ \beta$ are types I and II Errors respectively, the approximations to the stopping boundaries are $A \doteq \frac{\alpha}{1-\beta}\ \ and\ \ B \doteq \frac{1-\alpha}{\beta}$

(i)

$$L_n(\theta) = \frac{1}{\{ln(\theta)\}^n \prod_{i=1}^n x_i} \cdot (\frac{\theta - 1}{\theta})^{\sum_{i=1}^n x_i}, \quad for\ \theta > 1,$$

$$\lambda_n(\theta) = \frac{L_n(2)}{L_n(4)} = (\frac{ln4}{ln2})^n (\frac{1/2}{3/4})^{\sum x_i} = 2^n (\frac{2}{3})^{\sum x_i}$$

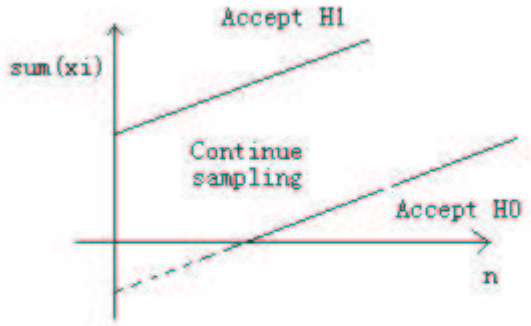
$A \doteq \frac{0.01}{0.99} = \frac{1}{99}$ and $B \doteq \frac{0.99}{0.01} = 99$ continue sampling which $\frac{1}{99} < 2^n(\frac{2}{3})^{\sum x_i} < 99$ i.e.

$$-ln99 < nln2 - (\sum x_i)(ln\frac{2}{3}) < ln99, \quad or$$

$$-4.5951 < 0.6931n - 0.4055\sum x_i < 4.5951, \quad i.e.$$

$$0.709n - 11.333 < \sum x_i < 1.709n + 11.333$$

(ii) Stop and decide for H_0 if $\sum x_i \leq 1.709n - 11.333$; and stop and decide for H_1 if $\sum x_i$



With each new observation,plot $\sum x_i$ against n,and stop as soon as one of acceptance boundaries is reached on this graph.

(iii)

$$Let\ Z = ln(\frac{P_0(x_i)}{P_1(x_i)}) = ln(2) + \frac{2}{3}x_i,\ (i = 1, 2, \cdots, n)$$

$$E[X] = \sum_{k=1}^{\infty} \frac{1}{ln(\theta)}(\frac{\theta - 1}{\theta})^k = \frac{1}{ln(\theta)} \cdot \frac{\theta - 1}{\theta} \cdot \frac{1}{1 - \frac{\theta - 1}{\theta}} = \frac{\theta - 1}{ln\theta}$$

$E_0[Z_i] = ln2 + \frac{2}{3} \cdot \frac{1}{ln2}$ $for\ H_0;\ this\ is$ 1.6549 Expected sample size on H_0 is

$$E[n] = \frac{\alpha lnA + (1 - alpha)lnB}{E_0(Z_i)}\ which\ is\ \frac{-0.01ln99 + 0.99ln99}{0.6549} = 2.72$$

say n is about 3.

3 Merits include: ML estimators have good asymptotic properties-asymptotic unbias, asymptotic efficiency, asymptotic normality. They are invariant and are functions of the minimal sufficient statistics;
Subject to regularity conditions, they have minimum variance if there exists an estimator which satisfies the Gamer Rao lower bound, or if they are unbiased; computing methods such

as Newton-Raphson or scoring can find solutions to normal equations if necessary;
Censored samples can be handled, and variable sample sizes can be used in related studies.
The likelihood argument has intuitive logical appeal. limitations include: ML estimators may be biased, which can be serious in small sample;
They can be quite difficult to complete directly; Their distributions in small samples can be intractable; Therefore limits for estimates in these cases are not easy to obtain;
The normal equations can be hard to solve in the case of several parameters, or for some special distributions leg cauchy or those like $exp(-\sqrt{x-\theta})$-or in some pathological situations (of stein; Neyman and Scott); analytical solutions may not give proper answers, as in the uniform distribution.

(i)

$$L(\sigma) = \prod_{i=1}^{n}(\frac{1}{\sigma\sqrt{2\pi}}exp\{-\frac{1}{2}(\frac{x_i}{\sigma})^2\}) = (2\pi)^{-\frac{1}{2}n}\sigma^{-n}exp(-\frac{1}{2\sigma^2}\sum x_i^2), \quad \sigma > 0$$

$$lnL(\sigma) = -\frac{1}{2}ln(2\pi) - nln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2, \quad and\ so\ \frac{dlnL}{d\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum x_i^2$$

$$\frac{d^2(lnL)}{d\sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum x_i^2, \quad Solving \quad \frac{dlnL}{d\sigma} = 0$$

gives

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum x_i^2}; \quad \frac{d^2(lnL)}{d\sigma^2} < 0$$

for this value,so this is a maximum.

(ii) The "Fisher Information" $I(\sigma) = E(-\frac{d^2(lnL)}{d\sigma^2}) = -\frac{n}{\sigma^2} + \frac{3}{\sigma^4}E[\sum x_i^2]$
Now $E[x^2]V[x]+[E(x)]^2 = \sigma^2(+0)$ in this distribution. Hence $I(\sigma) = -\frac{n}{\sigma^2}+\frac{3}{\sigma^4}\cdot n\sigma^2 = \frac{2n}{\sigma^2}$
The asymptotic distribution of $\hat{\sigma}$ will be $N(\sigma, \frac{\sigma^2}{2n})$, and an approximate 90% confidence interval is $\hat{\sigma} \pm 1.645\hat{\sigma}/\sqrt{2n}$

(iii) In this problem, $\frac{x}{\sigma} \sim N(0,1)$ and so $\frac{x^2}{\sigma^2} \sim \chi^2_{(1)}$; Therefore by independence of χ^2 distributions $\sum_{i=1}^{n} \frac{x_i^2}{\sigma^2} \sim \chi^2_{(n)}$ This sum of squares is a function of σ whose distribution does not depend on σ, and so is a pivotal quantity.
Hence a 90% confidence interval for σ is $R_1 < \frac{\sum x_i^2}{\sigma^2} < R_2$ where R_1 and R_2 are the lower and upper 5% points of χ^2_n if $n = 12,\ R_1 = 5.23,\ R_2 = 21.03$, and the interval is

$$5.23 < \frac{0.46}{\sigma^2} < 21.03, \quad i.e.\ for\ \sigma\ it\ is\ (0.15, 0.30)$$

4 (i) On the given NH, $H_0 : \lambda_1 = \lambda_1, with\ AN\ H_1 : \lambda_1 \neq \lambda_2$ the likelihood function is

$$L(\lambda_1, \lambda_2) = \prod_{i=1}^{m}(\lambda_1 e^{-\lambda_1 x_i}) \prod_{j=1}^{n}(\lambda_2 e^{-\lambda_2 y_j}), \quad (\lambda_1, \lambda_2 > 0)$$

which on H_1 is $\lambda_1^m \lambda_2^n e^{-\lambda_1 m\bar{x} - \lambda_2 n\bar{y}}$,where $\bar{x} = \frac{\sum x_i}{m}$, $\bar{y} = \frac{\sum y_j}{n}$ Then $lnL = mln\lambda_1 + nln\lambda_2 - \lambda_1 m\bar{x} - \lambda_2 n\bar{y}$,and

$$\frac{\partial}{\partial \lambda_1}(lnL) = \frac{m}{\lambda_1} - m\bar{x};\ \frac{\partial}{\partial \lambda_2}(lnL) = \frac{n}{\lambda_2} - n\bar{y}$$

Setting these derivatives equal to 0 gives $\hat{\lambda}_1 = 1/\bar{x},\ \hat{\lambda}_2 = 1/\bar{y}$. On H_0 L simplifies to gives $lnL = (m+n)ln\lambda - \lambda(m\bar{x} + n\bar{y})$ where $\lambda_1 = \lambda_2 = \lambda$, thus

$$\frac{\partial}{\partial \lambda}(lnL) = \frac{m+n}{\lambda} - (m\bar{x} + n\bar{y}) \ \ and \ \ \hat{\lambda} = \frac{m+n}{m\bar{x} + n\bar{y}}$$

The likelihood ratio statistic L_0/L_1 is

$$\wedge(x,y) = (\frac{m+n}{m\bar{x} + n\bar{y}})^{m+n} e^{-(m+n)} / (\bar{x}^{-m}\bar{y}^{-n}e^{-(m+n)}) = (\frac{\bar{y}}{\bar{x}})^n (\frac{m+n}{m + n\bar{y}/\bar{x}})^{m+n}$$

(ii) Using the given result that $M_x(t) = \frac{\lambda}{\lambda - t}$, and $m\bar{x} = \sum x_i$, we have

$$M_{m\bar{x}}(t) = \prod_{i=1}^{m} M_{x_i}(t) = (\frac{\lambda_1}{\lambda_1 - t})^m,\ for\ t < \lambda_1$$

and for $2\lambda_1 m\bar{x}$ the mgf is $M_{m\bar{x}}(2\lambda_1 t) = (1-2t)^{-m},\ t < 1/2$ By the uniqueness property of mgf's, it follows that $2\lambda_1 m\bar{x} \sim \chi^2_{(2m)}$; similarly $2\lambda_2 n\bar{y} \sim \chi^2_{(2n)}$ and thus

$$\frac{2\lambda_2 n\bar{y}/2n}{2\lambda_1 m\bar{x}/2m} \sim F_{(2n,2m)};\ \ i.e.\ under\ H_0\ \frac{\bar{y}}{\bar{x}} \sim F_{(2n,2m)}$$

(iii) The generalised likelihood ratio test has critical region $\{x : \bigwedge(x) \leq K\}$ for some K; and so the test of size will reject H_0 if $\bar{y}/\bar{x} \leq R_1$ *or* $\bar{y}/\bar{x} \geq R_2$ where R_1 R_2 are respectively the lower and upper $\frac{\alpha}{2}\%$ points of $F_{(2n,2m)}$. When $m = 37, n = 39$ and $\alpha = 0.05, R_1 = 0.6, R_2 = 1.6$ But $\bar{y}/\bar{x} = 0.9$, and so we cannot reject H_0 at the 5% level.

5 (i) $\hat{p} = \frac{180}{250} = 0.72$. The number killed , r is Bin(n,p) and for sufficiently larger, and not too extreme a value of p, this is approximated by $N(np, np(1-p))$; hence the proportion r/n is $N(p, \frac{p(1-p)}{n})$ and a 95% confidence interval for the true p based on the sample proportion is given as

$$P(-1.96 < \frac{\hat{p} - p}{\sqrt{p(1-p)/250}}) < 1.96 = 0.96$$

in which we must use $\hat{p}$ when estimating the standard error $\sqrt{p(1-p)/250}$. Hence the interval is

$$\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{250}} < p < \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{250}}$$

This approximate 95% interval has limits

$$0.72 \pm 1.96\sqrt{\frac{0.72 \times 0.28}{250}} = 0.72 \pm 0.056,\ \ i.e.(0.664, 0.776)$$

(ii) The prior distribution of p is $\pi(p) = 1, 0 < p < 1$ Also $R|P$ is Binomial (250,p). So the posterior distribution of p is

$$\Pi(p|R = 180) \propto 1 \cdot \begin{pmatrix} 250 \\ 180 \end{pmatrix} p^{180}(1-p)^{70} \propto p^{180}(1-p)^{70}, 0 < p < 1$$

This makes $p|R = 180 \;\; Beta(181, 71), i.e.$

$$\Pi(p|R = 180) = \frac{\Gamma(252)}{\Gamma(181)\Gamma(71)} p^{180}(1-p)^{70}, \;\; 0 < p < 1$$

(iii) $E[p|R = 180] = \frac{181}{252}$ *and* $V[p|R = 180] = \frac{181 \times 71}{252^2 \times 253}$; and so an approximate 95% confidence interval using normal theory will be

$$\frac{181}{252} \pm 1.96\sqrt{\frac{181 \times 71}{252^2 \times 253}} \;\; or \; 0.718 \pm 0.055, i.e. (0.663; 0.773)$$

(iv) The number killed out of is Binomial (4,p) and so $P(3) = 4p^3(1-p)$ (NOTE-for interest only-if we set p=0.72 this is 0.418)
The Bayesian prediction of probability is

$$E[4p^3(1-p)] = \int_0^1 4p^3(1-p)\frac{\Gamma(252)}{\Gamma(151)\Gamma(71)} p^{180}(1-p)^{70} dp$$

$$= 4\int_0^1 \frac{\Gamma(252)}{\Gamma(181)\Gamma(71)} p^{183}(1-p)^{71} dp$$

$$= \frac{4 \times 183 \times 182 \times 181 \times 71}{255 \times 254 \times 253 \times 252} = 0.4146$$

6 Suppose that independent random sample are available from two populations whose distributions are of the sample, although the family of distributions is not known (and will usually not be supposed symmetrical, so that a normal model would not be appropriate). The Mann-Whitney test compares the location parameters (e.g.medians)in this situation.
Given two sample A and B, compare each member of A in turn with each member of B; U_{AB} is the number of pairs in which the A-value is less than the B-value. If A contains in items and B contains n items,the number if different ways in which the A's and B's can be ordered in the combined sample of size(m+n)is $\begin{pmatrix} m+n \\ m \end{pmatrix}$. Under the Null Hypothesis H_0, that the location parameters of A and B are equal, each of these ways is equally likely. The observed value, u, of U_{AB} has expectation $\frac{1}{2}mn$; for $u \leq \frac{1}{2}mn$, its one-sided significance is the probability of U_{AB} *being* $\leq u$ *on* H_0, i.e. the number of orderings of the A's and B's such that U_{AB}, divided by $\begin{pmatrix} m+n \\ m \end{pmatrix}$, This is found by direct enumeration.(This significance has to be doubled for 2-sided Alternative Hypothesis)
The value in Table XIV are the largest values of w such that $P(U_{AB} < w)$ under H_0 is $\leq$ the given value of p.m=12, n=7 and the ordered combined data are:

70	83	85	94	97	101	104	107	107	113	118	119	123	124	129	132	134	146	161
B	*A*	*B*	*B*	*A*	*B*	*A*	*B*	*A*	*A*	*B*	*A*	*A*	*A*	*A*	*B*	*A*	*A*	*A*

with A denoting the high-protein weights.

$$U_{AB} = 0 + 1 + 1 + 2 + 3.5 + 5 + 9 = 21.5$$

Considering each B in order and counting A's below it .The tie between an A and a B leads to the value 3.5. The critical region at the 5% level for a 2-sided alterative is $U_{AB} < 19$,and at 10% is $U_{AB} < 22$. Hence there is evidence against H_0 at a probability level somewhere between 5% and 10% (nearer to 10% than 5%).

7 A decision rule is a function from the sample space to the action space. The risk of a decision rule δ at parameter value θ, $R_\delta(\theta)$, is the expected loss. A decision rule is called minimax if its risk R^* satisfies:

$$\sup_\theta R^*(\theta) \leq \sup_\theta R_\delta(\theta) \quad for\ all\ \delta \in D$$

Where D is the decision space.

(i)

$$P(X \geq k) = \sum_{j=k}^{\infty}(1-p)p^j = (1-p)p^k(1 + P + p^2 + \cdots) = (1-p)p^k\frac{1}{1-p} = p^k$$

The risk $R_k(\delta)$ of δ_k is

$$R_k(p) = \{ \begin{array}{ll} 4P(X < k) = 4(1-p^k) & for\ p > 0.7 \\ -8P(X < k) + 2P(X \geq k) = -8 + 10p^k & for\ p \geq 0.7 \end{array}$$

For $P > 0.7$,the maximum rise is $4(1 - 0.7^k)$since the risk is an increasing function of k. For $P \leq 0.7$, $-8+10p^k$ decreases as k increases and so the maximum risk is $-8+10(0.7^k)$. Thus the maximum risk are:

δ_0	δ_1	δ_2
2	1.2	2.04
(B)	(A)	(A)

so the minimax is δ_1.

(ii) The Bayes Risk

$$B(k) = \int_0^1 R_k(p)dp = \int_0^{0.7}(-8 + 10p^k)dp + \int_{0.7}^1 4(1-p^k)dp$$

$$= [-8p + \frac{10p^{k+1}}{k+1}]_0^{0.7} + [4(p - \frac{p^{k+1}}{k+1})]_{0.7}^1$$

$$= -5.6 + \frac{10(0.7)^{k+1}}{k+1} + 4 - \frac{4}{k+1} - 2.8 + \frac{4(0.7)^{k+1}}{k+1}$$

$$= -4.4 - \frac{4}{k+1} + \frac{14(0.7)^{k+1}}{k+1} \quad (k = 0, 1, 2, \cdots)$$

(iii)

$$B(k+1) - B(k) = \frac{-4}{k+2} + \frac{4}{k+1} + \frac{14(0.7)^{k+2}}{k+2} - \frac{14(0.7)^{k+1}}{k+1}$$

$$= \frac{4}{(k+2)(k+1)} + \frac{14(0.7)^{k+1}}{(k+1)(k+2)}\{4 - 14(0.7)^{k+1}(0.3k + 1.3)\} < 0$$

if and only if $14(0.7)^{k+1}(0.3k + 1.3) > 4$

Now $(0.7)^{k+1}(0.3k + 1.3)$ decreases as k increases, as may be shown by computing the first few value; and for k=5, 6 we have $(0.7)^6(0.3 \times 5 + 1.3) = 0.33$ *which is* $> \frac{4}{14}(0.286)$, while $(0.7)^7(0.3 \times 6 + 1.3) = 0.255$ *which is* $< \frac{4}{14}$, showing that $B(5) > B(6)$; but also from the second inequality which is violated we see $B(7) > B(6)$. Hence δ_6 achieves the smallest Bayes Risk.

8 For simple hypothesis $H_0 : \theta = \theta_0(N.H.)$ *and* $H_1 : \theta = \theta_1(A.H.)$ in a probability density (or mass)function $f(x, \theta)$, the Neyman-Pearson method uses the likelihood ratio test of the form

$$c = \{x : \frac{f(x, \theta_0)}{f(x, \theta_1)} \leq k\}$$

for some suitable k. H_0 is rejected when the AH gives the better explanation, i.e. k is small; the size of k is chosen to make the probability of falling in the critical region when $\theta = \theta_0$ equal to the specified significance level.

(i)

$$L(v) = \prod_{i=1}^{n} e^{-\lambda v^i}(\lambda v^i)^{x^i}/x_i! = exp(-\lambda \sum_{i=1}^{n} v^i)\lambda^{\sum_{i=1}^{n} x_i} v^{\sum_{i=1}^{n} i x_i} / \sum_{i=1}^{n} x_i! \quad v > 0$$

$$\bigwedge = \frac{L(1)}{L(2)} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{e^{-\lambda \sum 2^i}\lambda^{\sum x_i} 2^{\sum i x_i}} = \frac{e^{-n\lambda}}{[e^{-\lambda(2^{n+1}-2)} 2^{\sum i x_i}]}$$

Hence the critical region is $\{x : \sum_{i=1}^{n} i x_i \geq k'\}$ since $\bigwedge$ decreases as $\sum i x_i$ increases.

(ii) Write $Y = \sum_{i=1}^{n} i x_i$; under H_0 $E[Y] = \lambda \sum i = \frac{1}{2}\lambda n(n+1)$ and $V[Y] = \lambda \sum i^2 = \frac{1}{6}\lambda n(n+1)(2n+1)$. Calling these μ and σ^2, and using a continuity correction,

$$P(Y \geq k' | v = 1) \doteq \Phi(\frac{k' - 0.5 - \mu}{\sigma})$$

is the significance level.Hence k' is chosen by equating this to α.

(iii) For $\lambda = \frac{1}{3}$ and $n = 2$ observations, on the NH both x, and x_2 are Poisson (1/3).

$P(x_1 + 2x_2 > 2) = P(x_1 > 2) + P(x_1 = 1 \text{ or } 2)P(x_2 \geq 1) + P(x_1 = 0)P(x_2 \geq 2) =$
$1 - 0.7165 - 0.2388 - 0.0398 + (0.2388 + 0.0398)(1 - 0.7165) + 0.7165(1 - 0.7165 - 0.2388) =$

0.1156
This is the significance level.
On the AH,x_1 is Poisson(2/3) and x_2 is Poisson(4/3).The power is $P(x_1 + 2x_2 > 2) = 1-0.5134-0.3423-0.1141+(0.3423+0.1141)(1-0.2636)+0.5134(1-0.2636-0.3515 = 0.5639)$

Applied Statistics I

1 (i) $\alpha + 2\beta = \pi$. Assume that we measure A n times, and B and c altogether $12 - n$ times. Then

$$A_i = \alpha + \varepsilon_i \ (i = 1 \ to \ n), \quad E[\varepsilon_i] = 0, \ V[\varepsilon_i] = \sigma^2$$

and

$$B_j = \beta + \varepsilon_j \ (j = 1 \ to \ 12 - n), \quad E[\varepsilon_j] = 0, \ V[\varepsilon_j] = \sigma^2.$$

Minimize

$$\sum_{i=1}^{n} \varepsilon_i^2 + \sum_{j=1}^{12-n} \varepsilon_j^2, i.e. \ \sum_{i=1}^{n} (A_i - \alpha)^2 + \sum_{j=1}^{12-n} (B_j - \beta)^2$$

Using the constraint on $\alpha, \ \beta$ this is simplifies, to

$$S = \sum_{i=1}^{n} (A_i - \pi + 2\beta)^2 + \sum_{j=1}^{12-n} (B_j - \beta)^2$$

$$\frac{dS}{d\beta} = 4\sum_{i=1}^{n} (A_i - \pi + 2\beta) - 2\sum_{j=1}^{12-n} (B_j - \beta) = 0 \ for \ a \ minimum(or \ maximum)i.e.$$

$$2\sum_{i=1}^{n} A_i - 2n\pi + 4n\hat{\beta} - \sum_{j=1}^{12-n} B_j + (12 - n)\hat{\beta} = 0 \ or$$

$$2\sum_{i=1}^{n} A_i - 2n\pi - \sum_{j=1}^{12-n} B_j = (n - 12 - 4n)\hat{\beta}.Hence$$

$$\hat{\beta} = (2n\pi - 2\sum_{i=1}^{n} A_i + \sum_{j=1}^{12-n} B_j)/(3n + 12). \ Also \ \hat{\alpha} = \pi - 2\hat{\beta}$$

(ii) Now by independence,$Var[\sum A_i] = n\sigma^2$ and $Var[\sum B_j] = (12 - n)\sigma^2$ so

$$Var[\hat{\beta}] = \frac{4n\sigma^2 + (12 - n)\sigma^2}{(3n + 12)\sigma^2} = \frac{\sigma^2}{3n + 12}.$$

$$E[\hat{\beta}] = \frac{2n\pi - 2n\alpha + (12 - n)\beta}{3n + 12} = \frac{2n\pi - 2n(\pi - 2\beta) + (12 - n)\beta}{3n + 12} = \beta.$$

Hence

$$E[\hat{\alpha}] = \pi - 2E[\hat{\beta}] = \pi - 2\beta = \alpha$$

$$V[\hat{\alpha}] = 4V[\hat{\beta}] = \frac{4\alpha^2}{3n+12}$$

both $\hat{\alpha}$ and $\hat{\beta}$ are unbiased. By method P:

$$n = 4,\ so\ \hat{\beta} = \frac{1}{24}(8\pi - 2\sum_{i=1}^{4} A_i + \sum_{j=1}^{8} B_j);$$

$$\hat{\alpha} = \pi - 2\hat{\beta},\ V[\hat{\beta}] = \frac{\sigma^2}{24},\ and\ V[\hat{\alpha}] = \frac{\sigma^2}{6}$$

Using Q:

$$n = 6,\ so\ \hat{\beta} = \frac{1}{30}(12\pi - 2\sum_{i=1}^{6} + \sum_{j=1}^{6} B_j);$$

$$\hat{\alpha} = \pi - 2\beta,\ V[\hat{\beta}] = \frac{\sigma^2}{30},\ and\ V[\hat{\alpha}] = \frac{40^2}{30} = \frac{2\sigma^2}{15}$$

Q gives smaller variances.

(iii) If possible,minimum variance unbiased estimators are required. Choose n to minimize the variances.

2 (a) Strict stationarity is when the joint distribution of $X(t_1), \cdots, X(t_n)$ is the same as that for $X(t_1+\tau), \cdots, X(t_n+\tau)$ for all $t_1, \cdots, t_n$ and τ.

Weak stationarity has $E[X(t)]$ constant and the autocovariance for $X(t)$ and $X(t+\tau)$ depends only on the lag separation τ.
In practice, we mean a series with no trend, no seasonal effects and no cyclical changes, but only irregular fluctuations.

(b) Series 1. This is not stationary,but there is strong evidence of seasonality with large peaks at $12, 24, \cdots$ A model requires a seasonal term. Seasonal difference should lead to stationarity.
Series 2. This appears to be stationarity,since there is no obvious pattern,so a model need only contain 'white noise'.
Series 3. This is not stationary;there is strong evidence of a trend, the autocorrelation function is decaying slowly, and first differences will be useful.
Series 4. The autocorrelation function dies away quickly, suggesting an AR model. The

partial autocorrelation function cuts off at 1, or possibly 2, which suggests AR(1), or AR(2).
Series 5.This is not clear;neither function dies away quickly so ARMA(1,1) may be a possibility but would need investigating.

3 (a) (1) Leverage is the potential for influencing parameter estimates when a point is in a relatively extreme position in the x-space.

(ii) $H = X(X'X)^{-1}X'$ .In a model with s parameters, and with n data points, the diagonal elements of H may be compared with 2s/n, and values higher than this correspond to points with high leverage.

(b) (i) OD620 and OD740 have a strong positive linear relation, and for OD740 values less than $25(OD620 < 50)$ there is very little scatter about a line. Variability about the same line remains low throughout the range. However the data are heavily concentrate in the lower region of the graph, with relatively few values in the middle and just one high value.

(ii) If the linear regression is a good model, these residuals should show a "normally distributed" scatter about 0. The plot shows no real evidence to the contrary. However, the largest residuals are at large values of OD so there may be doubt whether the variance remains constant. Neither diagram throws doubt on linearity.

(iii) Point 33 was identified as having a large residual and also (because of its position) influence on parameter estimates. Without it, the line has a large intercept and lightly smaller gradient.R^2 is slightly less became the point removed made a large contribution to the original total sum of squares. The residual mean square is about 10% lower. These effects are because the point removed was above the original fitted line and had a high residual.
Also without the high-leverage point and its influence on the gradient,the new gradient has a larger standard error.
Some new 'unusual' points have appeared due either to their distance from other points to the change in position of the fitted line and the reduction in residual mean square.
The new model predicts most of the OD620 points better, and with more precision,but it should not be used for extrapolation to higher values as the these are now on the whole less satisfaction predicted. The new model is probably preferred.

4 (i) Plant k is a subsample from pot j receiving treatment i, the pots and the plants are all different for each treatment, and only the treatment effect may be called "fixed". Asmitable model is

$$y_{ijk} = \mu + \alpha_i + \beta_{(i)j} + \varepsilon_{(ij)k}$$

where μ is an overall mean effect
α_i is the effect of treatment i $(i = 1\ to\ 6)$; $\sum_{i=1}^{6} \alpha_i = 0$, $\beta_{(i)j}$ is the effect of the j^M unit receiving treatment i $(j = 1\ to\ 3)$ and $\varepsilon_{(ij)k}$ is the effect of the k^M subunit from unit$j$ /treatment$i$ $(k = 1\ to\ 4)$; $\beta_{(i)j}$ and $\varepsilon_{(ij)k}$ are mutually independent.

(ii) $G = 416.5,\ N = 72,\ G^2/N = 2409.3368.$
Total corrected sum of squares is therefore $2665.25 - \frac{G^2}{N} = 255.9132$ s.s. For treatments $= \frac{1}{12}(44.0^2 + \cdots + 95.0^2) - \frac{G^2}{N} = 179.6424.$ s.s. between all pots= $\frac{1}{4}(15.0^2 + 18.0^2 + \cdots + 29.0^2 + 35.0^2) = 2614.8125.$ so the(corrected), s.s. for pots within treatments is $2614.8125 - 2409.3368 - 179.6424 = 25.8333$. The Analysis of Variance becomes:

SOURCE OF VARIATION	*D.F.*	*SUM OF SQUARES*	*MEAN SQARE*
Between Treatments	5	179.64	35.93
Between pots	12	25.83	2.15
within treatments residual	54	50.44	0.934
Total	71	255.91	

(iii) (1) Null Hypothesis: all treatment effects α_i are 0, i.e. all treatment give the same mean μ. Alternative:not all α are zero. $F_{(5,12)}$ test this: $\frac{35.93}{2.15} = 16.71 * **$ so we reject the NH.

(2) Null Hypothesis: $\sigma_b^2 = 0$. Alternative: $\sigma_b^2 \neq 0,\ F_{(12,54)} = 2.30*$, so there is evidence that variability between pots is greater than that within.

(iv) The 5df for treatments could be split into individual contrasts,(1 to 3)versus(4 to 6) for low versus high temperature(1df), then linear and quadratic components of time within each temperature, 2df(low)+2df(high).
The constancy of variance over all the pot/plant/treatment groups might be examined.

5 (i) There are very few data. But there is an indication that house increases as condition improves. As a first approximation three straight lines might be fitted to relate price and floor area at the three conditions.

(ii) $y_{ij} = \alpha_i + \beta_i x_{ij} + \epsilon_{ij}$

$$y_{ij} = selling\ price,\ \alpha_i = intercept\ for\ condition\ i (i = 1, 2, 3),$$

$$\beta_i = slope\ for\ condition\ i, x_{ij} = floor\ area\ of\ j^M\ house\ at\ condition\ i$$

$$(j = 1\ to\ 3, for\ i = 1\ and\ 3;\ j = 1\ to\ 5\ for\ i = 2),$$

$$\epsilon_{ij} = random\ term\ of\ mean\ zero\ and\ variance\ \sigma^2 (all\ i,\ j)$$

(iii) The analysis of variance for price shows that the interaction term for condition and area is negligible.When it is removed (lower half of page 8)both main effects are very highly significant.

Leaving out either one of these increases the residual mean square substantially area even more so than condition.No interaction implies that three parallel lines could be used.

(iv) The model therefore becomes $y_{ij} = \alpha_i + \beta_i x_{ij} + \epsilon_{ij}$. If , S_{xy} represents the sum of products XY under condition 1, and, S_{xx} the corresponding sum of squares for x, the common slope β is found as $({}_1S_{xy} +_2 S_{xy} +_3 S_{xy})/({}_1S_{xx} +_2 S_{xx} +_3 S_{xx})$. Then α_i is estimated as $\bar{y}_i - \hat{\beta}\bar{x}_i$. Different statistical packages show this in different ways.

(v) Residual diagnosis and normal plots would be useful to homoscedasticity, normality and independence (but in this case we have very few data, so the plot of residuals against fitted values is likely to be the best).

6 (i) The given model is correct (lacks no important terms); x_{ij} fixed;$E[\epsilon] = 0,\ Var[\epsilon] = \sigma^2$, constants; all $\{e_{ij}\}$ are independent.Also ,in analysis, $\{\epsilon_{ij}\}$ are normally distributed.

(ii)

$$Y = X\beta + \epsilon,\ where\ Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix},\ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_p \end{pmatrix},\ \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdots \\ \epsilon_n \end{pmatrix},\ and$$

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{pmatrix},\ \hat{\beta} = (X'X)^{-1}X'Y$$

(iii) $(X'X)_{-1}$ may be very unstable, near-singular if two or more $x's$ are highly correlated. Studying the scatter grams for all the pairs of x-variables shows whether any are very highly correlated; if so ,one of the pair should be omitted. Sometimes principal components of the x's can be used instead.

(iv) R^2 increases(or stays the same) when the number of predictor variables increases; adjusted R^2 allows for the number of variables.

(v) Residuals $\hat{\epsilon}_i$ are $y_i - \hat{y}_i$; $\sum_i \hat{\epsilon}_i^2$ measures lack of fit of the model. They can be studied for evidence of failure of assumptions(i), and any pattern in the residuals can indicate terms to be added, or the need for a variance-stabilizing transformation, or possible lack of normality(symmetry). $\hat{\epsilon}_i$ may be plotted against y_i, or against x's , or against any other x's that might be suitable to include in the model, or against the order of observations in a time-series.

(vi) Various packages contain the following items for this purpose:
the hat-matrix elements $\{h_i\}$
student residuals; student deleted residuals
Cook's distance statistic
affits; dfbetas; covratio.

7 (a) Grand Total G=1196.3, N=$\sum n = 20$. $G^2/N = 71556.6845$. Hence Total ss=120.3055. s.s. for laboratories=$\frac{372.1^2}{6} + \frac{298.3^2}{5} + \frac{243.3^2}{4} - G^2/N = 87.5697$.

$SOURCE$	DF	SS	MS	
$Laboratories$	3	87.6597	29.19	$F_{(3,16)} = 14.27$
$Residual$	16	32.7358	2.046	
$Total$	19	120.3055		

The F-value give strong evidence of difference between laboratories. Means are:

$LAB3$	1	4	2
56.52	59.66	60.83	62.02

The data suggest that 3 gives lowest, and 2 highest, results; but this would need confirmation in further trials.(see iv).

(i) This model is suitable when only these four laboratories are being compared, not if they had been a random sample from a larger set.

(ii) If they were a sample from a lager set (population)the term for laboratories in the linear model which is the basis for analysis would be assumed $N(0, \sigma_l^2)$. The same analysis would be used , but we would conclude that $\sigma_l^2 \neq 0$, i.e. there is variation among laboratories carrying out these analysis.

(iii) If these were particular comparisons(contrasts)among the four given laboratories that were interesting in a fixed effect model we would us t-tests to make these comparisons. The contrasts must be planned before the analysis is done, not merely suggested by the data. For a random-effects model the only further calculation might be approximate confidence limits for σ_l^2.

(b) A variance-stabilizing transformation would be $\sqrt{y}$. Otherwise the necessary assumption of constant variance is violated. If the variance is stabilized, this does not guarantee (approximate) normality though it often improves the distributional assumption's validity. AGLM log-linear model would probably be better for making the distributional assumptions.

8 (i) The first 5 seem similar; so do the next 3(French, Spanish, Italian); Polish seems to stand by itself; The last 2 have some similarity.

(ii) The concordances are not fully standardized to be in (0,1), and so each entry is divided by 10 to give a proper measure of similarity.

(iii) Distance measures must satisfy:

$$d(X, Y) \geq 0, \ \textit{with equality when } X = Y;$$

$$d(X, Y) = d(Y, X); \ d(X, Z) + d(Y, Z) \geq d(X, Y).$$

The figures in Table 1.3 satisfy these conditions. Clustering proceeds by successively combining sets of individuals into groups; at each stage individuals or sets which are 'closest' are combined. Starting from the distance matrix, methods vary according to how we define distance between an individual and an existing grouping. Single-link clustering is attaching a new point to that group which contains the nearest point to it. Complete clustering attaches it to that group where the furthest existing member is as near as possible.

(iv) Euclidean(geometric)distance is the usual graphical idea: in two dimensions

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Single linkage begins with two groups at the same distance:
$N,\ D_a$ and $F_r,\ S_p,\ I$; next add to 1^{st} group E; then add to second group P; then G to 1^{st} group. Then these two groups are merged and D_a added. Complete linkage takes as the first two groups $N,\ D_a$ and $S_p,\ I$ to which add E and F_r at the same distance. Then form a new group $D_u,\ G$ and add P to $F_r,\ S_p,\ I$. Later $N,\ D_a,\ E$ and $F_r,\ S_p,\ I,\ P$ can be combined, then $H,\ F_i$ can be combined and finally these are added to $Du,\ G$. Average linkage (center of gravity) goes up to adding P to $Fr,\ Sp,\ I$ just as complete links but $Du,\ G$ only happens after this. This time $Du,\ G$ combines with $E,\ N,\ Da$ instead of $H,\ Fi$.
Note that we use only the first letter to judge similarity; it would be better to consider spelling (and sound). Hence the method is rather rough-and-ready, but it does indicate that $E,\ N,\ Da$ go together and so do $F_r,\ S_p,\ I$. The position of $D_u,\ G$ is arguable, and P is nearer to $Fr,\ Sp,\ I$ than first glance suggested. $H,\ Fi$ can be grouped rather weakly.

Applied Statistics II

1 (a) When v treatments are to be compared, and two systematic sources of variation exist among the units (plots) available, a Latin square design is one way of removing these sources. The design is in v rows and v columns, arranged so that each treatment occurs once in each row and once in each column.
A $B \times B$ square is

$$\begin{array}{ccc} A & C & B \\ B & A & C \\ C & B & A \end{array}$$

and a 4×4 is

$$\begin{array}{cccc} A & D & B & C \\ C & A & D & B \\ B & C & A & D \\ D & B & C & A \end{array}$$

There are v^2 units. Therefore when $v = Bor4$ we have a design that will usually be too small to give a precise experiment. As in the following example, more than one complete square will need to be used.

(b) The systematic variation between days is removed in rows, and that between rabbits in columns. The treatments may be split into a factorial scheme of main effects (preparations, doses) and interaction.

(i) Treatment totals are; $A,\ 486;\ B,\ 358;\ C,\ 480;\ D,\ 374$. $G = 1698$. $N = 32$. Correction term $G^2/N = 90100.125$.
Treatment s.s. $= \frac{1}{8}(486^2 + 358^2 + 480^2 + 374^2) - G^2/N = 1729.375$
Corrected total s.s. $= 6417.875$. s.s.Rows$= \frac{381^2+\cdots+410^2}{8} - G^2/N = 599.625$.
s.s.Columns $= \frac{1}{4}(160^2 + \cdots + 203^2) - G^2/N = 3244.375$.
Assuming that the two squares need not be separated, i.e. the rabbits all come from

the same population, we have:

$SOURCE$ $OF\ VARIATION$	DF	$SUM\ OF$ $SQUARES$	$MEAN$ $SQURE$	
$ROWS(DAYS)$	3	599.625	199.875	$F_{(3,18)} = 4.26$
$Columns(Rabbits)$	7	3244.375	463.482	$F_{(7,18)} = 9.88$
$Treatments(Doses)$	3	1729.375	576.458	
$Residual$	18	844.500	$46.9167 = s^2$	
$Total$	31	6417.875		

Clearly there is a day effect and a difference between rabbits.

(ii) The appropriate factorial set is :

$$(A, B)\ versus\ (C, D) - standard/test$$
$$(A, C)\ versus\ (B, D) - dose\ leves$$
$$(A, D)\ versus\ (B, C) - interaction$$

(iii) They may be computed as contrasts:

	A	B	C	D
$Preparations$	-1	-1	1	1
$Levels$	-1	1	-1	1
$Interaction$	-1	1	1	-1

or as sums of squares:

$$(A, B)\ vs\ (C, D) : \frac{844^2 + 854^2}{16} - \frac{1698^2}{32} = 3.125$$
$$(A, C)\ vs\ (B, D) : \frac{966^2 + 732^2}{16} - \frac{1698^2}{32} = 1771.125$$
$$(A, D)\ vs\ (B, C) : \frac{860^2 + 888^2}{16} - \frac{G^2}{N} = 15.125$$

Each has 1 d.f. Only the dose levels are significant(use $F_{(1,18)}$). The blood sugar is higher (0.1% level) at 0.6 than at 1.2.
Summary: there are differences between days, between rabbits and between doses, but not between preparations. Doses and preparations do not interact. (preparations/doses) do not interact.

2 (i) The eight treatment combinations used all had an odd number(1,3 or 5)of letters in common with ACE and with BDE ; ABCD is the generalized interaction of these two and therefore also appears in the defining contrast.

(ii) The alias structure is $I = ACE = BDE = ABCD$ and therefore

$$A = CE = ABDE = BCD$$
$$B = ABCE = DE = ACD$$
$$C = AE = BCDE = ABD$$
$$D = ACDE = BE = ABC$$
$$E = AC = BD = ABCDE$$
$$AB = BCE = ADE = CD$$
$$AD = CDE = ABE = BC$$

(iii) Labelling each alias group by its lead number in the list above, we compute from the eight treatment combinations used as follows:

	e	ad	bde	ab	cd	ace	bc	$abcde$	$value$	$effect\ estimate$
A	−	+	−	+	−	+	−	+	2.4	0.60
B	−	−	+	+	−	−	+	+	19.6	4.90
C	−	−	−	−	+	+	+	+	6.6	1.65
D	−	+	+	−	+	−	−	+	7.4	1.85
E	+	−	+	−	−	+	−	+	8.8	2.20
AB	+	−	−	+	+	−	−	+	−12.2	−3.05
AD	+	+	−	−	−	−	+	+	4.0	1.00
	8.7	12.0	17.5	11.0	9.0	13.0	16.1	17.7		

A is the difference between all with a and all without,etc. AB is the "interaction" or inner product of the A and B rows. With apparently, no proper replication no estimate of residual variability is possible and hence no tests of significance.

(iv) It is invalid to construct a residual by inspection of the data. Assumptions need to be made about which interactions may be neglected, usually only higher order ones unless some others have featured in earlier experiments. The aim of a properly designed experiment should be to estimate all main effects and lower order interactions(at least the 2-factor ones) about which information is not already available.

In a very small design such as a fractional factorial only large effects(the B and AB alias sets here) will be detectable, and there are no alias sets that contain only high order interactions. This design by itself can tell us very little.

3 (i) The score y_{ij} consists of a general average score μ plus a departure from this α_i representing the effect of the treatment given to group i plus a random natural variation term ϵ_{ij} for each individual.

(ii) The analysis proceeds by estimating the parameters μ, α_i by the method of least squares. $\hat{\mu} = \bar{\bar{y}}$, the overall mean of the N observations, and $\hat{\alpha}_i = \bar{y}_i - \bar{\bar{y}}$, where $\bar{y}_i$ is the mean for group i. The residual sum of squares is:

$$\sum_{i,j} e_{ij}^2 = \sum_{j=A}^{C} \sum_{j=i}^{n_i} (y_{ij} - \hat{\mu} - \hat{\alpha}_i)^2 = \sum \sum [(y_{ij} - \bar{\bar{y}}) - (\bar{y}_i - \bar{\bar{y}})]^2$$

Hence:

$$E[RSS] = E\sum_{i=A}^{C}\sum_{j=1}^{n_i}[(\epsilon_{ij} - \bar{\bar{\epsilon}}) - (\bar{\epsilon}_i - \bar{\bar{\epsilon}})]^2$$
$$= \sum_i \sum_j E[(\epsilon_{ij} - \bar{\bar{\epsilon}})^2 + (\bar{\epsilon}_i - \bar{\bar{\epsilon}})^2 - 2(\epsilon_{ij} - \bar{\bar{\epsilon}})(\bar{\epsilon}_i - \bar{\bar{\epsilon}})]$$
$$= \sum_i \sum_j [Var(\epsilon_{ij} + Var(\bar{\epsilon}_i) - 2COV(\epsilon_{ij}, \bar{\epsilon}_i)]$$
$$= \sum_i \sum_j (\sigma^2 + \frac{\sigma^2}{n_i} - 2\frac{\sigma^2}{n_i})$$

since $E(\epsilon_{ij}) = 0$ for all i, j; $V[\epsilon_{ij}] = \sigma^2$ for all i, j; and the $\{\epsilon_{ij}\}$ are uncorrelated for all $i \neq j$. So

$$E[RSS] = \sigma^2 \sum_{i=A}^{C}\sum_{n_i}^{j=1}(1 - \frac{1}{n_i}) = \sigma^2 \sum_{i=A}^{C} n_i(1 - \frac{1}{n_i}) = \sigma^2(N - 3)$$

where $N = \sum_i n_i$.

(iii) The total sum of squares with $(N-1)$ degrees of freedom can be partitioned into components due to the differences between the 3 treatment groups, with 2 d.f., and the residual within all treatments, with $(N-3)$ d.f.
Cochraus Theorem for m i.i.d.$N(0,1)$ variables $\{Z_i\}$ says that if $\sum_{i=1}^{m} Z_i^2 = Q_1 + Q_2 + \cdots + Q_s$, where $\{Q_j\}$ are quadratic forms with $m_1, m_2, \cdots, m_s$ d.f. (and $s \leq m$) then the $\{Q_j\}$ are independent $\chi^2_{(m_j)}$ variables if and only if $m = \sum_{j=1}^{s} m_j$.
Here SS for treatments and the total SS has $(N-1)$ d.f.; so $\frac{SS\ treatments}{\sigma^2}$ and $\frac{residual\ SS}{\sigma^2}$ are independent χ^2. Assuming normality of residuals $\{\epsilon_{ij}\}$, with constant σ^2, the ratio $\frac{SST/2}{RSS/(N-3)} \sim F_{(2,N-3)}$.

(iv) The pooled estimate of σ^2 is $s^2 = \frac{1}{34}\{10(24.25)^2 + 12(24.07)^2 + 12(17.71)^2\} = 488.1392$ A 95% confidence interval for $\alpha_A - \alpha_B$ is $\bar{y}_A - \bar{y}_B \pm t_{(5\%,34)}\sqrt{s^2(\frac{1}{11} + \frac{1}{13})}$ i.e. $-2.2 \pm 2.034 \times 9.051$ or -2.2 ± 18.4 or -20.6 to $+16.2$. Thus A and B do not appear to differ in effect (since 0 is in the interval). The contrast $\frac{1}{2}(\alpha_A + \alpha_B) - \alpha_C$ is estimated by $\frac{1}{2}(\bar{y}_A + \bar{y}_B) - \bar{y}_C$ and so has variance $\frac{1}{4}(\frac{\sigma^2}{11} + \frac{\sigma^2}{13}) + \frac{\sigma^2}{13} = \frac{17\sigma^2}{143}$, estimated as $\frac{17\sigma^2}{143} = 58.0305$. A 95% interval for this contrast is then $24.0 \pm 2.034\sqrt{58.0305}$ i.e. 24.0 ± 15.5 or 8.5 to 39.5. There is strong evidence that teaching plus treadmill training with or without exercise training, is a better programme of patient care.

4 (i) Having fitted a first-order(linear)model to the data, the directions in which y increases most rapidly for x, and x_2 are found, and a new center along this live is chosen for a second experiment in the series. This may be repeated until a line does not fit but a carve is needed. Then second-order models are used(see later). The idea is to approach the region of the maximum as quickly as possible.

(ii) $X_1 = \frac{x_1 - 90}{10}$ takes values ± 1. $Y_1 = \frac{y_1 - 20}{10}$ is also ± 1. Fit

$$y_i = \beta_0 + \beta_1 X_{1i} \beta_2 X_{2i} + \epsilon_i,\ \{\epsilon_i\}\ i.i.d.N(0, \sigma^2)$$

$$X = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}; \quad X'X = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}; \quad \hat{\beta} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix};$$

$$Y = \begin{bmatrix} 11 \\ 0 \\ 29 \\ 6 \\ 12 \end{bmatrix}; \quad \hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 1/5 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & 0 \\ -1 & -1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 11\ 0 \\ 29 \\ 6 \\ 12 \end{bmatrix} =$$

$$\begin{bmatrix} 1/5 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix} \begin{bmatrix} 58 \\ -34 \\ 24 \end{bmatrix} = \begin{bmatrix} 11.6 \\ -8.5 \\ 6.0 \end{bmatrix}, \text{ so that } y = 11.6 - 8.5X_1 + 6.0X_2.$$

(iii) $\frac{dy}{dx_1} = -8.5$ and $\frac{dy}{dx_2} = 6.0$, so these give the gradient of the steepest ascent line: as θ changes $(x_1, x_2) = (-8.5\theta, 6.0\theta)$ dies on this line.

(iv) In the original units, $(90 - 85\theta, 20 + 60\theta)$ gives the line. For the next 6 runs,

$$\theta = 0.3 \textit{ gives } x_1 = 64.5 \textit{ and } x_2 = 38;$$

$$\theta = 0.5 \quad x_1 = 47.5 \quad x_2 = 50;$$

$$\theta = 0.6 \quad x_1 = 39.0 \quad x_2 = 56;$$

$$\theta = 0.7 \quad x_1 = 30.5 \quad x_2 = 62;$$

$$\theta = 0.55 \quad x_1 = 43.25 \quad x_2 = 53;$$

$$\theta = 0.65 \quad x_1 = 34.75 \quad x_2 = 59,$$

so all are on the steepest ascent line.

(vi) The point of maximum response(39,56) could be taken as the center of a design for fitting a second-order surface. Either some more points could be added here and the two points just above and below it combined in an analysis; or a central composite design taking, say,(40.55) as center for simplicity, moving a short distance,such as ± 5 in x_1 and ± 3.5 in x_2, to give the corners.

5 (a)

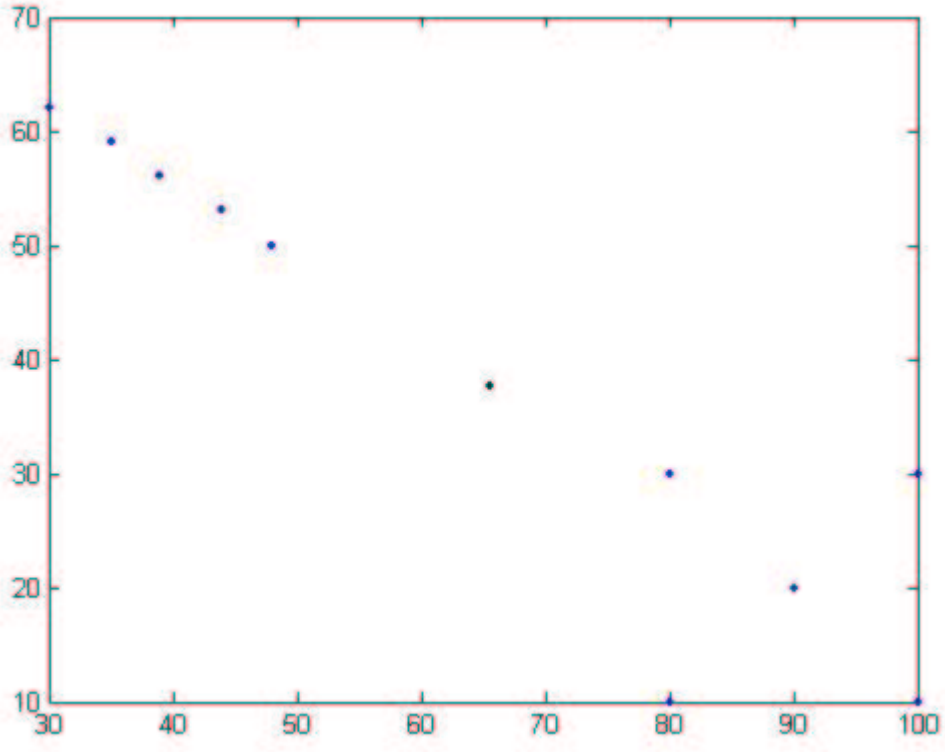
$$W_h = \textit{proportion of total population in statum } h;$$

$$N_h = \textit{total number of units in stratum } h;$$

$$n_h = \textit{number of units sampled in stratum } h;$$

$$P_h = \textit{the propotion of units with a particular attribute in stratum } h;$$

$$L = \textit{the number of strate.}$$


70
60
50
40
30
20
10
30
40
50
60
70
80
90
100

(b)(i) Minimize $Var(p_{st})$ subject to the constraint $C - c_0 = \sum_{h=1}^{L} c_h n_k$. By the lagrange multiplier technique, minimize $V + \lambda C$.

$$S = V + \lambda C = \sum_{h=1}^{L} \frac{W_h^2 P_h Q_h}{n_h} - \sum_{h=1}^{L} \frac{W_h^2 P_h Q_h}{N_h} + \lambda(c_0 + \sum_{h=1}^{L} c_h n_h)$$

$$\frac{\partial S}{\partial n_h} = -\frac{W_h^2 P_h Q_h}{n_h^2} + \lambda c_h = 0 \; if \; n_h = W_h \sqrt{\frac{P_h Q_h}{\lambda c_h}} \; for \; h = 1, 2, \cdots, L$$

(ii) If $N = total\ population\ size$, $W_h = N_h/N$. If V is fixed,

$$V = \sum_{h=1}^{L} W_h^2 P_h Q_h (\frac{1}{W_h} \sqrt{\frac{\lambda c_h}{P_h Q_h}} - \frac{1}{N_h}) = \sum_{h=1}^{L} W_h \sqrt{\lambda P_h Q_h c_h} - \sum_{h=1}^{L} W_h P_h Q_h / N$$

Hence

$$\sqrt{\lambda} = (V + \frac{1}{N} \sum_{h=1}^{L} W_h P_h Q_h) / (\sum_{h=1}^{L} W_h \sqrt{P_h Q_h c_h})$$

(c)(i)(ii) For the required width of confidence interval we must have $1.96\sqrt{V(P_{st})} = 0.1$, and therefore $V(P_{st}) = (\frac{0.1}{1.96})^2 = 0.0026031$

$Area$	W_h	$\sqrt{P_h Q_h / c_h}$	$n_h\sqrt{\lambda}$	$W_h P_h Q_h$	$W_h\sqrt{P_h Q_h C_h}$
1	0.58736	0.14434	0.084778	0.11013	0.76300
2	0.25814	0.16667	0.043023	0.06454	0.38721
3	0.10612	0.10000	0.010612	0.01698	0.16979
4	0.04837	0.08124	0.003930	0.00511	0.06287

For (ii) $\sqrt{\lambda} = \frac{0.0026031 + 0.19676/2026}{1.38287} = 0.0019526$, so $\lambda = 3.8127 \times 10^{-6}$. Using (i), the values of n_h are 43.42, 22.03, 5.43, 2.01, so these are rounded to 43, 22, 5, 2, total 72.

(iii) Total cost $= c_0 + 387 + 198 + 80 + 32 = c_0 + 697$ units. (Note: rounding n_1, n_3 up would add 25 units.)

6 (a) The target population is that about which we seek information, the study population that from which data are actually collected, the sampling frame the list of members of the population used when selection sample members(or equivalent, such as a map).

Sample random sampling is a method of selecting n members out of a population size N, such that any one of the $\begin{pmatrix} N \\ n \end{pmatrix}$ possible selections has the same probability of being chosen. Tn the sample random sampling with replacement, each population member has the same probability of being chosen at each draw, and this may be called an equal probability method.

(b) The question is vague, gives no units in which to answer, does not distinguish between beers, wines, spirits(for which units are different), relies on memory over a long period for any reasonably reliable answer. If a categorized answer is expected, "much" or "little" will have different interpretations according to age, sex, occupation, leisure activities etc. Possible replacements are:
(1) How often do you drink alcohol? (tick one box only)

$$regularly - every\ day\ (\ )$$

$$sometimes - not\ every\ day\ (\ )$$

$$very\ occasionally\ (\ )$$

$$have\ given\ up\ (\ )$$

$$have\ never\ drunk\ alcohol\ (\ )$$

(2) On a typical day, how much would you drink?

$$beer\ (points)\ 1/2\ (\ ),\ 1\ (\ ),\ 2\ (\ ),\ more\ (\ )$$

$$wine\ (glasses)\ 1\ (\ ),\ 2\ (\ ),\ 3\ (\ ),\ more\ (\ )$$

$$spirits\ (measures)\ 1\ (\ ),\ 2\ (\ ),\ more\ (\ )$$

Note: for an accurate assessment, several questions would be needed, but simple ones such as these may allow replies to be placed in categories.

(c) Quota sampling is quick, needing no prelimary selection, no preparation if sampling frame, no time revisiting members not available, can apply the ideas of stratification while collecting information, without needing to know exactly which units fall into which strata.

But there is no theoretical method of computing measures of variability, bias exists due to rather high refusal rates and to easy availability of people interviewed, and to subjective selection by the interviewer, and does not have randomization to help balance out uncontrolled factors.

(d) $N = 10000$, sample size is n, population $\sigma = 35.38$,width of 5% interval is to be 2×1.5. Assuming a normal distribution, the sample mean will be $N(\bar{Y}, (1 - \frac{n}{N})\frac{\sigma^2}{n})$, where $\bar{Y}$ is the true mean. So $1.5 = 1.96\sqrt{(1 - \frac{n}{10000}(\frac{(35.38)^2}{n}))}$ as a first approximation for n. Thus $(\frac{1.5}{1.96 \times 35.38})^2 = \frac{1}{n} - \frac{1}{10000}$ which gives $n = 1760.868$. There is no need to replace 1.96 by any t-value with a sample of this size, so take n=1761.

7 (i) The sample is a cluster sample because it consists of whole classes selected from among all the classes in the school. This process will be much simpler and less disruptive than locating individuals among all the classes.

(iii) If $R = \bar{Y}/\bar{M}$ then $r - R = \frac{\bar{y}}{\bar{m}} - R = \frac{\bar{y} - R\bar{m}}{\bar{m}}$. Write $d_i = y_i - Rm_i$, so that $\mu_d = \bar{Y} - R\bar{M} = 0$. Then $\bar{y} - R\bar{m} = \bar{d}$, and since it is based on a simple random sample $Var(\bar{d}) = E[\bar{d} - \mu_d]^2 = (1 - f)\frac{S_d^2}{n}$. Approximating $\bar{m}$ by $\bar{M}$ in the ratio, $V(\tau) = \frac{1}{\bar{M}^2}E[(\bar{y} - R\bar{m})^2]$ which is $\frac{(1-f)S_d^2}{\bar{M}^2 n} = \frac{1-f}{n\bar{M}^2}\sum_{i=1}^{N}\frac{(y_i - Rm_i)^2}{N-1}$

(iv) $\bar{y}_{cl} = \frac{13790}{270} = 51.074$. Estimate $\bar{M}$ by $\bar{m} = 27$. For each class, the value of $y_i - M_i\bar{y}$ is required, using $\bar{y}_{cl}$ as $\bar{y}$. These are:

1	2	3	4	5	6	7	8	9	10
28.854	49.188	−43.628	87.780	−209.332	−62.960	186.372	−110.406	6.706	67.466

The variance of these figures is $(111.6356)^2$. Hence $\hat{V}(\bar{y}_{cl}) = \frac{1}{27^2}(1 - \frac{10}{108})\frac{1}{10}(111.6356)^2 = 1.5512$, and the 95% limits are $51.074 \pm 1.96\sqrt{0.5512}$ i.e. 51.074 ± 2.441 *or*(48.63 *to* 53.52)

(v) If the interval is $\bar{y} \pm d$, then $d = 1.96\sqrt{Var(\bar{y})}$. This is to be 1, so $Var(\bar{y}) = (\frac{1}{1.96})^2 = 0.26031$. Also N=100, the average clusters size is unknown; use 27 as above. Similarly use $(111.6356)^2$ for the variance as above. These give $\frac{1}{27^2}(1 - \frac{n}{100})\frac{1}{n}(111.6356)^2 = 0.26031$, or $\frac{27^2 \times 0.26031}{111.6356} = \frac{1}{n} - \frac{1}{100}$ which requires n=39.64; so take n=40.

8

(i) standardization takes account of the different age-group patterns in different areas. Age standardization adjusts to show what the death rate would be if the proportions in the age groups were the same in each area. It takes no account of other factors affecting death rate. Direct standardization defines a standard population and applies to it different specific death rates for subgroups being compared. This gives the number of deaths expected in the standard population if the rates for the subgroups were to apply. Indirect standardization uses a set of specific death rates for the standard population and applies this to the subgroups. The actual number of deaths in each subgroup may then be compared with the number predicted from the standard rates.

(ii) The crude death rate is $\frac{actual\ number\ of\ deaths}{total\ population}$. For Scotland this is $\frac{16866}{5100000} = 0.003307$,i.e. 3.30 per thousand. For UK, $\frac{171179}{57649000} = 0.002969$,i.e. 2097 per thousand.

(iii)

Age	*UICpopulation*	$Age - specific\ death\ rates$	*Expected deaths*
under 35	28226	0.0072	203.227
35 − 44	7932	0.2639	2093.255
45 − 54	6593	1.2948	8536.616
55 − 64	5814	4.3687	25399.622
65 − 74	5.75	11.0794	56227.955
75 *and over*	4009	26.4634	106091.771

Age standardized mortality rate= $\frac{198552.446}{57649000} = 3.44$ per 1000.

(iv) Standardized mortality ratio= $\frac{observed\ number\ of\ deaths}{expected\ no.\ of\ deaths} \times 100$. Expected number of deaths

is found by using the UK mortality rates with the Scotland population.

Age	*UK Age specific death rate*	*population*	*expected deaths*
under 35	0.0068	2513	17.088
$35-44$	0.2011	701	140.971
$45-54$	0.9154	580	530.932
$55-64$	3.3566	537	1802.494
$65-74$	901367	441	4029.285
75 *and over*	24.3135	328	7974.828

Standardized mortality ratio= $\frac{16866}{14495.598} \times 100 = 116$. The mortality ratio in Scotland is 16% higher than for UK as a whole.

(v) Taking the UK as the standard population the age-specific death rates are as in (iv) and total expected deaths are 14495.598. Applied to the Scottish population figure gives an indirect standardized rate of $\frac{14495.598}{5100} = 2.84$ per 1000. This is what the Scottish figure would have been if its age specific pattern had been the same as the UK as a whole.