THE ROYAL STATISTICAL SOCIETY

GRADUATE DIPLOMA
Statistical Theory & Methods (2 papers)
Applied Statistics (2 papers)

SOLUTIONS 1999

Statistical Theory & Methods I

1. Bayes' Theorem. In a sample space S, suppose that the event $E_1, E_2, \cdots$ are such that $P(E_i) > 0$, all i, and $P(E_i \cap E_j) = 0$, all $i \neq j$, and $E_1 \cup E_2 \cup \cdots = S$. Let A be any event in S such that $P(A) > 0$. Then

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

$P(A) = 0.6$, $P(G) = 0.4$. For A, $X \sim N(5.2, 0.25^2)$ and for G, $X \sim N(5.7, 0.20^2)$.

(i) Let D be the event that the patient is diagnosed as Type A. We require to find $P(G|D)$.

$$\begin{aligned}
P(D|A) &= P(X < 5.5 \text{ in } N(5.2, 0.25^2)) \\
&= P(Z < \tfrac{0.3}{0.25} \text{ in } N(0,1)) = P(Z < 1.2) = 0.8849 \\
\text{Also } P(D|G) &= P(X < 5.5 \text{ in } N(5.7, 0.20^2)) \\
&= P(Z < \tfrac{-0.2}{0.20} \text{ in } N(0,1)) = P(Z < -1) = 0.1589. \\
\text{Hence } P(G|D) &= \frac{P(D|G)P(G)}{P(D|G)P(G) + P(D|A)P(A)} = \frac{0.1587 \times 0.4}{0.1587 \times 0.4 + 0.8849 \times 0.6} \\
&= \frac{0.06348}{0.06348 + 0.53094} = 0.1068.
\end{aligned}$$

(ii) $P(\text{Correctt diagnosis}) = P(\text{A diagnosed correctly}) + P(\text{G diagnosed correctly})$
$= P(A)P(X < c \text{ in } N(5.2, 0.25^2)) + P(G)P(X > c \text{ in } N(5.7, 0.20^2))$.

Let the $p.d.f.$ and $c.d.f.$ of $N(0,1)$ be $\phi(z), \Phi(z)$ respectively. Using $X = c$ as cut-off point, the probability is

$$P = 0.6\Phi(\frac{c-5.2}{0.25}) + 0.4\left\{1 - \Phi(\frac{c-5.7}{0.2})\right\}.$$

$$\begin{aligned}
\frac{dP}{dc} &= \frac{0.6}{0.25}\phi(\frac{c-5.2}{0.25}) - \frac{0.4}{0.2}\phi(\frac{c-5.7}{0.20}) \\
&= 2.4\exp(-\frac{1}{2}\left\{\frac{c-5.2}{0.25}\right\}^2) - 2\exp(-\frac{1}{2}\left\{\frac{c-5.7}{0.20}\right\}^2) \\
&= 0 \text{ when } 2.4\exp(-\frac{1}{2}\left\{\frac{c-5.2}{0.25}\right\}^2) = 2\exp(-\frac{1}{2}\left\{\frac{c-5.7}{0.20}\right\}^2)
\end{aligned}$$

i.e. $\exp(-\frac{1}{2}\left\{\frac{c-5.2}{0.25}\right\}^2) = \frac{5}{6}\exp(-\frac{1}{2}\left\{\frac{c-5.7}{0.20}\right\}^2)$

giving $-\frac{1}{2}(\frac{c^2 - 10.4c + 5.2^2}{0.25^2}) = \ln\frac{5}{6} - \frac{1}{2}(\frac{c^2 - 11.4c + 5.7^2}{0.20^2})$

or $\frac{1}{2}c^2(\frac{1}{0.2^2} - \frac{1}{0.25^2}) + \frac{1}{2}c(\frac{10.4}{0.25^2} - \frac{11.4}{0.2^2}) + \frac{1}{2}(\frac{5.7^2}{0.2^2} - \frac{5.2^2}{0.25^2}) = -0.182$

i.e. $4.5c^2 - 59.3c + (189.805 + 0.182) = 0$.

$$\begin{aligned} \text{Hence } c &= [59.3 \pm \sqrt{59.3^2 - 4 \times 4.5 \times 189.897}]/9 \\ &= 6.589 \pm \frac{1}{9}\sqrt{96.724} = 6.589 \pm 1.093 = 5.496 \text{ or } 7.682. \end{aligned}$$

Clearly the cut-off point must lie between the two means (5.2 and 5.7); and so $c \approx 5.5$.

2. (i) Let $Z = X + Y$. Then $Z = z$ if $X = x$ and $Y = z - x$ for $z = 0, 1, 2, \cdots, n + m$.

$$\begin{aligned} P(Z = z) &= \sum_{x=0}^{z} P(X = x \text{ and } Y = z - x) = \sum_{x=0}^{z} P(X = x)P(Y = z - x) \\ &= \sum_{x=0}^{z} \binom{n}{x} \theta^x (1-\theta)^{n-x} \binom{m}{z-x} \theta^{z-x} (1-\theta)^{(m-z+x)} \\ &= \theta^z (1-\theta)^{m+n-z} \sum_{x=0}^{z} \binom{n}{x}\binom{m}{z-x} \\ &= \binom{m+n}{z} \theta^z (1-\theta)^{m+n-z} \sim B(m+n, \theta). \end{aligned}$$

However, the final step depends on an algebraic result for $\binom{n}{r}$ which is not very "well known", and a more satisfactory proof is to use probability generating functions, as follows. This method also makes it clear that the result is only true when θ is the same in both distributions.
The $p.g.f.$ for $B(n, \theta)$ is $G(t) = p_0 t^0 + p_1 t^1 + \cdots + p_k t^k + \cdots$ which is

$$\begin{aligned} \sum_{r=0}^{n} \binom{n}{r} \theta^r (1-\theta)^{n-r} t^r &= \sum_{r=0}^{n} \binom{n}{r} (\theta t)^r (1-\theta)^{n-r} = \{\theta t + (1-\theta)\}^n \\ &= \{1 + \theta(t-1)\}^n. \end{aligned}$$

We require the result that for the sum of two independent random variables, $G_{x+y}(t) = G_X(t) G_y(t)$. Thus $G_{x+y}(t) = \{1 + \theta(t-1)\}^m \{1 + \theta(t-1)\}^n = \{1 + \theta(t-1)\}^{m+n}$
which is Binomial$(m+n, \theta)$ by the equivalence theorem between distributions and their $pgf's$,

(ii)

$$\begin{aligned} & P(X = x | X + Y = z) = P(X = x \cap Z = z)/P(Z = z) \\ = \;& P(Y = z - x \cap X = x)/P(Z = z) \\ = \;& \binom{n}{x} \theta^x (1-\theta)^{n-x} \binom{m}{z-x} \theta^{z-x} (1-\theta)^{m-z+x} / \binom{m+n}{z} \theta^z (1-\theta)^{m+n-z} \\ = \;& \binom{n}{x}\binom{m}{z-x} / \binom{m+n}{z}, \text{ a hypergeometric distribution.} \end{aligned}$$

(iii) $n = 10$, $m = 30$, $\theta = 0.1$, $z = 5$, $x = 2$, so the conditional distribution is $\dfrac{10!}{2!8!} \cdot \dfrac{30!}{3!27!} \cdot \dfrac{5!35!}{40!} = 0.278.$

3. (i) $f(y|X=x) = \dfrac{f(x,y)}{f(x)}$ and so $E[Y|X=x] = \displaystyle\int_{-\infty}^{\infty} y \cdot \frac{f(x,y)}{f(x)} dy.$

Therefore
$$E[h(x).E(Y|X=x)] = \int_{-\infty}^{\infty} \left\{ h(x) \int_{-\infty}^{\infty} y \frac{f(x,y)}{f(x)} dy \right\} f(x) dx$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x) \cdot y \cdot f(x,y) \cdot dy dx = E_x[h(x).Y].$$

(ii) $E[Y] = E\{E(Y|X=x)\} = E\{\alpha + \beta X\} = \alpha + \beta E[X].$
$E[XY] = E[X \cdot E(Y|X=x)] = E[\alpha X + \beta X^2] = \alpha E[X] + \beta E[X^2].$

$$\begin{aligned} P(X,Y) &= \frac{E[XY] - E[X]E[Y]}{\sqrt{V[X]V[Y]}} = \frac{\alpha E[X] + \beta E[X^2] - E[X](\alpha + \beta E[X])}{\sqrt{V[X]V[Y]}} \\ &= \frac{\beta \{E[X^2] - (E[X])^2\}}{\sqrt{V[X]V[Y]}} = \beta \sqrt{\frac{V[X]}{V[Y]}}. \end{aligned}$$

Hence $\beta = \rho(X,Y)\sqrt{\dfrac{V[Y]}{V[X]}}$; $\alpha = E[Y] - E[X]\rho(X,Y)\sqrt{\dfrac{V[Y]}{V[X]}}$.

(iii) Using the relation $V[Y] = V[E(Y|X)] + E[V(Y|X)]$, we have
$V[Y] = V[\alpha + \beta X] + E[\sigma^2] = \beta^2 V[X] + \sigma^2.$
So $\sigma^2 = V[Y] - \beta^2 V[X] = V[Y] - \rho^2 \dfrac{V[Y]}{V[X]} V[X] = (1-\rho^2)V[Y].$

4. (i)

$$\begin{aligned} F_1(x) &= P(X_1 \le x) = 1 - P(X_1 > x) \\ &= 1 - P(\text{all observations are greater than } x) \\ &= 1 - \{1 - F(x)\}^n. \end{aligned}$$

(ii)

$$\begin{aligned} F_j(x) &= P(X_j \le x) = P(\text{at least j observations } \le x) \\ &= \sum_{k=j}^{n} \binom{n}{k} \{F(x)\}^k \{1 - F(x)\}^{n-k}, \end{aligned}$$

the sum of probabilities in a binomial distribution with $p = F(x)$, since the sample is categorized into those $\le x$ and those $> x$. The individual terms in the sum are the $b_j(x)$ as defined, beginning from $k = j$ and continuing through $j+1, j+2, \cdots$ up to n.

(iii) For $U(0,1)$, $F(x) = x$, $0 \le x \le 1$.

Hence $b_j(x) = \binom{n}{j} x^j(1-x)^{n-j}$ and therefore

$$\begin{aligned}\frac{d}{dx}b_j(x) &= \binom{n}{j}\left\{jx^{j-1}(1-x)^{n-j} - x^j(n-j)(1-x)^{n-j-1}\right\}\\ &= \frac{n!}{(j-1)!(n-j)!}x^{j-1}(1-x)^{n-j} - \frac{n!}{j!(n-j-1)!}x^j(1-x)^{n-j-1}\end{aligned}$$

(iv)

$$\begin{aligned}f_j(x) &= \frac{d}{dx}F_j(x) = \frac{d}{dx}\{b_j(x)+b_{j+1}(x)+\cdots+b_n(x)\}\\ &= \frac{b_j(x)}{dx}+\frac{b_{j+1}(x)}{dx}+\cdots+\frac{b_n(x)}{dx}\\ &= \frac{n!x^{j-1}(1-x)^{n-j}}{(j-1)!(n-j)!} - \frac{n!x^j(1-x)^{n-j-1}}{j!(n-j-1)!} + \frac{n!x^j(1-x)^{n-j-1}}{j!(n-j-1)!}\\ &\quad -\frac{n!x^{j+1}(1-x)^{n-j-2}}{(j+1)!(n-j-2)!}+\cdots-\cdots+\frac{n!x^{n-2}(1-x)}{(n-2)!1!}\\ &\quad -\frac{n!x^{n-1}}{(n-1)!0!}+\frac{db_n(x)}{dx}\end{aligned}$$

But $b_n(x) = x^n$, so the last term is nx^{n-1} which cancels with the previous one.

Hence $f_j(x) = \dfrac{n!}{(j-1)!(n-j)!}x^{j-1}(1-x)^{n-j}, j = 1, 2, \cdots, n.$

5. The old and new variables are related by $X = UV$; $Y = (1-U)V$. The Jacobian of the transformation is $\begin{vmatrix} \frac{\partial x}{\partial U} & \frac{\partial x}{\partial V} \\ \frac{\partial y}{\partial U} & \frac{\partial y}{\partial V}\end{vmatrix} = \begin{vmatrix} V & U \\ -V & 1-U\end{vmatrix} = V.$

The joint distribution $f(X,Y) = \dfrac{\theta^{\alpha+\beta}e^{-\theta(x+y)}x^{\alpha-1}y^{\beta-1}}{T'(\alpha)T'(\beta)}$

which becomes

$$\begin{aligned}&\frac{v\cdot\theta^{\alpha+\beta}\cdot e^{-v\theta}\cdot v^{\alpha+\beta-2}\cdot u^{\alpha-1}(1-u)^{\beta-1}}{T'(\alpha)T'(\beta)} \qquad \begin{matrix}0<u<1\\ v>0\end{matrix}\\ &= \frac{\theta^{\alpha+\beta}}{T'(\alpha)T'(\beta)}\left\{u^{\alpha-1}(1-u)^{\beta-1}\right\}\left\{e^{-v\theta}v^{\alpha+\beta-1}\right\} \qquad \begin{matrix}0<u<1\\ v>0.\end{matrix}\end{aligned}$$

This is $\left\{\dfrac{\theta^{\alpha+\beta}}{T'(\alpha+\beta)}e^{-v\theta}v^{\alpha+\beta-1}\right\}\left\{\dfrac{T'(\alpha+\beta)}{T'(\alpha)T'(\beta)}u^{\alpha-1}(1-u)^{\beta-1}\right\} \quad \begin{matrix}0<u<1\\ v>0.\end{matrix}$

which shows

(1) that U, V are independent (by factorisation);
(2) that V is gamma with parameters $\theta, (\alpha+\beta)$;
(3) that U is $B(\alpha,\beta)$.

If X, Y are both exponential then $\alpha = \beta = 1$ and the distribution of the proportion U is $\dfrac{T^{'}(2)}{T^{'}(1)T^{'}(1)} u^0(1-u)^0 = 1 \quad (0 < u < 1)$,

i.e. uniform on $(0,1)$. This is the distribution W.

6. (i)

$$\begin{aligned} M_x(t) &= E[e^{xt}] = \int_0^\infty e^{xt} \cdot \theta e^{-\theta x} dx = \int_0^\infty \theta e^{-(\theta - t)x} dx \\ &= \theta \left[-\frac{1}{(\theta - t)} e^{-(\theta-t)x} \right]_0^\infty \quad (t < \theta) \\ &= \theta/(\theta - t). \qquad \text{[This may be written } 1/(1 - t/\theta)]. \end{aligned}$$

$\dfrac{d}{dt}(M_x(t)) = \theta/(\theta - t)^2$, $E[X] = M^{'}(0) = 1/\theta$

$\dfrac{d^2}{dt^2}(M_x(t)) = 2\theta/(\theta - t)^3$, $E[X^2] = M^{''}(0) = 2/\theta^2$

and $V[X] = \dfrac{2}{\theta^2} - (\dfrac{1}{\theta})^2 = 1/\theta^2$.

(ii) For the mgf of a (total $x\theta/\sqrt{n}$) we require

$$\begin{aligned} M_z(t) &= e^{-nt/\sqrt{n}} (M_x \{\theta t/\sqrt{n}\})^n \\ &= e^{-t\sqrt{n}} (\frac{1}{1 - t/\sqrt{n}})^n \end{aligned}$$

$$\begin{aligned} \log_e M_z(t) &= -t\sqrt{n} - n \log_e (1 + \{-t/\sqrt{n}\}) \\ &= -t\sqrt{n} - n \left\{ -\frac{t}{\sqrt{n}} - \frac{1}{2}\frac{t^2}{n} - \frac{1}{3}\frac{t^3}{n^{3/2}} - \cdots \right\} \\ &= \frac{1}{2}t^2 + \frac{1}{3}t^3/\sqrt{n} + \cdots o(1/\sqrt{n}) \cdots \\ &\to \frac{1}{2}t^2 \quad \text{as } n \to \infty \end{aligned}$$

Hence $M_z(t) \to e^{\frac{1}{2}t^2}$.

Therefore Z follows a standard normal distribution.

7. (a) $X \sim N(\mu, \sigma^2)$, so $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp(-\dfrac{1}{2}\left\{\dfrac{x-\mu}{\sigma}\right\}^2)$, $Z = \dfrac{X - \mu}{\sigma}$ is a monotonic function of x, with $\dfrac{dX}{dZ} = \sigma$.

So $g(z) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp(-\dfrac{1}{2}z^2) \cdot \sigma = \dfrac{1}{\sqrt{2\pi}} \exp(-\dfrac{1}{2}z^2)$, i.e. $N(0,1)$.

(b) (i) For a Poisson distribution with mean 2:

$r =$	0	1	2	3	4	5	
$\rho(r) =$	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	$\cdots$
$F(r) =$	0.1353	0.4060	0.6767	0.8571	0.9473	0.9834	$\cdots$

For random numbers up to 0.1353, take $r = 0$; for 0.1354 to 0.4060 take $r = 1$; for 0.4060 to 0.6767 take $r = 2$; etc.

The given numbers correspond to $r = 1, 1, 2, 4$.

(ii) Using the same "inverse cumulative distribution function" method as above, and the tables provided, the following standard normal values are obtained: $r = -1.07, -0.42, +0.45, +1.40$.

(iii) Given Z, the corresponding values of X are $X = \mu + \sigma Z$. Here $\mu = -3$ and $\sigma = 0.5$, so $x = -3.53, -3.21, -2.77, -2.30$.

(iv) Since $\chi^2_{(1)}$ is the square of a $N(0, 1)$, the following values are obtained: 1.145, 0.176, 0.203, 1.960.

8. The states of the Markov Chain are -2, -1, 0, 1, 2 since the game ends at the "absorbing barriers" ± 2.

The one-step transition matrix is

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1-\theta & 0 & \theta & 0 & 0 \\ 0 & 1-\theta & 0 & \theta & 0 \\ 0 & 0 & 1-\theta & 0 & \theta \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The two-step matrix is

$$P^2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1-\theta & \theta(1-\theta) & 0 & \theta^2 & 0 \\ (1-\theta)^2 & 0 & 2\theta(1-\theta) & 0 & \theta^2 \\ 0 & (1-\theta)^2 & 0 & \theta(1-\theta) & \theta \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The initial state is $X_0 = 0$, since the scores after the sixth point are equal. So X_{2m} must be either -2, 0 or +2. From the third column of P^2,

$$P(X_{2m} = 0) = 2\theta(1-\theta)P(X_{2(m-1)} = 0) \cdots\cdots (1)$$

From the fifth column of P^2,

$$P(X_{2m} = 2) = \theta^2 P(X_{2(m-1)} = 0) + P(X_{2(m-1)} = 2) \cdots\cdots (2)$$

As $X_0 = 0$, (1) gives $P(X_{2m} = 0) = \{2\theta(1-\theta)\}^m$.

From (2), $P(X_2 = 2) = \theta^2$; $P(X_4 = 2) = \theta^2 2\theta(1-\theta) + \theta^2; \cdots$

$P(X_{2m} = 2) = \theta^2 \left\{(2\theta(1-\theta))^{m-1} + \cdots + (2\theta(1-\theta)) + 1\right\}$.

which is a geometric series. Its limiting sum is

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

Statistical Theory & Methods II

1. If a random variable X has probability density (or mass) function $f(x;\theta)$ where $\theta = (\theta_1, \theta_2, \cdots, \theta_k)$ then the j^{th}-population moment of X is $\mu_j(\theta) = E[X^j]$, $j = 1, 2, \cdots$ as long as the expectation exists. Suppose that $(x_1, x_2, \cdots, x_n)$ is a random sample from X. Then the j^{th} sample moment is $m_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$ for $j = 1, 2, \cdots$. The method of moments estimator is given by solving the equations $\mu_j(\theta) = m_j$ for $j = 1, 2, \cdots, k$.

(i) $\mu_1 = E[X] = \int_0^\infty \frac{5x\theta^5}{(x+\theta)^6}dx = \theta^5 \left\{ \left[\frac{-x}{(x+\theta)^5}\right]_0^\infty + \int_0^\infty \frac{dx}{(x+\theta)^5} \right\}$

i.e. $\mu_1 = \theta^5 \left[\frac{-1}{4(x+\theta)^4}\right]_0^\infty = \theta/4.$

Also $m_1 = \bar{x}$. So $\bar{x} = \frac{1}{4}\hat{\theta}$, or $\hat{\theta} = 4\bar{x}$.

(ii) $E[\hat{\theta}] = 4E[\bar{x}] = \frac{4}{n}E[\sum_{i=1}^{n} x_i] = \frac{4}{n} \cdot n \cdot \frac{\theta}{4} = \theta.$

$V(\hat{\theta}) = 16V(\bar{x}) = \frac{16}{n}V(x)$, which is $\frac{16}{n}[E(x^2) - \{E(x)\}^2]$.

$$\begin{aligned} E[x^2] &= \int_0^\infty \frac{5x^2\theta^5}{(x+\theta)^5}dx = \theta^5 \left\{ \left[\frac{-x^2}{(x+\theta)^5}\right]_0^\infty + \int_0^\infty \frac{2x dx}{(x+\theta)^5} \right\} \\ &= \frac{1}{2}\theta^5 \left\{ \left[\frac{-x}{(x+\theta)^4}\right]_0^\infty + \int_0^\infty \frac{dx}{(x+\theta)^4} \right\} \\ &= \frac{1}{2}\theta^5 \left\{ \left[\frac{-1}{3(x+\theta)^3}\right]_0^\infty \right\} = \theta^2/6. \end{aligned}$$

$V[X] = \frac{\theta^2}{6} - \frac{\theta^2}{16} = \frac{5\theta^2}{48}$; so $V[\hat{\theta}] = \frac{16}{n} \cdot \frac{5\theta^2}{48} = \frac{5\theta^2}{3n}$.

Since $V(\hat{\theta}) \to 0$ as $n \to \infty$, $\hat{\theta}$ is consistent for θ.

(iii) $\ln L = n\ln 5 + 5n\ln\theta - 6\sum_{i=1}^{n}\ln(x_i + \theta)$, $[\theta > 0]$

$\frac{d}{d\theta}(\ln L) = \frac{5n}{\theta} - 6\sum_{i=1}^{n}\frac{1}{x_i+\theta}$, and

$\frac{d^2}{d\theta^2}(\ln L) = \frac{-5n}{\theta^2} + 6\sum_{i=1}^{n}\frac{1}{(x_i+\theta)^2}$.

The Cramèr-Rao lower bound is $-1/E[\frac{d^2(\ln L)}{d\theta^2}]$

$-E[\frac{d^2}{d\theta^2}(\ln L)] = \frac{5n}{\theta^2} - 6\sum_{i=1}^{n} E\left[\frac{1}{(x_i+\theta)^2}\right].$

Now $E[\frac{1}{(x+\theta)^2}] = \frac{5\theta^5 dx}{(x+\theta)^8} = \left[\frac{-5\theta^5}{7(x+\theta)^7}\right]_0^\infty = \frac{5}{7\theta^2}$.

Hence $-E[\frac{d^2}{d\theta^2}(\ln L)] = \frac{5n}{\theta^2} - \frac{30n}{7\theta^2} = \frac{5n}{7\theta^2}$ and so the lower bound is $7\theta^2/5n$.

The efficiency of $\hat{\theta}$ is therefore $\frac{7\theta^2}{5n}/\frac{3n}{5\theta^2} = \frac{21}{25}$.

2. Let X_i be the number of different plant species in area i ($i = 1$ to 150). Note that $\sum_{i=1}^{150} x_i = 4 \times 150 = 600$.

(i) $L(\alpha) = \prod_{i=1}^{150} \frac{-1}{\ln(1-\alpha)} \frac{\alpha^{x_i}}{x_i} = \{-\ln(1-\alpha)\}^{-150} \frac{\alpha^{600}}{\prod_{i=1}^{150} x_i}$, $0 < \alpha < 1$;

so $\ln L = -150\ln(-\ln(1-\alpha)) + 600\ln\alpha - \sum_{i=1}^{150} \ln x_i$.

$\frac{d}{d\alpha}(\ln L) = \frac{150}{(1-\alpha\ln(1-\alpha))} + \frac{600}{\alpha}$. Hence if the appropriate regularity conditions are satisfied, the $m.l.$ estimate $\hat{\alpha}$ satisfies $\frac{150}{(1-\hat{\alpha})\ln(1-\hat{\alpha})} + \frac{600}{\hat{\alpha}} = 0$.

(ii) To solve $\frac{d}{d\alpha}(\ln L) = 0$ by the Newton-Raphson method, we shall require $\frac{d^2}{d\alpha^2}(\ln L)$; this is

$$\begin{aligned} & \frac{150}{(1-\alpha)^2\{\ln(1-\alpha)\}^2} + \frac{150}{(1-\alpha)^2\ln(1-\alpha)} - \frac{600}{\alpha^2} \\ = \quad & \frac{150(1+\ln(1-\alpha))}{(1-\alpha)^2\{\ln(1-\alpha)\}^2} - \frac{600}{\alpha^2}. \end{aligned}$$

The iterative algorithm uses $\alpha_{n+1} = \alpha_n - \frac{\frac{d}{d\alpha}(\ln L)\mid_{\alpha=\alpha_n}}{\frac{d^2}{d\alpha^2}(\ln L)\mid_{\alpha=\alpha_n}}$ where the derivatives are evaluated at α_n, the n^{th} approximation to $\hat{\alpha}$ by the iterative method. Plotting $L(\alpha)$ against α could provide an initial value, α_0.

(iii) $E[X] = \sum_{k=1}^{\infty} \frac{-\alpha^k}{\ln(1-\alpha)} = \frac{1}{\ln(1-\alpha)} \sum_{k=1}^{\infty} \alpha^k = -\frac{\alpha}{(1-\alpha)\ln(1-\alpha)}$.

$E[-\frac{d^2}{d\alpha^2}(\ln L)] = \frac{-150\{1+\ln(1-\alpha)\}}{(1-\alpha)^2\{\ln(1-\alpha)\}^2} + \frac{150}{\alpha^2} \times 4$; but we are given $E[X] =$

4 and therefore the second term may be written as $\frac{150}{\alpha^2}(\frac{-\alpha}{(1-\alpha)\ln(1-\alpha)})$.

$$\begin{aligned}
\text{Thus}\quad E[-\frac{d^2}{d\alpha^2}(\ln L)] &= \frac{-150-150\ln(1-\alpha)}{(1-\alpha)^2(\ln(1-\alpha))^2} - \frac{150}{\alpha(1-\alpha)\ln(1-\alpha)} \\
&= \frac{-150\alpha - 150\alpha\ln(1-\alpha) - 150(1-\alpha)\ln(1-\alpha)}{\alpha(1-\alpha)^2(\ln(1-\alpha))^2} \\
&= \frac{-150\{\alpha+\ln(1-\alpha)\}}{\alpha(1-\alpha)^2\{\ln(1-\alpha)\}^2}.
\end{aligned}$$

When $\hat{\alpha} = 0.9$, this is $\frac{-150(0.9-2.3026)}{(0.9)(0.01)(2.3026)^2} = \frac{210.39}{4.7818} = 4409.01$.

The asymptotic distribution of $\hat{\alpha}$ is therefore $N(0.9; 1/4409.01)$.

The 99% confidence interval for α is approximately

$$\begin{aligned}
0.9 \pm 2.576\sqrt{\tfrac{1}{4409.01}} &= 0.9 \pm 0.039 \\
&= 0.86 \text{ to } 0.94 \text{ approx.}
\end{aligned}$$

3. (i) We wish to test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_m$ against $H_1 : \mu_k \neq \mu_l$ for at least one pair (k, l). Under H_1,

$$\begin{aligned}
L(\mu_1, \cdots, \mu_m) &= \prod_{i=1}^{m}\prod_{j=1}^{n}(2\pi\sigma^2)^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}(x_{ij}-\mu_i)^2\right\} \\
&= (2\pi\sigma^2)^{-\frac{1}{2}mn}\exp\left\{-\tfrac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\mu_i)^2\right\} \\
&\quad \text{for } -\infty < \mu_i < -\infty;\ i = 1, 2, \cdots, m
\end{aligned}$$

Hence $\ln L(\mu) = -\frac{1}{2}mn\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\mu_i)^2$

and $\frac{d}{d\mu_i}(\ln L) = \frac{-1}{2\sigma^2}(-2)\sum_{j=1}^{n}(x_{ij}-\mu_i)$ for $i = 1, 2, \cdots, m$.

$\frac{d}{d\mu_i}(\ln L) = 0$ gives $\hat{\mu_i} = \frac{1}{n}\sum_{j=1}^{n}x_{ij} = \bar{x}_i$ for $i = 1, 2, \cdots, m$.

Under H_0, $\ln L = -\frac{1}{2}mn\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\mu)^2$

so $\frac{d}{d\mu}(\ln L) = \frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\mu)$, which is 0 when $\hat{\mu} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij} = \bar{x}$.

The likelihood ratio test statistic is

$$\begin{aligned}
\Lambda(x) &= \frac{(2\pi\sigma^2)^{-\frac{1}{2}mn}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x})^2\right\}}{(2\pi\sigma^2)^{-\frac{1}{2}mn}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2\right\}} \\
&= \exp(-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x})^2-(x_{ij}-\bar{x}_i)^2\right\})
\end{aligned}$$

(ii) The critical region is $C : \{x \text{ such that } \Lambda(x) \leq k\}$ for some k.

Thus C is the values of x satisfying

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left\{(x_{ij}-\bar{x})^2-(x_{ij}-\bar{x}_i)^2\right\} \geq k'$$

(where $k' = -2\sigma^2 \ln k$). Using the given relation,

$$\begin{aligned}
\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)^2 &= \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x})^2+\sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{x}_i-\bar{x})^2-2\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x})(\bar{x}_i-\bar{x}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x})^2+n\sum_{i=1}^{m}(\bar{x}_i-\bar{x})^2-2n\sum_{i=1}^{m}(\bar{x}_i-\bar{x})^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x})^2-n\sum_{i=1}^{m}(\bar{x}_i-\bar{x})^2.
\end{aligned}$$

Therefore the region for C is $\left\{x : \sum_{i=1}^{m}(\bar{x}_i-\bar{x})^2 \geq k''\right\}$ where $k'' = k'/n$.

(iii) When H_0 is true, $\bar{X}_i \sim N(\mu, \sigma^2/n)$ for $i = 1, 2, \cdots, m$.

So $\sum_{i=1}^{m}(\bar{x}_i-\bar{x})^2/(\sigma^2/n)$ is distributed as $\chi^2_{(m-1)}$ on H_0.

A test of size α will reject H_0 if $\sum_{i=1}^{m}(\bar{x}_i-\bar{x})^2 \geq \dfrac{\sigma^2}{n}\chi^2_{(m-1,\alpha)}$

where the upper α% point of χ^2 is used.

When $\alpha = 0.05, m = 3, n = 7, \sigma^2 = 30$ and $\sum_{i=1}^{m}(\bar{x}_i-\bar{x})^2 = 112$, the critical value in the test is $\frac{30}{7} \times 5.99 = 25.67$, and 112 is much greater than this. Hence the evidence against H_0 is significant at (more than) the 5% level.

4. (i) The prior distribution of p is $\Pi(p) = 1$, $0 < p < 1$. Let X be the number of seeds out of L.S. that germinate. Then $X|p$ is binomial $(45, p)$; hence the

posterior distribution of p is

$$\begin{aligned}\Pi(p|X=25) &\propto 1 \times \left(\begin{smallmatrix}45\\25\end{smallmatrix}\right) p^{25}(1-p)^{20}\\ &\propto p^{25}(1-p)^{20} \quad 0<p<1.\end{aligned}$$

Thus $p|X=25$ is Beta$(26,21)$, so that

$$\Pi(p|X=25) \quad = \frac{T'(47)}{T'(26)T'(21)}p^{25}(1-p)^{20}, \quad 0<p<1.$$

(ii) $\ln\{\Pi(p|X=25)\} = const + 25\ln p + 20\ln(1-p)$

and $\frac{d}{dp}(\ln\Pi) = \frac{25}{p} - \frac{20}{1-p}$; $\frac{d^2}{dp^2}(\ln\Pi) = -\frac{25}{p^2} - \frac{20}{(1-p)^2} < 0$.

The mode is at $0 = \frac{25}{p} - \frac{20}{1-p}$ (and is a maximum)

i.e. $25(1-p) = 20p$ or $25 = 45p$, so $\hat{p} = \frac{25}{45} = \frac{5}{9}$.

If we consider the highest probability of being close to the true value, we could use the mode as a Bayes estimator of p.

(iii) With quadratic loss, the Bayes estimator of p is the expected value of p in the posterior distribution.

$$\begin{aligned}E[p|X=25] &= \int_0^1 \frac{T'(47)}{T'(26)T'(21)}p^{26}(1-p)^{20}dp\\ &= \left\{\int_0^1 \frac{T'(48)}{T'(27)T'(21)}p^{26}(1-p)^{20}dp\right\} \times \tfrac{T'(47)}{T'(48)} \cdot \tfrac{T'(27)}{T'(26)}\\ &= \frac{26}{47} \quad \text{since the integral } \{\} = 1.\end{aligned}$$

(iv) The variance in the posterior distribution is required.

$$\begin{aligned}E[p^2|X=25] &= \frac{T'(47)}{T'(26)T'(21)}p^{27}(1-p)^{20}dp\\ &= \frac{T'(47)}{T'(49)} \cdot \frac{T'(28)}{T'(26)} \int_0^1 \frac{T'(49)}{T'(28)T'(21)}p^{26}(1-p)^{20}dp\\ &= \frac{27\times 26}{48\times 47};\end{aligned}$$

Hence $V[p|X=25] = \frac{27\times 26}{48\times 47} - (\frac{26}{47})^2 = \frac{546}{48\times 47^2}$.

So that an approximate 95% Bayesian confidence interval for p is given by

$\frac{26}{47} \pm 1.96 \cdot \frac{1}{47}\sqrt{\frac{546}{48}} = 0.55 \pm 0.14$ i.e. 0.41 to 0.69.

5. If X denotes a random sample of observations, from a distribution with unknown parameter θ, in a parameter space $\mathcal{H}$, then any subset S_x of $\mathcal{H}$,

depending on X and such that $P(X : S_x \supset \theta) = 1 - \alpha$, is a $100(1-\alpha)\%$ confidence set for θ. (Thus a confidence interval is a special case)

(i) The distribution function of Y is

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(\max(x_i) \le y) = P(x_1, x_2, \cdots, x_n \le y) \\ &= \prod_{i=1}^{n} P(x_i \le y) \text{ by independence } = (\frac{y}{\theta})^n,\ 0 < y < \infty. \end{aligned}$$

So the $p.d.f.$ of Y is $f_Y(y) = \dfrac{ny^{n-1}}{\theta^n}$, $0 < y < \theta$.

(ii) Let $W = Y/\theta$. Then $F_W(w) = P(W \le w) = P(Y \le w\theta) = F_Y(w\theta)$, so that the $p.d.f.$ of W is $f_W(w) = \theta f_Y(w\theta) = nw^{n-1}$, $0 < w < 1$. Now Y/θ is a function of θ whose distribution does not depend on θ, i.e., it is a pivotal quantity.

(iii) Using this pivotal quantity, a family of $100(1-\alpha)\%$ confidence intervals for θ is $\{\theta : R_1 < Y/\theta < R_2\}$ where R_1, R_2 satisfy $P(R_1 < W < R_2) = 1 - \alpha$ (for $0 < \alpha < 1$).

(iv) Because $f_w(w) = nw^{n-1}(0 < w < 1)$, the shortest $100(1-\alpha)\%$ confidence interval will have $R_2 = 1$; thus R_1 must satisfy

$P(W > R_1) = 1 - \alpha$, i.e., $\displaystyle\int_{R_1}^{1} nw^{n-1}dw = 1 - \alpha$

or $[w^n]_{R_1}^1 = 1 - \alpha$, so that $1 - R_1^n = 1 - \alpha$ or $R_1^n = \alpha$.

Thus $R_1 = \alpha^{1/n}$.

Hence, the shortest $100(1-\alpha)\%$ confidence interval for θ is $(Y, Y\alpha^{-1/n})$, of length $Y(\alpha^{-1/n} - 1)$.

6. Suppose that $X_1, X_2, \cdots, X_n$ is a random sample from a symmetric distribution with median M. Then, Wilcoxon's signed ranks test can be used to test the Null Hypothesis $H_0 : M = m_0$ against the Alternative $H_1 : M \ne m_0$, where m_0 is a given value.

Let $D_i = X_i - m_0 (i = 1, 2, \cdots, n)$. Arrange $\{D_i\}$ in increasing order of absolute magnitude, and then allocate ranks $1, 2, \cdots, n$ according to the order of the D_i's. When there are tied ranks, use an average rank for all the tied D's. Let R_- and R_+ denote the sums of the negative and positive D_i's respectively, and let $T = \min(R_-, R_+)$. On H_0, there are 2^n possible sequences of positive and negative signs associated with the ranks, all of which are equally likely. Suppose $T = t$ is observed; its one-sided significance $= P(T \le t|H_0)$, which is the number of ways in which T can be $\le t$, divided by 2^n. This can be found by direct enumeration. For a 2-sided A.H. the probability is doubled. The values in Table XVII are the largest values of w

such that $P(T < w)$ under H_0 is less than or equal to the given value of p. The difference D_i between the two consumption rates for bird i ($i = 1$ to 8) gives the ordering:

D	-0.1	0.1	0.1	0.1	0.2	0.2	0.5	0.5
Rank	$2\frac{1}{2}$	$2\frac{1}{2}$	$2\frac{1}{2}$	$2\frac{1}{2}$	$5\frac{1}{2}$	$5\frac{1}{2}$	$7\frac{1}{2}$	$7\frac{1}{2}$,

where the first four share positions 1, 2, 3, 4 so have average rank $\frac{1}{4}(1+2+3+4)$; $R_- = 2.5$, $R_+ = 33.5$, so $T = 2.5$.
Table XVII gives the critical region at 5% for a 2-sided alternative as $T < 4$, and at 2% as $T < 2$. So there is evidence against H_0 at 5%, though not at 2%; we should reject the Null Hypothesis at the 5% level.

7 The size of a test is the probability of rejecting the N.H. when it is correct.
Suppose that we wish to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, and the likelihood function is $L(\theta)$. The Neyman-Pearson lemma states that the test with critical region of the form

$$C = \left\{x : \frac{L(\theta_i)}{L(\theta_1)} \le k\right\},$$

k chosen to make the test of size α, has the highest power among all tests of size $\le \alpha$.

(i). H_0 is "$\theta = \theta_0$", H_1 is "$\theta = \theta_1 > \theta_0$",

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta x_i \sqrt{2\pi}} \exp\left\{-\frac{1}{2}(\frac{\ln x_i}{\theta})^2\right\} = \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\frac{\ln x_i}{\theta})^2\right\}}{(2\pi)^{n/2}\theta^n \prod_{i=1}^{n} x_i}, \theta > 0.$$

The likelihood ratio is

$$\lambda = \frac{L(\theta_i)}{L(\theta_1)} = (\frac{\theta_1}{\theta_0})^n \exp\left\{-\frac{1}{2}(\frac{1}{\theta_0^2} - \frac{1}{\theta_1^2}) \sum (\ln(x_i))^2\right\},$$

and so the most powerful test is the likelihood ratio test with critical region $C = \{x : \lambda \le k\}$ for some k; that is,

$$C = \left\{x : \sum_{i=1}^{n} \{\ln(x_i)\}^2 \ge k'\right\},$$

where $k' = -\frac{2\theta_0^2\theta_1^2}{(\theta_1^2 - \theta_0^2)} \ln\left\{(\frac{\theta_0}{\theta_1})^n k\right\}$. Thus the test depends on $\sum_{i=1}^{n} \{\ln(x_i)\}^2$.

(ii). The distribution function of $Y = (\ln X)/\theta$ is

$$F_Y(y) = P(Y \leq y) = P((\ln X)/\theta \leq y) = P(X \leq e^{\theta y}) = F_X(e^{\theta y}).$$

So that *p.d.f.* is $f_Y(y) = \theta e^{\theta y} f_X(e^{\theta y}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$, $-\infty < y < \infty$, which is $N(0,1)$. Therefore $Y^2 = \chi^2_{(1)}$ and by the independence of $X_1, X_2, \cdots, X_n$ we have $\dfrac{\sum_{i=1}^{n}[\ln(x_i)]^2}{\theta^2} \sim \chi^2_{(n)}$.

(iii). On H_0, $\sum_{i=1}^{25}[\ln(x_i)]^2 \sim \chi^2_{(25)}$, and so the test of size 0.05 rejects H_0 if $\sum_{i=1}^{25}[\ln(x_i)]^2 \geq 37.65$.

(iv). On H_1, $\frac{1}{3}\sum_{i=1}^{25}[\ln(x_i)]^2 \sim \chi^2_{(25)}$ and the power of this test therefore is

$$P[\sum_{i=1}^{25}[\ln(x_i)]^2 \geq 37.65|\theta = 3] = P[\frac{1}{3}\sum_{i=1}^{25}[\ln(x_i)]^2 \geq 12.55|\theta = 3] = 0.98.$$

8. The Central Limit Theorem allows large samples of data to be studied as if they were normal, by examining either the mean or the total in the sample. If a distribution has mean μ and variance σ^2, then as sample sign $n \to \infty$ the distribution of the sample mean becomes approximately $N(\mu, \sigma^2/n)$, and of the sample total $N(n\mu, n\sigma^2)$. Therefore in large samples of data the CLT allows tests based on normal theory to be made and confidence intervals to be calculated, for means or totals.

   The size of n required for this to be satisfactory depends on how skew the original distribution is; when it is not very skew n need not be very large. So in relatively small samples from distributions that are not very skew the CLT allows us to carry out the standard methods and regard them as robust.

   In experimental design, normality is an assumption for the Analysis of Variance. Although samples are usually small, they may be based on records which themselves are the sum of many components, e.g. crop weights from plots each of a large number of plants. The CLT justifies assuming (approximate) normality for many items of biological data.

   Statistical Tables, especially for the t-test and for many nonparametric tests, need only be constructed for fairly small sample sizes because the normal approximations for these statistics are good for quite small samples; the relevant theory behind the derivation of the functions involved is subject to the

CLT mathematically. The same is true for the correlation coefficient.

It is also possible to give normal approximations to the Poisson and Binomial distributions when their parameters satisfy certain conditions; the same operation of the CLT applies. So again special tables are required only for cases where the approximations do not apply adequately.

Asymptotic results for large samples apply to maximum likelihood estimators; besides distribution theory that allows confidence intervals to be calculated using normal theory the asymptotic variance is the basis (by the Crawer-Rao bound) for assessing efficiency if other estimators.

The asymptotic distribution of $\ln \lambda$, where λ is the generalized likelihood ratio test statistic, under H_0 uses the CLT to support its proof; so also does the asymptotic theory for a posterior distribution.

Theoretical ways of examining approximations are to look at third and fourth moments of samples of data, or do a Monte Carlo study, or see whether the log likelihood function is quadratic. Practical ways are to construct (with the aid of suitable programs) dot-plots, histograms or box-and-whisker plots. Computer analysis can sometimes be done with and without transformation and the results compared. Residuals from fitted models may also be useful.

Applied Statistics I

1. (i). We are studying percentages, almost all of whose values are outside the range 20-80. 'Extreme' percentages do not have a variance that is even approximately constant and an inverse sine transformation greatly improves the validity of this assumption. The numbers upon which each percentage is based should be the same.

(ii). The factors are T = type of school ($B/M/G$); A = area ($S/M/N$),

$\sum x$	S	M	N	$TOTAL$	$\sum x^2$ $TOTAL$
B	6.4370	6.7945	5.7455	18.9770	69.1353
M	6.0050	5.6910	6.1630	17.8590	
G	5.3745	6.5455	6.5700	18.4900	$N = 45$
$TOTAL$	17.8165	19.0310	18.4785	55.3260	

Total sum of squares (corrected) $= 69.1353 - 55.3260^2/45 = 1.11383$

Area S.S. $= \frac{1}{15}(17.8165^2 + 19.0310^2 + 18.4785^2) - \frac{55.3260^2}{45} = 0.04930.$

Type S.S. $= \frac{1}{15}(18.9770^2 + 17.8590^2 + 18.4900^2) - \frac{55.3260^2}{45} = 0.04190.$

Area+Type+(Area $\times$ Type)$= \frac{1}{5}(6.4370^2+6.7945^2+\cdots+6.5455^2+6.5700^2)-\frac{55.3260^2}{45} = 0.36548.$

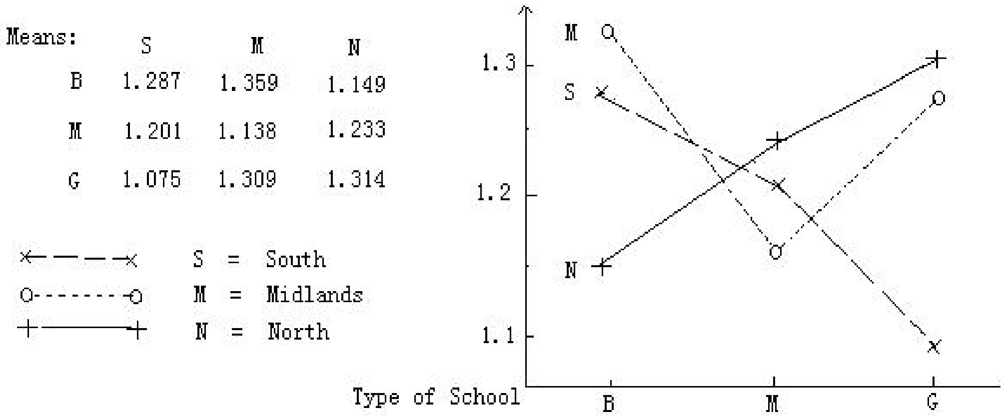
Analysis of Variance

$SOURCE$	$D.F.$	$S.S.$	$M.S.$	
$Type$	2	0.04190	0.0210	$f = 1 n.s.$
$Area$	2	0.04930	0.0247	$F_{(2,36)} = 1.19 n.s.$
$Type \times Area$	4	0.27428	0.0686	$F_{(4,36)} = 3.30^*$
$A + T + (A \times T)$	8	0.36548		
$Residual$	36	0.74835	$0.0208 = \hat{\sigma}^2.$	
$TOTAL$	44	1.11383		

There is evidence of interaction between area and type of school. (see next page)

2. (i) $f(x) = \frac{1}{x}$, $f'(x) = -\frac{1}{x^2}$. Hence $E[\frac{1}{x}] = f(\mu) = \frac{1}{\mu}$.
   $V[\frac{1}{x}] = \sigma^2(-\frac{1}{\mu^2})^2 = \sigma^2/\mu^4$. (Taylor Series Approximation).

(ii) The population consists of N animals, of whom m are marked. There are $\binom{N}{n}$ ways of selecting n from the whole population. The x marked ones can be selected from m in $\binom{m}{x}$ ways and the $(n-x)$ unmarked from $(N-m)$



in $\binom{N-m}{n-x}$ ways.

The probability that $X = x$ is the proportion (no of ways of making selection of $(x, n-x)$)$\div$(total no of ways of selecting n), and the numerator is the product of the expressions $\binom{m}{x}$ and $\binom{N-m}{n-x}$. Hence

$$P(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}, \quad \begin{array}{l}\text{for } \max(0, n-N+m) \leq x \leq \min(n, m) \\ \text{since } x \text{ can only take values in this range.}\end{array}$$

(iii) If we assume that the proportions marked in sample and population are the same, $\frac{x}{n} = \frac{m}{N}$, i.e. $\hat{N} = \frac{mn}{x}$.

Using this estimator, $E[\hat{N}] = mnE\left[\frac{1}{x}\right] = \frac{mn}{\mu} = \frac{mn}{\frac{mn}{N}} = N$ to first order, since $E[x] = \frac{mn}{N}$ in the hypergeometric distribution.

With the given expression for variance (σ^2),

$$\begin{aligned} Var[\hat{N}] &= (mn)^2 V\left[\frac{1}{x}\right] = m^2n^2\sigma^2/\mu^4 = m^2n^2 \cdot \frac{mn(N-m)(N-n)}{N^2(N-1)} \cdot \frac{N^4}{n^4m^4} \\ &= \frac{(N-m)(N-n)N^2}{(N-1)mn}, \end{aligned}$$

which to first approximation may be taken as $(N-m)(N-n)N/mn$.

(iv) $\hat{N} = \frac{(100)^2}{20} = 500$. $V[\hat{N}] = \frac{(\hat{N}-100)(\hat{N}-100)(500)}{(100)(100)} = (\frac{400}{100})^2(500) = 8000.$

Using a normal approximation, the 95% confidence interval for N is $\hat{N} \pm 1.96\sqrt{V(\hat{N})} = 500 \pm 1.96\sqrt{8000} = 500 \pm 175$ or $(325; 675)$.

3. (i) The matrix that has to be inverted can be near-singular. The estimates of coefficients become unstable and the variances large. It is difficult to select

a subset from the whole.

Some of the most highly correlated variables can be omitted. Otherwise principal component regression or ridge regression may be used.

(ii). With x_1 in the model, x_1^2 alone is not worth adding. But including x_2 and x_3 with x_1 makes a big improvements, increasing the regression S.S. by $35232 - 2847 = 32385$. $(a) : SS = 2847$; and $(b) : SS = 35232$.

Different slopes and intercepts is $(c) : SS = 35517$.

The simplest appropriate model is (b), with different intercepts.

Compare it with (c):

SOURCE OF VARIATION	$D.F.$	$S.S.$	$M.S.$	
Intercepts only	3	35232		
Slopes as well as intercepts	2	285	142.5	$F_{(2,21)} = 19.18^{***}$
Slopes and Intercepts	5	35517		
Residual	21	156	7.43	
TOTAL	26	35673		

However, there is a considerable improvement by including slopes also. Consider $(d) : S.S. = 35669$. The increase in S.S. of curvature over linearity is 35669-35517=152, with $3d.f.$; the corresponding M.S. is 50.667 and residual now is 156-152=4 with 21-3=18 $d.f.$ The new residual M.S. is $4/18 = 0.222$, and the improvement by including curvature is shown by $F_{(3,18)} = \frac{50.667}{0.222} = 228^{***}$, apparently a very great improvement.

But for (c), already $R^2 = \frac{35517}{35673} = 99.6\%$, so there is not a very great need for improvement. With adequate residual $d.f.$, as here, quadratic terms can be included; but if $d.f.$ were less it might be best to omit them.

Other useful information would include the raw data, some graphical plots of them, some standard diagnostic methods, the view of the engineer as to whether the quadratic terms are worth including, and any previous work on similar problems. Different intercepts in (b) can be examined by the model including x_1x_2 and x_1x_3.

4. (i) The Gauss-Markov Theorem for simple linear regression says that if $y_i = \alpha + \beta x_i + \epsilon_i (i = 1, \cdots, n)$, $E[\epsilon_i] = 0$, $Var[\epsilon_i] = \sigma^2$, all ϵ_i, ϵ_j uncorrelated then the least squares estimators of α and β are best linear unbiased estimators.

For the general linear model ("multiple regression") with the same conditions on $\{\epsilon_i : i = 1 \text{ to } n\}$, i.e., $E[\epsilon] = 0$ and $E[\epsilon\epsilon^T] = \sigma^2 I$, then in $Y = X\beta + \epsilon$, the least squares estimate $\hat{\beta} = (X^TX)^{-1}X^Ty$ gives the best linear unbiased estimate of $\beta = (\beta_1, \beta_2, \cdots)^T$.

These "best" estimators are minimum variance.

Hence least-squares provides estimates that are both unbiased and of minimum variance. In the case where $\{\epsilon_i\}$ follow normal distributions we also have maximum likelihood estimators by this method.

(ii). (a) If $Y = X\beta + \epsilon$, the least-squares estimator is the solution of

$\frac{d}{d\beta}(Y - X\beta)^T(Y - X\beta) = \frac{d}{d\beta}\epsilon^T\epsilon = 0$, i.e.

$\frac{d}{d\beta}(Y^TY - \beta^TX^TY - Y^TX\beta + \beta^TX^TX\beta) = 0,$

or $\frac{d}{d\beta}(Y^TY - 2\beta^TX^TY + \beta^TX^TX\beta) = 0.$

Hence $X^TY = X^TX\hat{\beta}$, so that $\hat{\beta} = (X^TX)^{-1}(X^TY)$.

$$\begin{aligned} E[\hat{\beta}] &= E[(X^TX)^{-1}(X^TY)] = E[(X^TX)^{-1}X^T(X\beta + \epsilon)] \\ &= \beta + E[(X^TX)^{-1}X^T\epsilon] = \beta \text{ since } E[\epsilon] = 0. \end{aligned}$$

$$\begin{aligned} V[\hat{\beta}] &= V[(X^TX)^{-1}(X^TY)] = (X^TX)^{-1}X^TV[Y]X(X^TX)^{-1} \\ &= (X^TX)^{-1}\sigma^2 \text{ since } V[Y] = V[\epsilon] = \sigma^2 I. \end{aligned}$$

(b).

$$\begin{aligned} \hat{\beta} &= \begin{bmatrix} 0.690129 & -0.083363 & -0.002234 \\ -0.083363 & 0.056302 & 0.000023 \\ -0.002234 & 0.000023 & 0.000009 \end{bmatrix} \begin{bmatrix} 604 \\ 791 \\ 146578 \end{bmatrix} \\ &= \begin{bmatrix} 23.4425 \\ -2.4451 \\ -0.0119 \end{bmatrix} \end{aligned}$$

$V(\beta_0) = 0.690129\hat{\sigma}^2$. We require the regression analysis of variance to estimate σ^2.

The regression S.S. is $[X^TY]^T\hat{\beta} = [604\ 791\ 146578] \begin{bmatrix} 23.4425 \\ -2.4451 \\ -0.0119 \end{bmatrix} = 10481$

and the residual S.S. is 11194 - 10481 =713. This has $(n-3) = 47 d.f.$ so the residual M.S. is $713/47 = 15.17$.

Hence $V(\beta_0) = 0.690129 \times 15.17 = 10.469$ and the 95% confidence interval is

$$\begin{aligned} 23.4425 \pm t_{(47)}\sqrt{10.469} &= 23.4425 \pm 2.01 \times 3.236 \\ &= 23.4425 \pm 6.5036 \\ &\ i.e.(16.94 \text{ to } 29.95). \end{aligned}$$

The S.S. with β_0 only is $604^2/50 = 7296$, so a full analysis of variance is:

SOURCE	$D.F.$	$S.S.$		
β_0	1	7296		
β_1, β_2 after β_0	2	3185	1592.5	$F_{(2,47)} = 105^{***}$
Residual	47	713	15.17	
	50	11194		

There is very strong evidence against the NH "$\beta_1 = \beta_2 = 0$".

5. (i) Conditions $C_i (i = 1, 2)$ is a "fixed" effect; batches are randomly selected from a wider population, and so will have a variance σ_b^2. The residual $\{\epsilon_{ijk}\}$ terms are independently distributed as $N(0, \sigma^2)$; μ is a general mean potency response, so batches may be assumed $N(0, \sigma_b^2)$.

(ii) The degrees of freedom for the items in the analysis are respectively 1, 4, 18; total 23.

We require $\sum_i \sum_j \sum_k (y_{ij\cdot} - y_{i\cdot\cdot})^2 = S$

$$\begin{aligned} y_{ij\cdot} - y_{i\cdot\cdot} &= \mu + C_i + b_j(i) + \epsilon_{ij\cdot} - \mu - C_i - t_0(i) - \epsilon_{i\cdot\cdot} \\ &= (b_j(i) - b.(i)) + (\epsilon_{ij\cdot} - \epsilon_{i\cdot\cdot}) \end{aligned}$$

Now $\sum_i \sum_j \sum_k (b_j(i) - b.(i))(\epsilon_{ij\cdot} - \epsilon_{i\cdot\cdot}) = 0$, since b and ϵ are independent.

Thus

$$\begin{aligned} S &= \sum_i \sum_j \sum_k (b_j(i) - b.(i))^2 + \sum_i \sum_j \sum_k (\epsilon_{ij\cdot} - \epsilon_{i\cdot\cdot})^2 \\ &= 4 \sum_i \sum_j (b_j(i) - b.(i))^2 + 4 \sum_i \sum_j (\epsilon_{ij\cdot} - \epsilon_{i\cdot\cdot})^2 \end{aligned}$$

and $E[S] = 4 \cdot 2 \cdot 2\sigma_b^2 + 4 \cdot 2 \cdot 2\sigma^2/4 = 16\sigma_b^2 + 4\sigma^2$

so that the expected mean square is $4\sigma_b^2 + \sigma^2$.

The S.S. between conditions is $\frac{T_A^2}{12} + \frac{T_B^2}{12} - \frac{G^2}{24}$

$= \frac{1}{12}(321^2 + 150^2) - \frac{471^2}{24} = 10461.75 - 9243.375 = 1218.375.$

The complete analysis of variance is

SOURCE	$D.F.$	$S.S.$	$M.S.$	F_{ratio}
Between conditions	1	1218.375	1218.375	$F_{(1,4)} = 15.75^{*}$
Within conditions between batches	4	309.500	77.375	13.96^{***}
Within batches	18	99.750	$5.542 = \hat{\sigma}^2$	
TOTAL	23	1627.625		

Conditions A lead to significantly higher potency than B (the significance is only at 5% because there are very few $d.f.$ for the test - $F_{(1,4)}$). The variation between batch potencies is very highly significant ($F_{(4,18)}$). The first test has N.H., "$C_1 = C_2$" and the second has N.H., "$\sigma_b^2 = 0$". The estimate of σ_b^2 is estimated as $\frac{1}{4}(77.375 - 5.542) = 17.96$, much larger than σ^2, $\hat{C}_A - \hat{C}_B$ is estimated as $\frac{1}{12}(321 - 150) = 14.25$.

6. (i) Suppose that $\{Y_t\}$ is a purely random process with mean 0 and variance σ^2. Then $\{X_t\}$ is $MA_{(q)}$ if

$$X_t = \beta_0 Y_t + \beta_1 Y_{t-1} + \cdots \beta_q Y_{t-q}.$$

Also, U_t is $AR(p)$ if $U_t = \alpha_1 U_{t-1} + \cdots + \alpha_p U_{t-p} + Y_t$.

(ii) $X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$

(a)

$$\begin{aligned} E[X_t X_t] &= E[(a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2})^2] = E[a_t^2] + \theta_1^2 E[a_{t-1}^2] + \theta_2^2 E[a_{t-2}^2] \\ &= \sigma^2(1 + \theta_1^2 + \theta_2^2) \text{ since } \{a_t\} \text{ are independent.} \end{aligned}$$

$E[X_t X_{t-k}] = E[(a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2})(a_{t-k} + \theta_1 a_{t-k-1} + \theta_2 a_{t-k-2})]$.

When $k = 1$, $E[X_t x_{t-1}] = (\theta_1 + \theta_1\theta_2)\sigma^2 = \theta_1(1 + \theta_2)\sigma^2$ by independence and for $k = 2$, $E[X_t X_{t-2}] = \theta_2\sigma_2$, all other terms being 0.

For $k \geq 3$, $E[X_t X_{t-k}] = 0$.

Therefore $\rho_1 = \dfrac{\theta_1(1 + \theta_2)}{1 + \theta_1^2 + \theta_2^2}$ and $\rho_2 = \dfrac{\theta_2}{1 + \theta_1^2 + \theta_2^2}$; also $\rho_k = 0$ for $k \geq 3$.

(b) $Z_t = \frac{1}{2}(X_t + X_{t-1})$, so $V[Z_t] = \frac{1}{4}V[X_t + X_{t-1}]$.

Hence

$$\begin{aligned} V[Z_t] &= \tfrac{1}{4}[V[X_t] + 2Cov[X_t, X_{t-1}] + V[X_{t-1}]] \\ &= \tfrac{1}{2}\sigma^2(1 + \theta_1^2 + \theta_2^2 + \theta_1(1 + \theta_2)) \end{aligned}$$

$\dfrac{\partial V}{\partial \theta_1} = \dfrac{1}{2}\sigma^2(2\theta_1 + 1 + \theta_2)$ and $\dfrac{\partial V}{\partial \theta_2} = \dfrac{1}{2}\sigma^2(2\theta_2 + \theta_1)$.

Setting these to 0, we have $2\theta_1 + 1 + \theta_2 = 0$ and $2\theta_2 + \theta_1 = 0$; therefore $\theta_1 = -2\theta_2$ and so $-4\theta_2 + 1 + \theta_2 = 0$ or $\theta_2 = 1/3$. This gives $\theta_1 = -2/3$.

7. (i) (a) The first principal component of the set of observations $X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$,

with p measurements on each of n units from a population is that linear combination $Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots a_{p1}X_p = a_1^{'}X$ whose sample variance $a_1^{'} \sum a_1$ is greatest among all possible vectors a_1 satisfying $a_1^{'}a_1 = 1$.

The second $Y_2 = a_{12}X_1 + a_{12}X_1 + \cdots + a_{12}X_1 = a_2^{'}x$ is orthogonal to the first component and has the greatest possible variance subject to this, i.e. satisfies $a_2^{'}a_2 = 1$ and $a_1^{'}a_2 = 0$.

When these are based on a variance-covariance matrix, the scales in which the x-variables are measured is important, but this is corrected for by using a correlation matrix. The variance-covariance matrix has a simple sampling distribution, but the components may be dominated by large measurements. The correlation matrix gives equal weight to all variables and provides linear combinations of scale free measurements.

(b) Principal components analysis can check the dimensionality of the data - do we really need p measurements to explain them or can fewer linear combinations be used, for example as the predictors in a regression analysis?

The transformation to orthogonal (uncorrelated) components can also be useful; and clusters of points, or outliers, can be examined.

The components with the smallest variances can also be helpful in identifying measurements which need not be taken in future.

(ii) Since there are 5 measured variables, the eigenvalues will add to 5. The first three contribute very nearly all of this; and in fact the first two contribute 80%. Therefore the dimensionality is not more than 3 and could perhaps be taken as 2. The five given x's would be expected to be quite highly correlated.

PC1 is a linear combination of all five, perhaps a "wealth index " but when the correlation matrix is used the first component is quite often of this sort.

PC2 is a contrast between (x_2, x_3) and (x_4, x_5); there is no obvious interpretation but it might be useful to plot the value of this contrast against PC1 for the set of observed units.

PC3, if used, gives a contrast between the first and second incomes.

8. A generalized linear model requires (1) a link function, (2) a linear predictor, (3) an error distribution.

Given a set of observation $y_1, y_2, \cdots, y_n$ from a distribution having density function $f(y_i, \eta_i, \phi)$, which is in the exponential family and includes normal, binomial, gamma and Poisson; $\eta_i = \sum_{j=1}^{p} \beta_j x_{ji}$, the linear predictor, and $\eta_i = E(y_i)$ in the simplest cases but need not be so in general. The link function relates η_{ij} to the mean μ_{ij}, and in the contingency table model given the linking function is $\log \mu_{ij}$, or $\log \lambda_{ijk}$ in this particular example. The right hand side is the linear predictor. The error distribution assumed is Poisson.

(ii) The levels of variables required are $i = 1$, $j = 0$, $k = 1$, so that $R_1 = -0.011$; $E_0 = +0.104$; $I_1 = +0.011$; $(RE)_{10} = -0.284$; $(RI)_{11} = +0.348$; $(EI)_{01} = +0.021$, giving

$$\begin{aligned} \log_e \lambda_{101} &= +2.953 - 0.011 + 0.104 - 0.284 + 0.345 + 0.021 \\ &= 3.142 \text{ and so } \lambda_{101} = 23.15. \end{aligned}$$

(iii) Fitting the main effects and (EI) does not reduce the deviance significantly ($\chi_1^2 = 34.94 - 31.96 = 2.98 n.s.$). Beginning with μ, R, E, I we may add (RE) or (RI); the first of these reduces the deviance by 12.40, the second by 18.82, both significant at 0.1%. Adding (RE) after (RI) reduces the deviance by 14.00-1.60 =12.40, again very highly significant. with $\mu, R, E, I, (RE), (RI)$ we have a small deviance that will not be improved by the 3-factor term. We need all these 6 terms in a model that explains the data satisfactorily.

1. (i) The total (corrected) sum of squares in the analysis of variance is $8988 - \frac{740^2}{64} = 431.75$, treatment S.S.$= \sum \frac{T_i^2}{8} - \frac{G^2}{N} = \frac{1}{8}(69196) - \frac{740^2}{64} = 93.25$ (batches); panel S.S.$= \frac{69066}{8} - \frac{740^2}{64} = 77.00$.

SOURCE	*D.F.*	*S.S.*	*M.S.*	
Replicates(Panels)	7	77.00	11.000	$F_{(7,49)} = 2.06$ *n.s.*
Treatments(Batches)	7	93.25	13.321	$F_{(7,49)} = 2.50^{*}$
Residual	49	261.50	5.337	
TOTAL	63	431.75		

There is no evidence of systematic panel differences. We can subdivide the "treatments" into single degrees of freedom.

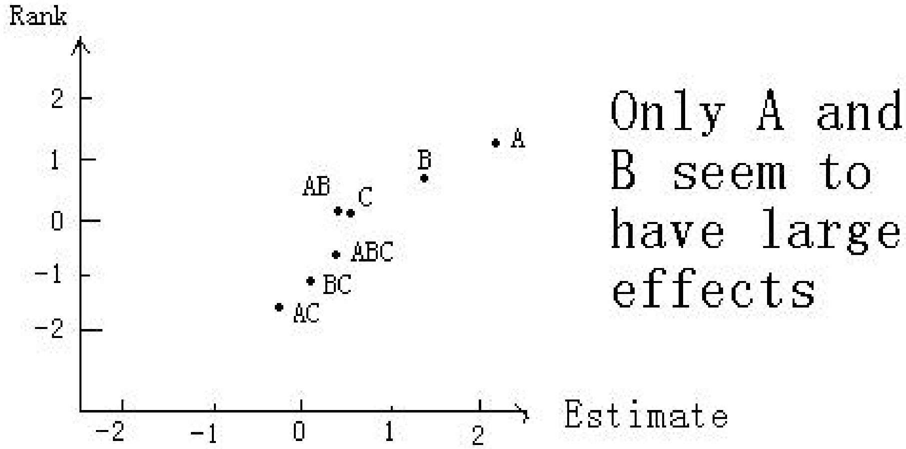
"Treatment"	(1)	*a*	*b*	*ab*	*c*	*ac*	*bc*	*abc*			
Total	78	97	88	108	81	97	89	102	Value	$\frac{\text{Value}^2}{64}$	$F_{(1,49)}$
A	−	+	−	+	−	+	−	+	68	72.25	13.54^{**}
B	−	−	+	+	−	−	+	+	34	18.06	3.38 *n.s.*
AB	+	−	−	+	+	−	−	+	−2	0.06	< 1
C	−	−	−	−	+	+	+	+	−2	0.06	< 1
AC	+	−	+	−	−	+	−	+	−10	1.56	< 1
BC	+	+	−	−	−	−	+	+	−8	1.00	< 1
ABC	−	+	+	−	+	−	−	+	−4	0.25	< 1

Only A(pan material)has a significant effect: "high level", i.e. aluminium, being better than glass. The only other effect worth any further experimentation would be B(stirring).

(ii) There is only one mix per recipe, and there is no true replication as the 8 samples from it are unlikely to be "independent". Also these scores may not be even approximately normally distributed.

(iii) The "effects" of each main effect and interaction can be estimated by dividing the values in the above table by 8, and the averages of these by dividing again by 4, since each effect is the average of four (+,-) comparisons. In fact it does not help to reduce all these comparisons to averages because there is no significance test that can be done on them (having no genuine residual *d.f.*).

	A	B	AB	C	AC	BC	ABC
Effect:	2.13	1.06	−0.06	−0.06	−0.31	−0.25	−0.12.



(iv) Averages are nearer to normality than individual data. This analysis gets round the problem of dependence among the eight 'replicate' ratings. For these reasons it may be thought better than that in (i).

2. (a) If block size is strictly limited, to k, and the number of treatment to be composed, v, is more than this, a balanced incomplete block will be useful when comparisons between pairs of treatment means are all equally important. Any pair of treatments occurs together the same number, λ, of times in a block.

The total number of unit plots is N, which is equal to rv but also to bk. Hence $(N =)rv = bk$.

Consider one particular treatment. It will occur together with others in a block $r(k-1)$ times; but it also appears λ times with each of the other $(v-1)$ treatments. Hence $\lambda(v-1) = r(k-1)$. But λ must be an integer. So $\lambda = \frac{r(k-1)}{v-1}$ is an integer.

(b) (i) $v = 5 = b$, $r = 4 = k$. $N = 20$. $\lambda = \frac{4\times 3}{4} = 3$.

(ii) $G = 4348$; $\sum x^2 = 955360$. Total $S.S. = 955360 - \frac{4348^2}{20} = 10104.8$.

$S.S.$ for batches (not adjusted for treatments)

$= \frac{1}{4}(863^2 + 838^2 + 835^2 + 801^2 + 906^2) - \frac{4348^2}{20} = 704.8$.

$B^{(i)}$ = total yield of all plots in all those blocks containing trt. i.

Treatment:	A	B	C	D	E	TOTAL
TOTAL	761	825	949	818	995	: 4348
$B^{(i)}$	3465	3442	3485	3487	3513	
$Q_i = kT_i - B^{(i)}$	−421	−142	311	−215	467	

$S.S.$ Treatments adjusted for Batches $= \sum Q_i^2/vk\lambda = 558440/60 = 9307.33$
Analysis of Variance.

SOURCE	$D.F.$	$S.S.$	$M.S.$	
Batches (ignoring treatment)	4	704.80		
Treatments adjusted for batches	4	9307.33	2326.8	$F_{(4,11)} = 276.2^{***}$
Residual	11	92.67	$8.425 = \hat{\sigma}^2$	
TOTAL	19	10104.80		

These is very strong evidence of treatment differences. (We cannot test batches because the above $S.S.$ is unadjusted.)

(iii) Means are $\hat{\mu} + Q_i/\lambda v$ and $\hat{\mu} = \frac{G}{N} = \frac{4348}{20} = 217.4$. Also the variance of a difference between any pair of means is $2k\hat{\sigma}^2/v\lambda = \frac{8}{15} \times 8.425 = 4.493$, so $S.E.$ is 2.12. Also $t_{(11,5\%)} = 2.201$. Any pair of means differing by more than $2.12 \times 2.201 = 4.67$ may be called significant.

Means are:	A	D	B	C	E
	189.33	203.07	207.93	238.13	248.53

A versus D: At 10% Cd, the addition of 10% Sn gives a significant rise in melting point.
A versus B: Without Sn, increasing Cd from 10% to 20% does the same.
All these three are very much less than C and E.
B versus C: Without Sn, increasing Cd from 20% to 30% gives a further significant rise in melting point.
C versus E: At 30% Cd, the addition of 10% Sn gives a significant rise in melting point.
Summary: Each increase in Cd or in Sn raises the melting point.

3. (i) Examining the 3-factor interaction first, it provides no evidence at all against the NH that $A * B * C$ is zero. For the 2-factor interactions, in the same way, $A * B$ and $A * C$ can be taken as zero. The F-value for $B * C$ is very large, and on the NH "$B * C = 0$" it has an extremely small p-value. Therefore B and C must be studied together. However, the main effect of A gives information and shows strong evidence for an increased yield when nitrogen is added:

	TOTAL(kg)	MEAN(kg)
plots with A	4832	302.0
without A	4499	281.2

We know from the p-value that these means differ at the 1% significance level.

Two-way table $B \times C$:

	TOTALS			MEANS	
	C low	high		C low	high
B low	851	2679	B low	106.4	334.9
high	2321	3480	high	290.1	435.0

The least significant difference between two of these means is

$$t_{(24)}\sqrt{\frac{2\times 335.906}{8}} = 9.164 \times \begin{cases} 2.064(5\%) = 18.91 \\ 2.797(1\%) = 25.63 \\ 3.745(0.1\%) = 34.32 \end{cases},$$

showing that all four means differ at 0.1%. B and C both give very large increases when used alone, but when together the effect is exceedingly large (4 times more yield than without either).

(ii) Random allocation provides a basis for any valid statistical test and also gives practical help in avoiding any systematic layout that could be said to bias results in favor of, or against, any particular treatment combination. (If a layout looks 'systematic' after randomization, this is due to sampling accident rather than deliberate choice.)

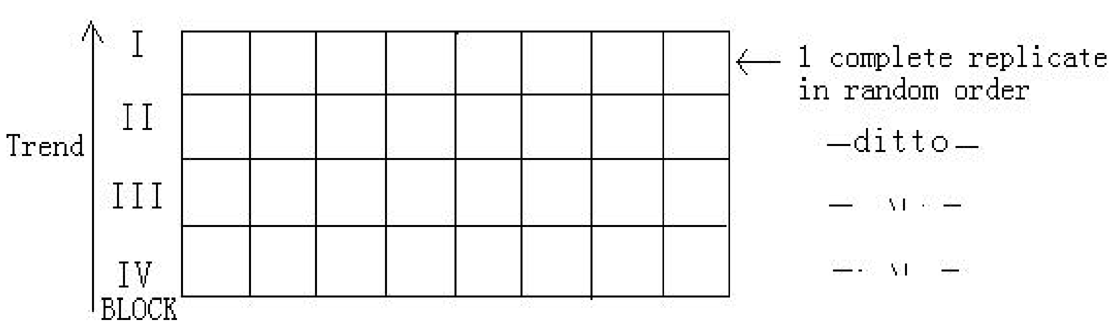
(iii) Within the available plots, allocate to them the numbers 01-32. Take pairs of random digits, and if any of 01-32 occur they immediately locate a plot. 33-64 have 32 subtracted, to give 01-32 again; 65-96 likewise have 64 subtracted. 00 and 97, 98, 99 are not used.

For example,, $8725037441182936555000\cdots$ gives

$87, 25, 03, 74, 41, 18, 29, 36, 55, 50, 00, \cdots$, which reduce to

$23, 25, 03, 10, 09, 18, 29, 04, 23, 18, 00, \cdots$

The first four of these carry treatment (1), the next four a, the next four b, and so on, until bc; then the four remaining must carry abc.



(If the trend had been the other way we could use two columns meet to one another as a block.)

This is a randomized complete block design.

4. (i) Consider stratum h: the sample mean from that stratum is an unbiased estimate of the population mean in the stratum. By taking a (0,1) random variable as the observation, with $y = 0$ if the accommodation is not rented and $y = 1$ if it is, suppose we obtain A_h rented in stratum h and a_h in a. Simple random sample from that stratum.

Then $\bar{Y}_h = \dfrac{A_h}{N_h} = P_h$ and $\bar{y_h} = \dfrac{a_h}{n_h} = p_h$.

Since $E(\bar{y}_h) = \bar{Y}_h$, it follows that $E(p_h) = P_h$.

For the whole city, the estimated mean from a stratified sample is $\sum_{h=1}^{L} \dfrac{N_h}{N} p_h \equiv p_{st}$

and

$$E(p_{st}) = \sum \frac{N_h}{N} E(p_h) = \sum \frac{N_h P_h}{N} = P.$$

(ii) In general, $V(\bar{y}_{st}) = \dfrac{1}{N^2} \sum_{h=1}^{L} N_h (N_h - n_h) \dfrac{S_h^2}{n_h}$. When the (0,1) variable above replaces y, $S_h^2 = \dfrac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{ih} - \bar{Y}_h)^2 = \dfrac{1}{N_h - 1} (\sum_{i=1}^{N_h} y_{ih}^2 - N_h \bar{Y}_h^2)$. Because $y = 0$ or 1, this is

$$\frac{1}{N_h - 1} (\sum_{i=1}^{N_h} y_{ih} - N_h \bar{Y}_h^2) = \frac{1}{N_h - 1} (N_h P_h - N_h P_h^2) = \frac{N_h}{N_h - 1} P_h (1 - P_h)$$

Thus $V(p_{st}) = \dfrac{1}{N^2} \sum_{h=1}^{L} \dfrac{N_h^2 (N_h - n_h)}{N_h - 1} \dfrac{P_h Q_h}{n_h}$, where $Q_h = 1 - P_h$.

In simple random sampling within a stratum,

$$E(s_h^2) = E[\frac{1}{n_h - 1} \sum_{i=1}^{N_h} (y_{ih} - \bar{y}_h)^2] = S_h^2; \text{ and } \frac{s_h^2}{n_k} = \frac{p_h q_h}{n_h - 1}$$

by the same argument used above for S_h^2.

Hence an unbiased estimator of $V(p_{st})$ is as given.

(iii)

Stratum	N_h	n_h	p_h	$W_h = N_h/N$	$\frac{p_h q_h}{n_h - 1}$	$1 - f_h$
< 50	1190	40	0.7500	0.5874	0.004808	0.9664
$50 - 100$	523	35	0.5143	0.2581	0.007347	0.9331
$100 - 200$	215	35	0.2000	0.1061	0.004706	0.8372
> 200	98	40	0.1250	0.0484	0.002804	0.5918

in which $f_h = \frac{n_h}{N_h}$, the sampling fraction in stratum h.

$$p_{st} = \sum_{h=1}^{L} \frac{N_h}{N} p_h = 0.6006 \text{ or } 60.06\%.$$

The estimate

$$\begin{aligned} V(p_{st}) &= \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 (1 - f_h) \frac{p_h q_h}{n_h - 1} = \sum_{h=1}^{L} W_h^2 (1 - f_h) \frac{p_h q_h}{n_h - 1} \\ &= 0.001603205 + 0.000456682 + 0.000044351 + 0.00388726 \\ &= 0.002108 \end{aligned}$$

giving a standard error of 0.0459.

(iv) A good sample allocation is $n_h \propto N_h S_h = N_h \sqrt{P_h Q_h}$, and using the sample estimates of P_h and $\sum n_h = 150$, we have the ratio 515.285: 261.393: 86.000: 32.410, so the scale factor is 150/895.088 giving 86; 44; 14; 6.
We have far too many in the third and fourth strata and only half of what we ought to have in the first.

5. (i) The words in italics are vague, with no precise meaning that will be understood by everyone, particularly in different age groups. No time period is suggested: is it over a year, or a month, or in winter, summer etc.?
The second question is not capable of an answer by everyone so will cause non-response. There are many more forms of exercise possible, and many more sports than are listed.

(ii) How often do you exercise (including training, playing sport, “keep fit” etc)

(see next page)

(iii) Personal interviews should gain a high response rate, and ensure that the questions are answered precisely and answers recorded properly; any misunderstandings can be dealt with. Interviewers must be careful not to induce bias by stressing any answer more than others or by suggesting answers.
Telephone “interviewing” is cheaper, but people are more difficult to obtain

PLEASE TICK ONE BOX

- [ ] more than twice a week
- [ ] twice a week
- [ ] once a week
- [ ] once a fortnight
- [ ] once a month
- [ ] less than once a month
- [ ] never

PLEASE TICK ALL THE FORMS OF EXERCISE YOU TAKE FROM THE FOLLOWING LIST:

☐	tennis	☐	football/ragby football
☐	squash	☐	hockey
☐	badminton	☐	cricket
☐	swimming	☐	riding
☐	walking	☐	running
☐	keep fit	☐	gym
☐	rowing	☐	sailing
☐	others-PLEASE LIST ____________		

[Note - others, such as skiing, may be included for some contries - and the list can be amended depending on local facilities.]

and the basic sampling frame may not be good. Response may be lower, and the questions usually must be fewer, because telephone interviewing annoys some people and cooperation is lower.

Postal questionnaires are cheapest. Problems of people not being at home are avoided. In a scattered area, the whole of it can be sampled, without excessive travelling as in personal interviewing. Response rate tens to be low initially and follow up is needed; more time must be allowed for collecting responses. Questions need to be simple and straightforward, and preferably not very many of them. An explanatory leaflet can be useful towards improving the quality of answers.

6. Denote the 1989 figures by y and the 1980 figures by x. Then $N = 19$, $n = 6$, $\sum_{i=1}^{6} y_i = 327$, $\sum_{i=1}^{6} x_i = 245$, $T_x = 674$ (for 1980).

   Hence in the sample $\bar{y} = \frac{327}{6} = 54.50$ and $\bar{x} = \frac{145}{6} = 40.83$.

   Also $\sum_{i=1}^{6} y_i^2 = 22131$, $\sum_{i=1}^{6} x_i^2 = 11991$, $\sum_{i=1}^{6} x_i y_i = 16196$.

(i) (a) Using a simple random sample of these 6 data for 1989,

$$\hat{Y} = N\bar{y} = 19 \times 54.5 = \$1035.50.$$

(b) Ratio estimator $\hat{Y}_R = T_x \cdot \bar{y}/\bar{x} = 674 \times \frac{54.50}{40.83} = \$899.58$.

(c) $\bar{Y}_{LR} = \bar{y} + b(\mu - \bar{x})$ where μ is the 1980 mean and

$$\begin{aligned}\hat{b} &= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\sum x_i y_i - \frac{1}{6}(\sum x_i)(\sum y_i)}{\sum x_i^2 - \frac{1}{6}(\sum x_i)^2} = \frac{2843.50}{1986.83} \\ &= 1.4312.\end{aligned}$$

$$\bar{Y}_{LR} = 54.50 + 1.4312(\frac{674}{19} - 40.83) = 54.50 - 7.67 = 46.83$$

and $\hat{Y}_{LR} = N\bar{y}_{LR} = \$889.85$.

(ii) There is a considerable difference between μ and $\bar{x}$, and the SRS estimator makes no allowance for this, which would be important assuming there is a relation between x and y. It is clear from the six sample pairs that this is so. We therefore gain information by using this relation.

Unless there is good reason to expect x and y to have a linear relation through the origin, linear regression should give a better estimate than ratio. In this case, there is little difference, but linear regression would be preferred.

(iii) Estimated $V(\hat{Y}) = N^2(1-f)s^2/n$, where $s^2 = \frac{1}{5}(\sum_{i=1}^{6} y_i^2 - \frac{(\sum y_i)^2}{6}) = 861.90$

i.e. $s = 29.358$. Hence $\hat{V}(\hat{y}) = 19^2(1 - \frac{6}{19})\frac{861.9}{6} = 35481.55$, SE= 188.4.

By the ratio method, estimated variance is

$$\begin{aligned}\hat{V}(\hat{y}_R) &= N^2(\frac{1-f}{n})\left\{s_y^2 - 2Rs_{xy} + R^2 s_x^2\right\} = 19^2\frac{(1-\frac{6}{19})}{6}\left\{861.90 - 2\frac{\bar{y}}{\bar{x}}s_{xy} + \frac{\bar{y}^2}{\bar{x}^2}s_x^2\right\} \\ &= 41.1667\left\{861.90 - 2\frac{1.3347}{5}(16196 - \frac{327 \times 245}{6}) + \frac{1.3347^2}{5}(11991 - \frac{245^2}{6})\right\} \\ &= 41.1667(861.90 - \frac{7090 \cdot 44}{5} + \frac{3539 \cdot 47}{5}) = 2128.57.\end{aligned}$$

Using linear regression, estimated variance is

$$\begin{aligned}V(\hat{Y}_{LR}) &= N^2(\tfrac{1-f}{n})\left\{s_y^2 - sts_{xy} + t^2 s_x^2\right\} = 41.1667\left\{861.90 - 2 \times 1.4312 s_{xy} + 1.4312^2 s_x^2\right\} \\ &= 41.1667(8619 - 2.8624 \times \tfrac{2843.60}{5} + 1.4312^2 \times \tfrac{1986.83}{5}) = 1975.64.\end{aligned}$$

Relative efficiency of $\hat{Y}_R$ to $\hat{Y}_{LR}$ is $\frac{1975.64}{2128.57} \times 100 = 92.8\%$.

(i.e. $\hat{Y}_{LR}$ to $\hat{Y}_R$ is 107.7%)

Hence linear regression is slightly more efficient.

Compared with SRS, $\frac{V(\text{ratio})}{V(\text{SRS})} = \frac{2128.57}{35481.55} = 6\%$, or the efficiency of SRS relative to ratio is only 6%, so the efficiency of ratio relative to SRS is 1666.7%.

$\frac{V(\text{LR})}{V(\text{SRS})} = \frac{1975.64}{35481.55} = 5.6\%$, so SRS is only 5.6% of the efficiency of LR, or LR efficiency relative to SRS is 1796.0%.

These results are in agreement with (ii).

7. (i) A first-order model is suitable in the early part of a research programme when the experimental region may not contain the maximum or minimum value which is being sought; the model $y = a + b_1 x_1 + b_2 x_2$ is fitted to response values y and the gradient coefficients t_1, t_2 show the directions in which the experimental values of x_1, x_2 should more for subsequent experiments.

An orthogonal design has a diagonal $(X'X)$ matrix which leads to easy arithmetic and independent estimates of the parameters. For a first-order model, an orthogonal design is rotatable and gives minimum-variance estimates of a, b_1, b_2. A rotatable design in general if $Var(\hat{y})$ depends only on the distance of the experimental point x from the design center.

(ii). (1) $X = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$, so that $X'X = \begin{pmatrix} 8 & & & 0 \\ & 8 & & \\ & & 8 & \\ 0 & & & 8 \end{pmatrix}$

and $V(b) = \sigma^2 (X'X)^{-1} = \sigma^2 \begin{pmatrix} 1/8 & & & 0 \\ & 1/8 & & \\ & & 1/8 & \\ 0 & & & 1/8 \end{pmatrix}$

where $b_0 \equiv a$ and b_1, b_2, b_3 refer to the three experimental x-variables. $Var(b_i) = \sigma^2/8$, $i = 0, 1, 2, 3$.

At distance ρ from the center, $\rho^2 = x_1^2 + x_2^2 + x_3^2$ and

$$V(\hat{y}) = V(b_0) + \sum_{i=1}^{3} x_i^2 V(b_i) = \frac{1}{8}\sigma^2(1 + \rho^2).$$

(2) $X = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$, and $X'X = \begin{pmatrix} 8 & & & 0 \\ & 4 & & \\ & & 4 & \\ 0 & & & 4 \end{pmatrix}$

with $(X'X)^{-1}\sigma^2 = \sigma^2 \begin{pmatrix} 1/8 & & & 0 \\ & 1/4 & & \\ & & 1/4 & \\ 0 & & & 1/4 \end{pmatrix}$

so that $V(b_0) = \sigma^2/8$ and $V(b_i) = \sigma^2/4$ for $i = 1, 2, 3$. Hence at distance ρ from the center, $V(\hat{y}) = \dfrac{\sigma^2}{8} + \dfrac{\rho^2\sigma^2}{4} = \dfrac{\sigma^2}{8}(1 + 2\rho^2)$.

(3) $X = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$, and $X'X = \begin{pmatrix} 8 & & & 0 \\ & 8 & & \\ & & 8 & \\ 0 & & & 8 \end{pmatrix}$

with $(X'X)^{-1}\sigma^2 = \sigma^2 \begin{pmatrix} 1/8 & & & 0 \\ & 1/8 & & \\ & & 1/8 & \\ 0 & & & 1/8 \end{pmatrix}$

and all $V(b_i)$ are $\sigma^2/8$. Hence at distance ρ from the center, $V(\hat{y}) = \dfrac{\sigma^2}{8}(1 + \rho^2)$.

Designs (1) and (3) are minimum-variance; design (2) is rotatable and orthogonal, like the others, but not minimum-variance. In fact the variance in (2) increased quite quickly the larger ρ is.

(iii). In (1) there is no replication so no "pure error" among the 4 $d.f.$ for residual. But it would be suitable where it is expected that further experiments will be needed anyway.

(2) has 3 "pure error" $d.f.$ for testing whether linearity is adequate, and 1 $d.f.$ for quadratic but no provision for interaction. Therefore it can be used

if curvature is suspected but no interaction.

In (3) the residual 4 $d.f.$ are not associated with interaction or quadratic terms, but the coefficients in a first-order model are capable of being tested. It is useful when quadratic or interaction effects are to be estimated, although the adequacy of a model cannot be tested.

8. (i) n_i is the number exposed to risk in the i^{th} time interval. It is $n_i - \frac{1}{2}(l_i + w_i)$, assuming a uniform distribution of loss within the interval.

$\hat{q}_i$ is the conditional proportion dying; $\hat{q}_i = d_i/n_i$ for $i = 1$ to $s-1$ and $\hat{q}_s = 1$ for the fast interval s, and is an estimate of the conditional probability of death in interval i given that the individual is exposed to risk of death in this interval.

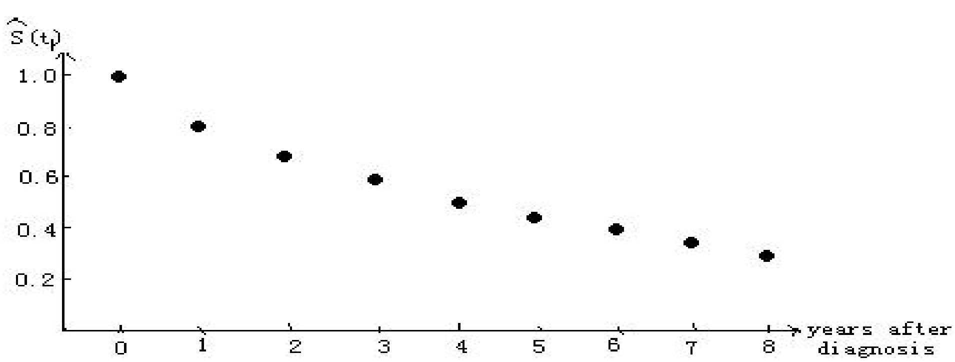
$\hat{p}_i$ is the conditional proportion surviving, $= 1 - \hat{q}_i$.

$\hat{S}(t_i)$ is the cumulative proportion surviving, an estimate of the survivorship function at time t_i ("cumulative survival rate"). $\hat{S}(t_1) = 1$, and $\hat{S}(t_i) = \hat{p}_{i-1}\hat{S}(t_{i-1})$ for $i = 2, 3, \cdots s$.

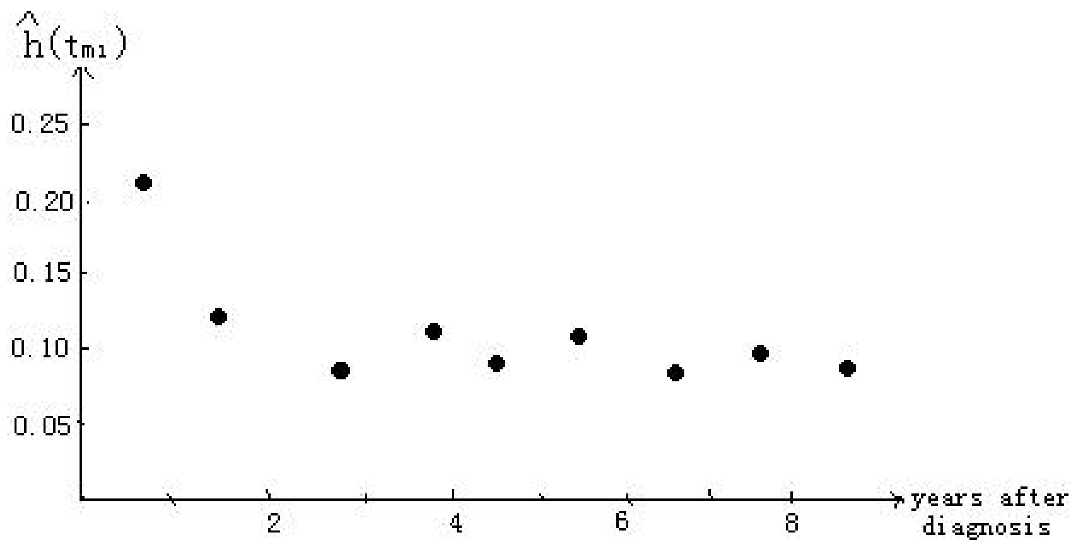
(ii)(iii).

Year	t_i	w_i	d_i	n_i'	n_i	$\hat{q}_i$	$\hat{p}_i$	$\hat{S}(t_i)$	$\hat{h}(t_{mi})$
0	0	0	456	2418	2418	0.1886	0.8114	1.0000	0.2082
1	9	30	226	1962	1942.5	0.1163	0.8837	0.8114	0.1235
2	10	12	152	1697	1686	0.0902	0.9098	0.7170	0.0945
3	0	23	171	1523	1511.5	0.1131	0.8869	0.6524	0.1199
4	9	15	135	1329	1317	0.1025	0.8975	0.5786	0.1080
5	10	97	125	1170	1116.5	0.1120	0.8880	0.5193	0.1186
6	25	108	83	938	871.5	0.0952	0.9048	0.4611	0.1000
7	15	87	74	722	671	0.1103	0.8897	0.4172	0.1167
8	8	60	51	546	512	0.0996	0.9004	0.3712	0.1048

(iv).



This shows a reasonably smooth curve, and the median survival time ($s = 0.5$) is about 5.3 years.



The death rate is highest in the first year after diagnosis. After this it fluctuates in the region of $0:1$, so that a patient who has survived one year has a better prognosis than at the beginning (subject to other factor such as ages, sex, racial group).