Statistics 583, Problem Set 6 Solutions

Wellner; 5/3/2000

1. Suppose that X_{ijk}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$ satisfy the general linear model with $\xi_{ijk} = \xi + \mu_i + \eta_j + \delta_{ij}$ where $\sum_i \mu_i = 0$, $\sum_j \eta_j = 0$, $\sum_j \delta_{ij} = 0$ for all i, and $\sum_i \delta_{ij} = 0$ for all j. (δ_{ij} is called the interaction effect of the ith row and the jth column.)
(a) Show that

$$
\begin{aligned}
S^2 &= \sum\sum\sum(X_{ijk} - \xi - \mu_i - \eta_j - \delta_{ij})^2 \\
&= \sum\sum\sum(X_{ijk} - \overline{X}_{ij\cdot})^2 \\
&\quad + \sum\sum\sum(\overline{X}_{ij\cdot} - \overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot j\cdot} + \overline{X}_{\cdots} - \delta_{ij})^2 \\
&\quad + \sum\sum\sum(\overline{X}_{i\cdot\cdot} - \overline{X}_{\cdots} - \mu_i)^2 \\
&\quad + \sum\sum\sum(\overline{X}_{\cdot j\cdot} - \overline{X}_{\cdots} - \eta_j)^2 \\
&\quad + \sum\sum\sum(\overline{X}_{\cdots} - \xi)^2
\end{aligned}
$$

where $\overline{X}_{ij\cdot} = \sum_k X_{ijk}/K$, and so on.
(b) Find the UMP invariant test of the hypothesis of no row effect $H_0 : \mu_1 = \ldots = \mu_I = 0$. What is the distribution of the test statistic under the general linear hypothesis – including the noncentrality parameter?
(c) Find the UMP invariant test of the hypothesis of no interaction effect $H_0 : \delta_{ij} = 0$ for all i, j. What is the distribution of the test statistic under the general linear hypothesis?

Solution: (a) This goes by straightforward addition, and subtraction, followed by algebra:

$$
\begin{aligned}
S^2 &= \sum\sum\sum(X_{ijk} - \xi - \mu_i - \eta_j - \delta_{ij})^2 \\
&= \sum\sum\sum \Big(X_{ijk} - \overline{X}_{ij\cdot} \\
&\qquad + \overline{X}_{ij\cdot} - \overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot j\cdot} + \overline{X}_{\cdots} - \delta_{ij} \\
&\qquad + \overline{X}_{i\cdot\cdot} - \overline{X}_{\cdots} - \mu_i \\
&\qquad + \overline{X}_{\cdot j\cdot} - \overline{X}_{\cdots} - \eta_j \\
&\qquad + \overline{X}_{\cdots} - \xi\Big)^2 \\
&= \sum\sum\sum \left(X_{ijk} - \overline{X}_{ij\cdot}\right)^2 \\
&\quad + \sum\sum\sum \left(\overline{X}_{ij\cdot} - \overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot j\cdot} + \overline{X}_{\cdots} - \delta_{ij}\right)^2 \\
&\quad + \sum\sum\sum \left(\overline{X}_{i\cdot\cdot} - \overline{X}_{\cdots} - \mu_i\right)^2 \\
&\quad + \sum\sum\sum \left(\overline{X}_{\cdot j\cdot} - \overline{X}_{\cdots} - \eta_j\right)^2 \\
&\quad + \sum\sum\sum \left(\overline{X}_{\cdots} - \xi\right)^2
\end{aligned}
$$

Since the cross terms all vanish.
(b) For testing the hypothesis of no row effect, the numerator of the $F-$statistic is

$$\begin{aligned}\sum\sum\sum \hat{\mu}_i^2 &= \sum\sum\sum(\overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot\cdot\cdot})^2 \\ &= JK\sum_{i=1}^{I}(\overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot\cdot\cdot})^2\end{aligned}$$

divided by its degrees of freedom, $I-1$. The denominator of the $F-$statistic is the error sum of squares

$$SS_e \equiv \sum\sum\sum\left(X_{ijk} - \overline{X}_{ij\cdot}\right)^2$$

divided by its degrees of freedom, $IJK - IJ = IJ(K-1)$. Thus the $F-$statistic for testing no row effect is

$$\begin{aligned}F &\equiv \frac{JK\sum_{i=1}^{I}(\overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot\cdot\cdot})^2/(I-1)}{\sum\sum\sum\left(X_{ijk} - \overline{X}_{ij\cdot}\right)^2/IJ(K-1)} \\ &\sim F_{I-1,IJ(K-1)}\end{aligned}$$

under H_0. Under the general hypothesis $F \sim F_{I-1,IJ(K-1)}(\delta^2)$ where the non-centrality parameter

$$\delta^2 = \frac{JK\sum_1^I \mu_i^2}{\sigma^2}.$$

(c) For testing the hypothesis of no interactions, the numerator of the $F-$statistic is

$$\begin{aligned}\sum\sum\sum \hat{\delta}_{ij}^2 &= \sum\sum\sum\left(\overline{X}_{ij\cdot} - \overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot j\cdot} + \overline{X}_{\cdot\cdot\cdot} - \delta_{ij}\right)^2 \\ &= K\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\overline{X}_{ij\cdot} - \overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot j\cdot} + \overline{X}_{\cdot\cdot\cdot} - \delta_{ij}\right)^2\end{aligned}$$

divided by its degrees of freedom, $(I-1)(J-1)$. The denominator of the $F-$statistic is the error sum of squares

$$SS_e \equiv \sum\sum\sum\left(X_{ijk} - \overline{X}_{ij\cdot}\right)^2$$

divided by its degrees of freedom, $IJK - IJ = IJ(K-1)$. Thus the $F-$statistic for testing no row effect is

$$\begin{aligned}F &\equiv \frac{K\sum_{i=1}^{I}\sum_{j=1}^{J}(\overline{X}_{ij\cdot} - \overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot j\cdot} + \overline{X}_{\cdot\cdot\cdot} - \delta_{ij})^2/(I-1)(J-1)}{\sum\sum\sum(X_{ijk} - \overline{X}_{ij\cdot})^2/IJ(K-1)} \\ &\sim F_{(I-1)(J-1),IJ(K-1)}\end{aligned}$$

under H_0. Under the general hypothesis $F \sim F_{(I-1)(J-1),IJ(K-1)}(\delta^2)$ where the non-centrality parameter

$$\delta^2 = \frac{K\sum_1^I\sum_{j=1}^J \delta_{ij}^2}{\sigma^2}.$$

2. Suppose that X_i, $i = 1, \ldots, I$ and Y_j, $j = 1, \ldots, J$ satisfy the general linear hypothesis with $E(X_i) = \alpha_1 + \beta_1 u_i$ and $E(Y_j) = \alpha_2 + \beta_2 v_j$ where the u_i and v_j are known and $\sum u_i = \sum v_j = 0$, $\sum u_i^2 = I$, $\sum v_j^2 = J$. Find the UMP invariant test of $H : \beta_1 = \beta_2$ versus $K : \beta_1 \neq \beta_2$. What is the noncentrality parameter δ^2 for your test?

Solution: Here $\underline{Z}^T = (\underline{X}^T, \underline{Y}^T)^T \sim N_n(\underline{\xi}, \sigma^2 I)$ with $n = I + J$, $k = 4$, and

$$\underline{\xi} = E(\underline{Z}) = A\underline{\theta} = \begin{pmatrix} 1 & u_1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & u_I & 0 & 0 \\ 0 & 0 & 1 & v_1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & v_J \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \end{pmatrix}.$$

The least squares estimator of θ is

$$\hat{\underline{\theta}}_n = (A^T A)^{-1} A^T \underline{Z} = \begin{pmatrix} \widehat{\alpha}_1 \\ \widehat{\beta}_1 \\ \widehat{\alpha}_2 \\ \widehat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \overline{X} \\ \overline{uX} \\ \overline{Y} \\ \overline{vY} \end{pmatrix}.$$

(a) Here the hypothesis is $H_0 :\ \beta_1 = \beta_2$, or $(0, 1, 0, -1)\underline{\theta} = 0$, so $r = 1$. In this case

$$\underline{\xi} = E(\underline{Z}) = A\underline{\theta} = \begin{pmatrix} 1 & 0 & u_1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & 0 & u_I \\ 0 & 1 & v_1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 1 & v_J \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{pmatrix}.$$

Therefore the least squares estimator of θ under the hypothesis is

$$\hat{\hat{\underline{\theta}}}_n = (A^T A)^{-1} A^T \underline{Z} = \begin{pmatrix} \widehat{\widehat{\alpha}}_1 \\ \widehat{\widehat{\alpha}}_2 \\ \widehat{\widehat{\beta}} \end{pmatrix} = \begin{pmatrix} \overline{X} \\ \overline{Y} \\ \frac{I}{I+J}\overline{uX} + \frac{J}{I+J}\overline{vY} \end{pmatrix}.$$

We then compute

$$\begin{aligned} ||\widehat{\underline{\xi}} - \widehat{\widehat{\underline{\xi}}}||^2 &= \sum_{i=1}^{I} (\widehat{\alpha}_1 + \widehat{\beta}_1 u_i - \widehat{\widehat{\alpha}}_1 - \widehat{\widehat{\beta}}_1 u_i)^2 + \sum_{j=1}^{J} (\widehat{\alpha}_2 + \widehat{\beta}_2 v_j - \widehat{\widehat{\alpha}}_2 - \widehat{\widehat{\beta}}_2 v_j)^2 \\ &= \sum_{i=1}^{I} u_i^2 (\widehat{\beta}_1 - \widehat{\widehat{\beta}}_1)^2 + \sum_{j=1}^{J} v_j^2 (\widehat{\beta}_2 - \widehat{\widehat{\beta}}_2)^2 \\ &= I(\widehat{\beta}_1 - \widehat{\widehat{\beta}})^2 + J(\widehat{\beta}_2 - \widehat{\widehat{\beta}})^2 \\ &= \frac{IJ}{I+J}(\overline{uX} - \overline{vY})^2 \end{aligned}$$

and

$$\begin{aligned}
||\underline{Z}-\widehat{\underline{\xi}}||^2 &= \sum_{i=1}^{I}(X_i-\widehat{\alpha}_1-\widehat{\beta}_1 u_i)^2+\sum_{j=1}^{J}(Y_j-\widehat{\alpha}_2+\widehat{\beta}_2 v_j)^2\\
&= \sum_{i=1}^{I}(X_i-\overline{X}-u_i\overline{uX})^2+\sum_{j=1}^{J}(Y_j-\overline{Y}-v_j\overline{vY})^2\\
&= \sum_{i=1}^{I}(X_i-\overline{X})^2-I(\overline{uX})^2+\sum_{j=1}^{J}(Y_j-\overline{Y})^2-J(\overline{vY})^2\,.
\end{aligned}$$

Therefore we find that

$$F=\frac{||\widehat{\underline{\xi}}-\widehat{\widehat{\underline{\xi}}}||^2/r}{||\underline{Z}-\widehat{\underline{\xi}}||^2/(n-k)}=\frac{\frac{IJ}{I+J}(\overline{uX}-\overline{vY})^2}{\frac{1}{I+J-4}(\sum_{i=1}^{I}(X_i-\overline{X})^2-I(\overline{uX})^2+\sum_{j=1}^{J}(Y_j-\overline{Y})^2-J(\overline{vY})^2)}$$

which has an $F_{1,I+J-4}$ distribution under the null hypothesis. Hence the UMP G-invariant test is: "reject H_0 if $F>F_{1,I+J-4,\alpha}$". The non-centrality parameter is

$$\begin{aligned}
\delta^2 &= \frac{||\underline{\xi}-\underline{\xi}^0||^2}{\sigma^2}\\
&= \frac{1}{\sigma^2}\left\{\sum_{1}^{I}(\alpha_1+\beta_1u_i-(\alpha_1+\overline{\beta}u_i))^2+\sum_{1}^{J}(\alpha_2+\beta_2v_j-(\alpha_2+\overline{\beta}v_j))^2\right\}\\
&= \frac{1}{\sigma^2}\left\{\sum_{1}^{I}(\beta_1-\overline{\beta})^2u_i)^2+\sum_{1}^{J}(\beta_2-\overline{\beta})^2v_j)^2\right\}\\
&= \frac{1}{\sigma^2}\frac{IJ}{I+J}(\beta_1-\beta_2)^2\,.
\end{aligned}$$

(b) In this case the null hypothesis is $H_0:\alpha_1=\alpha_2,\quad \beta_1=\beta_2$, or equivalently

$$\begin{pmatrix}1&0&-1&0\\0&1&0&-1\end{pmatrix}\theta=\underline{0}\,,$$

so $r=2$. Furthermore, under the null hypothesis we can write

$$\underline{\xi}=E(\underline{Z})=A\underline{\theta}=\begin{pmatrix}1&u_1\\ \cdot&\cdot\\ \cdot&\cdot\\ 1&u_I\\ 1&v_1\\ \cdot&\cdot\\ \cdot&\cdot\\ 1&v_J\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix},$$

and this leads by straightforward computation to the least squares estimator

$$\hat{\hat{\underline{\theta}}}_n=(A^TA)^{-1}A^T\underline{Z}=\begin{pmatrix}\widehat{\widehat{\alpha}}\\ \widehat{\widehat{\beta}}\end{pmatrix}=\begin{pmatrix}\frac{I}{I+J}\overline{X}+\frac{J}{I+J}\overline{Y}\\ \frac{I}{I+J}\overline{uX}+\frac{J}{I+J}\overline{vY}\end{pmatrix}.$$

We therefore find that

$$\begin{aligned}
||\underline{\widehat{\xi}} - \underline{\widehat{\widehat{\xi}}}||^2 &= \sum_{i=1}^{I}(\widehat{\alpha}_1 + \widehat{\beta}_1 u_i - \widehat{\widehat{\alpha}}_1 - \widehat{\widehat{\beta}}_1 u_i)^2 + \sum_{j=1}^{J}(\widehat{\alpha}_2 + \widehat{\beta}_2 v_j - \widehat{\widehat{\alpha}}_2 - \widehat{\widehat{\beta}}_2 v_j)^2 \\
&= \sum_{i=1}^{I}(\widehat{\alpha}_1 + \widehat{\beta}_1 u_i - \widehat{\widehat{\alpha}} - \widehat{\widehat{\beta}} u_i)^2 + \sum_{j=1}^{J}(\widehat{\alpha}_2 + \widehat{\beta}_2 v_j - \widehat{\widehat{\alpha}} - \widehat{\widehat{\beta}} v_j)^2 \\
&= \frac{IJ}{I+J}\left\{(\overline{X} - \overline{Y})^2 + (\overline{uX} - \overline{vY})^2\right\} .
\end{aligned}$$

Therefore the F statistic for this testing problem is

$$F = \frac{||\underline{\widehat{\xi}} - \underline{\widehat{\widehat{\xi}}}||^2/r}{||\underline{Z} - \underline{\widehat{\xi}}||^2/(n-k)} = \frac{\frac{IJ}{2(I+J)}\{(\overline{X} - \overline{Y})^2 + (\overline{uX} - \overline{vY})^2\}}{\frac{1}{I+J-4}(\sum_{i=1}^{I}(X_i - \overline{X})^2 - I(\overline{uX})^2 + \sum_{j=1}^{J}(Y_j - \overline{Y})^2 - J(\overline{vY})^2)}$$

which has an $F_{2,I+J-4}$ distribution under the null hypothesis. Hence the UMP G-invariant test is: "reject H_0 if $F > F_{2,I+J-4,\alpha}$".

3. What is the locally best rank test of $F = G$ against $G = (e^{\theta F} - 1)/(e^{\theta} - 1)$, $\theta > 0$? Of $F = G$ against $G = F/(e^{\theta}(1-F) + F)$?
What can you say about the power of these tests?

Solution: By Hoeffding's formula

$$P_\theta(\underline{Q} = \underline{q}) = \frac{1}{\binom{N}{n}} E_{uniform}\left\{\prod_{j=1}^{n} \psi'_\theta(U_{(q_j)})\right\}$$

where

$$\psi_\theta(u) = G_\theta \circ F^{-1}(u) = \frac{e^{\theta u} - 1}{e^{\theta} - 1}$$

for the first alternatives, and

$$\psi_\theta(u) = G_\theta \circ F^{-1}(u) = \frac{u}{e^{\theta}(1-u) + u}$$

in the case of the second type of alternative. In either case the locally most powerful rank test rejects for those values $\underline{q}$ of $\underline{Q}$ which make

$$\begin{aligned}
\frac{\partial}{\partial\theta} P_\theta(\underline{Q} = \underline{q})|_{\theta=0} &= \frac{1}{\binom{N}{n}} E_{uniform}\left\{\sum_{j=1}^{n} \frac{\partial}{\partial\theta}\psi'_\theta(U_{(q_j)})|_{\theta=0}\right\} \\
&= \sum_{j=1}^{n} E_{uniform}\left\{\frac{\partial}{\partial\theta}\psi'_\theta(U_{(q_j)})|_{\theta=0}\right\}
\end{aligned}$$

as large as possible. Hence it remains only to calculate

$$\phi(u) \equiv \frac{\partial}{\partial\theta}\psi'_\theta(u)|_{\theta=0}$$

and $E_{uniform}\phi(U_{(i)})$ for the two alternatives in question.
(i) In the first case,

$$\psi_\theta'(u) = \frac{\theta e^{\theta u}}{e^\theta - 1},$$

and straightforward calculation yields

$$\frac{\partial}{\partial\theta}\psi_\theta'(u) = e^{\theta u}\frac{(e^\theta - 1)(1 + \theta u - \theta) - \theta}{(e^\theta - 1)^2}.$$

By applying L'Hopital's rule twice, we find that

$$\frac{\partial}{\partial\theta}\psi_\theta'(u)|_{\theta=0} = u - \frac{1}{2}.$$

Since $E(U_{(i)}) = i/(N+1)$, the locally most powerful rank test of H versus this alternative K is the Wilcoxon test "reject H if $S_N = \sum_{j=1}^n Q_j > k_\alpha$.
(ii) In the second case,

$$\psi_\theta'(u) = \frac{e^\theta}{(e^\theta(1-u) + u)^2}.$$

Hence

$$\frac{\partial}{\partial\theta}\psi_\theta'(u)|_{\theta=0} = 2u - 1,$$

and again the locally most powerful rank test is the Wilcoxon rank sum test.
As for interpretations of these alternatives, first note that the functions $\psi_\theta(u)$ are distribution functions on $[0,1]$ with densities $\psi_\theta'(u)$. (i) This alternative is the simplest exponential family density related to the uniform$(0,1)$ distribution: the density is of the form $p_\theta(u) = \psi_\theta'(u) = c(\theta)\exp(\theta u)1_{[0,1]}(u)$.
(ii) For this family, note that

$$1 - \psi_\theta(u) = \frac{e^\theta(1-u)}{e^\theta(1-u) + u},$$

and hence the *odds ratio* is

$$\frac{1 - \psi_\theta(u)}{\psi_\theta(u)} = e^\theta\frac{1-u}{u} = e^\theta \cdot \text{the odds ratio for Uniform(0,1)}.$$

Thus this family is one with proportional odds ratios.

4. Suppose that $X_1, \ldots, X_n$ are independent Exponential(1) random variables. Let $Y_i \equiv X_{(i)}$, for $i = 1, \ldots, n$, denote the *order statistics* corresponding to $X_1, \ldots, X_n$.
(a) Show that the vector $(Y_1, \ldots, Y_n)$ has the same joint distribution as $(W_1, \ldots, W_n)$ where

$$W_i \equiv \sum_{j=1}^{i} Z_j/(n-j+1) \tag{1}$$

and $Z_1, \ldots, Z_n$ are i.i.d. Exponential(1).
(b) Use the result of (a) to compute $E(Y_i)$, $Var(Y_i)$, and $Cov(Y_i, Y_j)$ for any fixed i, j.

Solution: (a) Note that $0 \le W_1 \le \ldots \le W_n$ and

$$Z_i = (n-i+1)(W_i - W_{i-1}), \qquad i = 1, \ldots, n \tag{2}$$

(with $W_0 \equiv 0$). Let $g(\underline{Z}) \equiv \underline{W}$ be the map defined by (1) so that $g^{-1}(\underline{W}) = \underline{Z}$ is given in (2). Then the Jacobian of g^{-1} has entries n, $(n-1), \ldots, 1$ on the diagonal, entries $-(n-1), \ldots, -2, -1$ below the diagonal, and zero elsewhere. Hence $\det(J_{g^{-1}}) = \mathrm{tr}(J_{g^{-1}}) = n!$ and the density of $\underline{W}$ is given by

$$\begin{aligned}
f_{\underline{W}}(\underline{w}) &= f_{\underline{Z}}(g^{-1}(\underline{w}))\det(J_{g^{-1}}) \\
&= n! \prod_{i=1}^{n} \exp(-(n-i+1)(w_i - w_{i-1})) \\
&= n! \exp(-\sum_{i=1}^{n}(n-i+1)(w_i - w_{i-1})) \\
&= n! \exp(-\sum_{i=1}^{n}(\sum_{j=i}^{n} 1)(w_i - w_{i-1})) \\
&= n! \exp(-\sum_{j=1}^{n}\sum_{i \le j}(w_i - w_{i-1})) \\
&= n! \exp(-\sum_{j=1}^{n} w_j) = n! f(w_1)\cdots f(w_n)
\end{aligned}$$

on the set $0 \le w_1 \le \ldots \le w_n < \infty$ where $f(x) = \exp(-x)1_{[0,\infty)}(x)$ is the standard exponential density. Hence $\underline{Y} =_d \underline{Y} \equiv \underline{X}_{(\cdot)}$ where $X_1, \ldots, X_n$ are i.i.d. exponential(1).

(b) It follows immediately from (a) that

$$E(Y_i) = E(\sum_{j=1}^{i} \frac{Z_j}{n-j+1}) = \sum_{j=1}^{i} \frac{1}{n-j+1},$$

$$Var(Y_i) = Var(\sum_{j=1}^{i} \frac{Z_j}{n-j+1}) = \sum_{j=1}^{i} \frac{1}{(n-j+1)^2},$$

and

$$\begin{aligned}
Cov(Y_i, Y_j) &= Cov\left(\sum_{k=1}^{i} \frac{Z_k}{n-k+1}, \sum_{k'=1}^{i} \frac{Z_{k'}}{n-k'+1}\right) \\
&= \sum_{k=1}^{i \wedge j} \frac{1}{(n-k+1)^2}.
\end{aligned}$$

for any fixed i, j.