The Annals of Statistics
1975, Vol. 3, No. 2, 521-527

ON THE USE OF RANKS FOR TESTING THE COINCIDENCE OF SEVERAL REGRESSION LINES

BY J. N. ADICHIE

University of Nigeria, Nsukka

For several linear regression lines $Y_{ij} = \alpha_i + \beta_i(x_{ij} - x_{i.}) + Z_{ij}$, $i = 1, \cdots, k; j = 1, \cdots, n_i$, a statistic for testing $\alpha_i = \alpha$, $\beta_i = \beta$ is constructed based on the simultaneous ranking of all the observations. The asymptotic properties of the criterion are also studied. The results are, however, not directly applicable to the general design model $Y_{ij} = \alpha_i + \beta_i x_{ij} + Z_{ij}$, unless it is assumed that the group means $x_{i.}$ are all equal.

0. Introduction and summary. In two recent papers Sen (1969, 1972) considered several regression lines $Y_{ij} = \alpha_i + \beta_i x_{ij} + Z_{ij}$, $i = 1, \cdots, k; j = 1, \cdots, n_i$, and studied optimum rank score tests for the separate hypotheses H_1: $\beta_i = \beta$ (unknown), α_i are nuisance parameters and H_2: $\alpha_i = \alpha$ (unknown), β_i are nuisance parameters. In the present paper, we propose rank score tests that discriminate simultaneously against different β's and different α's. The methods of Sen could be used to construct rank order statistics for testing H_0: $\alpha_i = \alpha$, $\beta_i = \beta$. This would, however, involve the estimation of α and the β_i and the combination of the separate k rankings. The alternative method presented below makes use of the simultaneous ranking of all the observations, and involves the estimation of β only. However, our procedure is limited to designs where the group means $x_{i.}$ of x_{ij}'s are all equal.

The proposed test statistic is shown to have a limiting chi-square distribution under the hypothesis and a non-central chi-square under an appropriate sequence of alternatives. The asymptotic efficiency of the given procedure relative to the least squares procedure is also shown to be the familiar efficiency of rank score tests relative to the t-test in the two-sample problem.

1. Notation and assumptions. For each $i = 1, \cdots, k$, let $Y_{ij}, j = 1, \cdots, n_i$; be independent random variables. Also let x_{ij} be known constants that are not all equal. It is assumed that the distribution function $F_{ij}(\cdot)$ of Y_{ij} are given by

$$P(Y_{ij} < y) = F_{ij}(y) = F\{y - \alpha_i - \beta_i(x_{ij} - x_{i.})\} \tag{1.1}$$

where $x_{i.} = n_i^{-1} \sum_j x_{ij}$, F is continuous but its functional form need not be known, and α_i, β_i are the unknown parameters of interest. Our problem is to test the hypothesis

$$H_0: \alpha_i = \alpha \quad \text{(unknown)}, \qquad \beta_i = \beta \quad \text{(unknown)}, \tag{1.2}$$

against the set of alternatives that violate (1.2).

Received November 1972; revised January 1974.

AMS 1970 *subject classifications.* Primary 62G10, 62G20; Secondary 62E20, 62J05.

Key words and phrases. Linear rank statistic, score generating function, bounded in probability, least squares estimates, asymptotic efficiency.

REMARK. The model in (1.1) assumes that the x_{ij}'s have been centered about their group means $x_{i.}$. This assumption, which necessarily limits the scope of our procedure to specially balanced designs, is equivalent to the so-called "orthogonality condition" in the classical least squares regression theory. For testing hypothesis about β_i, test criteria are always available independent of α_i, and without extra condition on the distribution functions F_{ij}. On the other hand, tests about α_i usually depend on the β_i and require some symmetry condition. The symmetry effect may be achieved if either the F_{ij} are symmetric or the x_{ij} are symmetrically balanced as in (1.1) (see e.g., Hájek (1969), Theorem 3F).

In order to maintain the notation in Adichie (1974), we shall actually be considering sequences $\{Y_{nij}\}$ and $\{x_{nij}\}$, $n = 1, 2, \cdots$, of independent random variables, and constants respectively. However, for simplicity of notation, the dependence on n of Y_{ij}, x_{ij} and some of their function, will often be suppressed. We shall therefore write

$$(1.3) \qquad \begin{aligned} C_{ni}^2 &= \textstyle\sum_j (x_{ij} - x_{i.})^2 , \qquad \gamma_{ni} = C_{ni}^2/C_n^2 , \\ \lambda_{ni} &= (n_i/n) , \qquad \rho_{ni} = (n_i/C_n) , \qquad\qquad i = 1, \cdots, k , \end{aligned}$$

where

$$(1.4) \qquad C_n^2 = \textstyle\sum_i C_{ni}^2 ; \qquad n = \sum_i n_i .$$

For all the summations in this paper, i, s, and t go from 1 to k, while j or v goes from 1 to n_i or n_t. All limits are taken as $n \to \infty$. It is assumed that each of the quantities n_i and C_{ni}^2 increases with n in such a way that for each $i = 1, \cdots, k$,

$$(1.5) \qquad \begin{aligned} 0 < \gamma_0 &< (\sup_n \gamma_{ni}) < (1 - \gamma_0) < 1 , \\ 0 < \lambda_0 &\leqq (\sup_n \lambda_{ni}) \leqq (1 - \lambda_0) < 1 , \\ 0 < \rho_0 &\leqq (\sup_n \rho_{ni}) < K , \end{aligned}$$

and $\max(\gamma_0, \lambda_0) < (1/k)$. Throughout this paper K with or without subscripts will denote a generic constant. We shall write

$$(1.6) \qquad \begin{aligned} c_{sj}^{(i)} &= \gamma_{ni}(x_{sj} - x_{s.}) \qquad & s\,(\neq i) = 1, \cdots, k , \\ &= -(1 - \gamma_{ni})(x_{sj} - x_{s.}) \qquad & s = i , \end{aligned}$$

$$(1.7) \qquad \begin{aligned} d_{sj}^{(i)} &= 0 \qquad & s \neq i , \\ &= 1 \qquad & s = i , \end{aligned}$$

so that

$$(1.8) \qquad \bar{d}^{(i)} = n^{-1} \textstyle\sum_s \sum_j d_{sj}^{(i)} = \lambda_{ni} ; \qquad \bar{c}^{(i)} = 0 ,$$

$$(1.9) \qquad \textstyle\sum_s \sum_j (d_{sj}^{(i)} - \bar{d}^{(i)})^2 = \rho_{ni}(1 - \lambda_{ni})C_n^2 ,$$

and

$$(1.10) \qquad \textstyle\sum_s \sum_j (c_{sj}^{(i)} - \bar{c}^{(i)})^2 = \gamma_{ni}(1 - \gamma_{ni})C_n^2 .$$

2. The test statistic. Let $\phi(u)$, $0 < u < 1$, be a smooth non-decreasing function, and let the scores generated by ϕ be defined by

$$a_n(p) = \phi\{p/(n+1)\}\,, \qquad p = 1, \cdots, n\,. \tag{2.1}$$

Also let R_{ij} be the rank of Y_{ij} in the combined ranking of all the n observations. For the unknown β in (1.2), we shall require an estimate $\hat\beta$ defined in Adichie (1974). For ease of reference, the estimate is

$$\hat\beta = \tfrac{1}{2}(\beta_n{}^* + \beta_n{}^{**})\,, \tag{2.2}$$

where

$$\beta_n{}^* = \sup\{b : S_n(Y - bx) > 0\}\,, \qquad \beta_n{}^{**} = \inf\{b : S_n(Y - bx) < 0\}\,, \tag{2.3}$$

and $S_n(Y - bx)$ denotes the statistic

$$S_n(Y) = \textstyle\sum_i \sum_j (x_{ij} - x_{i.})a_n(R_{ij})\,, \tag{2.4}$$

when the observations Y_{ij} are replaced by $\{Y_{ij} - b(x_{ij} - x_{i.})\}$.

Now write $\hat Y_{ij} = \{Y_{ij} - \hat\beta(x_{ij} - x_{i.})\}$, and let $\hat R_{ij}$ be the rank of $\hat Y_{ij}$. For each $i = 1, \cdots, k$, define

$$\hat T_{\alpha ni} = \textstyle\sum_s \sum_j (d_{sj}^{(i)} - \bar d^{(i)})a_n(\hat R_{sj})\,, \tag{2.5}$$

$$\hat T_{\beta ni} = \textstyle\sum_s \sum_j c_{sj}^{(i)} a_n(\hat R_{sj})\,. \tag{2.6}$$

Also for each i, let

$$\hat V_{ni} = n^{-\frac{1}{2}}(\hat T_{\alpha ni}/A)\,; \qquad \hat U_{ni} = (\hat T_{\beta ni}/AC_{ni})\,, \tag{2.7}$$

where

$$A^2 = \textstyle\int \phi^2(u)\,du - \{\int \phi(u)\,du\}^2\,. \tag{2.8}$$

The proposed test statistic is

$$M_n = \textstyle\sum_i (\hat V_{ni}^2 + \hat U_{ni}^2)\,. \tag{2.9}$$

3. Asymptotic distribution of $\hat M_n$. We shall consider the limiting distribution, not only under the hypothesis (1.2), but also under a sequence of alternatives defined by

$$H_n : \alpha_i = \alpha + (\xi_i/C_n)\,; \qquad \beta_i = \beta + (\theta_i/C_n)\,, \tag{3.1}$$

where $|\xi_i| < K_2$ and $|\theta_i| < K_3$, $i = 1, \cdots, k$.

Now set $Y_{ij}^0 = \{Y_{ij} - \beta(x_{ij} - x_{i.})\}$, and let R_{ij}^0 be the rank of Y_{ij}^0. For the proof of the limiting distribution we shall need the following $2k$ statistics:

$$T_{\alpha ni}^0 = \textstyle\sum_s \sum_j \{d_{sj}^{(i)} - \bar d^{(i)}\}a_n(R_{sj}^0)\,, \tag{3.2}$$

$$T_{\beta ni}^0 = \textstyle\sum_s \sum_j c_{sj}^{(i)} a_n(R_{sj}^0)\,, \qquad i = 1, \cdots, k\,. \tag{3.3}$$

Observe that, although neither (3.2) nor (3.3) can be calculated, because they depend on the unobservable random variables Y_{ij}^0 their distributions are fairly well known both under H_0 and H_n. The main tool in the proof of the limiting distribution of $\hat M_n$ is the following lemma

LEMMA 3.1. *Let the score generating function* $\phi(u)$ $0 < u < 1$, *satisfy the following*

(i) $|\phi''(u)| < K_4$,

(ii) $\sup_y |\phi_y'\{F(y)\}| < K_5$.

Also let the regression constants be such that

(iii) $\{\max_{ij} |x_{ij} - x_{i.}|/C_n\} \to 0$,

and assume that the estimate $\hat{\beta}$ *defined in* (2.2) *is such that as* $n \to \infty$,

(iv) $|C_n(\hat{\beta} - \beta)|$ *is bounded in both* P_0 *and* P_n *probabilities. Then under* (1.5), *for each* $1 = 1, \cdots, k$,

$$\{(\hat{T}_{\alpha ni} - T^0_{\alpha ni})/C_n\} \to 0 \,, \tag{3.4}$$

$$\{\hat{T}^0_{\beta ni} - T^0_{\beta ni})/C_n\} \to 0 \,, \tag{3.5}$$

in both P_0 *and* P_n *probabilities, where* ϕ' *denotes the derivative, and* ϕ_y' *the derivative with respect to* y, *while* P_0 *and* P_n *denote probabilities under* (1.2) *and* (3.1) *respectively.*

PROOF. The detailed proof of (3.5), under a slightly different P_n has been given in Adichie (1974). The proof of (3.4) proceeds on similar lines, upon defining Håjek's projection statistics for T^0_{ni}, and noting that in view of (1.7), (1.8) and (1.9), the constants $d^{(i)}_{sj}$ satisfy condition (iii) of the lemma. The proof that $\hat{\beta}$-estimate satisfies condition (iv) of the lemma is similar to that given in Sen (1969).

LEMMA 3.2. *Let* $\hat{V}_{ni}$ *and* $\hat{U}_{ni}$ *be as defined in* (2.7), *and let*

$$\mu_{\alpha ni} = \sum_s \sum_j (d^{(i)}_{sj} - \bar{d}^{(i)}) \int \phi\{\bar{F}(y)\}\, dF_{sj}(y) \,, \tag{3.6}$$

$$\mu_{\beta ni} = \sum_s \sum_j c^{(i)}_{sj} \int \phi\{\bar{F}(y)\}\, dF_{sj}(y) \,, \tag{3.7}$$

where

$$\bar{F}(y) = n^{-1} \sum_s \sum_j F_{sj}(y) \,. \tag{3.8}$$

Then, under the conditions of Lemma 3.1

(i) $\hat{\mathbf{V}}_n' = (\hat{V}_{n1}, \cdots, \hat{V}_{nk})$ *is asymptotically* $N(\mathbf{0}, \Sigma_\alpha)$ *under* P_0, *and asymptotically* $N(\boldsymbol{\nu}_{\alpha n}, \Sigma_\alpha)$ *under* P_n;

(ii) $\hat{\mathbf{U}}_n' = (\hat{U}_{n1}, \cdots, \hat{U}_{nk})$ *is asymptotically* $N(\mathbf{0}, \Sigma_\beta)$ *under* P_0, *and asymptotically* $N(\boldsymbol{\nu}_{\beta n}, \Sigma_\beta)$ *under* P_n;

(iii) $\hat{\mathbf{V}}_n$ *and* $\hat{\mathbf{U}}_n$ *are asymptotically independent both under* P_0 *and* P_n;

where $\boldsymbol{\nu}'_{\alpha n} = (\nu_{\alpha n1}, \cdots, \nu_{\alpha nk})$, $\boldsymbol{\nu}'_{\beta n} = (\nu_{\beta n1}, \cdots, \nu_{\beta nk})$, *with*

$$\nu_{\alpha ni} = n^{-\frac{1}{2}}(\mu_{\alpha ni}/A) \,, \qquad \nu_{\beta ni} = (\mu_{\beta ni}/AC_{ni}) \,; \tag{3.9}$$

$$\Sigma_\alpha = (\sigma_{\alpha is}) \,; \qquad \sigma_{\alpha is} = \{\delta_{is} - (\lambda_{ni}\lambda_{ns})^{\frac{1}{2}}\} \,, \tag{3.10}$$

$$\Sigma_\beta = (\sigma_{\beta is}) \,; \qquad \sigma_{\beta is} = \{\delta_{is} - (\gamma_{ni}\gamma_{ns})^{\frac{1}{2}}\} \,, \tag{3.11}$$

and δ_{is} *is the Kronecker delta.*

PROOF. If the lemma is true for P_n then it is a fortiori true for P_0, so the proof is given for P_n. Without loss of generality, we take $\alpha = \beta = 0$, and by Lemma 3.1, we restrict attention to $\mathbf{V}_n{}^0$ and $\mathbf{U}_n{}^0$ defined through $T^0_{\alpha ni}$ and $T^0_{\beta ni}$. Now under (3.1) with $\alpha = \beta = 0$,

$$F^0_{ij}(y) = F\{y - (\xi_i/C_n) - (\theta_i/C_n)(x_{ij} - x_{i.})\}\,,$$

so that

$$\begin{aligned}\max\nolimits_{sjtvy} |F_{sj}(y) - F_{tv}(y)| &\leqq K_7|(\xi_t - \xi_s) + \theta_t(x_{tv} - x_{t.}) - \theta_s(x_{sj} - x_{s.})|/C_n \\ &\leqq 2K_7(K_5 + K_6)\max\nolimits_{sj}|x_{sj} - x_{s.}|/C_n \\ &\leqq K\max\nolimits_{sj}|x_{sj} - x_{s.}|/C_n\,.\end{aligned}$$

Also because of (1.5), (1.7), (1.8) and (1.9),

$$\begin{aligned}\max\nolimits_{sj}|d^{(i)}_{sj} - \bar{d}^{(i)}|/\{\textstyle\sum_s\sum_j(d^{(i)}_{sj} - \bar{d}^{(i)})^2\}^{\frac{1}{2}} \\ &= \max\{\lambda_{ni.}(1 - \lambda_{ni})\}/\{\rho_{ni}(1 - \lambda_{ni})\}^{\frac{1}{2}}C_n \\ &\leqq \{(\rho_0\lambda_0)^{-\frac{1}{2}}/C_n\} \leqq K_0\max\nolimits_{sj}|x_{sj} - x_{s.}|/C_n\,.\end{aligned}$$

for some appropriate K_0. It follows then from Theorem 2.2 of Hájek (1968), tnat under (2.1), $n_i^{-\frac{1}{2}}(T^0_{\alpha ni} - \mu_{\alpha ni})/A(1 - \lambda_{ni})^{\frac{1}{2}}$ is asymptotically $N(0, 1)$. Furthermore, any linear combination of the k statistics $T^0_{\alpha ni}$ is again a linear rank statistic whose constants satisfy condition (iii) of Lemma 3.1. Hence $\mathbf{V}_n{}^0$ under P_n is asymptotically normal with asymptotic mean $\boldsymbol{\nu}_{\alpha n}$. For the asymptotic covariance matrix, if we write

$$W_{\alpha ni} = n^{-\frac{1}{2}}\textstyle\sum_s\sum_j(d^{(i)}_{sj} - \bar{d}^{(i)})\psi\{F_{sj}(y)\}/A\,,$$

then arguments similar to those used in the proof of Theorem 2.2 of Hájek (1968) show that under (3.1),

$$\operatorname{Cov}(V^0_{ni}, V^0_{ns}) \sim \operatorname{Cov}(W_{\alpha ni}, W_{\alpha ns}) = -(\lambda_{ni}\lambda_{ns})^{\frac{1}{2}}\,,$$

where $\sim$ denotes asymptotic equivalence in the ratio sense. This establishes (i) of the lemma. The proof for (ii) is similar. Finally (iii) follows from the fact that

$$\textstyle\sum_s\sum_j c^{(i)}_{sj}(d^{(i)}_{sj} - \bar{d}^{(i)}) = 0\,.$$

The limiting distribution of $\hat{M}$ is given in the following

THEOREM 3.1. *Consider model* (1.1), *and assume that the conditions of Lemma* 3.1 *are satisfied. Then under* P_0, $\hat{M}_n$ *has asymptotically a chi-square distribution with* $2k - 2$ *degrees of freedom, and under* P_n, *a non-central chi-square distribution with* $2k - 2$ *degrees of freedom and non-centrality parameter given by*

$$\Delta_M = \textstyle\sum_i\{\rho_i(\xi_i - \bar{\xi})^2 + \gamma_i(\theta_i - \bar{\theta})^2\}[\int\psi_v{}^2(F(y)\,dF(y))/A]^2\,, \tag{3.12}$$

where $\bar{\xi} = \sum_i\lambda_i\xi_i$; $\bar{\theta} = \sum_i\gamma_i\theta_i$, *and* ρ_i, λ_i *are the limits of* ρ_{ni} *and* λ_{ni} *respectively.*

PROOF. Each of the covariance matrices (3.10) and (3.11) is singular of rank $(k - 1)$. On applying orthogonal transformations to the $\hat{V}_{ni}$ and $\hat{U}_{ni}$ it follows from Lemmas 3.1, and 3.2, that under (3.1) each of $\sum_i \hat{V}^2_{ni}$ and $\sum_i \hat{U}^2_{ni}$ has asymptotically a chi-square distribution with $(k - 1)$ degrees of freedom and non-centrality parameters

$$\Delta_v = \lim \textstyle\sum_i \nu^2_{\alpha ni}\,, \qquad \Delta_u = \lim \textstyle\sum_i \nu^2_{\beta ni}\,.$$

From (iii) of Lemma 3.2, the non-centrality parameter of $\hat{M}_n$ is $\Delta_v + \Delta_u = D_M$ say, so that

$$D_M = \lim \textstyle\sum_i (\nu^2_{\alpha ni} + \nu^2_{\beta ni})\,. \tag{3.13}$$

Upon expanding the quantities $\nu_{\alpha ni}$ and $\nu_{\beta ni}$ and integrating by parts, it is easily seen that D_M is equal to Δ_M given in (3.12). The proof is thus complete.

4. Asymptotic efficiency. The classical test statistic Q_n for the hypothesis (1.2) is based on the difference between the least squares estimates of α_i and β_i when (1.1) is true, and the estimates of α and β when (1.2) is true. Q_n is the variance ratio criterion which in this case becomes

$$Q_n = \textstyle\sum_i \{n_i(Y_{i.} - Y_{..})^2 + C^2_{ni}(\bar{\beta}_i - \bar{\bar{\beta}})^2\}/(2k-2)s_\varepsilon^2\,, \tag{4.1}$$

where the least squares estimates $\bar{\beta}_i$ and $\bar{\bar{\beta}}$ are given by

$$\bar{\beta}_i = C^{-2}_{ni}\{\textstyle\sum_j (x_{ij} - x_{i.})(Y_{ij} - Y_{i.})\}, \qquad \bar{\bar{\beta}} = \textstyle\sum_i \gamma_{ni}\bar{\beta}_i\,, \tag{4.2}$$

with $Y_{i.} = n^{-1}\sum_j Y_{ij}$, and s_ε^2 is the mean square due to error. If F is assumed to be normal, as in the classical case, then under H_0, Q_n has the variance-ratio distribution with $(2k - 2, n - 2k)$ degrees of freedom, and the test based on Q_n is in this case most powerful.

When the assumption of normality of F is dropped, the exact distribution of Q_n is not known. Although it can be shown that for any $F(y)$ for which

$$\sigma^2(F) = \{\textstyle\int y^2\,dF(y) - (\int y\,dF(y))^2\} < \infty\,,$$

$(2k - 2)Q_n$ under H_0 has asymptotically a chi-square distribution with $(2k - 2)$ degrees of freedom and under (3.1) has asymptotically a non-central chi-square distribution with $2k - 2$ degrees of freedom and non-centrality parameter,

$$\Delta_Q = \lim E_n\{(2k-2)s_\varepsilon^2 Q_n\}/\sigma^2(F)\,, \tag{4.3}$$

where the expectation E_n is taken with respect to P_n-probability distribution. Straightforward computations yield

$$\Delta_Q = \textstyle\sum_i \{\rho_i(\xi_i - \bar{\xi}_i)^2 + \gamma_i(\theta_i - \bar{\theta})^2\}/\sigma^2(F)\,, \tag{4.4}$$

where $\bar{\xi}$ and $\bar{\theta}$ are defined in (3.12). By the conventional method of measuring efficiency, the asymptotic efficiency of the $\hat{M}_n$-test relative to the Q_ν-test is therefore

$$\Delta_M/\Delta_Q = \{\sigma(F) \textstyle\int \phi_y'(F(y))\,dF(y)/A\}^2\,, \tag{4.5}$$

which is the familiar efficiency expression of rank score tests relative to the classical t-test in the two-sample problem.

If $Z_n = -2 \log L$, where L is the likelihood ratio criterion, it follows that for $F(y)$ with a finite Fisher information $I(F)$, the efficiency of the $\hat{M}_n$ test relative to the asymptotically optimum parametric Z_n-test is

$$\Delta_M/\Delta_Z = [\int \phi_y'\{F(y)\}\, dF(y)]^2/A^2I(F)\,, \tag{4.6}$$

which is unity if $A^2 = I(F)$.

Acknowledgment. I wish to thank the Associate Editor and the referee for their very useful comments on this paper.

REFERENCES

[1] Adichie, J. N. (1967). Estimates of regression based on rank tests. *Ann. Math. Statist.* **38** 894–904.

[2] Adichie, J. N. (1974). Rank score comparison of several regression parameters. *Ann. Statist.* **2** 396–402.

[3] Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.

[4] Hájek, J. (1969). *Nonparametric Statistics.* Holden-Day, San Francisco.

[5] Sen, P. K. (1969). On a class of rank order tests for the parallelism of several regression lines. *Ann. Math. Statist.* **40** 1668–1683.

[6] Sen, P. K. (1972). On a class of aligned rank order tests for the identity of the intercepts of several regression lines. *Ann. Math. Statist.* **43** 2004–2012.

Faculty of Science
University of Nigeria
Nsukka, Nigeria