VERSITA

Mathematica Slovaca

DOI: 10.2478/s12175-009-0154-z
Math. Slovaca **59** (2009), No. 6, 661–666

TESTING THE GENERAL LINEAR HYPOTHESIS VIA K. PEARSON'S CHI-SQUARED STATISTIC

Lynn Roy LaMotte

Dedicated to Professor Andrej Pázman on the occasion of his 70th birthday

(*Communicated by Gejza Wimmer*)

ABSTRACT. In a linear model $\boldsymbol{Y} \sim (X\boldsymbol{\beta}, \sigma^2 I)$, powers of tests of $\mathrm{H}_0 : H'X\boldsymbol{\beta} = 0$ are developed following Pearson's (1900) formulation. The class considered comprises all tests based on linear statistics $A'\boldsymbol{Y}$ that have expected value 0 under H_0. The standard F-statistic, which is in this class, has good power properties, but others may be preferred in some settings.

1. Introduction

A linear model for the mean vector $\boldsymbol{\mu}$ of the random n-vector $\boldsymbol{Y}$ is a linear subspace $\mathscr{S}$ spanned by the columns of the $n \times p$ matrix X. The object of inference in such models can be considered to be a set of linear functions of $\boldsymbol{\mu}$, $H'\boldsymbol{\mu}$, where H is a specified $n \times c$ matrix, columns of which are in $\mathscr{S}$. Assume for this discussion that $\boldsymbol{Y} \sim \mathbf{N}(\boldsymbol{\mu}, \sigma^2 \mathrm{I})$, $\boldsymbol{\mu} \in \mathscr{S}$, $\sigma^2 > 0$. For now, consider σ^2 to be known.

For testing $\mathrm{H}_0 : \ H'\boldsymbol{\mu} = 0$, the usual test statistic is an F-statistic with numerator sum of squares that can be formulated as $\boldsymbol{y}'\mathbf{P}_H\boldsymbol{y}$ and numerator degrees of freedom $\nu = \mathrm{tr}(\mathbf{P}_H)$. Here, $\mathbf{P}_H$ denotes the orthogonal projection matrix onto $\mathrm{sp}(H)$ and $\boldsymbol{y}$ is the realized value of $\boldsymbol{Y}$. Note that, since $\mathrm{sp}(H) \subset \mathrm{sp}(X)$, $\mathbf{P}_X - \mathbf{P}_H$ and $\mathbf{P}_H$ are orthogonal, symmetric, idempotent matrices.

2000 Mathematics Subject Classification: Primary 62F03.
Keywords: power, noncentrality parameter, alternative F statistics.

The sum of squares $\boldsymbol{y}'\mathbf{P}_H\boldsymbol{y}$ can be arrived at in several ways: as a likelihood-ratio statistic,

$$\frac{\sup_{\boldsymbol{\mu}\in\mathscr{S}_0} L(\boldsymbol{\mu})}{\sup_{\boldsymbol{\mu}\in\mathscr{S}} L(\boldsymbol{\mu})} = \exp\left(-\frac{\boldsymbol{y}'\mathbf{P}_H\boldsymbol{y}}{2\sigma^2}\right);$$

as a Pearson's chi-squared statistic based on $H'\hat{\boldsymbol{\mu}} = H'\mathbf{P}_X\boldsymbol{y}$,

$$(H'\hat{\boldsymbol{\mu}})'\left[\mathrm{Var}(H'\hat{\boldsymbol{\mu}})\right]^-(H'\hat{\boldsymbol{\mu}}) = \boldsymbol{y}'H(H'H)^-H'\boldsymbol{y}/\sigma^2 = \boldsymbol{y}'\mathbf{P}_H\boldsymbol{y}/\sigma^2;$$

or as the result of an analysis of variance, partitioning the squared norm of $\hat{\boldsymbol{\mu}}$ as

$$\hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\mu}} = \boldsymbol{y}'(\mathbf{P}_X)\boldsymbol{y} = \boldsymbol{y}'(\mathbf{P}_X - \mathbf{P}_H)\boldsymbol{y} + \boldsymbol{y}'\mathbf{P}_H\boldsymbol{y}.$$

The last seems to have been the approach that originally led to this standard statistic.

Pearson (1900) derived the probability distribution of the exponent in the multivariate normal probability density function. Let $\boldsymbol{U} \sim \mathbf{N}_n(\boldsymbol{\mu}, \Sigma)$. Pearson derived the probability density function of

$$\chi^2 = (\boldsymbol{U} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{U} - \boldsymbol{\mu}),$$

proposing that its right-tail probability might be used as a p-value in examining how well a model for $\boldsymbol{\mu}$ fits the observed data $\boldsymbol{u}$. He went on to apply this form to test the goodness of fit of observed frequencies to a given theoretical probability distribution.

Pearson didn't discuss how one might identify the statistic to use as $\boldsymbol{U}$ in a particular setting. Considering the model $\boldsymbol{Y} \sim \mathbf{N}(X\boldsymbol{\beta}, \sigma^2\mathrm{I})$, if we want to test the linear hypothesis $\mathrm{H}_0: \ H'X\boldsymbol{\beta} = 0$, we might base a chi-squared statistic on any one of several statistics. All we require is that the statistic be normally distributed and have expected value 0 under H_0.

In this paper, we consider choosing a linear function $A'\boldsymbol{Y}$ as $\boldsymbol{U}$ in Pearson's chi-squared statistic, with the object of finding one that has expected value 0 under the null hypothesis and that maximizes the noncentrality parameter (hence the power) if $\boldsymbol{\mu}$ is not in $\mathscr{S}_0 = \{\boldsymbol{\mu} = X\boldsymbol{\beta}: \ H'X\boldsymbol{\beta} = 0\}$.

2. Development

That $\mathrm{E}(A'\boldsymbol{Y}) = 0$ for all $\boldsymbol{\beta} \in \mathscr{S}_0$ requires that $A'X\boldsymbol{\beta} = 0$ when $H'X\boldsymbol{\beta} = 0$. This requires that columns of A be orthogonal to all vectors $X\boldsymbol{\beta} \in \mathrm{sp}(X)$ that are in $\mathrm{sp}(H)^\perp$. That is, columns of A must be in the orthogonal complement of $\mathrm{sp}(X) \cap \mathrm{sp}(H)^\perp$. It can be shown that

$$(\mathrm{sp}(X) \cap \mathrm{sp}(H)^\perp)^\perp = \mathrm{sp}(H) + \mathrm{sp}(X)^\perp,$$

that is, columns of A must be the sum of linear combinations of columns of H and vectors orthogonal to all columns of X. Since $\mathrm{sp}(H) \subset \mathrm{sp}(X)$, every vector in $\mathrm{sp}(H)$ is orthogonal to every vector in $\mathrm{sp}(X)^{\perp}$.

Based on Pearson's chi-squared statistic, the statistic formed from $A'\boldsymbol{y}$ is

$$\chi^2_A = (A'\boldsymbol{y})'(\sigma^2 A'A)^{-}(A'\boldsymbol{y}) = \boldsymbol{y}'\mathbf{P}_A\boldsymbol{y}/\sigma^2.$$

When H_0 is true, $\mathrm{E}(A'\boldsymbol{Y}) = 0$, and this statistic has a central chi-squared distribution with $\mathrm{tr}(\mathbf{P}_A)$ degrees of freedom. In general, the noncentrality parameter of this statistic is

$$\delta^2 = \boldsymbol{\beta}'X'\mathbf{P}_A X\boldsymbol{\beta}/\sigma^2.$$

The condition that $\mathrm{E}(A'\boldsymbol{Y}) = 0$ when H_0 is true requires only that columns of A be in the linear subspace $\mathrm{sp}(H) + \mathrm{sp}(X)^{\perp}$. It doesn't limit A otherwise, and it doesn't say anything about how many columns A might have. There are lots of possible choices for linear statistics that have expected value 0 when H_0 is true. All that distinguishes them is their noncentrality parameter and their degrees of freedom.

Consider then the power of the test that rejects H_0 if $\chi^2_A \geq Q_{\alpha,df}$, where $Q_{\alpha,df}$ is the upper α quantile of the central chi-squared distribution with df degrees of freedom. Ghosh (1973) showed that the power of the test is monotone increasing in the noncentrality parameter δ^2, and that it is decreasing in df. Proposition 1 establishes that the noncentrality parameter of χ^2_A cannot exceed the noncentrality parameter of $\boldsymbol{y}'\mathbf{P}_H\boldsymbol{y}/\sigma^2$.

Proposition 1. *For any matrix A such that* $\mathrm{sp}(A) \subset \mathrm{sp}(H) + \mathrm{sp}(X)^{\perp}$,

$$X'\mathbf{P}_H X - X'\mathbf{P}_A X$$

is nonnegative definite.

Proof. Denote the orthogonal projection matrix onto $\mathrm{sp}(X)^{\perp}$ by $\mathbf{P}_{X^{\perp}}$. Because $\mathrm{sp}(H) \subset \mathrm{sp}(X)$, $\mathrm{sp}(H)$ and $\mathrm{sp}(X)^{\perp}$ are orthogonal. Then the orthogonal projection matrix onto $\mathrm{sp}(H) + \mathrm{sp}(X)^{\perp}$ is $\mathbf{P}_H + \mathbf{P}_{X^{\perp}}$. Because $\mathrm{sp}(A) \subset \mathrm{sp}(H) + \mathrm{sp}(X)^{\perp}$,

$$\mathbf{P}_A = (\mathbf{P}_H + \mathbf{P}_{X^{\perp}})\mathbf{P}_A = \mathbf{P}_H\mathbf{P}_A + \mathbf{P}_{X^{\perp}}\mathbf{P}_A = \mathbf{P}_A\mathbf{P}_H + \mathbf{P}_A\mathbf{P}_{X^{\perp}},$$

the last because all three projection matrices are symmetric. Then

$$\mathbf{P}_A X = \mathbf{P}_A\mathbf{P}_H X$$

because $\mathbf{P}_{X^{\perp}}X = 0$. And

$$X'\mathbf{P}_A X = (\mathbf{P}_A X)'(\mathbf{P}_A X) = X'\mathbf{P}_H\mathbf{P}_A\mathbf{P}_H X,$$

and so

$$X'\mathbf{P}_H X - X'\mathbf{P}_A X = X'\mathbf{P}_H(\mathrm{I} - \mathbf{P}_A)\mathbf{P}_H X,$$

which is nnd. □

Whatever A we might consider, the power of the test based on $A'\boldsymbol{y}$ is greater if we eliminate any part of A in $\mathrm{sp}(X)^{\perp}$, because doing so reduces degrees of freedom without affecting the noncentrality parameter. What's left of A is all in $\mathrm{sp}(H)$. We can increase the noncentrality parameter by each additional dimension from $\mathrm{sp}(H)$ we add to $\mathrm{sp}(A)$. However, each additional degree of freedom reduces power, so there is a question of balancing these two. If it happens that we omit a dimension in $\mathrm{sp}(H)$ where $\boldsymbol{\beta}$ is such that $H'X\boldsymbol{\beta} = 0$ or is close to 0, then the gain in power due to losing the dimension may outweigh the loss due to the change in the noncentrality parameter in that direction of the $\boldsymbol{\beta}$ space. Omitting any part of the H space renders the test unable to detect departures from H_0 in some direction in the $\boldsymbol{\beta}$ space, though. Although there does not appear to be a globally optimum choice of A based on power alone, a conservative approach would be to expand A to include all of $\mathrm{sp}(H)$, that is, to choose $A = H$. This is the choice that yields the conventional test statistic χ^2_H.

3. Illustration

Consider a setting with $X = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are linearly independent, $\mathbf{1}'\boldsymbol{x}_1 = \mathbf{1}'\boldsymbol{x}_2 = 0$, and $\mathrm{H}_0 : \ \boldsymbol{\beta} = (\beta_1, \beta_2)' = 0$ is to be tested. That is, $H = X(X'X)^{-1}$, and $\mathrm{H}_0 : \ H'X\boldsymbol{\beta} = \boldsymbol{\beta} = 0$ is the hypothesis to be tested. We shall consider statistics $A'\boldsymbol{Y}$ with either $A = H$ or $A = \boldsymbol{h}$, where $\boldsymbol{h}$ is a vector in $\mathrm{sp}(H)$. The corresponding test statistics are $Q_2 = \boldsymbol{y}'\mathbf{P}_H\boldsymbol{y}$ and $Q_1 = \boldsymbol{y}'\mathbf{P}_{\boldsymbol{h}}\boldsymbol{y}$. Let $\boldsymbol{p} = X'\boldsymbol{h}$. The degrees of freedom and noncentrality parameters associated with these two statistics are 2 and

$$\delta^2_H = \boldsymbol{\beta}'(X'X)\boldsymbol{\beta}/\sigma^2$$

for Q_2 and 1 and

$$\delta^2_{\boldsymbol{h}} = \frac{(\boldsymbol{p}'\boldsymbol{\beta})^2}{\boldsymbol{p}'(X'X)^{-1}\boldsymbol{p}\sigma^2}$$

for Q_1.

Q_2 has power to detect departures from H_0 in any direction in the $\boldsymbol{\beta}$ space. Let λ_1 and λ_2, $\lambda_1 \geq \lambda_2$, denote the eigenvalues of $X'X$, and denote the corresponding eigenvectors by $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. As a function of $\boldsymbol{\beta}/\sigma$, the power function of Q_2 has elliptical contours with axes along $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$; lengths of the axes are proportional to $1/\lambda_1$ and $1/\lambda_2$, respectively. If $\lambda_1 = \lambda_2$, then the power function increases at the same rate in any direction from 0. If λ_2 is small, power increases slowly in the direction along $\boldsymbol{v}_2$; in the extreme, when $\lambda_2 = 0$, power doesn't increase at all along $\mathrm{sp}(\boldsymbol{v}_2)$ and the contours of the power function are straight lines parallel to $\mathrm{sp}(\boldsymbol{v}_2)$.

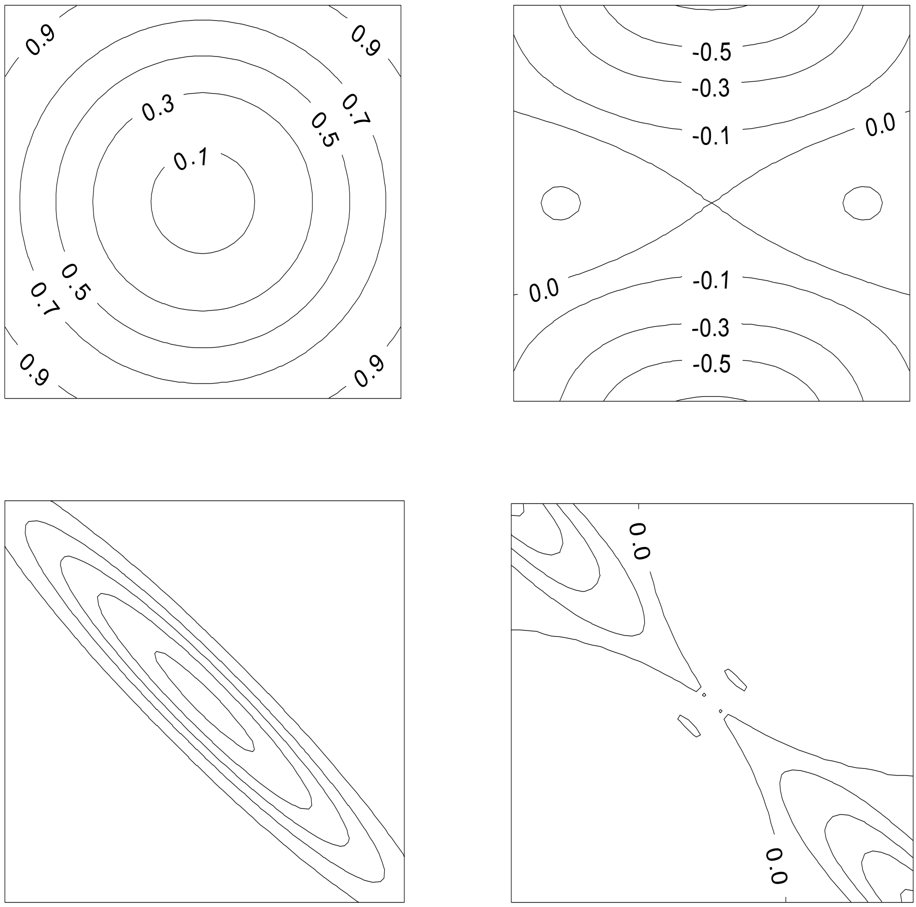

FIGURE 1. Comparisons of powers for two different 20×2 X matrices. Left: contours of the power function of Q_2. Right: contours of the difference (1 minus 2) between powers of Q_1 and Q_2

See Figure 1, where power functions for Q_2 and Q_1 are shown for two different X matrices. The direction $\boldsymbol{h}$ for Q_1 is chosen in the most favorable direction, namely $\boldsymbol{h} = \boldsymbol{v}_1$, the direction in which the power function for Q_2 is increasing most rapidly. In the top row, the two columns of X are practically orthogonal and the eigenvalues are nearly equal. In the bottom row, the eigenvalues are quite different, corresponding to a correlation between columns of X of about 0.9. Contours of the power functions for Q_2 are shown in the left panels of each row.

The right panels show contours of the difference, power for Q_1 minus power for Q_2. In the upper right panel, Q_1 has greater power than Q_2 throughout the region in the horizontal strip outlined by the 0.0 contour. In the lower right panel, the power function for Q_1 is greater than the power function for Q_2 in a broader region; Q_2 has greater power only in the upper-left and lower-right corners.

Considering the choice of a linear function $A'\boldsymbol{y}$ on which to base a chi-squared statistic for testing $\mathrm{H}_0 : \ H'X\boldsymbol{\beta} = 0$, choosing A such that $\mathrm{sp}(A) = \mathrm{sp}(H)$ ensures that the power function increases in every direction in the $\boldsymbol{\beta}$ space, so that, wherever $\boldsymbol{\beta}$ lies, the test has some power to detect departures from H_0. Depending on the configuration of the X matrix, though, the test based on $\boldsymbol{y}'\mathbf{P}_H\boldsymbol{y}$ may have very little power in some directions. In that case, it may not be unreasonable to choose a statistic $A'\boldsymbol{y}$ that has no power at all (that is, no greater than the level of significance α) in those directions, thus gaining power in directions better covered by X.

REFERENCES

[1] GHOSH, B. K. : *Some monotonicity theorems for chi-square, F, and t distributions with applications*, J. Roy. Statist. Soc. Ser. B **35** (1973), 480–492.

[2] PEARSON, K.: *On the criterion that a given set of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, Phil. Mag. (5) **50** (1900), 157–175.

Received 19. 5. 2008
Accepted 5. 8. 2008

Biostatistics Program
LSU Health Sciences Center School of Public Health
1615 Poydras Street
New Orleans, LA 70112
USA
E-mail: llamot@lsuhsc.edu