STATISTICAL RESEARCH REPORT No 13
Institute of Mathematics 1971
University of Oslo

ON THE HALF-NORMAL PLOT AND SOME RELATED MATTERS

by

Emil Spjøtvoll

ABSTRACT

It is shown that in the case when we can estimate or know the variance, the use of Daniels' halfnormal plot corresponds to a certain stepwise procedure where we at each step use an optimal or almost optimal test. Results are also obtained that indicates that stepwise regression analysis can be described in a similar way.

1. Introduction. Daniels' [4] half-normal plot has found wide use in statistical practice, but the theoretical properties of the procedure is not very well known, see, e.g. Birnbaum [3]. Daniels used the plot to simultaneously estimate the variance and to determine the non-null effects in 2^{p-q} experiments. In this paper we shall assume that the variance is known or that we have an independent estimate of the variance. We shall then show that use of the half-normal plot corresponds to a certain stepwise procedure where we at each step use an optimal test. Since we do assume that we know or can estimate the variance, we are only halfway to understand the theoretical properties of the half-normal plot. But even this case corresponds to situations in practice. Though we rarely know the variance, we can always estimate it using higher-order interactions.

We shall also give a result which indicates that the technique used in stepwise regression analysis can be described as a stepwise method where we at each step use an optimal test.

2. The case with known variance. Let $X_1,\ldots,X_p$ be p independent normal random variables with $EX_i = \mu_i$, $i = 1,\ldots,p$ and common known variance σ^2 . The problem is to decide which (if any) of the μ_i-s are different from zero. This we shall do in a stepwise manner by first determining if the μ_i with the largest absolute value is different from zero. If this is the conclusion, the next step is to determine if the μ_i with the largest absolute value of the remaining $p-1$ μ_i-s is different from zero, and we continue like this until the conclusion is that the remaining μ_i-s are not different from zero. We now proceed to the formal description of the method.

<u>Step 1.</u> Let $|\mu|_{(1)} < \dots < |\mu|_{(p)}$ be the ordered absolute values of $\mu_1,\dots,\mu_p$. Consider the following hypothesis testing problem with p alternatives

$$(1)\quad \begin{array}{l} H_1: \mu_1 = \dots = \mu_p = 0 \text{ against } K_{11}: |\mu_1| = |\mu|_{(p)} > 0 \text{ or } \dots \text{ or} \\ K_{1p}: |\mu_p| = |\mu|_{(p)} > 0 . \end{array}$$

A test of this problem has p elements $\varphi_1,\dots,\varphi_p$ where $\varphi_i(x)$ is the probability that the alternative K_{1i} is accepted when x is observed. For a general treatment of hypothesis testing problems with p alternatives see, e.g., Spjøtvoll [10].

We shall find the test which maximizes the minimum power or the average power (see [10] pp. 4-5 for the definition of average power) over alternatives with $\max_j |\mu_j| \geq \Delta$, where $\Delta > 0$. First consider the problem

$$H_1: \mu_1 = \dots = \mu_p = 0 \text{ against } K_{11}^*: |\mu_1| = \Delta, \mu_2 = \dots = \mu_p = 0$$

$$\text{or } \dots \text{ or } K_{1p}^*: \mu_1 = \dots = \mu_{p-1} = 0, |\mu_p| = \Delta .$$

For each alternative K_{1i}^* we use a least favorable distribution with probability $\frac{1}{2}$ of the point with $\mu_i = -\Delta$ and $\frac{1}{2}$ of the point with $\mu_i = \Delta$. Then the ratio of the density of the observations at K_{1i}^* and the density under H_1 is

$$(2)\quad \tfrac{1}{2}\exp(-\tfrac{1}{2}\frac{\Delta}{\sigma^2})\{\exp(-x_i\frac{\Delta}{\sigma^2}) + \exp(x_i\frac{\Delta}{\sigma^2})\}$$

By [10, Theorem 1 and remark on p. 10] the test which maximizes the average power over the alternatives K_{1i}^* is given by accepting the alternative K_{1i}^* corresponding to the i for which (2) is maximized, if this maximum is big enough. It easily seen by differentiation that (2) increases with x_i when $x_i > 0$ and decreases with x_i when $x_i < 0$. Furthermore (2) is symmetric in x_i around zero.

Hence (2) is maximized for the i which satisfy $|x_i| = \max_j |x_j|$, and (2) greater than a constant is the same as $\max_j |x_j|$ greater than a constant.

The level α test which maximizes the average power over the alternatives K_{1i}^* therefore rejects H_1 if

(3) $\max_j |X_j| > \sigma m(\alpha/2,p)$,

where $m(\alpha/2,p)$ is the upper $\alpha/2$-point of the distribution of the maximum of p independent standard normal random variables. Table 24 in [9] contains these α-points for p from 1 to 30 and for the usual α-values. If (3) holds the alternative K_{1i}^* with

$$|X_i| = \max_j |X_j| .$$

is accepted.

This test also maximizes the average power over all alternatives with $|\mu|_{(p)} \geq \Delta$, since it is easily seen that it has its smallest power over these alternatives for one of the alternatives $K_{1i}^*,\ldots,K_{1p}^*$. Any other test has smaller average power over these alternatives.

Because of the symmetri of the problem the minimum powers over the alternatives K_{1i}^* are equal, it therefore follows that the test also maximizes the minimum power over alternatives with $|\mu|_{(p)} \geq \Delta$ (see also [10], the Corollory p. 10).

If we do not reject the hypothesis H_1 , we stop the step-wise procedure and the conclusion is that we cannot reject the hypothesis that all the μ_i are equal to zero. If we reject the conclusion is that the μ_i corresponding to the index of $|X_i| = \max_j |X_j|$ is different for zero and also the largest in absolute value of the μ_i-s . We then go to

Step 2. We now assume that the conclusion of the first step is correct. Then we want to find out if any of the remaining p-1 μ_i-s are different from zero. Denote the remaining μ_i-s , $\mu_1',\ldots,\mu_{p-1}'$ and the corresponding random variables $X_1',\ldots,X_{p-1}'$. Then $X_1',\ldots,X_{p-1}'$ are independent normal with $EX' = \mu_i'$ and $\mathrm{Var}\, X_j' = \sigma^2$. (If the conclusion of step 1 is wrong then $X_1',\ldots,X_{p-1}'$ are not independent but the p-1 smallest observation from a random sample of p). Now consider the problem

$$H_2:\ \mu_1' = \ldots = \mu_{p-1}' = 0 \quad \text{against} \quad K_{21}: |\mu_1'| = |\mu'|_{(p-1)} \text{ or } \ldots \text{ or} \tag{4}$$
$$K_{2,p-1}:\ |\mu_{p-1}'| = |\mu'|_{(p-1)} ,$$

where $|\mu'|_{(1)} < \ldots < |\mu'|_{(p-1)}$ are the ordered absolute values of $\mu_1',\ldots,\mu_{p-1}'$. This is analogous to the problem (1) with p-1 instead of p . The test which maximizes the minimum and average power over alternatives with $|\mu'|_{(p-1)} \geq \Delta$ rejects when

$$\max_j |X_j'| > \sigma m(\alpha/2,p-1) , \tag{5}$$

and accepts the alternative K_{2i} corresponding to the largest $|X_i'|$.

If H_2 is not rejected then the conclusion is that $\mu(p) \neq 0$ and $\mu(1) = \quad = \mu(p-1) = 0$, and the stepwise procedure stops. If H_2 is rejected, then the conclusion is that $\mu(p) \neq 0$, $\mu(p-1) \neq 0$, and we have to proceed to step 3 to test the hypothesis H_3 that the remaining p-2 μ_i-s are equal to zero. We proceed like this until all hypotheses H_i are rejected or we stop because som hypothesis H_i is not rejected. If H_k is the first hypothesis which is not rejected, then the conclusion of the whole procedure is that $|\mu|_{(i)} \neq 0$, i=p, p-1 , p-k+1 and we cannot reject the hypothesis $|\mu|_{(1)} = \quad = |\mu|_{(p-k)} = 0$.

3. Unknown variance. Let the situation be as in Section 2, but let the variance σ^2 be unknown. Suppose that we have an estimate S^2 of σ^2 such that $\nu S^2/\sigma^2$ has a chi-square distribution with ν degrees of freedom and is independent of $X_1,\ldots,X_p$. It seems natural when we use a step-wise procedure analogous to that of Section to replace the test statistics (3) and (5) by

(6) $\max_j |X_j| > S\, m(\alpha/2,p,\nu)$

and

(7) $\max_j |X_j| > S\, m(\alpha/2,p-1,\nu)$,

respectively, where $m(\alpha/2,p,\nu)$ is the upper $\alpha/2$-point of the distribution of

$$\frac{\max_j X_j}{S}$$

when all $EX_j = 0$. To our knowledge this distribution is not tabulated. If the number ν of degrees of freedom is large $m(\alpha/2,p,\nu)$ should be close to $m(\alpha/2,p)$, otherwise one should expect that $m(\alpha/2,p,\nu) > m(\alpha/2,p)$.

It seems difficult to show an optimality property of the tests (6) and (7) in the same way as of the tests (3) and (5). They may not be optimal in the same sense as (3) and (5) as is indicated by a counterexample in Lehmann [8, pp. 1003-1006] for a situation which is somewhat similar to the one here.

4. Properties of the proposed stepwise procedure. We shall discuss the properties in relation to the situation where σ^2 is known, but the same results and conclusions also hold when σ^2 is unknown. The following holds

P{The conclusion is $\mu_1 = \ldots = \mu_p = 0$ when $\mu_1 = \ldots = \mu_p = 0$}

$= P\{ |X|_{(p)} \leq m(\alpha/2,p) \mid \mu_1 = \quad = \mu_p = 0\} = 1-\alpha$.

(8) P{One or more μ_i declared different from zero when $\mu_1 = \dots = \mu_p = 0\} = P\{|X|_{(p)} > m(\alpha/2,p) | \mu_1 = \dots = \mu_p = 0\} = \alpha$.

(9) P{Two or more μ_i declared different from zero when $|\mu|_{(1)} = \dots = |\mu|_{(p-1)} = 0$, $|\mu|_{(p)} > 0\}$ = P{The largest of $p-1$ independent $N(0,\sigma^2)$ variables is greater than $\sigma m(\alpha/2,p-1)\} \leq \alpha$.

Here we have used that $m(\alpha,p) < m(\alpha,q)$ when $p < q$.

(10) P{k+1 or more μ_i declared different from zero when $|\mu|_{(1)} = \quad = |\mu|_{(p-k)} = 0$ and $|\mu|_{(j)} > 0 \quad j = p-k+1,\dots,p.\}$ = P{The largest of $p-k$ independent $N(0,\sigma^2)$ variables is greater than $\sigma m(\alpha/2,p-k)\} \leq \alpha$

In fact (10) contains (8) and (9) as special cases. The bound α in (10) cannot be improved upon, when $|\mu|_{(j)} \to \infty$, $j = p-k+1,\dots,p$ the probability in (10) gets as close to α as we wish.

Stepwise procedure where we at each step use an optimal (in some sense) procedure is well known is statistical literature, see, e.g., Anderson [2] and Lehmann [7]. Anderson is also able to show that his procedure has some optimality properties when one considers the procedure as a whole and not only each individual step. Both the Anderson and the Lehmann procedure has the defect that one proceeds to the next step if a hypothesis is not rejected. At the next step one then assumes that the hypothesis is true. Clearly, there are cases when the hypothesis is not true, but where we have a very high probability of not rejecting and hence falsely going to the next step. In the present procedure we continue only if a hypothesis is rejected, and the maximum probability of going wrongly to the next step is then a small number α .

5. A stepwise procedure in regression analysis. Let the $n\times 1$ random Y be $N(X'\beta,\sigma^2 I)$, where X' is a known $n\times p$ matrix of rank p. The $p\times 1$ vector β is unknown. The variance σ^2 is known. The least square estimate

$$\hat\beta = (XX')^{-1}XY$$

of β is $N(\beta, A\,\sigma^2)$ where $A = (XX')^{-1}$. The problem is to determine which of the elements $\beta_1,\ldots,\beta_p$ of the vector β are different from zero. To do this we shall proceed in a way similar to what was done in Section 2.

Step 1. Let $\hat\beta_1,\ldots,\hat\beta_p$ be the elements of $\hat\beta$. The conditional distribution of $\hat\beta_i$ given $\hat\beta_{(i)} = (\hat\beta_1,\ldots,\hat\beta_{i-1},\hat\beta_{i+1},\ldots,\hat\beta_p)'$ is a normal distribution with expectation $\beta_i + d_i'(\hat\beta_{(i)}-\beta_{(i)})$ and variance $c_i^2\sigma^2$, where d_i and c_i can be found from the matrix A (see, e.g., [1] p. 28). Now consider the problem

$$H_1:\ \beta_1= \ldots =\beta_p = 0 \text{ against } K_{1i}:\ |\beta_1|/c_1 = \max_j\ |\beta_j|/c_j$$

$$\text{or}\ldots\text{or } K_{1p}:\ |\beta_p|/c_p = \max_j\ |\beta_j|/c_j\ .$$

This is similar to the problem (1) in Section 2. As in Section 2 we first consider the problem

$$H_1:\ \beta_1= \ldots =\beta_p = 0 \text{ against } K_{1i}^*:\ |\beta_1|/c_1 = \Delta,\ \beta_2= \ldots =\beta_p = 0$$

$$\text{or}\ldots\text{or } K_{1p}^*:\ |\beta_p|/c_p = \Delta,\ \beta_1= \ldots = \beta_{p-1} = 0\ .$$

Again we use for each alternative K_{1i}^* a least favorable distribution with probability ½ of the points with $\beta_i/c_i = \pm\Delta$. We can write the density of $\hat\beta_1,\ldots,\hat\beta_p$ as the product of the conditional density of $\hat\beta_i$ given the others, times the marginal density of the remaining $p-1$ $\hat\beta_j$-s. Then, since the latter is the same both under H_1 and K_{1i}^*, it follows that the ratio of the densities of

$\hat{\beta}_1,\dots,\hat{\beta}_p$ under K_{1i}^* and H_1 is equal to the ratio of the conditional densities of $\hat{\beta}_i$ under K_{1i}^* and H_1. The conditional density is $N(\beta_i + d_i' \hat{\beta}_{(i)}, c_1{}^2 \sigma^2)$ since $\beta_{(i)} = 0$ both under H_1 and K_{1i}^*. Using the least favorable distribution we find that the ratio is

$$\tfrac{1}{2} \exp(-\tfrac{1}{2} \frac{\Delta^2}{\sigma^2}) \{\exp(- \frac{\hat{\beta}_i - d_i'\hat{\beta}'_{(i)}}{c_i} \frac{\Delta}{\sigma^2}) + \exp(\frac{\hat{\beta}_i - d_i'\hat{\beta}_{(i)}}{c_i} \frac{\Delta}{\sigma^2})\} .$$

Arguing as after (2) in Section 2 we find that the test which maximizes the average power over the alternatives K_{1i}^* rejects H_1 when

$$\max_j \left| \frac{\hat{\beta}_j - d_j' \hat{\beta}_{(j)}}{c_j} \right| > \text{constant},$$

and accepts the alternative K_{1i}^* corresponding to the index i which satisfy

$$\left|\frac{\hat{\beta}_i - d_i' \hat{\beta}_{(i)}}{c_i}\right| = \max_j \left|\frac{\hat{\beta}_j - d_j' \hat{\beta}_{(j)}}{c_j}\right| .$$

This test does not generally maximize the average power over alternatives with $\max_j |\beta_j|/c_j = \Delta$. It can be seen that there exist alternatives to H_1 where $E(\hat{\beta}_i - d_i' \hat{\beta}_{(i)}) = 0$ and for such alternatives the power is equal to the level of the test.

In the situation when A is diagonal, however, the test maximizes the average power over all alternatives with $\max_j |\beta_j|/c_j = \Delta$. Then the test is to reject H_1 when

$$\max_j \frac{|\hat{\beta}_j|}{c_j} > \sigma m(\alpha/2,p)$$

and accept the alternative with

$$\frac{|\hat{\beta}_i|}{c_i} = \max_j \frac{|\hat{\beta}_j|}{c_j} .$$

This is the procedure usually employed by stepwise methods in regression analysis (see, e.g., [5]), whether A is diagonal or not.

If H_1 is rejected one proceeds to step 2 with $p-1$ of $\beta_1,\dots,\beta_{p-1}$, and goes through a similar procedure. One continues until at some step the null hypothesis is not rejected.

If σ is unknown, it should be replaced by an estimate as in Section 2.

Draper, Guttman and Kanemasu [6] have studied what happens at an individual step in stepwise regression analysis and their results indicate that $m(\alpha/2,p,\nu)$ is not far from the upper $\alpha/(2p)$-point of the t-distribution with ν degrees of freedom.

REFERENCES

[1] ANDERSON, T.W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

[2] ANDERSON, T.W. (1962). The choice of degree of a polynomial as a multiple decision procedure. Ann. Math. Statist. **33** 255-265.

[3] BIRNBAUM, A. (1959). On the analysis of factorial experiments without replication. Technometrics **1** 343-357.

[4] DANIELS, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. Technometrics **1** 311-341

[5] DRAPER, N.R. and SMITH, H. (1966). Applied Regression Analysis. Wiley, New York.

[6] DRAPER, N.R., GUTTMAN, I. and KANEMASU, H. (1971). The distribution of certain regression statistics. Biometrika **58** 295-298.

[7] LEHMANN, E.L. (1957). A theory of multiple decision problems, II. Ann. Math. Statist. **28** 547-572.

[8] LEHMANN, E.L. (1961). Some model I problems of selection. Ann. Math. Statist. **32** 990-1012.

[9] PEARSON, E.S. and HARTLEY, H.O. (1958). Biometrika Tables for Statisticians, 1. Cambridge University Press.

[10] SPJØTVOLL, E. (1971). Tests maximizing minimum power. Statistical Research Report, No. 2, 1971, University of Oslo.