CHAPTER 3

Selecting the normal populations in terms of means based on quantal response data .

In this chapter, as in the previous one the problem of selecting the best of K populations in terms of means is considered. However, unlike earlier, now we suppose that the data from each of the populations is as described in table T.1 of chapter 1, section 1 .

As explained in chapter 1, such data arises in bio-assays and the mean of normal distribution is equal to log ED50 of the substance under consideration. Since smaller the value of ED50, the better is the substance we compare the K substances in terms of smallness of log ED50's, since logarithm is strictly increasing function.

In view of the above application our problem is that of selecting the normal population with minimum value of mean, when data from each population is of quantal response type.

In section 1 below the problem of estimating log ED50 from quantal response data by the method of maximum likelihood is described. The asymptotic normality of the estimator is established for later use. In section 2 the estimator obtained above is used to select the normal population which has smaller mean of the K populations.

The probability of correct selection or the selections rule is given in both the cases

i) when variances of the normal distributions are equal and known.

ii) When variances are equal and unknown.

The case of equal variances only is considered in the dissertation because in the context of Bioassays this is equivalent to considering a parallel line assay or assuming that the true probit regression lines of K populations are parallel. The other case known as slope ratio assays is not considered.

In section 3, we give two numerical examples to illustrate the selection procedures described in section 2.

<u>3.1 Maximum likelihood estimators of log ED50</u>

Suppose that a given substance is administrated to groups of subjects at different doses. Let n_i be the number of subjects in the i^{th} group, which receive dose z_i (with $x = \log z_i$) and suppose that r_i individuals (out of n_i) show the desired response. The above information can be tabulated as below :

log dose	no. of subjects tested	no. of respondents	population of pre-respondent
x	n	r	p
x_1	n_1	r_1	p_1
x_2	n_2	r_2	p_2
⋮	⋮	⋮	⋮
x_ℓ	n_ℓ	r_ℓ	p_ℓ

In order to estimate log ED50, from the above data, we make the following assumptions.

a) The subjects respond independantly

b) The probability of response is the same for subjects which receive the same dose, and

c) The log tolerance distribution is of the form $G(\alpha + \beta x)$, so that the true probability of response at dose x_i is p_i given by

$$p_i = G(\alpha + \beta x_i) \qquad 3.1.1$$

where G is a completely known distribution.

In the above set up it is noted in chapter 1, that log ED50 is given by

$$\log ED50 = \frac{G^{-1}(0.5) - \alpha}{\beta} \qquad 3.1.2$$

So the problem of estimating log ED50 reduces to that of estimating α and β.

In the following we consider this problem. Here we have to consider two cases

i) α unknown and β known

ii) α unknown and β is also unknown.

Note that the case of α known and β unknown need not be considered. Because we want to compare $\mu_1, \mu_2, \ldots, \mu_k$ which, in turn, is equivalent to comparing $-\alpha_1/\beta$, $-\alpha_2/\beta, \ldots, -\alpha_k/\beta$, that is to find $\frac{1}{\beta}$ min ($-\alpha_1, -\alpha_2, \ldots, -\alpha_k$). So whether β is known or unknown, the ordering and hence selection do not change once the α_i's are known.

In the following discussion we consider the case when both parameters are unknown and only give the estimator in the other case as a remark.

Under the assumptions given above, it is clear that the likelihood function of (α,β) given the observations $r_1, r_2, \ldots, r_l$ can be written as

$$L = L(\alpha,\beta) = \prod_{i=1}^{l} \binom{n_i}{r_i} p_i^{r_i} (1-p_i)^{n_i - r_i} \qquad 3.1.3$$

Equivalently;

Log L = constant + $\Sigma r_i \log P_i + \Sigma(n_i - r_i)\log(1-P_i)$

where constant = $\prod_{i=1}^{l} \binom{n_i}{r_i}$ is independant of α and β.

We have to solve the likelihood equations

$$\frac{\partial \log L}{\partial \alpha} = 0 \qquad 3.1.4$$

and

$$\frac{\partial \log L}{\partial \beta} = 0 \qquad 3.1.5$$

We see that the likelihood equations donot directly involve α and β, whereas it involves P_i, the probability of response at dose x_i which is a function of α and β. Here

$$P_i = G(\alpha+\beta x_i) \qquad 3.1.6$$

The equations 3.1.4 and 3.1.5 are equivalent to

$$\sum \frac{\partial \log L}{\partial P_i} \frac{\partial P_i}{\partial \alpha} = 0 \qquad 3.1.7$$

and

$$\sum \frac{\partial \log L}{\partial P_i} \frac{\partial P_i}{\partial \beta} = 0 \qquad 3.1.8$$

Now

$$\frac{\partial \log L}{\partial P_i} = \frac{r_i - n_i P_i}{P_i(1-P_i)}$$

so that

$$\frac{\partial \log L}{\partial \alpha} = \Sigma \frac{r_i - n_i P_i}{P_i(1-P_i)} \frac{\partial P_i}{\partial \alpha}$$

$$= \Sigma \frac{r_i - n_i P_i}{P_i(1-P_i)} g_i \qquad 3.1.9$$

where

$$g_i = g\ (\alpha+\beta x_i) \qquad 3.1.10$$

and g is the density corresponding to distribution function G. Similarly

$$\frac{\partial \log L}{\partial \beta} = \Sigma \frac{(r_i - n_i P_i)}{P_i\ (1-P_i)} x_i g_i \qquad 3.1.11$$

In order to simplify equations 3.1.9 and 3.1.10, we write

$$r_i = n_i P_i \qquad 3.1.12$$

Where p_i is observed proportion of respondants and

$$w_i = \frac{g_i^2}{P_i(1- P_i)} \qquad 3.1.13$$

Note that p_i is unbiased estimator of P_i so that

$$E(p_i - P_i) = 0 \qquad 3.1.14$$

and

$$\text{Var}\,(p_i) = \frac{P_i(1-P_i)}{n_i} \qquad 3.1.15$$

for $i = 1,2,..,\ell$

Now equations 3.1.9 and 3.1.11 reduce to

$$\Sigma \frac{n_i w_i (p_i - P_i)}{g_i} = 0 \qquad 3.1.16$$

and

$$\Sigma \frac{n_i w_i (p_i - P_i) x_i}{g_i} = 0 \qquad 3.1.17$$

For further simplifications add the equation

$\Sigma\, n_i w_i y_i = \Sigma\, n_i w_i (\alpha + \beta x_i)$

to equation 3.1.16 and add

$\Sigma\, n_i w_i x_i y_i = \Sigma\, n_i w_i x_i (\alpha + \beta x_i)$ to equation 3.1.17.

We get

$$\Sigma\, n_i w_i \left(y_i + \frac{(p_i - P_i)}{g_i}\right) = \Sigma\, n_i w_i (\alpha + \beta x_i) \qquad 3.1.18$$

$$\Sigma\, n_i w_i x_i \left(y_i + \frac{p_i - P_i}{g_i}\right) = \Sigma\, n_i w_i x_i (\alpha + \beta x_i) \qquad 3.1.19$$

Put

$$\tilde{y}_i = y_i + \frac{(p_i - P_i)}{g_i} \qquad 3.1.20$$

So that the above equations can be written as

$$\Sigma\, n_i w_i \tilde{y}_i = \alpha\, \Sigma\, n_i w_i + \beta\, \Sigma\, n_i w_i x_i \qquad 3.1.21$$

$$\Sigma\, n_i w_i x_i \tilde{y}_i = \alpha\, \Sigma\, n_i w_i x_i + \beta\, \Sigma\, n_i w_i x_i^2 \qquad 3.1.22$$

The above two equations are in the form of usual normal equations except that the summands are multiplied by weights $n_i w_i$.

It is important to note that these weights involve the unknown parameters α and β. Also, the left hand side of the equations involve $\tilde{y}_i$, which is again a function of α and β. Hence unlike the usual case it is not possible to solve these equations in a straight forward way. We solve 3.1.21 and 3.1.22 using 3.1.10, 3.1.13 and 3.1.20 by iterative method. That is we start with some initial values for (α,β) say (α_o,β_o) and at the m^{th} stage solve for (α_m,β_m) from 3.1.21 and 3.1.22 using previous values $(\alpha_{m-1}, \beta_{m-1})$ to evaluate w_i and $\tilde{y}_i$. We continue the iterations until the successive values of (α,β) are to the desired accuracy.

We denote the values of $\tilde{y}_i$ and w_i evaluated at previous values of α and β by $\hat{y}_i$ and $\hat{w}_i$; so that for the current iteration equations 3.1.21 and 3.1.22 reduce to

$$\Sigma\, n_i \hat{w}_i \hat{y}_i = \alpha\, \Sigma\, n_i \hat{w}_i + \beta\, \Sigma\, n_i \hat{w}_i x_i \qquad 3.1.23$$

$$\Sigma\, n_i \hat{w}_i x_i \hat{y}_i = \alpha\, \Sigma\, n_i \hat{w}_i x_i + \beta\, \Sigma\, n_i \hat{w}_i x_i^2 \qquad 3.1.24$$

Now these equations can be easily solved for α and β.

Suppose that α and β denote the solutions to the equations at the final iteration. (The procedure can be shown to terminate with appropriate initial values). Then the solution can be written as

$$\hat{\alpha} = \hat{y} - \hat{\beta} x \qquad 3.1.25$$

and

$$\hat{\beta} = S_{xy} \,/\, S_{xx} \qquad 3.1.26$$

where

$$\bar{X} = \Sigma\, n_i \hat{w}_i x_i / \Sigma\, n_i \hat{w}_i \qquad 3.1.27$$

$$y = \Sigma\, n_i \hat{w}_i \hat{y}_i \,/\, \Sigma\, n_i \hat{w}_i \qquad 3.1.28$$

$$S_{xy} = \Sigma\, n_i \hat{w}_i (x_i - \bar{x})(y_i - y) / \Sigma\, n_i \hat{w}_i \qquad 3.1.29$$

$$S_{xx} = \Sigma\, n_i \hat{w}_i (x_i - \bar{x})^2 \,/\, \Sigma\, n_i \hat{w}_i \qquad 3.1.30$$

Remark : If β is known, the estimator of α is given by

$$\hat{\alpha} = \tilde{y} - \beta\bar{x} \qquad 3.1.31$$

Now we study the asymptotic properties of the estimator $(\hat{\alpha}, \hat{\beta})$ of (α, β).

In order to do so, first we note that (α, β) are the roots of the likelihood equation 3.1.4 and 3.1.5 obtained by iterative method.

Secondly, the likelihood equations have unique root in (α,β), because the matrix of second order derivatives is non-negative definite.

Therefore, Crammer's theorem as given in Theorem 41 on page 429 of Lehman (Theory of point estimation) is applicable.

We show below that information matrix is non-negative definite.

$$\text{LogL} = \text{constant} + \Sigma\, n_i \log P_i + \Sigma(n_i - r_i)\log(1-P_i)$$

$$\frac{\partial \log L}{\partial \alpha} = \Sigma \frac{n_i w_i (p_i - P_i)}{g_i} \frac{\partial P_i}{\partial \alpha}$$

$$\frac{\partial^2 \log L}{\partial \alpha^2} = \left[-\frac{r_i}{P_i^2}\frac{\partial P_i}{\partial \alpha} - \frac{(n_i - r_i)}{(1-P_i)^2}\frac{\partial P_i}{\partial x}\right]\frac{\partial P_i}{\partial \alpha} + \Sigma\left[\frac{r_i}{P_i} - \frac{(n_i - r_i)}{1-P_i}\right]\frac{\partial^2 P_i}{\partial \alpha^2}$$

$$- E\left(\frac{\partial^2 \log L}{\partial \alpha^2} \right) = \Sigma \left(\frac{n_i}{P_i} + \frac{n_i}{1-P_i} \right) \left(\frac{\partial P_i}{\partial \alpha} \right)^2$$

$$= \sum_{i=1}^{l} \left(\frac{n_i}{P_i} + \frac{n_i}{1-P_i} \right) g_i^2$$

$$= \sum_{i=1}^{l} n_i w_i \qquad 3.1.32$$

$$\frac{\partial \log L}{\partial \beta} = \sum_{r=1}^{l} n_i w_i \frac{(p_i - P_i)}{g_i} \frac{\partial P_i}{\partial \beta}$$

$$\frac{\partial^2 \log L}{\partial \beta^2} = \left[- \frac{r_i}{P_i^2} \frac{\partial P_i}{\partial \beta} - \frac{(n_i - r_i)}{(1-P_i)^2} \frac{\partial P_i}{\partial \beta} \right] \frac{\partial P_i}{\partial \beta}$$

$$+ \sum_{i=1}^{l} \left(\frac{r_i}{P_i} - \frac{(n_i - r_i)}{1-P_i} \right) \frac{\partial^2 P_i}{\partial \beta^2}$$

$$- E\left(\frac{\partial^2 \log L}{\partial \beta^2} \right) = \sum_i \left(\frac{n_i}{P_i} - \frac{n_i}{(1-P_i)} \right) \left(\frac{\partial P_i}{\partial \beta} \right)^2$$

$$= \Sigma \left(\frac{n_i}{P_i} - \frac{n_i}{1-P_i} \right) x_i^2 g_i^2$$

$$= \Sigma n_i w_i x_i^2 \qquad 3.1.33$$

$$- E\left(\frac{\partial^2 \log L}{\partial\alpha\ \partial\beta}\right) = \Sigma\left(\frac{n_i}{P_i} + \frac{n_i}{1-P_i}\right)\frac{\partial P_i}{\partial\alpha}\ \frac{\partial P_i}{\partial\beta}$$

$$= \Sigma\left(\frac{n_i}{P_i} + \frac{n_i}{1-P_i}\right) g_i^2\ x_i \qquad 3.1.34$$

$$= \Sigma\ n_i w_i x_i$$

The information matrix is

$$I\ (\alpha,\beta) = \begin{bmatrix} -E\left(\frac{\partial^2 \log L}{\partial\alpha^2}\right) & -E\left(\frac{\partial^2 \log L}{\partial\beta\,\partial\alpha}\right) \\ -E\left(\frac{\partial^2 \log L}{\partial\alpha\,\partial\beta}\right) & -E\left(\frac{\partial^2 \log L}{\partial\beta^2}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma n_i w_i & \Sigma n_i w_i x_i \\ \Sigma n_i w_i x_i & \Sigma n_i w_i x_i^2 \end{bmatrix}$$

Thus information matrix is non-negative definate and its inverse is given by

$$I^{-1}(\alpha,\beta) = \begin{bmatrix} \frac{1}{\Sigma n_i w_i} & \frac{1}{\Sigma n_i w_i x_i} \\ \frac{1}{\Sigma n_i w_i x_i} & \frac{1}{\Sigma n_i w_i x_i^2} \end{bmatrix} \qquad 3.1.35$$

Hence (α, β) as obtained in 3.1.25 and 3.1.26 is asymptotically normal. that is,

$$\sqrt{n} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix} \xrightarrow{D} Y$$

where

$$Y \sim N_2[\binom{0}{0} \quad I^{-1}(\alpha,\beta)] \qquad 3.1.36$$

In view of equation 3.1.2 we obtain the estimators of log ED50 as

$$\log[\widehat{ED50}] = \frac{G^{-1}(0.5) - \hat{\alpha}}{\beta} \qquad 3.1.37$$

Remark : We feel that the above estimator of log ED50, should not be called as Maximum likelihood estimators of log ED50, because the parameter log ED50 is not a one-to-one function of (α,β). However, we show below that the above estimator is best assymptotically normal.

For this purpose, we note that in the calcuations we can take $\bar{X} = 0$, otherwise we can fit the probit regression line of Y and $X-\bar{X}$ instead of an Y and X. The effect is that

$$G^{-1}(0.5) = \bar{x}$$

where the tolerance distribution is symmetric so that in this case

$$\log (ED50) = \frac{\bar{x} - \hat{\alpha}}{\beta} \qquad 3.1.38$$

where

$$\hat{\alpha} = \bar{\hat{y}} \qquad 3.1.39$$

For the estimator of (α,β) given in 3.1.39 and 3.1.26, we have

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = 0 \qquad 3.1.40$$

Hence

$$I^{-1}(\alpha,\beta) = \begin{bmatrix} \frac{1}{\Sigma n_i w_i} & 0 \\ 0 & \frac{1}{\Sigma n_i w_i x_i^2} \end{bmatrix} \qquad 3.1.41$$

Now we know that (α, β) are best assymptotically normal.

It is known that if $\hat{\Theta}$ is a CAN with asymptotic variance co-variance matrix Σ , estimator of Θ (a vector), and if $f(\Theta)$ is real valued function of Θ, $\frac{\partial f}{\partial \Theta}$ does not Vanish, then $f(\hat{\Theta})$ is CAN for $f(\Theta)$, that is

$$\sqrt{n}(f(\hat{\Theta})-f(\Theta)) \xrightarrow{D} Y$$

where

$$Y \sim N(0, \nu(\Theta))$$

where $\nu(\Theta) = (\nabla f)' \, \Sigma(\nabla f)$, $f = (\frac{\partial \Theta}{\partial \Theta_1}, \ldots, \frac{\partial \Theta}{\partial \Theta_p})$.

Hence we have

$$\sqrt{n}(f(\hat{\alpha},\hat{\beta}) - f(\alpha,\beta)) \xrightarrow{D} Y$$

where

$$Y \sim N(0, \nu(\alpha,\beta))$$

where

$$\nu(\alpha,\beta) = \left(\frac{\partial f}{\partial \alpha}, \frac{\partial f}{\partial \beta}\right) I^{-1}(\alpha,\beta) \left(\frac{\partial f}{\partial \alpha} \; \frac{\partial f}{\partial \beta}\right)'$$

here

$$f(\alpha,\beta) = \frac{G^{-1}(0.5) - \alpha}{\beta}$$

$$\partial f/\partial\alpha = -1/\beta$$

$$\partial f/\partial\beta = \frac{(G^{-1}(0.5) - \alpha)}{\beta^2}$$

$$\nu(\alpha,\beta) = \frac{1}{\beta^2 \Sigma n_i w_i} + \frac{1}{\Sigma n_i w_i x_i^2}\left(\frac{-(G^{-1}(0.5)-\alpha)}{\beta^2}\right)^2$$

$$= \frac{1}{\beta^2 \Sigma n_i w_i} + \frac{(G^{-1}(0.5)-\alpha)^2}{\beta^4 \Sigma n_i w_i x_i^2}$$

Since $(\hat{\alpha}, \hat{\beta})$ is BAN. $f(\hat{\alpha}, \hat{\beta})$ is also BAN.

Using this theory, we develope in section2 actual selection method for minimum of normal populations based on log ED50.

3.2 Selection of normal population with smallest mean based on ED50

We state the selection rule as among K populations the population with minimum value of log $\widehat{ED50}$ should be the best one. For this we separate out two differente cases regarding distributions of populations.

1) When variances of the populations are known and equal This in turn, means that

$\beta = 1/\sigma$ is known.

2) When variances of normal populations are equal and unknown. That is, $\beta = 1/\sigma$ is unknown.

CASE 1

Variance is known, that is β is known, Therefore, $\mu = -\alpha/\beta$. Hence

$$\sqrt{n}(\hat{\mu} - \mu) = -\sqrt{n}\left(\frac{\hat{\alpha}}{\beta} - \frac{\alpha}{\beta} \right)$$

$$= -\frac{\sqrt{n}}{\beta}(\hat{\alpha} - \alpha)$$

we know that

$$\sqrt{n}(\hat{\alpha} - \alpha) \sim N\left(0, \frac{1}{(\Sigma n_i w_i)} \right)$$

Since $\sqrt{n}(\hat{\alpha} - \alpha)$ is normal, $-\sqrt{n}(\hat{\alpha} - \alpha)$ is also normal, because of symmetricity of normal distribution.

Suppose we have K different populations with distributions $N(\mu_1,\sigma^2)$, $N(\mu_2, \sigma^2)$, $N(\mu_k, \sigma^2)$.

Then because of asymptotic properties

$$\hat{\mu}_j \sim N\left(\mu_j, \frac{\sigma^2}{C_j} \right), \qquad j = 1,2,\dots,k$$

where $C_j = (\Sigma\, n_i w_i)_j$ for j^{th} population.

Thus variance of $\hat{\mu}_j$ is dependent on C_j. Hence variance of $\hat{\mu}_j$'s are different. So we have arrived at a case of unequal means and unequal variances.

If C_j were known, then we can use the selection rule for unequal known variances case and compute the probability of correct selection. Since in our case C_j are unknown, we assume that C_j are known to be equal to the estimated value $(\Sigma \widehat{n_i w_i})_j$ and use the same rule as in unequal known variance case.

This is a common approach followed in the probit regression analysis while finding confidence limit for log ED50.

Let $\sigma_j^2 = \frac{\sigma^2}{C_j}$

Thus, we have now K populations with distributions $N(\mu_1,\sigma_1^2), N(\mu_2,\sigma_2^2),\ldots,N(\mu_k,\sigma_k^2)$ respectively, where variances are assumed to be known.

CASE 2

Variance is unknown, that is β is unknown.

Here $\hat{\mu} = -\hat{\alpha}/\hat{\beta}$

Hence $\sqrt{n}(\hat{\mu}-\mu) = -\sqrt{n}\left(\frac{\hat{\alpha}}{\hat{\beta}} - \frac{\alpha}{\beta}\right)$

adding and substracting $\alpha/\hat{\beta}$

$$\sqrt{n}(\hat{\mu}-\mu) = -\sqrt{n}\left(\frac{\hat{\alpha}}{\hat{\beta}} - \frac{\alpha}{\hat{\beta}} + \frac{\alpha}{\hat{\beta}} - \frac{\alpha}{\beta}\right)$$

$$= -\sqrt{n}\,\frac{1}{\hat{\beta}}(\hat{\alpha}-\alpha) - \sqrt{n}\,\alpha\left(\frac{1}{\hat{\beta}} - \frac{1}{\beta}\right)$$

$$= -\sqrt{n}\,\frac{1}{\hat{\beta}}(\hat{\alpha}-\alpha) - \sqrt{n}\,\frac{\alpha}{\hat{\beta}\,\beta}(\beta-\hat{\beta})$$

$$= \frac{-\sqrt{n}}{\hat{\beta}}(\hat{\alpha}-\alpha) + \frac{\sqrt{n}\,\alpha}{\hat{\beta}\,\beta}(\hat{\beta}-\beta) \qquad 3.2.1$$

Since $\sqrt{n}(\hat{\alpha} - \alpha) \sim N(O, \frac{1}{\Sigma n_i w_i})$ 3.2.2

and

$$\sqrt{n}(\hat{\beta} - \beta) \sim N(O, \frac{1}{\Sigma n_i w_i x_i^2}) \qquad 3.2.3$$

From the above three expressions we get

$$\sqrt{n}(\hat{\mu} - \mu) \sim N(O, \frac{1}{\beta^2 \Sigma n_i w_i}) + N(O, \frac{\alpha^2}{\beta^4 \Sigma n_i w_i x_i^2})$$

$$= N(O, \frac{1}{\beta^2} (\frac{1}{\Sigma n_i w_i} + \frac{\alpha^2}{\beta^2 \Sigma n_i w_i x_i^2})$$

Now substitute $\mu = - \alpha/\beta$ and

Hence $\mu^2 = \alpha^2/\beta^2$

also, $\sigma^2 = 1/\beta^2$

Therefore,

$$\sqrt{n}(\hat{\mu} - \mu) \sim N(O, \sigma^2(\frac{1}{\Sigma n_i w_i} + \frac{\mu^2}{\Sigma n_i w_i x_i^2})) \qquad 3.2.4$$

Let $d = \Sigma n_i w_i$

and $c = \Sigma n_i w_i x_i^2$

therefore,

$$\sqrt{n}(\hat{\mu} - \mu) \sim N(O, \sigma^2(1/d + \mu^2/c)) \qquad 3.2.5$$

As shown in earlier section μ is best asymptotic normal.

Therefore, we have

$$I^{-1}(\hat{\mu}) = \sigma^2(\mu^2/c + 1/d) \qquad 3.2.6$$

This variance is dependent on μ. Hence using variance stabilization transformation we find a function of μ, $g(\mu)$ such that

$$\sqrt{n}\,(g(\hat{\mu}) - g(\mu)) \xrightarrow{D} Y$$

where

$$Y \sim N(0, \sigma^2)$$

and

$$\sigma^2 = (g'(\mu))^2 I^{-1}(\hat{\mu}) \qquad 3.2.7$$

From 3.2.6 and 3.2.7

$$(g'(\mu)^2 = 1/(\mu^2/c + 1/d)]$$

$$\therefore \quad g'(\mu) = \sqrt{\frac{1}{\mu^2/c + 1/d}}$$

$$= \sqrt{\frac{cd}{d\mu^2+c}}$$

$$= \frac{\sqrt{c}}{\sqrt{\mu^2 + c/d}} \qquad 3.2.8$$

Hence

$$g(\mu) = \int g'(\mu)$$

From equation 3.2.8

$$= \sqrt{c}\int \frac{1}{\sqrt{\mu^2 + c/d}}\, d\mu \qquad 3.2.9$$

make a transformation of variable as

$$u = \mu + \sqrt{\mu^2 + c/d}$$

$$du = \left(1 + \frac{\mu}{\sqrt{\mu^2 + c/d}} \right) d\mu$$

$$= \frac{\mu + \sqrt{\mu^2 + c/d}}{\sqrt{\mu^2 + c/d}} \quad d\mu$$

This gives

$$\frac{du}{u} = \frac{d\mu}{\sqrt{\mu^2 + c/d}} \qquad 3.2.10$$

From 3.2.9 and 3.2.10 we get

$$g(\mu) = \sqrt{c} \int \frac{1}{u}\, du$$

$$= \sqrt{c} \quad \log u$$

$$= \sqrt{c} \log \left(\mu + \sqrt{\mu^2 + c/d}\right)$$

substituting value of c, we get,

$$g(\mu) = \sqrt{c}\, \Sigma\, n_i w_i x_i^2 \quad \log \left(\mu + \sqrt{\frac{\mu^2 + \Sigma n_i w_i x_i^2}{\Sigma\, n_i w_i}} \right)$$

Now since $g'(\mu)$ is a positive quantity, from 3.2.8 we know that $g(\mu)$ is a strictly **increasing function of μ.** Hence we can use these $g(\mu_i)$' s for comparison of populations instead of μ_i's .

Note : The above statement is made on the basis of selection theory discussed in 1.2, where Θ is equivalent to μ and Q is equivalent to $g(\mu)$.

3.3 Application of the theory

We consider two problems in bioassays and try to apply the selection procedures based on our discussion in this dissertation. The first assay is the assay of insulin which is taken from ' Statistical method in biological assay ' by finney (on page 474). The second assay is ' The experiment of comparison of 4 analgesics , taken from ' Probit Analysis ' by Finney.

Illustration 1 : The assay of insulin

This assay was conduced by mouse convulsion method by Hemmingsen and Krogh, 1926. At each of nine doses of standard preparation and five doses at the test prepartion batches of mice were injected with a dose of insulin and the number of mice showing the symptoms of collapse or convulsions were recorded. The doses of standard prepartion measured in units of 0.001 i.u. and doses of test preparation were measured in same units on the assumption that the potency was 20 i.u. per m.g. The following four columns of the table show the dose (Z), dose metameter X = log Z, the number of mice tested (n) and the number of respondants r.

	dose Z	dose X	no. of subjects tested n	no.of respondants r
STANDARD	3.4	0.53	33	0
	5.2	0.776	32	5
	7.0	0.845	38	11
	8.5	0.929	37	14
	10.5	1.0212	40	18
	13.0	1.114	37	21
	18.0	1.255	31	23
	21.0	1.322	37	30
	28.0	1.447	30	27
			315	
TEST	6.5	0.813	40	2
	10.0	1.000	30	10
	14.0	1.146	40	18
	21.5	1.33	35	21
	29.0	1.462	37	27
			182	

The final values $\hat{\alpha}$ and $\hat{\beta}$ are obtained after second iteration.

as $\hat{\alpha}_1 = 1.638$ $\hat{\beta} = 3.2$

$\hat{\alpha}_2 = 1.08$

where $\hat{\alpha}_1$ is $\hat{\alpha}$ for standard prepartion and $\hat{\alpha}_2$ is $\hat{\alpha}$ for test prepartion.

Now, we assume that this $\hat{\beta}$ is such that / $1/\hat{\beta}$ is true estimate of σ and further consider that **it is the** known value of σ.

We know that $\mu = - \alpha/\beta$.

But since here we are using tables values of Y, for the estimation of α, β, are given in terms of probits.

we have

$$\mu = 5 - \alpha/\beta$$

For standard prepartion

$$\hat{\mu}_1 = 5 - \frac{1.638}{3.2}$$

$$= 4.488$$

For test prepartion

$$\hat{\mu}_2 = 5 - \frac{1.08}{3.2}$$

$$= 4.66$$

The minimum value is $\hat{\mu}_1$ Therefore, standard prepartion should be the best population.

Here $X_1 \sim N(\mu_1, \sigma^2)$

$X_2 \sim N(\mu_2, \sigma^2)$

$$\therefore \quad \sqrt{n}\,(\hat{\mu} - \mu) \sim N((0, \frac{\sigma^2}{\Sigma n_i w_i})$$

Hence

$$\hat{\mu}_1 \sim N((\mu_1, \frac{\sigma^2}{n(\Sigma n_i w_i)_1})$$

where $(\Sigma\, n_i w_i)_j$ is the variance of α for j^{th} population.

$$\hat{\mu}_2 \sim N\left(\mu_2, \frac{\sigma^2}{n(\Sigma n_i w_i)_2} \right)$$

$$\sigma_1^2 = \frac{\sigma^2}{n(\Sigma n_i w_i)_1}$$

$$= 6.397 \times 10^{-6}$$

$$\sigma_2^2 = \frac{\sigma^2}{n(\Sigma n_i w_i)_2}$$

$$= 1.84409 \times 10^{-5}$$

This is the case of unequal means and unequal known variance.

Further

$$\mu_1 \sim N\,(\mu_1, \sigma_1^2)$$

and

$$\mu_2 \sim N\,(\mu_2, \sigma_2^2)$$

Here the variance are assumed to be known and(since this is an asympototic distribution) it is as if we have only 1 observation on each of above 2 populations. We also observe that $\sigma_1^2 \doteqdot \sigma_2^2$.

Let σ^2 be the mean of σ_1^2 and σ_2^2.

We know that

$$P_{LFC} = \int_{-\infty}^{\infty} \phi^{k-1} \left(y + \frac{\partial^* \sqrt{n}}{\sigma}\right) f(y)\, dy$$

$$= P^*$$

From 2.1.6 .

The value of h corresponding to k = 2 and P* = 0.999 from the table A-1 on page 400 of 'Selecting and ordering populations '.

$h = 2.326$

$$\partial^* = \frac{h \times \sigma}{\sqrt{n}}$$

here

$\sigma = 0.00352$

and

$\sqrt{n} = 1$

$\therefore \quad \partial^* = 0.00816$

In our problem $\hat{\partial} = \hat{\mu}_{[2]} - \hat{\mu}_{[1]}$

$= 0.172$

Therefore, we select the standard population as the best population with probability of correct selection at least 0.999.

II Experiment of comparisn of four analgestics.

Grewal tested four drugs morphine, amidone, phenadoxone and pethidine for their potenties. The technique was to record many standard electric shocks could be applied to the tail of a mouse before the mouse squeaked, a specified dose of drug was then administered and a new shock-count recorded. If the number of shocks increased by four or more, the mouse was classed as responding. Large number of mice (60-120) were tested at several doses of each drug.

The following table gives the data for this experiment.

Substance	The log dose x	no.of subjects tested n	no. of respondants r
morphine	0.18	103	19
	0.48	120	53
	0.78	123	83
amidone	0.18	60	14
	0.48	110	54
	0.78	100	81
phenadoxone	0.12	90	31
	0.18	80	54
	0.48	90	80

	x	n	r
pethidine	0.70	60	13
	0.88	85	27
	1.00	60	32
	1.18	90	55
	1.30	60	44

The final values of $\hat{\alpha},\hat{\beta}$ and $\hat{\mu}$ obtained after 2nd are as follows :

$$\hat{\beta} = 2.4811$$

	Morphine	Amidone	Phenadxone	Pethidine
$\hat{\alpha}$	3.596	3.844	4.971	2.422
$\hat{\alpha}/\hat{\beta}$	1.449	1.549	2.0035	0.97618
$\hat{\mu}$	3.551	3.451	2.9965	4.02382

Since here we are using table values of Y which are given in terms of probits, we have

$$\hat{\mu} = 5 - \hat{\alpha} / \hat{\beta}$$

The minimum value of μ is obtained for phenadoxone. Hence we can select it as the best drug. Let us find out the probability of correct selection for this drug.

we have

$$\hat{\mu}_1 \sim N\ (\mu_1,\ \sigma_1^2)$$
$$\hat{\mu}_2 \sim N\ (\mu_2,\ \sigma_2^2)$$
$$\hat{\mu}_3 \sim M\ (\mu_3,\ \sigma_3^2)$$
$$\hat{\mu}_4 \sim N\ (\mu_4,\ \sigma_4^2)$$

where

$$\sigma_i^2 = \frac{\sigma^2}{n(\Sigma n_i w_i)} \qquad \text{for } i = 1,2,3,4$$

We suppose that the variances σ_i^2 are known and are equal to σ_i^2 above.

So there are the populations with unequal means and unequal but known variances. Further for each population we have only one obeervation for σ_i^2.

	no.of subjects n_i	variance of α_i i.e $1/(\Sigma n_i w_i)$	σ_i^2
morphine	346	0.00512	6.75×10^{-7}
amidone	270	0.006426	8.47589×10^{-7}
phenadoxone	260	0.00715	9.43×10^{-7}
Pethidine	355	0.00479	6.318×10^{-7}

We assume that values of σ_1^2 are approximately equal. Let σ^2 be the mean of σ_1^2, σ_2^2, σ_3^2, σ_4^2.

$$\sigma^2 = \frac{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2}{4}$$

$$= 7.74325 \times 10^{-7}$$

Now, we can regard this problem as that of equal and known variances. Thus using 2.2.3 we determine values of P^* and δ^*.

For $k = 4$, $P^* = 0.999$

We have $h = 4.7987$

We know that $h = \dfrac{\delta^* \sqrt{n_o}}{\sigma_o}$

here $n_o = 1$

Hence $\delta^* = h\ \sigma_o$

$= 0.00429$

In our problem $\hat{\delta} = \hat{\mu}_{[2]} - \hat{\mu}_{[1]}$

$= 0.454$

Hence we select the drug 'Phenadoxone' as the best drug with probability of correct selection at least 0.999.

REFERENCES

1] D. J. Finney — ''Probit Analysis'' Griffin.

2] D. J. Finney — ''Statistical method in biological assay'' Griffin Lonson, 1952.

3] J. D. Gibbons, I. Olkin, M. Sobel — ''Selecting and ordering populations: ''A new statistical methodology'' John Wiley and Sons, 1977.

4] S. S. Gupta, S. Panchapakesan — ''Multiple decision procedures'' John Wiley and Sons, 1979.

5] E. J. Dudewitz — ''Introduction to Statistics and Probability'' Holt, Rinehart and Winston, New York.

6] M. K. Jain, S.R.K. Iyngar and R. K. Jain — ''Numerical Methods for Scientific and Engineering Computations'' Wiley Eastarn Limited, 1985.