DOUBLE SAMPLING WITH REGRESSION IN COMPARATIVE STUDIES OF CARCASS COMPOSITION

D. CONNIFEE\(^1\) AND M. A. MORAN\(^2\)

Statistics Department, Agricultural Institute, Dublin 2.

SUMMARY

The expense of complete dissection limits replication in studies of carcass composition of large meat producing animals. This paper examines a method of improving the precision of treatment comparisons by making less expensive measurements on additional experimental animals. Problems of estimation and hypothesis testing are considered and the distribution of a test statistic is examined by simulation. The allocation of resources between direct and concomitant measurements is discussed.

1. INTRODUCTION

Direct measurement of carcass composition involves carcass dissection. This dissection largely destroys the financial value of the carcass of meat producing animals and also requires expensive laboratory facilities and staff. Considerable meat research has been devoted to the derivation of multiple regression equations relating components of carcass composition such as total lean or percentage lean to more accessible measurements, e.g., composition of sample joints or external carcass measurements. Several authors including Timon and Bichard (1965) suggest that these equations be used to predict carcass composition and so avoid dissection costs in future animal production studies. If total dissection is absolutely out of the question then this is reasonable if some assumptions can be made. One obvious assumption is that the prediction relationship is substantially unaffected by experimental treatment. Harrington (1963) notes that constant terms in the regression equations differed from treatment to treatment although certain features such as slopes did exhibit some stability.

The object of the present paper is to present a potentially more satisfactory method for using concomitant variables in place of costly direct measurements. The method is a generalisation of double sampling techniques which are familiar in sample survey theory (Cochran [1963]). It involves complete dissection of a subsample of the carcasses from the treatments while measurements of the concomitant variables only are made on the remaining carcasses. This method avoids the inevitable bias which would

\(^1\) Present address: Operational Research Unit, Agricultural Institute, Dublin, Ireland.
\(^2\) Present address: Department of Applied Statistics, University of Reading, England.
result from the use of a single equation. In general, the same variance-covariance structure is assumed for all treatments. This assumption, which implies that the various regressions are parallel, simplifies the analysis and may be tested on the experimental data.

2. ESTIMATION

Suppose we have a completely randomized design with \( t \) treatments, each replicated \( r_1 \) times. Measurements on a single concomitant variable \( x \) are made on all \( r_1 t \) experimental units and will be denoted by \( x_{ij} \) for \( i = 1, 2, \ldots, r_1 \) and \( j = 1, 2, \ldots, t \). Measurements on the variable of direct interest \( y \) are made on a subsample of \( r_2 \) experimental units for each treatment and may be denoted by \( y_{ij} \) for \( i = 1, 2, \ldots, r_2 \) and \( j = 1, 2, \ldots, t \). We assume that \((x, y)\) have a bivariate normal distribution with variances \( \sigma_1^2, \sigma_2^2 \) respectively and correlation coefficient \( \rho \), which are the same for all treatments. The means of \( x \) and \( y \) for the various treatments will be denoted by \( \mu_{1j}, \mu_{2j} \) for \( j = 1, 2, \ldots, t \).

The log likelihood of the observations may be written as

\[
L = -K - r_1 t \log \sigma_1 - \sum_{i=1}^{r_1} \sum_{j=1}^{t} \left( \frac{(x_{ij} - \mu_{ij})^2}{2\sigma_1^2} \right) - r_2 t \log \sigma_2 - \sum_{i=1}^{r_1} \sum_{j=1}^{t} \left[ \frac{(y_{ij} - \mu_{2j} - \beta(x_{ij} - \mu_{1j}))^2}{2\sigma_2^2} \right]
\]

where \( \beta = \rho(\sigma_2/\sigma_1) \) is the regression coefficient of \( y \) on \( x \), \( \sigma_{21} = \sigma_2^2(1 - \rho^2) \) is the variance of \( y \) conditional on \( x \) and \( K \) is constant. The maximum likelihood (ML) estimators of the parameters are

\[
\hat{\mu}_{2j} = \bar{y}_j + \beta(\bar{x}_{2j} - \bar{x}_i); \quad \hat{\mu}_{1j} = \bar{x}_{ij}
\]

\[
\hat{\sigma}_1^2 = \frac{\sum_{i=1}^{r_1} \sum_{j=1}^{t} (x_{ij} - \bar{x}_i)^2}{r_1 t}; \quad \hat{\sigma}_2^2 = \frac{\sum_{i=1}^{r_1} \sum_{j=1}^{t} [y_{ij} - \bar{y}_j - \beta(x_{ij} - \bar{x}_j)]^2}{r_2 t}
\]

\[
\hat{\beta} = \frac{\sum_{i=1}^{r_1} \sum_{j=1}^{t} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_j)}{\sum_{i=1}^{r_1} \sum_{j=1}^{t} (x_{ij} - \bar{x}_j)^2}
\]

where

\[
\bar{x}_{ij} = \frac{\sum_{i=1}^{r_1} x_{ij}}{r_1}, \quad \bar{x}_i = \frac{\sum_{i=1}^{r_1} x_{ij}}{r_2} \quad \text{and} \quad \bar{y}_j = \frac{\sum_{i=1}^{r_2} y_{ij}}{r_2}
\]

The mean and variance of \( \hat{\mu}_{2j} \) are easily obtained. The expectation of \( \hat{\mu}_{2j} \) conditional on \( x \) is

\[
E_{x}(\hat{\mu}_{2j}) = \mu_{2j} + \beta(\bar{x}_{2j} - \mu_{1j}) + \beta(\bar{x}_{1i} - \bar{x}_{2j}) = \mu_{2j} + \beta(\bar{x}_{1i} - \bar{\mu}_{1i})
\]

Thus \( E(\hat{\mu}_{2j}) = E_2[E_{x}(\hat{\mu}_{2j})] = \mu_{2j} \).
These ML estimators have been obtained under the assumption of a common variance-covariance structure. The estimate $\mu_2$, still unbiased even if this assumption is not correct. If $\beta_j$ for $j = 1, 2, \cdots, t$ denote the separate regression coefficients of $y$ on $x$ for the $t$ treatments then

$$E_{x_{1\times}}(\hat{\beta}) = \frac{\sum_{i=1}^{t} \sum_{j=1}^{r} (x_{i,j} - \bar{x}_{2;i})^2 \beta_j}{\sum_{i=1}^{t} \sum_{j=1}^{r} (x_{i,j} - \bar{x}_{2;i})^2}.$$  

The sums of squares in this expression are distributed independently of $(\bar{x}_{1;i} - \bar{x}_{2;i})$ and consequently

$$E[\beta(\bar{x}_{1;i} - \bar{x}_{2;i})] = E_x[\beta(\bar{x}_{1;i} - \bar{x}_{2;i})] = E_x[((\bar{x}_{1;i} - \bar{x}_{2;i})^2 / \sigma_x^2) = 0$$

as $E_x(\bar{x}_{1;i} - \bar{x}_{2;i}) = 0$. Thus $E(\hat{\mu}_{2;i}) = E(y_i) = \mu_{2;i}.$

$$V(\hat{\mu}_{2;i}) = E_x[V_{x_{1\times}}(\hat{\mu}_{2;i})] + V_x[E_{x_{1\times}}(\hat{\mu}_{2;i})]$$

$$= E_x\left[\sigma_x^2(1 - \rho^2) \left(\frac{1}{r_2} + \frac{(\bar{x}_{1;i} - \bar{x}_{2;i})^2}{S_{xx}}\right)\right] + V_x[\mu_{2;i} + \beta(\bar{x}_{1;i} - \mu_{1;i})]$$

where

$$S_{xx} = \sum_{i=1}^{r} \sum_{j=1}^{r} (x_{i,j} - \bar{x}_{2;i})^2$$

$(\bar{x}_{1;i} - \bar{x}_{2;i})^2 / S_{xx}$ is distributed as $(1 - \rho^2) \frac{1}{r_2} - \frac{1}{r_1} \frac{1}{r_2 - 1}) F$

where $F$ is an $F$ variate with $1$ and $t(r_2 - 1)$ degrees of freedom (D.F.) with expectation $[t(r_2 - 1)]/[t(r_2 - 1) - 2]$. Thus

$$V(\hat{\mu}_{2;i}) = \sigma_x^2(1 - \rho^2) \left(\frac{1}{r_2} - \frac{1}{r_1} \frac{1}{r_2 - 1) - 2} \right) + \frac{\beta^2 \sigma_x^2}{r_1}$$

$$= \sigma_x^2(1 - \rho^2) \left(\frac{1}{r_2} - \frac{1}{r_1} \right) \left(1 + \frac{1}{r_2 - 1) - 2} \right) + \frac{\sigma_x^2}{r_1}.$$  

(2.1)

A simple unbiased estimator of $V(\hat{\mu}_{2;i})$ is given by

$$V(\hat{\mu}_{2;i}) = s_{2.1}(1 - \frac{1}{r_2} - \frac{1}{r_1} \left(1 + \frac{1}{r_2 - 1) - 2} \right) + \frac{s_x^2}{r_1}.$$  

(2.2)

where $s_{2.1}$ and $s_x^2$, the usual unbiased estimators of $s_{2.1}^2$ and $s_x^2$, are

$$s_{2.1}^2 = \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{i,j} - \bar{y}_i - \hat{\beta}(x_{i,j} - \bar{x}_{2;i}))^2 / t(r_2 - 1) - 1$$

and

$$s_x^2 = \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{i,j} - \bar{y}_i)^2 / t(r_2 - 1).$$
Treatment contrasts and variances of their estimators may therefore be unbiasedly estimated. The estimator of a treatment contrast \( \sum_{i=1}^{r} \lambda_i \mu_{2i} \) is \( \sum_{i=1}^{r} \lambda_i \mu_{2i} \) with a variance \( \sum_{i=1}^{r} \lambda_i^2 \sigma^2 \). The modifications for the case of a one-way classification with unequal replications are straightforward.

The estimator (2.2) is not the ML estimator of \( V(\mu_{2i}) \) as no use is made of the additional \( x \) measurements in estimating \( \sigma^2 \). It can be shown that the ML estimator, corrected for bias, is

\[
\delta^2_{2.1} \left( \frac{1}{s_1} - \frac{1}{r_1(t_1 - 1)} + \frac{(r_1 - r_2)(r_1 - r_2 - 1)}{r_1r_2(r_1 - 1)(r_2 - 1) - 2} \right) + \frac{\beta^2_2 \delta^2_1}{r_1} \quad (2.3)
\]

where \( s_1^2 \) is the estimator \( s_1^2 \) with the usual correction for bias.

3. HYPOTHESIS TESTING

The likelihood ratio criterion is an obvious way to obtain a test of the hypothesis \( H_0 : \mu_{21} = \mu_{22} = \cdots = \mu_{2i} \). The maximum of the likelihood function without constraints on the parameters is given by:

\[
L_1(\text{Max}) = K \delta_1^{-1} e^{-r_1 t_1 / 2} \delta_2^{-1} e^{-r_2 t_2 / 2}
\]

The ML estimators of the parameters under the null hypothesis \( H_0 : \mu_{21} = \mu_{22} = \mu_{2i} = \mu_2 \) say, are

\[
\bar{\mu}_2 = \bar{y} + \bar{\beta}(\bar{x}_1 - \bar{x}_2)
\]

where

\[
\bar{y} = \frac{\sum_{i=1}^{r} \sum_{i=1}^{r} y_{ij}}{r_2 t_1} ; \quad \bar{x}_1 = \frac{\sum_{i=1}^{r} \sum_{i=1}^{r} x_{ij}}{r_1 t_1} ; \quad \bar{x}_2 = \frac{\sum_{i=1}^{r} \sum_{i=1}^{r} x_{ij}}{r_2 t_1}
\]

\[
\delta^2_1 = \frac{\sum_{i=1}^{r} \sum_{i=1}^{r} (x_{ij} - \bar{\mu}_1)^2}{r_1 t_1} ; \quad \delta^2_{2.1} = \frac{\sum_{i=1}^{r} \sum_{i=1}^{r} (y_{ij} - \bar{\mu}_2 - \bar{\beta}(x_{ij} - \bar{\mu}_1))^2}{r_2 t_1}
\]

\[
\beta = \frac{\sum_{i=1}^{r} \sum_{i=1}^{r} (x_{ij} - \bar{\mu}_1)(y_{ij} - \bar{y})}{\sum_{i=1}^{r} \sum_{i=1}^{r} (x_{ij} - \bar{\mu}_1)(x_{ij} + \bar{x}_2 - \bar{x}_1 - \bar{\mu}_1)} \quad \text{and}
\]

\[
\bar{\mu}_{1i} = \frac{r_1}{\sigma^2_1} \bar{x}_{1i} + \frac{r_2}{\sigma^2_{2.1}} \bar{\beta} \left( \bar{\mu}_2 - \bar{y} + \bar{\beta}\bar{x}_2 \right).
\]

These equations must be solved iteratively. The maximum of the likelihood function under \( H_0 \) is given by

\[
L_0(\text{Max}) = K \delta_1^{-1} e^{-r_1 t_1 / 2} \sigma_2^{-1} e^{-r_2 t_2 / 2}
\]
DOUBLE SAMPLING WITH REGRESSION

and if $H_0$ holds, the quantity

$$-2 \log l = 2 \log \left\{ \frac{r_1 \log \frac{\hat{\sigma}_1}{\hat{\sigma}_1} + r_2 \log \frac{\hat{\sigma}_{1.1}}{\hat{\sigma}_{2.1}}}{\bar{\mu}} \right\}$$

is distributed asymptotically as $\chi^2$ with $t - 1$ d.f. Since the ML estimators under the null hypothesis cannot be obtained explicitly it is not possible to investigate the exact distribution of $-2 \log l$.

4. APPROXIMATE $t$ TEST

An approximate test of a treatment contrast may be obtained by comparing

$$\frac{\sum_{i=1}^{t} \lambda_i \bar{\mu}_{1i}}{\left[ \sum_{i=1}^{t} \lambda_i^2 \bar{V}(\bar{\mu}_{1i}) \right]^{1/2}} \tag{4.1}$$

to a $t$ distribution with appropriate degrees of freedom.

These were obtained by taking $\bar{V}(\bar{\mu}_{1i})$ as proportional to a $\chi^2$ variate with $f$ d.f. and equating the first two moments.

This gives

$$f = \frac{\left[ g_1 (1 - \rho^2) + g_2 \bar{\rho}^2 \right]}{t(r_2 - 1) - 1} + \frac{2g_1 g_2 (1 - \rho^2)}{t(r_2 - 1)} + \frac{g_2}{t(r_2 - 1)} \tag{4.2}$$

where

$$g_1 = \left( \frac{1}{\bar{r}_2} - \frac{1}{r_1} \right) \left[ 1 + \frac{1}{t(r_2 - 1) - 2} \right] \quad \text{and} \quad g_2 = \frac{1}{\bar{r}_1}$$

The degrees of freedom $f$ depend on $\rho$ and must be estimated from the experimental data. Table 1 gives values of $f$ (corrected to nearest integer) for varying values of $r_1$, $r_2$, $t$ and $\rho$.

This approximation by a $t$ variate assumes that the non-normality of the numerator of (4.1) can be neglected. The adequacy of the approximation was therefore investigated by simulation for the case of two treatments and a range of values of $r_1$, $r_2$ and $\rho$.

For comparing two treatments (4.1) has the form

$$\frac{(\bar{g}_1 - \bar{g}_2) + \bar{\beta} (\bar{x}_{11} - \bar{x}_{12}) - \bar{\beta} (\bar{x}_{21} - \bar{x}_{22})}{\left[ 2(g_1 \bar{\sigma}_{1.1}^2 + g_2 \bar{\sigma}_{2.1}^2) \right]^{1/2}} \tag{4.3}$$

1000 values of (4.3) were generated for each set of values of the parameters. Each value was compared with the percentage points of a $t$ distribution with degrees of freedom given by (4.2) where $\rho$ was estimated in the usual way. The proportions of times the value of (4.3) exceeded the upper 10%, 5% and 1% points of the corresponding $t$ distribution are given in Table 2. The general agreement appears quite good.
The power of the approximate $t$ test was also obtained on the same simulations for $(\mu_{21} - \mu_{22})/\sigma_2 = 0.0(0.5)3.0$. The corresponding power of the usual $t$ test based on observations on $y$ alone was also obtained using an algorithm by Cooper [1968]. Some comparisons for tests at the 5% level for various values of $r_1$, $r_2$ and $\rho$ are presented in Table 3.

It can be seen that use of a double sampling scheme is particularly valuable where observations on the variable of direct interest are difficult to obtain as compared with observations on the concomitant variable and $r_2$ is therefore small relative to $r_1$.

It may be noted that a comparison of (2.1) with the variance $\sigma_2^2/r_2$ of the usual estimator of $\mu_{21}$ based on $r_2$ observations shows that a loss of precision can result from double sampling if $\rho^2 < 1/[t(r_2 - 1) - 1]$. 

### TABLE 1

<table>
<thead>
<tr>
<th>$\rho^2$</th>
<th>$r_1=2r_2$</th>
<th>$t = 2$</th>
<th>$t = 3$</th>
<th>$t = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>24</td>
<td>36</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>.65</td>
<td>16</td>
<td>19</td>
<td>28</td>
<td>37</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>20</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>24</td>
<td>36</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>.75</td>
<td>16</td>
<td>19</td>
<td>28</td>
<td>37</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>20</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>23</td>
<td>34</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>.85</td>
<td>16</td>
<td>18</td>
<td>26</td>
<td>35</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>19</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>30</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>.95</td>
<td>16</td>
<td>15</td>
<td>23</td>
<td>31</td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>17</td>
<td>22</td>
<td></td>
</tr>
</tbody>
</table>
5. MULTIVARIATE CASE

Results for the multivariate case with \( p \) concomitant variables are obtained by a straightforward generalization of the univariate case. In the case of the one-way classification with equal replication the \( j \)th treatment mean of \( y \) will be estimated by

\[
\mu_{2j} = \bar{y}_{2j} + \sum_{k=1}^{p} \beta_k (\bar{x}_{1jk} - \bar{x}_{2jk})
\]

where \( \bar{x}_{1jk} \) and \( \bar{x}_{2jk} \) denote the means of observations on the \( k \)th concomitant variable for the total samples and subsamples respectively. The coefficients \( \{ \beta_k \} \) are obtained from a multiple regression of \( y \) upon the \( \{ x_i \} \) where parallel regression planes are fitted for all treatments.

<table>
<thead>
<tr>
<th>( p^2 )</th>
<th>( r_1 = 2r_2 )</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>.65</td>
<td>20</td>
<td>.101</td>
<td>.058</td>
<td>.007</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>.095</td>
<td>.055</td>
<td>.008</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>.104</td>
<td>.054</td>
<td>.008</td>
</tr>
<tr>
<td>.75</td>
<td>20</td>
<td>.098</td>
<td>.052</td>
<td>.008</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>.093</td>
<td>.045</td>
<td>.010</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>.101</td>
<td>.047</td>
<td>.003</td>
</tr>
<tr>
<td>.85</td>
<td>20</td>
<td>.097</td>
<td>.047</td>
<td>.008</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>.095</td>
<td>.047</td>
<td>.008</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>.095</td>
<td>.042</td>
<td>.003</td>
</tr>
<tr>
<td>.95</td>
<td>20</td>
<td>.099</td>
<td>.046</td>
<td>.009</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>.105</td>
<td>.042</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>.103</td>
<td>.053</td>
<td>.010</td>
</tr>
<tr>
<td>$\frac{\mu_2 - \mu_1}{\sigma_2}$</td>
<td>$r_1 = 12$</td>
<td>$r_2 = 6$</td>
<td>$r_1 = 16$</td>
<td>$r_2 = 4$</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>$\alpha = 0.10$</td>
<td>Usual</td>
<td>Approximate</td>
<td>Usual</td>
<td>Approximate</td>
</tr>
<tr>
<td></td>
<td>$t$ test</td>
<td>$t$ test</td>
<td>$t$ test</td>
<td>$t$ test</td>
</tr>
<tr>
<td>$\left(\frac{\sigma^2}{n^2} = 0.65\right)$</td>
<td>$\left(\frac{\sigma^2}{n^2} = 0.95\right)$</td>
<td>$\left(\frac{\sigma^2}{n^2} = 0.65\right)$</td>
<td>$\left(\frac{\sigma^2}{n^2} = 0.95\right)$</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.100</td>
<td>0.104</td>
<td>0.103</td>
<td>0.100</td>
</tr>
<tr>
<td>0.5</td>
<td>0.326</td>
<td>0.383</td>
<td>0.444</td>
<td>0.058</td>
</tr>
<tr>
<td>1.0</td>
<td>0.647</td>
<td>0.782</td>
<td>0.865</td>
<td>0.205</td>
</tr>
<tr>
<td>1.5</td>
<td>0.866</td>
<td>0.956</td>
<td>0.980</td>
<td>0.469</td>
</tr>
<tr>
<td>2.0</td>
<td>0.979</td>
<td>0.997</td>
<td>0.999</td>
<td>0.745</td>
</tr>
<tr>
<td>2.5</td>
<td>0.998</td>
<td>0.999</td>
<td>1.000</td>
<td>0.918</td>
</tr>
<tr>
<td>3.0</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.983</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha = 0.01$</th>
<th>Usual</th>
<th>Approximate</th>
<th>Usual</th>
<th>Approximate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$ test</td>
<td>$t$ test</td>
<td>$t$ test</td>
<td>$t$ test</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{\sigma^2}{n^2} = 0.65\right)$</td>
<td>$\left(\frac{\sigma^2}{n^2} = 0.95\right)$</td>
<td>$\left(\frac{\sigma^2}{n^2} = 0.65\right)$</td>
<td>$\left(\frac{\sigma^2}{n^2} = 0.95\right)$</td>
</tr>
<tr>
<td>0.0</td>
<td>0.100</td>
<td>0.112</td>
<td>0.101</td>
<td>0.100</td>
</tr>
<tr>
<td>0.5</td>
<td>0.267</td>
<td>0.346</td>
<td>0.481</td>
<td>0.039</td>
</tr>
<tr>
<td>1.0</td>
<td>0.513</td>
<td>0.695</td>
<td>0.876</td>
<td>0.114</td>
</tr>
<tr>
<td>1.5</td>
<td>0.754</td>
<td>0.914</td>
<td>0.990</td>
<td>0.255</td>
</tr>
<tr>
<td>2.0</td>
<td>0.910</td>
<td>0.984</td>
<td>1.000</td>
<td>0.451</td>
</tr>
<tr>
<td>2.5</td>
<td>0.977</td>
<td>1.000</td>
<td>1.000</td>
<td>0.656</td>
</tr>
<tr>
<td>3.0</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
<td>0.821</td>
</tr>
</tbody>
</table>
It may be shown, as in the univariate case, that
\[
V(\mu_{2i}) = \sigma^2_2 (1 - R^2) \left(\frac{1}{r_2} - \frac{1}{r_1}\right) \left[1 + \frac{p}{(t-1)(r_2 - 1) - p - 1}\right] + \frac{\sigma^2_2}{r_1}
\]
where \(R^2\) is the multiple correlation coefficient of \(y\) with the \(\{x_k\}\). An unbiased estimator of \(V(\mu_{2i})\) is obtained by substituting the usual unbiased estimators of \(\sigma^2_2 (1 - R^2)\) and \(\sigma^2_2\). Tests of hypotheses can be obtained as for the univariate case.

A ML estimator of \(\sigma^2_2\) corresponding to (2.3) with an adjustment for bias has been obtained. It is rather complicated in comparison to the simple estimator described above and its relative advantages are currently being investigated.

### 6. RANDOMIZED BLOCK DESIGN

The most common experimental design employed in carcass composition studies is the randomized block design. We assume here that observations on \(y\) are made on a random sample of \(r_2\) blocks. The usual linear model is assumed for \(y\) and each of the \(\{x_k\}\). It may be shown that

\[
\hat{\mu}_{2i} = \bar{y}_i + \sum_{k=1}^{t} \hat{\beta}_k (\bar{x}_{1ik} - \bar{x}_{2ik})
\]

is now the maximum likelihood unbiased estimator of a treatment mean \(\mu_{2i}\). The \(\{\hat{\beta}_k\}\) are the estimators of the partial regression coefficients and are obtained from the residual line of a multivariate analysis of covariance of \(y\) with the \(\{x_k\}\). The variance and estimated variance of treatment contrasts are as for the one-way classification except that the term \(t(r_2 - 1)\) should be replaced by \((t - 1)(r_2 - 1)\). In the case of a blocked experiment \(R^2\) denotes the multiple correlation of \(y\) with the \(\{x_k\}\) within a block. Since blocks are often based on animals’ initial weights, it should be noted that in such circumstances it is the partial correlations of \(y\) with the \(\{x_k\}\) conditional on initial weight, which are of interest. These may be considerably smaller than the corresponding total correlations.

### 7. COST CONSIDERATIONS

The choice of sample numbers \(r_1\) and \(r_2\) will be influenced by the costs of direct and concomitant measurements. Assuming that an experiment has a randomized block design and that \(p\) concomitant variables are measured the variance of an estimated treatment mean \(\mu_{2i}\) is

\[
V(\hat{\mu}_{2i}) = \sigma^2_2 (1 - R^2) \left(\frac{1}{r_2} - \frac{1}{r_1}\right) \left[1 + \frac{p}{(t-1)(r_2 - 1) - p - 1}\right] + \frac{\sigma^2_2}{r_1},
\]

which may be written as

\[
V(\hat{\mu}_{2i}) = \frac{\sigma^2_2 (1 - R^2) h}{r_2} + \frac{\sigma^2_2 (1 - h)}{r_1} + \frac{\sigma^2_2 hR^2}{r_1} \tag{7.1}
\]
where

$$h = 1 + \left[ \frac{p}{(t - 1)(r_2 - 1) - p - 1} \right].$$

Let the cost of measuring $p$ concomitant variables on one animal be $C_1$ and let $C_0$ be the additional cost of a direct observation. Then, if $C$ is the total resources available,

$$C = tr_2 C_0 + tr_1 C_1.$$  \hspace{1cm} (7.2)

The minimization (7.1) subject to (7.2) leads to cumbersome equations unless simplifying assumptions are made. If we assume $h \approx 1$ the optimum value of the ratio $r_2/r_1$ is

$$\left( \frac{1 - R^2 C_1}{R^2 C_0} \right)^{1/2}.$$ 

Thus double sampling could be considered worthwhile if $R^2 > C_1/(C_0 + C_1)$. This is only a rough indication and the value of $h$ will indicate whether a more thorough investigation is required. We have assumed here that there is no additional cost in making concomitant measurements if a direct measurement is made. This is usually the case in carcass composition work. If there was such a cost then the result obtained would be equivalent to the usual result in sampling theory.

8. WORKED EXAMPLE

This worked example is based on data taken from experiments reported by Harte [1966] in which the carcass composition of several breeds of beef cattle were compared. For sake of simplicity, data for two treatments and one concomitant variable are considered. The variable of interest, total chemical lean ($y$), was measured on six animals in each treatment. The concomitant variable, chemical lean in the 8th and 9th ribs ($x$), was measured on all twelve animals on each treatment. Measurements on both variables are expressed in kilograms.

The quantities used in estimating the treatment difference and testing its significance are

$$r_1 = 12; \quad r_2 = 6,$$

$$\bar{x}_{11} = 3.84; \quad \bar{x}_{21} = 3.80; \quad y_{21} = 85.15,$$

$$\bar{x}_{12} = 2.99; \quad \bar{x}_{22} = 3.03; \quad y_{22} = 80.65.$$

The sums of squares required may be obtained from an analysis of variance and covariance of the data from completely dissected animals which are as presented in Table 4.
TABLE 4

ANALYSIS OF VARIANCE AND COVARIANCE: DATA FROM COMPLETELY DISSECTED ANIMALS

<table>
<thead>
<tr>
<th>Source</th>
<th>D.F.</th>
<th>$s_{yy}$</th>
<th>$s_{xy}$</th>
<th>$s_{xx}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>1</td>
<td>60.75</td>
<td>10.39</td>
<td>1.78</td>
</tr>
<tr>
<td>Residual</td>
<td>10</td>
<td>168.91</td>
<td>5.59</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Thus

$$s_2^2 = \frac{(168.91)}{10} = 16.89; \quad s_{2.1}^2 = \frac{1}{9} \left[ 168.91 - \frac{(5.59)^2}{0.45} \right] = 11.05$$

$$\beta = \frac{5.59}{0.45} = 12.42; \quad \beta^2 = \frac{(5.59)^2}{(0.45)(168.9)} = 0.41.$$

Estimated treatment difference

$$= (85.15 - 80.65) + 12.42[(3.84 - 2.99) - (3.80 - 3.03)]$$

$$= 5.49$$

$$g_1 = \left( \frac{1}{k} - \frac{1}{k^2} \right)(1 + \frac{1}{k}) = 0.094; \quad g_2 = \frac{1}{k} = 0.083$$

and estimated standard error

$$= \left[ 2[(0.094)(11.05) + (0.083)(16.89)] \right]^{1/2} = 2.21$$

Approximate $t$ value $= \frac{5.49}{2.21} = 2.48$.

Substitution of .41 for $\rho$ in (3.2) gives 12 D.F. to nearest integer.

If the measurements on $x$ had not been used, the estimated treatment difference, its standard error, and associated $t$ value (10 D.F.) would have been 4.50, 2.37, and 1.90, respectively.

9. CONCLUDING REMARKS

For sake of simplicity, we confine remarks to the case of a single concomitant variable. The statistical model, on which this paper is based, assumes that $(x, y)$ has a separate bivariate normal distribution for each treatment group but makes the simplifying assumption of a common variance-covariance matrix. This assumption, which implies that the regression coefficients of $y$ on $x$ are equal for all treatments, may be tested on the experimental data. It should be noted that complete stability of the regression relationship in the coefficient and the constant term from treatment to
treatment would imply that variation in the dependent variable conditional on the regressor variable was independent of treatment effect. In carcass composition studies \( x \) and \( y \) are distinct carcass measurements, often part and whole, and there is no reason to suppose that a treatment effect on one should predetermine the effect on the other. So if there are treatment effects it seems likely that regression constants will differ. If \( x \) alone has been measured, treatment effects on \( y \) cannot be estimated correctly from the observed effects on \( x \).

We have assumed the pairs of treatment means \( (\mu_{1i}, \mu_{2i}) \) to be separate and distinct and no simplifying structure was postulated for them. If a functional relationship between \( \mu_{1i} \) and \( \mu_{2i} \) was assumed, direct predictions of treatment effects on \( \mu_{2i} \) from observed effects on \( \mu_{1i} \) would be much more plausible. The existence of a functional relationship between \( \mu_{1i} \) and \( \mu_{2i} \) would be reasonable were it known that \( x \) and \( y \) basically measure the same quantity apart from a scale difference. The comparison of such scales of measurement have been thoroughly discussed by Cochran [1943]. In carcass work, however, there rarely seems reason to suppose a functional relationship between \( x \) and \( y \).

A considerable amount of carcass research, to date, has been directed towards the identification of characteristics which are highly correlated with carcass components of particular importance. There is less information available on the invariance of relationships. The recognition of the existence or non-existence of such invariance is most important and would repay investigation. This paper, for example, assumed a common variance-covariance structure for all treatments. Is this always plausible, and if not, for what characteristics and situations is it plausible?

**DOUBLE ECHANTILLONNAGE AVEC REGRESSION DANS DES ETUDES COMPARATIVES DE COMPOSITION DE CARCASSES**

**RESUME**

Les dépenses de dissections complètes limitent les répétitions dans les études de composition de carcasse des animaux de boucherie. On examine dans ce papier une méthode pour améliorer la précision des comparaisons de traitement en faisant des mesures moins coûteuses sur des animaux additionnels. On considère les problèmes d'estimation et de test d'hypothèse et on examine par simulation la distribution d'une statistique de ce test. On discute le partage des ressources entre mesures directes et concomitantes.

**REFERENCES**


Harrington, G. [1963]. Statistical problems arising in body composition studies, with special reference to the separation of technical errors and biological variation, and other


Received March 1971, Revised March 1972

Key Words: Double sampling; Regression; Covariance; Incomplete data; Carcass composition.