

Kansas State University Libraries

**New Prairie Press**

---

Conference on Applied Statistics in Agriculture

1997 - 9th Annual Conference Proceedings

---

## COVARIANCE ADJUSTMENT IN STUDIES INVOLVING OBSERVATIONAL FACTORS OR COVARIATES INFLUENCED BY TREATMENTS

O. B. Allen

I. B. Mandell

J. W. Wilton

I. B. Buchanan-Smith

*See next page for additional authors*

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

---

### Recommended Citation

Allen, O. B.; Mandell, I. B.; Wilton, J. W.; and Buchanan-Smith, I. B. (1997). "COVARIANCE ADJUSTMENT IN STUDIES INVOLVING OBSERVATIONAL FACTORS OR COVARIATES INFLUENCED BY TREATMENTS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1300>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).

---

**Author Information**

O. B. Allen, I. B. Mandell, J. W. Wilton, and I. B. Buchanan-Smith

## COVARIANCE ADJUSTMENT IN STUDIES INVOLVING OBSERVATIONAL FACTORS OR COVARIATES INFLUENCED BY TREATMENTS

O.B. Allen, I.B. Mandell, J.W. Wilton and J.B. Buchanan-Smith  
University of Guelph  
Guelph, ON N1G 2W1 Canada

### Abstract

We extend the definition of adjusted treatment means in the analysis of covariance to deal with the case where some of the covariates are influenced by treatments or where some of the factors are observational. In these cases, comparison of treatment means adjusted to a common value of the covariate may be inappropriate. Partially adjusted means are defined and it is shown that special cases include the usual adjusted means (adjusted to a common value for each of the covariates) and unadjusted means. In fact, in a multifactorial experiment, one can, by appropriate choice of adjustment, compare adjusted means for one factor but unadjusted means for the second factor. Partially adjusted means can be computed by any linear models software which will estimate linear combinations of the parameters.

**Key Words:** Covariance adjustment, observational factor, partially adjusted mean, LSMEAN, breed×diet interaction.

### 1. Introduction

The analysis of covariance generally refers to the fitting of a linear statistical model that includes both factors (such as treatments or blocks) and regressor variables. It is used for a variety of reasons (Cochran, 1957). In many instances it is useful to obtain estimates of the mean for each of the levels of the factors, adjusted for the covariates in the model. Covariance adjusted means are usually understood to be the model based estimate of the mean for each of the levels of the treatment, with the covariate evaluated at its overall mean. The analysis of covariance is most commonly used when the factors are randomized and the covariate measured before the treatments are applied as a means of increasing the precision of the treatment comparisons. In this case, the means adjusted for the covariate can be substantially more precise than the means unadjusted for the covariate.

In this paper, we wish to examine the definition of adjusted means in the case when some of the factors are not randomized or where the covariate is influenced by some of the treatment factors. Urquhart (1982) considered the two factor case, where a single covariate was influenced by one factor (herd) but not by the second factor (diet). He considered adjusted means evaluated at the herd means. We will define a general quantity which we will call a partially adjusted mean and argue that, in some settings, partially adjusted means may be more appropriate than the usual adjusted mean.

## 2. Steer Trial

In order to motivate the ideas, consider a steer trial designed to compare two breed groups, each receiving one of three dietary treatments. One breed group was the result of a rotational breeding scheme involving three medium framed breeds (Hereford, Angus, Saler) and the second breed group resulted from a rotational breeding scheme involving three large framed breeds (Charolais, Simmental, Maine-Anjou). We will refer to these subsequently simply as breeds. Spring born steers from each breed were randomly allocated to either ad libitum feeding or to a backgrounded feeding management after weaning at about 1 November. The first management group moved directly to the feedlot and were fed ad libitum. The second management group was placed on a restricted level of intake for 112 days, after which they moved to the feedlot and began ad libitum feeding. The third management group involved Fall born calves which moved directly onto ad libitum feeding after weaning at about 30 March. The steers were slaughtered when they had achieved approximately 7 mm of fat cover at the  $\frac{3}{4}$  position over the longissimus muscle at the interface of the 12<sup>th</sup> and 13<sup>th</sup> rib. The response considered in this paper is hot carcass weight. It is very natural, in comparing the dietary treatments, to include weaning weight (INITWT) as a covariate.

## 3. Partially Adjusted Means

We will initially entertain the following model to explain hot carcass weight,  $Y_{ijk}$  :

$$Y_{ijk} = \mu + \alpha_i + \delta_j + \gamma_{ij} + \beta X_{ijk} + e_{ijk} \quad (1)$$

where  $i = 1, 2$  indexes the breeds (large frame, medium frame);  $j = 1, 2, 3$  indexes management groups (backgrounded spring born, ad libitum spring born and ad libitum fall born);  $k = 1, \dots, r$  indexes the steers within each dietary-breed group.  $X_{ijk}$  is the weaning weight for the  $k$ -th steer of the  $i$ -th breed group, assigned to the  $j$ -th management group. The errors,  $e_{ijk}$ , are assumed to be uncorrelated with variance  $\sigma^2$ .

The usual covariate adjusted mean for each breed-diet combination is:

$$\begin{aligned} \hat{\mu}_{ij} &= \hat{\mu} + \hat{\alpha}_i + \hat{\delta}_j + \hat{\gamma}_{ij} + \hat{\beta} \bar{X}_{ij} \dots \\ &= \bar{Y}_{ij.} - \hat{\beta} (\bar{X}_{ij.} - \bar{X} \dots) \end{aligned} \quad (2)$$

The variance of the difference between any two of these adjusted means is

$$\text{Var} (\hat{\mu}_{ij} - \hat{\mu}_{lm}) = \left\{ \frac{2}{r} + \frac{(\bar{X}_{ij.} - \bar{X}_{lm.})^2}{E_{xx}} \right\} \sigma^2$$

where  $E_{xx}$  is the error sum of squares resulting from a two way analysis of variance of the covariate

(Snedecor and Cochran, 1980, chapter 18). The second term in the expression above represents the penalty that must be paid for comparing adjusted means.

Inclusion of weaning weight as a covariate is likely to increase the precision of the comparison of the management groups substantially. On the other hand, it may not be relevant to compare the breed means adjusted to the overall average weaning weight, since the average weaning weights for the two breeds are substantially different and this difference is inherently part of the breed effect. Thus, the comparison of the hot carcass weights of two breeds with weaning weight evaluated at a common overall average is not a comparison that would occur naturally and is likely to be biologically meaningless.

More generally, consider evaluating the adjusted breed-management group means at the covariate value  $x_{ij}$ , a value that may differ for each breed-management group combination

$$\begin{aligned}\hat{\mu}_{ij} &= \hat{\mu} + \hat{\alpha}_i + \hat{\delta}_j + \hat{\gamma}_{ij} + \hat{\beta}x_{ij} \\ &= \bar{Y}_{ij} - \hat{\beta}(\bar{X}_{ij} - x_{ij})\end{aligned}\tag{3}$$

We will refer to these as partially adjusted means. The variance of the difference of two partially adjusted means is

$$\text{Var}(\hat{\mu}_{ij} - \hat{\mu}_{kl}) = \left\{ \frac{2}{r} + \frac{[\bar{X}_{ij} - \bar{X}_{kl} - (x_{ij} - x_{kl})]^2}{E_{xx}} \right\} \sigma^2\tag{4}$$

These partially adjusted means include several special cases of interest. If  $x_{ij} = \bar{X}_{i.}$ , we have the usual adjusted means (1). If  $x_{ij} = \bar{X}_{ij}$ , the partially adjusted means become the unadjusted means  $\bar{Y}_{ij}$ .

Urquhart (1982) proposed choosing  $x_{ij} = \bar{X}_{i.}$ . In this case, the difference in the adjusted means for two management groups within the same breed takes the same form as for the usual adjusted means (2)

$$\hat{\mu}_{ij} - \hat{\mu}_{ik} = \bar{Y}_{ij} - \bar{Y}_{ik} - \hat{\beta}(\bar{X}_{ij} - \bar{X}_{ik})\tag{5}$$

However, the difference in the adjusted means for the two breeds receiving the same management is

$$\hat{\mu}_{ij} - \hat{\mu}_{kj} = \bar{Y}_{ij} - \bar{Y}_{kj} - \hat{\beta} \{ \bar{X}_{ij} - \bar{X}_{kj} - (\bar{X}_{i.} - \bar{X}_{k.}) \}\tag{6}$$

Since the management groups were randomized, the breed difference in the mean values for the covariate is likely to be similar for each of the management groups, provided the covariate is measured prior to randomizing. In this case, the adjustment term in (6) is likely to be small.

Even in the case when the partially adjusted means are evaluated at each breed-management group covariate mean (the second special case above), producing unadjusted means, there are advantages to including the covariate in the model. Only by including the covariate in the model is

one able to explore the nature and strength of the relationship between response and covariate. Furthermore, the error variance is reduced from  $\sigma_w^2$  (the estimate based on a two-factor analysis of variance) to  $(1 - r^2)\sigma_w^2$ , where  $r$  is the correlation estimated from the error line of the analysis of covariance (Snedecor and Cochran, 1980, chapter 18). In this case, the difference in the breed groups is estimated by  $\bar{Y}_{ij} - \bar{Y}_{kl}$  whether the covariate is included or not. However, the standard error we associate with this difference can be dramatically different depending on whether the covariate is included in the model or not. This highlights the fact that while the estimate is the same with or without the covariate, the basis of the inference is quite different. In the case in which the covariate is included in the model, inferences are conditional on the covariate. It is this conditioning that leads to the increase in precision.

We have not addressed the issue of the choice of covariates. The above discussion assumes that covariates that have been included in the model are statistically significant and biologically meaningful. One has, in the choice of covariates, to find a compromise between parsimony and realism. Including too many covariates of marginal significance can lead to a model in which the standard error (4) is larger than it would be if the model had fewer covariates. A fuller discussion of the issues involved in the choice of covariates can be found in Draper and Smith (1981).

### 3.1 Main Effects Model

Consider a model in which there is no breed-diet interaction

$$Y_{ijk} = \mu + \alpha_i + \delta_j + \beta X_{ijk} + e_{ijk} \quad (7)$$

and where all of the assumptions remain the same as for model (1). The partially adjusted means are now defined as

$$\begin{aligned} \hat{\mu}_{ij} &= \hat{\mu} + \hat{\alpha}_i + \hat{\delta}_j + \hat{\beta}x_{ij} \\ &= \bar{Y}_{i..} + \bar{Y}_{.j.} - \bar{Y}_{...} + \hat{\beta}(\bar{X}_{i..} + \bar{X}_{.j.} - \bar{X}_{...} - x_{ij}) \end{aligned} \quad (8)$$

If the covariate  $x_{ij}$  is evaluated at the breed mean  $\bar{X}_{i..}$ , then the difference of the partially adjusted means for two management groups is

$$\hat{\mu}_{ij} - \hat{\mu}_{ik} = \bar{Y}_{.j.} - \bar{Y}_{.k.} - \hat{\beta}(\bar{X}_{.j.} - \bar{X}_{.k.}) \quad (9)$$

Equation (9) is the same as would be obtained using the traditional definition of adjusted means (2). On the other hand, the difference in the partially adjusted means for the two breeds is the difference in the unadjusted means

$$\hat{\mu}_{ij} - \hat{\mu}_{kj} = \bar{Y}_{i..} - \bar{Y}_{k..}$$

This adjustment for the covariate is likely the most useful in practice since it produces the usual

covariance adjustment, with the corresponding increase in precision, for the randomized factor but does not adjust the observational factor to biologically unreasonable values for the covariate.

### 3.2 Partially Adjusted Main Effect Means

Having defined appropriate partially adjusted means for each breed-management group, we can define appropriate main effect means (termed population marginal means by Searle, Speed and Milliken, 1980) for the breeds averaged over the management groups and for the management groups averaged over the breeds. The partially adjusted main effect mean for breed  $i$  is defined to be

$$\hat{\mu}_i = \frac{1}{2} (\hat{\mu}_{i1} + \hat{\mu}_{i2})$$

for  $i = 1, 2$ . Similarly, the partially adjusted main effect mean for the  $j$ -th management group is

$$\hat{\mu}_j = \frac{1}{2} (\hat{\mu}_{1j} + \hat{\mu}_{2j})$$

for  $j = 1, 2, 3$ .

### 3.3 Factor-Covariate Interactions

The covariate may interact with one or more of the factors. For example, weaning weight may interact with breed group. This can be incorporated into the model by allowing the slope of the regression on weaning weight to depend on breed group

$$Y_{ijk} = \mu + \alpha_i + \delta_j + \beta_i X_{ijk} + e_{ijk} \quad (10)$$

and where all of the definitions are the same as for model (1). The partially adjusted means for the breed-diet combinations are obtained by evaluating  $X$  at  $x_{ij}$ ,

$$\begin{aligned} \hat{\mu}_{ij} &= \hat{\mu} + \hat{\alpha}_i + \hat{\delta}_j + \hat{\beta}_i x_{ij} \\ &= \bar{Y}_{i..} + \bar{Y}_{.j.} - \bar{Y}_{...} - \hat{\beta}_i (\bar{X}_{i..} - x_{ij}) - \frac{1}{I} \sum_k \hat{\beta}_k (\bar{X}_{kj.} - \bar{X}_{k..}) \end{aligned} \quad (11)$$

where  $I$  is, in general, the number of levels of the breed factor. Then the estimated difference between management groups  $j$  and  $h$  for breed  $i$  is

$$\hat{\mu}_{ij} - \hat{\mu}_{ih} = \bar{Y}_{.j.} - \bar{Y}_{.h.} + \hat{\beta}_i (x_{ij} - x_{ih}) - \frac{1}{I} \sum_k \hat{\beta}_k (\bar{X}_{kj.} - \bar{X}_{kh.}) \quad (12)$$

and the estimated difference between breeds  $i$  and  $g$  in management group  $j$  is

$$\hat{\mu}_{ij} - \hat{\mu}_{gj} = \bar{Y}_{i..} - \bar{Y}_{g..} - \hat{\beta}_i (\bar{X}_{i..} - x_{ij}) + \hat{\beta}_g (\bar{X}_{g..} - x_{gj}) \quad (13)$$

If the partially adjusted means are adjusted to the breed means ( $x_{ij} = \bar{X}_{i..}$ ), the management effect becomes

$$\bar{Y}_{.j.} + \bar{Y}_{.h.} - \frac{1}{I} \sum_k \hat{\beta}_k (\bar{X}_{kj.} - \bar{X}_{kh.}) \quad (14)$$

and the breed effect is estimated by the difference in the unadjusted means,  $\bar{Y}_{i..} - \bar{Y}_{g..}$ .

#### 4. Extensions

These ideas can be extended to any number of factors and any number of covariates. The model can include arbitrary interactions among factors, among interactions and between factors and interactions. When the cells formed by the intersection of the factors are unequally replicated, partially adjusted means are defined in terms of the model parameters and estimated as the appropriate linear combinations of the least squares estimates of these parameters.

Following is the suggested algorithm for estimating partially adjusted means. First consider the atomic cells formed by taking the intersection of the levels of each of the factors. For each atomic cell, the investigator must decide at what values to evaluate each of the covariates. The atomic cell means are then estimated, using the fitted linear model and evaluating the covariates at the chosen values. This defines the partially adjusted cell means,  $\hat{\mu}_{ij...z}$ . Main effect or population marginal means are then defined as an average of these atomic cell means, averaged over one or more factors. Partially adjusted means are estimable whenever the corresponding LSMEANS in SAS/STAT (SAS Institute Inc., 1989) are estimable.

#### 5. The Steer Trial Revisited

The steer trial described above was conducted over 3 years, with a new group of steers each year. There were a total of 252 steers used, in total. In addition to the dietary management factor described above, steers were switched from an alfalfa/grass silage diet to a high corn diet, at 28 day intervals, from day 0 to day 196 of ad libitum feeding. Because the steers ended trial when they had achieved approximately 7 mm of fat cover, the time on the high corn diet was random. In fact, some steers achieved 7 mm of backfat before their scheduled switch to corn. Because of these complications, it was decided to include total time on ad libitum feed, in days (DAYSFED) and the percent of this diet which was made up of corn (PCCORN), as two additional covariates. Although the intent was to slaughter at 7 mm of fat cover, this did not always occur. In fact, some steers were slaughtered at a fat cover substantially different from 7 mm. The actual fat cover at slaughter (GRADEFAT) was included as a fourth covariate.

Following the algorithm suggested above, we must decide at what values for each of the covariates we wish to evaluate the cell means formed by the intersection of the levels of year, diet and breed. Weaning weight was evaluated at the average for fall or spring born calves of each breed group. Days on ad libitum feeding and percent of diet composed of corn were evaluated at the mean for ad libitum or backgrounded diet for each breed group. Fat cover was evaluated at 7 mm for all



groups since this was the target, even though the average was 7.9 mm. The same values for each of the four covariates were chosen for each of the three years. Years are replicates in this experiment. A priori, there is no reason to assume that covariate values will be consistently higher or lower one year compared to another year. The chosen values are given in Table 1.

The trial was run over three years. While years was formally regarded as a fixed factor, there was no interest in estimating breed-management group means separately for each year and so main effect means were computed for each breed-management group combination, averaged over the three years. The response variable reported here is hot carcass weight. Unadjusted means, LSMEANS (SAS Institute Inc., 1989) and partially adjusted means are given in Table 2, for the two spring-born breeds receiving an ad libitum diet and in Table 3 for the large rotational breed for each of the three management groups. The partially adjusted means are more similar to the unadjusted means than to the LSMEANS for this example, although this is not always the case. For the comparison of the breeds, the difference in the LSMEANS is not significant whereas the difference in the partially adjusted means is much larger. For the comparison of the backgrounded and ad libitum groups (S1 vs S2), just the opposite is true.

The standard error for the comparison of unadjusted means will always be smaller than for LSMEANS or partially adjusted means, since there is no penalty for adjustment. (This is true only if the unadjusted means and their standard errors are computed from a model that conditions on the covariates, however.) The standard error for the comparison of partially adjusted means may be larger or smaller than the corresponding comparison of LSMEANS. For example, for the comparison of breeds in Table 2 the LSMEANS has the smaller standard error whereas for the comparison of backgrounding vs ad libitum feeding in Table 3, the difference of the partially adjusted means have the smaller standard error.

## 6. Conclusions

Partially adjusted means are proposed as a generalization to the usual adjusted treatment means in analysis of covariance. It includes as special cases, both classical adjusted means and unadjusted means. In instances where the covariate is influenced by a factor, it may be more appropriate to adjust means for the factor levels to values for the covariate which differ across the levels of the factor. A greater responsibility then falls to the investigator to choose the most appropriate levels of each covariate to which to adjust each cell formed by the intersection of the factors. However, this is preferable to computing classical adjusted means which are not meaningful.

## Acknowledgements

The authors acknowledge the support of the Ontario Ministry of Agriculture, Food and rural Affairs for the steer trial and the Natural Sciences and Engineering Research Council of Canada.

## References

- Cochran, W.G. 1957. Analysis of covariance: Its nature and uses. *Biometrics* **13**, 261-281.
- Draper and Smith. 1981. *Applied Regression Analysis, Second Edition*. New York: Wiley.
- SAS Institute Inc. 1989. *SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2*. Cary, NC:SAS Institute Inc.
- Searle, S.R., F.M. Speed and G.A. Milliken. 1980. Population marginal means in the linear model: An alternative to least squares means. *American Statistician* **34**, 216-221.
- Snedecor, G.W. and W.G. Cochran. 1980. *Statistical Methods. Seventh Edition*. Ames, IA:The Iowa State University Press.
- Urquhart, N.S. 1982. Adjustment in covariance when one factor affects the covariate. *Biometrics* **38**, 651-660.

## Appendix

Typical SAS code is given below, using proc glm, to compute partially adjusted means, for the case in which there are two factors (breed and diet) and one covariate (weaning weight). The first two estimate statements compute the partially adjusted means and their standard errors for the large and small rotational breed groups receiving diet 1. The mean for the large rotational breed group is adjusted to a weaning weight of 300 kg, whereas the small rotational breed group is adjusted to a mean of 270. The third estimate statement computes the partially adjusted main effect mean for the large rotational breed group and the fourth the partially adjusted main effect mean for diet 1.

```
proc glm;
  class breed diet;
  model gain = breed diet breed*diet initwt/solution I;
  estimate 'partially adj mean for LR-diet 1' intercept 1 breed 1
    0 diet 1 0 breed*diet 1 0 0 0 initwt 300;
  estimate 'partially adj mean for SR-diet 1' intercept 1 breed 0
    1 diet 1 0 breed*diet 0 0 1 0 initwt 270;
  estimate 'partially adj mean for LR' intercept 1 breed 1 0 diet
    .5 .5 breed*diet .5 .5 0 0 initwt 300;
  estimate 'partially adj mean for diet 1' intercept 1 breed .5 .5
    diet 1 0 breed*diet .5 0 .5 0 initwt 285;
```

**Table 1:** Values to which the covariates weaning weight (INITWT), total days on ad libitum feed (DAYSFED), percent of the diet made up of corn (PCCORN) and the fat cover at slaughter, in mm (GRADEFAT) are adjusted in computing partially adjusted means. LR and SR are the large and small rotational breed groups, respectively, and F2, S1 and S2 refer to fall born calves on an ad libitum diet, spring born calves on a backgrounded diet and spring born calves on an ad libitum diet.

	F2		S1		S2	
	LR	SR	LR	SR	LR	SR
INITWT	292.6	268.7	299.5	281.0	299.5	281.0
DAYSFED	250.7	203.5	203.5	158.1	250.7	203.5
PCCORN	44.4	39.4	34.9	29.0	44.4	39.4
GRADEFAT	7.0	7.0	7.0	7.0	7.0	7.0

**Table 2:** The unadjusted means, LSMEANS (SAS Institute Inc., 1989), partially adjusted means and the standard error of the breed difference for the trait hot carcass weight in large and small rotational breed groups, spring born, ad-libitum fed

<u>Mean</u>	<u>Large Breed</u>	<u>Small Breed</u>	<u>St. Error Diff</u>
Unadjusted	395	316	5.9
LSMEAN	342	330	6.8
Partially Adjusted	401	330	7.1

**Table 3:** The unadjusted means, LSMEANS (SAS Institute Inc., 1989) and partially adjusted means for the trait hot carcass weight, in fall born ad-libitum fed (F2), spring born backgrounded (S1) and the spring born ad libitum (S2) calves, for the large rotational breed group. The standard error of the difference is given for the comparison of spring born and fall born ad libitum groups and for the comparison of backgrounded and ad libitum spring born calves.

<u>Mean</u>	<u>F2</u>	<u>S1</u>	<u>S2</u>	<u>SE(F2-S2)</u>	<u>SE(S1-S2)</u>
Unadjusted	388	390	395	5.6	5.4
LSMEAN	349	379	342	5.9	6.1
Partially Adjusted	394	394	402	6.4	5.9