Chapter 3
Nonfull Rank Linear Models and Cell Means Models

Much of Sections 2.1 and 2.2 apply to both full rank and nonfull rank linear models. In this chapter we often assume $\boldsymbol{X}$ has rank $r < p \le n$.

3.1 Nonfull Rank Linear Models

Definition 3.1. The **nonfull rank linear model** is $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{X}$ has rank $r < p \le n$, $\boldsymbol{X}$ is an $n \times p$ matrix, $E(\boldsymbol{e}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}$.

Nonfull rank models are often used in experimental design models. Much of the nonfull rank model theory is similar to that of the full rank model, but there are some differences. Now the generalized inverse $(\boldsymbol{X}^T\boldsymbol{X})^-$ is not unique. Similarly, $\hat{\boldsymbol{\beta}}$ is a solution to the normal equations, but depends on the generalized inverse and is not unique. Some properties of the least squares estimators are summarized below. Let $\boldsymbol{P} = \boldsymbol{P}_{\boldsymbol{X}}$ be the projection matrix on $C(\boldsymbol{X})$. Recall that projection matrices are symmetric and idempotent but singular unless $\boldsymbol{P} = \boldsymbol{I}$. Also recall that $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$, so $\boldsymbol{X}^T\boldsymbol{P} = \boldsymbol{X}^T$.

Theorem 3.1. Let $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{X}$ has rank $r < p \le n$, $E(\boldsymbol{e}) = \boldsymbol{0}$, and $\text{Cov}(\boldsymbol{e}) = \sigma^2\boldsymbol{I}$.

i) $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T$ is the unique projection matrix on $C(\boldsymbol{X})$ and does not depend on the generalized inverse $(\boldsymbol{X}^T\boldsymbol{X})^-$.

ii) $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{Y}$ does depend on $(\boldsymbol{X}^T\boldsymbol{X})^-$ and is not unique.

iii) $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{P}\boldsymbol{Y}$, $\boldsymbol{r} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}$ and $RSS = \boldsymbol{r}^T\boldsymbol{r}$ are unique and so do not depend on $(\boldsymbol{X}^T\boldsymbol{X})^-$.

iv) $\hat{\boldsymbol{\beta}}$ is a solution to the *normal equations*: $\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y}$.

v) $\text{Rank}(\boldsymbol{P}) = r$ and $\text{rank}(\boldsymbol{I} - \boldsymbol{P}) = n - r$.

vi) $MSE = \dfrac{RSS}{n-r} = \dfrac{\boldsymbol{r}^T\boldsymbol{r}}{n-r}$ is an unbiased estimator of σ^2.

vii) Let the columns of $\boldsymbol{X}_1$ form a basis for $C(\boldsymbol{X})$. For example, take r linearly independent columns of $\boldsymbol{X}$ to form $\boldsymbol{X}_1$. Then $\boldsymbol{P} = \boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T$.

Proof. Parts i) follows from Theorem 2.2 a), b). For part iii), $\boldsymbol{P}$ and $\boldsymbol{I}-\boldsymbol{P}$ are projection matrices and projections $\boldsymbol{Pw}$ and $(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{w}$ are unique since projection matrices are unique. For ii), since $(\boldsymbol{X}^T\boldsymbol{X})^-$ is not unique, $\hat{\boldsymbol{\beta}}$ is not unique. Note that iv) holds since $\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{P}\boldsymbol{Y} = \boldsymbol{X}^T\boldsymbol{Y}$ since $\boldsymbol{PX} = \boldsymbol{X}$ and $\boldsymbol{X}^T\boldsymbol{P} = \boldsymbol{X}^T$. From the proof of Theorem 2.2, if $\boldsymbol{M}$ is a projection matrix, then $rank(\boldsymbol{M}) = tr(\boldsymbol{M}) =$ the number of nonzero eigenvalues of $\boldsymbol{M} = rank(\boldsymbol{X})$. Thus v) holds. vi) $E(\boldsymbol{r}^T\boldsymbol{r}) = E(\boldsymbol{e}^T(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{e}) = tr[(\boldsymbol{I}-\boldsymbol{P})\sigma^2\boldsymbol{I})] = \sigma^2(n-r)$ by Theorem 2.5. Part vii) follows from Theorem 2.2. □

Definition 3.2. Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be constant vectors. Then $\boldsymbol{a}^T\boldsymbol{\beta}$ is **estimable** if there exists a linear unbiased estimator $\boldsymbol{b}^T\boldsymbol{Y}$ so $E(\boldsymbol{b}^T\boldsymbol{Y}) = \boldsymbol{a}^T\boldsymbol{\beta}$.

The term "estimable" is misleading since there are nonestimable quantities $\boldsymbol{a}^T\boldsymbol{\beta}$ that can be estimated with biased estimators. For full rank models, $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable for any $p\times 1$ constant vector $\boldsymbol{a}$ since $\boldsymbol{a}^T\hat{\boldsymbol{\beta}}$ is a linear unbiased estimator of $\boldsymbol{a}^T\boldsymbol{\beta}$. See the Gauss Markov Theorem (Full Rank Case) 2.22. Estimable quantities tend to go with the nonfull rank linear model. We can avoid nonestimable functions by using a full rank model instead of a nonfull rank model (delete columns of $\boldsymbol{X}$ until it is full rank). From Chapter 2, the linear estimator $\boldsymbol{a}^T\boldsymbol{Y}$ of $\boldsymbol{c}^T\boldsymbol{\theta}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{c}^T\boldsymbol{\theta}$ if $E(\boldsymbol{a}^T\boldsymbol{Y}) = \boldsymbol{c}^T\boldsymbol{\theta}$, and if for any other unbiased linear estimator $\boldsymbol{b}^T\boldsymbol{Y}$ of $\boldsymbol{c}^T\boldsymbol{\theta}$, $V(\boldsymbol{a}^T\boldsymbol{Y}) \le V(\boldsymbol{b}^T\boldsymbol{Y})$. Note that $E(\boldsymbol{b}^T\boldsymbol{Y}) = \boldsymbol{c}^T\boldsymbol{\theta}$.

Since $r \le p \le n$, the model is full rank in the following theorem if $r = p$. Then the next theorem shows that the least squares estimator of an estimable function $\boldsymbol{a}^T\boldsymbol{\beta}$ is $\boldsymbol{a}^T\hat{\boldsymbol{\beta}} = \boldsymbol{b}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{b}^T\boldsymbol{PY}$.

Theorem 3.2. Let $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{X}$ has rank $r \le p \le n$, $E(\boldsymbol{e}) = \boldsymbol{0}$, and $\text{Cov}(\boldsymbol{e}) = \sigma^2\boldsymbol{I}$.

a) The quantity $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable iff $\boldsymbol{a}^T = \boldsymbol{b}^T\boldsymbol{X}$ iff $\boldsymbol{a} = \boldsymbol{X}^T\boldsymbol{b}$ (for some constant vector $\boldsymbol{b}$) iff $\boldsymbol{a} \in C(\boldsymbol{X}^T)$.

b) Let $\hat{\boldsymbol{\theta}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\theta} = \boldsymbol{X}\boldsymbol{\beta}$. Suppose there exists a constant vector $\boldsymbol{c}$ such that $E(\boldsymbol{c}^T\hat{\boldsymbol{\theta}}) = \boldsymbol{c}^T\boldsymbol{\theta}$. Then among the class of linear unbiased estimators of $\boldsymbol{c}^T\boldsymbol{\theta}$, the least squares estimator $\boldsymbol{c}^T\hat{\boldsymbol{\theta}}$ is the unique BLUE.

c) **Gauss Markov Theorem**: If $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable and a least squares estimator $\hat{\boldsymbol{\beta}}$ is any solution to the normal equations $\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y}$, then $\boldsymbol{a}^T\hat{\boldsymbol{\beta}}$ is the unique BLUE of $\boldsymbol{a}^T\boldsymbol{\beta}$.

Proof. a) If $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable, then $\boldsymbol{a}^T\boldsymbol{\beta} = E(\boldsymbol{b}^T\boldsymbol{Y}) = \boldsymbol{b}^T\boldsymbol{X}\boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$. Thus $\boldsymbol{a}^T = \boldsymbol{b}^T\boldsymbol{X}$ or $\boldsymbol{a} = \boldsymbol{X}^T\boldsymbol{b}$. Hence $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable iff $\boldsymbol{a}^T = \boldsymbol{b}^T\boldsymbol{X}$ iff $\boldsymbol{a} = \boldsymbol{X}^T\boldsymbol{b}$ iff $\boldsymbol{a} \in C(\boldsymbol{X}^T)$.

For part b), we use the proof from Seber and Lee (2003, p. 43). Since $\hat{\boldsymbol{\theta}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{P}\boldsymbol{Y}$, it follows that $E(\boldsymbol{c}^T\hat{\boldsymbol{\theta}}) = E(\boldsymbol{c}^T\boldsymbol{P}\boldsymbol{Y}) = \boldsymbol{c}^T\boldsymbol{P}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{c}^T\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{c}^T\boldsymbol{\theta}$. Thus $\boldsymbol{c}^T\hat{\boldsymbol{\theta}} = \boldsymbol{c}^T\boldsymbol{P}\boldsymbol{Y} = (\boldsymbol{P}\boldsymbol{c})^T\boldsymbol{Y}$ is a linear unbiased estimator of $\boldsymbol{c}^T\boldsymbol{\theta}$. Let $\boldsymbol{d}^T\boldsymbol{Y}$ be any other linear unbiased estimator of $\boldsymbol{c}^T\boldsymbol{\theta}$. Hence $E(\boldsymbol{d}^T\boldsymbol{Y}) = \boldsymbol{d}^T\boldsymbol{\theta} = \boldsymbol{c}^T\boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in C(\boldsymbol{X})$. So $(\boldsymbol{c}-\boldsymbol{d})^T\boldsymbol{\theta} = 0$ for all $\boldsymbol{\theta} \in C(\boldsymbol{X})$. Hence $(\boldsymbol{c}-\boldsymbol{d}) \in [C(\boldsymbol{X})]^\perp$ and $\boldsymbol{P}(\boldsymbol{c}-\boldsymbol{d}) = \boldsymbol{0}$, or $\boldsymbol{P}\boldsymbol{c} = \boldsymbol{P}\boldsymbol{d}$. Thus $V(\boldsymbol{c}^T\hat{\boldsymbol{\theta}}) = V(\boldsymbol{c}^T\boldsymbol{P}\boldsymbol{Y}) = V(\boldsymbol{d}^T\boldsymbol{P}\boldsymbol{Y}) = \sigma^2\boldsymbol{d}^T\boldsymbol{P}^T\boldsymbol{P}\boldsymbol{d} = \sigma^2\boldsymbol{d}^T\boldsymbol{P}\boldsymbol{d}$. Then $V(\boldsymbol{d}^T\boldsymbol{Y}) - V(\boldsymbol{c}^T\hat{\boldsymbol{\theta}}) = V(\boldsymbol{d}^T\boldsymbol{Y}) - V(\boldsymbol{d}^T\boldsymbol{P}\boldsymbol{Y}) = \sigma^2[\boldsymbol{d}^T\boldsymbol{d} - \boldsymbol{d}^T\boldsymbol{P}\boldsymbol{d}] = \sigma^2\boldsymbol{d}^T(\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{d} = \sigma^2\boldsymbol{d}^T(\boldsymbol{I}_n - \boldsymbol{P})^T(\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{d} = \boldsymbol{g}^T\boldsymbol{g} \geq 0$ with equality iff $\boldsymbol{g} = (\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{d} = \boldsymbol{0}$, or $\boldsymbol{d} = \boldsymbol{P}\boldsymbol{d} = \boldsymbol{P}\boldsymbol{c}$. Thus $\boldsymbol{c}^T\hat{\boldsymbol{\theta}}$ has minimum variance and is unique.

c) Since $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable, $\boldsymbol{a}^T\hat{\boldsymbol{\beta}} = \boldsymbol{b}^T\boldsymbol{X}\hat{\boldsymbol{\beta}}$. Then $\boldsymbol{a}^T\hat{\boldsymbol{\beta}} = \boldsymbol{b}^T\hat{\boldsymbol{\theta}}$ is the unique BLUE of $\boldsymbol{a}^T\boldsymbol{\beta} = \boldsymbol{b}^T\boldsymbol{\theta}$ by part b). $\square$

Remark 3.1. There are several ways to show whether $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable or nonestimable. i) For the full rank model, $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable: use the BLUE $\boldsymbol{a}^T\hat{\boldsymbol{\beta}}$. Let $\hat{\boldsymbol{\theta}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ be the least squares estimator of $\boldsymbol{X}\boldsymbol{\beta}$ where $\boldsymbol{X}$ has full rank p. a) $\boldsymbol{c}^T\hat{\boldsymbol{\theta}}$ is the unique BLUE of $\boldsymbol{c}^T\boldsymbol{\theta}$. b) $\boldsymbol{a}^T\hat{\boldsymbol{\beta}}$ is the BLUE of $\boldsymbol{a}^T\boldsymbol{\beta}$ for every vector $\boldsymbol{a}$.

Now consider the nonfull rank model. ii) If $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable: use the BLUE $\boldsymbol{a}^T\hat{\boldsymbol{\beta}}$.

iii) There are two more ways to check whether $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable.

a) If there is a constant vector $\boldsymbol{b}$ such that $E(\boldsymbol{b}^T\boldsymbol{Y}) = \boldsymbol{a}^T\boldsymbol{\beta}$, then $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable.

b) If $\boldsymbol{a}^T = \boldsymbol{b}^T\boldsymbol{X}$ or $\boldsymbol{a} = \boldsymbol{X}^T\boldsymbol{b}$ or $\boldsymbol{a} \in C(\boldsymbol{X}^T)$, then $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable. Then $\boldsymbol{b}^T\boldsymbol{Y}$ is a linear unbiased estimator of $\boldsymbol{a}^T\boldsymbol{\beta}$, and the least squares estimator $\boldsymbol{b}^T\boldsymbol{P}\boldsymbol{Y} = \boldsymbol{a}^T\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator (BLUE) in that $V(\boldsymbol{a}^T\hat{\boldsymbol{\beta}}) = V(\boldsymbol{b}^T\boldsymbol{P}\boldsymbol{Y}) \leq V(\boldsymbol{b}^T\boldsymbol{Y})$.

3.2 Cell Means Models

Nonfull rank models are often used for experimental design models, but cell means models have full rank. The cell means models will be illustrated with the one way Anova model. See Problem 3.9 for the cell means model for the two way Anova model.

Definition 3.3. Models in which the response variable Y is quantitative, but all of the predictor variables are qualitative are called *analysis of variance* (ANOVA or Anova) models, *experimental design* models, or *design of experiments* (DOE) models. Each combination of the levels of the predictors gives a different distribution for Y. A predictor variable W is often called a factor and a factor level a_i is one of the categories W can take.

The one way Anova model is used to compare p treatments. Usually there is replication and $H_0: \mu_1 = \mu_2 = \cdots = \mu_p$ is a hypothesis of interest.

Investigators may also want to rank the population means from smallest to largest.

Definition 3.4. Let $f_Z(z)$ be the pdf of Z. Then the family of pdfs $f_Y(y) = f_Z(y - \mu)$ indexed by the *location parameter* μ, $-\infty < \mu < \infty$, is the *location family* for the random variable $Y = \mu + Z$ with *standard pdf* $f_Z(z)$.

Definition 3.5. A *one way fixed effects Anova model* has a single qualitative predictor variable W with p categories a_1, ..., a_p. There are p different distributions for Y, one for each category a_i. The distribution of

$$Y|(W = a_i) \sim f_Z(y - \mu_i)$$

where the location family has second moments. Hence all p distributions come from the same location family with different location parameter μ_i and the same variance σ^2.

Notation. It is convenient to relabel the response variable $Y_1, ..., Y_n$ as the vector $\boldsymbol{Y} = (Y_{11}, ..., Y_{1,n_1}, Y_{21}, ..., Y_{2,n_2}, ..., Y_{p1}, ..., Y_{p,n_p})^T$ where the Y_{ij} are independent and $Y_{i1}, ..., Y_{i,n_i}$ are iid. Here $j = 1, ..., n_i$ where n_i is the number of cases from the ith level where $i = 1, ..., p$. Thus $n_1 + \cdots + n_p = n$. Similarly use double subscripts on the errors. Then there will be many equivalent parameterizations of the one way fixed effects Anova model.

Definition 3.6. The *cell means model* is the parameterization of the one way fixed effects Anova model such that

$$Y_{ij} = \mu_i + e_{ij}$$

where Y_{ij} is the value of the response variable for the jth trial of the ith factor level. The μ_i are the unknown means and $E(Y_{ij}) = \mu_i$. The e_{ij} are iid from the location family with pdf $f_Z(z)$ and unknown variance $\sigma^2 = \text{VAR}(Y_{ij}) = \text{VAR}(e_{ij})$. For the normal cell means model, the e_{ij} are iid $N(0, \sigma^2)$ for $i = 1, ..., p$ and $j = 1, ..., n_i$.

The cell means model is a linear model (without intercept) of the form $\boldsymbol{Y} = \boldsymbol{X}_c \boldsymbol{\beta}_c + \boldsymbol{e} =$

$$
\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1,n_1} \\ Y_{21} \\ \vdots \\ Y_{2,n_2} \\ \vdots \\ Y_{p,1} \\ \vdots \\ Y_{p,n_p} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{1,n_1} \\ e_{21} \\ \vdots \\ e_{2,n_2} \\ \vdots \\ e_{p,1} \\ \vdots \\ e_{p,n_p} \end{bmatrix}. \tag{3.1}
$$

Notation. Let $Y_{i0} = \sum_{j=1}^{n_i} Y_{ij}$ and let

$$\hat{\mu}_i = \overline{Y}_{i0} = Y_{i0}/n_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}. \tag{3.2}$$

Hence the "dot notation" means sum over the subscript corresponding to the 0, e.g. j. Similarly, $Y_{00} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} Y_{ij}$ is the sum of all of the Y_{ij}.

Let $\boldsymbol{X}_c = [\boldsymbol{v}_1 \ \ \boldsymbol{v}_2 \ \ \cdots \ \ \boldsymbol{v}_p]$, and notice that the indicator variables used in the cell means model (3.1) are $\boldsymbol{v}_{hk} = x_{hk} = 1$ if the hth case has $W = a_k$, and $\boldsymbol{v}_{hk} = x_{hk} = 0$, otherwise, for $k = 1, ..., p$ and $h = 1, ..., n$. So Y_{ij} has $x_{hk} = 1$ only if $i = k$ and $j = 1, ..., n_i$. The model can use p indicator variables for the factor instead of $p-1$ indicator variables because the model does not contain an intercept. Also notice that $(\boldsymbol{X}_c^T \boldsymbol{X}_c) = \mathrm{diag}(\mathrm{n}_1, ..., \mathrm{n}_\mathrm{p})$,

$$E(\boldsymbol{Y}) = \boldsymbol{X}_c \boldsymbol{\beta}_c = (\mu_1, ..., \mu_1, \mu_2, ..., \mu_2, ..., \mu_p, ..., \mu_p)^T,$$

and $\boldsymbol{X}_c^T \boldsymbol{Y} = (Y_{10}, ..., Y_{10}, Y_{20}, ..., Y_{20}, ..., Y_{p0}, ..., Y_{p0})^T$. Hence $(\boldsymbol{X}_c^T \boldsymbol{X}_c)^{-1} = \mathrm{diag}(1/\mathrm{n}_1, ..., 1/\mathrm{n}_\mathrm{p})$ and the OLS estimator

$$\hat{\boldsymbol{\beta}}_c = (\boldsymbol{X}_c^T \boldsymbol{X}_c)^{-1} \boldsymbol{X}_c^T \boldsymbol{Y} = (\overline{Y}_{10}, ..., \overline{Y}_{p0})^T = (\hat{\mu}_1, ..., \hat{\mu}_p)^T.$$

Thus $\hat{\boldsymbol{Y}} = \boldsymbol{X}_c \hat{\boldsymbol{\beta}}_c = (\overline{Y}_{10}, ..., \overline{Y}_{10}, ..., \overline{Y}_{p0}, ..., \overline{Y}_{p0})^T$. Hence the ijth fitted value is

$$\hat{Y}_{ij} = \overline{Y}_{i0} = \hat{\mu}_i \tag{3.3}$$

and the ijth residual is

$$r_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \hat{\mu}_i. \tag{3.4}$$

Since the cell means model is a linear model, there is an associated response plot and residual plot. However, many of the interpretations of the OLS quantities for Anova models differ from the interpretations for MLR models.

First, for MLR models, the conditional distribution $Y|\boldsymbol{x}$ makes sense even if $\boldsymbol{x}$ is not one of the observed $\boldsymbol{x}_i$ provided that $\boldsymbol{x}$ is not far from the $\boldsymbol{x}_i$. This fact makes MLR very powerful. For MLR, at least one of the variables in $\boldsymbol{x}$ is a continuous predictor. For the one way fixed effects Anova model, the p distributions $Y|\boldsymbol{x}_i$ make sense where $\boldsymbol{x}_i^T$ is a row of $\boldsymbol{X}_c$.

Also, the OLS MLR ANOVA F test for the cell means model tests $H_0 : \boldsymbol{\beta}_c = \boldsymbol{0} \equiv H_0 : \mu_1 = \cdots = \mu_p = 0$, while the one way fixed effects ANOVA F test given after Definition 3.10 tests $H_0 : \mu_1 = \cdots = \mu_p$.

Definition 3.7. Consider the one way fixed effects Anova model. The *response plot* is a plot of $\hat{Y}_{ij} \equiv \hat{\mu}_i$ versus Y_{ij} and the *residual plot* is a plot of $\hat{Y}_{ij} \equiv \hat{\mu}_i$ versus r_{ij}.

The points in the response plot scatter about the identity line and the points in the residual plot scatter about the $r = 0$ line, but the scatter need not be in an evenly populated band. A *dot plot* of $Z_1, ..., Z_m$ consists of an axis and m points each corresponding to the value of Z_i. The response plot consists of p dot plots, one for each value of $\hat{\mu}_i$. The dot plot corresponding to $\hat{\mu}_i$ is the dot plot of $Y_{i1}, ..., Y_{i,n_i}$. The p dot plots should have roughly the same amount of spread, and each $\hat{\mu}_i$ corresponds to level a_i. If a new level a_f corresponding to $\boldsymbol{x}_f$ was of interest, hopefully the points in the response plot corresponding to a_f would form a dot plot at $\hat{\mu}_f$ similar in spread to the other dot plots, but it may not be possible to predict the value of $\hat{\mu}_f$. Similarly, the residual plot consists of p dot plots, and the plot corresponding to $\hat{\mu}_i$ is the dot plot of $r_{i1}, ..., r_{i,n_i}$.

Assume that each $n_i \geq 10$. Under the assumption that the Y_{ij} are from the same location family with different parameters μ_i, each of the p dot plots should have roughly the same shape and spread. This assumption is easier to judge with the residual plot. If the response plot looks like the residual plot, then a horizontal line fits the p dot plots about as well as the identity line, and there is not much difference in the μ_i. If the identity line is clearly superior to any horizontal line, then at least some of the means differ.

Definition 3.8. An **outlier** corresponds to a case that is far from the bulk of the data. Look for a large vertical distance of the plotted point from the identity line or the $r = 0$ line.

Rule of thumb 3.1. Mentally add 2 lines parallel to the identity line and 2 lines parallel to the $r = 0$ line that cover most of the cases. Then a case is an outlier if it is well beyond these 2 lines.

This rule often fails for large outliers since often the identity line goes through or near a large outlier so its residual is near zero. A response that is far from the bulk of the data in the response plot is a "large outlier" (large in magnitude). Look for a large gap between the bulk of the data and the large outlier.

Suppose there is a dot plot of n_j cases corresponding to level a_j that is far from the bulk of the data. This dot plot is probably not a cluster of "bad outliers" if $n_j \geq 4$ and $n \geq 5p$. If $n_j = 1$, such a case may be a large outlier.

The assumption of the Y_{ij} coming from the same location family with different location parameters μ_i and the same constant variance σ^2 is a big assumption and often does not hold. Another way to check this assumption is to make a box plot of the Y_{ij} for each i. The box in the box plot corresponds to the lower, middle, and upper quartiles of the Y_{ij}. The middle quartile is just the sample median of the data m_{ij}: at least half of the $Y_{ij} \geq m_{ij}$ and at least half of the $Y_{ij} \leq m_{ij}$. The p boxes should be roughly the same length and the median should occur in roughly the same position (e.g. in the center) of each box. The "whiskers" in each plot should also be roughly similar. Histograms for each of the p samples could also be made. All of the histograms should look similar in shape.

Example 3.1. Kuehl (1994, p. 128) gives data for counts of hermit crabs on 25 different transects in each of six different coastline habitats. Let Z be the count. Then the response variable $Y = \log_{10}(Z + 1/6)$. Although the counts Z varied greatly, each habitat had several counts of 0 and often there were several counts of 1, 2, or 3. Hence Y is not a continuous variable. The cell means model was fit with $n_i = 25$ for $i = 1, ..., 6$. Each of the six habitats was a level. Figure 3.1a and b shows the response plot and residual plot. There are 6 dot plots in each plot. Because several of the smallest values in each plot are identical, it does not always look like the identity line is passing through the six sample means $\overline{Y}_{i0}$ for $i = 1, ..., 6$. In particular, examine the dot plot for the smallest mean (look at the 25 dots furthest to the left that fall on the vertical line FIT ≈ 0.36). Random noise (jitter) has been added to the response and residuals in Figure 3.1c and d. Now it is easier to compare the six dot plots. They seem to have roughly the same spread.

The plots contain a great deal of information. The response plot can be used to explain the model, check that the sample from each population (treatment) has roughly the same shape and spread, and to see which populations have similar means. Since the response plot closely resembles the residual plot in Figure 3.1, there may not be much difference in the six populations. Linearity seems reasonable since the samples scatter about the identity line. The residual plot makes the comparison of "similar shape" and "spread" easier.

Definition 3.9. a) The *total sum of squares*

$$SSTO = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{00})^2.$$

b) The *treatment sum of squares*

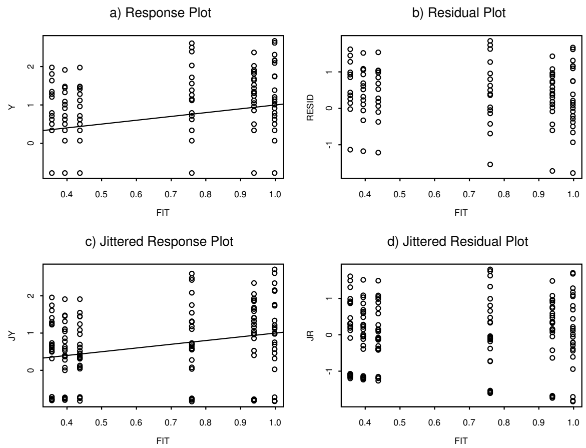

Fig. 3.1 Plots for Crab Data

$$SSTR = \sum_{i=1}^{p} n_i(\overline{Y}_{i0} - \overline{Y}_{00})^2.$$

c) The residual sum of squares or *error sum of squares*

$$SSE = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i0})^2.$$

Definition 3.10. Associated with each SS in Definition 3.9 is a *degrees of freedom* (df) and a *mean square* = *SS*/*df*. For SSTO, df = $n-1$ and $MSTO = SSTO/(n-1)$. For SSTR, df = $p-1$ and $MSTR = SSTR/(p-1)$. For SSE, $df = n - p$ and $MSE = SSE/(n-p)$.

Let $S_i^2 = \sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i0})^2/(n_i - 1)$ be the sample variance of the ith group. Then the MSE is a weighted sum of the S_i^2:

$$\hat{\sigma}^2 = MSE = \frac{1}{n-p}\sum_{i=1}^{p}\sum_{j=1}^{n_i} r_{ij}^2 = \frac{1}{n-p}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i0})^2 =$$

$$\frac{1}{n-p}\sum_{i=1}^{p}(n_i - 1)S_i^2 = S_{pool}^2$$

where S_{pool}^2 is known as the pooled variance estimator.

The ANOVA F test tests whether the p means are equal. If H_0 is not rejected and the means are equal, then it is possible that the factor is unimportant, but **it is also possible that the factor is important but the level is not**. For example, the factor might be type of catalyst. The yield may be equally good for each type of catalyst, but there would be no yield if no catalyst was used.

The ANOVA table is the same as that for MLR, except that SSTR replaces the regression sum of squares. The MSE is again an estimator of σ^2. The ANOVA F test tests whether all p means μ_i are equal. Shown below is an ANOVA table given in symbols. Sometimes "Treatment" is replaced by "Between treatments," "Between Groups," "Model," "Factor," or "Groups." Sometimes "Error" is replaced by "Residual," or "Within Groups." Sometimes "p-value" is replaced by "P", "$Pr(> F)$," or "PR > F." The "p-value" is nearly always an estimated p-value, denoted by pval.

Summary Analysis of Variance Table

Source	df	SS	MS	F	p-value
Treatment	$p-1$	SSTR	MSTR	F_0=MSTR/MSE	for H_0:
Error	$n-p$	SSE	MSE		$\mu_1 = \cdots = \mu_p$

Here is the 4 step fixed effects one way ANOVA F test of hypotheses.
i) State the hypotheses $H_0 : \mu_1 = \mu_2 = \cdots = \mu_p$ and H_A: not H_0.
ii) Find the test statistic $F_0 = MSTR/MSE$ or obtain it from output.
iii) Find the pval from output or use the F–table: pval =

$$P(F_{p-1,n-p} > F_0).$$

iv) State whether you reject H_0 or fail to reject H_0. If the pval $\leq \delta$, reject H_0 and conclude that the mean response depends on the factor level. (Hence not all of the treatment means are equal.) Otherwise fail to reject H_0 and conclude that the mean response does not depend on the factor level. (Hence all of the treatment means are equal, or there is not enough evidence to conclude that the mean response depends on the factor level.) Give a nontechnical sentence.

Rule of thumb 3.2. If

$$\max(S_1, ..., S_p) \leq 2 \min(S_1, ..., S_p),$$

then the one way ANOVA F test results will be approximately correct if the response and residual plots suggest that the remaining one way Anova model assumptions are reasonable. See Moore (2007, p. 634). If all of the $n_i \geq 5$, replace the standard deviations by the ranges of the dot plots when examining the response and residual plots. The range $R_i = \max(Y_{i,1}, ..., Y_{i,n_i}) - \min(Y_{i,1}, ..., Y_{i,n_i})$ = length of the ith dot plot for $i = 1, ..., p$.

The assumption that the zero mean iid errors have constant variance $V(e_{ij}) \equiv \sigma^2$ is much stronger for the one way Anova model than for the multiple linear regression model. The assumption implies that the p population distributions have pdfs from the same location family with different means $\mu_1, ..., \mu_p$ but the same variances $\sigma_1^2 = \cdots = \sigma_p^2 \equiv \sigma^2$. The one way ANOVA F test has some resistance to the constant variance assumption, but confidence intervals have much less resistance to the constant variance assumption. Consider confidence intervals for μ_i such as $\overline{Y}_{i0} \pm t_{n_i-1,1-\delta/2}\sqrt{MSE}/\sqrt{n_i}$. MSE is a weighted average of the S_i^2. Hence MSE overestimates small σ_i^2 and underestimates large σ_i^2 when the σ_i^2 are not equal. Hence using $\sqrt{MSE}$ instead of S_i will make the CI too long or too short, and Rule of thumb 3.2 does not apply to confidence intervals based on MSE.

All of the parameterizations of the one way fixed effects Anova model yield the same predicted values, residuals, and ANOVA F test, but the interpretations of the parameters differ. The cell means model is a linear model (without intercept) of the form $\boldsymbol{Y} = \boldsymbol{X}_c\boldsymbol{\beta}_c + \boldsymbol{e} =$ that can be fit using OLS. The OLS MLR output gives the correct fitted values and residuals but an incorrect ANOVA table. An equivalent linear model (with intercept) with correct OLS MLR ANOVA table as well as residuals and fitted values can be formed by replacing any column of the cell means model by a column of ones $\mathbf{1}$. Removing the last column of the cell means model and making the first column $\mathbf{1}$ gives the model $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1}x_{p-1} + e$ given in matrix form by (3.5) below.

It can be shown that the OLS estimators corresponding to (3.5) are $\hat{\beta}_0 = \overline{Y}_{p0} = \hat{\mu}_p$, and $\hat{\beta}_i = \overline{Y}_{i0} - \overline{Y}_{p0} = \hat{\mu}_i - \hat{\mu}_p$ for $i = 1, ..., p-1$. The cell means model has $\hat{\beta}_i = \hat{\mu}_i = \overline{Y}_{i0}$.

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1,n_1} \\ Y_{21} \\ \vdots \\ Y_{2,n_2} \\ \vdots \\ Y_{p,1} \\ \vdots \\ Y_{p,n_p} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{1,n_1} \\ e_{21} \\ \vdots \\ e_{2,n_2} \\ \vdots \\ e_{p,1} \\ \vdots \\ e_{p,n_p} \end{bmatrix}. \tag{3.5}$$

Definition 3.11. A **contrast** $C = \sum_{i=1}^{p} k_i \mu_i$ where $\sum_{i=1}^{p} k_i = 0$. The estimated contrast is $\hat{C} = \sum_{i=1}^{p} k_i \overline{Y}_{i0}$.

If the null hypothesis of the fixed effects one way ANOVA test is not true, then not all of the means μ_i are equal. Researchers will often have hypotheses, before examining the data, that they desire to test. Often such a hypothesis can be put in the form of a contrast. For example, the contrast $C = \mu_i - \mu_j$ is used to compare the means of the ith and jth groups while the contrast $\mu_1 - (\mu_2 + \cdots + \mu_p)/(p-1)$ is used to compare the last $p-1$ groups with the 1st group. This contrast is useful when the 1st group corresponds to a standard or control treatment while the remaining groups correspond to new treatments.

Assume that the normal cell means model is a useful approximation to the data. Then the $\overline{Y}_{i0} \sim N(\mu_i, \sigma^2/n_i)$ are independent, and

$$\hat{C} = \sum_{i=1}^{p} k_i \overline{Y}_{i0} \sim N\left(C, \sigma^2 \sum_{i=1}^{p} \frac{k_i^2}{n_i}\right).$$

Hence the standard error

$$SE(\hat{C}) = \sqrt{MSE \sum_{i=1}^{p} \frac{k_i^2}{n_i}}.$$

The degrees of freedom is equal to the MSE degrees of freedom $= n - p$.

Consider a family of null hypotheses for contrasts $\{Ho : \sum_{i=1}^{p} k_i \mu_i = 0$ where $\sum_{i=1}^{p} k_i = 0$ and the k_i may satisfy other constraints$\}$. Let δ_S denote the probability of a type I error for a single test from the family where a type I error is a false rejection. The **family level** δ_F is an upper bound on the (usually unknown) size δ_T. Know how to interpret $\delta_F \approx \delta_T =$
P(of making at least one type I error among the family of contrasts).

Two important families of contrasts are the family of all possible contrasts and the family of pairwise differences $C_{ij} = \mu_i - \mu_j$ where $i \neq j$. The Scheffé multiple comparisons procedure has a δ_F for the family of all possible contrasts, while the Tukey multiple comparisons procedure has a δ_F for the family of all $\binom{p}{2}$ pairwise contrasts.

3.3 Summary

1) The **nonfull rank linear model:** suppose $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ where $\boldsymbol{X}$ has rank $r < p$ and $\boldsymbol{X}$ is an $n \times p$ matrix.

i) $\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T$ is the unique projection matrix on $C(\boldsymbol{X})$ and does not depend on the generalized inverse $(\boldsymbol{X}^T\boldsymbol{X})^{-}$.

ii) $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{Y}$ does depend on $(\boldsymbol{X}^T\boldsymbol{X})^-$ and is not unique.

iii) $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y}$, $\boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Y}$ and $RSS = \boldsymbol{e}^T\boldsymbol{e}$ are unique and so do not depend on $(\boldsymbol{X}^T\boldsymbol{X})^-$.

iv) $\hat{\boldsymbol{\beta}}$ is a solution to the *normal equations*: $\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y}$.

v) It can be shown that $\text{rank}(\boldsymbol{P}_{\boldsymbol{X}}) = r$ and $\text{rank}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}}) = n - r$.

vi) Let $\hat{\boldsymbol{\theta}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\theta} = \boldsymbol{X}\boldsymbol{\theta}$. Suppose there exists a constant vector $\boldsymbol{c}$ such that $E(\boldsymbol{c}^T\hat{\boldsymbol{\theta}}) = \boldsymbol{c}^T\boldsymbol{\theta}$. Then among the class of linear unbiased estimators of $\boldsymbol{c}^T\boldsymbol{\theta}$, the least squares estimator $\boldsymbol{c}^T\hat{\boldsymbol{\theta}}$ is BLUE.

vii) If $\text{Cov}(\boldsymbol{Y}) = \text{Cov}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}$, then $MSE = \dfrac{RSS}{n-r} = \dfrac{\boldsymbol{e}^T\boldsymbol{e}}{n-r}$ is an unbiased estimator of σ^2.

viii) Let the columns of $\boldsymbol{X}_1$ form a basis for $C(\boldsymbol{X})$. For example, take r linearly independent columns of $\boldsymbol{X}$ to form $\boldsymbol{X}_1$. Then $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T$.

2) Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be constant vectors. Then $\boldsymbol{a}^T\boldsymbol{\beta}$ is **estimable** if there exists a linear unbiased estimator $\boldsymbol{b}^T\boldsymbol{Y}$ so $E(\boldsymbol{b}^T\boldsymbol{Y}) = \boldsymbol{a}^T\boldsymbol{\beta}$.

3) The quantity $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable iff $\boldsymbol{a}^T = \boldsymbol{b}^T\boldsymbol{X}$ iff $\boldsymbol{a} = \boldsymbol{X}^T\boldsymbol{b}$ (for some constant vector $\boldsymbol{b}$) iff $\boldsymbol{a} \in C(\boldsymbol{X}^T)$.

4) If $\boldsymbol{a}^T\boldsymbol{\beta}$ is estimable and a least squares estimator $\hat{\boldsymbol{\beta}}$ is any solution to the normal equations $\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y}$. Then $\boldsymbol{a}^T\boldsymbol{\beta}$ is unique and $\boldsymbol{a}^T\hat{\boldsymbol{\beta}}$ is the BLUE of $\boldsymbol{a}^T\boldsymbol{\beta}$.

5) The term "estimable" is misleading since there are nonestimable quantities $\boldsymbol{a}^T\boldsymbol{\beta}$ that can be estimated with biased or nonlinear estimators.

6) Estimable quantities tend to go with the nonfull rank linear model. Can avoid nonestimable functions by using a full rank model instead of a nonfull rank model (delete columns of $\boldsymbol{X}$ until it is full rank).

7) The linear estimator $\boldsymbol{a}^T\boldsymbol{Y}$ of $\boldsymbol{c}^T\boldsymbol{\theta}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{c}^T\boldsymbol{\theta}$ if $E(\boldsymbol{a}^T\boldsymbol{Y}) = \boldsymbol{c}^T\boldsymbol{\theta}$, and if for any other unbiased linear estimator $\boldsymbol{b}^T\boldsymbol{Y}$ of $\boldsymbol{c}^T\boldsymbol{\theta}$, $V(\boldsymbol{a}^T\boldsymbol{Y}) \le V(\boldsymbol{b}^T\boldsymbol{Y})$. Note that $E(\boldsymbol{b}^T\boldsymbol{Y}) = \boldsymbol{c}^T\boldsymbol{\theta}$.

8) Let $\hat{\boldsymbol{\theta}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ be the least squares estimator of $\boldsymbol{X}\boldsymbol{\beta}$ where $\boldsymbol{X}$ has full rank p. a) $\boldsymbol{c}^T\hat{\boldsymbol{\theta}}$ is the unique BLUE of $\boldsymbol{c}^T\boldsymbol{\theta}$. b) $\boldsymbol{a}^T\hat{\boldsymbol{\beta}}$ is the BLUE of $\boldsymbol{a}^T\boldsymbol{\beta}$ for every vector $\boldsymbol{a}$.

9) In experimental design models or design of experiments (DOE), the entries of $\boldsymbol{X}$ are coded, often as $-1, 0$ or 1. Often $\boldsymbol{X}$ is not a full rank matrix.

10) Some DOE models have one Y_i per $\boldsymbol{x}_i$ and lots of $\boldsymbol{x}_i$'s. Then the response and residual plots are used like those for MLR.

11) Some DOE models have n_i Y_i's per $\boldsymbol{x}_i$, and only a few distinct values of $\boldsymbol{x}_i$. Then the response and residual plots no longer look like those for MLR.

12) A *dot plot* of $Z_1, ..., Z_m$ consists of an axis and m points each corresponding to the value of Z_i.

13) Let $f_Z(z)$ be the pdf of Z. Then the family of pdfs $f_Y(y) = f_Z(y - \mu)$ indexed by the *location parameter* μ, $-\infty < \mu < \infty$, is the *location family* for the random variable $Y = \mu + Z$ with *standard pdf* $f_Z(y)$. A one way fixed effects ANOVA model has a single qualitative predictor variable W with p

categories a_1, ..., a_p. There are p different distributions for Y, one for each category a_i. The distribution of

$$Y|(W = a_i) \sim f_Z(y - \mu_i)$$

where the location family has second moments. Hence all p distributions come from the same location family with different location parameter μ_i and the same variance σ^2. The one way fixed effects normal ANOVA model is the special case where $Y|(W = a_i) \sim N(\mu_i, \sigma^2)$.

14) The *response plot* is a plot of $\hat{Y}$ versus Y. For the one way Anova model, the response plot is a plot of $\hat{Y}_{ij} = \hat{\mu}_i$ versus Y_{ij}. Often the identity line with unit slope and zero intercept is added as a visual aid. Vertical deviations from the identity line are the residuals $e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \hat{\mu}_i$. The plot will consist of p dot plots that scatter about the identity line with similar shape and spread if the fixed effects one way ANOVA model is appropriate. The ith dot plot is a dot plot of $Y_{i,1}, ..., Y_{i,n_i}$. Assume that each $n_i \geq 10$. If the response plot looks like the residual plot, then a horizontal line fits the p dot plots about as well as the identity line, and there is not much difference in the μ_i. If the identity line is clearly superior to any horizontal line, then at least some of the means differ.

The *residual plot* is a plot of $\hat{Y}$ versus e where the residual $e = Y - \hat{Y}$. The plot will consist of p dot plots that scatter about the $e = 0$ line with similar shape and spread if the fixed effects one way ANOVA model is appropriate. The ith dot plot is a dot plot of $e_{i,1}, ..., e_{i,n_i}$. Assume that each $n_i \geq 10$. Under the assumption that the Y_{ij} are from the same location scale family with different parameters μ_i, each of the p dot plots should have roughly the same shape and spread. This assumption is easier to judge with the residual plot than with the response plot.

15) Rule of thumb: Let R_i be the range of the ith dot plot = $\max(Y_{i1}, ..., Y_{i,n_i}) - \min(Y_{i1}, ..., Y_{i,n_i})$. If the $n_i \approx n/p$ and if $\max(R_1, ..., R_p) \leq 2\min(R_1, ..., R_p)$, then the one way ANOVA F test results will be approximately correct if the response and residual plots suggest that the remaining one way ANOVA model assumptions are reasonable. Confidence intervals need stronger assumptions.

16) Let $Y_{i0} = \sum_{j=1}^{n_i} Y_{ij}$ and let

$$\hat{\mu}_i = \overline{Y}_{i0} = Y_{i0}/n_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}.$$

Hence the "dot notation" means sum over the subscript corresponding to the 0, e.g. j. Similarly, $Y_{00} = \sum_{i=1}^{p}\sum_{j=1}^{n_i} Y_{ij}$ is the sum of all of the Y_{ij}. Be able to find $\hat{\mu}_i$ from data.

17) The **cell means model** for the fixed effects one way Anova is $Y_{ij} = \mu_i + \epsilon_{ij}$ where Y_{ij} is the value of the response variable for the jth trial of the

ith factor level for $i = 1, ..., p$ and $j = 1, ..., n_i$. The μ_i are the unknown means and $E(Y_{ij}) = \mu_i$. The ϵ_{ij} are iid from the location family with pdf $f_Z(z)$, zero mean and unknown variance $\sigma^2 = V(Y_{ij}) = V(\epsilon_{ij})$. For the normal cell means model, the ϵ_{ij} are iid $N(0, \sigma^2)$. The estimator $\hat{\mu}_i = \overline{Y}_{i0} = \sum_{j=1}^{n_i} Y_{ij}/n_i = \hat{Y}_{ij}$. The ith residual is $e_{ij} = Y_{ij} - \overline{Y}_{i0}$, and $\overline{Y}_{00}$ is the sample mean of all of the Y_{ij} and $n = \sum_{i=1}^p n_i$. The total sum of squares SSTO $= \sum_{i=1}^p \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{00})^2$, the treatment sum of squares SSTR $= \sum_{i=1}^p n_i (\overline{Y}_{i0} - \overline{Y}_{00})^2$, and the error sum of squares SSE = RSS $= \sum_{i=1}^p \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i0})^2$. The MSE is an estimator of σ^2. The Anova table is the same as that for multiple linear regression, except that SSTR replaces the regression sum of squares and that SSTO, SSTR and SSE have $n-1$, $p-1$ and $n-p$ degrees of freedom.

Summary Analysis of Variance Table

Source	df	SS	MS	F	p-value
Treatment	$p-1$	SSTR	MSTR	F_0=MSTR/MSE	for H_0:
Error	$n-p$	SSE	MSE		$\mu_1 = \cdots = \mu_p$

18) Shown is a one way ANOVA table given in symbols. Sometimes "Treatment" is replaced by "Between treatments," "Between Groups," "Model," "Factor" or "Groups." Sometimes "Error" is replaced by "Residual," or "Within Groups." Sometimes "p-value" is replaced by "P", "$Pr(> F)$" or "PR > F." SSE is often replaced by RSS = residual sum of squares.

19) In matrix form, the cell means model is the linear model without an intercept (although $\mathbf{1} \in C(\boldsymbol{X})$), where $\boldsymbol{\mu} = \boldsymbol{\beta} = (\mu_1, ..., \mu_p)^T$, and $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\mu} + \boldsymbol{\epsilon} =$

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1,n_1} \\ Y_{21} \\ \vdots \\ Y_{2,n_2} \\ \vdots \\ Y_{p,1} \\ \vdots \\ Y_{p,n_p} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1,n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2,n_2} \\ \vdots \\ \epsilon_{p,1} \\ \vdots \\ \epsilon_{p,n_p} \end{bmatrix}.$$

20) For the cell means model, $\boldsymbol{X}^T\boldsymbol{X} = diag(n_1, ..., n_p)$, $(\boldsymbol{X}^T\boldsymbol{X})^{-1} = diag(1/n_1, ..., 1/n_p)$, and $\boldsymbol{X}^T\boldsymbol{Y} = (Y_{10}, ..., Y_{p0})^T$. So $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\mu}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ $= (\overline{Y}_{10}, ..., \overline{Y}_{p0})^T$. Then $\hat{\boldsymbol{Y}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{\mu}}$, and $\hat{Y}_{ij} = \overline{Y}_{i0}$. Hence the ijth residual $e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \overline{Y}_{i0}$ for $i = 1, ..., p$ and $j = 1, ..., n_i$.

21) In the response plot, the dot plot for the jth treatment crosses the identity line at $\overline{Y}_{j0}$.

22) The one way Anova F test has hypotheses $H_0 : \mu_1 = \cdots = \mu_p$ and H_A: not H_0 (not all of the p population means are equal). The one way Anova table for this test is given above 18). Let RSS = SSE. The test statistic

$$F = \frac{MSTR}{MSE} = \frac{[RSS(H) - RSS]/(p-1)}{MSE} \sim F_{p-1,n-p}$$

if the ϵ_{ij} are iid $N(0, \sigma^2)$. If H_0 is true, then $Y_{ij} = \mu + \epsilon_{ij}$ and $\hat{\mu} = \overline{Y}_{00}$. Hence $RSS(H) = SSTO = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{00})^2$. Since $SSTO = SSE + SSTR$, the quantity $SSTR = RSS(H) - RSS$, and $MSTR = SSTR/(p-1)$.

23) The one way Anova F test is a large sample test if the ϵ_{ij} are iid with mean 0 and variance σ^2. Then the Y_{ij} come from the same location family with the same variance $\sigma_i^2 = \sigma^2$ and different mean μ_i for $i = 1, ..., p$. Thus the p treatments (groups, populations) have the same variance $\sigma_i^2 = \sigma^2$. The $V(\epsilon_{ij}) \equiv \sigma^2$ assumption (which implies that $\sigma_i^2 = \sigma^2$ for $i = 1, ..., p$) is a much stronger assumption for the one way Anova model than for MLR, but the test has some resistance to the assumption that $\sigma_i^2 = \sigma^2$ by 15).

24) Other design matrices $\boldsymbol{X}$ can be used for the full model. One design matrix adds a column of ones to the cell means design matrix. This model is no longer a full rank model.

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1,n_1} \\ Y_{21} \\ \vdots \\ Y_{2,n_2} \\ \vdots \\ Y_{p,1} \\ \vdots \\ Y_{p,n_p} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1,n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2,n_2} \\ \vdots \\ \epsilon_{p,1} \\ \vdots \\ \epsilon_{p,n_p} \end{bmatrix}.$$

25) A full rank one way Anova model with an intercept adds a constant but deletes the last column of the $\boldsymbol{X}$ for the cell means model. Then $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{Y}$ and $\boldsymbol{\epsilon}$ are as in the cell means model. Then $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_{p-1})^T = (\mu_p, \mu_1 - \mu_p, \mu_2 - \mu_p, ..., \mu_{p-1} - \mu_p)^T$. So $\beta_0 = \mu_p$ and $\beta_i = \mu_i - \mu_p$ for $i = 1, ..., p-1$.

It can be shown that the OLS estimators are $\hat{\beta}_0 = \overline{Y}_{p0} = \hat{\mu}_p$, and $\hat{\beta}_i = \overline{Y}_{i0} - \overline{Y}_{p0} = \hat{\mu}_i - \hat{\mu}_p$ for $i = 1, ..., p-1$. (The cell means model has $\hat{\beta}_i = \hat{\mu}_i = \overline{Y}_{i0}$.) In matrix form the model is shown above.

Then $\boldsymbol{X}^T\boldsymbol{Y} = (Y_{00}, Y_{10}, Y_{20}, ..., Y_{p-1,0})^T$ and

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} n & n_1 & n_2 & n_3 & \cdots & n_{p-2} & n_{p-1} \\ n_1 & n_1 & 0 & 0 & \cdots & 0 & 0 \\ n_2 & 0 & n_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ n_{p-2} & 0 & 0 & 0 & \cdots & n_{p-2} & 0 \\ n_{p-1} & 0 & 0 & 0 & \cdots & 0 & n_{p-1} \end{bmatrix} = \begin{bmatrix} n & (n_1 \ n_2 \ \cdots \ n_{p-1}) \\ \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_{p-1} \end{pmatrix} & diag(n_1, ..., n_{p-1}) \end{bmatrix}.$$

$$\text{Hence} \quad (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1} = \frac{1}{\mathrm{n_p}} \begin{bmatrix} 1 & -1 & -1 & -1 & \cdots & -1 & -1 \\ -1 & 1+\frac{n_p}{n_1} & 1 & 1 & \cdots & 1 & 1 \\ -1 & 1 & 1+\frac{n_p}{n_2} & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ -1 & 1 & 1 & 1 & \cdots & 1+\frac{n_p}{n_{p-2}} & 1 \\ -1 & 1 & 1 & 1 & \cdots & 1 & 1+\frac{n_p}{n_{p-1}} \end{bmatrix} =$$

$$\frac{1}{n_p} \begin{bmatrix} 1 & -\boldsymbol{1}^T \\ -\boldsymbol{1} & \boldsymbol{1}\boldsymbol{1}^T + diag(\frac{n_p}{n_1}, ..., \frac{n_p}{n_{p-1}}) \end{bmatrix}.$$

This model is interesting since the one way Anova F test of $H_0 : \mu_1 = \cdots = \mu_p$ versus H_A : not H_0 corresponds to the MLR Anova F test of $H_0 : \beta_1 = \cdots = \beta_{p-1} = 0$ versus H_A : not H_0.

26) A contrast $\theta = \sum_{i=1}^p c_i\mu_i$ where $\sum_{i=1}^p c_i = 0$. The estimated contrast is $\hat{\theta} = \sum_{i=1}^p c_i\overline{Y}_{i0}$. Then $SE(\hat{\theta}) = \sqrt{MSE}\sqrt{\sum_{i=1}^p \frac{c_i^2}{n_i}}$ and a $100(1-\delta)\%$ CI for θ is $\hat{\theta} \pm t_{n-1,1-\delta/2}SE(\hat{\theta})$. CIs for one way Anova are less robust to the assumption that $\sigma_i^2 \equiv \sigma^2$ than the one way Anova F test.

27) Two important families of contrasts are the family of all possible contrasts and the family of pairwise differences $\theta_{ij} = \mu_i - \mu_j$ where $i \neq j$. The Scheffé multiple comparisons procedure has a δ_F for the family of all possible contrasts while the Tukey multiple comparisons procedure has a δ_F for the family of all $\binom{p}{2}$ pairwise contrasts.

3.4 Complements

Section 3.2 followed Olive (2017a, ch. 5) closely. The one way Anova model assumption that the groups have the same variance is very strong. Chapter 9 shows how to use large sample theory to create better one way MANOVA type tests, and better one way Anova tests are a special case. The tests tend to be better when all of the n_i are large enough for the CLT to hold for each $\overline{Y}_{io}$. Also see Rupasinghe Arachchige Don and Olive (2019).

3.5 Problems

3.1. When $\boldsymbol{X}$ is not full rank, the projection matrix $\boldsymbol{P}_X$ for $C(\boldsymbol{X})$ is $\boldsymbol{P}_X = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ where $\boldsymbol{X}' = \boldsymbol{X}^T$. To show that $C(\boldsymbol{P}_X) = C(\boldsymbol{X})$, you can show that a) $\boldsymbol{P}_X\boldsymbol{w} = \boldsymbol{X}\boldsymbol{y} \in C(\boldsymbol{X})$ where $\boldsymbol{w}$ is an arbitrary conformable constant vector, and b) $\boldsymbol{X}\boldsymbol{y} = \boldsymbol{P}_X\boldsymbol{w} \in C(\boldsymbol{P}_X)$ where $\boldsymbol{y}$ is an arbitrary conformable constant vector.

a) Show $\boldsymbol{P}_X\boldsymbol{w} = \boldsymbol{X}\boldsymbol{y}$ and identify $\boldsymbol{y}$.

b) Show $\boldsymbol{X}\boldsymbol{y} = \boldsymbol{P}_X\boldsymbol{w}$ and identify $\boldsymbol{w}$. Hint: $\boldsymbol{P}_X\boldsymbol{X} = \boldsymbol{X}$.

3.2. Let $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T$ be the projection matrix onto the column space of $\boldsymbol{X}$. Using $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$, show $\boldsymbol{P}$ is idempotent.

3.3. Suppose that $\boldsymbol{X}$ is an $n \times p$ matrix but the rank of $\boldsymbol{X} < p < n$. Then the normal equations $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{Y}$ have infinitely many solutions. Let $\hat{\boldsymbol{\beta}}$ be a solution to the normal equations. So $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{Y}$. Let $\boldsymbol{G} = (\boldsymbol{X}'\boldsymbol{X})^{-}$ be a generalized inverse of $(\boldsymbol{X}'\boldsymbol{X})$. Assume that $E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}$ and $\text{Cov}(\boldsymbol{Y}) = \sigma^2\boldsymbol{I}$. It can be shown that all solutions to the normal equations have the form $\boldsymbol{b_z}$ given below.

a) Show that $\boldsymbol{b_z} = \boldsymbol{G}\boldsymbol{X}'\boldsymbol{Y} + (\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{I})\boldsymbol{z}$ is a solution to the normal equations where the $p \times 1$ vector $\boldsymbol{z}$ is arbitrary.

b) Show that $E(\boldsymbol{b_z}) \neq \boldsymbol{\beta}$.

(Hence some authors suggest that $\boldsymbol{b_z}$ should be called a solution to the normal equations but not an estimator of $\boldsymbol{\beta}$.)

c) Show that $\text{Cov}(\boldsymbol{b_z}) = \sigma^2\boldsymbol{G}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{G}'$.

d) Although $\boldsymbol{G}$ is not unique, the projection matrix $\boldsymbol{P} = \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}'$ onto $\mathcal{C}(\boldsymbol{X})$ is unique. Use this fact to show that $\hat{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{b_z}$ does not depend on $\boldsymbol{G}$ or $\boldsymbol{z}$.

e) There are two ways to show that $\boldsymbol{a}'\boldsymbol{\beta}$ is an estimable function. Either show that there exists a vector $\boldsymbol{c}$ such that $E(\boldsymbol{c}'\boldsymbol{Y}) = \boldsymbol{a}'\boldsymbol{\beta}$, or show that $\boldsymbol{a} \in \mathcal{C}(\boldsymbol{X}')$. Suppose that $\boldsymbol{a} = \boldsymbol{X}'\boldsymbol{w}$ for some fixed vector $\boldsymbol{w}$. Show that $E(\boldsymbol{a}'\boldsymbol{b_z}) = \boldsymbol{a}'\boldsymbol{\beta}$.

(Hence $\boldsymbol{a}'\boldsymbol{\beta}$ is estimable by $\boldsymbol{a}'\boldsymbol{b_z}$ where $\boldsymbol{b_z}$ is any solution of the normal equations.)

f) Suppose that $\boldsymbol{a} = \boldsymbol{X}'\boldsymbol{w}$ for some fixed vector $\boldsymbol{w}$. Show that $Var(\boldsymbol{a}'\boldsymbol{b_Z}) = \sigma^2 \boldsymbol{w}'\boldsymbol{P}\boldsymbol{w}$.

3.4. Let $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ where $E(\boldsymbol{e}) = \boldsymbol{0}$, $\text{Cov}(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_n$, and $\boldsymbol{X}$ has full rank. Let $\boldsymbol{a}$ be a constant vector. (Hint: full rank model formulas are rather simple.)

a) Find $E(\boldsymbol{a}^T\hat{\boldsymbol{\beta}})$.

b) Is $\boldsymbol{a}^T\boldsymbol{\beta}$ estimable? Explain briefly.

3.5. Let $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ where $\boldsymbol{Y} = (Y_1, Y_2, Y_3)'$, $\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 2 & 4 \end{bmatrix}$, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, $E(\boldsymbol{e}) = \boldsymbol{0}$, and $\text{Cov}(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}$.

a) Find $[C(\boldsymbol{X}')]$.

Show whether or not the following functions are estimable.

b) $5\beta_1 + 10\beta_2$

c) β_1

d) $\beta_1 - 2\beta_2$

3.6. Let $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ where $E(\boldsymbol{e}) = \boldsymbol{0}$, $\text{Cov}(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_n$, and $\boldsymbol{X}$ has full rank. Note that $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$. Assume $\boldsymbol{X}$ is a constant matrix.

a) Find $E(Y_i)$.

b) Is $E(Y_i)$ estimable? Explain briefly.

3.7. An overparameterized two way Anova model is $Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_{ij} + e_{ijk}$ for $i = 1, ..., a$ and $j = 1, ..., b$ and $k = 1, ..., m$. Suppose $a = 2$, $b = 2$, and $m = 2$. Then

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix} = \boldsymbol{X} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \tau_{11} \\ \tau_{12} \\ \tau_{21} \\ \tau_{22} \end{bmatrix} + \begin{bmatrix} e_{111} \\ e_{112} \\ e_{121} \\ e_{122} \\ e_{211} \\ e_{212} \\ e_{221} \\ e_{222} \end{bmatrix}.$$

a) Give the matrix $\boldsymbol{X}$.

b) We can write the above model as $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$. This model is **not full rank**. What is the projection matrix $\boldsymbol{P}$ (onto the column space of $\boldsymbol{X}$)? Hint: $\boldsymbol{X}^T\boldsymbol{X}$ is singular, so use the generalized inverse.

3.8. Suppose that $\boldsymbol{Y} = (Y_1, Y_2)'$, $\text{Var}(\boldsymbol{Y}) = \sigma^2 \boldsymbol{I}$, $E(Y_1) = E(Y_2) = \beta_1 - 2\beta_2$. Show whether or not the following functions are estimable. Hint $E(\boldsymbol{Y}) = \boldsymbol{X\beta}$, so find $\boldsymbol{X}$.

a) β_1

b) β_2

c) $-\beta_1 + 2\beta_2$

d) $4\beta_1 - 8\beta_2$

3.9. The cell means model for the two way Anova model is $Y_{ijk} = \mu_{ij} + e_{ijk}$ for $i = 1, ..., a$ and $j = 1, ..., b$ and $k = 1, ..., m$. Suppose $a = 2$, $b = 2$, and $m = 2$. Then

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix} = \boldsymbol{X} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} + \begin{bmatrix} e_{111} \\ e_{112} \\ e_{121} \\ e_{122} \\ e_{211} \\ e_{212} \\ e_{221} \\ e_{222} \end{bmatrix}.$$

a) Give the matrix $\boldsymbol{X}$.

b) Suppose that a full rank cell means two way Anova model is written in matrix form as $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$. What is the vector of residuals $\boldsymbol{r}$?

3.10. Note that $C(\boldsymbol{X'X}) = C(\boldsymbol{X'})$ since $C(\boldsymbol{X'X}) \subseteq C(\boldsymbol{X'})$ and $rank(\boldsymbol{X'X}) = rank(\boldsymbol{X'})$.

Use this result to explain why there is always a solution $\hat{\boldsymbol{\beta}}$ to the normal equations:

$$\boldsymbol{X'X}\hat{\boldsymbol{\beta}} = \boldsymbol{X'Y}.$$

3.11. An alternative parameterization of the one way Anova model is $Y_{ij} = \mu + \alpha_i + e_{ij}$ for $i = 1, ..., p$ and $j = 1, ..., n_i$. Hence $\mu_i = \mu + \alpha_i$. Suppose $p = 3$ and $n_i = 2$. Then

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} = \boldsymbol{X} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{31} \\ e_{32} \end{bmatrix}.$$

Give the matrix $\boldsymbol{X}$.

3.12^Q. Consider the linear regression model $Y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$ or $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ where $\boldsymbol{Y} \sim N_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$. Assume $\boldsymbol{X}$ is $n \times p$ with $rank(\boldsymbol{X}) = r \leq p$.

a) Give expressions for SSE and SSR using matrix notation.

b) Find $E(SSE)$ and $E(SSR)$.

c) Find the distribution of i) SSE, ii) SSR, and iii) MSR/MSE under the assumption $\beta_2 = \cdots = \beta_p = 0$.

3.13^Q. Consider the linear regression model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ where $\boldsymbol{Y} \sim N_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$. Assume $\boldsymbol{X}$ is $n \times p$ with $rank(\boldsymbol{X}) = r \leq p$.

a) i) Define what is meant by an estimable linear function of $\boldsymbol{\beta}$.

ii) Write down the least squares estimator of an estimable function of $\boldsymbol{\beta}$.

iii) Write down an unbiased estimator of σ^2.

b) Show the estimators of part a) ii) and iii) are unbiased.
c) State the Gauss Markov Theorem.
d) Give expressions for SSE and SSR using matrix notation.

3.14^Q. Let $E(\boldsymbol{Y}) = \boldsymbol{X\beta}$ where $\boldsymbol{Y}$ is 3×1, $\boldsymbol{X}$ is 3×2, and $\boldsymbol{\beta}$ is 2×1. Let

$$i)\ \boldsymbol{X} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad \text{ii)}\ \boldsymbol{X} = \begin{bmatrix} 3 & 6 \\ 2 & 4 \\ 1 & 2 \end{bmatrix}.$$

a) In each of cases i) and ii), state whether $\boldsymbol{\beta}$ is estimable andexplain your answer.

b) If the answer is "yes," then determine the matrix $\boldsymbol{B}$ in $\hat{\boldsymbol{\beta}} = \boldsymbol{BY}$.

c) If the answer is "no," then produce one estimable parametric function and its unbiased estimator.