Notes for 751-752
Sections 6-7

Martin Lindquist*

September 13, 2017

*Thanks to Brian Caffo and Ingo Ruczinski for supplying their notes, upon which much of this document is based.

6 Residuals and variability

6.1 Residuals

The residuals represent the variability left unexplained by the projection of $\mathbf{y}$ onto the linear space spanned by the design matrix. The residuals are othogonal to the space spanned by the design matrix, and thus are othogonal to the design matrix itself.

We define the residuals as follows:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}.$$

Thus, the least squares solution can be thought of as minimizing the squared norm of the residuals. Notice further that expanding the column space of $\mathbf{X}$ by including new linearly independent variables, leads to a decrease in the the norm of the residuals. In other words, by adding non-redundant regressors, we necessarily remove residual variability. Furthermore, if $\mathbf{X}$ is $n \times n$ of full rank, then the residuals are all zero, since $\mathbf{y} = \hat{\mathbf{y}}$.

We can re-express the residuals as follows:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}(\mathbf{X}\mathbf{X})^{-1}\mathbf{X}\mathbf{y} = \{\mathbf{I} - \mathbf{X}(\mathbf{X}\mathbf{X})^{-1}\mathbf{X}\}\mathbf{y}.$$

Thus multiplication by the matrix $\mathbf{I} - \mathbf{X}(\mathbf{X}\mathbf{X})^{-1}\mathbf{X}$ transforms the vector $\mathbf{y}$ into its residual. This matrix is interesting for several reasons. First, note that $\{\mathbf{I} - \mathbf{X}(\mathbf{X}\mathbf{X})^{-1}\mathbf{X}\}\mathbf{X} = 0$, thus making the residuals orthogonal to any vector, $\mathbf{X}\boldsymbol{\gamma}$, in the space spanned by the columns of $\mathbf{X}$. Secondly, it is both symmetric and idempotent.

A consequence of the orthogonality is that if an intercept is included in the model, the residuals must sum to 0. Specifically, since the residuals are orthogonal to any column of $\mathbf{X}$, $\mathbf{e}'\mathbf{J}_n = 0$.

6.2 Partitioning variability

In this section we discuss partitioning the variability in the data. Before continuing, it is useful to note that the mean centered version of $\mathbf{y}$, $\mathbf{y} - \mathbf{J}_n\bar{y}$ can be written as:

$$\begin{aligned}
\tilde{\mathbf{y}} &= \mathbf{y} - \mathbf{J}_n\bar{y} \\
&= \mathbf{y} - \mathbf{J}_n(\mathbf{J}_n'\mathbf{J}_n)^{-1}\mathbf{J}_n'\mathbf{y} \\
&= \left\{\mathbf{I} - \mathbf{J}_n(\mathbf{J}_n'\mathbf{J}_n)^{-1}\mathbf{J}_n'\right\}\mathbf{y}.
\end{aligned}$$

In other words, multiplication by the matrix $\left\{\mathbf{I} - \mathbf{J}_n(\mathbf{J}_n'\mathbf{J}_n)^{-1}\mathbf{J}_n'\right\}$ centers vectors. This can be very handy for centering matrices as well. For example, if $\mathbf{X}$ is an $n \times p$ matrix then the matrix $\tilde{\mathbf{X}} = \left\{\mathbf{I} - \mathbf{J}_n(\mathbf{J}_n'\mathbf{J}_n)^{-1}\mathbf{J}_n'\right\}\mathbf{X}$ is the matrix with every column centered.

Continuing with partitioning the variability, for convenience, define $\mathbf{H}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}\mathbf{X})^{-1}\mathbf{X}$ and $\mathbf{H}_{\mathbf{J}} = \mathbf{J}_n(\mathbf{J}_n'\mathbf{J}_n)^{-1}\mathbf{J}_n'$. We define the total sums of squares as

$$\text{SS}_{Tot} = ||\mathbf{y} - \bar{y}\mathbf{J}_n||^2 = \mathbf{y}'(\mathbf{I} - \mathbf{H}_{\mathbf{J}})\mathbf{y}.$$

This is an unscaled measure of the total variability in the sample. Given a design matrix, $\mathbf{X}$, define the residual sums of squares as

$$\text{SS}_{Res} = ||\mathbf{y} - \hat{\mathbf{y}}||^2 = \mathbf{y}'(\mathbf{I} - \mathbf{H}_{\mathbf{X}})\mathbf{y}$$

and the regression sums of squares as

$$\text{SS}_{Reg} = ||\hat{\mathbf{y}} - \mathbf{J}_n\bar{y}||^2 = \mathbf{y}'(\mathbf{H}_{\mathbf{X}} - \mathbf{H}_{\mathbf{J}})\mathbf{y}.$$

The latter equality is obtained by the following. First note that since $(\mathbf{I} - \mathbf{H}_{\mathbf{X}})\mathbf{J}_n = 0$ (since $\mathbf{X}$ contains an intercept) we have that $\mathbf{H}_{\mathbf{X}}\mathbf{J}_n = \mathbf{J}_n$ and then $\mathbf{H}_{\mathbf{X}}\mathbf{H}_{\mathbf{J}_n} = \mathbf{H}_{\mathbf{J}_n}$ and $\mathbf{H}_{\mathbf{J}_n} = \mathbf{H}_{\mathbf{J}_n}\mathbf{H}_{\mathbf{X}}$. Also, note that $\mathbf{H}_{\mathbf{X}}$ is symmetric and idempotent. Now we can perform the following manipulation

$$\begin{aligned}
||\hat{\mathbf{y}} - \mathbf{J}_n\bar{y}||^2 &= \mathbf{y}'(\mathbf{H}_{\mathbf{X}} - \mathbf{H}_{\mathbf{J}})'(\mathbf{H}_{\mathbf{X}} - \mathbf{H}_{\mathbf{J}})\mathbf{y} \\
&= \mathbf{y}'(\mathbf{H}_{\mathbf{X}} - \mathbf{H}_{\mathbf{J}})(\mathbf{H}_{\mathbf{X}} - \mathbf{H}_{\mathbf{J}})\mathbf{y} \\
&= \mathbf{y}'(\mathbf{H}_{\mathbf{X}} - \mathbf{H}_{\mathbf{J}}\mathbf{H}_{\mathbf{X}} - \mathbf{H}_{\mathbf{X}}\mathbf{H}_{\mathbf{J}} + \mathbf{H}_{\mathbf{J}})\mathbf{y} \\
&= \mathbf{y}'(\mathbf{H}_{\mathbf{X}} - \mathbf{H}_{\mathbf{J}})\mathbf{y}.
\end{aligned}$$

Using this identity we can now show that

$$\begin{aligned}
\text{SS}_{Tot} &= \mathbf{y}'(\mathbf{I} - \mathbf{H}_{\mathbf{J}})\mathbf{y} \\
&= \mathbf{y}'(\mathbf{I} - \mathbf{H}_{\mathbf{X}} + \mathbf{H}_{\mathbf{X}} - \mathbf{H}_{\mathbf{J}})\mathbf{y} \\
&= \mathbf{y}'(\mathbf{I} - \mathbf{H}_{\mathbf{X}})\mathbf{y} + \mathbf{y}'(\mathbf{H}_{\mathbf{X}} - \mathbf{H}_{\mathbf{J}})\mathbf{y} \\
&= \text{SS}_{Res} + \text{SS}_{Reg}
\end{aligned}$$

Thus our total sum of squares partitions into the residual and regression sums of squares. We define the coefficient of determination

$$R^2 = \frac{\text{SS}_{Reg}}{\text{SS}_{Tot}}.$$

as the proportion of the total variability explained by our model. Via our equality above, this is guaranteed to be between 0 and 1. Here high values imply that the explanatory variables are useful in explaining the response and low values that the explanatory variables are not useful.

Note that SS_{Tot} only depends on the response variable and not on the model formulation. Hence, it is equal for all regression models. Adding additional explanatory variables to a multiple regression model can only lower SS_{Reg}, and thus lead to an increase in the value of R^2. Since R^2 can be made large by including more (and sometimes unimportant) explanatory variables, it is sometimes modified to adjust for the number of variables included in the model. This allows us to balance model parsimony with explanatory power.

The ratio of the sum of squares to the 'degrees of freedom' (corresponding to the dimensions of the respective subspaces) gives the mean squares:

$$\begin{aligned}\mathrm{MS}_{Tot} &= \frac{\mathrm{SS}_{Tot}}{n-1}\\ \mathrm{MS}_{Res} &= \frac{\mathrm{SS}_{Res}}{n-p}\\ \mathrm{MS}_{Reg} &= \frac{\mathrm{SS}_{Reg}}{p-1}\end{aligned}$$

The adjusted coefficient of multiple determination, uses the mean squares instead of the sums of square, i.e.

$$R_a^2 = 1 - \frac{\mathrm{MS}_{Res}}{\mathrm{MS}_{Tot}} = 1 - \left(\frac{n-1}{n-p}\right)\frac{\mathrm{SS}_{Res}}{\mathrm{SS}_{Tot}}.$$

Since the term includes the number of model parameters, p, it penalizes for model complexity.

6.3 Coding example

```
> summary(lm(y ~ X))

Call:
lm(formula = y ~ X)

Residuals:
     Min       1Q   Median       3Q      Max
-15.2743  -5.2617   0.5032   4.1198  15.3213

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       66.91518   10.70604   6.250 1.91e-07 ***
XAgriculture      -0.17211    0.07030  -2.448  0.01873 *
XExamination      -0.25801    0.25388  -1.016  0.31546
XEducation        -0.87094    0.18303  -4.758 2.43e-05 ***
XCatholic          0.10412    0.03526   2.953  0.00519 **
XInfant.Mortality  1.07705    0.38172   2.822  0.00734 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 7.165 on 41 degrees of freedom
Multiple R-squared:  0.7067,Adjusted R-squared:  0.671
```

```
F-statistic: 19.76 on 5 and 41 DF,  p-value: 5.594e-10


> anova(fit)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
X          5 5072.9 1014.58  19.761 5.594e-10 ***
Residuals 41 2105.0   51.34
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> SSreg = anova(fit)[1,2]
> SSres = anova(fit)[2,2]
> SStot = SSres + SSreg

> 1-SSres/SStot
[1] 0.706735
```

7 Conceptual examples of least squares

In this section we discuss some conceptual examples of least squares. This illustrates the flexibility of this approach and how it allows us to effectively analyze different types of data. First, we revisit some of the models previously used, before moving on to introduce some new ones.

7.1 Mean only regression

First let us revisit mean only regression, which we recall can be expressed as as $\mathbf{y} = \mathbf{J}_n\mu + \epsilon$. Placing this into the GLM framework, our design matrix is $\mathbf{X} = \mathbf{J}_n$. Our coefficient estimate is therefore:

$$\hat{\mu} = (\mathbf{J}_n'\mathbf{J}_n)^{-1}\mathbf{J}_n'\mathbf{y} = \bar{y}.$$

7.2 Regression through the origin

Next, we revisit the regression through the origin problem, i.e., $\mathbf{y} = \mathbf{x}\beta$. Here the design matrix is $\mathbf{X} = \mathbf{x}$. Our coefficient estimate is therefore:

$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} = \frac{\langle \mathbf{y}, \mathbf{x} \rangle}{||\mathbf{x}||^2}.$$

7.3 Linear regression

In the case of simple linear regression $\mathbf{y} = \mathbf{J}_n\beta_0 + \mathbf{x}\beta_1$, the design matrix is $\mathbf{X} = [\mathbf{J}_n \ \mathbf{x}]$. Now, the estimate of $\beta = [\beta_0 \ \beta_1]'$ can be obtained through the equations

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Here the term $\mathbf{X}'\mathbf{X}$ is a 2×2 matrix and easily invertible. It is thus easy (though somewhat tedious) to show that this solution corresponds to the one we have previously obtained. We leave this as an exercise to the reader.

7.4 ANOVA

Analysis of Variance (ANOVA) is a technique for comparing the means across multiple groups. For example, we may be interested in determining whether the cholesterol levels (y) differ between subjects in a drug group and a control group. There are multiple ways to formulate this model. Here we use the following:

$$y_{ij} = \alpha_i + \epsilon_{ij} \qquad \text{for } \ j = 1, \ldots n_i, \ \ i = 1, 2$$

or

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \end{pmatrix}$$

Note here the first column codes whether an observation belongs to the first group, and the second column whether it belongs to the second group. These types of indicator variables, or 'dummy variables' are often used to denote the values of a categorical variable.

To estimate α_1 and α_2, note the following results:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}$$

and

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} n_1\bar{\mathbf{y}}_1 & 0 \\ 0 & n_2\bar{\mathbf{y}}_2 \end{pmatrix}.$$

Now the solution can be obtained by computing $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, which gives us

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{y}}_1 \\ \bar{\mathbf{y}}_2 \end{pmatrix}$$

We can generalize this to more than two treatment groups as follows. Assume J groups, each with K observations. Let $\mathbf{y} = [y_{11}, \dots y_{JK}]$ and our design matrix look like

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \dots & \dots & \vdots \\ 0 & 0 & \dots & 1 \\ \vdots & \dots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \mathbf{I}_K \otimes \mathbf{J}_n,$$

where $\otimes$ is the Kronecker product. That is, our $\mathbf{y}$ arises out of J groups where there is K measurements per group. Let $\bar{y}_j$ be the mean of the $\mathbf{y}$ measurements in group j. Then

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} K\bar{y}_1 \\ \vdots \\ K\bar{y}_J \end{bmatrix},$$

Note also that $\mathbf{X}'\mathbf{X} = K\mathbf{I}$. Therefore, $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\bar{y}_1 \ldots, \bar{y}_J)'$. Thus, if our design matrix parcels $\mathbf{y}$ into groups, the coefficients are the group means.

7.5 Computing

The data set `PlantGrowth` in `R` contains results from an experiment to compare yields (as measured by dried weight of plants) obtained under a control and two different treatment conditions. The data consists of 30 observations on 2 variables weight and group. The levels of group are 'ctrl', 'trt1', and 'trt2'.

```
> fit = lm(weight ~ group -1, data = PlantGrowth)
> summary(fit)

Call:
lm(formula = weight ~ group - 1, data = PlantGrowth)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0710 -0.4180 -0.0060  0.2627  1.3690

Coefficients:
          Estimate Std. Error t value Pr(>|t|)
groupctrl   5.0320     0.1971   25.53   <2e-16 ***
grouptrt1   4.6610     0.1971   23.64   <2e-16 ***
grouptrt2   5.5260     0.1971   28.03   <2e-16 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 0.6234 on 27 degrees of freedom
Multiple R-squared:  0.9867,Adjusted R-squared:  0.9852
F-statistic: 665.5 on 3 and 27 DF,  p-value: < 2.2e-16
```

7.6 ANCOVA

Next, we consider analysis of covariance. This allows us to compare differences in means between two or more groups while taking into account the variability of other variables, called covariates. Suppose we have data on two variables $\mathbf{x}$ and $\mathbf{y}$ collected for two separate groups (A and B). Let $\mathbf{x} = (\mathbf{x}_1 \ \mathbf{x}_2)'$, where $\mathbf{x}_1$ are the observations associated with group A and $\mathbf{x}_2$ those associated with group B.

We can write this as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & x_{11} \\ 1 & 0 & x_{12} \\ \vdots & \dots & \vdots \\ 1 & 0 & x_{1n} \\ 0 & 1 & x_{21} \\ 0 & 1 & x_{22} \\ \vdots & \dots & \dots \\ 0 & 1 & x_{2n} \end{bmatrix} = [\mathbf{I}_2 \otimes \mathbf{1}_n \ \ \mathbf{x}].$$

In this setting we seek to project $\mathbf{y}$ onto the space spanned by two groups and a regression variable. This is effectively equivalent to fitting two parallel lines to the data. Let $\boldsymbol{\beta} = (\mu_1 \ \mu_2 \ \beta)' = (\boldsymbol{\mu}' \ \beta)'$. Denote the outcome vector, $\mathbf{y}$, as comprised of y_{ij} for $i = 1, 2$ and $j = 1, \dots, n$ stacked in the relevant order. Now, imagine holding β fixed. We now want to solve:

$$||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2 = ||\mathbf{y} - \mathbf{x}\beta - (\mathbf{I}_2 \otimes \mathbf{1}_n)\boldsymbol{\mu}||^2 \tag{1}$$

Then we are in the case of the previous section and the best estimate of $\boldsymbol{\mu}$ are the group means $\frac{1}{n}(\mathbf{I}_2 \otimes \mathbf{J}_n)'(y - \mathbf{x}\boldsymbol{\beta}) = (\bar{y}_1 \ \bar{y}_2)' - (\bar{x}_1 \ \bar{x}_2)'\beta$ where $\bar{y}_i$ and $\bar{x}_i$ are the group means of $\mathbf{y}$ and $\mathbf{x}$ respectively. Then we have that (1) satisfies:

$$\begin{aligned} (1) \ \ &\geq \ \ ||\mathbf{y} - \mathbf{x}\beta - (\mathbf{I}_2 \otimes \mathbf{1}_n)\{(\bar{y}_1 \ \bar{y}_2)' - (\bar{x}_1 \ \bar{x}_2)'\beta\}||^2 \\ &= \ \ ||\tilde{\mathbf{y}} - \tilde{\mathbf{x}}\beta||^2 \end{aligned}$$

where $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{x}}$ are the group centered versions of $\mathbf{y}$ and $\mathbf{x}$. (That is $\tilde{y}_{ij} = y_{ij} - \bar{y}_i$, for example.) This is equivalent to the regression through the origin problem yielding the solution

$$\hat{\beta} = \frac{\sum_{ij}(y_{ij} - \bar{y}_i)(x_{ij} - \bar{x}_i)}{\sum_{ij}(x_{ij} - \bar{x}_i)^2} = p\hat{\beta}_1 + (1 - p)\hat{\beta}_2$$

where

$$p = \frac{\sum_j (x_{1j} - \bar{x}_1)^2}{\sum_{ij}(x_{ij} - \bar{x}_i)^2}$$

and

$$\hat{\beta}_i = \frac{\sum_j (y_{ij} - \bar{y}_i)(x_{ij} - \bar{x}_i)}{\sum_j (x_{ij} - \bar{x}_i)^2}.$$

This implies that the estimated slope is a convex combination of the group-specific slopes weighted by the variability in the $\mathbf{x}$'s within the group. Furthermore, $\hat{\mu}_i = \bar{y}_i - \bar{x}_i\hat{\beta}$ and thus

$$\hat{\mu}_1 - \hat{\mu}_2 = (\bar{y}_1 - \bar{y}_2) - (\bar{x}_1 - \bar{x}_2)\hat{\beta}.$$

7.7 Computing

We illustrate simple linear regression using the `mtcars` data set that is directly available in R. The data was extracted from the 1974 Motor Trend US magazine, and consists of gas consumption (`mpg`) and 10 other aspects of automobile design and performance for a total of 32 cars. Here we focus on how mileage depends upon transmission type (`am` / 0 = automatic, 1 = manual), controlling for the weight of the car (`wt`).

First fit ANOVA, ignoring weight.

```
> fit = lm(mpg~factor(am) - 1,data = mtcars)
> summary(fit)

Call:
lm(formula = mpg ~ factor(am) - 1, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-9.3923 -3.0923 -0.2974  3.2439  9.5077

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
factor(am)0   17.147      1.125   15.25 1.13e-15 ***
factor(am)1   24.392      1.360   17.94  < 2e-16 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 4.902 on 30 degrees of freedom
Multiple R-squared:  0.9487,Adjusted R-squared:  0.9452
F-statistic: 277.2 on 2 and 30 DF,  p-value: < 2.2e-16
```

Now fit ANCOVA, controlling for weight.

```
> fit = lm(mpg~factor(am) +wt - 1,data = mtcars)
> summary(fit)

Call:
lm(formula = mpg ~ factor(am) + wt - 1, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5295 -2.3619 -0.1317  1.4025  6.8782

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
factor(am)0  37.3216     3.0546  12.218 5.84e-13 ***
factor(am)1  37.2979     2.0857  17.883  < 2e-16 ***
wt           -5.3528     0.7882  -6.791 1.87e-07 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 3.098 on 29 degrees of freedom
Multiple R-squared:  0.9802,Adjusted R-squared:  0.9781
F-statistic: 478.1 on 3 and 29 DF,  p-value: < 2.2e-16
```

7.8 Including interactions

In the ANCOVA model we used an indicator variable to model differences in the intercept between groups. Sometimes we also want the slopes of the regression model to differ between groups. This can be done by including an interaction term together with the indicator variable in the model.

Suppose we have data on two variables $\mathbf{z}$ and $\mathbf{y}$ collected for two groups (A and B). Let $\mathbf{x}_1$ be equal to 1 if the observation belongs to group A and 0 if it belongs to group B. Similarly, let $\mathbf{x}_2$ be equal to 1 if the observation belongs to group B and 0 if it belongs to group A. Let $\mathbf{z} = (\mathbf{z}_1\ \mathbf{z}_2)'$, where $\mathbf{z}_1$ are the observations associated with group A and $\mathbf{z}_2$ those associated with group B. Consider the following model with interactions:

$$\mathbf{y} = \mathbf{J}_n\mu_1 + \mathbf{x}_2\mu_2 + \mathbf{z}\beta_1 + \mathbf{z} * \mathbf{x}_2\beta_2 + \epsilon$$

We can fit this model using the following design matrix:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & z_{11} & 0 \\ 1 & 0 & z_{12} & 0 \\ \vdots & \dots & \vdots & \dots \\ 1 & 0 & z_{1n} & 0 \\ 1 & 1 & z_{21} & z_{21} \\ 1 & 1 & z_{22} & z_{22} \\ \vdots & \dots & \dots & \dots \\ 1 & 1 & z_{2n} & z_{2n} \end{bmatrix}.$$

The model allows both the slopes and intercepts to vary between groups. It can be fit in the same manner as described above.

7.9 Computing

```
> fit = lm(mpg ~ factor(am) + wt + factor(am)*wt, data = mtcars)
> summary(fit)

Call:
lm(formula = mpg ~ factor(am) + wt + factor(am) * wt, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6004 -1.5446 -0.5325  0.9012  6.0909

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     31.4161     3.0201  10.402 4.00e-11 ***
factor(am)1     14.8784     4.2640   3.489  0.00162 **
wt              -3.7859     0.7856  -4.819 4.55e-05 ***
factor(am)1:wt  -5.2984     1.4447  -3.667  0.00102 **
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 2.591 on 28 degrees of freedom
Multiple R-squared:  0.833,Adjusted R-squared:  0.8151
F-statistic: 46.57 on 3 and 28 DF,  p-value: 5.209e-11
```