Chapter 4: Multiple Linear Regression

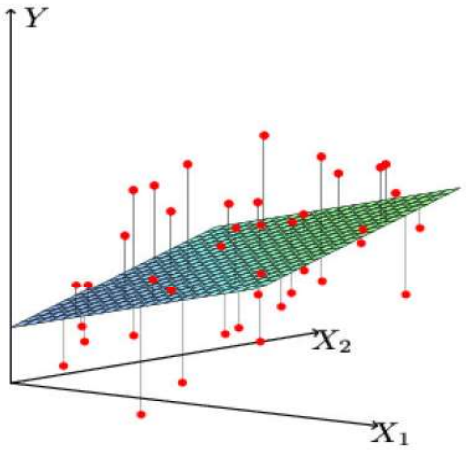
CZ

Fall, 2014

An Example

Suppose that a researcher is studying factors that might affect blood pressures for women aged 45 to 65 years old. The y-variable is blood pressure. Suppose that two predictor variables (x-variables) of interest are age and body mass index (BMI, calculated as $weight/height^2$). The general structure of a linear multiple regression model for this situation would be

$$\text{Blood Pressure} = \beta_0 + \beta_1 \cdot \text{age} + \beta_2 \cdot \text{BMI} + \text{error}$$



Notation for Multiple Regression Model

Model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i$$

or $$E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}.$$

- β_0: (intercept) The mean y when all x variables are 0.
- β_j, $j = 1, \ldots, p$: The average amount of change in y when x_j increases by one unit and other x-variables remain the same.
- $\beta_j = 0$ for $j \geq 1$ means ?

Note: x_{ij}

- i: the ith individual or unit
- j: the jth predictor for subject i
- p: number of predictors in the model

Matrix Notation for Multiple Regression Model

Consider the multiple regression models for all the individuals in the sample (size n), with intercept and p predictor variables:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + ... + \beta_p x_{1p} + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + ... + \beta_p x_{2p} + \epsilon_2 \\ &... \\ y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon_i \\ &... \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + ... + \beta_p x_{np} + \epsilon_n \end{aligned}$$

Matrix Notation for Multiple Regression Model

Regression model: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ or $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, where

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ .. \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & ... & x_{1p} \\ 1 & x_{21} & x_{22} & ... & x_{2p} \\ 1 & x_{31} & x_{32} & ... & x_{3p} \\ . & .. & .. & . & ... \\ 1 & x_{n1} & x_{n2} & ... & x_{np} \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ .. \\ \beta_p \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ . \\ \epsilon_n \end{pmatrix}$$

Matrix Notation for Multiple Regression Model

$\mathbf{X}$ is called design matrix. Each row of $\mathbf{X}$ corresponds to one subject/unit/observation, each column (except the 1st) corresponds to a predictor. The 1st column is the intercept. And, this form is the same as the way we input data in R.

Matrix Notation for Multiple Regression Model: Example

Consider a toy example with a data set of size 4, we have response Y and two predictors X_1, X_2 and consider fitting a multiple regression model. Based on the data given in the following table, how can we represent the models using matrix notations?

Y	12	17	15	11
X_1	3	5	4	2
X_2	1	1	2	2

In other words, for the matrix notation representation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, what are the exact forms of the design matrix and vectors based on the data?

Assumptions Regarding the Error Terms

(Recall the assumptions in simple linear regression)

i ϵ is a random vector with expectation:

$$E(\boldsymbol{\epsilon}) = 0$$

ii the variance-covariance matrix:

$$\text{Var}(\boldsymbol{\epsilon}) = \begin{pmatrix} \sigma^2 & 0 & ... & 0 \\ 0 & \sigma^2 & ... & 0 \\ .. & .. & . & .. \\ 0 & 0 & ... & \sigma^2 \end{pmatrix}.$$

i.e.

$$\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$$

where $\mathbf{I}$ is $n \times n$ identity matrix.

Normality Assumption

iii We need the normality assumption for inference (CI, PI, t-test, F-test)

$$\boldsymbol{\epsilon} \sim MVN(0, \sigma^2 \mathbf{I})$$

MVN: multivariate normal distribution.

Note: if $\boldsymbol{\epsilon} \sim MVN(0, \sigma^2 \mathbf{I})$, $\epsilon_i \sim N(0, \sigma^2)$. The reverse is not necessarily true.

Estimation

- Least squares criterion:

$$\min \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - ... - \beta_p x_{ip})^2$$

- In matrix form:

$$\min (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathbf{T}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

- Solution:

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ .. \\ .. \\ \hat{\beta}_p \end{pmatrix} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}.$$

Require $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ be full rank.

Proof

Estimation

$$\widehat{\beta} = (\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{X}^\mathbf{T}\mathbf{Y}.$$

- Property 1: $E(\widehat{\beta}) = \beta$
- Property 2: $Var(\widehat{\beta}) = \sigma^2(\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}$
- Property 3: Under normality assumption about random errors, we have:

$$\widehat{\beta} \sim MVN(\beta, \sigma^2(\mathbf{X}^\mathbf{T}\mathbf{X})^{-1})$$

Proof

Sampling Properties about $\widehat{\mathbf{Y}}$

Fitted value:
$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + ... + \hat{\beta}_p x_{ip},\ i = 1, \ldots, n$; or $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\beta}$.

$$\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\beta} = \mathbf{X}(\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{X}^\mathbf{T}\mathbf{Y} := \mathbf{H}\mathbf{Y}$$

H is called the hat matrix. **H** is idempotent and symmetric.

- $E(\widehat{\mathbf{Y}}) = E(\mathbf{Y})$
- $\text{Var}(\widehat{\mathbf{Y}}) = \sigma^2\mathbf{H}$
- Under normality assumption, $\widehat{\mathbf{Y}} \sim MVN(\mathbf{E}(\mathbf{Y}), \sigma^2\mathbf{H})$

Proof

Residuals

Residual:
$r_i = y_i - \hat{y}_i$, $i = 1, \ldots, n$; or $\mathbf{r} = \mathbf{Y} - \widehat{\mathbf{Y}}$ where $\mathbf{r} = (r_1, \ldots, r_n)^T$.

$$\mathbf{r} = \mathbf{Y} - \widehat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}(\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{X}^\mathbf{T}\mathbf{Y} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

The matrix $\mathbf{I} - \mathbf{H}$ is idempotent and symmetric. Properties:

- $\mathbf{X}^\mathbf{T}\mathbf{r} = 0$
- $\hat{\mathbf{Y}}^T\mathbf{r} = 0$
- $E(\mathbf{r}) = 0$
- $\text{Var}(\mathbf{r}) = \sigma^2(\mathbf{I} - \mathbf{X}(\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{X}^\mathbf{T}) = \sigma^2(\mathbf{I} - \mathbf{H})$
- Under normality assumption, $\mathbf{r} \sim MVN(0, \sigma^2(\mathbf{I} - \mathbf{H}))$

Proof

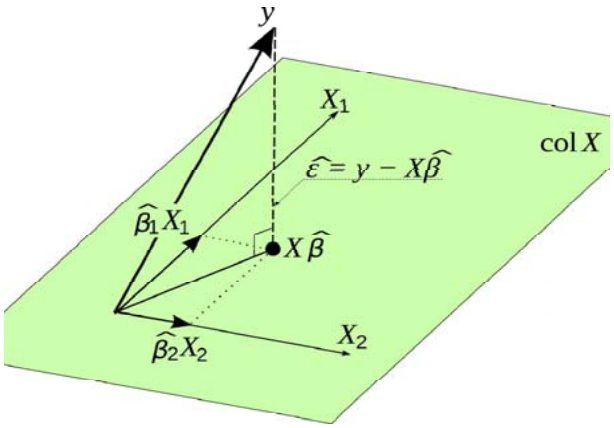
An Estimator of σ^2

An unbiased estimator of σ^2:

$$\hat{\sigma}^2 = \frac{1}{n-p-1}\sum_{i=1}^{n} r_i^2$$

Proof:

Another way to look at regression problems



Inference: Testing Significance of Each β Coefficient

To assess whether a particular x-variable is making an important contribution to the model. That is,

- given the presence of the other x-variables in the model, does a particular x-variable help us to explain more about the y-variable?

What does given the presence of the other x-variables in the model mean?

For example, suppose that we have 3 variables in the model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

To determine whether x_1 is a useful predictor variable in this model, we could test

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned}$$

If the null hypothesis is true, y and x_1 are not significantly related, or x_1 is not important when x_2 and x_3 are in the model.

Sampling Properties of $\hat{\beta}$ and $\hat{\sigma}^2$

- $\hat{\boldsymbol{\beta}} \sim MVN(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\mathbf{T}\mathbf{X})^{-1})$
- $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent
- $(n-p-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-p-1)$
- For $j = 1, \ldots, p$, we have:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 v_{jj}}} \sim t_{n-p-1}$$

where v_{jj} is the (j,j)th element in matrix $(\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}$.

Proof

Inference: Testing Significance of Each β Coefficient

$$H_0 : \beta_j = 0 \text{ vs. } H_a : \beta_j \neq 0$$

Carry out the test:

$$T = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{n-p-1} \text{ under } H_0.$$

where $se(\hat{\beta}_j) = \hat{\sigma}\sqrt{v_{jj}}$.

1. **critical value approach:** if $|T_0| > t_{n-p-1,\alpha/2}$, reject H_0.
2. **p-value approach:** if p-value $= P(|T| > |T_0|) < \alpha$, reject H_0.

where T_0 is the observed value of T using the given sample.

Confidence Interval for $E(y)$

Suppose that a researcher is studying factors that might affect blood pressures for women aged 45 to 65 years old. The y-variable is blood pressure. Two predictor variables (x-variables) of interest are age and body mass index. The multiple regression model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon.$$

Now we are interested in constructing a 95% confidence interval for $E(y)$ when a woman is 65 year-old with a BMI value of 25:

$$E(y) = \beta_0 + \beta_1 \times 65 + \beta_2 \times 25 = \mathbf{a}^{\mathbf{T}}\beta$$

where $\mathbf{a} = (1, 65, 25)^T$.

Confidence Interval for $E(y)$

$$\hat{y} = \mathbf{a}^T\widehat{\beta}$$

Properties:

- $E(\hat{y}) = E(y)$ where $E(y) = \mathbf{a}^T\beta$
- $Var(\hat{y}) = \sigma^2(\mathbf{a}^{\mathbf{T}}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{a})$
- Under normality assumption, $\hat{y} \sim N(E(y), \sigma^2(\mathbf{a}^{\mathbf{T}}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{a}))$
- $100(1-\alpha)\%$ confidence interval for $E(y)$ at a given vector of values for predictors, $\mathbf{a}$, is

$$\hat{y} \pm t_{n-p-1,\alpha/2}\sqrt{\hat{\sigma}^2(\mathbf{a}^T(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{a})}$$

where $\hat{\sigma}^2$ is the unbiased estimator of σ^2 and $t_{n-p-1,\alpha/2}$ is the upper $100(\alpha/2)$th percentile for t_{n-p-1} .

Prediction Interval for new y

Now, we are interested in predicting the blood pressure for a new case with a vector of predictor values, $\mathbf{a}_p$ (e.g., $\mathbf{a}_p = (1, 65, 25)^T$)

$$y_p = \mathbf{a}_p^T\beta + \epsilon_p, \quad \hat{y}_p = \mathbf{a}_p^T\widehat{\beta}$$

Properties of $y_p - \hat{y}_p$:

- $E(y_p - \hat{y}_p) = 0$
- $Var(y_p - \hat{y}_p) = \sigma^2(1 + \mathbf{a}_p^T(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{a}_p)$
- Under normality assumption, $y_p - \hat{y}_p \sim N(0, \sigma^2(1 + \mathbf{a}_p^T(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{a}_p))$
- $100(1-\alpha)\%$ prediction interval for y_p at a given vector of values for predictors, $\mathbf{a}_p$, is

$$\hat{y}_p \pm t_{n-p-1,\alpha/2}\sqrt{\hat{\sigma}^2(1 + \mathbf{a}_p^T(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{a}_p)}$$

Analysis of Variance (ANOVA) Table

As in simple regression, the **analysis of variance (ANOVA) table** for a multiple regression model displays quantities that measure how much of the variability in the y-variable is explained and how much is not explained by x-variables.
Reminder: An underlying conceptual idea for the construction of the analysis of variance table is

$$SST = SSR + SSE$$

ANOVA Table

Souce	DF	SS	MS	F
Regression	p	$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	MSR=SSR/p	MSR/MSE
Error	$n-p-1$	$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	MSE=SSE/(n-p-1)	
Total	$n-1$	$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$		

The computation of the table is identical with the simple regression model.

Uses of ANOVA Table: F test

- The F statistic in the analysis of variance can be used to test whether the y-variable is related to at least one x-variables in the model. Specifically,

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = ... = \beta_p = 0 \\ H_a &: \text{at least one of the } \beta_i \neq 0, \text{ for } j = 1, ..., p. \end{aligned}$$

1. The null hypothesis means that the **y-variable is not related to any of the x-variables in the model**.
   The alternative hypothesis means that the **y-variable is related to one or more of the x-variables in the model**.
2. $F = MSR/MSE \sim F(p, n-p-1)$ under H_0.
   If $F_0 > F_\alpha(p, n-p-1)$, reject H_0.
3. p-value$= P(F > F_0)$. Usually, if $p - value < 0.05$, reject H_0, and conclude that y is related to at least one of the x-variables in the model.

Uses of ANOVA Table: MSE and R^2

- MSE is the unbiased estimator of the error variance σ^2.
- Coefficient of Determination:

$$R^2 = \frac{SSR}{SST}.$$

Still represents the proportion of the total variation in y that is explained by the multiple regression model.

Practice

1. Complete the following ANOVA table for a multiple linear regression model if the number of x-variables is 3, and the number of observations in the sample is 35.

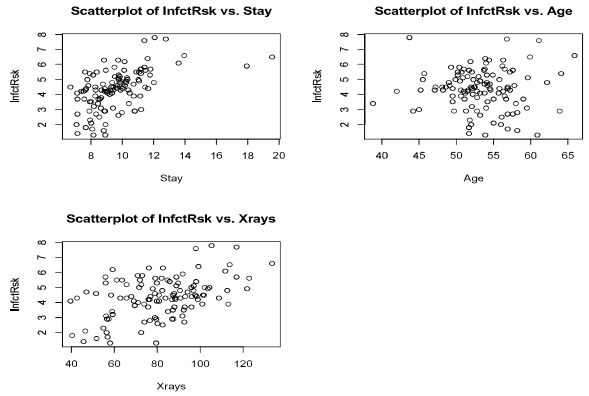
Source	DF	SS	MS	F
Regression		322.1		
Error				NA
Total		547.6	NA	NA

Example

Data from n = 113 hospitals in the United States are used to assess factors related to the likelihood that a hospital patients acquires an infection while hospitalized. The variables here are y =infection risk, x_1 =average length of patient stay, x_2 =average patient age, x_3 =measure of how many x-rays are given in the hospital.
Note: sample size $n = 113$, number of predictors= 3.

Example

Step 1. Check individual scatter plots of y versus x_i, $i = 1, 2, 3$.



Example

Step 2. Now, simply include all of the three predictor variables into the model and fit the multiple regression model:

```
> fm=lm(InfctRsk~Stay+Age+Xrays,data=senic)
> summary(fm)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.001162   1.314724   0.761 0.448003
Stay         0.308181   0.059396   5.189 9.88e-07 ***
Age         -0.023005   0.023516  -0.978 0.330098
Xrays        0.019661   0.005759   3.414 0.000899 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1

Residual standard error: 1.085 on 109 degrees of
freedom
Multiple R-squared: 0.363,   Adjusted R-squared: 0.3455
F-statistic:  20.7 on 3 and 109 DF,  p-value: 1.087e-10
```

Example

Step 3. Inference. The hypothesis testing for:

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \beta_3 = 0 \\ H_a &: \text{at least one of the } \beta_j \neq 0, \text{ for } j = 1, 2, 3. \end{aligned}$$

- $F_0 = 20.7 \Rightarrow F_0 > F_{0.05}(3, 109) = 2.69$, so reject H_0.
- p-value$= P(F \geq 20.7) \approx 0$, so reject H_0.

Example

Step3. Inference. The hypothesis testing for β_1:

$$\begin{aligned} H_0 : \beta_1 &= 0 \\ H_a : \beta_1 &\neq 0 \end{aligned}$$

- $T_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = 5.19 \Rightarrow |T_0| > t_{113-4,0.025} = 1.98$, so reject H_0.
- p-value$= P(|T| \geq 5.19) \approx 0$, so reject H_0.

Example

Step 3. Inference. The hypothesis testing for β_2:

$$\begin{aligned} H_0 : \beta_2 &= 0 \\ H_a : \beta_2 &\neq 0 \end{aligned}$$

- $T_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = -0.98 \Rightarrow |T_0| < t_{113-4,0.025} = 1.98$, so fail to reject H_0.
- p-value$= P(|T| \geq 0.98) \approx 0.33$, so fail to reject H_0.

Thus we cannot reject the null hypothesis $H_0 : \beta_2 = 0$, so we cannot conclude that "Age" is a useful predictor **within this model** (i.e. given the presence of the other two predictors).

Example

Step 3. Inference. The hypothesis testing for β_3:

$$\begin{aligned} H_0 : \beta_3 &= 0 \\ H_1 : \beta_3 &\neq 0 \end{aligned}$$

In a similar fashion, we can also obtain the p-value$= 0.001 < 0.05$. Therefore, X-rays is useful for predicting y with the other two variables in the model.

Example

Step 4. Remove $x_2 = Age$ from the model and refit the model.

```
> fm=lm(InfctRsk~Stay+Xrays,data=senic)
> summary(fm)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.150603   0.585036  -0.257 0.797331
Stay         0.295845   0.058030   5.098 1.44e-06 ***
Xrays        0.020227   0.005728   3.531 0.000606 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1

Residual standard error: 1.085 on 110 degrees of
freedom
Multiple R-squared: 0.3574,  Adjusted R-squared: 0.3457
F-statistic: 30.59 on 2 and 110 DF,  p-value: 2.734e-11
```


Example

Step 5. Interpret the results.

- The fitted regression equation is
  Infection Risk $= -0.151 + 0.296 \times$ Stay $+ 0.020 \times$ X-rays
- Interpret $\hat{\beta}_1$: When the number of X-rays is held constant, the estimated infection risk increases by 0.296% with one day increase in the average length of stay.
- The value of $R^2 = 35.7\%$ means that the model (the two x-variables) explains 35.7% of the observed variation in infection risk.
- The value $\hat{\sigma} = 1.085$ is the estimated standard deviation of the errors.

Testing Linear Constraints

- Suppose we have a model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

- Want to test hypothesis

$$H_0 : \beta_1 = 2\beta_2, \beta_3 = 0$$

- Write this as $\mathbf{A}\beta = 0$

$$\begin{pmatrix} 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Testing Linear Constraints

- Testing l linear constraints:

$$H_0 : \mathbf{A}\beta = 0, H_a : \mathbf{A}\beta \neq 0$$

  where $\mathbf{A}$ is an $l \times (p+1)$ matrix of rank l.
- In the previous example, full model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

  Restricted model:

$$y = \beta_0 + \beta_2(2x_1 + x_2) + \epsilon$$

Practice

Write the following hypothesis testing problems in the form of linear constraints $\mathbf{A}\beta = 0$ (assume there are three predictors in the study).

- $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ vs. H_a: at least one of $\beta_i, i = 1, 2, 3$, is not zero.
- $H_0 : \beta_1 = \beta_2 = \beta_3$ vs. H_a: at least two of $\beta_i, i = 1, 2, 3$, are not equal.
- $H_0 : \beta_1 = 0, \beta_2 = \frac{\beta_0+\beta_3}{2}$ vs. $H_a : \beta_1 \neq 0 \quad \text{or} \quad \beta_2 \neq \frac{\beta_0+\beta_3}{2}$

Testing linear constraints

Use additional sums of squares principle:

- Let
$$C(\mathbf{X}) = \{\beta_0 1 + \beta_1 x_1 + \cdots + \beta_p x_p\}$$
be the column space spanned by $\mathbf{X}$
- Let
$$C_A(\mathbf{X}) = \{\beta_0 1 + \beta_1 x_1 + \cdots + \beta_p x_p | \mathbf{A}\beta = 0\}$$
be the sub column space spanned by $\mathbf{X}$ which satisfy constraints
- Let $\widehat{\mathbf{Y}}$ be the orthogonal projection of $\mathbf{Y}$ onto $C(\mathbf{X})$ and $\widehat{\mathbf{Y}}_A$ be the orthogonal projection of $\mathbf{Y}$ onto $C_A(\mathbf{X})$
- If H_0 is true, we would expect $\widehat{\mathbf{Y}} - \widehat{\mathbf{Y}}_A$ to be small.
- Use $\|\widehat{\mathbf{Y}} - \widehat{\mathbf{Y}}_A\|^2 = (\widehat{\mathbf{Y}} - \widehat{\mathbf{Y}}_A)^T(\widehat{\mathbf{Y}} - \widehat{\mathbf{Y}}_A)$ to define test statistic. This is called additional sum of squares.

Testing linear constraints

Theorem: If $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ $\epsilon \sim N(0, \sigma^2 I)$. If hypothesis $\mathbf{A}\beta = 0$ holds (where $\mathbf{A}$ is $l \times (p+1)$ matrix of rank l), we have

1. $\frac{\mathbf{Y}-\mathbf{Y}_A}{\sigma^2}^2 \sim \chi^2_l$
2. $\|\widehat{\mathbf{Y}} - \widehat{\mathbf{Y}}_A\|^2$ is independent of $\hat{\sigma}^2$
3. Using above,
$$F = \frac{\|\widehat{\mathbf{Y}} - \widehat{\mathbf{Y}}_A\|^2/l}{\hat{\sigma}^2} \sim F(l, n-p-1)$$

Make conclusions about testing $H_0 : \mathbf{A}\beta = 0$:

- **critical value approach:** if $F_0 > F_\alpha(l, n-p-1)$, reject H_0.
- **p-value approach:** if p-value $= P(F > F_0) < \alpha$, reject H_0.

Connection to ANOVA

- Consider the full model
$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$$
- The restricted model under $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$ is
$$y = \beta_0 + \epsilon$$
- What is the LSE of β_0? What is $\hat{\mathbf{Y}}_A$?

Connection to ANOVA

- The additional sum of squares $\|\widehat{\mathbf{Y}} - \widehat{\mathbf{Y}}_A\|^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$, which is SSR in the ANOVA table.
- The F-test is:

$$F = \frac{\|\widehat{\mathbf{Y}} - \widehat{\mathbf{Y}}_A\|^2/l}{\hat{\sigma}^2} = \frac{SSR/p}{MSE} = MSR/MSE \sim F(p, n-p-1)$$

Under H_0.

General Linear F-test

Assume we obtain two ANOVA tables for the full model and restricted model, respectively.

- $\|\widehat{\mathbf{Y}} - \widehat{\mathbf{Y}}_A\|^2 = SSE(Restricted) - SSE(full) = SSR(full) - SSR(Restricted)$
- The "general linear" F-statistic for testing $H_0 : \mathbf{A}\beta = 0$ is computed by

$$F = \frac{SSR(Full) - SSR(Restricted)}{(\text{regression df}(Full) - \text{regression df}(Restricted))} / MSE(Full)$$

Under H_0, $F \sim F(df_1, df_2)$.

df_1: regression df(Full) - regression df(Restricted)
df_2: error df(Full)
critical value approach: if $F_0 > F_\alpha(df_1, df_2)$, reject H_0.
p-value approach: if p-value $= P(F > F_0) < \alpha$, reject H_0.

Example

Dataset = variables possibly relating to blood pressures of people who have moved from rural high altitude areas to urban lower altitude areas.

Y = Blood pressure
X_1 = Age
X_2 = Years in urban area
X_3 = X_1/X_2 = fraction of life in urban area
X_4 = weight (kg)
X_5 = height (m)
X_6 = Chin skinfold
X_7 = Forearm skinfold
X_8 = calf skinfold
X_9 = resting pulse rate

Multiple regression results, model has all 9 x-variables

Regression Analysis: Systol_BP versus Age, Years, ...

The regression equation is

Systol_BP = 147 - 1.12 Age + 2.46 Years - 115 fraclife + 1.41 Weight
- 0.0346 Height - 0.944 Chin - 1.17 Forearm - 0.159 Calf
+ 0.115 Pulse

Predictor	Coef	SE Coef	T	P
Constant	146.82	48.97	3.00	0.006
Age	-1.1214	0.3274	-3.43	0.002
Years	2.4554	0.8146	3.01	0.005
fraclife	-115.29	30.17	-3.82	0.001
Weight	1.4139	0.4310	3.28	0.003
Height	-0.03464	0.03686	-0.94	0.355
Chin	-0.9437	0.7410	-1.27	0.213
Forearm	-1.171	1.193	-0.98	0.335
Calf	-0.1587	0.5372	-0.30	0.770
Pulse	0.1146	0.1704	0.67	0.507

Step1: Set up H_0 of interest

According to the individual T tests, the last five predictors are not significant. Therefore, we are interested in whether we could discard those five predictors all together in order to simplify the model. In other words, we want to test

$$H_0 : \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$$

- If this null is not rejected, we don't have significant evidence that any of the variables X_5 to X_9 contribute to the prediction/explanation of y, so we can discard all of them.

Step2: fit both full model and restricted model

Full Model
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	4358.85	484.32	6.46	0.000
Residual Error	29	2172.59	74.92		
Total	38	6531.44			

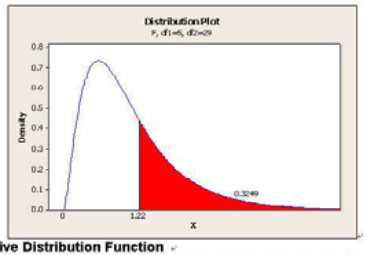
Reduced Model
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	3901.72	975.43	12.61	0.000
Residual Error	34	2629.71	77.34		
Total	38	6531.44			

- **Full Model:** includes all nine variables. $SSR(full) = 4358.85$, reg $df(full) = 9$, $MSE(full) = 74.92$, error $df(full) = 29$.
- **Restricted Model:** includes only the variables X_1 to X_4. $SSR(restricted) = 3901.73$ and reg $df(restricted) = 4$.
- $F_0 = \frac{4358.85 - 3901.73}{(9-4) \times 74.92} = 1.22$.

Step3: Make Conclusion

- Critical-Value Approach: If $F_0 > F_{0.05}(9-4, 29)$, then reject H_0. We have $F_{0.05}(9-4, 29) = 2.55$. Since observed $F_0 < 2.55$, we do not have enough evidence to reject H_0.
- P-value Approach: If p-value$= P(F > F_0) < 0.05$, then reject H_0, where $F \sim F(9-4, 29)$. Here, p-value$= 0.325 > 0.05$.



Cumulative Distribution Function

F distribution with 5 DF in numerator and 29 DF in denominator

X	P(X <= x)
1.22	0.675115