Multiple Regression

Dan Frey
Associate Professor of Mechanical Engineering and Engineering Systems

Plan for Today

- Multiple Regression
  - Estimation of the parameters
  - Hypothesis testing
  - Regression diagnostics
  - Testing lack of fit
- Case study
- Next steps

The Model Equation

For a single variable $Y = \alpha + \beta x + \varepsilon$

For multiple variables $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

α is renamed β_0

$\boxed{p = k + 1}$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

These 1's allow β_0 to enter the equation without being mult by x's

The Model Equation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

Each row of $\mathbf{X}$ is paired with an observation

Each column of $\mathbf{X}$ is paired with a coefficient

$$E(\varepsilon_i) = 0$$

$$Var(\varepsilon_i) = \sigma^2$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_k \end{bmatrix}$$

There are n observations of the response

There are k coefficients

Each observation is affected by an independent homoscedastic normal variates

Accounting for Indices

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

nx1 nxp px1 nx1

$p = k + 1$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Concept Question

Which of these is a valid $\mathbf{X}$ matrix?

$$\mathbf{X} = \begin{bmatrix} 1 & 5.0m & 0.3\sec \\ 1 & 7.1m & 0.2\sec \\ 1 & 3.2m & 0.7\sec \\ 1 & 5.4m & 0.4\sec \end{bmatrix}$$

A

$$\mathbf{X} = \begin{bmatrix} 1 & 5.0m & 0.3m \\ 1 & 7.1V & 0.2V \\ 1 & 3.2\sec & 0.7\sec \\ 1 & 5.4A & 0.4A \end{bmatrix}$$

B

$$\mathbf{X} = \begin{bmatrix} 1 & 5.0m & 0.1\sec \\ 1 & 7.1m & 0.3\sec \end{bmatrix}$$

C

1) A only
2) B only
3) C only
4) A and B
5) B and C
6) A and C
7) A, B, & C
8) None
9) I don’t know

Adding h.o.t. to the Model Equation

Each row of $\mathbf{X}$ is paired with an observation

You can add interactions

You can add curvature

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}x_{12} & x_{11}^{\ 2} \\ 1 & x_{21} & x_{22} & x_{21}x_{22} & x_{21}^{\ 2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}x_{n2} & x_{n1}^{\ 2} \end{bmatrix}$$

There are n observations of the response

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{12} \\ \beta_{11} \end{bmatrix}$$

Estimation of the Parameters β

Assume the model equation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

We wish to minimize the sum squared error

$$L = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)^T \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)$$

To minimize, we take the derivative and set it equal to zero

$$\left.\frac{\partial L}{\partial \boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}$$

The solution is

$$\boxed{\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}}$$

And we define the fitted model

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

MathCad Demo Montgomery Example 10-1

Done in MathCad:

DCM Example 10-1

$ORIGIN := 1$

Viscosity of a Polymer	Reaction temperature	Catalyst feed rate

$$y := \begin{pmatrix} 2256 \\ 2340 \\ 2426 \\ 2293 \\ 2330 \\ 2368 \\ 2250 \\ 2409 \\ 2364 \\ 2379 \\ 2440 \\ 2364 \\ 2404 \\ 2317 \\ 2309 \\ 2328 \end{pmatrix} \qquad X := \begin{pmatrix} 1 & 80 & 8 \\ 1 & 93 & 9 \\ 1 & 100 & 10 \\ 1 & 82 & 12 \\ 1 & 90 & 11 \\ 1 & 99 & 8 \\ 1 & 81 & 8 \\ 1 & 96 & 10 \\ 1 & 94 & 12 \\ 1 & 93 & 11 \\ 1 & 97 & 13 \\ 1 & 95 & 11 \\ 1 & 100 & 8 \\ 1 & 85 & 12 \\ 1 & 86 & 9 \\ 1 & 87 & 12 \end{pmatrix}$$

$$X := \text{augment}\left[X, \overrightarrow{\left(X^{\langle 2 \rangle}\right)^2}\right]$$ disabled

$p := \text{cols}(X)$ $\quad p = 3$

$n := \text{rows}(X)$ $\quad n = 16$

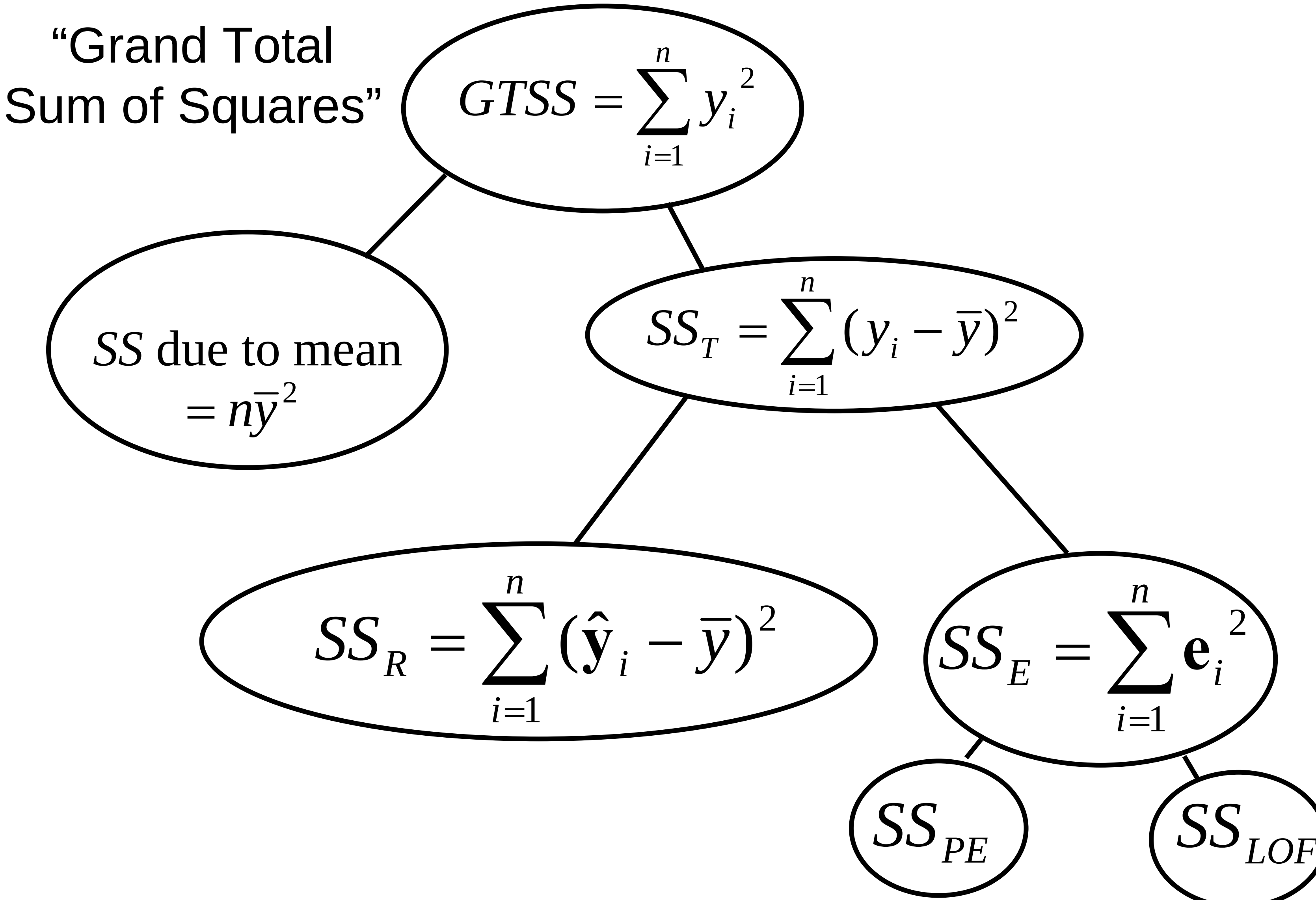
$k := p - 1$

$$\beta\_hat := \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot y \qquad \beta\_hat = \begin{pmatrix} 1.566 \times 10^3 \\ 7.621 \\ 8.585 \end{pmatrix}$$

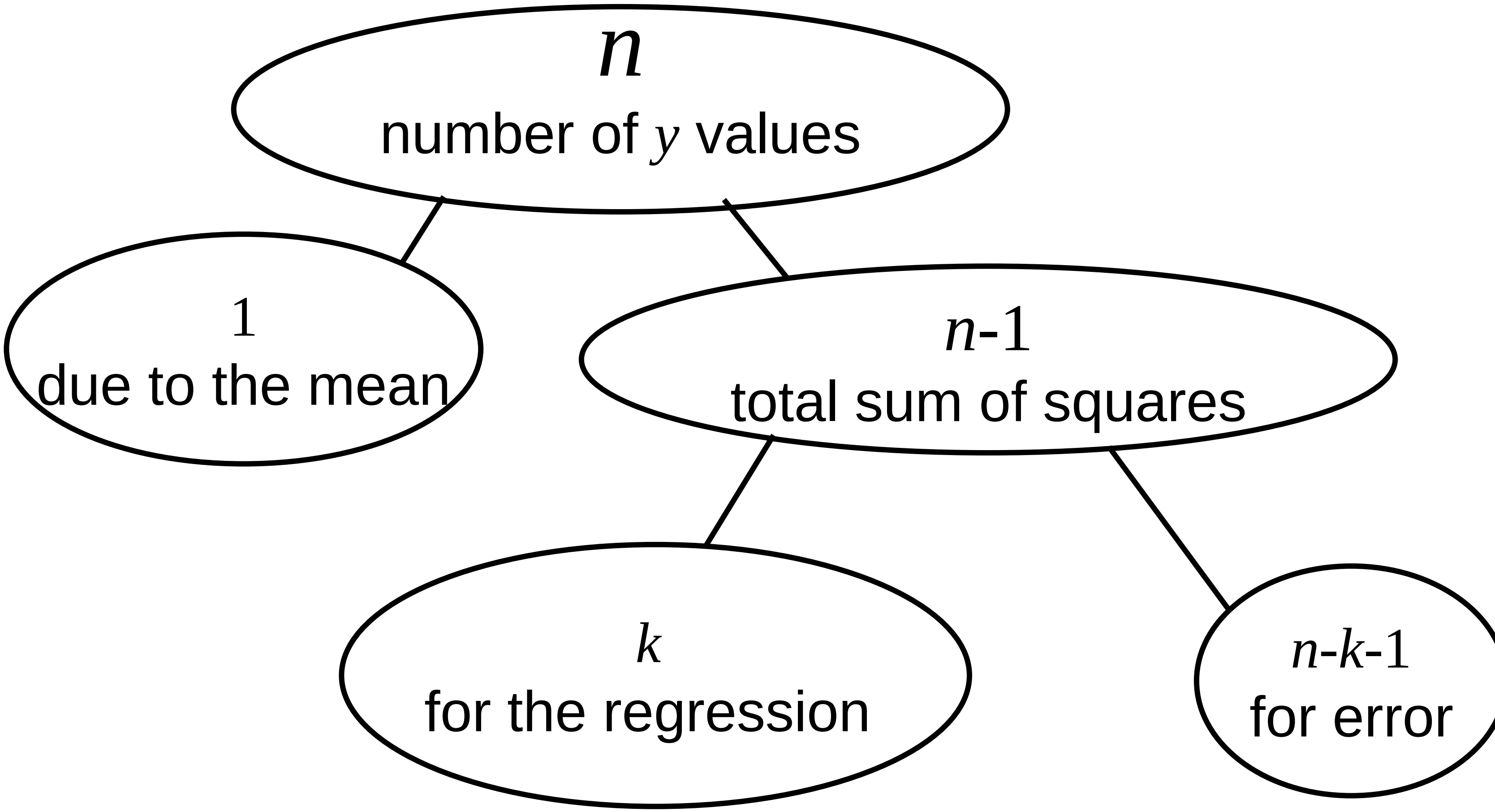
$y\_hat := X \cdot \beta\_hat$ $\qquad \varepsilon := y - y\_hat$

Montgomery, D. C., 2001, *Design and Analysis of Experiments*, John Wiley & Sons.

Breakdown of Sum Squares



Breakdown of DOF



Estimation of the Error Variance σ^2

$$\boldsymbol{\varepsilon} \sim N(0, \sigma)$$

Remember the the model equation $\quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

If assumptions of the model equation hold, then

$$E\left(SS_E/(n-k-1)\right) = \sigma^2$$

So an unbiased estimate of σ^2 is

$$\boxed{\hat{\sigma}^2 = SS_E/(n-k-1)}$$

a.k.a. “coefficient of multiple determination”

R^2 and Adjusted R^2

What fraction of the total sum of squares (SS_T) is accounted for jointly by all the parameters in the fitted model?

$$R^2 \equiv \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

R^2 can only rise as parameters are added

$$R^2_{adj} \equiv 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} = 1 - \left(\frac{n-1}{n-p}\right)(1-R^2)$$

R^2_{adj} can rise or drop as parameters are added

Back to MathCad Demo Montgomery Example 10-1

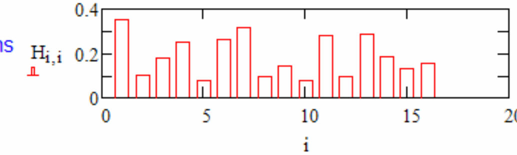
$$F := \frac{SS_R}{SS_E} \cdot \frac{n-k-1}{k}$$

$F = 82.505 \qquad qF(0.05, k, n-k-1) = 0.051$

$$\begin{vmatrix} \text{"reject Ho"} & \text{if } F > qF(0.05, k, n-k-1) \\ \text{"acccept Ho"} & \text{otherwise} \end{vmatrix} = \text{"reject Ho"}$$

$$H := X \cdot \left(X^T \cdot X\right)^{-1} \cdot X^T \qquad 2 \cdot \frac{p}{n} = 0.375 \qquad i := 1 .. n$$

Influence of the observations



Covariance of the residuals

$$\frac{SS_E}{n-k-1} \cdot (\text{identity}(n) - H) =$$

	1	2	3	4	5
1	174.074	-25.35	17.472	-32.12	-11.55
2	-25.35	240.182	-25.095	3.015	-10.307
3	17.472	-25.095	220.326	17.077	-11.703
4	-32.12	3.015	17.077	200.413	-28.569
5	-11.55	-10.307	-11.703	-28.569	247.028
6	-13.233	-39.278	-47.08	34.612	-3.194
7	-89.303	-26.083	14.074	-28.608	-11.11
8	0.567	-22.163	-33.688	3.028	-13.462
9	18.594	-5.781	-23.693	-25.044	-23.292
10	1.128	-12.506	-21.896	-18.033	-19.257
11	44.511	-0.523	-32.286	-25.032	-26.447
12	9.581	-13.972	-28.691	-11.008	-18.377
13	-9.007	-40.011	-50.478	38.124	-2.754
14	-19.441	0.816	6.884	-56.654	...

Montgomery, D. C., 2001, *Design and Analysis of Experiments*, John Wiley & Sons.

Why Hypothesis Testing is Important in Multiple Regression

- Say there are 10 regressor variables
- Then there are 11 coefficients in a linear model
- To make a fully 2nd order model requires
  - 10 curvature terms in each variable
  - 10 choose 2 = 45 interactions
- You'd need 68 samples just to get the matrix $\mathbf{X}^T\mathbf{X}$ to be invertible
- You need a way to discard insignificant terms

Test for Significance of Regression

The hypotheses are

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

$$H_1: \beta_j \neq 0 \text{ for at least one } j$$

The test statistic is

$$\boxed{F_0 = \frac{SS_R / k}{SS_E / (n-k-1)}}$$

Reject H_0 if

$$F_0 > F_{\alpha,k,n-k-1}$$

Test for Significance
Individual Coefficients

The hypotheses are $$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

The test statistic is $$\boxed{t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}}$$

$$C = \left(\mathbf{X}^T \mathbf{X}\right)^{-1}$$

Standard error

$$\sqrt{\hat{\sigma}^2 C_{jj}}$$

Reject H_0 if $|t_0| > t_{\alpha/2, n-k-1}$

Test for Significance of Groups of Coefficients

Partition the coefficients into two groups $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}$ $\begin{matrix} \leftarrow \text{to be removed} \\ \leftarrow \text{to remain} \end{matrix}$

Reduced model $\mathbf{y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$

$$\boxed{\begin{matrix} H_0 : \boldsymbol{\beta}_1 = \mathbf{0} \\ H_1 : \boldsymbol{\beta}_1 \neq \mathbf{0} \end{matrix}}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cancel{x_{12}} & \cdots & \cancel{x_{1k}} \\ 1 & x_{21} & \cancel{x_{22}} & \cdots & \cancel{x_{2k}} \\ \vdots & \vdots & \cancel{\vdots} & \ddots & \cancel{\vdots} \\ 1 & x_{n1} & \cancel{x_{n2}} & \cdots & \cancel{x_{nk}} \end{bmatrix} \Rightarrow \mathbf{X}_2$$

Basically, you form $\mathbf{X}_2$ by removing the columns associated with the coefficients you are testing for significance

Test for Significance Groups of Coefficients

Reduced model $\mathbf{y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$

The regression sum of squares for the reduced model is $SS_R(\boldsymbol{\beta}_2) = \mathbf{y}^T\mathbf{H}_2\mathbf{y} - n\bar{y}^2$

Define the sum squares of the removed set given the other coefficients are in the model $SS_R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2) \equiv SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta}_2)$

The partial F test

$$\boxed{F_0 = \frac{SS_R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)/r}{SS_E/(n-p)}}$$

Reject H_0 if $F_0 > F_{\alpha,r,n-p}$

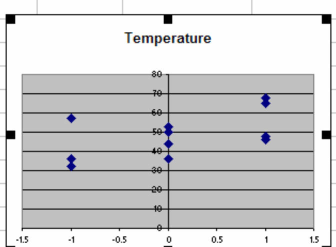
Excel Demo -- Montgomery Ex10-2

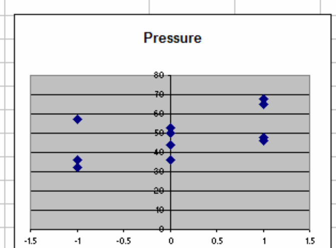
Regression Statistics	
Multiple R	0.907831
R Square	0.824157
Adjusted R Squa	0.723675
Standard Error	5.978135
Observations	12

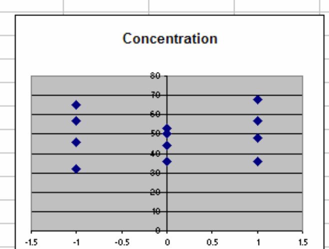
ANOVA

	df	SS	MS	F	ignificance F
Regression	4	1172.5	293.125	8.202032	0.008857
Residual	7	250.1666667	35.7381		
Total	11	1422.666667			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	Ipper 95.0%
Intercept	49.33333	1.725738857	28.58679	1.65E-08	45.25261	53.41405	45.25261	53.41405
X Variable 1	5.625	2.113589815	2.661349	0.032404	0.627158	10.62284	0.627158	10.62284
X Variable 2	10.625	2.113589815	5.026992	0.001518	5.627158	15.62284	5.627158	15.62284
X Variable 3	1.125	2.113589815	0.53227	0.611009	-3.87284	6.122842	-3.87284	6.122842
X Variable 4	-0.875	2.113589815	-0.41399	0.691273	-5.87284	4.122842	-5.87284	4.122842



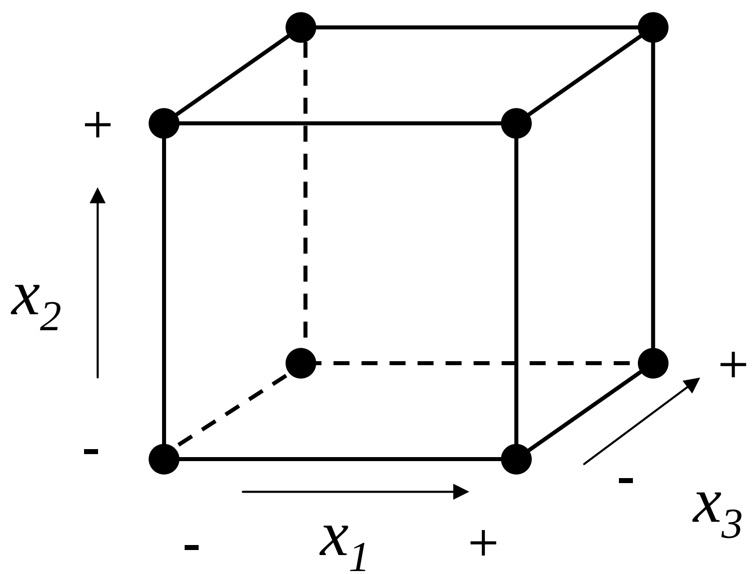




Montgomery, D. C., 2001, *Design and Analysis of Experiments*, John Wiley & Sons.

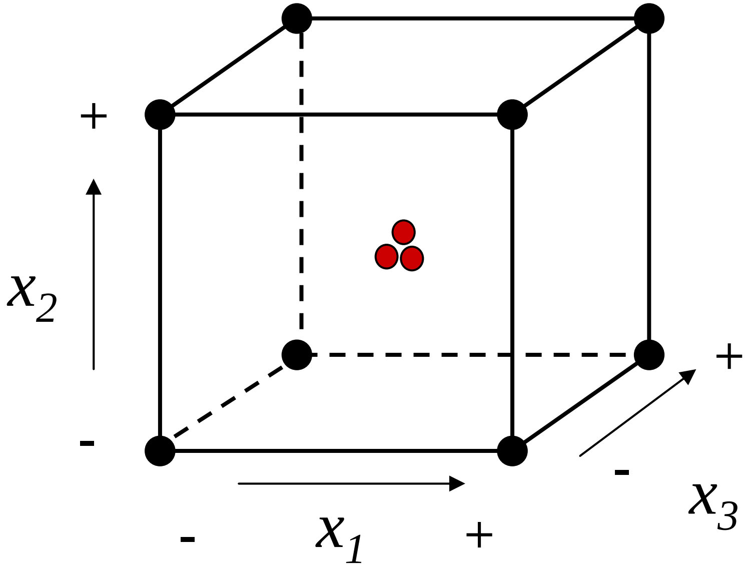
Factorial Experiments

Cuboidal Representation



Exhaustive search of the space of discrete 2-level factors is the full factorial 2^3 experimental design

Adding Center Points



Center points allow an experimenter to check for curvature and, if replicated, allow for an estimate of pure experimental error

Plan for Today

- Mud cards
- Multiple Regression
  - Estimation of the parameters
  - Hypothesis testing
  - Regression diagnostics
  - Testing lack of fit
- Case study
- Next steps

The "Hat" Matrix

Since $$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

and $$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

therefore $$\hat{\mathbf{y}} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

So we define $$\mathbf{H} \equiv \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$$

Which maps from observations $\mathbf{y}$ to predictions $\hat{\mathbf{y}}$

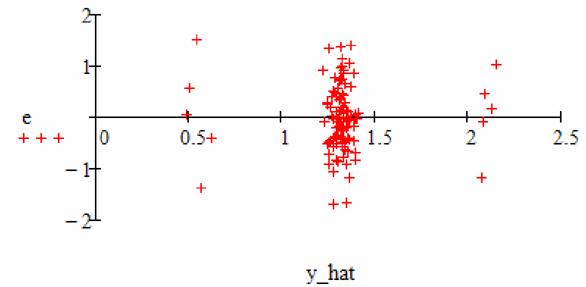
$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$$

Influence Diagnostics

- The relative disposition of points in x space determines their effect on the coefficients
- The hat matrix $\mathbf{H}$ gives us an ability to check for leverage points
- h_{ij} is the amount of leverage exerted by point $\mathbf{y}_j$ on $\hat{\mathbf{y}}_i$
- Usually the diagonal elements ~p/n and it is good to check whether the diagonal elements within 2X of that

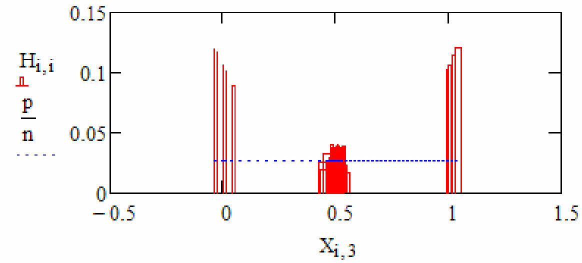
MathCad Demo on Distribution of Samples and Its Effect on Regression

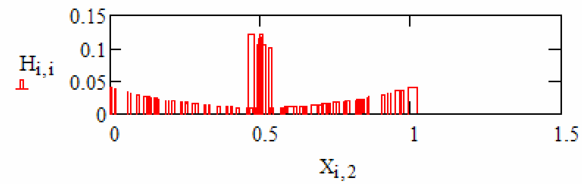
Plot the residuals



$$H := X\cdot\left(X^T\cdot X\right)^{-1}\cdot X^T \qquad i := 1..110$$

Influence of the observations





Standardized Residuals

The residuals are defined as $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$

So an unbiased estimate of σ^2 is $\hat{\sigma}^2 = SS_E/(n-p)$

The standardized residuals are defined as $\mathbf{d} = \dfrac{\mathbf{e}}{\hat{\sigma}}$

If these elements were z-scores then with probability 99.7%

$$-3 < d_i < 3$$

Studentized Residuals

The residuals are defined as $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$

therefore $\mathbf{e} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}$

So the covariance matrix of the residuals is

$$\mathrm{Cov}(\mathbf{e}) = \sigma^2 \mathrm{Cov}(\mathbf{I} - \mathbf{H})$$

The studentized residuals are defined as

$$r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2 (1 - h_{ii})}}$$

If these elements were z-scores then with probability 99.7%

???? $-3 < r_i < 3$

Testing for Lack of Fit
(Assuming a Central Composite Design)

- Compute the standard deviation of the center points and assume that represents the MS_{PE}

$$MS_{PE} = \frac{\sum_{\text{center points}} \left(y_i - \bar{y}\right)}{n_C - 1}$$

$$MS_{LOF} = \frac{SS_{LOF}}{p}$$

$$SS_{PE} = (n-1)MS_{PE}$$

$$\boxed{F_0 = \frac{MS_{LOF}}{MS_{PE}}}$$

$$SS_{PE} + SS_{LOF} = SS_E$$

Concept Test

- You perform a linear regression of 100 data points (n=100). There are two independent variables x_1 and x_2. The regression R^2 is 0.72. Both β_1 and β_2 pass a t test for significance. You decide to add the interaction x_1x_2 to the model. Select all the things that cannot happen:

1) Absolute value of β_1 decreases

2) β_1 changes sign

3) R^2 decreases

4) β_1 fails the t test for significance

Plan for Today

- Mud cards
- Multiple Regression
  - Estimation of the parameters
  - Hypothesis testing
  - Regression diagnostics
  - Testing lack of fit
- Case study
- Next steps

Scenario

- The FAA and EPA are interested in reducing CO2 emissions
- Some parameters of airline operations are thought to effect CO2 (e.g., Speed, Altitude, Temperature, Weight)
- Imagine flights have been made with special equipment that allowed CO2 emission to be measured (data provided)
- You will report to the FAA and EPA on your analysis of the data and make some recommendations

Phase One

- Open a Matlab window
- Load the data (load FAAcase3.mat)
- Explore the data

Phase Two

- Do the regression
- Examine the betas and their intervals
- Plot the residuals

```
y=[CO2./ground_speed];
ones(1:3538)=1;
X=[ones' TAS alt temp weight];
[b,bint,r,rint,stats] = regress(y,X,0.05);
yhat=X*b;
plot(yhat,r,'+')
```

```
dims=size(X);
i=2:dims(1)-1;
climb(1)=1;
climb(dims(1))=0;
des(1)=0;
des(dims(1))=1;
climb(i)=(alt(i)>(alt(i-1)+100))|(alt(i+1)>(alt(i)+100));
des(i)=(alt(i)<(alt(i-1)-100))|(alt(i+1)<(alt(i)-100));
for i=dims(1):-1:1
if climb(i)|des(i)
        y(i,:)=[]; X(i,:)=[]; yhat(i,:)=[]; r(i,:)=[];
end
end
hold off
plot(yhat,r,'or')
```

This code will remove the points at which the aircraft is climbing or descending

Try The Regression Again on Cruise Only Portions

- What were the effects on the residuals?
- What were the effects on the betas?

```
hold off
[b,bint,r,rint,stats] = regress(y,X,0.05);
yhat=X*b;
plot(yhat,r,'+')
```

See What Happens if We Remove Variables

- Remove weight & temp
- Do the regression (CO_2 vs TAS & alt)
- Examine the betas and their intervals

```
[b,bint,r,rint,stats] = regress(y,X(:,1:3),0.05);
```

Phase Three

- Try different data (flight34.mat)
- Do the regression
- Examine the betas and their intervals
- Plot the residuals

```
y=[fuel_burn];
ones(1:34)=1;
X=[ones' TAS alt temp];
[b,bint,r,rint,stats] = regress(y,X,0.05);
yhat=X*b;
plot(yhat,r,'+')
```

Adding Interactions

```
X(:,5)=X(:,2).*X(:,3);
```

This line will add a interaction

What's the effect on the regression?

Case Wrap-Up

- What were the recommendations?
- What other analysis might be done?
- What were the key lessons?

Next Steps

- Wenesday 25 April
  - Design of Experiments
  - Please read "Statistics as a Catalyst to Learning"
- Friday 27 April
  - Recitation to support the term project
- Monday 30 April
  - Design of Experiments
- Wednesday 2 May
  - Design of Computer Experiments
- Friday 4 May?? Exam review??
- Monday 7 May – Frey at NSF
- Wednesday 9 May – Exam #2