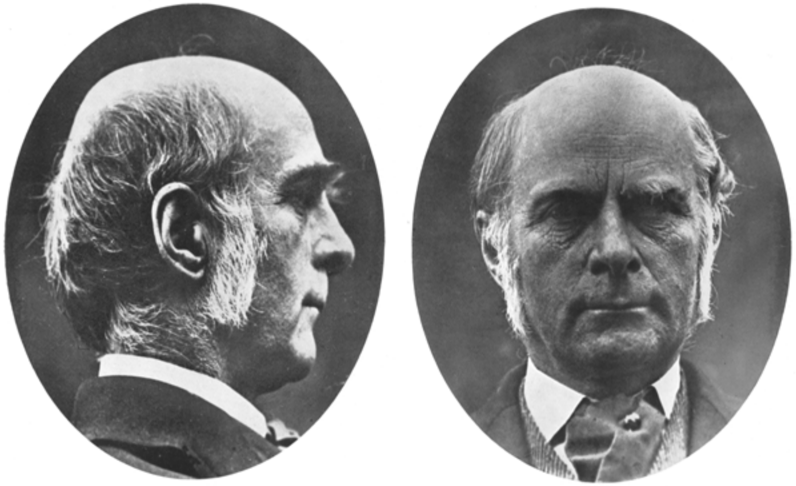
Multiple Linear Regression

- First-Order model with 2 predictor variables
- General linear regression model
- The model in matrix notation
- Estimation of regression coefficients
- ANOVA
- Testing of individual parameter coefficient
- R square and adjusted R square
- Variable selection

Historical Origins of Regression Models

- First introduced by Sir Francis Galton in the latter part of the 19th century
- Galton had studied the relation between heights of parents and children and noted that the heights of children of both tall and short parents appeared to "revert" or "regress" to the mean of the group.



geographer, meteorologist, tropical explorer, founder of differential psychology, inventor of fingerprint identification, pioneer of statistical correlation and regression …

First-Order Model with two predictor variables

Model:

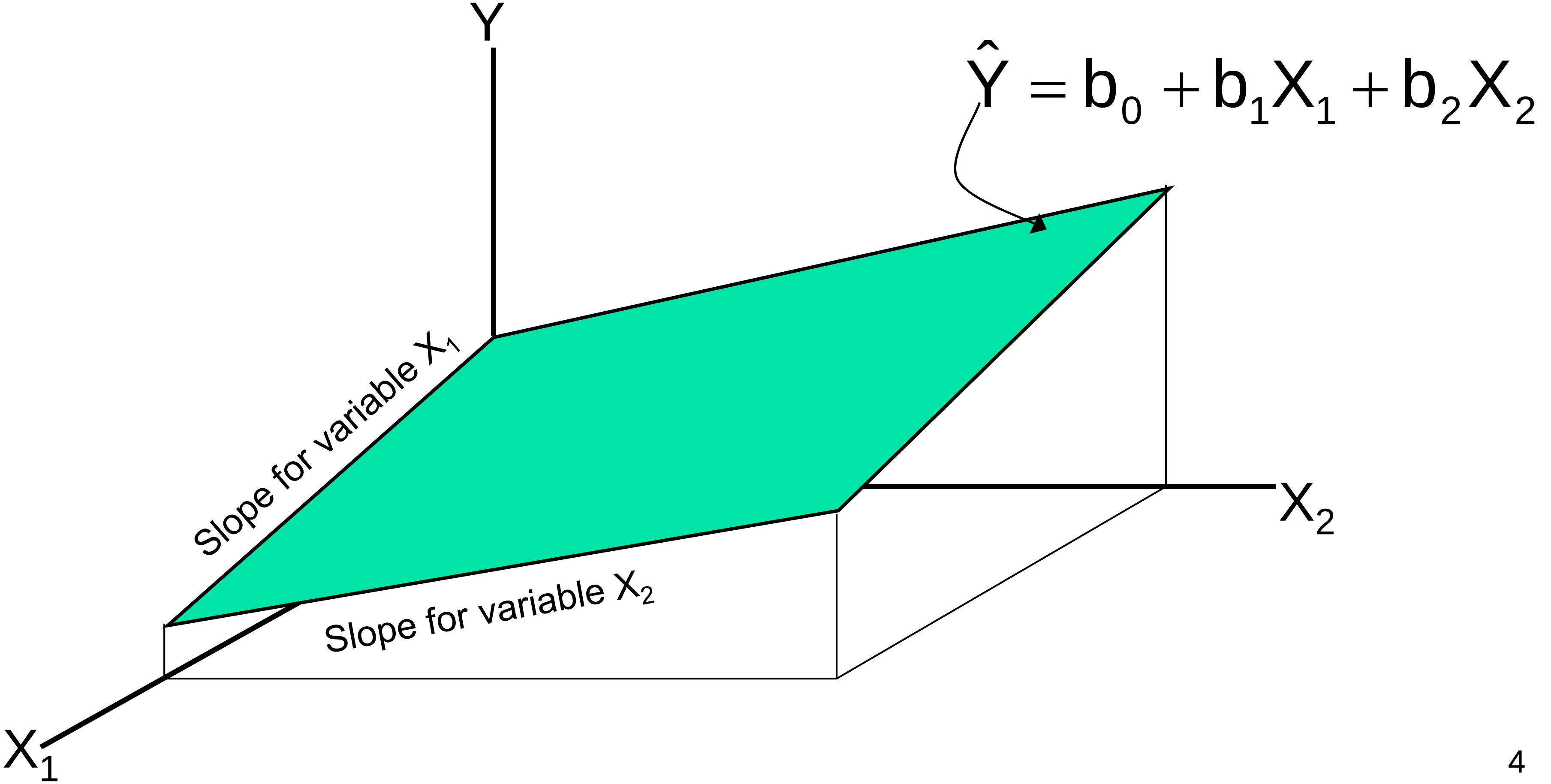
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

<u>Assumptions</u>:

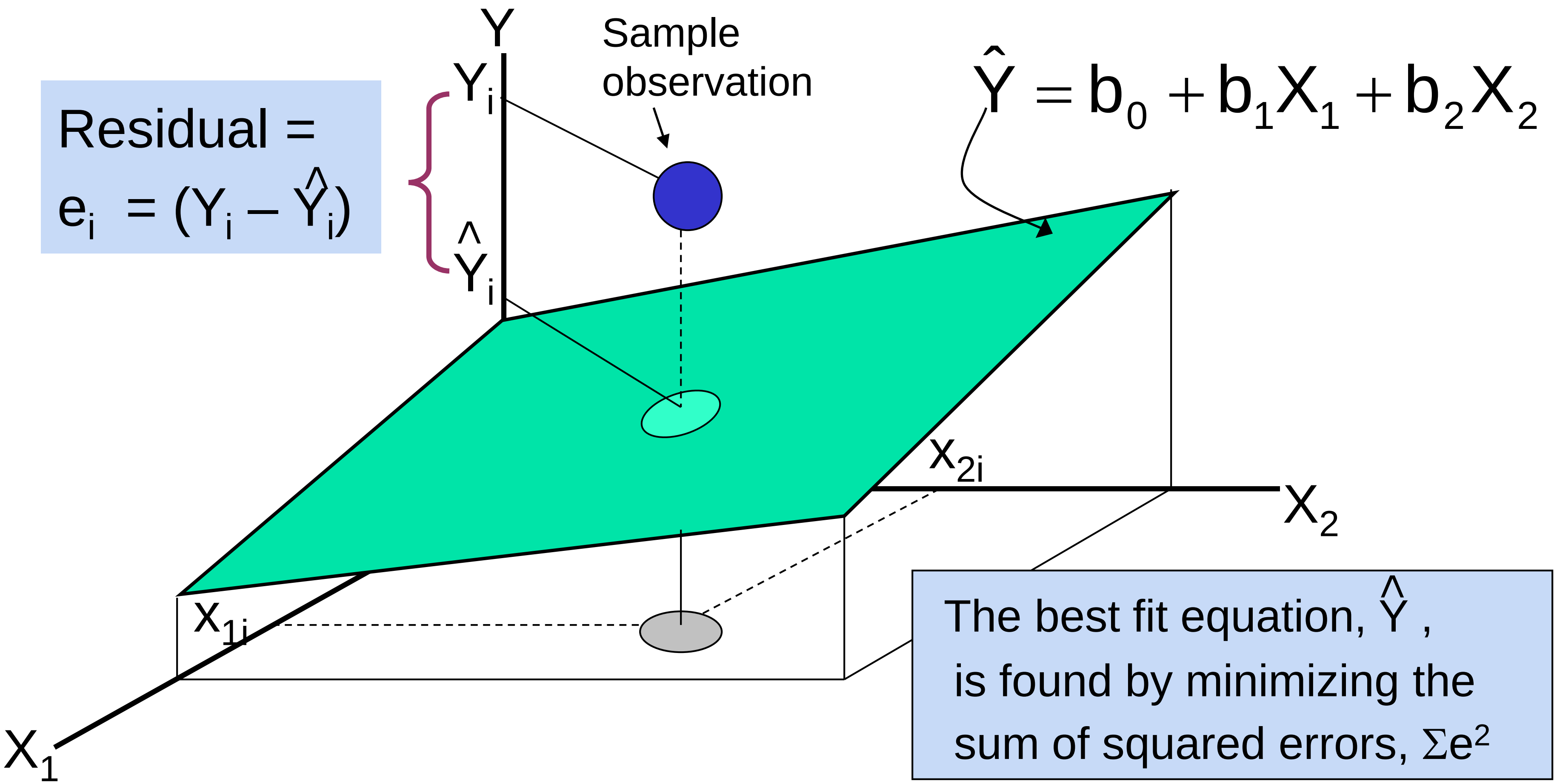
- $E(\varepsilon_i) = 0$
- Variance $(\varepsilon_i) = \sigma^2$
- Covariance $(\varepsilon_i, \varepsilon_j) = 0$

So, $$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Response Surface (Plane)



The regression equation



$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

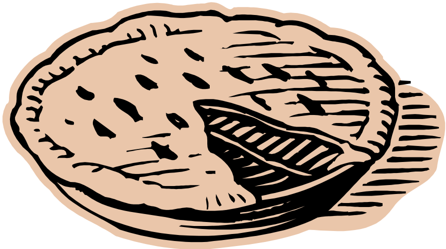
Residual =

$e_i = (Y_i - \hat{Y}_i)$

The best fit equation, $\hat{Y}$,
is found by minimizing the
sum of squared errors, Σe^2

Example (Pie sales)

- A distributor of frozen desert pies wants to evaluate factors thought to influence demand
  - Dependent variable: Pie sales (units per week)
  - Independent variables: Price (in $)
    Advertising ($100's)
- Data are collected for 15 weeks

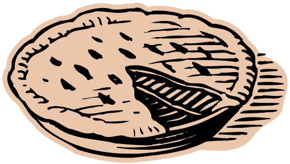


Pie Example (cont'd)

Week	Pie Sales	Price ($)	Advertising ($100s)
1	350	5.50	3.3
2	460	7.50	3.3
3	350	8.00	3.0
4	430	8.00	4.5
5	350	6.80	3.0
6	380	7.50	4.0
7	430	4.50	3.0
8	470	6.40	3.7
9	450	7.00	3.5
10	490	5.00	4.0
11	340	7.20	3.5
12	300	7.90	3.2
13	440	5.90	4.0
14	450	5.00	3.5
15	300	7.00	2.7

Multiple regression equation:

$$\widehat{\text{Sales}} = b_0 + b_1\,(\text{Price}) + b_2\,(\text{Advertising})$$



Computer Output

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.722[a]	.521	.442	47.46341

a. Predictors: (Constant), Advertising, Price

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	29460.027	2	14730.013	6.539	.012[b]
	Residual	27033.306	12	2252.776		
	Total	56493.333	14			

a. Dependent Variable: Pie Sales

b. Predictors: (Constant), Advertising, Price

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	306.526	114.254		2.683	.020	57.588	555.464
	Price	-24.975	10.832	-.461	-2.306	.040	-48.576	-1.374
	Advertising	74.131	25.967	.570	2.855	.014	17.553	130.709

a. Dependent Variable: Pie Sales

Regression equation and its interpretation

$$\widehat{\text{Sales}} = 306.526 - 24.975(\text{Price}) + 74.131(\text{Advertising})$$

where
Sales is in number of pies per week
Price is in $
Advertising is in $100's.

b_1 = -24.975: sales will decrease, on average, by 24.975 pies per week for each $1 increase in selling price, net of the effects of changes due to advertising

b_2 = 74.131: sales will increase, on average, by 74.131 pies per week for each $100 increase in advertising, net of the effects of changes due to price

Regression equation and its interpretation

- The parameters $ß_1$ and $ß_2$ are sometimes called partial regression coefficients because they reflect the partial effect of one predictor variable when the other predictor variable is included in the model and is held constant.
- The two predictor variables are said to have additive effects or not to interact.

General linear regression model ($X_1, \ldots, X_k$) do not need to represent different predictor variables)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{i,k} + \varepsilon_i$$

- If the variables $X_1, \ldots, X_k$ represent k different predictor variables, the general linear regression model is a first-order model in which there are no interaction effects between the predictor variables.
- Qualitative predictor variables Vs Quantitative predictor variables. Eg. Gender [Indicator variables]
- Polynomial regression: contain squared and higher – order terms of the predictor variables.
- Transformed variables
- Interaction effects
- Combination of cases

General linear regression model in matrix notation

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

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,k} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

General linear regression model in matrix notation

Assumptions

$$\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

Hence

$$\mathbf{Y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

General linear regression model in matrix notation

<u>Expected value of **Y**</u>

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$$

<u>Variance-covariance matrix of **Y**</u>

$$Var(\mathbf{Y}) = \sigma^2\mathbf{I}$$

General linear regression model in matrix notation

<u>Estimation of regression coefficients (least squares estimate)</u>

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

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

General linear regression model in matrix notation

Fitted values

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

Residuals

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$$

Sum of Squares of Error (SSE)

$$SSE = (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}})$$

Is the Model Significant?

- F-Test for Overall Significance of the Model
- Shows if there is a linear relationship between all of the X variables considered together and Y
- Use F test statistic
- Hypotheses:

  H_0: $\beta_1 = \beta_2 = \dots = \beta_k = 0$ (no linear relationship)

  H_1: at least one $\beta_i \neq 0$ (at least one independent variable affects Y)

ANOVA table

Source	DF	SS	MS	F
Regression	k	SSR = SST-SSE	MSR=SSR/k	MSR/MSE
Error	$n - k - 1$	SSE	MSE=SSE/($n - k - 1$)	
Total	$n - 1$	SST		

Test of an individual parameter coefficient

H_0: $\beta_i = 0$
H_1: $\beta_i \neq 0$

Test statistics: T = $\dfrac{b_i}{s\{b_i\}}$

Decision rule: Reject Ho if $|T| \geq t_{\alpha/2, n-(k+1)}$

Coefficient of Multiple Determination

- Reports the proportion of total variation in Y explained by all X variables taken together

$$R^2 = r^2_{Y.12..k} = \frac{\text{SSR}}{\text{SST}} = \frac{\text{regression sum of squares}}{\text{total sum of squares}}$$

Adjusted R^2 (R_a^2)

- R^2 never decreases when a new X variable is added to the model
  - This can be a disadvantage when comparing models
- What is the net effect of adding a new variable?
  - We lose a degree of freedom when a new X variable is added
  - Did the new X variable add enough explanatory power to offset the loss of one degree of freedom?

Adjusted R^2 (R_a^2)

- Shows the proportion of variation in Y explained by all X variables adjusted for the number of X variables used

$$R_a^2 = 1 - \left[(1 - R^2) \left(\frac{n-1}{n-(k+1)} \right) \right]$$

(where n = sample size, k = number of independent variables)

- Penalize excessive use of unimportant independent variables
- Smaller than R^2
- Useful in comparing among models

Variable screening

- Goal is to develop a model with the best set of independent variables
  - Easier to interpret if unimportant variables are removed
  - Lower probability of collinearity
- Stepwise regression procedure
  - Provide evaluation of alternative models as variables are added (or withdrawn)
- Best-subset approach
  - Try all combinations and select the best using various criteria, such as the highest adjusted R^2

AIC_p and SBC_p

$$AIC_p = n \ln SSE_p - n \ln n + 2p$$

$$SBC_p = n \ln SSE_p - n \ln n + p \ln n$$

Both selection criteria penalize models having large numbers of predictors. For $n \geqq 8$, the second criterion tends to favor more parsimonious models.

Example: Boston Housing Data

Data: https://github.com/QuantLet/MVA-ToDo/tree/master/QID-1615-MVAlinregbh

(X_1 to X_{14}, description of the 14 variables on p.561 of the textbook)

- Download the data as bostonh.txt
- Transform the data
  - (R-code)

```
x <- read.table("bostonh.txt")
xt <- x
xt[, 1]   = log(x[, 1])
xt[, 2]   = x[, 2]/10
xt[, 3]   = log(x[, 3])
xt[, 5]   = log(x[, 5])
xt[, 6]   = log(x[, 6])
xt[, 7]   = ((x[, 7]^2.5))/10000
xt[, 8]   = log(x[, 8])
xt[, 9]   = log(x[, 9])
xt[, 10]  = log(x[, 10])
xt[, 11]  = exp(0.4 * x[, 11])/1000
xt[, 12]  = x[, 12]/100
xt[, 13]  = sqrt(x[, 13])
xt[, 14]  = log(x[, 14])
```

Example: Boston Housing Data

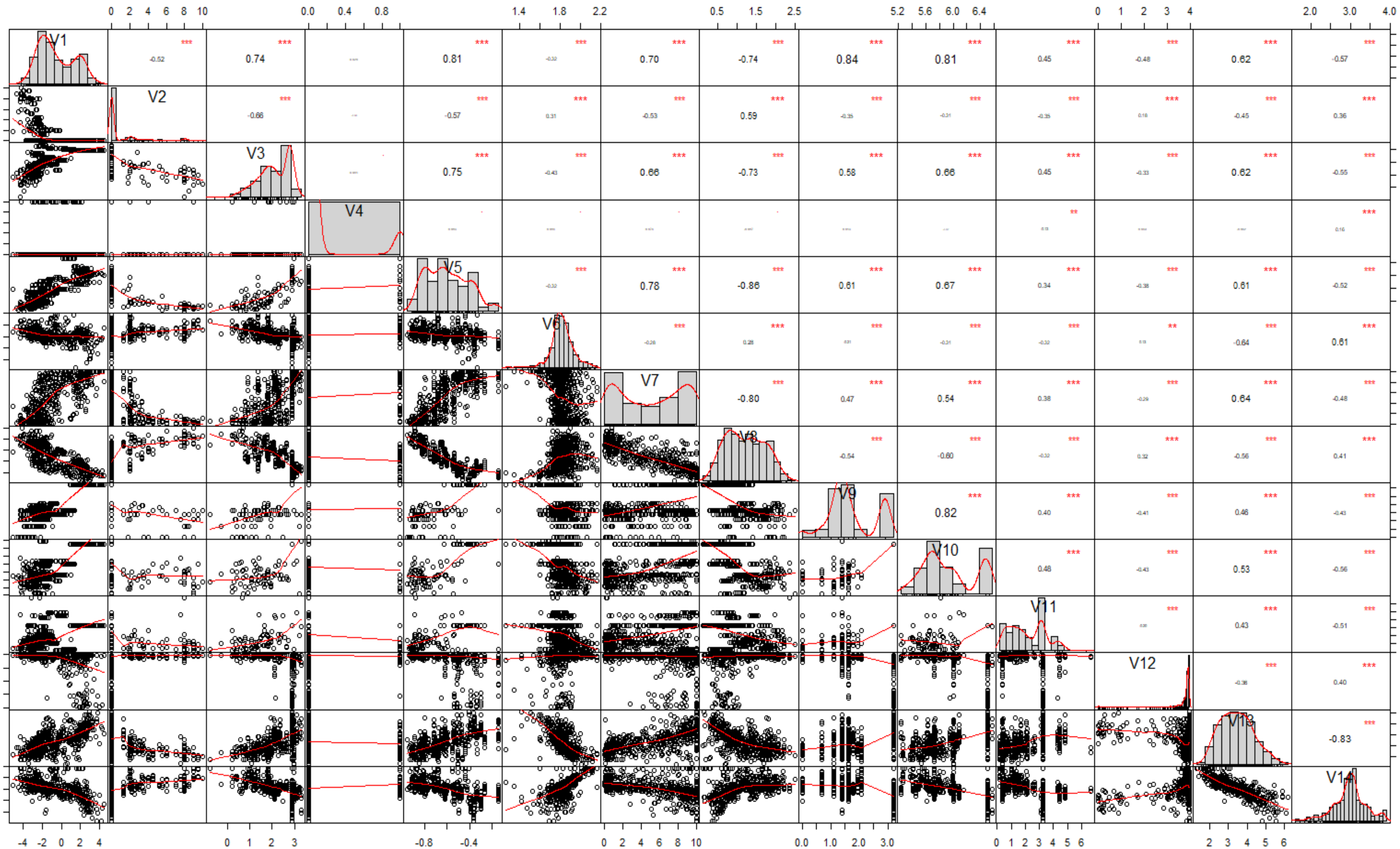
Correlation matrix

(R Code)

- [Install package: PerformanceAnalytics]

```
library(PerformanceAnalytics)

chart.Correlation(xt, method="pearson", histogram=TRUE, pch=16)
```


V1
V2
V3
V4
V5
V6
V7
V8
V9
V10
V11
V12
V13
V14
0.74
0.81
0.70
-0.74
0.84
0.81
0.62
0.75
0.66
-0.73
0.58
0.66
0.62
0.78
-0.86
0.61
0.67
0.61
0.61
-0.80
0.54
0.64
0.82
0.53
-0.83

Example: Boston Housing Data

Regression with the transformed data: V14 (dependent variable) v1-v13 (independent variables)

(R Code)

```
fit <- lm(V14~.,data=xt)
summary(fit)
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.176874   0.379017  11.020  < 2e-16 ***
V1          -0.014606   0.011650  -1.254 0.210527    
V2           0.001392   0.005639   0.247 0.805121    
V3          -0.012709   0.022312  -0.570 0.569195    
V4           0.109980   0.036634   3.002 0.002817 ** 
V5          -0.283112   0.105340  -2.688 0.007441 ** 
V6           0.421108   0.110175   3.822 0.000149 ***
V7           0.006403   0.004863   1.317 0.188536    
V8          -0.183154   0.036804  -4.977 8.97e-07 ***
V9           0.068362   0.022473   3.042 0.002476 ** 
V10         -0.201832   0.048432  -4.167 3.64e-05 ***
V11         -0.040017   0.008091  -4.946 1.04e-06 ***
V12          0.044472   0.011456   3.882 0.000118 ***
V13         -0.262615   0.016091 -16.320  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2008 on 492 degrees of freedom
Multiple R-squared:  0.765,	Adjusted R-squared:  0.7588 
F-statistic: 123.2 on 13 and 492 DF,  p-value: < 2.2e-16
```

Example: Boston Housing Data

Variable selection

(R Code)

```
library(MASS)
start = lm(V14 ~ 1, data=xt)
fitAll = lm(V14 ~ ., data=xt)
stepwise <- step(start, direction="both",scope=formula(fitAll))
```

Example: Boston Housing Data

```
Start:  AIC=-904.37
V14 ~ 1

       Df Sum of Sq    RSS      AIC
+ V13   1    57.432 26.944 -1479.98
+ V6    1    31.442 52.935 -1138.28
+ V1    1    27.149 57.227 -1098.83
+ V10   1    26.195 58.181 -1090.46
+ V3    1    25.886 58.491 -1087.78
+ V5    1    22.401 61.976 -1058.49
+ V11   1    21.794 62.583 -1053.56
+ V7    1    19.611 64.766 -1036.21
+ V9    1    15.930 68.446 -1008.25
+ V8    1    13.889 70.487  -993.38
+ V12   1    13.661 70.715  -991.75
+ V2    1    11.139 73.237  -974.01
+ V4    1     2.117 82.259  -915.23
<none>                84.376  -904.37

Step:  AIC=-1479.98
V14 ~ V13

       Df Sum of Sq    RSS      AIC
+ V11   1     2.348 24.596 -1524.11
+ V10   1     1.615 25.330 -1509.25
+ V12   1     1.060 25.884 -1498.28
+ V6    1     0.981 25.964 -1496.74
+ V4    1     0.976 25.969 -1496.64
+ V1    1     0.399 26.545 -1485.53
+ V8    1     0.338 26.607 -1484.36
+ V9    1     0.313 26.631 -1483.89
+ V7    1     0.269 26.675 -1483.06
+ V3    1     0.234 26.711 -1482.38
<none>                26.944 -1479.98
+ V5    1     0.021 26.924 -1478.37
+ V2    1     0.010 26.934 -1478.17
- V13   1    57.432 84.376  -904.37
```

Example: Boston Housing Data

```
Step:  AIC=-1524.11
V14 ~ V13 + V11

       Df Sum of Sq    RSS     AIC
+ V12   1     0.890 23.706 -1540.8
+ V6    1     0.803 23.793 -1538.9
+ V10   1     0.676 23.921 -1536.2
+ V4    1     0.656 23.941 -1535.8
+ V7    1     0.564 24.032 -1533.8
+ V8    1     0.563 24.033 -1533.8
+ V2    1     0.161 24.435 -1525.4
<none>                24.596 -1524.1
+ V1    1     0.058 24.538 -1523.3
+ V9    1     0.034 24.562 -1522.8
+ V3    1     0.007 24.589 -1522.3
+ V5    1     0.001 24.596 -1522.1
- V11   1     2.348 26.944 -1480.0
- V13   1    37.986 62.583 -1053.6

Step:  AIC=-1540.76
V14 ~ V13 + V11 + V12

       Df Sum of Sq    RSS     AIC
+ V6    1    1.0900 22.616 -1562.6
+ V8    1    0.8244 22.882 -1556.7
+ V7    1    0.6711 23.035 -1553.3
+ V4    1    0.6233 23.083 -1552.2
+ V10   1    0.3226 23.384 -1545.7
+ V2    1    0.1641 23.542 -1542.3
<none>                23.706 -1540.8
+ V5    1    0.0529 23.653 -1539.9
+ V9    1    0.0087 23.698 -1539.0
+ V1    1    0.0083 23.698 -1538.9
+ V3    1    0.0019 23.704 -1538.8
- V12   1    0.8900 24.596 -1524.1
- V11   1    2.1781 25.884 -1498.3
- V13   1   31.0378 54.744 -1119.3
```

Example: Boston Housing Data

```
Step:  AIC=-1562.58
V14 ~ V13 + V11 + V12 + V6

       Df Sum of Sq    RSS     AIC
+ V8    1    0.6459 21.970 -1575.2
+ V4    1    0.5384 22.078 -1572.8
+ V10   1    0.3683 22.248 -1568.9
+ V7    1    0.3591 22.257 -1568.7
+ V2    1    0.1804 22.436 -1564.6
<none>              22.616 -1562.6
+ V5    1    0.0165 22.600 -1561.0
+ V3    1    0.0113 22.605 -1560.8
+ V1    1    0.0002 22.616 -1560.6
+ V9    1    0.0002 22.616 -1560.6
- V6    1    1.0900 23.706 -1540.8
- V12   1    1.1771 23.793 -1538.9
- V11   1    1.9527 24.569 -1522.7
- V13   1   14.8598 37.476 -1309.0

Step:  AIC=-1575.24
V14 ~ V13 + V11 + V12 + V6 + V8

       Df Sum of Sq    RSS     AIC
+ V10   1    1.0159 20.954 -1597.2
+ V5    1    0.6357 21.335 -1588.1
+ V4    1    0.3739 21.596 -1581.9
+ V1    1    0.3739 21.596 -1581.9
+ V3    1    0.2053 21.765 -1578.0
+ V9    1    0.0943 21.876 -1575.4
<none>              21.970 -1575.2
+ V7    1    0.0049 21.966 -1573.3
+ V2    1    0.0041 21.966 -1573.3
- V8    1    0.6459 22.616 -1562.6
- V6    1    0.9115 22.882 -1556.7
- V12   1    1.4081 23.378 -1545.8
- V11   1    2.1779 24.148 -1529.4
- V13   1   14.7438 36.714 -1317.4
```

Example: Boston Housing Data

```
Step:  AIC=-1597.2
V14 ~ V13 + V11 + V12 + V6 + V8 + V10

       Df Sum of Sq    RSS     AIC
+ V4    1    0.3448 20.610 -1603.6
+ V5    1    0.2815 20.673 -1602.0
+ V9    1    0.2714 20.683 -1601.8
<none>              20.954 -1597.2
+ V3    1    0.0345 20.920 -1596.0
+ V2    1    0.0166 20.938 -1595.6
+ V7    1    0.0048 20.950 -1595.3
+ V1    1    0.0010 20.953 -1595.2
- V12   1    0.7902 21.745 -1580.5
- V6    1    0.9020 21.857 -1577.9
- V10   1    1.0159 21.970 -1575.2
- V11   1    1.2048 22.159 -1570.9
- V8    1    1.2935 22.248 -1568.9
- V13   1   13.5776 34.532 -1346.4

Step:  AIC=-1603.59
V14 ~ V13 + V11 + V12 + V6 + V8 + V10 + V4

       Df Sum of Sq    RSS     AIC
+ V5    1    0.3392 20.270 -1610.0
+ V9    1    0.2335 20.376 -1607.4
<none>              20.610 -1603.6
+ V3    1    0.0746 20.535 -1603.4
+ V2    1    0.0219 20.588 -1602.1
+ V1    1    0.0057 20.604 -1601.7
+ V7    1    0.0008 20.609 -1601.6
- V4    1    0.3448 20.954 -1597.2
- V12   1    0.7429 21.353 -1587.7
- V6    1    0.8593 21.469 -1584.9
- V10   1    0.9867 21.596 -1581.9
- V11   1    1.0399 21.650 -1580.7
- V8    1    1.0605 21.670 -1580.2
- V13   1   13.4205 34.030 -1351.8
```

Example: Boston Housing Data

```
Step:  AIC=-1609.99
V14 ~ V13 + V11 + V12 + V6 + V8 + V10 + V4 + V5

       Df Sum of Sq    RSS     AIC
+ V9    1    0.2914 19.979 -1615.3
<none>              20.270 -1610.0
+ V7    1    0.0297 20.241 -1608.7
+ V3    1    0.0268 20.244 -1608.7
+ V1    1    0.0110 20.259 -1608.3
+ V2    1    0.0027 20.268 -1608.1
- V5    1    0.3392 20.610 -1603.6
- V4    1    0.4025 20.673 -1602.0
- V10   1    0.6118 20.882 -1596.9
- V12   1    0.6699 20.940 -1595.5
- V6    1    0.8869 21.157 -1590.3
- V11   1    1.0888 21.359 -1585.5
- V8    1    1.2798 21.550 -1581.0
- V13   1   12.0721 32.343 -1375.6

Step:  AIC=-1615.32
V14 ~ V13 + V11 + V12 + V6 + V8 + V10 + V4 + V5 + V9

       Df Sum of Sq    RSS     AIC
<none>              19.979 -1615.3
+ V1    1    0.0561 19.923 -1614.7
+ V7    1    0.0497 19.929 -1614.6
+ V3    1    0.0319 19.947 -1614.1
+ V2    1    0.0179 19.961 -1613.8
- V9    1    0.2914 20.270 -1610.0
- V4    1    0.3624 20.341 -1608.2
- V5    1    0.3971 20.376 -1607.4
- V12   1    0.7468 20.726 -1598.8
- V6    1    0.7824 20.761 -1597.9
- V10   1    0.8916 20.871 -1595.2
- V11   1    1.1229 21.102 -1589.7
- V8    1    1.2990 21.278 -1585.4
- V13   1   12.1901 32.169 -1376.3
```

Example: Boston Housing Data (Final Model)

R Code: summary(stepwise)

```
Call:
lm(formula = V14 ~ V13 + V11 + V12 + V6 + V8 + V10 + V4 + V5 + V9, data = xt)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.01206 -0.10847 -0.00662  0.11665  0.77767 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.158186   0.362788  11.462  < 2e-16 ***
V13         -0.258773   0.014875 -17.396  < 2e-16 ***
V11         -0.041047   0.007774  -5.280 1.94e-07 ***
V12          0.048139   0.011180   4.306 2.01e-05 ***
V6           0.466807   0.105917   4.407 1.28e-05 ***
V8          -0.185537   0.032671  -5.679 2.31e-08 ***
V10         -0.209594   0.044550  -4.705 3.30e-06 ***
V4           0.108655   0.036227   2.999  0.00284 ** 
V5          -0.305541   0.097307  -3.140  0.00179 ** 
V9           0.049184   0.018286   2.690  0.00739 ** 
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 0.2007 on 496 degrees of freedom
Multiple R-squared:  0.7632,	Adjusted R-squared:  0.7589 
F-statistic: 177.6 on 9 and 496 DF,  p-value: < 2.2e-16
```