Ch2 Regression

Contents

Textbook: James et al. ISLR 2ed.

2 Subsection

[ToC]

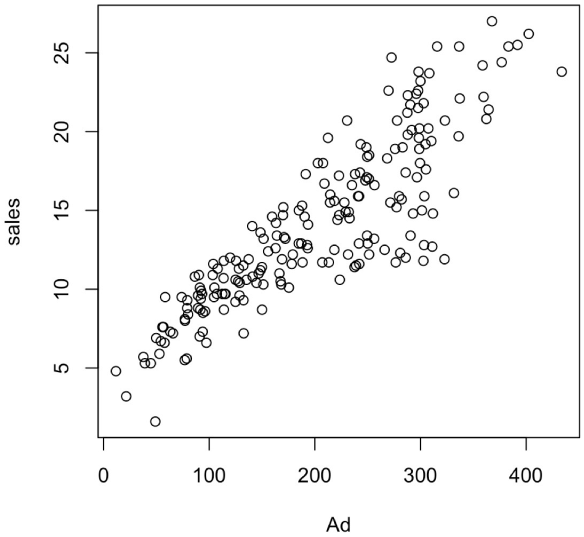
A.1 Sinple Linear Regression (Advertising Data)

```
Advertising.csv from ISLR web site. 4 variables are recorded.
Sales
TV
radio
newspaper
```

- Let Ad = TV + radio + newspaper (Total Spending)
- Then we are thinking:

$$Sales = f(Ad)$$

A.2 Scatterplot



A.3 Questions

1. Is there a relationship between advertising budget and sales?
2. If so, what is the form of the relationship?
3. How strong is the relationship between advertising budget and sales?
4. How accurately can we estimate the effect?
5. How accurately can we predict future sales?

Simple Linear Regression

A.5 SLR

- There must be a relationship between Y and X.

$$Y = f(X) + \epsilon$$

- Assumes that $f(\cdot)$ is a linear function, and there is an normally distributed error term

$$Y = \beta_0 + \beta_1 X + \epsilon, \qquad \epsilon \sim N(0, \sigma^2)$$

- Estimate parameters β_0 and β_1 by minimizing sum of squared residuals (**Residual Sum of Squares**)

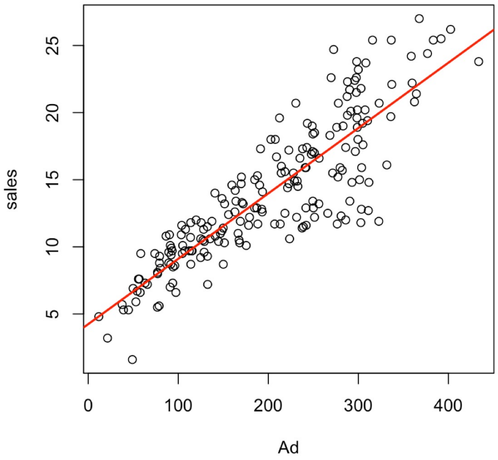
$$\text{RSS} = \sum_{i=1}^{n} (Y_i - \hat{Y})^2$$

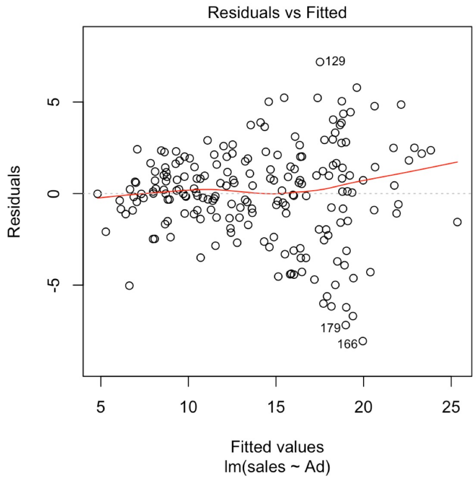
  It is called **ErrorSS** in some books.

- Estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are best estimators possible. (Minimum Variance Unbiased Estimators)

- Formula for $\hat{\beta}_1$ and $\hat{\beta}_0$:

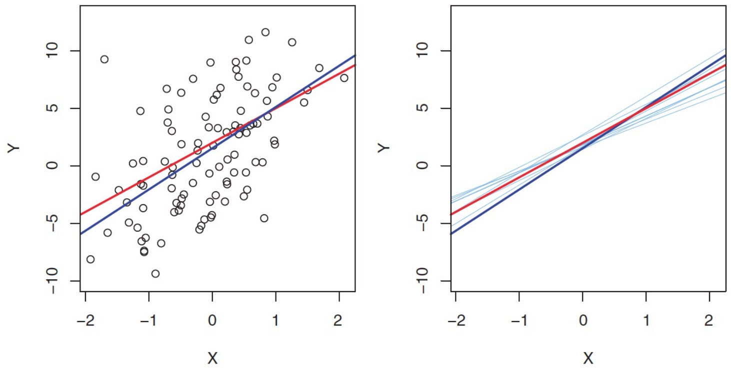
$$\hat{\beta}_1 = r s_y / s_x \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$





- is assumption true?
- how off are the estimators?
- how accurate is the prediction?

A.6 Simulation




```
X = rnorm(30, 3, 5)
Y = 4+1.5* X + rnorm(30, 0, 5)

plot(X, Y, xlim=c(-6, 11), ylim=c(-10, 25))
abline(a=4, b=1.5, col="blue", lwd=2)
m1 <- lm(Y~X)
abline(m1, col="red")
```

A.7 Reducible vs Irreducible Error

- True relationship

$$Y = f(X) + \epsilon$$

- We are trying to estimate f
- Reducible Error:

$$|f(X) - \hat{f}(X)|$$

- Irreducible Error:

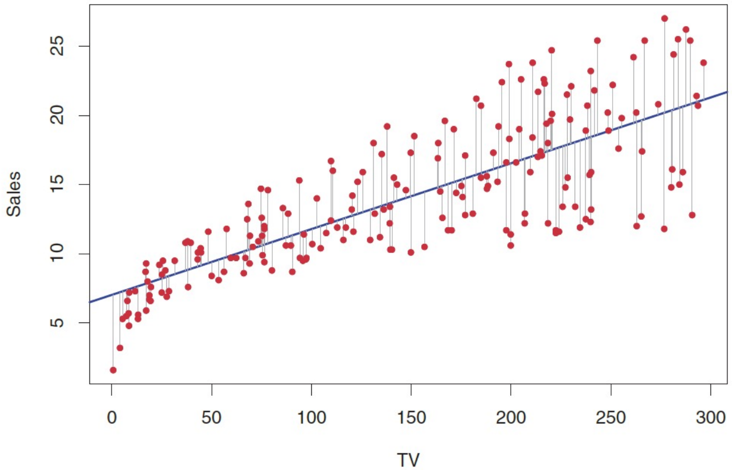
$$\epsilon$$

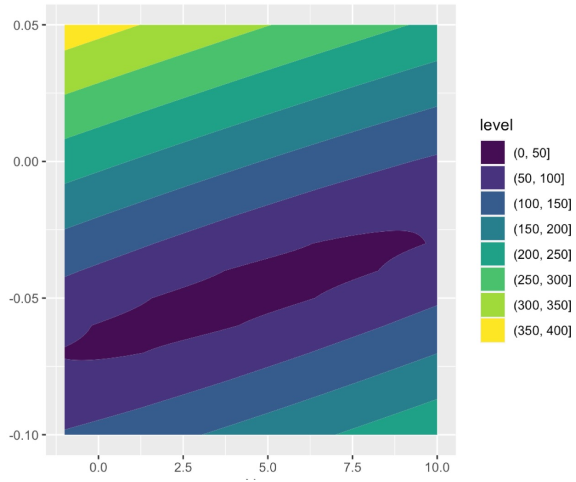
A.8 Minimizing RSS

- $$\text{RSS} = \sum_{i=1}^{n} (Y_i - \hat{Y})^2$$

- $\hat{\beta}_0$ and $\hat{\beta}_1$ were calculated by minimizing RSS analytically.

$$\hat{\beta}_1 = r s_y / s_x \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$


Sales
TV
0
50
100
150
200
250
300
5
10
15
20
25


level
(0, 50]
(50, 100]
(100, 150]
(150, 200]
(200, 250]
(250, 300]
(300, 350]
(350, 400]
0.05
0.00
-0.05
-0.10
0.0
2.5
5.0
7.5
10.0

A.9 MULTIPLE LINEAR REGRESSION

Multiple Linear Regression

A.10 MLR

- Want to guess the next Y as accurate as possible

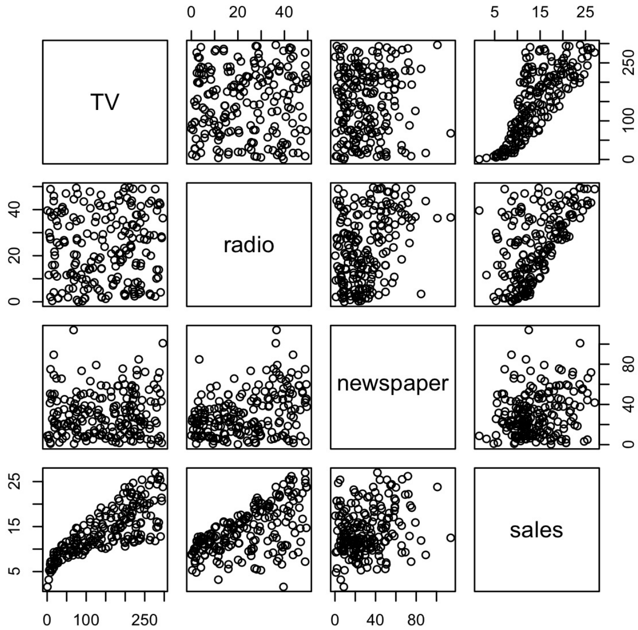
$$\texttt{sales} = \beta_0 + \beta_1\,\texttt{TV} + \beta_2\,\texttt{radio} + \beta_3\,\texttt{newspaper} + \epsilon, \qquad \epsilon \sim N(0, \sigma^2)$$

- Estimate parameters by minimizing

$$\text{RSS} = \sum_{i=1}^{n}(Y_i - \hat{Y})^2$$

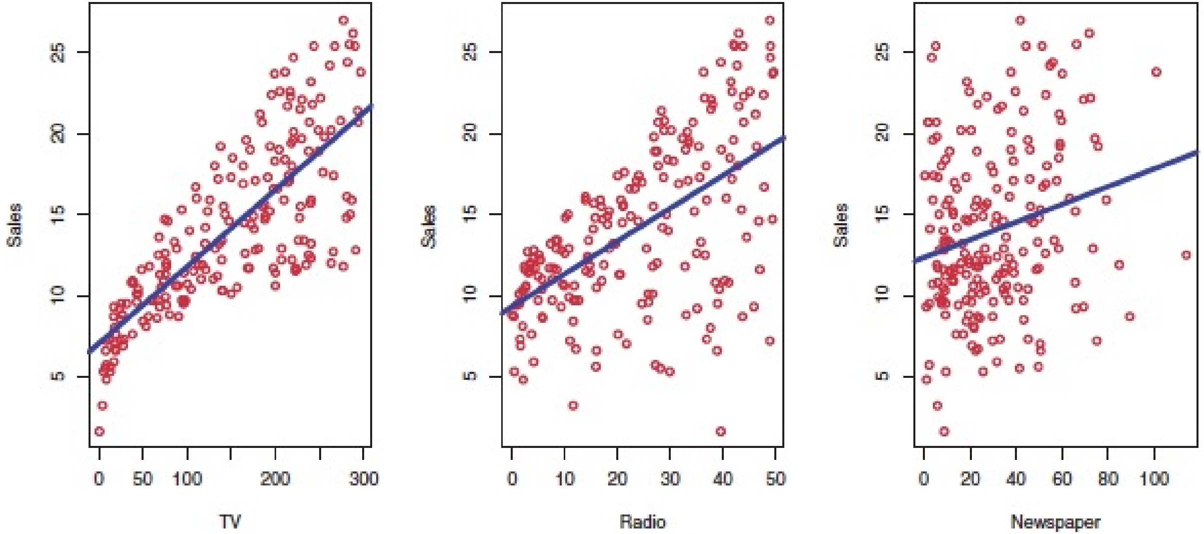
- Formula for $\hat{\boldsymbol{\beta}} = (\beta_0, \beta_1, \beta_2, \beta_3)'$:

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


TV
radio
newspaper
sales
0
20
40
5
15
25
250
100
0
40
20
0
80
40
0
25
15
5
0
100
250
0
40
80

`pairs()`

A.11 Last row

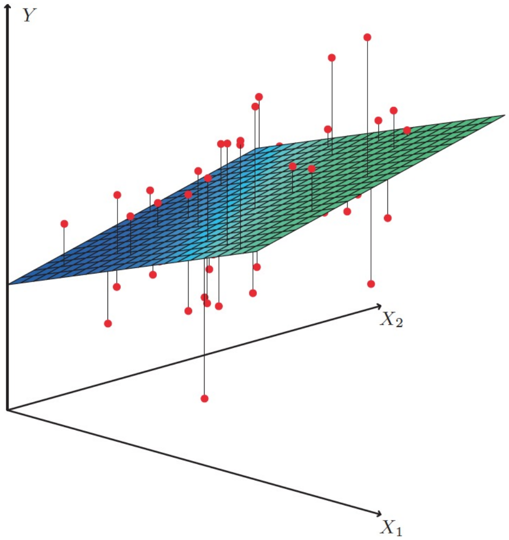

Sales
25
20
15
10
5
0
50
100
200
300
TV
Sales
25
20
15
10
5
0
10
20
30
40
50
Radio
Sales
25
20
15
10
5
0
20
40
60
80
100
Newspaper

```
Model2 = lm(sales ~ TV + radio + newspaper)
summary(Model2)

    Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
    (Intercept)  2.938889   0.311908   9.422   <2e-16 ***
    TV           0.045765   0.001395  32.809   <2e-16 ***
    radio        0.188530   0.008611  21.893   <2e-16 ***
    newspaper   -0.001037   0.005871  -0.177     0.86
    ---
    Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Residual standard error: 1.686 on 196 degrees of freedom
    Multiple R-squared:  0.8972,Adjusted R-squared:  0.8956
    F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

A.12 Model Without Newspaper

```
lm(formula = sales ~ TV + radio)

    Coefficients:
                Estimate Std. Error t value Pr(>|t|)
    (Intercept)  2.92110    0.29449   9.919   <2e-16 ***
    TV           0.04575    0.00139  32.909   <2e-16 ***
    radio        0.18799    0.00804  23.382   <2e-16 ***
    ---
    Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Residual standard error: 1.681 on 197 degrees of freedom
    Multiple R-squared:  0.8972,Adjusted R-squared:  0.8962
    F-statistic: 859.6 on 2 and 197 DF,  p-value: < 2.2e-16
```

With only TV and Radio as predictor It's like fitting plane in 3-d space



A.13 The Marketing Questions

1. Is there a relationship between advertising sales and budget?
2. How strong is the relationship?
3. Is it important to advertise in newspaper?
4. Is all predictor important, or just a subset?
5. Which media contribute to most to the sales? How much?
6. How accurately can we predict future sales?
7. Is the relationship linear?
8. Is there synergy among the advertising media?

A.14 Q1. Is there a relationship between advertising sales and budget?

- At least one X useful?
- In SLR, we only need to test $\beta_1 = 0$.
- For MLR, we have to test $\beta_1 = \beta_2 = \beta_3 = 0$.
- Use F-statistic

$$F = \frac{(TSS - RSS)/p}{RSS/(n - p - 1)}$$

and p is the number of variables (excludng intercept).

```
summary(Model2)

# Residual standard error: 1.686 on 196 degrees of freedom
# Multiple R-squared:  0.8972,Adjusted R-squared:  0.8956
# F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

- We can test SUBSET of parameters ($\beta_2 = \beta_3 = 0$) by

$$F = \frac{(RSS_0 - RSS)/q}{RSS/(n-p-1)}$$

  Where RSS_0 is the RSS from the model using $\beta_2 = \beta_3 = 0$, and q is the number of suppressed parameters ($q = 2$ in this case).

- Why test as a whole when you can do the t-test individually? (important when p is large)

A.15 Q2. How strong is the relationship?

How good is Model Fit?

- Coefficient of Determination

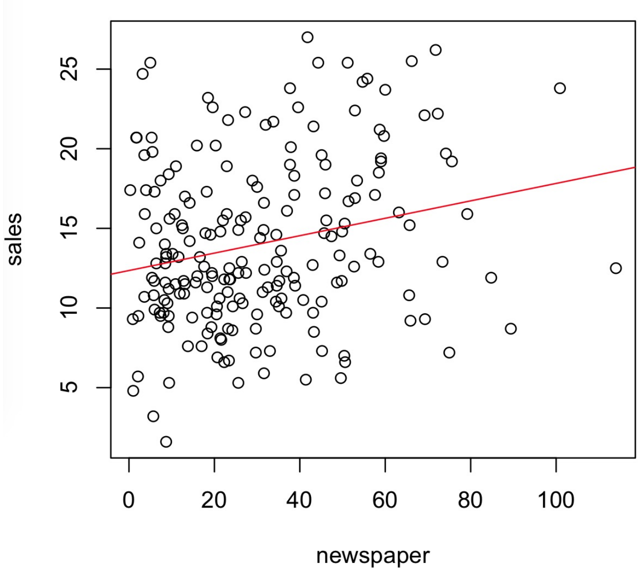
$$\text{TSS} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2, \qquad \text{RSS} = \sum_{i=1}^{n}(Y_i - \hat{Y})^2, \qquad R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$$

- $R^2 = .89719$ without `newspaper`
- $R^2 = .8972$ with `newspaper`
- MSE (RSE in ISLR) estimates σ^2 and represents irreducible error.
- With p predictors,

$$MSE = \sqrt{\frac{1}{n-p-1}RSS}$$

A.16 Q3. newspaper? (Confounding Effect)

```
Model3 = lm(sales ~ newspaper)
summary(Model3)

    Coefficients:
                Estimate Std. Error t value Pr(>|t|)
    (Intercept) 12.35141    0.62142   19.88  < 2e-16 ***
    newspaper    0.05469    0.01658    3.30  0.00115 **

    Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Residual standard error: 5.092 on 198 degrees of freedom
    Multiple R-squared:  0.05212,Adjusted R-squared:  0.04733
    F-statistic: 10.89 on 1 and 198 DF,  p-value: 0.001148
```



sales
newspaper
0
20
40
60
80
100
5
10
15
20
25

```
cor(Adv)   # correlation matrix of each column

          TV         radio      newspaper  sales
TV        1.00000000 0.05480866 0.05664787 0.7822244
radio     0.05480866 1.00000000 0.35410375 0.5762226
newspaper 0.05664787 0.35410375 1.00000000 0.2282990
sales     0.78222442 0.57622257 0.22829903 1.0000000
```

- Mutiple Regression suggests `newspaper` has no effect
- In Simple Regression, `newspaper` gets credit through `radio` because of the correlation.
- Other examples of confounding variables (lurking variables) includes: (shark attack vs ice cream sales, vocabulary score vs num of cavity)

A.17 Q4. All predictors or just a few?

- In general, have to try out many models, and use some kind of criteria to pick the best
- Mallow's C_p, AIC, BIC, Adjusted R^2. (more in Ch6)
- There's 2^p models with p predictors. $2^3 = 8, 2^{10} = 1024, 2^{30} = 1,073,741,824$.
- Forward, Backward, Mixed selection

A.18 Q5. Effect of each medium?

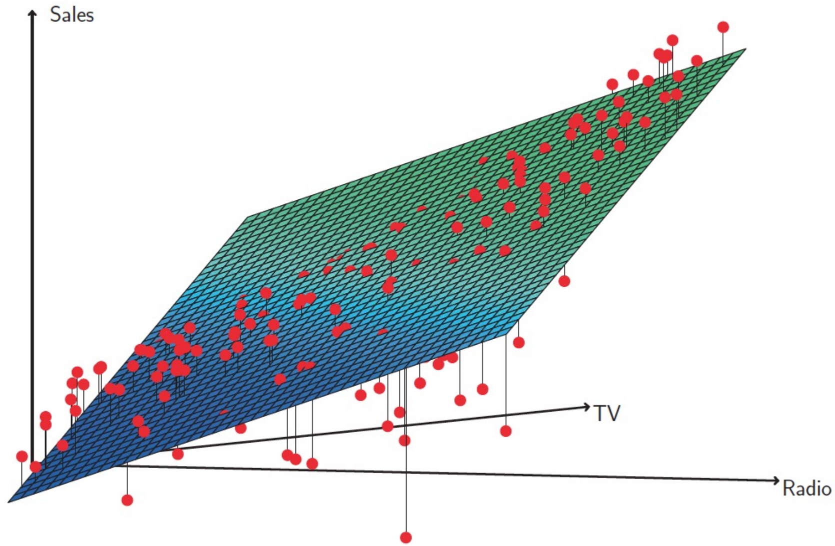
We can construct CI for parameters.

```
For the Advertising data, the 95\% CI are:
(0.043, 0.049) for TV,
(0.172, 0.206) for radio,
(-0.013, 0.011) for newspaper.

    Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
    (Intercept)  2.938889   0.311908   9.422   <2e-16 ***
    TV           0.045765   0.001395  32.809   <2e-16 ***
    radio        0.188530   0.008611  21.893   <2e-16 ***
    newspaper   -0.001037   0.005871  -0.177     0.86

    Residual standard error: 1.686 on 196 degrees of freedom
```

A.19 Q6. Prediction Accuracy?

- We can get $\hat{f}(X)$ using estimated β_i.
- There could be model bias
- Get CI for parameters, and PI for predictions

```
Residual standard error: 1.686 on 196 degrees of freedom
```


Sales
TV
Radio

```
newAdv = data.frame(TV=c(50, 60), radio=c(20, 10), newspaper=c(0, 0))
newAdv
   TV radio newspaper
 1 50    20         0
 2 60    10         0

predict(Model2, newdata=newAdv, interval="confidence")
        fit      lwr      upr
 1 8.997722 8.515752 9.479692
 2 7.570068 7.099337 8.040800

predict(Model2, newdata=newAdv, interval="prediction")
        fit      lwr      upr
 1 8.997722 5.638898 12.35655
 2 7.570068 4.212838 10.92730
```

A.20 Q7. Is the relationship linear?

- If the relationships are linear, then the residual plots should display no pattern. Needs transformation?

A.21 Q8. Is there synergy among the advertising media?

- The standard linear regression model assumes an additive relationship between the predictors and the response.
- Including an interaction term in the model results in a substantial increase in R^2, from around 90% to almost 97

A.22 Inference vs Prediction

- Inference
  Which media contribute to sales?
  Which media generate the biggest boost in sales?
  How much increase in sales is associated with a given increase in TV advertising?

- Prediction
  If we spend \$A on TV ads and \$B on Radio ads, how much Sales would we get?

A.23 Flexibility and Interpretability trade-off