NOTES 4

Linear Regression

Acknowledgement: some of the contents are borrowed with or without modification from *An Introduction to Statistical Learning, with applications in R* (Springer, 2013) with permission from the authors: G. James, D. Witten, T. Hastie and R.Tibshirani.

Linear Regression

- Linear regression is a simple approach to supervised learning. It assumes that the dependence of Y on (X_1, X_2, $X_{3...}$) is linear.

- (X_1, X_2, $X_{3...}$) are called as inputs, in dependent variables or predictors. They are the variables we use to estimate Y.

- Y is the response, dependent variable, or outcome. It's the variable we want to predict.

- All the predictors can be denoted as a single input vector X = (X_1, X_2, $X_{3...}$). So the model can be denoted as:

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

where ϵ is a mean-zero random error term, typically independent of X

We Can Use a Linear Model to

- Predict the values of Y with new *X*.

- Find out which variables in (X_1, X_2, $X_{3\ldots}$) help predict Y and how?
  - Variables may or may not significantly affect Y
  - Variables may positively or negatively affect Y

- So how to develop a linear model?

① Simple Linear Model

- Use one single predictor

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

β_0

where β_0 and β_1 are unknown constants that represent the *intercept* and *slope*, also known as *coefficients* or *parameters*, and ϵ is the error term.

β_1

- Given some estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ for the model coefficients, we predict Y using

predict value

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x,$$

where $\hat{y}$ indicates a prediction of Y on the basis of $X = x$. The *hat* symbol denotes an estimated value.

Which one has a larger impact on Sales ?

An Example---Advertising.csv

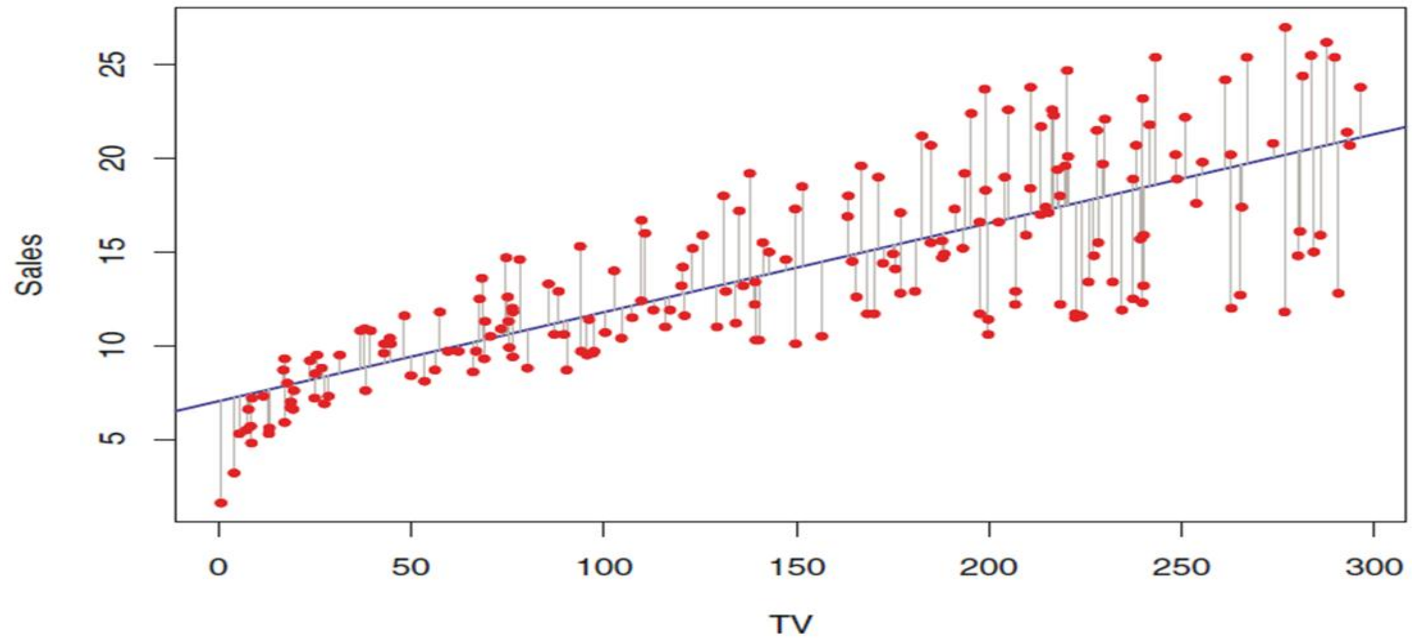
- The sales of one product in 200 different markets, along with advertising budgets for the product in each of those markets for three different media: TV, radio, and newspaper

TV	Radio	Newspaper	Sales
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	9.3
151.5	41.3	58.5	18.5
180.8	10.8	58.4	12.9
8.7	48.9	75	7.2
⋮	⋮	⋮	⋮

The sales is in thousands of units, and the TV is in thousands of dollars

Q1: Is there a relationship between advertising budget and sales? If so, how strong is it?

Only consider one media, TV



- Develop a model that describes Sales as a function of the TV advertising budget. Sales≈ f (TV)

$$\texttt{sales} = \beta_0 + \beta_1 \times \texttt{TV} + \epsilon$$

Coefficients Estimation

- We have training data

$$(x_1, y_1),\ (x_2, y_2), \ldots,\ (x_n, y_n)$$

- Estimates coefficients by finding the coefficient values that *minimize* the residual sum of squares (RSS) $|\hat{y}_i - y_i|$

- Residual is the difference between the observed response value and the predicted response value

i is the observation index $\boxed{e_i = y_i - \hat{y}_i}$

- $\mathrm{RSS} = e_1^2 + e_2^2 + \cdots + e_n^2$ which is equal to

$$\mathrm{RSS} = (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \ldots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2$$

R raw results

```
> summary(M1)

Call:
lm(formula = Sales ~ TV)

Residuals:
    Min      1Q  Median      3Q     Max
-8.3860 -1.9545 -0.1913  2.0671  7.2124

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 7.032594   0.457843   15.36   <2e-16 ***
TV          0.047537   0.002691   17.67   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.259 on 198 degrees of freedom
Multiple R-squared: 0.6119,     Adjusted R-squared: 0.6099
F-statistic: 312.1 on 1 and 198 DF,  p-value: < 2.2e-16
```

reference

if P > 1 in this case,
⇒ insignificant

It's significant when we set significant level as ______ / given P value

※ P value level varies among different industries

Results of Sales~TV

- Using R, we get

	Coefficient	Std.Error	t-statistic	p-value
Intercept	7.033	0.458	15.36	< 0.0001
TV	0.048	0.003	17.67	< 0.0001

- For every $1000 spent on TV advertising, sales on average increase by (1000 × 0.048) = 48 units
- If the p-value is small, the coefficient is likely to be significant
  - a small p-value indicates that there is an association between the predictor and the response, that is $\beta_1 \neq 0$.
  - How to define "small"?
    - Usually, a significant level at least 0.05 is taken, 0.001 is preferred.
    - If the p-value is above 0.05, we can not reject the hypothesis that $\beta_1 = 0$

Assess the Model Performance

- Residual standard error: measures a lack of fit

$$\boxed{\mathrm{RSE} = \sqrt{\frac{1}{n-2}\mathrm{RSS}} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}$$

A smaller RSE is preferred

- R^2 Statistics: the proportion of variance explained
  - scale independent, always in [0,1] $\boxed{RSS = e_1^2 + e_2^2 + \dots + e_n^2}$

$$\boxed{R^2 = \frac{\mathrm{TSS} - \mathrm{RSS}}{\mathrm{TSS}} = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}}}$$

A larger R^2 value is preferred

x_i

If $RSS \to 0$
$R^2 \to 1$
Then model is 100% fitch the data

  - where TSS (*total sum of squares*) $\mathrm{TSS} = \sum(y_i - \bar{y})^2$
  - the proportion of variability in Y that can be explained using X

Results of Sales~TV

Measure	Value
Residual standard error	3.26
R^2	0.612

predicted

- It indicates that actual sales in each market deviate from the true regression line by approximately 3,260 units, on average.

?

61.2% ??

- Under two-thirds of the variability in sales is explained by this linear regression on TV.

- A larger R^2 value is usually preferred.

Multiple Linear Model

- The model f(X) is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

- We interpret β_j as the average effect on Y of a one unit increase in X_j , holding all other predictors fixed.

- Minimizing RSS to get all coefficients

$$\begin{aligned} \mathrm{RSS} &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})^2. \end{aligned}$$

The Advertising Example

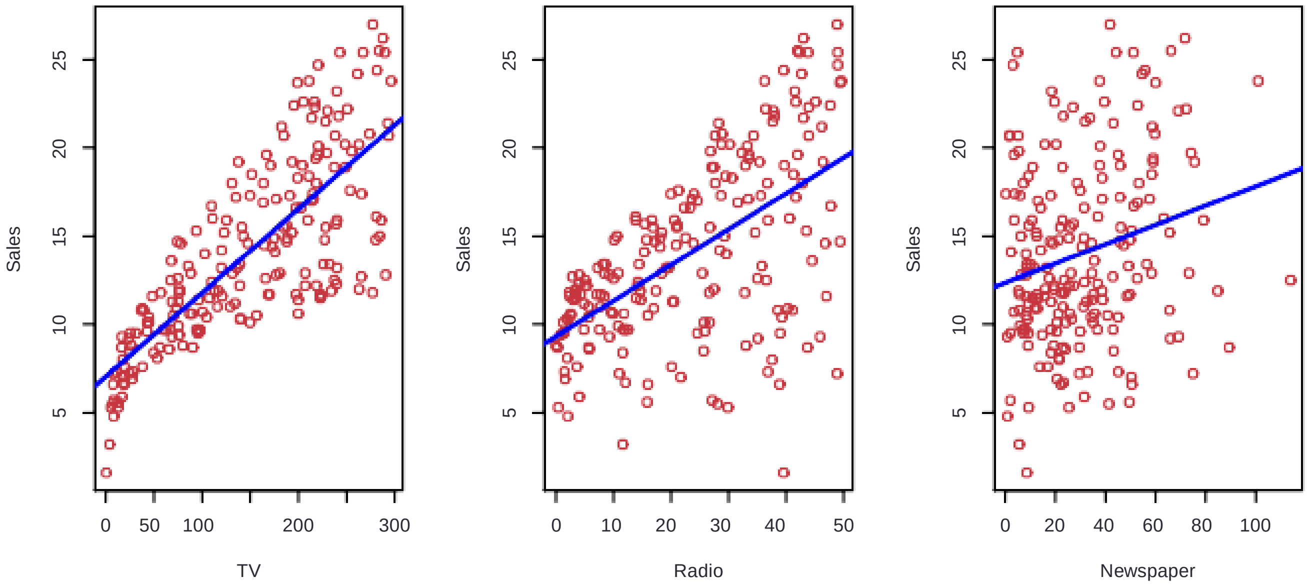


Figure : 2.1 *Y* = Sales plotted against TV, Radio and Newspaper advertising budgets.

Simple regression of sales on radio

	Coefficient	Std. error	t-statistic	p-value
Intercept	9.312	0.563	16.54	< 0.0001
radio	0.203	0.020	9.92	< 0.0001

significant & positive

Simple regression of sales on newspaper

	Coefficient	Std. error	t-statistic	p-value
Intercept	12.351	0.621	19.88	< 0.0001
newspaper	0.055	0.017	3.30	< 0.0001

The Advertising Example Continued

- In the advertising example, the model becomes

$$\text{sales} = \beta_0 + \beta_1 \times \text{TV} + \beta_2 \times \text{radio} + \beta_3 \times \text{newspaper} + \epsilon$$

- The coefficient β_1 tells us the expected change in sales per unit change of the TV budget, with all other predictors held fixed.

- With R

```
M2=lm(Sales~TV+Radio+Newspaper)
summary(M2)
```

R raw results

```
> summary(M2)

Call:
lm(formula = Sales ~ TV + Radio + Newspaper)

Residuals:
    Min      1Q  Median      3Q     Max
-8.8277 -0.8908  0.2418  1.1893  2.8292

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
```

negative

$> 0.05 \Rightarrow$ insignificant

why?

Results of Sales~TV+Radio+Newspaper

	Coefficient	Std. Error	t-statistic	p-value
Intercept	2.939	0.3119	9.42	< 0.0001
TV	0.046	0.0014	32.81	< 0.0001
radio	0.189	0.0086	21.89	< 0.0001
newspaper	-0.001	0.0059	-0.18	0.8599

- Are the simple and the multiple regression coefficients the same?

No.

???

Correlation

Eg: smoking is correlated with alcoholism, but doesn't cause alcoholism.
While smoking causes a increase in the risk of developing lung cancer.

- The correlation between two variables is calculated as:

Correlation (relationship)
≠ Causation
(Cause & Effect)

$$\mathrm{Cor}(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}},$$

The impact from radio makes newspaper look significant in simple linear model.

- With R cor(Ad)

	TV	radio	newspaper	sales
TV	1.000	0.055	0.057	0.782
radio		1.000	0.354	0.576
newspaper			1.000	0.228
sales				1.000

- newspaper gets “credit” for the effect of radio on sales

Model Fit: Advertising Example

- Is at least one of the predictors $X_1, X_2, \ldots, X_p$ useful in predicting the response?

- Hypothesis Testing:
  - Null Hypothesis: H_0: $\beta_0 = \beta_1 = \ldots \beta_p = 0$ H_a: at least one is not 0
  - Use F statistics. To reject H_0, F should be greater than 1

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

- Model Fit

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

Quantity	Value
Residual standard error	1.69
R^2	0.897
F-statistic	570

Subset Selection

- Do all predictors help to explain Y , or is only a subset of the predictors useful?

- Subset Selection: identify a subset of the p predictors that are related to the response. Fit a model using least squares on the reduced set of variables.
  - Best Subset Selection: computational extensive
  - Stepwise Selection
    - Forward
    - Backward
    - Mix: Go Forward, any time try to eliminate useless predictor

Best Subset Selection

- Fit a separate least squares regression for each possible combination of the *p* predictors and pick up the best one

- Let M_0 denote the *null model*, which contains no predictors. This model simply predicts the sample mean for each observation.

- For $k = 1, 2, \ldots, p$:
  - (a) Fit all $\binom{p}{k}$ models that contain exactly *k* predictors.

  - (b) Pick the best among these $\binom{p}{k}$ models, and call it M_k. Here *best* is defined as having the smallest RSS, or equivalently largest R^2.

- Select a single best model from among $M_0, \ldots, M_p$ using cross validated prediction error, *Cp* (AIC), BIC, or adjusted R^2.

Stepwise Selection: Forward Selection

- Let M_0 denote the *null model*, which contains no predictors.
- For $k = 0, 1, \ldots p-1$:
  - (a) Consider all $p - k$ models that augment the predictors in M_k with one additional predictor
  - (b) Choose the *best* among these $p - k$ models, and call it M_{k+1}. Here *best* is defined as having the smallest RSS, or equivalently largest R^2.
- Select a single best model from among $M_0, \ldots, M_p$ using cross validated prediction error, Cp (AIC), BIC, or adjusted R^2.

Stepwise Selection: Backward Selection

- Let M_p denote the *full model*, which contains all p predictors.

- For $k = p, p\text{-}1, \ldots 1$:

  - (a) Consider all k models that consider all but one of the predictors in M_k, for a total of k-1 predictor

  - (b) Choose the *best* among these k models, and call it M_{k-1}. Here *best* is defined as having the smallest RSS, or equivalently largest R^2.

- Select a single best model from among $M_0, \ldots, M_p$ using cross validated prediction error, *Cp* (AIC), BIC, or adjusted R^2.

Model Selection Criteria

- The model containing all of the predictors will always have the smallest RSS and the largest R^2, since these quantities are related to the training error.

- We wish to choose a model with low test error, not a model with low training error.

- RSS and R^2 are not suitable for selecting the best model among a collection of models with different numbers of predictors.

use adjusted R^2

- We can indirectly estimate test error by making an *adjustment* to the training error to account for the bias due to overfitting.

Overfitting

- When a given method yields a small training MSE but a large test MSE, we are said to be *overfitting* the data.

- Mean Square Error (MSE):a measure of the overall model fit

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2,$$

C_p, AIC, BIC, and Adjusted R

- *Mallow's C_p*

$$C_p = \frac{1}{n}\left(\text{RSS} + 2d\hat{\sigma}^2\right)$$

- Akaike information criterion (AIC)

$$\text{AIC} = \frac{1}{n\hat{\sigma}^2}\left(\text{RSS} + 2d\hat{\sigma}^2\right)$$

- Bayesian information criterion (BIC)

$$\text{BIC} = \frac{1}{n}\left(\text{RSS} + \log(n)d\hat{\sigma}^2\right)$$

  - Since $\log n > 2$ for any $n > 7$, it generally places a heavier penalty on models with many variables

if $d \uparrow$, $R^2 \downarrow$
prefer a larger R^2
$\rightarrow$ prefer fewer predictors

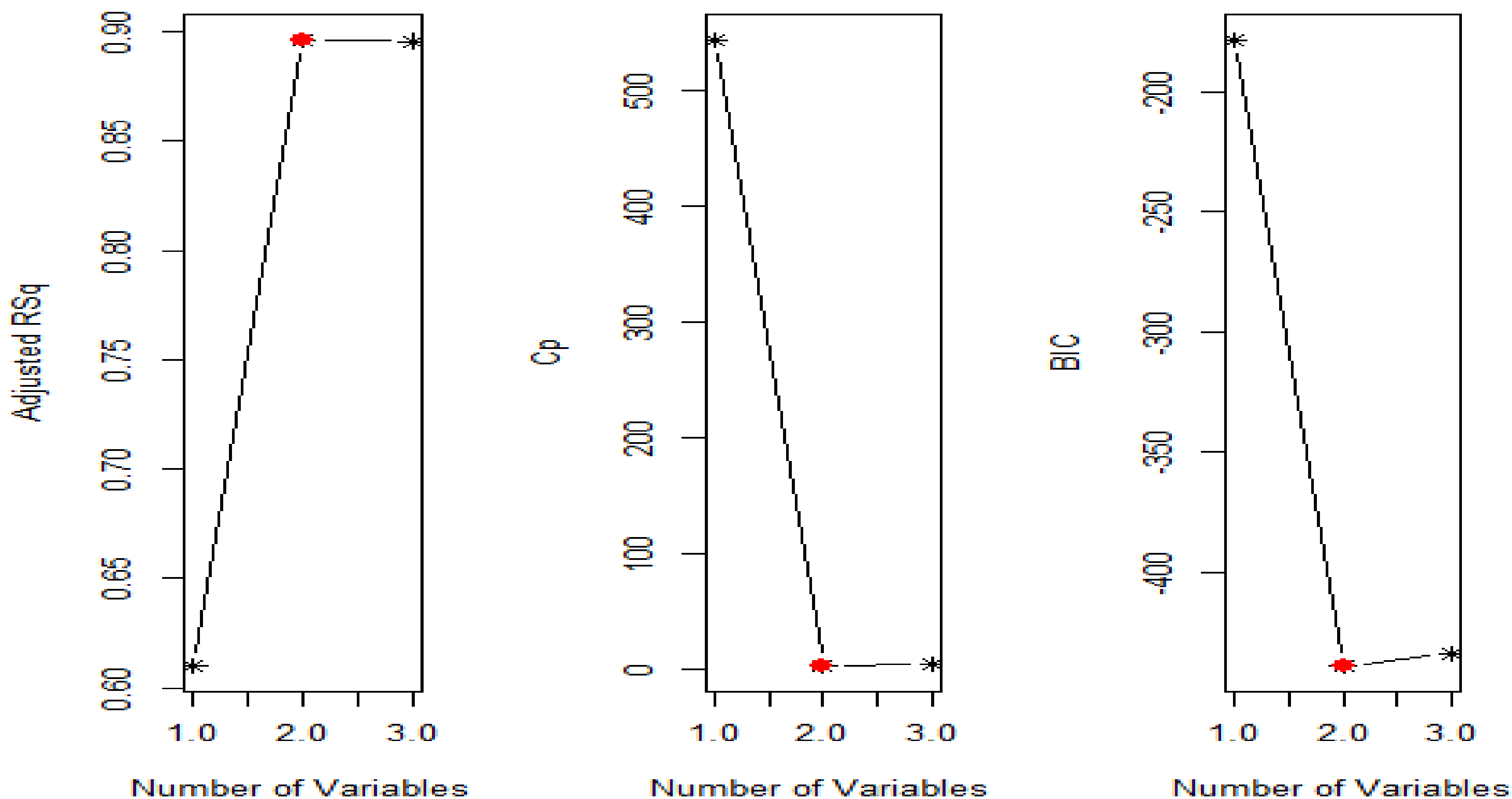
- *Adjusted R^2*

$$\text{Adjusted } R^2 = 1 - \frac{\text{RSS}/(n-d-1)}{\text{TSS}/(n-1)}$$

$\hat{\sigma}^2$ is an estimate of the variance of the error associated with each response measurement, d is the total number of parameters

R OUTPUT

```
> coef(reg.best,2)
(Intercept)          TV       Radio
 2.92109991  0.04575482  0.18799423
```



Other Practical Issues

- Interaction effect
- Model assumption diagnostics
- Collinearity
- Non-linearity
- Qualitative Predictors

Interaction

- Suppose that spending money on radio advertising actually increases the effectiveness of TV advertising, so that the slope term for TV should increase as radio increases.

- In this situation, given a fixed budget of $100, 000, spending half on radio and half on TV may increase sales more than allocating the entire amount to either TV or to radio.

$$\texttt{sales} = \beta_0 + \beta_1 \times \texttt{TV} + \beta_2 \times \texttt{radio} + \beta_3 \times (\texttt{radio} \times \texttt{TV}) + \epsilon = \beta_0 + (\beta_1 + \beta_3 \times \texttt{radio}) \times \texttt{TV} + \beta_2 \times \texttt{radio} + \epsilon$$

With R

```
M3=lm(Sales~TV*Radio)
summary(M3)
```

```
> summary(M3)

Call:
lm(formula = Sales ~ TV * Radio)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.3366 -0.4028  0.1831  0.5948  1.5246 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 6.750e+00  2.479e-01  27.233   <2e-16 ***
TV          1.910e-02  1.504e-03  12.699   <2e-16 ***
Radio       2.886e-02  8.905e-03   3.241   0.0014 ** 
TV:Radio    1.086e-03  5.242e-05  20.727   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9435 on 196 degrees of freedom
Multiple R-squared: 0.9678,     Adjusted R-squared: 0.9673 
F-statistic:  1963 on 3 and 196 DF,  p-value: < 2.2e-16
```

Model Fit – Again!

- The results in this table suggests that interactions are important.

- The p-value for the interaction term TV×radio is extremely low, indicating that there is strong evidence $\beta_3 \neq 0$

- The R^2 for the interaction model is 96.8%, compared to only 89.7% for the model that predicts sales using TV and radio without an interaction term.

	Coefficient	Std.Error	t-statistic	p-value
Intercept	6.750	0.248	27.23	< 0.0001
TV	0.019	0.002	12.70	< 0.0001
radio	0.029	0.009	3.24	0.0014
TV×radio	0.001	0.000	20.73	< 0.0001

Results Interpretation

- The coefficient estimates suggest that an increase in TV advertising of $1,000 is associated with increased sales of

$$(\hat{\beta}_1 + \hat{\beta}_3 \times \text{radio}) \times 1000 = 19 + 1.1 \times \text{radio} \text{ units.}$$

- An increase in radio advertising of $1, 000 will be associated with an increase in sales of

$$(\hat{\beta}_2 + \hat{\beta}_3 \times \text{TV}) \times 1000 = 29 + 1.1 \times \text{TV} \text{ units.}$$

- Sometimes it is the case that an interaction term has a very small p-value, but the associated main effects (in this case, TV and radio) do not.

- If include an interaction in a model, also include the main effects, even if the p-values associated with their coefficients are not significant.

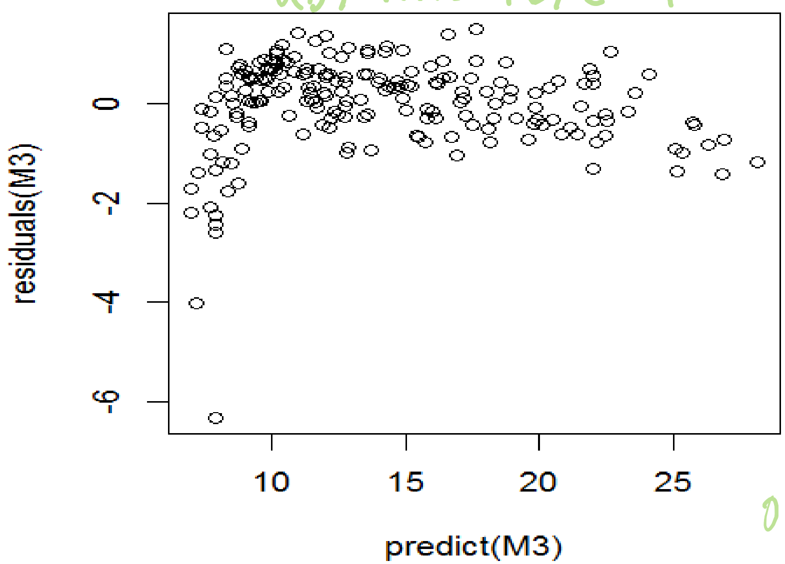
Outliers

- Can be identified using residual plots
- Plot the studentized residuals the predicted (fitted) values
- Rule of thumb: absolute value greater than 3 is an outlier
- Error in the dataset or missing predictor?

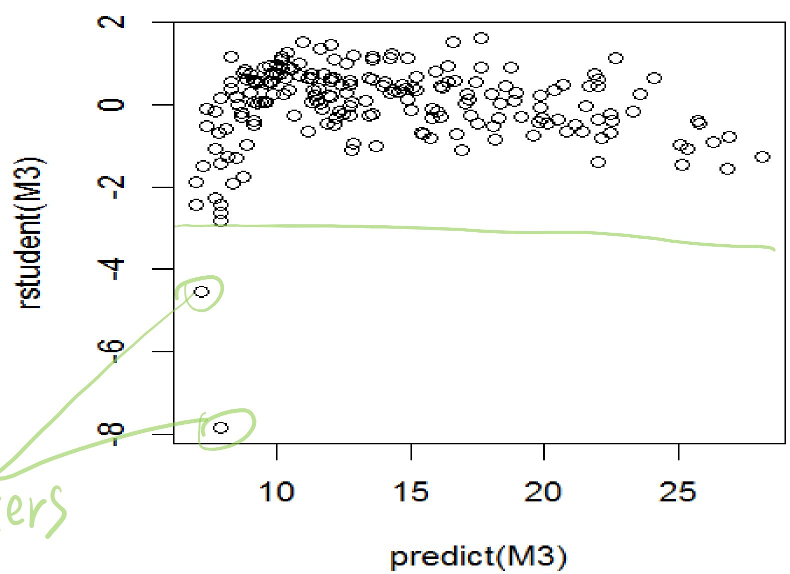
With R:

```
plot(predict(M3), residuals(M3))
plot(predict(M3), rstudent(M3))
```

absolute residual



outliers

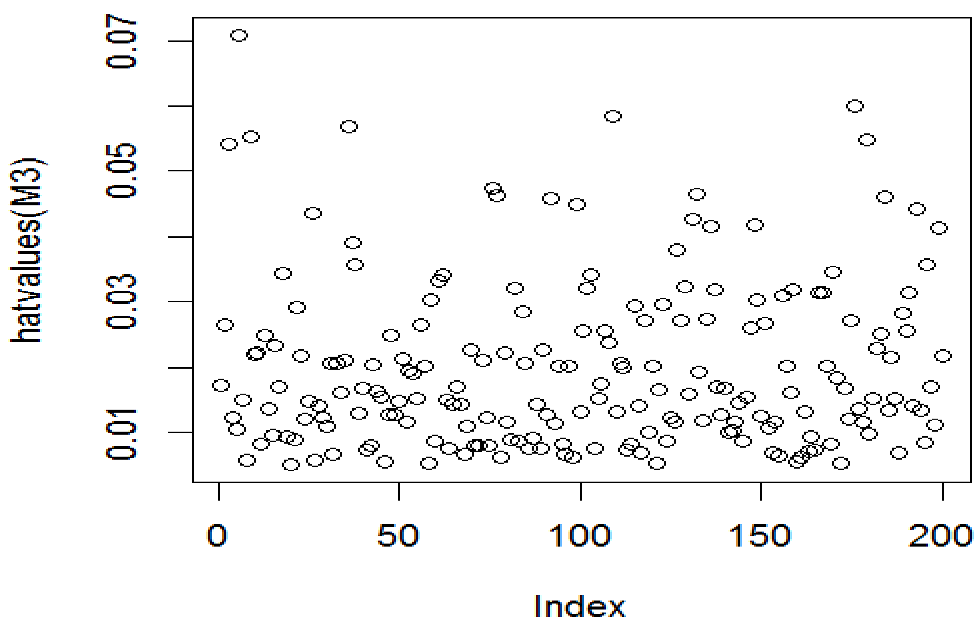


High leverage Points

- Can be identified using leverage statistic, usually should be noted if $\boxed{>2*(p+1)/n}$
- With R

number of predictors

```
plot(hatvalues(M3))
which.max(hatvalues(M3))
```



? ? ?

Non-constant Variance

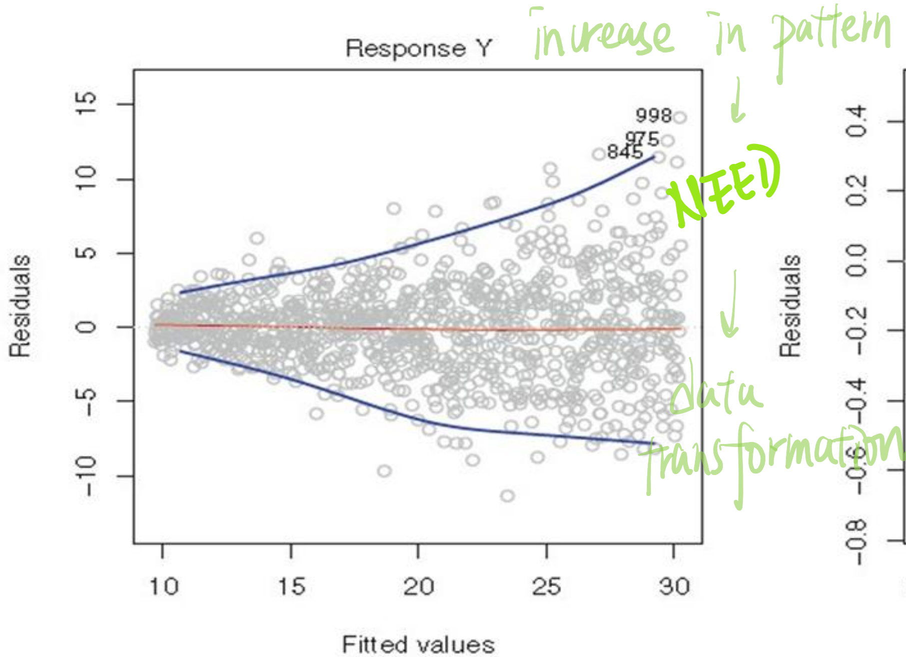


FIGURE 3.11. *Residual plots. In each plot, the red line is a smooth fit to the residuals, intended to make it easier to identify a trend. The blue lines track the outer quantiles of the residuals, and emphasize patterns.* Left: *The funnel shape indicates heteroscedasticity.* Right: *The predictor has been log-transformed, and there is now no evidence of heteroscedasticity.*

- In linear regression, we assume the error terms have a constant variance $\sigma^2 = \mathrm{Var}(\epsilon)$
- Consider data transformation

Collinearity

interrelation between 2 variables
↓ the accuracy of the model

- Two or more predictor variables are closely related to one another.
- Reduces the accuracy of the estimates of the regression coefficients
- Interpretations become hazardous: when Xj changes, everything else changes.

mathmatical way to define:

- Can be detected by *variance inflation factor* $$VIF(\hat{\beta}_j) = \frac{1}{1 - R^2_{X_j|X_{-j}}}$$

?

  - Smallest possible value is 1, which indicates no collinearity
  - Rule of thumb: greater than 5 or 10 indicates a problematic amount of collinearity

When do we need to use adjusted R^2
compare model with different num. of predictors.

```
library(car)
library(MASS)
library(nnet)
vif(M2)
```

```
> vif(M2)
      TV    Radio Newspaper
1.004611 1.144952  1.145187
```

Polynomial Model

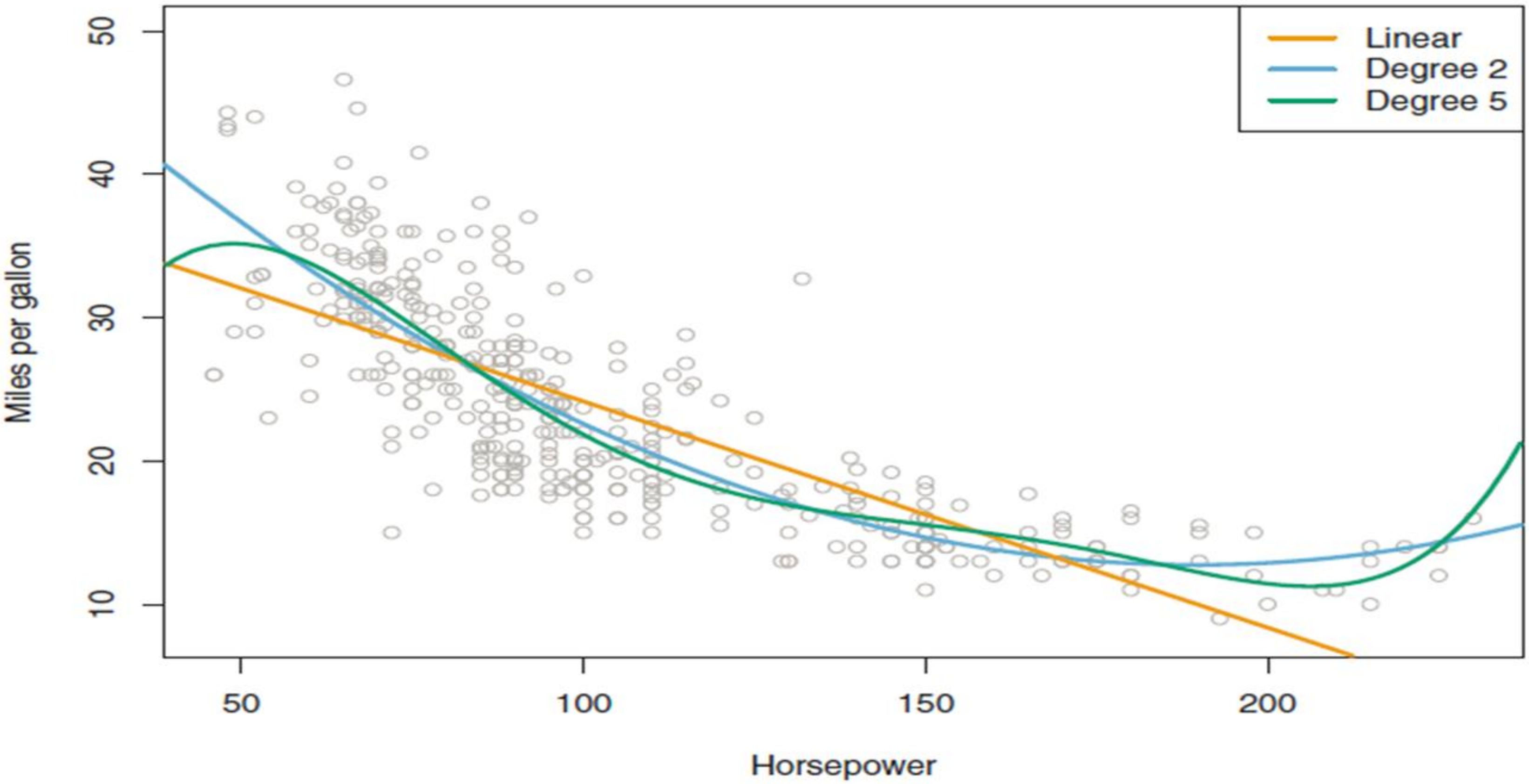


Figure 3.8 Polynomial regression model for Auto data. The linear regression fit for a model that includes horsepower (Orange), horsepower^2(Blue), horsepower^5 (Green).

?

Non-linear Relationship

The figure suggests that

$$mpg = \beta_0 + \beta_1 \times horsepower + \beta_2 \times horsepower^2 + \epsilon$$

may provide a better fit.

	Coefficient	Std.Error	t-statistic	p-value
Intercept	56.900	1.800	31.6	< 0.0001
horsepower	-0.466	0.031	-15.0	< 0.0001
horsepower^2	0.001	0.000	10.1	< 0.0001

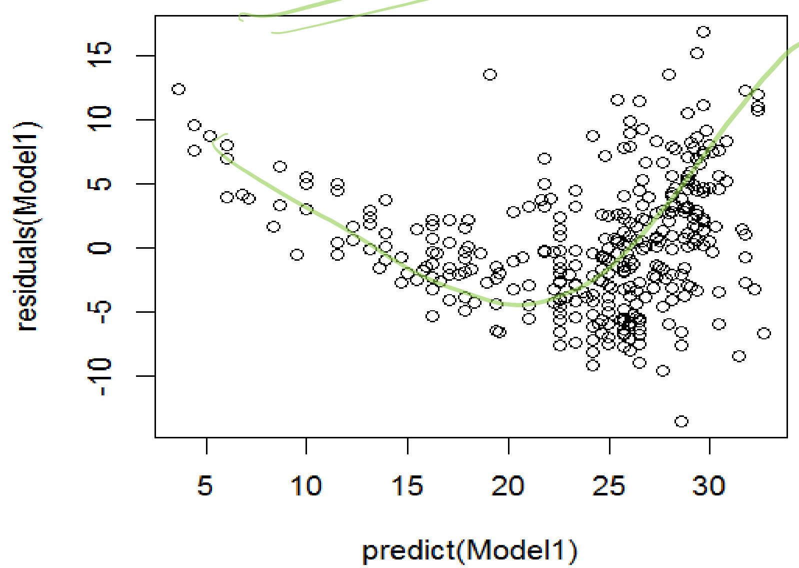
both are significant

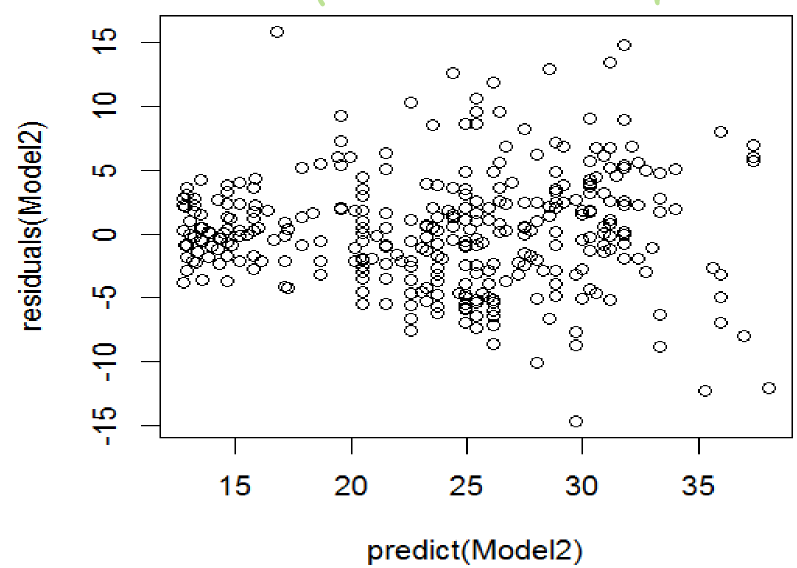
Non-linearity

- Residual plots can be used to identify non-linearity
- Plot the residuals versus the predicted (fitted) values
- Check if there is a presence of a pattern

If there's a curve in residule plot → non-linearity

(add the horsepower square)





Qualitative Predictors

- Some predictors are not quantitative but are qualitative, taking a discrete set of values.
- These are also called categorical predictors or factor variables. We can create a dummy variable to encode it.
- The level with no dummy variable is denoted as baseline

Have 1 categorical variables.

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is female} \\ 0 & \text{if } i\text{th person is male,} \end{cases}$$

$X_1 = \begin{cases} A & X_{11} \\ B & X_{12} \\ C & \end{cases}$ 0

~~1, 2, 3~~ — value differences.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is female} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is male.} \end{cases}$$

$\begin{cases} X_{11} & 1 & \\ X_{12} & 0 & 1 \end{cases}$
B C

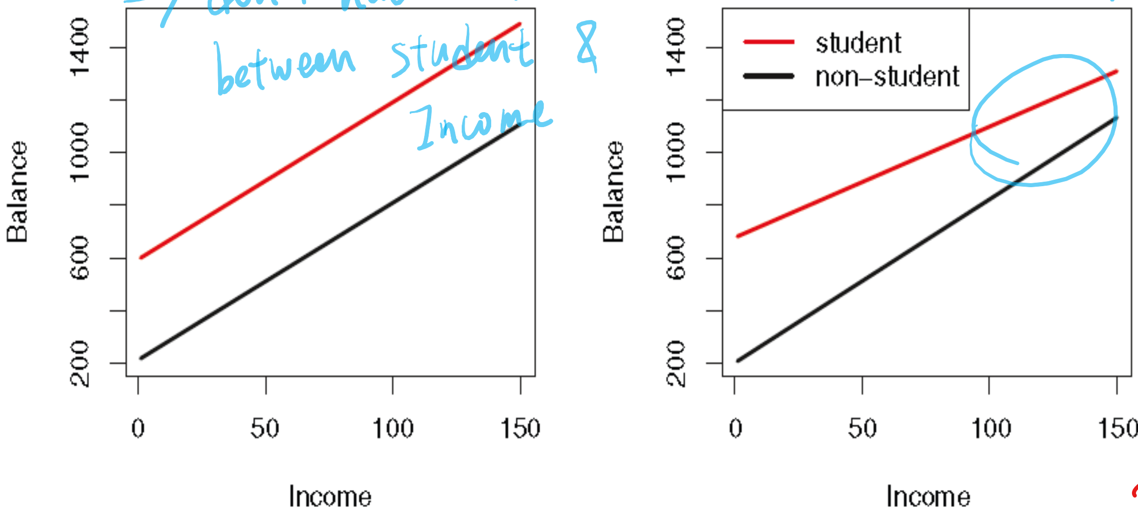
$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is Asian} \\ \beta_0 + \beta_2 + \epsilon_i & \text{if } i\text{th person is Caucasian} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is African American} \end{cases}$$

Interaction between Qualitative and Quantitative Variables

parallel ⇒ don't have interaction between student & Income

not parallel

interaction betw



- Interactions can be applied to qualitative variables, or to a combination of quantitative and qualitative variables

?? what about 2 quantitative variables?

FIGURE 3.7. *For the* Credit *data, the least squares lines are shown for prediction of* balance *from* income *for students and non-students.* Left: *The model (3.34) was fit. There is no interaction between* income *and* student. Right: *The model (3.35) was fit. There is an intera*

student

$$\texttt{balance}_i \approx \beta_0 + \beta_1 \times \texttt{income}_i + \begin{cases} \beta_2 & \text{if } i\text{th person is a student} \\ 0 & \text{if } i\text{th person is not a student} \end{cases}$$

$$= \beta_1 \times \texttt{income}_i + \begin{cases} \beta_0 + \beta_2 & \text{if } i\text{th person is a student} \\ \beta_0 & \text{if } i\text{th person is not a student.} \end{cases}$$

$$\texttt{balance}_i \approx \beta_0 + \beta_1 \times \texttt{income}_i + \begin{cases} \beta_2 + \beta_3 \times \texttt{income}_i & \text{if student} \\ 0 & \text{if not studen} \end{cases}$$

Model Output

```
Call:
lm(formula = Balance ~ Income * Student + Limit + Rating, data = credit)

Residuals:
    Min       1Q   Median       3Q      Max
-192.03   -76.82   -14.99    57.14   299.47

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -509.12714   20.87263 -24.392  < 2e-16 ***
Income              -8.00028    0.24450 -32.721  < 2e-16 ***
StudentYes         409.65125   26.92456  15.215  < 2e-16 ***
Limit                0.12360    0.02741   4.509 8.60e-06 ***
Rating               2.16731    0.40750   5.318 1.76e-07 ***
Income:StudentYes    0.28775    0.44809   0.642    0.521
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 101 on 394 degrees of freedom
Multiple R-squared: 0.9524,     Adjusted R-squared: 0.9518
F-statistic:  1575 on 5 and 394 DF,  p-value: < 2.2e-16
```

insignificant

Summary

- *Is there a relationship between advertising sales and budget?* P16. code as long as ~~F-statistic~~ is significant
  F > 1, has a significant relationship
- *How strong is the relationship?* both RSE, R^2 — larger, the better (>80%)
- *Which media contribute to sales?* significant p value.
- *How large is the effect of each medium on sales?*
  slope
- *Is there synergy among the advertising media?*
  interaction P32. code TV: Radio P-value