UNIVERSITY OF ARKANSAS

DASC 4113 Machine Learning

Ukash Nakarmi

Lecture 3

Linear Regression

Learning Objectives

In this class, we will learn about following concepts:

- Simple Regression and Multiple Linear Regression
- How to assess the estimated regression parameters and regression model
- How to understand/explore the relation/significance between predictors and response
- How to understand the relation between the predictors(if any) and enforce it in our regression model

Linear Regerssion

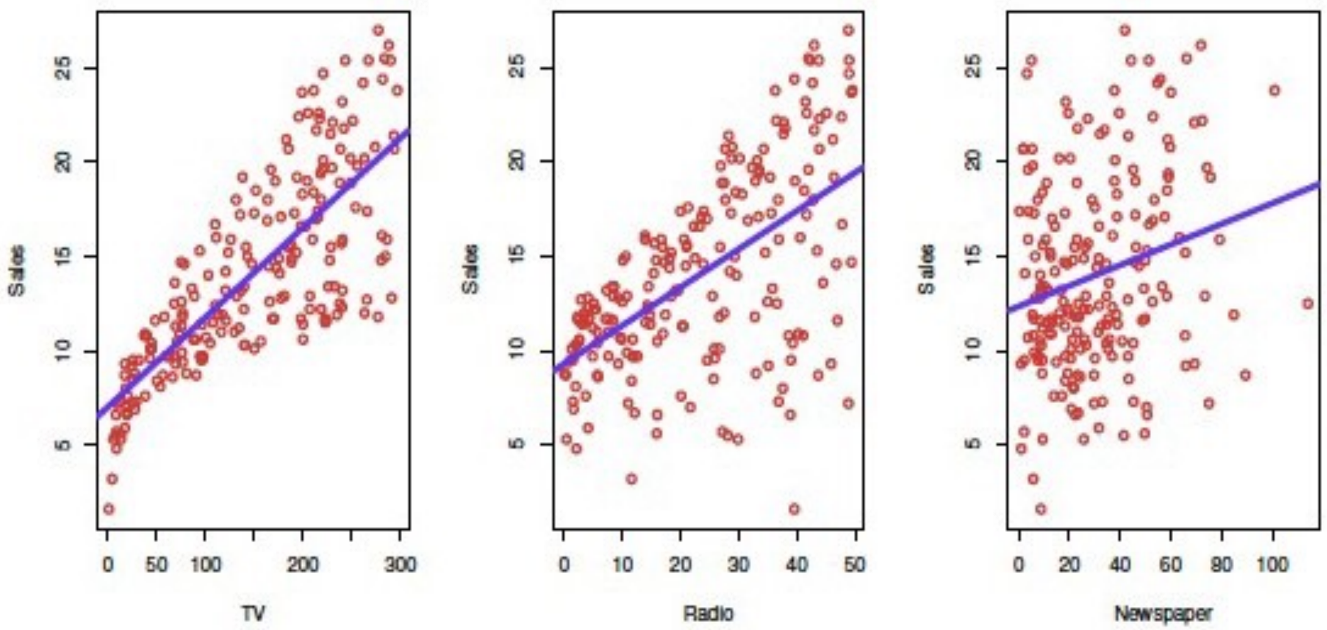
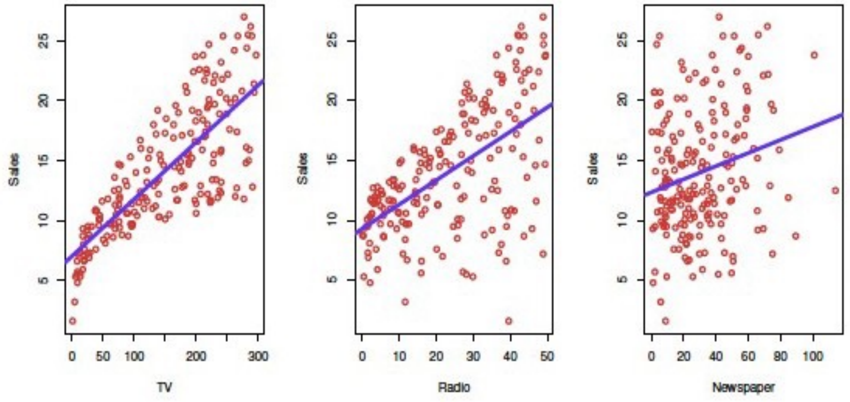


Fig 2.1 Advertisement budget vs Sales

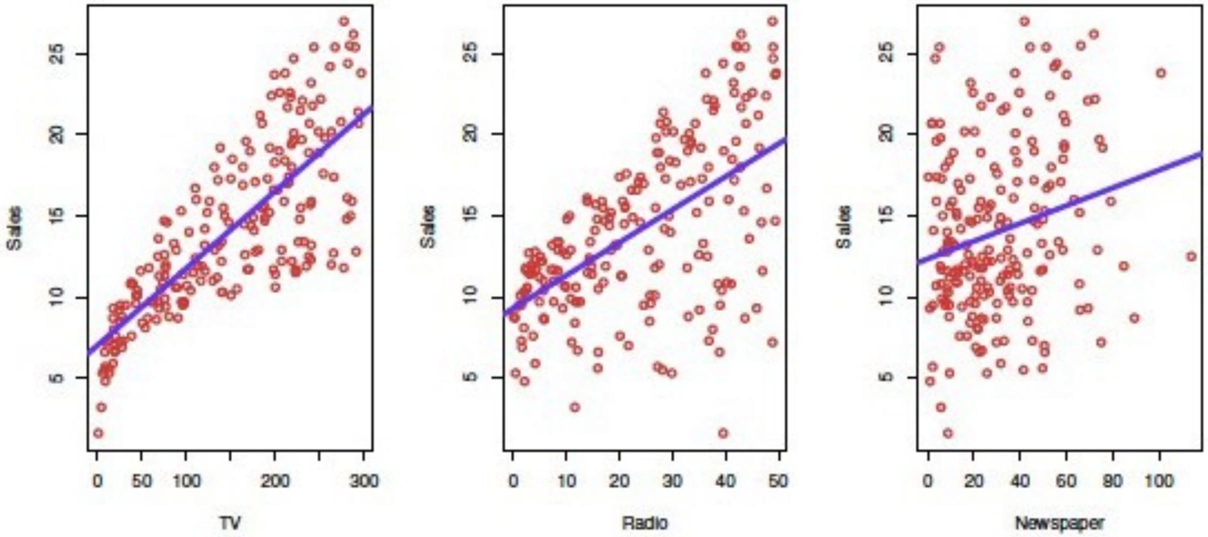
Linear Regression



Some questions that we can ask:

1. Is there any relation between advertisement budget and sales?
2. How strong is the relation?
3. Which media are associated with sales?
4. How strong is the association between each medium and sales?
5. How accurately can we predict future sales?
6. Is the relation linear?
7. Is there a synergy between the advertising media?

Linear Regre



Is there a synergy between the advertising media?

Q. Estimated sales when advertisement budget is $100k in TV media?

Q. Estimated sales when advertisement budget is $50k in TV media?

Q. Estimated sales when advertisement budget is $50k in Radio media?

Simple Linear Regression (Regression with 1 variable)

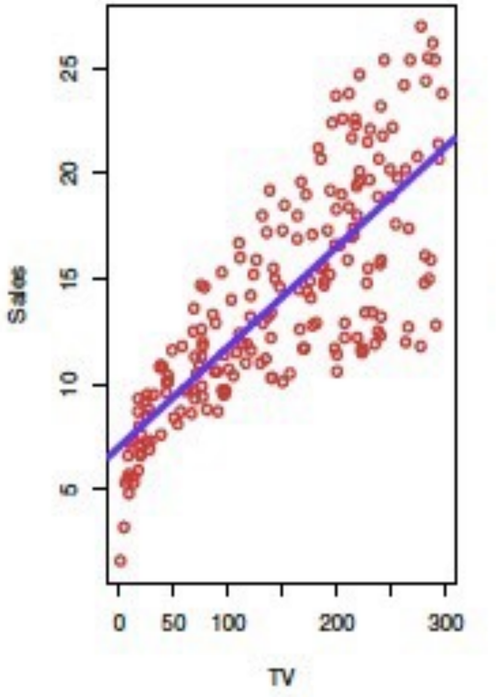
1. Assumption of the functional form of the relation.

$$Y \approx \beta_0 + \beta_1 X.$$

Example:

$$\texttt{sales} \approx \beta_0 + \beta_1 \times \texttt{TV}.$$

Parameters to estimate:
β_0 and β_1



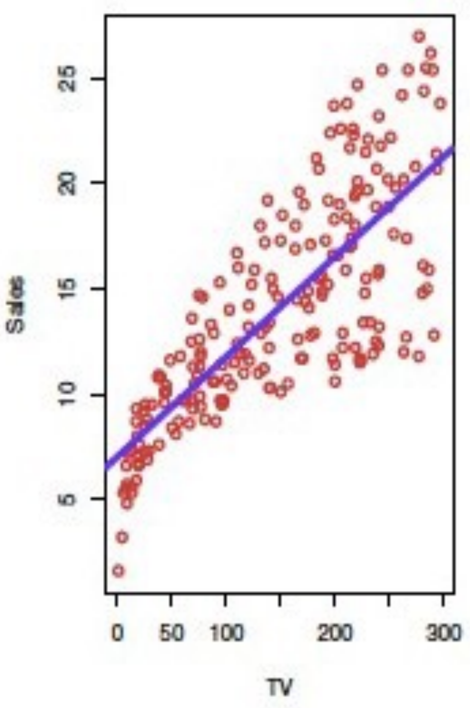
Q. What do β_0 and β_1 represent in the geometric sense?

Simple Linear Regression

What do β_0 and β_1 represent in the geometric sense?

Parameters of simple linear regression : β_0 and β_1

=> Intercept and Slope of a line.



Simple Linear Regression

2. Estimating Coefficients

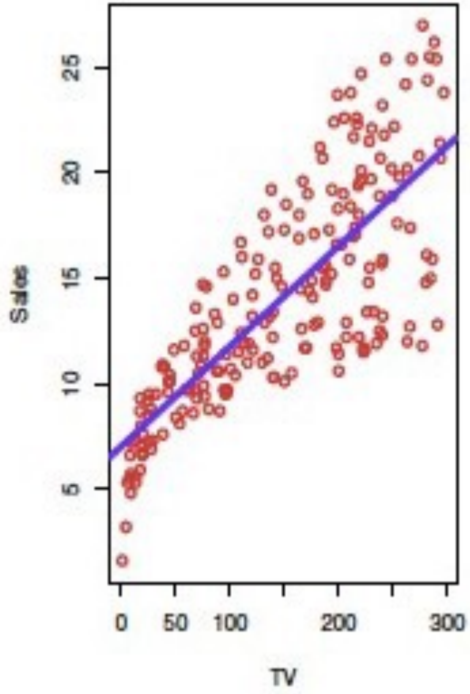
$$Y \approx \beta_0 + \beta_1 X.$$

Given a n training samples in a training data set,

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

Estimate:

$$y_i \approx \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad \forall (x_i, y_i) \in Training\ Data,\ i = 1,2,\ldots n$$



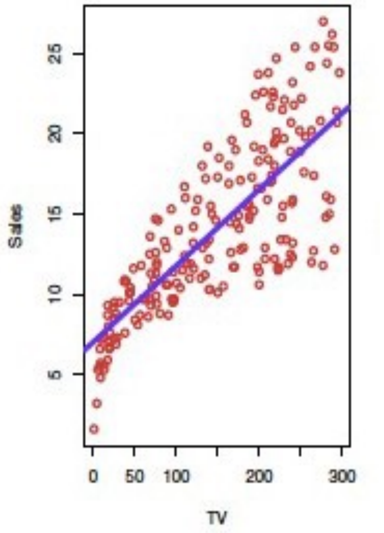
Job: Find Intercept β_0 and Slope β_1

Simple Linear Regression

2. Estimating Coefficients

$$Y \approx \beta_0 + \beta_1 X.$$

Job: Find Intercept β_0 and Slope β_1



Estimating Coefficients through **Least Squares**

Let $\hat{y}$ be predictions using estimates $\hat{\beta}_0$ and $\hat{\beta}_1$

For each i^{th} sample: $$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Residual/Error for i^{th} sample: $$e_i = y_i - \hat{y}_i$$

Simple Linear Regression

Residual Sum of Squares:

$$\mathrm{RSS} = e_1^2 + e_2^2 + \cdots + e_n^2$$

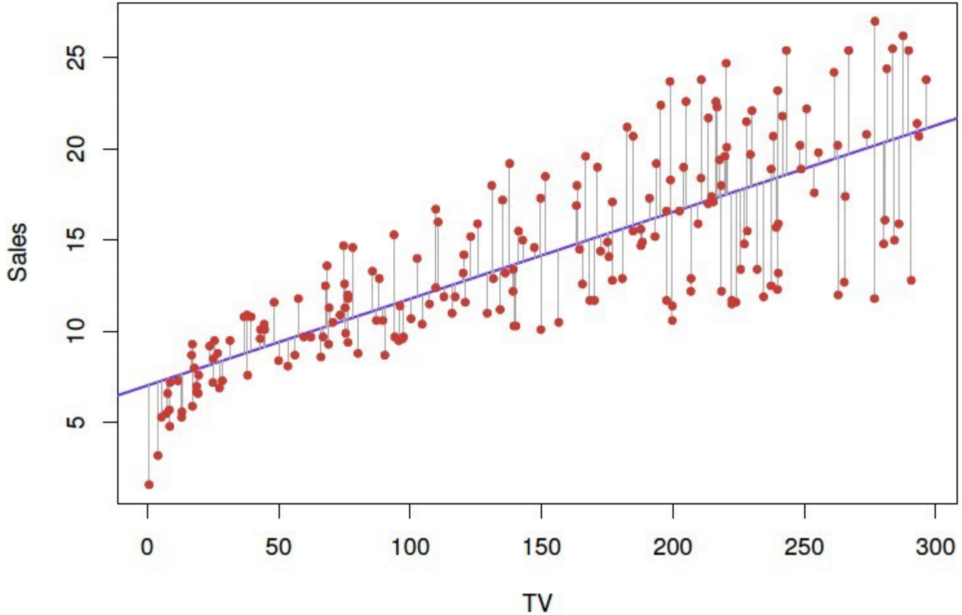
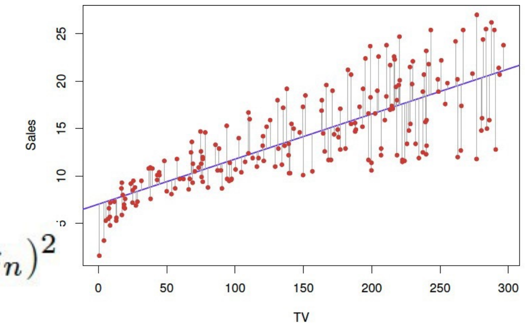


Fig 3.1 Advertisement budget vs Sales, linear regression and error terms

$$\mathrm{RSS} = (y_1 - \hat\beta_0 - \hat\beta_1 x_1)^2 + (y_2 - \hat\beta_0 - \hat\beta_1 x_2)^2 + \cdots + (y_n - \hat\beta_0 - \hat\beta_1 x_n)^2$$

Goal: Find β_0 and that β_1 minimizes RSS

Simple Linear Regression



Residual Sum of Squares:

$$\mathrm{RSS} = e_1^2 + e_2^2 + \cdots + e_n^2$$

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

Goal: Find β_0 and that β_1 minimizes RSS

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

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

$$\bar{y} \equiv \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\bar{x} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i$$

Simple Linear Regression

In this example:

$\beta_0 = 7.03$
$\beta_1 = 0.0475$

We started with these questions:

Is there a relation?
(Yes/No, Positive/Negative)

How strong is the relation ?

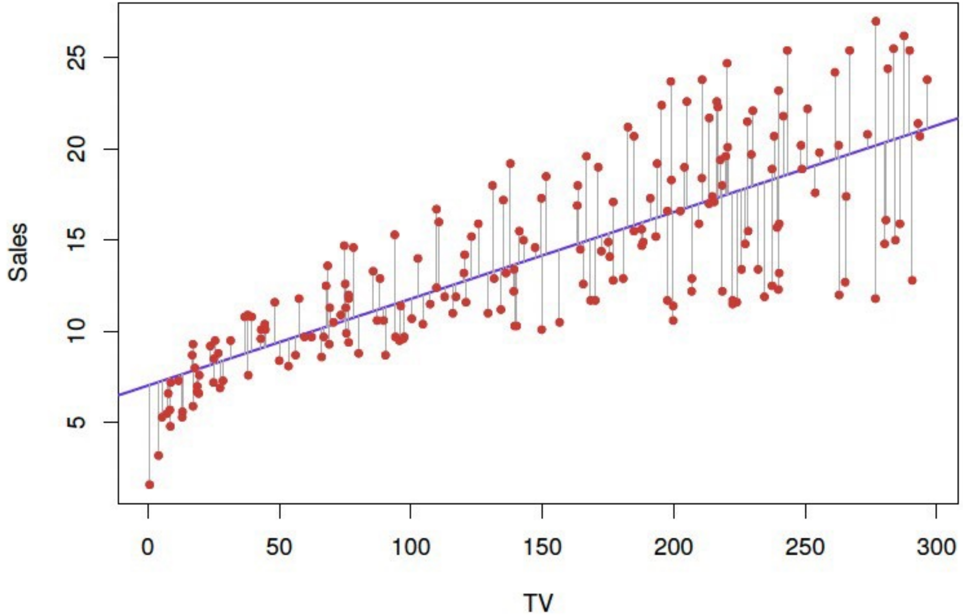


Fig 3.1 Advertisement budget vs Sales, linear regression and error terms

Q. If we increase the advertisement budget by 500, how many more units of sales can we expect?

Simple Linear Regression

Assessing the accuracy of Coefficients Estimates

Population regression line

$$f(X) = 2 + 3X$$

Observed Sample Data

$$Y = 2 + 3X + \epsilon$$

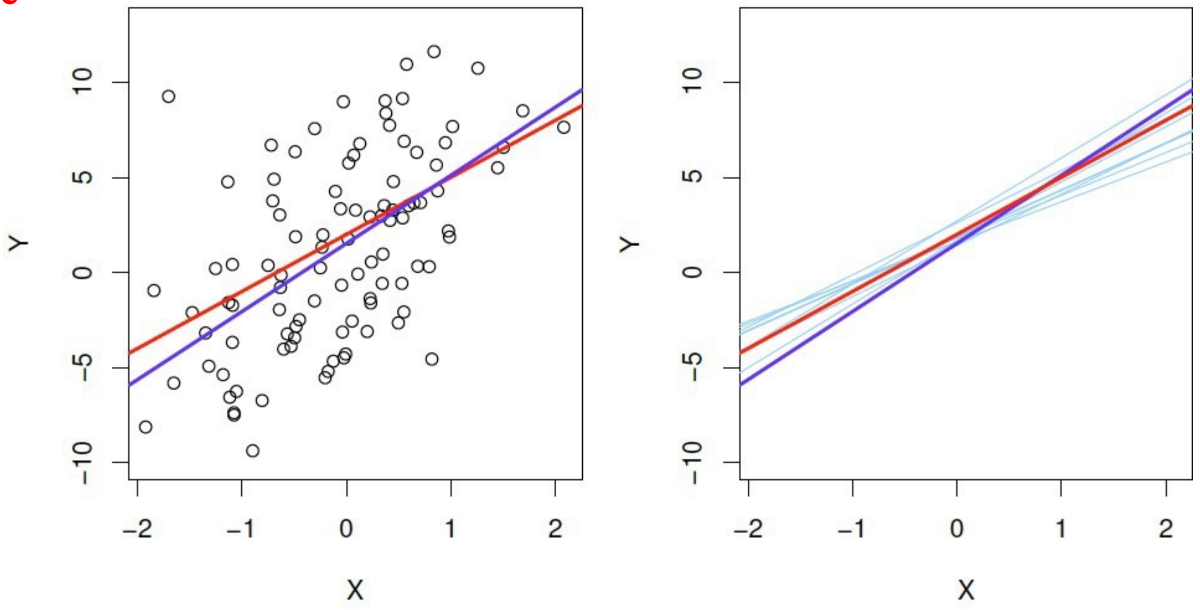


Fig 3.3 Left: Population regression line (True Relation), Observed Sample Data, Estimated Regression Line. | Right: True regression line and many estimate regression lines for several training dataset.

Simple Linear Regression

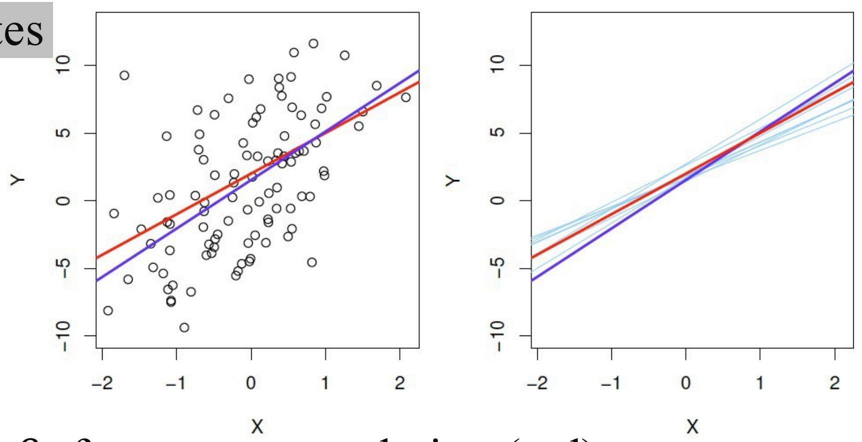
Assessing the accuracy of Coefficients Estimates

Population regression line

$$f(X) = 2 + 3X$$

Observed Sample Data

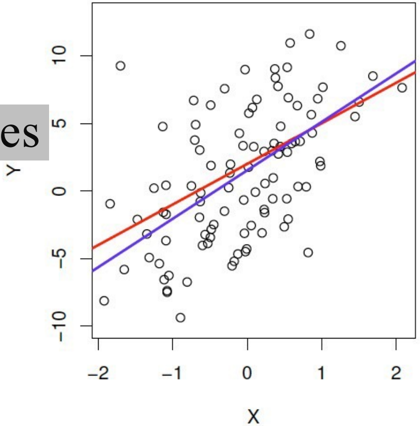
$$Y = 2 + 3X + \epsilon$$

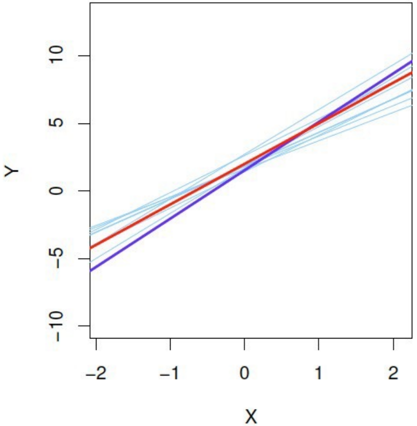


- We could get different estimates of β_0 and β_1 for same true relation (red) depending on the training data set.
- How do we assess the parameters β_0 and β_1 we estimated ?

Simple Linear Regression

Assessing the accuracy of Coefficients Estimates





- How do we assess the parameters β_0 and β_1 we estimated ?

=> Standard Error (SE)

Let's take an example of Sample Mean to understand SE

** Sample mean is an Unbiased Estimator:

Q. What do we mean by sample mean is an Unbiased Estimator?

Simple Linear Regression

Assessing the accuracy of Coefficients Estimates

Q. What do we mean by sample mean is an Unbiased Estimator?

- On average sample mean $\hat{u}$ is equal to the population mean u.
- Unbiased estimator does not systematically over or underestimate the true parameter.

- Calculate the average of the sample means of many training dataset, then we expect that average to be equal to the population mean.

- Estimation of $\hat{u}$ from one training data set maybe however over or under- estimate the true u.

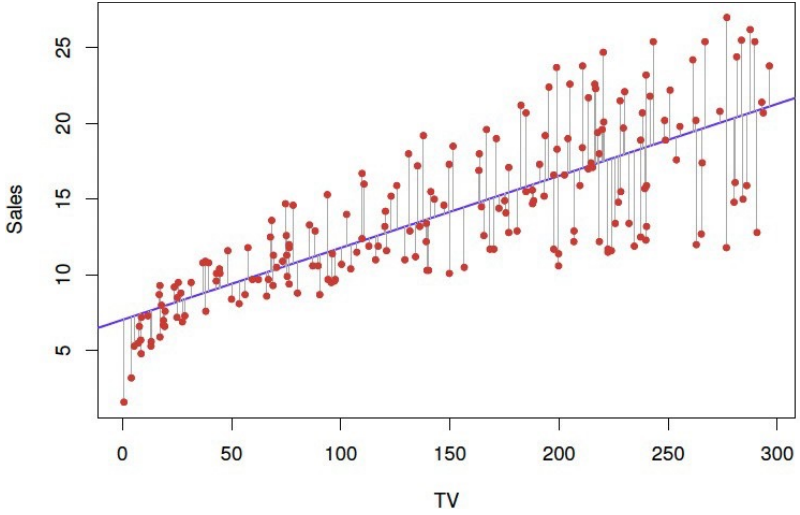
- Standard Error (SE) of $\hat{u}$

$$\mathrm{Var}(\hat{\mu}) = \mathrm{SE}(\hat{\mu})^2 = \frac{\sigma^2}{n},$$

- σ is the standard deviation of each of realization of $\hat{y}$ of Y.
- As n increases, SE decreases.

Simple Linear Regression

Standard Error (SE) of β



$$\mathrm{SE}(\hat{\beta}_0)^2 = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

$$\mathrm{SE}(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\sigma^2 = Var\ (\epsilon)\ .$$

The relation is strictly true only when error ϵ_i for each observation are independent and uncorrelated.

Q. Are error uncorrelated in the figure above?

Simple Linear Regression

Confidence of Interval

Standard Error (SE) can be used to compute confidence intervals.

For example: A 95% confidence interval is defined as a range of values such that with 95% probability the range will contain the true unknown parameter.

In Linear regression:
A 95% confidence interval for β_1 approximately takes the form: $\hat{\beta}_1 \pm 2 \cdot \mathrm{SE}(\hat{\beta}_1).$

i.e. There is approximately 95 % chance that that true β_1 is contained within $\left[\hat{\beta}_1 - 2 \cdot \mathrm{SE}(\hat{\beta}_1),\ \hat{\beta}_1 + 2 \cdot \mathrm{SE}(\hat{\beta}_1)\right]$

Similarly, for β_0

$$\hat{\beta}_0 \pm 2 \cdot \mathrm{SE}(\hat{\beta}_0).$$

Conclusion: If we can estimate SE then we can assess the parameters β_0 and β_1

Simple Linear Regression

Conclusion: If we can estimate SE then we can assess the parameters β_0 and β_1

$$\mathrm{SE}(\hat{\beta}_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right]$$

$$\mathrm{SE}(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

But σ is unknown in general.
We estimate σ from the data (Termed as Residual Standard Error)

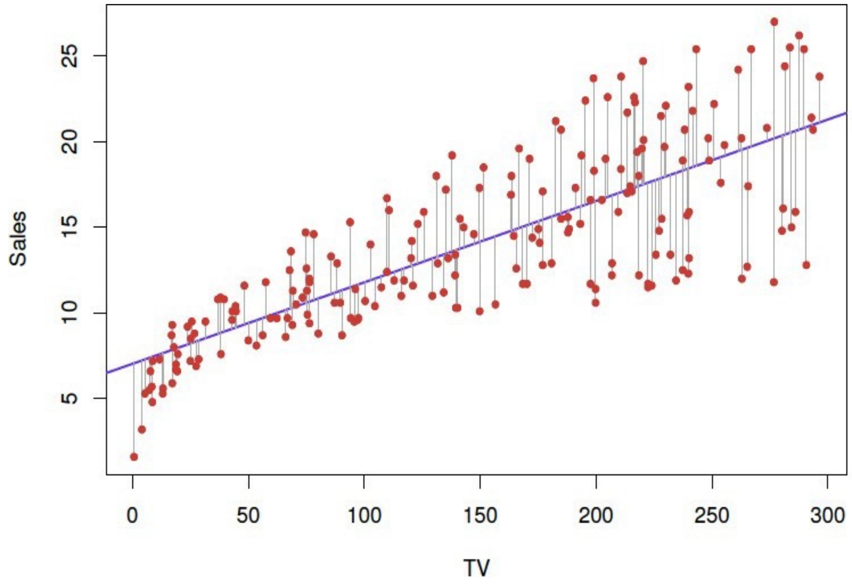
$$\mathrm{RSE} = \sqrt{\mathrm{RSS}/(n-2)}.$$

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

Simple Linear Regression

For our example, $\beta_0 = 7.03$, SE = 0.4578
$\beta_1 = 0.0475$, SE = 0.0027

95 %confidence interval of β_0 is [6.130, 7.935]
95 %confidence interval of β_1 is [0.042, 0.053]



Using these confidence intervals, we can conclude:

With 95% of confidence when can say that:
When there is no budget for advertisement, on average the sales will be 6130 units to 7935 units .
For each \$1000 increase in advertisement budget, on average there will be increase in sales between 42 t0 53 units.

Simple Linear Regression

- SE can also be used for hypothesis testing.
- Hypothesis testing can be used to understand relation between X and Y.

$$Y \approx \beta_0 + \beta_1 X.$$

H_0 : There is no relationship between X and Y ⇒ $H_0 : \beta_1 = 0$

H_a : There is some relationship between X and Y ⇒ $H_a : \beta_1 \neq 0,$

To test Null Hypothesis, we measure how far/close is the estimate $\hat{\beta}_1$ from 0

But how far is far, and how close is close? Suppose we have $\hat{\beta}_1 = 0.047$, is it far or close to zero?

For this, we use t-statistic: Measure of the departure of the estimated value of a parameter from its hypothesized value to its standard error.

$$t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)},$$

For Our Null Hypothesis, hypothesized value of $\beta_!$ =0

Simple Linear Regression

t-statistic: Measure of the departure of the estimated value of a parameter from its hypothesized value to its standard error.

$$t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)},$$

For Null Hypothesis, hypothesized value of β_1 =0

H_0 : There is no relationship between X and Y $\quad H_0 : \beta_1 = 0$

- If $\mathrm{SE}(\hat{\beta}_1)$ is small, then even relatively small value of $\hat{\beta}_1$ suggest strong evidence that β_1 is non-zero (That is, there is relation between that input and output).

- If $\mathrm{SE}(\hat{\beta}_1)$ is is large, then we need relatively large $\hat{\beta}_1$ to reject null hypothesis.

Simple Linear Regression

t-statistic: Measure of the departure of the estimated value of a parameter from its hypothesized value to its standard error.

$$t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)},$$

For Our Null Hypothesis, hypothesized value of $\beta_1 = 0$

$$H_0 : \beta_1 = 0$$

p-value: The probability of observing t-statistic greater or equal to some value. (t-statistic calculated under Null hypothesis.)

In General,

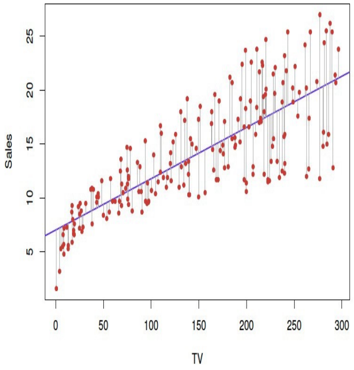
Lower the p-value: Null-Hypothesis is unlikely => There is a strong relation between X and true Y.

For our TV-budget and Sales example

	Coefficient	Std. error	t-statistic	p-value
Intercept	7.0325	0.4578	15.36	< 0.0001
TV	0.0475	0.0027	17.67	< 0.0001

Simple Linear Regression

Assessing the accuracy of the Model



Assess Using two Quantities:

1. Residual Standard Error(RSE): Measure of the Lack of Fit of the Model.

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

- Greater the RSE value, more is the Lack of Fit.
- Smaller RSE refers to better fit of the model.

Quantity	Value
Residual standard error	3.26

Residual Standard Error(RSE) for our simple linear model on TV budget vs Sales in thousand

What does RSE = 3.26 mean in this case?

=> Even if we exactly know the true value of parameters, on average, we would still have an offset of 3260 units on any prediction of sales based on the TV budget. (Because of ϵ)

UNIVERSITY OF ARKANSAS

Simple Linear Regression

Assessing the accuracy of the Model

Assess Using two Quantities:

2. R^2 Statistic : Measure of proportion of Variance of Y explained.

Bounds the quantity within 0-1

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

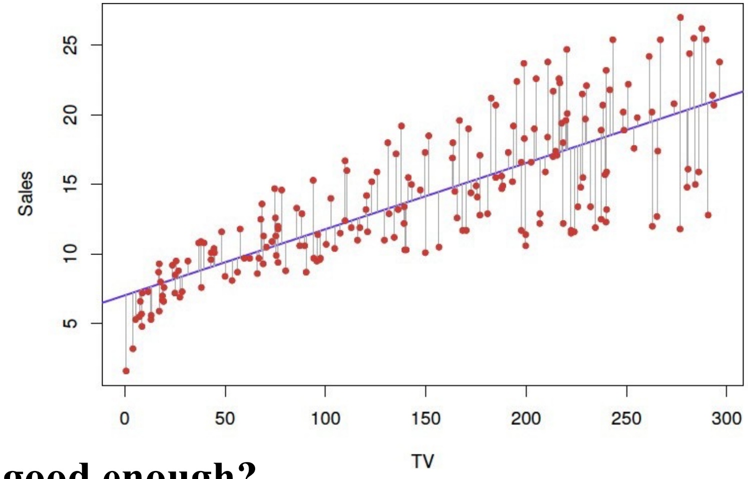
$\mathrm{TSS} = \sum (y_i - \bar{y})^2$: Total Sum of Squares: Measures the total variance in the response Y.

RSS: Measures the amount of variability (error) unexplained after the regression. ($\mathit{pertaining\ to}\ \epsilon$ or wrong model)

- So, R^2 measures the proportion of variability that can be explained using Y.
- Smaller value (close to 0) means does not explains the variability in the response Y.
- Could mean: The model is poor, or the irreducible error $\mathrm{Var}(\epsilon)$ is too high.

Simple Linear Regression

Quantity	Value
Residual standard error	3.26
R^2	0.612



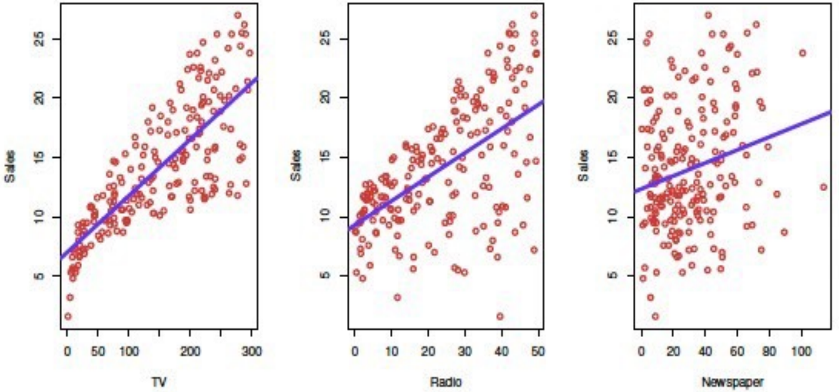
Q. So how big of a R and how small of a RSE is good enough?

Depends on the application.

For Simple Linear Regression, R^2 Statistic = Squared correlation between X and Y

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

Multiple Linear Regression



General Form: $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon,$$

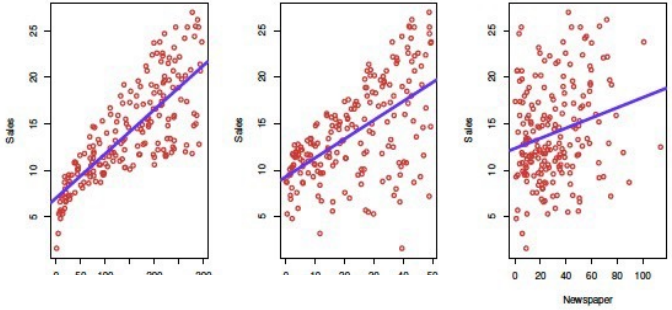
Our Example: $$\texttt{sales} = \beta_0 + \beta_1 \times \texttt{TV} + \beta_2 \times \texttt{radio} + \beta_3 \times \texttt{newspaper} + \epsilon.$$

Prediction and Parameter Estimation: $$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p.$$

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

Minimize

Multiple Linear Regression



SLR and MLR might tell different stories.

Example: Newspaper budget vs Sales Regression

Table 3.3 SLR Newspaper adv. Budget vs. sales

	Coefficient	Std. error	t-statistic	p-value
Intercept	12.351	0.621	19.88	< 0.0001
newspaper	0.055	0.017	3.30	0.00115

Table 3.4 MLR in Adv Data

	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

Multiple Linear Regression

	p-value
Intercept	< 0.0001
newspaper	0.00115

Table 3.3 SLR Newspaper adv. Budget vs. sales

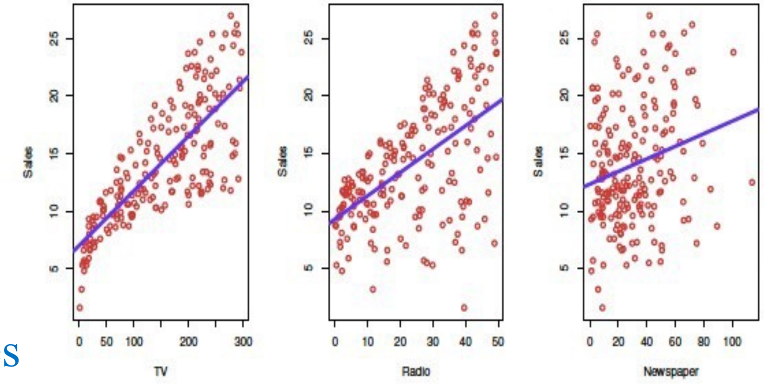


Table 3.4 MLR

	p-value
Intercept	< 0.0001
TV	< 0.0001
radio	< 0.0001
newspaper	0.8599

Q. Why do we get different p-values for Newspaper in SLR and MLR?

Q. What do different p-value for newspaper signify?

Multiple Linear Regression

Q. Why do we get different p-values for Newspaper in SLR and MLR?

SLR

In SLR settings, the slope represent combine effect of increase in all 3 media of advertisement. In MLR, the slope term (β_3)represent effect of increase in Newspaper advertisement budget only.

Q. What do the different p-values signify?

- By increasing advertisement budget in Newspaper only, we may not get expected increase in sales.
- Newspaper budget may have some relation with other advertisement media.
- High correlation between radio and newspaper budget suggest that, for some advertisement market when we increase the radio advertisement budget, we also increased Newspaper budget.
- The apparent significance of Newspaper advertisement on sales might be coming from this fact.

	TV	radio	newspaper	sales
TV	1.0000	0.0548	0.0567	0.7822
radio		1.0000	0.3541	0.5762
newspaper			1.0000	0.2283
sales				1.0000

Multiple Linear Regression

MLR allows us to discuss following questions:

- Is at least one of the predictors useful in predicting the response?
- Do all predictors help to explain Y or only some of the predictors?
- How well does the model fit the data?
- Given the set of predictor values, what response values should we predict, and how accurate is our prediction?

Multiple Linear Regression

- Is at least one of the predictors useful in predicting the response?

Similarly, as in Simple Linear Regression: We do Hypothesis Testing using F-statistic

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

$$H_a : \text{ at least one } \beta_j \text{ is non-zero.}$$

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

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

Table 3.6 : Mult. Linear regression assess parameters of Adv. data

- If there is no relationship between the response and input, F-statistic will be close to 1.
- To reject Null Hypothesis: We need a higher value of F.

Multiple Linear Regression

- Is at least one of the predictors useful in predicting the response?

We can also perform Null Hypothesis test over subset of inputs:

$$H_0: \quad \beta_{p-q+1} = \beta_{p-q+2} = \cdots = \beta_p = 0,$$

$$F = \frac{(\mathrm{RSS}_0 - \mathrm{RSS})/q}{\mathrm{RSS}/(n-p-1)}.$$

RSS_0 is Residual sum of square of the model we are testing.

- If there is no relationship between the response and input, F-statistic will be close to 1.
- To reject Null Hypothesis: We need a higher value of F. (i.e. to claim there is some relation between at least 1 input variable and the output.)

Multiple Linear Regression

- How well does the model fit the data?

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

R^2: Value closer to 1 suggest model explains the large portion of the variance in the data.

Q. Does larger R^2 (closer to 1) value always means the better model?

Multiple Linear Regression

Q. Does larger R^2 (closer to 1) value always means the better model?

$R^2 = 0.89719$: Model that used only TV and Radio Budget
$R^2 = 0.8972$: Model that used all three TV, Radio and Newspaper Budget

- Including Newspaper improved our R^2 of the model.
- But we established Newspaper do not significant role in Sales (high p-value)
- Adding a new input variable always increases R^2 in the training data but not necessarily in the test data.
- This in a sense is like overfitting the training data. (Why??)

RSE is a better metric for making such decisions instead. (Try compare RSE of model with TV only, TV and Radio and TV, Radio and Newspaper all three.)

Multiple Linear Regression

- Given the set of predictor values, what response values should we predict, and how accurate is our prediction?

Answered through: Confidence Interval and Prediction Interval

Other Considerations in Regression Model

Qualitative Input Variables (Qualitative Predictors)

When input takes Yes/No form (Factors)
Examples:

- Investigate difference in credit card balance between those who own a house and and who don't.
- Investigate some response based on whether a person is from the south or not and west or not.

Other Considerations in Regression Model

Predictors with only two levels

Example: Investigate difference in credit card balance between those who own a house and who don't.

Address it by creating an Indicator variable(dummy variable)

$$x_i = \begin{cases} 1 & \text{if } i\text{th person owns a house} \\ 0 & \text{if } i\text{th person does not own a house} \end{cases}$$

Indicator variable

Other Considerations in Regression Model

- Address it by creating an Indicator variable(dummy variable)

Indicator variable

$$x_i = \begin{cases} 1 & \text{if } i\text{th person owns a house} \\ 0 & \text{if } i\text{th person does not own a house} \end{cases}$$

- Use the Indicator variable as a predictor in the regression.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person owns a house} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person does not.} \end{cases}$$

Recall: In simple linear regression with quantitative variables, β_0 and β_1 correspond to intercept and slope.

Q: Would we still have same intercept and slope interpretation of β_0 and β_1 as in Simple Linear Regression?

Other Considerations in Regression Model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person owns a house} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person does not.} \end{cases}$$

Q: Would we still have same intercept and slope interpretation of β_0 and β_1 ?

	Coefficient	Std. error	t-statistic	p-value
Intercept	509.80	33.13	15.389	< 0.0001
own[Yes]	19.73	46.05	0.429	0.6690

Table 3.7, Regression coeffs. and parameters for credit card balance difference data.

β_0 : Note: it solely governs value of y when x_i takes value 0 (do not own a house) Represents average credit card loan of someone who do not owns a house.

β_1 : We know, x_i can take value 0/1 (based on owns or does not). β_1 is the average amount of increase in the loan (y) if a person owns a house.

Other Considerations in Regression Model

Alternative way to define a two step Indicator Variable

$$x_i = \begin{cases} 1 & \text{if } i\text{th person owns a house} \\ -1 & \text{if } i\text{th person does not own a house} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person owns a house} \\ \beta_0 - \beta_1 + \epsilon_i & \text{if } i\text{th person does not own a house.} \end{cases}$$

Q: Will β_0 and β_1 value be different than when we defined Indicator variable as 0/1?

Q: Do β_0 and β_1 have same interpretation as when we defined Indicator variable as 0/1 level?

Other Considerations in Regression Model

Qualitative Predictors with more than two levels:

Define additional Indicator Variable

$$x_{i1} = \begin{cases} 1 & \text{if } i\text{th person is from the South} \\ 0 & \text{if } i\text{th person is not from the South} \end{cases}$$

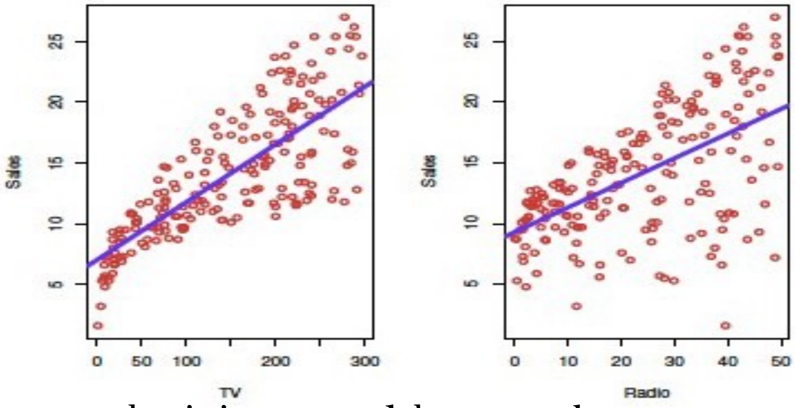
$$x_{i2} = \begin{cases} 1 & \text{if } i\text{th person is from the West} \\ 0 & \text{if } i\text{th person is not from the West} \end{cases}$$

Regression Takes the form:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is from the South} \\ \beta_0 + \beta_2 + \epsilon_i & \text{if } i\text{th person is from the West} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is from the East.} \end{cases}$$

Extension of the Linear Model

Recall our Synergy Effect Discussion.



Q. Given the budget is $100k, what advertisement decision would we make based on the graphs above?

A) Spend all on Radio

B) Spend some on TV and some on the radio

Q. What different assumptions did we make in each of these decisions ?

A)For same amount of ad budget, Radio has much more sales. Relation between adv. budget in the radio and sales is not affected by the advertisement budget in the TV.

B) Maybe affected. So, lets divide the budget.

Extension of the Regression Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$$

No interaction between the Input variables.

In many applications, there may be interaction between the input variables.

Examples:

- Interaction between Radio and TV budget on Sales
- Interaction between the number of production lines and number of workers in a factory.

(Interaction: Changing one input would have effect on the other input or not)

Introduce an Interaction variable

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \underbrace{X_1 X_2}_{\text{Interaction variable}} + \epsilon.$$

Extension of the Regression Model

Introduce an Interaction variable

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \underbrace{X_1 X_2}_{\text{Interaction variable}} + \epsilon.$$

$$\begin{aligned} Y &= \beta_0 + (\beta_1 + \beta_3 X_2) X_1 + \beta_2 X_2 + \epsilon \\ &= \beta_0 + \tilde{\beta}_1 X_1 + \beta_2 X_2 + \epsilon \end{aligned}$$

$$\tilde{\beta}_1 = \beta_1 + \beta_3 X_2.$$

The coeff. of X1 is now dependent on X2.

The relation between Y and X1 is no longer a constant. (Think in terms of partial derivatives).

Extension of the Regression Model

Introduce an Interaction variable

Example: Sales vs radio and TV advertisement

$$
\begin{aligned}
\texttt{sales} &= \beta_0 + \beta_1 \times \texttt{TV} + \beta_2 \times \texttt{radio} + \beta_3 \times (\texttt{radio} \times \texttt{TV}) + \epsilon \\
&= \beta_0 + (\beta_1 + \beta_3 \times \texttt{radio}) \times \texttt{TV} + \beta_2 \times \texttt{radio} + \epsilon.
\end{aligned}
$$

β_3 can be interpreted as increase in the effectiveness of TV advertising associated with one-unit increase in radio budget.

We can draw similar conclusion by grouping radio and interaction variable (radio$\times$ TV)

Take Away!

- Simple Linear Regression considers only 1 input variable.
- We can assess coefficient of regression by calculating Standard Error and Confidence interval.
- We can assess the significance of predictors by assessing t-statistic and corresponding p-values.
- Model accuracy can be assessed by RSE and R2.
- Multiple Linear Regression considers many input variables.
- Input variables in regression can be qualitative.
- Interaction between input variables can be enforced by introducing Interaction variable in the regressor.

Python code base for the book

https://github.com/JWarmenhoven/ISLR-python