

# Creating Statistical Models

---



**Matthew Renze**

DATA SCIENCE CONSULTANT

@matthewrenze [www.matthewrenze.com](http://www.matthewrenze.com)



# Overview



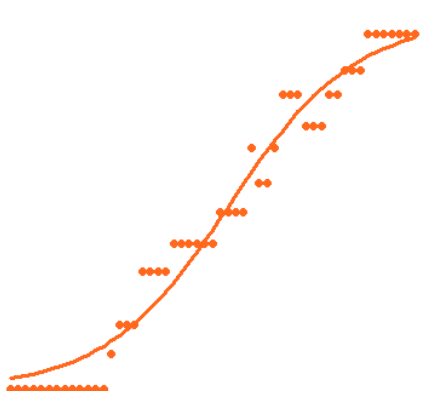
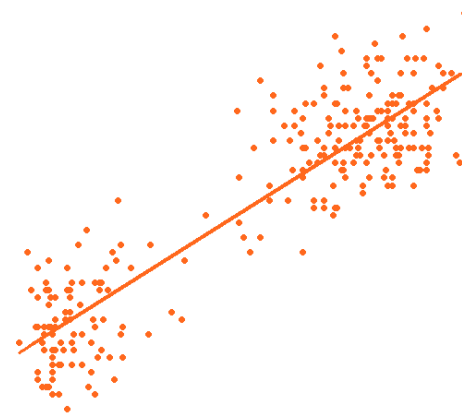
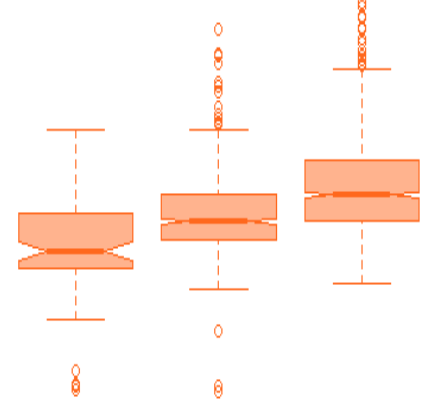
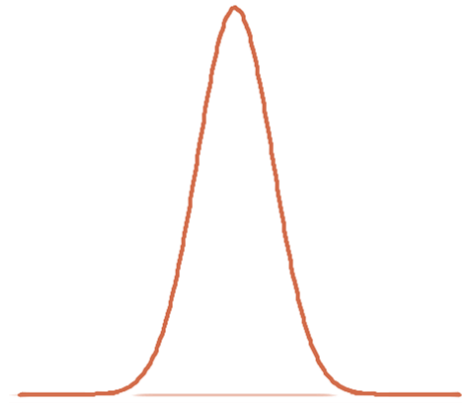
**Creating Statistical Models**

**Demo**



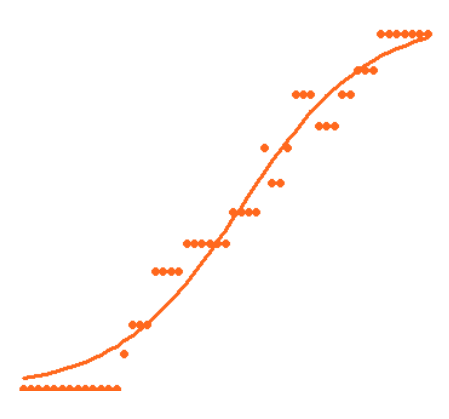
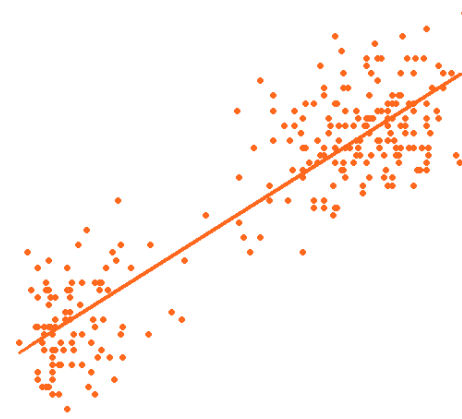
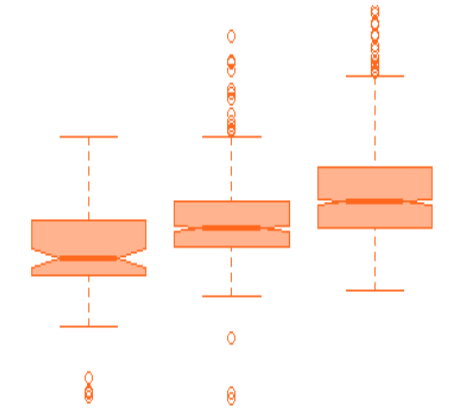
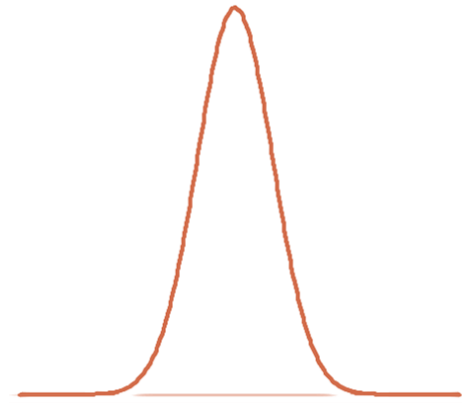
# Statistical Model

Mathematical equation  
Approximation of reality



# Statistical Model

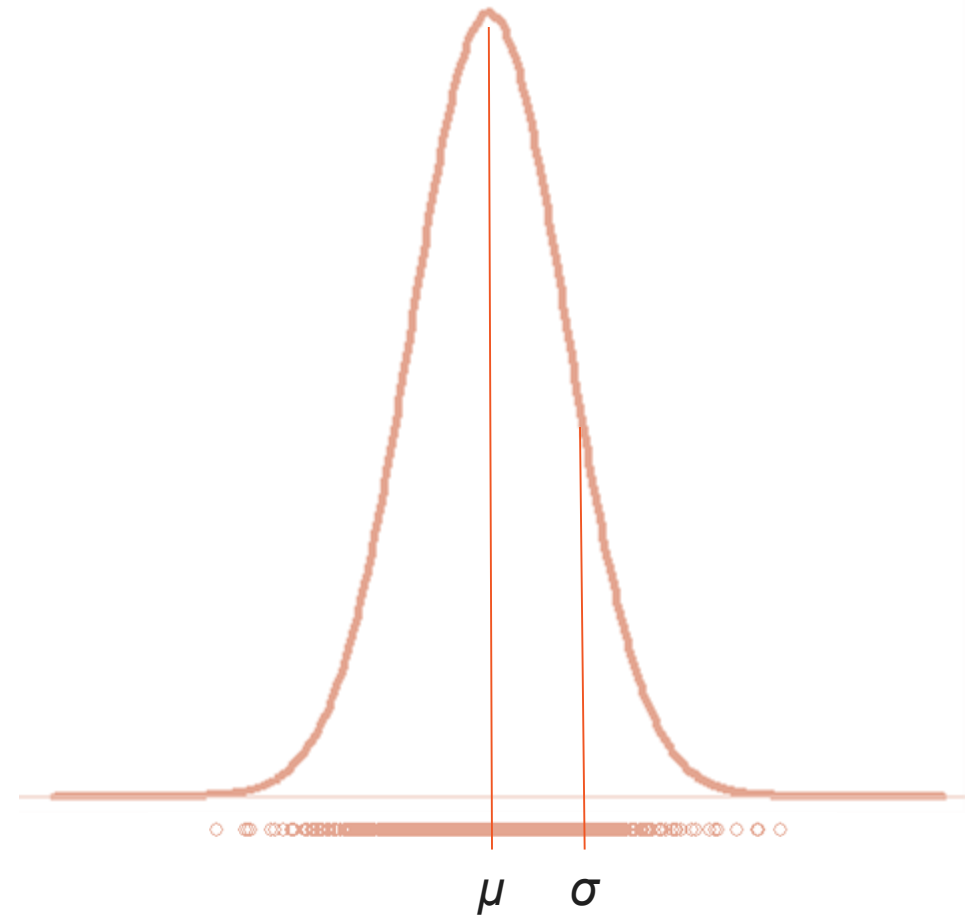
Description  
Inference  
Comparison  
Prediction



# Statistical Model

Parametric

Non-parameteric



# Types of Statistical Models

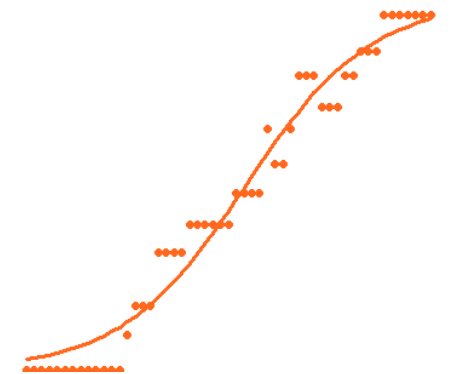
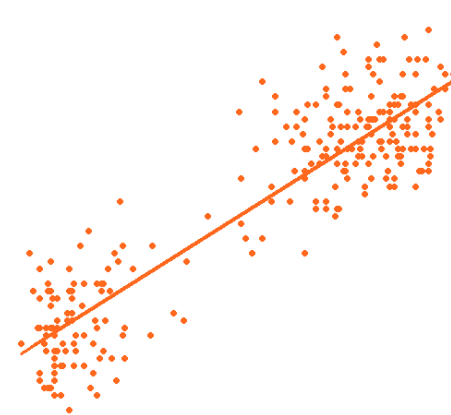
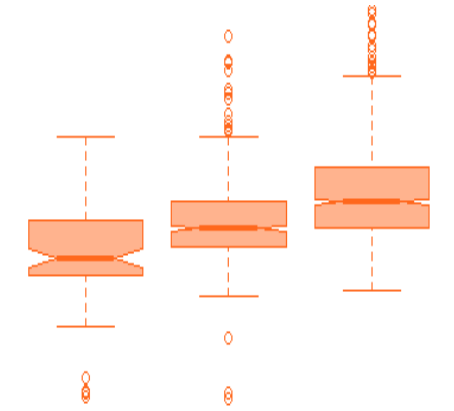
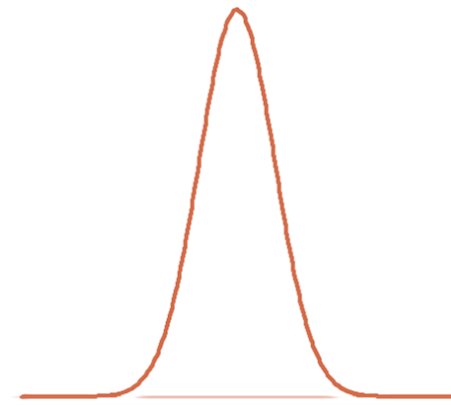
Probability density function

Analysis of variance

Simple linear regression

Non-linear regression

Bayesian network



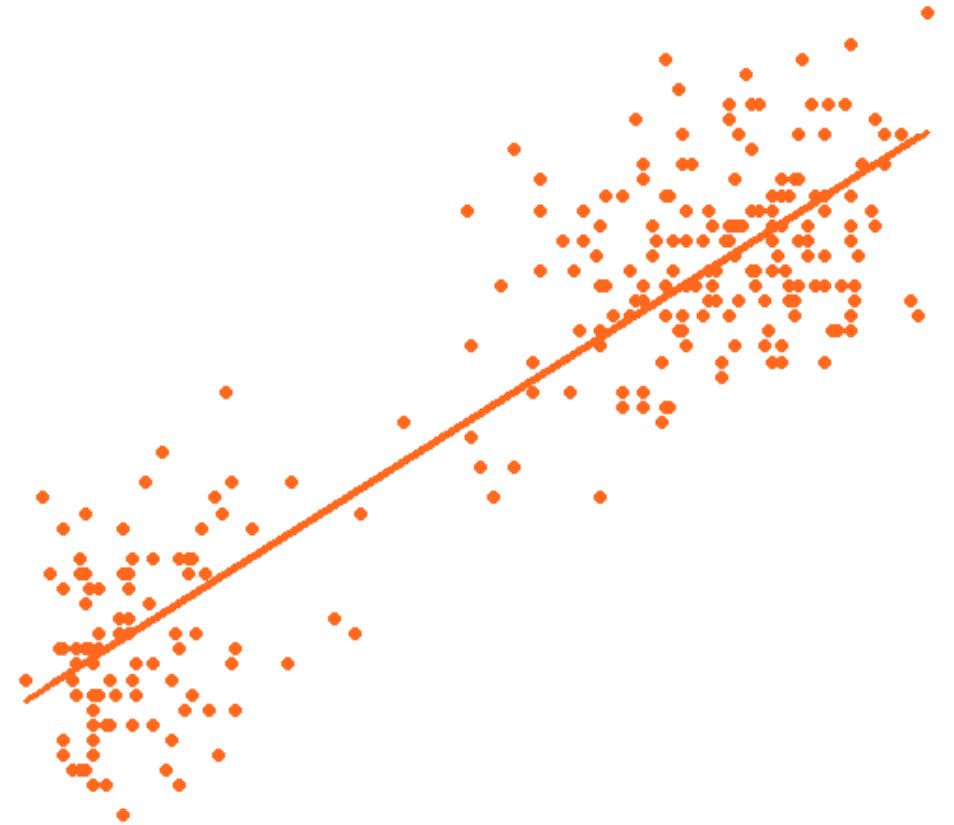
# Simple Linear Regression

Relationship

Linear model

Explanatory variable

Outcome variable



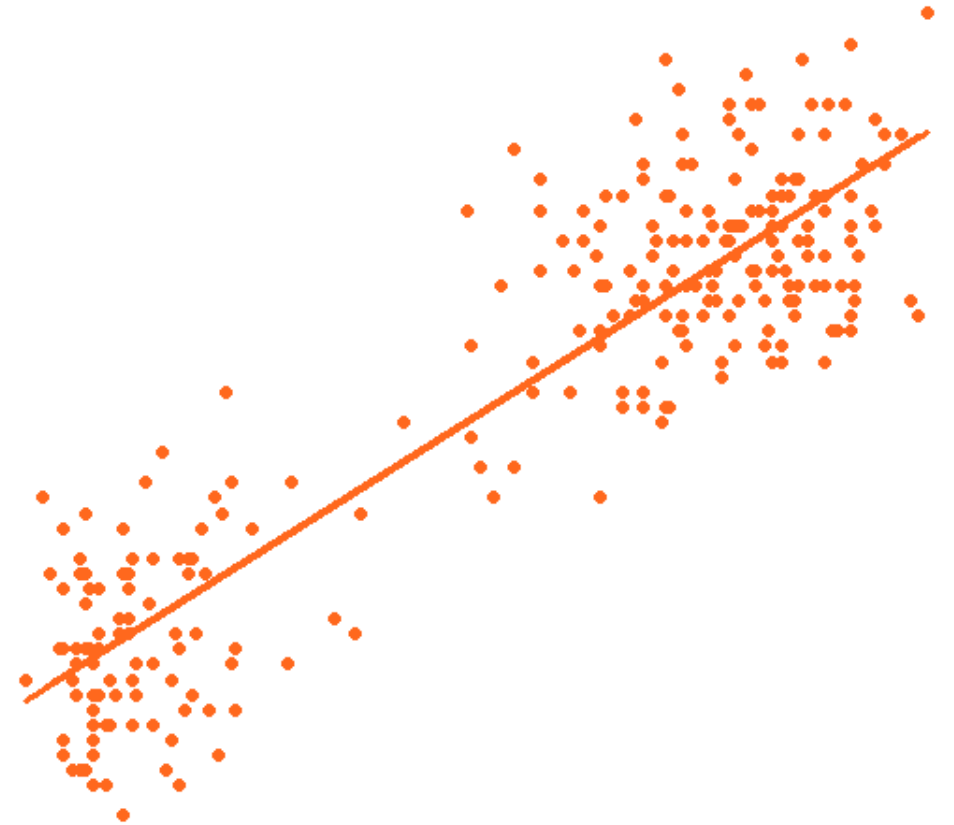
# Simple Linear Regression

Linear predictor function

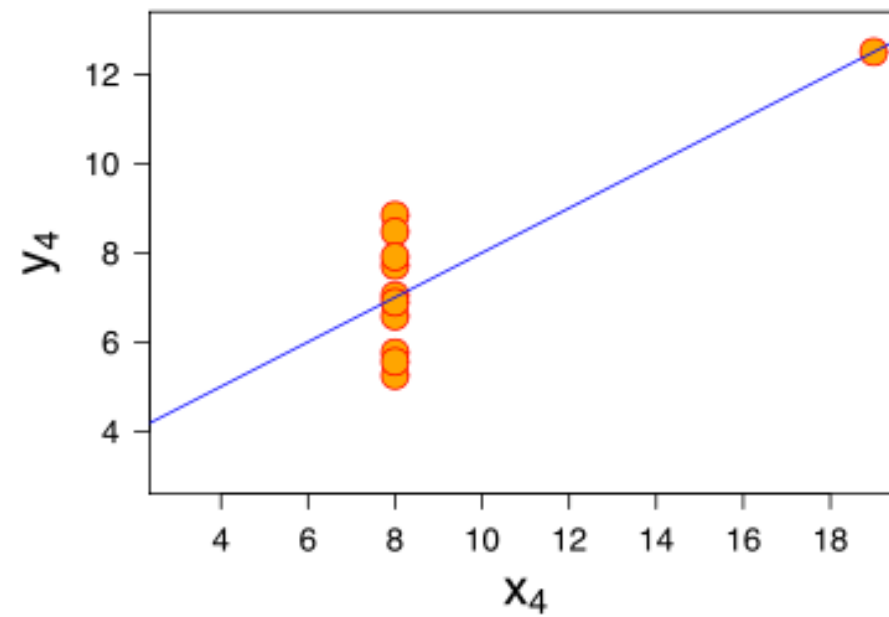
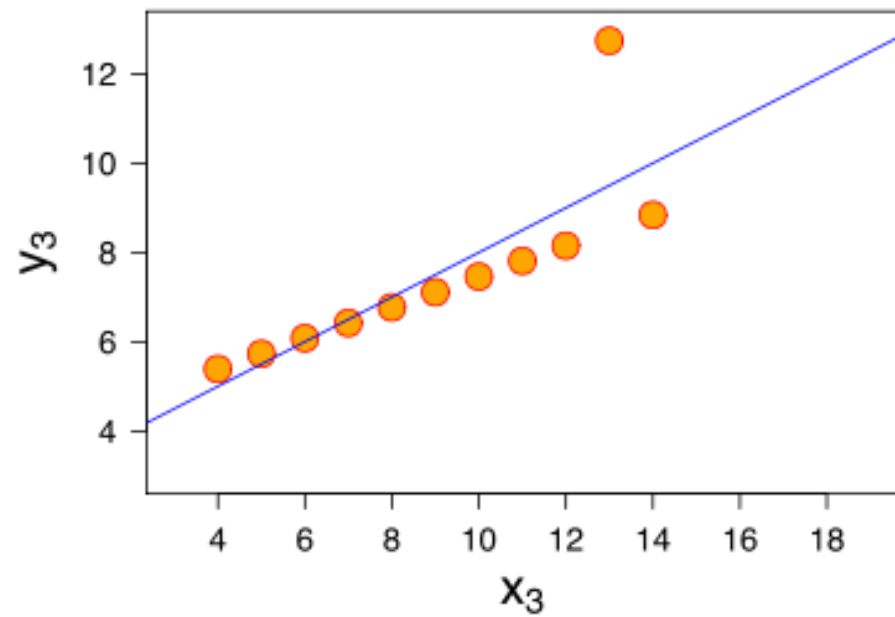
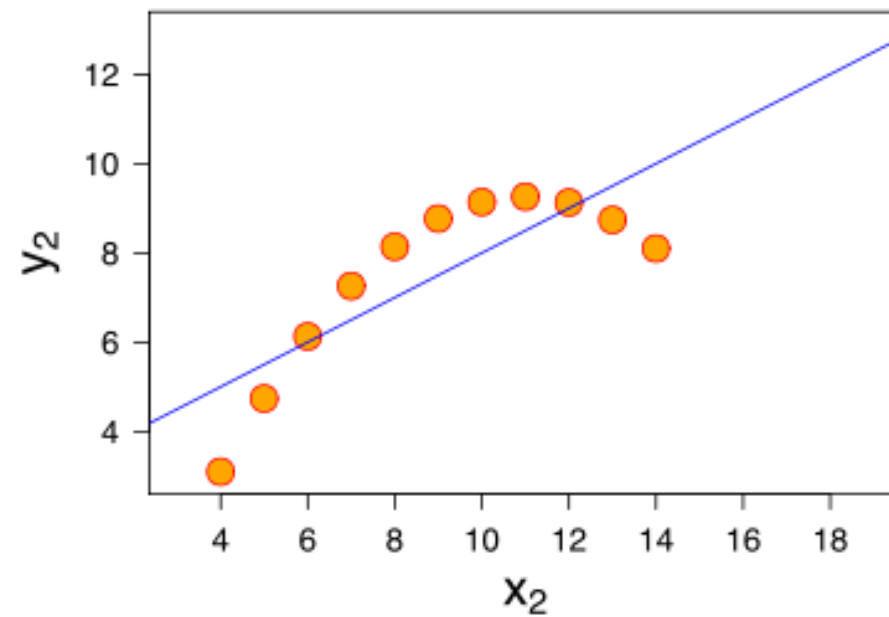
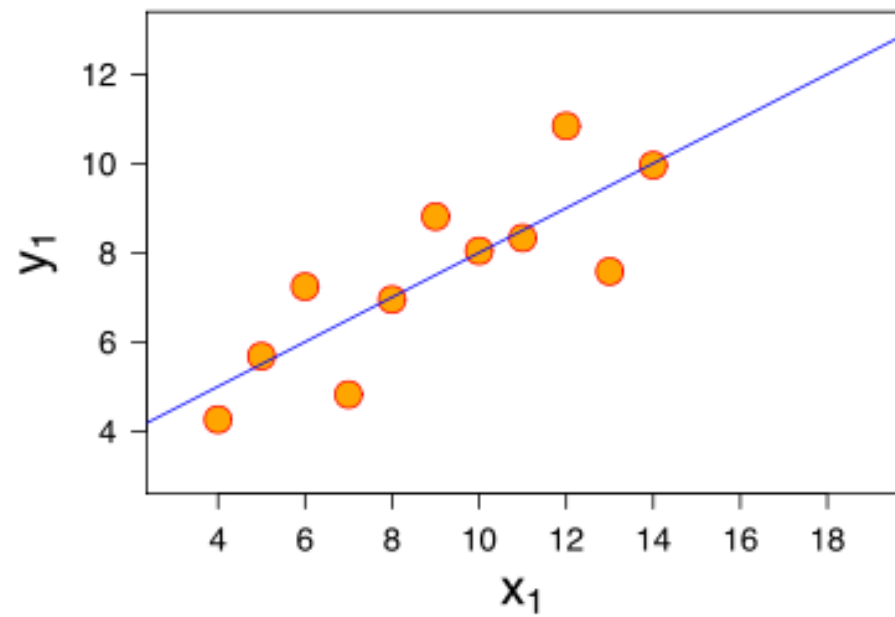
$$y = m \cdot x + b$$

Parameters estimated

Relies on assumptions







# Iris Data Set



**Iris Setosa**



**Iris Versicolor**



**Iris Virginica**

# Iris Data Set

Fisher's Iris Data				
Species	Petal Length	Petal Width	Sepal Length	Sepal Width
setosa	1.1	0.1	4.3	3.0
setosa	1.4	0.2	4.4	2.9
setosa	1.3	0.2	4.4	3.0
setosa	1.3	0.2	4.4	3.2
setosa	1.3	0.3	4.5	2.3
...		...	...	...











1. How are petal length and petal width related?
2. Can we predict new petal width values?



# Load the Data



# Create a Scatterplot



# Create a Linear Model





# Summarize the Model



# Draw Linear Regression Line

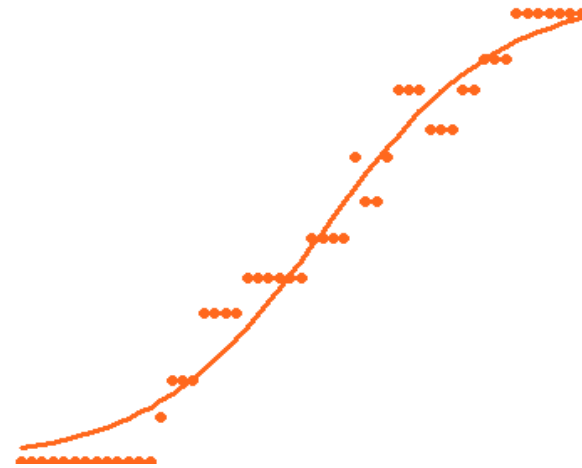
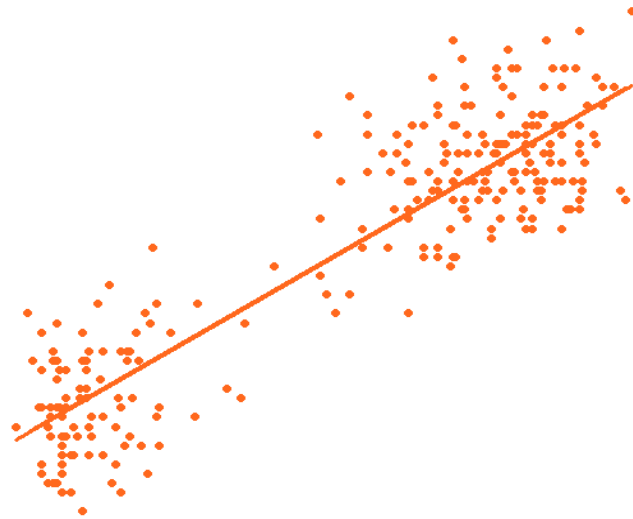
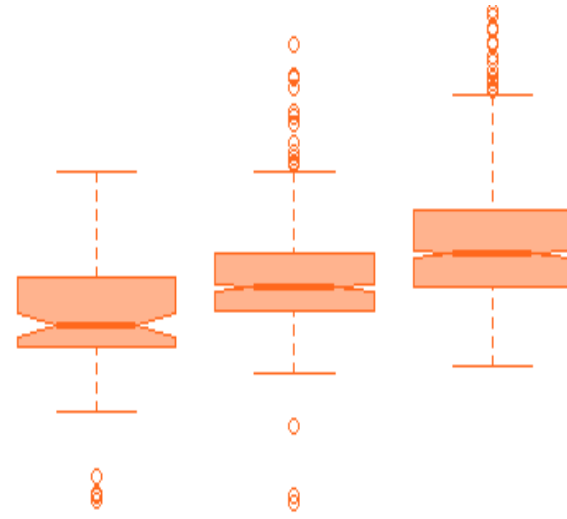
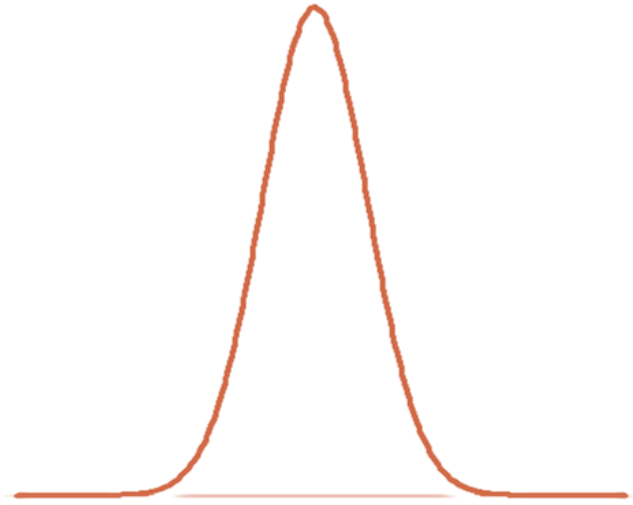


# Get Correlation Coefficient



# Predict New Values





# Summary



**Creating Statistical Models**

**Demo**

