Modeling For Inference

Modeling for inference

- Descriptive or causal inference
- Focus on understanding the data generating process and its component parts
- Concerned with statistical significance
- Theory driven
- Hypothesis testing
- Regression
- White box

Modeling for prediction

- Forecasting
- Not concerned with model components
- Concerned with out of sample prediction
- Less theory driven
- Wider variety of methods used
- Black box

Inferential Modeling With Regression

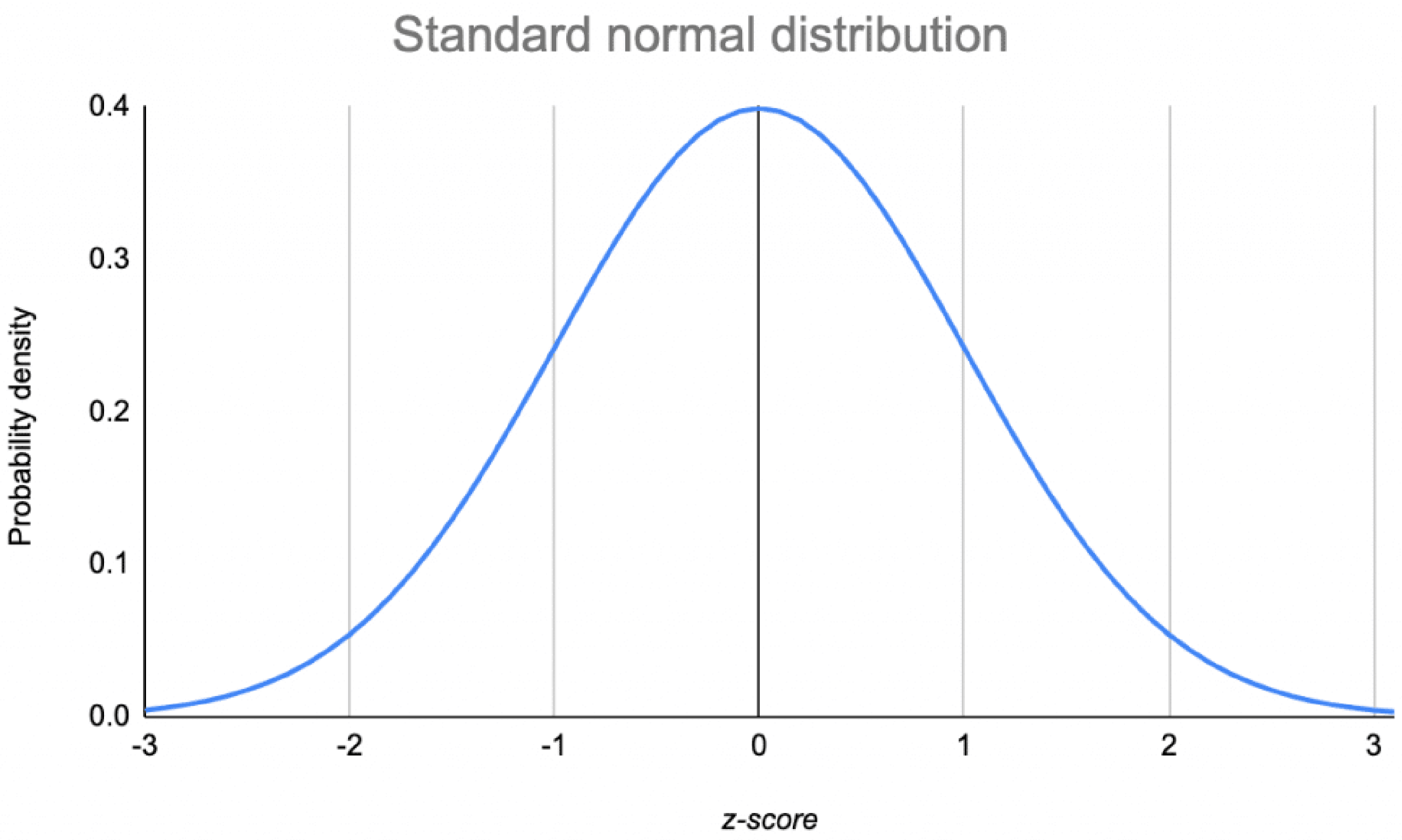
Constant/Intercept

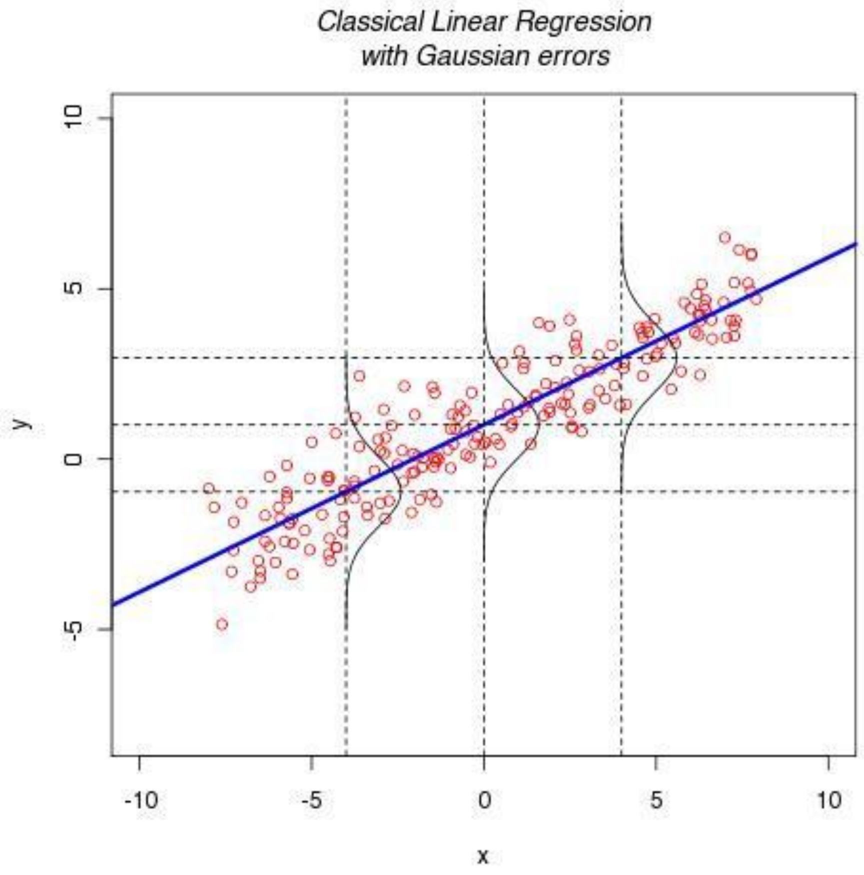
Independent Variable

$$\boldsymbol{Y}_i = \beta_0 + \beta_1 \boldsymbol{X}_i$$

Dependent Variable

Slope/Coefficient


Standard normal distribution
Probability density
0.4
0.3
0.2
0.1
0.0
-3
-2
-1
0
1
2
3
z-score



Classical Linear Regression
with Gaussian errors
y
x
10
5
0
-5
-10
-5
0
5
10

Generalized Linear Models

GLMs

1. An exponential distribution for modeling Y
2. A linear predictor
3. A link function

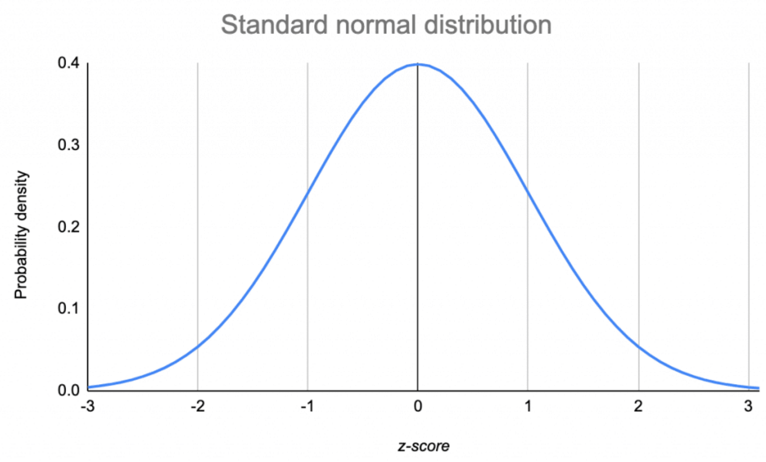
A distribution for Y

Probability Distribution: *A mathematical function that gives the probability of a random phenomenon occurring in terms of its sample space.*

A distribution for Y

Probability Distribution: *A mathematical function that gives the probability of a random phenomenon occurring in terms of its sample space.*

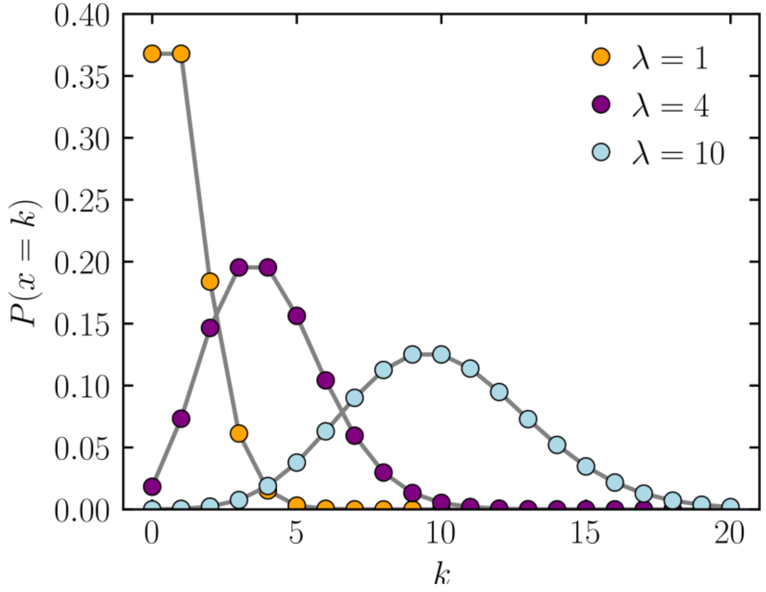
- Normal: Continuous random variables



A distribution for Y

Probability Distribution: *A mathematical function that gives the probability of a random phenomenon occurring in terms of its sample space.*

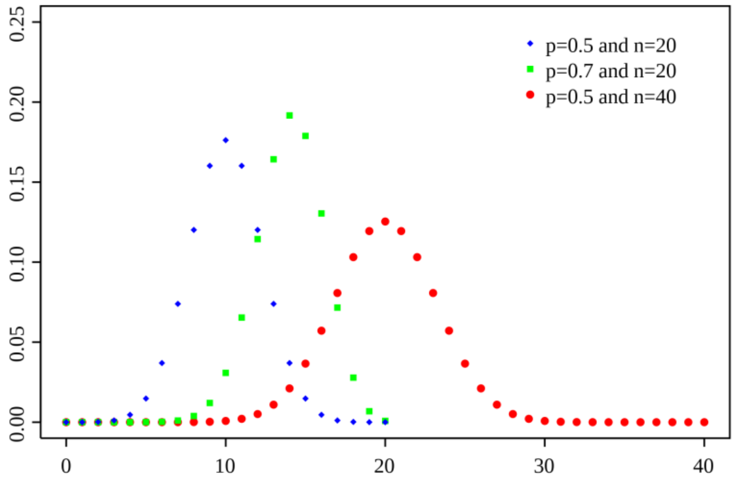
- Normal: Continuous random variables
- Poisson and negative binomial: Count of events



A distribution for Y

Probability Distribution: *A mathematical function that gives the probability of a random phenomenon occurring in terms of its sample space.*

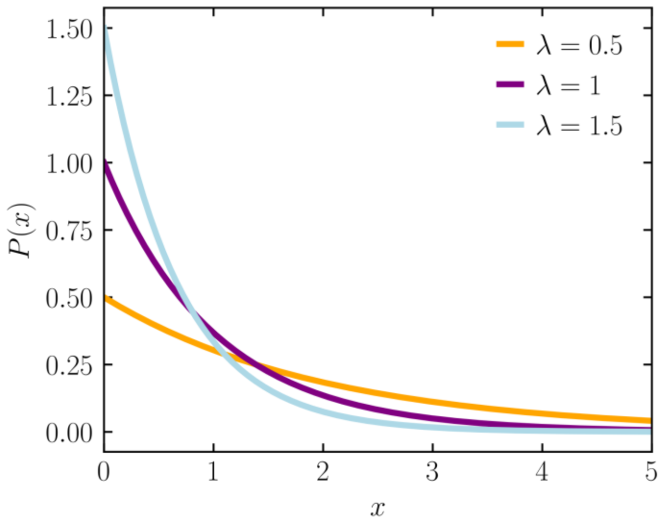
- Normal: Continuous random variables
- Poisson and negative binomial: Count of events
- Binomial: Number of positive results in independent experiments.



A distribution for Y

Probability Distribution: *A mathematical function that gives the probability of a random phenomenon occurring in terms of its sample space.*

- Normal: Continuous random variables
- Poisson and negative binomial: Count of events
- Binomial: Number of positive results in independent experiments.
- Exponential, Gamma, Weibull: Time to event.



A linear predictor

$$\beta_0 + \beta_1 X_i$$

A link function

- “Links” the linear predictor to the probability distribution.

Given link function g:

$$E(Y|X) = \mu = g^{-1}(\beta_0 + \beta_1 X_i)$$

Distribution	Support of distribution	Typical uses	Link name	Link function, $\mathbf{X}\boldsymbol{\beta} = g(\mu)$	Mean function
Normal	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbf{X}\boldsymbol{\beta} = \mu$	$\mu = \mathbf{X}\boldsymbol{\beta}$
Exponential	real: $(0, +\infty)$	Exponential-response data, scale parameters	Negative inverse	$\mathbf{X}\boldsymbol{\beta} = -\mu^{-1}$	$\mu = -(\mathbf{X}\boldsymbol{\beta})^{-1}$
Gamma					
Inverse Gaussian	real: $(0, +\infty)$		Inverse squared	$\mathbf{X}\boldsymbol{\beta} = \mu^{-2}$	$\mu = (\mathbf{X}\boldsymbol{\beta})^{-1/2}$
Poisson	integer: $0, 1, 2, \ldots$	count of occurrences in fixed amount of time/space	Log	$\mathbf{X}\boldsymbol{\beta} = \ln(\mu)$	$\mu = \exp(\mathbf{X}\boldsymbol{\beta})$
Bernoulli	integer: $\{0, 1\}$	outcome of single yes/no occurrence	Logit	$\mathbf{X}\boldsymbol{\beta} = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(\mathbf{X}\boldsymbol{\beta})}{1+\exp(\mathbf{X}\boldsymbol{\beta})} = \frac{1}{1+\exp(-\mathbf{X}\boldsymbol{\beta})}$
Binomial	integer: $0, 1, \ldots, N$	count of # of "yes" occurrences out of N yes/no occurrences		$\mathbf{X}\boldsymbol{\beta} = \ln\left(\frac{\mu}{n-\mu}\right)$	
Categorical	integer: $[0, K)$	outcome of single K-way occurrence		$\mathbf{X}\boldsymbol{\beta} = \ln\left(\frac{\mu}{1-\mu}\right)$	
	K-vector of integer: $[0, 1]$, where exactly one element in the vector has the value 1				
Multinomial	*K*-vector of integer: $[0, N]$	count of occurrences of different types (1 .. *K*) out of *N* total *K*-way occurrences			

GLMs

1. An exponential distribution for modeling Y
2. A linear predictor
3. A link function

GLMs

1. An exponential distribution for modeling Y ⟸ Theory driven
2. A linear predictor
3. A link function

GLMs

1. An exponential distribution for modeling Y ⟵ Theory driven
2. A linear predictor
3. A link function ⟵ Done by R

GLMs

1. An exponential distribution for modeling Y ⟵ Theory driven
2. A linear predictor ⟵ The hard part
3. A link function ⟵ Done by R

How is your degree different from a stats or information science degree?

What makes modeling human behavior so difficult?

Vocabulary

- Inferential modeling
- Predictive modeling
- GLM
- Probability distribution
- Link function
- Normal, Poisson, Binomial, and Exponential distributions