Probability Distributions

Francisco Luquero
Epicentre

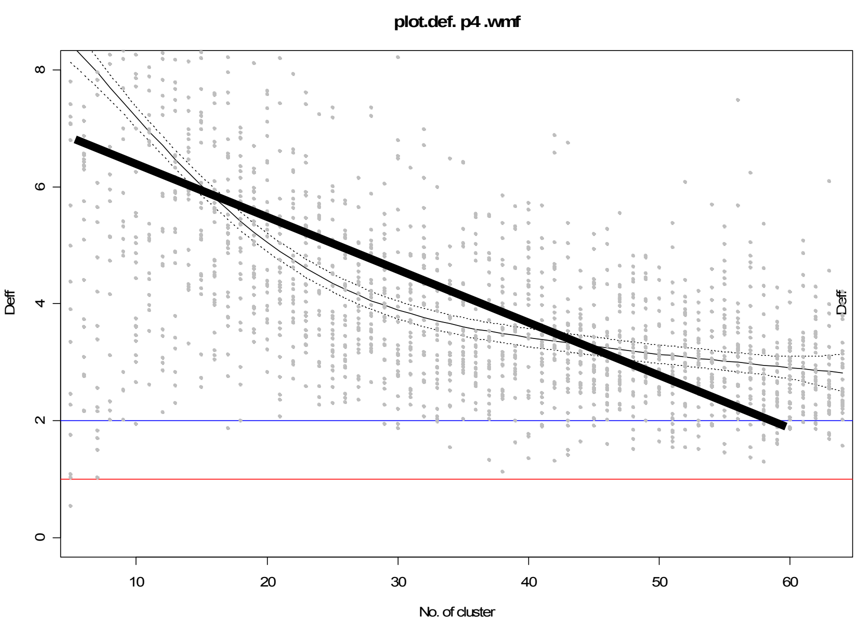
Madrid, June 2009

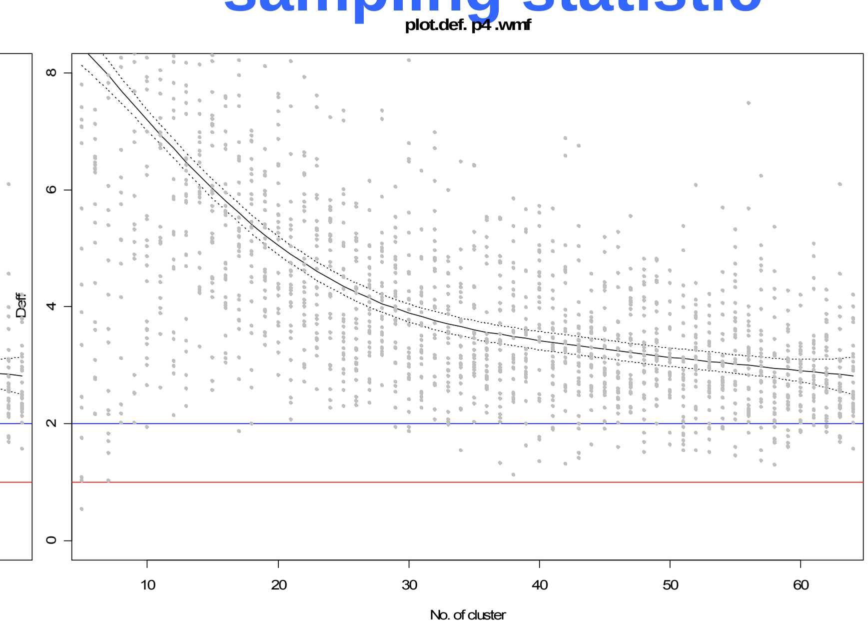
Objectives

- To understand the concept of probability distribution
- To associate common probability distributions to certain types of variables

Parametric Vs. non parametric aproach

- **Parametric: decision making method where the distribution of the sampling statistic is known**
- **Non-Parametric: decision making method which does not require knowledge of the distribution of the sampling statistic**





Parametric Vs.
non parametric aproach

- **Parametric: decision making method where the distribution of the sampling statistic is known**

- **Non-Parametric: decision making method which does not require knowledge of the distribution of the sampling statistic**

Probability Mass Function

Let X be a discrete random variable with possible values x_0, x_1, x_2, x_3, ……x_k and the corresponding probabilities $p(x_0)$, $p(x_1)$, $p(x_2)$, $p(x_3)$, ……. $P(x_k)$.

Example: imagine X to be number of children with possible values 0, 1, 2, 3, ……10

For any i,

$$p(x_i) \geq 0 \text{ and } \sum_{i=0}^{k} p(x_i) = 1.$$

Then $p(x_i)$ is a probability mass function.

Let X be a discrete random variable and its corresponding probabilities are:

$p(0) = 0.04$

$p(1) = 0.18$

$p(2) = 0.27$

$p(3) = 0.25$

$p(4) = 0.15$

$p(5) = 0.07$

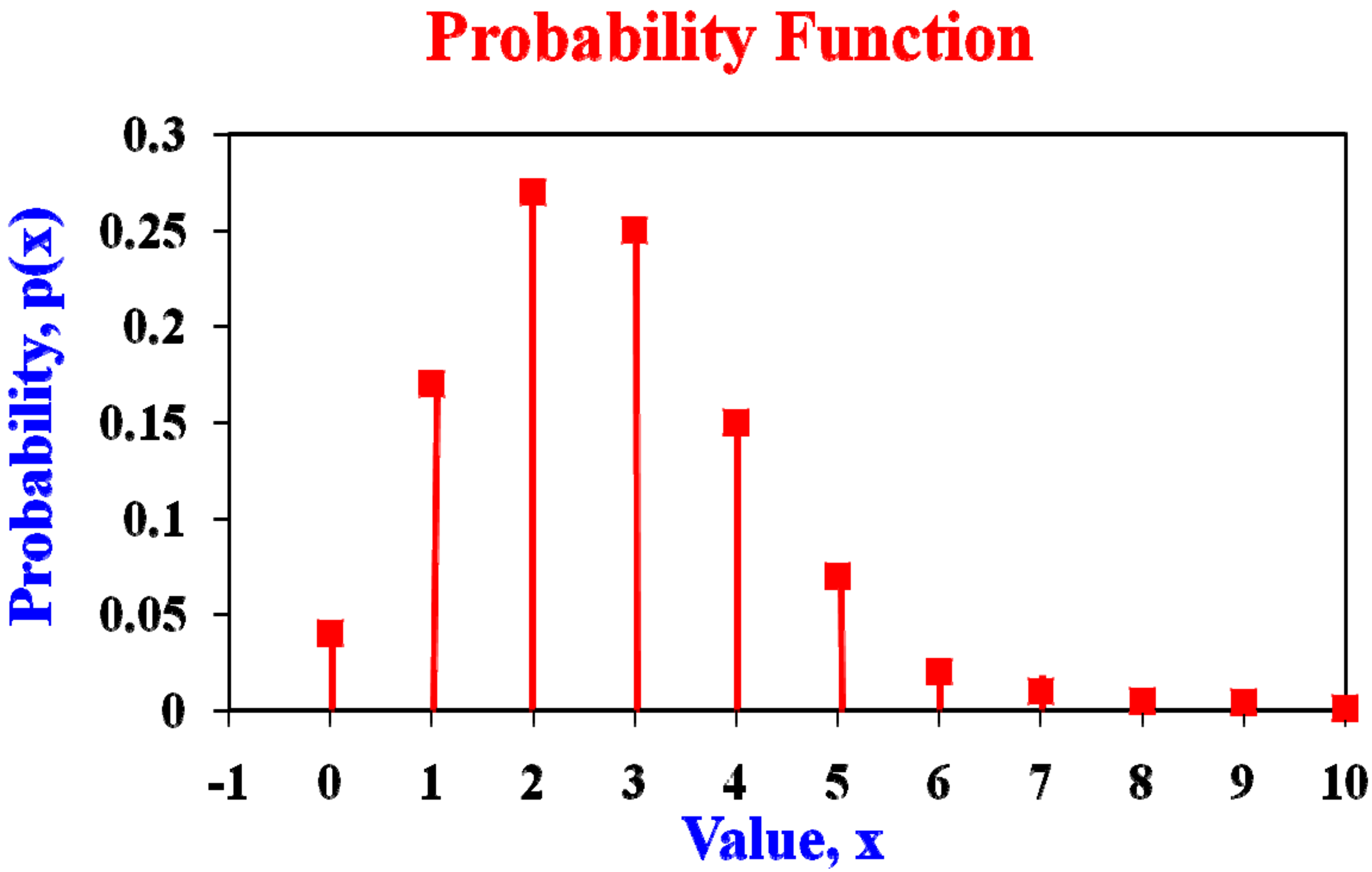
$p(6) = 0.02$

$p(7) = 0.01$

$p(8) = 0.005$

$p(9) = 0.004$

$p(10) = 0.001$

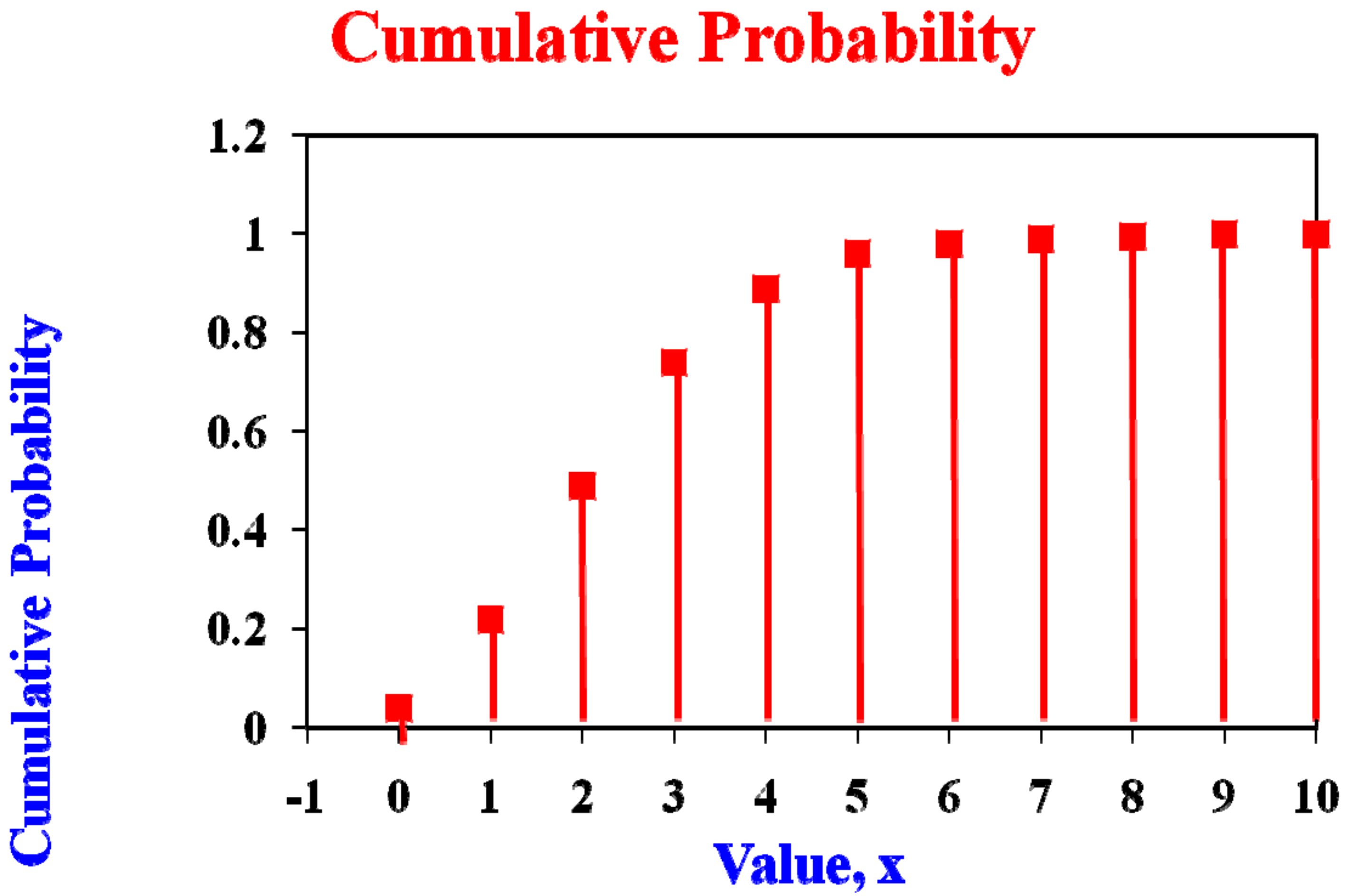

Probability Function
0.3
0.25
0.2
0.15
0.1
0.05
0
Probability, p(x)
-1
0
1
2
3
4
5
6
7
8
9
10
Value, x

Cumulative probabilities are found by adding individual probabilities.

$$\Pr[X \le x] = \sum_{x_i \le x} p(x_i)$$

$$\Pr[X \le 3] = p(0) + p(1) + p(2) + p(3)$$

x_i	$p(x_i)$	$Pr[X \leq x_i]$
0	0.040	0.040
1	0.180	0.220
2	0.270	0.490
3	0.250	0.740
4	0.150	0.890
5	0.070	0.960
6	0.020	0.980
7	0.010	0.990
8	0.005	0.995
9	0.004	0.999
10	0.001	1.000



Cumulative Probability
1.2
1
0.8
0.6
0.4
0.2
0
-1
0
1
2
3
4
5
6
7
8
9
10
Value, x
Cumulative Probability

Two important concepts

- Expectation
- Variance

Two important concepts

- Expectation
- Variance

Expectation of a Discrete Random Variable

The expected value is the mean of all possible results for an infinite number of trials.

The expected value of a random variable, X is denoted by E(X).

$$E(X) = \sum_{\text{all } x_i} (x_i) p(x_i)$$

x_i	$p(x_i)$	$Pr[X \le x_i]$
0	0.040	0.040
1	0.180	0.220
2	0.270	0.490
3	0.250	0.740
4	0.150	0.890
5	0.070	0.960
6	0.020	0.980
7	0.010	0.990
8	0.005	0.995
9	0.004	0.999
10	0.001	1.000

Expectation of a Discrete Random Variable

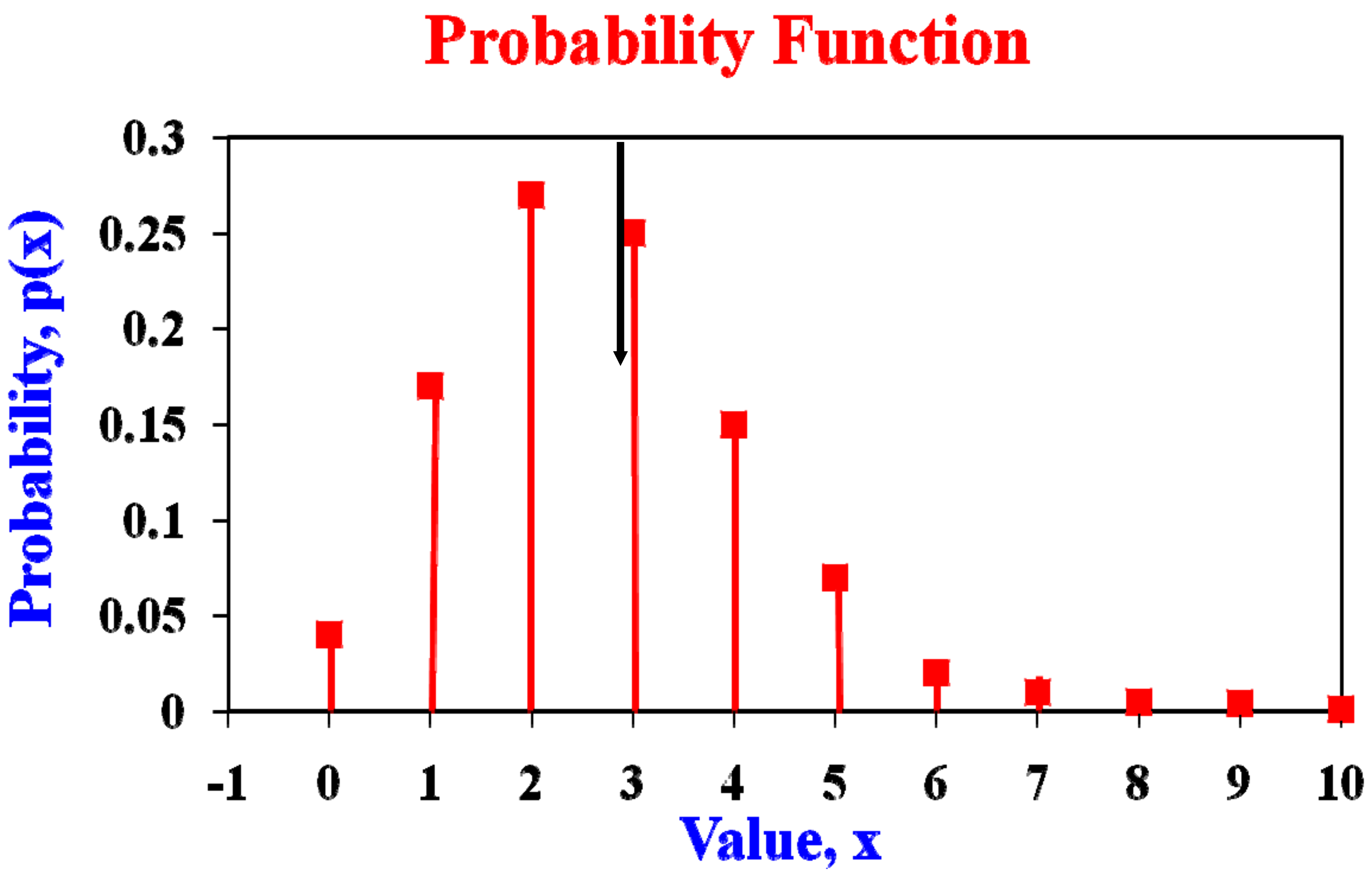
For the example of the discrete random variable, X

$$\begin{aligned}E(X) = (0)\times p(0) + (1)\times p(1) + (2)\times p(2) \\ + (3)\times p(3) + (4)\times p(4) + (5)\times p(5) \\ + (6)\times p(6) + (7)\times p(7) + (8)\times p(8) \\ + (9)\times p(9) + (10)\times p(10)\end{aligned}$$

Expectation of a Discrete Random Variable

$$
\begin{aligned}
\mathbf{E}(\mathbf{X}) = &(0)\times 0.04+(1)\times 0.18+(2)\times 0.27 \\
&+(3)\times 0.25+(4)\times 0.15+(5)\times 0.07 \\
&+(6)\times 0.02+(7)\times 0.01+(8)\times 0.005 \\
&+(9)\times 0.004+(10)\times 0.001
\end{aligned}
$$

$$\mathbf{E}(\mathbf{X}) = 2.696$$

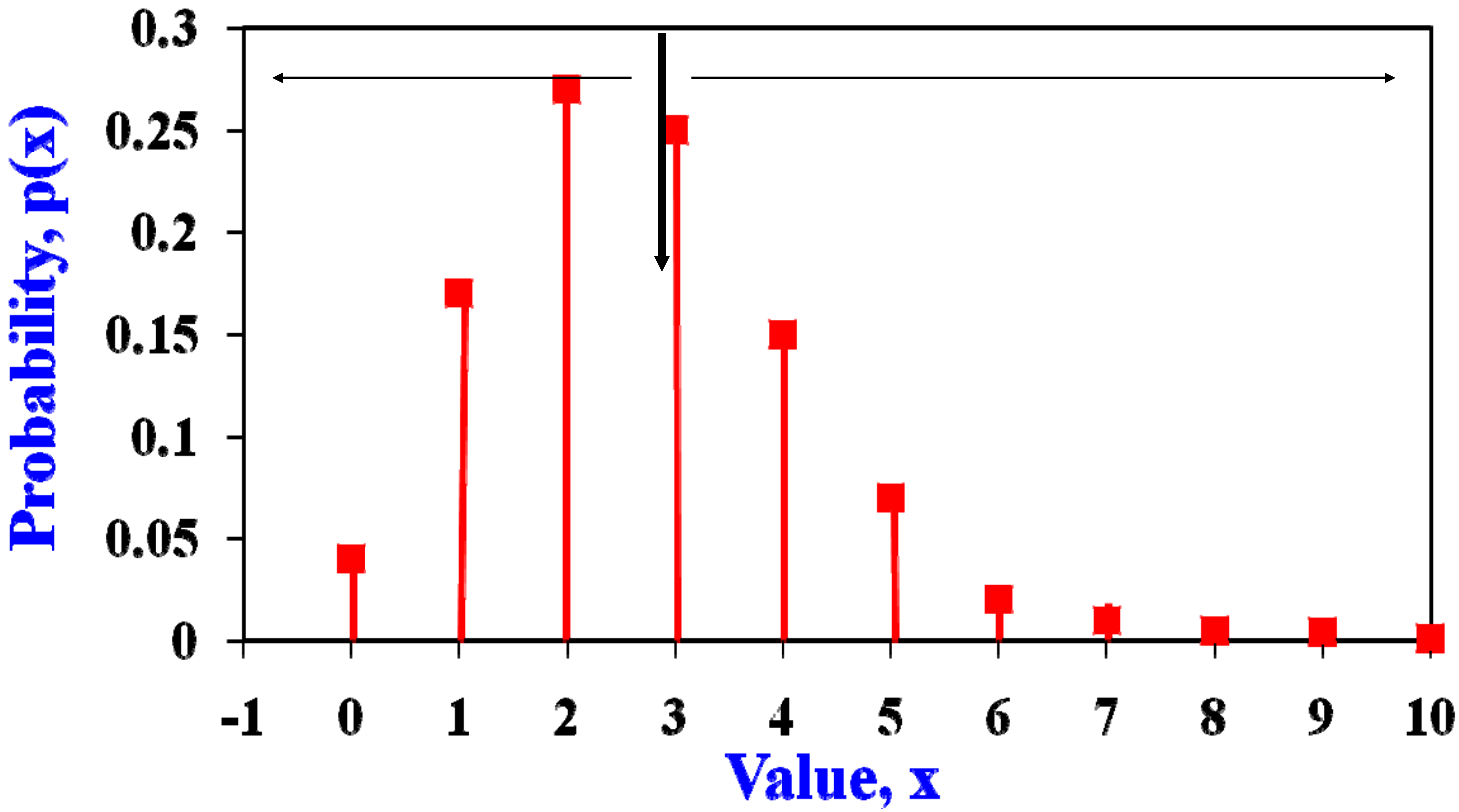


Probability Function
Probability, p(x)
0.3
0.25
0.2
0.15
0.1
0.05
0
-1
0
1
2
3
4
5
6
7
8
9
10
Value, x

Two important concepts

- Expectation
- Variance

Probability Function



Variance of a Discrete Random Variable

$$\sigma_x^2 = \frac{\sum_{i=1}^{N} \left(x_i - \mu_x\right)^2}{N}$$

$$\left(\sum_{i=1}^{N} \left(x_i - \mu_x\right)^2\right) / N$$

Variance of a Discrete Random Variable

$$\left(\sum_{i=1}^{N} \left(x_i - \mu_x \right)^2 \right) / N$$

is the mean value of $\left(x_i - \mu_x \right)^2$

$$\sigma_x^2 = E(X - \mu_x)^2$$

Two important concepts

- Expectation: more probable value

- Variance: dispersion of the values

Three main distribution (regression)

- Gaussian (normal): linear regression
- Binomial: logistic regresion
- Poisson: poisson regresion

Other Probability Distributions: Special Cases

- Gaussian Distribution:
  - The normal distribution or Gaussian distribution is a continuous probability distribution
  - Describes data that clusters around a mean or average.
  - The graph of the associated probability density function is bell-shaped, with a peak at the mean
  - It is known as the Gaussian function or bell curve.

  - The normal distribution can be used to describe, at least approximately, any variable that tends to cluster around the mean.

Why transform data?

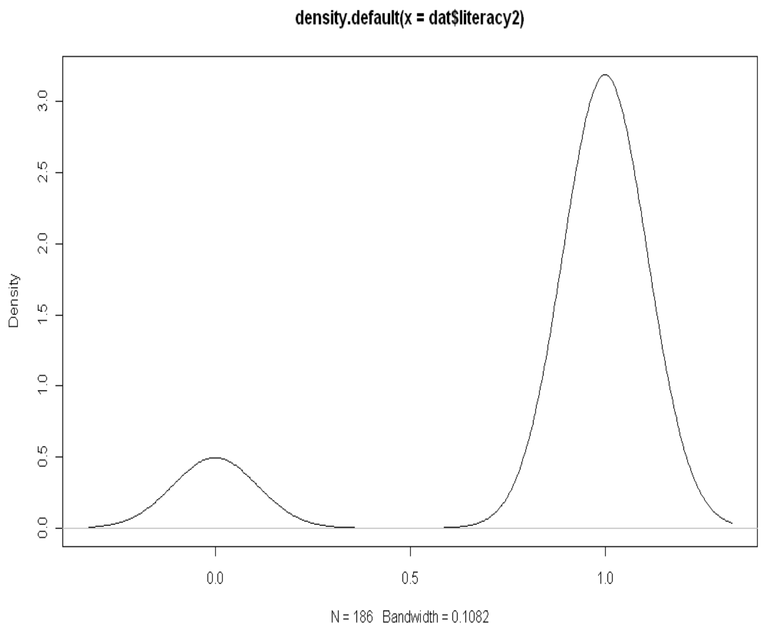
- In some instances it can help us better examine a distribution
- Many statistical models are based on the mean and thus require that the mean is an appropriate measure of central tendency (*i.e.*, the distribution is approximately normal)
- Linear regression assumes that the relationship between two variables is linear. Often we can “straighten” a nonlinear relationship by transforming one or both of the variables
  - Often transformations will ‘fix’ problem distributions so that we can use least-squares regression
  - When transformations fail to remedy these problems, another option is to use nonparametric regression, which makes fewer assumptions about the data

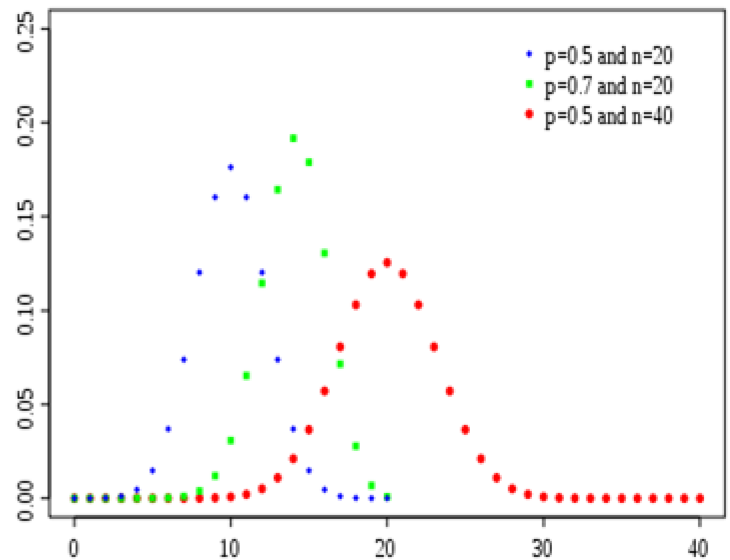
Power transformations for quantitative variables

- Although there are an infinite number of functions *f(x)* that can be used to transform a distribution in practice only a relatively small number are regularly used
- For quantitative variables one can usually rely on the “family” of powers and roots: x^2 or $x^{0.5}$
- When *p* is negative, the transformation is an inverse power:
- When *p* is a fraction, the transformation represents aroot:

Other Probability Distributions: Special Cases

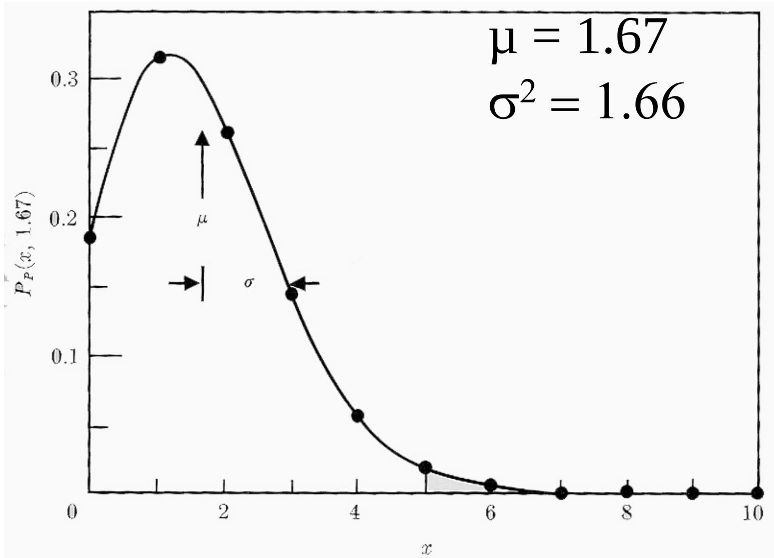
- Binomial Distribution:
  - The binomial distribution is the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments
  - Each of these success have probability p.
  - Such a success/failure experiment is also called a Bernoulli experiment or Bernoulli trial.
  - In fact, when n = 1, the binomial distribution is a Bernoulli distribution

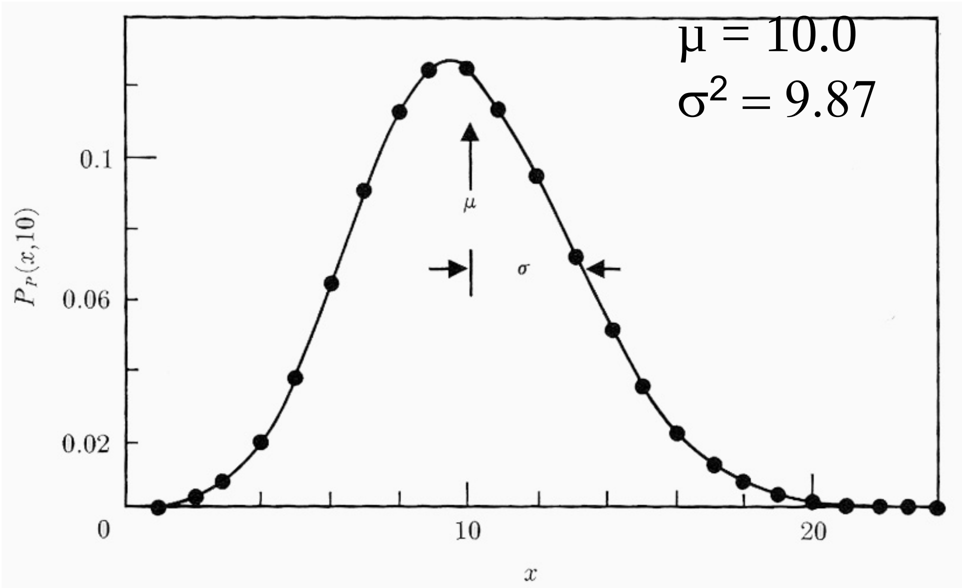




Other Probability Distributions: Special Cases

- Poisson Distribution:
  - The Poisson distribution' is a discrete probability distribution
  - Expresses the probability of a number of events occurring in a fixed period of time
  - The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area or volume.
  - It is an approximation to the binomial distribution for the special case when the average number of successes is very much smaller than the possible number
  - The possible values are higher than 0
  - The mean (λ) and the variance (λ) are the same





Variables vs Distributions

Type of Variable		Distribution
Qualitative	Dichotomous	Binomial
Quantitative	Continuous	Gaussian
Quantitative	Discrete	Poisson