Inspire…Educate…Transform.

Foundations of Statistics and Probability for Data Science

Probability Distributions: Discrete and Continuous, Sampling Distribution of Means, CLT

Prof Anuradha Sharma

December 9, 2018

MATERIAL CONTENT FROM Dr. SRIDHAR PAPPU

Analyzing attributes

PROBABILITY DISTRIBUTIONS

Describing a Distribution – Summary of Moments – [Revision]

Measure	Formula	Description
Mean (μ)	$E(X)$	Measures the center of the distribution of X
Variance (σ^2)	$E[(X-\mu)^2]$	Measures the spread of the distribution of X about the mean
Skewness	$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$	Measures asymmetry of the distribution of X
Kurtosis (excess)	$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]-3$	Measures 'tailed'ness of the distribution of X and useful in outlier identification

Measure	Formula	Description
		Measures the center of the distribution of X
		Measures the spread of the distribution of X about the mean
Skewness		Measures asymmetry of the distribution of X
Kurtosis (excess)		Measures 'tailed'ness of the distribution of X and useful in outlier identification

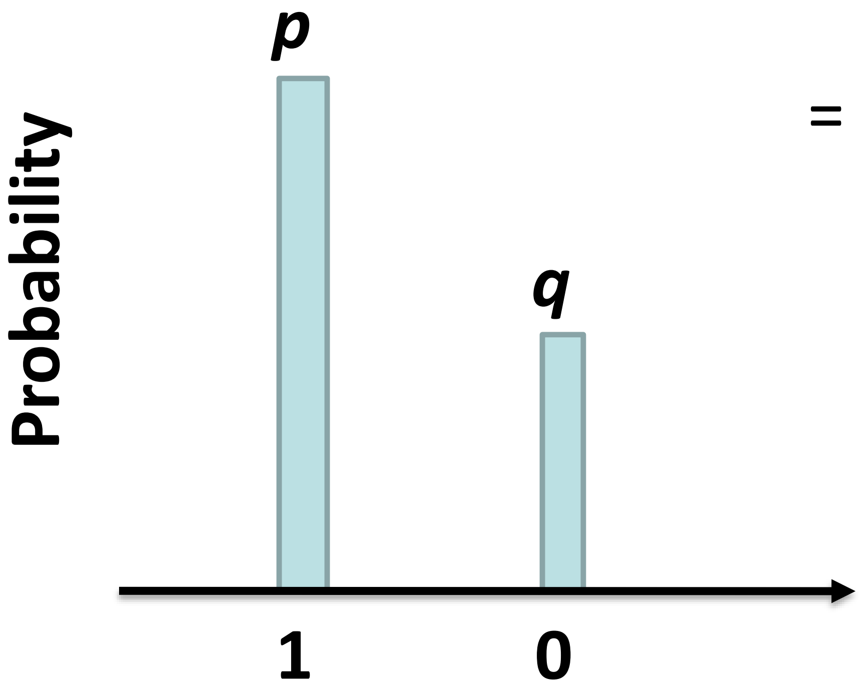
SOME COMMON DISTRIBUTIONS

Bernoulli

There are two possibilities (loan taker or non-taker) with probability p of success and $1-p$ of failure

- Expectation: p
- Variance: $p(1-p)$ or pq, where $q=1-p$

Bernoulli



$\Rightarrow$ has two values 1 and 0

$\Rightarrow$ Whenver we have two value success is defined as "1" and failure "0"

$$= Expectation, E(X) = \sum x_i P(x_i) =$$

$\Rightarrow x_i$ has two values 1 and 0

$\Rightarrow$ Whenver we have two value success is defined as "1" and failure "0"

$$= 1 * p + 0 * q = p$$

$$Variance, Var = \sum (x_i - \mu)^2 P(x_i)$$

$$= (1-p)^2 * p + (0-p)^2 * (1-p)$$
$$= p(1-p)$$

Geometric Distribution

Number of independent and identical (i.i.d) Bernoulli trials needed to get ONE success, e.g., number of people I need to call for the first person to accept the loan.

Geometric Distribution

- PMF*, P(X=r) = **q*q*q.....(**r-1 times) * **p** [i.i.d]

$PMF^* \; P(X=r) = q^{r-1}p$ — *(r-1)* failures followed by ONE success.

$P(X > r) = q^r$ — Probability you will need more than *r* trials to get the first success.

CDF**, $P(X \leq r) = 1 - q^r$ — Probability you will need *r* trials or less to get your first success.

Note : P(X>r) = 1- P(X<=r)

$$E(X) = \frac{1}{p} \qquad Var(X) = \frac{q}{p^2}$$

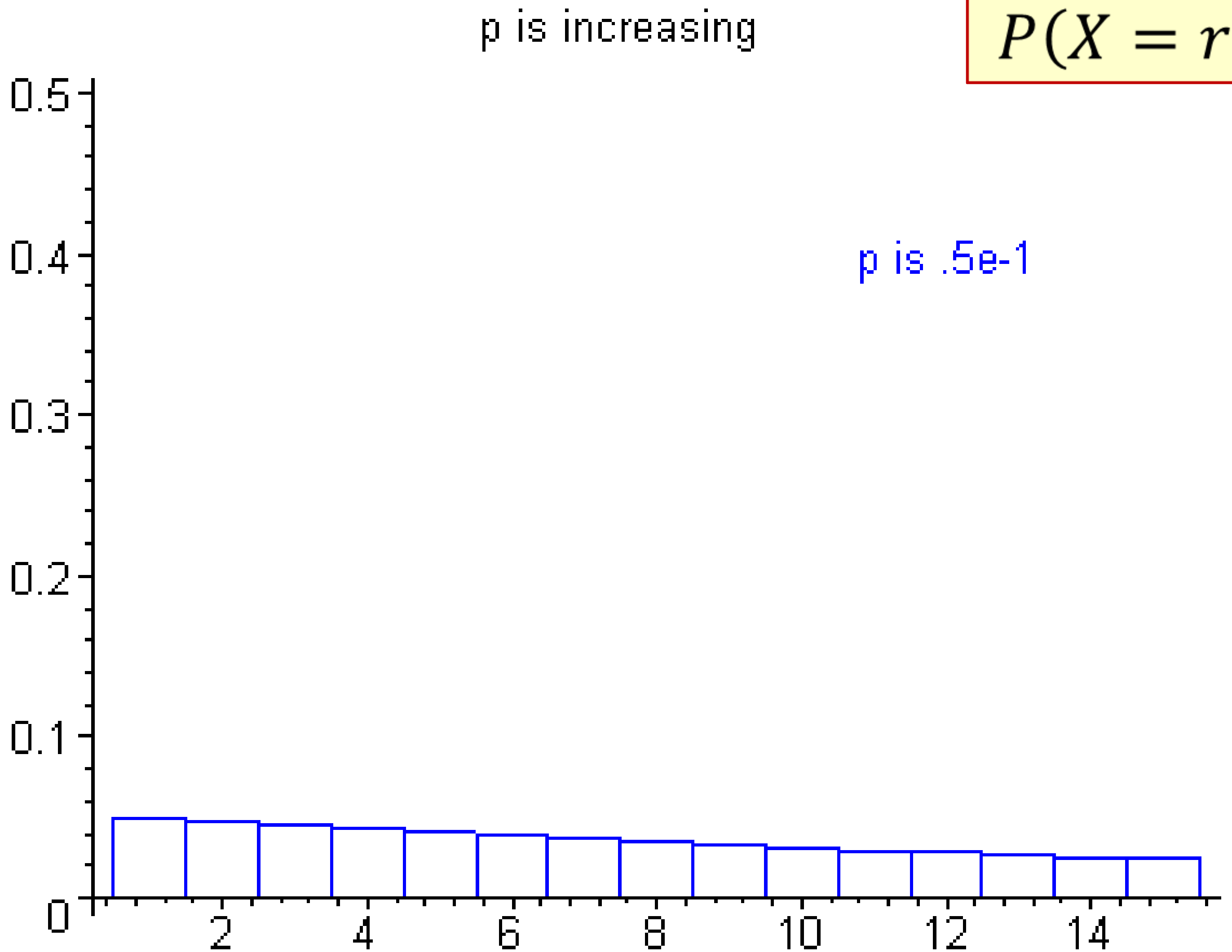
- Probability Mass Function ** Cumulative Distribution Function
- X is the random variable

Geometric Distribution

- You run a series of independent trials.
- There can be either a success or a failure for each trial, and the probability of success is the same for each trial.
- The main thing you are interested in is how many trials are needed in order to get the first successful outcome.
- Geometric Distributions and other distributions that we discuss are called “Parametric Distributions”
  - A parameter is needed to define the distribution
  - Recall Parameter from Day 1 Statistic Class referring to “Population”

X~Geo(p)

$$P(X = r) = q^{r-1}p$$



Ref: http://personal.kenyon.edu/hartlaub/MellonProject/Geometric2.html
Last accessed: June 12, 2015

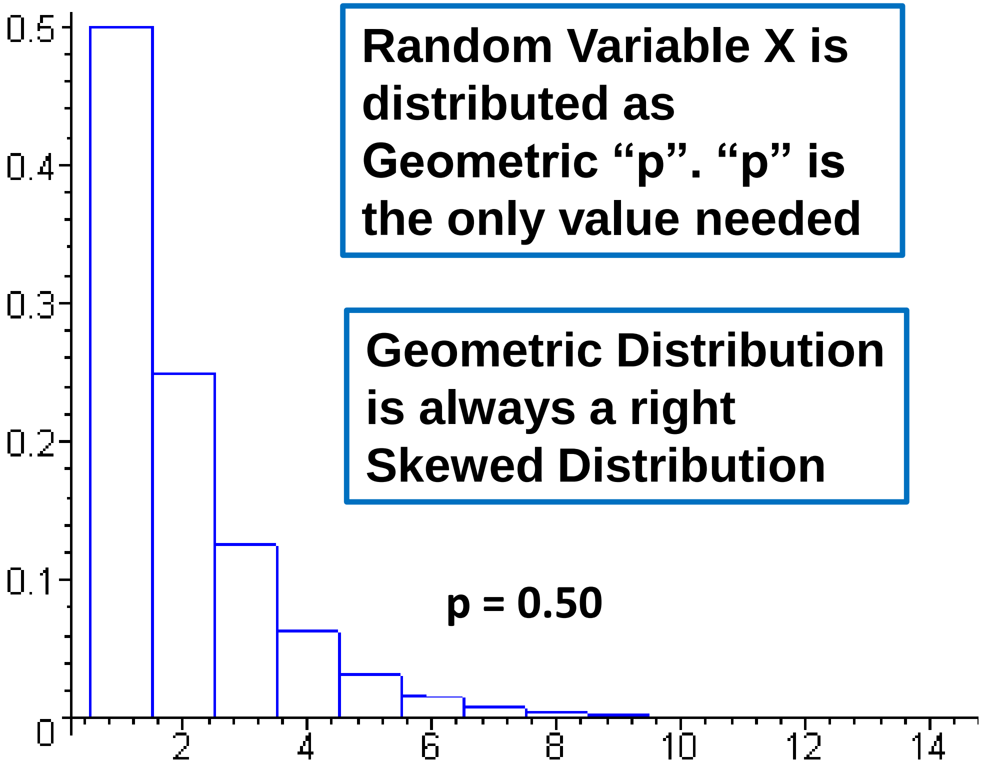
X~Geo(p)

$$P(X = r) = q^{r-1}p$$

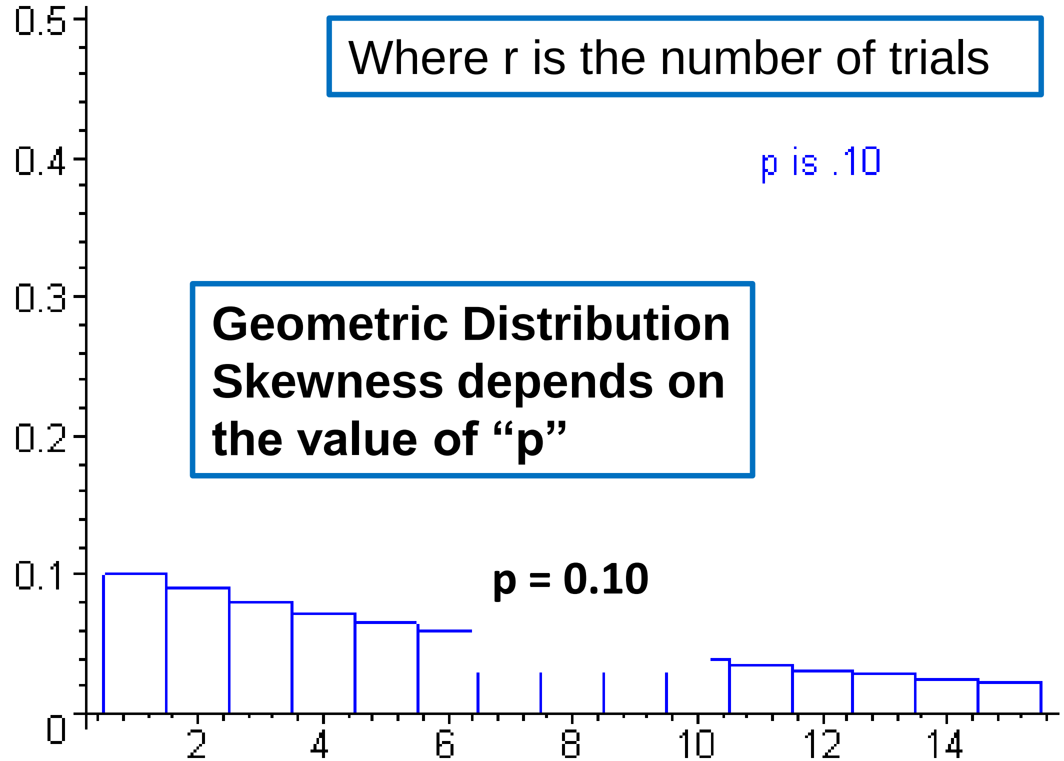
Where r is the number of trials

Random Variable X is distributed as Geometric "p". "p" is the only value needed

Geometric Distribution is always a right Skewed Distribution



Geometric Distribution Skewness depends on the value of "p"



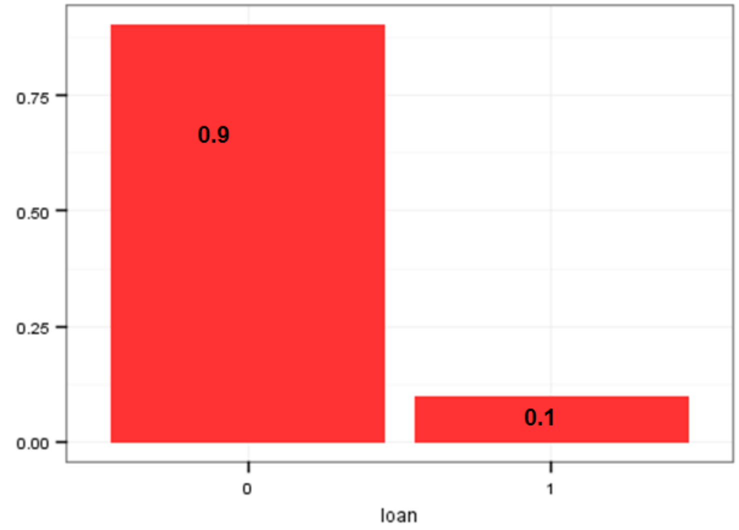
p=0.5, **q=1-0.5 = 0.5**
Probability of Success in first Trial (r=1)
P(X=1) = $q^{1-1}p = q^0p = p =$ **0.5**
Probability of Success in Second Trial (r=2)
P(X=2) = $q^{2-1}p = q^1p = q*p = 0.5*0.5 =$**0.25**

p=0.1, **q=1-0.1 = 0.9**
Probability of Success in first Trial (r=1)
P(X=1) = $q^{1-1}p = q^0p = p =$ **0.1**
Probability of Success in Second Trial (r=2)
P(X=2) = $q^{2-1}p = q^1p = q*p = 0.9*0.1 =$**0.09**

Ref: http://personal.kenyon.edu/hartlaub/MellonProject/Geometric2.html
Last accessed: December 09, 2017

Binomial Distribution

If I randomly pick 10 people, what is the probability that I will get exactly



- 0 person will take a loan = 0.9*0.9*--(10 times)
- **0 person will take a loan =**
- **1 person will take a loa**n = (first person can take a loan) OR (second person can take a loan) OR (third person can take a loan) OR … (10 times)
- **1 person will take a loan =**
- **2 people will take a loan =**
- **3 people will take a loan =**
- **And so on …………**

Binomial Distribution

- If there are two possibilities with probability *p* for success and *q* for failure, and if we perform *n* trials, the probability that we see *r* successes is

PMF, $P(X = r) = C_r^n p^r q^{n-r}$

CDF, $P(X \le r) = \sum_{i=0}^{r} C_i^n p^i q^{n-i}$

Where $\boldsymbol{C_r^n} = \frac{\boldsymbol{n}!}{(\boldsymbol{n}-\boldsymbol{r})! * \boldsymbol{r}!}$ Where n! = n*(n-1)*(n-2)*....*1

For example

$\mathbf{C_3^5} = \frac{5!}{(5-3)! * 3!}. = \frac{5!}{2! * 3!} = \frac{5*4*3*2*1}{(2*1) * (3*2*1)} = \mathbf{10}$

Binomial Distribution

- $E(X) = np$ $\qquad\qquad\qquad\qquad Var(X) = npq$

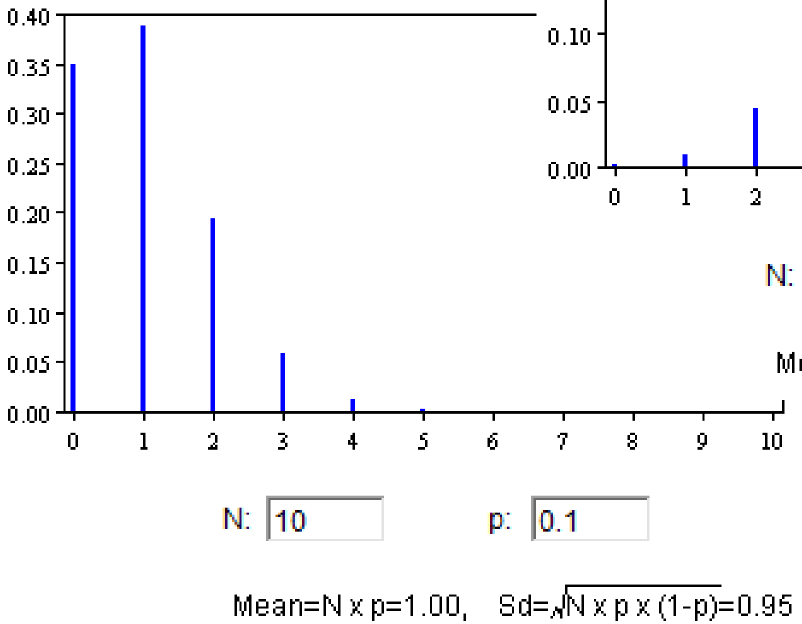
When to use?

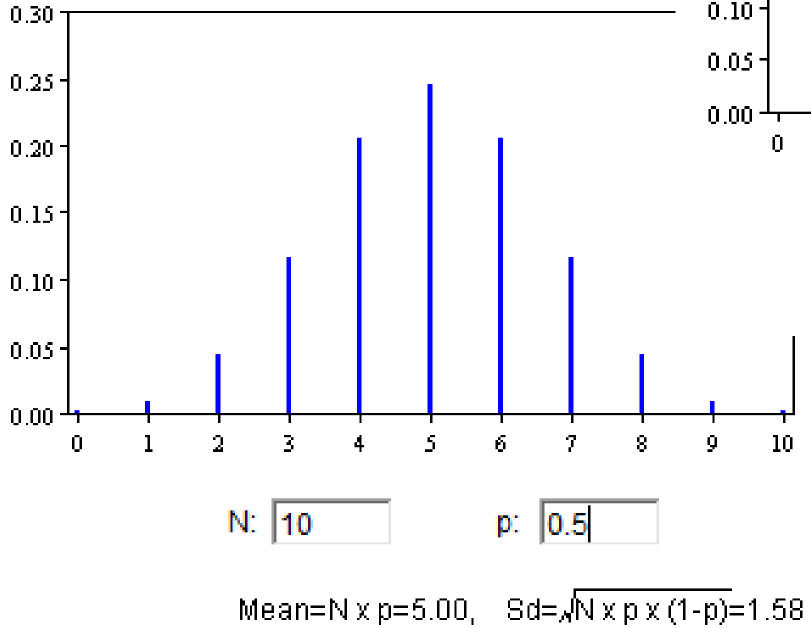
- You run a series of independent trials.
- There can be either a success or a failure for each trial, and the probability of success is the same for each trial.
- There are a finite number of trials, and you are interested in the number of successes or failures.

X~B(n,p)

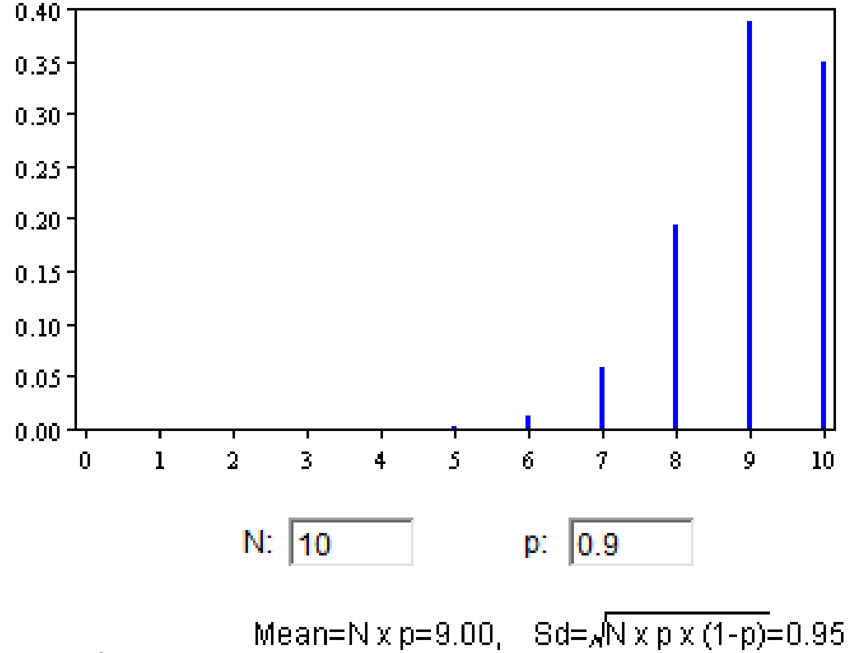
$$P(X = r) = C_r^n p^r q^{n-r}$$

Binomial Distribution is right Skewed if p is low





Binomial Distribution is symmetric if p is 0.5



Binomial Distribution is left Skewed if p is low

Ref: http://onlinestatbook.com/2/probability/binomial_demonstration.html
Last accessed: December 09, 2017 on Safari

Poisson Distribution

French pronunciation: [pwasɔ̃]; in English often rendered /ˈpwɑːsɒn/ - Wikipedia

Binomial: We are interested in number of successes/events (discrete) occurring randomly in fixed *number of trials* (discrete).

Poisson: We are interested in number of successes/events (discrete) occurring randomly in fixed *duration or space* (continuous).

Poisson Distribution

- No. of deaths by horse and mule kicking between 1875-1894 in the Prussian army (http://blog.minitab.com/blog/quality-data-analysis-and-statistics/no-horsing-around-with-the-poisson-distribution-troops)
- No. of birth defects
- No. of defects in a batch of semiconductor wafers
- No. of typing errors per page
- No. of insurance claims (or policies sold) per week
- No. of vehicles passing through a busy traffic junction per minute
- No. of car accidents per hour

Poisson Distribution

- Probability of getting 15 customers requesting for loans in a given day, given on average we see 10 customers
  $\lambda = 10 \; and \; r = 15$

  PMF, $\mathrm{P(X = r)} = \frac{e^{-\lambda}\lambda^{r}}{r!}$.

  Where "r! " is read as r Factorial

  For example "4!" read as 4 Factorial

  4! = 4*3*2*1

  CDF, $\mathrm{P(X \leq r)} = e^{-\lambda} \sum_{i=0}^{r} \frac{\lambda^{i}}{i!}$

Poisson Distribution

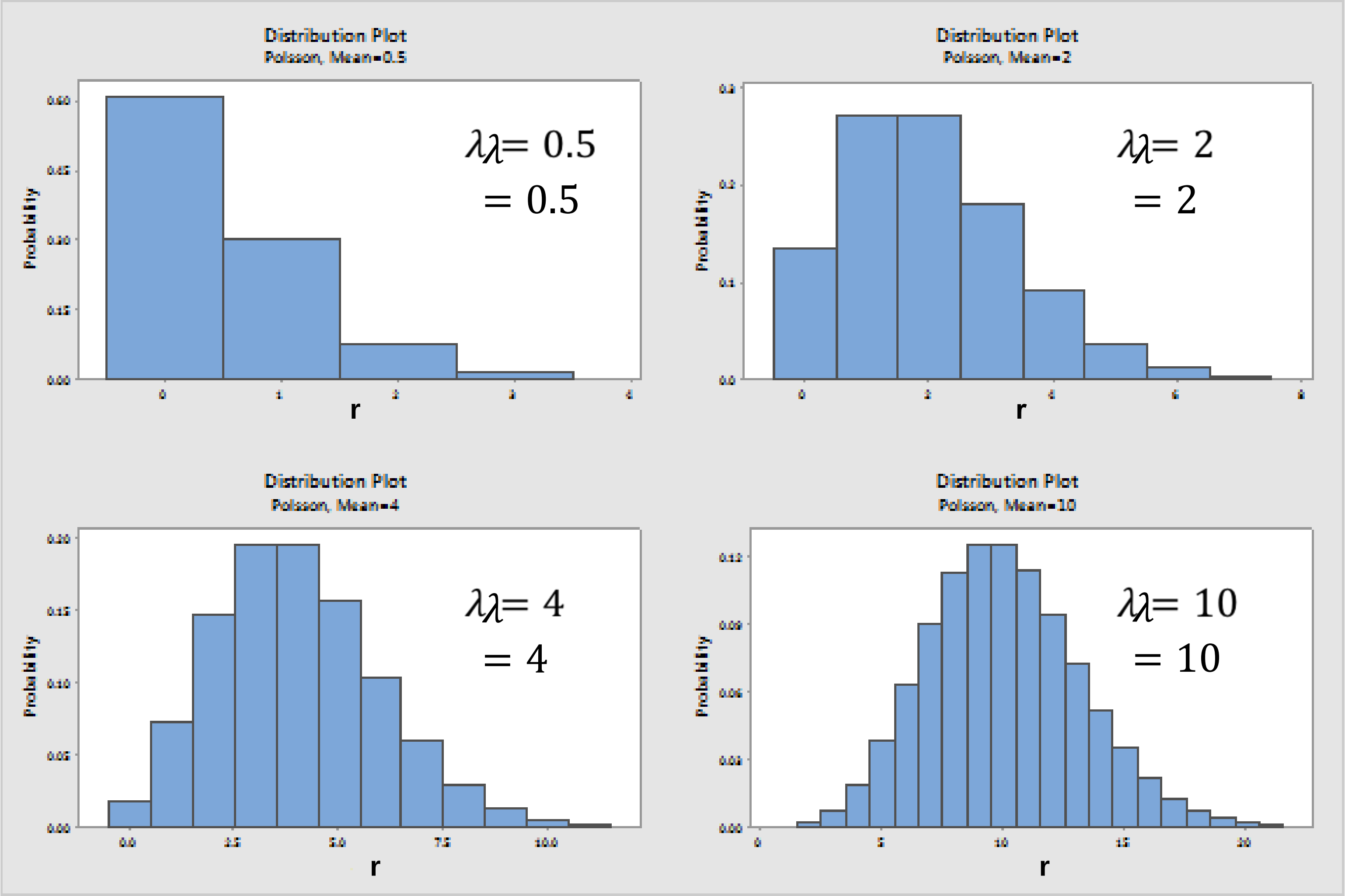
- $E(X) = \lambda$
  $Var(X) = \lambda$

When to use?

- Individual events occur at random and independently in a given interval (time or space).
- You know the mean number of occurrences, λ, in the interval or the rate of occurrences, and it is finite.

X~Po(λ)

$$P(X = r) = \frac{e^{-\lambda}\lambda^{r}}{r!}$$



Ref: http://blog.minitab.com/blog/quality-data-analysis-and-statistics/no-horsing-around-with-the-poisson-distribution-troops
Last accessed: March 02, 2018

Poisson Distribution

- Limiting case of Binomial distribution when $n \to \infty$ (infinite trials) and $p \to 0$ (infinitesimally small probability, i.e., "rare" events).
- As a rule of thumb, if $n > 50$ and $p < 0.1$, Binomial can be approximated by Poisson, i.e., $np \to \lambda$.
- That is, Poisson distribution is used to model occurrences of events that ***could*** happen a very large number of times (large *n*), but ***actually*** happen very rarely (small *p*).

Poisson Distribution

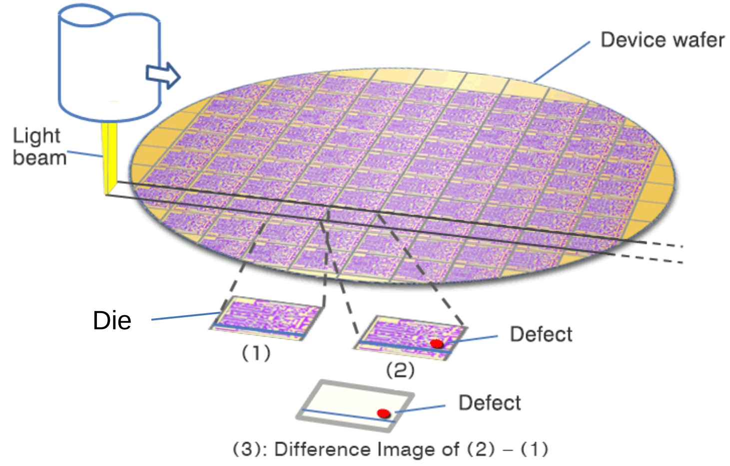
Example

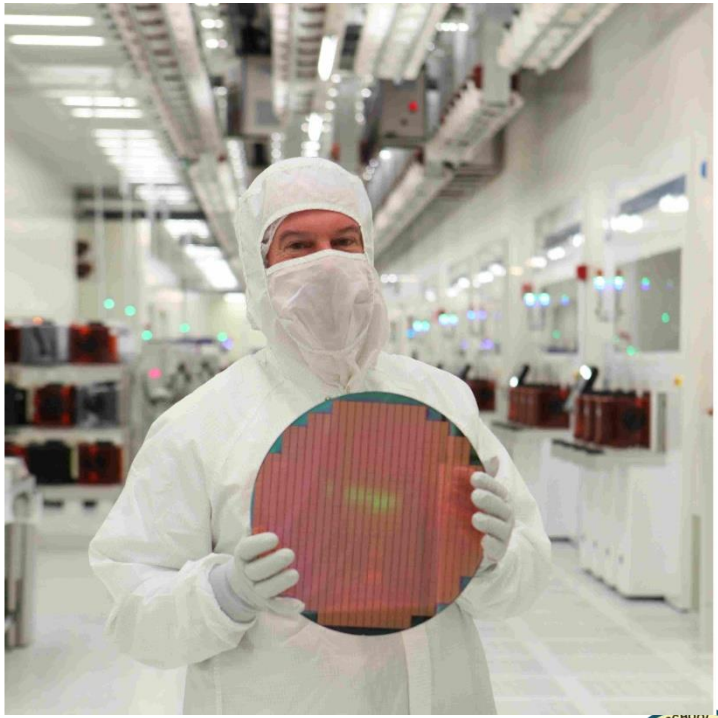
In a tie-breaking T20 Super Over, there are fixed number of opportunities to hit a six, and the probability of hitting a six is very high. So, the number of sixes in a T20 Super Over is **Binomial**.

On the other hand, in a cricket Test Match, a six can be hit almost every few minutes, but a six is probably hit once in a few hours. So, the number of sixes in a Test Match is **Poisson**.

A company makes semiconductor wafers. The probability of a defective die on the wafer is 0.001. What is the probability that a random sample of 500 dies will contain exactly 5 defective dies?

What distribution is this?





Poisson Distribution

Approach 1: Binomial

- $n = 500,\ p = 0.001,\ r = 5$

$$P(X = r) = C_r^n p^r q^{n-r}$$

$^{500}C_5 * (0.001)^5 * (1-0.001)^{495} = 0.00156$

Approach 2: Poisson - $n > 50$ and $p < 0.1$

$$n=500 \text{ and } p = 0.001$$

$\lambda = np = 0.5, r = 5$ **[using np since we are equating Binomial and Poisson]**

$$\mathrm{P(X = r)} = \frac{e^{-\lambda}\lambda^r}{r!}$$

$$\frac{2.718^{-0.5}0.5^5}{5!} = 0.00158$$

Note: e = 2.718

Poisson Distribution

The probability that no customer will visit the store in one day

$$P(X=0) = \frac{e^{-\lambda}\lambda^{0}}{0!} = e^{-\lambda}$$

Note that

- . = 1 (anything to the power of 0 is 1)
- 0! = 1 (0 factorial equals 1)

Probability that no customer will visit in *n* days

$$e^{-n\lambda}$$

Exponential Distribution

Probability that a customer will visit in *n* days:

$$1 - e^{-n\lambda}$$

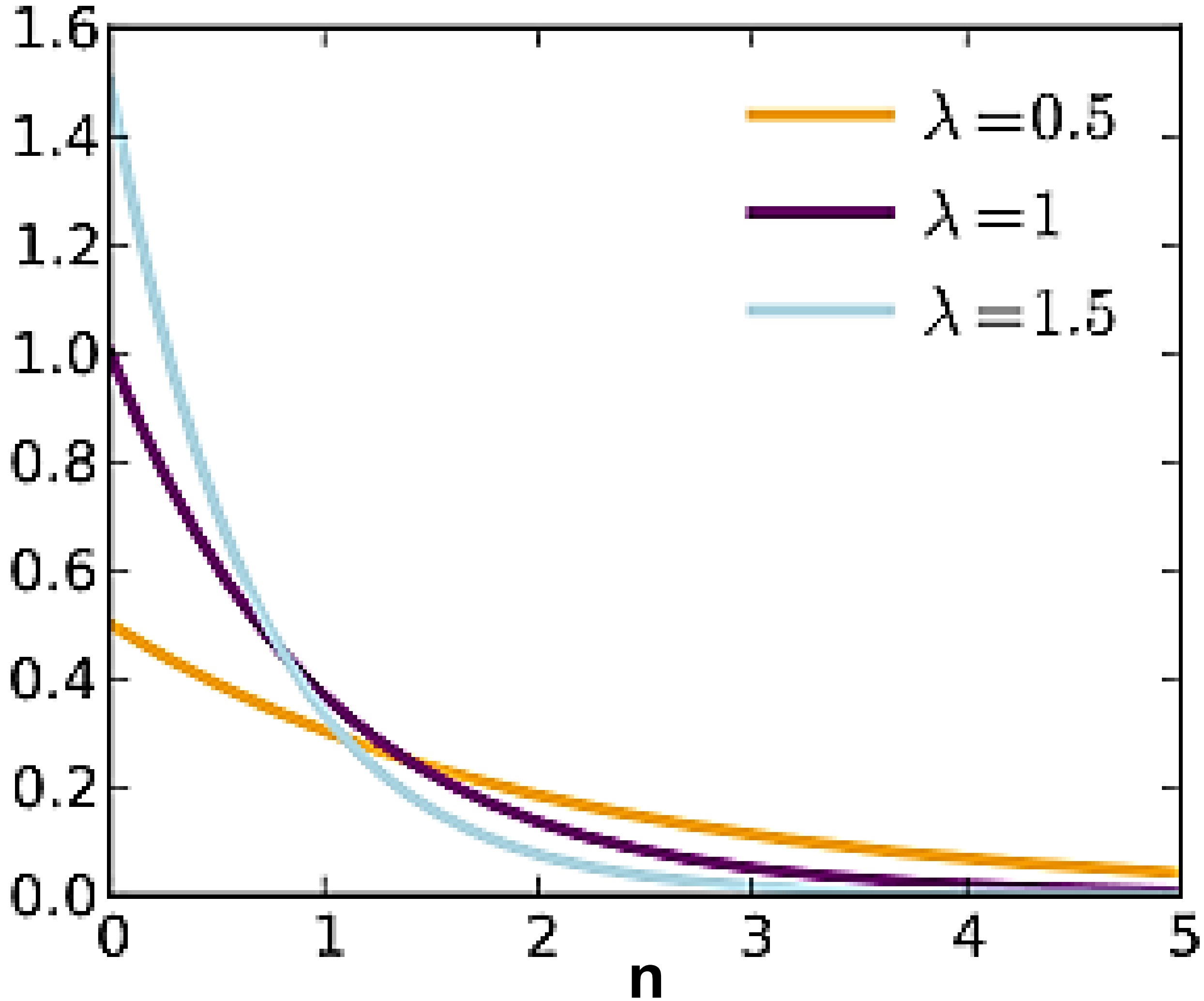
$1 - e^{-n\lambda}$

$$CDF = 1 - e^{-n\lambda}, n \geq 0$$

$$PDF = \lambda e^{-n\lambda}, n \geq 0$$

X~Exp(λ)

$$PDF = \lambda e^{-n\lambda}, n \geq 0$$



Ref: http://en.wikipedia.org/wiki/Exponential_distribution
Last accessed: June 12, 2015

Exponential Distribution

- Poisson process
- Continuous analog of Geometric distribution

$$E(X) = \frac{1}{\lambda} \qquad Var(X) = \frac{1}{\lambda^2}$$

Probability Distributions (Discrete)

Geometric:	For estimating number of attempts before first success
Binomial:	For estimating number of successes in *n* attempts
Poisson:	For estimating *n* number of events in a given time period when on average we see *m* events

Probability Distributions (Continuous)

Exponential: Time between events

Normal :

Z :

T :

(Chi-squared) :

F :

Probability Distributions – Discrete

Distribution	Geometric	Binomial	Poisson
Type	Discrete	Discrete	Discrete
Representation	X~Geo(p)	X~B(n,p)	X~Po(λ)
Explanation	For estimating number of attempts before first success	For estimating number of successes in "n" attempts	For estimating "n" number of events in a given time period when on average we see "m" events
Expected Value			
Variance			
Probability Mass Function (PMF) P(X=x)			
Cumulative Distribution Function (CDF). P(X<=x)			
P(X>x) = 1- P(X<=x)			

Probability Distributions – Continuous

BREAK

Distribution	Exponential
Type	Continuous
Representation	X~Exp(λ)
Explanation	Time between events
Expected Value	
Variance	
Probability Density Function (PDF) P(X=x)	
Cumulative Density Function (CDF) P(X<=x)	
P(X>x) = 1- P(X<=x)	

Distribution	Exponential
Type	Continuous
Representation	X~Exp(λ)
Explanation	Time between events
Expected Value	$E(X) = \frac{1}{\lambda}$
Variance	$Var(X) = \frac{1}{\lambda^2}$
Probability Density Function (PDF) P(X=x)	P(X=x) = $\lambda e^{-n\lambda}, n \geq 0$
Cumulative Density Function (CDF) P(X<=x)	$1 - e^{-n\lambda}, n \geq 0$
P(X>x) = 1- P(X<=x)	$e^{-n\lambda}, n \geq 0$

Distribution	Exponential
Type	Continuous
Representation	X~Exp(λ)
Explanation	Time between events
Expected Value	$E(X) = \frac{1}{\lambda}$
Variance	$Var(X) = \frac{1}{\lambda^2}$
Probability Density Function (PDF) P(X=x)	P(X=x) = $\lambda e^{-n\lambda}, n \geq 0$
Cumulative Density Function (CDF) P(X<=x)	$1 - e^{-n\lambda}, n \geq 0$
P(X>x) = 1- P(X<=x)	$e^{-n\lambda}, n \geq 0$

Distribution	Exponential
Type	Continuous
Representation	X~Exp(λ)
Explanation	Time between events
Expected Value	$E(X) = \frac{1}{\lambda}$
Variance	$Var(X) = \frac{1}{\lambda^2}$
Probability Density Function (PDF) P(X=x)	P(X=x) = $\lambda e^{-n\lambda}, n \geq 0$
Cumulative Density Function (CDF) P(X<=x)	$1 - e^{-n\lambda}, n \geq 0$
P(X>x) = 1- P(X<=x)	$e^{-n\lambda}, n \geq 0$

Distribution	Exponential
Type	Continuous
Representation	X~Exp(λ)
Explanation	Time between events
Expected Value	$E(X) = \frac{1}{\lambda}$
Variance	$Var(X) = \frac{1}{\lambda^2}$
Probability Density Function (PDF) P(X=x)	P(X=x) = $\lambda e^{-n\lambda}, n \geq 0$
Cumulative Density Function (CDF) P(X<=x)	$1 - e^{-n\lambda}, n \geq 0$
P(X>x) = 1- P(X<=x)	$e^{-n\lambda}, n \geq 0$

Distribution	Exponential
Type	Continuous
Representation	X~Exp(λ)
Explanation	Time between events
Expected Value	$E(X) = \frac{1}{\lambda}$
Variance	$Var(X) = \frac{1}{\lambda^2}$
Probability Density Function (PDF) P(X=x)	P(X=x) = $\lambda e^{-n\lambda}, n \geq 0$
Cumulative Density Function (CDF) P(X<=x)	$1 - e^{-n\lambda}, n \geq 0$
P(X>x) = 1- P(X<=x)	$e^{-n\lambda}, n \geq 0$

Probability Distributions - Scenarios

Identify the distribution and calculate expectation, variance and the required probabilities.

Q1. A man is bowling. The probability of him knocking all the pins over is 0.3. If he has 10 shots, what is the probability he will knock all the pins over less than 3 times?

Binomial Distribution

Because $n = 10$ shots (fixed)
$p=0.3$,

$P(X<3) = ?$

Probability Distributions - Scenarios

$X \sim B(10,0.3)$; n=10, p=0.3, q=1-0.3=0.7, r=0, 1, 2 (< 3)

$E(X) = np = 10*0.3 = 3$

$Var(X) = npq = 2.1$

$P(X=0) = 0.028$; $P(X=1) = 0.121$; $P(X=2) = 0.233$

$P(X<3) = P(X=0) + P(X=1) + P(X=2)$

$P(X<3) = 0.028 + 0.121 + 0.233 = 0.382$

Probability Distributions - Scenarios

Identify the distribution and calculate expectation, variance and the required probabilities.

Q2. On average, 1 bus stops at a certain point every 15 minutes. What is the probability that no buses will turn up in a single 15 minute interval?

Poisson Distribution

$\lambda=1$, $r=0$

Probability Distributions - Scenarios

X ~ Po(1); λ=1, r=0

E(X) = λ = 1

Var(X) = λ = 1

$$P(X = r) = \frac{e^{-\lambda}\lambda^{r}}{r!}$$

P(X= 0) = $\frac{e^{-1}1^{0}}{0!}$

P(X=0) = 0.368

Note that

- 1^{0}. = 1 (anything to the power of 0 is 1)
- 0! = 1 (0 factorial = 1)

Probability Distributions - Scenarios

Identify the distribution and calculate expectation, variance and the required probabilities.

Q3. 20% of cereal packets contain a free toy. What is the probability you will need to open fewer than 4 cereal packets before finding your first toy?

Geometric Distribution

$p = 20\% = 0.2$

Probability Distributions - Scenarios

X ~ Geo(0.2); p=0.2, q=1-0.2=0.8, r<4 or $\leq$3

$E(X) = \frac{1}{p} = 5$

$Var(X) = \frac{q}{p^2} = 20$

$$P(X \leq r) = 1 - q^r$$

$P(X \leq 3) = 0.488$

Probability Distributions - Scenarios

- Products produced by a machine has a 3% defective rate.
  a) What is the probability that the first defective occurs in the fifth item inspected
  b) What is the probability that the first defective occurs in the first five inspections?

Geometric Distribution

$p = 3\% = 0.03$, $q = 0.97$

a) $P(X=r) = q^{r-1}.p$

P(X=5) $= 0.97^5 . 0.03 =$ **0.265**

b) $P(X<= 5) = 1-q^r$

P(X<=5) $= 1- 0.97^5 =$ **0.1412**

0.000608677

Probability Distributions - Scenarios

- Suppose 14 students each have a .6 probability of passing statistics. What's the probability that 3 or more will pass?

Binomial Distribution

p = 0.6, q=0.4, n=14, r=3

P(X>=3) = 1- P(X<3)
P(X>=3) = 1- [P(X=0) + P(X=1) + P(X=2)]

$$P(X = r) = {}^{n}C_{r} p^{r} q^{n-r}$$

P(X>=3) = 1 – [0.0006]
P(X>=3) = 0.9994

Poisson Distribution Formula Differences?

- $P(X = r) = \frac{e^{-\lambda}\lambda^r}{r!}$ or $\frac{e^{-\lambda t}(\lambda t)^r}{r!}$?

Suppose births in a hospital occur randomly at an average rate of 1.8 births per hour. What is the probability of 5 births in a given 2 hour interval?

$$P(X = 5) = \frac{e^{-3.6}3.6^5}{5!} \; or \; \frac{e^{-1.8*2}(1.8*2)^5}{5!} \; ?$$

If you use 1.8, use t=2 in the second formula. Alternatively, you could say that since the average is 1.8 per hour, it is 3.6 per 2 hours (the interval of interest).

Poisson Distribution Formula Differences?

- $P(X = r) = \dfrac{e^{-\lambda}\lambda^r}{r!}$ or $\dfrac{e^{-\lambda t}(\lambda t)^r}{r!}$

Now suppose head injury patients (due to not wearing helmets) arrive in Hospital A randomly at an average rate of 0.25 patients per hour, and in Hospital B randomly at an average rate of 0.75 per hour. What is the probability of more than 3 such patients arriving in a given 2 hour interval in both hospitals together?

What is the probability distribution?

$X \sim Po(\lambda_1)$ and $Y \sim Po(\lambda_2)$

$X + Y \sim Po(\lambda_1 + \lambda_2)$

What are if we use first formula?

What are λ_1 and λ_2 if we use first formula?

$\lambda_1 = 0.5$ and $\lambda_2 = 1.5$. - This is because of the 2 hour interval

$\lambda_1 + \lambda_2 = 2$

$$P(X + Y > 3) = P(X + Y = 4) + P(X + Y = 5) + P(X + Y = 6) + \cdots$$

$$P(X + Y > 3) = 1 - [P(X + Y \leq 3)$$

$$= 1 - (P(X + Y = 0) + P(X + Y = 1) + P(X + Y = 2) + P(X + Y = 3)]$$

Continued on next slide

Poisson Distribution Formula Differences?

$$P(X=r) = \frac{e^{-\lambda}\lambda^{r}}{r!} \ or \ \frac{e^{-\lambda t}(\lambda t)^{r}}{r!}$$

We use . Since t = 2 hrs. Also = 2

We use $P(X=r) = \frac{e^{-\lambda t}(\lambda t)^{r}}{r!}$. Since t = 2 hrs. Also λ = 2

$$P(X+Y>3) = P(X+Y=4) + P(X+Y=5) + P(X+Y=6) + \cdots =$$

$P(X$+Y >3) $= 1 - P(X+Y \leq 3)$

$P(X$+Y >3) $=$

P(X+Y >3) = $1 - [P(X+Y=0) + P(X+Y=1) + P(X+Y=2) + P(X+Y=3)]$

P(X+Y >3)

= 1- ($\frac{e^{-2}(2*2)^{0}}{0!} + \frac{e^{-2}(2*2)^{1}}{1!} + \frac{e^{-2}(2*2)^{2}}{2!} + \frac{e^{-2}(2*2)^{3}}{3!}$) =. **0.5665**

= 1- (+ + +) =. **0.5665**

Poisson or Exponential?

Given a Poisson process:

- The *number* of events in a given time period
- The *time* until the first event
- The *time* from now until the next occurrence of the event
- The *time interval* between two successive events

Poisson

Exponential

Poisson or Exponential?

A tech support center of a company receives 5 calls per hour on an average.

a) What is the probability that the center will receive 8 calls in the next hour?
b) What is the probability that more than 30 minutes will elapse between calls?
c) What is the probability that more than 30 minutes and less than 45 minutes will elapse between calls?

a) Poisson Distribution

$$P(X = 8) = \frac{e^{-5}5^{8}}{8!} = 0.065$$

b) Exponential Distribution

$$P(Time\ between\ calls > 30) = \int_{0.5}^{\infty} \lambda e^{-\lambda T} dT = -e^{-\lambda T}\Big]_{0.5}^{\infty} = e^{-5*0.5}$$

$$= 0.082$$

Poisson or Exponential?

A tech support center of a company receives 5 calls per hour on an average.

a) What is the probability that the center will receive 8 calls in the next hour?
b) What is the probability that more than 30 minutes will elapse between calls?
c) What is the probability that more than 30 minutes and less than 45 minutes will elapse between calls?

c) Exponential Distribution

$$P(Time\ between\ calls > 30\ and < 45) = \int_{0.5}^{0.75} \lambda e^{-\lambda T}\, dT = -e^{-\lambda T}\Big]_{0.5}^{0.75}$$

$$= -e^{-5*0.75} + e^{-5*0.5}$$

$$=0.058$$

Probability Distributions

Babyboom Data - Excel

Forty-four babies -- a new record -- were born in one 24-hour period at the Mater Mothers' Hospital in Brisbane, Queensland, Australia, on December 18, 1997. For each of the 44 babies, *The Sunday Mail* recorded the time of birth, the sex of the child, and the birth weight in grams.

Probability Distributions

Determine the distributions for the following scenarios for this dataset:

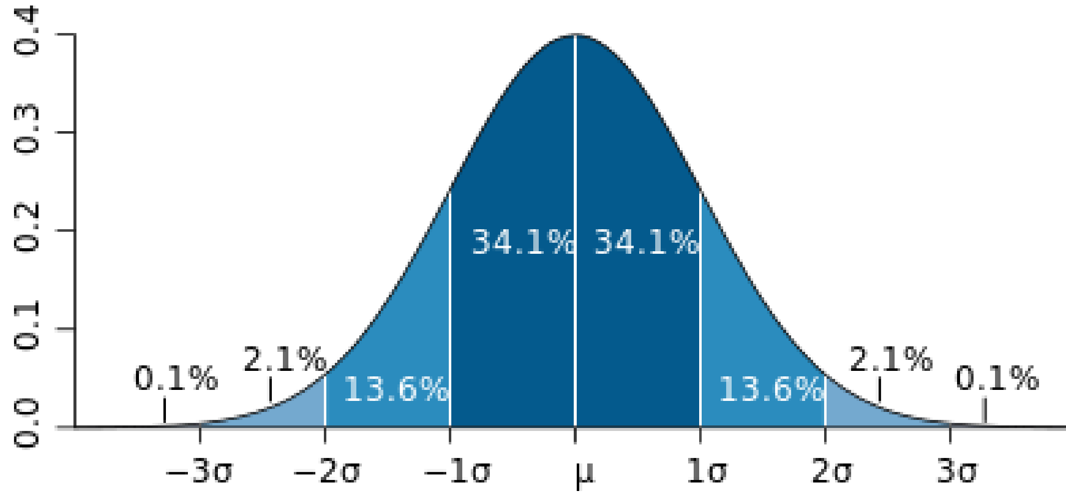
1. Probability of observing at least 26 boys in 44 births assuming equal probability of a boy or a girl being born.
2. Probability that 3 births occur before the birth of a girl.
3. Probability of 4 births per hour given 44/24 = 1.83 births per hour on average.
4. Probability that more than 60 minutes will elapse between births.

1. Binomial; 2. Geometric; 3. Poisson; 4. Exponential

NORMAL DISTRIBUTION

Normal (Gaussian) Distribution

- Mean = Median = Mode
- 68-95-99.7 empirical rule
- Zero Skew and Kurtosis
- X ~ N(μ, σ^2)

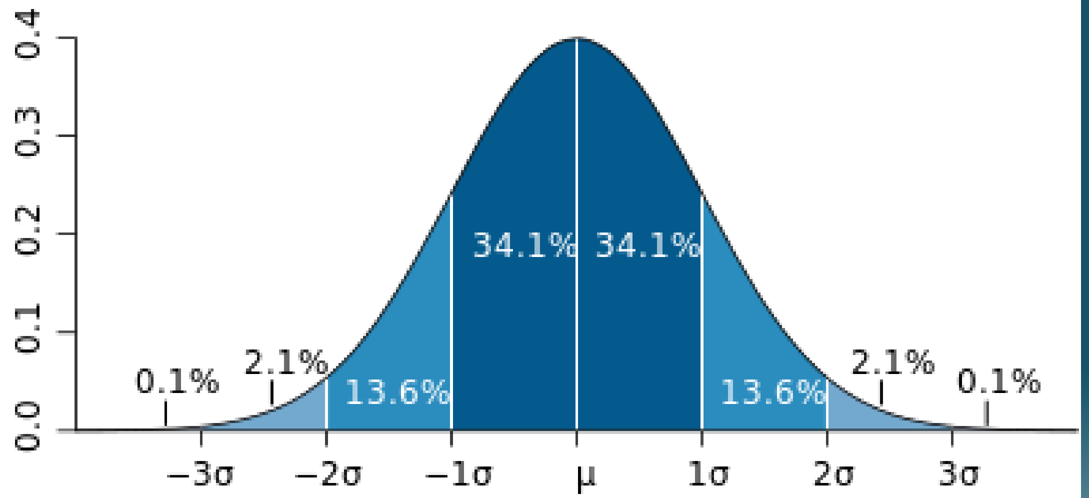


- Shaded area gives the probability that X is between the corresponding values

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Normal Distribution - 68-95-99.7 empirical rule

- If the data is normally distributed then
  - 68% of the data is within the ± one standard deviation (±1σ) from the mean
  - 95% of the data is within the ± two standard deviations (±2σ) from the mean
  - 99.7% of the data is within the ± three standard deviations (±3σ) from the mean
- Shaded area gives the probability that X is between the corresponding values

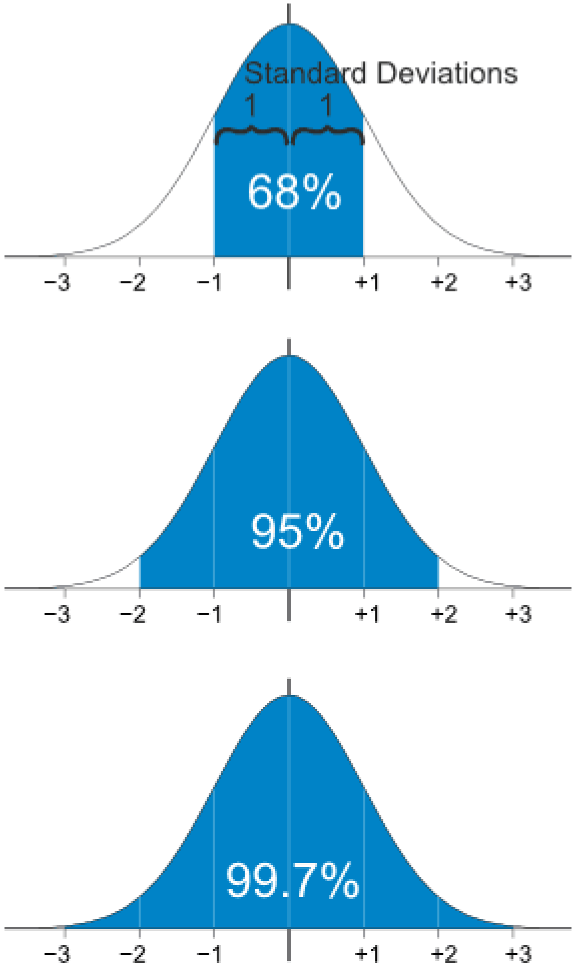


PDF

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Measures of Spread (Dispersion)

You know the 68-95-99.7 rule.



A company produces a valve that is specified to weigh 1500g, but there are imperfections in the process. While the mean weight is 1500g, the standard deviation is 300g.

Mean (μ) = 1500g and Standard deviation (σ) $= 300$

Q1. What is the range of weights within which 95% of the valves will fall?

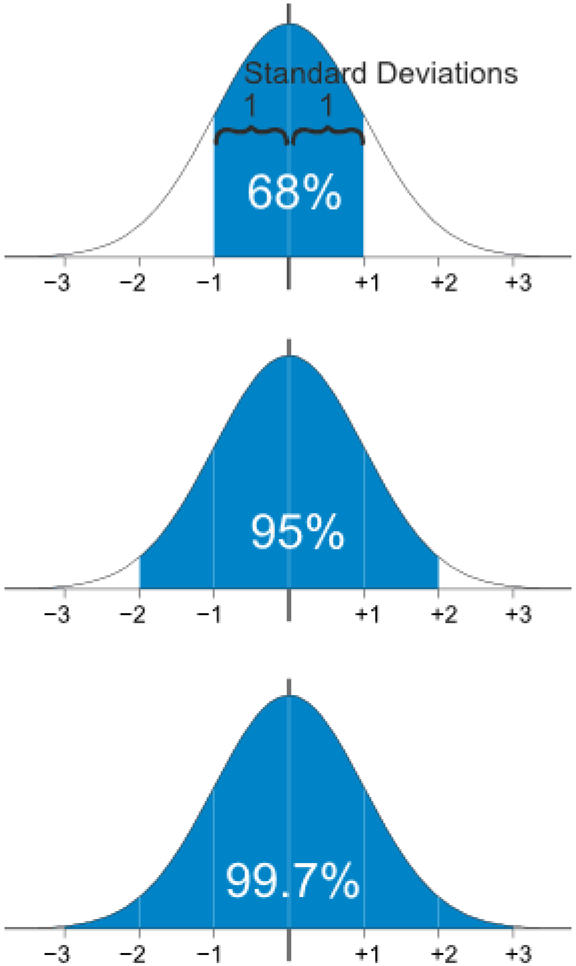
Ans) 95% of the data will be between $\pm 2\sigma$ from the mean . So 1500 $\pm$ (2*300) = Between 900g and 2100g

Q2. Approximately 16% of the weights will be more than what value?

Ans) 32% of the data will be outside $\pm 1\sigma$ from the mean, because 68% of the data is between $\pm 1\sigma$. Since it is symmetrical we have 16% of the weights on each side outside of $\pm 1\sigma$. So 1500 + (1*300) = 1800 g. So 16% of the weights will be greater than 1800g

Measures of Spread (Dispersion)

You know the 68-95-99.7 rule.



A company produces a valve that is specified to weigh 1500g, but there are imperfections in the process. While the mean weight is 1500g, the standard deviation is 300g.

Mean (μ) = 1500g and Standard deviation (σ) $= 300$

Q3. Approximately 0.15% of the weights will be less than what value?

Ans) 0.3% (100 – 99.7) of the data will be outside $\pm 3\sigma$ from the mean, because 99.7% of the data is between $\pm 3\sigma$. Since it is symmetrical we have 0.15% of the weights on each side outside of $\pm 3\sigma$. So 1500 - (3*300) = 600 g. So 0.15% of the weights will be less than 600g

Sample Software Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.717055011
R Square	0.514167888
Adjusted R Square	0.494734604
Standard Error	4.21319131
Observations	27

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	469.6573265	469.6573265	26.4581054	2.57053E-05
Residual	25	443.7745253	17.75098101		
Total	26	913.4318519			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99.0%*	*Upper 99.0%*
Intercept	-4.154014573	2.447784673	-1.697050651	0.102104456	-9.195321476	0.88729233	-10.97705723	2.669028089
Big Mac Price ($)	3.547427488	0.689658599	5.143744297	2.57053E-05	2.127049014	4.967805962	1.625048409	5.469806567

Sample Software Output

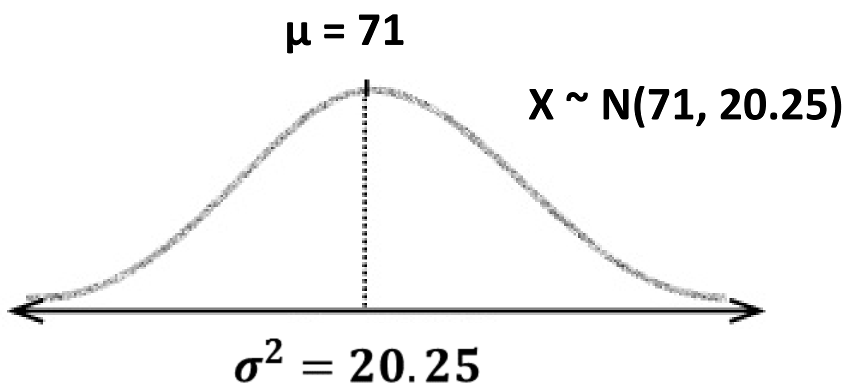
```
Call:
glm(formula = Response ~ Age, family = "binomial", data = flierresponse)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.95015  -0.32016  -0.05335   0.26538   1.72940

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -20.40782    4.52332  -4.512 6.43e-06 ***
Age           0.42592    0.09482   4.492 7.05e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 123.156  on 91  degrees of freedom
Residual deviance:  49.937  on 90  degrees of freedom
AIC: 53.937

Number of Fisher Scoring iterations: 7
```

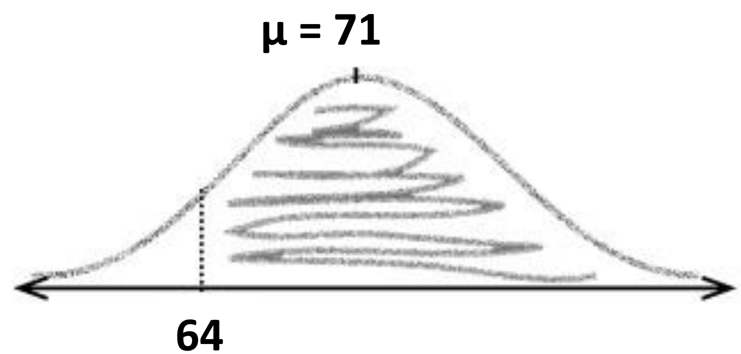
Calculating Normal Probabilities

Step 1: Determine the distribution

Julie wants to marry a person taller than her and is going on blind dates. The mean height of the 'available' guys is 71" and the variance is 20.25 inch^2 (yuck!).

Oh! By the way, Julie is 64" tall.





Variance = 20.25
Std Deviation = sqrt(20.25) = **4.5 inches**

Calculating Normal Probabilities

Step 2: Standardize to Z ~ N(0,1)

1. Move the mean

   This gives a new distribution

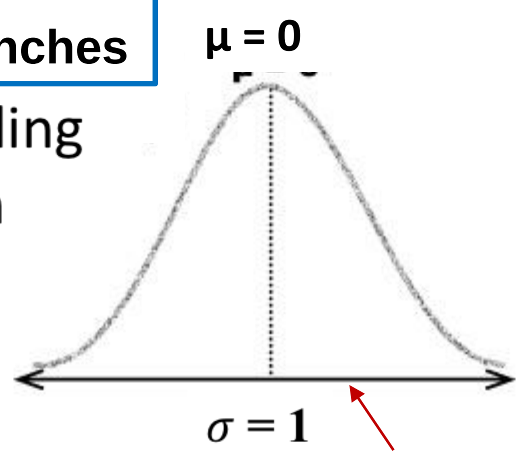
   X-71 ~ N(0,20.25)

Random variable is **x**, the ***actual*** heights of available guys

Variance = 20.25 inches^2
Std Deviation = sqrt(20.25) = **4.5 inches**

2. Squash the width by dividing by the standard deviation

This gives us $\frac{X-71}{4.5} \sim N(0,1)$

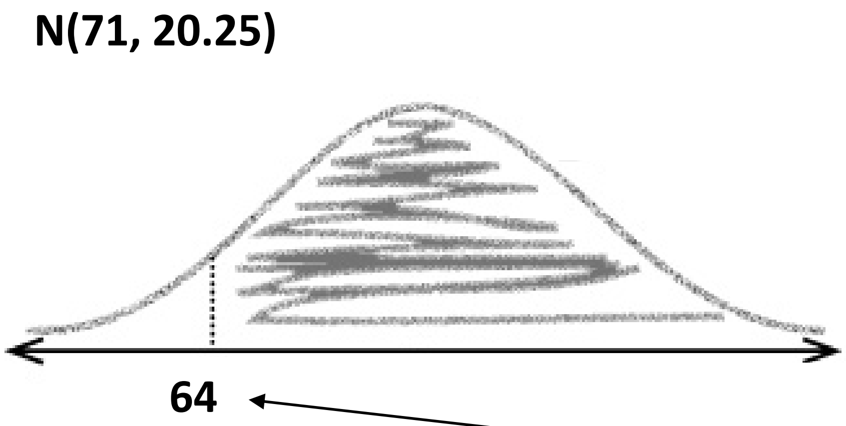


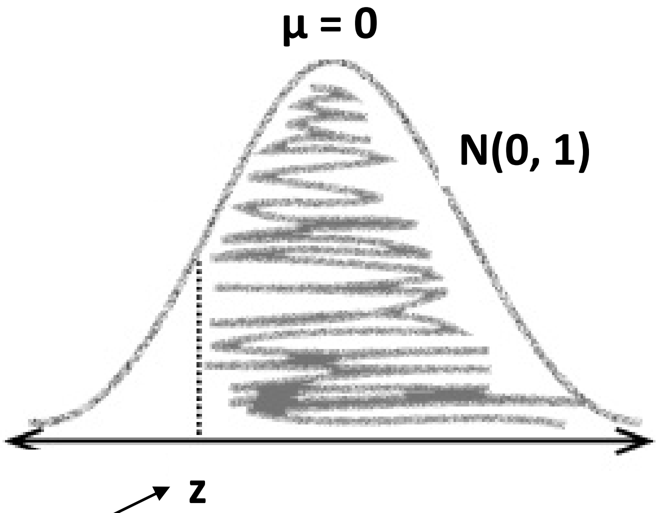
$Z = \frac{X-\mu}{\sigma}$ is called the Standard Score or the z-score.

Random variable is **z**, the the ***standardized*** heights of available guys

Calculating Normal Probabilities

Step 2: Standardize to Z ~ N(0,1)





$$z = \frac{64 - 71}{4.5} = -1.56$$

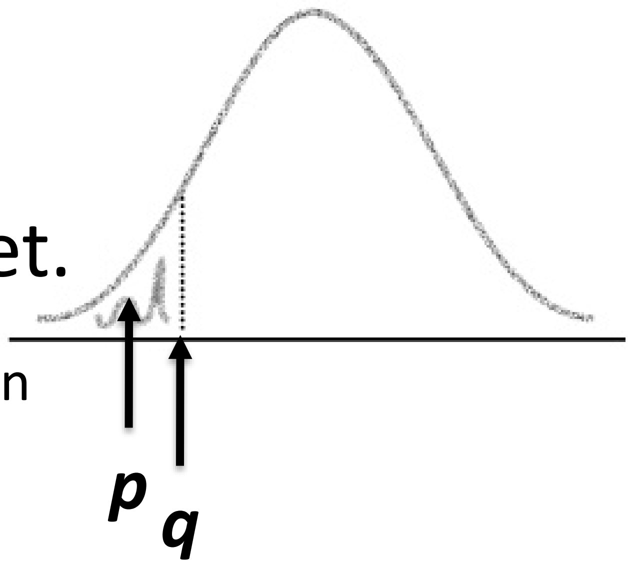
Julie is 64” tall, i.e., she is 1.56 standard deviations shorter than the average height of the available guys.

Calculating Normal Probabilities

Step 3: Look up the probability in the tables

Note the tables give P(Z<z).

In R functions, the distribution is abbreviated and prefixed with an alphabet.



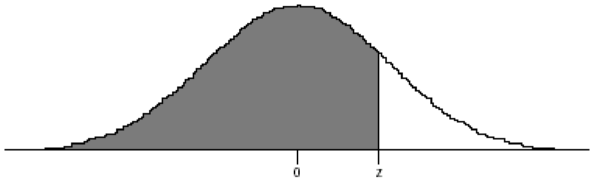
pnorm: Probability (Cumulative Distribution Function, CDF) in a *Normal Distribution*

qnorm: Quantile (Inverse CDF) in a *Normal Distribution* – The value corresponding to the desired probability.

Calculating Normal Probabilities

Step 3: Look up the probability in the tables

Note the tables give P(Z<z).



$z = \frac{64-71}{4.5} = -1.56$ in the case of our problem.

P(Z>-1.56) = 1- P(Z<-1.56)
= 1-0.0594 = 0.9406

Normal Deviate z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-4.0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.7	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379

Calculating Normal Probabilities

Step 3: Get the probability from R

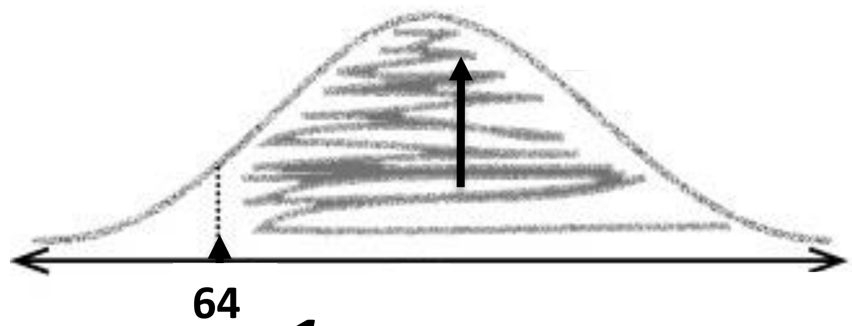
1-pnorm(64, mean=71, sd=sqrt(20.25))
or
1-pnorm(64, 71, 4.5)

Answer: 1-0.0599 = 94.01%

qnorm(0.0599, 71, 4.5)

Answer: 64

N(71, 20.25)



1-pnorm

qnorm

Attention Check

Q. What is the standard score or Z Score for N(10,4), value 6?

$$Z = \frac{X - \mu}{\sigma}$$

$$\sigma = \text{sqrt}(4) = 2$$

$$z = \frac{6-10}{2} = -2$$

Remember X ~ N(μ, σ^2)

Variance is specified we need to get Std Deviation

Q. The standard score of value 20 is 2. If the variance is 16, what is the mean

$$\boldsymbol{\sigma = \text{sqrt}(16) = 4}$$

$$Z = \frac{X - \mu}{\sigma}$$

$$2 = \frac{20 - \mu}{4} \quad \therefore \quad \mu = 20 - 8 = 12$$

Attention Check

Q. Julie just realized that she wants her date to be taller when she is wearing her heels, which are 5” high. Find the new probability that her date will be taller.

New height for Julie = 64+5 = 69 in

$$\therefore z = \frac{69-71}{4.5} = -0.44;$$

P(Z<-0.44) = 0.33,
P(Z<-0.44) = 0.33,
P(Z>-0.44) = 0.67 or 67%
∴ P(Z>-0.44) = 0.67 or 67%

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

1-pnorm(69, 71, 4.5). This gives P(X>69) = 67%

Attention Check

Q. Julie wants to have at least 80% probability of finding the right guy. What is the maximum size of heels she can wear?

A. qnorm(0.20, 71, 4.5). This gives a value of 67.2”. As Julie is 64” tall, the maximum heel size she should wear is about 3”.

Attention Check

Q. Julie is convinced of the dangers of high heels and decides to stick with only 1" heels. What is the probability of finding the right guy now?

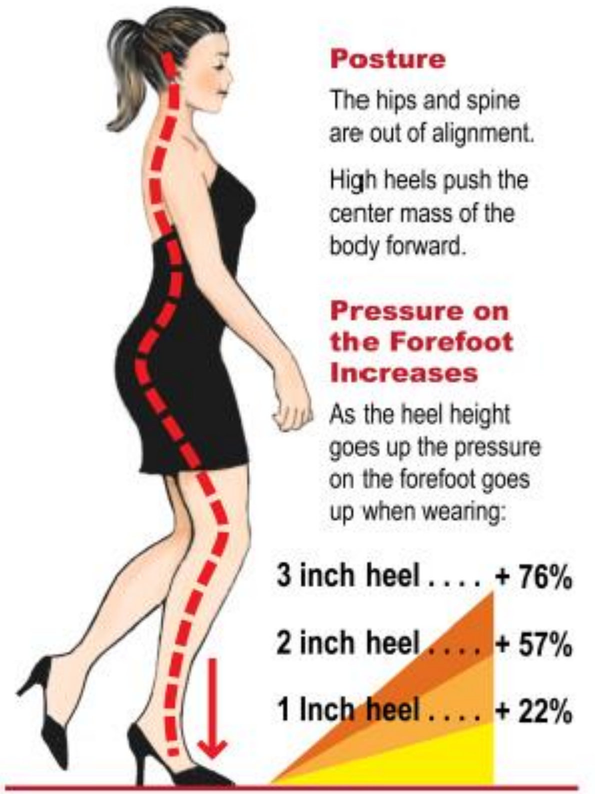


Figure 2

A. 1-pnorm(65, 71, 4.5). This gives a P(X>65)=90.9%.

PRIYANKA PRAVEEN
DECCAN CHRONICLE

FLAT REFUSAL

It's not just high heels that can be a pain, flat footwear is equally damaging

Almost everyone's favourite pair of 'killer' high heels have been notorious for bad posture and foot aches amongst other issue. Now reports say that its simple cousin — the flats — aren't really goody two shoes either.

Even celebrities like Victoria Beckham, who swear by their stilettos, have on quite a few occasions traded them for a pair of flats, but doctors feel that this really might not be the best thing for our feet. From agonising pain, spinal damage and even disorders — flats, are responsible for a host of problems.

"Our foot consists of the toes, the arch and the heel, this mechanism works so well that when we walk our entire weight is distributed equally," explains Dr Mithin Aachi, Senior Orthopedician. "The arch is what helps with the equal distribution of weight and so when we wear flat footwear unequal distribution of weight takes place and undue stress is put on the heel. This leads to several problems including plantar fasciitis and an inflammation of the thick band of tissues that connects the heel and the toes," he adds. In such cases, the pain is, several times, unbearable. Dr Praveen Rao, Orthopedic Surgeon, says, "When this happens, people find it difficult to walk after sitting for a long time."

Flats can cause spinal problems and inflammation of the thick band of tissues that connects the heel and the toes

Apart from pain, the lack of a cushioning and an arch in these footwear can eventually lead to spine troubles. "Since the pressure is on the heel, the gait of the person changes over the years and that leads to spinal problems and causes severe pain," explains Dr Rao.

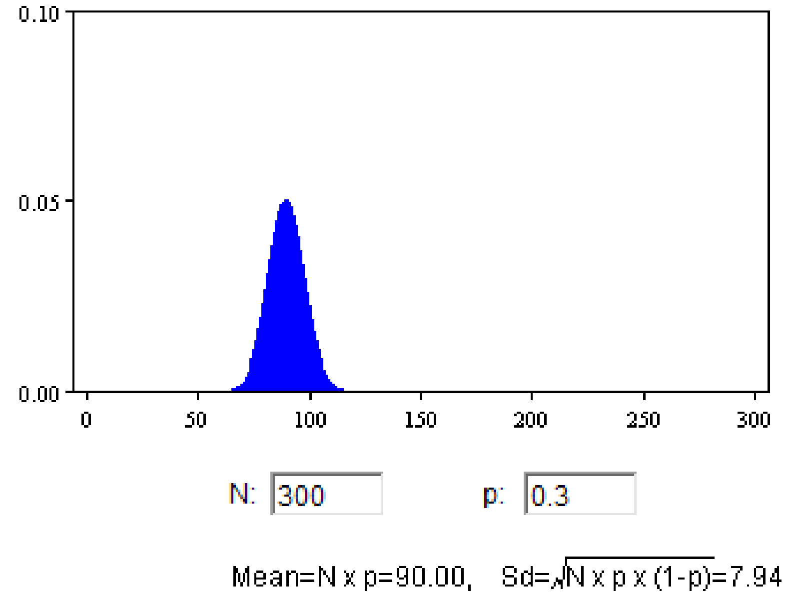
Doctors believe that we need to find a middle ground. "It's okay to wear high heels once in a while and since flats are more convenient, you can wear them occasionally, but you will need to find a balance. It helps to take a 'foot holiday' once a week by giving flats and heels a break and opting for an arched and cushioned footwear," explains Dr Aachi.

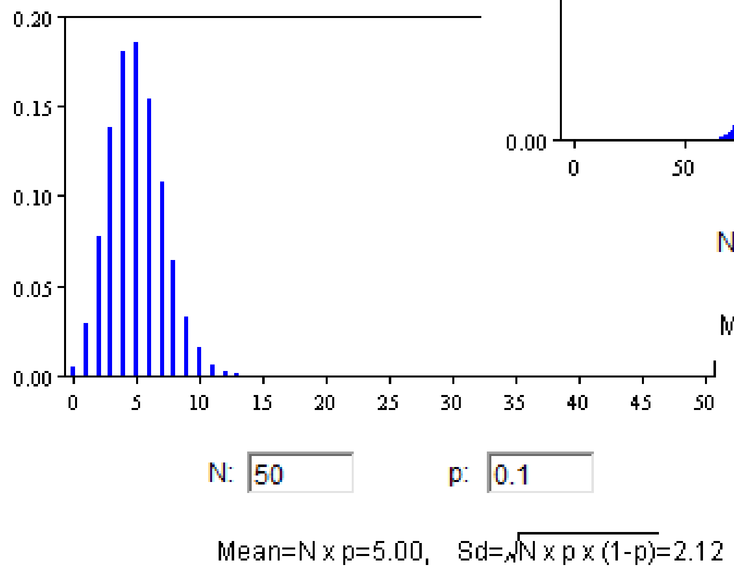
So, is there an ideal heel height that one needs to follow? "There isn't a number as such, but heels above one inch should be avoided regularly. Also wearing cushioned footwear with a small block-heel sometimes is fine," adds Dr. Rao.

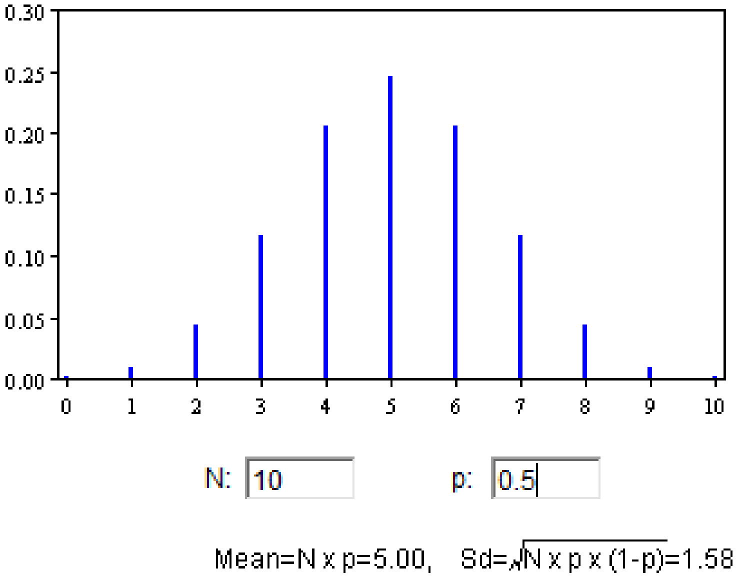
ALL TOO FLAT: Wearing flats regularly can be bad for your feet

Normal Distribution

Binomial distribution can be approximated to a Normal distribution if np>5 and nq>5.

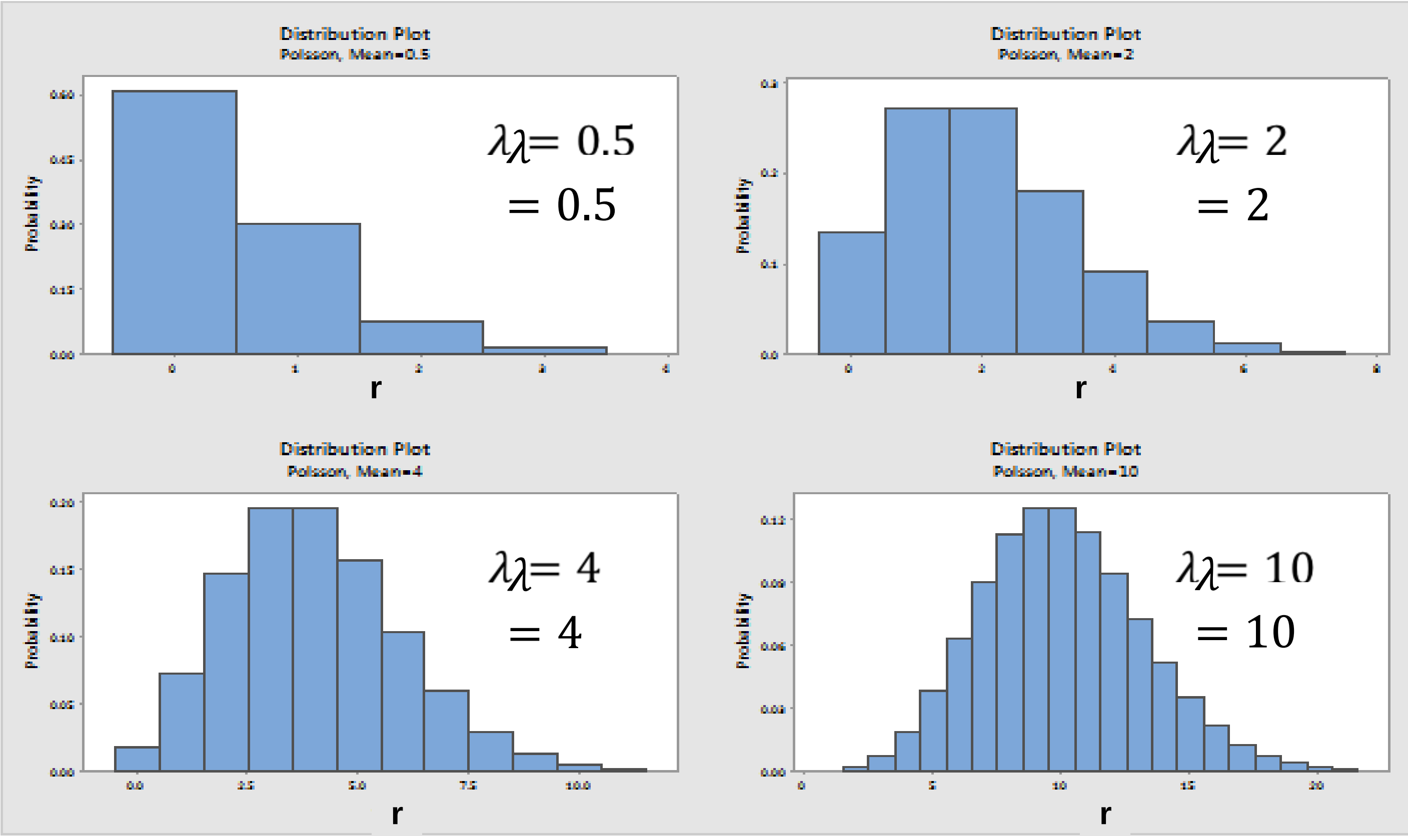






Normal Distribution

Poisson distribution can be approximated to a Normal distribution when $\lambda > 15$.



Normal Distribution

You have designed a new game, Angry Buds. The key to success is that it should not be so difficult that people get frustrated, nor should it be so easy that they don't get challenged. Before building the new level, you want to know what the <u>mean</u> and <u>standard deviation</u> are of the number of minutes people take to complete level 1. You know the following:

1. The # of minutes follows a normal distribution.
2. The probability of a player playing for less than 5 minutes is 0.0045.
3. The probability of a player playing for less than 15 minutes is 0.9641.

Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

$P(X < 5) = 0.0045$

$z_1 = -2.61$

Rcode: qnorm(0.0045,0,1)

Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

$P(X < 15) = 0.9641$

$z_2 = 1.8$

Rcode: qnorm(0.9641,0,1)

Normal Distribution

$-2.61 = \frac{5-\mu}{\sigma}$ and $1.8 = \frac{15-\mu}{\sigma}$

Solving for the above 2 equations, we get

$\mu = 5 + 2.61\sigma$

$\mu = 15 - 1.8\sigma$

Subtracting the two, we get

$0 = -10 + 4.41\sigma \Rightarrow \ \sigma = 10 \div 4.41 = 2.27$

Substituting this value of σ in either of the above 2 equations, we get

$\mu = 5 + 2.61 * 2.27 = 10.925$

Normal Distribution

ON PAGE 13

NO ACHHE DIN FOR SALARIED CLASS

● Employees in India are likely to get an average salary hike of just 9.4 per cent this year.

performer." Moreover, the bell curve is sharpening with a significant drop in the percentage of people in the highest rating.

India sees lower salary hike for 2 straight yrs

PAWAN BALI | DC
NEW DELHI, FEB. 27

No *"acche din"* for the salaried class as Indians will only get single digit hike in salaries in two consecutive years of 2017 and 2018. Lower income growth for two consecutive years has happened for the first time in 22 years, according to HR consultancy firm Aon Hewitt's.

In 2017, the average salaries hike was 9.3 per cent and in 2018 it is projected to remain at the same level at 9.4 per cent despite forecasts of improvement in macro-economic situation, according to Aon Hewitt's annual India salary increase survey.

The need for cost prudence in the wake of ongoing economic uncertainty came across as the single most critical factor for rationalisation of salary budgets, it said.

Moreover, the attrition rate in India is seeing a continuous dip, indicating fewer opportunity to move as the economic got hit by the double whammy of note ban and GST. Overall, attrition has come down from an average of 20 per cent in the previous decade to 15.9 per cent in 2017, said the survey.

"As per the survey, companies in India gave an average pay increase of 9.3 per cent during 2017 marking a departure from the double digit increments given by organisations since the inception of this study," said the company.

The survey was initiated in 1995-1996. Aon believes average pay increases in India will remain between 9.4-9.6 per cent.

It said that the focus on performance is getting sharper year-on-year. "A top performer is getting an average salary increase of 15.4 per cent, approximately 1.9 times the pay increase for an average performer." Moreover, the bell curve is sharpening with a significant drop in the percentage of people in the highest rating.

In the last 10 years salary hikes have been seeing a somewhat downward bias. In 2007, the average salary hike was 15.1 per cent, which went down to 6.6 per cent in 2009 after the financial crisis. In 2010, it again rose sharply to 11.7 per cent and in 2011 it further went up to 12.6 per cent. However, between 2012 to 2016 salary hikes have been around 10 per cent.

The survey said that over the years, with increasing pressure on compensation budgets, there is an emerging focus on rationalisation of budgets.

"Companies are increasingly taking into account the base effect e.g., pay increases for top and senior management is consistently going down," it said.

The study analysed data across 1,000 plus companies from more than 20 industries.

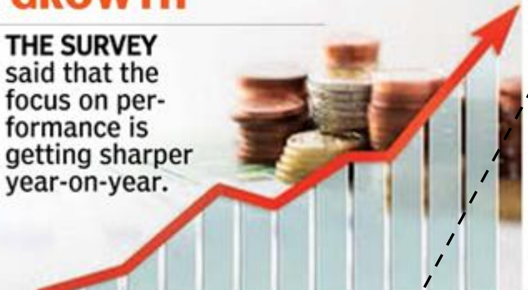
The survey said that sectors such as professional services, consumer internet firms, life sciences, automotive and consumer products continue to project a double digit salary increase for 2018.

Consumer internet firms however, over the past three years have seen a significant drop of 250 basis points, from 12.9 per cent to 10.4 per cent projected for 2018. Engineering services, financial institutions and cement industry is going to see the slowest hike in salaries in 2018.

SLOWER GROWTH

THE NEED FOR cost prudence in the wake of ongoing economic uncertainty came across as the single most critical factor for rationalisation of salary budgets, the survey said.

THE SURVEY said that the focus on performance is getting sharper year-on-year.

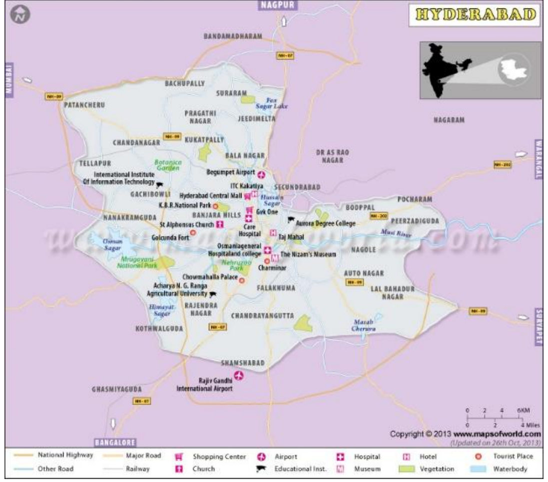


Source: Deccan Chronicle, Hyderabad edition, Feb 28, 2018
Last accessed: March 02, 2018

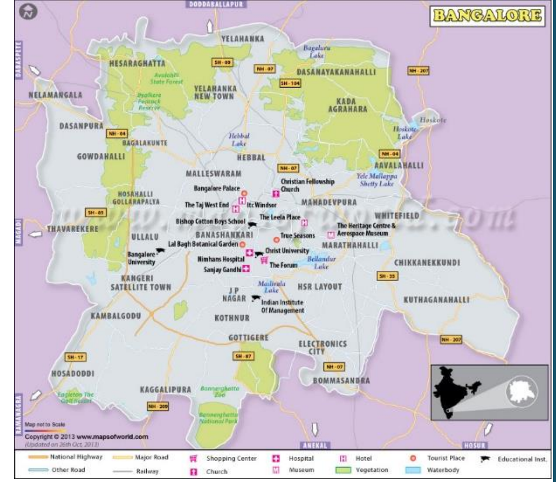
Recall Day 3

End of Day 3

- Revision
  - Expectation and Variance - Properties
  - Skewness and Kurtosis
- Discrete Probability Functions
  - Bernoulli's Experiment, Geometric, Binomial, Poisson
- Continuous Density Functions
  - Exponential
  - Normal Distribution (68-95-99.7 empirical rule)
  - Mean = Median = Mode, Zero Skew and Kurtosis, $X \sim N(\mu, \sigma 2)$
- Z distribution
- One distribution morphing into another distribution
  - Binomial to Normal - if np>5 and nq>5
  - Poisson to Normal - when $\lambda > 15$.



Inspire...Educate...Transform.



HYDERABAD
2nd Floor, Jyothi Imperial, Vamsiram Builders, Old Mumbai Highway, Gachibowli, Hyderabad - 500 032
+91-9701685511 (Individuals)
+91-9618483483 (Corporates)

BENGALURU
Floors 1-3, L77, 15th Cross Road, 3A Main Road, Sector 6, HSR Layout, Bengaluru – 560 102
+91-9502334561 (Individuals)
+91-9502799088 (Corporates)

Social Media

Web:	http://www.insofe.edu.in
Facebook:	https://www.facebook.com/insofe
Twitter:	https://twitter.com/Insofeedu
YouTube:	http://www.youtube.com/InsofeVideos
SlideShare:	http://www.slideshare.net/INSOFE
LinkedIn:	http://www.linkedin.com/company/international-school-of-engineering