Confusion Matrix And Distributions

Confusion Matrix

- Buyer or Non-Buyer
- A retail store’s marketing team uses analytics to predict who islikely to buy a newly introduced high-end (read “expensive”)product. Indicate which measure is more important for the business to track and explain why. Calculate other measures also.
-

Buyer or Not		Actual		Total
		Negative	Positive	
Predicted	Negative	725	158	883
	Positive	75	302	377
Total		800	460	1260

Confusion Matrix

- **Buyer or Non-Buyer**
- State/Calculate:
- TP = 302   TN = 725   FP = 75   FN = 158
- Should the business be more worried about FP or FN or equally worried about both of them? Why?
- **FN**. If the model predicts that the person will not buy, the product will not be marketed to him/her, and the business will lose...er, business. FP is not such a big worry since only the cost of a phone call, SMS or sending a catalog will be lost.

Confusion Matrix

What is more important: Recall, Precision or Accuracy?

$$Recall\ (or\ Sensitivity) = \frac{TP}{TP + FN} = \frac{302}{460} = 65.6\%$$

$$Precision = \frac{TP}{TP + FP} = \frac{302}{377} = 80.1\%$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} = \frac{1027}{1260} = 81.5\%$$

$$Specificity = \frac{TN}{TN + FP} = \frac{725}{800} = 90.6\%$$

$$F_1 = \frac{2 * Recall * Precision}{Recall + Precision} = \frac{2 * 0.656 * 0.801}{0.656 + 0.801} = 72.1\%$$

Confusion Matrix

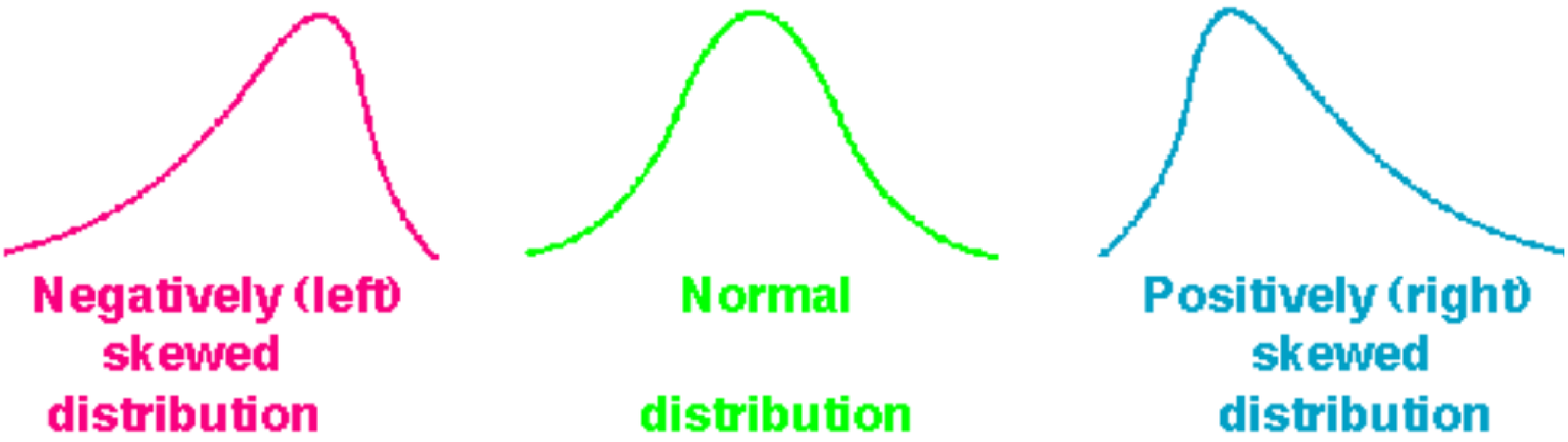
Rheumatoid arthritis is prevalent among 0.75% of adult Indian population. Assuming a test with a certain sensitivity and specificity gives the following results, what are the values for these metrics?

Rheumatoid Arthritis Diagnostic Test		**Predicted**		**Total**
		Positive	Negative	
Actual	Positive	20	10	30
	Negative	318	3652	3970
Total		338	3662	4000

$$Recall\ (Sensitivity) = \frac{20}{30} = 0.67$$

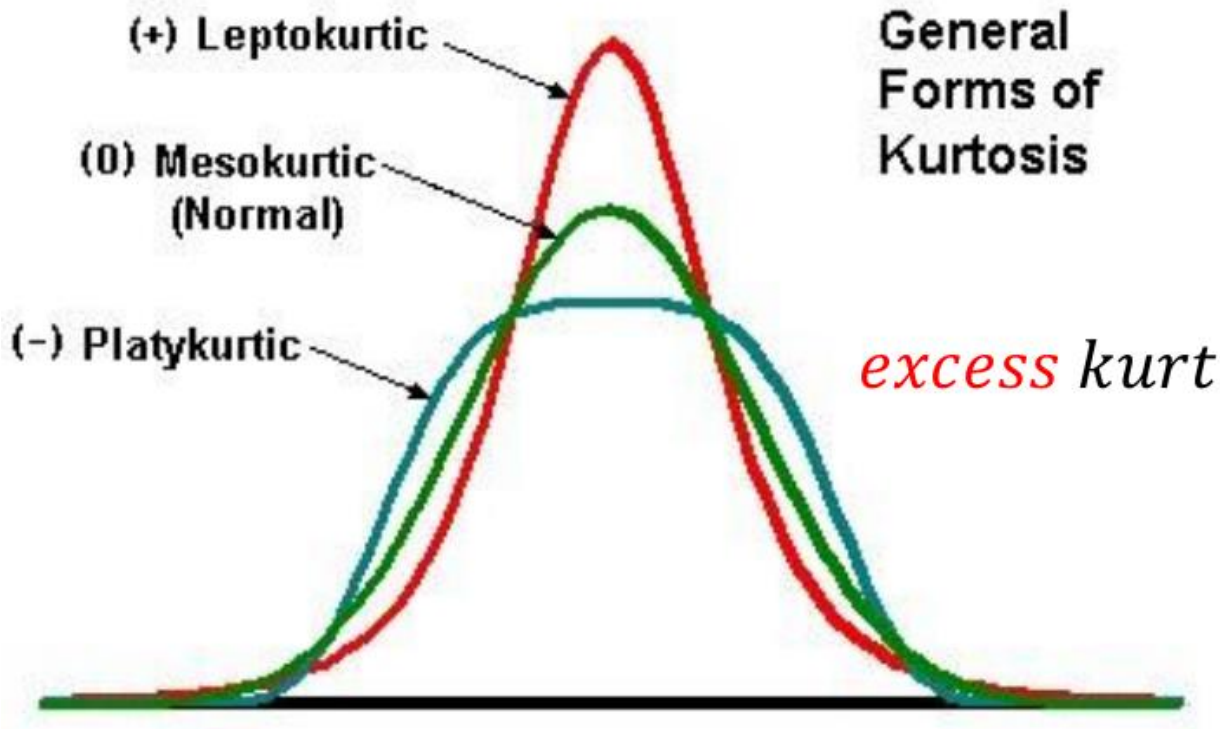
$$Specificity = \frac{3652}{3970} = 0.92$$

Understanding the shape of a PDF - Skewness

- A measure of symmetry. Negative skew indicates mean is less than median, and positive skew means median is less than mean.
- 

$$skew(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

Understanding the shape of a PDF - Kurtosis

- A measure of the 'peaked'ness and 'tailed'ness of the data distribution. Negative kurtosis means a flat distribution with light tails. Positive kurtosis means a peaked distribution with heavy tails.
- 

$$\textit{excess}\ kurt(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3$$

Describing a Distribution – Summary of Moments

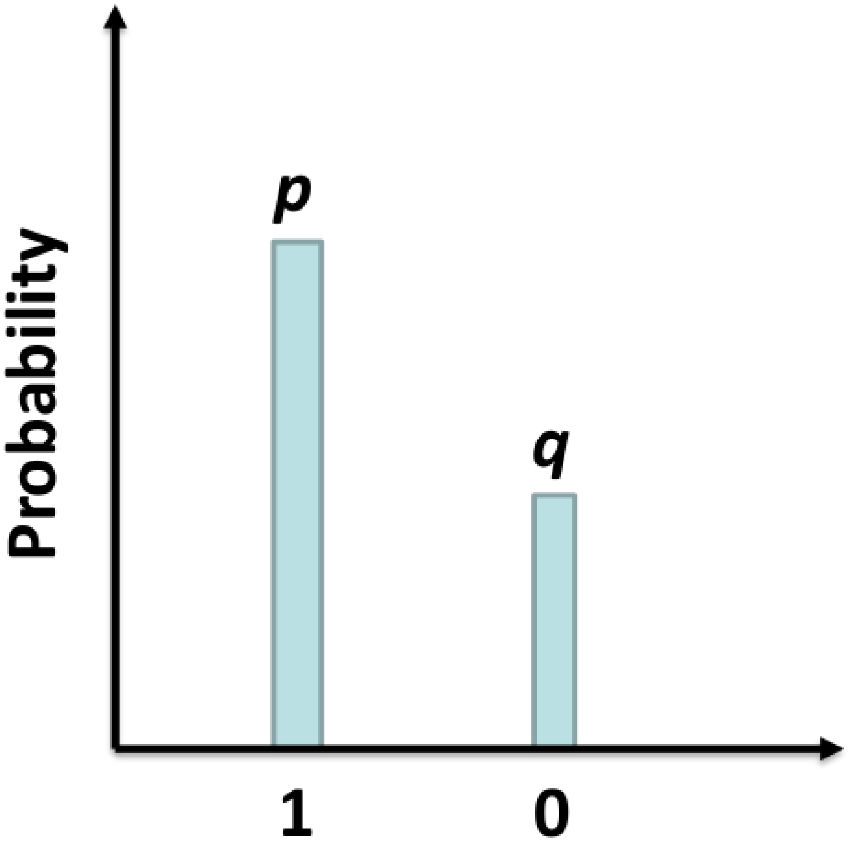
Measure	Formula	Description
Mean (μ)	$E(X)$	Measures the centre of the distribution of X
Variance (σ^2)	$E[(X-\mu)^2]$	Measures the spread of the distribution of X about the mean
Skewness	$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$	Measures asymmetry of the distribution of X
Kurtosis (excess)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



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

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

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

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

Geometric Distribution

- Number of independent and identical 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*, $P(X = r) = q^{r-1}p$ — *(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.

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

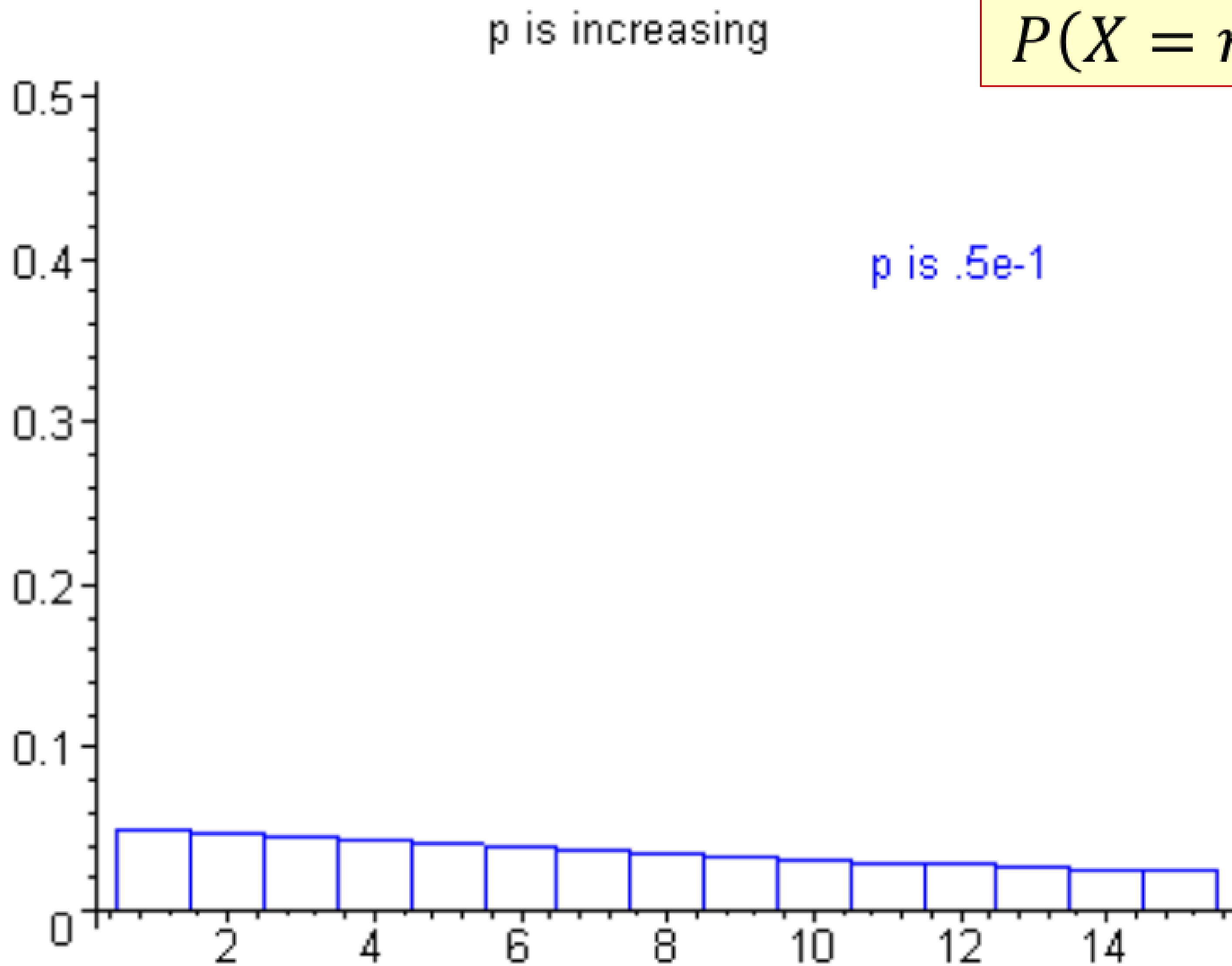
* Probability Mass Function ** Cumulative Distribution Function

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.

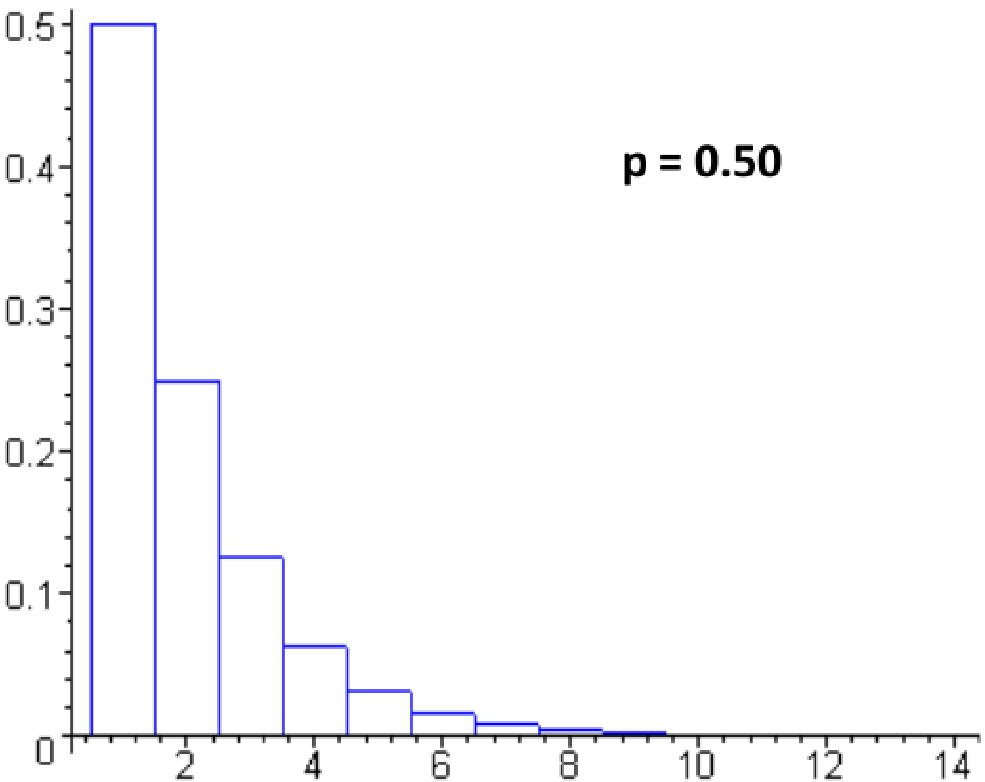
X~Geo(p)

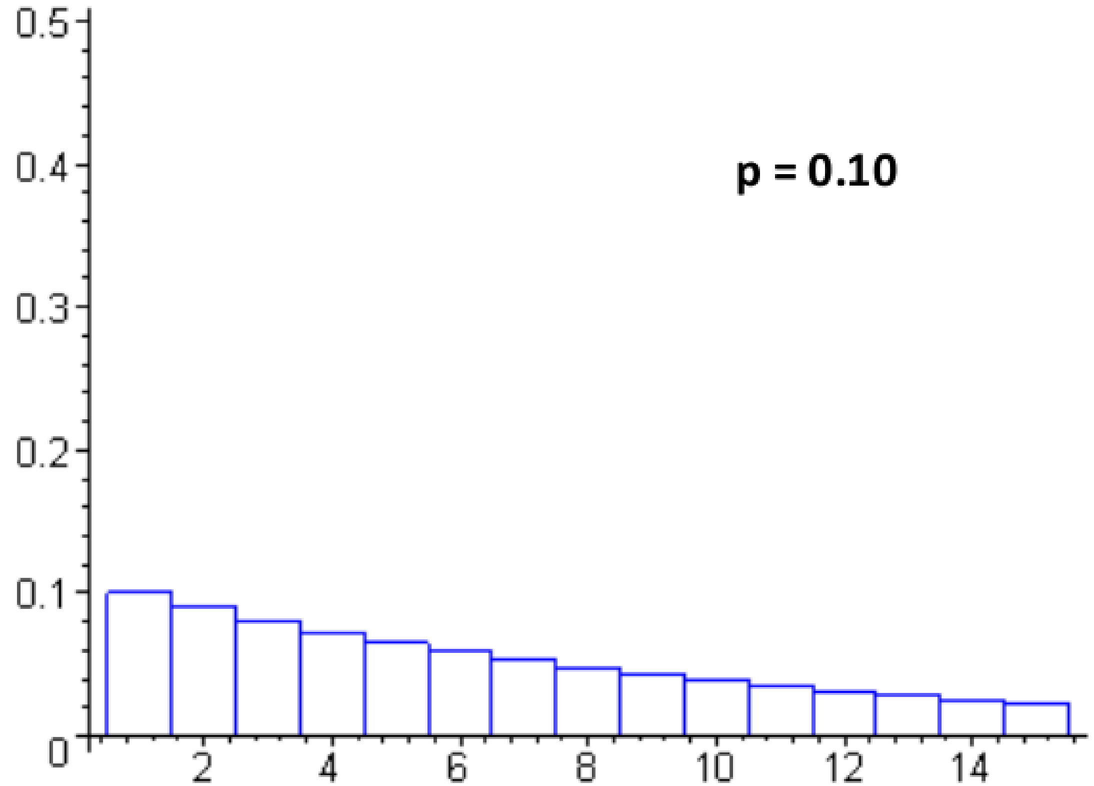
$$P(X = r) = q^{r-1}p$$



Ref: http://personal.kenyon.edu/hartlaub/MellonProject/Geometric2.html

X~Geo(p)



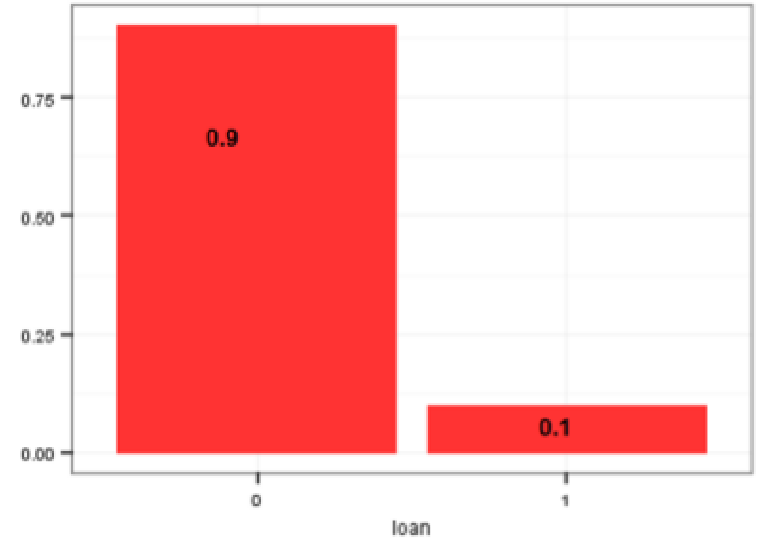


$$\boxed{P(X = r) = q^{r-1}p}$$

Ref: http://personal.kenyon.edu/hartlaub/MellonProject/Geometric2.html

Binomial Distribution

- If I randomly pick 10 people, what is the probability that I will get exactly
-
    - 0 loan takers = 0.9^{10}
    - 1 loan taker = $10 * 0.1^1 * 0.9^9$
    - 2 loan takers = $C_2^{10} * 0.1^2 * 0.9^8$



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

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

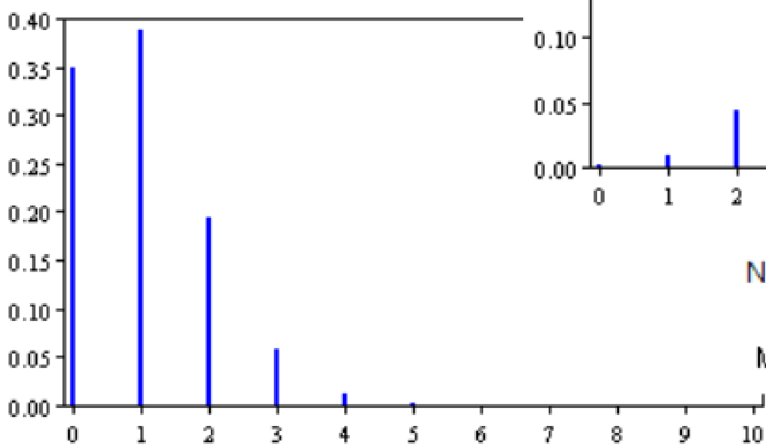
Binomial Distribution

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

When to use?

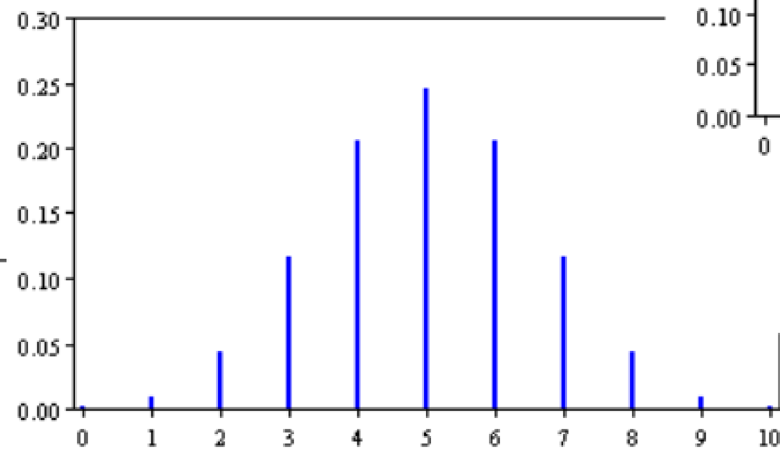
- You run a series of independent trials.
- There can be either a success or a failure for each trial, and the probability of success is the same for each trial.
- There are a finite number of trials, and you are interested in the number of successes or failures.

X~B(n,p)



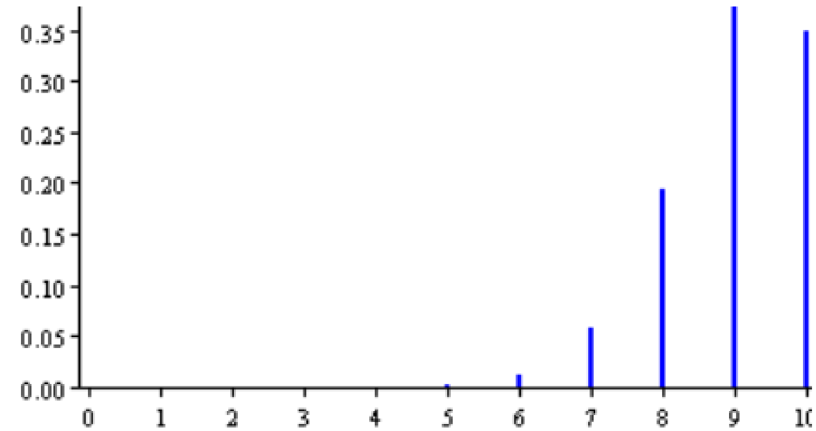
N: 10 p: 0.1

Mean=N x p=1.00, Sd=$\sqrt{\text{N x p x (1-p)}}$=0.95



N: 10 p: 0.5

Mean=N x p=5.00, Sd=$\sqrt{\text{N x p x (1-p)}}$=1.58



N: 10 p: 0.9

Mean=N x p=9.00, Sd=$\sqrt{\text{N x p x (1-p)}}$=0.9

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

Ref: http://onlinestatbook.com/2/probability/binomial_demonstration.html

Poisson Distribution

- Probability of getting 15 customers requesting for loans in a given day given on average we see 10 customers

$$\lambda = 10 \; and \; r = 15$$

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

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

Poisson Distribution

$E(X) = \lambda$ Can be equated to *np* of Binomial if *n* is large (>50) and *p* is small (<0.1)

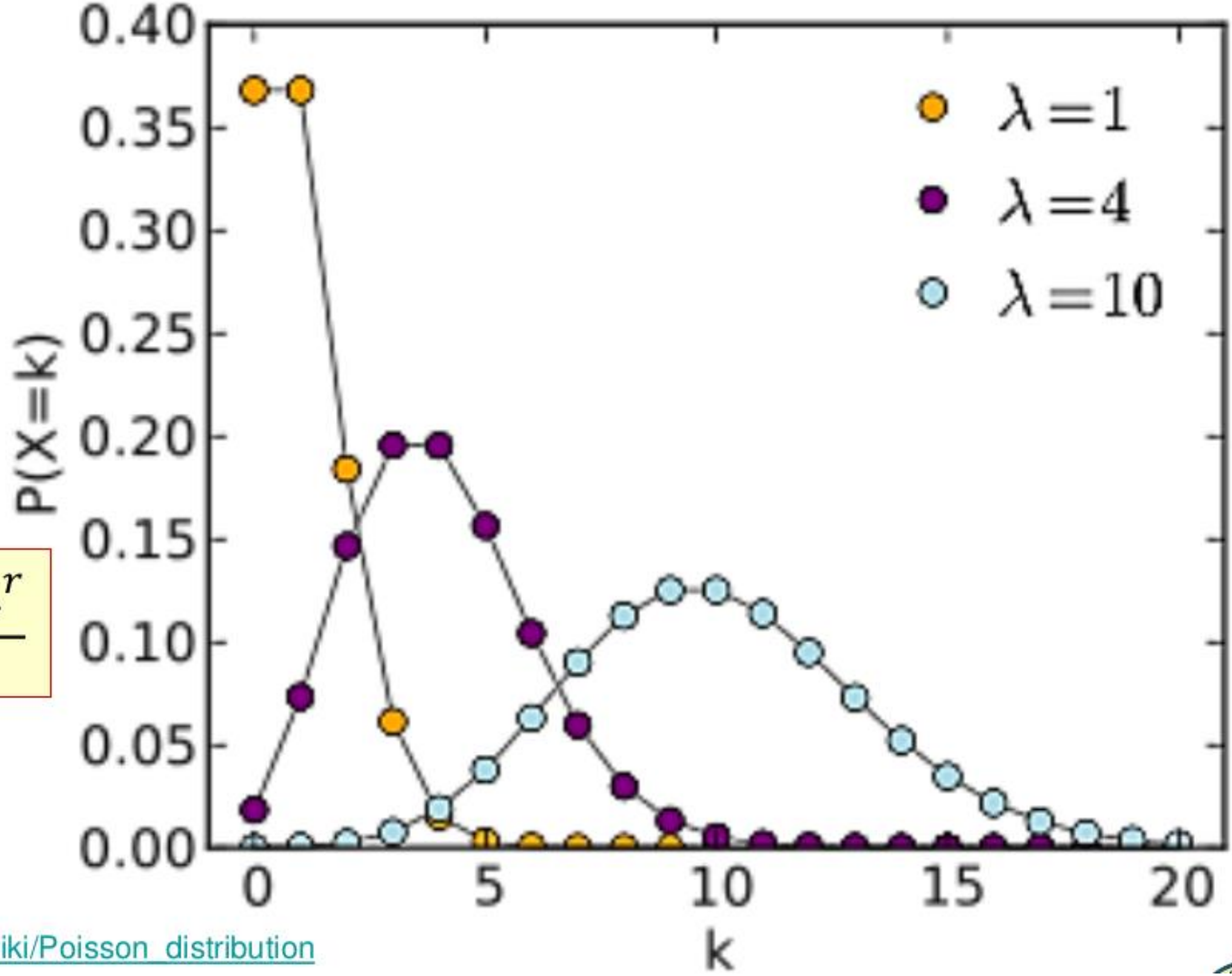
$Var(X) = \lambda$ Can be equated to *npq* of Binomial in the above situation.

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: https://en.wikipedia.org/wiki/Poisson_distribution

Poisson Distribution

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

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

Probability that she will not have a customer for *n* days

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

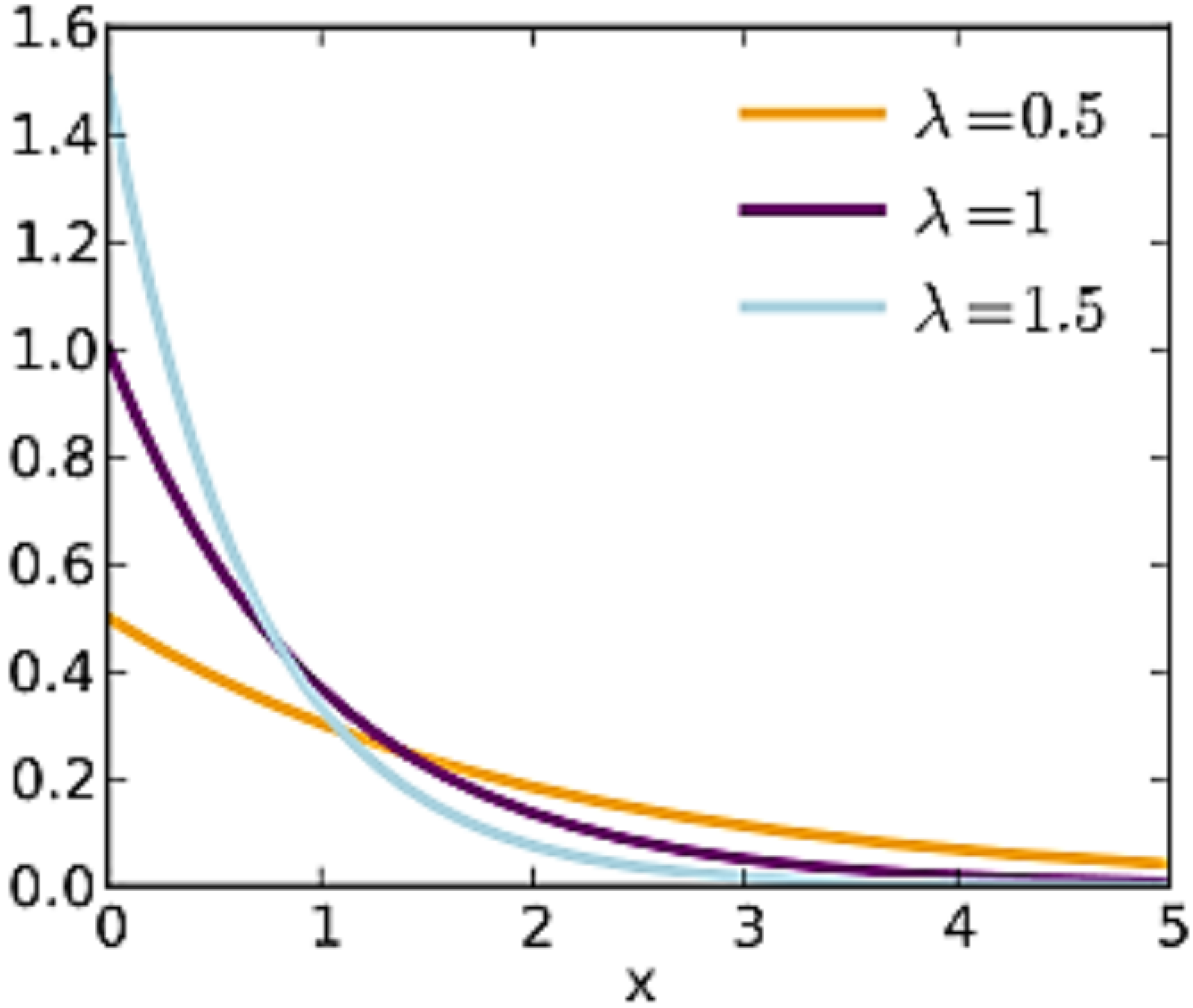
Exponential Distribution

Probability that a customer will visit in *n* days: $1 - e^{-n\lambda}$

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

X~Exp(λ)

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



Ref: http://en.wikipedia.org/wiki/Exponential_distribution

Exponential Distribution

- Poisson process
- Continuous analog of Geometric distribution

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

Distributions

- 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
- Exponential: Time between events

Probability Distributions

Here are a few 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?

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?

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?

Probability Distributions

Solutions

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?

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

E(X) = np = 3

Var(X) = npq = 2.1

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

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

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

Probability Distributions

Solutions

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?

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) = 0.368

Probability DistributionU+0073

Solutions

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?

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

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?

What is λ?

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

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?

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

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 λ_1 and λ_2 if we use first formula?

$$\lambda_1 = 0.5 \text{ and } \lambda_2 = 1.5$$
$$\lambda_1 + \lambda_2 = 2$$

$$P(X+Y > 3) = P(X+Y=4) + P(X+Y=5) + P(X+Y=6) + \cdots$$
$$= 1 - P(X+Y \le 3) = 1 - (P(X+Y=0) + P(X+Y=1) + P(X+Y=2) + P(X+Y=3))$$

Poisson or Exponential?

Given a Poisson process:

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

Exponential

Poisson or Exponential?

The tech support centre of a computer retailer receives 5 calls per hour on an average. What is the probability that the centre will receive 8 calls in the next hour? What is the probability that more than 30 minutes will elapse between calls?

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

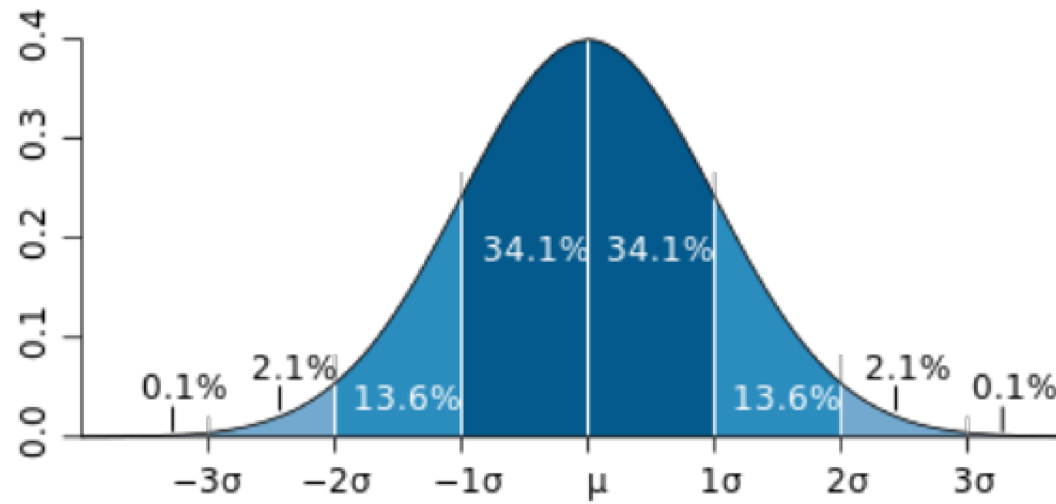
$$P(Time\ between\ calls > 0.5) = \int_{0.5}^{\infty} \lambda e^{-\lambda T} dT$$

$$= -e^{-\lambda T}\Big]_{0.5}^{\infty} = e^{-5*0.5} = 0.082$$

NORMAL DISTRIBUTION

Normal (Gaussian) Distribution

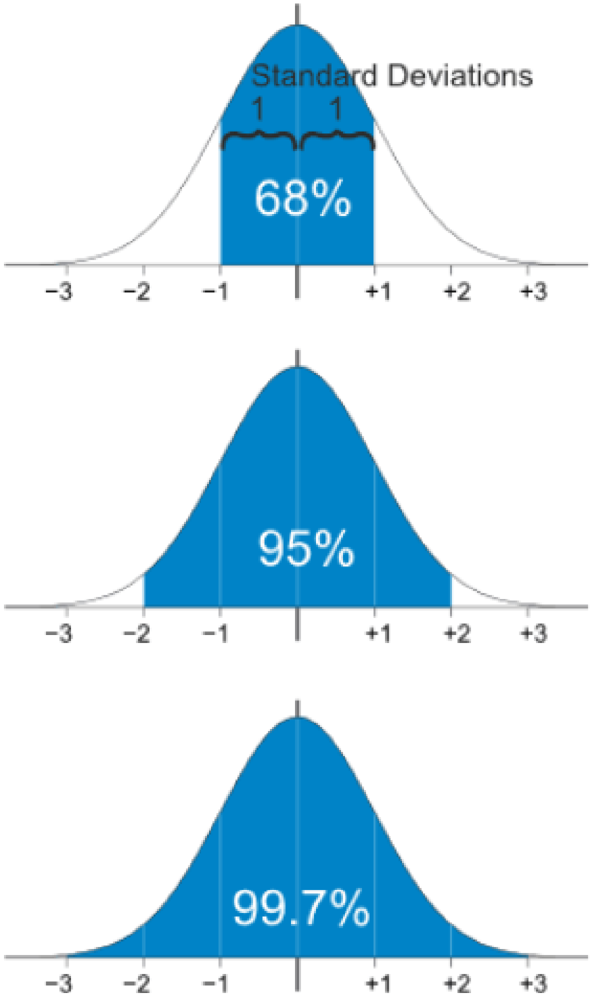
- Mean = Median = Mode
- 68-95-99.7 empirical rule
- Zero Skew and Kurtosis
- X ~ N(μ, σ^2)
- Shaded area gives the probability that X is between the corresponding values



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

Measures of Spread (Dispersion)

You know the 68-95-99.7 rule.



A company produces a lightweight 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.

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

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

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

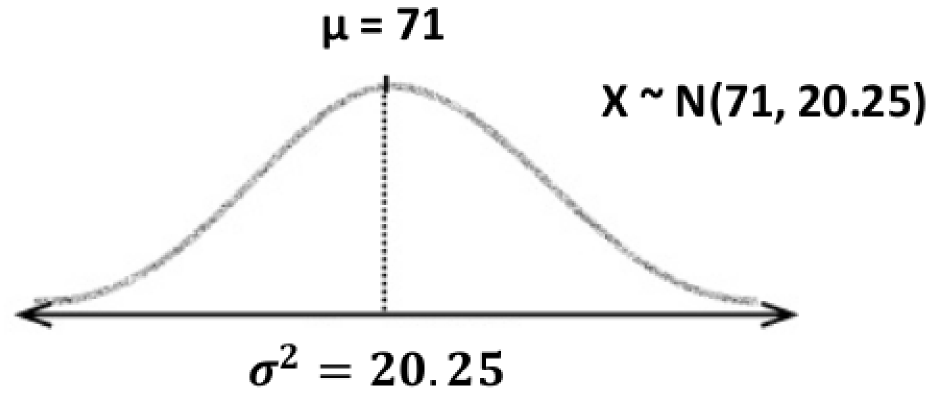
Image source: http://www.mathsisfun.com/data/standard-normal-distribution.html

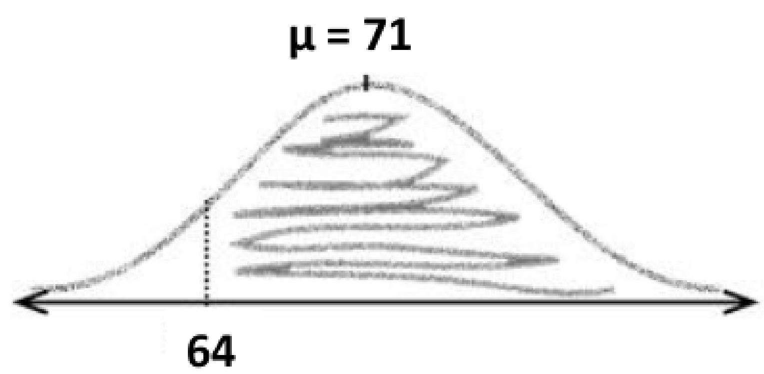
Calculating Normal Probabilities

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.





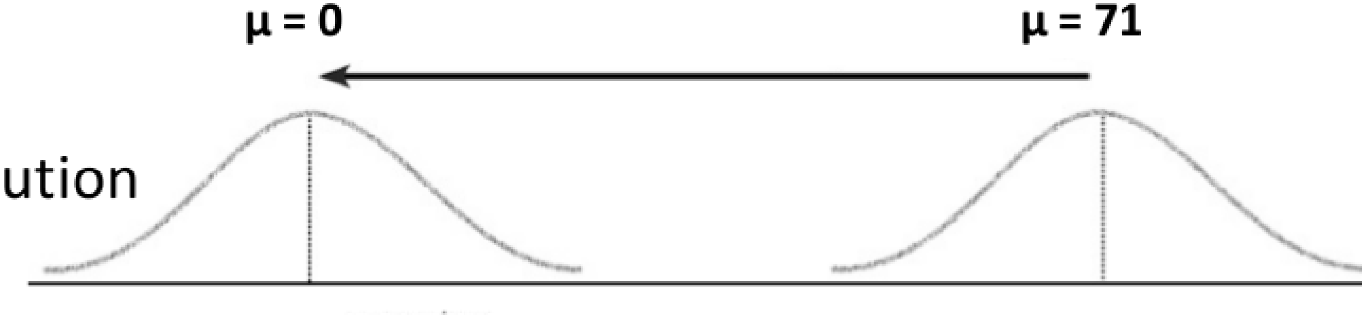
Calculating Normal Probabilities

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

1. Move the mean

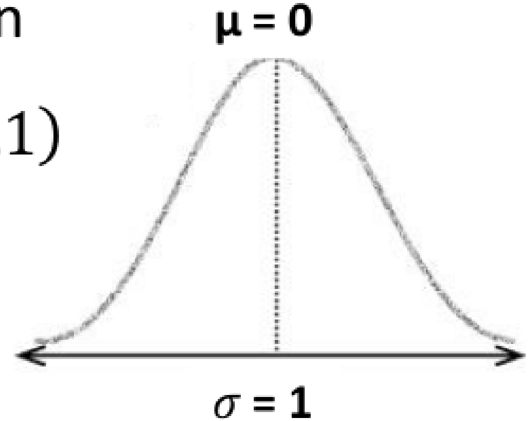
   This gives a new distribution

   X-71 ~ N(0,20.25)



2. Squash the width by dividing by the standard deviation

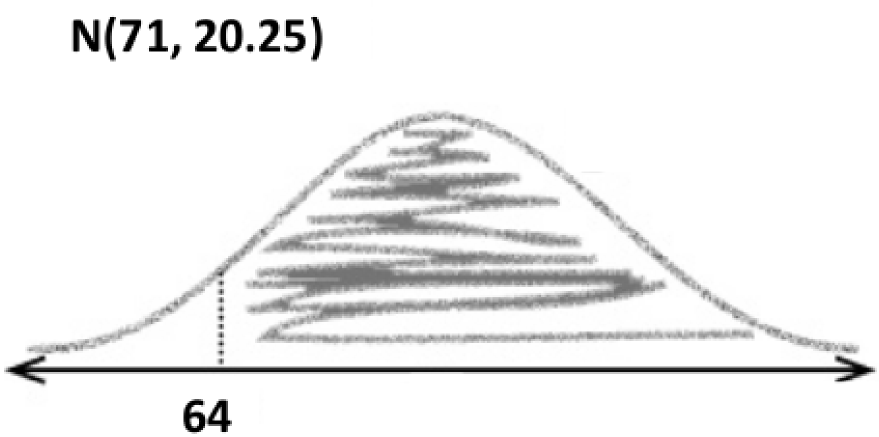
   This gives us $\frac{X-71}{4.5} \sim N(0,1)$

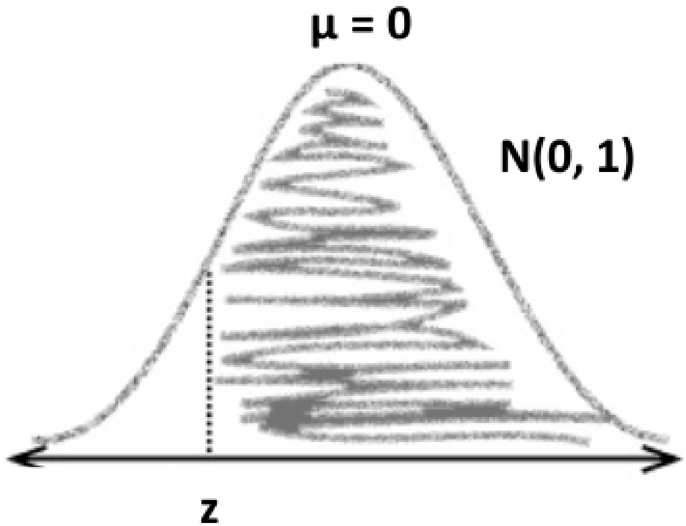


$Z = \frac{X-\mu}{\sigma}$ is called the Standard Score or the z-score.

Calculating Normal Probabilities

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





Note: R does this step internally but you must understand the concept of z-score as this is fundamental to most statistical thinking

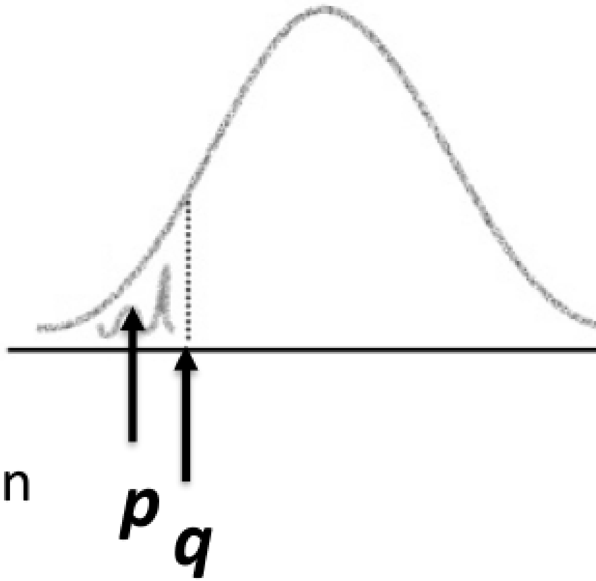
$Z = \frac{64-71}{4.5} = -1.56$ in the case of our problem.

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.

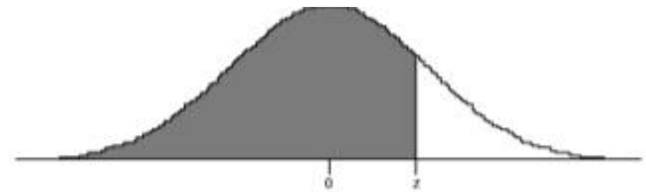
Calculating Normal Probabilities

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

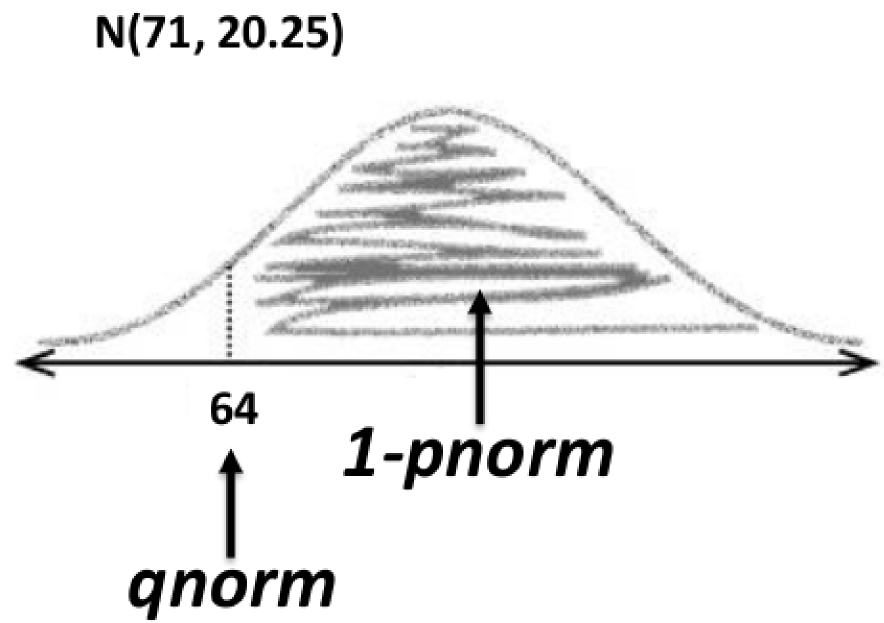
Step 3: Get the probability from R

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



Attention Check

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

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

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

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

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.

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

A. $z = \frac{69-71}{4.5} = -0.44$; P(Z<-0.44) = 0.33, ∴ P(Z>-0.44) = 0.67 or 67%

A. 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?

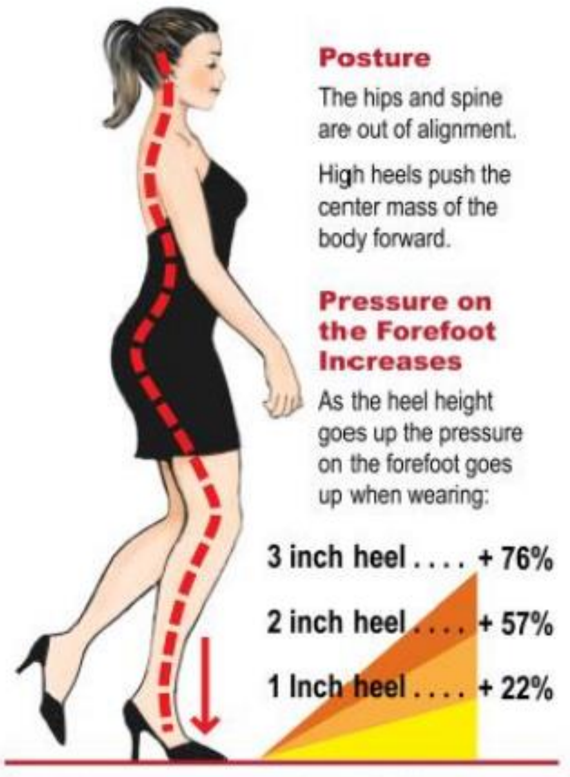


Figure 2

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