Introduction to Healthcare Supply Chain (HCSC) Analytics

FINNISH NATIONAL AGENCY FOR EDUCATION

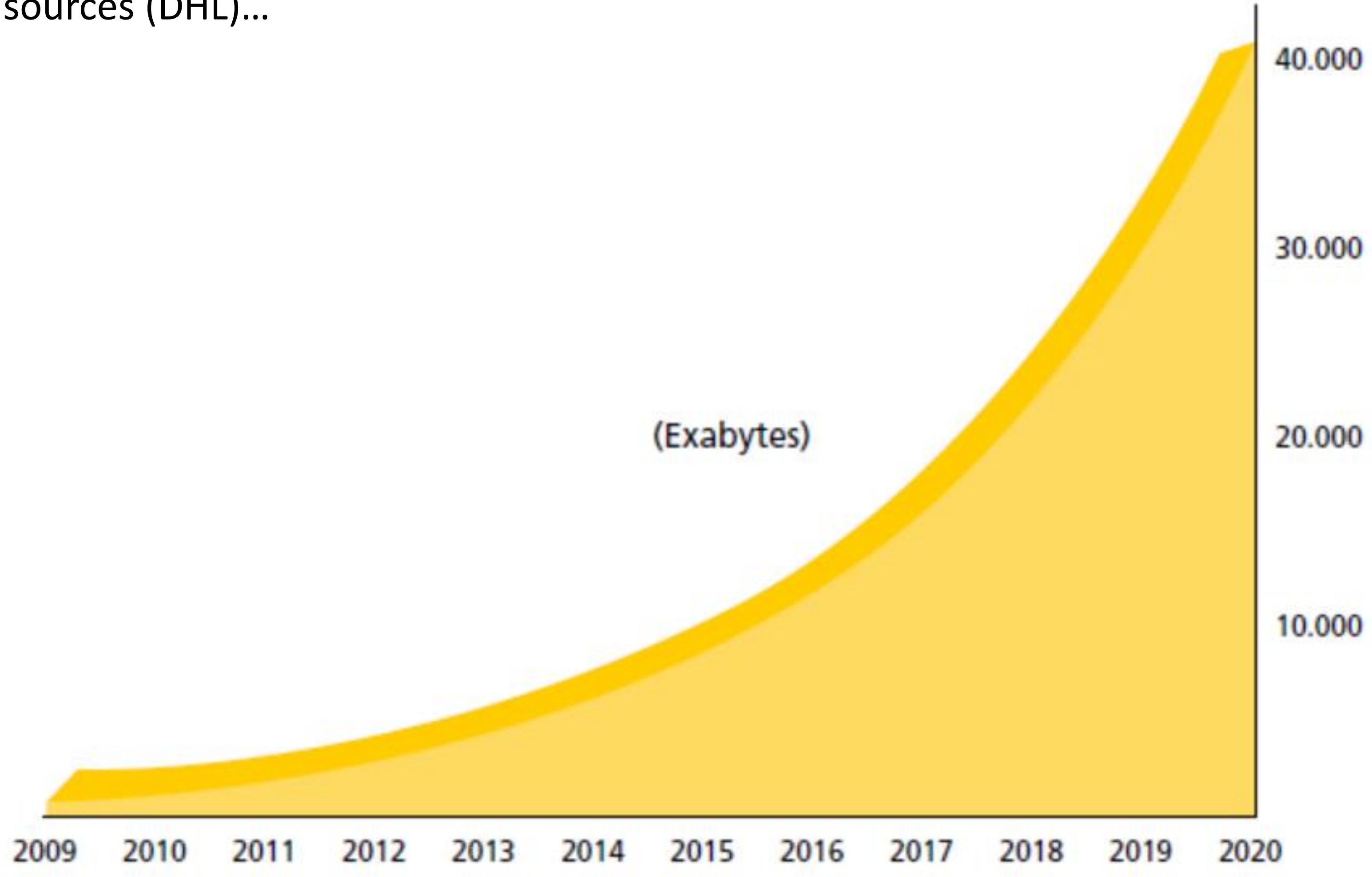
Co-funded by the Erasmus+ Programme of the European Union

Industry 4.0

Data every where…

Gigamon Blog

The massive deployment of connected devices such as cars, smartphones, RFID readers, webcams, and sensor networks adds a huge number of autonomous data sources (DHL)…



Exponential data growth between 2010 and 2020; **Source:** IDC's Digital Universe Study, sponsored by EMC, December 2012

Where is the data processing?

Gigamon Blog

Data processing…(e.g. Artificial Intelligence)

dataset - experiences

algorithms

human instructions

needed

computer…

capable to:

learning

decision making

Process efficiency

One purpose alone

HR

FR

Time

MR

Data processing…(e.g. Artificial Intelligence)

dataset - experiences

algorithms

human instructions

needed

Is this sufficient to make decisions?

computer…

capable to:

learning

decision making

Process efficiency

HR

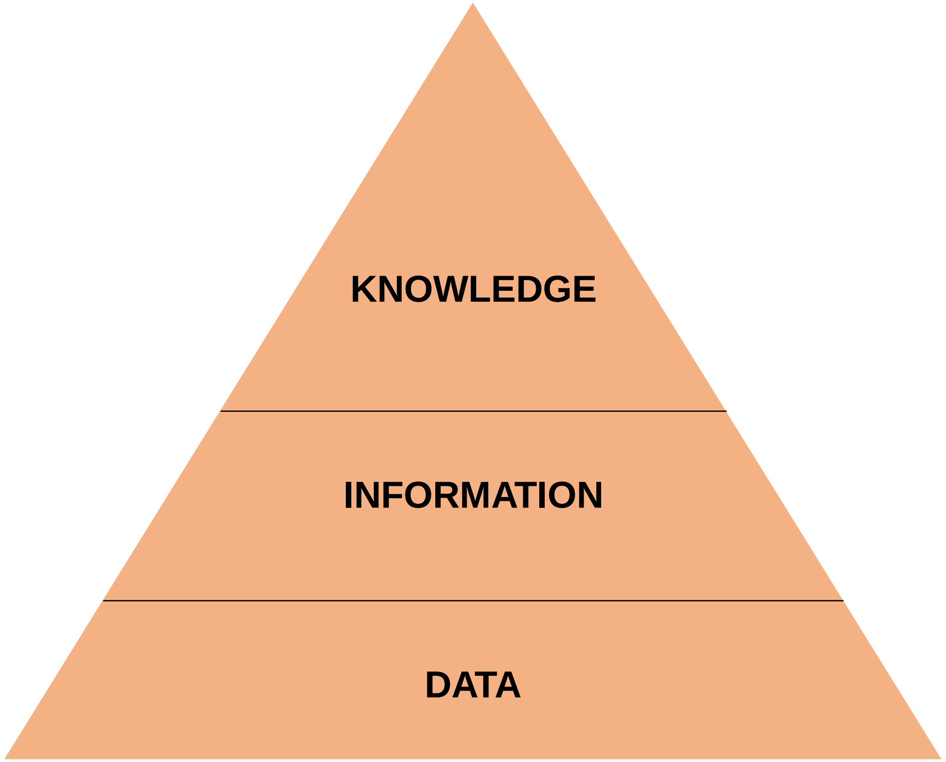
FR

Time

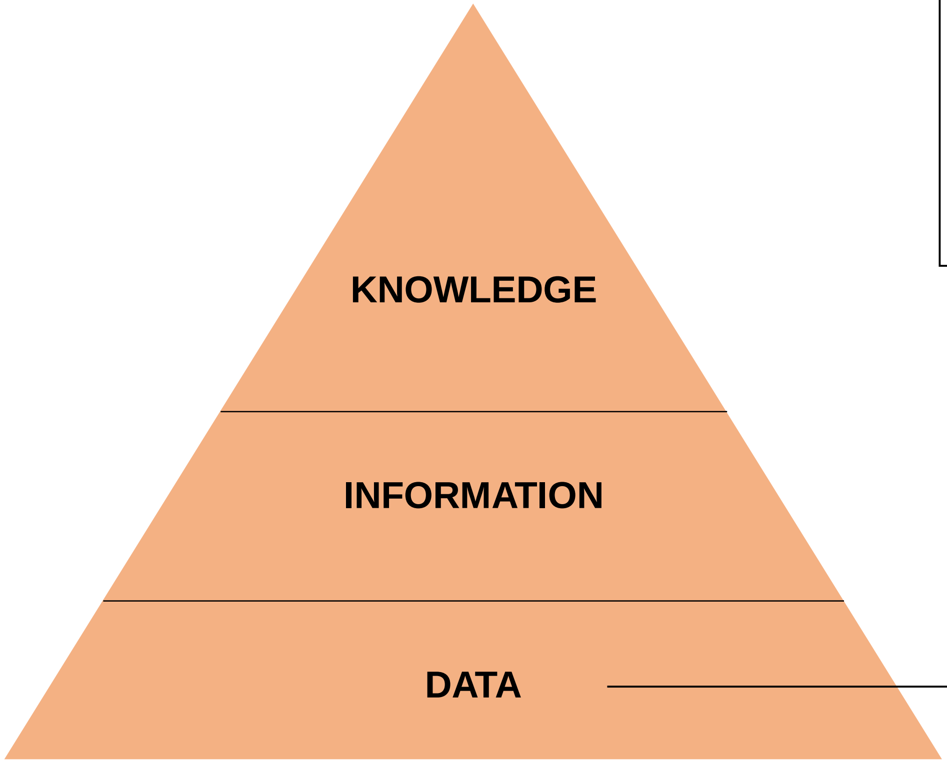
MR

One purpose alone

The great confusion…

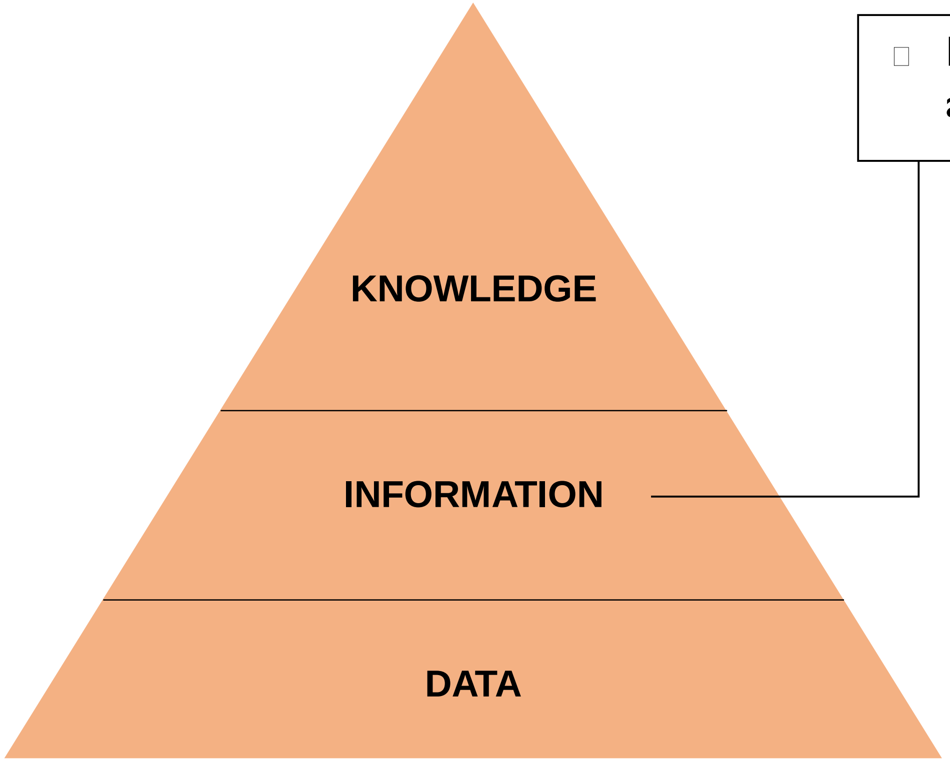


The great confusion…



- Data **are** raw facts and figures that on their own have no meaning
- These can be any alphanumeric characters i.e. text, numbers, symbols

The great confusion…

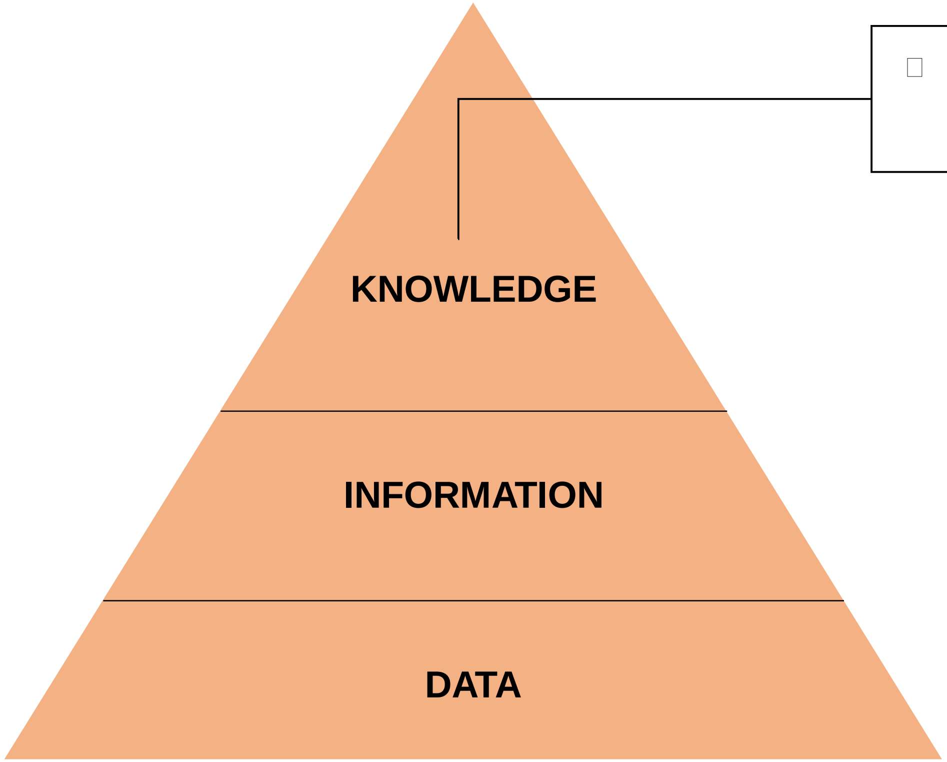


- Data that has been processed within a context or form to give it meaning

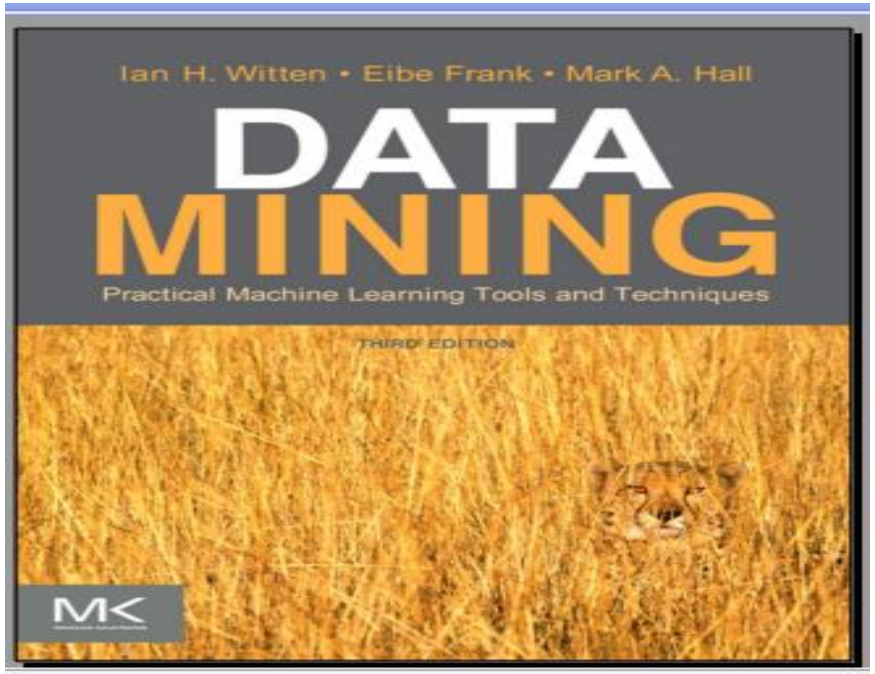
Responses to the market research question – "Would you buy brand x at price y?"

Yes, Yes, No, Yes, No, Yes, No, Yes, No, Yes, Yes

The great confusion…



- Knowledge is the understanding of rules needed to interpret information

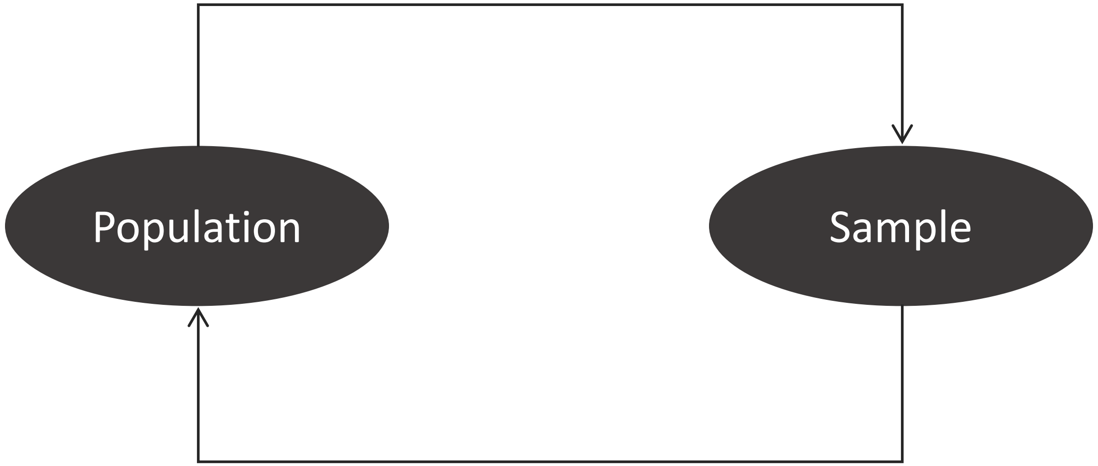


HCSC Analytics…data processing for decision making aimed to:

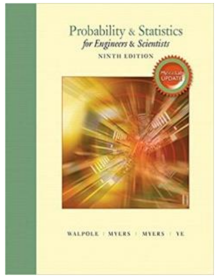
Hierarchical level	Facility	Supply and Inventory	Transportation	Customer service
Strategic	- Facility location - Capacity setting - Technology selection - Process configuration - Setting the IT system for planning and controlling	- Defining the inventory policy - Identify the supplier list and select the best ones - Product design - Choose the IT system (S&I) - Warehouse design - Material handling system	- Transportation mode - IT system (Trans)	- Define the service policy and strategy - Portfolio indicators
Tactical	- Capacity planning during mid term	- Purchase planning (procurement) - Definition of supplies - Planning the inventory level - Planning the safety stock	- Transportation system capacity - Fleet routing - Transportation planning during mid term	- Demand projection during mid term - Advertisement planning
Operative	- Order scheduling - Production execution - Order control - Maintenance planning	- Order dispatching and packing - Material requirement planning - Purchase control - Stock control - Discharge and loading operations	- Delivery planning - Vehicle routing - Control of transport operations	- Demand projection (short term) - Tracking the customer service indicators - Loyalty activities

Transforming data into predictive insights…

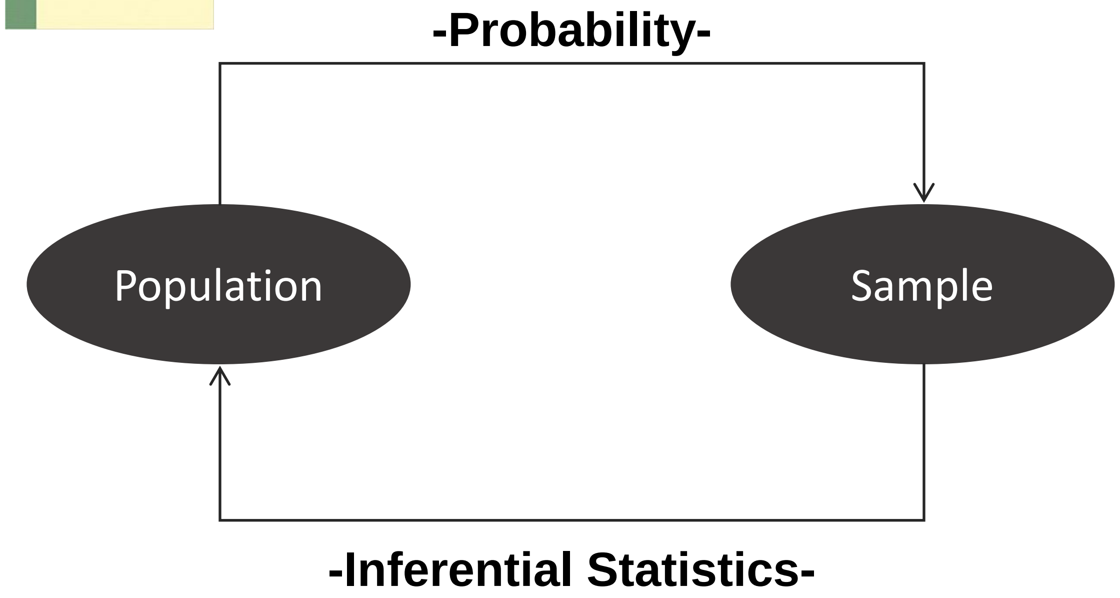
-Probability-

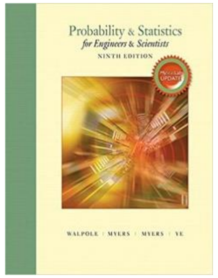


-Inferential Statistics-

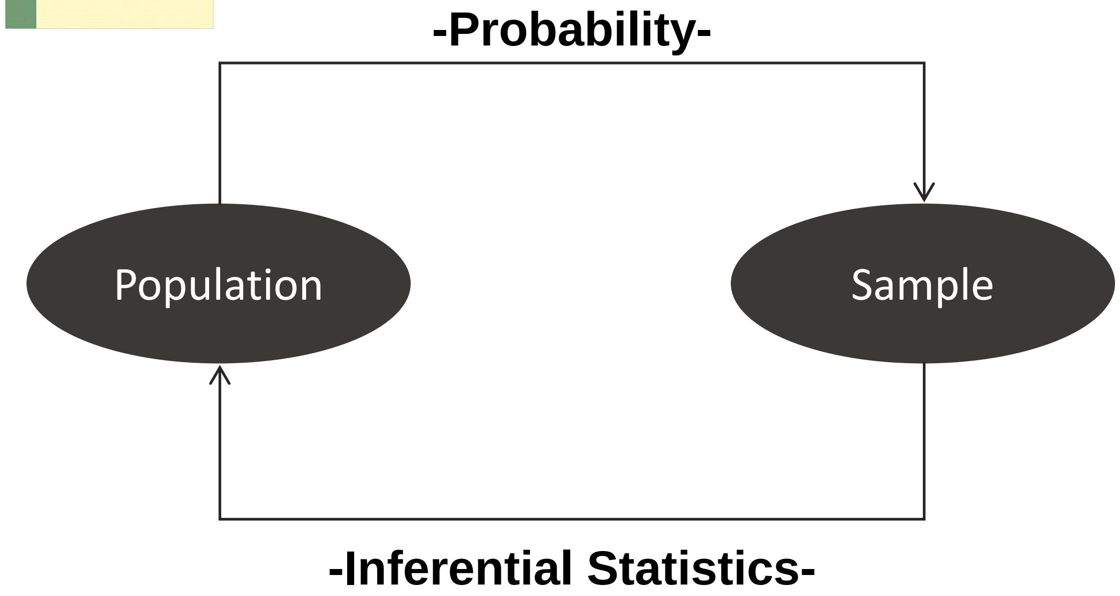


…elements in probability allow us to draw conclusions about characteristics of hypothetical data taken from the population, based on known features of the population.



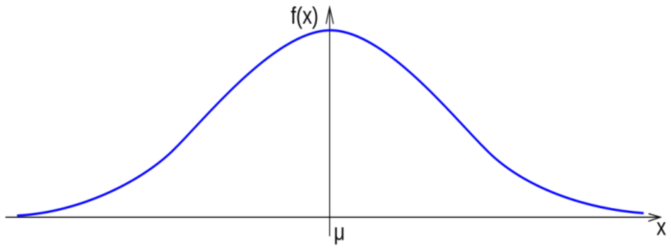


…elements in probability allow us to draw conclusions about characteristics of hypothetical data taken from the population, based on known features of the population.



Most useful aspect of
"theory of probabilities"
in data analytics

Probability distribution



Probability distribution

is a mathematical formula that describes the probabilities of occurrence of different possible outcomes in an experiment.

Probability distribution → is a mathematical formula that describes the probabilities of occurrence of different possible outcomes in an experiment.

finite number of outcomes → **Discrete** [predicting Success / Failure]

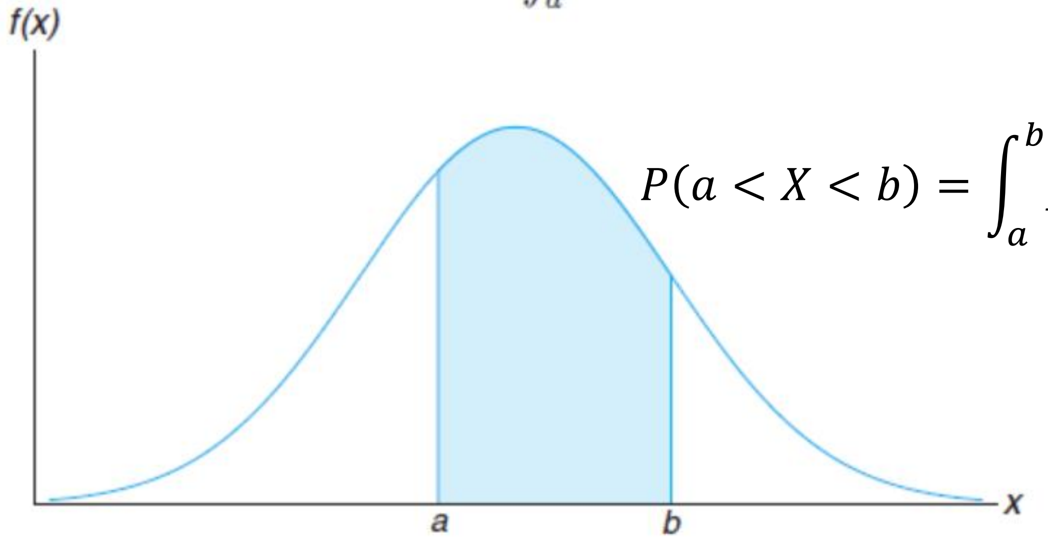
- ❑ Binomial Distribution
- ❑ Poisson Distribution
- ❑ Bernoulli Distribution
- ❑ Geometric Distribution
- ❑ Others.

infinite number of outcomes → **Continuous** [predicting cost]

- ❑ Normal Distribution
- ❑ Uniform Distribution
- ❑ Chi-squared Distribution
- ❑ Exponential Distribution
- ❑ Others.

Continuous Probability Distribution

$$P(a < X < b) = \int_a^b f(x)\, dx.$$

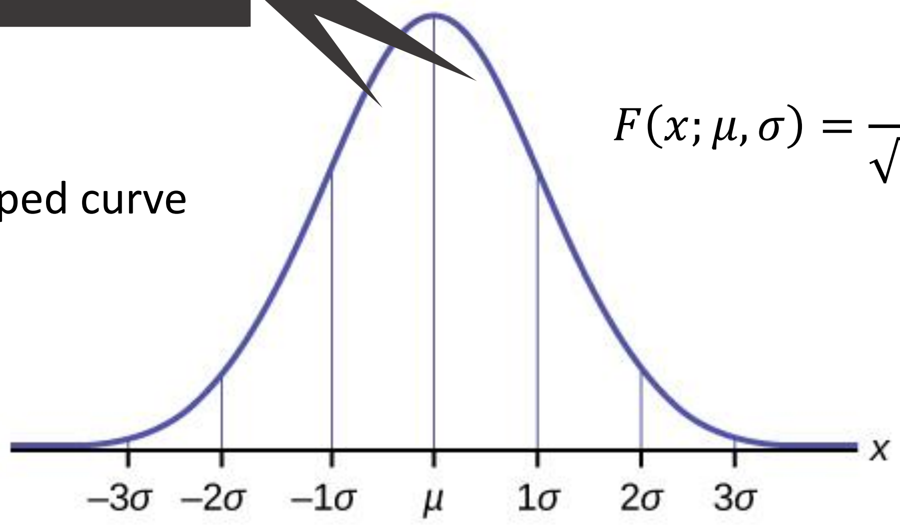


$$P(a < X < b) = \int_a^b f(x)\, dx$$

Continuous Probability Distribution (Normal)

Highly probable…logical explanation

…bell-shaped curve



Density function…

$$F(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad , \qquad -\infty < x < +\infty$$

x: random variable

μ: mean

Deeper study later

σ: standard deviation

π: 3.14159 …

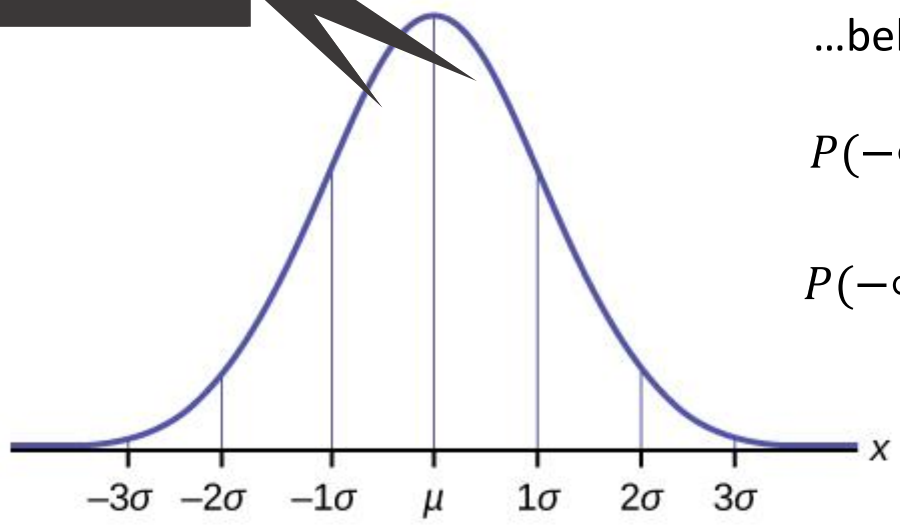
e: 2.71828 …

Normal Distribution: describes many phenomena that occur in nature, industry, and research

$\boldsymbol{X \sim N(\mu, \sigma^2)}$ **…** $\boldsymbol{\mu \pm \sigma}$

Continuous Probability Distribution (Normal)

Highly probable…logical explanation



Properties…

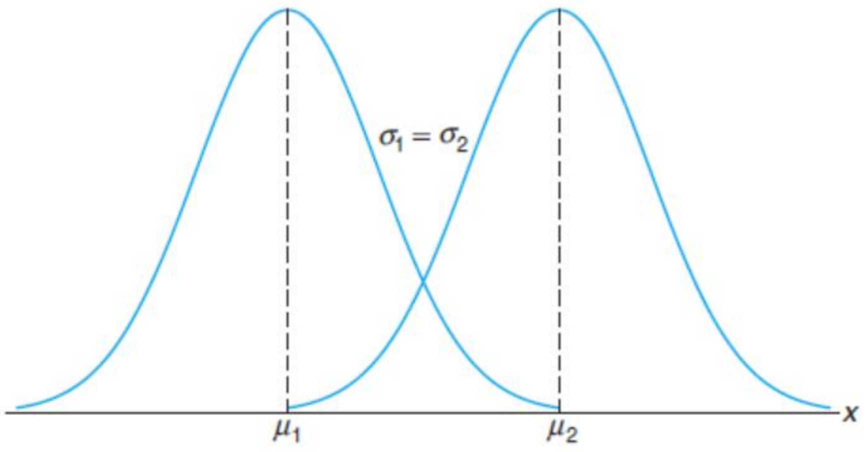
…bell-shaped curve

$$P(-\infty < X < +\infty) = \int_{-\infty}^{+\infty} f(x)\,dx = 1$$

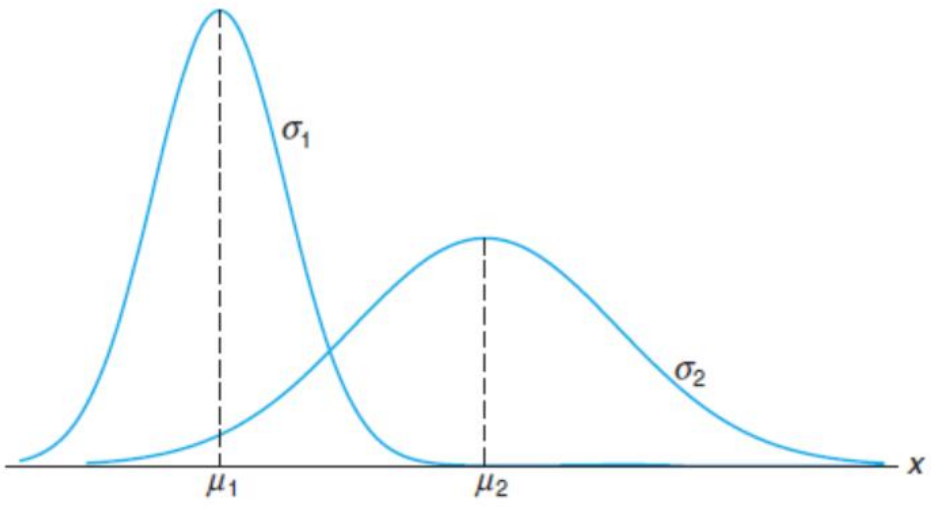
$$P(-\infty < X < \mu) = P(\mu < X < +\infty)$$

Normal Distribution: describes many phenomena that occur in nature, industry, and research

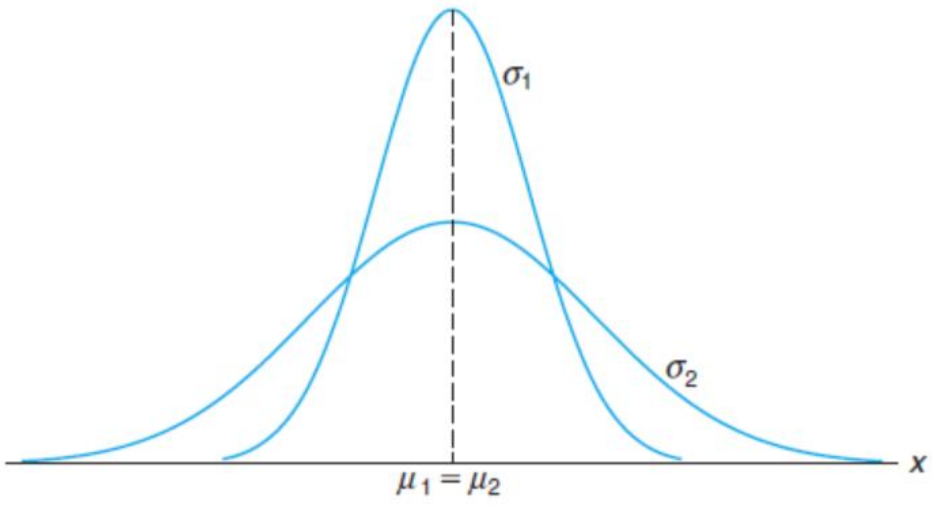
Continuous Probability Distribution (Normal)



Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$.



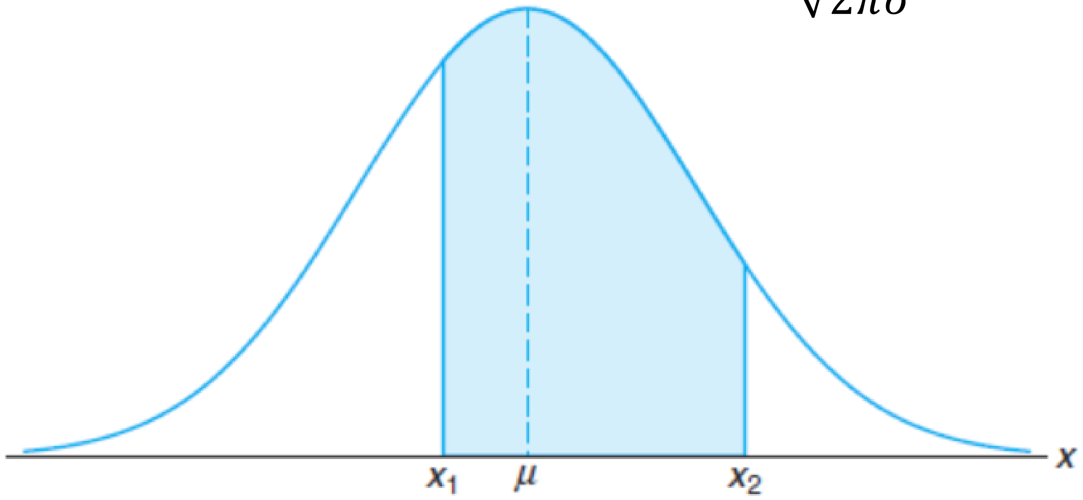
Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$.



Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$.

Continuous Probability Distribution (Normal)

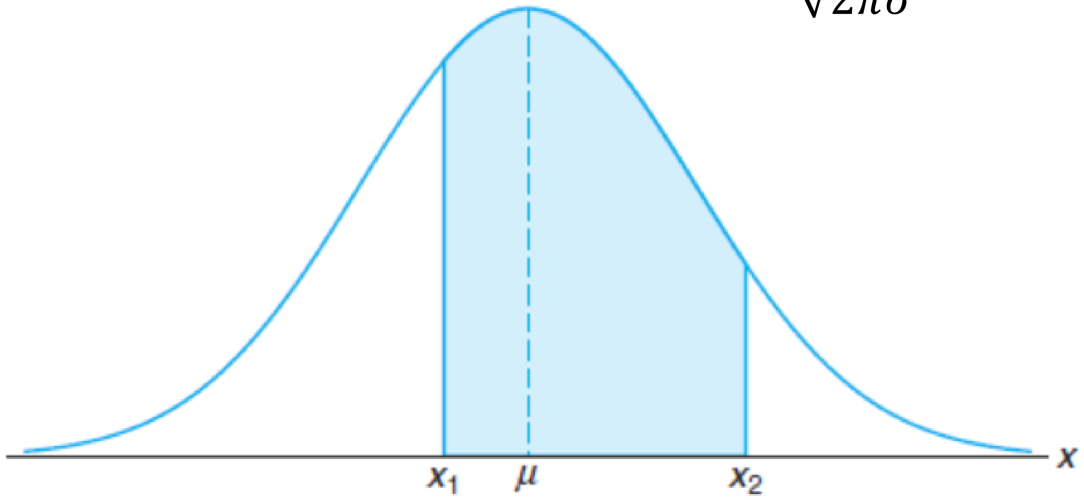
$$F(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad , \qquad -\infty < x < +\infty$$



Computing the probability values

Continuous Probability Distribution (Normal)

$$F(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad -\infty < x < +\infty$$

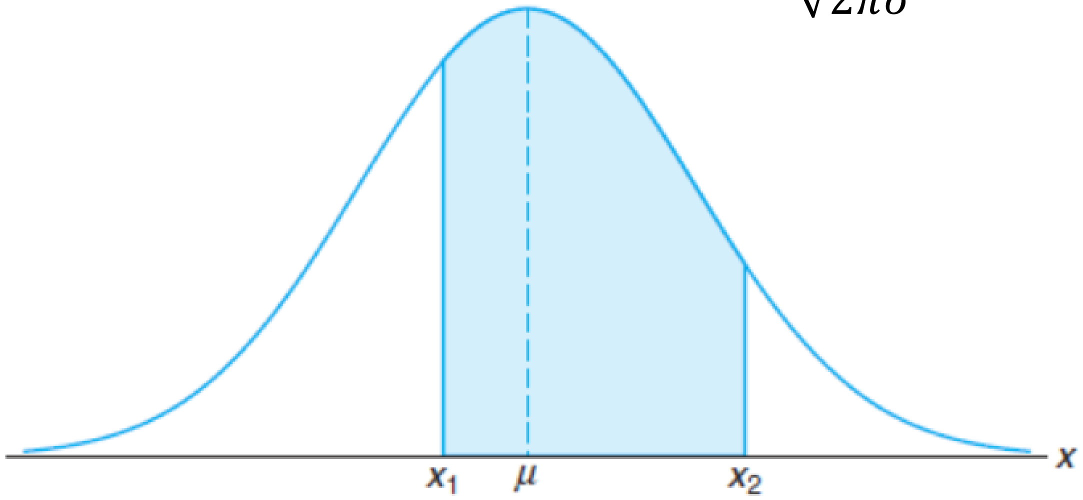


Computing the probability values

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi\sigma}} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$ …hard to solve

Continuous Probability Distribution (Normal)

$$F(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad -\infty < x < +\infty$$



Computing the probability values

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi\sigma}} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

…hard to solve

Z-score

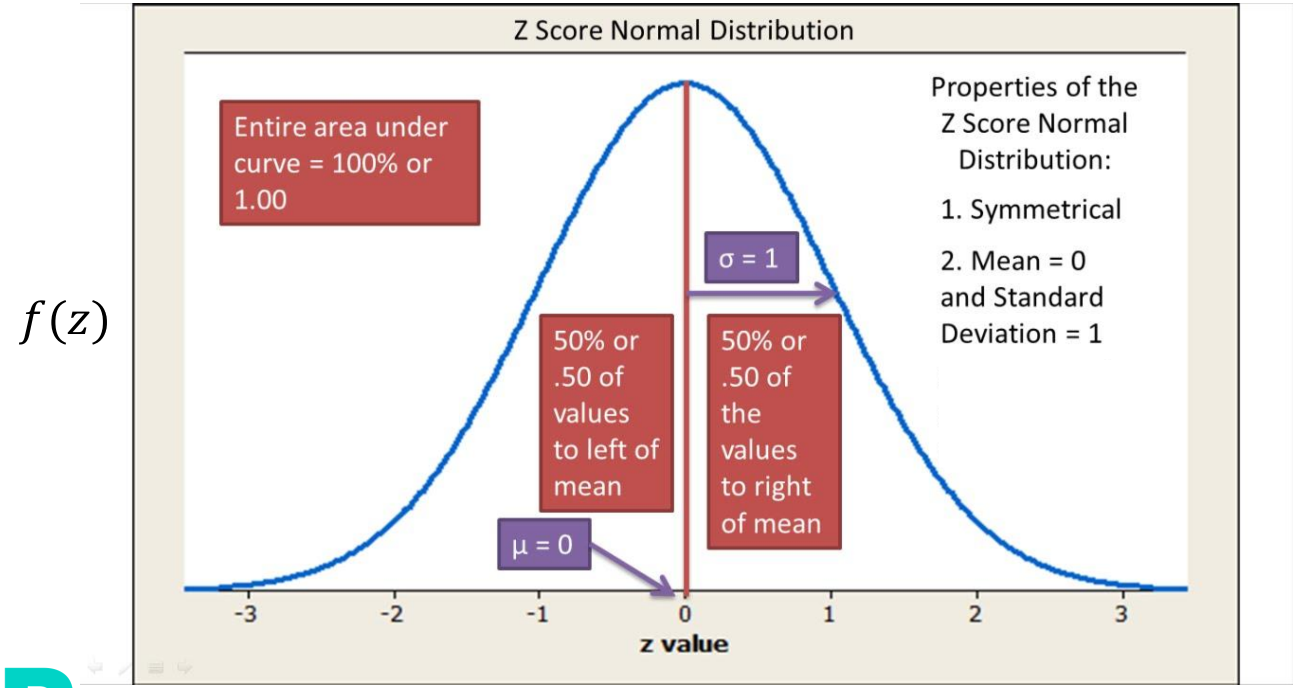
Z-score benefits…

- **Standardize** all the observations of any normal random variable *X* into a new set of observations;
- **Reduce the compl**exity of computing the probability;
- Make possible the **statistical comparison** between to random variables;
- The **tabulation** of Normal Distribution exists for Z-score only.

Z-score formula…

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

The distribution of a normal random variable with mean 0 and variance 1 is called a **standard normal distribution**.



Practicing calculations of probabilities using the Normal Distribution…

$N(\mu, \sigma^2)$ **practical examples…**

Empirical evidences show that certain supplier can provide an important medical device within a normal distributed delivery time (with $\mu = 12$ and $\sigma^2 = 4$, days and squared-days, respectively). For the firm that receives the devices, more 15 days of lead time would make almost impossible to serve their customers. The main question is: how likely is that delivery time overcomes 15 days?

Historical dataset provide sufficient evidence to assume our oxygenated water demand is normally distributed, with mean 300 Kgs and standard deviation of 25 Kgs. After a discussion with the financial department, we realize that for overcoming the breaking-even point our sales should be between 250 and 325 kilograms. How probable it is that our sales are between 250 and 325 kilograms?

$N(\mu, \sigma^2)$ **practical examples…**

Empirical evidences show that certain supplier can provide an important medical device within a normal distributed delivery time (with $\mu = 12$ and $\sigma^2 = 4$, days and squared-days, respectively). For the firm that receives the devices, more 15 days of lead time would make almost impossible to serve their customers. The main question is: how likely is that delivery time overcomes 15 days?

$$Z = \frac{x(15) - \mu(12)}{\sigma(\sqrt{4} = 2)} = \frac{15 - 12}{2} = \frac{3}{2} = 1.5$$

$N(\mu, \sigma^2)$

practical examples…

Empirical evidences show that certain supplier can provide an important medical device within a normal distributed delivery time (with $\mu = 12$ and $\sigma^2 = 4$, days and squared-days, respectively). For the firm that receives the devices, more 15 days of lead time would make almost impossible to serve their customers. The main question is: how likely is that delivery time overcomes 15 days?

$$Z = \frac{x(15) - \mu(12)}{\sigma(\sqrt{4} = 2)} = \frac{15 - 12}{2} = \frac{3}{2} = 1.5$$

Compute:

P (Z >= 1.5) = ?

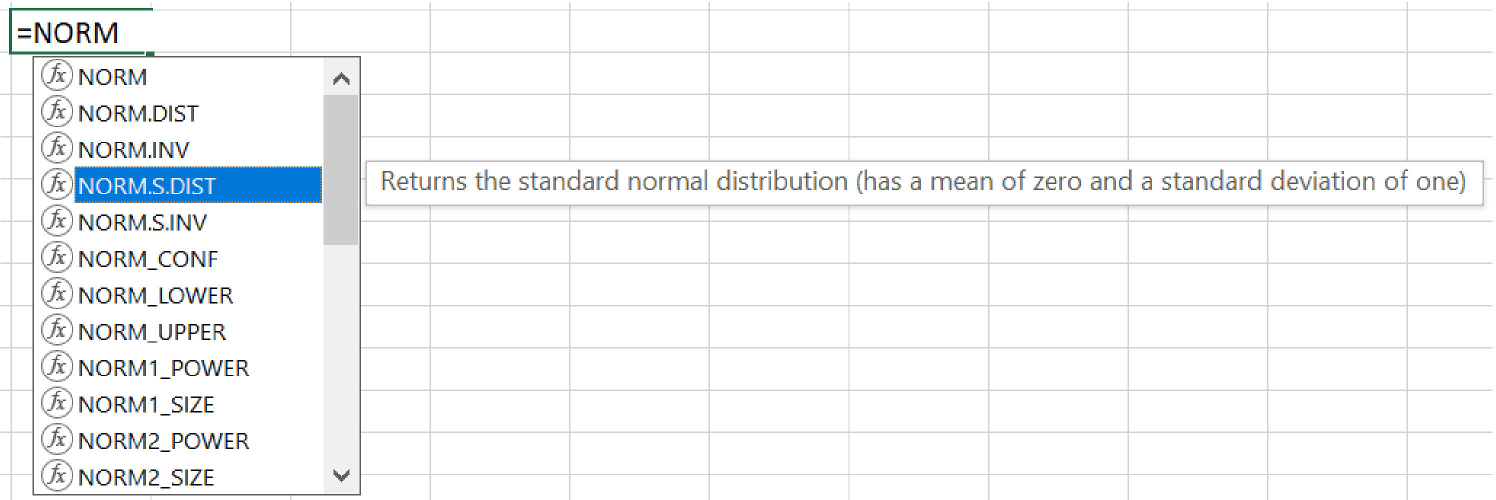
$N(\mu, \sigma^2)$ practical examples…

Empirical evidences show that certain supplier can provide an important medical device within a normal distributed delivery time (with $\mu = 12$ and $\sigma^2 = 4$, days and squared-days, respectively). For the firm that receives the devices, more 15 days of lead time would make almost impossible to serve their customers. The main question is: how likely is that delivery time overcomes 15 days?

Compute:

P (Z >= 1.5) = ?

Excel…



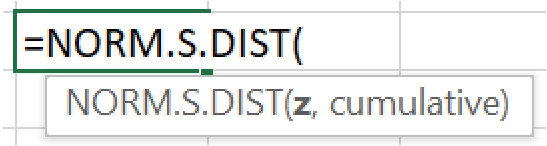
$N(\mu, \sigma^2)$ **practical examples…**

Empirical evidences show that certain supplier can provide an important medical device within a normal distributed delivery time (with $\mu = 12$ and $\sigma^2 = 4$, days and squared-days, respectively). For the firm that receives the devices, more 15 days of lead time would make almost impossible to serve their customers. The main question is: how likely is that delivery time overcomes 15 days?

Compute:

P (Z >= 1.5) = ?

Excel…



$N(\mu, \sigma^2)$ **practical examples…**

Empirical evidences show that certain supplier can provide an important medical device within a normal distributed delivery time (with $\mu = 12$ and $\sigma^2 = 4$, days and squared-days, respectively). For the firm that receives the devices, more 15 days of lead time would make almost impossible to serve their customers. The main question is: how likely is that delivery time overcomes 15 days?

Compute:

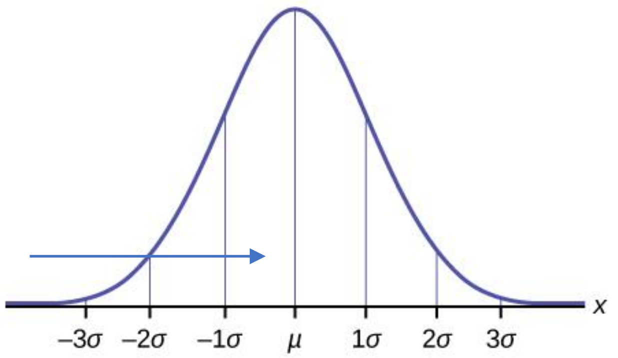
P (Z >= 1.5) = ?

Excel…

=NORM.S.DIST(

NORM.S.DIST(**z**, cumulative)

Cumulative = 1



The area below the curve from the left asymptote (bell) to the define z-value

$N(\mu, \sigma^2)$ practical examples…

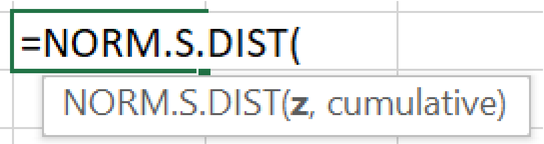
Empirical evidences show that certain supplier can provide an important medical device within a normal distributed delivery time (with $\mu = 12$ and $\sigma^2 = 4$, days and squared-days, respectively). For the firm that receives the devices, more 15 days of lead time would make almost impossible to serve their customers. The main question is: how likely is that delivery time overcomes 15 days?

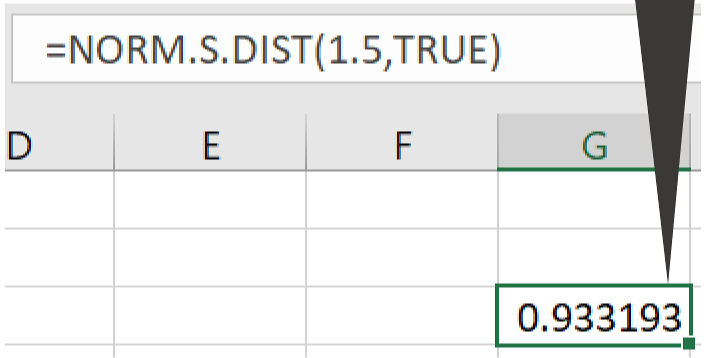
Cumulative = 1

Compute:

P (Z >= 1.5) = ?

Excel…





$N(\mu, \sigma^2)$ **practical examples…**

Empirical evidences show that certain supplier can provide an important medical device within a normal distributed delivery time (with $\mu = 12$ and $\sigma^2 = 4$, days and squared-days, respectively). For the firm that receives the devices, more 15 days of lead time would make almost impossible to serve their customers. The main question is: how likely is that delivery time overcomes 15 days?

Compute:

P (Z >= 1.5) = ?

Excel…

	0.933193
Prob:	=1-G3

Prob:	0.066807

Historical dataset provide sufficient evidence to assume our oxygenated water demand is normally distributed, with mean 300 Kgs and standard deviation of 25 Kgs. After a discussion with the financial department, we realize that for overcoming the breaking-even point our sales should be between 250 and 325 kilograms. How probable it is that our sales are between 250 and 325 kilograms?

Compute:

$P(250 \leq X \leq 325) = ?$

Historical dataset provide sufficient evidence to assume our oxygenated water demand is normally distributed, with mean 300 Kgs and standard deviation of 25 Kgs. After a discussion with the financial department, we realize that for overcoming the breaking-even point our sales should be between 250 and 325 kilograms. How probable it is that our sales are between 250 and 325 kilograms?

Compute:

P (-2 ≤ Z ≤ +1) = ?

$$Z_{250} = \frac{x(250) - \mu(300)}{\sigma(25)} = \frac{-50}{25} = -2$$

$$Z_{325} = \frac{x(325) - \mu(300)}{\sigma(25)} = \frac{25}{25} = +1$$

Historical dataset provide sufficient evidence to assume our oxygenated water demand is normally distributed, with mean 300 Kgs and standard deviation of 25 Kgs. After a discussion with the financial department, we realize that for overcoming the breaking-even point our sales should be between 250 and 325 kilograms. How probable it is that our sales are between 250 and 325 kilograms?

$$Z_{250} = \frac{x(250) - \mu(300)}{\sigma(25)} = \frac{-50}{25} = -2$$

$$Z_{325} = \frac{x(325) - \mu(300)}{\sigma(25)} = \frac{25}{25} = +1$$

Compute:

P (-2 ≤ Z ≤ +1) =

P (Z >= -2) – P (Z>= +1)

Historical dataset provide sufficient evidence to assume our oxygenated water demand is normally distributed, with mean 300 Kgs and standard deviation of 25 Kgs. After a discussion with the financial department, we realize that for overcoming the breaking-even point our sales should be between 250 and 325 kilograms. How probable it is that our sales are between 250 and 325 kilograms?

$$Z_{250} = \frac{x(250) - \mu(300)}{\sigma(25)} = \frac{-50}{25} = -2$$

$$Z_{325} = \frac{x(325) - \mu(300)}{\sigma(25)} = \frac{25}{25} = +1$$

Compute:

P (-2 ≤ Z ≤ +1) =

P (Z >= -2) – P (Z>= +1)

... 0.8186

=NORM.S.DIST(1,TRUE) − =NORM.S.DIST(-2,TRUE)

NORM.S.DIST(z, cumulative)

$N(\mu, \sigma^2)$ practical examples…

Empirical evidences show that certain supplier can provide an important medical device within a normal distributed delivery time (with $\mu = 12$ and $\sigma^2 = 4$, days and squared-days, respectively). For the [illegible] that receives the devices, more 15 days of lead time would make almost impossible to serve their customers. The main question is: how likely is that delivery time overcomes 15 days?

How do I know this?

Historical dataset provide sufficient evidence to assume our oxygenated water demand is normally distributed, with mean 300 Kgs and standard deviation of 25 Kgs. After a discussion with the financial department, we realize that for overcoming the breaking-even point our sales should be between 250 and 325 kilograms. How probable it is that our sales are between 250 and 325 kilograms?

The *goodness of fit test*…

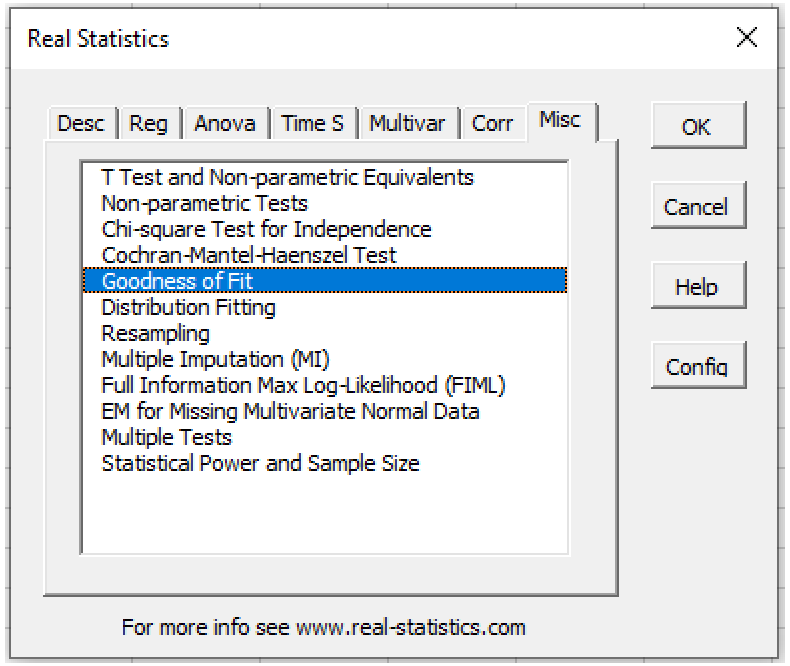
The *goodness of fit test* is used to *test* if sample data *fits* a distribution from a certain population (i.e. a population with a normal distribution or one with a Weibull distribution).

Professional software

Charles Zaiontz

Surgical gloves
12.8
11.51
19.32
18.4
14.34
17.2
18.78
12.69
16.09
14.06
14.64
14
17.88
15.49
18.61
11.48
12.58

…carry it out in Excel…

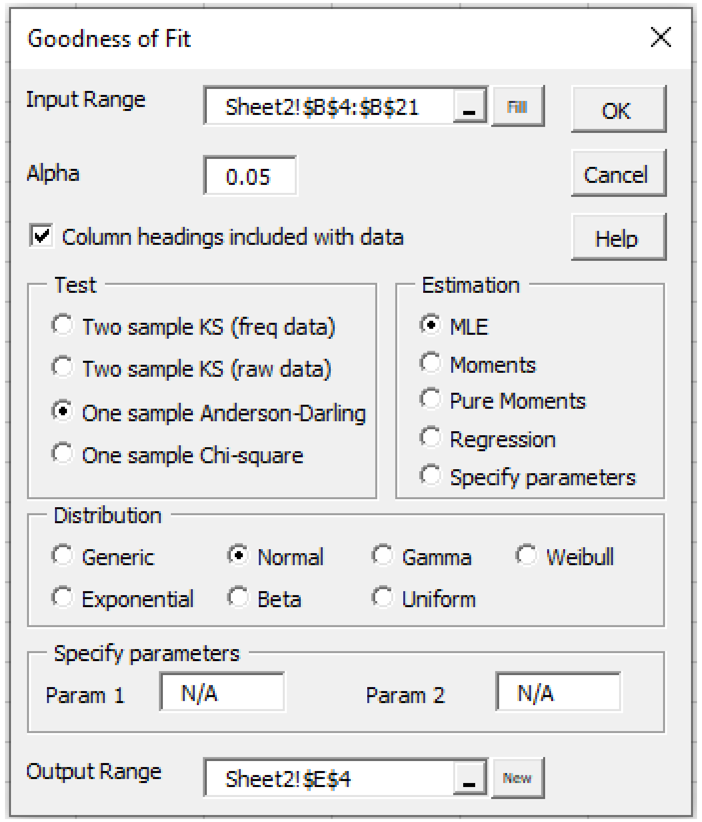


The *goodness of fit test* is used to *test* if sample data *fits* a distribution from a certain population (i.e. a population with a normal distribution or one with a Weibull distribution).

Professional software

Surgical gloves
12.8
11.51
19.32
18.4
14.34
17.2
18.78
12.69
16.09
14.06
14.64
14
17.88
15.49
18.61
11.48
12.58

...carry it out in Excel...



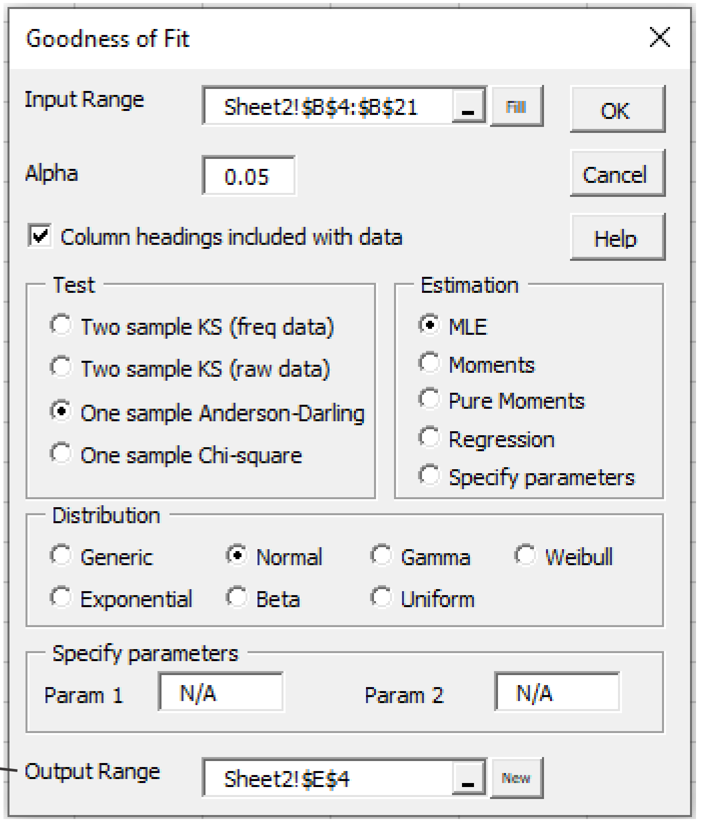
The *goodness of fit test* is used to *test* if sample data *fits* a distribution from a certain population (i.e. a population with a normal distribution or one with a Weibull distribution).

Professional software

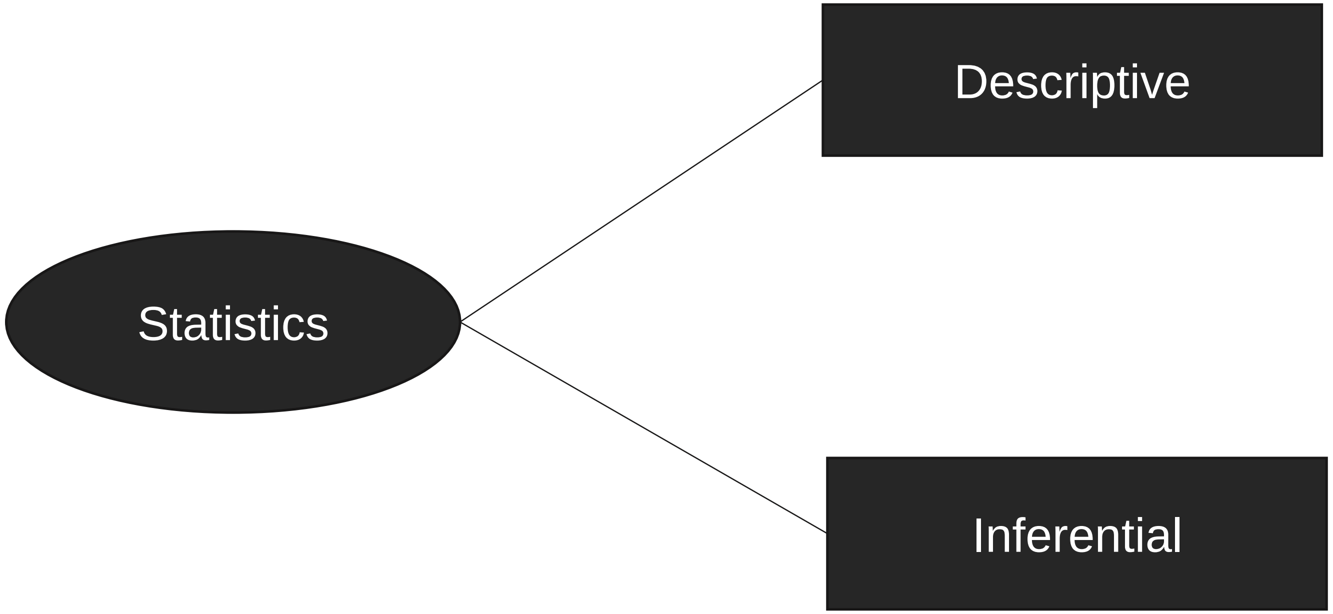
...carry it out in Excel...

Surgical gloves
12.8
11.51
19.32
18.4
14.34
17.2
18.78
12.69
16.09
14.06
14.64
14
17.88
15.49
18.61
11.48
12.58

Anderson-Darling Test			
Alpha	0.05	mean	15.28647
Distrib	Normal	std dev	2.586869
Method	MLE		
AD stat	0.509352		
p-value	0.19793		
crit value	0.713947		



Refreshing statistics…

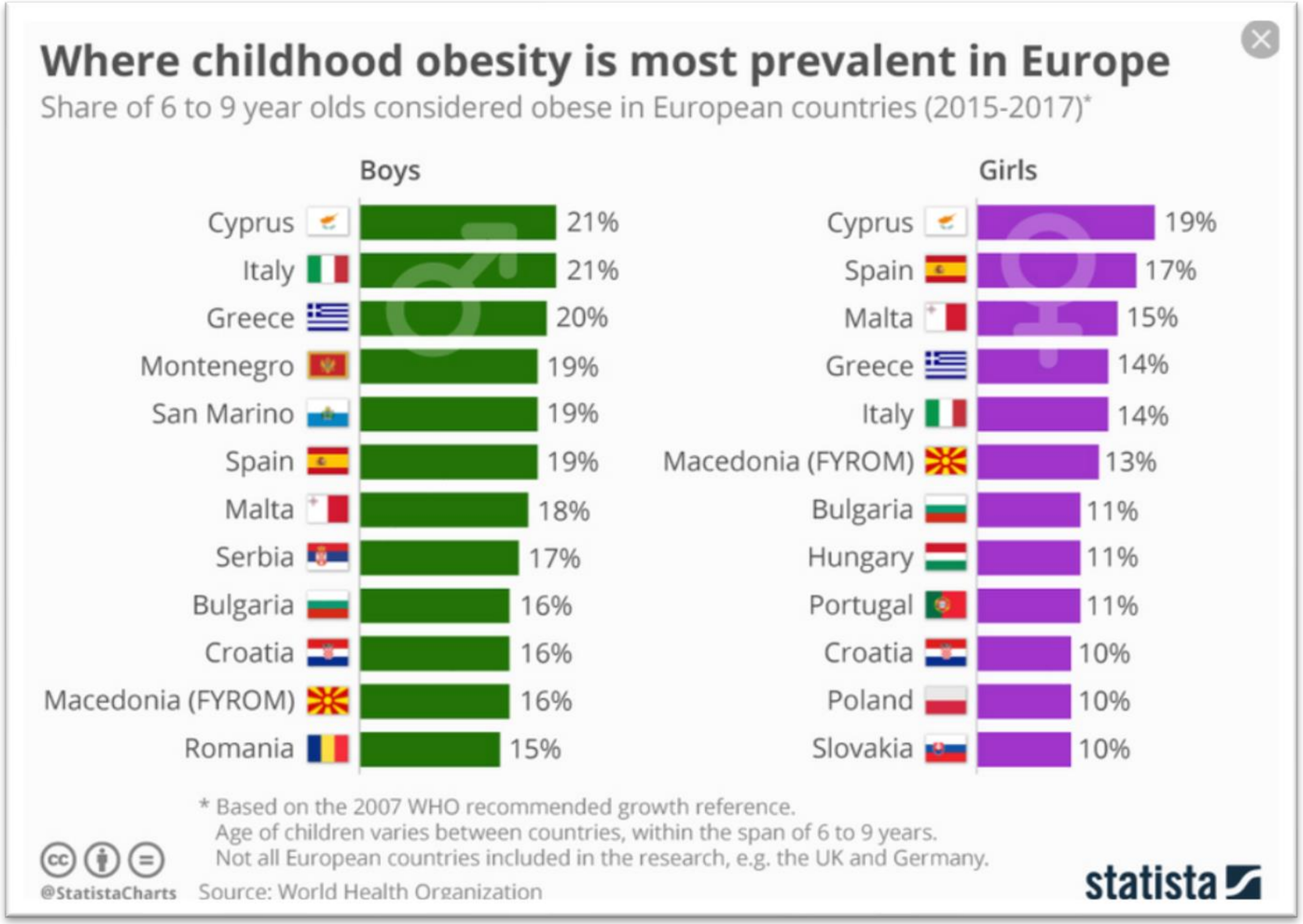

Statistics
Descriptive
Inferential

Descriptive

Descriptive statistics is the term given to the analysis of data that helps to describe, show or summarize data in a meaningful way such that, for example, patterns might emerge from the data.

Example of descriptive results

e.g.



Elements

Measures of central tendency: these are ways of describing the central position of a frequency distribution for a group of data

Mean (average, geometric, harmonic)

Median

Mode

Later in Excel

Elements

Measures of central tendency: these are ways of describing the central position of a frequency distribution for a group of data

Mean (average, geometric, harmonic)

Median

Mode

Later in Excel as well

working with samples…no compensation

$$G = \sqrt[n]{x_i \cdot x_{i+1} \cdot x_{i+2} \cdot \dots \cdot x_n}$$

Elements

Measures of central tendency: these are ways of describing the central position of a frequency distribution for a group of data

Mean (average, geometric, harmonic)

Median

Mode

Examine in details

$$H = \frac{N}{\sum_{i=1}^{n} 1/X_i}$$

working with samples…less important the positive extreme values

Using the measures of central tendency in forecasting…

Date	Surgical gloves		Mean(average)	Errors
Jan-17	12.8		16.74529412	=ABS(C3-E3
Feb-17	11.51		16.74529412	ABS(number)
Mar-17	19.32		16.74529412	
Apr-17	18.4		16.74529412	
May-17	14.34		16.74529412	
Jun-17	31		16.74529412	
Jul-17	18.78		16.74529412	
Aug-17	12.69		16.74529412	
Sep-17	16.09		16.74529412	
Oct-17	14.06		16.74529412	
Nov-17	14.64		16.74529412	
Dec-17	25		16.74529412	
Jan-18	17.88		16.74529412	
Feb-18	15.49		16.74529412	
Mar-18	18.61		16.74529412	
Apr-18	11.48		16.74529412	
May-18	12.58		16.74529412	

Using the measures of central tendency in forecasting…

Date	Surgical gloves	Mean(average)	Errors	Geometric-Mean	Errors
Jan-17	12.8	16.74529412	3.945294	=GEOMEAN(C3:C19)	
Feb-17	11.51	16.74529412	5.235294	GEOMEAN(number1, [number2], ...)	
Mar-17	19.32	16.74529412	2.574706	16.13918802	
Apr-17	18.4	16.74529412	1.654706	16.13918802	
May-17	14.34	16.74529412	2.405294	16.13918802	
Jun-17	31	16.74529412	14.25471	16.13918802	
Jul-17	18.78	16.74529412	2.034706	16.13918802	
Aug-17	12.69	16.74529412	4.055294	16.13918802	
Sep-17	16.09	16.74529412	0.655294	16.13918802	
Oct-17	14.06	16.74529412	2.685294	16.13918802	
Nov-17	14.64	16.74529412	2.105294	16.13918802	
Dec-17	25	16.74529412	8.254706	16.13918802	
Jan-18	17.88	16.74529412	1.134706	16.13918802	
Feb-18	15.49	16.74529412	1.255294	16.13918802	
Mar-18	18.61	16.74529412	1.864706	16.13918802	
Apr-18	11.48	16.74529412	5.265294	16.13918802	
May-18	12.58	16.74529412	4.165294	16.13918802	

$$G = \sqrt[n]{x_i \cdot x_{i+1} \cdot x_{i+2} \cdot \cdots \cdot x_n}$$

Using the measures of central tendency in forecasting…

$$H = \frac{N}{\sum_{i=1}^{n} {}^{1}/_{X_i}}$$

Date	Surgical gloves	Mean(average)	Errors	Geometric-Mean	Errors	Harmonic-Mean	Errors
Jan-17	12.8	16.74529412	3.945294	16.13918802	3.339188	=HARMEAN(C3:C19)	
Feb-17	11.51	16.74529412	5.235294	16.13918802	4.629188	HARMEAN(**number1**, [number2], ...)	
Mar-17	19.32	16.74529412	2.574706	16.13918802	3.180812	15.6313613	
Apr-17	18.4	16.74529412	1.654706	16.13918802	2.260812	15.6313613	
May-17	14.34	16.74529412	2.405294	16.13918802	1.799188	15.6313613	
Jun-17	31	16.74529412	14.25471	16.13918802	14.86081	15.6313613	
Jul-17	18.78	16.74529412	2.034706	16.13918802	2.640812	15.6313613	
Aug-17	12.69	16.74529412	4.055294	16.13918802	3.449188	15.6313613	
Sep-17	16.09	16.74529412	0.655294	16.13918802	0.049188	15.6313613	
Oct-17	14.06	16.74529412	2.685294	16.13918802	2.079188	15.6313613	
Nov-17	14.64	16.74529412	2.105294	16.13918802	1.499188	15.6313613	
Dec-17	25	16.74529412	8.254706	16.13918802	8.860812	15.6313613	
Jan-18	17.88	16.74529412	1.134706	16.13918802	1.740812	15.6313613	
Feb-18	15.49	16.74529412	1.255294	16.13918802	0.649188	15.6313613	
Mar-18	18.61	16.74529412	1.864706	16.13918802	2.470812	15.6313613	
Apr-18	11.48	16.74529412	5.265294	16.13918802	4.659188	15.6313613	
May-18	12.58	16.74529412	4.165294	16.13918802	3.559188	15.6313613	

Using the measures of central tendency in forecasting…

Date	Surgical gloves		Mean(average)	Errors		Geometric-Mean	Errors		Harmonic-Mean	Errors
Jan-17	12.8		16.74529412	3.945294		16.13918802	3.339188		15.6313613	2.831361
Feb-17	11.51		16.74529412	5.235294		16.13918802	4.629188		15.6313613	4.121361
Mar-17	19.32		16.74529412	2.574706		16.13918802	3.180812		15.6313613	3.688639
Apr-17	18.4		16.74529412	1.654706		16.13918802	2.260812		15.6313613	2.768639
May-17	14.34		16.74529412	2.405294		16.13918802	1.799188		15.6313613	1.291361
Jun-17	31		16.74529412	14.25471		16.13918802	14.86081		15.6313613	15.36864
Jul-17	18.78		16.74529412	2.034706		16.13918802	2.640812		15.6313613	3.148639
Aug-17	12.69		16.74529412	4.055294		16.13918802	3.449188		15.6313613	2.941361
Sep-17	16.09		16.74529412	0.655294		16.13918802	0.049188		15.6313613	0.458639
Oct-17	14.06		16.74529412	2.685294		16.13918802	2.079188		15.6313613	1.571361
Nov-17	14.64		16.74529412	2.105294		16.13918802	1.499188		15.6313613	0.991361
Dec-17	25		16.74529412	8.254706		16.13918802	8.860812		15.6313613	9.368639
Jan-18	17.88		16.74529412	1.134706		16.13918802	1.740812		15.6313613	2.248639
Feb-18	15.49		16.74529412	1.255294		16.13918802	0.649188		15.6313613	0.141361
Mar-18	18.61		16.74529412	1.864706		16.13918802	2.470812		15.6313613	2.978639
Apr-18	11.48		16.74529412	5.265294		16.13918802	4.659188		15.6313613	4.151361
May-18	12.58		16.74529412	4.165294		16.13918802	3.559188		15.6313613	3.051361
			Average-all-errors:	3.737993			3.631033			3.595374

Small errors in predictions…

Outliers

Elements

Measures of spread: these are ways of summarizing a group of data by describing how spread out the scores are.

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Standard deviation

Variance

Range

CV

Later in Excel…and Real Stat Add-Ins

The coefficient of variation (CV) represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from each other.

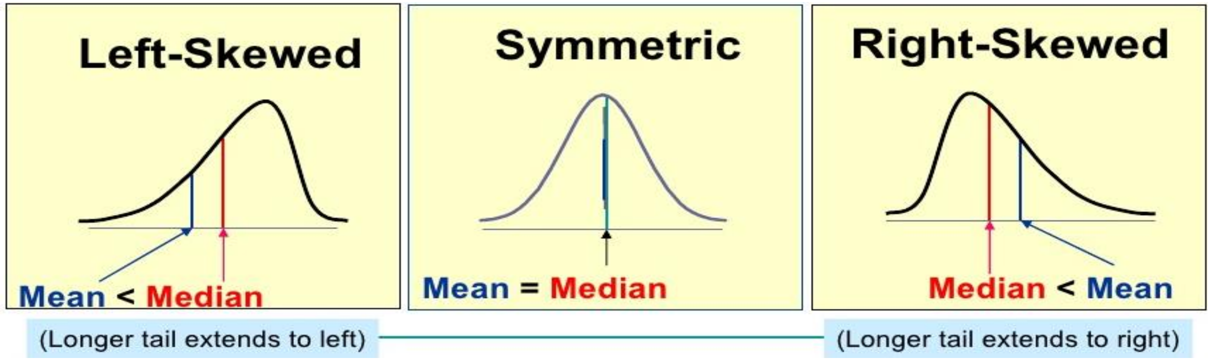
$$CV = \frac{S}{\bar{\bar{X}}}$$

Useful in risk analysis…

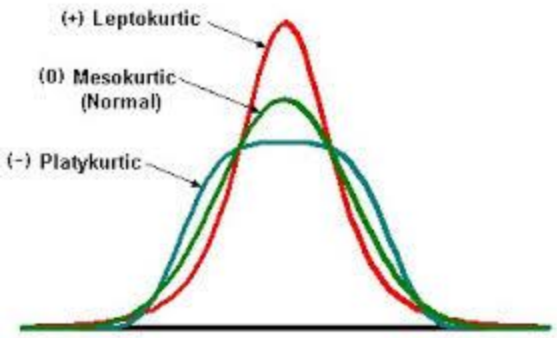
Supplier 1(Kgs) *vs* Supplier 2 (units)

Measures of Distribution

Skewness: distribution (aggregations of observations) can be spread around both sides of the central tendency.

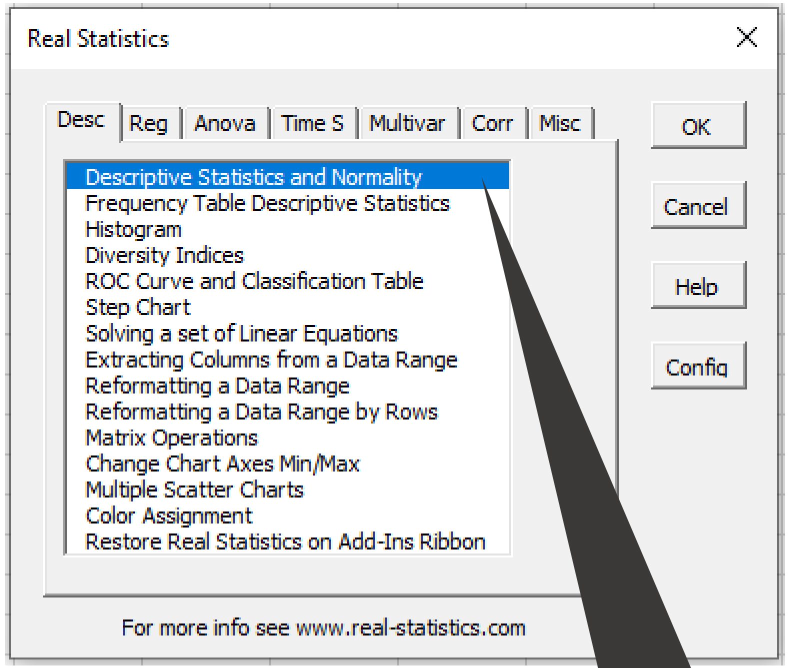


Kurtosis: is the measure of the peak of a distribution, and indicates how high is around the mean.



Week	Patients arrivals
1	101
2	112
3	89
4	105
5	92
6	81
7	105
8	104
9	138
10	109
11	97
12	110
13	115
14	127
15	107
16	90
17	86
18	110
19	99
20	75
21	112
22	98
23	91
24	64
25	98
26	113
27	114
28	151
29	102
30	109
31	114
32	82
33	94
34	93
35	90
36	98
37	82
38	110
39	121
40	107
41	110
42	102
43	119
44	111
45	105
46	103
47	113
48	86
49	107
50	91
51	90
52	81

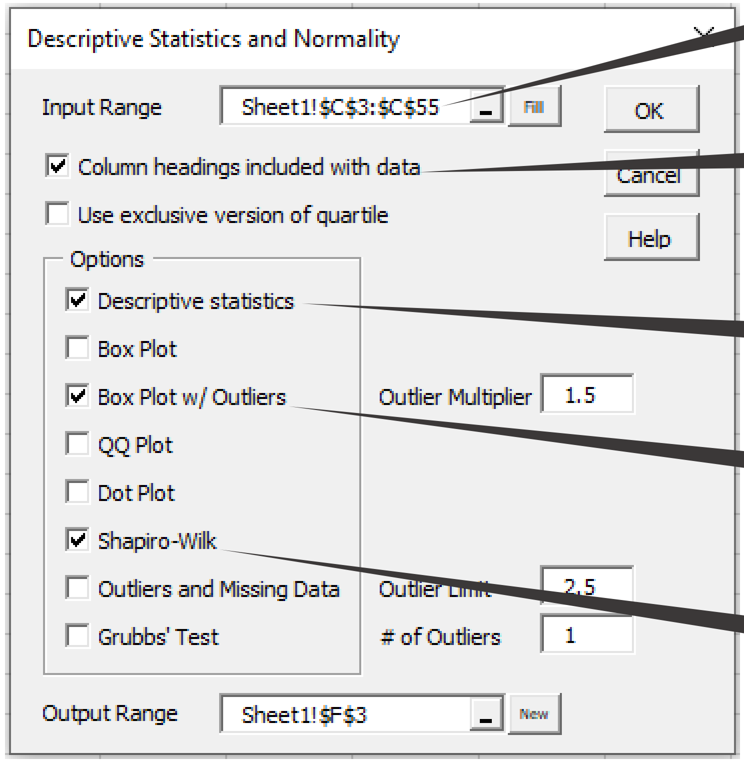
Analyzing the weekly patient arrivals…



Comprehensive module for descriptive statistics

Week	Patients arrivals
1	101
2	112
3	89
4	105
5	92
6	81
7	105
8	104
9	138
10	109
11	97
12	110
13	115
14	127
15	107
16	90
17	86
18	110
19	99
20	75
21	112
22	98
23	91
24	64
25	98
26	113
27	114
28	151
29	102
30	109
31	114
32	82
33	94
34	93
35	90
36	98
37	82
38	110
39	121
40	107
41	110
42	102
43	119
44	111
45	105
46	103
47	113
48	86
49	107
50	91
51	90
52	81

Analyzing the weekly patient arrivals…



Data array…

Include the cell “patient arrivals”

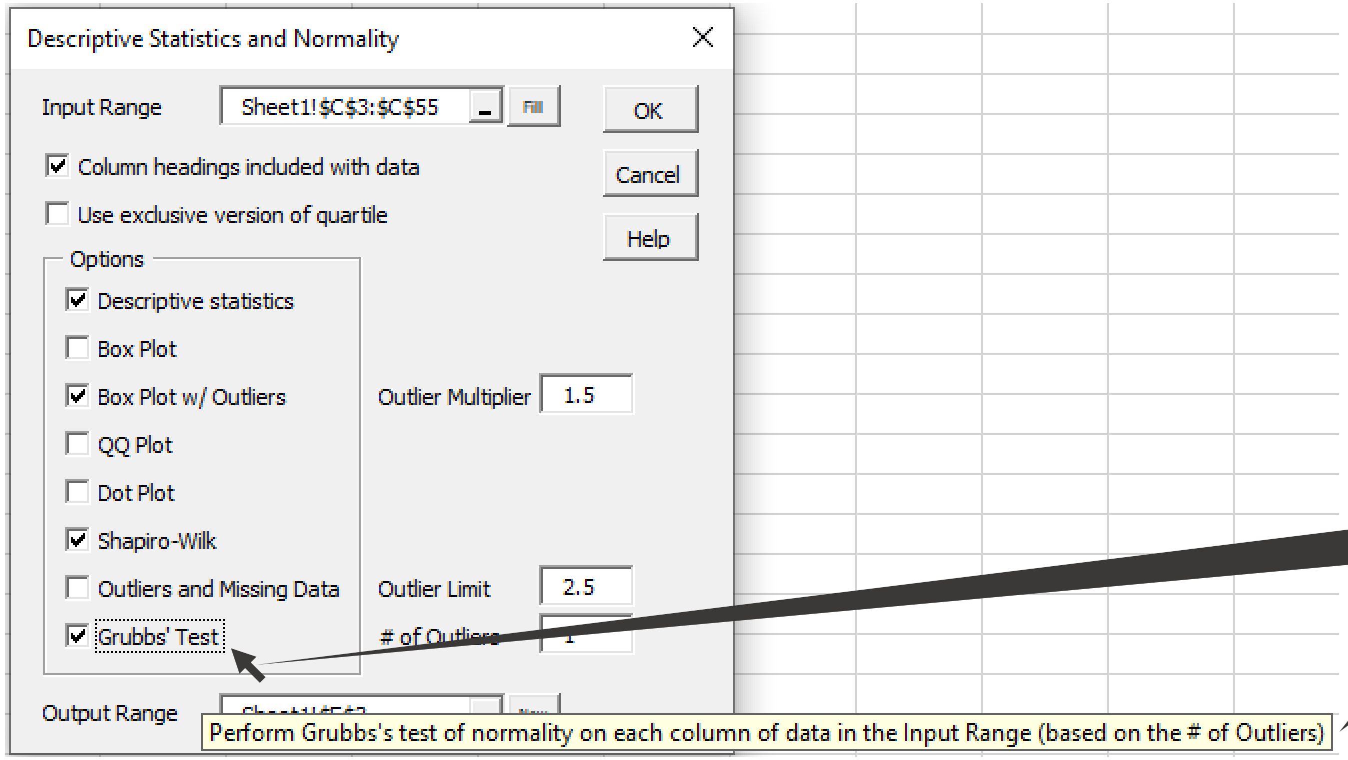
All descriptive statistics…

Box Plot depicting the outliers

Another Normality-test…

Week	Patients arrivals
1	101
2	112
3	89
4	105
5	92
6	81
7	105
8	104
9	138
10	109
11	97
12	110
13	115
14	127
15	107
16	90
17	86
18	110
19	99
20	75
21	112
22	98
23	91
24	64
25	98
26	113
27	114
28	151
29	102
30	109
31	114
32	82
33	94
34	93
35	90
36	98
37	82
38	110
39	121
40	107
41	110
42	102
43	119
44	111
45	105
46	103
47	113
48	86
49	107
50	91
51	90
52	81

Analyzing the weekly patient arrivals…



Use the pointer…searching choice´s meaning…

Analyzing the weekly patient arrivals…

Descriptive Statistics	
	Patients arrivals
Mean	102.1730769
Standard Error	2.122692368
Median	103.5
Mode	110
Standard Deviation	15.30695235
Sample Variance	234.3027903
Kurtosis	1.534557486
Skewness	0.397113782
Range	87
Maximum	151
Minimum	64
Sum	5313
Count	52
Geometric Mean	101.0481308
Harmonic Mean	99.90476696
AAD	11.62795858
MAD	9.5
IQR	19.25

102 patients arrive…on average

$$SE = \frac{S}{\sqrt{n}}$$

↓(data better distributed)

The most repeated value

A little high…

Symmetric respect to the mean

Analyzing the weekly patient arrivals…

Descriptive Statistics	
	Patients arrivals
Mean	102.1730769
Standard Error	2.122692368
Median	103.5
Mode	110
Standard Deviation	15.30695235
Sample Variance	234.3027903
Kurtosis	1.534557486
Skewness	0.397113782
Range	87
Maximum	151
Minimum	64
Sum	5313
Count	52
Geometric Mean	101.0481308
Harmonic Mean	99.90476696
AAD	11.62795858
MAD	9.5
IQR	19.25

Average of the Absolute Deviation…

$$AAD = \frac{1}{n}\sum|x_i - \bar{x}|$$

Analyzing the weekly patient arrivals…

Descriptive Statistics	
	Patients arrivals
Mean	102.1730769
Standard Error	2.122692368
Median	103.5
Mode	110
Standard Deviation	15.30695235
Sample Variance	234.3027903
Kurtosis	1.534557486
Skewness	0.397113782
Range	87
Maximum	151
Minimum	64
Sum	5313
Count	52
Geometric Mean	101.0481308
Harmonic Mean	99.90476696
AAD	11.62795858
MAD	9.5
IQR	19.25

Median Absolute Deviation…

Median $\{|x_i - \tilde{x}| : x_i \text{ in } S\}$

where $\tilde{x}$ = median of the data elements in S.

HELP

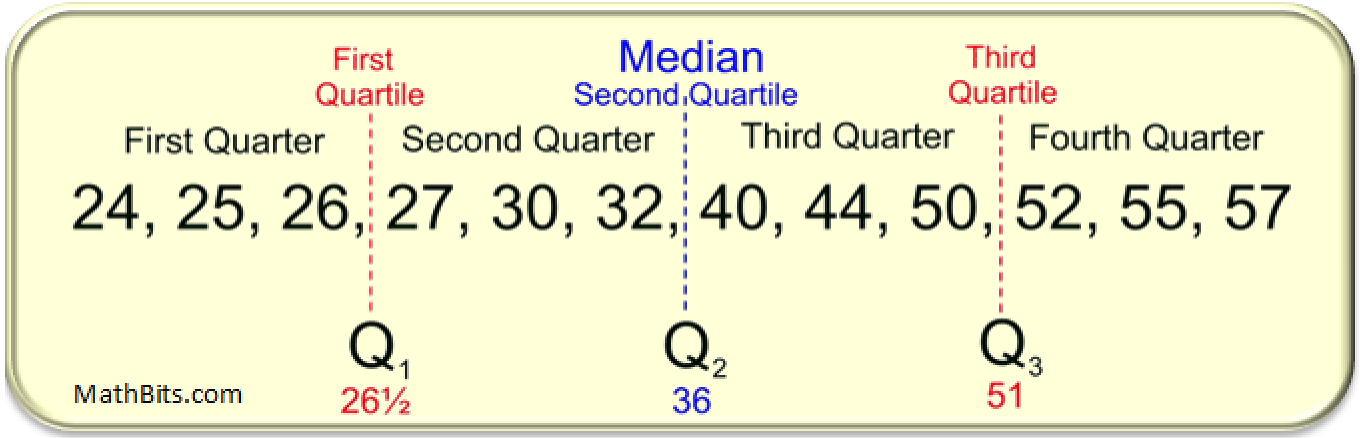
Analyzing the weekly patient arrivals…

Descriptive Statistics	
	Patients arrivals
Mean	102.1730769
Standard Error	2.122692368
Median	103.5
Mode	110
Standard Deviation	15.30695235
Sample Variance	234.3027903
Kurtosis	1.534557486
Skewness	0.397113782
Range	87
Maximum	151
Minimum	64
Sum	5313
Count	52
Geometric Mean	101.0481308
Harmonic Mean	99.90476696
AAD	11.62795858
MAD	9.5
IQR	19.25

Inter-quartile Range…

Descriptive Stats

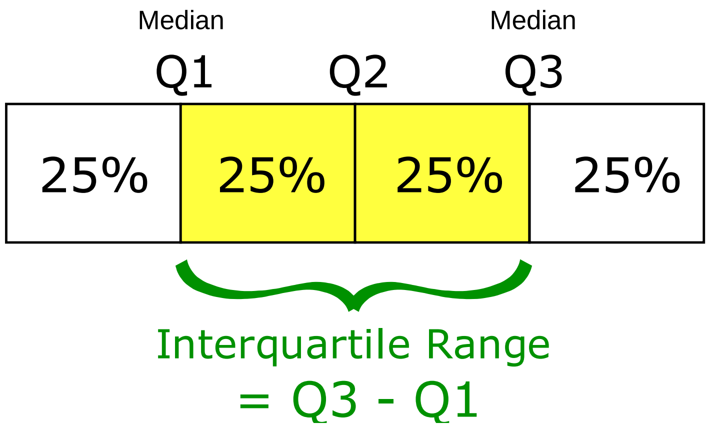
Inter-quartile Range…

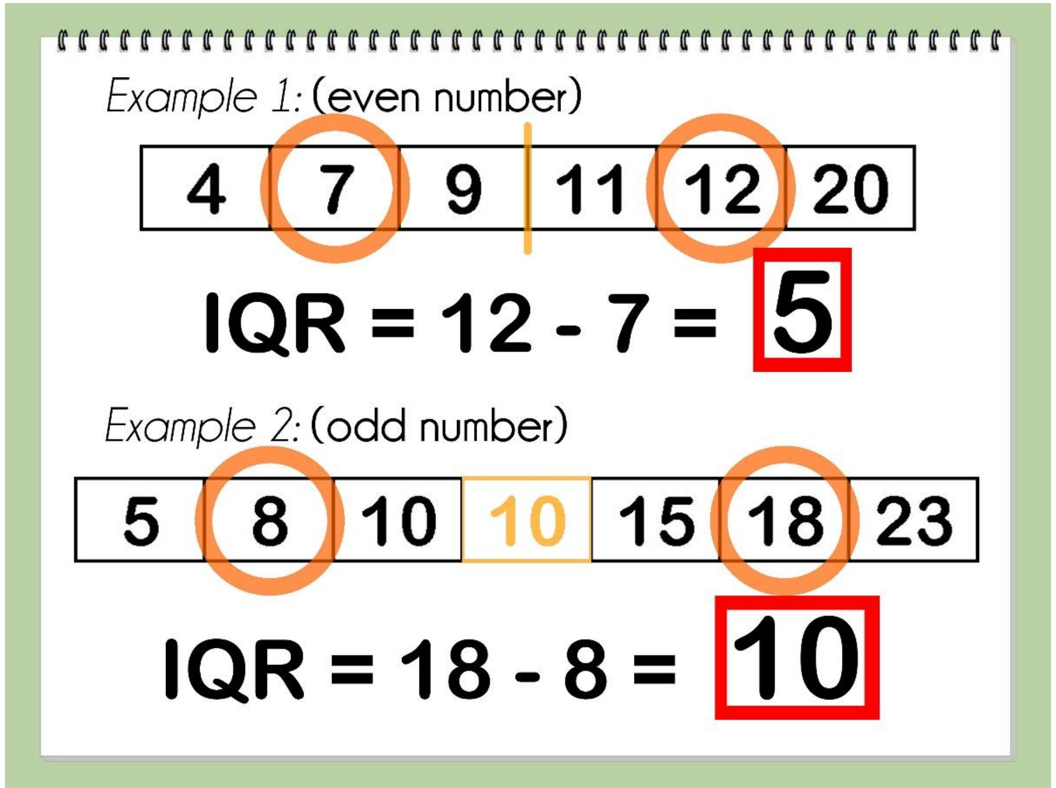


Sorting data

Descriptive Stats

Inter-quartile Range…



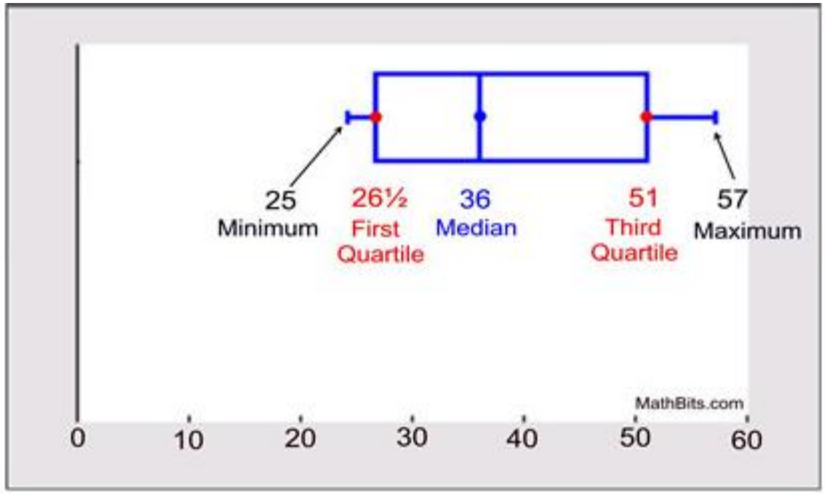


https://www.pinterest.com/pin/203928689353407866/

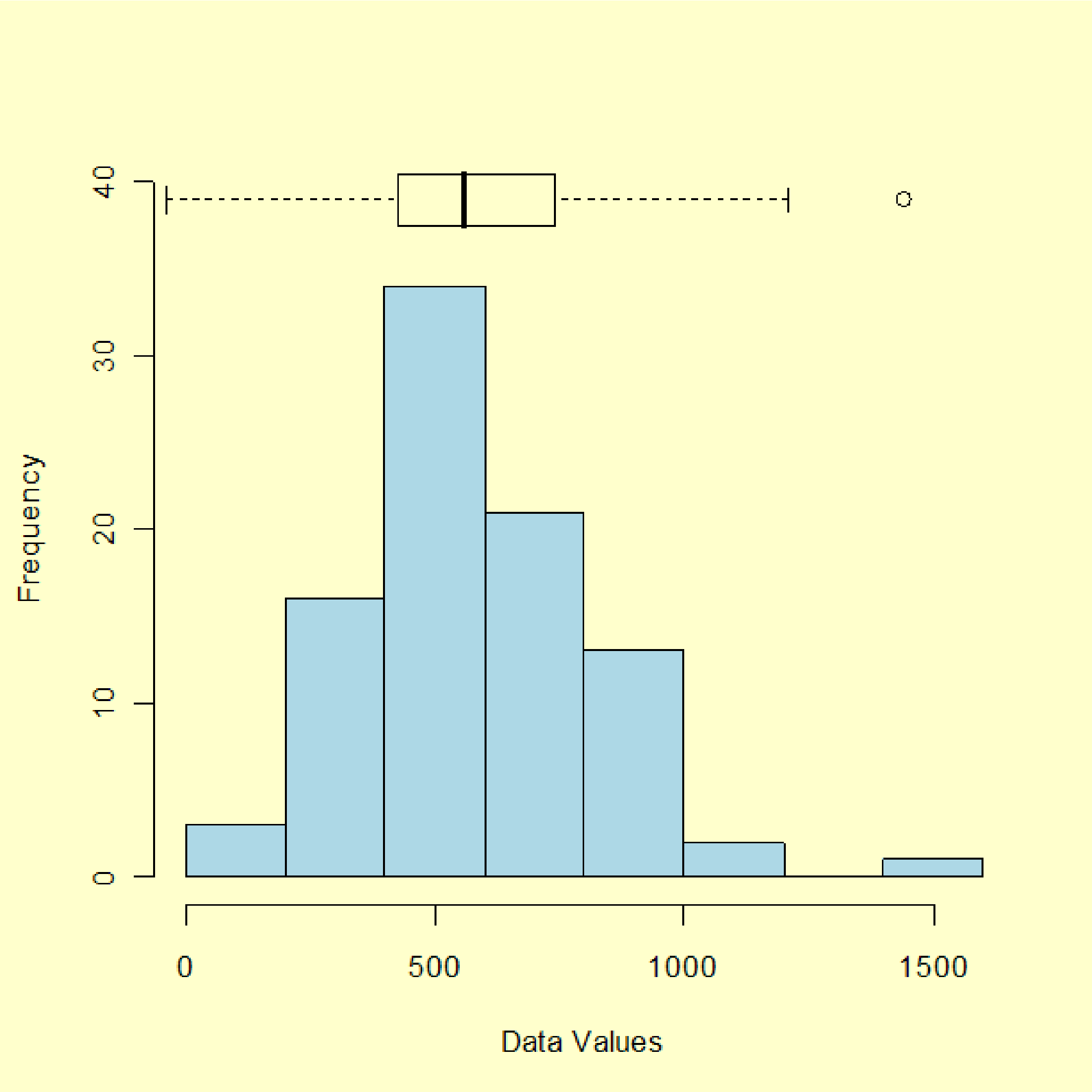
Descriptive Stats

Outliers are:
greater than $Q_3 + (1.5 \cdot IQR)$
(referred to as the upper fence)
or less than $Q_1 - (1.5 \cdot IQR)$
(referred to as the lower fence)

Box plot

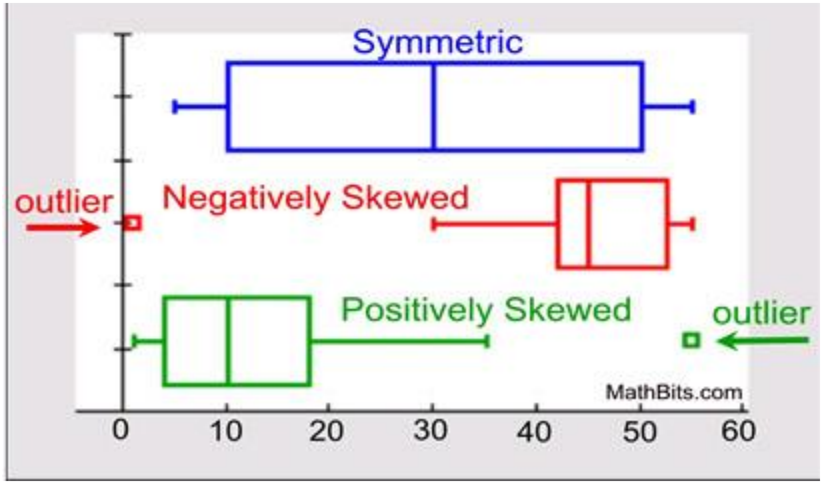


Descriptive Stats

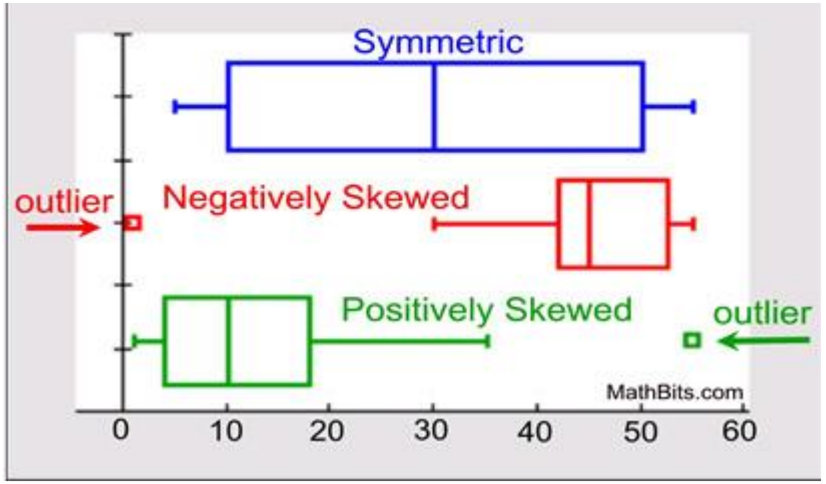


https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm

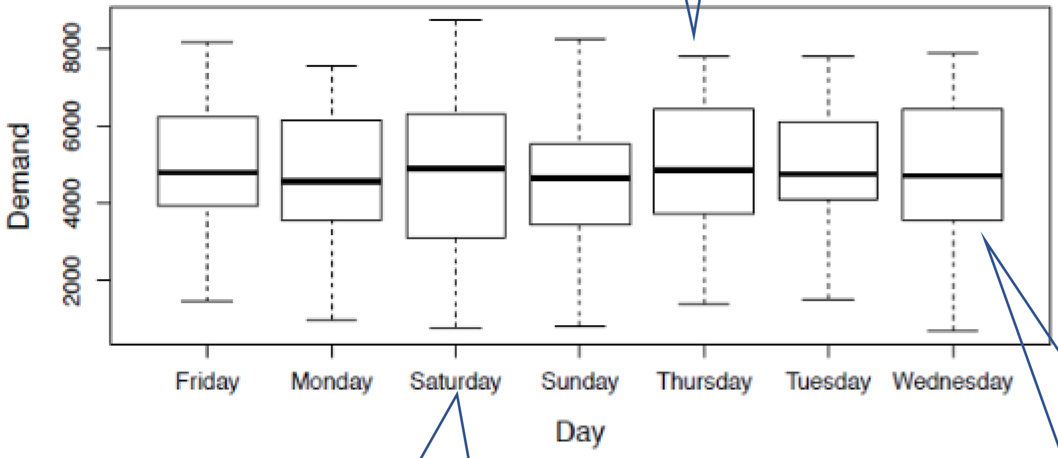
Descriptive Stats



Descriptive Stats



Very stable demand…

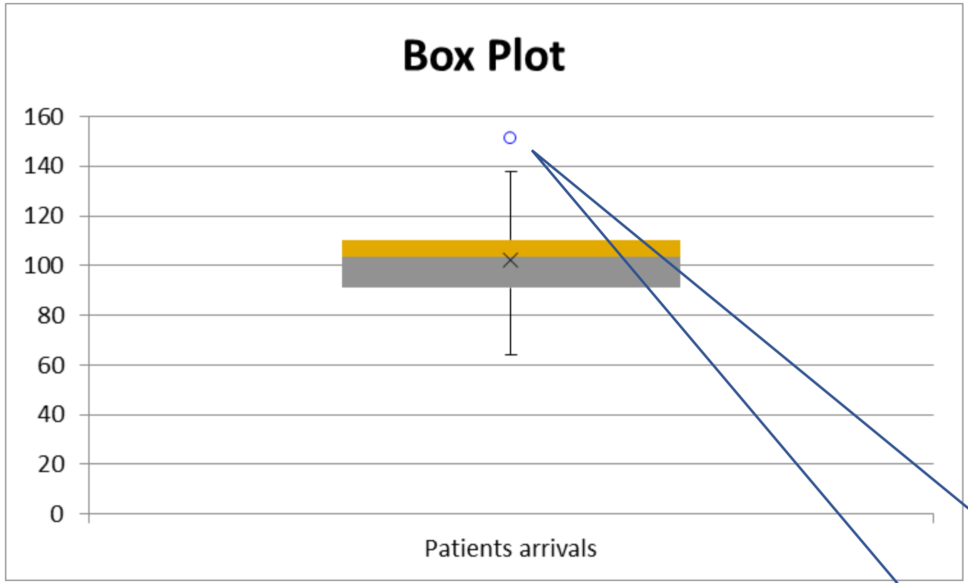


Somehow bias…

Somehow bias…

Analyzing the weekly patient arrivals…

Multiplier	1.5
	Patients arrivals
Min	64
Q1-Min	27
Med-Q1	12.5
Q3-Med	6.75
Max-Q3	27.75
Mean	102.1731
Min	64
Q1	91
Median	103.5
Q3	110.25
Max	138
Mean	102.1731
Grand Min	0
Outliers	151



One outlier…the data could be removed from the dataset

Analyzing the weekly patient arrivals…

Shapiro-Wilk Test	
	Patients arrivals
W-stat	0.96806
p-value	0.174667
alpha	0.05
normal	yes

d'Agostino-Pearson	
DA-stat	5.134768
p-value	0.076736
alpha	0.05
normal	yes

There is normality

1) Define the hypothesis
2) Identify the proper statistical test
3) Compute the p-value
4) Compare p-value against an "acceptable significance value(α)"…then make a decision…

If p-value ≤ α Then
the null hypothesis is ruled out, and the alternative hypothesis is valid.
Else
The null hypothesis is valid

H0: The data follow a Normal Distribution
H1: The data do not follow a Normal Distribution

HELP

Descriptive-Stats in R…