

# **MAT 7003 : Mathematical Foundations**

**(for Software Engineering)**

**J Paul Gibson, A207**

`paul.gibson@it-sudparis.eu`

<http://www-public.it-sudparis.eu/~gibson/Teaching/MAT7003/>

## **Probability and Statistics**

<http://www-public.it-sudparis.eu/~gibson/Teaching/MAT7003/L8-ProbabilityAndStatistics.pdf>

### **QUESTION:**

What do you know about –

- Probability?
- Statistics ?
- The relationship between them?

## Problem 1:

There are 3 boxes in which I place (without you seeing) a prize.

You pick one of the boxes (your goal is to end up with the box containing the prize)

I then open one of the other two boxes and show you that it is empty.

I then offer you the chance to switch boxes (without looking in the one in front of me or the one in front of you)

Should you swap boxes, if you wish to maximise your chances of winning the prize?

## Problem 2:

A man has two children.

One of them is a boy.

What's the probability that the other one is a boy?

### Problem 3:

There are two players or teams. Each has two cards, one marked 'Defect', the other 'Co-operate'. There is a neutral banker, who pays out or collects payments depending on the two cards played. Each player or team decides on a single card to play and gives it to the banker. The banker then reveals both cards.

Here's the scoring system:

Both play the 'Co-operate' card - Banker pays each £300.

Both play the 'Defect' card - Banker collects £10

One of each card - Banker pays 'Defect' £500, but collects £100 from 'Co-operate'.

Question: What is best strategy to winning most money?

### Problem 4:

The three dice game:

Player 1 throws 2 6-sided dice and adds them to get their score

Player 2 throws 1 6-sided dice and multiplies the answer by 2 to get their score

If both scores are greater than 10 then the match is a draw

If both scores are the same then the match is a draw

Otherwise the highest scoring total wins

Question: Who has the best chance to win – player 1 or player 2?

## **Problem 5: Is the dice loaded?**

I roll a 6-sided dice 20 times and I never roll a 6.

Do you think the dice is fair?

Should you bet on the next roll being a 6?

## **Probability Theory:**

### **Approaches of Assigning Probabilities:**

There are three approaches of assigning probabilities, as follows:

#### **1. Classical Approach:**

Classical probability is predicated on the assumption that the outcomes of an experiment are equally likely to happen.

$P(X) = \text{Number of favorable outcomes} / \text{Total number of possible outcomes}$

Note that we can apply the classical probability when the events have the same chance of occurring (called equally likely events), and the set of events are mutually exclusive and collectively exhaustive.

## Probability Theory:

### 2. Relative Frequency Approach:

Relative probability is based on cumulated historical data. The following equation is used to assign this type of probability:

$P(X) = \text{Number of times an event occurred in the past} / \text{Total number of opportunities for the event to occur}$

Note that relative probability is not based on rules or laws but on what has happened in the past. For example, your company wants to decide on the probability that its inspectors are going to reject the next batch of raw materials from a supplier. Data collected from your company record books show that the supplier had sent your company 80 batches in the past, and inspectors had rejected 15 of them. By the method of relative probability, the probability of the inspectors rejecting the next batch is  $15/80$ , or 0.19. If the next batch is rejected, the relative probability for the subsequent shipment would change to  $16/81 = 0.20$ .

## Probability Theory:

### 3. Subjective Approach:

The subjective probability is based on personal judgment, accumulation of knowledge, and experience. For example, medical doctors sometimes assign subjective probabilities to the length of life expectancy for people having cancer.

## Probability Theory:

**Discrete probability theory** deals with events that occur in **countable** sample spaces.

Examples: Throwing **dice**, experiments with **decks of cards**, and **random walk**.

**Classical definition:** Initially the probability of an event to occur was defined as number of cases favorable for the event, over the number of total outcomes possible in an equiprobable sample space.

For example, if the event is "occurrence of an even number when a **die** is rolled", the probability is given by  $\frac{3}{6} = \frac{1}{2}$ , since 3 faces out of the 6 have even numbers and each face has the same probability of appearing.

## Probability Theory:

The modern definition starts with a **set** called the **sample space**, which relates to the set of all *possible outcomes* in classical sense, denoted by  $\Omega = \{x_1, x_2, \dots\}$ . It is then assumed that for each element  $x \in \Omega$ , an intrinsic "probability" value  $f(x)$  is attached, which satisfies the following properties:

1.  $f(x) \in [0, 1]$  for all  $x \in \Omega$  ;
2.  $\sum_{x \in \Omega} f(x) = 1$  .

That is, the probability function  $f(x)$  lies between zero and one for every value of  $x$  in the sample space  $\Omega$ , and the sum of  $f(x)$  over all values  $x$  in the sample space  $\Omega$  is equal to 1. An **event** is defined as any **subset**  $E$  of the sample space  $\Omega$ . The **probability** of the event  $E$  is defined as

$$P(E) = \sum_{x \in E} f(x) .$$

So, the probability of the entire sample space is 1, and the probability of the null event is 0.

## Probability Theory: Some Terminology

### Experiment:

Experiment is an activity that is either observed or measured, such as tossing a coin, or drawing a card.

### Event (Outcome):

An event is a possible outcome of an experiment. For example, if the experiment is to sample six lamps coming off a production line, an event could be to get one defective and five good ones.

### Elementary Events:

Elementary events are those types of events that cannot be broken into other events. For example, suppose that the experiment is to roll a die. The elementary events for this experiment are to roll a 1 or a 2, and so on, i.e., there are six elementary events (1, 2, 3, 4, 5, 6). Note that rolling an even number is an event, but it is not an elementary event, because the even number can be broken down further into events 2, 4, and 6.

## Probability Theory: Some Terminology

### Sample Space:

A sample space is a complete set of all events of an experiment. The sample space for the roll of a single die is 1, 2, 3, 4, 5, and 6. The sample space of the experiment of tossing a coin three times is:

First toss.....T T T T H H H H  
Second toss.....T T H H T T H H  
Third toss.....T H T H T H T H

Sample space can aid in finding probabilities. However, using the sample space to express probabilities is hard when the sample space is large. Hence, we usually use other approaches to determine probability.

## Probability Theory: Some Terminology

### Unions & Intersections:

An element qualifies for the **union** of  $X, Y$  if it is in either  $X$  or  $Y$  or in both  $X$  and  $Y$ . For example, if  $X=(2, 8, 14, 18)$  and  $Y=(4, 6, 8, 10, 12)$ , then the union of  $(X,Y)=(2, 4, 6, 8, 10, 12, 14, 18)$ . The key word indicating the union of two or more events is *or*.

An element qualifies for the **intersection** of  $X, Y$  if it is in *both*  $X$  and  $Y$ . For example, if  $X=(2, 8, 14, 18)$  and  $Y=(4, 6, 8, 10, 12)$ , then the intersection of  $(X,Y)=8$ . The key word indicating the intersection of two or more events is *and*.

## Probability Theory: Some Terminology

### Mutually Exclusive Events:

Those events that cannot happen together are called mutually exclusive events. For example, in the toss of a single coin, the events of heads and tails are mutually exclusive. The probability of two mutually exclusive events occurring at the same time is zero

## Probability Theory: Some Terminology

### Independent Events:

Two or more events are called independent events when the occurrence or nonoccurrence of one of the events does not affect the occurrence or nonoccurrence of the others. Thus, when two events are independent, the probability of attaining the second event is the same regardless of the outcome of the first event. For example, the probability of tossing a head is always 0.5, regardless of what was tossed previously. Note that in these types of experiments, the events are independent if sampling is done with *replacement*.

## Probability Theory: Some Terminology

### Collectively Exhaustive Events:

A list of collectively exhaustive events contains all possible elementary events for an experiment. For example, for the die-tossing experiment, the set of events consists of 1, 2, 3, 4, 5, and 6. The set is collectively exhaustive because it includes all possible outcomes. Thus, all sample spaces are collectively exhaustive.

## Probability Theory: Some Terminology

### Complementary Events:

The complement of an event such as A consists of all events *not included* in A. For example, if in rolling a die, event A is getting an odd number, the complement of A is getting an even number. Thus, the complement of event A contains whatever portion of the sample space that event A does not contain.

## Probability Theory: Some Laws

### The Additive Law:

#### A. General Rule of Addition:

when two or more events will happen at the same time, and the events *are not* mutually exclusive, then:

$$P(X \text{ or } Y) = P(X) + P(Y) - P(X \text{ and } Y)$$

For example, what is the probability that a card chosen at random from a deck of cards will either be a king or a heart?

$$P(\text{King or Heart}) = P(X \text{ or } Y) = 4/52 + 13/52 - 1/52 = 30.77\%$$

## Probability Theory: Some Laws

### General Rule of Multiplication:

when two or more events will happen at the same time, and the events *are* dependent, then the general rule of multiplication law is used to find the joint probability:

$$P(X \text{ and } Y) = P(X) \cdot P(Y|X)$$

For example, suppose there are 10 marbles in a bag, and 3 are defective. Two marbles are to be selected, one after the other without replacement. What is the probability of selecting a defective marble followed by another defective marble?

Probability that the first marble selected is defective:  $P(X)=3/10$

Probability that the second marble selected is defective:  $P(Y)=2/9$

$$P(X \text{ and } Y) = (3/10) \cdot (2/9) = 7\%$$

## Probability Theory: Some Laws

### The Conditional Law:

Conditional probabilities are based on knowledge of one of the variables. The conditional probability of an event, such as X, occurring given that another event, such as Y, has occurred is expressed as:

$$P(X|Y) = P(X \text{ and } Y) / P(Y) = \{P(X) \cdot P(Y|X)\} / P(Y)$$

Note that when using the conditional law of probability, you always divide the joint probability by the probability of the event after the word given. Thus, to get  $P(X$  given  $Y)$ , you divide the joint probability of X and Y by the unconditional probability of Y. In other words, the above equation is used to find the conditional probability for any two *dependent* events.

When two events, such as X and Y, are *independent* their conditional probability is calculated as follows:

$$P(X|Y) = P(X) \text{ and } P(Y|X) = P(Y)$$

## Permutations and Combinations

**Question:** What do you know about permutations and combinations?

And the relationship between them?

## Permutations and Combinations

$${}_n P_r$$

**Theorem 1** The number of  $r$ -permutations of set with  $n$  distinct elements is

$${}_n P_r = n(n-1)(n-2) \cdots (n-r+1)$$

### Using Factorials

$${}_n P_r = \frac{n!}{(n-r)!}$$

In particular, note that  ${}_n P_n = n!$ .

### Example

How many ways are there to select a first-prize winner, a second-prize winner, and a third prize winner from 100 different people who have entered a contest?

*Solution:*

- The order in which people are selected matters so we want a 3-permutation.
- ${}_{100} P_3 = 100 \cdot 99 \cdot 98 = 970,200$

## Permutations and Combinations

A **permutation** of a set of distinct objects is an ordered arrangement of these objects. We also are interested in ordered arrangements of some of the elements of a set. An ordered arrangement of  $r$  elements of a set is called an  **$r$ -permutation**.

Let  $S = \{1; 2; 3\}$ . The arrangement/sequence 3, 1, 2 is a permutation of  $S$ . The arrangement 3, 2 is a 2-permutation of  $S$ .

## Permutations and Combinations

An  **$r$ -combination** of elements of a set is an unordered selection of  $r$  elements from the set. Thus, an  $r$ -combination is simply a subset of the set with  $r$  elements.

### Example

Let  $S$  be the set  $\{1, 2, 3, 4\}$ . Then  $\{1, 3, 4\}$  is a 3-combination from  $S$ . The 3-combination  $\{1, 3, 4\}$  is identical to the 3-combination  $\{4, 1, 3\}$  even though the 3-permutations 134 and 413 are different.

$${}_n C_r$$

The number of  $r$ -combinations of a set with  $n$  distinct elements is denoted by  ${}_n C_r$ .

**Theorem 2** *The number of  $r$ -combinations of a set with  $n$  elements, where  $n$  is a nonnegative integer and  $r$  is an integer with  $0 \leq r \leq n$ , equals*

$${}_n C_r = \frac{n!}{r!(n-r)!}$$

## Permutations and Combinations

### Example

How many ways are there to select five players from a 10-member tennis team to make a trip to a match at another school?

*Solution:* The answer is given by the number of 5-combinations of a set with ten elements.

$${}_{10}C_5 = \frac{10!}{5!5!} = 252$$

## Statistical Distributions

**Question :** What so you know about different distribution functions/curves?

# Statistical Values

## The Mean and Mode

The *sample mean* is the *average* and is computed as the sum of all the observed outcomes from the sample divided by the total number of events. We use  $\bar{x}$  as the symbol for the sample mean. In math terms,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x$$

where  $n$  is the *sample size* and the  $x$  correspond to the *observed values*.

The *mode* of a set of data is the number with the highest frequency.

The *population mean* is the average of the entire population and is usually impossible to compute. We use the Greek letter  $\mu$  for the population mean.

The *median* is the middle score. If we have an even number of events we take the average of the two middles.

## Statistical Distributions: Variance, Standard Deviation and Coefficient of Variation

The mean, mode, median, do a nice job in telling where the center of the data set is, but often we are interested in more.

For example, a pharmaceutical engineer develops a new drug that regulates iron in the blood. Suppose she finds out that the average sugar content after taking the medication is the optimal level. This does not mean that the drug is effective. There is a possibility that half of the patients have dangerously low sugar content while the other half have dangerously high content. Instead of the drug being an effective regulator, it is a deadly poison. What the pharmacist needs is a measure of how far the data is spread apart. This is what the variance and standard deviation do.

## Statistical Distributions: Variance, Standard Deviation and Coefficient of Variation

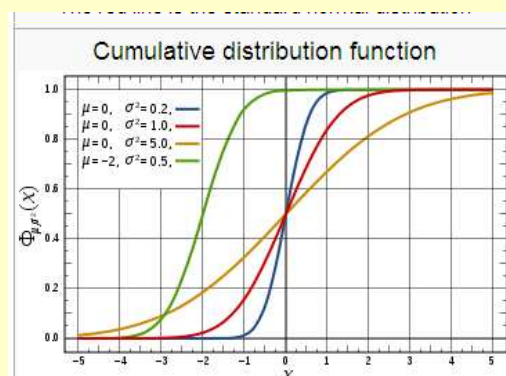
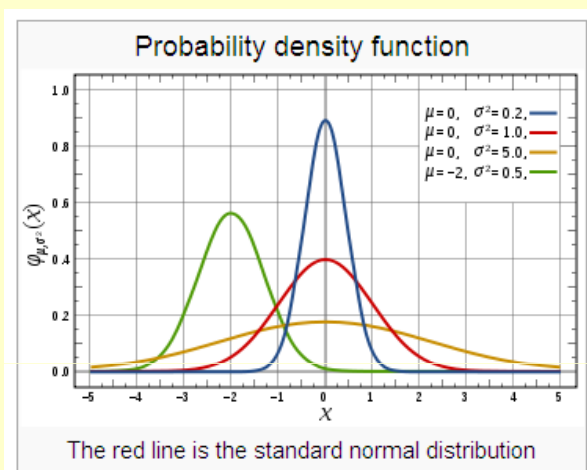
We define the *variance* to be

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

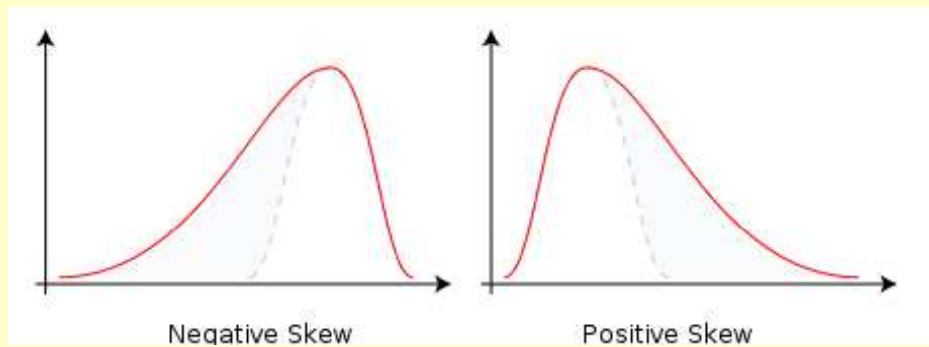
and the *standard deviation* to be

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

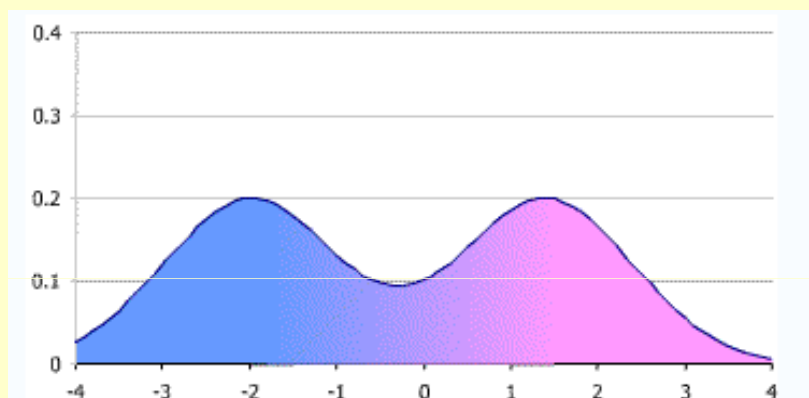
## Statistical Distributions: normal curves



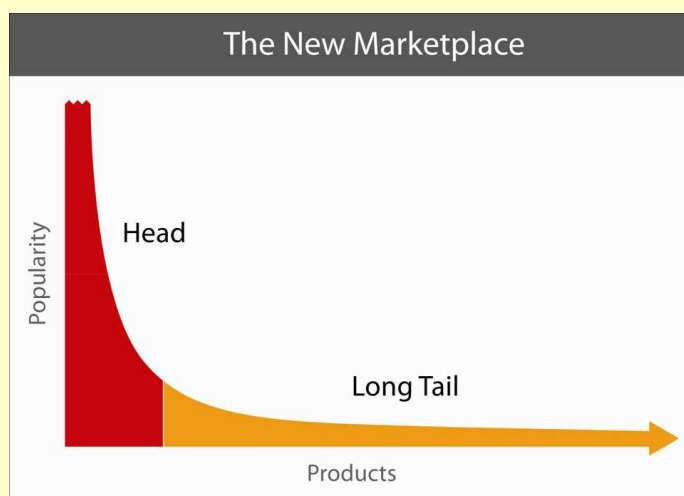
## Statistical Distributions: skewed curves



## Statistical Distributions: bimodal curves



## Statistical Distributions: long tail curves



## Statistical Distributions: correlation

In statistics, correlation and dependence are any of a broad class of statistical relationships between two or more random variables or observed data values.

Familiar examples of dependent phenomena include the correlation between the physical statures of parents and their offspring, and the correlation between the demand for a product and its price.

Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. For example, an electrical utility may produce less power on a mild day based on the correlation between electricity demand and weather.

Correlations can also suggest possible causal, or mechanistic relationships; however statistical dependence is not sufficient to demonstrate the presence of such a relationship.

# Statistical Distributions: correlation

## Correlation Example

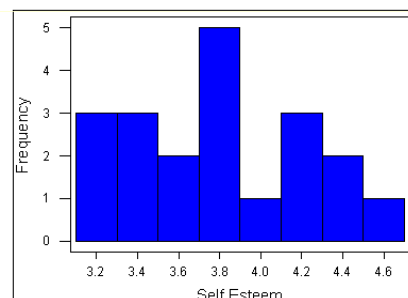
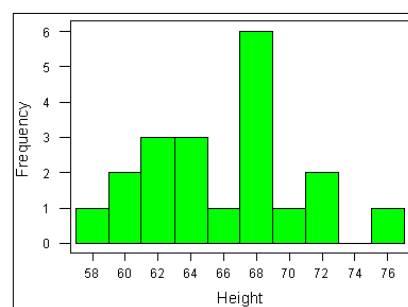
Let's assume that we want to look at the relationship between two variables, height (in inches) and self esteem.

Perhaps we have a hypothesis that how tall you are effects your self esteem (incidentally, I don't think we have to worry about the direction of causality here -- it's not likely that self esteem causes your height!).

Let's say we collect some information on twenty individuals (all male -- we know that the average height differs for males and females so, to keep this example simple we'll just use males). Height is measured in inches. Self esteem is measured based on the average of 10 1-to-5 rating items (where higher scores mean higher self esteem). Here's the data for the 20 cases (don't take this too seriously -- I made this data up to illustrate what a correlation is):

## Statistical Distributions: correlation example

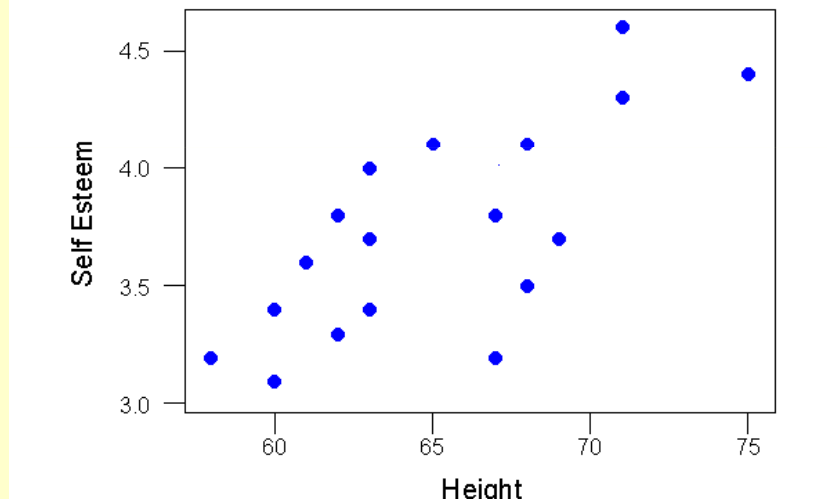
Person	Height	Self Esteem
1	68	4.1
2	71	4.6
3	62	3.8
4	75	4.4
5	58	3.2
6	60	3.1
7	67	3.8
8	68	4.1
9	71	4.3
10	69	3.7
11	68	3.5
12	67	3.2
13	63	3.7
14	62	3.3
15	60	3.4
16	63	4.0
17	65	4.1
18	67	3.8
19	63	3.4
20	61	3.6



## Statistical Distributions: correlation example

Variable	Mean	StDev	Variance	Sum	Minimum	Maximum	Range
Height	65.4	4.40574	19.4105	1308	58	75	17
Self Esteem	3.755	0.426090	0.181553	75.1	3.1	4.6	1.5

Finally, we'll look at the simple bivariate (i.e., two-variable) plot:



## Statistical Distributions: a standard correlation formula

Correlation is a measure of association between two variables. The variables are not designated as dependent or independent. The two most popular correlation coefficients are: Spearman's correlation coefficient rho and Pearson's product-moment correlation coefficient.

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

Where:

- N = number of pairs of scores
- $\sum xy$  = sum of the products of paired scores
- $\sum x$  = sum of x scores
- $\sum y$  = sum of y scores
- $\sum x^2$  = sum of squared x scores
- $\sum y^2$  = sum of squared y scores

NOTE: statistics tools/packages exist for calculating this « *automatically* »

=0.73 in our example, which is a *fairly strong* positive relationship

## Statistical Distributions: Testing the Significance of a Correlation

Once you've computed a correlation, you can determine the probability that the observed correlation occurred by chance. That is, you can conduct a significance test.

Most often you are interested in determining the probability that the correlation is a real one and not a chance occurrence.

In this case, you are testing the mutually exclusive hypotheses:

Null Hypothesis:	$r = 0$
Alternative Hypothesis:	$r \neq 0$

## Statistical Distributions: Testing the Significance of a Correlation

The easiest way to test this hypothesis is to find a statistics book/package that has a table of critical values of  $r$ .

As in all hypothesis testing, you need to first determine the significance level. Here, I'll use the common significance level of  $\alpha = .05$ . This means that I am conducting a test where the odds that the correlation is a chance occurrence is no more than 5 out of 100.

Before I look up the critical value in a table I also have to compute the degrees of freedom or  $df$ . The  $df$  is simply equal to  $N-2$  or, in this example, is  $20-2 = 18$ .

Finally, I have to decide whether I am doing a one-tailed or two-tailed test. In this example, since I have no strong prior theory to suggest whether the relationship between height and self esteem would be positive or negative, I'll opt for the two-tailed test. With these three pieces of information -- the significance level ( $\alpha = .05$ ), degrees of freedom ( $df = 18$ ), and type of test (two-tailed) -- I can now test the significance of the correlation I found: in this case the critical value is .4438.

As  $0.73 > .4438$  the correlation is significant

## Regression

Simple regression is used to examine the relationship between one dependent and one independent variable. After performing an analysis, the regression statistics can be used to predict the dependent variable when the independent variable is known.

### Regression goes beyond correlation by adding prediction capabilities.

People use regression on an intuitive level every day. In business, a well-dressed man is thought to be financially successful. A mother knows that more sugar in her children's diet results in higher energy levels. The ease of waking up in the morning often depends on how late you went to bed the night before. Quantitative regression adds precision by developing a mathematical formula that can be used for predictive purposes.

## TO DO - *Probability and statistics for game analysis*

In a game of noughts and crosses.

If 2 players play completely randomly (correctly following the rules of the game, but showing no other intelligence regarding where/how to play at each turn) then :

- What is the probability that the player who starts wins the game?
- What is probability that the player who goes second wins the game?
- What is probability that the game ends in a draw?

Calculate the probabilities ( $\pm 0.1$ ), and test your answer through a computer simulation

