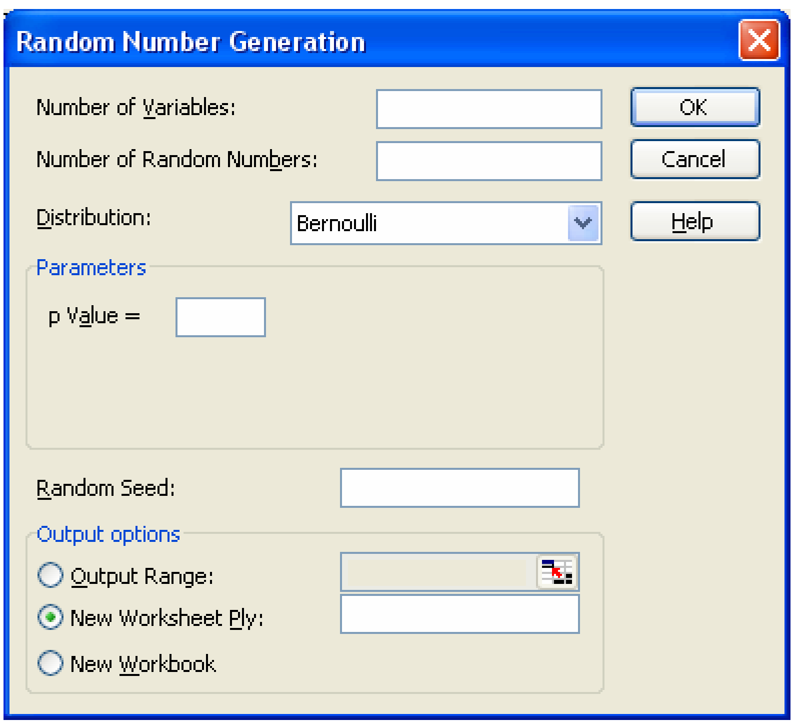
We shall understand the distribution of proportions by looking at a process called Bernouli Process

The particular process connected to the probability theory that I am going to focus on is called “Bernoulli Process.”

Make sure to read about such a process and include more details from the Website http://en.wikipedia.org/wiki/Bernoulli_trial

For such a process, we make many independent trials, and term the outcomes only as “success” or “failure” and the probability of “success” in each trial stays the same. The number of trials is generally denoted by n and the probability of success in any trial is denoted by p.
Example: Let us say we would like to observe an experimental illustration of Bernouli Process for 10000 tosses of a fair coin. One will need a lot of patience and resources to perform such a trial.

If we use the “Data Analysis” feature of Excel, we can choose the random number generator, and then choose Bernoulli as shown below

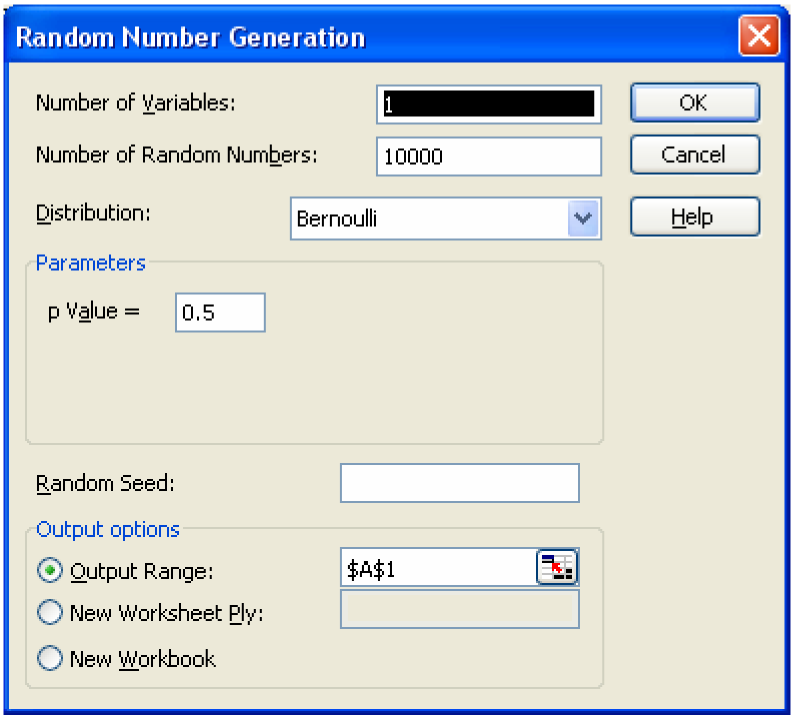


For 10000 trials, we shall need 10000 numbers of type Bernoulli with p=.5

If we get a head, we shall treat our outcome as having a numerical value 1

If we do not get a head, we shall treat our outcome as having a numerical value 0

Let us fill in the dialogue box in the manner shown below



Here are a few numbers that show up on the top of the spreadsheet,

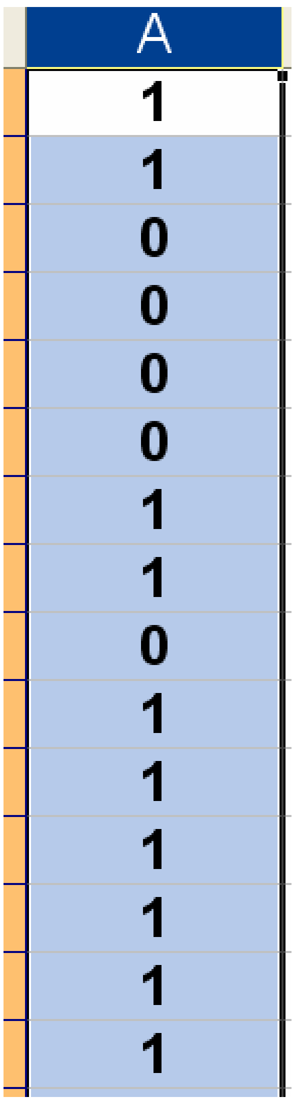
1
1
0
0
0
0
1
1
0
1
1
1
1
1
1
1

Note that out of 16 outcomes, 11 outcomes have shown a “head” which is of course much larger than 50% of 10 that is 8.

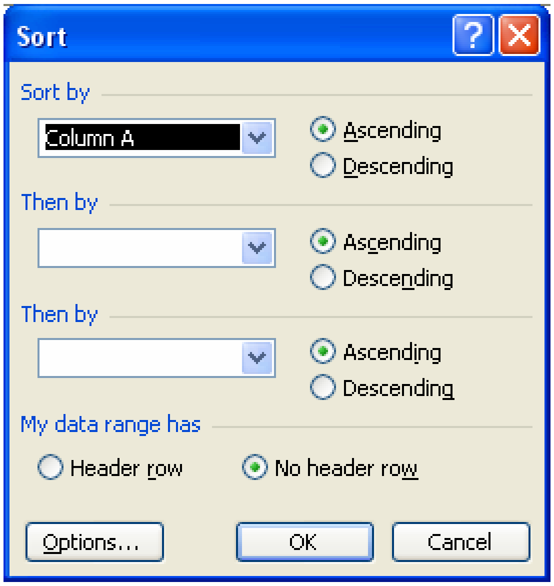
But we count the number of heads (that is the number of 1s) from all the 10000 outcomes, the situation will be different.

We can have computer do this calculation in the following manner

Click on the letter A on the top



Click on Data and then on Sort and follow the following



The data will be sorted

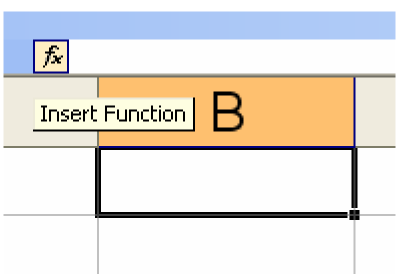
To compute the number of heads out of 10000,

We can either do an eyeball count,

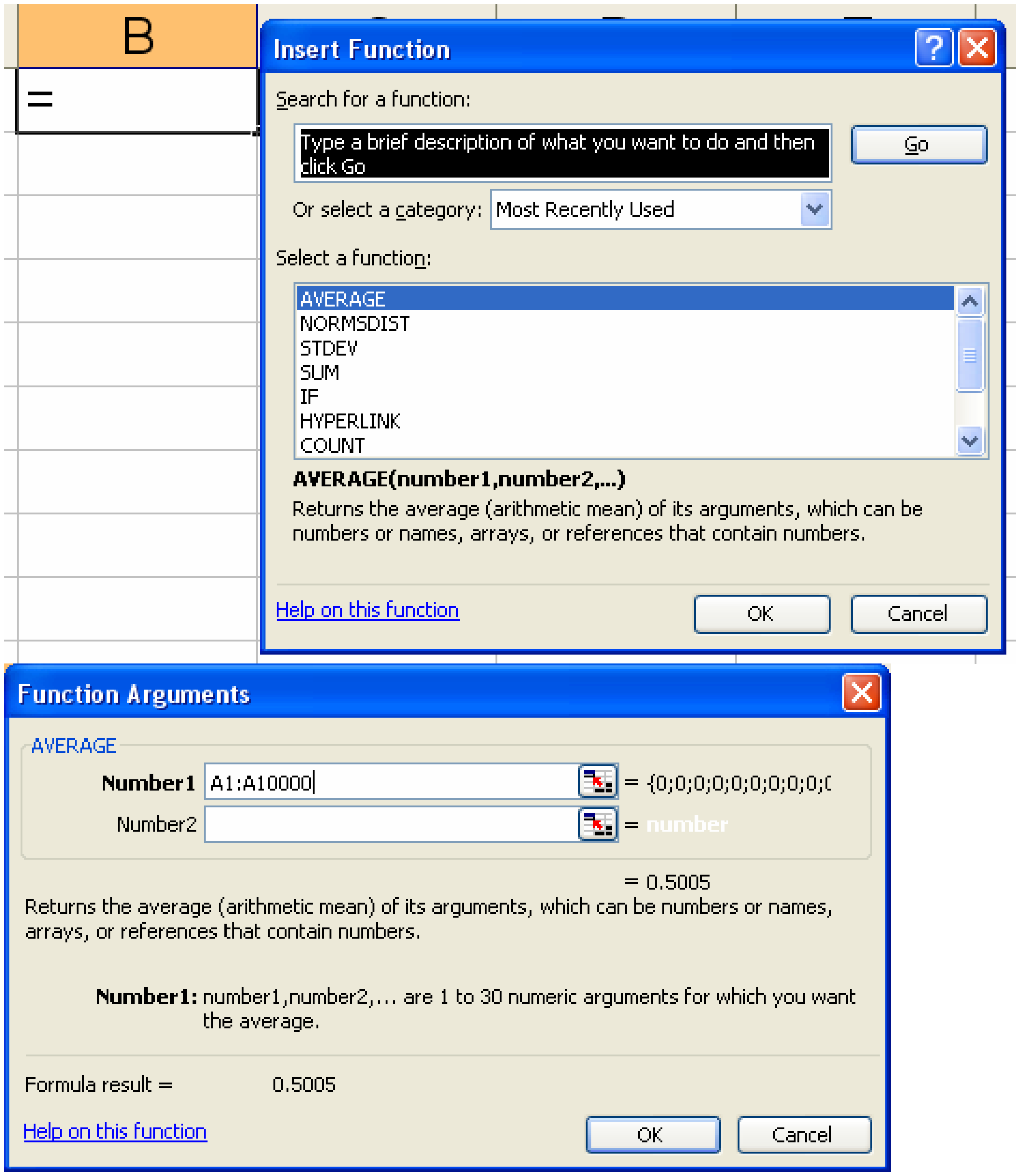
OR we can compute the average of these 0s and 1s and the average will tell us the proportion of the trials that are HEADS

We can do this by following the steps shown below

CLICK ON AN EMPTY CELL, LIKE THE ONE BELOW B



Click on fx, the Insert Function option



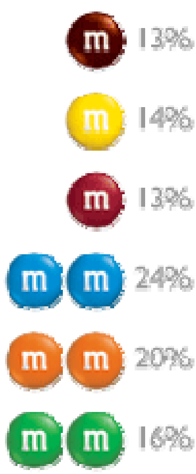
See that the average is 0.5005

Which means that 0.5005times 10000 that is 5005 of the 10000 outcomes did show a heads

What is saw above is the feature of variability

Example 2:

The famous candy manufacturer M&Ms gives the following information for the percentage of various colors in the bags.



We can interpret this information that the bags of candies will be filled of candies of various colors from a large bin of candies.

Assume that many of us have bags of candies that each contains 200 candies. Would the percentage of "blue" will be the same in each bag, Of course not, as the simulation of bags in a manner shown in the video

Video1

In the proportions

At

http://www.mathgurukul.org/Videotable.htm

shows.

Example 3:

An interesting application of such a distribution of proportions can be seen in the following mathematical procedure that can be found in any book on Basic Statistics

Suppose that we would like to know the percentage of population in the nation that thinks that Economic Progress should be the top priority in the planning process.

Since it is not possible to ask each individual adult of their opinion, we can take a simple random sample of size n and ask if they think that Economic Progress should be the top priority in the planning process.

If the number of such people in the population is x, then
$\frac{x}{n}$ is the proportion of people in this sample think that
Economic Progress should be the top priority in the planning process.

We can use the result from this sample to estimate the proportion of people in the entire population that think that
Economic Progress should be the top priority in the planning process.

If we would like our confidence level to be 95% the procedure will work as follows

The estimate is $\frac{x}{n}$ and the margin of error is $1.96\sqrt{\frac{\left(\frac{x}{n}\right)\left(1-\frac{x}{n}\right)}{n}}$

For example, if 288 of 400 people think that
Economic Progress should be the top priority in the planning process,

Then the estimate of such proportion in the population is $\frac{288}{400}=0.72$

With a margin of error $1.96\sqrt{\frac{.72(1-.72)}{400}}\cong 0.044$

We can say with 95% confidence that the proportion of people in the entire population who think that
Economic Progress should be the top priority in the planning process is in the interval

$$\left(.72-.044,.72+.044\right)=\left(0.676,0.764\right)$$

Example 4:

A simulation to show the meaning of a 95% confidence interval

Assume that we have a population in which the proportion of "successes" is 0.59
We are going to take 1000 samples of size 250 each from this population, denote "success" by 1 and then construct 1000 intervals with 95% confidence and see how many of these 1000 captured the proportion in the population that is 0.59

Look at the

Video2

In the proportions

At

http://www.mathgurukul.org/Videotable.htm

shows.

We see that 27+36=63 out of 1000 missed the true proportion,

937 of 1000 captured the true proportion

Now the procedure should show a capture rate of 95% but we saw 93.7% only. Whether this number is significantly lower, such questions can be answered by using the hypothesis testing procedures that we shall do in later chapters.