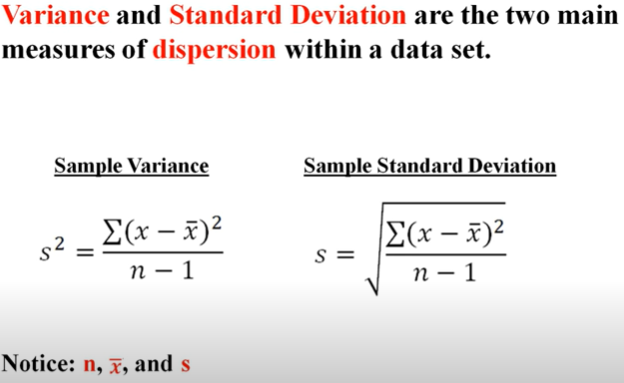
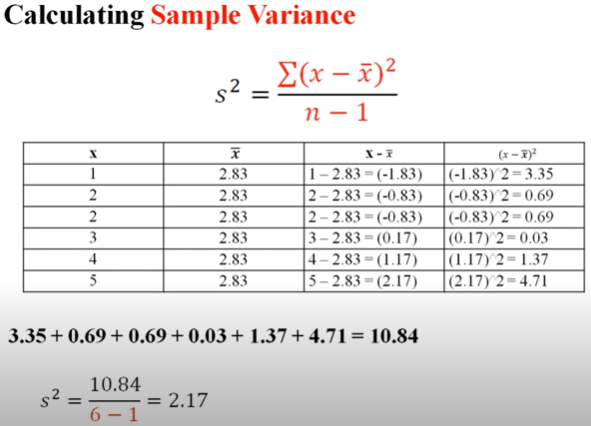
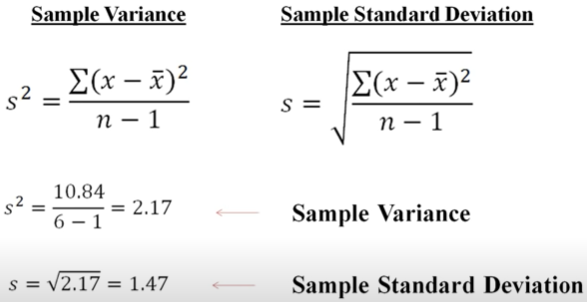
***Statistics Notes – 2***

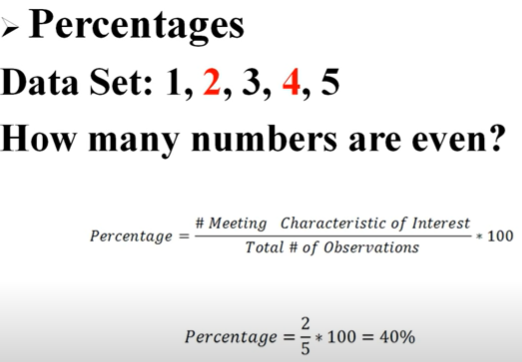
***Measures of Dispersion: Variance and Standard Deviation of a Sample.***





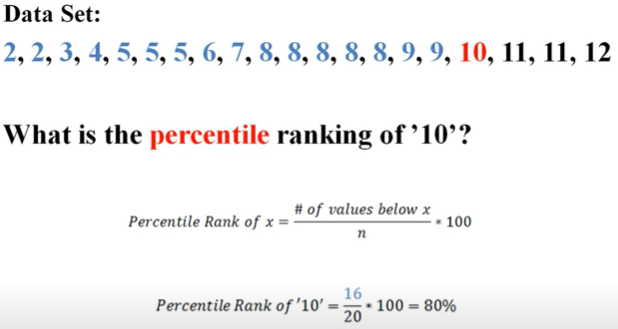


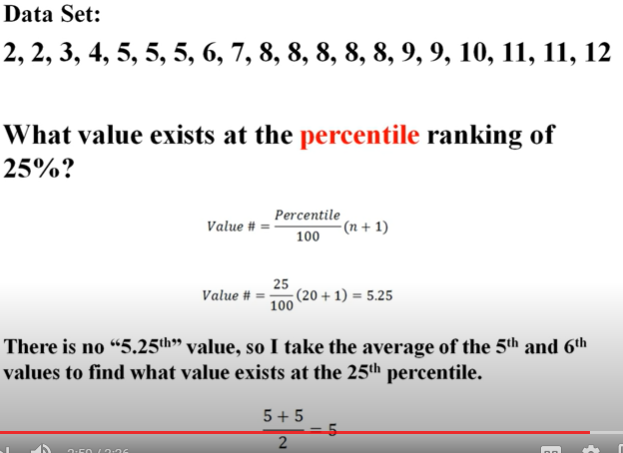
***Percentiles and Quartiles***

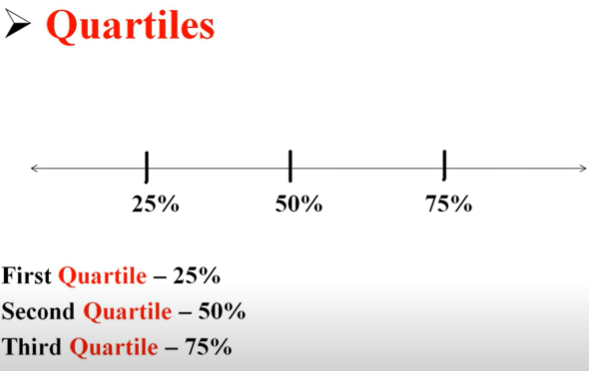


A **Percentile** is a value below which a certain percentage of observation lies.

What percent values are below that ranking?



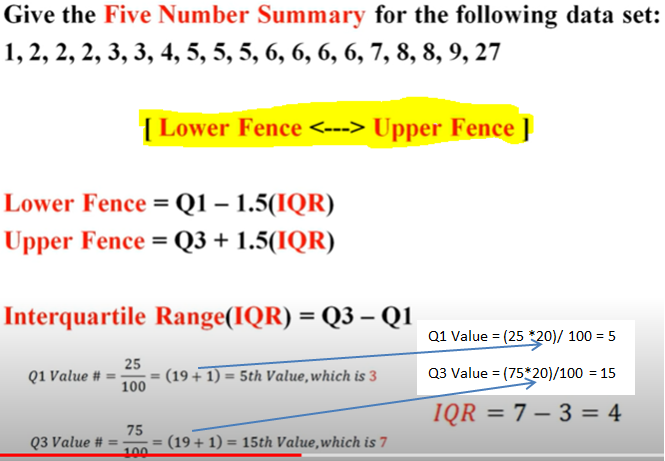




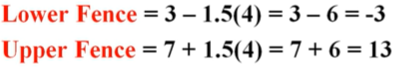
***The Five Number Summary, interquartile Range (IQR), and Boxplots***

Then Five number Summary is a method of summarizing a distribution of data. **1) Minimum, 2) first Quartile (A1), 3) Median, 4) Third Quartile (Q3), 5) Maximum**.

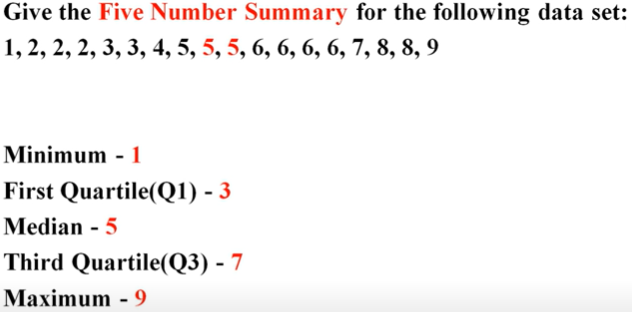
In the below example if we see then 27 number seems an outlier so we can remove the outlier which with below criteria (highlighted in yellow) i.e. if anything goes out of this fence then we can remove it.



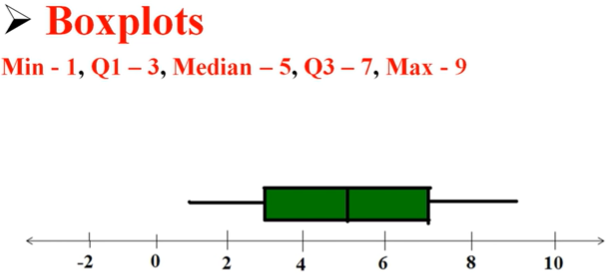
As per below calculation if anything does not fall between -3 to 13 is an outlier.



Now are able to derive this below value.

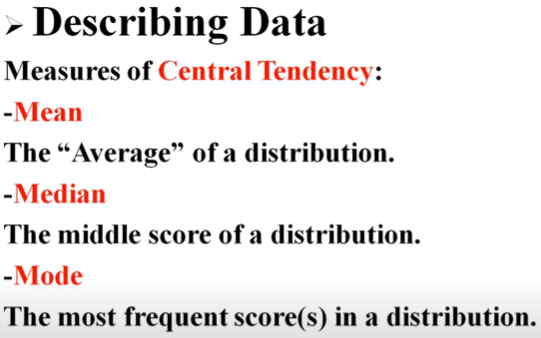


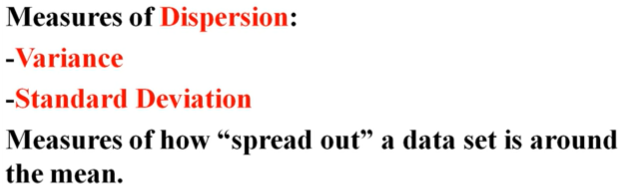
On the basis of above derived values we can draw Box Plot as below:

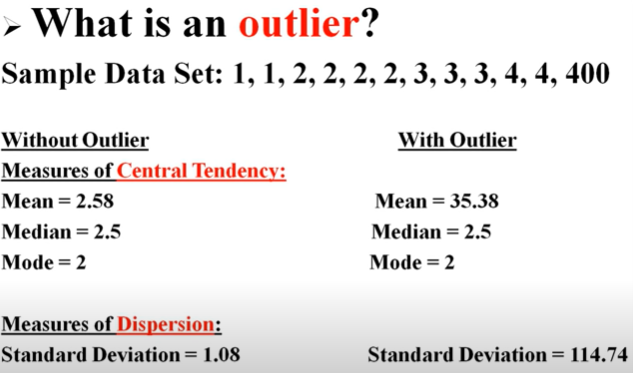


***The Effects of Outliers***

Below we will see what **outliers** can do in distribution of data

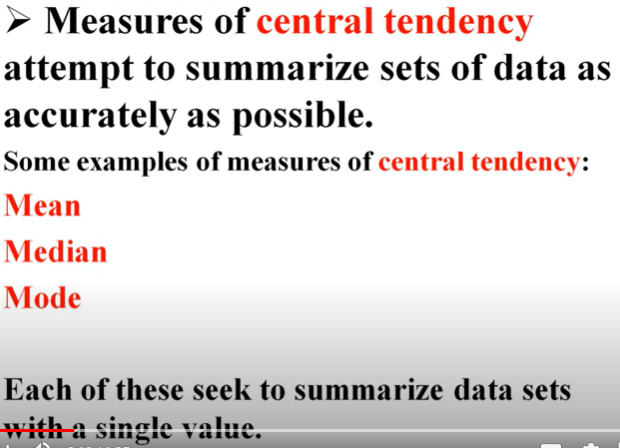




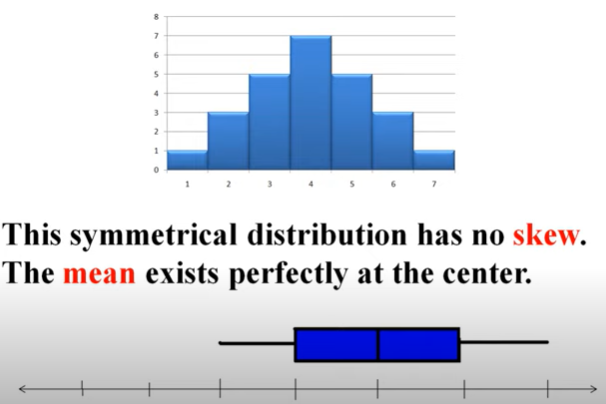


After seeing the difference in Mean and in standard deviation while computation either we remove the outlier or we do not consider mean in the calculation and use median or mode

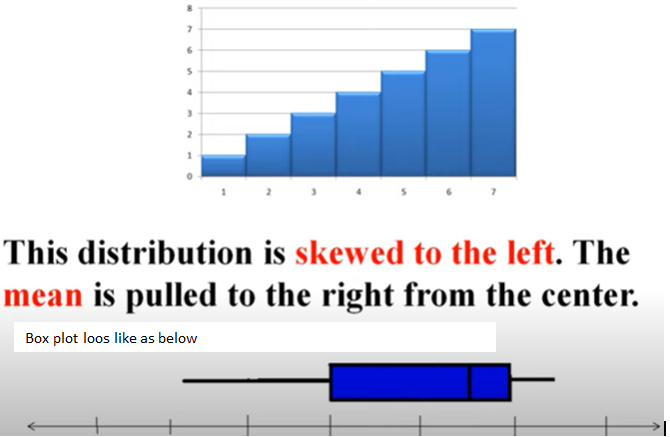
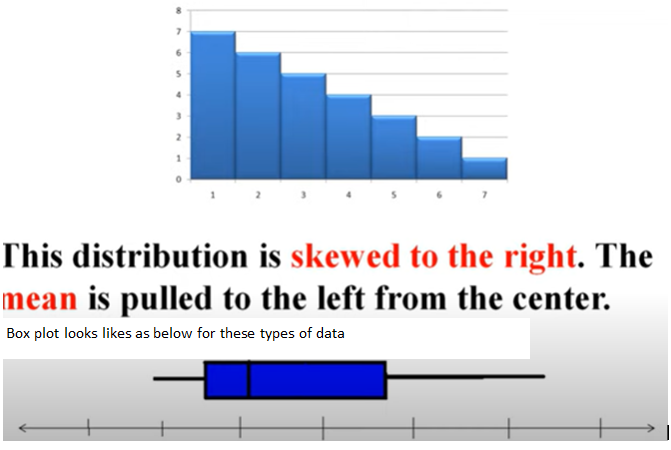
***SKEWNESS***.





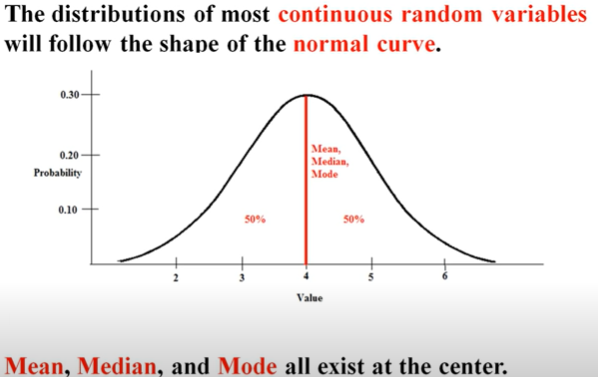


On the above diagram if draw box plot then it will derive something like this

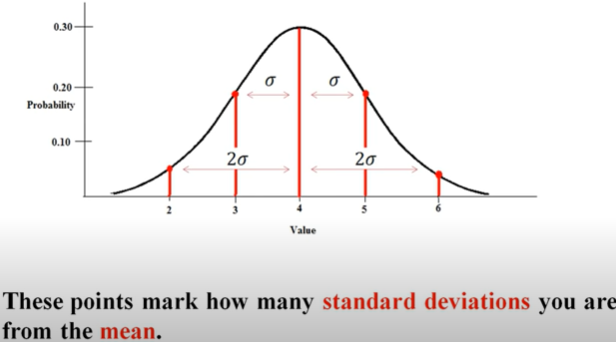


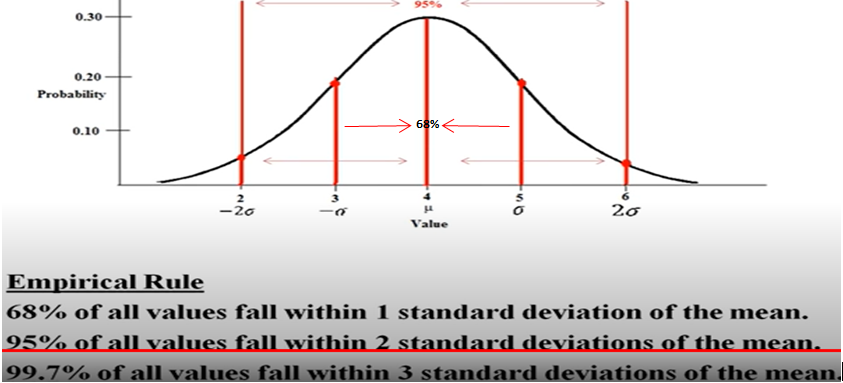
In Hypothesis testing Skewness is very important.

The ***Normal Curve*** and ***Empirical Rule***

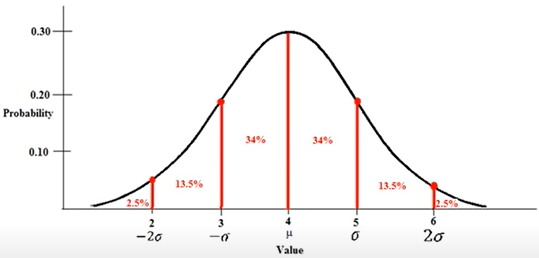


Graph changes direction at inflection point. If we mark one inflection point from the mean then it will mark in standard deviation from the mean.

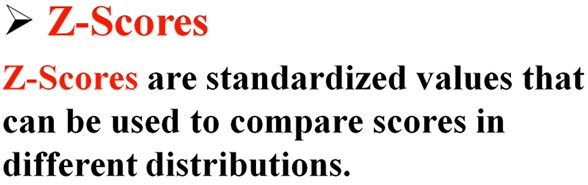




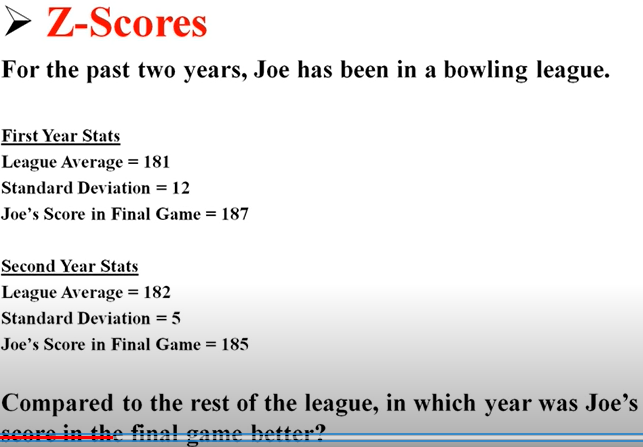
In below example we have distributed the 68% score between 34% and 34%. And using this Empirical rule we can find out what area are associated between different distances without doing complicated calculation.



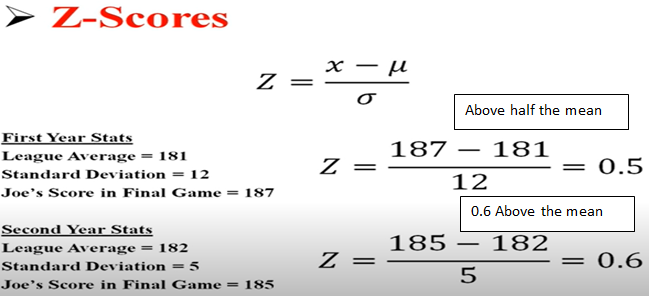
***Z-Scores Part 1***

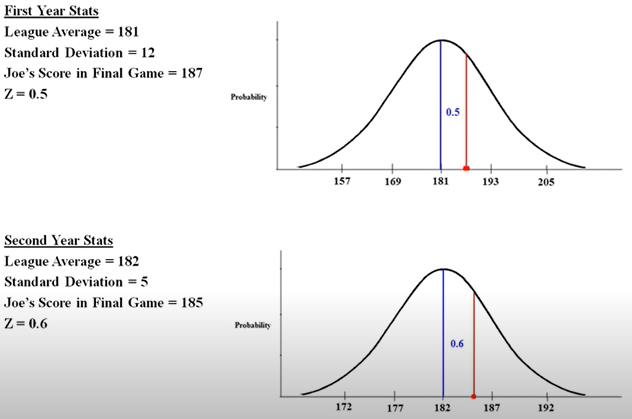






What Z-scores tell us that how far above or below we are it’s the mean (how many Standard Deviation above or below that’s then mean).

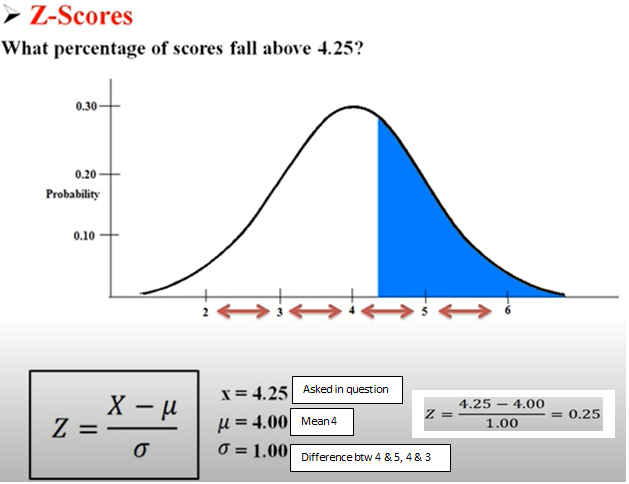




As per above graph comparing to other distribution 185 score is higher than score 187, as 187 score 0.5 standard deviation above the mean whereas 185 scores 0.6 SD above the mean.

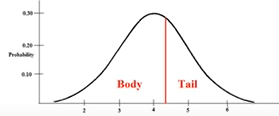
***Z-Scores Part 2***

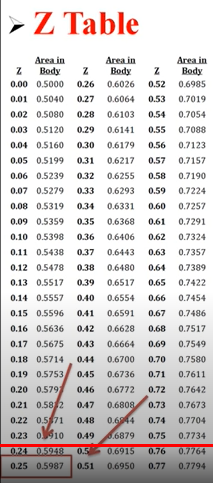
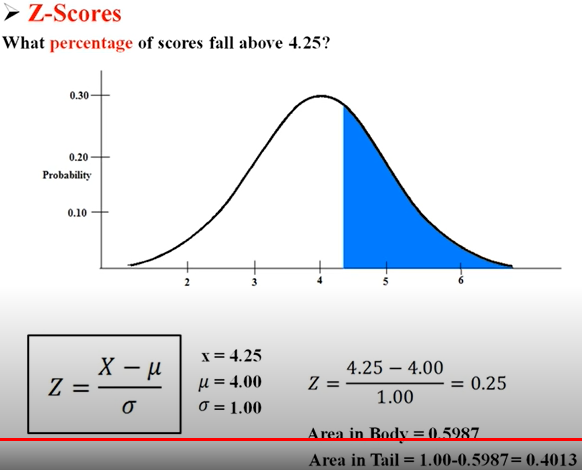
In the past we have learned Empirical Rule for Normal Distributed data.



As per above 4.25 is 0.25 above the mean.

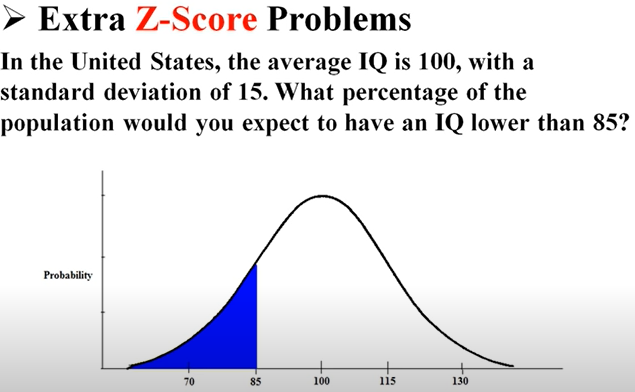
Always in the distribution in the bigger area call as Body and small area call tail.

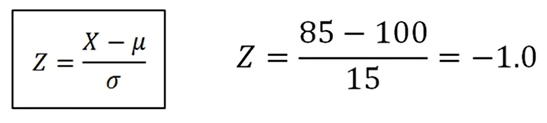


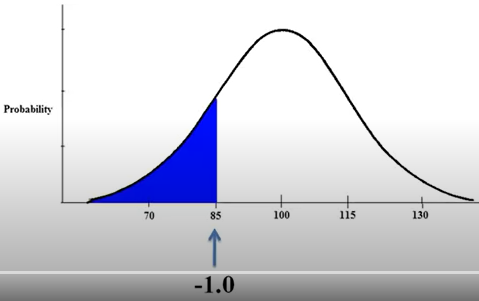
In the above we want to know above the 4.25 so if whole area is of 100% then what is the blue highlighted then check calculation of Tail in above screen shot. So the answer for the question that score fall above of 4.25 is 0.4013 or 40%. That is the use of Z-score to find the percentage.

1 extra question:

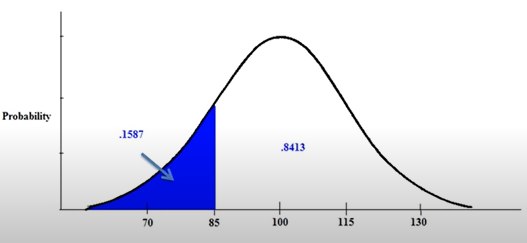
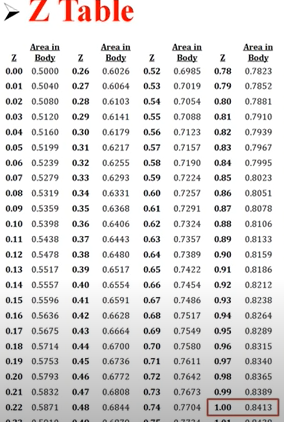




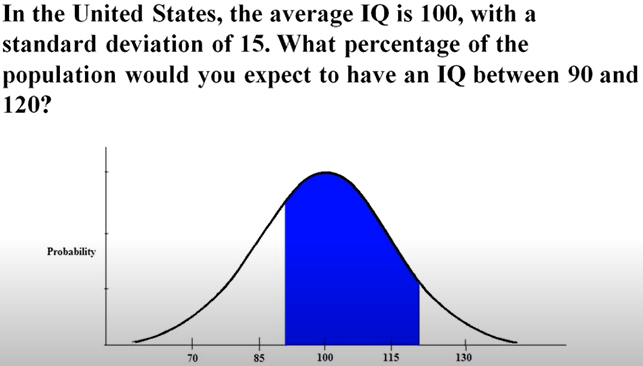
-1 means 1 standard deviation is below the “mean”. As mentioned in below diagram.



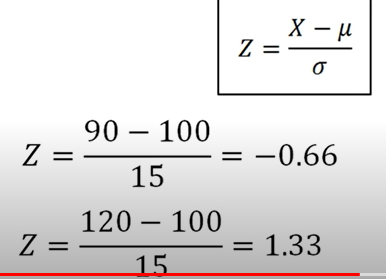
If we calculate then 1 - 0.8413 = 0.1587 so the answer - % of IQ lower than 85 is 16%

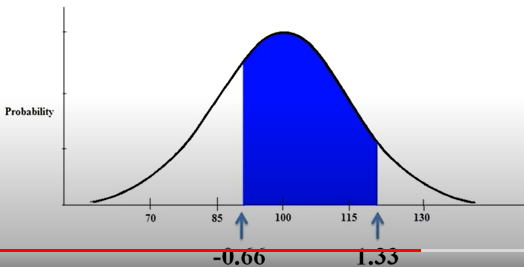


2 extra question:



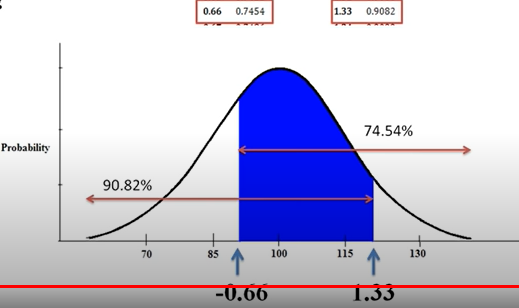
We will calculate Z for both 90 and 120





Look at the below Z-table.



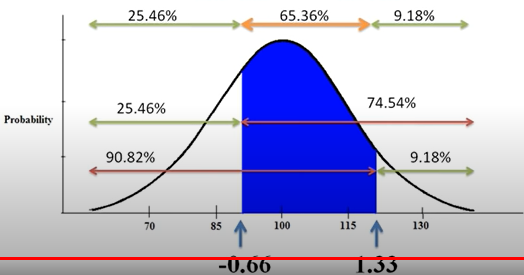


We need to calculate the area covered in blue for that lets calculated the area as below which is not covered for 74.54% and 90.82% as per below diagram.

For 74.54%: - 1 – 0.7454 = 0.2546 i.e. 25.46%

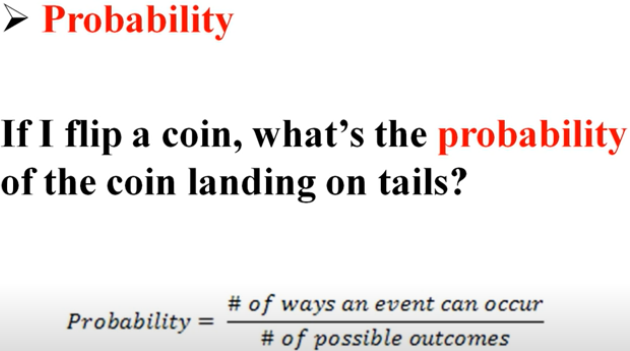
For 90.82%:- 1 – 0.9082 = 0.0918 i.e. 9.18%

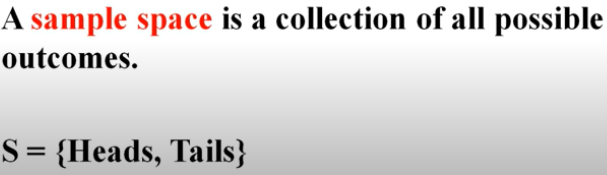
As per calculation 0.2546 + 0.0918 = 0.3464 so 1 – 0.3464 = 0.6536 i.e. 65.36% is the population we would expect to have an IQ between 90 and 120.

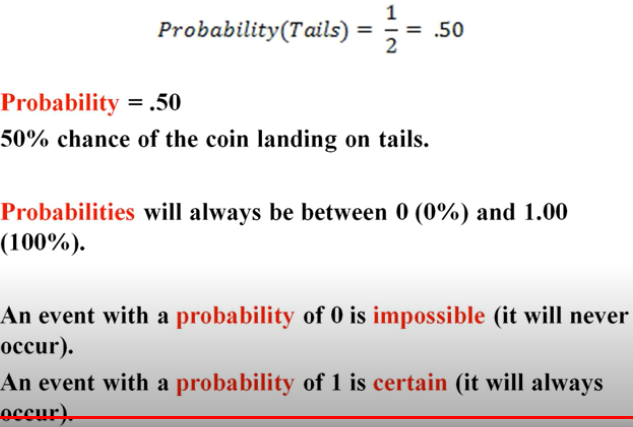


***Basics of Probabilities***

Probabilities are a measure of the likelihood of an event.





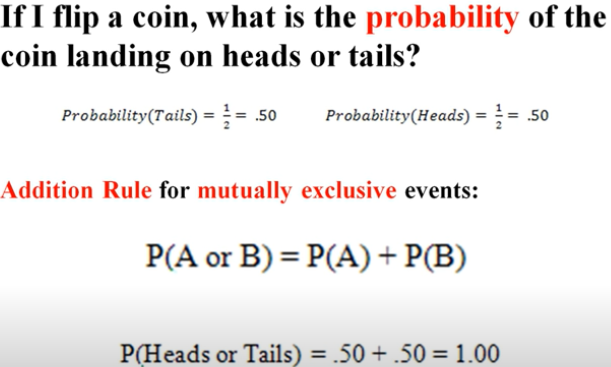


**Probability Additional Rules**:

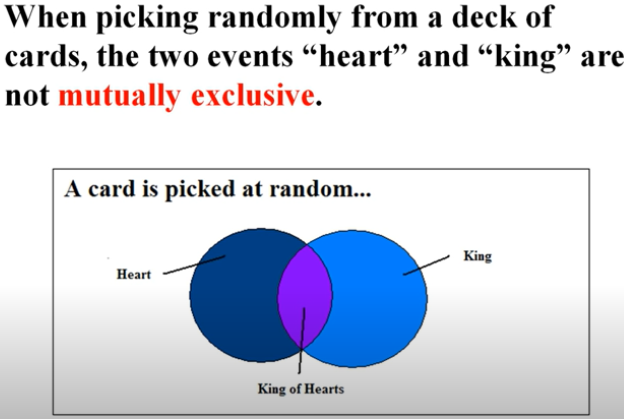
When both the events are not getting at the same time then it is “**mutually exclusive**”



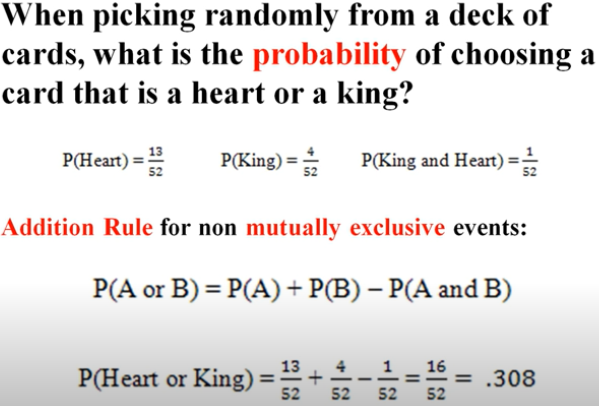
Mathematical Calculation of mutually exclusive events:

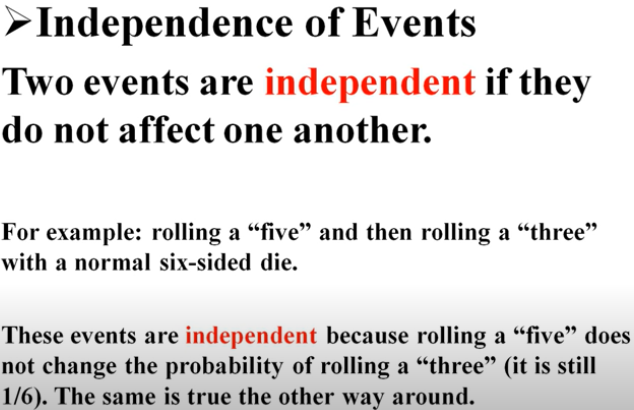


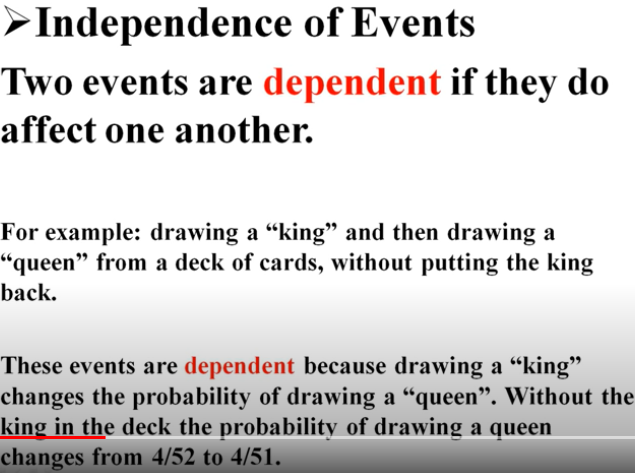
In the below example if we see heart card, King and King of Heart cards these events can occure individually or at the same time.

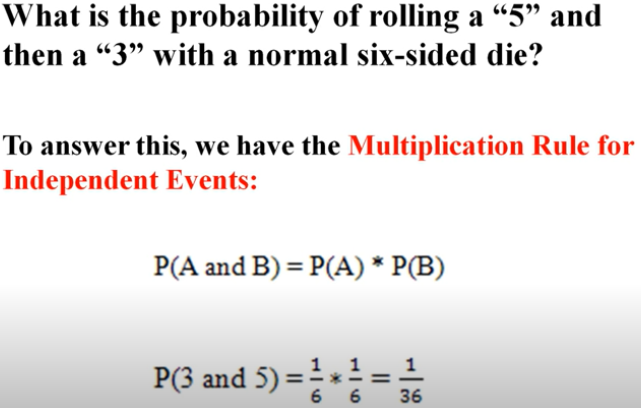


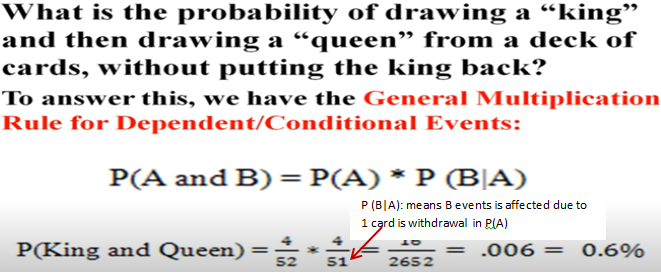
Mathematical calculation of Not Mutually exclusive:



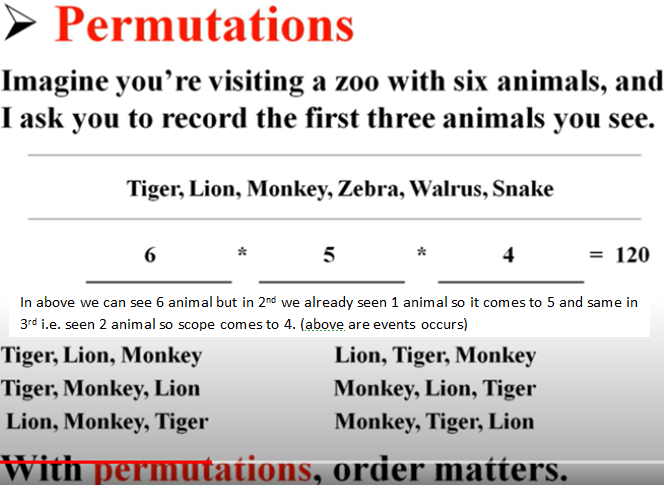
**Probability Multiplication Rules (Probability “and”).** In this event we multiply both the events.



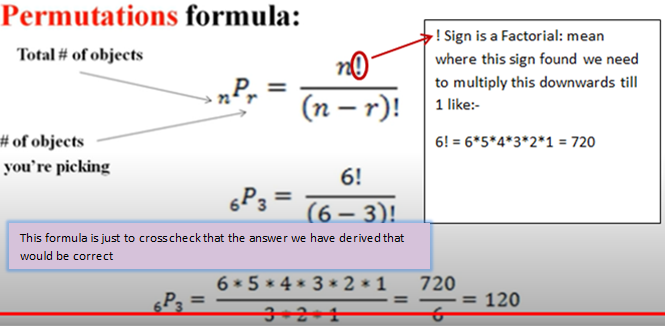




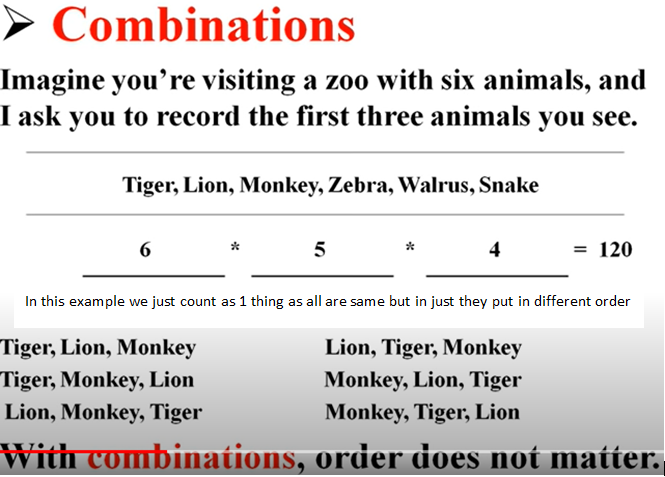
***Permutation***

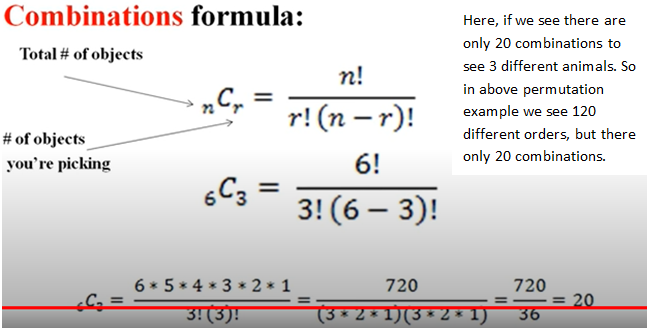


In the above example we don’t care about the combination only order matters same order should not be repeat.



**Probabilities – Combinations**





***Discrete and Continuous Random Variable***

