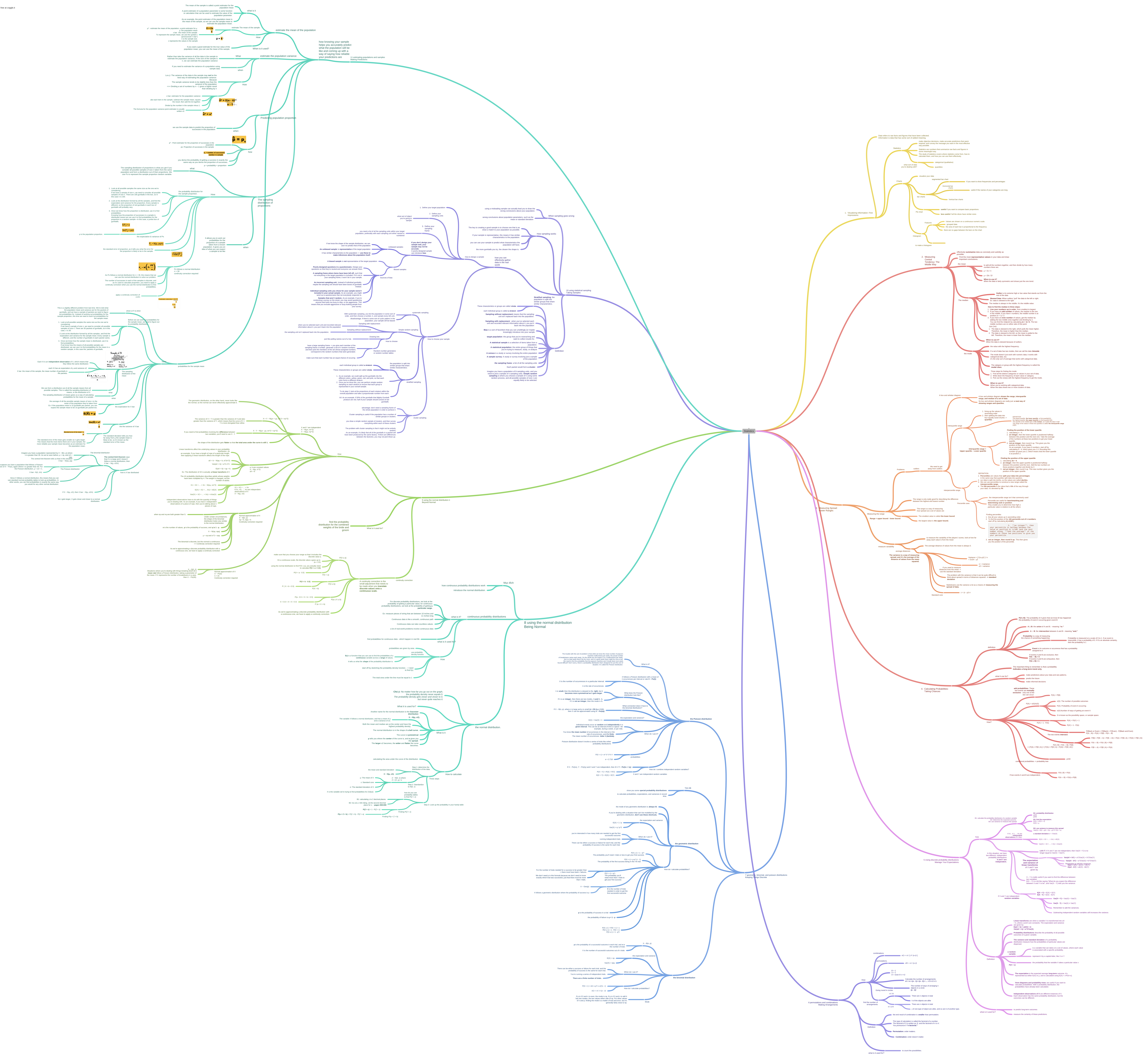
coggle
made for free at coggle.it


Statistics
1. Visualizing Information: First Impressions
Statistics are numbers that summarize raw facts and figures in some meaningful way.
The study of statistics covers where statistics come from, how to calculate them, and how you can use them effectively.
what sort of data you're dealing with?
categorical (qualitative)
quantities
Charts
visualize your data
bar charts
segmented bar chart
Horizontal bar charts
Vertical bar charts
Pie chart
useful if you want to compare basic proportions
less useful if all the slices have similar sizes
Histogram
Features
Values are shown on a continuous numeric scale.
grouped data
the area of each bar is proportional to the frequency
there are no gaps between the bars on the chart
to make a histogram
2. Measuring Central Tendency: The Middle Way
the mean
The median
3. Measuring Spread: Power Ranges
Interquartile range = Upper quartile – Lower quartile
Problems
outliers
We need to get away from outliers
interpercentile range.
Percentile uses
Measuring the range
Range = upper bound - lower bound
The range is only really good for describing the difference between the highest and lowest number
The range is a way of measuring how spread out a set of values are
The smallest value is called the lower bound
the largest value is the upper bound.
measure variability
to measure the variability of the players' scores, look at how far away each value is from the mean
average distances
The average distance of values from the mean is always 0.
The variance is a way of measuring spread, and it's the average of the distance of values from the mean squared.
Variance = [Σ(x-μ)2] / n = Σx2/n - μ2
If you need to measure distances from the mean -> use the standard deviation
Statisticians use the variance a lot as a means of measuring the spread of data.
Standard core
z = (x - μ)/ σ
4. Calculating Probabilities: Taking Chances
5 Using discrete probability distributions Manage Your Expectations
6 permutations and combinations Making Arrangements
How
combinations
nCr = n! / [r!* (n-r)!]
permutations.
nPr = n! / (n-r)!
0! = 1
1! = 1
n! = even if n >=2
Calculate the number of arrangements
n! = n × (n - 1) × (n - 2) × ... × 3 × 2 × 1
Going round in circles
The number of ways of arranging n objects in a circle
(n - 1)!
find the number of arrangements
n! / k!
There are n objects in total
k of the objects are alike
n! / j! k!
There are n objects in total
j of one type of object are alike, and so are k of another type.
Definition
the end result of combination is smaller than permutation.
This type of calculation is called the factorial of a number.
The factorial of 3 is written as 3!, and the factorial of n is n!.
You pronounce it "n factorial."
Permutation: order matters.
Combination: order doesn't matter.
what is it used for?
to count the possibilities.
7 geometric, binomial, and poisson distributions Keeping Things Discrete
8 using the normal distribution Being Normal
Mục đích
how continuous probability distributions work
introduce the normal distribution.
continuous probability distributions
what is it?
For discrete probability distributions, we look at the probability of getting a particular value; for continuous probability distributions, we look at the probability of getting a particular range.
Ex: measure pieces of string that are between 10 inches and 11 inches long
Continuous data is like a smooth, continuous path
Continuous data can take countless values.
a lot of real-world problems involve continuous data
what is it used for?
find probabilities for continuous data - which happen in real life
How
probabilities are given by area
use probability density function.
f(x) is a function that you can use to find the probabilities of a continuous variable across a range of values.
It tells us what the shape of the probability distribution is.
start off by sketching the probability density function. - > need to find f(x)
The total area under the line must be equal to 1
9 using the normal distribution ii Beyond Normal
What is it used for?
find the probability distribution for the combined weights of the bride and groom
continuity correction
A continuity correction is the small adjustment that needs to be made when you translate discrete values onto a continuous scale.
P(X ≤ a)
P(X < a)
P(X < a - 0.5)
P(X ≥ b)
P(X > b - 0.5)
P(X > b)
P(a ≤ X ≤ b)
P(a - 0.5 < X < b + 0.5)
P (a < X < b)
10 using statistical sampling Taking Samples
When sampling goes wrong
using a misleading sample can actually lead you to draw the wrong conclusions about your population.
wrong conclusions about population parameters, such as the mean or standard deviation
How sampling works
The key to creating a good sample is to choose one that is as close a match to your population as possible.
If your sample is representative, this means it has similar characteristics to the population.
you can use your sample to predict what characteristics the population will have.
the more gumballs you try, the clearer the shape is.
how you can effectively gather data in the real world
How to design a sample
1. Define your target population
2. Define your sampling units
what sort of object you're going to sample
3. Define your sampling frame
you need a list of all the sampling units within your target population, preferably with each sampling unit either named or numbered
If you don't design your sample well, your sample may not be accurate.
A poorly designed sample can introduce bias
Unbiased samples
If we know the shape of the sample distribution, we can use it to predict that of the population
An unbiased sample is representative of the target population
it has similar characteristics to the population -> use these to make inferences about the population itself.
Biased samples
A biased sample is not representative of the target population
Sources of bias
Poorly designed questions in a questionnaire. Design your questions so that they're neutral and everyone can answer them
A sampling frame where items have been left off, such that not everything in the target population is included. If it's not in your sampling frame, it won't be in your sample.
An incorrect sampling unit. Instead of individual gumballs, maybe the sampling unit should have been boxes of gumballs instead.
Individual sampling units you chose for your sample weren't included in your actual sample. As an example, you might send out a questionnaire that not everybody responds to.
Samples that aren't random. As an example, if you're conducting a survey on the street, you may avoid questioning anyone that looks too busy to stop, or too aggressive. This means that you exclude aggressive or busy-looking people from your survey.
How to choose your sample
systematic sampling
With systematic sampling, you list the population in some sort of order, and then choose a number, k, and sample every kth unit.
disadvantage. If there's some sort of cyclic pattern in the population, your sample will be biased
Simple random sampling
Sampling with replacement
when you've selected each unit and recorded relevant information about it, you put it back into the population
Sampling without replacement
How to choose
Drawing lots
just like pulling names out of a hat.
Random number generators
Definition
Stratified sampling: the population is split into similar groups that share similar characteristics.
These characteristics or groups are called strata
each individual group is called a stratum
Sampling without replacement means that the sampling unit isn't replaced back into the population.
Sampling with replacement : when you've selected each unit and recorded relevant information about it, you put it back into the population
Bias is a sort of favoritism that you can unwittingly (or maybe knowingly) introduce into your sample
target population: the group that you're reseraching and want to collect results for.
A statistical sample is a selection of items taken from a population.
A statistical population: the entire group of things that you're trying to measure, study, or analyze
11 estimating populations and samples Making Predictions
how knowing your sample helps you accurately predict what the population will be like and coming up with a way of saying how reliable your predictions are