Online Resource

# Chapter 3: Variables

## Normal distribution

Imagine you have a walk to the local supermarket. Every time a red car passes you, your happiness is increased slightly. Every time a white car passes you, your happiness is decreased slightly. Every time, a passer-by smiles at you, you get slightly happier; and you get slightly grumpier every time a passer-by ignores you. Every time you see a flower, you feel slightly happier; every time you see a nettle, you feel slightly less happy.

At the end of your walk, your happiness level will be the product of hundreds of small, independent random nudges one way or the other. Due to the rules of mathematics (the Central Limit Theorem), these nudges will combine to produce a normal distribution of happiness. A normal distribution is the consequence of the combined effects of lots of independent causes.

A normal distribution is completely described by two numbers: its mean and its standard deviation. Our example allows us to see why.

**Mean:** If we had hundreds of different people walking that journey at different times, then the average happiness would be fixed by the ratio of red to white cars, whether passers-by are typically smilers and how well people tend their gardens. These fix the mean happiness at the end of the walk.

**Standard deviation:** The only other aspect of the situation that matters is just how many cars, how many people, how many gardens are passed. Each of these determines how what the spread of so-called “happinesses” could be. If the journey only ever involves 2 cars, 1 passer-by and 1 garden, then it is impossible to become more elated or grumpy than 4 (2+1+1) nudges either side of neutral. If the journey involves 200 cars, 75 passers-by and 42 gardens, then in theory you could be extremely lucky and get 317 positive nudges (200+75+42) or extremely unlucky and get 317 negative nudges. The range of “happinesses” that are produced depends on the number of different factors causing it.

For this reason, the normal distribution only has two parameters: the mean and the standard deviation. The formula for it is this:



In this formula, *x* is any value we might be interested in; *e* is a constant 2.718, *μ* is the mean of the distribution and *σ* is the standard deviation.

## Skewed distributions

A skewed distribution is often encountered in statistical textbooks, as we have ourselves included one in Figure 3.7. The explanation of a normal distribution, which has no skew, is very simple and universal. The existence of skewed distributions appears to spoil things.

Our first thought might be that skew could happen if the positive nudges were all stronger than the negative nudges. This doesn’t produce any skew – it just shifts the mean happiness upwards.

Our next thought does produce skew: that the nudge you get from a positive event has more effect the more happy you are at that moment. If I have had 1 positive nudge, then getting another one lifts my happiness from 1 to 2.2 (an increase of 1.2 rather than 1) and then if I get another, my happiness score goes up from 2.2 to 3.4 (an increase of 1.4).

If this is what is happening, then we can say that happiness is a skewed variable.

One last thought leads to something different. Don’t forget that to measure what is happening here, we are asking people how happy they feel. We can produce a skewed distribution of measurements very easily just by supposing that the feeling of being happy produces that multiplying effect. After all, they were walking to a supermarket: feeling more happy when you arrive is fairly noteworthy.

If that is what is happening, then we can say that our measurement skews the distribution of happiness.

Later on, resources linked to Chapter 9 will show ways to convert a skewed distribution into a more symmetric one.

## Extra activity

*Could be a randomised series of flashcards with multiple choice answers. Could simply be a long table to fill out and box goes green if correct answer is entered. Provide the variable name, and the method, and students pick the type.*

**Identify the variable type for each of the listed variables:**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Method** | **Type of measurement?**  *Interval/categorical/ordinal* |
| Drug trial type | Organised into control and experimental groups | Categorical |
| Arctic ice melt | Measured in centimetres | Interval |
| Class ranking | Students arranged by their academic position in the class | Ordinal |
| Eyewitness recognition | Number of attributes recognised | Interval |
| Weight loss | Measured in kilograms | Interval |
| Scenario type | Participants first experience one scenario, and then another | Categorical |
| Diet | Low-carb, low-fat, Mediterranean | Categorical |
| Memory | Items remembered on a tray | Interval |
| Self-rated stress | Likert scale | Ordinal |
| Degree class | Where a first class degree is higher than a 2:1 degree | Ordinal |