

## Coding data in SPSS:

Here are the biographical details from ten vampires. Your task is to construct an SPSS spreadsheet containing all of these data. This will give you practice at "coding" data in SPSS. (Similar logic also applies to coding data in Excel, although the details are a little different).

Use a separate row for each "case" (each vampire). Each column contains the data for a single variable (an attribute for which you have data). Thus you could have a column labelled "year\_born" which would contain the year in which each vampire was born, like so:

	name	year_born	birthplace	current residence
1	Vlad	1433.00		
2	Terence	1948.00		
3	Daisy	1872.00		
4	Dragos	1546.00		
5	Dracul	1322.00		
6	Chlamidiah	1789.00		
7	Mogranu	1966.00		
8	Wilbur	1973.00		
9	Davina	1967.00		
10	Sharon	1933.00		

So row 1 contains the data for "Vlad". At present, we've entered data for two variables, "name" and "year\_born". A third variable, "birthplace" is ready for data entry. Add further columns so that you can include all of the data on these ten vampires.

Here's a list of variables that you could use:

- Name
- year born
- birthplace
- current residence
- number of years dead
- age when turned into a vampire
- number of victims
- method of killing victims (violent versus non-violent)
- blood tastiness rating
- feeding frequency

most feared death  
blood volume drunk  
current occupation

### Levels of measurement - a topic that will return to haunt you...

Some of these variables are "categorical" in nature, while others represent "quantitative" measurements. "Name", "birthplace", "current residence", "method of killing victims", "most feared death" and "current occupation" are **categorical** or **nominal** variables. For these variables, the values that can be obtained are qualitatively different from each other, and arithmetic cannot be performed on them. For example, for "birthplace", we could in principle have two or more people who come from a particular place (as in this example, where two vampires come from London). So we could count up how many *examples* of each category occur. However we cannot *add* "London" to "Eastbourne" in any meaningful way, because you can't do arithmetic on names.

The other variables in this list are **quantitative** in nature: they are measurements on a scale, and as such you could do arithmetic with them. For example "blood volume drunk" is an amount of blood. As such, it makes sense to perform arithmetic on the various values. We can make statements such as the following: Vlad drinks 7 pints a week and Terence drinks 6, so together they drink 13 pints a week; Vlad drink a pint more a week than Terence; on average, between them they drink 6.5 pints per week; and so on.

Within this list of quantitative variables, you could make a further distinction between variables that are measured on an inherently continuous scale, and variables that are measured on a scale with discrete values. "Blood volume drunk" is a measurement on a continuous scale - although we've measured blood consumption to the nearest pint, in principle one of our vampires could drink any amount of blood, and we could measure this to any level of precision that we wanted because the scale is inherently continuous (e.g. we've measured blood consumption to the nearest pint, but we could have measured it *in principle* to the nearest millilitre, yoctolitre<sup>1</sup> or whatever).

The remaining variables ("year born", "number of years dead", "age when turned into a vampire", "number of victims", "blood tastiness rating" and "feeding frequency") are all quantitative in nature, but measured on scales that are inherently discontinuous. You can feed on 6 people a week, or 5 people a week, but it doesn't really make much sense if you claim to feed on 5.5 people a week. Likewise it makes no sense to say that you feed 3.6 times a week. (Note that if you average together the data measured on these kinds of scales, you will get these kinds of fractions emerging:

---

<sup>1</sup> A yoctolitre is equivalent to a cubic nanometre. It's an extremely tiny amount, corresponding to the portion of tomato ketchup that you get in a nouvelle cuisine restaurant.

the classic example is the statement that the average family has 2.5 children, which literally makes no sense because you can only have whole children).

The important distinction to make is between categorical and quantitative measurements, because these have to be treated in very different ways for the purposes of data analysis.

### **Coding variables in SPSS:**

To set up these variables, open SPSS and click on the "variable view" tab in the bottom left corner of the screen. Initially, variables have scary-looking names like "VAR00001". To create a new variable, simply replace this name with a new one, something that makes sense to you. (SPSS will only accept single-word variable names, so if you want to use more than one word in a variable name, such as "year born", you will need to join the words together with the underscore character. That's the long dash to the left of the "+" character on a PC keyboard.). Then click on the "data view" tab to get to the spreadsheet into which you can enter the data.

While you are in "variable view", you should decide whether each variable is "string" or "numeric". A "string" variable is inherently categorical in nature, such as a person's name or the place in which they were born. It's called a string because each piece of data consists of a string of characters (letters and/or numbers) which will be treated as text. A "numeric" variable is something that can be measured meaningfully in a quantitative fashion, such as "age", "years dead" or "number of victims".

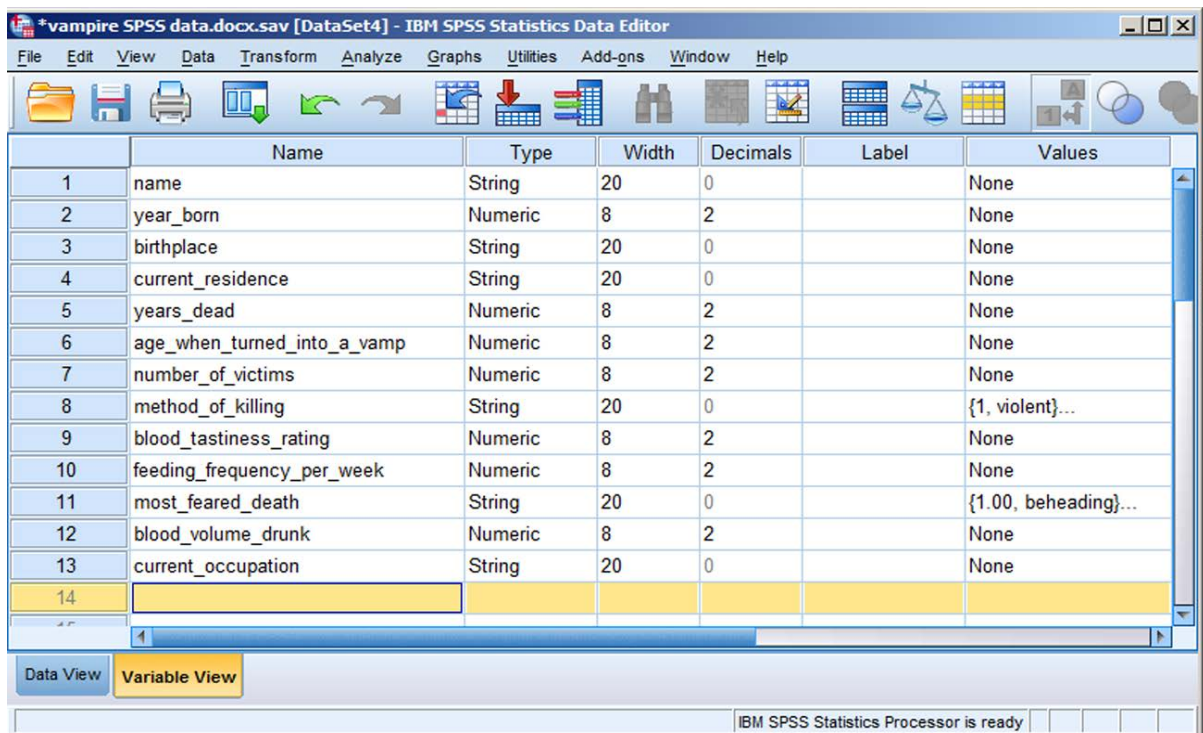
If you don't tell it otherwise, SPSS will try to work out for itself whether a variable is string or numeric. Normally this works OK, but it's best to explicitly tell SPSS whether a variable is string or numeric. Otherwise you can run into problems if you are using numbers as codes for attributes which differ qualitatively (so that the numbers are really names rather than truly numerical values).

Here's an example to make this clear. How do we code "most feared death"? You could laboriously type in each entry, in words. However a quicker way, especially when you have a lot of data to enter, is to use numerical codes to signify each value. Here, I've arbitrarily coded "beheading" as "1", "staking" as "2", "exposure to sunlight" as "3" and "drowning" as "4". Then I can simply type "1", "2", etc. instead of having to type in the words. SPSS will know that "4" represents "drowning", and so on.

The important point to remember is that these numerical codes look like numbers, but they aren't really numbers at all: they are labels. It comes back to the distinction I made earlier between categorical and continuous variables. It really doesn't make any sense to try to do arithmetic with them. You can't add "1" and "2" together to get "3". ("Beheading" + "staking" does not equal "exposure to sunlight"). If you tell SPSS

that this variable is a string, SPSS will not allow you to fall into the mistake of trying to do arithmetic with this column of data. if you don't tell SPSS that it's a string variable, SPSS will think these labels really are numbers, and will (stupidly) allow you to treat them as such. So if you ask for descriptive statistics for this variable, SPSS will unintelligently tell you that the average "most feared death" is 2.4 - which is a completely meaningless statement!

So, in short, it's good practice to make sure that each variable is designated correctly as a string variable or as a numerical value, to ensure that SPSS treats it in an appropriate fashion. Note that the default width of a string in SPSS is eight characters: that's often not enough (for example, Davina's occupation would end up as "TV prese" instead of "TV presenter") so I simply changed the length of the string variables to 20 characters in the "Width" column in "variable view".



### **Biographical details of the vampires:**

**Vlad** was born in Brzoi, in the Transylvanian Alps in 1433. He has been a member of the undead since he was 41. He currently lives in Sittingbourne, where he works as a TV repair-man. He has killed 3000 people during his time as a vampire, and his preferred method of slaughter is simply to jump on his victim and rip their throat out. He drinks about three times a week, depending on how much TV repair business comes his way. He likes to take about seven pints each time. On a 1-7 scale, where 1 is "highly unpleasant" and 7 is "highly pleasant", Vlad rates the tastiness of human blood as "6". He most fears death by staking.

**Terence** was born in Barnsley in 1948. He became a vampire when he was 22. He now lives in Surbiton, working as a bus driver. His current kill total is 2,160. He likes to decimate whole bus-loads of passengers whenever he's working a night shift (i.e. about once a week, when he drinks forty pints in one feeding frenzy). Like Vlad, throat-ripping is his preferred method of despatch. He too would give human blood a rating of "6" on the tastiness scale. He most fears death by beheading.

**Daisy** was born in London in 1872. She became a vampire when she was 46. She now lives in Eastbourne, where she is a care worker. She has killed 800 people, mainly old age pensioners: she drugs them first and then drinks their blood while they are unconscious. Daisy likes to feed regularly, consuming two pints of blood once a week if possible. She would give human blood a rating of "3" on the tastiness scale. She most fears death by staking.

**Dragos** was born in Fioresti, Moldova, in 1546, and was turned into a vampire when he was 32. He lives in Ibiza, working as a tour guide. He has killed 8000 people. These days his preferred method is to lock the disco doors and then slaughter everyone inside, drinking about seventy pints of blood at a time. (It's a small disco). He gives human blood a rating of "7" on the tastiness scale. He drinks four times a week. He most fears death by exposure to sunlight, which is a little ironic given where he chooses to live.

**Dracul** was born in Surbiton in 1322, and was turned into a vampire when he was 56. He now lives in a remote castle in darkest Edgbaston, and works as an IT consultant. He has killed 8,652 people. Usually he talks to them about computers until they fall asleep, and then he bites them. He likes to drink six pints at a time, and gives human blood a rating of "5" on the tastiness scale. he feeds three times a week. He most fears death by staking.

**Chlamidiah** was born in Petrozani, in Romania, in 1789. She was turned into a vampire when she was 23. She has killed 7,400 people. She currently lives in Burwash, Kent, where she works as one of those slightly annoying people who stop you in the street and try to make you sign up to give charitable donations to a charity you've never heard of, for the rest of your life. She procures victims by obtaining their name and address in the guise of doing her job. She then turns up at their house

later that day and savagely attacks them, drinking nine pints in one go. She feeds once a week, and rates human blood as "6" on the tastiness scale. She most fears death by drowning.

**Mogranu** was born in Littlehampton in 1966, and has been a vampire since he was 35. He works in the bedding department of Debenhams department store, Crawley. He has killed only 1000 people, usually by luring them into trying out a bed in his store and then biting them once they have fallen asleep. He likes to drink five pints at a time, three times a week. He rates human blood as "5" on the tastiness scale. he most fears death by exposure to sunlight.

**Wilbur** was born in Bognor Regis, in 1973. He has been a vampire since he was 19. He works as a salesman in PC World in Bognor Regis. As you might expect from his occupation, he is the most evil of the vampires in this list. He has killed 10,000 people, attacking them savagely as they wander back to their car after purchasing a computer from his store. (This also conveniently reduces the number of "returns" to the store). He normally feeds seven times a week. He rates human blood as a "7" on the tastiness scale, and always drains his victim of the full nine pints. He most fears death by beheading.

**Davina** was born in London in 1967, and has been a vampire since she was 22. She works as a TV presenter, mainly in London. Her method of killing is to chat to the victim until they fall asleep, and then bite them. Although she likes to drink five times a week, she has only actually killed 653 people. Most of the time she only takes two pints from each victim; this doesn't kill them but merely turns them into vampires too, so she has lots of friends and followers. She rates human blood as "4" on the tastiness scale. Her most feared fate is death by exposure to sunlight.

**Sharon** was born in Chelsea in 1933. She has been a vampire since she was 49. She works in a KFC in Chelmsford. She has killed 1056 people. She drinks every day, and takes nine pints from each victim. (Then she falls over into the gutter and is taken home by her vampire friends, mumbling "never again"). Her preferred method of obtaining victims is to sell them a drugged "Bucket of Chicken Bits" after the pubs close, and then follow them home. That way, she can feed from an entire unconscious family in one sitting. She rates human blood as "7" on the tastiness scale. She most fears death by exposure to sunlight.