

## Problem sheet 4 :: experimental design



Tea is better than coffee. Let's just get that out of the way before we start. ~~Many Most~~ ALL British people think this. Even those who say the exact opposite agree really, they're just trying to be provocative and confrontational due to consuming too much caffeine. Yes, it may look like pretty much every other building you come across these days is a Starbucks, but tea is still more popular. Tea doesn't need a global empire shoving it in people's faces. Not after the last one, anyway.

The above paragraph is obviously exaggerated for comic effect (but only slightly), but it can't be underestimated how important tea is to many people in Britain (and beyond of course). And because it's so important, how it's made becomes a serious issue. How long you leave the tea to brew, whether to put sugar in, what type of tea to use, and perhaps the biggest cause of disputes: if you put milk in your tea (which you should) do you put it in the cup before or after the boiling water?

Many arguments have been had about this. If anything is going to kick off another civil war in the UK, it is probably going to be this. What most people don't know is that scientists settled this debate. Supposedly.

To test the recipe for the perfect cup of tea put forward in 1946 by George Orwell himself, Dr Stapley of Loughborough University established that putting the milk in after the boiling water is incorrect, as it causes the milk to heat unevenly (as opposed to pouring the water on top of it). This uneven heating of the milk causes the proteins in it to denature, meaning they lose their structure and "clump", affecting the taste and contributing to that skin you get on the top. So when someone says they can tell if you put the milk in first or second in the tea you've made for them just by tasting it, turns out they probably can.

Inspired from <https://www.theguardian.com/science/brain-flapping/2014/oct/03/how-to-make-tea-science-milk-first>

## Questions

Four questions, each worth two marks with two marks for attendance.

We want to reproduce the experiment by Dr Stapley, i.e. we want to find out which method is the best between adding the milk before or after the boiling water. By “best” we mean that the tea will have a better taste.

Q1. Create an experimental design to test this, i.e. (a) what are your independent and dependent variables and how to you gather the feedback from the participants (what question do you ask; (b) are you performing a between or within-subject experiment and why; (c) what is the task, i.e. what are you asking the participants to do; (d) are you using any kind of counterbalancing?; (e) how many participants do you think you will need and how many trial per participants you will do?

**Note that many answers where possible as the goal was to make you think about designing an experiment and the problems we can encounter. If you have proposed something different you will still have full points.**

### Variables:

One independent variable with two levels = Milk timing (before or after)

One ordinal dependent variable = tastiness score. You can ask the participants to reply to a questionnaire question: on a scale from 1 to 5 how much would you rate the tastiness of the cup you just drunk? 1 not at all 2 not really 3 undecided 4 somewhat 5 very

### Task:

Make participants drink a cup of tea, and then make them rate on a scale from 1 (not tasty) to 5 (very tasty) the tastiness of the cup they have just drunk. Options: make them clean their mouth in between etc.

### Design type and participants:

There are different ways of doing, possibly the best one is to do a between-subject where you have many participants testing one cup each. This will avoid possible biases, e.g. imagine drinking one cup give you a taste in the mouth that subsequently biases other trial you do after.

Another way of doing would be within-subject in which participants have to test a randomly generated cup every other morning. In such case you would need less participants (e.g. 12). You can repeat the experiment over many trial (many days) in order to increase your chance to find signal in the noise.

### Counterbalancing:

If you opt for a within subject study, you will need to counterbalance but you should not use something that the participants can easily predict.

**In all case you will have to do a blind test! Because if the participants see you when preparing the test, they may be biased! You might also need to make them taste the drink blinded (in case the milk timing has an effect of the overall drink coloration).**

### Trial and participants:

If you are using a between subject design you will need twice more participants. Rule for a simple experiment like that in within is around 12-16 participants. Can you design allow you to make multiple trials (=make the task multiple times)? If so do it, it will make the data more reliable.

Q2. Let's suppose you have asked participants to answer a questionnaire using a Likert scale for each cup of tea drunk. The scale is from 1 (not tasty) to 5 (very tasty). What kind of distribution you expect your data to follow: normal or skewed? Why?

It is more likely that your data will be skewed as it is mostly the case with Likert data (especially if you take participants living in UK who will likely put high grades anyway because they like tea!)

Q3. You run your normality test and observe that assumption of normality is rejected for your Likert data, so your data is skewed. What statistical test will you be using to compare your two conditions (milk before, milk after)?

Likert scales are ordinal data but the beauty of them is that we can consider them as continuous variable. So we can use the graph at the end of this document. If you are using a between-subject = Mann-Whitney Test. If you are using a within-subject design = Wilcoxon Signed rank test

Q4. You are wondering what the effect of would be pouring the milk and the boiling water *at the same time*. You run the same experiment again but this time with 3 conditions: milk before, milk after or milk at-the-same-time. What do you need to change in your experimental design and why? What statistical test will you be using to compare your three conditions?

You need to add a 3<sup>rd</sup> condition so if you do a within subjects design you will need to counterbalance the conditions with a Latin square, i.e. if you have 12 participants, 4 of the participants start with 'milk before' then 'milk after' then 'together'; 4 of the participants start with 'together' then 'milk before' then 'milk after'; 4 of the participants start with 'milk after' then 'together' then 'milk before'. If you are using a between-subject = Kruskal-Wallis test. If you are using a within-subject design = Friedman test

## Note on pairwise experiment

Some of you proposed a very valid way to do the experiment: make the participants taste the two cup of tea and tell which one they preferred. These types of experiment are called pairwise experiments and we won't cover them in this unit. However, it is good to know that they are quite used when researchers need to retrieve a subjective information but still have it as a metric. So, if you want to compare 3 conditions, you just need to make sure that the participants compared them all. Following that you have a table of preferences which you can input into a Bradley-Terry model, which is another kind of statistical test handling this type of data. The Bradley-Terry model can create a metric from the comparison data and can thus allow comparing the metric of the different condition. If you want to see an example of how it can be used look at this [http://www.anneroudaut.fr/papers/roudaut\\_chi17.pdf](http://www.anneroudaut.fr/papers/roudaut_chi17.pdf)

## Useful

