Understanding the data from a two-group experiment

The scenario

In the task below, imagine that you are testing some new drugs to see if they are effective in reducing anxiety. In each experiment, half your participants are given a drug: the Drug group. The remaining participants in each experiment are given a placebo pill: the Control group. After a month, each participant rates their anxiety on a scale of 1-10. So the independent variable (IV) is whether the participants received a drug or a placebo. The dependent variable (DV) is their anxiety ratings at the end of a month.

**Each graph below depicts** **two experiments**, Experiment 1 in the left panel and Experiment 2 in the right panel. Both Experiments test the effect of a new drug on anxiety. On the graphs, **the black squares** **indicate the mean (average) anxiety ratings for each group in each experiment**. **Each grey dot indicates the anxiety rating for a *single* participant**.

Your task

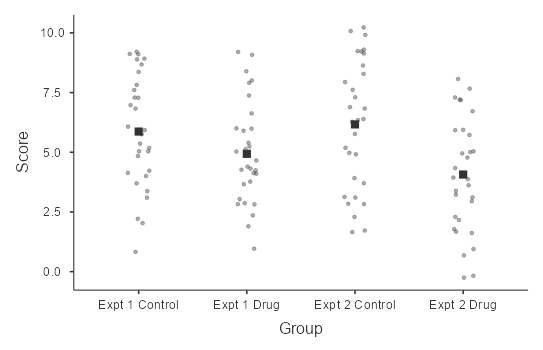
For each of the five graphs A-E below, decide which of Experiment 1 (on the left) or Experiment 2 (on the right) seems to provide the strongest evidence that a drug is effective in reducing anxiety. In other words, is Experiment 1 the most convincing that the Drug group gave lower anxiety ratings than the Control group? Or is Experiment 2 more convincing?

1. First, go through the graphs quickly and write down a guess as to which experiment provides the strongest evidence of a difference – see what your intuition tells you. Remember that the grey dots are each an individual’s anxiety rating and the black squares are the average for each group.
2. Second, have a look at the **Principles** that appear after the Graphs. Then think about the guesses you made. Revise your thoughts if necessary.

**Once you have made your final judgment about which Experiment is most convincing, submit a short voice note for each of Graphs A-E explaining your reasoning.**

*(Only a statistical test (e.g., a t-test in Jamovi) will tell you whether there is a ‘significant difference’ between the two groups. The task you are about to undertake is only to give you a sense of the factors that lead to a significant difference.)*

Graph A

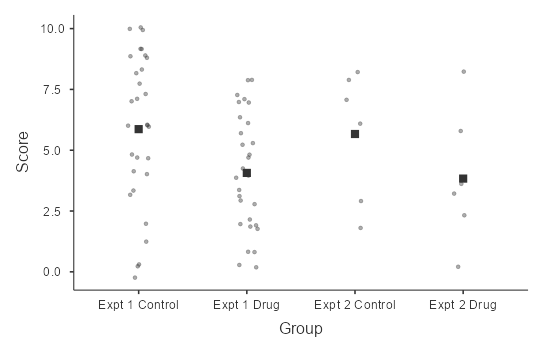


Answer for ChatGPT:

The mean difference is greater for Experiment 2 than Experiment 1, so Experiment 2 provides stronger support for the idea that the drug is effective. The two experiments are similar in terms of the overlap in scores, the range in scores and the number of participants tested. So the mean score will have an important impact on the results of the analysis.

Decoder (Expt 2vs6 – mean; Jamovi: expt 1 is N.S., Expt 2 is significant)

Graph B

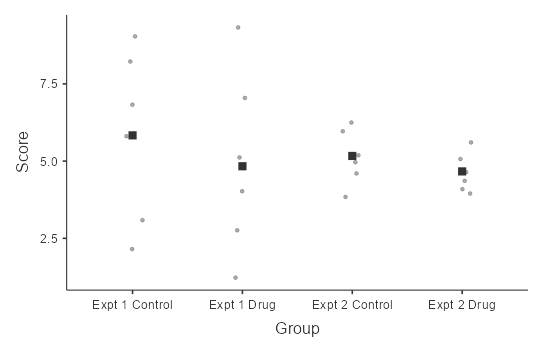


Answer for ChatGPT:

The number of participants tested in Experiment 1 is much higher than in Experiment 2. So the evidence for a benefit of the Drug in Experiment 2 is stronger than in Experiment 1. Other factors are roughly similar: the variance/range is similar, the overlap in scores is similar and the difference in means is similar.

Decoder (Expt 6vs5; Jamovi: 6 is sig, 5 is marginal)

Graph C

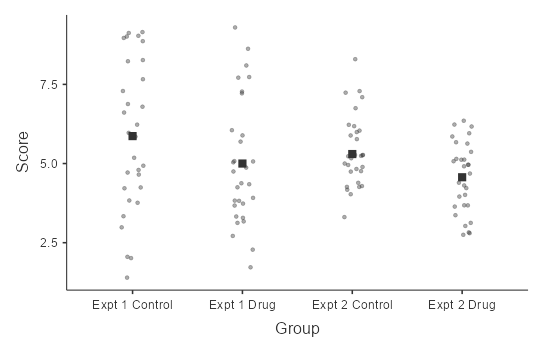


Answer for ChatGPT

There is slightly more evidence for a difference in Experiment 2, but the two Experiments are similarly weak. There is a larger mean difference in Experiment 1. But the variance/range is much smaller in Experiment 2. In both experiments, there is large overlap in scores between the two groups. Lastly, only a small number of participants were tested in both experiments.

Experiment 1vs3 (both non-significant – large overlap and small N)

Graph D

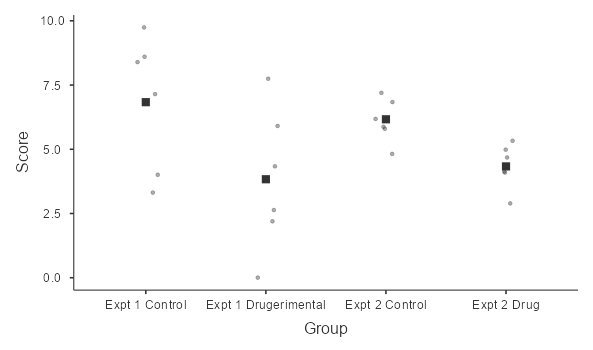


Answer for ChatGPT

There is less overlap between the two groups in Experiment 2. The mean difference is a bit smaller in Experiment 2 than Experiment 1, but the data overlap more in Experiment 1, which is more important. The variance is also smaller in Experiment 2. The number of participants tested is the same in the two experiments, so that isn’t a factor.

Decoder: Experiment 2 vs 4 (1 is NS, 2 is p=0.01)

Graph E



Answer for ChatGPT:

Experiment 2 provides better evidence for the benefit of the drug tested. This is because there is less overlap between the groups in Experiment 2 than Experiment 1. Relatedly, the variance/range is much smaller in Experiment 2. The mean difference is greater in Experiment 1, but the overlap is more important. There are only a few participants tested here – the same number in each experiment – it would be much better to test more people.

Expt 5 vs 7 (1 is non-sig; 2 is sig)

**The second part: the Principles**

*Keep these principles in mind when you have another look at the graphs before making your voice note.*

1. **Difference in means**: If the mean anxiety in the Drug group is much lower than the mean anxiety in the Control group, that suggests (all else being equal) that the drug might be effective in reducing anxiety.
2. **Overlap in scores for the two samples**: This is more important than the difference in means. Look at the extent to which the *individual* scores for the two groups overlap. Even if the *mean* anxiety scores for the two groups are quite different, the two groups might look very similar when you consider all of the individual scores as a cluster. Your task here is to consider all of the data together and judge whether the two groups of participants look different in their ratings.
3. **Variance/range of scores in each group**: This is related to the overlap in the scores. For a given difference in the means, if the individual scores are tightly clustered around the mean (most participants give roughly the same scores), then there will be less overlap between the groups and there is more likely to be a difference. Smaller range/variance means, all else being equal, better evidence for a difference.
4. **Size of the sample/number of participants**: All else being equal, the more participants you test (the more grey dots there are), the more evidence you’ll have for any differences between the groups.