

sign test

paired t-tests revision

motivating example:

- does biochemical substance have an inhibitive effect on muscular growth?

10 rats:

- one hind leg muscle was injected with the substance
- the other leg had a placebo
- 6 months and weighed the muscles in grams



in r:

- make a dataframe, col 1 is the substance measures, col 2 is the placebo
- use mutate to add a column where we have the difference between placebo and substance
- then we extract the mean of the difference and the standard deviation
- we then run a one sample t-test with alternative being "greater" as we want to show if the substance has a inhibitive effect (makes muscles smaller, or if placebo is bigger)



formally:



if we wanted to look at the critical value to which we would compare our t-statistic, proposing that the distribution follows a t-distribution, we would look to see the value of t on a t-distribution where the area under the curve to its left would be equal to 0.95. here, this would be 1.833, and our test statistic was larger than that.



in terms of the rejection region, we can map back to the original scale and write the rejection region in terms of the sample mean. we take our observed mean and multiple it by $c \times \text{standard error}$, where c is the critical value



here, this means we would reject the null if the difference was greater than 0.19 grams (this is enough to be significant at the 5% level.)

in terms of a confidence interval - confidence interval refers to the range of plausible values that the proposed population parameter would be in, to NOT reject the null.



so here, we got our one-sided confidence interval to be $(0.06, \infty)$. now because our null hypothesised population mean difference is 0, it is not inside of this range. so we reject null.

checking for normality

one of our assumptions when conducting our t-tests is that the samples are normally distributed / data is sampled from a normal population. if we have a large enough sample size, then the normality assumption is not as important as we rely on the central limit theorem to ensure our test statistic at least approximately follows a t-distribution.

however, in small samples, it might be difficult to tell whether or not our sample comes from a normal population.

we check this normality assumption using:

- Q-Q plots (visually assessing normality) - plotting actual data against what the data would look like under a normal distribution
- Box plots (looking for symmetry) - and also histograms, but just looking for symmetry

checking for normality: box plots

we want to check if our median (the line in the middle) is in the middle (roughly) between the second and third quartile. also check if the whiskers are roughly the same length.