

Statistics Notes for Master's Thesis at ZHAW (in progress)

Jürgen Degenfellner

October 4, 2023

1 General statistical considerations

Note: The following notes are by no means to be considered as conclusive, will be **expanded** over time as problems come along, and are based, among others, on the following publications:

- *Moving to a World Beyond “ $p < 0.05$ ”* [WSL19]
This paper is quite readable, and in particular, points 1 and 2 (Don't Say "Statistically Significant") are relevant. I recommend reading this from time to time when doing statistics. Some Don'ts from point 1 (in the paper):
 1. Don't base your conclusions solely on whether an association or effect was found to be “statistically significant” (i.e., the p-value passed some **arbitrary** threshold such as $p < 0.05$).
 2. Don't believe that an association or effect exists just because it was statistically significant.
 3. Don't believe that an association or effect is absent just because it was not statistically significant.
 4. Don't believe that your p-value gives the probability that chance alone produced the observed association or effect or the probability that your test hypothesis is true.
 5. Don't conclude anything about scientific or practical importance based on statistical significance (or lack thereof).
- *A Dirty Dozen: Twelve P-Value Misconceptions* [Goo08]
In this paper, some misconceptions about the p-value are not so obvious and can surprise even experienced researchers. This is a definite reading recommendation.

1. If $P=.05$, the null hypothesis has only a 5% chance of being true.
2. A nonsignificant difference (e.g., $P > .05$) means there is no difference between groups.
3. A statistically significant finding is clinically important.
4. Studies with P values on opposite sides of .05 are conflicting.
5. Studies with the same P value provide the same evidence against the null hypothesis.
6. $P=.05$ means that we have observed data that would occur only 5% of the time under the null hypothesis.
7. $P=.05$ and $P < .05$ mean the same thing.
8. P values are properly written as inequalities (e.g., “ $P < .02$ ” when $P=.015$)
9. $P=.05$ means that if you reject the null hypothesis, the probability of a type I error is only 5%.
10. With a $P=.05$ threshold for significance, the chance of a type I error will be 5%.
11. You should use a one-sided P value when you don’t care about a result in one direction, or a difference in that direction is impossible.
12. A scientific conclusion or treatment policy should be based on whether or not the P value is significant.

From the misconceptions listed, probably numbers 2, 7, 9 and 10 are most surprising.

- *Adjusting for multiple testing—when and how?* [BL01]
There many papers on this topic, a nice overview is given by this paper. You can read it, but the more important thing is to remember to use correction for multiple p-values. Producing many p-values of independent tests results in the fact that approximately 5% of them are smaller than the **arbitrary** threshold of 0.05. You can use R (function *p.adjust*) or online tools (MultipleTesting.com [MWG21]) for the correction. The most important thing to remember is:
Avoid p-hacking at any cost - it is the cardinal sin of statistics. Results of implicit or explicit hypothesis tests, that were not named in the sample size calculation/analysis plan should be interpreted as *exploratory*.
- *Bayesian estimation supersedes the t test* [Kru13]
John K. Kruschke discusses a Bayesian approach to data analysis as an

alternative to traditional t-tests and Null Hypothesis Significance Testing (NHST) in general. He argues that Bayesian estimation methods provide richer information and are more intuitive compared to frequentist methods like the t-test.

In this paper, Kruschke introduces a Bayesian method for comparing two groups. Unlike the t-test, which only tests for a difference in means between groups, the Bayesian method provides a full distribution of credible values for the difference in means, variances, and effect sizes, making the results more interpretable and informative.

Kruschke also emphasizes that Bayesian methods allow for the incorporation of prior information, which can be beneficial in many real-world situations where some prior knowledge exists. Moreover, he provides examples and software code to encourage the adoption of Bayesian methods in the practice of data analysis.

Overall, Kruschke's paper advocates for a shift from traditional frequentist methods like the t-test to Bayesian methods, which he argues are more intuitive, informative, and adaptable to real-world circumstances.

Admittedly, this paper is rather long and comparatively technical, but worth the read if one has the time. An eye-opener was the dependence of the p-value on sampling intentions. The Bayesian framework is at least a valuable addition (if not replacement of) to the classical statistical framework of NHST.

2 Descriptives and visualisations

2.1 Boxplot vs. violin plots

Apart from the usual descriptives (mean, median, standard deviation, interquartile range....), one should also understand the distributions of the data by looking at the raw values. This can be done, for instance, with a nice looking violin plot (figure 2) which looks more informative compared to a basic boxplot (figure 1).

Especially when dealing with small sample sizes, it makes sense to present the data thoroughly since too much summarization in form of a boxplot could lead to a false impression of the data.

2.2 Histograms vs. density plots plus boxplots

Especially for smaller sample sizes, a more informative boxplot is advised, see figures 3 and 4. This gives a richer impression of the data. Histograms and

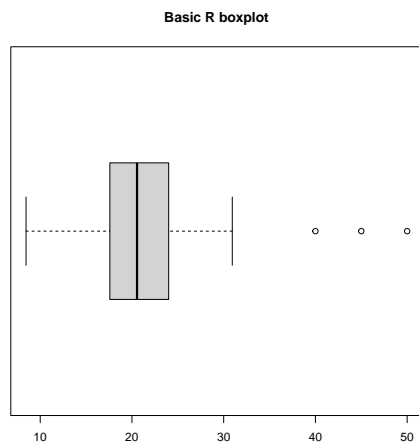


Figure 1: Boxplot

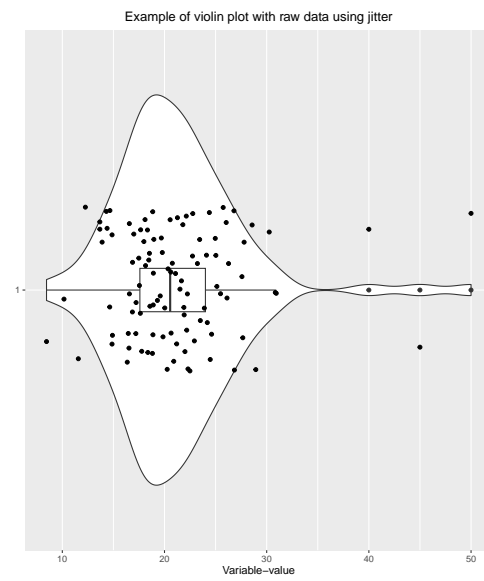


Figure 2: Violin Plot

boxplots for small sample sizes are problematic.

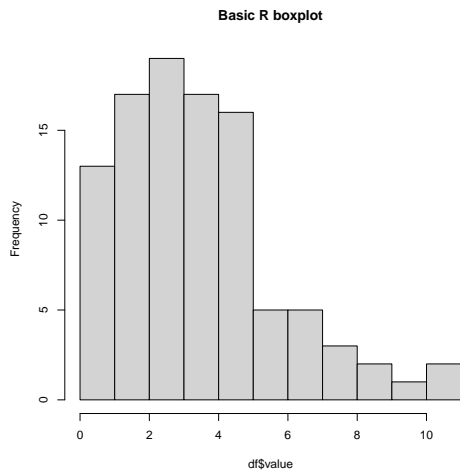


Figure 3: Histogram

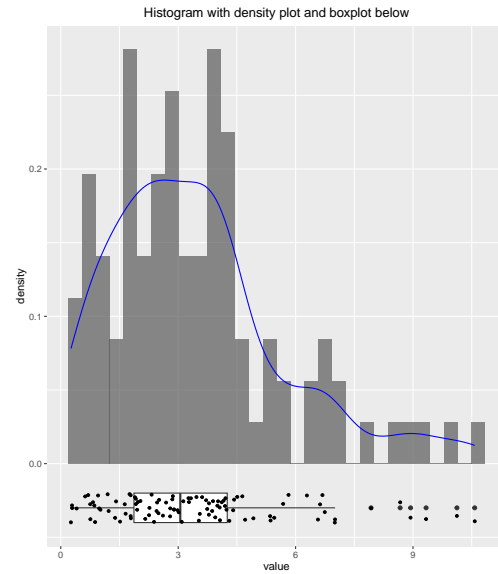


Figure 4: Violin Plot

References

- [BL01] R Bender and S Lange. Adjusting for multiple testing—when and how? *J Clin Epidemiol*, 54(4):343–349, Apr 2001.
- [Goo08] Steven Goodman. A dirty dozen: Twelve p-value misconceptions. *Seminars in Hematology*, 45(3):135–140, 2008. Interpretation of Quantitative Research.
- [Kru13] J.K. Kruschke. Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2):573–603, May 2013. Epub 2012 Jul 9.
- [MWG21] Otilia Menyhart, Boglárka Weltz, and Balázs Györfly. Multipletesting.com: A tool for life science researchers for multiple hypothesis testing correction. *PLOS ONE*, 16(6):1–12, 06 2021.
- [WSL19] Ronald L. Wasserstein, Allen L. Schirm, and Nicole A. Lazar. Moving to a world beyond “ $p < 0.05$ ”. *The American Statistician*, 73(sup1):1–19, 2019.