P values

Peter Geelan-Small, Stats Central

28/02/2020

P values

- What are they?
- Where are they used?
- Where do they come from?
- What do they mean?
- How are they *misused*?

Reporting research findings

Zaslavsky, K. et al. 2019. SHANK2 mutations associated with autism spectrum disorder cause hyperconnectivity of human neurons. *Nature Neuroscience* 22:556-564.

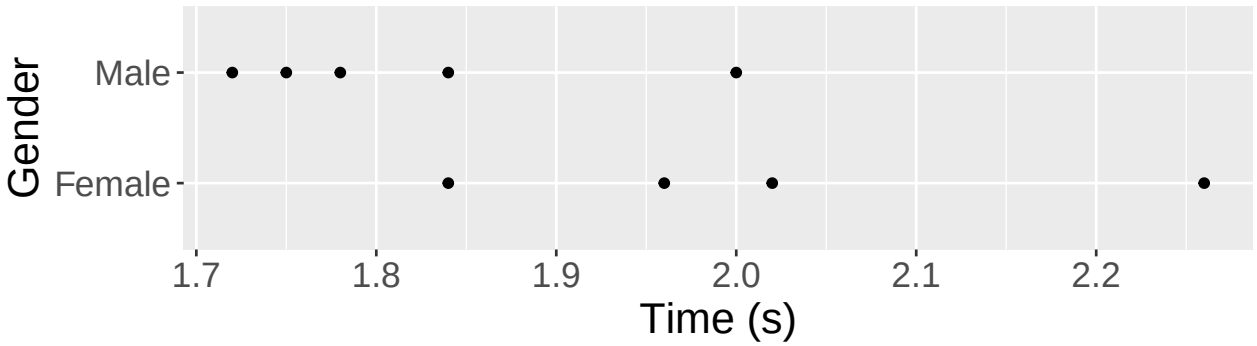
- Figure 4a, p. 6 of 19
- Supplementary Table 10, p. 45 of 55
- People report their research results using:
  - graphs
  - numerical estimates (with estimates of uncertainty like standard errors)
  - P values
- Search for "significant"
- *You'll need to use P values (correctly!) in your report*

What's the idea behind a P value?

Example

To investigate effects of alcohol on young adults, reaction times of nine volunteers to a prompt on a computer screen were measured. Times were precisely measured by computer software under the same conditions for all volunteers. All volunteers had zero blood alcohol readings. There were four females and five males. The data are shown below. (Source: MacGillivray, H et al. (2013), pp. 130-132)

Gender						Mean
Female	1.96	1.84	2.02	2.26		2.020
Male	1.75	1.78	1.72	2.00	1.84	1.818



What's the idea behind a P value?

Question: Is there a difference between female and male reaction time?

What can we use as a typical value of reaction time? The mean!

1. We assume female and male reaction time is the same (Null hypothesis).
2. Shuffle the female and male labels on the data values to get all possible combinations.
3. Calculate the mean for the female and male groups in each combination.
4. Find the difference between the female and male means for each combination.
5. Look at the pattern of values of all these differences.
6. See how unusual the difference is that we got from our sample data.

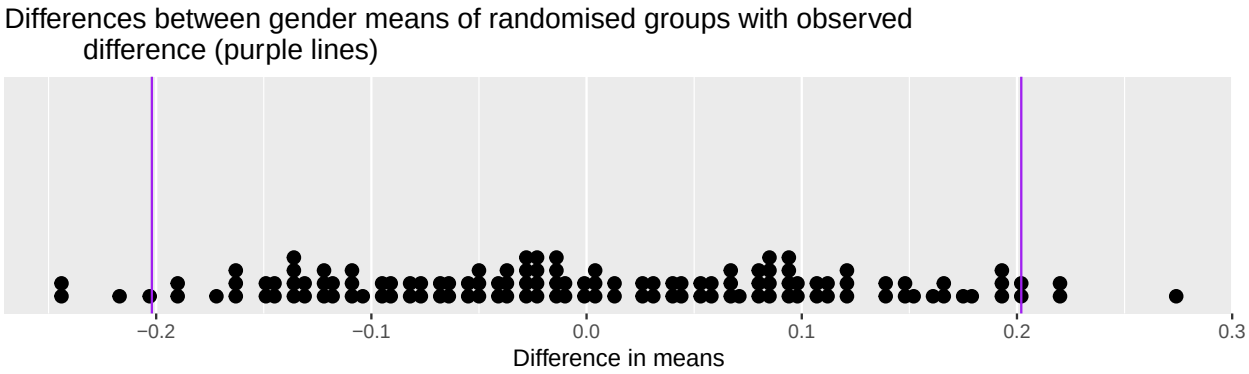
What's the idea behind a P value?

For example, two sets of shuffled labels give these new groupings:

Time	1.96	1.84	2.02	2.26	1.75	1.78	1.72	2.00	1.84
Original	F	F	F	F	M	M	M	M	M
Shuffled	F	F	M	M	M	M	M	F	F
Shuffled	M	F	F	F	M	M	M	F	M

What's the idea behind a P value?

Observed difference in means from sample (Male - Female) = 2.020 - 1.818 = 0.202



- How unusual is our sample?
- Is there some number that describes how unusual our sample difference is?

What's the idea behind a P value?

Is the average reaction time of females and males different?

- How many of the differences are greater than or equal to 0.202?

There are 5 differences greater than or equal to what we got in our sample.

- Allowing for the other end of the scale, how many differences are less than or equal to -0.202 (i.e. same size but other direction)?

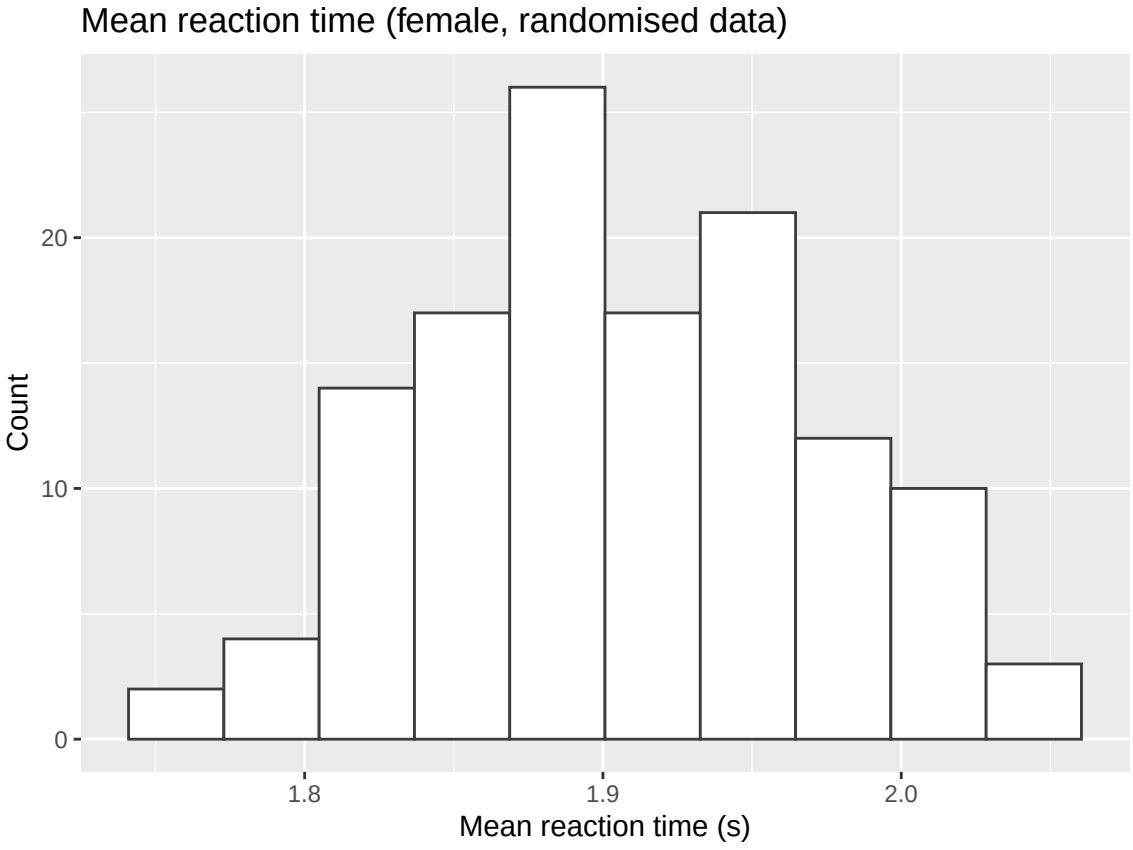
There are 4 differences greater than or equal in size but *negative* compared to what we got in our sample.

What is the proportion of differences that are at least as large in size as we got?

$(5 + 4)/126 = 0.071$

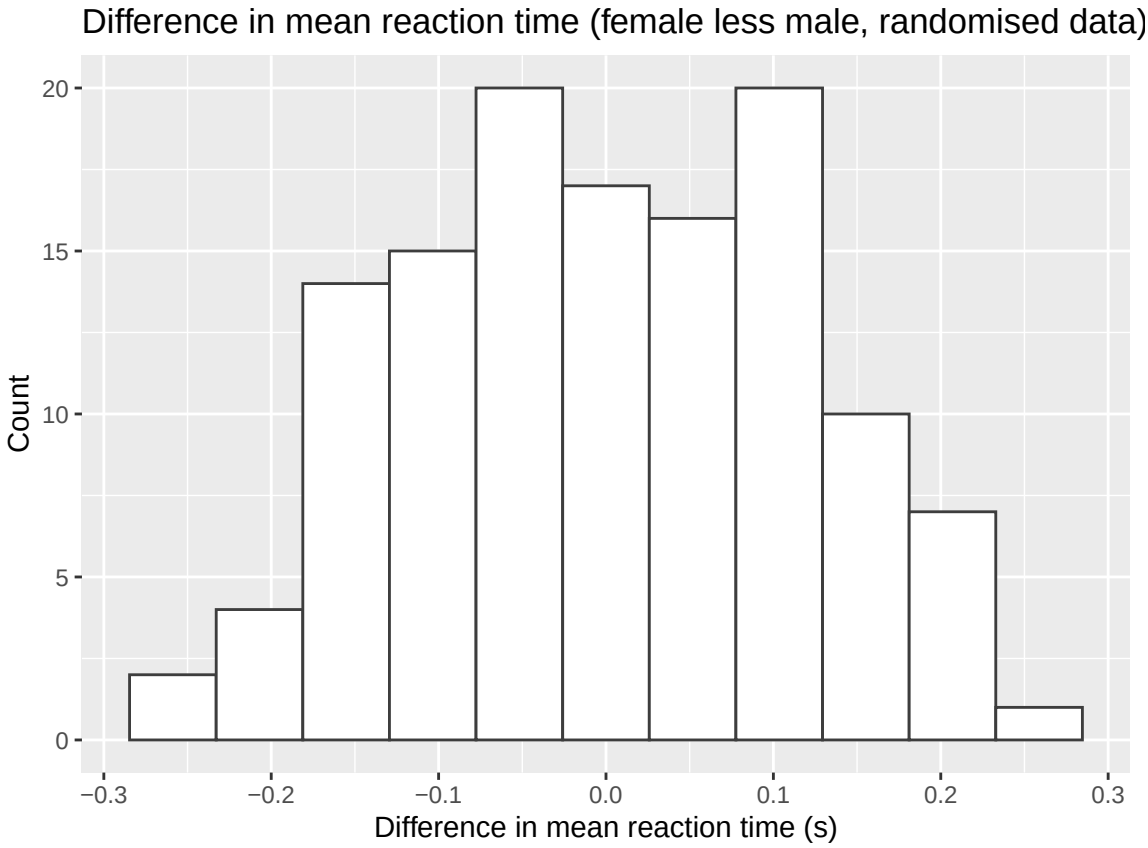
Are the group or treatment means different?

- With larger samples, the number of combinations increases very quickly.
- Look at the pattern of means for female reaction time (randomised data).
- Means often come in a fairly *symmetric* pattern



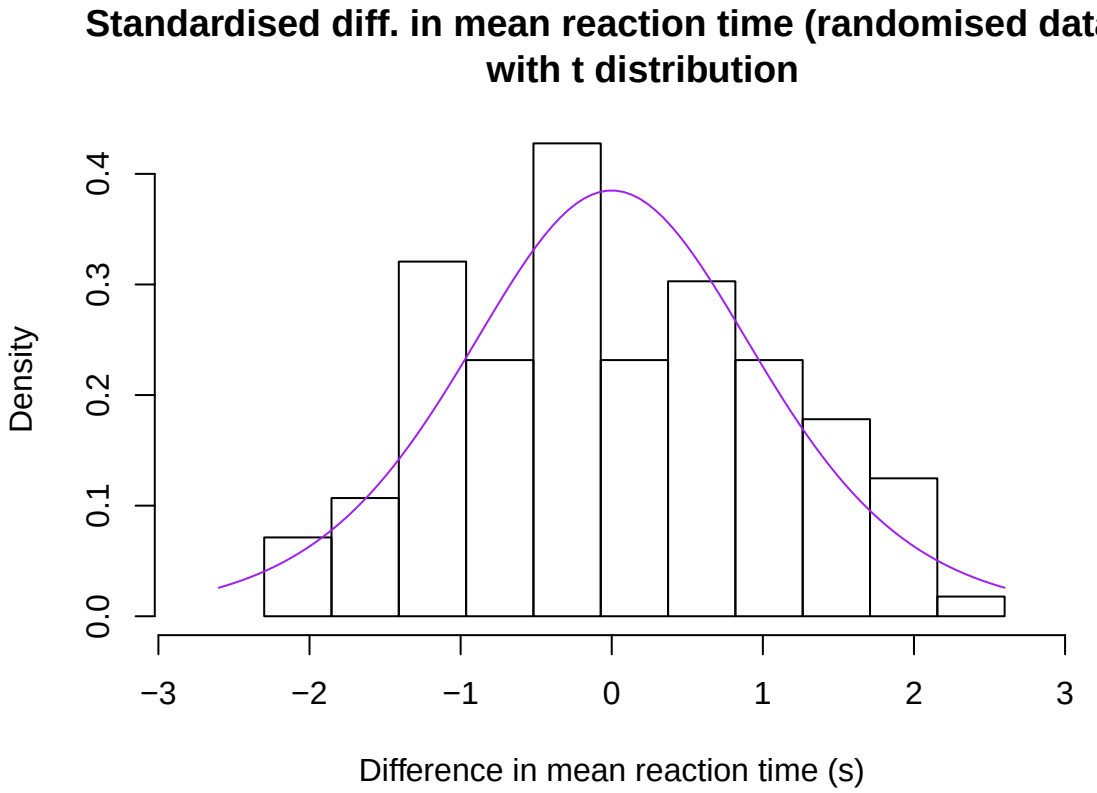
Are the group or treatment means different?

- Differences of means also often come in a fairly *symmetric* pattern



Are the group or treatment means different?

- Because means often come in a symmetric pattern, we can use another method to answer our question.
- We can use a ready-made pattern: T distribution



Are the group or treatment means different?

If we use a T distribution to fit to the patttern of means, we can use a T test to answer our question.

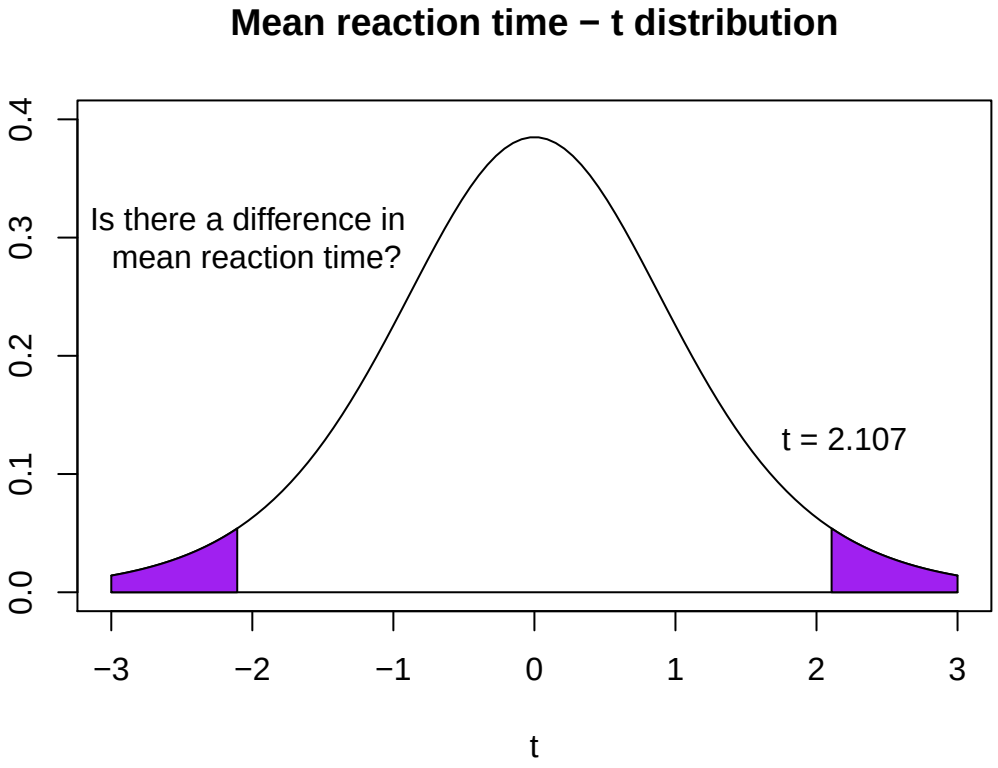
```
##
##  Two Sample t-test
##
## data:  Time by Gender
## t = 2.1077, df = 7, p-value = 0.07304
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.02462358  0.42862358
## sample estimates:
## mean in group Female   mean in group Male
##                2.020                1.818
```

$P = 0.073$

Very similar to what we got using the "shuffling"technique

Are the group or treatment means different?

"P = 0.073" means that the chance of getting the data that we got or a larger difference is 0.073.



P value - meaning

A P value is the chance of getting the data we got or something more extreme on the basis that we assume the null hypothesis is true.

A P value is **not**:

- the probability that the null hypothesis is true
- the probability that our results were due to chance

A small P value is **not**

- evidence that we've made an important finding - statistical significance is not the same as scientific or clinical importance

P values - a topic of debate

Google: American Statistical Society P value

New England Journal of Medicine,

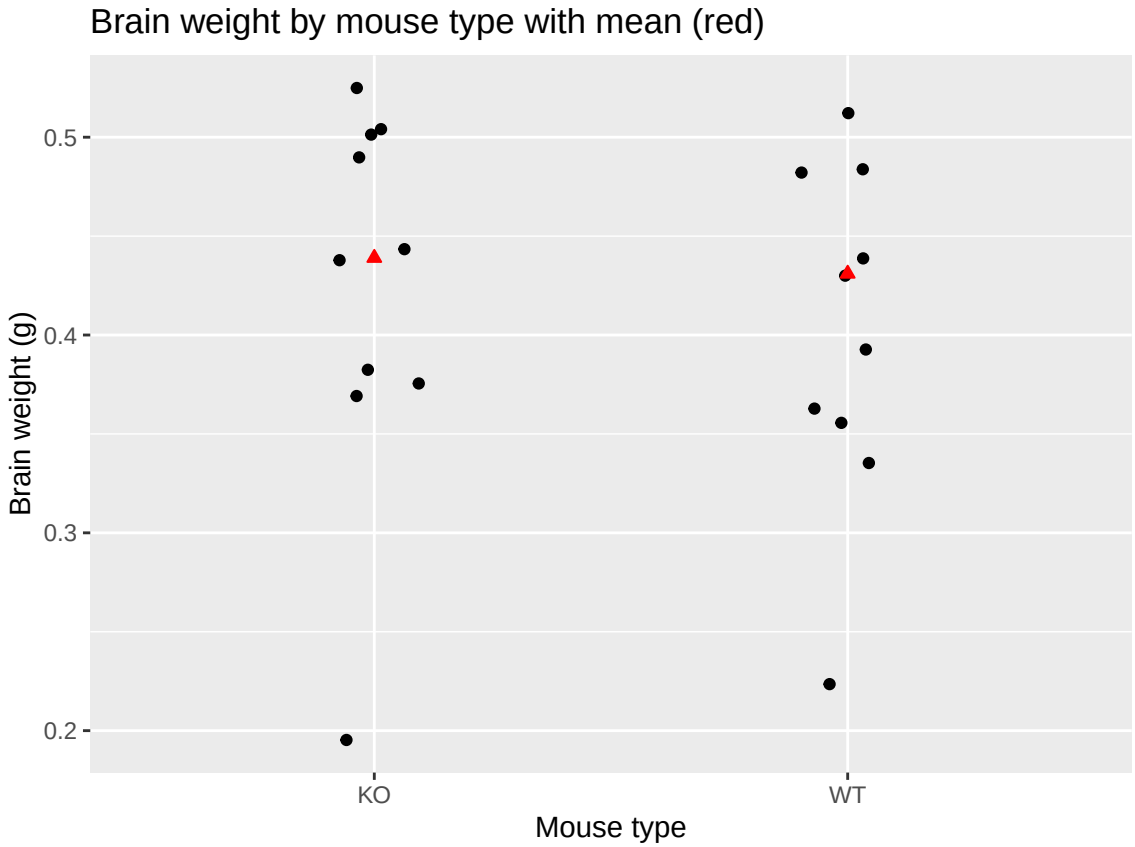
New Guidelines for Statistical Reporting in the Journal

https://www.nejm.org/doi/full/10.1056/NEJMe1906559

- P values give no information about the size of an effect
- P values *and* effect sizes with confidence intervals should *both* be reported

P values - Wrap-up questions

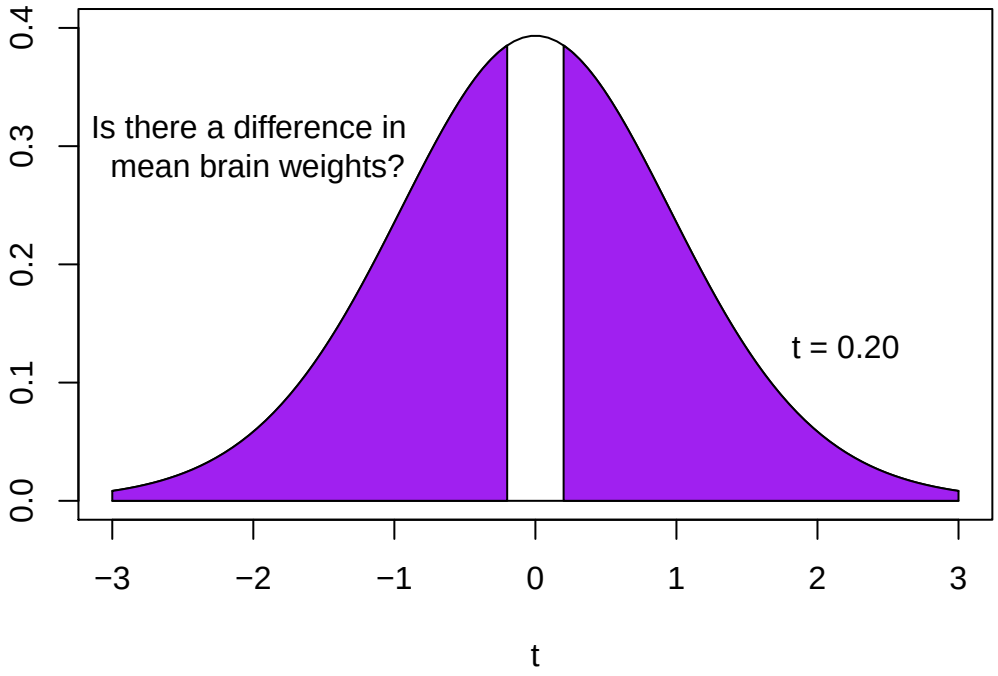
Plot shows brain weight of diseased mice of two types from a study of autoimmune encephalomyelitis development (Welsh et al. PLOS Biology https://doi.org/10.1371/journal.pbio.3000590 18 February, 2020)



- How different do the group means look?
- Assuming the group means are equal, how unusual are these results?
- Would you expect a large or small P value if we tested the hypothesis that the group means were equal?

P values - Wrap-up questions

Mean brain weight (t = 0.20 is standardised difference in weight



P values - Wrap-up questions

```
##
##  Two Sample t-test
##
## data:  weight by mouse_type
## t = 0.19597, df = 18, p-value = 0.8468
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.07776406  0.09376406
## sample estimates:
## mean in group KO mean in group WT
##            0.439            0.431
```

- Assuming the KO and WT groups have the same mean, the (standardised) difference we got from our data does not look very unusual - that difference is pretty close to zero on the plot on the previous slide.
- So, assuming equal means for the KO and WT groups, the chance of getting the difference in the two means that we got from our data is quite large. From the t test, P = 0.85, which is a large P value.
- There is no evidence here against the null hypothesis that the two groups have equal means.

P values - Wrap-up questions

Zaslavsky, K. et al. 2019. SHANK2, Supplementary Figure 3c

- How different do the group means look?
- Assuming the group means are equal, how unusual are these results?
- Would you expect a large or small P value if we tested the hypothesis that the group means were equal?

Useful references

- Good statistics textbooks

MacGillivray H et al. (2013) *Mind on Statistics.* Cengage Learning Australia, Melbourne, 2nd ed. (Available as e-book in UNSW library)

Quinn GP & Keough MJ (2002) *Experimental Design and Data Analysis for Biologists.* Cambridge University Press, Cambridge. (Available as e-book in UNSW library)

- Up-to-date book - topics include RNA-Seq, flow-cytometry, taxa abundances, imaging data and single cell measurements (examples are given using R)

Holmes S & Huber W (2019) *Modern Statistics for Modern Biology.* URL: http://web.stanford.edu/class/bios221/book/index.html

Foulkes AS (2009) *Applied Statistical Genetics with R: For Population-based Association Studies.* Springer, New York. (Available as e-book in UNSW library)

Laird NM & Lange C (2011) *The Fundamentals of Modern Statistical Genetics.* Springer, New York. (Available as e-book in UNSW library)

continued

Useful references (continued)

- Motley collection of useful articles

Greenland S et al. (2016) Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 31:337–350 DOI 10.1007/s10654-016-0149-3

Vaux D (2014) Basic statistics in cell biology. *Annu. Rev. Cell Dev. Biol.* 30:23–37 DOI 10.1146/annurev-cellbio-100913-013303

- Guidelines on how to report statistical results

Lang TA & Altman DG (2016) Statistical analyses and methods in the published literature: The SAMPL guidelines. *Medical Writing* 25:31-36. (Not available through UNSW library. Available here (if you can't get it, let me know and I'll send it): https://journal.emwa.org/statistics/)

Lang T (2016) Never P alone: The value of estimates and confidence intervals. *Medical Writing* 25:17:21. (Available at url above.)

Harrington D et al. (2019) New guidelines for statistical reporting in the Journal. *New England Journal of Medicine* 381:285-286