[redacted]

UNDERSTANDING NULL HYPOTHESIS TESTS, T TESTS

[redacted]

Learning Outcomes

By the end of this lecture, you should be able to:

- State statistical hypotheses (H_0 and H_a) in symbolic and sentence form by identifying the relevant claims in a scenario, and selecting an appropriate parameter
- Understand the definition of a P-value in null hypothesis testing
- describe and evaluate the model conditions for the t confidence interval and t-test for μ

Case Study: *Problem and Plan*

Does aromatherapy foot massage reduce systolic blood pressure (SBP) in adults?

Healthy Japanese adults

↓

Baseline data

Group A	**Group B**
Foot massage	Non-intervention
↓	↓
Week 4 data
Non-intervention	Foot massage
Week 8 data

Data Structure:

Analyse the design in terms of:

- Number and identity of comparison groups
- Type of comparison groups (matched vs. independent)
- Types of variables

Eguchi et al. 2016. doi:10.1371/journal.pone.0151712

Statistical hypotheses

Mathematical statements about the unknown parameter, based on the research claim/prediction

Null hypothesis (H_0):

- Assumed to be true for purpose of test
- Provides a value for the parameter (i.e. $\theta = c$)
- Assumes no difference, no relationship, no effect, etc.

Alternative hypothesis (H_A):

- What we conclude is true if there is evidence against H_0
- Can be:
  - one-tailed/sided: $\theta > c$ (right tailed) OR $\theta < c$ (left tailed)
  - two-tailed/sided: $\theta \neq c$)

Case Study: *hypotheses for t-test of μ*

Variable? *differences (after–before) in systolic blood pressure*

What parameter? *mean*

Null hypothesis (H_0): $\mu_{diff} = 0$

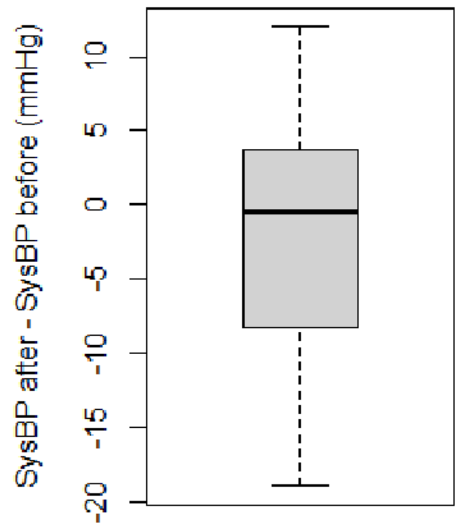
The mean difference in systolic blood pressure is zero.

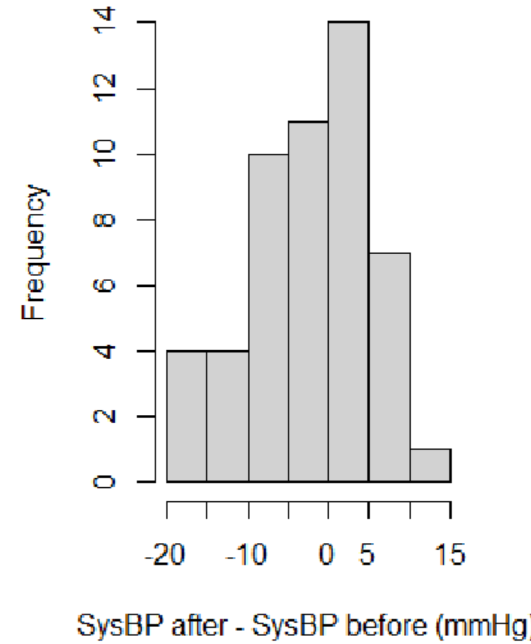
Alternative hypothesis (H_A): $\mu_{diff} < 0$

The mean difference in systolic blood pressure is less than zero.

Case Study: *Data*

```
> summary(diff)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-19.000  -8.250  -0.500  -2.225   3.750  12.000
> sd(diff)
[1] 7.647427
> length(diff)
[1] 51
```





Data manipulation required calculating:

- mean SBP ("mSBP") from two measurements per individual for each time point
- differences in mSBP per individual to compare *foot massage* (*'after'*) minus *no intervention* ('before') (*paired comparison groups*)
- differences in SBP paid attention to the group (*treatment order randomization*)

Model for t CI & test for μ

- Sample data is an SRS with replacement
  - Use your knowledge of the sampling design
- Sample observations come from a population distribution that is Normal
  - If the sample distribution appears Normal, we infer the population distribution is Normal; **use Normal QQ plots**

t procedures are robust to *some* deviations from Normality, provided we have a large sample (n) from a roughly symmetric, unimodal population

Is our statistic *unusual*?

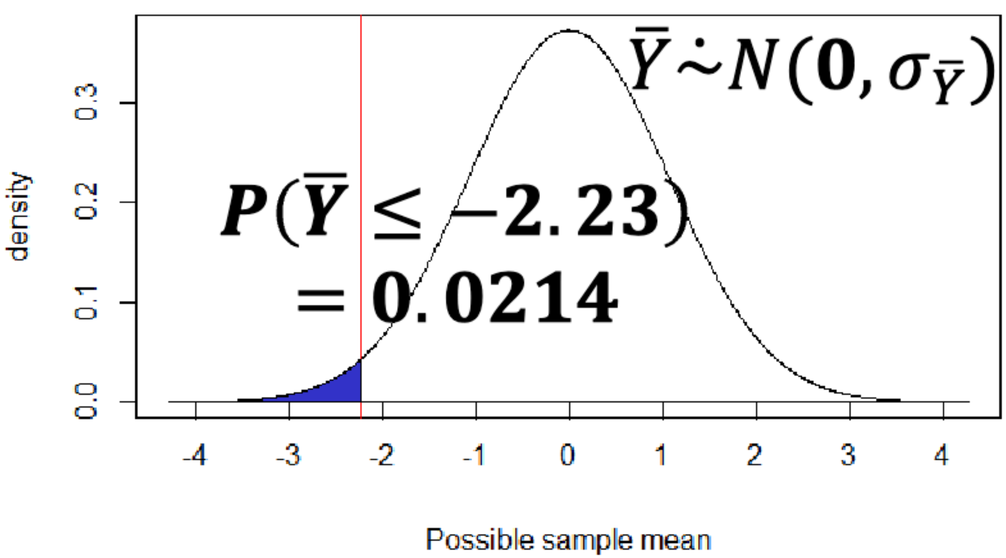
What is the probability of observing a statistic as extreme or more extreme than we did, if H_0 is true?

Choose a model: Assume* $Y \sim N(\mathbf{0}, \sigma)$

H_0: $\mu_{diff} = 0$

H_A: $\mu_{diff} < 0$

Sampling distribution of means (n=51), under H0



Given our original claim, **very small** values (in the *left tail*) would be evidence against H_0 but consistent with the H_A

P-value

*probability of observing sample results as extreme or more extreme (**as relevant to H_A**), if the null model is correct*

Draw a conclusion: If the probability—under a given assumption—is small, then we have evidence that the assumption is not be correct.

Consider:

- Quality of data (sample and study design!)
- Possible implications of an incorrect conclusion

‘Traditional’ conclusion vocabulary

- α is the ‘significance level’: an arbitrary cut-off probability (e.g. 0.05) that we choose as a threshold ‘defining’ too unlikely to happen just due to chance (i.e. sampling error) alone
- ‘statistically significant’ results are more extreme than we would expect due to chance alone.

We avoid using such ‘bright line rules’ and vocabulary to help prevent errors/incorrect beliefs about the scientific process *(as per Wasserstein & Lazar 2016)*

Wasserstein & Lazar (2016). The ASA’s statement on p-values: Context, process, and purpose. The American Statistician 70(2): 192.