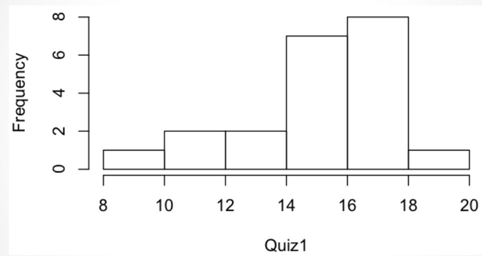
Inference: Fisher's Exact p-values

STA 320
Design and Analysis of Causal Studies
Dr. Kari Lock Morgan and Dr. Fan Li
Department of Statistical Science
Duke University

Quiz 1 Grades




```
> summary(Quiz1)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   8.00   15.00   16.00   15.33   17.00   19.00
```

Review of Quiz

- Potential outcomes: what would spring GPA be if student tents AND what would it be if student doesn't tent?
- Covariates must be pre-treatment
- Assumptions:
  - SUTVA no interference: Y_i not affected by W_j
  - individualistic: W_i not affected by X_j or Y_j
  - unconfounded: W_i not affected by Y_i or Y_j, conditional on **X**
- W_i can depend on W_j
- Context: don't just recite definitions

Review of Last Class

- Randomizing units to treatments creates balanced treatment groups
- Placebos and blinding are important
- Four types of classical randomized experiments:
  - Bernoulli randomized experiment
  - Completely randomized experiment
  - Stratified randomized experiment
  - Paired randomized experiment

Diet Cola and Calcium

- Does drinking diet cola leach calcium from the body?
- 16 healthy women aged 18-40 were randomly assigned to drink 24 ounces of either diet cola or water
- Their urine was collected for 3 hours, and calcium excreted was measured (in mg)
- Is there a significant difference?

Diet Cola and Calcium

Drink	Calcium Excreted
Diet cola	50
Diet cola	62
Diet cola	48
Diet cola	55
Diet cola	58
Diet cola	61
Diet cola	58
Diet cola	56
Water	48
Water	46
Water	54
Water	45
Water	53
Water	46
Water	53
Water	48

Test Statistic

- A test statistic, T, can be any function of:
  - o the observed outcomes, $\mathbf{Y}^{obs}$
  - o the treatment assignment vector, **W**
  - o the covariates, **X**
- The test statistic must be a scalar (one number)
- Examples:
  - o Difference in means
  - o Regression coefficients
  - o Rank statistics
  - o See chapter 5 for a discussion of test statistics

Diet Cola and Calcium

- Difference in sample means between treatment group (diet cola drinkers) and control group (water drinkers)

$$T^{obs} \equiv \bar{Y}_T^{obs} - \bar{Y}_C^{obs}$$
$$= \frac{\sum_{i=1}^{n} W_i Y_i(1)}{N_T} - \frac{\sum_{i=1}^{n} (1-W_i) Y_i(0)}{N_C}$$
$$= 6.875 \text{ mg}$$

Key Question

- Is a difference of 6.875 mg more extreme than we would have observed, just by random chance, if there were no difference between diet cola and water regarding calcium excretion?
- What types of statistics would we see, just by the random assignment to treatment groups?

p-value

- T: A random variable
- T^{obs}: the observed test statistic computed in the actual experiment
- The p-value is the probability that T is as extreme as T^{obs}, if the null is true
- GOAL: Compare T^{obs} to the distribution of T under the null hypothesis, to see how extreme T^{obs} is
- SO: Need distribution of T^{obs} under the null

Sir R.A. Fisher

Randomness

- In Fisher's framework, the only randomness is the treatment assignment: **W**
- The potential outcomes are considered fixed, it is only random which is observed
- The distribution of T arises from the different possibilities for **W**
- For a completely randomized experiment, N choose N_T possibilities for **W**

Sharp Null Hypothesis

- Fisher's sharp null hypothesis is there is no treatment effect:
- H_0: $Y_i(0) = Y_i(1)$ for all i
- Note: this null is stronger than the typical hypothesis of equality of the means
- Advantage of Fisher's sharp null: under the null, all potential outcomes "known"!

Diet Cola and Calcium

- There is NO EFFECT of drinking diet cola (as compared to water) regarding calcium excretion
- So, for *each person* in the study, their amount of calcium excreted would be the same, whether they drank diet cola or water

Sharp Null Hypothesis

- Key point: under the sharp null, the vector $\mathbf{Y}^{obs}$ does not change with different **W**
- Therefore we can compute T exactly under the null for each different **W**!
- Assignment mechanism completely determines the distribution of T under the null
- (why is this not true without sharp null?)

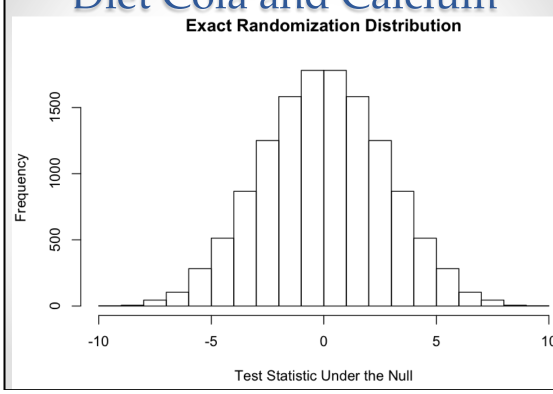
Randomization Distribution

- The randomization distribution is the distribution of the test statistic, *T*, assuming the null is true, over all possible assignment vectors, **W**
- For each possible assignment vector, compute *T* (keeping Y^{obs} fixed, because we are assuming the null)
- The randomization distribution gives us *exactly* the distribution of T, assuming the sharp null hypothesis is true

Diet Cola and Calcium

- 16 choose 8 = 12,870 different possible assignment vectors
- For each of these, calculate *T*, the difference in sample means, keeping the values for calcium excretion fixed

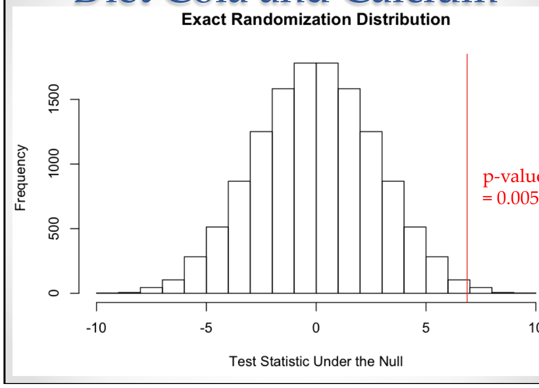
Diet Cola and Calcium



Exact p-value

- From the randomization distribution, computing the p-value is straightforward:
- The exact p-value is the proportion of test statistics in the randomization distribution that are as extreme as T^{obs}
- This is exact because there are no distributional assumptions – we are using the exact distribution of T

Diet Cola and Calcium



Diet Cola and Calcium

- If there were no difference between diet cola and water regarding calcium excretion, only 5/1000 of all randomizations would lead to a difference as extreme as 6.875 mg (the observed difference)
- Drinking diet cola probably does leach calcium from your body!

Notes

- This approach is completely nonparametric – no model specified in terms of a set of unknown parameters
- We don't model the distribution of potential outcomes (they are considered fixed)
- No modeling assumptions or assumptions about the distribution of the potential outcomes

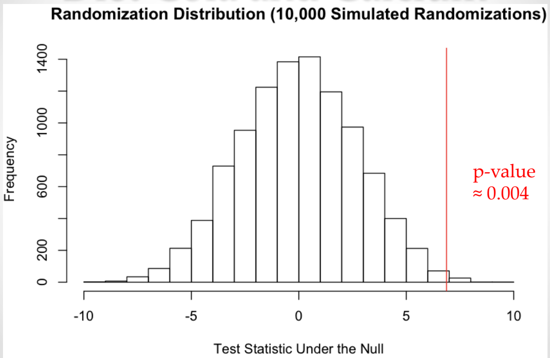
Approximate p-value

- For larger samples, the number of possible assignment vectors (N choose N_T) gets very large
- Enumerating every possible assignment vector becomes computationally difficult
- It's often easier to simulate many (10,000? 100,000?) random assignments

Approximate p-value

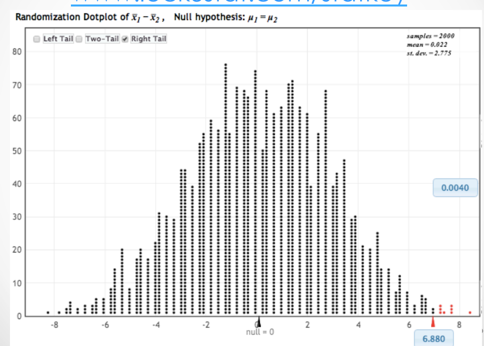
- Repeatedly randomize units to treatments, and calculate test statistic keeping $\mathbf{Y}^{obs}$ fixed
- If the number of simulations is large enough, this randomization distribution will look very much like the exact distribution of T
- Note: estimated p-values will differ slightly from simulation to simulation. This is okay!
- The more simulations, the closer this approximate p-value will be to the exact p-value

Diet Cola and Calcium



StatKey

www.lock5stat.com/statkey



To Do

- Read Ch 5
- Bring laptops to class Wednesday
- HW 2 due next Monday