STAT 500

Randomization Based Inference

Scenario

- Randomized Experiment
  - One Factor with 2 Levels
  - Two Treatments

- Randomly assign experimental units to one of two treatment groups.

Research Question

- Is there a difference in the value of the response variable between the two treatments?

- Source of inference
  - Random assignment of experimental units to treatments

Notation

- Parameters
  - Treatment 1
    - $\boldsymbol{\mu_1}$ = mean response for Treatment 1
    - $\boldsymbol{\sigma_1^2}$ = variance of response for Treatment 1
    - $\boldsymbol{\sigma_1}$ = std. dev. of response for Treatment 1
  - Treatment 2
    - $\boldsymbol{\mu_2}$ = mean response for Treatment 2
    - $\boldsymbol{\sigma_2^2}$ = variance of response for Treatment 2
    - $\boldsymbol{\sigma_2}$ = std. dev. of response for Treatment 2

Notation

- Data
  - $\boldsymbol{Y_{11}, Y_{12}, \ldots, Y_{1n_1}}$

    value of response variable for $\boldsymbol{n_1}$ experimental units receiving treatment 1.
  - $\boldsymbol{Y_{21}, Y_{22}, \ldots, Y_{2n_2}}$

    value of response variable for $\boldsymbol{n_2}$ experimental units receiving treatment 2.

Notation

- Summary Statistics
  - Treatment 1

    $\overline{Y}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} Y_{1j}$

    $S_1^2 = \frac{1}{n_1-1}\sum_{j=1}^{n_1}(Y_{1j} - \overline{Y}_1)^2$ $\qquad$ $S_1 = \sqrt{\frac{1}{n_1-1}\sum_{j=1}^{n_1}(Y_{1j} - \overline{Y}_1)^2}$

  - Treatment 2

    $\overline{Y}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{2j}$

    $S_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(Y_{2j} - \overline{Y}_2)^2$ $\qquad$ $S_2 = \sqrt{\frac{1}{n_2-1}\sum_{j=1}^{n_2}(Y_{2j} - \overline{Y}_2)^2}$

Methods of Analysis

- Reach conclusions and make recommendations using:
  - Visual displays
  - Point estimation: estimates for $\mu_1,\ \mu_2,\ \sigma_1, \sigma_2,$ $\mu_1 - \mu_2,$ etc.
  - Interval estimation: confidence intervals for $\mu_1 - \mu_2$ and other quantities
  - Tests of hypotheses ($\mu_1 = \mu_2$?)
- Types of inference
  - Randomization (design-based)
  - Model-based (relies on the specification of a model)

Randomization as a Basis for Inference

- Used for randomized experiments
- Use the probability distribution imposed by the **random** assignment of units to treatment groups
  - Under the null hypothesis

    H_0 : *treatments have the same effect*

    the response provided by any particular unit does not depend on the assigned treatment ($\Rightarrow \mu_1 = \mu_2$)
  - Is the observed difference $\bar{y}_1 - \bar{y}_2$ inconsistent with H_0?
  - Compare $\bar{y}_1 - \bar{y}_2$ with differences in sample means for all other possible random assignments of units to treatment groups (What if H_0 is true?)

Randomization Test, An Example

- Suppose we want to test whether a drug affects the running ability of rats.
- We randomly divide a group of eight rats into two groups of four.
  - Each rat in one group is injected with the drug.
  - Each rat in the other group is injected with a control substance.
- Then the running time before rest (in minutes) is measured for each rat.

Rats Running Study

Running Time in minutes (Hypothetical Data)

Control:	9	12	14	17
Drug:	18	21	23	26

- The average running time is 13 for the control group, and 22 for the drug group.

- Is this difference caused by the drug?

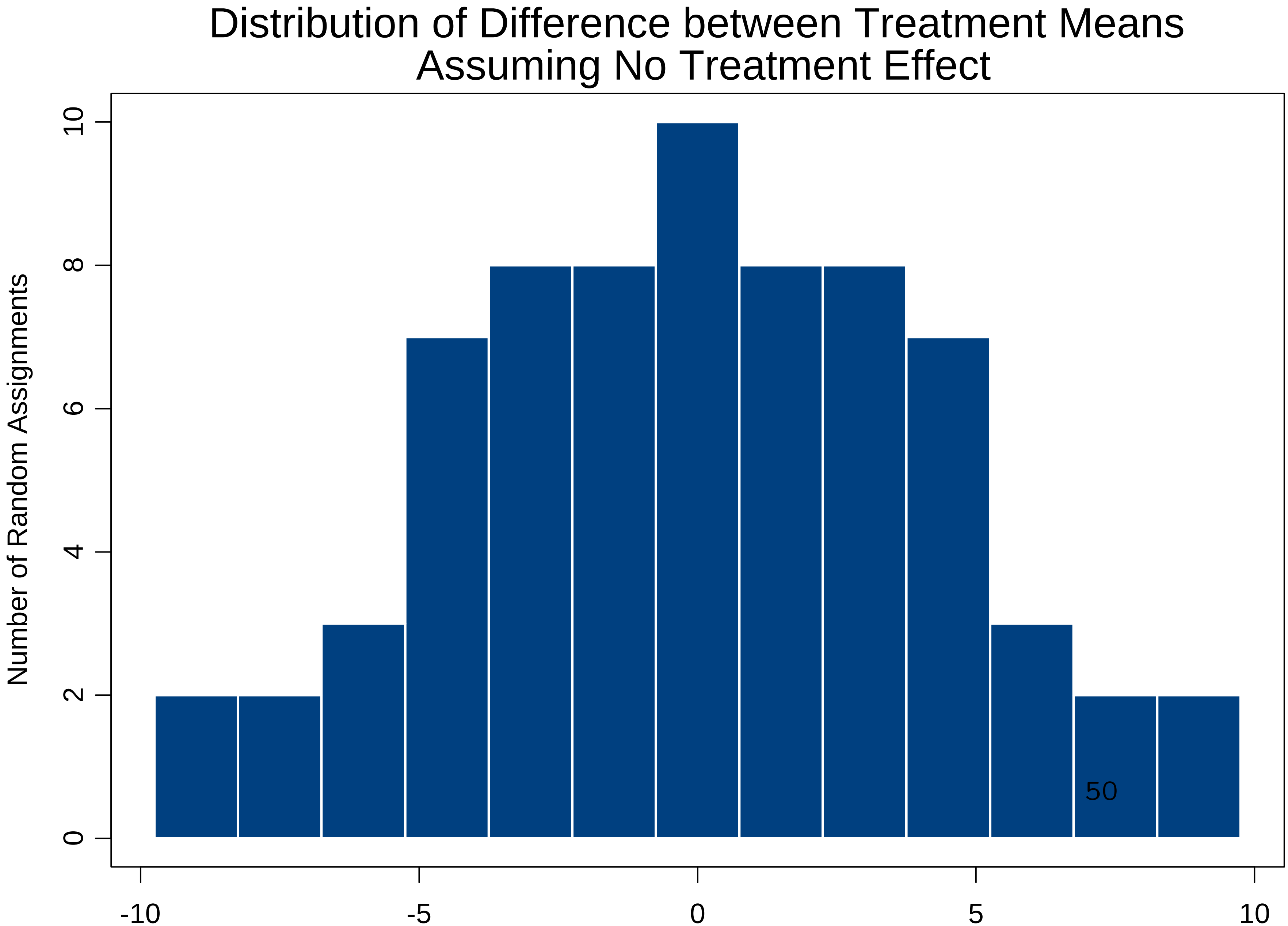
Rats Running Study

- Clearly there is some natural variation in the response variable (not due to treatment) because the running times differ among rats **within each treatment group**.

- Maybe the observed difference (22-13=9) showed up simply because it happened that the rats with better endurance were chosen for injection with the drug.

- What is the chance of seeing such a large difference in treatment means if the drug has no effect?

Random Assignment	Control				Drug				Difference in Averages
1	9	12	14	17	18	21	23	26	9.0
2	9	12	14	18	17	21	23	26	8.5
3	9	12	14	21	17	18	23	26	7.0
4	9	12	14	23	17	18	21	26	6.0
5	9	12	14	26	17	18	21	23	4.5
6	9	12	17	18	14	21	23	26	7.0
7	9	12	17	21	14	18	23	26	5.5
8	9	12	17	23	14	18	21	26	4.5
9	9	12	17	26	14	18	21	23	3.0
10	9	12	18	21	14	17	23	26	5.0
11	9	12	18	23	14	17	21	26	4.0
12	9	12	18	26	14	17	21	23	2.5
13	9	12	21	23	14	17	18	26	2.5
14	9	12	21	26	14	17	18	23	1.0
15	9	12	23	26	14	17	18	21	0.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
69	18	21	23	26	9	12	14	17	-8.5
70	18	21	23	26	9	12	14	17	-9.0


Distribution of Difference between Treatment Means
Assuming No Treatment Effect
Number of Random Assignments
0
2
4
6
8
10
-10
-5
0
5
10
50

Rats Running Study

- Only 2 of the 70 possible random assignments would have led to a difference between treatment means as large as 9.

- Thus, under the assumption of no drug effect, the chance of seeing a difference as large as we observed was $2/70 = 0.0286$.

- Because 0.0286 is a small probability, we have reason to attribute the observed difference to the effect of the drug rather than a coincidence due to the way we assigned our experimental units to treatment groups.

Motivation and Creativity

The Statistical Sleuth, Section 1.1

T. Amabile, *J. Per. and Soc.l Psych.*, 48(2), 1985, 393-99

- Experimental units: experienced creative writers

- Treatments: questionnaires on motivation for writing given at the beginning of the study (SS, page 3)

  - intrinsic motivation (enjoyment, satisfaction, etc...)

  - extrinsic motivation (jobs, financial rewards, etc.)

- Random assignment: (24 intrinsic, 23 extrinsic)

- Response: Creativity displayed in writing a Haiku style poem on laughter $\Rightarrow$ average of evaluations by 12 poets on a 40 point scale

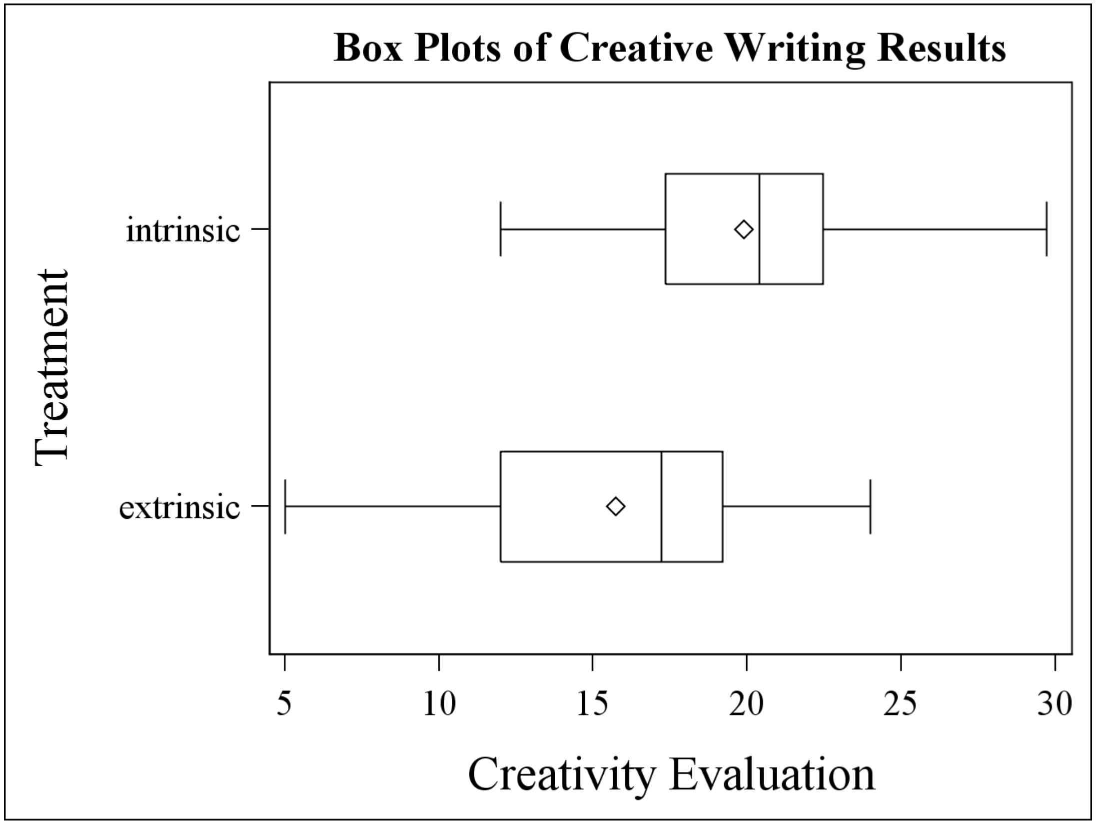
Creative Writing Study

Observed data

intrinsic:	12.0	12.0	12.9	13.6	16.6	17.2
	17.5	18.2	19.1	19.3	19.8	20.3
	20.5	20.6	21.3	21.6	22.1	22.2
	22.6	23.1	24.0	24.3	26.7	29.7
extrinsic:	5.0	5.4	6.1	10.9	11.8	12.0
	12.3	14.8	15.0	16.8	17.2	17.2
	17.4	17.5	18.5	18.7	18.7	19.2
	19.5	20.7	21.2	22.1	24.0	

Creative Writing Study

- Data display (histograms, boxplots, stem-leaf plots)



Creative Writing Study

- Five summary statistics:

  Treatment 1: min=12.0 Q1=17.35 median(Q2)=20.40
  Q3=22.45 max=29.70

  Treatment 2: min= 5.0 Q1=12.00 median(Q2)=17.20
  Q3=19.20 max=24.00

- Sample means and standard deviations

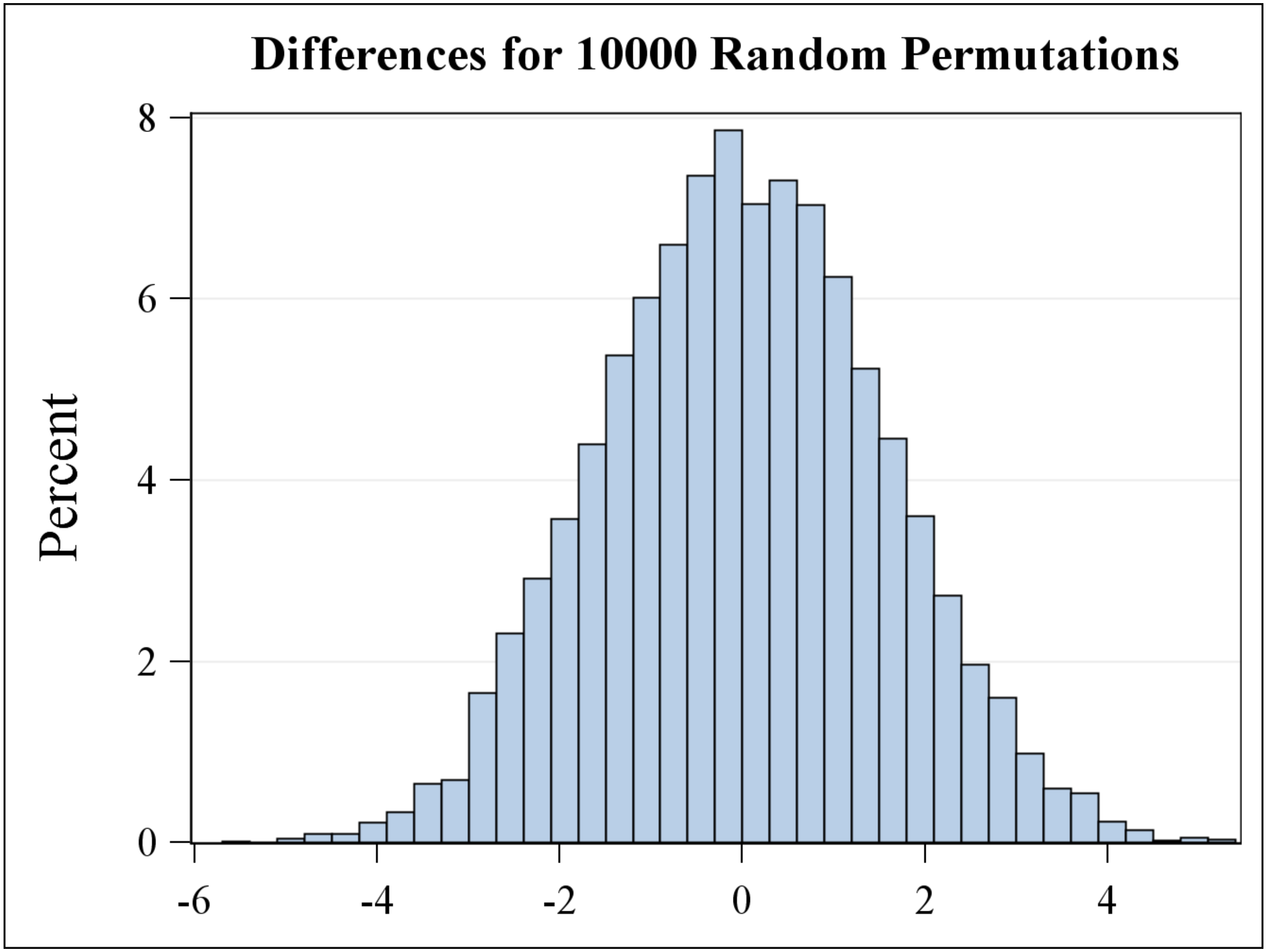
  $$\bar{y}_1 = 19.8875, \qquad s_1 = 4.4418$$

  $$\bar{y}_2 = 15.7391, \qquad s_2 = 5.2526$$

- Observed difference in sample means is 4.1484

Randomization Test

- H_o : Treatments 1 and 2 have the same effect on creativity
- H_a : Treatments 1 and 2 do not have the same effect on creativity
- 1.6×10^{13} possible random assignments
- Sample of 10000 randomization assignments of subjects to treatments (assume the null hypothesis is true)
  - 50 of 10000 randomizations have values as large as 4.1484 or as small as -4.1484)
  - extremely unlikely to see a difference this big by chance (two-tailed p-value = .0050)


Differences for 10000 Random Permutations
Percent
8
6
4
2
0
-6
-4
-2
0
2
4

Randomization Test

- Conclusions
  - questionnaire on intrinsic rewards leads to more creative writing in these students
  - not a random sample ... can't necessarily infer that this is true in a larger population
- The randomization test is also called the permutation test

General Comments

- The randomization test (permutation test) depends on identifying units to permute, which should be the units in the experiment that are **exchangeable under the null hypothesis**, determined by the design of the experiment and the factor(s) being tested.

Randomization Confidence Intervals

- A confidence interval can be constructed from the set of null hypotheses that are not rejected by the randomization test.
- Consider a possible value for $\delta = \mu_1 - \mu_2$
- Subtract δ from every value in sample 1
- Perform a randomization test to determine if $Y_{1j} - \delta$'s have a different mean than Y_{2j}'s
- If H_o is not rejected at the α level, put δ in the $(1-\alpha)\times 100\%$ confidence interval