- office hours M 12–1, W after class

STA 305/1004

Jan. 18, 2016

- Read Week 2 Class notes.
- Read Week 1 Class notes
- Try practice problems.

TODAY'S CLASS

- The concepts of: Randomization, Blocking, Replication
- Summaries of sample populations
- Hypothesis testing via randomization

RANDOMIZED EXPERIMENTS AND OBSERVATIONAL STUDIES

- A technical definition of an observational study is given by Imbens and Rubin (2015)
- The process that determines which experimental units receive which treatments is called the assignment mechanisim.
- When the assignment mechanism is unknown then the design is called an observational study.

RANDOMIZED EXPERIMENTS AND OBSERVATIONAL STUDIES

In randomized experiments (pg. 20, Imbens and Rubin, 2015): "... the assignment mechanism is under the control of the experimenter, and the probability of any assignment of treatments across the units in the experiment is entirely knowable before the experiment begins."

Treatment Assignment

Suppose, for example, that we have two breast cancer patients and we want to randomly assign these two patients to two treatments (A and B). Then how many ways can this be done?

1. patient 1 receives A and patient 2 receives A
2. patient 1 receives A and patient 2 receives B
3. patient 1 receives B and patient 2 receives A
4. patient 1 receives B and patient 2 receives B

- There are 4 possible treatment assignments.
- The probability of a treatment assignment is 1/4,
- The probability that an individual patient receives treatment A (or B) is 1/2.
- In general, if there are N experimental units then there are 2^N possible treatment assignments (provided there are two treatments).

Assuming 2 treatments

Treatment Assignment

A treatment assignment vector records the treatmemnt that each experimental unit is assigned to receive. If $N = 2$ then the possible treatment assignment vectors are:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

where 1= treatment A, and 0=treatment B.

Treatment Assignment

- It wouldn't be a very imformative expriment if both patients received A or both received B.
- Therefore, it makes sense to rule out this scenario.
- We want to assign treatments to patients such that one patient receives A and the other receives B.
- The possibile treatment assignments are:

1. patient 1 receives A and patient 2 receives B or (in vector notation) $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$.
2. patient 1 receives B and patient 2 receives A or (in vector notation) $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

- In this case the probability of a treatment assignment is $1/2$, and the probability that an individual patient receives treatment A (or B) is still $1/2$.

RANDOMIZED EXPERIMENTS AND OBSERVATIONAL STUDIES

Randomized experiments are currently viewed as the most credible basis for determining cause and effect relationships. Health Canada, the U.S. Food and Drug Administration, European Medicines Agency, and other regulatory agencies all rely on randomized experiments in their approval processes for pharmaceutical treatments.

RANDOMIZATION

- The primary objective in the design of experiments is the avoidance of bias or systematic error (Cox and Reid, 2005).
- One way to avoid bias is to use randomization.

RANDOMIZATION

- Create two groups that are the same except for treatment assignment.

- Applied to the allocation of experimental units to treatments.

- Provides protection to experimenter against variables unknown to experimenter but may impact the response.

- Reduces influence of subjective judgement in treatment allocation.

RANDOMIZATION

- National supported work demonstration program (NSW) included a randomized experiment to evaluate the effect of on the job training on unemployment. (Ref: Rosenbaum, pg. 22- 28)
- Treatment: work experience in form of subsidized employment then individuals transitioned to unsubsidized employment.
- Control: standard social programs

RANDOMIZATION

- The response was earnings ($) in 1978.
- Later in course we will compare this with observational studies.
- So participants were matched on pre-treatment covariates.
- Results in 185 treated men matched to 185 treated controls.

RANDOMIZATION

Covariate	Group	
Age (Mean)	Treated Control	25.82 25.70
Years of Education (Mean)	Treated Control	10.35 10.19
Black (%)	Treated Control	84% 85%
Married (%)	Treated Control	19% 20%
Earnings in $ 1974 (Mean)	Treated Control	2096 2009

The two groups look similar before treatment

REPLICATION

Consider the two scenarios:

1. Suppose that 10 batches of a chemical were produced using a new production method and yield is measured.

2. Suppose that 10 samples were taken from the same batch of a chemical produced using a new production method and yield is measured.

1. is concerned with the variation in independent batches of the chemical and 2. reflects the variation in measurement process

1. is an example of replication and 2. is an example of repeated measures.

BLOCKING

When block is size 2 then paired design

- To block an experiment is to divide the observations into groups called blocks so that observations in a block are collected under relatively similar conditions.

- Suppose that the yield of a manufacturing process for penicillin varies a lot depending on how much of a certain raw material is used in the process. To compare four variants of the manufacturing process we might randomize within blocks of the raw material.

BLOCKING

- NSW experiment: assume we paired similar men.
- One member of each pair was randomized to subsidized employment.
- The pair of men would form a block.
- Paired experiments are a form of blocking.

EXAMPLE: WHEAT YIELD

- Is one fertilizer better than another in terms of yield?
- What is the outcome variable? → yield
- What are factor of interest? → fert < A, B

EXAMPLE: WHEAT YIELD

Experimental material?

1	2	3	4	5	6
7	8	9	10	11	12.

EXAMPLE: WHEAT YIELD

- How should we assign treatments/factor levels to plots?
- We want to make sure that we can identify the treatment effect in the presence of other sources of variation.
- What other (besides fertilizer) potential sources could cause variation in wheat yield?

More sun light, different soil,
etc...

EXPERIMENTAL DESIGN

- Assigning treatments randomly avoids any pre-experimental bias.

- 12 playing cards, 6 red, 6 black were shuffled (7 times??) and dealt

1st card black $\longrightarrow$ 1st plot gets B

2nd card red $\longrightarrow$ 2nd plot gets A

3rd card black $\longrightarrow$ 3rd plot gets B

Completely randomized design

R R R B B B Assign

1

B B B R R R

R B B R R R Assign

2

R B B B B R

$\binom{12}{6}$

WHEAT YIELD DATA

B 26.9	A 11.4	B 26.6	A 23.7	B 25.3	B 28.5
B 14.2	A 17.9	A 16.5	A 21.1	B 24.3	A 19.6

- Evidence that fertilizer type is a source of yield variation?
- Evidence about differences between two populations is generally measured by comparing summary statistics across two sample populations.
- A *statistic* is any computable function of the observed data.

SUMMARIZING A SAMPLE DISTRIBUTION

Empirical distribution: $P(a,b] = \#(a < y_i \leq b)/n$

Empirical CDF (cumulative distribution function)

$\hat{F}(y) = \#(y_i \leq y)/n = P(-\infty, y]$

Histograms

Boxplots

EMPIRICAL CDF

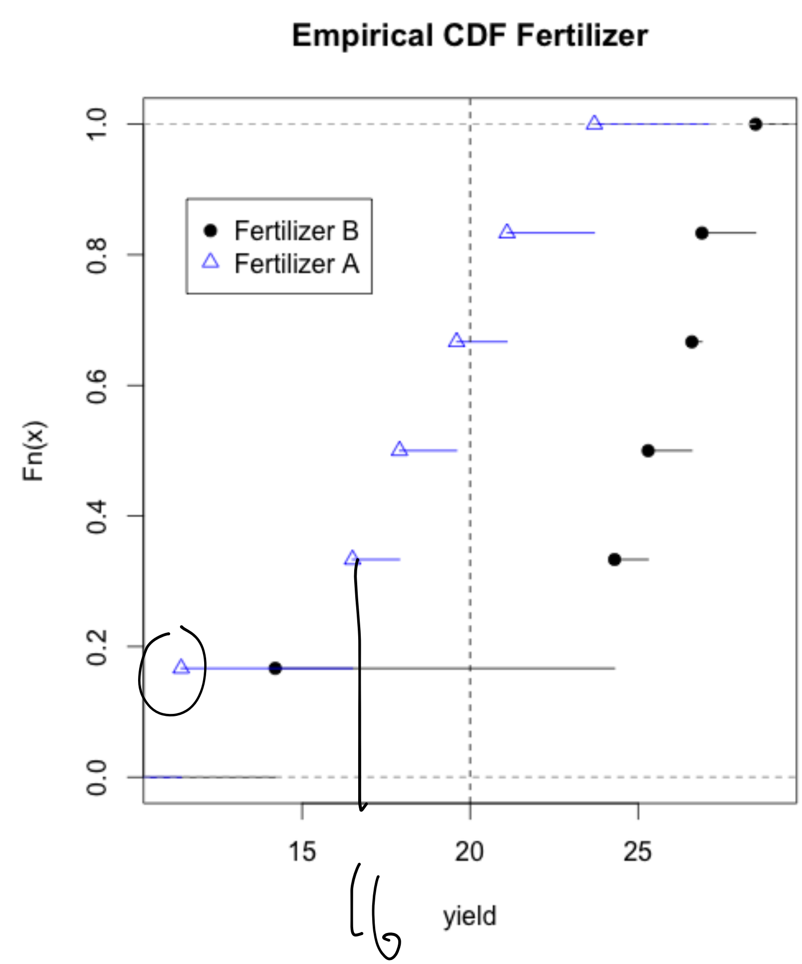
Which fertilizer produces a higher yield?

Respond at **PollEv.com/nathantaback**

Text **NATHANTABACK** to **37607** once to join, then **A or B**

Fertilizer A	10	**A**
Fertilizer B	53	**B**

Total Results: 0



SUMMARIZING A SAMPLE DISTRIBUTION

Sampe mean: $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Sample median: A value $y_{0.5}$ such that

$$y_{(0.5)} = \hat{F}^{-1}(0.5) = \min_{x\in\mathbb{R}}\{\hat{F}^{-1}(x) \geq 0.5\} = \min_{x\in\mathbb{R}}\{\#(y_i \leq x)/n \geq 0.5\}$$

To find the median, sort the data in increasing order, and call these values $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$. If there are no ties then for n odd the median is $y_{\left(\frac{n+1}{2}\right)}$. If n is even then all numbers between $y_{\left(\frac{n}{2}\right)}$ and $y_{\left(\frac{n+1}{2}\right)}$ are medians.

SUMMARIES OF SCALE

sample variance and standard deviation

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2, \ s = +\sqrt{s^2}$$

interquartile range: $[y_{(.25)}, y_{(.75)}]$

95% interval: $[y_{(.025)}, y_{(.975)}]$

SUMMARIES IN R

All of these summaries easily obtained in R

plots that received fert. A

mean yield fert. A

median yield fert. A

sd yield fert. A

1st and 3rd quartiles for yield fert. A: IQR=20.7-16.9=3.8

95% interval for yield fert. A

```
> yA <- c(11.4,23.7,17.9,16.5,21.1,19.6)
> yB <- c(26.9,26.6,25.3,28.5,14.2,24.3)
>
> mean(yA)
[1] 18.36667
> mean(yB)
[1] 24.3
> median(yA)
[1] 18.75
> median(yB)
[1] 25.95
> sd(yA)
[1] 4.234934
> sd(yB)
[1] 5.151699
> quantile(yA,prob=c(0.25,0.75))
   25%    75%
16.850 20.725
> quantile(yB,prob=c(0.025,0.975))
   2.5%   97.5%
15.4625 28.3000
```

RESULTS

- So there is a difference in mean yield for these fertilizers.
- Would you recommend B over A for future plantings?
- Do you think these results generalize to a larger population?

HYPOTHESIS TESTING VIA RANDOMIZATION

- Are the observed differences in yield due to fertilizer type?
- Are the observed differences in yield due to plot-to-plot variation?

HYPOTHESIS TESTING VIA RANDOMIZATION

Hypothesis tests:

- H_0 (null hypothesis): Fertilizer type does not affect yield.
- H_1 (alternative hypothesis): Fertilizer type does affect yield.

A statistical hypothesis evaluates the compatibility of H_0 with the data

TEST STATISTICS AND NULL DISTRIBUTIONS

- We can evaluate H_0 by answering:
- Is a mean difference of -5.93 plausible/probable if H_0 true?
- Is a mean difference of -5.93 large compared to experimental noise?

TEST STATISTICS AND NULL DISTRIBUTIONS

Compare $\bar{y}_a - \bar{y}_b = -5.93$ (observed difference in the experiment) to values of $\bar{y}_a - \bar{y}_b$ that could have been observed if H_0 were true.

Hypothetical values of $\bar{y}_a - \bar{y}_b$ that could have been observed under H_0 are referred to as samples from the null distribution.

TEST STATISTICS AND NULL DISTRIBUTIONS

$\bar{y}_a - \bar{y}_b$ is a function of the outcome of the experiment.

If a different experiment were performed we would get a different value of $\bar{y}_a - \bar{y}_b$.

TEST STATISTICS AND NULL DISTRIBUTIONS

We observed $\bar{y}_a - \bar{y}_b = -5.93$.

If there was no difference between fertilizers then what other possible values of $\bar{y}_a - \bar{y}_b$ could have been observed?

EXPERIMENTAL PROCEDURE AND POTENTIAL OUTCOMES

Shuffled cards were dealt B, R, B, R, ..., fertilizers assigned to subplots

B	A	B	A	B	B
B	A	A	A	B	A

Crops were grown and yields obtained:

B 26.9	A 11.4	B 26.6	A 23.7	B 25.3	B 28.5
B 14.2	A 17.9	A 16.5	A 21.1	B 24.3	A 19.6

EXPERIMENTAL PROCEDURE AND POTENTIAL OUTCOMES

- Imagine re-doing the experiment if H_0 is true (no treatment effect)

B	A	B	B	A	A
A	B	B	A	A	B

- Crops are grown and wheat yields obtained:

B 26.9	A 11.4	B 26.6	B 23.7	A 25.3	A 28.5
A 14.2	B 17.9	B 16.5	A 21.1	A 24.3	B 19.6

EXPERIMENTAL PROCEDURE AND POTENTIAL OUTCOMES

- Under this hypothetical assignment $\bar{y}_a - \bar{y}_b = 1.07$
- This represents an outcome of the experiment in a universe where:

1. The treatment assignment is: B, A, B, B, A, A, A, B, B, A, A, B
2. H_0 is true.

THE NULL DISTRIBUTION

- What outcomes would we see if H_0 is true?
- Compute $\bar{y}_a - \bar{y}_b$ for each possible treatment assignment.

B 26.9	A 11.4	B 26.6	A 23.7	B 25.3	B 28.5
B 14.2	A 17.9	A 16.5	A 21.1	B 24.3	A 19.6

How many ways could treatments have been assigned if the null hypothesis is true?

Respond at **PollEv.com/nathantaback** Text **NATHANTABACK** to **37607** once to join, then **A, B, C, or D**

$\binom{12}{6}$ 924 **A**

$12!$ **B**

$6!$ **C**

2^{12} **D**

THE NULL DISTRIBUTION

- For each one of these compute $\delta = \bar{y}_a - \bar{y}_b$
- $\{\delta_1, \delta_2, \ldots, \delta_{924}\}$ enumerates all potential pre-randomisation outcomes assuming no treatment effect.
- Since each treatment assignment is equally likely the null distribution, a probability distribution of experimental results if H_0 were true can be be easily described as

$$\hat{F}\left(x \mid H_0\right) = P\left(\left(\bar{y}_a - \bar{y}_b\right) < x \mid H_0\right) = \#\left\{\delta_k \leq x\right\}/924$$

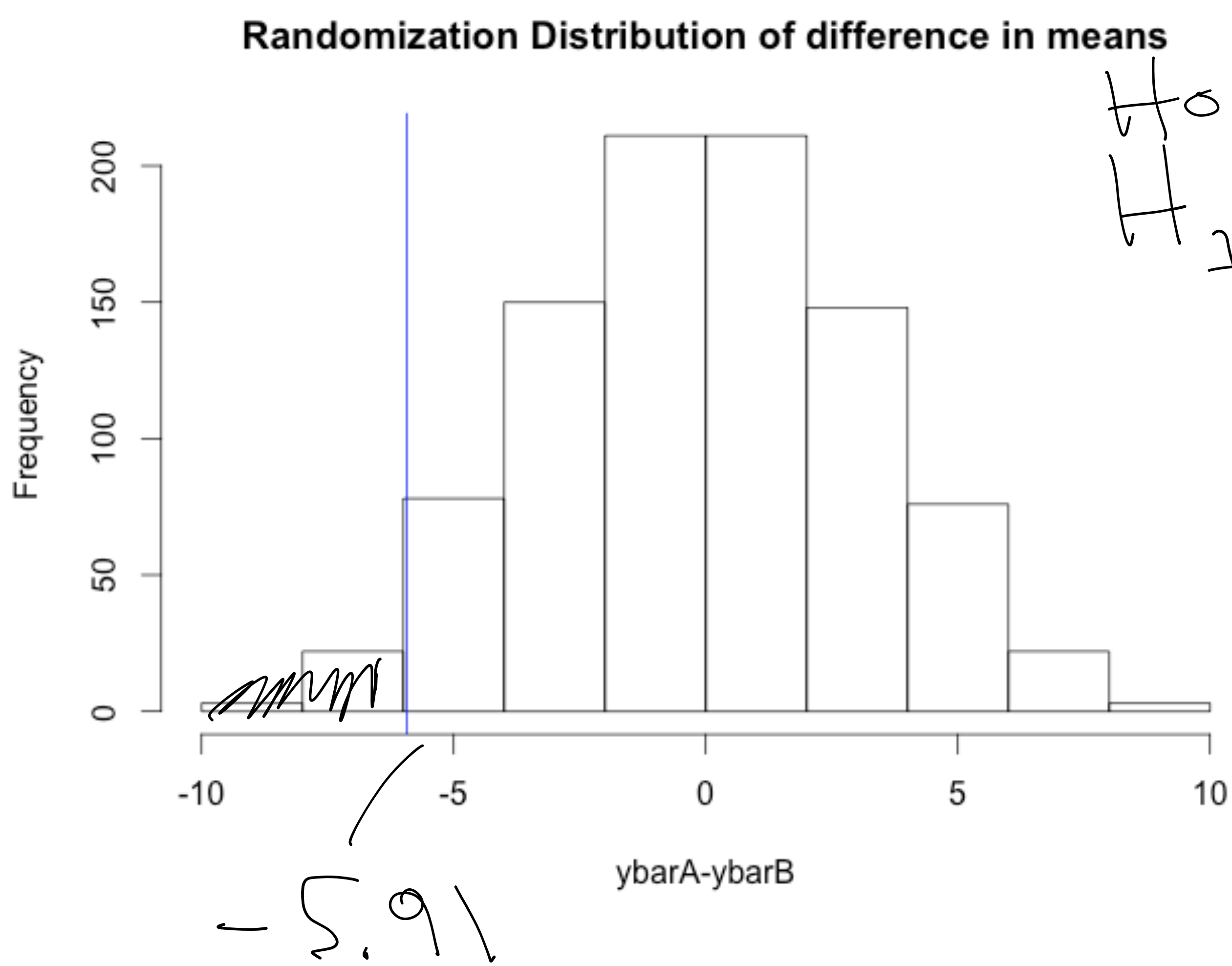
- Called the randomization distribution.

NULL DISTRIBUTION OF FERTILIZER EXAMPLE

μ – mean

$H_0: \mu_A = \mu_B$

$H_1: \mu_B > \mu_A$



HYPOTHESIS TESTING

Is there any contradiction between H_0 and the observed data?

Calculate $P\left(\left(\bar{y}_a - \bar{y}_b\right) < -5.93 \mid H_0\right)$ assume H_0 is true

$\mu_A = \mu_B$

HYPOTHESIS TESTING

- A p-value is the probability, under the null hypothesis of obtaining a more extreme than the observed result.

$$P - value = P\left(\left(\bar{y}_a - \bar{y}_b\right) < -5.93 \mid H_0\right)$$

- Small P-value implies evidence **against** null hypothesis.

does not mean H_0 is

- Large P-value implies *no* evidence **against** null hypothesis. true

- If the P-value is large does this imply that the null is true?

RANDOMIZATION TEST

- We could calculate the difference in means for every possible way to split the data into two samples of size 6.
- This would result in $\binom{12}{6} = 924$ differences.