Stat 411/511

RANDOMIZATION

Oct 14th 2015

Charlotte Wickham

stat511.cwick.co.nz

Today

Finish up last time.

An example from Sleuth.

Randomization.

Two sample t-test in R

```
> t.test(city_mpg ~ trans, data = mpg_sample,
         var.equal = TRUE)

   Two Sample t-test
                                        2*(1 - pt(2.7278, 50))

data:  city_mpg by trans
t = -2.7278, df = 50, p-value = 0.008774
alternative hypothesis: true difference in means
   is not equal to 0
95 percent confidence interval:
 -5.2757919 -0.8011311
sample estimates:
  mean in group auto mean in group manual
            17.38462             20.42308
```

Statistical Summary

There is convincing evidence that the mean fuel efficiency of automatic cars manufactured in 2012 is not equal to the mean fuel efficiency of manual cars manufactured in 2012 (two sample t-test, two-sided p-value = 0.009).

The mean fuel efficiency of automatic cars manufactured in 2012 is estimated to be 3.0 mpg lower than the mean fuel efficiency of manual cars manufactured in 2012.

With 95% confidence the mean fuel efficiency of automatic cars is between 0.8 and 5.3 mpg lower than the population mean fuel efficiency of manual cars manufactured in 2012.

t-tools summary so far

The t-tools are motivated by the random sampling models (paired or two sample).

Which t-tool is appropriate (paired or two sample) depends on the design of the study.

The sampling distributions of the t-ratios are known exactly if you also assume Normal populations (and in the two sample case, equal population standard deviations).

Our conclusions are about the parameters of the populations (mean difference or difference in means).

What if you don't have random samples?

Often people proceed with the t-tools anyway.

The conclusions rely on an additional assumption,

"our data **is just like a random sample** from a population of interest"

This assumption is always suspect, and any deviations can lead to significant bias and misleading conclusions.

Arguments for why your **"not random"** sample is just like a **random** sample cannot be backed up statistically.

There is one situation where the t-tools can be used without random sampling, but they become an approximation
this is where we are heading this week....

Some interesting reading about non-random samples:

http://www.stat.berkeley.edu/~census/berk2.pdf

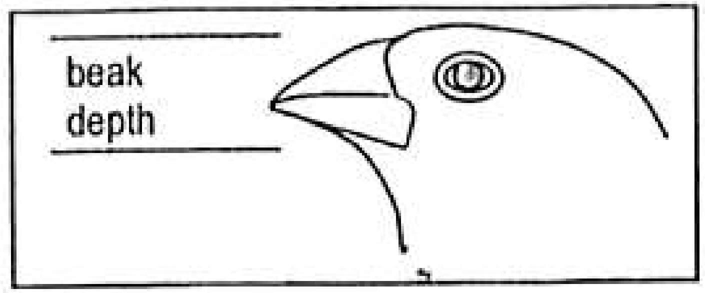
Conventional statistical inferences (e.g., formulas for the standard error of the mean, t-tests, etc.) depend on the assumption of random sampling. This is not a matter of debate or opinion; it is a matter of mathematical necessity.[3] When applied to convenience samples, the random sampling assumption is not a mere technicality or a minor revision on the periphery; the assumption becomes an integral part of the theory.

First a note about case0201 (3rd edition)

In the 1980s, biologists Peter and Rosemary Grant and colleagues found what Pearson had been looking for. Over the course of 30 years, the Grants' research team caught and measured all the birds from more than 20 generations of finches on the Galápagos island of Daphne Major. In one of those years, 1977, a severe drought caused vegetation to wither, and the only remaining food source was a large, tough seed, which the finches ordinarily ignored. Were the birds with larger and stronger beaks for opening these tough seeds more likely to survive that year and did they tend to pass this characteristic to their offspring?

The Grants measured beak depths (height of the beak at its base) of all 751 Daphne Major finches the year before the drought (1976) and all 89 finches captured the year after the drought (1978). Display 2.1 shows side-by-side stem-and-leaf diagrams comparing the 89 post-drought finch bill depths with an equal-sized random sample of the pre-drought bill depths. (For the full set of 1976 finches, see Exercise 2.18.) Is there evidence of a difference between the population distributions of beak depths in 1976 and 1978? (The data were read from a histogram in P. Grant, 1986, *Ecology and Evolution of Darwin's Finches*, Princeton University Press, Princeton, N.J.)

DISPLAY 2.1 Beak depths (mm) of Darwin finches on Daphne Major in 1976, pre-drought, and 1978, post-drought



1976
Average: 9.4697 mm
SD: 1.0353 mm
n: 89

a random sample

1978
Average: 10.1382 mm
SD: 0.9065 mm
n: 89

all birds in the population

1976	Stem	1978
2	6	
8		
411	7	1
98		9
44420	8	044
9999977765555		778
444321111100000	9	0011123344
99999888887777655		566666777789999
44443332211111000	10	00022222333334444
8766655555		55555566666777778999
440	11	00001111134444
7		5667

Legend: |11| 0 = 11.0 mm

A few different ways to proceed:

There's no sampling, so there's no sampling variability. We have both populations, calculate the means and compare.

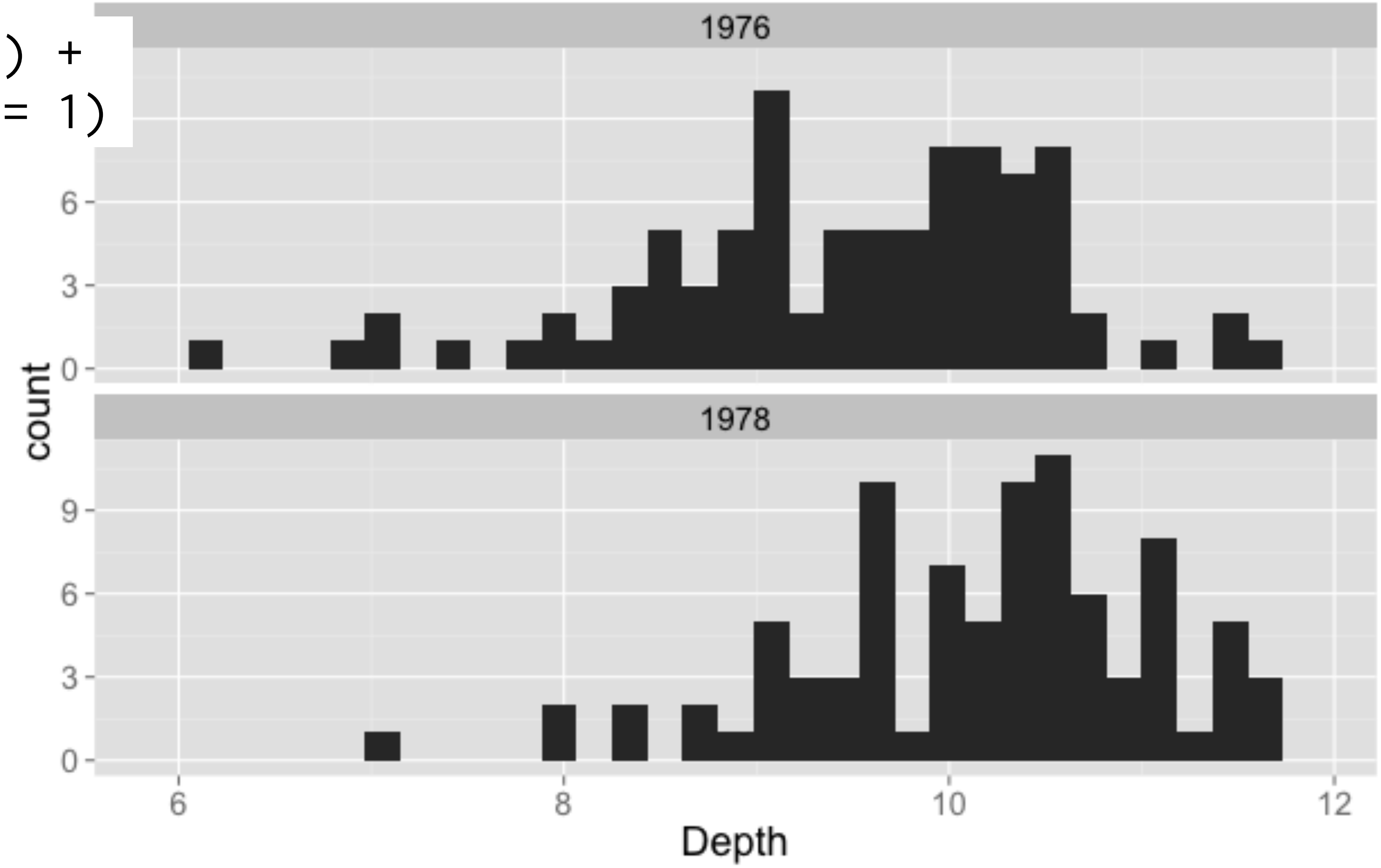
Assume the population is 1976 finches. If there is no natural selection, the 1978 finches are like a single random sample from the 1976 finches. one-sample t-test or one-sample exact test (we could find the sampling distribution of the sample average of a sample of size 89 exactly because we know the population, rather than assuming it's normal and doing a t-test)

There is variability but it isn't due to sampling. The finches in 1976 and 1978 are the result of some random process, that we can assume works like taking two random samples from two imaginary populations. two-sample t-test

Sleuth with the addition of subsampling 1976

The appropriate analysis depends on which assumptions you think are justifiable.

```
qplot(Depth, data = case0201) +
    facet_wrap(~ Year, ncol = 1)
```




```
> t.test(Depth ~ Year, data = case0201, var.equal = TRUE)

    Two Sample t-test

data:  Depth by Year
t = -4.5833, df = 176, p-value = 8.65e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9564088 -0.3806698
sample estimates:
mean in group 1976 mean in group 1978
          9.469663          10.138202
```

Your turn

Fill in the blanks.

There is _________ evidence that the mean _________ of ______________________ is not equal to the mean __________ of _____________________ (two sample t-test, two-sided p-value = _____).

The mean ____________ of ______________ is estimated to be _____ _____ _________ than the mean ______________ of _____________________.

With 95% confidence the mean __________ of __________________ is between ____ and ____ _____ ________ than the mean ____________ of _____________.

Where we are imagining the finch's on Daphne Major each year are like samples from population distributions for each year.

One sided versus two-sided p-values

not to be confused with one sample vs two sample t-test

Null: the mean difference is equal to zero, $\mu = 0$

Alternative: the mean difference **is not equal** zero, $\mu \neq 0$

two-sided

Null: the difference in means is equal to zero, $\mu_1 = \mu_2$

Alternative: the difference in means **is not equal** zero, $\mu_1 \neq \mu_2$

Null: the mean difference is equal to zero, $\mu = 0$

Alternative: the mean difference **is greater than** zero, $\mu > 0$

one-sided

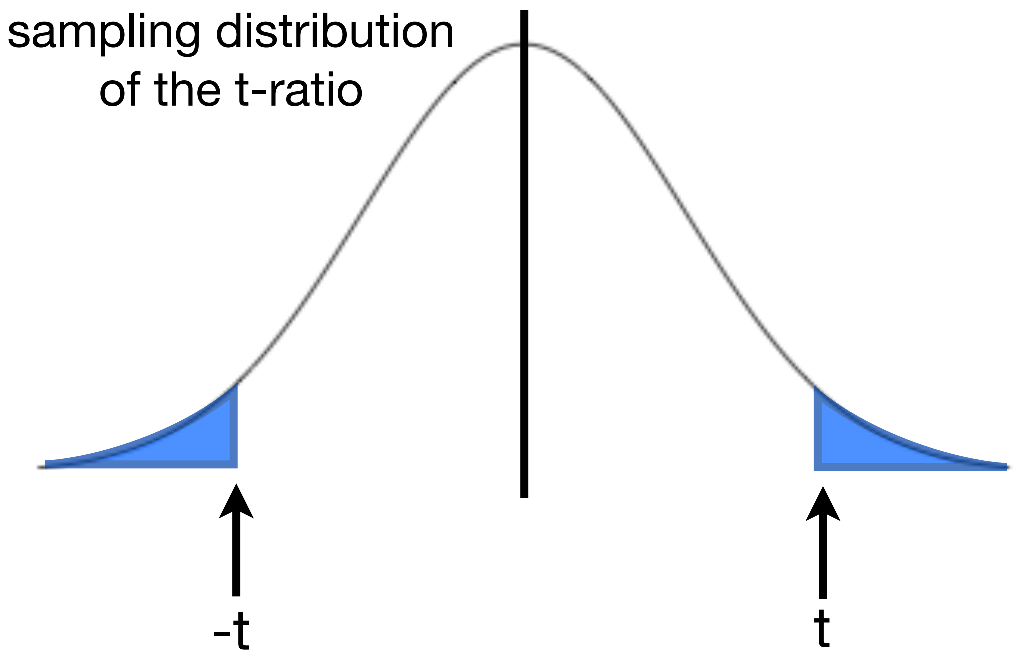
Null: the difference in means is equal to zero, $\mu_1 = \mu_2$

Alternative: the mean for population 1 **is greater than** the the mean for population 2, $\mu_1 > \mu_2$

or less

Review: OpenIntro 4.3.4

two-sided



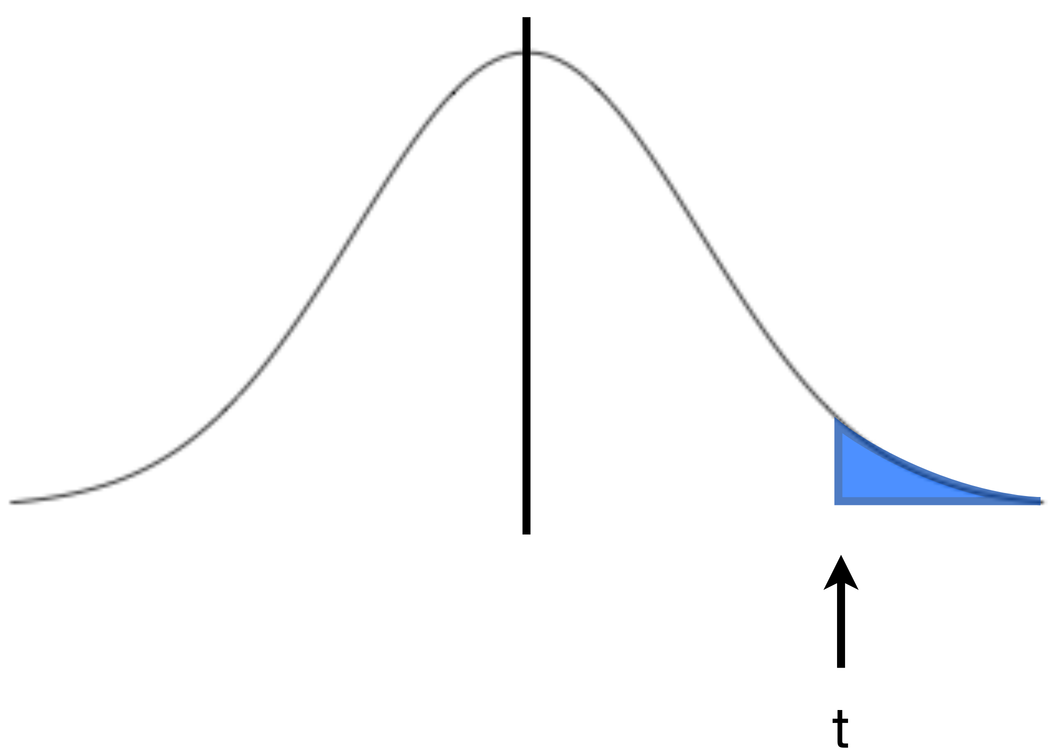
We don’t specify a direction so a more extreme test statistic could be:

observing a more extreme difference between the two groups in the same direction as we observed

or a more extreme difference between the two groups in the opposite direction to what we observed.

We need the area in blue.

one-sided



We specify a direction so a more extreme test statistic could only be:

observing a more extreme difference between the two groups in the direction we specified.

We need the area in blue.

In general, you need a good reason to specify a one-sided test, and you need to do so before seeing your data.

Forget about t-tests

Forget about random sampling

just for now....

A different mechanism of chance

The Randomized Experiment

Causal inference is using our data to make inferences about cause and effect relationships.

This is statistically justified as long as:

experimental units are **randomly assigned** to the **treatments of interest**.

We call a study in which experimental units are randomly assigned to treatment a **randomized experiment**.

Display 1.1 **p. 2**

Creativity scores in two motivation groups, and their summary statistics

Does intrinsic motivation improve creativity?

	Motivation Group		
	Assigned randomly by researcher		
Intrinsic		***Extrinsic***	
12.0	20.5	5.0	17.4
12.0	20.6	5.4	17.5
12.9	21.3	6.1	18.5
13.6	21.6	10.9	18.7
16.6	22.1	11.8	18.7
17.2	22.2	12.0	19.2

The intrinsic group has an average creativity score 4.1 points higher than the extrinsic group

	Intrinsic		Extrinsic	
Sample Size:	24		23	
Average:	19.88	-	15.74	= 4.1
Sample Standard Deviation:	4.44		5.25	

Read: 1.1.1 in Sleuth

Display 1.2

p. 3

Questionnaires given creative writers, to rank intrinsic and extrinsic reasons for writing

INSTRUCTIONS: Please rank the following list of reasons for writing, in order of personal importance to you (1 = highest, 7 = lowest).

__ You get a lot of pleasure out of reading something good that you have written.
__ You enjoy the opportunity for self-expression.
__ You achieve new insights through your writing.
__ You derive satisfaction from expressing yourself clearly and eloquently.
__ You feel relaxed when writing.
__ You like to play with words.
__ You enjoy becoming involved with ideas, characters, events, and images in your writing.

*List of **extrinsic** reasons for writing* (below)

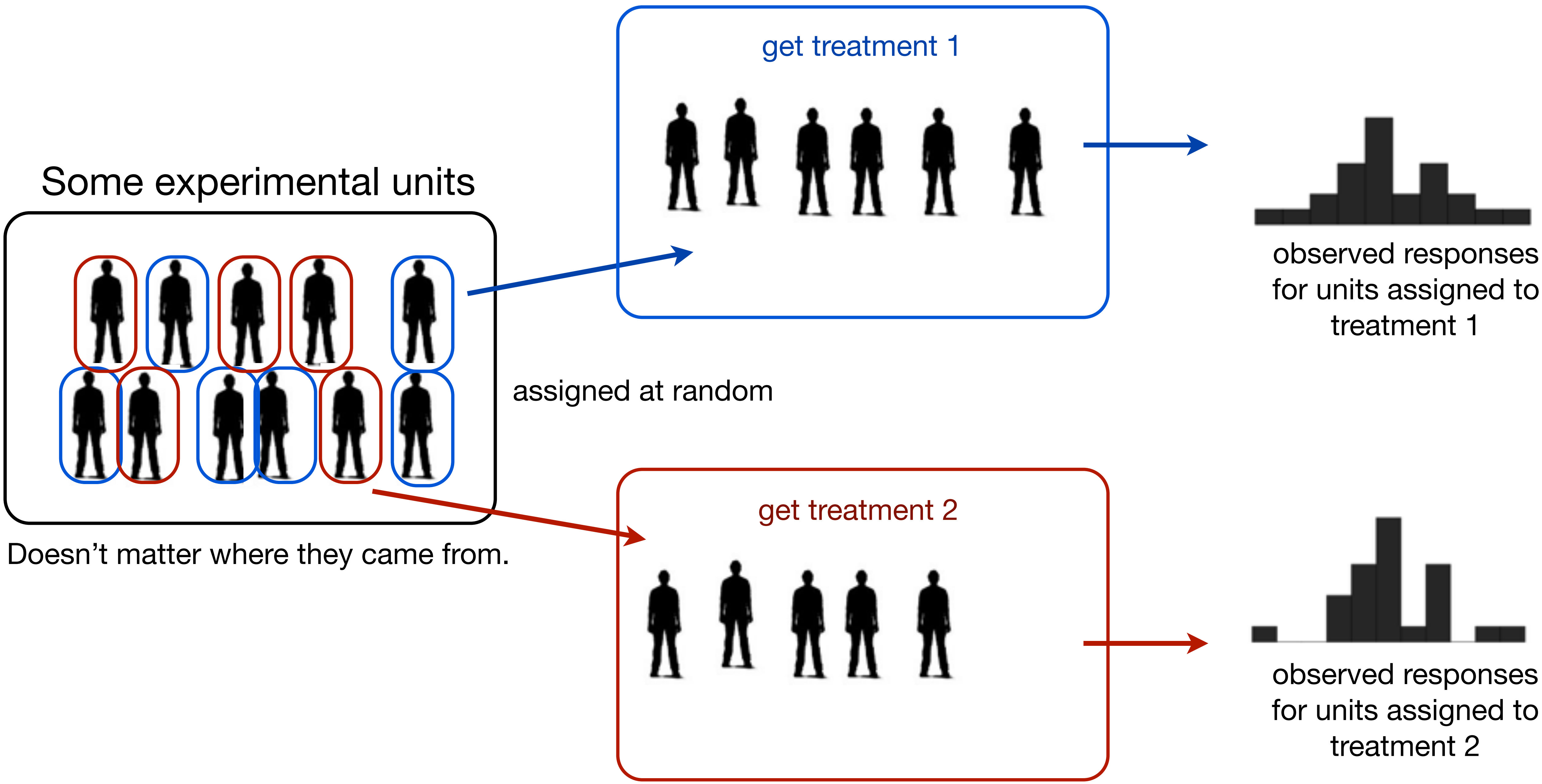
*List of **intrinsic** reasons for writing* (above)

INSTRUCTIONS: Please rank the following list of reasons for writing, in order of personal importance to you (1 = highest, 7 = lowest).

__ You realize that, with the introduction of dozens of magazines every year, the market for free-lance writing is constantly expanding.
__ You want your writing teachers to be favorably impressed with your writing talent.
__ You have heard of cases where one bestselling novel or collection of poems has made the author financially secure.
__ You enjoy public recognition of your work.
__ You know that many of the best jobs available require good writing skills.
__ You know that writing ability is one of the major criteria for acceptance into graduate school.
__ Your teachers and parents have encouraged you to go into writing.

The randomized experiment model

Key idea: there is no population, and no sampling!



Chance only enters through the random assignment of units to treatments

Randomization Distribution

The randomization distribution is the histogram of all values for the statistic from all possible ways the experimental units could have been randomly assigned to groups.

In the sampling model, the reason there is variability in a sample statistic is because we induced variability by taking a random sample. We describe the variability using the sampling distribution of the statistic.

In the randomized experiment model, the only reason we see variability in group statistics is because we induced variability by randomly assigning people to groups. We describe the variability using the randomization distribution of the statistic.

In randomized experiments it's the relationship between the randomization distribution and the effect of the treatment that allow us to make inferences.

Remember: Statistical testing

1. Set up the null hypothesis (and alternative hypothesis)

2. Calculate the **test statistic**

3. Evaluate the evidence against the null hypothesis by comparing the test statistic to test statistics expected under the null hypothesis, the **null distribution**.

The evidence is summarized by a **p-value**, the probability we would see such an extreme test-statistic if the null hypothesis is true.

4. **If the p is low, the null must go!**

Reject or fail to reject the null hypothesis

To do a test all we really need to know is the null distribution.
I.e. the randomization distribution if the null was true.