ST 411/511 Outline 1

Reading assignment: Chapter 1. This will be largely review. You should have covered most of the material in your introductory statistics class.

Chapter 1 Drawing Statistical Conclusions

Case Study 1.1.1 What are the effects of intrinsic vs. extrinsic motivation on creativity?

47 subjects randomly assigned to two groups. Each group fills out a diff. questionnaire. The each subject writes a haiku.

DISPLAY 1.2 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.* — *List of **intrinsic** 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 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 best-selling 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.

12 poets evaluate haikus & assign a creativity score to each. Creativity measurement for one subject is average of 12 scores.

Statistics

Columns of data

Visualize analyze

A story revealed.

5 syllable

7

5

DISPLAY 1.1 Creativity scores in two motivation groups, and their summary statistics

	Intrinsic group		Extrinsic group	
	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
	17.5	22.6	12.3	19.5
	18.2	23.1	14.8	20.7
	19.1	24.0	15.0	21.2
	19.3	24.3	16.8	22.1
	19.8	26.7	17.2	24.0
	20.3	29.7	17.2	
Sample Size:	24		23	
Average:	19.88		15.74	
Sample Standard Deviation:	4.44		5.25	

Summary stats

Case Study 1.1.2 Did Harris Trust and Savings Bank discriminate by paying higher salaries to men than women between 1969 and 1977?

all of them

DISPLAY 1.3 Starting salaries ($U.S.) for 32 male and 61 female clerical hires at a bank

Males			Females					
4,620	5,700	6,000	3,900	4,500	4,800	5,220	5,400	5,640
5,040	6,000	6,000	4,020	4,620	4,800	5,220	5,400	5,700
5,100	6,000	6,000	4,290	4,800	4,980	5,280	5,400	5,700
5,100	6,000	6,300	4,380	4,800	5,100	5,280	5,400	5,700
5,220	6,000	6,600	4,380	4,800	5,100	5,280	5,400	5,700
5,400	6,000	6,600	4,380	4,800	5,100	5,400	5,400	5,700
5,400	6,000	6,600	4,380	4,800	5,100	5,400	5,400	6,000
5,400	6,000	6,840	4,380	4,800	5,100	5,400	5,520	6,000
5,400	6,000	6,900	4,440	4,800	5,100	5,400	5,520	6,120
5,400	6,000	6,900	4,500	4,800	5,160	5,400	5,580	6,300
	6,000	8,100						6,300

Differences between the two case studies:

1. Census vs. sample

all starting salaries

some subset of a pop'n (e.g. college undergrads)

2. Randomized experiment vs. observational study.

Creativity study assigned subjects randomly to groups

Groups determined by subjects.

3. "Response" is quite different in each study. (Response = what's measured on each subject)

4. Sample size in each group

Creativity study - almost balanced

Salary study - many more females than males

5. Creativity score is based on poets' opinions. But it's average of 12. Responses depend on 12 poets chosen.

Considerations for ***Scope of Inference***: What can we infer from the study?

Random sample? If so, then inference applies to popn sampled.
If not specified as random sample, then don't assume it is.
It may be reasonable to assume subjects are representative of some popn. But clearly state this.

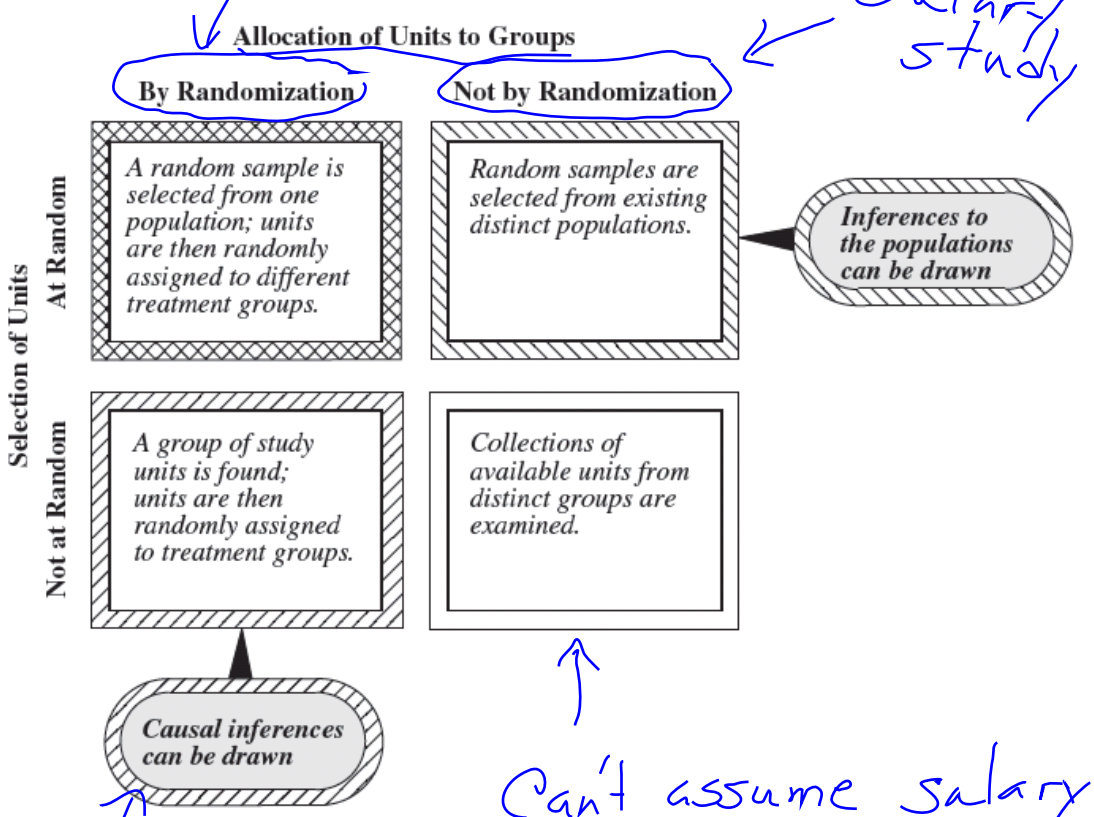
Subjects randomly assigned to groups?

If so, can make causal inference, i.e. inferred diffs. between groups caused by treatments.

Creativity study

Salary study

DISPLAY 1.5 Statistical inferences permitted by study designs



Can say treatments caused diff. in creativity score.

Can't assume salary diff. caused by gender.
Can only say there's an association between salary and gender other factors may be responsible for diff. in salary (e.g. experience).

when subjects not randomly drawn from pop'n, can argue they're representative of a larger pop'n, but be cautious and explicit that's what you're doing.

Statistical Hypothesis Test Logic

Review

1. Based on the research question, select the null hypothesis H_0. "dull"

Typically, no diff. between groups.
Alternative hypothesis H_A is negation of H_0.

2. Select a test statistic. Something we can calculate from data, to help decide if H_0 is plausible or not.
e.g. diff. in sample means.

3. Determine (or approximate) the **sampling distribution** of the test statistic when H_0 is true.

think of histogram of test stat. if study were repeated many times.

observed

4. Compare the test statistic calculated from the data to the sampling distribution when H_0 is true.

truth when H_0 of no diff is true.
Are observed data consistent with H_0?
p-value measures how incompatible H_0 & obs. data are.

Creativity Study Hypothesis Test - Permutation test

Null hypothesis:

H_0: no diff. between groups

H_A: some difference between groups

Test statistic: Diff. in sample means

$\bar{Y}_1 - \bar{Y}_2$

$\bar{Y}_i$ = sample mean in i^{th} group

Sampling distribution of test statistic under null hypothesis:

If H_0 is true, can randomly shuffle obs. data into two groups.

Calculate diff. in sample means.

Do this many times to simulate repeating the experiment many times.

There are finitely many ways to reshuffle, but that number is huge.

DISPLAY 1.1 Creativity scores in two motivation groups, and their summary statistics

Intrinsic group		Extrinsic group	
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
17.5	22.6	12.3	19.5
18.2	23.1	14.8	20.7
19.1	24.0	15.0	21.2
19.3	24.3	16.8	22.1
19.8	26.7	17.2	24.0
20.3	29.7	17.2	
Sample Size:	24		23
Average:	19.88		15.74
Sample Standard Deviation:	4.44		5.25

$\bar{Y}_1 - \bar{Y}_2 = 4.14$

A Random Group Assignment: One shuffle

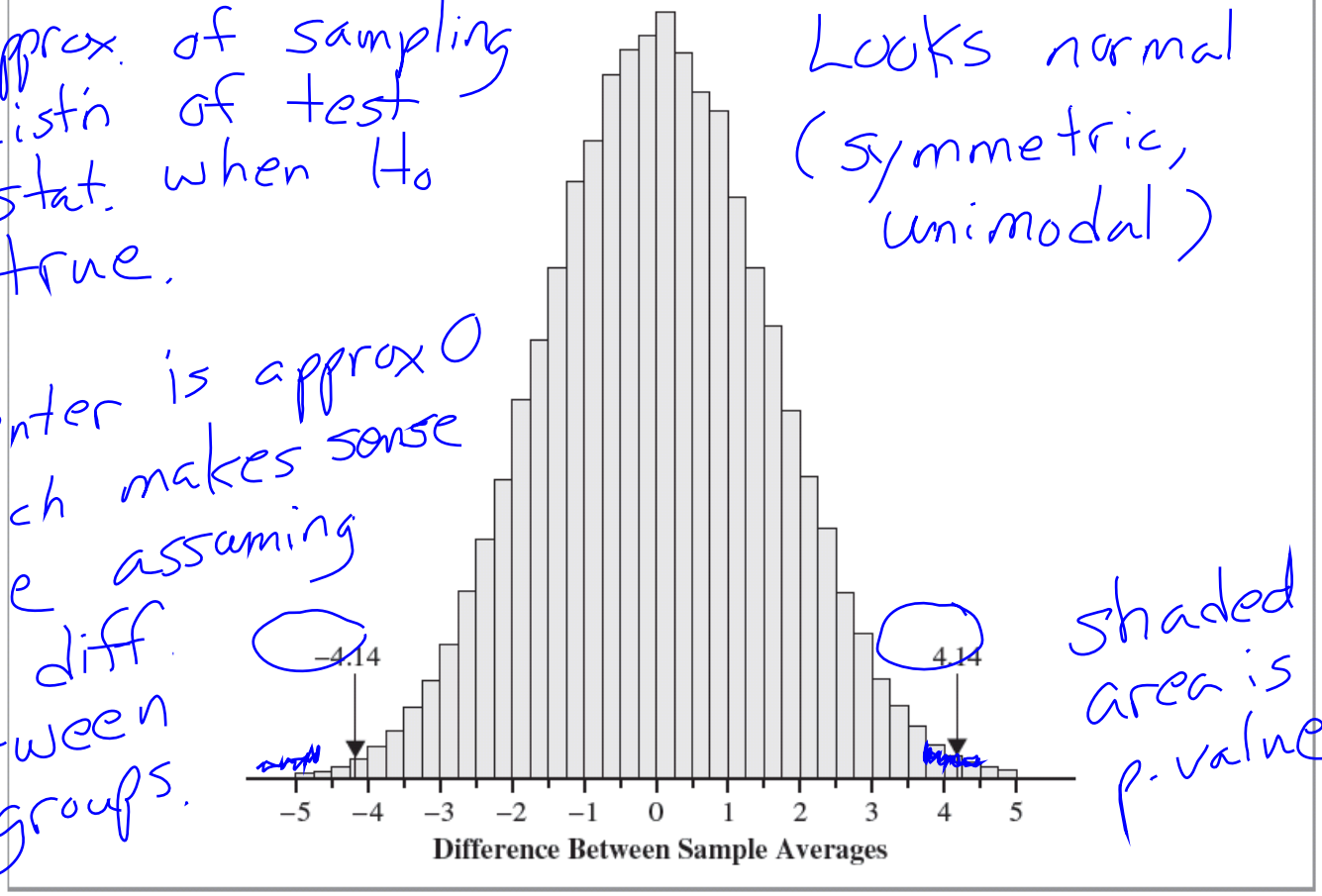
Intrinsic		Extrinsic	
17.2	20.7	5.0	18.2
17.5	21.3	22.1	17.2
16.8	18.7	18.7	17.2
20.5	22.1	12.0	19.8
20.3	10.9	15.0	19.5
20.6	16.6	19.3	18.5
24.0	11.8	21.2	12
22.6	21.6	13.6	12.3
12.9	6.1	19.1	23.1
19.2	17.4	29.7	24.0
24.3	12	5.4	22.2
26.7	17.5	14.8	
Average:	18.30142		17.38696

For this regrouping, $\bar{Y}_1 - \bar{Y}_2 = 0.91$

500,000

Repeat random group assignment many times, keeping track of the difference in group averages.

DISPLAY 1.8 Histogram of differences between group averages, from 500,000 regroupings of the creativity study data



Approx. of sampling distn of test stat. when H_0 true.

Looks normal (symmetric, unimodal)

Center is approx 0 which makes sense since assuming no diff between groups.

shaded area is p-value

Observed test stat. (4.14) is very unusual if H_0 true because it's in the far tail of the distribution.

Conclude H_0 not plausible given the data.

This is a randomization or permutation test. It doesn't require as many assumptions as t-test. But H_0 is more specific (no diff. between groups).

Creativity Study Hypothesis Test Redux

t-test

Null hypothesis: No diff. in pop'n means.

μ_1 = pop'n mean for intrinsic group

μ_2 = " " extrinsic group

$H_0: \mu_1 - \mu_2 = 0$ $H_A: \mu_1 - \mu_2 \neq 0$

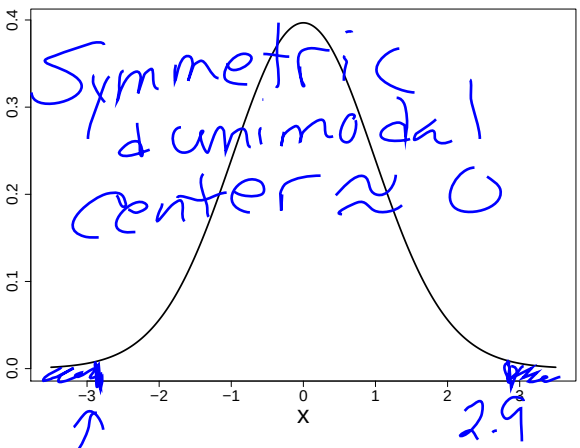
Test statistic:

$$t\text{-stat} = \frac{\bar{Y}_1 - \bar{Y}_2}{SE(\bar{Y}_1 - \bar{Y}_2)}$$

SE = "standard error" which is an estimate of standard deviation

So t-stat is in std. dev. units, so it's on a standard scale.

Sampling distribution of test statistic under null hypothesis:



If assumptions of t-test are met (Ch. 3) and H_0 true t-stat has a t dist'n with df degrees of freedom = 45 (more later)

p-value = shaded area

A t-test using R: You'll do this and more in Lab!

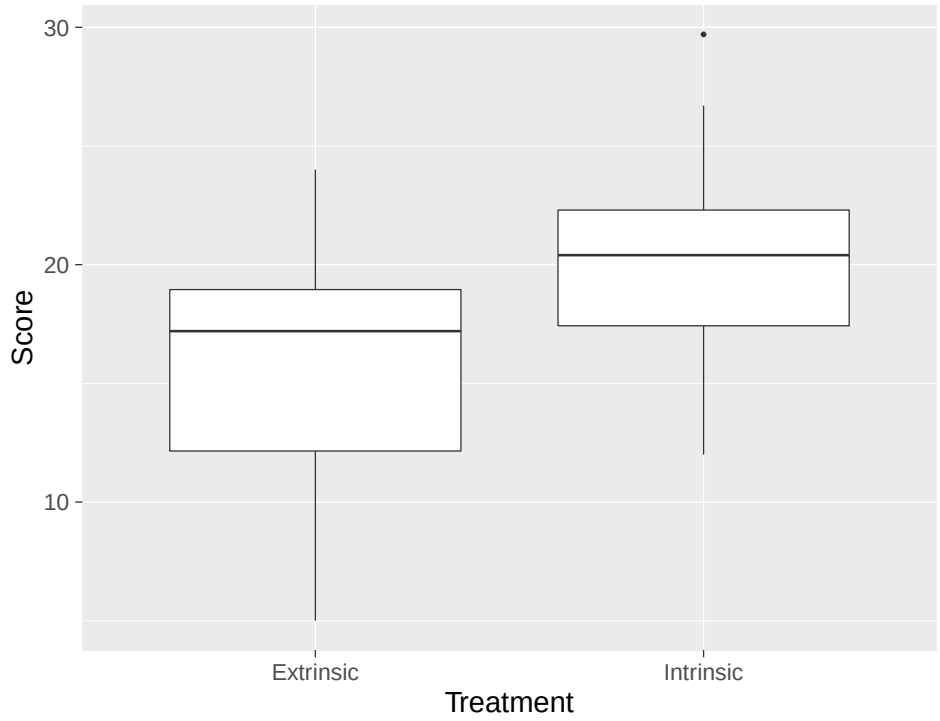
```
> # Load a couple of R packages of extra functions and data.
> library(ggplot2) # Lots of nice plotting functions.
> library(Sleuth3) # The textbook's datasets. (Note: case sensitive)
```

```
> case0101 # View the data for case study 1.1
   Score Treatment
1    5.0 Extrinsic
2    5.4 Extrinsic
3    6.1 Extrinsic
4   10.9 Extrinsic
5   11.8 Extrinsic
```

... (I snipped out most of the R output to save space.)

```
45  24.3 Intrinsic
46  26.7 Intrinsic
47  29.7 Intrinsic
```

```
> # qplot()="quick plot" from the ggplot2 library.
> qplot(Treatment, Score, data=case0101, geom="boxplot")
```



```
> # Use a t-test to compare means.
> t.test(Score~Treatment, data=case0101, var.equal=TRUE)

        Two Sample t-test

data:  Score by Treatment
t = -2.9259, df = 45, p-value = 0.005366
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.996973 -1.291432
sample estimates:
mean in group Extrinsic mean in group Intrinsic
               15.73913                19.88333
```

Function to do t-test

formula with response on left and grouping var. on right.

data frame

assume pop'n variances equal

2-sided test, so don't need alternative =

47-2

how consistent obs. data are to H_0

output

95% CI for $\mu_1 - \mu_2$

2-sided test

$\bar{Y}_1$

$\bar{Y}_2$

This line tells us how R orders groups.

See end of syllabus for details on writing a Stat. conclusion.

since p-value is small

Statistical conclusion from the *Sleuth*: This experiment provides strong evidence that receiving the "intrinsic" rather than the "extrinsic" questionnaire caused students in this study to score higher on poem creativity (two-sided p-value = 0.005 from a two-sample t-test as an approximation* to a randomization test). The estimated treatment effect—the increase in score attributed to the "intrinsic" questionnaire is 4.1 points (95% confidence interval: 1.3 to 7.0 points) on a 0–40-point scale.

Since CI is all negative.

CI endpoints are rounded and reported as pos. numbers.

Since subjects randomly assigned to groups

pt. est: 19.88333 - 15.73913 (rounded)

* R.A. Fisher (not obvious)

General Process:

1. Begin with a **research question**. More specific is better.

2. Collect **data**. ← Harder than it sounds. (see ST 431/531 - Sampling)

3. **Analyze** data.

   (a) Select H_0 and H_A based on research question. ✓

   (b) Select a test statistic e.g. t-stat

   (c) Compare observed test statistic to its sampling distribution under H_0 to assess whether observed data are consistent with H_0.

4. **Communicate** results. ← Important, and a focus of this class.

See end of syllabus for examples and guidelines.