



Sampling and Probability

Chapter 5

Sampling & Elections

- Problems with predicting elections:
 - Sample sizes are too small
 - Samples are biased (also tied to that SD thing!)
 - Samples were not independent
 - <http://www.fivethirtyeight.com>

Samples and Their Populations

- Decision making
 - The risks and rewards of sampling

Risks

- The sample might not represent the larger population.
- We might not know that the sample is misleading.
- We might reach inaccurate conclusions.
- We might make decisions based on this bad information.

Rewards

- The sample represents the larger population.
- We increase our level of confidence in our own findings.
- We reach accurate conclusions at a very low cost.
- We remain open-minded because we know samples can mislead us.
- We make wiser decisions based on the available evidence.

How to Sample

- So how do we get these people?
 - Random samples
 - Convenience samples

Random Sample

- Every member of the populations has an equal chance of being selected into the study.
- Random samples are almost never used in the social sciences – hard to access to the whole population from which to select the sample.

And ethics and stuff.

Variation & Random Sampling

- Convenience sample
 - Is one that uses participants who are readily available
 - Intro to Psyc participant pool
- Why would you use this instead of full random sampling?

Limitation of Convenience Sampling

- Generalizability – the ability to apply findings from one sample or in one context to other samples or contexts (external validity)
 - Can be improved with replication

Replication

- Replicability crisis?
 - Generally, replications are duplication of results but with different context or sample characteristics
 - Replication Q:
 - Which are you more likely to believe:
 - An effect that replicates 9/10 times
 - An effect that replicates 25/50 times

Biased Sampling

- Testimonials as Evidence? Use a volunteer sample of one person.
 - Convenience sampling where participants actively choose to be in the study



Check Your Learning

- Was random assignment used? Could it have been?
 - A health psychologist examined whether postoperative recovery time was less among patients who received counseling prior to surgery than among those who did not.
 - A clinical psychologist studied whether people with diagnosed personality disorder were more likely to miss therapy appointments than were people without diagnosed personality disorders.

Random Assignment

- All participants have an equal chance of being assigned to any level of the independent variable.
- Random selection is almost never used, but random assignment is frequently used.

Sampling Probability Quiz

- Do you agree?
 - “That woman has been playing that slot machine without success for two hours and she just quit; let’s play that one—it’s going to pay off soon.”
 - “My next-door neighbor has three boys and she’s pregnant again. This one is bound to be a girl.”

Probability

- Confirmation bias – only attending to evidence that confirms our beliefs (which means ignoring disconfirming evidence)

Probability

- Illusory correlation – believing an association between variables exists when it does not
 - Stereotypes?

Probability

- Probability – likelihood of an event occurring out of all possible events
 - So what's the probability of lefties?
 - What about our class?

Probability

- Subjective interpretations (personal probability)
 - Your judgment of a likelihood
- Objective interpretations (expected relative frequency probability)
 - The likelihood after testing many times

Probability

- Trials – all the times you test something
- Outcome – result of the trial
- Success – particular outcome we are looking for
 - So, Amanda Knox as a $\frac{2}{3}$ probability of being convicted

Calculating Probability

- Step 1. Determine the total number of trials.
- Step 2. Determine the number of these trails that are “successful” outcomes.
- Step 3. Divide the number of successful outcomes by the number of trials.

Independence

- Outcome of each trial is unrelated to outcome of previous trials.
- Gambler's fallacy is the opposite of independence:
 - The mistaken notion that the probability of a particular event changes with a long string of the same event.

Inferential Statistics

- Use rules of probability to test hypotheses
 - So it's call Hypothesis Testing
- Use probability to make decisions
 - Although ... not quite like you'd think.

Developing Hypotheses

- Usually you start by thinking about your variables/levels
 - Control group
 - Experimental group
 - Or two variables you want to correlate

Developing Hypotheses

- Then you frame those groups into TWO hypotheses
 - Null - There is no difference between levels, no relationship between variables
 - Research - There is a difference between levels, relationship between variables.
- Why two hypotheses?
 - Sometimes you predict a direction, more on that later.

Making a Decision about Hypotheses

- Reject the null hypothesis
 - Conclude that you found a difference (statistically significant)
- Fail to reject the null hypothesis
 - Conclude that you did not find a difference (not statistically significant)
- Why is it all about the null?!
 - NHST – Null hypothesis significance testing

TABLE 5-1.

The null hypothesis posits no difference, on average, whereas the research hypothesis posits a difference of some kind. There are only two decisions we can make. We can fail to reject the null hypothesis if the research hypothesis is *not* supported, or we can reject the null hypothesis if the research hypothesis *is* supported.

Hypothesis		Decision
Null Hypothesis	No mean change or difference	Fail to reject the null hypothesis (if research hypothesis is not supported)
Research Hypothesis	Mean change or difference	Reject the null hypothesis (if research hypothesis is supported)

An example

- We wanted to know if attendance in PASS sessions would lower the DFW rate for traditionally hard courses.

An example

- IV: Pass session attendance
 - Levels: Yes or No
 - NOIR: Nominal
- DV: DFW rate in percentage
 - NOIR: Ratio

An example

- Null hypothesis:
 - There is not difference in DFW rates between people who attended and did not attend PASS sessions.
 - OR
 - No sessions DFW = Sessions DFW

An example

- Research hypothesis
 - There is a difference in DFW rates for those who attended sessions versus not.
 - OR
 - No sessions DFW \neq sessions DFW

An example

Course	SI-Course	SI-Course
	Never Attended	Attended at Least Once
Biomedical Science	40.40	21.10
Chemistry	32.20	15.70
Economics	34.30	22.70
Psychology	21.10	7.40

An Example

- If we reject the null
 - We are supporting the idea that there is a difference (mainly a decrease) in DFW rates for those who attended PASS sessions.

An Example

- If we fail to reject the null
 - We have failed to find a difference between DFW rates ... did not support the research hypothesis.
That may be due to:
 - This sample
 - There really isn't a difference
 - Chance

What has this got to do with probability?

- We determine if we are going to reject or fail to reject by calculating the probability of the null hypothesis.
 - Remember it's call Null Hypothesis Significance Testing, so we test if the null is true.
 - So we want SMALL probabilities.

Type I and Type II Errors

- Statistical Inferences Can Be Wrong
- Type I errors (alpha)
 - Sins of commission – rejecting the null hypothesis when it is true
 - Saying that something happened when it didn't
- Type II errors (beta)
 - Sins of omission – failing to reject the null hypothesis when it is false
 - Saying that nothing happened when it did

Prevalence of Type I Errors

- Positive outcomes are more likely to be reported than null results.
 - Remember the study you picked on which was more likely?
 - Ways to test the rates of Type I errors, as well as the “file drawer” problem