

# Module 4a – Sampling and Intro to Inference ("Producing Data")

[Review Against All Odds: Unit 16](#)

[Review Against All Odds: Unit 17](#)

# Stat Procedure Diagram – Where we are

		<b>Descriptive Statistics (Describing Pops or Samples)</b>		<b>Inferential Statistics (from Samples)</b>
	Variable Types	Display	Describe	Estimation
Univariate	categorical (nominal or ordinal)*	Bar Graph/Pie Chart	Counts/Percentages	When binary/dichotomous: Confidence interval for proportions
	quantitative/continuous	Histogram/Stem & Leaf Box Plot	Mean/St Dev (normal) Median/Min, Q1, Q3, Max (skewed)	Confidence interval for means
		Display	Describe	Significance Tests/Hypothesis Tests
Bivariate	2 categorical	Tables or Bar Graphs	Two-way tables/Crosstabulation	Chi-square test (for goodness of fit)
	1 categorical, 1 quant.	Bar Graphs	Comparison of means/averages	T-test (one sample/group, two samples/groups) ANOVA (two or more samples/groups)
	2 quant.	Scatterplot	Correlation Coef. (Coef. of determ)/ Regression Line	T-test for correlation
		Display	Describe	Significance Tests/Hypothesis Tests
Multivariate	Response Variable is Quant.	-	Ordinary Least Squares Regression (OLS)	F-test for overall model T-tests for each explanatory variable
	Response Variable is categorical (dichotomous)	-	Logistic Regression	Chi-square tests of significance

NOTE: Items highlighted in yellow are covered in this course.

\*When a categorical variable has two categories, it is called dichotomous.

# Where do we get data?

- Two ways (Ch 8 and Ch 9) are from ***observational studies*** and ***experimental studies***
- One common type of observational study is ***Surveying***
- In either case, collecting data usually involves selecting information from a ***sample*** to reflect a ***population of interest***
  - *This is inference (or estimation)*

# Sample Types

(probability vs non-probability samples)

- ***Probability samples:*** you know the probability that any one individual in the population of interest will be selected.
  - Gold Standard: ***The Simple Random Sample (SRS)***
- ***Nonprobability samples:*** you do not know the probability...
  - *Voluntary Response Samples (ex: phone; Internet)*
  - *Convenience Samples*
  - *Snowball Samples*

# Can our sample results (data) be generalized/inferred? (in surveys)

- *Non-probability sampling introduces **Sampling Bias/Error (Bias #1)***

- We won't know if our sample responses "represent" the population.
  - Problem of non-response (who is selected/who responds)
  - What is our *response rate*?

- ***Response Bias/Error (Bias #2)*** (not necessarily deliberate)

- Mis-reporting/Misunderstanding/Mis-remembering

- ***Measurement Bias/Error (Bias #3)*** (not necessarily deliberate)

- Poor survey instrumentation (questions) or faulty record keeping
- Biased wording, Double Barreled Questions, etc.

# Sampling Error (Random Error)

- ***Random Error*** is one more form of “error”
  - But it isn’t bias and it isn’t a mistake. It’s error due to normal patterns of variation.
  - This is where the “error” in “Margin of Error” comes from.
  - We can approximate and account for this kind of error in inferential statistics (but not for biases 1, 2, and 3).

# Simple Random Sample in Practice

- You want to select 25 students with meal plans to ask questions about the cafeteria food
  - Step 1: Start with a full list of your population of interest.
  - Step 2: Number them from, for example, 001 to 879.
  - Step 3: Select three digit numbers from a “random digits table” to determine 25 individuals who will be in your sample.
- Each person’s probability of being selected is  $25/879$  (2.8% chance)
- NOTE: ***With replacement*** means the same person can be selected multiple times; ***Without replacement*** means each person can only be selected once.

# Selecting sample of 25...

Start anywhere on the table and begin collecting three digit numbers between 000 and 879

At line 101: 192; 239; 503; 405; 756; 287; 139; 640; ~~912~~...

We don't have a 912, so we throw it out and keep going until we get a sample of 25...

...531; 425; etc.

TABLE B								
Random digits								
Line								
101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893
108	60940	72024	17868	24943	61790	90656	87964	18883
109	36009	19365	15412	39638	85453	46816	83485	41979
110	38448	48789	18338	24697	39364	42006	76688	08708
111	81486	69487	60513	09297	00412	71238	27649	39950
112	59636	88804	04634	71197	19352	73089	84898	45785