Cause-Effect Relationships

- When studying the relationship between two variables, it is important whether or not a cause-effect relationship exists.

Example: GPA ⟶ Driving Record

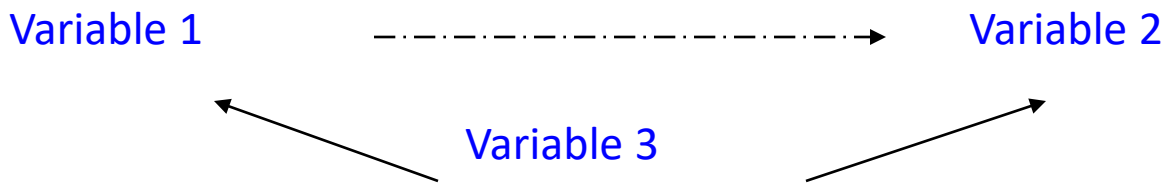
Explanatory Variable (independent) — Response Variable (dependent)

- Association between variables does not imply that a cause-effect relationship exists.

  (it merely means that the two variables tend to vary in sync with each other).

Confounding

- Confounding exists when two variables are associated due to a third hidden variable (lurking variable) has a simultaneous impact on both variables.

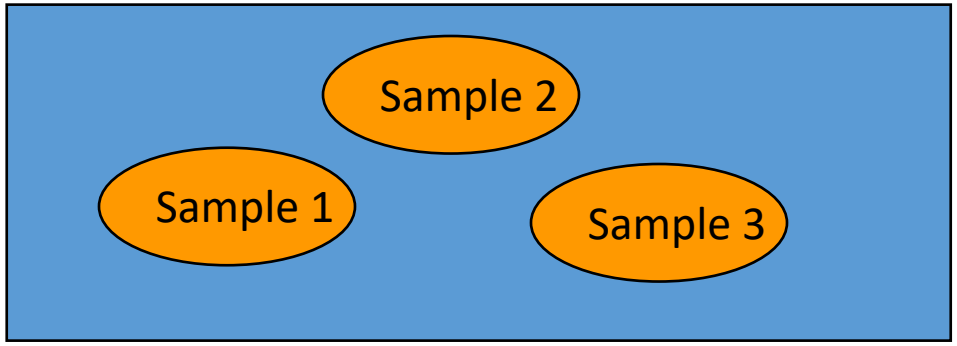


- Confounding often arise in purely observational studies.
- Confounding leads to an apparent, but not necessarily real cause-effect relationship.

Sampling Design

Population: The collection
of all subjects/cases of interest to our study

Sample: The collection of subjects/cases
actually used in our study



Sampling Design: The procedure by which we select the subjects/cases in the sample (including the size of the sample)

Sampling Designs

- Simple Random Sample (SRS)
- Stratified Random Sample
- Voluntary Response Sample
- Systematic Sample
- Multistage Sample

Simple Random Sample

SIMPLE RANDOM SAMPLE

A **simple random sample (SRS)** of size n consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected.

Population:
Alice
Bob
Carol
Dennis
Evelyn

SRS: *n=2*

Possible Samples (A=Alice, B=Bob etc.)

AB	BC	CD	DE
AC	BD	CE	
AD	AE		
AE			

Random Digits (Table B)

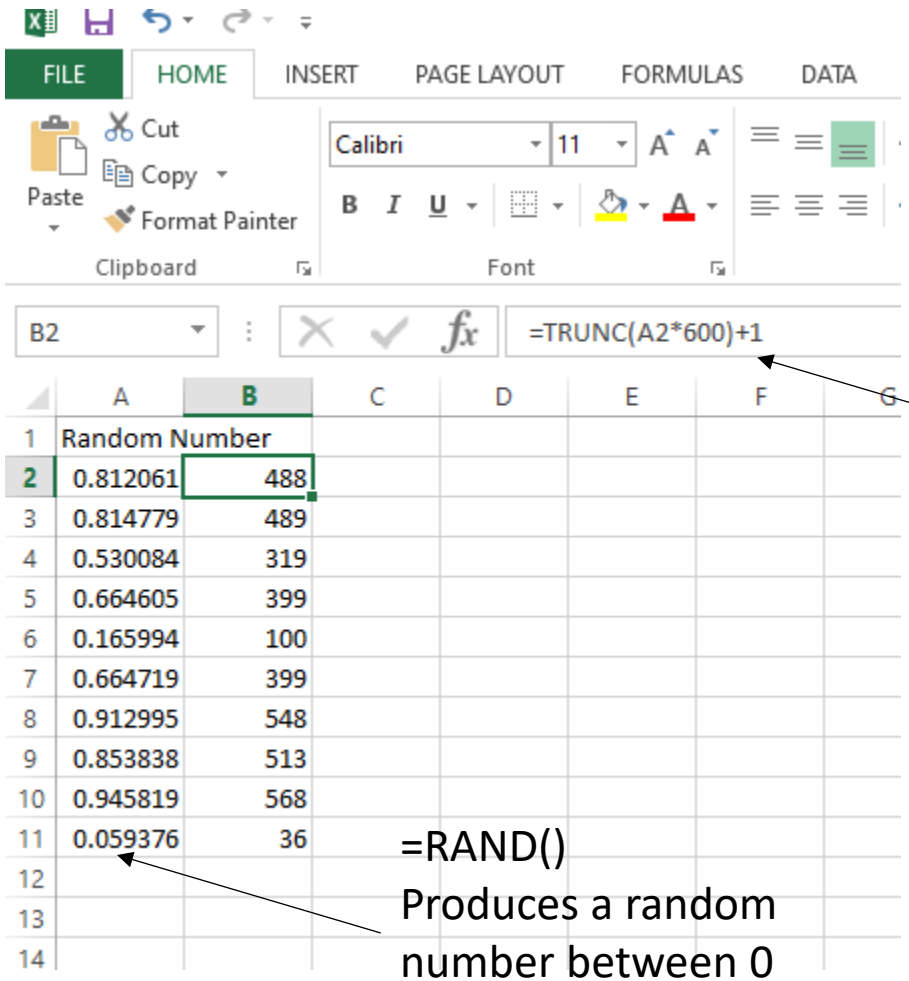
TABLE B Random Digits

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
113	62568	70206	40325	03699	71080	22553	11486	11776
114	45149	32992	75730	66280	03819	56202	02938	70915
115	61041	77684	94322	24709	73698	14526	31893	32592
116	14459	26056	31424	80371	65103	62253	50490	61181
117	38167	98532	62183	70632	23417	26185	41448	75532
118	73190	32533	04470	29669	84407	90785	65956	86382
119	95857	07118	87664	92099	58806	66979	98624	84826
120	35476	55972	39421	65850	04266	35435	43742	11937
121	71487	09984	29077	14863	61683	47052	62224	51025
122	13873	81598	95052	90908	73592	75186	87136	95761
123	54580	81507	27102	56027	55892	33063	41842	81868
124	71035	09001	43367	49497	72719	96758	27611	91596
125	96746	12149	37823	71868	18442	35119	62103	39244
126	96927	19931	36089	74192	77567	88741	48409	41903
127	43909	99477	25330	64359	40085	16925	85117	36071
128	15689	14227	06565	14374	13352	49367	81982	87209
129	36759	58984	68288	22913	18638	54303	00795	08727
130	69051	64817	87174	09517	84534	06489	87201	97245
131	05007	16632	81194	14873	04197	85576	45195	96565
132	68732	55259	84292	08796	43165	93739	31685	97150

Population Size: N=600

Sample Size: n=10

Random Digits (Excel)



Population Size: *N*=600

Sample Size: *n*=10

=TRUNC()
Drops the decimal part

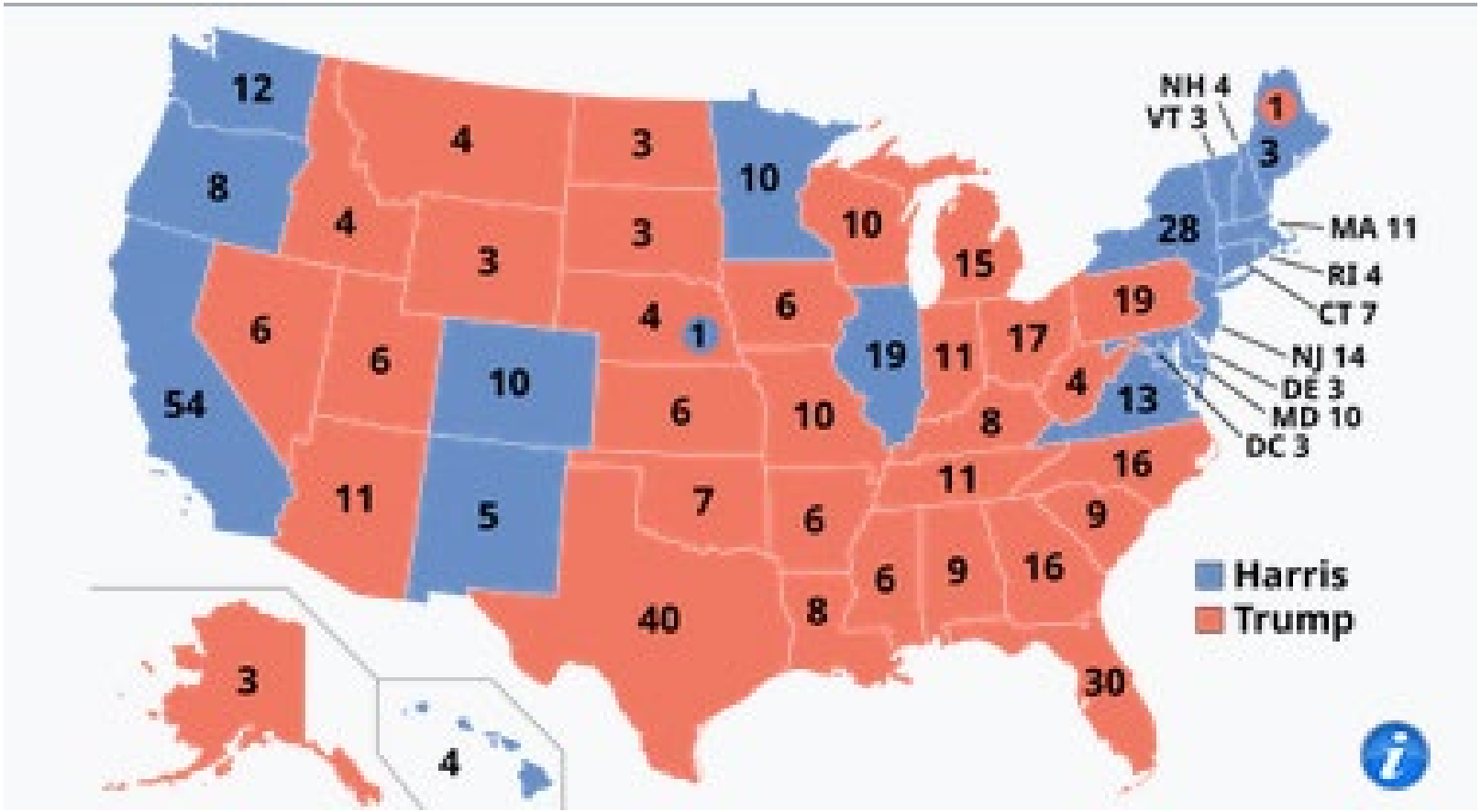
=RAND()
Produces a random number between 0 and 1

Stratified Random Sample

STRATIFIED RANDOM SAMPLE

To select a **stratified random sample,** first divide the population into groups of similar individuals, called **strata.** Then choose a separate SRS in each stratum and combine these SRSs to form the full sample.

Electoral College



538 Electoral Votes

Select a stratified sample of approx. n=1000

Voluntary Response Sample

VOLUNTARY RESPONSE SAMPLE

A **voluntary response sample** consists of people who choose themselves by responding to a general appeal. Voluntary response samples are biased because people with strong opinions, especially negative opinions, are most likely to respond.

Common Issues: Undercoverage, nonresponse, response bias

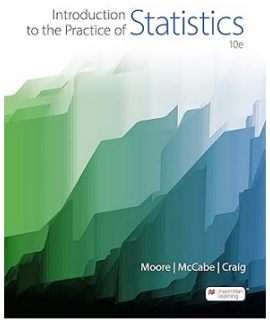
Survey Instrument: SurveyMonkey

Eastern Washington University
Spring 2020 Student Survey Results

Survey Start Date: April 24, 2020
Survey End Date: May 1, 2020

Survey Recipients: 9,996 students
Survey Invitations Opened: 2,886 (28.9%)
Survey Responses: 1,038 students (10.38%)
Complete Responses = 74.7% Partial Responses = 25.3%

Voluntary Reviews



Introduction to the Practice of Statistics 10th Edition, Kindle Edition

by DAVID S. MOORE (Author) | Format: Kindle Edition

3.9 ★★★★☆ 26 ratings | See all formats and editions

Statistics is more than just a collection of techniques and formulas. Introduction to the Practice of Statistics shows you how to produce and interpret data in the context of the real world—doing the same type of work as statisticians do in all kinds of businesses and institutions do every day.

ISBN-13	Edition	Publisher	Accessibility	Publication date
978-1319377656	10th	W.H. Freeman	Learn more	January 22, 2021

See all details

Due to its large file size, this book may take longer to download

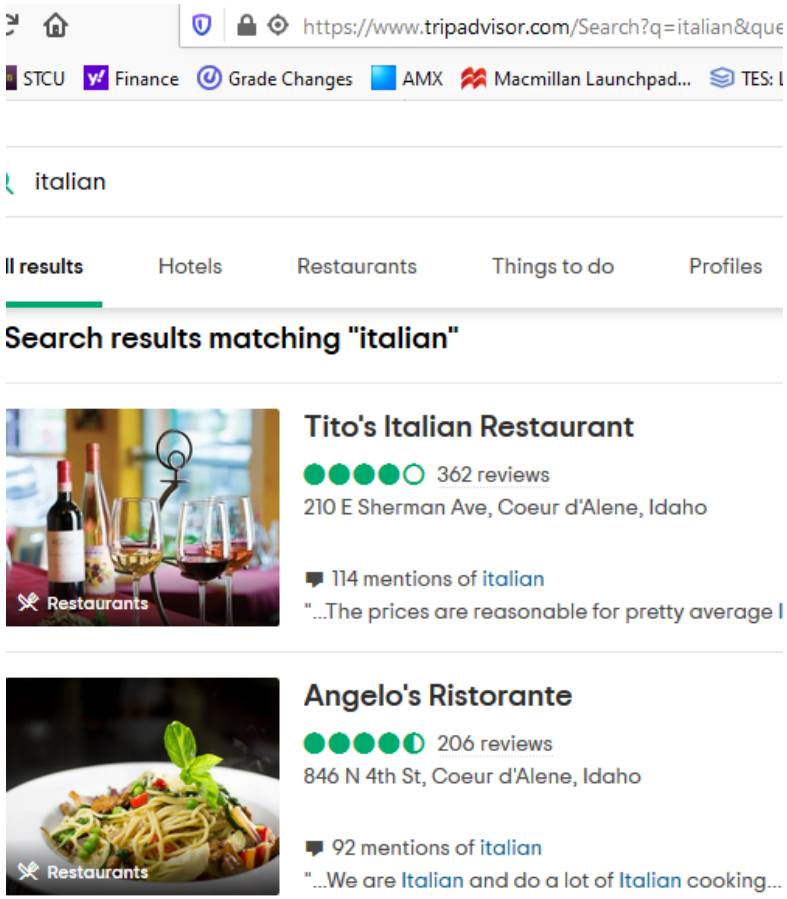
Customer reviews

★★★★☆ 3.9 out of 5

26 global ratings

5 star	59%
4 star	10%
3 star	16%
2 star	0%
1 star	15%

How customer reviews and ratings work



Other Sampling Designs

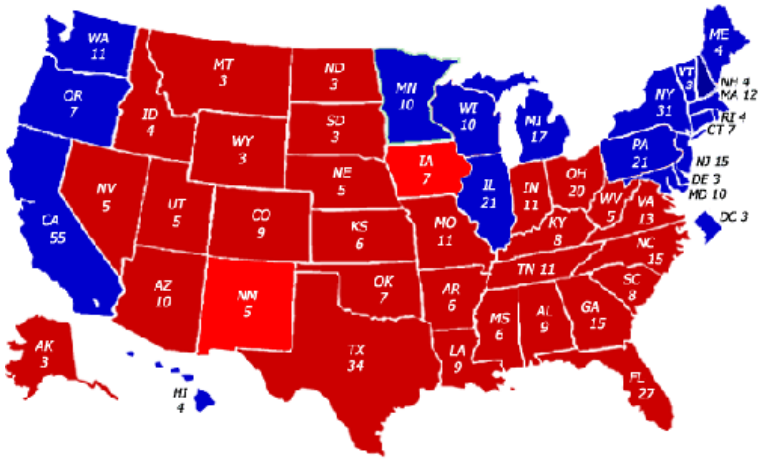
- Systematic Sampling

  Subjects are arranged in some natural order. Every k^{th} subject is selected.

Other Sampling Designs

- Multistage Sampling

  Smaller groups of subjects within the population are chosen in stages resulting in clusters of individuals.



Randomly select 5 states

Randomly select 3 precincts within each state

Randomly select 20 voters within each precinct