Economics 120A

6. Data Generation

Graham Elliott

Data Generation

We have now seen that under some assumptions we can understand the sample average and likely errors through using our probability tools.

These assumptions matter, and are the focus of this section of the course.

So what can go wrong?

We made the assumption that the data was a VSRS - basically this means that each observation is coming from the same distribution each time and that each observation could have come from the distribution with the probabilities suggested by the distribution.

We also assumed that our data comes from an X random variable that describes what we are interested in. It is common that we actually fail to measure what we care about, resulting in correct math but incorrect findings of the study.

Data Generation

This leads to a number of considerations.

1. We must be able to answer "Is the sample representative of the population that I care about?", in order to determine how far the results from the analysis of the data can be generalized.

2. Are the data i.i.d. from this representative sample? The desire for a VSRS suggests that we should collect our data in order to maximize the possibility that these assumptions underlying our statistical analysis are true.

If we are in charge of collecting data, then we may want to use techniques that ensure that it is a VSRS. Outside this framework the methods are quite complicated (but quite often solvable, the techniques are just more complicated).

Representativeness of the Data

A common problem in the interpretation of results is that the data is drawn in a way that is not representative of the problem that we are trying to solve. Alternatively readers of statistical results try to generalize the results in unreasonable directions.

Interpreting it as though it was leads to mistaken interpretations that could have been easily avoided.

This can arise many ways, both in informal statistics as well as formal statistics.

A dichotomy that is quite useful is to consider two concepts of the ‘validity’ of the results of a study:

1. Internal Validity .
2. External Validity.

Representativeness of the Data

1. Internal Validity

Internal validity suggests that there were no issues with the design of the study that results in the findings not being valid.

For the subpopulation I am looking at.

2. External Validity

External validity considers the question of whether or not the study can generalize beyond the situation of the study.

Representativeness of the Data

Internal Validity

Internal validity suggests that there were no issues with the design of the study that results in the findings not being valid.

Examples:
(a) Confirmation Bias
(b) Self Selection Bias
(c) Selection Bias
(d) Omitted effects/confounding

What you want to think about is if I had a target population, and someone tried to replicate my study with the same target population, would they get the same results.

Representativeness of the Data

External Validity

External validity considers the question of whether or not the study can generalize beyond the situation of the study.

A common dichotomy is
(a) Population
A study might be good for a subpopulation (done properly, internally valid) but does not extend to other subpopulations.

(b) Ecological
Does the study generalize to real life situations (Hawthorne effect for example)

(c) Temporal
Would the study still be valid at another time period.

What you want to think about is if I did the study properly with a different subpopulation, would they get the same results.

Self Selection Bias

Self Selection bias arises when the sample is allowed to 'select itself' and so we end up with a disproportionate sample of the population we are trying to measure.

e.g. Phone in polls (Should Mexican Avocados be banned?)
We have a random variable we want to measure from (X), where

$$X = \begin{cases} 0 & No \\ 1 & Yes \end{cases}$$

however participants must choose to call in to register their opinion, so

$$\mathrm{Y} = \begin{cases} 0 & No\ Call \\ 1 & Call \end{cases}$$

describes the random variable that measures whether a person will call in or not.

Self Selection Bias

Suppose that the joint distribution for (X,Y) is

X\Y	0	1	P(X=x)
0	0.7	0.1	0.8
1	0.1	0.1	0.2
P(Y=y)	0.8	0.2	

We are interested in the marginal distribution of X, i.e. the probability that a person wants to ban avocados.

P[X=1] = 0.2

But is this what the data from the poll measures?

Self Selection Bias

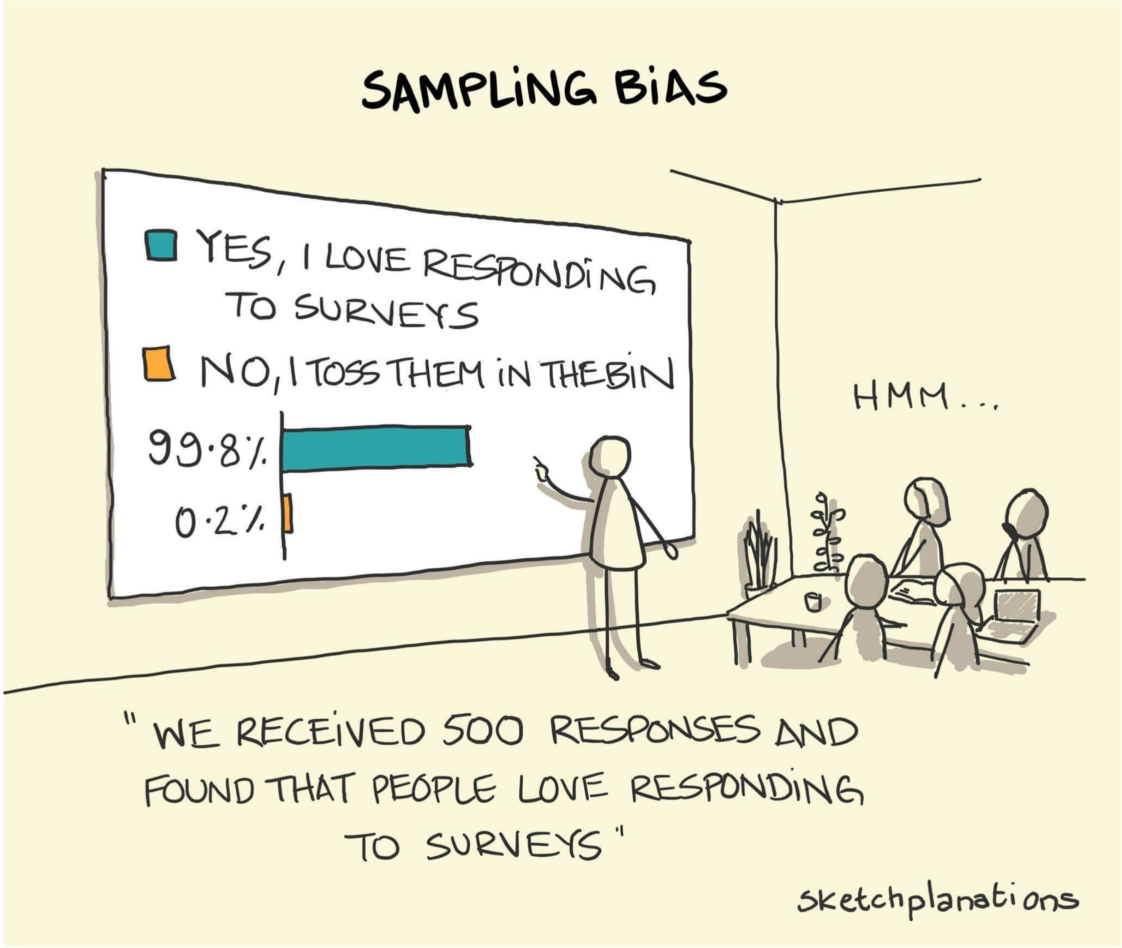
The poll measures the conditional distribution of X given Y=1, i.e. that people actually call in.

What is this?

X	$P(X \mid Y=1)$
0	$\frac{P(X=0, Y=1)}{P(Y=1)} = \frac{0.1}{0.2} = \frac{1}{2}$
1	$\frac{P(X=1, Y=1)}{P(Y=1)} = \frac{0.1}{0.2} = \frac{1}{2}$

The poll is biased in the sense that even with correct statistical work on the data we obtain, we do not get a number that is likely to be near what we want to measure.

Selection Bias



Self Selection Bias

There is the possibility that the conditional distribution that we observe and the marginal distribution are the same.

This is Independence

$x \mid y$	0	1	$P(X=x)$
0	0.64	0.16	0.8
1	0.16	0.04	0.2
$P(Y=y)$	0.8	0.2	

$$P(X=1 \mid Y=1) = \frac{0.04}{0.2} = 0.2$$

We use randomness to try and ensure this. If we cannot use randomness to select participants, we need to argue that this effect is either small or nonexistent.

SAT state averages comparison

Each year, to great consternation, the College board announces state average SAT scores.

There has of course been much effort to try and link school funding, schooling methods to outcomes we can measure such as course grades.

These are simply sample statistics, which we can think as being outcomes of a population random variable. We can apply our methods to understanding such reported information.

	State	Average Score	Rank
3	Arkansas	1194	16.0
4	California	1057	38.0
21	Massachusetts	1184	19.0
22	Michigan	1031	43.0
23	Minnesota	1263	1.0
24	Mississippi	1202	14.0
25	Missouri	1219	10.0

* 2021 numbers

SAT state averages comparison

Some conclusions from these types of numbers:

Conservatives (columnist George Will and Christian Coalition separately) Highest ranking schools get lowest public funding so this points out that it is a big government/public funding issue.

William J. Bennet (former US secretary of Education, backed up by his successor). Government money has not helped education levels, as states with larger resources do worse.

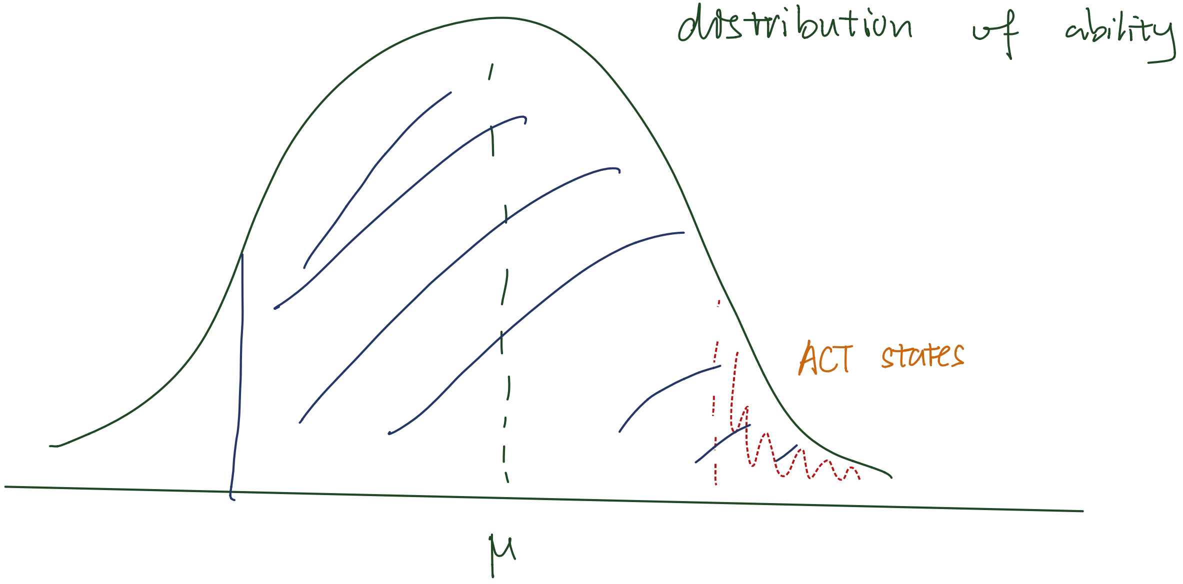
August 2005 press release from Burmeister, State Superintendent of Wisconsin Dept Public Instruction "Our students, overall, scored amongst the very best of the nation" revealing the state averages for the SAT.

Moynihan (senator and trained sociologist). Schools near Canada do better, maybe should send students there.

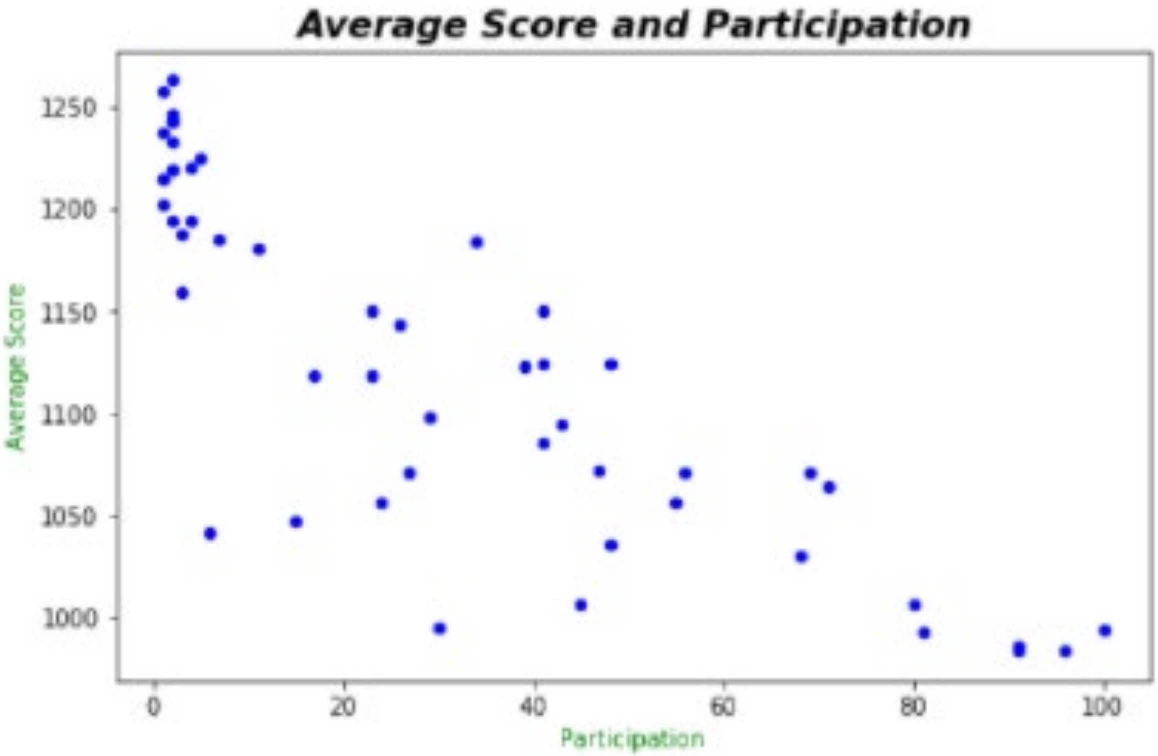
Before saying this we might ask : Is the data representative of cross state differences in ability?

SAT state averages comparison

Suppose all states were the same, with a distribution of abilities (SAT scores) with the same mean.



SAT state averages comparison



SAT state averages comparison

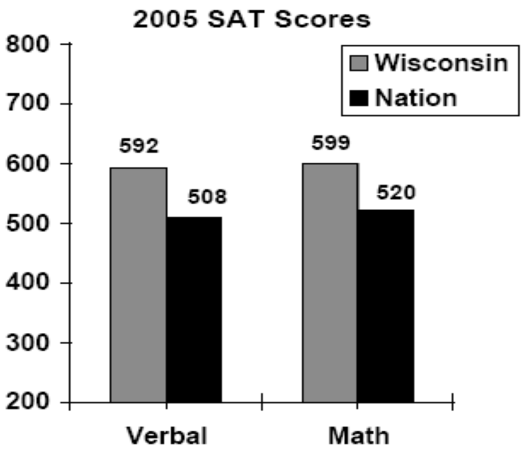
Policy responses.

Legislation was submitted in South Carolina to limit students' eligibility for the test. Now, I cannot believe that preparing for and taking an extra test lowers any particular students education (it probably helps), so this legislation seems to be aimed at trying to give the IMPRESSION that education levels are higher at the expense of most likely lowering them a little.

Georgia school board policy of not giving poor students the school code to enter on the exam. This way the students score does not get counted in the state (or school) average.

And Wisconsin?

When they reported their rankings on the ACT (their main test) they only compared themselves with the states that had more than 50% participation. So we know they knew the game they were playing!



Selection Bias

It is not just when participants select themselves into the sample that causes a bias, we can have selection bias because it is hard to obtain a random sample of participants that accurately represents our target population.

Examples:

(a) Management or financial studies of firms – often only have data for firms that do not go out of business.

(b) Studies of crime – only have reported crime and information on captured criminals, not all crimes and all criminals.

(c) Loss of participants in studies through attrition. Suppose you are studying the effects of exercise on a health outcome over a lengthy period of time. Dropouts from the study might be those that have on average worse health (so they are more likely to find participation onerous).

Extendability / External Validity.

If we cannot measure precisely the population (i.e. the one described by X) we often measure a subpopulation (this would be X|Y, conditioning on being in this subpopulation) or even worse a 'similar' population or a population in a 'similar' situation.

We can ask if results derived from a subpopulation or a similar population or situation 'extend' to the population of interest.

Example of a Subpopulation : In the past many medical studies only looked at white males – do the results extend to women or non-white people?

Example of a similar population: More common now is to run the medical study in a poorer country (East Europe etc.) as it is cheaper – do these results extend to people in the US?

Example of a similar situation: Economics experiments are usually with artificial rather than real markets – do these results extend to real market behavior?

Extendability – Violence Studies

Often it is hard to measure what we want. In studies of the effect of exposure to violent media on actual violence, how can we measure this?

Hard to measure lifetime exposure, most people do not act violently ever.

Extendability

There are many other examples. It is always a good starting point when reading a study or reported study to ask yourself

"How could they accurately measure that effect".

For example:

(a) How can we know the long term health effects of lots of new technology in food, heavy use of cellphones etc.
(b) Nature vs Nurture measurements.

Sometimes there are clear problems with extendability. Agressively cutting state welfare benefits might have strong effects on savings if recipients move to another state, but clearly this result does not extend to doing the same nationally.

Types of Studies

We can better think about whether or not we have representativeness and a VSRS through consideration of how the data was collected.

This leads to a classification of statistical studies into types.

They are

$X \mid Y = 1$

(1) Experiments
(2) Surveys
(3) Observational Studies

Each of these have meanings in statistics that are similar but more carefully defined than how we use them in conversation.

Experiments

Experiments are distinguished by the feature that the researcher both directly applies and measures the effect of some `treatment' on the units (or subjects) of the study.

So they
(a) Apply the effect we are trying to measure, so they are able to control the sample and the effect;
(b) Measure directly the effect we want to measure.

Because in an experiment we can directly control which subjects are in the experiment, we can use randomization to ensure that we obtain a VSRS and this also helps with representativeness.

Because we directly impose and measure the effect we are interested in, we can be much more sure of representativeness.

Examples

(a) Effectiveness of a Drug. A doctor may give a number of patients with a condition a drug to combat the condition. They then record how well the patient responds. Often two different levels of the drug (usually zero level which is a placebo) are given to different patients and the effect recorded. The `treatment' here is the drug being administered.

(b) Effect of speed laws on traffic fatalities. The city may impose a different speed limit on a particular highway and measure the effect of this on fatalities. The treatment here is changing the speed limit.

(c) Public goods provision Method: give some subjects tokens which can be spent on public goods or private goods. They get to keep the private goods but share the total public goods (plus a bit extra). It is best for everybody if all tokens are put in public goods, best for an individual to put all in private and everyone else in public, but worst for everyone if everyone puts their tokens into private goods. The treatment here is the payoff structure.

Examples

(d) Effect of monetary policy on economy. My NSF proposal to take over a number of small countries central bank and repeatedly shock the economy by changing the interest rates is approved. I record the effect on income, spending etc. to test our basic theories. The treatment here is the change in the interest rate.

(e) Project STAR 11,600 students in Tennessee were randomly assigned to different class sizes (small, regular, regular with teaching aide) for grades Kindergarten through 3rd. Teachers were also randomly assigned. The results were that there was a higher chance for those in the smaller class to go on to take the SAT. The effects were bigger for minorities and students from poor families. The treatment here is the size of the class.

(f) Minneapolis Domestic Crime Experiment. In this experiment police entertained three response options - arrest, removal of the offender for 8 hours, or mediation. The main outcome variable was re-offence rates (the police being called to the same place), which happened 18% of the time. It turns out that the 'lighter' touch involved a greater chance of re-offence.

Key Issues

For Representativeness:

Randomization ensures that the populations are similar so we are only measuring the effect of the treatment.

Double-blind experiments help with representativeness since there is less chance that knowledge of the treatment effects outcomes.

Compliance is important for representativeness. A lack of compliance means that the treatment could be different to what the researcher expects.

Still can have issues with extendability.

For VSRS:

Randomization basically ensures that we obtain a VSRS.

Surveys

Surveys are distinguished by the features that the researcher can no longer impose a treatment that they want to measure (often the concept of treatment in surveys is hard to think about) but they can choose which observations enter the sample.

So they
(a) Do not impose any treatment
(b) Can use selection of the observations to ensure that they are measuring what they want to measure.

Because we have control over the observations, we can use randomness. This helps with both ensuring a VSRS as well as helping with obtaining a representative sample.

Because we do not impose the treatment directly we need to be more careful about representativeness of the sample.

Examples

(a) Polls. Polls are typically surveys, we can use randomization to help ensure that we have a representative sample from our target population X. There is no real treatment here.

(b) Lottery Study. Survey lottery winners on the amount they won, their spending habits before and after the winnings (the idea is to try and get data to test hypotheses about consumption). The 'treatment' here is the lottery win, but the researcher doesn't get to impose this. The researcher can however choose which lottery winners are in the sample.

Key Issues

Randomization into the sample ensures that the random variable we are measuring and the random variable describing participation are independent. This helps with representativeness.

This does not guarantee representativeness – we might sample from an unrepresentative sub population.

The way in which the survey is worded and operated also can impact representativeness. (Foreign Aid example).

Non-response bias a big problem for representativeness.

Lying in answers is a big problem for representativeness (e.g. sensitive questions).

Observational Studies

Observational studies are distinguished by the features that the researcher can no longer impose a treatment that they want to measure and they also cannot choose which observations enter the sample.

So they
(a) Do not impose any treatment
(b) Do not select the observations themselves.

Typically the data is collected for other reasons, then we use it to try and learn about some feature.

Because we have no control over the observations, we cannot use randomness. Thus helps we cannot be sure of either a VSRS or a representative sample.

Because we do not impose the treatment directly we need to be careful about representativeness of the sample.

Here we need to spend a lot of effort convincing ourselves that the study is valid.

Examples

(a) Consumption study A researcher collects information on incomes and consumption of people from credit card data. The problem here is that if you want to find out how an increase in income affects consumption habits, it may be that the consumption habits were moving to the more expensive requiring a second job and higher income, not the other way around. This is quite different from the lottery survey, where changes in income are not planned.

(b) Effect of Monetary policy on GNP We collect interest rate data and income data and examine the correlation. But is a correlation due to an increase in interest rates decreasing income and thus monetary policy works? Or is it a case that when income falls, interest rates rise as capital is hard to find?

(c) Effect of Treatment for Heart Attacks. New expensive equipment administered quickly appears to help with survivability of heart attacks.

Examples

(d) Effect of Kindergarten on Crime Rates. New Hampshire state legislator Bob Kingsbury claims that sending kids to kindergarten causes crime rates to skyrocket. He collected data on areas in NH that have kindergartens or not and looked at crime rates in the areas. He claims areas with kindergartens have 400% greater crime. His theory is that mothers should stay at home with kids, outsourcing care leads to dysfunctional kids.

Key Issues

Unfortunately, it is observational studies that we are mostly concerned with in economics and in the social sciences more generally. This is really why we know so little about the economy whereas in other areas scientists can learn fairly fast (even if it is expensive).

Most studies with observational data spend a lot of time justifying representativeness, and most criticisms are along the lines that what we think we measure and what we measure are different.

For example in the Kindergarten example – are we measuring the effect of Kindergarten or the cumulative effects of all of the differences between areas in NH that have and do not have Kindergartens?

Examples

"Effect of Dextromethorphan, Diphenhydramine, and Placebo on Nocturnal Cough and Sleep Quality for Coughing Children and Their Parents," July 2004 issue of Pediatrics.

-Study volunteers ages two to 18 were recruited from patients with upper respiratory infections at two practices affiliated with Penn State Hershey Medical Center.
-parents were asked to answer questions about their children's conditions, such as cough severity, cough frequency, and the effect of the cough on sleep, using a seven-point symptom severity scale.
-Symptoms, as reported from the parents, had to reach a certain threshold for children to be admitted to the study. Then, each child was randomly assigned to receive dextromethorphan (often abbreviated as "DM"), diphenhydramine (an over-the-counter antihistamine) or placebo.
-a second survey was administered to parents asking the same questions as the day before.

All groups (including placebo) had similar positive effects.

Examples

Experiment, Survey or Observational Study?

To think about representativeness, first ask what we are trying to measure.

Representative?

Parents asked to evaluate improvement, no clear measure (qualitative measure)

Examples: Men want male children

Data Sources Divorce Data and child data -- Census 1940-2000. 1980, 1985, 1990 and 1995, the CPS Fertility Supplements, that reports the complete fertility and marital history of respondents. California Vital Statistics dataset 1989-2001.

Findings: Mothers with girls are significantly more likely to be divorced than mothers with boys. The effect is quantitatively substantial, explaining a 4% to 8% rise in divorce rates in the U.S

Divorced mothers with all-girl offspring less likely to remarry, when they do remarry they are more likely to get a second divorce. Women with only girls are substantially more likely to have never been married than women with only boys.

No relationship between sex of first child and marriage when child arrives later. When single mothers have had an ultrasound, they are less likely to be married (at birth) when the child is female.

In families with at least two children, the probability of having another child is higher for all-girl families that all-boy families.

Examples: Men want male children

Experiment, Survey or Observational Study?

Representativeness?

Is whether child is M or F random?

Very Simple Random Sample?

Establishing Causality

Examples
- Drug causes cure
- Budget deficit causes recession
- Tight monetary policy causes low inflation
- TV violence causes real world violence.

What we mean by causality is that if we increase X then Y changes.

There are two essential problems here

(1) if X and Y are positively correlated, is X causing Y or Y causing X? Correlation is not enough to tell us here.

(2) if X and Y are positively correlated, is it X causing Y to increase or some other factor Z causing both X and Y to increase.