Measurement and Sampling

Measurement: Concepts

A **hypothesis** is a statement about a possible relationship

Measurement: Concepts

A **hypothesis** is a statement about a possible relationship

"Strong economic performance causes more support for democracy"

Measurement: Concepts

A **hypothesis** is a statement about a possible relationship

"Strong economic performance causes more support for democracy"

In order to test a hypothesis, we must move from concepts to **measurement**

Measurement: Concepts

A **hypothesis** is a statement about a possible relationship

"Strong economic performance causes more support for democracy"

In order to test a hypothesis, we must move from concepts to **measurement**

Concept	Measurement
Economic Performance	- Average household income - Unemployment rate
Support for Democracy	- Electoral participation rates - Survey responses

Measurement: Concepts

Most measurements will be imperfect proxies of a concept.

Measurement: Concepts

Most measurements will be imperfect proxies of a concept.

We can think of this as basic **measurement error**

Measurement: Concepts

Most measurements will be imperfect proxies of a concept.

We can think of this as basic **measurement error**

When evaluating measurements, we should think about:

Measurement: Concepts

Most measurements will be imperfect proxies of a concept.

We can think of this as basic **measurement error**

When evaluating measurements, we should think about:

Validity:

Measurement: Concepts

Most measurements will be imperfect proxies of a concept.

We can think of this as basic **measurement error**

When evaluating measurements, we should think about:

Validity:

- The concept measures what it is supposed to measure

Measurement: Concepts

Most measurements will be imperfect proxies of a concept.

We can think of this as basic **measurement error**

When evaluating measurements, we should think about:

Validity:

- The concept measures what it is supposed to measure

Reliability:

Measurement: Concepts

Most measurements will be imperfect proxies of a concept.

We can think of this as basic **measurement error**

When evaluating measurements, we should think about:

Validity:

- The concept measures what it is supposed to measure

Reliability:

- The same result is found if the measurement is repeated (by someone else)

Measurement: Concepts

Most measurements will be imperfect proxies of a concept.

We can think of this as basic **measurement error**

When evaluating measurements, we should think about:

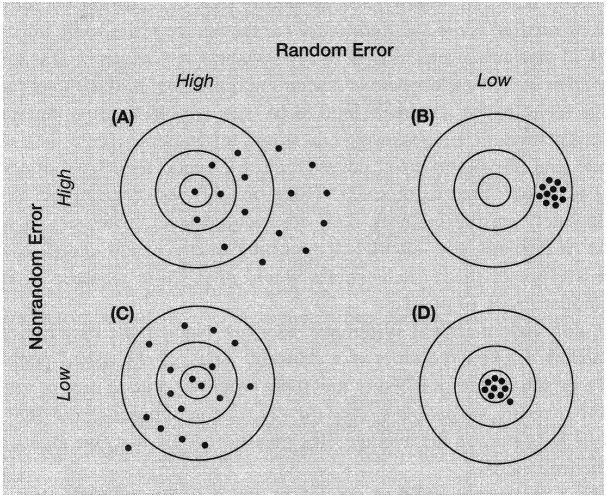
Validity:

- The concept measures what it is supposed to measure

Reliability:

- The same result is found if the measurement is repeated (by someone else)

Measurement and Error



Measurement Difficulty

Some concepts are easier to measure than others

Measurement Difficulty

Some concepts are easier to measure than others

Easily Validated — Tough to Validate

Age: Date of Birth

Measurement Difficulty

Some concepts are easier to measure than others

Easily Validated — Tough to Validate

Age: Date of Birth

Wealth: Annual Income

Measurement Difficulty

Some concepts are easier to measure than others

Easily Validated — Tough to Validate

Age: Date of Birth

Wealth: Annual Income

Support for Democracy

Measurement Difficulty

Some concepts are easier to measure than others

Easily Validated — Tough to Validate

Age: Date of Birth

Wealth: Annual Income

Support for Democracy

Support for Immigration

Measurement Difficulty

Some concepts are easier to measure than others

Easily Validated — Tough to Validate

Age: Date of Birth

Wealth: Annual Income

Support for Democracy

Support for Immigration

How do we measure most of these things?

Measurement Difficulty

Some concepts are easier to measure than others

Easily Validated — Tough to Validate

Age: Date of Birth

Wealth: Annual Income

Support for Democracy

Support for Immigration

How do we measure most of these things? Surveys!

Survey sampling

Reminder: Learning about a fairly large population by interviewing a small fraction of it.

Survey sampling

Reminder: Learning about a fairly large population by interviewing a small fraction of it.

- Assuming the subset is randomly selected

Survey sampling

Reminder: Learning about a fairly large population by interviewing a small fraction of it.

- Assuming the subset is randomly selected
- Assuming the subset is representative

How do we actually run a survey?

Two common methods:

How do we actually run a survey?

Two common methods:

- Probability sampling: every unit of a target population has a known AND non-zero probability of being selected

How do we actually run a survey?

Two common methods:

- Probability sampling: every unit of a target population has a known AND non-zero probability of being selected
- Simple *random sampling* (SRS): selects a pre-determined number of respondents from a target population with each potential respondent having an equal chance of being selected

Assessing a sample: Representativeness

How similar is the sample to the 'real-world?'

Assessing a sample: Representativeness

How similar is the sample to the 'real-world?'

- If we repeat the sampling procedure many times, the features of each resulting sample would not be exactly the same as those of the population, but on average (across all the samples) would be identical.

Assessing a sample: Representativeness

How similar is the sample to the 'real-world?'

- If we repeat the sampling procedure many times, the features of each resulting sample would not be exactly the same as those of the population, but on average (across all the samples) would be identical.

We infer population characteristics using those of a representative sample obtained through probability sampling procedures.

Assessing a sample: Representativeness

How similar is the sample to the 'real-world?'

- If we repeat the sampling procedure many times, the features of each resulting sample would not be exactly the same as those of the population, but on average (across all the samples) would be identical.

We infer population characteristics using those of a representative sample obtained through probability sampling procedures.

- Candidate support

Assessing a sample: Representativeness

How similar is the sample to the 'real-world?'

- If we repeat the sampling procedure many times, the features of each resulting sample would not be exactly the same as those of the population, but on average (across all the samples) would be identical.

We infer population characteristics using those of a representative sample obtained through probability sampling procedures.

- Candidate support
- Health

Assessing a sample: Representativeness

How similar is the sample to the 'real-world?'

- If we repeat the sampling procedure many times, the features of each resulting sample would not be exactly the same as those of the population, but on average (across all the samples) would be identical.

We infer population characteristics using those of a representative sample obtained through probability sampling procedures.

- Candidate support
- Health
- TV and movie preferences

Who to sample? Sampling frames

- Sampling frame: a list of individuals in the population to sample from

Who to sample? Sampling frames

- Sampling frame: a list of individuals in the population to sample from
  - Addresses

Who to sample? Sampling frames

- Sampling frame: a list of individuals in the population to sample from
  - Addresses
  - Phone numbers (RDD)

Who to sample? Sampling frames

- Sampling frame: a list of individuals in the population to sample from
  - Addresses
  - Phone numbers (RDD)
  - Census block

Who to sample? Sampling frames

- Sampling frame: a list of individuals in the population to sample from
  - Addresses
  - Phone numbers (RDD)
  - Census block
- But massive amounts of missing data

Who to sample? Sampling frames

- Sampling frame: a list of individuals in the population to sample from
  - Addresses
  - Phone numbers (RDD)
  - Census block
- But massive amounts of missing data
  - Cell phones, multiple units at each location/number

Who to sample? Sampling frames

- Sampling frame: a list of individuals in the population to sample from
  - Addresses
  - Phone numbers (RDD)
  - Census block
- But massive amounts of missing data
  - Cell phones, multiple units at each location/number

A more complex process

- *Multi-stage cluster sampling*: proceeds in multiple stages by sampling larger units first and then randomly selecting smaller units within each of the selected larger units.

Who to sample? Sampling frames

- Sampling frame: a list of individuals in the population to sample from
  - Addresses
  - Phone numbers (RDD)
  - Census block
- But massive amounts of missing data
  - Cell phones, multiple units at each location/number

A more complex process

- *Multi-stage cluster sampling*: proceeds in multiple stages by sampling larger units first and then randomly selecting smaller units within each of the selected larger units.
  - A random selection from a random subset is just as random as an initial random selection

What happens if people we select don't respond?

Things that can go wrong:

What happens if people we select don't respond?

Things that can go wrong:

- *Unit non-response*: the failure to reach selected units

What happens if people we select don't respond?

Things that can go wrong:

- *Unit non-response*: the failure to reach selected units
- *Item non-response*: respondents refuse to answer certain survey questions.

What happens if people we select don't respond?

Things that can go wrong:

- *Unit non-response*: the failure to reach selected units
- *Item non-response*: respondents refuse to answer certain survey questions.
- *Social desirability bias*: the problem where respondents choose an answer that is seen as socially desirable regardless of what their truthful answer is