EPSE 592: Design & Analysis of Experiments

Ed Kroc

University of British Columbia

ed.kroc@ubc.ca

January 9, 2020

Syllabus and Course Policies

- Check website: ekroc.weebly.com → "Teaching"

8 pillars of statistics

- Design
- Sampling
- Measurement
- Estimation
- Inference
- Modelling
- Computation
- Communication

Note: some items may be repeated or concurrent

Each item will require different assumptions which must be validated/checked!

Experiments

- Statement of problem
- Choice of response (dependent) variable
- Selection of factors/descriptors (independent variables) to be varied
- Choice of levels for these factors
    - quantitative or qualitative
    - fixed or random
- Number of observations to be taken
- Order of experimentation, choice of sample units
- Method(s) of randomization to be used
- Mathematical model to describe experiment
- Hypotheses to be tested

Only after completing all these steps do we begin *data collection, processing, computation, analysis, interpretation, and communication*

Randomization

Many different kinds of randomization; two most important:

- random selection of experimental units from some population (*random sampling*)
- random assignment and application of experimental units to different factor levels (*random assignment to treatment*)

Random sampling allows for *generalizability*.
Random assignment to treatment allows for *causal* conclusions.

Complete randomization is the gold-standard.

Unfortunately, often infeasible or impossible; only partial or restricted randomization possible.

Examples

What kind of randomizations are possible?

- Laboratory biology:
  - Subject 12 individuals, 3 each from 1 of 4 different species, to three different noise treatments and assess their change in corticosteroid levels.

Examples

What kind of randomizations are possible?

- Laboratory biology:
  - Subject 12 individuals, 3 each from 1 of 4 different species, to three different noise treatments and assess their change in corticosteroid levels.
- Ecology:
  - Select 8 acreages in shrubland at varying distances from main roads and assess ecological health via biomass and biodiversity metrics.

Examples

What kind of randomizations are possible?

- Laboratory biology:
  - Subject 12 individuals, 3 each from 1 of 4 different species, to three different noise treatments and assess their change in corticosteroid levels.
- Ecology:
  - Select 8 acreages in shrubland at varying distances from main roads and assess ecological health via biomass and biodiversity metrics.
- Pedagogy:
  - Subject 10 sections of the same class to either an "active learning" or a "traditional learning" strategy for the entire term.

Examples

What kind of randomizations are possible?

- Laboratory biology:
  - Subject 12 individuals, 3 each from 1 of 4 different species, to three different noise treatments and assess their change in corticosteroid levels.
- Ecology:
  - Select 8 acreages in shrubland at varying distances from main roads and assess ecological health via biomass and biodiversity metrics.
- Pedagogy:
  - Subject 10 sections of the same class to either an "active learning" or a "traditional learning" strategy for the entire term.
- Psychology:
  - Subject 30 people struggling with depression to either a one-on-one or a group therapy intervention for 3 months and assess change in sleeping habits.

Control

"Control" can refer to one of two ideas:

(1) creating a *control group* that is statistically identical to the *experimental group* in every way *except for the application of the experimental treatment*: e.g.
   - randomly select $2N$ experimental units from study population
   - randomly assign N units to receive the experimental treatment (*experimental group*) while the other N units receive no treatment or placebo (*control group*)

Control

"Control" can refer to one of two ideas:

(1) creating a *control group* that is statistically identical to the *experimental group* in every way *except for the application of the experimental treatment*: e.g.

- randomly select $2N$ experimental units from study population
- randomly assign N units to receive the experimental treatment (*experimental group*) while the other N units receive no treatment or placebo (*control group*)

(2) controlling/fixing a property (e.g. factor level) of an experimental unit: e.g.

- randomly select experimental units from a study (sub)-population with only certain fixed factor levels (e.g. only patients with high blood pressure)
- experiment is then *restricted* to this (sub)-population

Examples

Are control groups feasible for these examples? What about controlling certain factor levels?

- Laboratory biology:
  - Subject 12 individuals, 3 from 1 of 4 different species, to three different noise treatments and assess their change in corticosteroid levels.

Examples

Are control groups feasible for these examples? What about controlling certain factor levels?

- Laboratory biology:
  - Subject 12 individuals, 3 from 1 of 4 different species, to three different noise treatments and assess their change in corticosteroid levels.
- Ecology:
  - Select 8 acreages in shrubland at varying distances from main roads and assess ecological health via biomass and biodiversity metrics.

Examples

Are control groups feasible for these examples? What about controlling certain factor levels?

- Laboratory biology:
  - Subject 12 individuals, 3 from 1 of 4 different species, to three different noise treatments and assess their change in corticosteroid levels.
- Ecology:
  - Select 8 acreages in shrubland at varying distances from main roads and assess ecological health via biomass and biodiversity metrics.
- Pedagogy:
  - Subject 10 sections of the same class to either an "active learning" or a "traditional learning" strategy for the entire term.

Examples

Are control groups feasible for these examples? What about controlling certain factor levels?

- Laboratory biology:
  - Subject 12 individuals, 3 from 1 of 4 different species, to three different noise treatments and assess their change in corticosteroid levels.
- Ecology:
  - Select 8 acreages in shrubland at varying distances from main roads and assess ecological health via biomass and biodiversity metrics.
- Pedagogy:
  - Subject 10 sections of the same class to either an "active learning" or a "traditional learning" strategy for the entire term.
- Psychology:
  - Subject 30 people struggling with depression to either a one-on-one or a group therapy intervention for 3 months and assess change in sleeping habits.

Causality & causal inference

Three necessary ingredients:

- Randomization
  - Random assignment to and application of treatment
  - (Randoms sampling allows results to generalize)
- Manipulation
  - Researcher manipulates levels (treatments) of interest
- Control
  - Control group essential (placebo or no treatment)

All *confounding* factors should balance equally across treatments and controls; complete randomization ensures this (for large and representative enough sample sizes)

Confounding

A *confounding (lurking) variable* is one that is associated to both the dependent (response) and independent variable(s). These are pre-existing, unaccounted for differences between the treatment groups.

- Often try to account for confounding variables by incorporating them into an analysis.
- However, unless subjects are randomly assigned to treatments and control, we can *never* be sure that unknown or unaccounted for confounders are not distorting our results.
- Randomization of treatment eliminates the possibility of confounding.
  - *Very* difficult to accomplish in practice.

Example: confounding in a clinical trial

- Suppose we want to assess the efficacy of a new drug to treat high blood pressure. From a registry of patients with no known major comorbidities, we randomly select 50 to receive the new drug, and 50 to receive the current "best" medication.

Example: confounding in a clinical trial

- Suppose we want to assess the efficacy of a new drug to treat high blood pressure. From a registry of patients with no known major comorbidities, we randomly select 50 to receive the new drug, and 50 to receive the current "best" medication.
  - Will experimental units (patients) always take their assigned treatment?

Example: confounding in a clinical trial

- Suppose we want to assess the efficacy of a new drug to treat high blood pressure. From a registry of patients with no known major comorbidities, we randomly select 50 to receive the new drug, and 50 to receive the current "best" medication.
    - Will experimental units (patients) always take their assigned treatment?
    - Will experimental units always stay on their assigned treatment?

Example: confounding in a clinical trial

- Suppose we want to assess the efficacy of a new drug to treat high blood pressure. From a registry of patients with no known major comorbidities, we randomly select 50 to receive the new drug, and 50 to receive the current "best" medication.
  - Will experimental units (patients) always take their assigned treatment?
  - Will experimental units always stay on their assigned treatment?
- Even the best controlled experiments usually suffer from some degree of confounding.

Observational studies

- An *observational study* is usually defined as one where random assignment of the treatment of interest to the experimental units is not possible.

- Ubiquitious in the social sciences, but also very common in the natural sciences.

Experimental vs. observational paradigm

	Experimental studies	Observational studies
Confounding variables	No (ideally)	Yes
Control group	Yes	No
Manipulating factor levels	Yes	No
Type of inference	Causal (ideally)	Correlational
Generalizability	Depends...	Depends...

- There also exist hybrid study designs (quasi-experiments):
  - E.g. We may manipulate some factor levels in an observational study
  - E.g. We may assign a control group (i.e. one that receives no treatment) in an observational study

Generalizability

	Random assignment	No random assignment	
Random sampling	Causal inf. Generalizable to pop.	Correlational inf. Generalizable to pop.	Generalizability
No random sampling	Causal inf. only for sample	Correlational inf. only for sample	No generalizability

Generalizability

	Random assignment	No random assignment	
Random sampling	Causal inf. Generalizable to pop.	Correlational inf. Generalizable to pop.	Generalizability
No random sampling	Causal inf. only for sample	Correlational inf. only for sample	No generalizability

- ideal experiment
- good observational (and quasi-experimental) designs
- most experiments
- pilot studies (narrow scope)

Probability and Statistics review...

Probability & statistics

- **Probability** is a mathematical abstraction. It is usually used to describe the likelihood, or chance, of observing some event; e.g. sun or rain tomorrow, the height of the tallest person in the room.

Probability & statistics

- **Probability** is a mathematical abstraction. It is usually used to describe the likelihood, or chance, of observing some event; e.g. sun or rain tomorrow, the height of the tallest person in the room.
  - The most essential feature of a random phenomenon is its *distribution*; i.e. a function that tells us how likely is each possible outcome of the random phenomenon.

Probability & statistics

- **Probability** is a mathematical abstraction. It is usually used to describe the likelihood, or chance, of observing some event; e.g. sun or rain tomorrow, the height of the tallest person in the room.
  - The most essential feature of a random phenomenon is its *distribution*; i.e. a function that tells us how likely is each possible outcome of the random phenomenon.
- **Statistics** can be thought of as the discipine of *applied probability*. More specifically, it is the study of *uncertainty*.

Probability & statistics

- **Probability** is a mathematical abstraction. It is usually used to describe the likelihood, or chance, of observing some event; e.g. sun or rain tomorrow, the height of the tallest person in the room.
  - The most essential feature of a random phenomenon is its *distribution*; i.e. a function that tells us how likely is each possible outcome of the random phenomenon.
- **Statistics** can be thought of as the discipine of *applied probability*. More specifically, it is the study of *uncertainty*.
  - It is the goal of descriptive statistics to describe some random distribution via a *sample* of observations; e.g. sample mean as a "typical" value of the distribution.
  - It is the goal of inferential statistics to infer properties of some random distribution via a *sample* of observations and to *quantify* our confidence in these inferences; e.g. a 95% confidence interval for the mean.

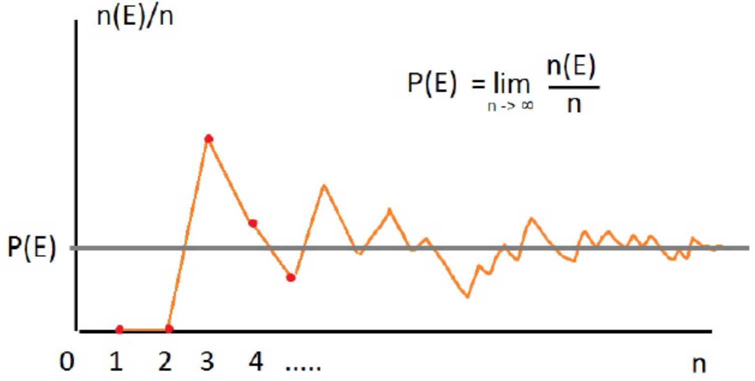
A formal definition of the probability of an event E

Consider repeating a random experiment (observation) many times; e.g. flipping a coin. If the sequence of trials are *independent* (i.e. the result of one trial of the experiment (observation) does not affect the result of any other trials), then the probability that an event E will occur is given by

$$\Pr(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

where n is the number of repetitions, and $n(E)$ is the number of times the event E occurs in n repetitions. $\Pr(E)$ is the limiting or long-run relative frequency.

A formal definition of the probability of an event E



Note that we must always have $0 \leqslant \Pr(E) \leqslant 1$.

Other definitions of probability?

- Does the previous definition always make sense? Counterexamples?

Conditional Probability

- Experiment: toss a fair coin twice: set of all outcomes is $S = \{HH,\ HT,\ TH,\ TT\}$
- $\Pr(\textit{tossing 2 heads}) = 1/4$ (all outcomes equally likely)

Conditional Probability

- Experiment: toss a fair coin twice: set of all outcomes is $S = \{HH,\ HT,\ TH,\ TT\}$
- $\Pr(\textit{tossing } 2 \textit{ heads}) = 1/4$ (all outcomes equally likely)
- Suppose you are given *additional information* that at least 1 head was tossed in this experiment. What is the probability of tossing 2 heads, *given this extra information*?

Conditional Probability

- Experiment: toss a fair coin twice: set of all outcomes is $S = \{HH,\ HT,\ TH,\ TT\}$
- $\Pr(\textit{tossing 2 heads}) = 1/4$ (all outcomes equally likely)
- Suppose you are given *additional information* that at least 1 head was tossed in this experiment. What is the probability of tossing 2 heads, *given this extra information*?
- With the given information, the only possibilites left from S are: $\{HH,\ HT,\ TH\}$.

Conditional Probability

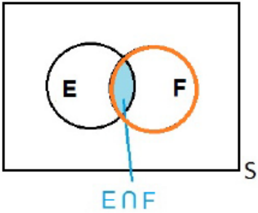
- Experiment: toss a fair coin twice: set of all outcomes is $S = \{HH,\ HT,\ TH,\ TT\}$
- $\Pr(\textit{tossing } 2 \textit{ heads}) = 1/4$ (all outcomes equally likely)
- Suppose you are given *additional information* that at least 1 head was tossed in this experiment. What is the probability of tossing 2 heads, *given this extra information*?
- With the given information, the only possibilites left from S are: $\{HH,\ HT,\ TH\}$.
- Thus, the probability that we toss 2 heads, *given that at least 1 head was tossed*, is...?

Conditional Probability

- Experiment: toss a fair coin twice: set of all outcomes is $S = \{HH,\ HT,\ TH,\ TT\}$
- $\Pr(\textit{tossing } 2 \textit{ heads}) = 1/4$ (all outcomes equally likely)
- Suppose you are given *additional information* that at least 1 head was tossed in this experiment. What is the probability of tossing 2 heads, *given this extra information*?
- With the given information, the only possibilites left from S are: $\{HH,\ HT,\ TH\}$.
- Thus, the probability that we toss 2 heads, *given that at least 1 head was tossed*, is...?
- $= \frac{1}{3}$.

Conditional Probability

- In general, given that an event F has occurred, the probability that another event E occurs is called the *conditional probability of E given F*.



- Notation and formula:

$$\Pr(E \mid F) = \frac{\Pr(E \cap F)}{\Pr(F)} = \frac{\Pr(E \text{ and } F)}{\Pr(F)}$$

Conditional Probability

Note:

$$\Pr(E \mid F) \neq \Pr(F \mid E)$$

If you assume different information, you should not expect to end up with the same results!

Independence of Events

Definition

Two events E and F are said to be **independent** if and only if $\Pr(E \mid F) = \Pr(E)$ or $\Pr(F \mid E) = \Pr(F)$.

- By definition of conditional probability then, we have

$$\Pr(E \cap F) = \Pr(E) \cdot \Pr(F) \quad \text{if and only if } E,\ F \text{ are independent.}$$

- This definition matches with our intuition: if two events are independent, then the fact that one event happens should *not* have any affect on how likely the other event is to happen.

The Prosecutor's Fallacy

The Prosecutor's Fallacy is a common probability *misconception*: the fallacy is thinking that $\Pr(A \cap B)$ is the same as $\Pr(A \mid B)$.

- This is obviously false! Only true if $\Pr(B) = 1$ or if $\Pr(A \cap B) = 0$. Recall that

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

- This fallacy is quite common and can have many distressing consequences...

The Case of Sally Clark

- In 1998, Sally Clark was accused of murdering her two infant sons. One died in 1996 at eleven weeks old. The second died a year later at eight weeks of age.

- Sir Roy Meadow, pediatrician and expert witness for the prosecution, testified that the chance of two children in the same family dying from Sudden Infant Death Syndrome (SIDS) was about $(1/8500)^2$, or 1 in 73 million.

- On the strength of this testimony alone, Clark was convicted in 1999. The Royal Statistical Society then pointed out the flaws in the argument. What are they?

The Case of Sally Clark

- Flaw #1: The events of two *siblings* dying from SIDS are *not* independent. There is a genetic component! In reality, the probability of two children from the same family dying of SIDS is much closer to $1/8500$ than to $(1/8500)^2$.

- Flaw #2: Meadow confused the conditional and unconditional probabilities (the Prosecutor's Fallacy).

  Let I : event that Clark is innocent of murder, E : event of two dead children (the evidence).

  We know that in general,

$$\Pr(I \mid E) \neq \Pr(E \cap I).$$

The Case of Sally Clark

Now,

$$\begin{aligned}\Pr(I \mid E) &= \frac{\Pr(I \cap E)}{\Pr(E)} \\ &= \frac{\Pr(I \cap E)}{\Pr(\{I \cap E\} \cup \{I^c \cap E\})} \\ &= \frac{\Pr(I \cap E)}{\Pr(I \cap E) + \Pr(I^c \cap E)}\end{aligned}$$

What are the events $I \cap E$ and $I^c \cap E$?

- $I \cap E$ is the event of the two chidren dying by SIDS.
- $I^c \cap E$ is the event of the two children dying by murder.

Double SIDS is rare, but double murder is much, much rarer! So,

$$\Pr(I^c \cap E) \ll \Pr(I \cap E).$$

The Case of Sally Clark

$$\Pr(I^c \cap E) \ll \Pr(I \cap E)$$

means:

$$\Pr(I \cap E) + \Pr(I^c \cap E) \ll \Pr(I \cap E) + \Pr(I \cap E)$$

$$\frac{1}{\Pr(I \cap E) + \Pr(I^c \cap E)} \gg \frac{1}{\Pr(I \cap E) + \Pr(I \cap E)}$$

$$\frac{\Pr(I \cap E)}{\Pr(I \cap E) + \Pr(I^c \cap E)} \gg \frac{\Pr(I \cap E)}{\Pr(I \cap E) + \Pr(I \cap E)}$$

$$\Pr(I \mid E) \gg \frac{1}{2}$$

The Case of Sally Clark

$$\Pr(I^c \cap E) \ll \Pr(I \cap E)$$

means:

$$\begin{aligned}\Pr(I \mid E) &= \frac{\Pr(I \cap E)}{\Pr(I \cap E) + \Pr(I^c \cap E)} \\ &\gg \frac{\Pr(I \cap E)}{\Pr(I \cap E) + \Pr(I \cap E)} = \frac{1}{2}\end{aligned}$$

So, $\Pr(I \mid E) \approx 1$!

Moral of the story 1: **circumstantial evidence of a rare event is very weak evidence.**

Moral of the story 2: **conditional information is radically different from unconditional information.**

The Case of Sally Clark

- Sally Clark's conviction was overturned in 2003, after she had already spent four years in jail.
- In prison, Clark developed psychological problems and an alcohol dependency.
- Sally Clark died of alcohol poisoning in 2007.

Random Variables

Definition

A **random variable** is a function that maps events in a sample space to the real numbers. We use uppercase letters to denote a random variable, and lowercase letters to denote sample realizations of that random variable.

Random Variables

Definition

A **random variable** is a function that maps events in a sample space to the real numbers. We use uppercase letters to denote a random variable, and lowercase letters to denote sample realizations of that random variable.

Example: Toss a fair coin three times:

$$S = \{HHH,\ HHT,\ HTH,\ THH,\ HTT,\ THT,\ TTH,\ TTT\}$$

We can define a random variable X to be the number of heads observed in the three tosses:

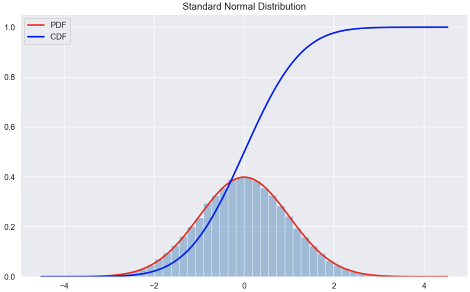
Event	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X = x$	3	2	2	2	1	1	1	0

Random Variables

- All r.v.'s come equipped with a *cumulative distribution function (CDF)* that lets us figure out probabilties of events.
- The CDF simply *accumulates* probabilties up to a certain value:

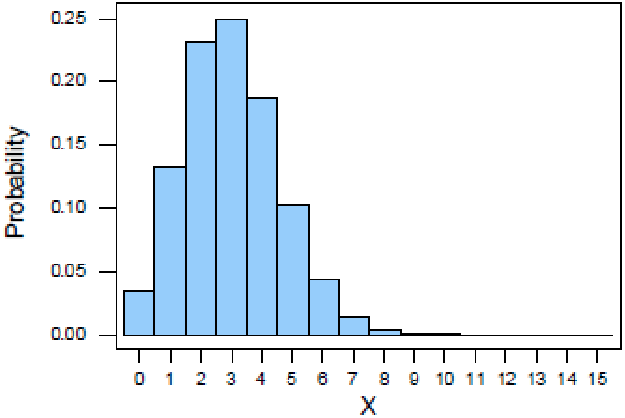
$$\Pr(X \leqslant x),$$

where X is the random variable, and $x \in \mathbb{R}$.



- Discrete r.v.'s also have a *probability mass function (PMF)* that tells us the probabilities of single events:

$$\Pr(X = x).$$

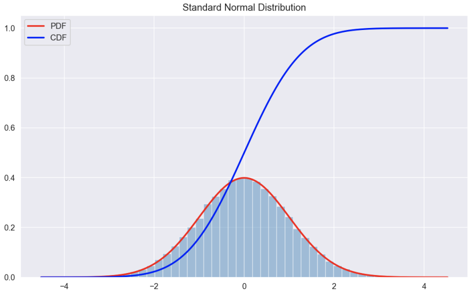


Random Variables

- Continuous r.v.'s instead have a *probability density function (PDF)*, denoted $f(x)$, that allows us to write:

$$\Pr(a \leqslant X \leqslant b) = \int_a^b f(x)dx,$$

for any $a, b \in \mathbb{R}$.



Expectation of a Random Variable

Definition

The **expectation** of X (discrete) is given by

$$\mathbb{E}(X) = \sum_x x \cdot \Pr(X = x).$$

The **expectation** of X (continuous) is given by

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx.$$

The expectation of X is also referred to as the **expected value** of X or the **mean** of X. We often denote the mean by μ or μ_X.

Note: the *expectation* generalizes the idea of the *simple average* of a bunch of numbers.

Variance and Standard Deviation of a Random Variable

Definition

We define the **variance** of a r.v. X as

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mu)^2].$$

For X discrete, this is

$$\mathrm{Var}(X) = \sum_x (x - \mu)^2 \Pr(X = x).$$

For X continuous, this is

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

Note: the *variance* of a random variable quantifies how likely it is that the random variable takes on values *away* from its mean/expectation.

Variance and Standard Deviation of a Random Variable

Definition

The **standard deviation** of X is

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}.$$

The standard deviation has the *same units* as the random variable itself.

- We often denote the variance of X by σ^2 or σ_X^2, and the standard deviation of X by σ or σ_X.
- Standard deviations give the same information as variances (just different units).
- Standard deviations are easier to interpret, but variances are easier to work with mathematically.