Research Methodology for Public Policy

Lecture 6 — Probability and Confidence Intervals

Dr. Colin Kuehl

2026-07-14

Northern Illinois University

SCHOOL OF PUBLIC POLICY
CHIANG MAI UNIVERSITY

Welcome Back & Agenda

Quick Recap: Measurement

- A good concept must be **concrete** (clearly defined) and must **vary** across the units we study
- Operationalization turns an abstract concept into a concrete, measurable indicator
- Measurement validity and reliability: are we measuring what we think we're measuring, and would we get the same answer again?
- A measure can be reliable without being valid, and (less obviously) hard to be valid without being reliable
- Today we shift from *measuring* variables to *reasoning about uncertainty* once we have them . . . before talking about statistical relationships tomorrow

Today

Today's Agenda

- Why probability matters for policy research
- Basic probability rules: simple, joint, union, conditional
- Random variables and probability distributions
- The normal distribution and the 68/95/99.7 rule
- Sampling distributions and the Central Limit Theorem
- Confidence intervals: constructing and interpreting one
- Z-scores: standardizing any score

Thinking Probabilistically

Why Probability, in a Policy Methods Course?

- Almost every interesting policy phenomenon is **probabilistic**, not deterministic
- "All else equal, X causes Y more often" — not "X always causes Y"
- Statistics is about understanding the *distribution* of outcomes we could have seen, not just the one we did see
- Every confidence interval, p-value, and margin of error you will ever read in a policy report rests on the ideas in today's lecture

Thinking Probabilistically

In addition to thinking differently about causality, this is the second big shift in thinking I hope you take away from this course.

Three Ways We Use the Word "Probability"

1. **Classical**: derived from logic — a fair coin lands heads with probability $1/2$
2. **Frequentist**: derived from repetition — 620 of 1000 trials landed a certain way, so $\widehat{p} = 0.62$
3. **Subjective / Bayesian**: a degree of belief, updated by evidence — "there's a 50% chance of a recession next year"

This course mostly uses the **frequentist** framework: probability as a long-run relative frequency.

Ie we assign probabilities based on the observed data

Basic Probability Rules

Four Building-Block Definitions

QUANT Ch. 4 distinguishes four basic kinds of probability that recur throughout the course:

- **Simple probability**: the chance that a single event occurs
- **Conditional probability**: the chance of an event occurring, *given* that another event has already occurred
- **Joint probability**: the chance that two events *both* occur
- **Union probability**: the chance that *at least one* of two events occurs

Simple Probability

- The most basic case: $P(A)$, the chance a single event A occurs
- Classical example: a fair six-sided die shows a 4, $P(4) = 1/6$
- Policy example: a randomly selected respondent approves of a policy, $P(\text{approve})$
- Always bounded: $0 \leq P(A) \leq 1$, and the probabilities of all possible outcomes sum to 1

Joint and Union Probability

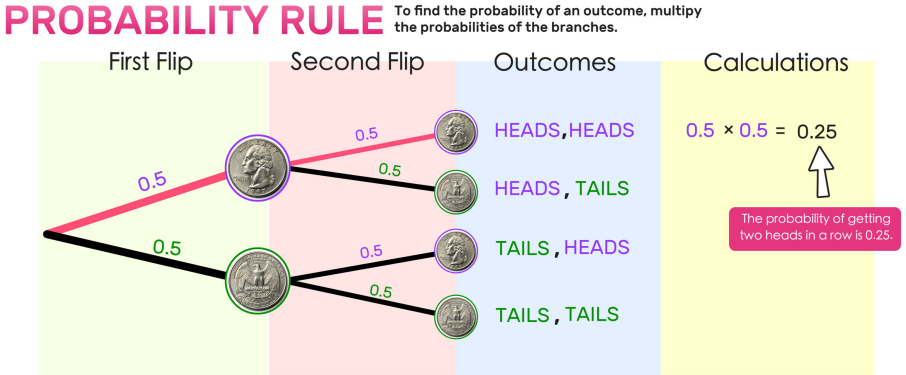
Joint probability

- Chance that *both* A and B occur: $P(A \text{ and } B)$
- Example: a respondent is both female *and* a Democrat
- If A and B are independent: $P(A \text{ and } B) = P(A) \times P(B)$

Union probability

- Chance that *at least one* of A or B occurs: $P(A \text{ or } B)$
- Example: a respondent is female *or* a Democrat (or both)
- If A and B are mutually exclusive: $P(A \text{ or } B) = P(A) + P(B)$

A Joint-Probability Illustration: Two Coin Flips



- Each branch multiplies the probabilities along the path: $0.5 \times 0.5 = 0.25$
- This is the **Bernoulli process** in miniature — more on that next

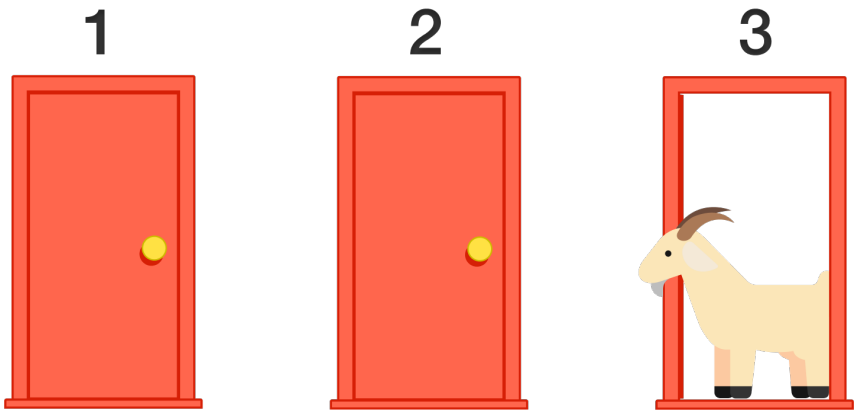
Conditional Probability

The probability of event A, *given* that event B has occurred:

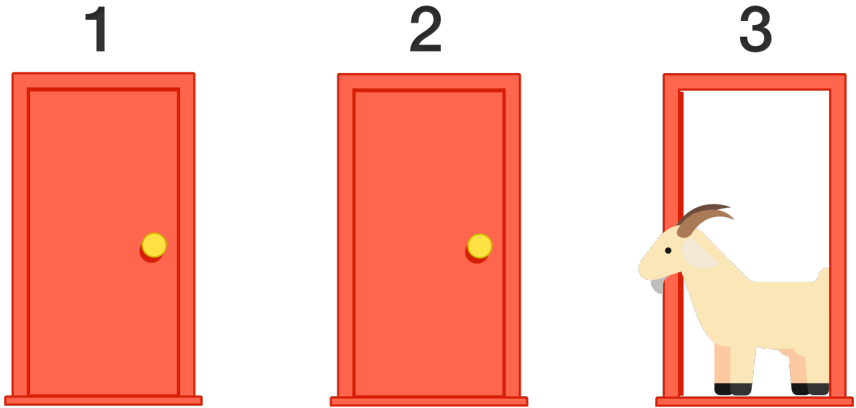
$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

- Example: the probability of interstate war, *given* that both countries are democracies
- Example: the probability of voting for a candidate, *given* gender or income
- **Conditional probability is the foundation of almost every interesting empirical claim in policy research — "given X, what is the chance of Y?"**

Probability in Practice: The Monty Hall Problem



Probability in Practice: The Monty Hall Problem



- A classic, famously counter-intuitive conditional probability puzzle — switching doors after a reveal changes your odds from $1/3$ to $2/3$
- A good reminder: intuition about probability is often wrong; the math is worth doing carefully

Independence

- Events A and B are **independent** if knowing A occurred tells you nothing about B: $P(A \mid B) = P(A)$
- We often *assume* independence of observations — e.g., one respondent's opinion doesn't depend on their neighbor's, one country's GDP doesn't depend on others in its region,
- This assumption often fails in practice (e.g., repeated measures on the same unit over time) — a recurring theme later in the course

From Probability to Distributions of Variables

The Bernoulli Process

Repeated Trials with Two Outcomes

- A **Bernoulli process** is a sequence of trials, each with only two possible outcomes (success/failure) and a constant probability of success
- QUANT Ch. 4 illustrates this with the simplest possible case: repeated flips of a fair coin
- Each flip is independent of the last — the coin has no memory
- This is the building block behind countless policy-relevant binary outcomes: pass/fail a bill, win/lose an election, default/repay a loan

Why the Bernoulli Process Matters

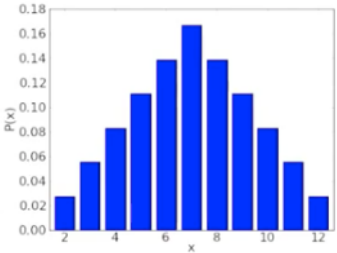
- Many policy-relevant outcomes are binary: did a country experience civil conflict in a given year? Did an applicant receive a visa? Did a project get funded?
- Treating repeated binary outcomes as a Bernoulli process lets us calculate the probability of any specific *sequence* or *count* of outcomes
- It is also the conceptual seed of the sampling-distribution logic we build toward today — many small random "coin flips" of sampling error, added together, start to look *normal*
- Similarly, asking people survey questions or other policies of interest, will lead us to a distribution

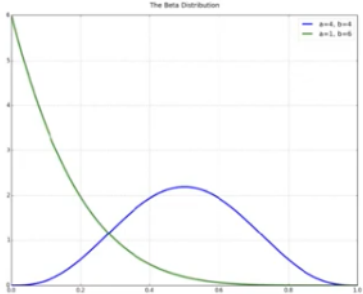
Random Variables and Distributions

What Is a Random Variable?

- A **random variable** maps outcomes of a random process onto numbers
- e.g., a coin flip $\rightarrow \{0, 1\}$; a survey response $\rightarrow$ a 1–7 scale; GDP growth $\rightarrow$ a percentage
- Two defining features:
  - **Expectation** $E[X]$: the long-run average value
  - **Variance**: how spread out the possible values are around that average
- A confusing term (why random again?) but the underlying idea is important

Probability Distributions

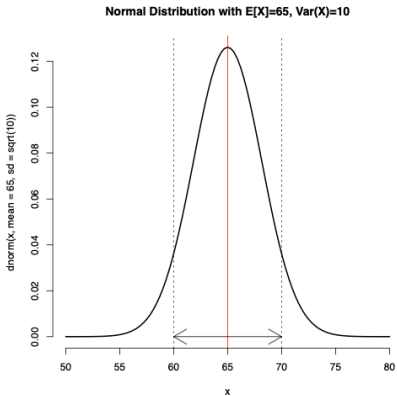




- A distribution describes every possible value a variable can take *and* its associated probability
- **Discrete** distributions (dice, counts) vs. **continuous** distributions (height, income)
- Probabilities must be between 0 and 1, and must sum to 1

The Normal Distribution

The Normal Distribution



The Most Important Distribution in Statistics

- The **normal distribution** is symmetric, bell-shaped, and fully described by just two numbers: its mean μ and standard deviation σ
- Many naturally occurring variables approximate it: height, standardized test scores, measurement error
  - We mostly assume that most variables have distributions that approximate the normal distribution.
- More importantly for us: the *sampling distribution* of a sample mean tends toward normal, even when the underlying variable does not (according to the Central Limit Theorem, coming up)

Standard deviation

The standard deviation is a measure of how spread out a set of values is around their mean. It tells you, roughly, how far a typical observation falls from the average.

Standard deviation formula

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

- $\bar{x}$

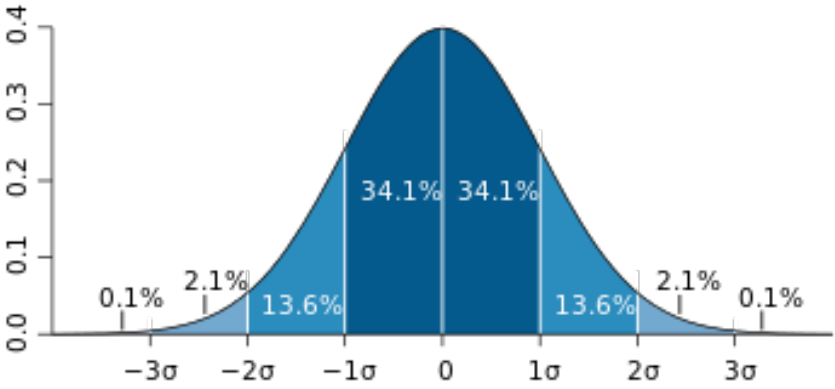
  Sample Mean -

- x_i

  is each observation

-

The Empirical Rule: 68/95/99.7



- Roughly **68.26%** of observations fall within 1 standard deviation of the mean
- Roughly **95.44%** fall within 2 standard deviations
- Roughly **99.72%** fall within 3 standard deviations

Why This Rule Is So Useful

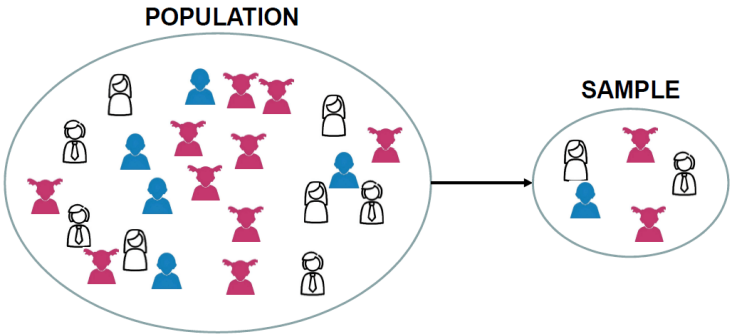
- If we know only the mean and standard deviation of a normal variable, we already know *most* of its shape
- We can say how unusual any particular value is, without needing the full distribution(if we know mean and sd)
- This is the conceptual basis for everything from "is this score unusually high?" to "is this estimate statistically significant?"

From One Sample to the Distribution of the Mean

A Thought Experiment

- Suppose you could sample 20 people's heights, take the mean, and do this over... and over... and over
- Each sample mean is a little different — noise from *who* happened to be sampled
- The **sampling distribution** describes how much those means would vary, if you could repeat the sampling
- In real research, you almost never get to repeat the sampling — you see *one* mean, from *one* sample

Population and Sample



- The **population** is the full set of units we care about
- The **sample** is the subset we actually observe
- Inference means using the sample to learn about the population we can't fully see

Why This Matters

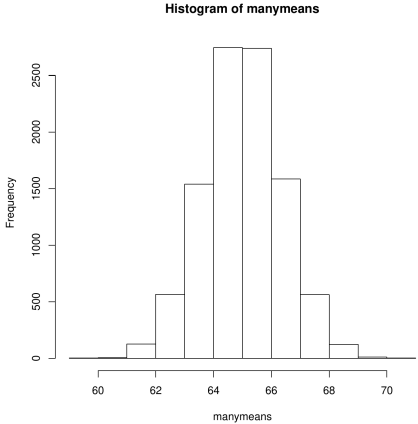
- The sampling distribution of the mean depends on only two things: N (sample size) and the underlying mean and variance of the population
- This lets us describe how much our estimate *could* have varied — without ever re-sampling
- This is the theoretical engine behind confidence intervals and (next lecture) hypothesis tests

The Central Limit Theorem

The Central Limit Theorem (CLT)

- The mean of a sufficiently large number of independent draws from *any* distribution will itself be approximately **normally distributed**
- This holds *regardless of the shape of the original distribution* — the remarkable part
- As N grows, this sampling distribution becomes more tightly concentrated around the true population mean

Seeing the CLT in Action



- Thousands of simulated sample means, each computed from its own random sample, pile up into a bell shape
- This happens *even if* the original variable being sampled is not normally

Central Limit Theorem is a big deal

Big picture: - CLT justifies just of normal distribution - Explains why larger samples are better (we're more confident that mean and SD of sample)

The Standard Error

- The standard deviation of the sampling distribution of the mean has its own name: the **standard error (SE)**

$$SE = \frac{s}{\sqrt{n}}$$

- Larger $N \rightarrow$ smaller $SE \rightarrow$ sample means cluster more tightly around the true population mean
- The standard error is what lets us turn "I have one sample mean" into "here is a range of plausible values for the population mean"

Confidence Intervals

Inference: From Sample to Population

- Inferential statistics: deriving knowledge about a population from a sample of that population.
- We almost never observe the entire population we care about; we rely on a sample and use probability theory to bridge the gap
- The best estimates we have of our population parameters are our sample statistics, and we never know with certainty how good that estimate is.

Why We Need an Interval, Not Just a Point

- A single sample mean is our *best guess* at the population mean — but it's almost certainly not exactly right
- , if we are estimating a population parameter using a sample statistic, we will want to know how much confidence to place in that estimate.
- A confidence interval reports a *range*, plus how much confidence we have that the range captures the truth

Building a Confidence Interval

A confidence interval gives a range of plausible values for an unknown population quantity, based on one sample:

$$\text{Estimate} \pm (\text{critical value}) \times (\text{standard error})$$

- For a 95% CI on a sample mean: $\bar{X} \pm 1.96 \times SE$, where $SE = \sigma/\sqrt{N}$
- A confidence interval will report both a range for the estimate and the probability the population value falls in that range.
- We need to think not just in terms of the point estimate(mean) but also its variation

Critical Values for Common Confidence Levels

TABLE 4.6 Calculating Confidence Intervals for Large Samples

Confidence level	Critical value	Confidence interval	Example $\hat{\beta}_1 = 0.41$ and $\text{se}(\hat{\beta}_1) = 0.1$
90%	1.64	$\hat{\beta}_1 \pm 1.64 \times \text{se}(\hat{\beta}_1)$	$0.41 \pm 1.64 \times 0.1 = 0.246$ to 0.574
95%	1.96	$\hat{\beta}_1 \pm 1.96 \times \text{se}(\hat{\beta}_1)$	$0.41 \pm 1.96 \times 0.1 = 0.214$ to 0.606
99%	2.58	$\hat{\beta}_1 \pm 2.58 \times \text{se}(\hat{\beta}_1)$	$0.41 \pm 2.58 \times 0.1 = 0.152$ to 0.668

- Higher confidence (99% vs. 90%) requires a *wider* interval, all else equal
- Larger $N \rightarrow$ smaller $SE \rightarrow$ narrower, more precise interval at any given confidence level

Interpreting a Confidence Interval

- A 95% CI means: if we repeated this sampling process many times, about 95% of the intervals we constructed would contain the true population value
- Usually can think in terms of "there's a 95% chance the true value is in *this* interval"

A Quick Check

- A poll reports 52% support, with a margin of error of ± 3 points
- The margin of error *is* (approximately) the 95% CI half-width
- What does this tell us — and not tell us — about the true level of public support?

From Confidence Intervals to Hypothesis Tests

- The same logic — estimate, standard error, critical value — sets up next lecture's topic: testing whether an effect is "real" or could plausibly be chance
- "Hypothesis testing is where our ideas meet the real world."
- A preview: we will ask not just "what's a plausible range for this estimate?" but "is zero (no effect) inside or outside that range?"

Research Tool: Z-Scores

Standardizing a Score

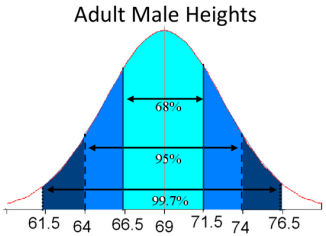
The **Z-score** converts any raw score into standard-deviation units above or below the mean:

$$Z = \frac{X - \mu}{\sigma}$$

- X: the raw score
- μ: the mean of the distribution
- σ: the standard deviation of the distribution

Z-Scores Visualized: Adult Male Heights

Z-Scores are measurements of how far from the center (mean) a data value falls.



Ex: A man who stands 71.5 inches tall is 1 standard deviation ABOVE the mean.(*z-score = 1*)

Ex: A man who stands 64 inches tall is 2 standard deviations BELOW the mean. (*z-score = -2*)

- A height of 71.5 inches is exactly 1 SD above the mean ($Z = 1$); 64 inches is 2 SD below the mean ($Z = -2$)
- The same logic works for any normally distributed variable — IQ, height, or (later

Why Standardize At All?

- Z-scores let us compare scores from *different* distributions on a common scale
- A Z-score of 1.5 always means "1.5 SD above the mean," whether we're talking about height, income, or test scores
- This is exactly the conversion we will need later to translate "how many standard errors away is my estimate" into a probability

Key Terms & Looking Ahead

Key Terms from Today

- Simple / joint / union probability
- Conditional probability
- Independence
- Bernoulli process
- Random variable
- Normal distribution / empirical rule
- Z-score
- Central Limit Theorem
- Standard error
- Confidence interval

Before Next Session

Lecture 7: Thinking about Relationships and Regression

- Read: QUANT, Ch. 5 (*Inference*, remainder) & Ch. 6 (*Bivariate Regression*)
- Read: EMPS, Ch. 5 (*Description*), if time allows

Discussion

Questions?

Thank you — see you in the Practicum this afternoon.

See You in the Practicum