CHAPTER 3

# Research Design Fundamentals

GOVT 10: Quantitative Political Analysis

> This chapter addresses one of the most fundamental questions in social science: how do we design studies that allow us to draw valid conclusions about political phenomena? We explore the hierarchy of research designs, from experiments that provide the strongest causal evidence to observational studies that describe patterns without establishing causation. The key insight is that random assignment, when possible, eliminates confounding by creating groups that are identical in expectation on all characteristics, both observed and unobserved. When randomization is impossible, clever observational designs can sometimes approximate experimental conditions, though they require stronger assumptions.

## 1 The Challenge of Learning from Observation

For decades, the Drug Abuse Resistance Education (DARE) program was one of the most popular anti-drug programs in American schools. Police officers visited classrooms, warned students about drugs, and taught them refusal skills. Politicians loved DARE. Parents loved DARE. It seemed obviously effective, and schools across America adopted it enthusiastically.

But when researchers finally conducted rigorous studies, they found something surprising: DARE had no effect on drug use. Some studies even suggested it might slightly increase drug use among participants. How could such an intuitive, popular program fail? And more importantly, why did it take so long to discover this?

The answer lies in research design. For years, the evidence for DARE came from observational studies that compared schools with the program to schools without it. But schools that adopted DARE might have been systematically different from those that did not. Perhaps they had more motivated principals, more engaged parents, or were already low in drug use. The apparent success of DARE was actually a reflection of which schools chose to adopt it, not the effect of the program itself.

Only when researchers conducted randomized experiments, randomly assigning some schools to receive DARE and others not to, did the truth emerge. The program simply did not work. This story illustrates the central challenge we face when trying to learn about politics and policy: observation alone cannot tell us why patterns exist. Correlation is not causation, and to make causal claims, we need careful research design.

## 2  The Research Design Hierarchy

Research designs can be organized into a hierarchy based on their ability to establish causation. At the top sit experiments, which provide the strongest causal evidence. In the middle are observational designs that attempt to approximate experimental conditions through clever exploitation of natural variation. At the bottom are purely descriptive studies that document patterns without claiming causation.

| Design Type | Key Feature | Causal Inference |
|---|---|---|
| Experimental | Random assignment | Strong |
| Observational-Causal | Natural variation | Moderate |
| Observational-Descriptive | Measurement only | None |

Table 1: Hierarchy of Research Designs

Understanding this hierarchy helps us evaluate the evidence behind political claims. When someone says that a policy "works" or that two things are causally related, we should ask what research design supports that claim. The answer determines how much confidence we should place in the conclusion.

## 3  Experimental Designs

In an experiment, the researcher controls who receives a treatment and who does not. The defining feature is random assignment: each unit, whether a person, district, or country, has an equal probability of being assigned to the treatment or control group. This simple mechanism has profound implications for causal inference.

### 3.1  Why Random Assignment Matters

Random assignment serves a crucial purpose: it creates groups that are, on average, identical in every way except for the treatment. Consider a study testing whether a persuasive message changes attitudes toward immigration. Without random assignment, we might compare people who received the message to those who did not. But people who seek out political information might already hold different attitudes. The groups differ before the treatment is even applied.

With random assignment, any differences between the groups are due to chance alone, and these differences average out with large enough samples. This is the Law of Large Numbers in action. Let us simulate this process to see how it works.

```
citizens %>%
  group_by(treatment) %>%
  summarise(
    n = n(),
```

```
    avg_age = mean(age),
    avg_education = mean(education),
    avg_interest = mean(political_interest),
    avg_income = mean(income_bracket),
    .groups = "drop"
  )
```

```
# A tibble: 2 x 6
  treatment     n avg_age avg_education avg_interest avg_income
      <dbl> <int>   <dbl>         <dbl>        <dbl>      <dbl>
1         0   502    44.8          3.01         4.03       2.95
2         1   498    44.4          3.10         3.96       3.01
```

> **What this code does:** We group our simulated citizens by treatment status (0 = control, 1 = treatment) and calculate the average of each characteristic within each group. The remarkable similarity between the two columns demonstrates the power of random assignment. Despite having 1,000 different individuals with varying characteristics, the treatment and control groups look nearly identical.

The treatment and control groups are nearly identical on age, education, political interest, and income. This balance extends to characteristics we did not measure. If we had recorded religious affiliation, personality traits, or childhood experiences, those would be balanced too. Random assignment balances everything, whether we observe it or not.

## 3.2   The Gold Standard

Experiments are considered the gold standard for causal inference for three interconnected reasons. First, random assignment eliminates confounding by making treatment assignment independent of all other variables. There is no way for a characteristic to be associated with both the treatment and the outcome if treatment was assigned by a coin flip. Second, the groups are comparable by construction. We do not need to assume they are similar; we created them to be similar. Third, the control group shows us the counterfactual: what would have happened to the treated units if they had not been treated.

## 3.3   Limitations of Experiments

Despite their power, experiments have important limitations. Ethical constraints prevent us from randomly assigning people to smoke cigarettes, experience poverty, or live under authoritarian rule. Practical constraints make it impossible to randomly assign countries to different political systems or states to different election rules. Laboratory settings may not reflect real-world behavior, creating questions about whether results generalize beyond the experimental context. These limitations do not diminish the value of experiments; they simply remind us that no single research design answers all questions.

# 4  Observational-Causal Designs

When experiments are impossible or impractical, researchers turn to observational designs that attempt to approximate experimental conditions using naturally occurring variation. These designs require stronger assumptions than experiments, but they allow us to study questions that cannot be randomized.

## 4.1  Natural Experiments

A natural experiment occurs when some external event creates quasi-random assignment to treatment and control conditions. The researcher does not control the assignment, but nature, policy, or circumstance does so in a way that approximates randomization.

Consider military conscription lotteries. In 1969, the U.S. military conducted a lottery to determine draft eligibility for the Vietnam War. Men born on different days were assigned different draft numbers. Those with low numbers were likely to be drafted; those with high numbers were not. Because birthday is essentially random with respect to other characteristics, comparing men with low versus high draft numbers approximates an experiment. Researchers have used this natural experiment to study the long-term effects of military service on earnings, education, and political attitudes.

The key requirement for a natural experiment is that the assignment mechanism must be truly random or as-if random. If people can manipulate their assignment or if assignment correlates with other important characteristics, the design fails.

## 4.2  Difference-in-Differences

Difference-in-differences compares changes over time in a treatment group to changes over time in a control group. The logic is elegant: if both groups would have followed similar trajectories without the treatment, then any deviation from parallel trends in the treatment group can be attributed to the treatment itself.
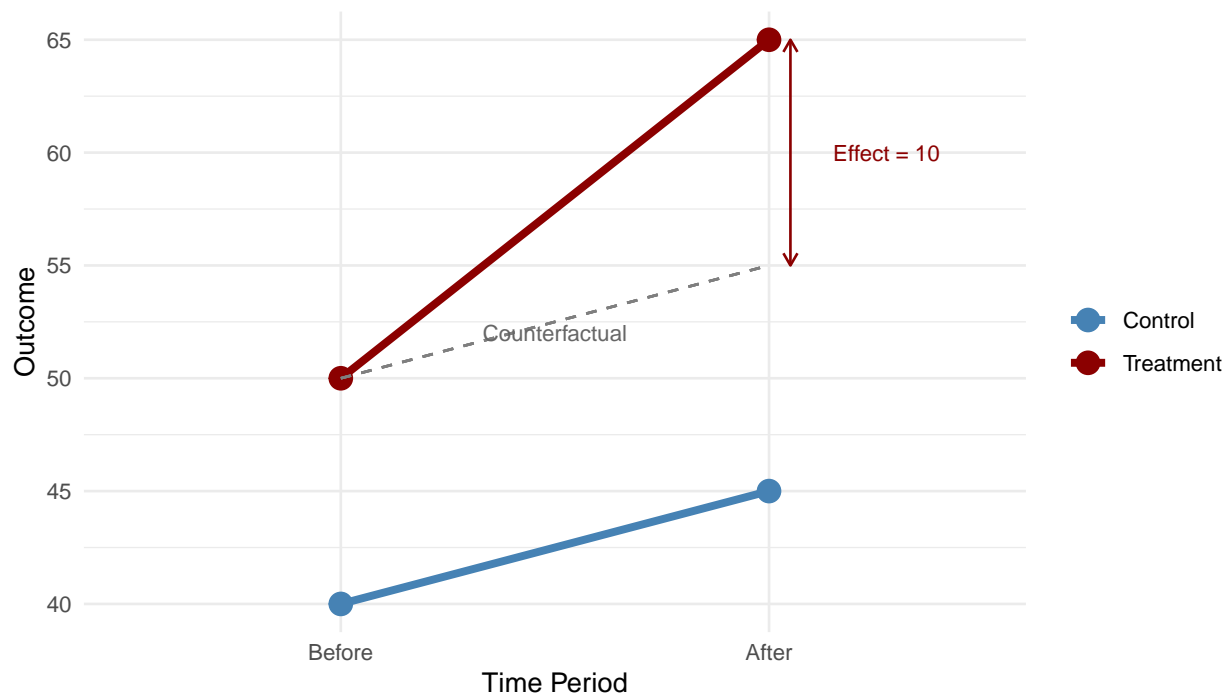
Figure 1: Figure 3.1: The difference-in-differences design. The dashed line shows where the treatment group would have been without the treatment, assuming parallel trends. The treatment effect is the gap between the actual outcome and this counterfactual.

> **DiD Calculation: The Four-Cell Scaffold**
>
> Think of DiD as filling in a 2×2 table and doing simple subtraction.
>
> |  | **Before** | **After** |
> | --- | --- | --- |
> | **Control Group** | 40 | 45 |
> | **Treatment Group** | 50 | 65 |
>
> **Step 1:** Calculate change for Treatment group: $65 - 50 = 15$
>
> **Step 2:** Calculate change for Control group: $45 - 40 = 5$
>
> **Step 3:** Subtract: Treatment change $-$ Control change $= 15 - 5 =$ **10**
>
> This "difference in differences" is your treatment effect estimate.
>
> **Alternative formula:** $(Y_{T,After} - Y_{T,Before}) - (Y_{C,After} - Y_{C,Before})$
>
> Or equivalently: $(Y_{T,After} - Y_{C,After}) - (Y_{T,Before} - Y_{C,Before})$

The calculation proceeds in steps. First, we measure how much the treatment group changed: from 50 to 65, an increase of 15 points. Then we measure how much the control group changed: from 40 to 45, an increase of 5 points. The control group change represents what would have happened anyway, the secular trend. The treatment effect is the difference beyond this trend: 15 minus 5 equals 10 points.

> **The Parallel Trends Assumption**
>
> DiD only works if the treatment group would have followed the same trajectory as the control group without the treatment. This is called the "parallel trends" assumption.
>
> **Red flags that parallel trends may fail:**
>
> - Groups were trending differently before the treatment
>
> - Treatment was adopted in response to unusual trends
>
> - Groups faced different shocks during the study period
>
> **You can never prove parallel trends,** but you can check if trends were parallel before treatment. If pre-treatment trends diverge, DiD is not credible.

### 4.3   Regression Discontinuity

Regression discontinuity exploits situations where treatment is assigned based on whether a unit falls above or below some cutoff value. Students with test scores above 80 receive a scholarship; those below do not. Candidates with vote shares above 50 percent win elections; those below lose. Policies take effect when pollution exceeds a threshold.

The insight is that units just above and just below the cutoff are nearly identical in all respects except their treatment status. A student scoring 80.1 is essentially the same as one scoring 79.9 in terms of ability, motivation, and background. The only difference is that one received the scholarship and one did not. By comparing outcomes for units just above and just below the cutoff, we can estimate the treatment effect.

## 5   The Logic of Random Assignment

Understanding why random assignment enables causal inference requires grappling with the fundamental problem that plagues all observational research: confounding.

### 5.1   The Confounding Problem

Confounding occurs when a third variable causes both the treatment and the outcome, creating a spurious association between them. Suppose we observe that people who exercise regularly live longer than people who do not. Does exercise cause longevity? Perhaps, but people who exercise also tend to eat better, avoid smoking, have higher incomes, and have jobs that allow leisure time. These confounding factors could explain the association without exercise itself having any causal effect.

```
cor(confounding_example$exercise, confounding_example$health)
```

```
[1] 0.4159324
```

> **What this code does:** We calculate the correlation between exercise and health in simulated data where both are caused partly by socioeconomic status. The correlation of about 0.42 reflects both the true causal effect of exercise and the confounding through socioeconomic status. Observational data cannot separate these two components.

Exercise and health are correlated, but we cannot determine how much of this correlation reflects a causal effect of exercise versus confounding through socioeconomic status or other variables. This is the fundamental problem with observational data.

## 5.2 How Randomization Solves the Problem

Random assignment solves the confounding problem by breaking the link between treatment and all potential confounders. When we flip a coin to assign treatment, the coin does not know about socioeconomic status, motivation, prior health, or any other characteristic. The assignment is independent of everything.

```
c(cor(experiment$treatment, experiment$ses),
  cor(experiment$treatment, experiment$motivation),
  cor(experiment$treatment, experiment$prior_health))
```

```
[1] -0.01985166 -0.01581015 -0.01459225
```

> **What this code does:** We calculate correlations between random treatment assignment and three potential confounders: socioeconomic status, motivation, and prior health. All correlations are essentially zero, demonstrating that random assignment is independent of these variables. Any differences we observe in outcomes between treatment and control groups must be caused by the treatment itself.

All correlations are near zero. The treatment is independent of potential confounders. This independence is not an assumption we hope holds; it is a mathematical consequence of the randomization procedure.

## 5.3 Seeing Randomization in Action: A Visual Comparison

To understand why random assignment works, let us compare two scenarios. In the first scenario, individuals choose whether to participate in a job training program (self-selection). In the second scenario, participation is randomly assigned. We will see how these two approaches lead to dramatically different conclusions about the program's effectiveness.

> **Why the estimates differ so dramatically:** Under self-selection, the observed earnings gap between trainees and non-trainees is about $10,000-12,000, but only $2,000 is the true causal effect. The rest is selection bias: motivated people both choose training AND would have earned more anyway. Under random assignment, the earnings gap is approximately $2,000, which correctly estimates the true causal effect. Random assignment works because it breaks the link between motivation and training.

We can examine this more directly by looking at the average motivation level in each group under both scenarios.
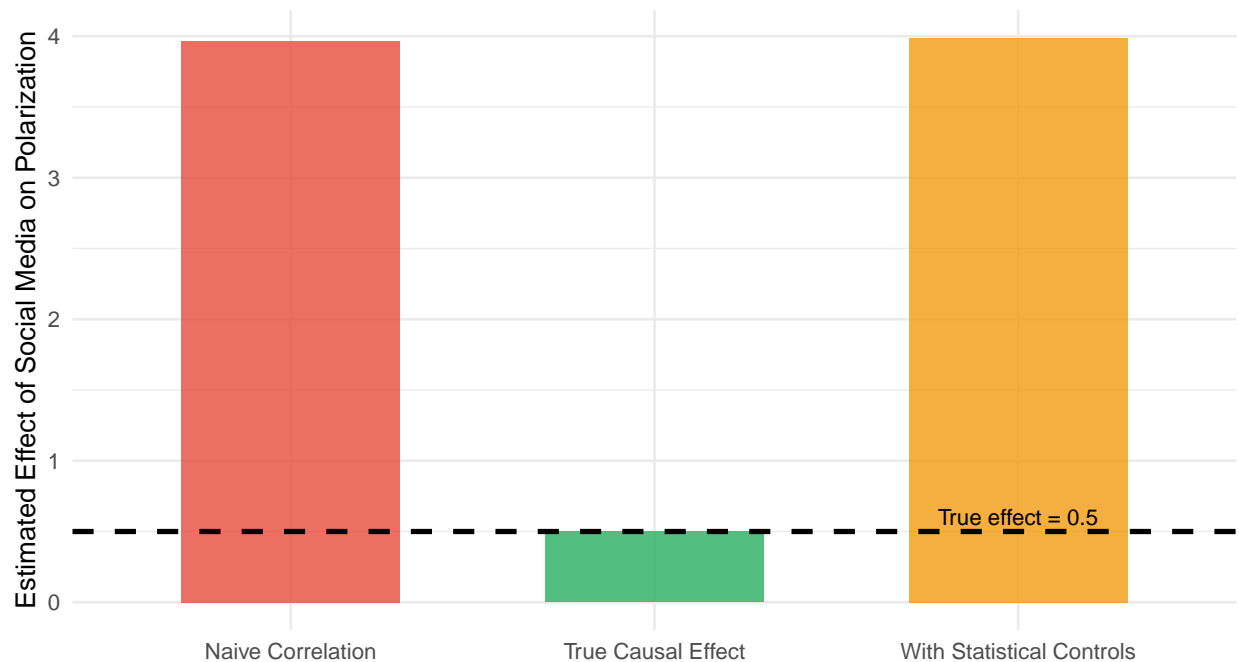
```r
# Compare motivation levels between groups
comparison_data %>%
  group_by(Scenario, Training) %>%
  summarise(avg_motivation = mean(Motivation), .groups = "drop") %>%
  pivot_wider(names_from = Training, values_from = avg_motivation)
```

```
# A tibble: 2 x 3
  Scenario            `0`   `1`
  <chr>             <dbl> <dbl>
1 Random Assignment  50.3  51.0
2 Self-Selection     43.0  58.4
```

> **The key insight:** Under self-selection, trainees have much higher motivation than non-trainees (about 55 vs 45). This imbalance confounds our estimate. Under random assignment, both groups have nearly identical motivation (both around 50). The coin flip knows nothing about motivation, so it cannot create imbalance. This is why only random assignment allows us to interpret group differences causally.

## 5.4 Another Example: Why Control Variables Are Not Enough

Students sometimes think statistical control (adding variables to regression) can substitute for random assignment. This example shows why it cannot. Imagine we want to know if social media use causes political polarization. We observe that heavy social media users hold more extreme views, but we suspect that already-polarized people seek out social media to confirm their beliefs.



**Why statistical control fails:** The true causal effect of social media on polarization is 0.5 points per hour of use. The naive correlation (about 1.5) vastly overstates this because people with extreme prior beliefs both use more social media AND are more polarized. Adding controls for education and age barely changes the estimate because the real confounders (prior beliefs, enjoyment of conflict) are unobserved. Only an experiment that randomly assigns social media exposure could recover the true effect. Statistical control adjusts only for what we measure, not for what we miss.

### The Unobserved Confounder Problem

No matter how many variables you control for in observational data, there may always be an unobserved confounder biasing your results. Random assignment solves this because it balances EVERYTHING, whether you measure it or not. The coin flip does not care about unmeasured personality traits, childhood experiences, or genetic predispositions. This is why experiments are the gold standard: they eliminate confounding from both observed and unobserved variables.

# 6 Internal and External Validity

Every research design involves trade-offs between two types of validity. Understanding these trade-offs helps us design better studies and interpret results more carefully.

## 6.1 Internal Validity

Internal validity asks whether we can trust the causal claim within this particular study. A study has high internal validity if we can be confident that differences in outcomes are caused by the treatment rather than by confounding factors, selection bias, or other threats.

Threats to internal validity include confounding (uncontrolled variables affecting both treatment and outcome), selection bias (non-random selection into treatment), attrition (differential dropout between groups), and measurement error (mismeasuring treatment or outcome). Experiments excel at internal validity because random assignment addresses the first two threats by design.

## 6.2 External Validity

External validity asks whether the results generalize beyond this specific study. A study has high external validity if findings apply to other populations, settings, and times. A laboratory study of American college students may not generalize to the broader American population, let alone to citizens of other countries.

Threats to external validity include sample bias (the study sample differs from the target population), artificiality (laboratory settings do not reflect real-world conditions), context dependence (results specific to time, place, or culture), and Hawthorne effects (people behave differently when observed).

## 6.3 The Fundamental Trade-Off

There is often tension between internal and external validity. Laboratory experiments maximize internal validity through control but may sacrifice external validity through artificiality. Field studies in natural settings maximize external validity but may sacrifice internal validity by giving up control.
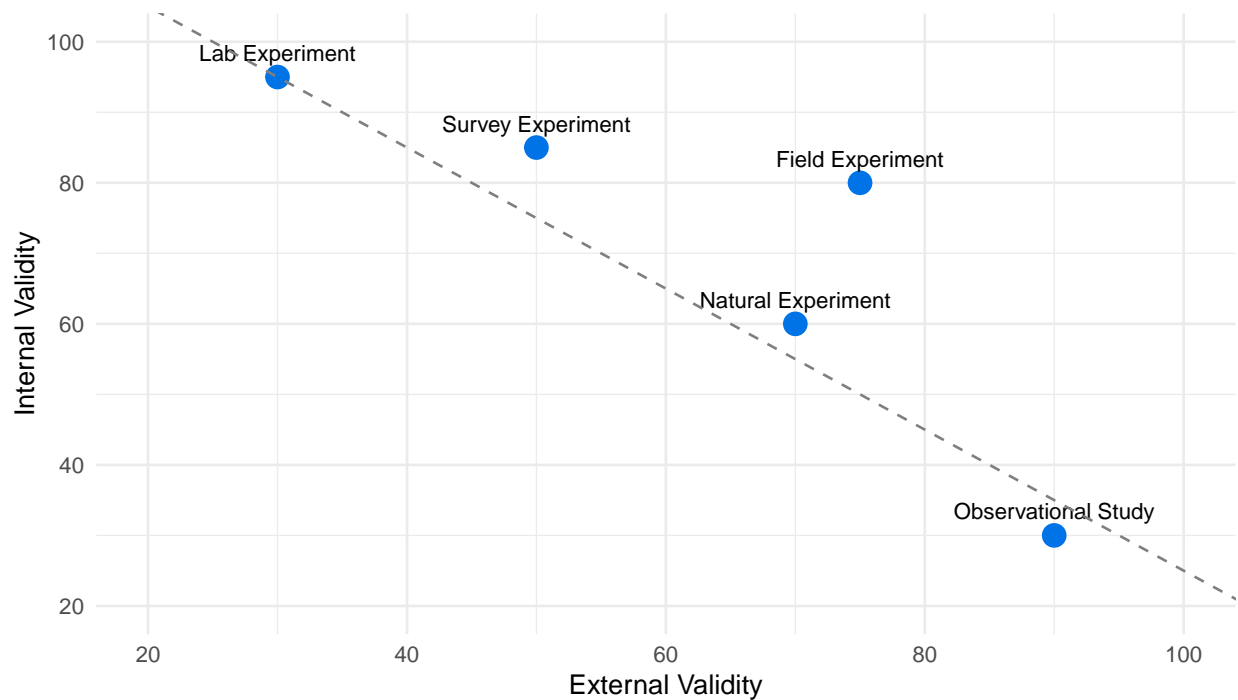
Figure 2: Figure 3.2: The internal-external validity trade-off. Different research designs occupy different positions along this trade-off. The goal is to maximize both types of validity, but this is often impossible.

The goal is to maximize both types of validity, but complete success is often impossible. Researchers must make strategic choices based on their research questions. Sometimes internal validity matters more; sometimes external validity takes priority. The best research programs use multiple designs, triangulating findings across different approaches.

# 7 Working with Research Design in R

R provides tools for implementing and analyzing research designs. The key operations involve creating treatment variables, checking balance between groups, and calculating treatment effects.

## 7.1 Creating Treatment Variables

```
experiment_data %>% count(treatment)

# A tibble: 2 x 2
  treatment     n
  <chr>     <int>
1 Control     251
2 Treatment   249
```

## 7.2   Checking Balance

After random assignment, we should verify that the treatment and control groups are similar on observed characteristics. Large imbalances might indicate a problem with randomization or simply bad luck with a small sample.

```
experiment_data %>%
  group_by(treatment) %>%
  summarise(
    n = n(),
    mean_baseline = mean(baseline_outcome),
    sd_baseline = sd(baseline_outcome),
    .groups = "drop"
  )
```

```
# A tibble: 2 x 4
  treatment       n mean_baseline sd_baseline
  <chr>       <int>         <dbl>       <dbl>
1 Control       251          49.4        9.44
2 Treatment     249          50.0       10.0
```

**Reading a Balance Table: What to Check**

**Step 1: Check sample sizes (n).** Are the groups roughly equal? With random 50/50 assignment, groups should be similar. Large differences suggest a problem with randomization.

**Step 2: Compare means.** Are the group means similar? Small differences (a few percentage points or units) are expected by chance. Large differences suggest imbalance.

**Step 3: Compare standard deviations.** Is variation similar in both groups? If one group is much more variable, groups may be systematically different.

**Step 4: Rule of thumb.** Means within 0.1-0.2 standard deviations of each other = good balance. Means more than 0.25 standard deviations apart = concerning.

**What good balance looks like:**

- n similar (e.g., 250 vs 250, not 350 vs 150)

- Means very close (e.g., 50.1 vs 49.8)

- SDs similar (e.g., 10.2 vs 10.5)

## 7.3   Calculating Treatment Effects

```
experiment_data %>%
  group_by(treatment) %>%
  summarise(mean_outcome = mean(outcome), .groups = "drop")
```

```
# A tibble: 2 x 2
  treatment mean_outcome
  <chr>            <dbl>
1 Control           49.0
2 Treatment         54.9
```

> **What this code does:** We calculate the average outcome in each group. The difference between these means is the estimated treatment effect. In this simulated example, the treatment group averages about 5 points higher than the control group, which matches the true effect we built into the simulation.

# 8   Common Mistakes and How to Avoid Them

**Mistake 1: Treating Observational Correlations as Causal**

The most pervasive error in quantitative research is interpreting observational correlations as causal relationships. When someone claims that a policy works because places with the policy have better outcomes, ask: why did those places adopt the policy? If adoption is related to the outcome through some third factor, the correlation is confounded. Only randomization or a credible quasi-experimental design can support causal claims.

**Mistake 2: Ignoring Selection Bias**

Selection bias occurs when the process determining who receives treatment is related to the outcome. People who choose to attend college differ from those who do not in ways that affect their subsequent earnings. Countries that adopt democracy differ from those that remain autocratic in ways that affect their subsequent economic growth. Any analysis that ignores selection will confuse selection effects with treatment effects.

**Mistake 3: Confusing Statistical Control with Experimental Control**

Including variables in a regression does not eliminate confounding the way random assignment does. Statistical control only adjusts for measured variables; it cannot adjust for unmeasured confounders. Moreover, controlling for variables that are themselves affected by the treatment can introduce new biases. The phrase "controlling for" should not inspire the same confidence as "randomly assigning."

**Mistake 4: Overinterpreting Natural Experiments**

Natural experiments are valuable, but they require careful scrutiny. Was the assignment

mechanism truly random? Could people anticipate or manipulate their assignment? Are there any ways the assignment could correlate with outcomes through channels other than the treatment? A lottery is only as good as its implementation, and many supposed natural experiments turn out on closer inspection to have significant problems.

# 9  R Functions Reference

| Function | Purpose | Example |
|---|---|---|
| `sample()` | Random assignment | `sample(c(0,1), n, replace=TRUE)` |
| `mutate()` | Create variables | `mutate(treatment = sample(...))` |
| `if_else()` | Binary condition | `if_else(x > 50, "High", "Low")` |
| `case_when()` | Multiple conditions | `case_when(x < 30 ~ "Low", ...)` |
| `group_by()` | Group data | `group_by(treatment)` |
| `summarise()` | Calculate statistics | `summarise(mean = mean(x))` |
| `count()` | Count observations | `count(treatment)` |

Table 2: Key R Functions for Research Design

# 10  Practice Problems

### Question 1

A researcher finds that students who attend private schools score higher on standardized tests than students who attend public schools. Why might this not mean private schools cause higher test scores? Describe at least two potential confounders and explain how they could produce the observed correlation without any causal effect of school type.

### Question 2

A city implements a new policing strategy in some neighborhoods but not others. Crime falls in the treated neighborhoods. A politician claims the strategy caused the reduction. What questions would you ask to evaluate this claim? What research design would provide stronger evidence?

### Question 3

Explain why we cannot use random assignment to study whether democracy causes eco-

nomic growth. What observational approaches might researchers use instead, and what assumptions would each require?

### Question 4 (Computational)

Create a simulated dataset of 400 voters with age, party identification, and a randomly assigned treatment indicator. Calculate the mean age separately for the treatment and control groups. Are the groups balanced? Run the simulation 10 times with different seeds and observe how balance varies.

### For Further Study

This chapter has introduced the fundamental concepts of research design. Students seeking deeper understanding should explore Angrist and Pischke's *Mostly Harmless Econometrics*, which provides an accessible treatment of modern causal inference methods. For natural experiments specifically, Dunning's *Natural Experiments in the Social Sciences* offers guidance on finding and evaluating quasi-random variation. Pearl's *The Book of Why* provides a philosophical foundation for thinking about causation more broadly.