Causal Inference and Bivariate Relationships

David A. Hughes, Ph.D.

Auburn University at Montgomery

david.hughes@aum.edu

April 11, 2022

Overview

Introduction to causality

- "Causality" refers to the effect one variable (an independent variable) has on another variable (dependent variable).
- That is, when the independent variable increases, the dependent variable either increases or decreases, independent of other factors.
- A "null" relationship occurs when the dependent variable is invariant to the independent variable.

Causal inference

Generally, there are three elements to drawing a valid causal inference:

- Time-ordering
- Association
- Elimination of other, confounding factors

Analysis of association

Our methods will largely depend upon the level of measurement of our variables, but our interests are relatively constant.

- Direction of association
- Strength of association
- Statistical significance

Elimination of other, confounding factors

- Our research design allows us to address spuriousness and minimize the risks of drawing a mistaken inference.
- Generally, experimental methods are the gold standard for minimizing such risks.
- In much of what we do, however, we are left with observational data.
- Regardless of our methodology, however, we'll make use of statistical theories of inference.

Hypothesis-testing

- How do we know whether an observed relationship is "real"?
- We start by specifying the expected relationship between two variables (the alternative hypothesis).
- An alternative hypothesis has three elements: (1) Dependent variable, (2) Independent variable, and (3) Anticipated association between them.
- We then specify what a null relationship would look like between the two variables (the null hypothesis).

Statistical significance

- Once our hypotheses are specified, we need some way to test them.
- We make use of a concept known as "statistical significance" to achieve this end.
- When we hypothesis-test in statistics, we recover p-values, which exist between 0 and 1.
- Crudely, when $p \leq 0.05$, we reject the null hypothesis. Otherwise, we fail to reject it.

Making sense of uncertainty

- Suppose we wanted to know how tall loblolly pine trees grow.
- If we measured the height of every single tree, this would be quite easy.
- But we can't (or at the very least, we shouldn't).

Making sense of uncertainty (contd.)

- We don't know the average height of *all* pines (μ).
- But we have the average height of a *sample* of them ($\bar{x}$).
- So we marshal our uncertainty from our samples and make probabilistic statements about the likely average height of the population of pines.

What is hypothesis-testing?

- We start with a hypothesis: e.g., Loblolly > 100ft
- We then specify our "null" and "alternative" hypotheses (H_0 and H_a, respectively).
- At the end of the day, we either "reject" or "fail to reject" the null hypothesis.

How does it work? An example

- Suppose we hypothesize that the average loblolly is *at least* 100 ft.
- A simple random sample of 9 trees yields:
  - $\bar{x} = 110$
  - $\sigma = 15$
- Can we claim with confidence that $\mu > 100$?

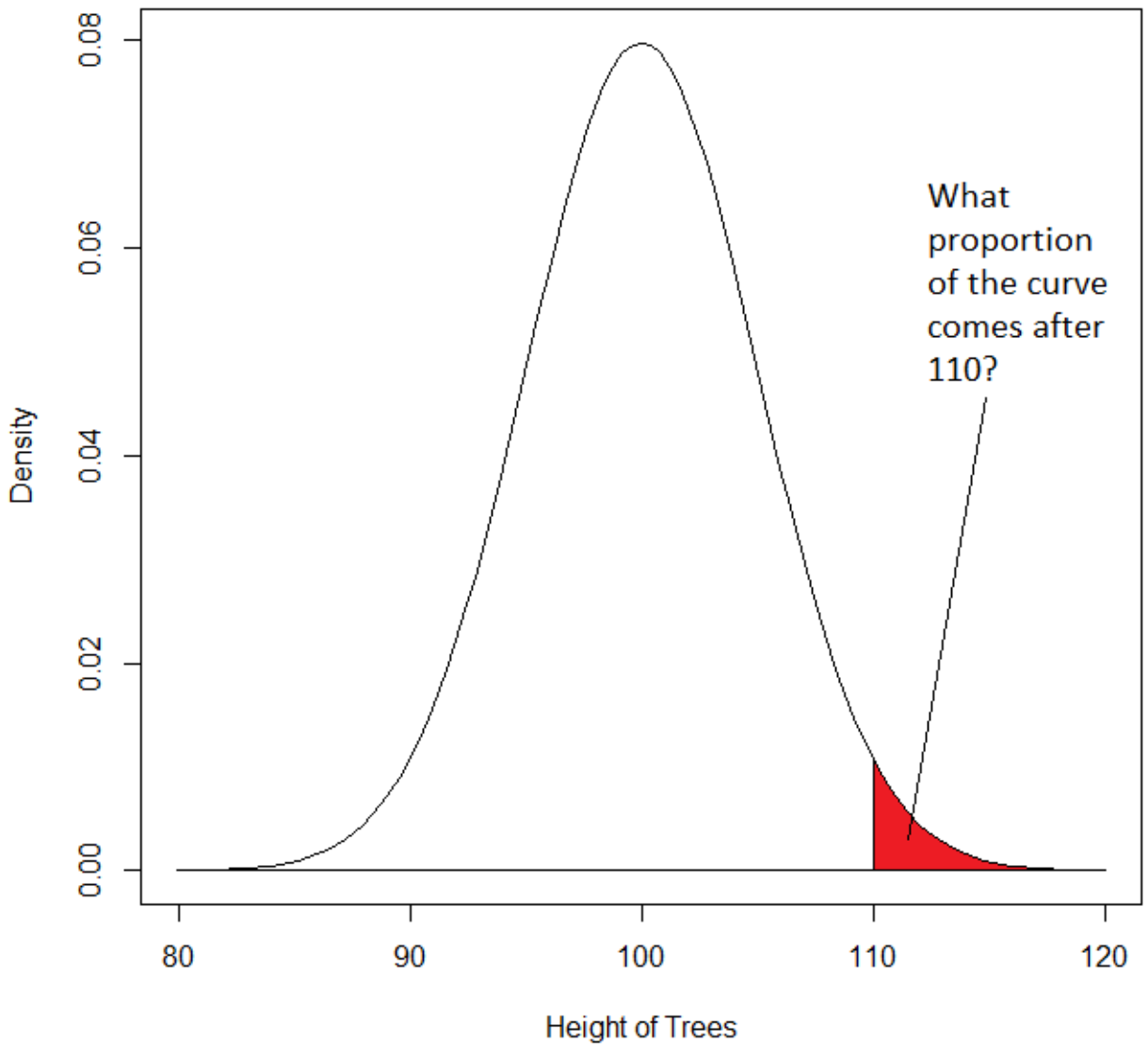
Hacking the sampling distribution

- We want to know the likelihood that we observed $\bar{x} = 110$, if the "real" mean was actually 100.
- If μ is really 100, then the sampling distribution tells us that increasingly larger values of $\bar{x}$ will be very unlikely.
- We want to calculate the number of standard errors $\bar{x}$ is from H_0 and calculate the area under the curve.

Calculate the standard error for our sample of trees

- We found: $\bar{x} = 110$ and $\sigma = 15$
- And $\hat{\sigma} = \frac{\sigma}{\sqrt{n}}$.
- Therefore, $\hat{\sigma} = \frac{15}{\sqrt{9}} = 5$.

What's the probability we drew $\bar{x} = 110$ by chance?



The z-distribution

- The z-distribution is a standard normal distribution.
- A z-score is the number of standard errors an observation is from H_0.
- It is calculated with the following formula:

$$z = \frac{\bar{x} - \mu}{\hat{\sigma}}.$$

Making probabilistic statements with distributions

- We can use z-scores to make probabilistic statements.
- The proportion of the z-distribution above/below our z-score is the probability we observed that figure by chance.
- This proportion is known as a p-value. Every z-score has a corresponding p-value.

So when do we know to reject the null?

- "Critical values" help us evaluate our hypotheses, α.
- Your α specifies how small of a p-value you demand before you reject H_0.
- Therefore, α is a measure of your willingness to accept risk.
- Most commonly, scholars set $\alpha = 0.05$.

Hypothesis-testing with pine trees

- Two ways to hypothesis-test using the z-distribution
  - Compare your p-value to α.
  - Compare the absolute value of your z-score to a relevant threshold.
- Let's try this with our sample of 9 trees.

Review: The steps for hypothesis-testing

1. State the null and alternative hypotheses.
2. Choose the α level.
3. Choose a one or two-tailed test.
4. Find the z-score (the test statistic).
5. Compare this to the critical value you established.
6. "Reject" or "fail to reject" the null.

When good hypotheses go bad...

- A "Type I Error" occurs when we reject the null hypothesis, but we should not have.
- A "Type II Error" occurs when we fail to reject the null hypothesis, but we should not have.
- The probability we commit a Type I Error is increasing in α, vice versa Type II.

Comparing two groups

- Suppose:
  - H_a: $\bar{X}_1 > \bar{X}_2$
  - H_0: $\bar{X}_1 = \bar{X}_2$
- We hypothesis-test using a "difference-of-means" test
- The test:

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Comparing two groups: An example

- We hypothesize that loblollies are taller than long-leaf pine trees.
- We gather data from 10 trees (5 loblollies, 5 long-leafs).
- We find:
  - $\bar{X}_{\text{Lob}} = 115$; $\bar{X}_{\text{Long}} = 110$
  - $\sigma^2_{\text{Lob}} = 10$; $\sigma^2_{\text{Long}} = 20$
- What's the likelihood that the population of loblollies are, in fact, taller than the population of long-leafs?

An example using the t-distribution

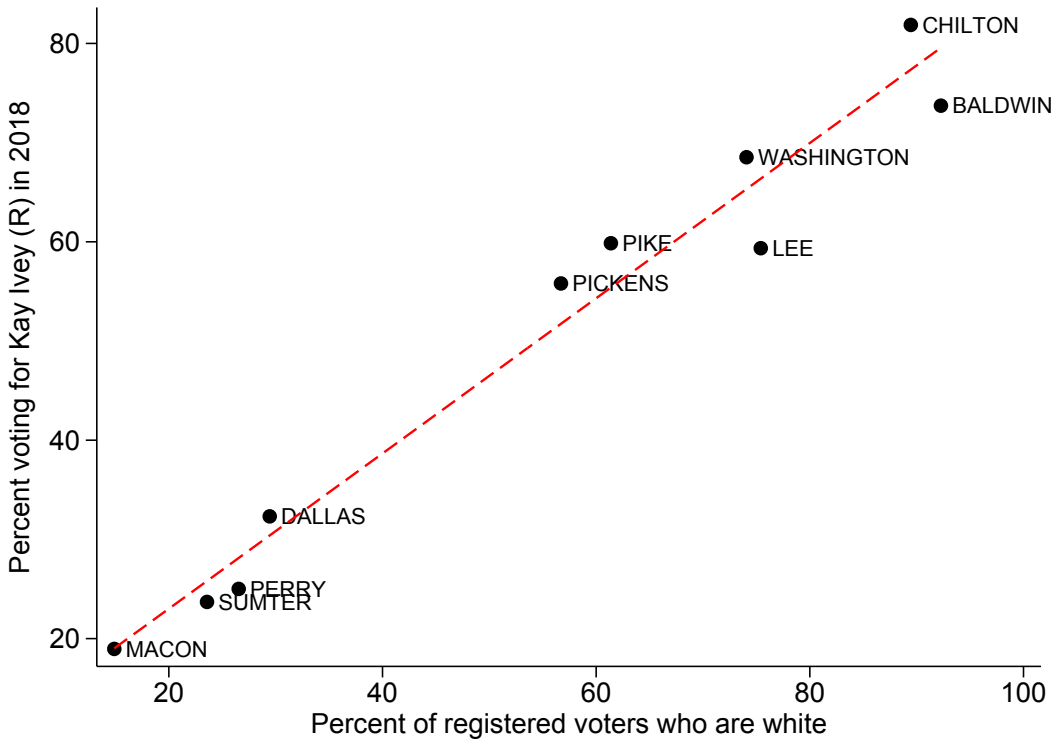
- You gather data from 10 Alabama counties, 5 in the black belt, 5 not ($n = 10$).
- You hypothesize the counties not in the black belt are more Republican.
- Therefore, you're conducting a difference of means test.
- Because we're comparing *two means*, we have two constraints ($k = 2$).
- Therefore, $df = 10 - 2 = 8$.

What about other types of relationships?

- So far, we've simply been examining differences in means. Hence, the independent variable is dichotomous.
- Oftentimes, however, we're interested in more complex relationships.
- What do we do when we have variables measured at other levels?

A simple example

- H_a: Pct. White $\rightarrow$ GOP Vote
- H_0: No relationship
- How *strong* is this relationship, and which hypothesis is more valid?



Pearson's correlation coefficient

- Pearson's correlation, $r \in [-1, 1]$, is a measure of .
- Formally:

$$r = \frac{\sum(Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

- We can check for statistical significance using a t-test:

$$t = r\sqrt{\frac{n-2}{1-r^2}}.$$

An example with two continuous-level variables

- Suppose we observe individuals' height in inches, $H = \{60, 66, 62, 72\}$ and weight in pounds, $W = \{150, 180, 130, 200\}$.
- Then we have $\bar{H} = 65$ and $\bar{W} = 165$.
- Let's use our formula for r to determine the strength of association between H and W.

An example (Continued)

Obs.	H	W	$\bar{H}$	$\bar{W}$	$(H_i - \bar{H})$	$(W_i - \bar{W})$
1	60	150	65	165	-5	-15
2	66	180	65	165	1	15
3	62	130	65	165	-3	-35
4	72	200	65	165	7	35

Obs.	$(H_i - \bar{H})^2$	$(W_i - \bar{W})^2$	$(H_i - \bar{H})(W_i - \bar{W})$
1	25	225	75
2	1	225	15
3	9	1225	105
4	49	1225	245
Sum	84	2900	440

An example (Continued)

$$r = \frac{440}{\sqrt{84 \times 2900}} = \frac{440}{493.56} = 0.89$$

$$z = 0.89\sqrt{\frac{4-2}{1-0.89^2}} = .89\sqrt{\frac{2}{0.21}} = 0.89(3.09) = 2.75$$

Non-linear correlations

- Much of the data we deal with aren't measured continuously.
- We need similar methods of measuring correlations, strengths of association, and statistical significance for these variables too.
- Much of our analysis will rely upon cross-tabulations (crosstabs) and χ^2 tests.

Cross-tabulations

- Useful when examining the relationship between categorical variables (why not scatterplots?).
- We can array the observations across variables' categories to uncover a relationship.

Example of a crosstab

Suppose we surveyed 100 educators about their careers and their income, we created a crosstab, and we found:

	Public	Private	Total
Low income	60	10	70
High income	5	25	30
Total	65	35	100

Table: Effects of job type on income

What do we see?

Analysis of independence in crosstabs

- The question is, are our observations independent of chance, or is there some pattern here?
- We can go ahead and state H_a, H_0, α, etc.
- Calculate degrees of freedom (don't include row/column totals): $df = (\#\text{Columns} - 1) \times (\#\text{Rows} - 1)$
- Calculate expected frequencies. For each cell: $(\text{Row Total} \times \text{Column Total})/n$.
- Then:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e},$$

  where f_o is each outcome, and f_e is its expectation.
- Finally, consult a χ^2 table.

Back to our previous example

	Public sector	Private sector	Total
Low income	f_o: 60 f_e: 46	f_o: 10 f_e: 25	70
High income	f_o: 5 f_e: 20	f_o: 25 f_e: 11	30
Total	65	35	100

Table: Effects of job type on income

$$\begin{aligned}
\chi^2 &= \frac{(60-46)^2}{53} + \frac{(10-25)^2}{18} + \frac{(5-20)^2}{23} + \frac{(25-11)^2}{8}, \\
&= \frac{196}{46} + \frac{225}{25} + \frac{225}{20} + \frac{196}{11}, \\
&= 42.33.
\end{aligned}$$

Crosstabs with ordinal data

- So far, we've established how to analyze statistical significance on categorical data—be they ordinal or nominal.
- But we'd also like some measure of *strength* of association (like Pearson's r).
- For ordinal data, that measure is sometimes just called γ (read, "gamma").

Example of ordinal crosstabs

Suppose we examine 100 individuals' political ideology as a function of their religious affiliation and found perfect separation. If this were a Pearson's r, we'd have gotten an $r = 1.00$. Deviations from perfect separation would diminish that relationship.

	Not Evangelical	Evangelical	Total
Liberal	30	0	30
Conservative	0	70	70
Total	30	70	100

Table: Effects of religion on ideology

Goodman and Kruskal's gamma

- Gamma ($\gamma \in [-1, +1]$) is calculated by looking at "concordant" and "discordant" pairs in the cross-tab.

$$\gamma = \frac{C - D}{C + D}.$$

- A pair of observations is concordant if the subject who is higher on one variable is also higher on the other. Otherwise, they are discordant.

Example of ordinal crosstabs

Let's make the relationship a little more complicated. Is the relationship statistically significant?

	Not Evangelical	Evangelical	Total
Liberal	30	10	40
Conservative	20	40	60
Total	50	50	100

Table: Effects of religion on ideology

Relationships between nominal data

- Ordinal measures of association are inappropriate if our cross-tab consists of nominal-level data (unless they're dichotomous).
- We could simply stop with the χ^2 test statistic and say, "There's a relationship."
- But that's a little unfulfilling.
- The ϕ coefficient can be helpful here: $\phi = \sqrt{\frac{\chi^2}{n}}$. So can Cramer's V, each of which varies between 0 and 1 where greater values reflect better fit.

An example of nominal data in a crosstab

Suppose you're interested in the number of children respondents have as a function of their religious affiliation

	Catholic	Protestant	Neither	Total
None	154	317	326	797
One	102	210	147	459
Two	162	377	194	733
Three	113	251	103	467
Four or more	118	216	77	411
Total	649	847	1,371	2,867

$\chi^2 = 92.99, \quad p < 0.000, \quad \phi = 0.18$

Table: Effects of religion on child-bearing (Source: GSS 2016)

Conclusion

- Often when we are presented with survey research data, we are shown how groups vary with respect to variables of interest.
- But when should we believe these differences, and when should we chalk them up to mere chance?
- The purpose of this unit is to help you understand whether those patterns are "statistically significant," which is to say, different enough to be unlikely due to chance.