

POLS 641

Introductory Political Analysis

Class 7: Bivariate Regression

Dr. Colin Kuehl
Northern Illinois University

October 13, 2026

Overview

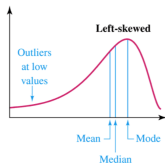
- Course Housekeeping & Problem Set # 2 Feed back
- Review of Hypothesis Testing and P-Values
- Measuring Association
- The logic of regression
- Interpreting regression

Course Housekeeping

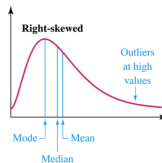
- Problem set # 3 due tomorrow
- Datacamp next week
- Next Week Research in Action: Kerner and Sumner. 2026
"Gender-Role Traditionalism and Opposition to Trade in the United States"

Pset #2 Describing Distributions

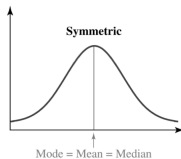
. . . describe the distribution



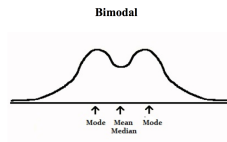
(a)



(b)



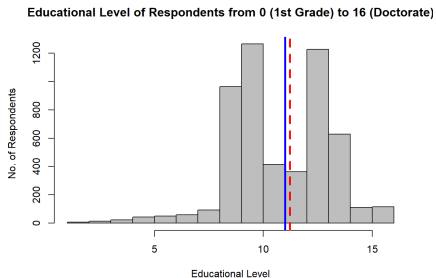
(c)



(d)

Describing Distributions

. . . describe the distribution



P-Set # 2

5. A political analyst is interested in the outcome of the 2020 federal elections for the House, the Senate, and the Presidency in the United States. Specifically, she hypothesizes that the election will result in a divided government (i.e., in one party failing to gain control of all three branches of government). Assuming that only the Democratic and Republican parties compete in the election, what is the sample space of election outcomes for the three branches?

Presidency, Senate, House

- DDD
- DRD
- DDR
- DRR
- RDR
- RDD
- RRD
- RRR

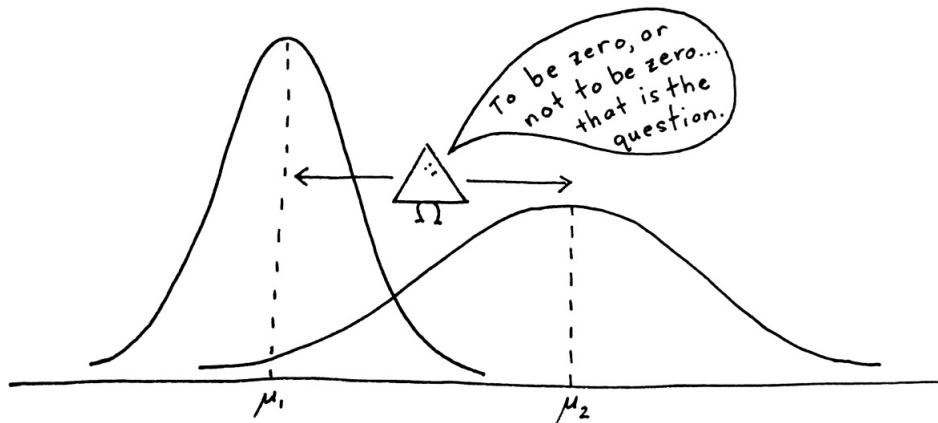
P-Set # 2

- 1 If you repeatedly draw a sample and take a mean, how will the distribution of the mean change if the variance of the underlying population is small versus big? If your sample size is small versus big? Explain.
- 2 A theory friend of yours asks "What is a random variable and why do you care?" In your own words, how would you answer?
- 3 Slightly bemused, your theory friend claims he's always hearing forms of the term "random" thrown around by positivists. Explain in a way a theorist would understand the difference between a random variable, randomization, and a random sample.
- 4 You're on a roll now. In your own words, explain the difference between these terms and put them in a logical sequence: Estimate, Estimand, Estimator. Give one example of each.
- 5 What concept(s), learned thus far in class, are you struggling the most with?
- 6 Did you find this problem set easier or harder than problem set # 1? Please feel free to explain/vent.

P-Set # 2 - General Advice

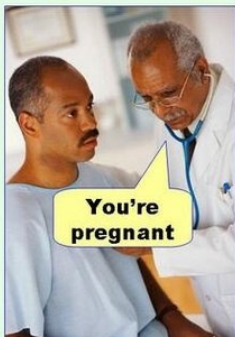
- Fully answer each question
- Read through your pdf output
- # note your code to remind your future self what you are doing
- Take a break or ask a friend when stuck

Review: Differences between Groups

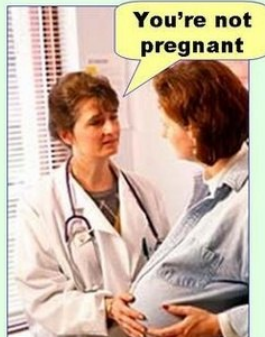


		REALITY	
		NULL HYPOTHESIS	
		TRUE	FALSE
STUDY FINDINGS	TRUE		Type II error (β) 'False negative'
	FALSE	Type I error (α) 'False positive'	

Type I error
(false positive)



Type II error
(false negative)



Significance

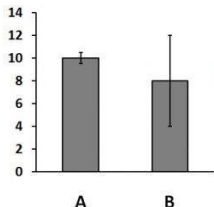
Significance Level/Critical Value - How unlikely a result has to be under the null hypothesis for us to reject the null hypothesis.

We can never rule out that observed result is happening by chance, but we can give a probability to the that likelihood

A T-test

MR T TEST

"THAT AIN'T SIGNIFICANT, FOOL!"



The t-test gives us a t-statistic that we then compare to our critical value.

In practice today we just look at the p-value

p-value

- probability of observing a coefficient as high as we actually observed if the null hypothesis were true
- Probability of committing a type one Error
- Probability happening by chance
- Usually an item is statistically significant if less than .05 (ie less than our critical value)
- We indicate using ***

Measuring Association

A T-test tells us if there is statistically significant difference between two groups.

How can we quantify the relationship between two continuous variables?

- Covariance
- Correlation
- **Regression**

Covariance

The first measure of association we will talk about is **covariance**. (See Appendix in Textbook).

It will help us determine whether two variables are:

- positively associated: when X is higher, we expect Y is usually higher
- negatively associated: when X is higher, we expect Y is usually lower
- not associated: when X is higher, it doesn't tell us anything about Y .

A Moment for Theory

In the Magical World of the Random Variable

Two random variables X and Y have a distribution $p(X, Y)$ that describes how often you see any combination of them. The covariance of X and Y , $\text{cov}(X, Y)$ is a property of this distribution.

Sample Covariance

We will often be interested in sample *estimates* of the covariance (aka the sample covariance). We will use samples of data: $(X_1, Y_1), \dots, (X_N, Y_N)$, from which we can make such an estimate using a *sample estimator*, $\widehat{\text{Cov}}(X, Y)$.

Covariance

As the name suggests, it is analogous to variance but associating two variables (co-). How two things vary together.

The real covariance is a function of the distributions:

In an actual sample using an observed sample, use analogous estimators:

- sample variance: $\widehat{Var}(X) = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$
- sample covariance: $\widehat{Cov}(X, Y) = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})$

Correlation

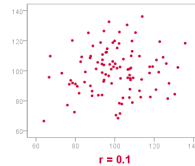
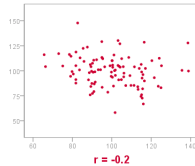
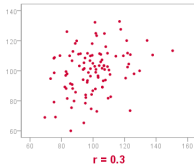
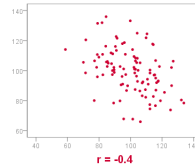
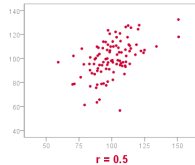
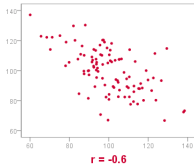
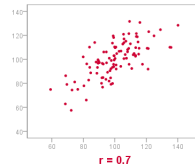
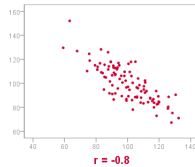
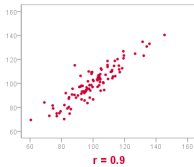
Problem with covariance: scale is not very natural. Wouldn't it be easier if it went between -1 and +1?

Correlation is a rescaled version such that:

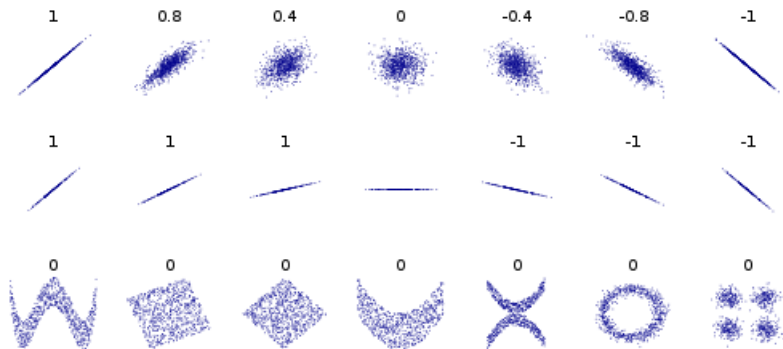
- a perfect positive relationship has $\text{cor}(X, Y) = 1$
- a perfect negative relationship has $\text{cor}(X, Y) = -1$
- two perfectly unrelated variables have $\text{cor}(X, Y) = 0$.

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

Correlation



Non-linear



Correlation does *not* capture all meaningful relationships between X and Y

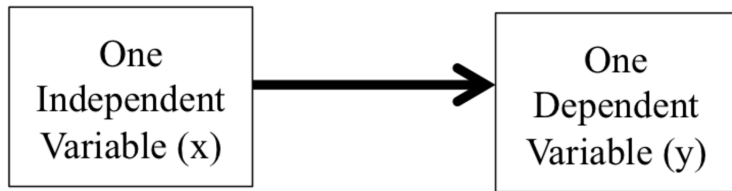
Regression has same issue

Correlation

- Correlation tells us:
 - whether a relationship is positively or negatively associated
 - a general measure of the strength of that relationship
- But doesn't tell us:
 - if the relationship is significant (There is no critical value)
 - the precise magnitude of the relationship
 - *the effect of other variables*
- That is why we usually use correlation

Our Focus Today: Bivariate Regression

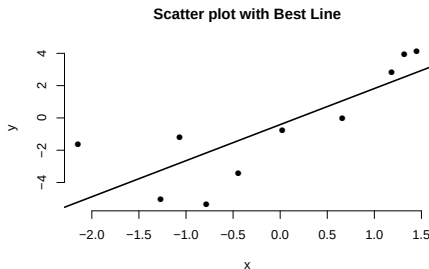
Single-Variable Regression:



Output: $y = f(x)$

How to Measure Association 3: Bivariate Regression

We might be interested in how big a change in Y we expect for a given change in X ?



Turns out this slope is:

$$\hat{\beta}_1 = \frac{\text{cov}(X, Y)}{\text{Var}(X)}$$

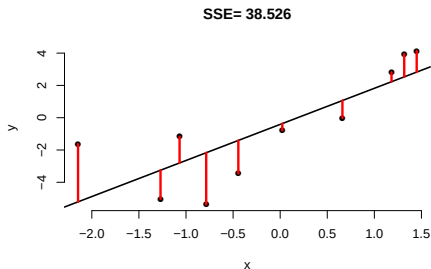
This is the same as Equation 3.4 in the textbook, just written differently.

Regression is a workhorse of statistics!

Usually think of in terms of the *line that best fits the data*.

- $\hat{Y}_i = \beta_0 + X_i\beta_1$ for an intercept β_0 and slope β_1
- Then think of this $\hat{Y}_i$ as a prediction for Y_i that takes X_i into account.
- ie Choosing β_0 and β_1 to minimize sum of squared errors:
- Suppose we guess β_0 and β_1 . How do we assess whether it is a good prediction?
- We use the sum of squared errors (called residuals in the textbook) to see how well we are doing: $SSE = \sum_{i=1}^N (Y_i - (\beta_0 + \beta_1 X_i))^2$

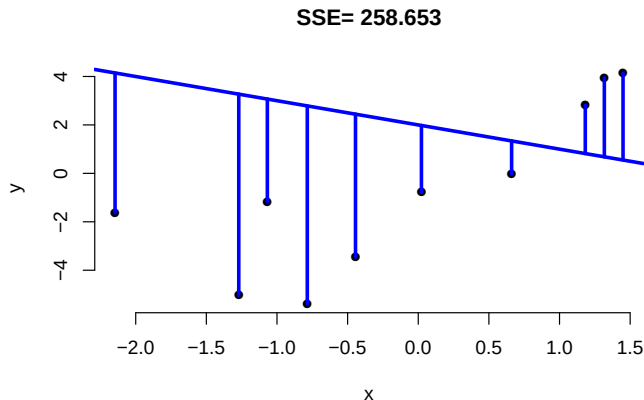
Best Fitting Line



The residual is the distance between the line and any given point in our data.

The SSE takes those residuals, squares them, and adds them up. The regression shows us the best fitting line in terms of sum of squared errors (SSE).

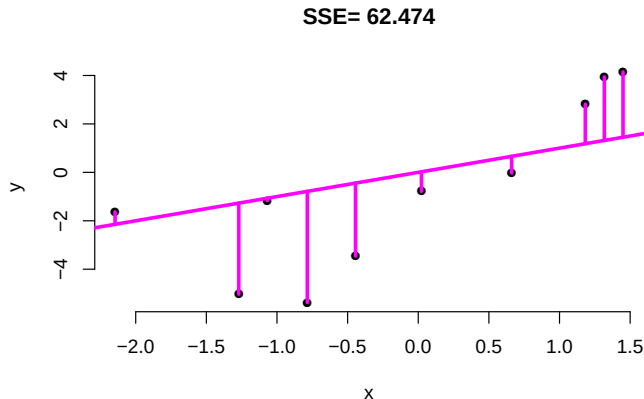
Compare to this



This is a pretty terrible fit.

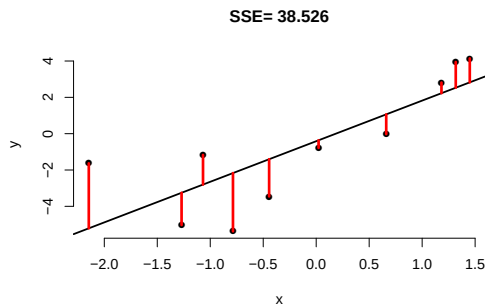
Sum of squared errors (SSE) is larger: 259 compared to 39

And this



This is not totally unreasonable, but could be improved
SSE=62, not as bad as last one, but not as good as 39.

Let's go back to the best fit



For this best fitting line, $\beta_0 = -.41$, $\beta_1 = cov(x, y) / var(x) = 2.24$

Ordinary Least Squares Regression: Some Math

The predicted values or fitted values $\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 x_j$. The differences between the observations y_j and the predicted values $\hat{y}_j$ are called the residuals (errors), i.e.

$$r_j = y_j - \hat{y}_j = y_j - \hat{\beta}_0 - \hat{\beta}_1 x_j.$$

The sum of squared residuals (SSE) is defined as

$$SSE = \sum_{j=1}^n (y_j - \hat{\beta}_0 - \hat{\beta}_1 x_j)^2 = \sum_{j=1}^n y_j^2 - \hat{\beta}_0 \sum_{j=1}^n y_j - \hat{\beta}_1 \sum_{j=1}^n x_j y_j = S_{yy} - \hat{\beta}_1 S_{xy} =$$

SSE measures the amount of variability that the linear model can not explain.

Ordinary Least Squares Regression: Some Math

The total sum of squares (SST) is defined as

$$SST = \sum_{j=1}^n (y_j - \bar{y})^2 = S_{yy}.$$

SST measures the total amount of variation in observed y values. Hence $\frac{SSE}{SST}$ measures the proportion of total variation that can not be explained by the linear regression. It can be shown that $0 \leq \frac{SSE}{SST} \leq 1$.
Proof.

$$\min_{\beta_0, \beta_1} \sum_{j=1}^n (y_j - \beta_0 - \beta_1 x_j)^2 \leq \min_{\beta_0} \sum_{j=1}^n (y_j - \beta_0)^2 \leq \sum_{j=1}^n (y_j - \bar{y})^2.$$

Ordinary Least Squares Regression: Some Math

Key Takeaway

The equation of the line of best fit minimizes the residuals.

Luckily, our best friend $\mathbb{R}$ does all of this for us

To visualize check out: `https://seeing-theory.brown.edu/regression-analysis/`

Our Key Model: Regression(Again, and not for the last time)

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- Y_i = dependent variable, outcome variable, the thing we want to predict
- X_i = independent variable, explanatory variable, the thing that predicts the DV
- ϵ_i = the error term, the part of the DV our IV doesn't explain, everything NOT in our model (usually a lot!). Note: the error term is not directly observable.

Key Assumptions(for now)

- **Linear**
- constant
- continuous

Interpreting Regression Outputs

REMEMBER THIS

Our core statistical model is

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

1. β_1 , the slope, indicates how much change in Y (the dependent variable) is expected if X (the independent variable) increases by one unit.
2. β_0 , the intercept, indicates where the regression line crosses the Y -axis. It is the value of Y when X is zero.
3. β_1 is almost always more interesting than β_0 .

β_1 - slope coefficient, relationship between X & Y

β_0 - constant; value of Y when X is zero (intercept)

Regression in Practice

We think in terms of a model for the data:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

- X is the *independent variable* or *predictor* or *covariate*, and
- Y is the *dependent variable*
- β_0 is the “intercept”: the value of Y_i we will guess at $X = 0$.
- β_1 is the “slope”: a one-unit change in X is associated with a β_1 change in Y .
- ϵ_i called the “residual”

Then we use data to estimate this model(plug values into our model):

- we estimate $\hat{\beta}_0, \hat{\beta}_1$
- $\hat{Y}_i$ is our best guess of each Y_i . You can construct $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$,
- the “fitted residual”, $Y_i - \hat{Y}_i = \hat{\epsilon}_i$

Regression in R

Typically though, we use an R function called `lm`, for “linear model”:

```
> lm.out = lm(y~x)
```

```
> summary(lm.out)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.4144	0.6969	-0.595	0.56853
x	2.2350	0.5922	3.774	0.00544 **

Note: Just like the equation for a regression, Y goes first. In other functions in R, like `plot`, X goes first. Think about the order in which you type things.

The intercept (β_0) and slope (β_1) are called *coefficients*, and listed under “estimate” column.

What we can interpret:

- 1 Significance - p-value
- 2 Sign - coefficient positive or negative
- 3 Size - coefficient's value
- 4 Substantive - stated in real world terms (terms an art historian would understand)

An example

Enough of x and y , let's look at *yearsschool* and *growth*

- We'll call *yearsschool* the independent variable/ predictor, and *growth* the dependent variable/ outcome
- We say we “regress (outcome) on (predictor)”, so we “regress *growth* on *yearsschool*”.

```
> lm.out=lm(dat$growth~dat$yearsschool)
> ## Or, same thing this way:
> lm.out=lm(growth~yearsschool, data=dat)
> summary(lm.out)
Call:
lm(formula = growth ~ yearsschool, data = dat)
```

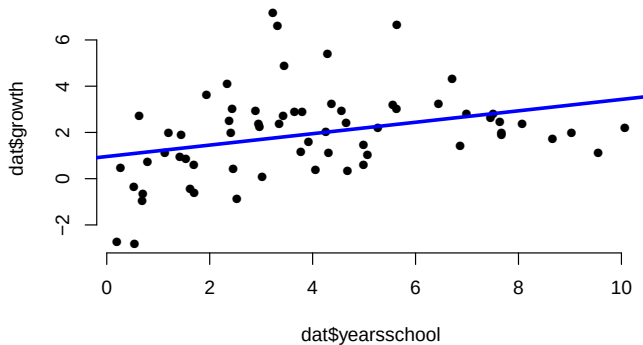
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.95829	0.41846	2.290	0.02538 *
yearsschool	0.24703	0.08873	2.784	0.00708 **

Residual standard error: 1.804 on 63 degrees of freedom
Multiple R-squared: 0.1096, Adjusted R-squared: 0.09543
F-statistic: 7.752 on 1 and 63 DF, p-value: 0.007077

What Your Regression is Saying

```
plot(dat$yearsschool, dat$growth, pch=16, frame.plot=FALSE)  
abline(lm.out, lwd=3, col=4)
```



Interpretation of the Regression Table

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.95829	0.41846	2.290	0.02538 *
yearsschool	0.24703	0.08873	2.784	0.00708 **

Residual standard error: 1.804 on 63 degrees of freedom

Multiple R-squared: 0.1096, Adjusted R-squared: 0.09543

F-statistic: 7.752 on 1 and 63 DF, p-value: 0.007077

- How to interpret the intercept?
At yearsschool = 0, we expect growth of 0.96.
- How to interpret the *yearsschool* coefficient estimate? *A one year change in yearsschool is associated with a 0.25 unit change in expected growth*
- Note that the units in the above statement reflect the original units of the *X* and the *Y*
- The “R-squared” tells us how much of the variation of *Y* (*growth*) is “explained” by *X* (*yearsschool*). We’ll talk more about this next week.

What we can interpret:

- 1 Significance - p-value
- 2 Sign - coefficient positive or negative
- 3 Size - coefficient's value
- 4 Substantive - stated in real world terms

Important You will need to be able to translate regression outcomes across:

- 1 Output table
- 2 Equation/Math
- 3 Visuals/Graph
- 4 Words/Writing

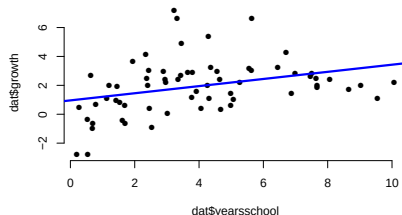
Interpretation Example

```
> lm.out=lm(growth~yearsschool, data=dat)
> summary(lm.out)
Call:
lm(formula = growth ~ yearsschool, data = dat)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.95829	0.41846	2.290	0.02538 *
yearsschool	0.24703	0.08873	2.784	0.00708 **

$$\hat{Y}_i \text{growth} = \beta_0 + \beta_1 X_i \text{yearsschool} + \epsilon_i$$
$$\hat{Y}_i \text{growth} = .958 + .247 X_i \text{yearsschool} + \epsilon_i$$



The effect of years in school is statistically significant at the .01 level. An increase of one year of schooling corresponds to a .25% increase in GDP growth.

This means that on average for every four years of schooling, for example from middle school to high school, we would expect a full percentage point increase in GDP.

Key Terms

- Covariance
- Correlation
- Regression
- OLS - Ordinary Least Squares
- Coefficient
- Residual
- Fitted Value
- Sum of Squared Errors

- Reading for Next Week: Bailey Ch. 5 "Multivariate OLS"