SOC 3811/5811:
BASIC SOCIAL STATISTICS

Associations Between Continuous Variables

Associations Between Variables

Between now and the 3rd exam we will focus on measuring the association between two variables, X & Y

1. When X is discrete and Y is continuous, we will use "analysis of variance" techniques (Last Tuesday)
2. When X and Y are both discrete, we will use cross-tabular and χ^2 analyses (Last Thursday)
3. When X and Y are both continuous, we will use correlation & regression analyses (This Week)

Correlation and Regression: Review

Correlation (r_{YX}) measures the strength and direction of the association between continuous variables Y and X

Correlation and Regression: Review

Bivariate regression involves drawing a line through the points on the scatterplot such that the sum of the squared prediction errors equals zero

Regression analysis allows us to…

- …quantify the degree to which variability in Y is "explained by" variability in X (using R^2_{YX});
- …describe how, on average, the response variable (Y) is related to the predictor variable (X) (using b_{YX}); and
- …make predictions about the value of the response variable (Y) given a specified value of the predictor variable (X)

Inferences About Associations

So far, we have been describing associations between continuous variables using sample data

We usually want to make inferences about associations between continuous variables in the population from which the sample data were drawn

For example:

- Is the relationship observed in the sample data strong enough to confidently conclude that there is a relationship in the population?
- What can we infer about the likely values of the slope in the population based on the slope in the sample?

Inferences About Associations

Hypothesis Tests About ρ^2_{YX}

R^2_{YX} is a sample estimate of population parameter ρ^2_{YX}

If ρ^2_{YX} equals zero, then X does nothing to explain variability in Y

Hypothesis Tests About ρ_{YX}

r_{YX} is a sample estimate of population parameter ρ_{YX}

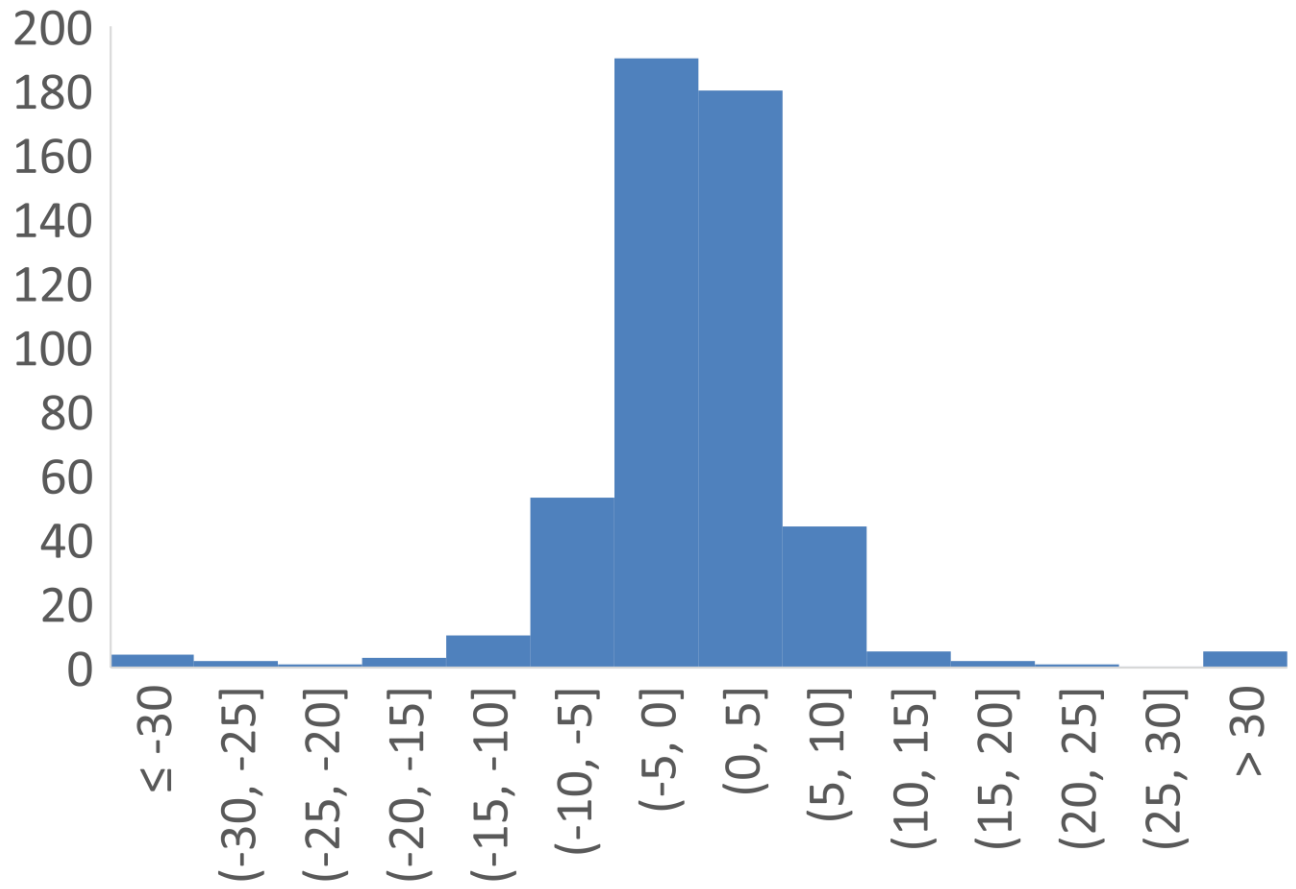
If ρ_{YX} equals zero, then there is no correlation between X and Y

Hypothesis Tests About Slope β_{YX}

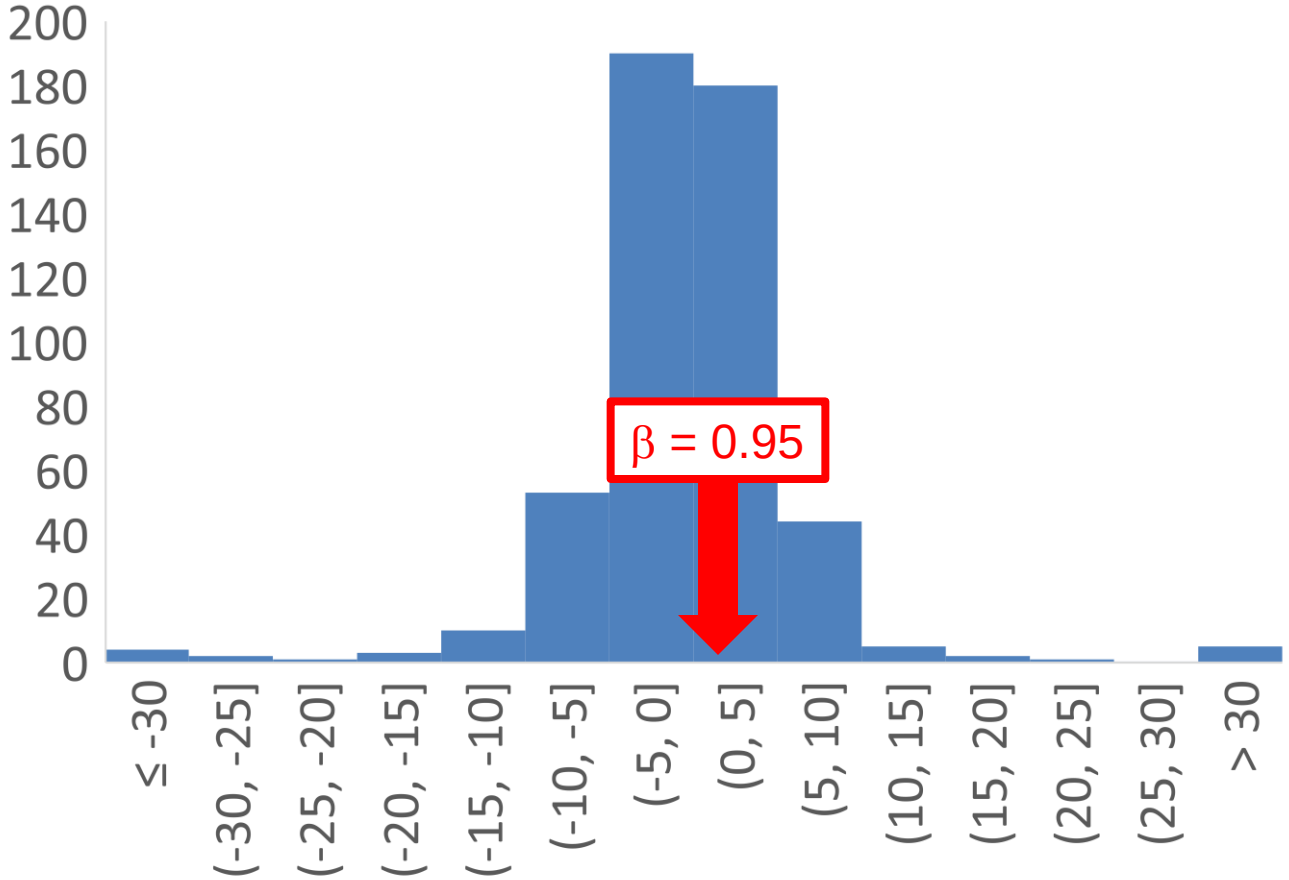
b_{YX} is a sample estimate of population parameter β_{YX}

If β_{YX} equals zero, then the regression of Y on X has a zero slope

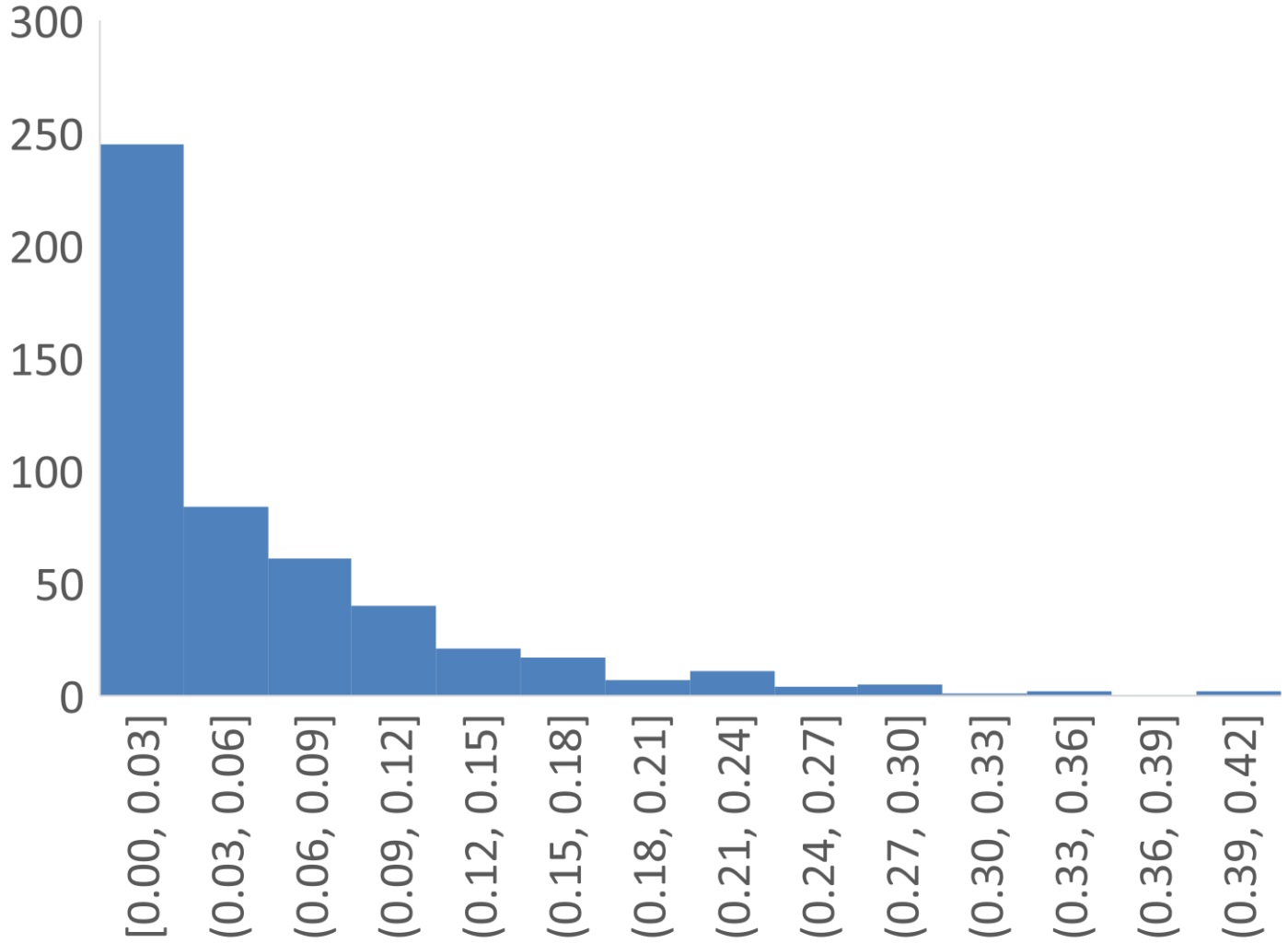
b_{YX} Values from 500 Random Samples (in each, n=20):
Regressions of Y="Total (Minus October) Snowfall" on X="October Snowfall"
in Minneapolis, 1883-2019



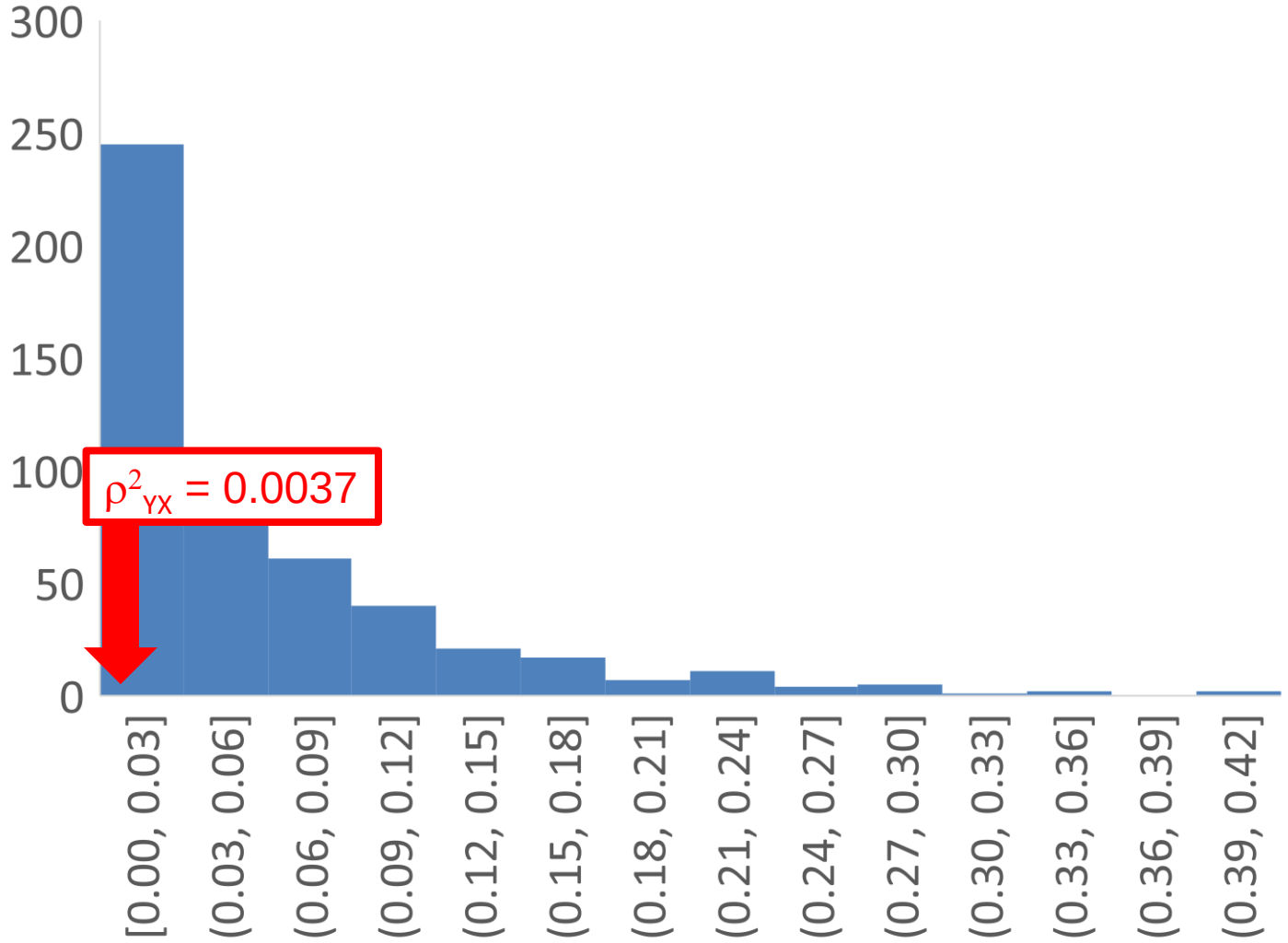
b_{YX} Values from 500 Random Samples (in each, n=20):
Regressions of Y="Total (Minus October) Snowfall" on X="October Snowfall"
in Minneapolis, 1883-2019



R^2_{YX} Values from 500 Random Samples (in each, n=20):
Regressions of Y="Total (Minus October) Snowfall" on X="October Snowfall"
in Minneapolis, 1883-2019



R^2_{YX} Values from 500 Random Samples (in each, n=20):
Regressions of Y="Total (Minus October) Snowfall" on X="October Snowfall"
in Minneapolis, 1883-2019



Inferences About Associations

Hypothesis Tests About ρ^2_{YX}

R^2_{YX} is a sample estimate of population parameter ρ^2_{YX}

If ρ^2_{YX} equals zero, then X does nothing to explain variability in Y

Hypothesis Tests About ρ_{YX}

r_{YX} is a sample estimate of population parameter ρ_{YX}

If ρ_{YX} equals zero, then there is no correlation between X and Y

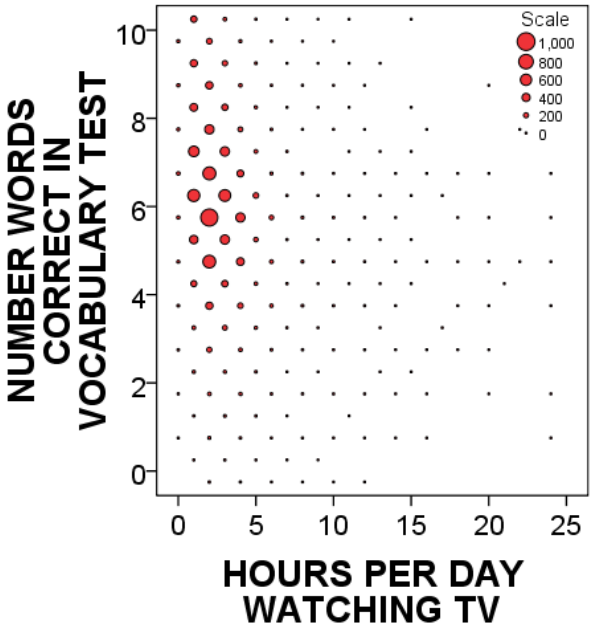
Hypothesis Tests About Slope β_{YX}

b_{YX} is a sample estimate of population parameter β_{YX}

If β_{YX} equals zero, then the regression of Y on X has a zero slope

Example

The scatterplot below relates GSS respondents' hours per week watching TV (X) & their vocabulary test score (Y)



$\bar{Y} = 6.01$ $\bar{X} = 2.95$

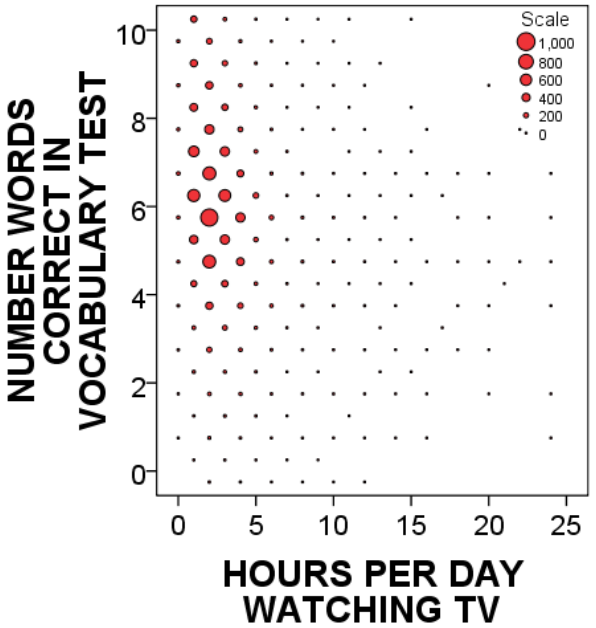
$s_Y = 2.13$ $s_X = 2.31$

$r_{YX} = -0.193$ $n = 15{,}357$

Note: These are the only statistics required to estimate the least squares regression equation and to perform the tests described today

Example

The scatterplot below relates GSS respondents' hours per week watching TV (X) & their vocabulary test score (Y)



$$b_{YX} = r_{YX}\frac{s_Y}{s_X}$$

$$b_{YX} = -0.193\frac{2.13}{2.31} = -0.178$$

$$a = \bar{Y} - b\bar{X}$$

$$a = 6.01 + 0.178(2.95)$$

$$a = 6.535$$

$$R^2_{YX} = r^2_{YX} = 0.193^2 = 0.037$$

Assumptions of the Regression Model

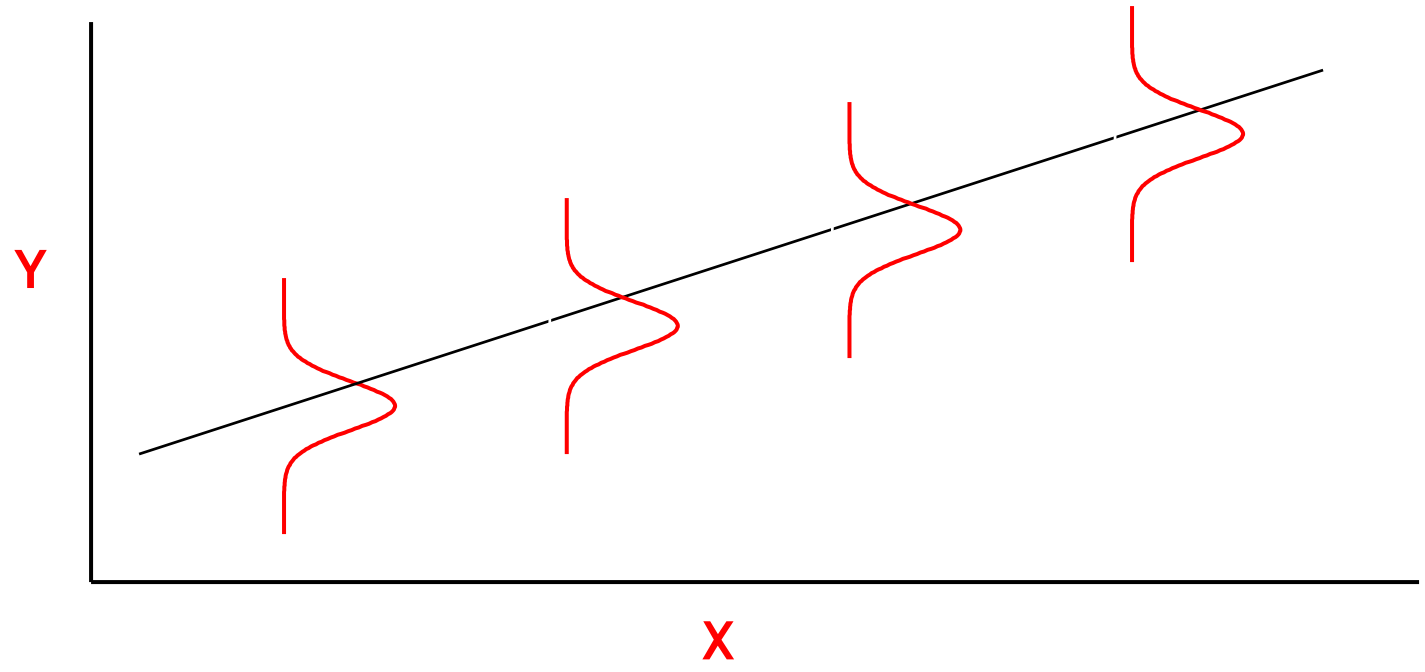
Assumptions of the Regression Model

Assumptions we make when using sample-based regression models to make inferences about associations in the population:

1. The functional form of the relationship between X and Y is appropriately specified; usually this means checking for linearity
2. There are no extreme outliers
3. The variability of the prediction errors is constant across the observed values of X (assumption of homoskedasticity)
4. The values of Y are normally distributed at each value of X (assumption of normality)
5. The observations are independent

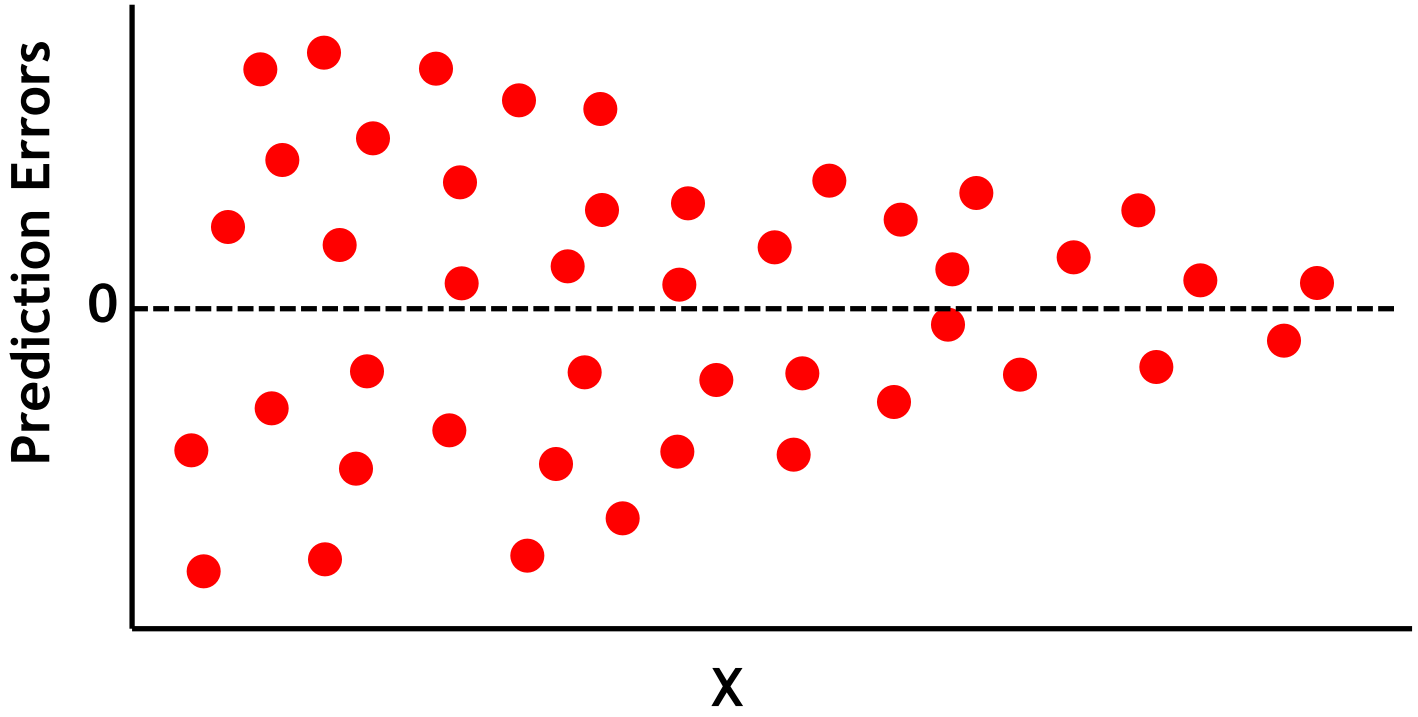
Assumptions of the Regression Model

Graphical summary of the first four assumptions:



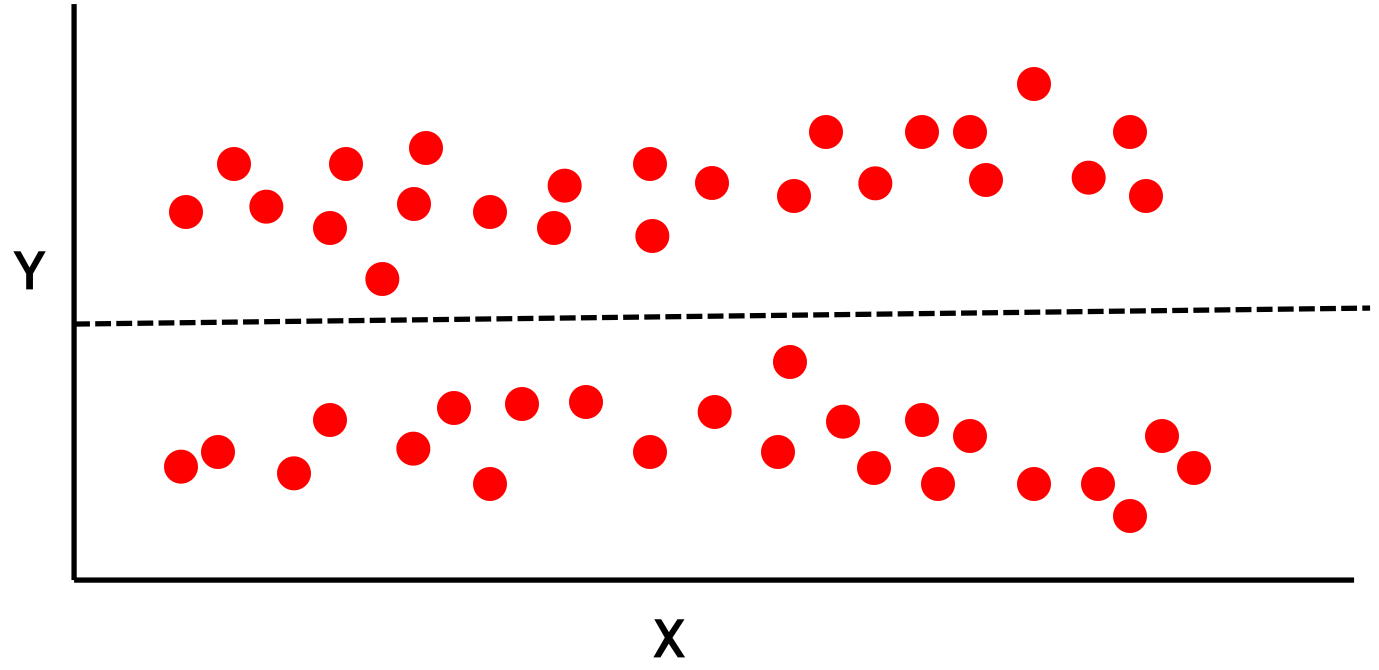
Assumptions of the Regression Model

What does it look like if variability of the prediction errors are not constant across all values of X? (*Note*: This is just one possibility)



Assumptions of the Regression Model

What does it look like if the values of Y are not normally distributed at all values of X in the population? (*Note*: This is just one possibility)



Assumptions of the Regression Model

The first three assumptions…

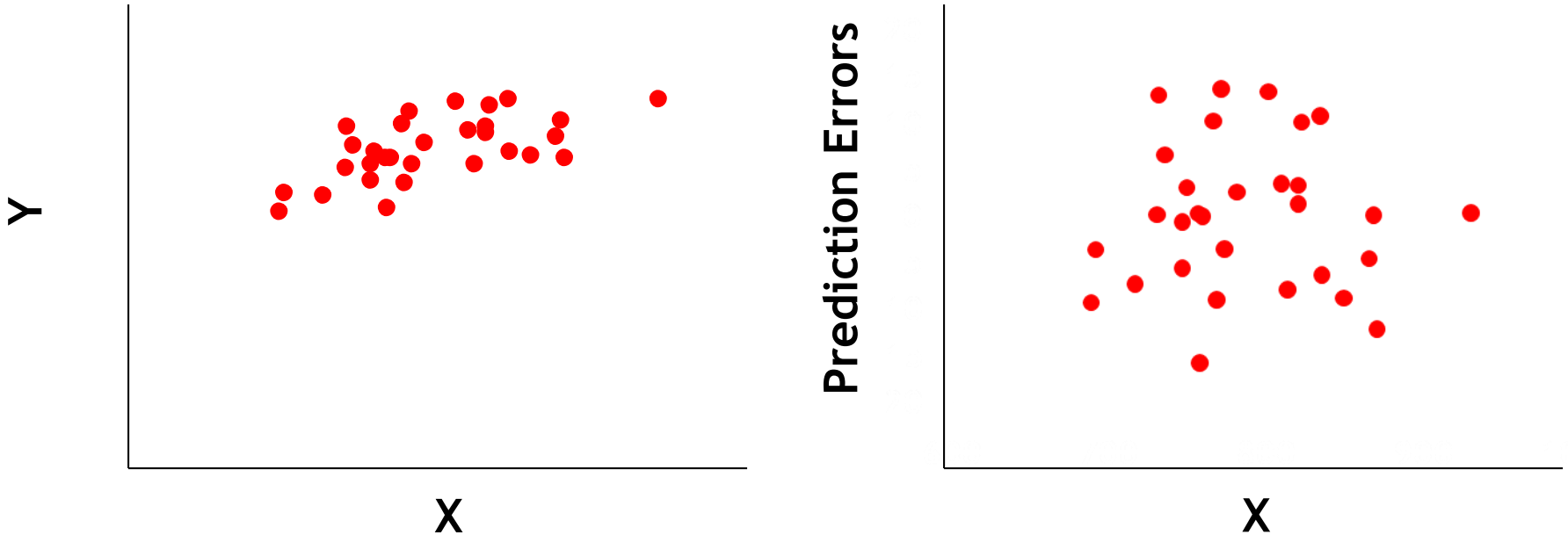
1. The functional form of the relationship between X and Y is appropriately specified; usually this means checking for linearity
2. There are no extreme outliers
3. The variability of the prediction errors is constant across the observed values of X

…can be checked using two plots

First, create a scatterplot with X on the horizontal axis and Y on the vertical axis

Second, create a scatterplot with X on the horizontal axis and the prediction errors on the vertical axis

Assumptions of the Regression Model



1. Is the association between X and Y plausibly linear?
2. Are there extreme outliers?
3. Is the variability of the prediction errors (the residuals) constant across the range of X?

Assumptions of the Regression Model

The fourth assumption…

4. The values of Y are normally distributed at each value of X

…can be checked by examining a histogram of the prediction errors (or residuals)

The fifth assumption…

5. The observations are independent

…is a matter of appropriate research design

Assumptions of the Regression Model

In later courses you may learn about more sophisticated techniques for diagnosing violations of the assumptions of the linear regression model

For now, just be aware of these assumptions

If the assumptions are not met, then hypothesis tests about ρ^2_{YX}, ρ_{YX}, and β_{YX} are generally invalid

Inferences About Associations

Hypothesis Testing in 6 Steps

1. State the null (H_0) and alternative (H_1) hypotheses
2. Check that the sample data conform to basic assumptions; if they do not, then do not go any further
3. Choose an α probability level … that is, a probability associated with incorrectly rejecting the null hypothesis
4. Determine the "critical value" … that is, how large the test statistic must be in order to reject the null hypothesis at the given α level
5. Calculate the test statistic … F or Z or t, depending
6. Compare the test statistic to the critical value

Inferences About ρ^2_{YX}

Inferences About ρ^2_{YX}

State the null (H_0) and alternative (H_1) hypotheses

H_0: $\rho^2_{YX} = 0$

H_1: $\rho^2_{YX} > 0$

This is a one-sided test (with no <) because ρ^2_{YX} cannot possibly be less than zero

Failing to reject the null means failing to reject the hypothesis that X explains none of the variation in Y

Inferences About ρ^2_{YX}

Check that the sample data conform to basic assumptions; if they do not, then do not go any further

The assumptions of the regression model described earlier must hold for hypothesis tests about ρ^2_{YX} to be valid

Inferences About ρ^2_{YX}

Choose an α probability level … that is, a probability associated with incorrectly rejecting the null hypothesis

Let's choose α=0.01

Inferences About ρ^2_{YX}

Determine the "critical value" … that is, how large the test statistic must be in order to reject the null hypothesis at the given α level

The hypothesis test for ρ^2_{YX} is (as described below) an F test with $df_{NUM}=1$ and $df_{DENOM}=n-2$

In our example, we want $F_{1,15355}$ for $\alpha=0.01$ which is 6.63

We will thus reject H_0 if our F statistic exceeds 6.63

Critical Values of F

(α=0.01)

DENOMINATOR Degrees of Freedom	NUMERATOR Degrees of Freedom																	
	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	100	200	∞
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6157.28	6208.73	6260.65	6286.78	6302.52	6334.11	6349.97	6365.86
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.43	99.45	99.47	99.47	99.48	99.49	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	26.87	26.69	26.50	26.41	26.35	26.24	26.18	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.20	14.02	13.84	13.75	13.69	13.58	13.52	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.72	9.55	9.38	9.29	9.24	9.13	9.08	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.56	7.40	7.23	7.14	7.09	6.99	6.93	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.31	6.16	5.99	5.91	5.86	5.75	5.70	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.52	5.36	5.20	5.12	5.07	4.96	4.91	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.96	4.81	4.65	4.57	4.52	4.41	4.36	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.56	4.41	4.25	4.17	4.12	4.01	3.96	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.25	4.10	3.94	3.86	3.81	3.71	3.66	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.01	3.86	3.70	3.62	3.57	3.47	3.41	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.82	3.66	3.51	3.43	3.38	3.27	3.22	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.66	3.51	3.35	3.27	3.22	3.11	3.06	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.52	3.37	3.21	3.13	3.08	2.98	2.92	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.41	3.26	3.10	3.02	2.97	2.86	2.81	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.31	3.16	3.00	2.92	2.87	2.76	2.71	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.23	3.08	2.92	2.84	2.78	2.68	2.62	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.15	3.00	2.84	2.76	2.71	2.60	2.55	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.09	2.94	2.78	2.69	2.64	2.54	2.48	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.03	2.88	2.72	2.64	2.58	2.48	2.42	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	2.98	2.83	2.67	2.58	2.53	2.42	2.36	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	2.93	2.78	2.62	2.54	2.48	2.37	2.32	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	2.89	2.74	2.58	2.49	2.44	2.33	2.27	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.85	2.70	2.54	2.45	2.40	2.29	2.23	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.81	2.66	2.50	2.42	2.36	2.25	2.19	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.78	2.63	2.47	2.38	2.33	2.22	2.16	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.75	2.60	2.44	2.35	2.30	2.19	2.13	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.73	2.57	2.41	2.33	2.27	2.16	2.10	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.70	2.55	2.39	2.30	2.25	2.13	2.07	2.01
31	7.53	5.36	4.48	3.99	3.67	3.45	3.28	3.15	3.04	2.96	2.68	2.52	2.36	2.27	2.22	2.11	2.04	1.98
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93	2.65	2.50	2.34	2.25	2.20	2.08	2.02	1.96
33	7.47	5.31	4.44	3.95	3.63	3.41	3.24	3.11	3.00	2.91	2.63	2.48	2.32	2.23	2.18	2.06	2.00	1.93
34	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89	2.61	2.46	2.30	2.21	2.16	2.04	1.98	1.91
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.60	2.44	2.28	2.19	2.14	2.02	1.96	1.89
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.52	2.37	2.20	2.11	2.06	1.94	1.87	1.80
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.42	2.27	2.10	2.01	1.95	1.82	1.76	1.68
75	6.99	4.90	4.05	3.58	3.27	3.05	2.89	2.76	2.65	2.57	2.29	2.13	1.96	1.87	1.81	1.67	1.60	1.52
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.22	2.07	1.89	1.80	1.74	1.60	1.52	1.43
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.13	1.97	1.79	1.69	1.63	1.48	1.39	1.28
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.04	1.88	1.70	1.59	1.52	1.36	1.25	1.00

Inferences About ρ^2_{YX}

Calculate the test statistic

Remember from before…

$$\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)^2=\sum_{i=1}^{N}\left(\hat{Y}_i-\bar{Y}\right)^2+\sum_{i=1}^{N}\left(Y_i-\hat{Y}_i\right)^2$$

Total Sum of Squares

Regression Sum of Squares

Error Sum of Squares

$SS_{TOTAL} = SS_{REGRESSION} + SS_{ERROR}$

We defined $R^2_{YX}=\dfrac{SS_{TOTAL}-SS_{ERROR}}{SS_{TOTAL}}=\dfrac{SS_{REGRESSION}}{SS_{TOTAL}}$

Inferences About ρ^2_{YX}

Calculate the test statistic

The F statistic here is

$$F_{1,N-2} = \frac{SS_{REGRESSION}/1}{SS_{ERROR}/n-2} = \frac{MS_{REGRESSION}}{MS_{ERROR}}$$

There is an analogy between this and ANOVA … in both cases we are asking whether variation in Y can be attributed to individuals' values on X

If X and Y are associated, then $MS_{REGRESSION}$ (and thus F) will be larger

Inferences About ρ^2_{YX}

Calculate the test statistic

Computationally: $SS_{TOTAL} = (s^2_Y)(n-1)$

$SS_{REGRESSION} = (R^2_{YX})(SS_{TOTAL})$

$SS_{ERROR} = SS_{TOTAL} - SS_{REGRESSION}$

In our example: $SS_{TOTAL} =$

$s_Y = 2.13$ $SS_{REGRESSION} =$

$r_{YX} = -0.193$ $SS_{ERRO_R} =$

$n = 15{,}357$

$$F_{1,N-2} = \frac{SS_{REGRESSION}/1}{SS_{ERROR}/n-2} = \frac{MS_{Regression}}{MS_{Error}} =$$

Inferences About ρ^2_{YX}

Calculate the test statistic

Computationally:

$$SS_{TOTAL} = (s_Y^2)(n-1)$$

$$SS_{REGRESSION} = (R_{YX}^2)(SS_{TOTAL})$$

$$SS_{ERROR} = SS_{TOTAL} - SS_{REGRESSION}$$

In our example:

$s_Y = 2.13$

$r_{YX} = -0.193$

$n = 15{,}357$

$$SS_{TOTAL} = (2.13^2)(15{,}357 - 1) = 69{,}668.6$$

$$SS_{REGRESSION} = (0.193^2)(69{,}668.6) = 2{,}595.1$$

$$SS_{ERROR} = 69{,}668.6 - 2{,}595.1 = 67{,}073.5$$

$$F_{1,N-2} = \frac{SS_{REGRESSION}/1}{SS_{ERROR}/n-2} = \frac{MS_{Regression}}{MS_{Error}} = \frac{2{,}595.1/1}{67{,}073.5/15{,}355} = 594.1$$

Inferences About ρ^2_{YX}

Compare the test statistic to the critical value

- If the test statistic is larger than the critical value, then reject H_0
- If the test statistic is less than or equal to the critical value, then do not reject H_0

We can restate the hypotheses:

H_0: $\rho^2_{YX} = 0$ ➔ Fail to reject H_0 if $F \leq 6.63$

H_1: $\rho^2_{YX} > 0$ ➔ Reject H_0 if $F > 6.63$

Since F=594.1, we reject H_0 … so it appears that in the population X and Y are associated, such that X accounts for some of the variability in Y

Worksheet

Mean of X:	6.50	SD of X:	2.95
Mean of Y:	7.15	SD of Y:	1.46
r_{XY}:	0.51		
n:	20		

Test the hypothesis that ρ^2_{YX} --- the population proportion of variation in Y explained by X --- is zero in the population; use α=0.05

Inferences About Correlation (ρ_{YX})

Inferences About Correlation (ρ_{YX})

Since r_{YX} is just the square root of R^2_{YX} (in the case of bivariate regression), a hypothesis test about ρ_{YX} will yield the same result as a hypothesis test about ρ^2_{YX}

Another way to directly test hypotheses about the correlation coefficient ρ_{YX} is to utilize the r-to-Z transformation:

$$Z_r = \left(\frac{1}{2}\right) \ln\left(\frac{1+r_{YX}}{1-r_{YX}}\right)$$

The following test statistic has a standard normal distribution:

$$Z = \frac{Z_r - 0}{\sqrt{1/n-3}}$$

Inferences About Correlation (ρ_{YX})

State the null (H_0) and alternative (H_1) hypotheses

H_0: $\rho_{YX} = 0$

H_1: $\rho_{YX} \neq 0$

This is a two-sided test since ρ_{YX} can range from -1 to +1

Failing to reject the null means failing to reject the hypothesis that X and Y are uncorrelated in the population

Inferences About Correlation (ρ_{YX})

Check that the sample data conform to basic assumptions; if they do not, then do not go any further

The assumptions of the regression model described earlier must hold for hypothesis tests about ρ_{YX} to be valid

Inferences About Correlation (ρ_{YX})

Choose an α probability level … that is, a probability associated with incorrectly rejecting the null hypothesis

Let's choose α=0.05

Inferences About Correlation (ρ_{YX})

Determine the "critical value" … that is, how large the test statistic must be in order to reject the null hypothesis at the given α level

Given α=0.05 for this two-sided test, the critical value Z* equals 1.96

Inferences About Correlation (ρ_{YX})

Calculate the test statistic

$r_{YX} = -0.193$

$n = 15{,}357$

$$Z_r = \left(\frac{1}{2}\right) \ln\left(\frac{1 + r_{YX}}{1 - r_{YX}}\right) =$$

$$Z = \frac{Z_r - 0}{\sqrt{1/n - 3}} =$$

Inferences About Correlation (ρ_{YX})

Calculate the test statistic

$r_{YX} = -0.193$

$n = 15{,}357$

$$Z_r = \left(\frac{1}{2}\right)\ln\left(\frac{1 + r_{YX}}{1 - r_{YX}}\right) = \left(\frac{1}{2}\right)\ln\left(\frac{1 - 0.193}{1 + 0.193}\right) = -0.195$$

$$Z = \frac{Z_r - 0}{\sqrt{1/n - 3}} = \frac{-0.195 - 0}{\sqrt{1/15{,}354}} = -24.2$$

Inferences About Correlation (ρ_{YX})

Compare the test statistic to the critical value

- If the test statistic is larger than the critical value, then reject H_0
- If the test statistic is less than or equal to the critical value, then do no reject H_0

We can restate the hypotheses:

H_0: $\rho_{YX} = 0$ ➔ Fail to reject H_0 if $|Z| \leq 1.96$

H_1: $\rho_{YX} \neq 0$ ➔ Reject H_0 if $|Z| > 1.96$

Since Z=-24.2, we reject H_0 … so it appears that in the population X and Y are correlated

Worksheet

Mean of X:	6.50	SD of X:	2.95
Mean of Y:	7.15	SD of Y:	1.46
r_{XY}:	0.51		
n:	20		

Test the hypothesis that ρ_{YX} --- the population correlation between X and Y --- is zero in the population; use α=0.05

Inferences About Slope (β_{YX})

Inferences About Slope (β_{YX})

The formula for the sample prediction equation is:

$$\hat{Y}_i = a + b_{YX} X_i$$

The formula for the population prediction equation is:

$$\hat{Y}_i = \alpha + \beta_{YX} X_i \quad \text{(sometimes written as } E(Y_i) = \alpha + \beta_{YX} X_i \text{)}$$

We use b_{YX} as an estimate of β_{YX}

Because b_{YX} is a sample estimate, we know that they would vary from sample to sample if we were to take repeated samples of the same size from the population

Inferences About Slope (β_{YX})

The sampling distribution of b_{YX} is normally distributed, centered over β_{YX}, with variance:

$$\sigma_b^2 = \frac{\sigma_e^2}{\sum_{i=1}^{N}(X_i - \overline{X})^2}$$

where σ^2_e is the population variance of the prediction errors

We can use MS_{ERROR} as a sample estimate of σ^2_e and the denominator can be re-expressed as $(s^2_X)(n-1)$, so the standard error of the sampling distribution of b_{YX} is

$$s_b^2 = \sqrt{\frac{MS_{Error}}{(s_X^2)(n-1)}}$$

Inferences About Slope (β_{YX})

State the null (H_0) and alternative (H_1) hypotheses

H_0: $\beta_{YX} = 0$

H_1: $\beta_{YX} \neq 0$

This is normally a two-sided test, although it needn't be

Inferences About Slope (β_{YX})

Check that the sample data conform to basic assumptions; if they do not, then do not go any further

The assumptions of the regression model described earlier must hold for hypothesis tests about β_{YX} to be valid

Inferences About Slope (β_{YX})

Choose an α probability level … that is, a probability associated with incorrectly rejecting the null hypothesis

Let's go with α=0.05

Inferences About Slope (β_{YX})

Determine the "critical value" … that is, how large the test statistic must be in order to reject the null hypothesis at the given α level

Because we are using sample-based estimates of the variability in the sampling distribution of b_{YX}, we will conduct a t (instead of a Z) test

Because MS_{ERROR} has n-2 degrees of freedom, we will select a critical value of t with df=n-2

For a two-sided test with α=0.05 and n-2=15,355 degrees of freedom, the critical value t*=1.962

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.

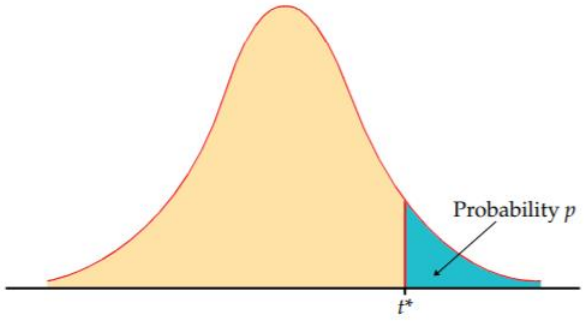


TABLE D

t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%

Inferences About Slope (β_{YX})

Calculate the test statistic

$b_{YX} = -0.178$

$n = 15{,}357$

$s_X = 2.31$

The test statistic t with df=N-2 equals:

$$t_{n-2} = \frac{b_{YX} - 0}{s_b} = \frac{b_{YX} - 0}{\sqrt{\frac{MS_{Error}}{(s_X^2)(n-1)}}}$$

In our example:

$$F_{1,N-2} = \frac{SS_{Regression}/1}{SS_{Error}/n-2} = \frac{MS_{Regression}}{MS_{Error}} \frac{2{,}595.1/1}{67{,}073.5/15{,}355} = \frac{2595.1}{4.368} = 594.1$$

$$t_{n-2} =$$

Inferences About Slope (β_{YX})

Calculate the test statistic

$b_{YX} = -0.178$

$n = 15{,}357$

$s_X = 2.31$

The test statistic t with df=N-2 equals:

$$t_{n-2} = \frac{b_{YX} - 0}{s_b} = \frac{b_{YX} - 0}{\sqrt{\frac{MS_{Error}}{(s_X^2)(n-1)}}}$$

In our example:

$$F_{1,N-2} = \frac{SS_{Regression}/1}{SS_{Error}/n-2} = \frac{MS_{Regression}}{MS_{Error}} \frac{2{,}595.1/1}{67{,}073.5/15{,}355} = \frac{2595.1}{4.368} = 594.1$$

$$t_{n-2} = \frac{-0.178 - 0}{\sqrt{\frac{4.368}{(2.31^2)(15{,}356)}}} = -24.4$$

Inferences About Slope (β_{YX})

Compare the test statistic to the critical value

- If the test statistic is larger than the critical value, then reject H_0
- If the test statistic is less than or equal to the critical value, then do no reject H_0

We can restate the hypotheses:

H_0: $\beta_{YX} = 0$ ➔ Fail to reject H_0 if $|t| \leq 1.962$

H_1: $\beta_{YX} \neq 0$ ➔ Reject H_0 if $|t| > 1.962$

Since t = -24.4, we reject H_0

Worksheet

Mean of X:	6.50	SD of X:	2.95
Mean of Y:	7.15	SD of Y:	1.46
r_{XY}:	0.51		
n:	20		

Test the hypothesis that β_{YX} --- the population slope relating Y to X --- is zero in the population; use α=0.05

Confidence Intervals for β_{YX}

Inferences About Slope (β_{YX})

Using the sample estimate (b_1) of β_{XY} and the standard error of the sampling distribution of β_{YX} (above), we can compute confidence intervals for β_{YX}

The standard error is:

$$s_b = \sqrt{\frac{MS_{ERROR}}{(s_X^2)(n-1)}}$$

So the confidence interval can be expressed as:

$$C.I. = b_1 \pm t * \sqrt{\frac{MS_{ERROR}}{(s_X^2)(n-1)}}$$

Inferences About Slope (β_{YX})

For our example, a 95% confidence interval would be:

$$\text{C.I.} = b_1 \pm 1.96\sqrt{\frac{MS_{ERROR}}{(s_X^2)(n-1)}} =$$

$$b_{YX} = -0.178$$

$$n = 15{,}357$$

$$s_X = 2.31$$

$$F_{1,N-2} = \frac{MS_{Regression}}{MS_{Error}} = \frac{2595.1}{4.368} = 594.1$$

Inferences About Slope (β_{YX})

For our example, a 95% confidence interval would be:

$$\text{C.I.} = b_1 \pm 1.96\sqrt{\frac{MS_{ERROR}}{(s_X^2)(n-1)}} = -0.178 \pm 1.96\sqrt{\frac{4.368}{(2.31^2)(15{,}356)}}$$

$$\text{C.I.} = -0.178 \pm 1.96(0.0074)$$

$$\text{C.I.} = -0.178 \pm 0.014$$

…so we are 95% certain that β_{YX} falls in within the interval -0.192 to -0.164

$$b_{YX} = -0.178$$

$$n = 15{,}357$$

$$s_X = 2.31$$

$$F_{1,N-2} = \frac{MS_{Regression}}{MS_{Error}} = \frac{2595.1}{4.368} = 594.1$$

Worksheet

Mean of X:	6.50	SD of X:	2.95
Mean of Y:	7.15	SD of Y:	1.46
r_{XY}:	0.51		
n:	20		

Construct a 95% confidence interval for β_{YX} --- the population slope relating Y to X

Standardized Regression Coefficients

Standardized Coefficients (β^*_{YX})

For a variety of reasons researchers often like to express the slope of the regression line in standardized terms

This is useful when:

- The metric of X is in an arbitrary scale, or a scale that is not intrinsically meaningful
- We want to better understand the magnitude of the association between X and Y

Instead of asking…

- "How many units does Y change as a result of a one unit change in X?"

we might ask,

- "How many standard deviations does Y change as a result of a one standard deviation change in X?"

Standardized Coefficients (β^*_{YX})

The standardized slope, or beta coefficient (or beta weight) is expressed as

$$\beta^*_{YX} = (b_{YX})\left(\frac{s_X}{s_Y}\right)$$

In bivariate regression, the standardized slope thus equals the correlation, r_{YX}

In our example:

$$\beta^*_{YX} = (-0.178)\left(\frac{2.31}{2.13}\right) = -0.193$$

Want More?

David Lane's Books

http://onlinestatbook.com/2/regression/regression.html

http://davidmlane.com/hyperstat/prediction.html

Stat Trek

http://stattrek.com/regression/linear-regression.aspx

Lowry's Book (Chapter 3)

http://vassarstats.net/textbook/

Dallal's Book (see "Simple Linear Regression" section)

http://www.jerrydallal.com/LHSP/LHSP.htm