SOC 3811/5811:
BASIC SOCIAL STATISTICS

Three Variable Relationships and Multiple Regression

Multiple Regression

We have reviewed regression techniques for describing the association between two continuous variables

However, we also talked about spuriousness … a threat to our ability to infer the causal impact of X on Y due to confounding variable(s) Z

How do we “statistically control” for Z using regression techniques?

Multiple Regression

Example: Why are some occupations (e.g., authors, machinists) considered to be more prestigious than others?

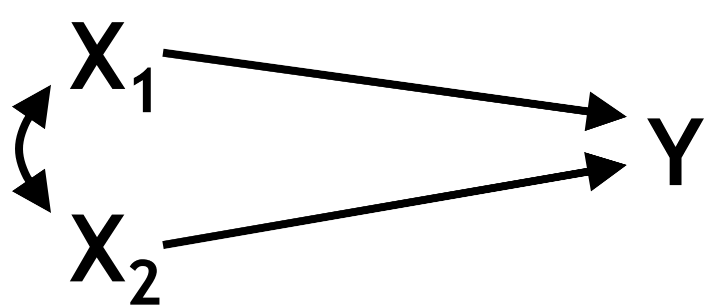
Y = The prestige accorded to 45 occupations

X_1 = How much education it requires to hold that occupation

X_2 = How well that occupation pays

What is the independent effect of X_1 on Y?

What is the independent effect of X_2 on Y?



Multiple Regression

Example: Why are some occupations (e.g., authors, machinists) considered to be more prestigious than others?

Y = The prestige accorded to 45 occupations

X_1 = How much education it requires to hold that occupation

X_2 = How well that occupation pays

Descriptive Statistics (always start by looking at descriptives)

	Y	X_1	X_2	Mean	SD
Y	1.00			47.7	31.5
X_1	0.85	1.00		52.6	29.8
X_2	0.84	0.73	1.00	41.9	24.4

Multiple Regression

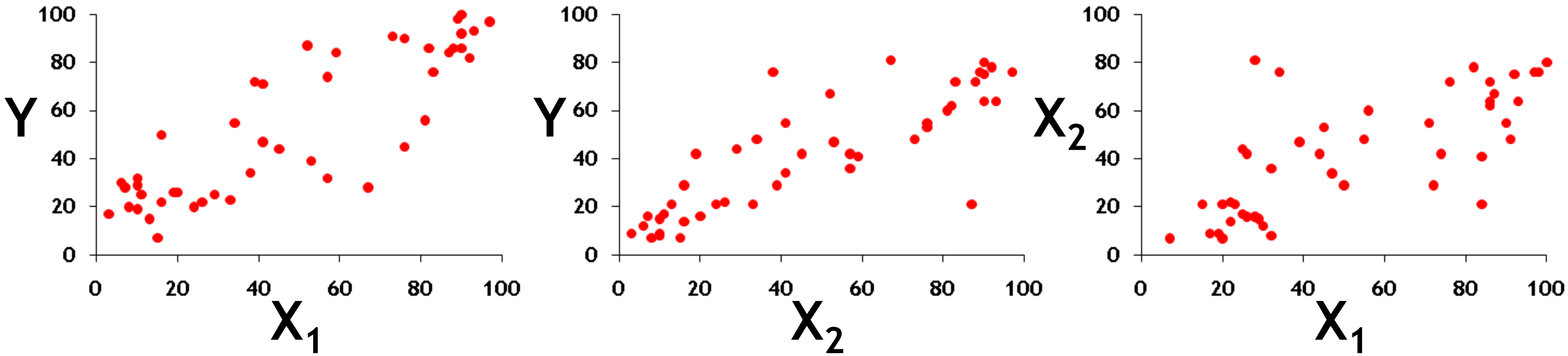
Example: Why are some occupations (e.g., authors, machinists) considered to be more prestigious than others?

Y = The prestige accorded to 45 occupations

X_1 = How much education it requires to hold that occupation

X_2 = How well that occupation pays

Bivariate Scatterplots (always start by looking at bivariate plots)



Multiple Regression

Example: Why are some occupations (e.g., authors, machinists) considered to be more prestigious than others?

Y = The prestige accorded to 45 occupations

X_1 = How much education it requires to hold that occupation

X_2 = How well that occupation pays

$$\hat{Y}_i = a + b_{YX_1} X_{1i} = 0.284 + 0.902 X_{1i}$$

$$\hat{Y}_i = a + b_{YX_2} X_{2i} = 2.457 + 1.080 X_{2i}$$

…but we <u>know</u> that neither slope (b_{YX_1} or b_{YX_2}) represents the "effects" of X_1 or X_2 because of confounding in the relationships between Y and the X's

Multiple Regression Analysis

Multiple Regression Analysis

“a statistical technique for estimating the relationship between a continuous dependent variable and two or more continuous or discrete independent, or predictor, variables”

For today, we will limit ourselves to…

…two predictor variables

…continuous predictor variables

Extensions to 3+ predictor variables and to discrete predictor variables will be natural extension of what we cover today

Multiple Regression Analysis

The population regression equation:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

The population prediction equation:

$$\hat{Y}_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i}$$

The sample regression equation:

$$Y_i = a + b_1 X_{1i} + b_2 X_{2i} + e_i$$

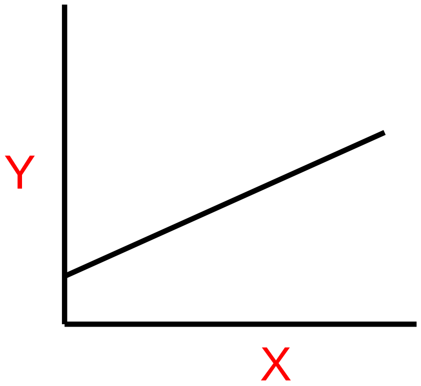
The sample prediction equation:

$$\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i}$$

Multiple Regression Analysis

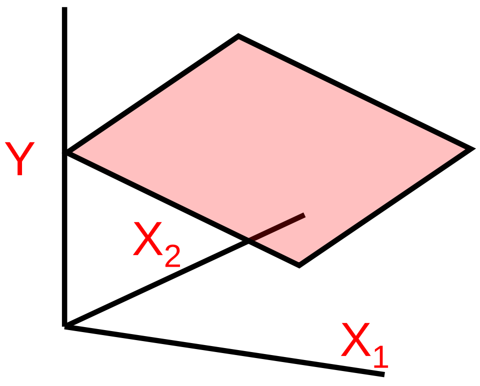
The bivariate regression prediction equation describes a 2-dimensional line

$$\hat{Y}_i = a + b_1 X_{1i}$$



The multivariate (2 independent variable) prediction equation describes a 3-dimensional plane

$$\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i}$$



Multiple Regression Analysis

The ordinary least squares (OLS) method is used to estimate a, b_1, and b_2 … again, this method minimizes the sum of the squared residuals (or prediction errors)

To compute a, b_1, and b_2 we only need the sample means, the standard deviations, and the correlations

$$b_1 = \left(\frac{s_Y}{s_{X_1}}\right)\frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{1 - r_{X_1X_2}^2}$$

$$b_2 = \left(\frac{s_Y}{s_{X_2}}\right)\frac{r_{YX_2} - r_{YX_1} r_{X_1X_2}}{1 - r_{X_1X_2}^2}$$

$$a = \overline{Y} - \left(b_1\overline{X}_1 + b_2\overline{X}_2\right)$$

Multiple Regression Analysis

Example:

$$b_1 = \left(\frac{s_Y}{s_{X_1}}\right)\frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{1 - r^2_{X_1X_2}} = \left(\frac{31.5}{29.8}\right)\frac{0.85 - (0.84)(0.73)}{1 - 0.73^2} = 0.546$$

$$b_2 = \left(\frac{s_Y}{s_{X_2}}\right)\frac{r_{YX_2} - r_{YX_1} r_{X_1X_2}}{1 - r^2_{X_1X_2}} = \left(\frac{31.5}{24.4}\right)\frac{0.84 - (0.85)(0.73)}{1 - 0.73^2} = 0.599$$

$$a = \overline{Y} - \left(b_1\overline{X}_1 + b_2\overline{X}_2\right) = 47.7 - \left[(0.546)(52.6) + (0.599)(41.9)\right]$$

$$= -6.065$$

so…

$$\hat{Y}_i = \text{-}6.065 + 0.546X_{1i} + 0.599X_{2i}$$

Multiple Regression Analysis

Example:

Compare the equations for the two bivariate models…

$$\hat{Y}_i = 0.284 + 0.902X_{1i}$$

$$\hat{Y}_i = 2.457 + 1.080X_{2i}$$

…to the prediction equation for the multivariate model:

$$\hat{Y}_i = -6.065 + 0.546X_{1i} + 0.599X_{2i}$$

The coefficient for X_1 is reduced by about 40% and the coefficient for X_2 is reduced by about 45%

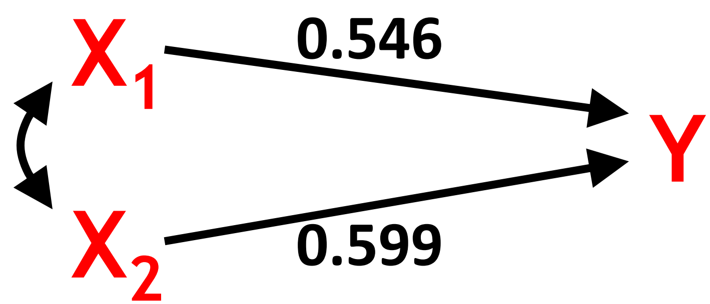
Multiple Regression Analysis

Example:

$$\hat{Y}_i = 0.284 + 0.902X_{1i}$$

$$\hat{Y}_i = 2.457 + 1.080X_{2i}$$

$$\hat{Y}_i = -6.065 + 0.546X_{1i} + 0.599X_{2i}$$



Interpreting Multiple Regression Coefficients

How are a, b_1, and b_2 interpreted in the equation:

$$\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i}$$

Intercept a:

The predicted value of Y when both X_1 and X_2 equal 0

Multiple regression coefficient (or slope) b_1:

The expected change in Y associated with a one unit increase in X_1, *controlling for* X_2

Multiple regression coefficient (or slope) b_2:

The expected change in Y associated with a one unit increase in X_2, *controlling for* X_1

Interpreting Multiple Regression Coefficients

Example: $\hat{Y}_i = -6.065 + 0.546X_{1i} + 0.599X_{2i}$

Intercept a:

When both occupational education (X_1) and occupational earnings (X_2) equal 0, we expect prestige (Y) to equal -6.065

Multiple regression coefficient (or slope) b_1:

Holding constant occupational earnings (X_2), a one unit increase in occupational education (X_1) is associated with a 0.546 increase in Y

Multiple regression coefficient (or slope) b_2:

Holding constant occupational education (X_1), a one unit increase in occupational earnings (X_2) is associated with a 0.599 increase in Y

Worksheet

Example: How is income affected by education and IQ?

Y = The adult income of 1,000 people (in $1,000s)

X_1 = The number of years of school they completed

X_2 = Their IQ

Descriptive Statistics

	Y	X_1	X_2	Mean	SD
Y	1.00			35.0	12.0
X_1	0.50	1.00		12.0	3.0
X_2	0.30	0.60	1.00	100.0	15.0

Compute and **interpret** the intercept and slopes of the multiple regression prediction equation

Coefficient of Determination

As in the bivariate case we can use R^2 to express the proportion of variation in Y that is accounted for by the predictor variables

Because, at worst, a predictor variable can explain none of the variation in Y, it follows that the addition of a second predictor variable to a bivariate regression model will either leave R^2 unchanged or increase it

Computationally, in the model with two predictors:

$$R^2_{Y \bullet X_1 X_2} = \frac{r^2_{YX_1} + r^2_{YX_2} - 2 r_{YX_1} r_{YX_2} r_{X_1 X_2}}{1 - r^2_{X_1 X_2}}$$

(What if X_1 and X_2 are uncorrelated?)

Coefficient of Determination

Example:

In two separate bivariate regression models of Y on X_1 and (separately) Y on X_2, we would see that

$$R^2_{Y \bullet X_1} = 0.85^2 = 0.72$$

$$R^2_{Y \bullet X_2} = 0.84^2 = 0.71$$

But in the multiple regression model

$$R^2_{Y \bullet X_1 X_2} = \frac{0.85^2 + 0.84^2 - 2(0.85)(0.84)(0.73)}{1 - 0.73^2} = 0.83$$

Testing Hypotheses about $\rho^2_{Y \bullet X1X2}$

Do the (in this case two) predictor variables collectively explain **any** of the variation in Y?

We use $R^2_{Y \bullet X1X2}$ to estimate $\rho^2_{Y \bullet X1X2}$

As before, another way to express $R^2_{Y \bullet X1X2}$ is:

$$R^2_{Y \bullet X_1 X_2} = \frac{SS_{REGRESSION}}{SS_{TOTAL}}$$

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

Testing Hypotheses about $\rho^2_{Y \bullet X1X2}$

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
6. Compare the test statistic to the critical value

Testing Hypotheses about $\rho^2_{Y \bullet X1X2}$

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

H_0: $\rho^2_{Y \bullet X1X2} = 0$

H_1: $\rho^2_{Y \bullet X1X2} > 0$

This is a one-sided test (with no <) because $\rho^2_{Y \bullet X1X2}$ cannot possibly be less than zero

Failing to reject the null means failing to reject the hypothesis that X_1 and X_2 (collectively) explain none of the variation in Y

Testing Hypotheses about $\rho^2_{Y \bullet X1X2}$

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 $\rho^2_{Y \bullet X1X2}$ to be valid

Testing Hypotheses about $\rho^2_{Y \bullet X1X2}$

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

Let's choose $\alpha=0.05$

Testing Hypotheses about $\rho^2_{Y \bullet X1X2}$

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 $\rho^2_{Y \bullet X1X2}$ is (as described below) an F test with $df_{NUM}=2$ (the number of predictors in the model) and $df_{DENOM}=N-3$ (N-1 minus the number of predictors in the model)

In our example, we want $F_{2,42}$ for $\alpha=0.05$ … so 3.23

We will thus reject H_0 if our F statistic exceeds 3.23

Critical Values of F
(α=0.05)

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	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95	248.01	250.10	251.14	251.77	253.04	253.68	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.45	19.46	19.47	19.48	19.49	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.62	8.59	8.58	8.55	8.54	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.75	5.72	5.70	5.66	5.65	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.50	4.46	4.44	4.41	4.39	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.81	3.77	3.75	3.71	3.69	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.38	3.34	3.32	3.27	3.25	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.08	3.04	3.02	2.97	2.95	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.86	2.83	2.80	2.76	2.73	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.70	2.66	2.64	2.59	2.56	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.57	2.53	2.51	2.46	2.43	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.47	2.43	2.40	2.35	2.32	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.38	2.34	2.31	2.26	2.23	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.31	2.27	2.24	2.19	2.16	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.25	2.20	2.18	2.12	2.10	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.19	2.15	2.12	2.07	2.04	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.15	2.10	2.08	2.02	1.99	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.11	2.06	2.04	1.98	1.95	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.07	2.03	2.00	1.94	1.91	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.04	1.99	1.97	1.91	1.88	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.18	2.10	2.01	1.96	1.94	1.88	1.84	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	1.98	1.94	1.91	1.85	1.82	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.13	2.05	1.96	1.91	1.88	1.82	1.79	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.94	1.89	1.86	1.80	1.77	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.09	2.01	1.92	1.87	1.84	1.78	1.75	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.90	1.85	1.82	1.76	1.73	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.06	1.97	1.88	1.84	1.81	1.74	1.71	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.87	1.82	1.79	1.73	1.69	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.03	1.94	1.85	1.81	1.77	1.71	1.67	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.84	1.79	1.76	1.70	1.66	1.62
31	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15	2.00	1.92	1.83	1.78	1.75	1.68	1.65	1.61
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	1.99	1.91	1.82	1.77	1.74	1.67	1.63	1.59
33	4.14	3.28	2.89	2.66	2.50	2.39	2.30	2.23	2.18	2.13	1.98	1.90	1.81	1.76	1.72	1.66	1.62	1.58
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	1.97	1.89	1.80	1.75	1.71	1.65	1.61	1.57
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	1.96	1.88	1.79	1.74	1.70	1.63	1.60	1.56
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.74	1.69	1.66	1.59	1.55	1.51
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.78	1.69	1.63	1.60	1.52	1.48	1.44
75	3.97	3.12	2.73	2.49	2.34	2.22	2.13	2.06	2.01	1.96	1.80	1.71	1.61	1.55	1.52	1.44	1.39	1.34

Testing Hypotheses about $\rho^2_{Y\bullet X1X2}$

Calculate the test statistic

The F statistic when there are two predictors is

$$F_{2,N-3} = \frac{SS_{REGRESSION}/2}{SS_{ERROR}/N-3} = \frac{MS_{REGRESSION}}{MS_{ERROR}}$$

Computationally:

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

$$SS_{REGRESSION} = (R^2_{Y\bullet X_1X_2})(SS_{TOTAL})$$

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

Testing Hypotheses about $\rho^2_{Y\bullet X1X2}$

Calculate the test statistic

In our example:

$$SS_{TOTAL} = (s_Y^2)(N-1) = (31.5^2)(45-1) = 43{,}659$$

$$SS_{REGRESSION} = (R^2_{Y\bullet X_1X_2})(SS_{TOTAL}) = (0.83)(43{,}659) = 36{,}236$$

$$SS_{ERROR} = SS_{TOTAL} - SS_{REGRESSION} = 43{,}659 - 36{,}236 = 7{,}423$$

so

$$F_{2,N-3} = \frac{SS_{REGRESSION}/2}{SS_{ERROR}/N-3} = \frac{36{,}236/2}{7{,}423/42} = 102.5$$

Testing Hypotheses about $\rho^2_{Y \bullet X1X2}$

Compare the test statistic to the critical value

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

We can restate the hypotheses:

- H_0: $\rho^2_{Y \bullet X1X2} = 0$ ➔ Fail to reject H_0 if $F \leq 3.23$
- H_1: $\rho^2_{Y \bullet X1X2} > 0$ ➔ Reject H_0 if $F > 3.23$

Since F=102.5, we reject H_0 … so it appears that in the population X_1 and X_2 (in combination) account for some of the variability in Y

Worksheet

Example: How is income affected by education and IQ?

Y = The adult income of 1,000 people (in $1,000s)

X_1 = The number of years of school they completed

X_2 = Their IQ

Descriptive Statistics

	Y	X_1	X_2	Mean	SD
Y	1.00			35.0	12.0
X_1	0.50	1.00		12.0	3.0
X_2	0.30	0.60	1.00	100.0	15.0

Test the hypothesis that $\rho^2_{Y \bullet X1X2} = 0$ … or, that X_1 and X_2 explain none of the variability in Y (Note: $R^2_{Y \bullet X1X2} = 0.25$); **use $\alpha = 0.05$**

Testing Hypotheses about β_1 & β_2

Can we conclude that β_1 and/or β_2 are different from 0?

We use b_1 and b_2 to estimate β_1 and β_2, respectively

In the bivariate model the variance of the sampling distribution of slope b was

$$s_b^2 = \frac{MS_{ERROR}}{\left(s_X^2\right)\left(N-1\right)}$$

In the model with two predictor variables the variances of the sampling distributions of b_1 and b_2 are

$$s_{b_1}^2 = \frac{MS_{ERROR}}{\left(s_{X_1}^2\right)\left(N-1\right)\left(1-R_{X_1 \bullet X_2}^2\right)} \qquad s_{b_2}^2 = \frac{MS_{ERROR}}{\left(s_{X_2}^2\right)\left(N-1\right)\left(1-R_{X_2 \bullet X_1}^2\right)}$$

Testing Hypotheses about β_1 & β_2

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 … t
6. Compare the test statistic to the critical value

Testing Hypotheses about β_1 & β_2

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

H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$

H_0: $\beta_2 = 0$
H_1: $\beta_2 \neq 0$

These are both two-sided tests

For each, failing to reject H_0 means failing to reject the hypothesis that there is no net association between Y and the corresponding X variable

Testing Hypotheses about β_1 & β_2

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_1 and β_2 to be valid

Testing Hypotheses about β_1 & β_2

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

Let's choose $\alpha=0.05$

Testing Hypotheses about β_1 & β_2

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 β_1 and β_2 are t tests with N-3 degrees of freedom (because MS_{ERROR} has N-3 degrees of freedom when there are two predictor variables)

In our example, we want t_{N-3} for α=0.05 which is close to 2.021 (because N-3 is 42 and thus close to 40)

For each hypothesis test we will thus reject H_0 if our t statistic exceeds 2.021 in absolute value

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%
	Confidence level C											

Testing Hypotheses about β_1 & β_2

Calculate the test statistic

The t statistic for β_1 is

$$t_{N-3} = \frac{b_1 - 0}{s_{b_1}} = \frac{b_1 - 0}{\sqrt{\frac{MS_{ERROR}}{\left(s_{X_1}^2\right)(N-1)\left(1-R_{X_1 \bullet X_2}^2\right)}}} = \frac{0.546}{0.099} = 5.52$$

The t statistic for β_2 is:

$$t_{N-3} = \frac{b_2 - 0}{s_{b_2}} = \frac{b_2 - 0}{\sqrt{\frac{MS_{ERROR}}{\left(s_{X_2}^2\right)(N-1)\left(1-R_{X_2 \bullet X_1}^2\right)}}} = \frac{0.599}{0.120} = 4.99$$

Testing Hypotheses about β_1 & β_2

Compare the test statistic to the critical value

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

We can restate the hypotheses:

$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
$H_1: \beta_1 \neq 0$	$H_1: \beta_2 \neq 0$

Since our values of t exceed our critical value t* (2.021) for both hypothesis tests, we reject the null hypothesis that β_1=0 and the null hypothesis that β_2=0

Worksheet

Example: How is income affected by education and IQ?

Y = The adult income of 1,000 people (in $1,000s)

X_1 = The number of years of school they completed

X_2 = Their IQ

Descriptive Statistics

	Y	X_1	X_2	Mean	SD
Y	1.00			35.0	12.0
X_1	0.50	1.00		12.0	3.0
X_2	0.30	0.60	1.00	100.0	15.0

Test the hypotheses that β_1 and β_2 equal zero (Note: Use b_1 and b_2 from above; MS_{error}=108.2 and $R^2_{Y \bullet X1X2}$ = 0.25); **α = 0.05**

Partial Correlation

Earlier we talked about the correlation coefficient, r, as a measure that describes the strength and direction of the association between two continuous variables

If r_{YX1} represents the bivariate correlation between Y and X_1, then $r_{YX1 \bullet X2}$ represents the partial correlation between Y and X_1 that persists after controlling for X_2

In the context of a regression model with two explanatory variables, the partial correlation between Y and X_1 is

$$r_{YX_1 \bullet X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{1 - r_{YX_2}^2}\sqrt{1 - r_{X_1X_2}^2}}$$

Partial Correlation

Example:

The bivariate correlation between Y and X_1 is 0.85

The partial correlation between Y and X_1 net of X_2 is

$$r_{YX_1 \bullet X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{1-r_{YX_2}^2}\sqrt{1-r_{X_1X_2}^2}} = \frac{(0.85)-(0.84)(0.73)}{\sqrt{1-.84^2}\sqrt{1-0.73^2}} = 0.64$$

The bivariate correlation between Y and X_2 is 0.84

The partial correlation between Y and X_2 net of X_1 is

$$r_{YX_2 \bullet X_1} = \frac{r_{YX_2} - r_{YX_1} r_{X_1X_2}}{\sqrt{1-r_{YX_1}^2}\sqrt{1-r_{X_1X_2}^2}} = \frac{(0.84)-(0.85)(0.73)}{\sqrt{1-.85^2}\sqrt{1-0.73^2}} = 0.61$$

Testing Hypotheses about $r_{YX1 \bullet X2}$

Hypotheses tests about partial correlation coefficients are identical to hypothesis tests for the corresponding regression coefficient

If we reject the hypothesis that β_1 equals zero in the population, we are simultaneously rejecting the null hypothesis that $\rho_{YX1 \bullet X2}$ equals zero

Likewise, if we reject the hypothesis that β_2 equals zero in the population, we are simultaneously rejecting the null hypothesis that $\rho_{YX2 \bullet X1}$ equals zero

Want More?

David Lane’s Books

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

Dallal’s Book (see “Simple Linear Regression” section)

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

(Look under “multiple linear regression”)

Biddle’s Book:

http://www.biddle.com/documents/bcg_comp_chapter4.pdf

Another good overview:

http://www.amstat.org/publications/jse/v16n3/datasets.kuiper.html