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?

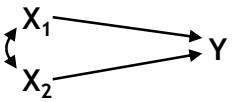
Multiple Regression

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)

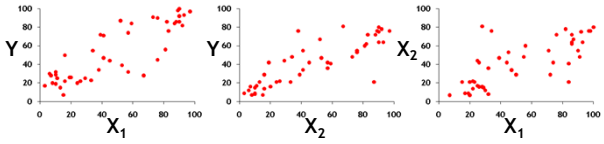
	Y	X_1	X_2	Mean	SD
Y	1.00			47.7	31.5
X_1	0.85	1.00		52.6	29.8
X_2	0.84	0.73	1.00	41.9	24.4

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

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

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

…but we know that neither slope (b_{YX_1} or b_{YX_2}) represents the "effects" of X_1 or X_2 because of spuriousness 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:

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

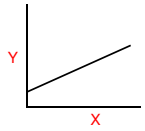
The sample prediction equation:

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

Multiple Regression Analysis

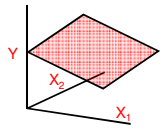
The bivariate regression prediction equation describes a 2-dimensional line

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



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

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

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

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

Multiple Regression Analysis

Example:

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

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

$$a = \bar{Y} - (b_1\bar{X}_1 + b_2\bar{X}_2) = 47.7 - [(0.546)(52.6) + (0.599)(41.9)]$$
$$= -6.065$$

so…

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

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%

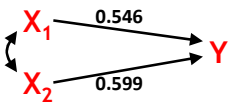
Multiple Regression Analysis

Example:

$$\hat{Y}_i = 0.284 + 0.902X_{1i}$$ $X_1 \xrightarrow{0.902} Y$

$$\hat{Y}_i = 2.457 + 1.080X_{2i}$$ $X_2 \xrightarrow{1.080} Y$

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



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

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:

$$R^2_{Y\bullet X_1X_2} = \frac{r^2_{YX_1} + r^2_{YX_2} - 2r_{YX_1}r_{YX_2}r_{X_1X_2}}{1 - r^2_{X_1X_2}}$$

(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

$$R^2_{Y\bullet X_1X_2} = \frac{0.85^2 + 0.84^2 - 2(0.85)(0.84)(0.73)}{1 - 0.73^2} = 0.83$$

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:

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

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 α=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 α=0.05 which is close to 3.32

We will thus reject H_0 if our F statistic exceeds 3.32

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:

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

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

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

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

Calculate the test statistic

In our example:

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

$$SS_{REGRESSION} = (R^2_{Y \bullet X_1 X_2})(SS_{TOTAL}) = (0.83)(43{,}659) = 36{,}236$$

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

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 $F \leq 3.32$
- H_1: $\rho^2_{Y \bullet X1X2} > 0$ ➔ Reject H_0 if $F > 3.32$

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

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

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

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

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_0: $\beta_2 = 0$
H_1: $\beta_1 \neq 0$	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 α=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

Testing Hypotheses about β_1 & β_2

Calculate the test statistic

The t statistic for β_1 is

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

The t statistic for β_2 is:

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

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

$$r_{YX_1 \bullet X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1 X_2}}{\sqrt{1 - r_{YX_2}^2}\sqrt{1 - r_{X_1 X_2}^2}}$$

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

$$r_{YX_1 \bullet X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1 X_2}}{\sqrt{1 - r_{YX_2}^2}\sqrt{1 - r_{X_1 X_2}^2}} = \frac{(0.85) - (0.84)(0.73)}{\sqrt{1 - .84^2}\sqrt{1 - 0.73^2}} = 0.64$$

The bivariate correlation between Y and X_2 is 0.84

The partial correlation between Y and X_2 net of X_1 is

$$r_{YX_2 \bullet X_1} = \frac{r_{YX_2} - r_{YX_1} r_{X_1 X_2}}{\sqrt{1 - r_{YX_1}^2}\sqrt{1 - r_{X_1 X_2}^2}} = \frac{(0.84) - (0.85)(0.73)}{\sqrt{1 - .85^2}\sqrt{1 - 0.73^2}} = 0.61$$

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

BREAK

Multiple Regression with *k* Independent Variables

We've seen how to estimate regression models that include two continuous predictor variables (X_1 and X_2) and a continuous response variable (Y)

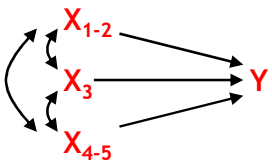
Extension #1: Models with *k* continuous predictor variables (X_1 through X_k) and a continuous response variable (Today)

Extension #2: Models with *k* predictor variables—some of which are continuous and some of which are discrete—and a continuous response variable (Bonus Material)

Extension #3: Models with *k* predictor variables and a discrete response variable (SOC 5811 and 8811)

Multiple Regression with *k* Independent Variables

Example: How do state educational resources ($X_{1\text{-}2}$), state education policies (X_3), and state economic conditions ($X_{4\text{-}5}$) influence states' high school graduation rates (Y)?



Multiple Regression with *k* Independent Variables

Example: How do state educational resources ($X_{1\text{-}2}$), state education policies (X_3), and state economic conditions ($X_{4\text{-}5}$) influence states' high school graduation rates (Y)?

(always start by looking at descriptive statistics)

Descriptive Statistics

	N	Mean	Variance
Y = Estimated Completion Rate, 2002	51	72.704	63.446
X1 = Per Pupil Expenditures, 2002	51	67.684	194.999
X2 = Pupil-Teacher Ratios, 2002	51	12.633	11.061
X3 = Carn. Units Required for Graduation, 2002	51	17.980	51.780
X4 = Poverty Rate, 2002	51	11.369	11.667
X5 = Unemployment Rate, 2002	51	4.522	.809
Valid N (listwise)	51		

D.C. is a state!

Multiple Regression with *k* Independent Variables

Example: How do state educational resources ($X_{1\text{-}2}$), state education policies (X_3), and state economic conditions ($X_{4\text{-}5}$) influence states' high school graduation rates (Y)?

(always start by looking at descriptive statistics)

Correlations

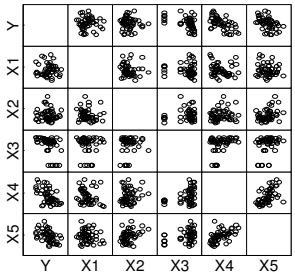
	Y	X1	X2	X3	X4	X5
Estimated Completion Rate, 2002	1.000					
Per Pupil Expenditures, 2002	-.004	1.000				
Pupil-Teacher Ratios, 2002	-.052	-.116	1.000			
Carn. Units Req. for Grad., 2002	-.436	-.057	.125	1.000		
Poverty Rate, 2002	-.574	-.196	.155	.327	1.000	
Unemployment Rate, 2002	-.475	-.121	.359	.289	.549	1.000

Multiple Regression with *k* Independent Variables

Example: How do state educational resources ($X_{1\text{-}2}$), state education policies (X_3), and state economic conditions ($X_{4\text{-}5}$) influence states' high school graduation rates (Y)?

Bivariate Scatterplots

(always start by looking at bivariate plots)



Multiple Regression with *k* Independent Variables

The population regression equation:

$$Y_i = \alpha + \sum_{j=1}^{k} \beta_j X_{ji} + \varepsilon_i$$

The population prediction equation:

$$\hat{Y}_i = \alpha + \sum_{j=1}^{k} \beta_j X_{ji}$$

The sample regression equation:

$$Y_i = a + \sum_{j=1}^{k} b_j X_{ji} + e_i$$

The sample prediction equation:

$$\hat{Y}_i = a + \sum_{j=1}^{k} b_j X_{ji}$$

Prediction Equations

Model with ONE Independent Variable

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

Model with TWO Independent Variables

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

Model with *k* Independent Variables

$$\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + \ldots + b_k X_{ki}$$

$$\hat{Y}_i = a + \sum_{j=1}^{k} b_j X_{ji}$$

Multiple Regression with *k* Independent Variables

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

The computational formulas for a and b_1 through b_k are complex when k>2

The computational formulas for a and b_1 through b_k are based on the correlation of the response and predictor variables and on their means and variances

Interpreting Multiple Regression Coefficients

How are a and b_k interpreted in the equation:

$$\hat{Y}_i = a + \sum_{j=1}^{k} b_{kj} X_{ji}$$

Intercept *a*:

The intercept, *a*, equals the predicted value of Y when each of the k predictor variables (X_1 through X_k) equal 0

Multiple regression coefficient (or slope) b_k:

Multiple regression coefficient b_k represents the expected change in Y associated with a one unit increase in X_k, controlling for all other predictors in the model

Interpreting Multiple Regression Coefficients

Example: How do state educational resources ($X_{1\text{-}2}$), state education policies (X_3), and state economic conditions ($X_{4\text{-}5}$) influence states' high school graduation rates (Y)?

$$\hat{Y}_i = 98.728 - 0.064X_{1i} + 0.274X_{2i} - 0.285X_{3i} - 0.933X_{4i} - 2.086X_{5i}$$

$$\hat{Y}_i = 98.728 - 0.064(\text{Pupil - Teacher Ratio}_i) + 0.274(\text{Per - Pupil Expenditure}_i) - 0.285(\text{Carnegie Units}_i) - 0.933(\text{Poverty Rate}_i) - 2.086(\text{Unemployment Rate}_i)$$

Coefficient of Determination

As before, 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 more predictor variables to the model will either leave R^2 unchanged or increase it

Adjusted R^2(R^2_{adj})

A critique of the traditional R^2 measure is that it may increase with the addition of more predictor variables simply because of chance (random) covariation between Y and the additional predictors

The **adjusted coefficient of determination** (R^2_{adj}) takes into account the number of independent variables relative to the number of observations … essentially rewarding parsimony in model specification

Computationally:

$$R^2_{adj} = R^2_{Y \bullet X_1 \ldots X_k} - \left(\frac{(k)\left(1 - R^2_{Y \bullet X_1 \ldots X_k}\right)}{N - k - 1} \right)$$

Worksheet

Below are the means and standard deviations of scores on the second exam ("Exam"), the total points earned on problem sets ("ProblemSets"), in-class worksheet scores ("InClass"), and lab worksheet scores ("InLab"); the latter three are measured at the time the 2nd exam was taken

Variable	Obs	Mean	Std. Dev.	Min	Max
Exam	155	78.96129	11.871	0	90
ProblemSets	155	166.7452	37.3687	0	200
InClass	155	69.09677	11.69125	18	78
InLab	155	35.84516	6.896397	0	40

Worksheet

Below are the correlations between these variables

	Exam	Proble~s	InClass	InLab
Exam	1.0000			
ProblemSets	0.5028	1.0000		
InClass	0.2897	0.6574	1.0000	
InLab	0.3308	0.6246	0.5801	1.0000

Worksheet

Here are the results of a regression of exam score on the other variables. Interpret the intercept, the slopes, & R^2

```
Number of obs =     155

R-squared     =  0.2572
Adj R-squared =  0.2425
```

Exam	Coef.
ProblemSets	.1676782
InClass	-.0886353
InLab	.0891642
_cons	53.93007

Testing Hypotheses about $\rho^2_{Y \bullet X1 \ldots Xk}$

Do the k predictor variables collectively explain any of the variation in Y?

We use $R^2_{Y \bullet X1 \ldots Xk}$ to estimate $\rho^2_{Y \bullet X1 \ldots Xk}$

As before, another way to express $R^2_{Y \bullet X1 \ldots Xk}$ is:

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

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

Testing Hypotheses about $\rho^2_{Y \bullet X1 \ldots Xk}$

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 $\rho^2_{Y \bullet X1 \ldots Xk}$

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

H_0: $\rho^2_{Y \bullet X1 \ldots Xk} = 0$
H_1: $\rho^2_{Y \bullet X1 \ldots Xk} > 0$

This is a one-sided test (with no <) because $\rho^2_{Y \bullet X1 \ldots Xk}$ cannot possibly be less than zero

Failing to reject the null means failing to reject the hypothesis that the k predictor variables collectively explain none of the variation in Y

Testing Hypotheses about $\rho^2_{Y \bullet X1 \ldots Xk}$

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 $\rho^2_{Y \bullet X1 \ldots Xk}$ to be valid

Testing Hypotheses about $\rho^2_{Y \bullet X1 \ldots Xk}$

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

Let's choose α=0.05

Testing Hypotheses about $\rho^2_{Y \bullet X1 \ldots Xk}$

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 $\rho^2_{Y \bullet X1 \ldots Xk}$ is (as described below) an F test with $df_{NUM}=k$ and $df_{DENOM}=N—k—1$

In our example, we want $F_{5,45}$ for $\alpha=0.05$ which is close to 2.45

We will thus reject H_0 if our F statistic exceeds 2.45

Testing Hypotheses about $\rho^2_{Y \bullet X1 \ldots Xk}$

Calculate the test statistic

The F statistic when there are k predictors is

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

Computationally:

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

$$SS_{REGRESSION} = (R^2_{Y \bullet X_1 \ldots X_k})(SS_{TOTAL})$$

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

Testing Hypotheses about $\rho^2_{Y \bullet X1 \ldots Xk}$

Calculate the test statistic

In our example:

$$SS_{TOTAL} = (s^2_Y)(N-1) = (63.446)(51-1) = 3,172.3$$

$$SS_{REGRESSION} = (R^2_{Y \bullet X_1 \ldots X_k})(SS_{TOTAL}) = (0.448)(3,172.3) = 1,421.19$$

$$SS_{ERROR} = SS_{TOTAL} - SS_{REGRESSION} = 3,172.3 - 1,421.19 = 1,751.11$$

so

$$F_{5,45} = \frac{SS_{REGRESSION}/5}{SS_{ERROR}/45} = \frac{1,421.19/5}{1,751.11/45} = 7.30$$

Testing Hypotheses about $\rho^2_{Y \bullet X1 \ldots Xk}$

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: $\rho^2_{Y \bullet X1 \ldots Xk} = 0$ ➔ Fail to reject H_0 if $F \leq 2.45$
- H_1: $\rho^2_{Y \bullet X1 \ldots Xk} > 0$ ➔ Reject H_0 if $F > 2.45$

Since F=7.30, we reject H_0 … so it appears that in the population the k predictors (in combination) account for some of the variability in Y

Worksheet

Here are the results of a regression of exam score on the other variables. Test the hypothesis that $\rho^2=0$; use $\alpha=0.05$

```
Number of obs =     155

R-squared     =  0.2572
Adj R-squared =  0.2425
```

Exam	Coef.
ProblemSets	.1676782
InClass	-.0886353
InLab	.0891642
_cons	53.93007

Variable	Obs	Mean	Std. Dev.
Exam	155	78.96129	11.871
ProblemSets	155	166.7452	37.3687
InClass	155	69.09677	11.69125
InLab	155	35.84516	6.896397

Testing Hypotheses about β_k

Can we conclude that β_k is different from 0?

We use b_k to estimate β_k

In the model with k=2 predictors the variance of the sampling distribution of slopes b_1 and b_2 were

$$s^2_{b_1} = \frac{MS_{ERROR}}{(s^2_{X_1})(N-1)(1-R^2_{X_1 \bullet X_2})} \qquad s^2_{b_2} = \frac{MS_{ERROR}}{(s^2_{X_2})(N-1)(1-R^2_{X_2 \bullet X_1})}$$

In the model with k predictor variables the variances of the sampling distributions of b_k is

$$s^2_{b_k} = \frac{MS_{ERROR}}{(s^2_{X_k})(N-1)(1-R^2_{X_k \bullet X_1, \ldots X_{k-1}})}$$

Testing Hypotheses about β_k

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_k

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

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

This is a two-sided tests (although it needn't be)

Failing to reject H_0 means failing to reject the hypothesis that there is no net association between Y and X_k

Testing Hypotheses about β_k

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_k to be valid

Testing Hypotheses about β_k

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

Let's choose α=0.05

Testing Hypotheses about β_k

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_k is a t tests with N-k-1 degrees of freedom (because MS_{ERROR} has N-k-1 degrees of freedom)

In our example, we want t_{45} for α=0.05 which is close to 2.021 (because N-k-1 is 45 and thus close to 40)

For hypothesis tests about β_k we will thus reject H_0 if our t statistic exceeds 2.021 in absolute value

Testing Hypotheses about β_k

Calculate the test statistic

The t statistic for β_k is

$$t_{N-k-1} = \frac{b_k - 0}{s_{b_1}} = \frac{b_k - 0}{\sqrt{\frac{MS_{ERROR}}{(s^2_{X_k})(N-1)(1-R^2_{X_k \bullet X_1 \ldots X_{k-1}})}}}$$

Testing Hypotheses about β_k

Calculate the test statistic

In our example…

Coefficients[a]

	Unstandardized Coefficients		Std. Coef.		
	B	Std. Error	Beta	t	Sig.
(Constant)	98.73	7.21		13.7	.00
Per Pupil Expenditures, 2002	-.06	.06	-.11	-.99	.33
Pupil-Teacher Ratios, 2002	.27	.29	.11	.96	.34
Carn. Units Req. for Grad., 2002	-.29	.13	-.26	-2.17	.03
Poverty Rate, 2002	-.93	.32	-.40	-2.91	.01
Unemployment Rate, 2002	-2.09	1.25	-.24	-1.66	.10

a. Dependent Variable: Estimated Completion Rate, 2002

Testing Hypotheses about β_k

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: $\beta_k = 0$
- H_1: $\beta_k \neq 0$

Our values of t exceed our critical value t* (2.021) for two of the five hypothesis tests about b_k, so we reject the null hypothesis in those instances

Worksheet

Here are the results of the regression of exam score on the other variables. Test the hypothesis that $\beta_{ProblemSets}=0$; use $\alpha=0.05$

Source	SS	df	MS
Model	5582.60155	3	1860.86718
Residual	16119.1662	151	106.749445
Total	21701.7677	154	140.92057

```
Number of obs =     155
F(  3,   151) =   17.43
Prob > F      =  0.0000
R-squared     =  0.2572
Adj R-squared =  0.2425
Root MSE      =  10.332
```

Exam	Coef.	Std. Err.
ProblemSets	.1676782	.0322029
InClass	-.0886353	.098682
InLab	.0891642	.1614245
_cons	53.93007	5.308435

BONUS TOPIC #1

Comparing Nested Equations

We can use hypothesis tests about specific slopes, b_k, to assess whether particular X variables add to the predictive power of the regression model

If we reject H_0 for β_k, we are concluding that X_k is significantly associated with Y net of the other covariates in the model … and thus that our ability to predict Y is improved by including X_k in the model

Sometimes, however, we are interested in assessing the contribution of theoretically derived groups of X variables to our ability to predict Y

Comparing Nested Equations

In the example we've been considering, there are three groups of predictors of states' graduation rates (Y):

- state educational resources (X_1 and X_2)
- state education policies (X_3)
- state economic conditions (X_4 and X_5)

For example, do state educational resources significantly add to the predictive power of the model (as compared to a model that does not include these predictors)?

Comparing Nested Equations

The nested equations below allow us to making useful statements about the contributions of groups of variables to our ability to predict state graduation rates

	Model 1	Model 2	Model 3	Model 4
Pup.-T Ratio	-0.01	—	—	-0.06
P.P. Exp.	-0.13	—	—	0.27
Carn. Units	—	-0.48**	—	-0.29*
Poverty	—	—	-1.04**	-0.93**
Unempl.	—	—	-2.03	-2.09
Constant	74.68**	81.38**	93.75**	98.73**
R^2	0.003	0.190	0.366	0.448
F (df_1,df_2)	0.066 (2,48)	11.478**(1,49)	13.831**(2,48)	7.301**(5,45)

* = p < 0.05 ; ** = p < 0.01

Comparing Nested Equations

In this example we can say that Models 1, 2, and 3 are nested within Model 4 … they contain a subset of the predictors included in Model 4

	Model 1	Model 2	Model 3	Model 4
Pup.-T Ratio	-0.01	—	—	-0.06
P.P. Exp.	-0.13	—	—	0.27
Carn. Units	—	-0.48**	—	-0.29*
Poverty	—	—	-1.04**	-0.93**
Unempl.	—	—	-2.03	-2.09
Constant	74.68**	81.38**	93.75**	98.73**
R^2	0.003	0.190	0.366	0.448
F (df_1,df_2)	0.066 (2,48)	11.478**(1,49)	13.831**(2,48)	7.301**(5,45)

* = p < 0.05 ; ** = p < 0.01

Comparing Nested Equations

The F tests for each individual model indicate whether the X variables in that model improve our ability to predict Y relative to the null model (with no X variables)

	Model 1	Model 2	Model 3	Model 4
Pup.-T Ratio	-0.01	—	—	-0.06
P.P. Exp.	-0.13	—	—	0.27
Carn. Units	—	-0.48**	—	-0.29*
Poverty	—	—	-1.04**	-0.93**
Unempl.	—	—	-2.03	-2.09
Constant	74.68**	81.38**	93.75**	98.73**
R^2	0.003	0.190	0.366	0.448
F (df_1,df_2)	0.066 (2,48)	11.478**(1,49)	13.831**(2,48)	7.301**(5,45)

* = p < 0.05 ; ** = p < 0.01

Comparing Nested Equations

A different question is whether a particular subset of X variables adds to our ability to predict Y relative to a model that contains a different subset of X variables

	Model 1	Model 2	Model 3	Model 4
Pup.-T Ratio	-0.01	—	—	-0.06
P.P. Exp.	-0.13	—	—	0.27
Carn. Units	—	-0.48**	—	-0.29*
Poverty	—	—	-1.04**	-0.93**
Unempl.	—	—	-2.03	-2.09
Constant	74.68**	81.38**	93.75**	98.73**
R^2	0.003	0.190	0.366	0.448
F (df_1,df_2)	0.066 (2,48)	11.478**(1,49)	13.831**(2,48)	7.301**(5,45)

* = p < 0.05 ; ** = p < 0.01

Comparing Nested Equations

Call the model that contains the full set of X variables the complete model; it has k_2 impendent variables

Call a model that contains a subset of those X variables the reduced model; it contains k_1 independent variables

Question: Does the addition of the k_2-k_1 new predictor variables in the complete model improve our ability to predict Y (relative to the reduced model)?

(If k_2-k_1 equals one, then we are just adding one new X variable and we can answer this question with a hypothesis test about the coefficient for that variable)

Comparing Nested Equations

In general, when the reduced model is nested within the complete model we test the hypothesis that the new additional variables in the complete model add to the predictive power of the null model

We make this comparison using an F statistic that is based on the change in R^2 between the reduced model (R^2_1) and the complete model (R^2_2)

$$F_{(k_2-k_1),(N-k_2-1)} = \frac{(R_2^2 - R_1^2)/(k_2 - k_1)}{(1-R_2^2)/(N-k_2-1)}$$

Comparing Nested Equations

In our example: Does the addition of the state educational resource variables and the state education policies variables (in Model 4) add to the predictive power of Model 3?

If we use α=0.05, the critical value F* has $df_{NUM}=k_2-k_1=$ 5-2=3 and $df_{DENOM}=N-k_2-1=51-5-1=45$, so F*=2.84

The F statistic equals $F_{3,45}=\frac{(0.448-0.366)/(5-2)}{(1-0.448)/(51-5-1)}=2.228$

We fail to reject H_0 … the addition of these 3 variables does not add to the predictive power of Model 3

Comparing Nested Equations

We thus conclude that Model 4 fits no better than Model 3, even though we observe a statistically significant coefficient (for "Carnegie Units") in Models 2 and 4

	Model 1	Model 2	Model 3	Model 4
Pup.-T Ratio	-0.01	—	—	-0.06
P.P. Exp.	-0.13	—	—	0.27
Carn. Units	—	-0.48**	—	-0.29*
Poverty	—	—	-1.04**	-0.93**
Unempl.	—	—	-2.03	-2.09
Constant	74.68**	81.38**	93.75**	98.73**
R^2	0.003	0.190	0.366	0.448
F (df_1,df_2)	0.066 (2,48)	11.478**(1,49)	13.831**(2,48)	7.301**(5,45)

* = p < 0.05 ; ** = p < 0.01

BONUS TOPIC #2

Regression vs. ANOVA

The regression techniques we have used thus far require continuous predictor variables

It would be wrong—technically and conceptually—to simply enter nominal or ordinal variables as predictor variables, since it is wrong to compute means and standard deviations for these variables

For a discrete variable X that has j categories, we can construct j dummy variables—each of which has possible values 0 and 1—and each of which indicates whether an individual falls into a particular category of X

Regression vs. ANOVA

For example, X might indicate father's education, which is a discrete measure of whether fathers (a) did not finish high school, (b) finished high school but went no further, or (c) completed at least some college

From this we can construct j=3 three dummy variables:

	X_1	X_2	X_3
Father: < H.S.	1	0	0
Father: = H.S.	0	1	0
Father: > H.S.	0	0	1

Notice that knowing the value of j—1 of the X_j values allows you to infer the value of the j^{th} X value

Regression vs. ANOVA

If we then regress child's education (Y) on j—1 of these dummy variables we observe:

$$\hat{Y}_i = 11.776 + 1.697X_2 + 3.098X_3$$

Compare these results to the means of Y by level of father's education:

Mean of Child's Education (Y)

Father: < H.S.	11.776	$\hat{Y}_i = 11.776 + 1.697(0) + 3.098(0)$
Father: = H.S.	13.473	$\hat{Y}_i = 11.776 + 1.697(1) + 3.098(0)$
Father: > H.S.	14.871	$\hat{Y}_i = 11.776 + 1.697(0) + 3.098(1)$

Regression vs. ANOVA

If we then regress child's education (Y) on j—1 of these dummy variables we observe: $\hat{Y}_i = 11.776 + 1.697X_2 + 3.098X_3$

This regression model is exactly equivalent to an ANOVA in which we investigate the association between discrete variable X and continuous variable Y

The F statistic for this regression model is identical to the F statistic for the ANOVA relating Y to X

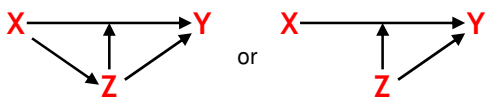
However, the regression framework gives us the ability to control for additional independent variables ... Doing so is known as ANalysis of COVAriance (ANCOVA)

BONUS TOPIC #3

Interaction Effects

In some cases the association between X and Y may be different across levels of Z (or "conditioned by" Z)

In these cases we say that there is an interaction between X and Z … Z is known as a moderating variable

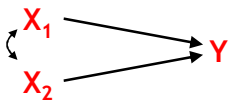


Interaction Effects

How do we model this in the regression context?

Imagine that we regress a continuous measure of education (Y) on a continuous measure of father's education (X_1) and a dummy variable for gender (X_2) that equals 0 for women and 1 for men

That model looks like this:

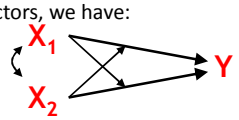


Interaction Effects

To allow for an "interaction effect" — such that the effect of X_1 on Y varies across levels of X_2 and the effect of X_2 on Y varies across levels of X_1 — we add an interaction term

An interaction term between X_1 and X_2 is simply a new variable creating by multiplying X_1 by X_2

If we add the interaction term to the model that includes X_1 and X_2 as predictors, we have:



Interaction Effects

The prediction equation for this new model is

$$\hat{Y}_i = a + b_1 \text{Father's Educ.}_i + b_2 \text{Male}_i + b_3 \text{Inter.}_i$$

where "Inter" is the variable that equals $X_1 \times X_2$

If we estimate this model we get

$$\hat{Y}_i = 9.37 + (0.34)\text{Father's Educ.}_i + (0.05)\text{Male}_i + (0.02)\text{Inter.}_i$$

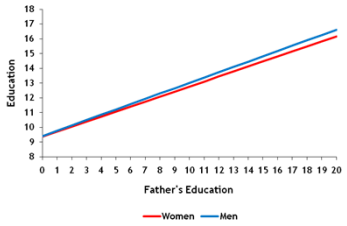
What is the "effect" of a one unit increase in father's education? What is the "effect" of being male?

What is the predicted value of Y for a man whose father completed 10 years of school? What about for a woman whose father completed 10 years of school?

Interaction Effects

It is almost always easiest to think about interaction effects if you make a graph of predicted values

Below are predicted values of education (Y) by father's education (X_1) and gender (X_2)



Interaction Effects

How do we know whether the interaction term improves the power of the model to predict Y? We may theorize that an interaction effect exists … but how do we test the hypothesis that this is true in the data?

Option #1: Look at the statistical significance of the coefficient for the interaction term

Option #2: Treat the model w/o the interaction term as a reduced model that is nested within a full model that does include the interaction term … conduct an F test .

Interaction Effects

How do we know whether the interaction term improves the power of the model to predict Y? We may theorize that an interaction effect exists … but how do we test the hypothesis that this is true in the data?

Option #1: In our example, the test statistic *t* for the interaction term is 2.48 … we would reject H_0 at α=0.05

Option #2: In our example, the test statistic $F_{1,27356}$ for the improvement in fit of the full model relative to the reduced model is 6.15 … we would reject H_0 at α=0.05

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