


Multiple Linear Regression: John McGready




Relating a Continuous Outcome to More than One Predictor: Multiple Linear Regression

John McGready
Johns Hopkins University

Lecture Topics
<ul style="list-style-type: none"> Why multiple linear regression? Interpreting coefficients from multiple linear regression Statistical inference on multiple regression coefficients ANOVA as a linear regression model Statistical interaction and linear regression

2



Section A

Why (Multiple) Linear Regression?

Multiple Linear Regression
<ul style="list-style-type: none"> Recall the arm circumference examples from lecture 4 At the end of the analysis, we found a positive statistically significant association between arm circumference and height <ul style="list-style-type: none"> With SLR, we estimated the magnitude of this association (slope) and put 95% confidence limits on it We also estimated that child height explained 46% of the variability in arm circumference At the end of the analysis, we found a negative, but non-statistically significant association between arm circumference and female sex

4

Questions
<ul style="list-style-type: none"> Let's start by expanding on the SLR model that included only height as a predictor <pre> regress armcirc height Source SS df MS Number of obs = 150 ----- ----- Model 148.874597 1 148.874597 F(1, 148) = 124.30 Residual 177.263335 148 1.19772523 Prob > F = 0.0000 ----- ----- Total 326.137932 149 2.18884518 R-squared = 0.4565 Adj R-squared = 0.4528 Root MSE = 1.0944 -----+----- armcirc Coef. Std. Err. t P> t [95% Conf. Interval] -----+----- height -1579469 .0141671 11.15 0.000 .1299511 -1859428 2.695906 .8774225 3.07 0.003 .9620116 4.4298 </pre>

5

Questions
<ul style="list-style-type: none"> Is it possible that other factors, besides child height, explain variability in arm circumference as well (for example, child weight) <ul style="list-style-type: none"> Can any of these other factors explain some of the variability not explained by height? Is it possible that the relationship between arm circumference and height is different after accounting for weight or sex differences by height

6

Multiple Linear Regression: John McGready

Arm Circumference and Weight

- SLR model that includes only weight (kg) as a predictor

```

. regress armcirc weight
-----+-----
Source |      SS      df       MS              Number of obs = 150
-----+-----
Model | 228.648565      1 228.648565          F( 1, 148) = 347.11
Residual | 97.4893667    148  .65871937          Prob > F      = 0.0000
-----+-----
Total | 326.137932    149  2.18884518          R-squared     = 0.7011
                                           Adj R-squared = 0.6991
                                           Root MSE    = .81161

-----+-----
armcirc |      Coef.   Std. Err.      t    P>|t|   [95% Conf. Interval]
-----+-----
weight |   .8013117   .0430096    18.63  0.000   .7163195   .8863038
 _cons |   7.792012  .2574701    30.26  0.000   7.28322   8.300805
    
```

- Relationship is positive, statistically significant
- The R^2 is .70 indicating that weight explains an estimated 70% of the variability in arm circumference

7

Questions

- How can we estimate the relationship between arm circumference and both weight and height in the same analysis?
- Answer: multiple linear regression
- We can estimate an equation of the following type:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$
 - Where x_1 is height (cm) and x_2 is weight (kg)

8

Arm Circumference and Weight

- MLR model that includes both weight and height as a predictor

```

. regress armcirc height weight
-----+-----
Source |      SS      df       MS              Number of obs = 150
-----+-----
Model | 250.505703      2 125.252851          F( 2, 147) = 243.44
Residual | 75.632229     147  .514504959          Prob > F      = 0.0000
-----+-----
Total | 326.137932    149  2.18884518          R-squared     = 0.7681
                                           Adj R-squared = 0.7649
                                           Root MSE    = .71729

-----+-----
armcirc |      Coef.   Std. Err.      t    P>|t|   [95% Conf. Interval]
-----+-----
height |  -1.589225   .2438288    -6.52  0.000  -2.071086  -1.107364
weight |   1.402873   .0998159    14.05  0.000   1.205614   1.600133
 _cons |  14.10371   .9947526    14.18  0.000   12.13785   16.06958
    
```

9

Results

- This model is given by:

$$\hat{y} = 14.1 + -0.16 \times height + 1.40 \times weight$$
- Together, height and weight explain an estimated 76% of the variability in arm circumference
- Both height and weight are statistically significant predictors of arm circumference in this model
- How to interpret in the estimated relationship between arm circumference and the height ($\hat{\beta}_1 = -0.16$) and weight ($\hat{\beta}_2 = 1.40$) predictors in an MLR model that includes both?

10

Slope of Height in MLR of Arm Circ on Height, Weight

- The slope estimate for height is $\hat{\beta}_1 = -0.16$
 - Still there is an estimated mean difference in arm circumference between two groups who differ by one cm in height, but are of the same weight
 - This is called the weight-adjusted association between arm circumference and height
 - This result estimates that groups of children who differ by 1 cm in height but are the same weight will differ in arm circumference on average by -0.16 cm, taller to shorter
 - If we were to compare two groups of children of the same weight, but who differ by 1 cm in height, the taller group would have arm circumference reduced by 0.16 on average
- The 95% for the true height-adjusted arm circumference and weight association is -0.21 to -0.11

11

Recapping the Results

- So far we have results from three models: arm circumference on height, arm circumference on weight, and arm circumference on height and weight

	Unadjusted	Unadjusted	Adjusted
	SLR1: x = height	SLR2: x = weight	MLR: x ₁ = height, x ₂ = weight
Height	0.16 (0.13-0.19)		-0.16 (-0.21, -0.11)
Weight		0.80 (0.72 - 0.89)	1.40 (1.20, 1.40)
R ²	0.46	0.70	0.76

12

Multiple Linear Regression: John McGready

Arm Circumference and Weight

- MLR model that includes weight, height, and child sex as predictors (sex = 1 for females)

```

. regress armcirc height weight sex
-----+-----
Source |      SS      df       MS              Number of obs =   150
-----+-----
Model | 253.753079      3   84.5843596          F( 3, 146) = 170.61
Residual | 72.3848531    146   .495786665          Prob > F      = 0.0000
-----+-----
Total | 326.137932    149   2.18884518          R-squared     = 0.7781
                                          Adj R-squared = 0.7735
                                          Root MSE    = .70412
-----+-----
armcirc |      Coef.   Std. Err.      t    P>|t|   [95% Conf. Interval]
-----+-----
height | -1.711503   .0244074    -7.01  0.000   -2.193876   -1.229129
weight |  1.459699   .1004677    14.53  0.000    1.26114    1.658258
sex |    .3018584 .1179463     2.56  0.012    .0687557   .5349611
 _cons |  14.37543   .9822445    14.64  0.000   12.43418   16.31668
    
```

13

Arm Circumference and Weight

- Interesting: the unadjusted (crude) relationship between arm circumference and sex showed a negative, but non-statistically significant association between arm circumference and female sex
- After adjusting for both height and weight, the association between arm circumference and being female is positive, and statistically significant
 - Female children have arm circumferences of 0.30 cm greater on average when compared to male children of the same height and weight (95% CI: 0.07 cm to 0.53 cm)

14

Arm Circumference and Weight

- MLR model that includes weight, height, and child sex as predictors (sex = 1 for females)

```

. regress armcirc height weight sex
-----+-----
Source |      SS      df       MS              Number of obs =   150
-----+-----
Model | 253.753079      3   84.5843596          F( 3, 146) = 170.61
Residual | 72.3848531    146   .495786665          Prob > F      = 0.0000
-----+-----
Total | 326.137932    149   2.18884518          R-squared     = 0.7781
                                          Adj R-squared = 0.7735
                                          Root MSE    = .70412
-----+-----
armcirc |      Coef.   Std. Err.      t    P>|t|   [95% Conf. Interval]
-----+-----
height | -1.711503   .0244074    -7.01  0.000   -2.193876   -1.229129
weight |  1.459699   .1004677    14.53  0.000    1.26114    1.658258
sex |    .3018584 .1179463     2.56  0.012    .0687557   .5349611
 _cons |  14.37543   .9822445    14.64  0.000   12.43418   16.31668
    
```

15

Arm Circumference and Weight

- Can also use results of this MLR for estimating mean arm circumferences for children given height, weight, and sex

$$\hat{y} = 14.4 + -0.17x_1 + 1.46x_2 + 0.30x_3$$

- Here: \hat{y} is estimated mean arm circumference, x_1 = height (cm), x_2 = weight (kg), x_3 = sex (1 = female)

16