


Simple Linear Regression: John McGready



Section B

Linear Regression: Motivating Example

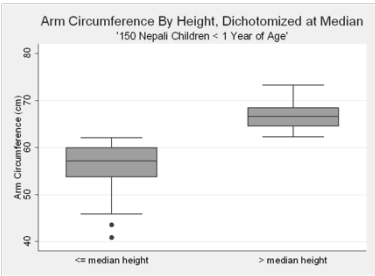
Example: Arm Circumference and Height

- Data on anthropomorphic measures from a random sample of 150 Nepali children [0, 12) months old
- Question: what is the relationship between average arm circumference and height
- Data:
 - Arm circumference: mean 12.4 cm, SD 1.5 cm, range 7.3 cm - 15.6 cm
 - Height: mean 61.6 cm, SD 6.3 cm, range 40.9 cm - 73.3 cm

2

Approach 1: Arm Circumference and Height

- Dichotomize height at median, compare mean arm circumference with t-test and 95% CI



3

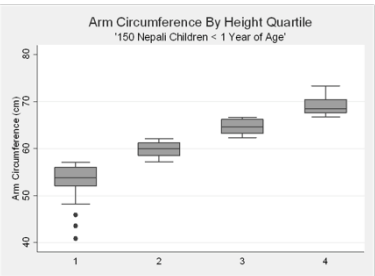
Approach 1: Arm Circumference and Height

- Potential advantages:
 - We know how to do it!
 - Gives a single summary measure (sample mean difference) for quantifying the arm circumference/height association
- Potential disadvantages:
 - Throws away a lot of information in the height data that was originally measured as continuous
 - Only allows for a single comparison between two crudely defined height categories

4

Approach 2: Arm Circumference and Height

- Categorize height into four categories by quartile, compare mean arm circumference with ANOVA, 95% CIs



5

Approach 2: Arm Circumference and Height

- Potential advantages:
 - We know how to do it!
 - Uses a less crude categorization of height than the previous approach of dichotomizing
- Potential disadvantages:
 - Still throws away a lot of information in the height data that was originally measured as continuous
 - Requires multiple summary measures (six sample mean differences between each unique combination of height categories) to quantify arm circumference/height relationship
 - Does not exploit the structure we see in the previous boxplot: as height increases so does arm circumference

6

Simple Linear Regression: John McGready

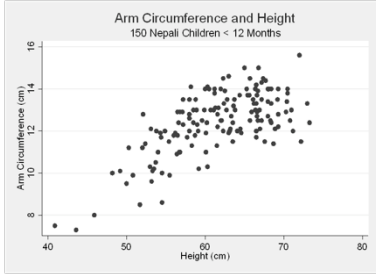
Approach 3: Arm Circumference and Height

- What about treating height as continuous when estimating the arm circumference/height relationship
- Linear regression is a potential option: allows us to associate a continuous outcome with a continuous predictor via a line
 - The line estimates the mean value of the outcome for each continuous value of height in the sample used
 - Makes a lot of sense: but only if a line reasonably describes the outcome/predictor relationship
- Linear regression can also use binary or categorical predictors (will show later in this set of lectures)

7

Visualizing Arm Circumference and Height Relationship

- A useful visual display for assessing the nature of association between two continuous variables: a scatterplot



8

Visualizing Arm Circumference and Height Relationship

- Question: does a line reasonably describe the general shape of the relationship between arm circumference and height?
- We can estimate a line, using the computer (details to come in subsequent lecture section)
- The line we estimate will be of the form:

$$\hat{y} = \beta_0 + \beta_1 x$$
- Here \hat{y} is the average arm circumference for a group of children all of the same height, x

9

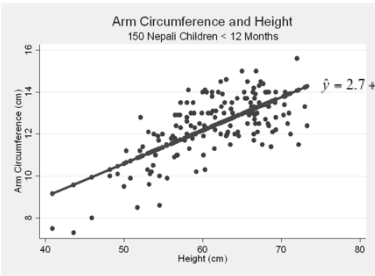
Example: Arm Circumference and Height

- Equation of regression line relating estimated mean arm circumference (cm) to height (cm): from Stata
 - $\hat{y} = 2.7 + 0.16x$
 - Here, \hat{y} = estimated average arm circumference (like what we previously would call \bar{y}), x = height, $\beta_0 = 2.7$ and $\beta_1 = 0.16$
 - This is the estimated line from the sample of 150 Nepali children

10

Example: Arm Circumference and Height

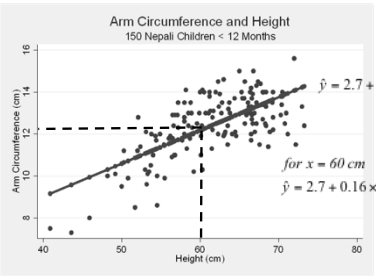
- Scatterplot with regression line superimposed



11

Example: Arm Circumference and Height

- Estimated mean arm circumference for children 60 cm in height

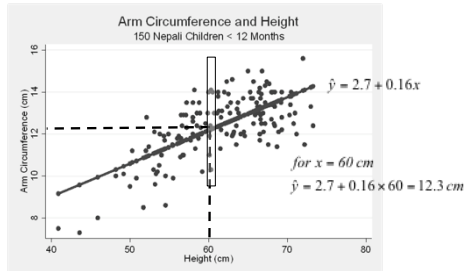


12

Simple Linear Regression: John McGready

Example: Arm Circumference and Height

- Notice, most points don't fall directly on the line: we are estimating the mean arm circumference of children 60 cm tall: observed points vary about the estimated mean



13

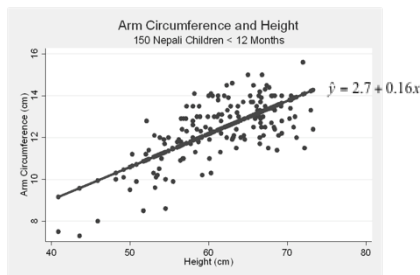
Example: Arm Circumference and Height

- How to interpret estimated slope?
 - $\hat{y} = 2.7 + 0.16x$
 - Here, $\hat{\beta}_1 = 0.16$
 - Two ways to say the same thing:
 - $\hat{\beta}_1$ is the average change in arm circumference for a one-unit (1 cm) increase in height
 - $\hat{\beta}_1$ is the mean difference in arm circumference for two groups of children who differ by one-unit (1 cm) in height, taller to shorter
 - These results estimate that the mean difference in arm circumferences for a one cm difference in height is 0.16 cm, with taller children having greater average arm circumference

14

Example: Arm Circumference and Height

- This mean difference estimate is constant across the entire height range in the sample: definition of a slope of a line



15

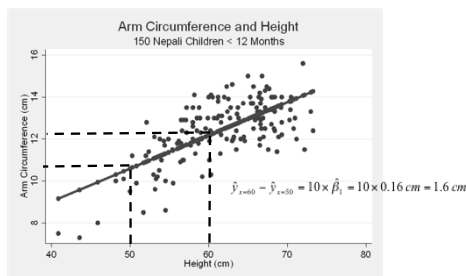
Example: Arm Circumference and Height

- What is the estimated mean difference in arm circumference for:
 - Children 60 cm tall versus children 59 cm tall?
 - Children 25 cm tall versus children 24 cm tall?
 - Children 72 cm tall versus children 71 cm tall?
 - Etc.?
 - Answer is the same for all of the above: 0.16 cm

16

Example: Arm Circumference and Height

- What is estimated mean difference in arm circumference for . . .
 - Children 60 cm tall versus children 50 cm tall?



17

Example: Arm Circumference and Height

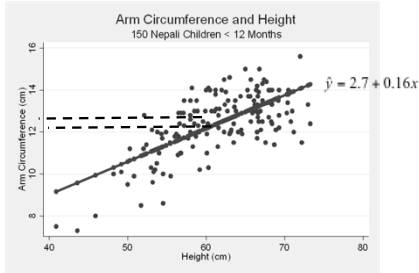
- What is estimated mean difference in arm circumference for . . .
 - Children 90 cm tall versus children 89 cm tall?
 - Children 34 cm tall versus children 33 cm tall?
 - Children 110 cm tall versus children 109 cm tall?
 - Etc.?
 - This is a trick question!

18

Simple Linear Regression: John McGready

Example: Arm Circumference and Height

- The range of observed heights in the sample is 40.9 cm - 73.3 cm: our regression results only apply to the relationship between arm circumference and height for this height range



19

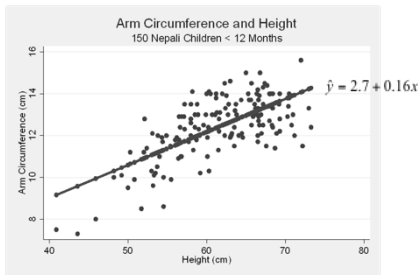
Example: Arm Circumference and Height

- How to interpret estimated intercept?
 - $\hat{y} = 2.7 + 0.16x$
 - Here, $\hat{\beta}_0 = 2.7\text{cm}$
 - This is the estimated y when $x = 0$: the estimated mean arm circumference for children 0 cm tall
 - Does this make sense given our sample?
 - As we noted before, the estimate of mean arm circumferences only applies to observed height range
 - Frequently, the scientific interpretation of the intercept is scientifically meaningless
 - But this intercept is necessary for fully specifying the equation of a line and making estimates of mean arm circumference for groups of children with heights in the sample range

20

Example: Arm Circumference and Height

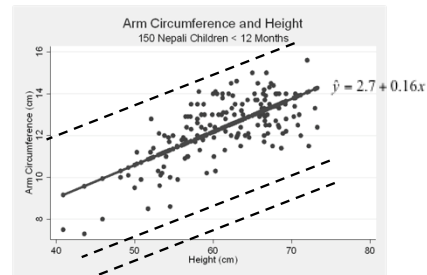
- Notice that $x = 0$ is not even on this graph (the vertical axis is at $x = 39$)



21

Example: Arm Circumference and Height

- Notice that $x = 0$ is not even on this graph (the vertical axis is at $x = 39$)



22