JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section 6d: Practice Problem Solutions

John McGready
Johns Hopkins University

Example: Wages and Education Level

1. Recall equation of regression line relating estimated mean hourly wages (U.S. $, 1985) to years of education: from Stata

$$\hat{y} = -0.75 + 0.75x$$

   – This regression is based on a random sample of 534 U.S. workers in 1985. The standard error estimate of the slope, $S\hat{E}(\hat{\beta}_1) = 0.08$

   a) Estimate a 95% CI for the true population slope relating hourly wages to years of education

   The resulting 95% CI can be estimated by $\hat{\beta}_1 \pm 2S\hat{E}(\hat{\beta}_1)$.
   *This yields 0.75±2*0.08, which gives a 95% CI of $0.59/hr to $0.91/hr.*

Example: Wages and Education Level

1. Recall equation of regression line relating estimated mean hourly wages (U.S. $, 1985) to years of education: from Stata

$$\hat{y} = -0.75 + 0.75x$$

   – This regression is based on a random sample of 534 U.S. workers in 1985. The standard error estimate of the slope, $\hat{SE}(\hat{\beta}_1) = 0.08$

b) What is the estimated 95% confidence interval for the mean difference in hourly wages (in year, 1985) for persons with 16 years of education versus 12 years of education?

Recall, the resulting estimated mean difference is given by $4\hat{\beta}_1$. *A 95% for 4B*$_1$ *is* $4\hat{\beta}_1 \pm 2\hat{SE}(4\hat{\beta}_1) \rightarrow 4\hat{\beta}_1 \pm 2 \times 4\hat{SE}(\hat{\beta}_1)$*Notice this is equivalent to* $4\left(\hat{\beta}_1 \pm 2\hat{SE}(\hat{\beta}_1)\right)$, *which can be easily obtained by multiplying the endpoints of your answer to part A by 4: (4*0.59, 4*0.91) gives a 95% CI of ($2.36/hr, $3.64/hr).*

Example: Arm Circumference and Sex

2. Recall the regression relating arm circumference to child's sex for the random sample of 150 Nepali children less than 12 months old.

$$\hat{y} = 12.5 + -0.13x$$

- The estimated standard error of the slope estimate is $S\hat{E}(\hat{\beta}_1) = 0.24$. In this example, x is the binary variable for sex, coded as a 1 for female children, and 0 for male children. Suppose x was coded as 1 for male children and 0 for female children.

a) What would the resulting 95% confidence interval for the true population slope be?

*As shown in the previous section, the resulting slope estimate with the reverse coding of x would be 0.13. The standard error will not be affected by the change in coding and the 95% would be estimated by 0.13±2*0,024, or (-0.35, 0.61).*