**Biostatistics 140.653**
**Third Term, 2021**
**February 15, 2020**
**Quiz 1 Solution**

**The purpose of this quiz is to assess your knowledge of the course materials covered during the first two weeks of class and covered in Problem Set 1.**

**Instructions:**

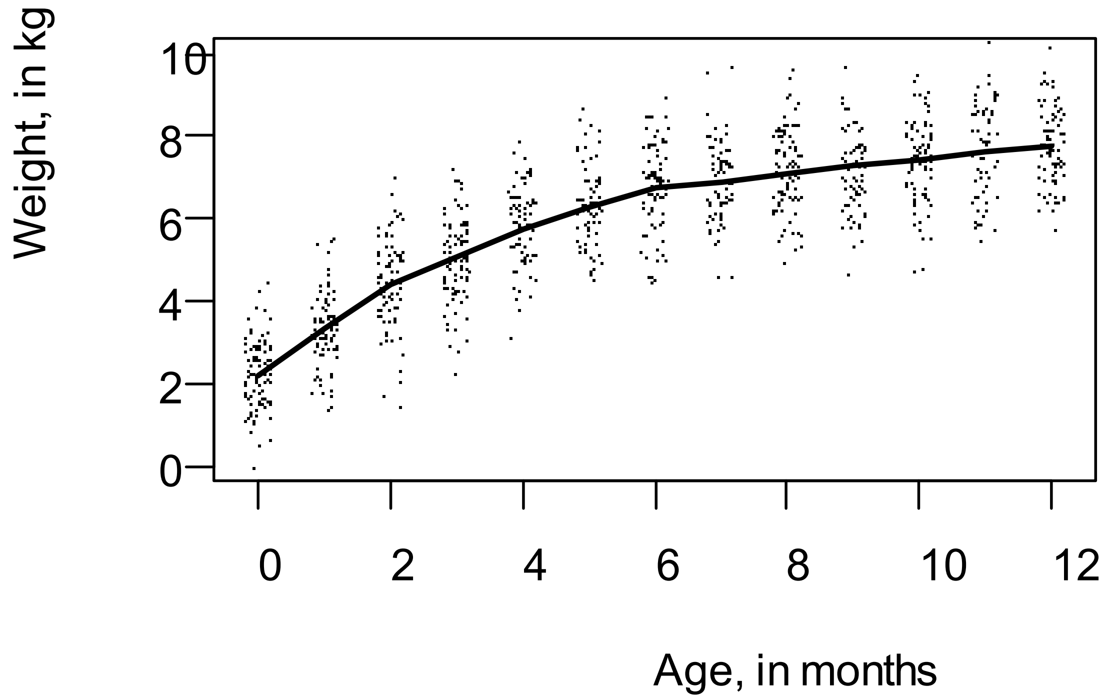
- **This is an open book quiz; you may consult your course notes and handouts.**
- **You should not discuss this quiz with any other student during Monday Feb 15th.**
- **This quiz is designed to be completed in 20-30 minutes.**
- **You may provide your solution by editing the word version of this quiz, annotating the pdf version of this quiz or writing your solution on paper and submitting a picture of your solution.**

**By signing my name, I enter agree to abide by the instructions above and the Johns Hopkins University School of Public Health Academic Code:**

**Name (Print): ________________________________________**

**Signature: __________________________________________**

**Below find a plot of weight against age (points include horizontal jitter) and a fitted curve estimated using a particular linear regression model with a subset of observations from the Nepali Children's Anthropometry Dataset that you used in Problem Set 1.**




1. Write the **multiple linear regression equation** for the fitted line **using specific numeric values** (not letters) for the coefficients. (Hint: if you site down the line, you will see 3 knots at 2, 4, and 6 months)

NOTE: Your numeric answers may vary from mine but I am looking for the correct procedure for estimation; not exactly the same numeric answers

The regression model can be written as a linear spline model with knots at 2, 4, and 6 months:

$$\boldsymbol{Y_i = B0 + B1age_i + B2(age_i - 2)^+ + B3(age_i - 4)^+ + B4(age_i - 6)^+ + \varepsilon_i}$$

To estimate the parameters, we need to estimate the average weight when age = 0, and compute the slope in each bin of age:

Average weight at age = 0: 2.2 kg
Monthly change in weight for ages 0 to 2 months: (4.2 – 2.2) / 2 = 1
Monthly change in weight for ages 2 to 4 months: (5.7 – 4.2) / 2 = 0.75
Monthly change in weight for ages 4 to 6 months: (6.5 – 5.7) / 2 = 0.4
Monthly change in weight for ages 6 to 12 months: (7.7 – 6.5) / 6 = 0.2

So from this information, we can estimate B0, B1, …, B4:

$\boldsymbol{\widehat{B0} = 2.2}$
$\boldsymbol{\widehat{B1} = 1}$

$$\widehat{B2} = 0.75 - \widehat{B1} = -0.25$$
$$\widehat{B3} = 0.4 - 0.75 = -0.35$$
$$\widehat{B4} = 0.2 - 0.4 = -0.2$$

So the multiple linear regression equation for the fitted line can be written as:

$$E(Y_i|age_i) = 2.2 + age_i - 0.25(age_i - 2)^+ - 0.35(age_i - 4)^+ - 0.2(age_i - 6)^+$$

2. An estimate of the residual standard deviation is (choose single best answer)
(a). 0.2 kg
(b). 0.8 kg
(c). 2 kg
(d). 8 kg
(e). 20 kg

In the classic MLR model, we have Y = mean(X) + residual so that the standard deviation of the residuals represents spread in the values of Y about the mean (which depends on X).

To obtain the 0.8kg as the best answer, we note that the majority of the children's weights are within 1.5 to 2kg of the mean weight at any age. We would expect that 95% of weights to fall within +/- 2 x SD of weights at any age, indicating that the SD of weights at any age is roughly 0.75 to 1kg.