Data Analysis and Statistical Inference
Dr. Çetinkaya-Rundel
Duke University

In-Video Quiz Questions for Unit 7: Part 4 – (1) Diagnostics for MLR

(01:04) – slide 2, after "So each numerical explanatory variable needs to be linearly related to the response variable."

1. Which of the following in a residuals plot indicates a linear relationship between the variables involved?

(a) fan shaped pattern
(b) linear pattern
(c) increasing pattern
(d) decreasing pattern
(e) no pattern, i.e. complete random scatter

(06:33) – slide 8, after "what do we need in terms of the sampling of the data to obtain independent observations."

2. Which of the following is not useful for checking for independence of observations in the sample?

(a) random sample
(b) if sampling without replacement, sample size is less than 10% of the population
(c) if running an experiment, random assignment
(d) sample size greater than 30

(07:18) – slide 9, after "so it appears that any sort of time series structure is not a consideration for this dataset."

3. Which of the following is not a condition that needs to be met for multiple linear regression?

(a) Nearly normal residuals with mean 0
(b) Nearly normal response variable with mean 0
(c) Constant variability of residuals
(d) Independent residuals
(e) Each (numerical) variable linearly related to outcome

Data Analysis and Statistical Inference
Dr. Çetinkaya-Rundel
Duke University

Answers:

1. e

Explanation: If the variables are linearly associated, the residuals plot should show a completely random scatter since the linear model will be able to capture the trend in the data and leave nothing behind for the residuals.

2. d

Explanation: Sample size > 30 is useful for nearly normal sampling distribution of the mean in cases where the population distribution is not normal, however it has nothing to do with independence of observations.

3. b

Explanation: There is no restriction on the mean of the response variable (that would be very restrictive anyway).