Epi 3 Check on Learning (COL) Quiz

COL quizzes are *NOT* intended to be extremely challenging. Rather, the goal of COL quizzes are to simply assess students’ basic comprehension of the assigned materials and provide them with feedback early and without high stakes.

**Quiz Instructions:** Please complete this check on learning quiz after reviewing all of the required readings for this module. You may take this quiz as many times as you like.

# 

# Question ideas:

* Question about central limit theorem and normal distribution (BOW, p. 54).
* Regression to the mean (BOW, p. 56).
* Think “due to” and say “associated with.” (BOW, p. 72)
* Path diagrams are a precursor to modern DAGs (BOW, p. 77, sort of – not these exact words).
* Causal discovery vs. causal effect estimation (BOW, p. 79)

# Q1. Why use regression

[Multiple Choice]

The readings for this module discussed some limitations of simple stratification techniques that can be overcome with regression techniques. Which of the following was not one of them?

|  |  |
| --- | --- |
|  | It is difficult to simultaneously adjust for many covariates using simple stratification techniques. |
|  | Simple stratification techniques allow only for adjustment of categorical covariates. |
|  | Simple stratification techniques can fail to return results when the strata are too numerous and the data become sparse. |
| ✅ | Simple stratification techniques require advanced statistical analysis software, making them less accessible to many epidemiologists. |

**Feedback**:

“Stratification-based methods, however, have practical limitations for multivariate adjustment:

1. Although they can be used to adjust for several covariates simultaneously, adjustment is carried out only for the association between one independent variable and an outcome at a time. For example, to assess the association of oral contraceptives with myocardial infarction while controlling for age and educational level, it would be necessary to create one 2 × 2 table for oral contraceptives vis-à-vis myocardial infarction for each stratum defined by combining age groups and educational levels. If the exposure of interest were educational level and the covariates to be adjusted for were age and oral contraceptive use, however, a new set of 2 × 2 tables would have to be created (representing education vs myocardial infarction for each stratum defined by categories of age and oral contraceptive use).
2. These methods allow adjustment only for categorical covariates (e.g., gender); continuous covariates need to be categorized, as age was in the example shown in Table 7-2. Residual differences within these somewhat arbitrarily defined categories may in turn result in residual confounding (Section 7.6 and Chapter 5, Section 5.5.4).
3. Data become sparse when the strata are too numerous. For the direct method, for example, if the sample size of a given stratum is 0, no corresponding stratum-specific rate is available for application to the standard population in that stratum; as a result, the adjusted rate becomes undefined.

Thus, in practice, stratification methods are usually limited to simultaneous adjustment for a few categorical confounders (usually one or two), with a small number of categories each. When simultaneous adjustment for multiple covariates (including continuous variables) is needed, methods based on multiple-regression techniques are typically used.

Szklo, Moyses, Nieto, F. Javier. Epidemiology (Kindle Locations 7170-7184). Jones & Bartlett Learning. Kindle Edition.

# Q2. Common types of regression

[Multiple Answer]

Which of the following are a type of frequently used regression model discussed in the readings for this module?

|  |  |
| --- | --- |
| ✅ | Linear regression |
| ✅ | Logistic regression |
| ✅ | Proportional hazards (Cox) regression |
| ✅ | Poisson regression |

**Feedback**:

The sections that follow describe four of the most frequently used regression models for multivariate adjustment in epidemiology: (1) linear regression, used when the outcome is continuous (e.g., blood pressure); (2) logistic regression, preferentially used when the outcome is categorical (cumulative incidence, prevalence); (3) proportional hazards (Cox) regression, used in survival analysis; and (4) Poisson regression, used when incidence rate (based on person-time) is the outcome of interest.

Szklo, Moyses, Nieto, F. Javier. Epidemiology (Kindle Locations 7193-7196). Jones & Bartlett Learning. Kindle Edition.

# 

# Q3. Model formulas

[Multiple Choice]

The reading shows the mathematical equations used to described linear regression models. For example, the reading shows the following linear regression equation in one example:

What do the authors call (pronounced beta zero)?

|  |  |
| --- | --- |
|  | The outcome |
| ✅ | The intercept |
|  | The regression coefficient |
|  | The constant |

**Feedback**:

“ is the intercept that is, the estimated value of y when x = 0.”

Szklo, Moyses, Nieto, F. Javier. Epidemiology (Kindle Locations 7275-7276). Jones & Bartlett Learning. Kindle Edition.

# Q4. The Quincunx

[Multiple Choice]

On February 9, 1877, Francis Galton spoke at the Royal Institution of Great Britain in London’s Friday Evening Discourse. The title of his talk was the “Typical Laws of Heredity.” Galton used a simple apparatus during his talk that evening to help demonstrate some of his assertions. The machine consisted of a “triangular array of pins or pegs, into which small metal balls can be inserted through an opening at the top.” What is the name of the apparatus?

|  |  |
| --- | --- |
|  | An abacus |
|  | A difference engine |
| ✅ | A Quincunx |
|  | A printing press |

**Feedback**:

“His [Galton’s] experimental apparatus for the evening was a curious contraption that he called a quincunx… consists of a triangular array of pins or pegs, into which small metal balls can be inserted through an opening at the top.” (p. 54)

Pearl J, Mackenzie D. *The Book of Why: The New Science of Cause and Effect*. Basic Books; 2018.

# Q5. Sewell Wright

[Multiple Choice]

\_\_\_\_\_ was the first person to develop a mathematical method for answering causal questions from data, known as path diagrams.

|  |  |
| --- | --- |
|  | R. A. Fisher |
|  | Francis Galton |
|  | Karl Pearson |
| ✅ | Sewell Wright |

**Feedback**:

“Sewell Wright was the first person to develop a mathematical method for answering causal questions from data, known as path diagrams.” (p. 74)

Pearl J, Mackenzie D. *The Book of Why: The New Science of Cause and Effect*. Basic Books; 2018.