Epi 3 Lab Warm-up

At the beginning of each in-person lab session students will complete a warm-up together, along with the instructor, as a class using a series of interactive poll questions that provide immediate feedback. These activities will not be graded. However, participating gives students an opportunity to self-assess their comprehension and ask questions.

The lab warm-up questions are placed into an interactive poll question format on the [Socrative](https://www.socrative.com/) platform.

# 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.