ST 4242
Lecture 2

Jialiang Li

Independence assumption

- What is the definition of independence?
- Advantage of independence.
- What is the meaning for uncorrelated observations?
- When do we encounter a violation of data independence?
- Why do longitudinal data not satisfy the independence assumption?

Statistical independence

- In probability theory, to say that two events are **independent,** intuitively means that the occurrence of one event makes it neither more nor less probable that the other occurs.
- Example:
- - The event of getting a 1 the first time a die is rolled and the event of getting a 1 the second time are *independent*.
- - By contrast, the event of getting a 1 the first time a die is rolled and the event that the sum of the numbers seen on the first and second trials is 3 are *dependent*.

Statistical independence

- Two events A and B are **independent** if and only if $\Pr(A \cap B) = \Pr(A)\Pr(B)$.
- Two random variables X and Y are independent if and only if for any numbers a and b the events $\{X \leq a\}$ (the outcomes where X being less than or equal to a) and $\{Y \leq b\}$ are independent events as defined above.

Properties

- If X and Y are independent, then the expectation operator E has the nice property
  - $E[X\ Y] = E[X]\ E[Y]$,
- and for the variance we have
  - $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$,

Related to distributions

- Furthermore, random variables X and Y with distribution functions $F(x)$ and $F(y)$, and probability densities $f(x)$ and $f(y)$, are independent if and only if the combined random variable (X,Y) has a joint distribution
    - $F(x,y) = F(x)F(y)$,
- or equivalently, a joint density
    - $f(x,y) = f(x)f(y)$.

Uncorrelaedness

- In probability theory and statistics, two real-valued random variables are said to be **uncorrelated** if their covariance is zero.
- Uncorrelated random variables have a correlation coefficient of zero, except in the trivial case when both variables have variance zero (are constants). In this case the correlation is undefined.

Uncorrelatedness

- If X and Y are independent, then they are uncorrelated. However, not all uncorrelated variables are independent. For example, if X is a continuous random variable uniformly distributed on $[-1, 1]$ and $Y = X$ *square*, then X and Y are uncorrelated even though X determines Y.
- Uncorrelatedness is a relation between only two random variables. By contrast, independence can be a relationship between more than two.

FEATURES OF LONGITUDINAL DATA

- Defining feature: repeated observations on individuals, allowing the direct study of change.
- Note that the measurements are commensurate, i.e. the same variable is measured repeatedly.
- Longitudinal data require sophisticated statistical techniques because the repeated observations are usually (positively) correlated.
- Sequential nature of the measures implies that certain types of correlation structures are likely to arise.
- Correlation must be accounted for to obtain valid inferences.

Advantage of longitudinal data

- Repeated observations are not perfectly correlated.
- more powerful than cross sectional data for a fixed number of subjects.
- each subject can serve as his or her own control. Intra-subject variability is substantially less than inter-subject variability.

Examples

- See text book and author’s slides. 2-19.