Estimating the effect of an experimental treatment[1]

STA305 Winter 2014

[1] See last slide for copyright information.

Sources

You don't need to look at these.

- *Theory of the design of experiments* (Cox and Reid, 2000)
- *Sampling design and analysis* (Lohr, 2009)

Assumption of unit-treatment additivity

Experimental units are randomly assigned to either a treatment condition or a control condition.

- Say the treatment has an effect.
- What effect?
- Suppose the treatment adds the same constant Δ to all values of the response variable in the experimental condition.
- Cox and Reid call this the "Assumption of unit-treatment additivity."
- Certainly it's not the only possibility.
- But it's very standard.

Random assignment is like sampling from a finite population

Use sample survey notation (Lohr, 2009)

- We have N experimental units.
- Sample n without replacement for the experimental group.
- For $i = 1, \ldots, N$ let

$$Z_i = \begin{cases} 1 & \text{if unit } i \text{ is chosen} \\ 0 & \text{if unit } i \text{ is not chosen} \end{cases}$$

- $E(Z_i) = P(Z_i = 1) = \frac{n}{N}$
- $Var(Z_i) = \frac{n}{N}\left(1 - \frac{n}{N}\right)$
- $Cov(Z_i, Z_j) = -\frac{n}{N}\left(1 - \frac{n}{N}\right)/(N-1)$

More definitions and properties

$Z_i = 1$ if unit i is selected, zero otherwise.

If all experimental units were in the control condition, their response variable values would have been $y_1, \ldots, y_N$.

$$\overline{y}_u = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{N} Z_i y_i, \quad S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y}_u)^2$$

$$E(\overline{y}) = \overline{y}_u \qquad Var(\overline{y}) = \frac{S^2}{n}\left(1 - \frac{n}{N}\right)$$

- $\overline{y}_1 = \frac{1}{n}\sum_{i=1}^{N} Z_i(y_i + \Delta)$
- $\overline{y}_2 = \frac{1}{N-n}\sum_{i=1}^{N}(1 - Z_i)y_i$

Have $E(\overline{y}_1 - \overline{y}_2) = \Delta$ and $Var(\overline{y}_1 - \overline{y}_2) = \frac{S^2}{n\left(1-\frac{n}{N}\right)}$.

Under the randomization model

- $E(\overline{y}_1 - \overline{y}_2) = \Delta$
- So $\overline{y}_1 - \overline{y}_2$ is an *unbiased estimator* of the treatment effect Δ.
- More later on the precision of this estimate

Random sampling model

- Suppose the N experimental units actually are a simple random sample from some large population.
- Approximately, the observed values of the response variable are independent and identically distributed random variables.
- Now we'll call the total sample size n.
- Random assignment of n_1 units to the treatment condition yields two independent random samples from the same distribution, with expected value μ and variance σ^2.
- Often assumed normal.
- $n_1 + n_2 = n$.
- The assumption of unit-treatment additivity says the treatment adds the constant Δ to the n_1 observations in the treatment condition.
- Want to estimate and test hypotheses about Δ.

Estimating the treatment effect

$$\overline{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}(Y_{i,1} + \Delta) \qquad \overline{Y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_{i,2}$$

- $\hat{\Delta} = \overline{Y}_1 - \overline{Y}_2$ is an unbiased estimator of Δ.
- $Var(\overline{Y}_1 - \overline{Y}_2) = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$
- Want the estimate to be as precise as possible.
  - Make σ^2 small somehow.
  - Make n_1 and n_2 big.
  - For fixed $n_1 + n_2 = n$, choose n_1 to minimize $Var(\hat{\Delta})$.

Extension to more treatments

- Can have p treatment conditions (including control).
- Effects $\Delta_1, \ldots, \Delta_{p-1}$

Dummy variable regression

A very good way to write the model

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i$$

Make a table.

Copyright Information