3) Experiment Randomization and Design

Vitor Kamada

January 2018

Randomized Experiments vs Observational Studies

Cochran (1972, 2015): "*randomized experiments as settings where the the assignment mechanism does not depend on characteristics of the units, either observed or unobserved, and the researcher has control over the assignments*".

(Rosenbaum, 1995; Imbens and Rubin, 2015): *In observational studies, the researcher does not have control over the assignment mechanism, and the assignment mechanism may depend on observed and or unobserved characteristics of the units in the study*".

Athey & Imbens (2016): Experimental Lalonde Data

	Average				
Covariate	Treated	Controls	Difference	s.e.	exact p-value
African-American	0.84	0.83	0.02	(0.04)	0.700
Hispanic	0.06	0.11	-0.05	(0.03)	0.089
age	25.8	25.0	0.8	(0.7)	0.268
education	10.3	10.1	0.3	(0.2)	0.139
married	0.19	0.15	0.045	(0.04)	0.368
no-degree	0.71	0.84	-0.13	(0.04)	0.002
earnings 1974	2.10	2.11	-0.01	(0.50)	0.983
unemployed 1974	0.71	0.75	-0.04	(0.04)	0.329
earnings 1974	1.53	1.27	0.27	(0.31)	0.387
unemployed 1975	0.60	0.69	-0.09	(0.05)	0.069

Stitch a Collar: Standard vs Ergonomic Workplace

Stitch <-
read.table("http://www.stat.umn.edu/~gary/book/fcdae.data/exmpl2.1",
header=TRUE)

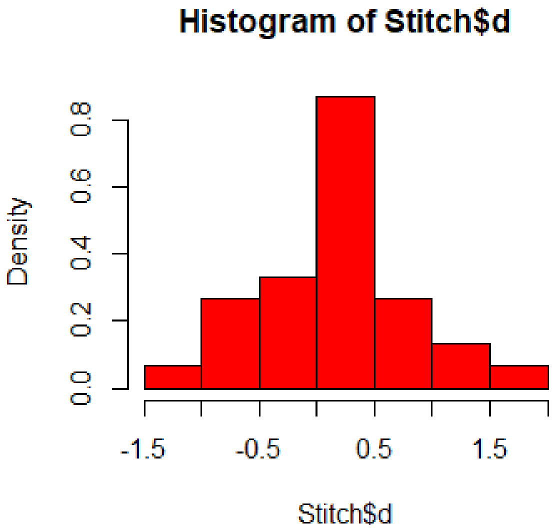
Stitch$d <- Stitch$std - Stitch$ergo

summary(Stitch$d)

Statistic	N	Mean	St. Dev.	Min	Max
std	30	4.956	0.488	4.360	6.390
ergo	30	4.781	0.482	3.870	5.590
d	30	0.175	0.645	−1.080	1.750

Bezjak and Knez (1995)

hist(Stitch$d, freq=FALSE, col="red")



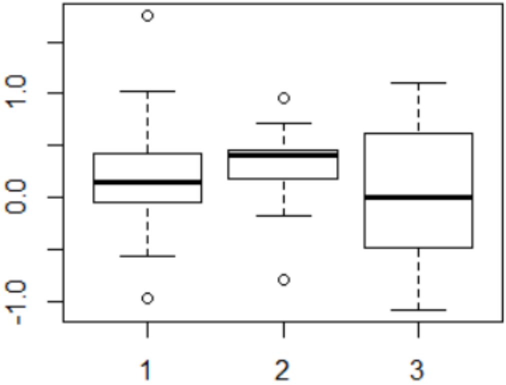
boxplot(G1, G2, G3)

d <- Stitch$std - Stitch$ergo

G1 <- d[1:10]

G2 <- d[11:20]

G3 <- d[21:30]



Paired T-Test

$$H_0 : \mu = 0 \text{ vs } H_A : \mu > 0$$

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2}$$

Result: Paired T-Test

t.test(d,alt="great")

$$t = 1.49,\ df = 29,\ \text{p-value} = 0.0735$$

t.test(G1,alt="great")

$$t = 0.92514,\ df = 9,\ \text{p-value} = 0.1895$$

t.test(G2,alt="great")

$$t = 1.8305,\ df = 9,\ \text{p-value} = 0.05021$$

t.test(G3,alt="great")

$$t = 0.10466,\ df = 9,\ \text{p-value} = 0.4595$$

Absorption of Phosphorus by Rumex Acetosa

15 Days				28 Days			
4.3	4.6	4.8	5.4	5.3	5.7	6.0	6.3

Two-Sample t-Test

$$H_0 : \mu_1 = \mu_2 \text{ vs } H_A : \mu_1 < \mu_2$$

$$t = \frac{\bar{y}_{1\bullet} - \bar{y}_{2\bullet}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_p = \sqrt{\frac{\sum_{i=1}^{n_1}(y_{1i} - \bar{y}_{1\bullet})^2 + \sum_{i=1}^{n_2}(y_{2i} - \bar{y}_{2\bullet})^2}{n_1 + n_2 - 2}}$$

t.test(Data$y~Data$days, var.equal=TRUE, alternative="less")

t = -3.3273, df = 6, p-value = 0.00793

$$t = \frac{\bar{y}_{1\bullet} - \bar{y}_{2\bullet}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\frac{4.775 - 5.825}{.446\sqrt{\frac{1}{4} + \frac{1}{4}}} = -3.33$$