Lecture Six: Controlling Sources of Variation: Paired Comparison Design

Example: Boys' Shoe Problem

- Two shoe materials: A, B
- B is cheaper, but possibly, wears away faster
- Response: Y is amount of shoe wear in *mm.* $\mu_A = E(Y_A)$, $\mu_B = E(Y_B)$.
- $H_0 : \mu_A = \mu_B$ H_1: $\mu_A \neq \mu_B$

Design 1: *Completeley Randomized Design*
Randomly assign 5 boys to A shoes, 5 boys to B shoes.
Test statistic

$$t(\mathbf{y_A}, \mathbf{y_B}) = \frac{\bar{y}_A - \bar{y}_B}{s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

Reject H_0 if the test statistic is large (Note: one-sided test).
Sources of variation?
Boys.... If S_p is very large, it may swamp any difference between A and B (see picture).
Solution: Make experimental units as similar as possible.

Design 2: *Randomized Complete Block Design*
Each of 10 boys is randomly assigned to either

- A on right foot; B on left foot
- B on right foot; A on left foot

(see plot)

Naive (incorrect) Analysis
10 A observations; 10 B observations

$$\begin{aligned} t(\mathbf{y_A}, \mathbf{y_B}) &= \frac{\bar{y}_A - \bar{y}_B}{s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \\ &= \frac{11.04 - 10.63}{1.109} = 0.369 \end{aligned}$$

$$Pr(T_{18} \geq 0.369) = 0.358$$

So (apparently) we don't reject H_0.....
But on closer inspection:

- Assumptions made: independence, equal variances, normality
- *But* are the observations independent? If we know that 'John' wore down his left sole, does that give us no information about his right sole?

- The t-statistic compares: $\bar{y}_A - \bar{y}_B$ to $s_p\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}$. But what is s_p estimating?

$$s^2_{\text{pooled}} = \frac{\sum_{i=1}^{n_A}(y_{iA}-\bar{y}_A)^2 + \sum_{i=1}^{n_B}(y_{iB}-\bar{y}_B)^2}{(n_A-1)+(n_B-1)}$$

  This would make sense as an estimate of the variability, if the left and right feet of each boy were subjected to independent sources of wear! But clearly this is not the case....

We can make these intuitions more precisely by thinking about the following model:

$$\begin{aligned} Y_{iA} &= \mu_A + \beta_i + \epsilon_{iA} \\ Y_{iB} &= \mu_B + \beta_i + \epsilon_{iB} \end{aligned}$$

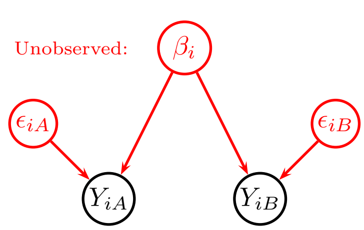
where Y_{iA} is the wear on the A shoe for boy i, and likewise for Y_{iB}. These equations form a *statistical model*, for the responses, breaking the response into three components.

- Treatment effect: μ_A, μ_B;
- Boy effects: β_i;
- 'Noise': ϵ_{iA}, ϵ_{iB}.
- Note that the model assumes things behave *additively*, i.e. it assumes that any difference in wear between the materials is the same regardless of whether the boy is a 'low' wearer (β_i small) or a 'high' wearer (β_i large).

Suppose that we make the following assumptions:

- $\beta_1, \ldots, \beta_{10} \sim_{\text{i.i.d.}} N(0, \tau^2)$
- $\epsilon_{1A}, \ldots, \epsilon_{10A}, \epsilon_{1B}, \ldots, \epsilon_{10B} \sim_{\text{i.i.d.}} N(0, \sigma^2/2)$

Why $\sigma^2/2$? This will be clear shortly.



- Q. Are the 20 observations independent?

  A.

$$\begin{aligned} \text{Cov}(Y_{iA}, Y_{iB}) &= E\left((Y_{iA} - \mu_A)(Y_{iB} - \mu_B)\right) \\ &= E\left((\beta_i + \epsilon_{iA})(\beta_i + \epsilon_{iB})\right) \\ &= \text{Cov}(\beta_i, \beta_i) \\ &= \tau^2 \neq 0 \end{aligned}$$

  thus the observations *within* a boy (i.e. from the same boy) are *not* independent in this model (unless $\tau^2 = 0$... which would mean?)

- Q. What is s_p in the two-sample t-test estimating?

  A. s_p^2 estimates $V(Y_{iB}) = V(Y_{iA}) = V(\beta_i + \epsilon_{iA}) = \tau^2 + \sigma^2/2$

- Q. What is the two-sample t-test doing?

  A. It is comparing $\bar{y}_B - \bar{y}_A$ to (an estimate of) $\tau^2 + \sigma^2/2$ (times $\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$).

  But this means that if τ^2 is very large then no difference will be detected.

- Q. Didn't we justify the t-test as an approximation to a randomization distribution, where all possible assignments were possible?

  A. Yes. Here we have restricted the possible assignments (every boy gets one A shoe and one B shoe) so we would have a different randomization distribution. (More on this later.)

Paired Comparisons

It's clearly time for a new idea:

Boy	y_A	y_B	$y_B - y_A$
1	13.2	14.0	0.8
2	8.2	8.8	0.6
⋮	⋮	⋮	⋮
10	13.3	13.6	0.3
			mean: 0.41

Consider

$$\begin{aligned} D_i = Y_{iB} - Y_{iA} &= (\mu_B + \beta_i + \epsilon_{iB}) - (\mu_A + \beta_i + \epsilon_{iA}) \\ &= (\mu_B - \mu_A) + (\epsilon_{iB} - \epsilon_{iA}) \end{aligned}$$

Under our (additive) 'boy effect' model:

- $E(D_i) = \mu_B - \mu_A$
- $V(D_i) = \sigma^2/2 + \sigma^2/2 = \sigma^2$

Let's revisit our hypothesis test:

Under H_0, $\mu_A = \mu_B$, so $D_1, \ldots, D_{10} \sim_{\text{iid}} N(0, \sigma^2)$.

So we can use our one-sample t-test:

$$t_d(\mathbf{y_A}, \mathbf{y_B}) = \frac{\bar{d}}{s_d/\sqrt{10}}$$

where

$$s_d^2 = \frac{1}{n_d - 1} \sum_{i=1}^{n_d} (d_i - \bar{d})^2$$

and $n_d = n_A = n_B$. Under the null hypothesis $t_d(\mathbf{y_A}, \mathbf{y_B})$ will have a t-distribution on $n_d - 1$ d.f.

Boys' shoe example:

$\bar{d} = 0.41$, $s_d = 0.12$, $n_d = 10$, so $t_d(\mathbf{y_A}, \mathbf{y_B}) = 3.5$, and

$$Pr(T_9 \geq 3.5) = 0.0042$$

Hence we have lots of evidence against the Null hypothesis in favour of the alternative that the wear on B was greater.

Note that this worked because there was greater variability 'between' boys, than within. However, had we paired the data in an arbitrary manner, s_d would not have been much less than s_p, yet we would have reduced our d.f. from 18 to 9.

Summary

This design is a *paired comparison design*, which is a type of *randomized block design*.

- A block is a plot of experimental material or subgroup of units that is more homogeneous than the whole.
  - In the boys' shoe problem: each 'boy' formed an experimental 'block'.
  - Observations 'within' a boy are more similar than those between.
- Blocking:
  - Each block receives each treatment (more on this later)
  - Allows us to 'subtract out' variation between blocks
- Randomization:
  - Treatments are randomly assigned *within* a block
  - This allows the use of a (suitably modified) randomization test
  - Allows causality to be inferred.

The sound-bite version:

Block what you can; Randomize what you cannot!

i.e. *Block* to control for large, known, sources of variation,
Randomize to eliminate bias from unknown sources of variation.