(7.1)

Paired Comparisons, Block Designs

(Secs 2.5, 4.1)

Recall: A disadvantage to a CRD is that it does not account for differences in experimental units ← def. (treatment applied, generalizations made) 1.

Paired Design: make comparisons between ~~treatments~~ factor levels within matched pairs (blocks)

examples: (1) treatment = drug protocol
block = patients with similar characteristics 2.

(2) treatment = fertilizer combination
block = soil type

(3) treatment = measurement method
block = specimen
i.e., R method, M method, measure octane rating of a gasoline blend

larger ↓ $\frac{R+M}{2}$

Application: Handout 7, Example 7.1

hardness testing machine, two different tips

research question: Does tip have an effect on hardness measurement?

(Is there a systematic difference between the machine tips?) 3.

(7.2)

Data:

Factor level		Block $j=1$	$j=2$	$\cdots$	$j=n$
	$i=1$	Y_{11}	Y_{12}	$\cdots$	Y_{1n}
	$i=2$	Y_{21}	Y_{22}	$\cdots$	Y_{2n}

model: $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$ $\begin{cases} i = 1,2 \\ j = 1,\dots,n \end{cases}$ 4.

- Y_{ij}: ith level, jth block
- τ_i: effect for ith level (factor)
- β_j: effect for jth level (block)

fixed effects: τ_1, τ_2 $\left(\sum_{i=1}^{2} \tau_i = 0\right)$

random effects: $\beta_j \sim N(0, \sigma^2_\beta)$ ← between block variance 4.

random errors: $\varepsilon_{ij} \sim N(0, \sigma^2)$ ← within block (measurement) variance

(think of random effects as representing the experimental units) 5.

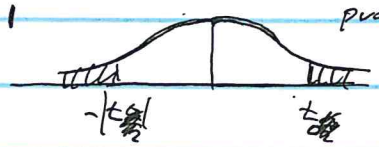
example 7.1: between blocks = differences between metal specimens
within blocks = differences in repeat measurements

Hypothesis Testing, $H_0: \tau_1 = \tau_2 = 0$

Approach #1 → Let $D_j = Y_{1j} - Y_{2j}$ for $j = 1,\dots,n$ 6.

test statistic → $t_0 = \dfrac{\bar{D}}{S_D/\sqrt{n}}$ (one sample t-statistic, using the differences)

reference distribution → t_{n-1}



See Handout for R function guide code (paired.test)

(7.3)

example 7.1 (see R output)

Y_{1j}	7	3	3	4	8	3	2	9	5	4
Y_{2j}	6	3	5	3	8	2	4	9	4	5
D_j	1	0	-2	1	0	1	-2	0	1	-1

$\bar{D} = -0.10$, $S_D = 1.20$, $t_0 = \frac{-0.10}{1.20/\sqrt{10}} = -0.26$

see HW (*)

The experiment finds that machine tip does not have an effect on hardness measurement.

$CI = [-0.96, 0.76]$

---

compare Block Design to Completely Randomized Design

$t_0 = \frac{\bar{D}}{S_d/\sqrt{n}}$ $(df = n-1)$ $\quad t_0 = \frac{\bar{Y}_1 - \bar{Y}_2}{S_p\sqrt{\frac{2}{n}}}$ $(df = 2(n-1))$

t_0 for CRD has more degrees of freedom. 7.

$Var(\bar{D}) = \frac{2\sigma^2}{n}$, $Var(\bar{Y}_1 - \bar{Y}_2) = \frac{2(\sigma^2 + \sigma_\beta^2)}{n}$

HW (*)

A blocking design is better than a CRD when between block variance is large relative to within block variance. 7.

Here, between blocks is differences between specimens, and within blocks is differences in repeat measurements.

Randomized Complete Block Design

see R otput example 7.2

$a = 4$

treatment = extrusion pressure (8500, 8700, 8900, 9100 psi)

block = batch of resin material

response = success proportion

(7.4)

Data:

factor level	Block $j=1$	$\cdots$	Block $j=b$	
$i=1$	y_{11}	$\cdots$	y_{1b}	$\bar{y}_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
$i=a$	y_{a1}	$\cdots$	y_{ab}	$\bar{y}_{a\cdot}$
	$\bar{y}_{\cdot 1}$	$\cdots$	$\bar{y}_{\cdot b}$	$\bar{y}_{\cdot\cdot}$

(*) HW

model: $$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \end{cases}$$

where 8.

$$\beta_j \sim N(0, \sigma_\beta^2), \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

parameter estimates:

$$\hat{\mu} = \bar{y}_{\cdot\cdot}, \quad \hat{\tau}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}, \quad \hat{\beta}_j = \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}$$

$$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j = \bar{y}_{i\cdot} + \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}$$

$$y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}$$ (recall an interaction effect)

(*) HW

mean squares:

$$MS_{tr} = \frac{b \sum_{i=1}^{a} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{a-1}$$

$$MS_{BL} = \frac{a \sum_{j=1}^{b} (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2}{b-1}$$

$$MS_E = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2}{(a-1)(b-1)}$$ 9.

$$E(MS_{tr}) = \frac{b \sum_i \tau_i^2}{a-1} + \sigma^2, \quad E(MS_E) = \sigma^2$$

$$E(MS_{BL}) = a\sigma_\beta^2 + \sigma^2$$ 10.

test statistic: $$F_0 = \frac{MS_{tr}}{MS_E}$$, compare to $F_{a-1,(a-1)(b-1)}$

(7.5)

Example 7.2 (see R output)

interaction.plot(batch, pressure, success)

Why does it make sense for interaction to measure error variance?

We are testing whether or not an observed effect is generalizable to a larger population. ~~This depends on~~ How the effect depends on the experimental units determines how ~~accurately~~ well the effect can be generalized.

Main effects, (i.e., Aggregate Effects) ~~estimated~~

Since levels for a block (exp. units) are not identifiable, we can only estimate the main effect of the ~~factor~~ treatment variable.

more R output

options (contrasts = ...) ← define model parameters

"lme4"

random.mod = lmer(Y ~ (1|B) + A) ← linear mixed effects

random effect, fixed effect

"lmerTest"

anova (random.mod)

HW (*)

$F_0 = 8.11$, $df_s = (3, 15)$, $P = .002$

The experiment finds that extrusion pressure has an effect on the success proportion for vascular grafts.

investigate further: Parameter estimates, pairwise comparisons

see output

$\hat{\tau}_1 = 3.021$, $\hat{\tau}_2 = 1.887$, $\hat{\tau}_3 = -0.879$, $\hat{\tau}_4 = -4.029$

11.

Fisher Comparisons

HW 2(d) (*)

pressure	level			
8500	1	~~A~~	C	
8700	2	A ~~B~~	C	B
8900	3	~~B~~ C		B
9100	4	~~C~~		A

(7.6)

Variance components:

$$\hat{\sigma}^2 = MS_E \quad , \quad \hat{\sigma}^2_\beta = \frac{MS_{BL} - MS_E}{a}$$

$\hat{\sigma}^2 = MS_E$: estimate of within block variance

$\hat{\sigma}^2_\beta = \frac{MS_{BL} - MS_E}{a}$: estimate of between block variance

R output / ~~matlab output~~:

example 7.2

$$\hat{\sigma}^2_\beta = 7.781 \quad , \quad \hat{\sigma}^2 = 7.326$$

HW 2(e)

Back to example 7.1

Suppose we run the analysis as a RCBD.

R output

trtmnt → factor with two levels (tip 1, tip 2)

block → specimen (10 exp. units)

y → hardness measurement

```
rcbd.mod = aov( y ~ block + trtmnt )
summary (rcbd.mod)
```

↙ OK, as long as we are not interested in estimating block variances

$F_0 = 0.07$, $p = .7976$

$(t_0 = -0.26, \quad p = .7976)$

HW 2(b) (*)

Note that $t_0^2 = F_0$. A paired difference analysis is equivalent to a block design analysis, when $a = 2$.

End Notes, #7

1. Recall the definition of experimental unit serves two purposes. The first is in terms of how the experiment is conducted (how the treatment is applied). The second is in terms of how the results are analyzed (how we generalize to the larger population). Thinking about experimental units in this rather basic sense will help us understand what is coming later.

2. It may be better for us to think about studies where the <u>same</u> patient can take both drug protocols. This would not be possible in many situations, but certain non-invasive procedures could lend themselves to a paired comparison.

3. The nature of getting different measurements implies that device has some "effect". Thus, we use <u>effect</u> formally to be a systematic difference.

(for example, in measuring octane for a gasoline blend, method M gives systematically larger measurements than method R $\mu_M > \mu_R$)

4. Because we have multiple measurements for each block (exp unit), we can include a term for block effect.

But because batch levels have no identifiable features, we model batch level effects through a probability distribution.

5. In this example, the metal specimens are the exp. unit. We want to generalize the sample comparison between tips to the larger population of all metal specimens.

6. From the model for Y_{ij}, we can write

$$D_j = Y_{1j} - Y_{2j} = (\mu + \tau_1 + \beta_j + \varepsilon_{1j}) - (\mu + \tau_2 + \beta_j + \varepsilon_{2j})$$

$$= (\tau_1 - \tau_2) + (\varepsilon_{1j} - \varepsilon_{2j})$$

By removing block effects from the comparison, we have reduced the sampling variance.

7. But we have also reduced the sample size. Thus, we only want to use paired comparisons when we can reduce variance. Since block variance is explained in the model, the blocking design is better when our explained variance is large.

8. We can easily extend the model from a paired comparison to modeling a treatment effect with $a > 2$ levels.

9. Algebraic formulas for mean squares follow naturally from the parameter estimates. We will use this connection to simplify derivations in future sections.

10. Note how each MS term estimates the corresponding effect, subject to measurement error variance σ^2.

11. $\hat{\tau}_4$ does not appear on the R output. But since $\sum \hat{\tau}_i = 0$, we can compute $\hat{\tau}_4 = -(\hat{\tau}_1 + \hat{\tau}_2 + \hat{\tau}_3)$

12. These are derived from the expected mean squares.

$E(MS_E) = \sigma^2$, so MS_E estimates σ^2.

$E(MS_{BL}) = a\sigma_\beta^2 + \sigma^2$. We need to subtract an estimate of σ^2 to get at σ_β^2.

Thus, $\hat{\sigma}_\beta^2 = \dfrac{MS_{BL} - MS_E}{a}$ estimates σ_β^2.

13. When $a = 2$, we can run the model as a paired comparison, or as a more general block design. We get equivalent results.