01NAEX - Lecture 03
Factorial design

Jiri Franc

Czech Technical University
Faculty of Nuclear Sciences and Physical Engineering
Department of Mathematics

Introduction to Blocking, Nuisance factors, and Factorial Design
D. C. Montgomery, Design and Analysis of Experiment: Section 4.1

- The randomized complete block design (RCBD).
- Extension of the single factor ANOVA to the RCBD.
- General principles of factorial experiments.
- The two factor factorial with fixed effects.
- The ANOVA for factorials.
- Extensions to more than two factors.
- Quantitative and qualitative factors, response curves and surfaces.
- Estimation of Sample Size.
- Other blocking scenarios (Latin square designs, ...) and Incomplete Block design will be covered next lesson.

Introduction to Blocking and Nuisance Factors

Blocking is a technique for dealing with nuisance factors

- A nuisance factor is a factor that probably has some effect on the response, but it s of no interest to the experimenter however, the variability it transmits to the response needs to be minimized.
- Typical nuisance factors include batches of raw material, operators, pieces of test equipment, time (shifts, days, etc.), different experimental units.
- Failure to block is a common flaw in the design of a experiment and data analysis.

If the nuisance variable is known and controllable, we use blocking:

- If the nuisance factor is known and uncontrollable, sometimes we can use the analysis of covariance to remove the effect of the nuisance factor from the analysis.
- If the nuisance factor is unknown and uncontrollable, we hope that randomization balances out its impact across the experiment.
- Sometimes several sources of variability are combined in a block, so the block becomes an aggregate variable.

When to Block vs When to Just Randomize

Nuisance vs noise

- **Nuisance factor** = known source of systematic variability.
- **Noise** = unknown, many tiny influences you can't name.

Decision guide

- **Known & assignable** (day/operator/lot) $\Rightarrow$ **block** it.
- **Unknown/weak** $\Rightarrow$ rely on **randomization**, check residuals vs run.
- **Trade-off:** Blocking uses $b-1$ **df**; does it noticeably **reduces** MS_E?.

Practical checklist

- Define blocks so each block is **as homogeneous as possible**.
- **Randomize within each block**. (Pre-print run sheets.)
- If β^{block} looks tiny and insignificant, consider CRD.

Fun fact: If **Monday** are a nuisance factors, block it.

Q: In a thermal test, which would you block - **oven**, **operator**, or **day**? Why?

RCBD Model & ANOVA

Model (no treatment-by-block interaction):
$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$, $i = 1, \ldots, a$ (treatments), $j = 1, \ldots, b$ (blocks),
$\sum \tau_i = \sum \beta_j = 0$.

Source	df	SS	MS / Test
Treatments (τ)	$a - 1$	SS_T	$MS_T = SS_T/df_T$
Blocks (β)	$b - 1$	SS_B	$MS_B = SS_B/df_B$
Error	$abn - a - b + 1$	SS_E	$MS_E = SS_E/df_E$

Tests: $F_T = MS_T/MS_E$ (treatment), $F_B = MS_B/MS_E$ (block).
N: If treatment-by-block interaction is significant, include it.

What Blocking Buys You

Model: $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$, $\varepsilon_{ij} \sim (0, \sigma^2)$, blocks variability $\mathrm{Var}(\beta_j) = \sigma_\beta^2$.

Without block (CRD): block shifts leak into error $\Rightarrow$ error variance $\approx \sigma^2 + \sigma_\beta^2$.
With block (RCBD): the block mean is removed $\Rightarrow$ error variance stays σ^2.

Gain in signal-to-noise:

$$F = \frac{MS_\tau}{MS_E} \text{ increases with: Gain} \approx \frac{\sigma^2 + \sigma_\beta^2}{\sigma^2} = 1 + \frac{\sigma_\beta^2}{\sigma^2}.$$

Example: If measurement noise $\sigma = 1$ and day-to-day shifts $\sigma_\beta = 2$,

$$\text{Gain} \approx 1 + \frac{4}{1} = 5 \times \text{ (much larger } F \text{ with blocks).}$$

- Always **randomize within blocks** and check residuals vs run order.

Q: If σ_β^2 is nearly zero, what 's the cost/benefit of blocking? *Hint:* You spend df but gain little precision.

The Hardness Testing Example

- We wish to determine whether different tips produce different (mean) hardness reading on a tester machine.
- 4 different types of tips: conical, ball, cylinder, and cube.
- The test coupons (experimental unit) are a source of nuisance variability.
- Experimental error measures both random error and variability between coupons.

- To conduct this experiment as a RCBD, assign all 4 tips to each coupon.
- Each coupon is called a **block**; that is, it is a more homogeneous experimental unit on which to test the tips.
- Variability between blocks can be large, variability within a block should be relatively small.

In general, a block is a specific level of the nuisance factor. A complete replicate of the basic experiment is conducted in each block and all runs within a block are randomized.

The Hardness Testing Example

Randomized Complete Block Design for the Hardness Testing Experiment

Test Coupon (Block)			
1	**2**	**3**	**4**
Tip 3	Tip 3	Tip 2	Tip 1
Tip 1	Tip 4	Tip 1	Tip 4
Tip 4	Tip 2	Tip 3	Tip 2
Tip 2	Tip 1	Tip 4	Tip 3

In completely randomized single-factor design, 16 different metal test coupons would be required, one for each run in the design.

In RCBD, the variability between coupons from the experimental error is removed.

Only randomization of treatments (Tips) is within the blocks, i.e. blocks represent a restriction on randomization.

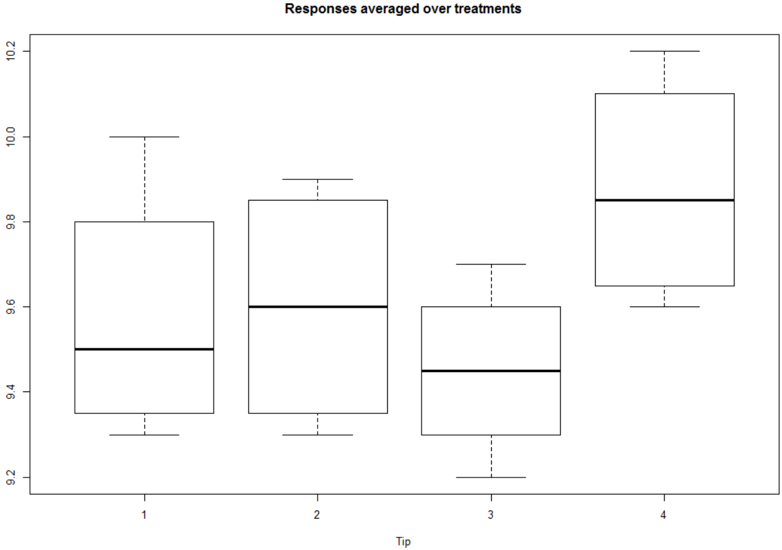
The Hardness Testing Example

The two-way structure of the experiment (Tips x Blocks)

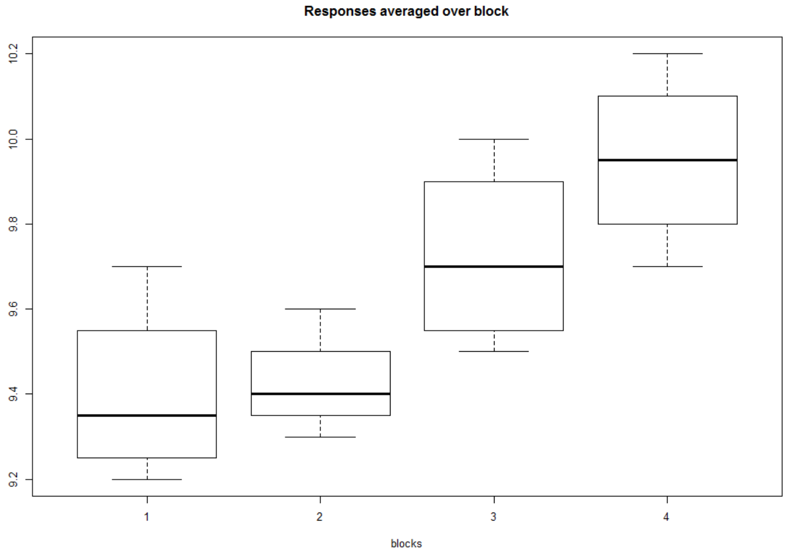
Tip	Test Coupon (Block)			
Number	1	2	3	4
1	9.3	9.4	9.6	10.0
2	9.4	9.3	9.8	9.9
3	9.2	9.4	9.5	9.7
4	9.7	9.6	10.0	10.2

- We are interested in testing the equality of treatment means (again).
- We have to remove the variability associated with the nuisance factor (the blocks).
- Analysis is identical to the two-factor factorial design without replication.

The Hardness Testing Example

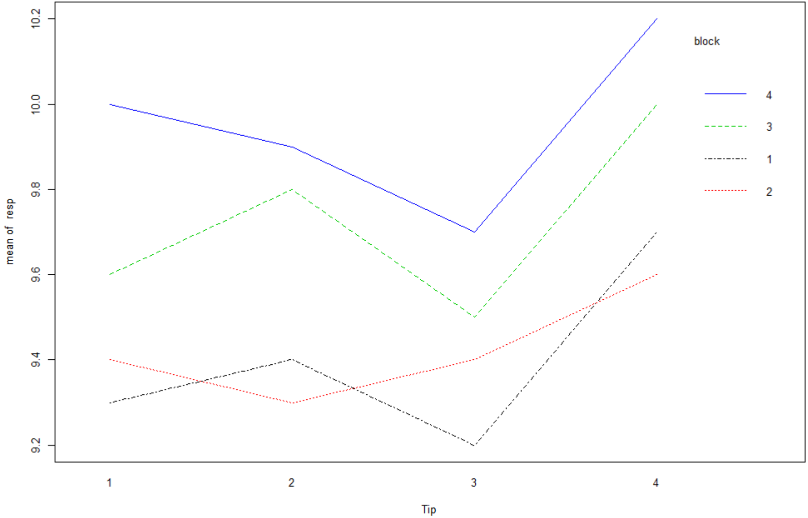


The Hardness Testing Example



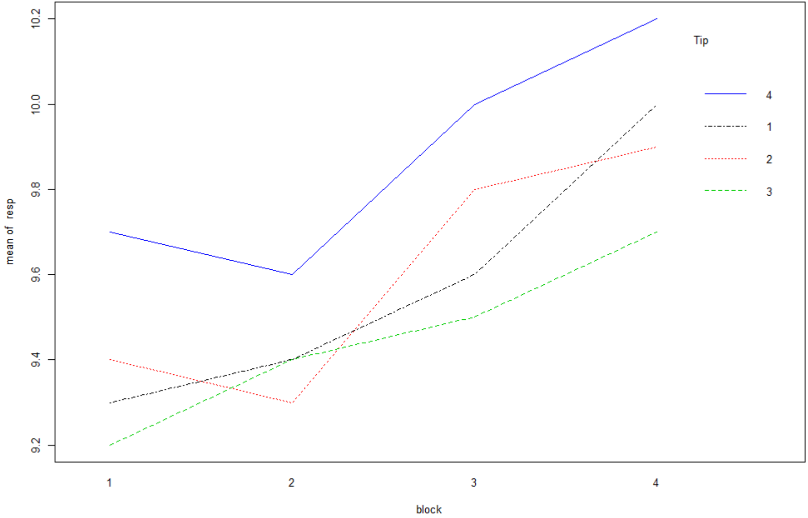
The Hardness Testing Example

Interaction plot over Tips.



The Hardness Testing Example

Interaction plot over Blocks.



ANOVAfor the RCBD

Effects model for the RCBD:

$$y_{ij} = \mu_{ij} + \epsilon_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad i = 1, 2, \ldots, a \quad j = 1, 2, \ldots, b,$$

where μ is an overall mean, τ_i is the effect of the ith treatment, β_j is the effect of the jth block and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ random error term.

Analysis of Variance for a Randomized Complete Block Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}}$	$a-1$	$\frac{SS_{\text{Treatments}}}{a-1}$	$\frac{MS_{\text{Treatments}}}{MS_E}$
Blocks	SS_{Blocks}	$b-1$	$\frac{SS_{\text{Blocks}}}{b-1}$	
Error	SS_E	$(a-1)(b-1)$	$\frac{SS_E}{(a-1)(b-1)}$	
Total	SS_T	$N-1$		

$$\begin{aligned} SS_E &= SS_T - SS_{Treatments} - SS_{Blocks} \\ &= \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2 - b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 - a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2 \end{aligned}$$

The Hardness Testing Example

```
summary(aov(resp~block+Tip,Data1))
            Df Sum Sq Mean Sq F value   Pr(>F)
block        3  0.825 0.27500   30.94 4.52e-05 ***
Tip          3  0.385 0.12833   14.44 0.000871 ***
Residuals    9  0.080 0.00889
```

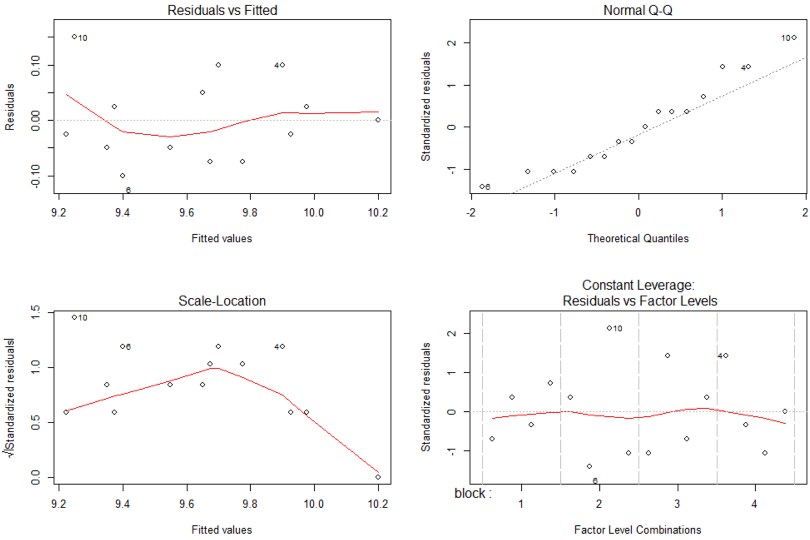
The results obtained from this experiment and RCBD differ from results obtained from Completely Randomized Design.

```
summary(aov(resp~Tip,Data1))
            Df Sum Sq Mean Sq F value Pr(>F)
Tip          3  0.385 0.12833   1.702   0.22
Residuals   12  0.905 0.07542
```

Note, that the MS for Residuals has more than tenfold. All of the variability due to blocks is now in the error term.

The Hardness Testing Example

Model adequacy checking:



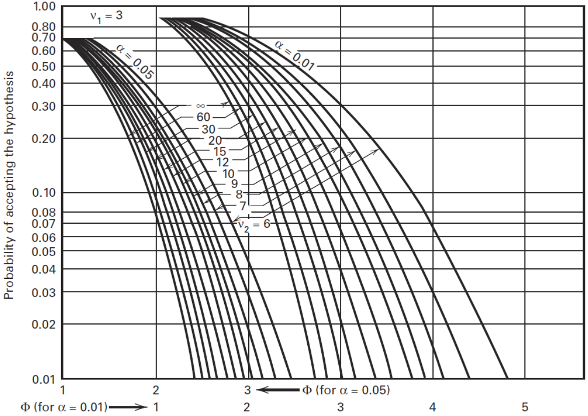
The Hardness Testing Example

Suppose that we would like to be able to determine the appropriate number of blocks to run if we are interested in detecting a true max difference in readings of 0.4 with a high probability and the estimate of the standard deviation is 0.1.

$$\Phi^2 = \frac{bD^2}{2a\sigma^2} \quad \text{,where } \Phi^2 \text{ is related to noncentral r.v. } F_0 = \frac{MS_{Treatments}}{MS_{Error}}$$

```
> b      = seq(3,8,by=1)  # number of blocks
> a     = 4               # number of treatment levels
> D     = 0.4             # max difference in group means
> sigma = 0.1             # standard deviation
>
> Fi_sq = b*(0.4)^2/(2*a*(0.1^2))
> Fi    = sqrt(Fi_sq)
> Fi
[1] 2.449490 2.828427 3.162278 3.464102 3.741657 4.000000
> powers <- power.anova.test(groups=a,n=b,between.var = (D^2/2)/(a),
  within.var=sigma^2,sig.level=.05)$power
> rbind(b , Fi, powers)
            [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
b      3.0000000 4.0000000 5.0000000 6.0000000 7.0000000 8.0000000
Fi     2.4494897 2.8284271 3.1622777 3.4641016 3.7416574 4.0000000
powers 0.8009998 0.9490241 0.9891378 0.9979684 0.9996559 0.9999461
```

OC Curves for the Fixed-Effects Model ANOVA



$\nu_1 = (a-1)$ Numerator df , $\nu_2 = (a-1)(b-1)$ Denominator df

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
a	4.0000000	4.0000000	4.0000000	4.0000000	4.0000000	4.0000000
b	3.0000000	4.0000000	5.0000000	6.0000000	7.0000000	8.0000000
Fi	2.4494897	2.8284271	3.1622777	3.4641016	3.7416574	4.0000000
powers	0.8009998	0.9490241	0.9891378	0.9979684	0.9996559	0.9999461

Power Planning - What Are We Guaranteeing?

Goal: Choose n so the test rejects with probability $\geq 1 - \beta$ when *meaningful* differences exist (at level α).

Two common settings

- **RCBD (Treatment + Block):** Power the **treatment** main effect A. Factor B is a nuisance that reduces noise.
- **Two-treatment factorial:** Power A, B, and **interaction** $A \times B$.

Detectable size(s): Specify minimally important differences:

$$D_A, \quad D_B, \quad D_{AB} \text{ (difference-of-differences).}$$

Q: Is your study

- confirmatory (stricter Family Wise Error Rate (FWER) control)?
- exploratory (report CIs with less stringent False Discovery Rate (FDR) control)?

RCBD (Treatment + Block) & Operating Characteristic (OC)

Model (balanced, n per cell):

$$y_{ijr} = \mu + \tau_i + \beta_j + \varepsilon_{ijr}, \quad i = 1..a,\ j = 1..b,\ r = 1..n, \quad \varepsilon_{ijr} \sim N(0, \sigma^2),$$

$\sum_i \tau_i = \sum_j \beta_j = 0.$

Test of interest (treatments): $H_0 : \tau_1 = \cdots = \tau_a = 0$ with

$$F_0 = \frac{MS_A}{MS_E}, \qquad \mathrm{df}_A = a - 1, \quad \mathrm{df}_E = \begin{cases} (a-1)(b-1), & n = 1 \\ ab(n-1), & n > 1 \end{cases}$$

Under H_1: $F_0 \sim F_{\text{num df, den df}}(\lambda)$ (noncentral F).

Power at level α:

$$1 - \beta \;=\; \Pr\Big(F_{\text{num df, den df}}(\lambda) \;>\; F_{\alpha;\,\text{num df, den df}}\Big).$$

Scaled distance (OC parameter): We use

$$\Phi^2 \equiv \frac{\lambda}{\text{num df}},$$

so OC tables/curves are indexed by $(\alpha, \mathrm{df}_{\text{num}}, \mathrm{df}_{\text{den}}, \Phi^2)$.
Practical view: Bigger Φ^2 (or λ) $\Rightarrow$ larger separation from H_0 $\Rightarrow$ higher power.

Noncentrality for Treatment (RCBD)

Let $\mu_{i\cdot}$ be the block-averaged mean for treatment i and $\bar{\mu}$ the grand mean. For balanced RCBD:

$$\lambda_A = \frac{nb}{\sigma^2} \sum_{i=1}^{a} (\mu_{i\cdot} - \bar{\mu})^2, \qquad \Phi_A^2 = \frac{\lambda_A}{a-1}.$$

Design target (pairwise): Detect *any* pairwise difference $\geq D_A$.
Least favorable configuration (hardest to detect):

$$\mu_{1\cdot} = +D_A/2, \quad \mu_{2\cdot} = -D_A/2, \quad \mu_{3\cdot} = \cdots = \mu_{a\cdot} = 0 \Rightarrow \sum (\mu_{i\cdot} - \bar{\mu})^2 = D_A^2/2.$$

Therefore

$$\boxed{\Phi_A^2 = \frac{nb\, D_A^2}{2(a-1)\sigma^2}}, \qquad \lambda_A = (a-1)\Phi_A^2 = \frac{nb\, D_A^2}{2\sigma^2}.$$

Check: If $b = 1$, this reduces to the one-way formula.

Choosing n in RCBD (From OC Curves)

Given α, target power, a, b, and pilot $\hat{\sigma}^2$:

Recipe

1. From OC tables (or software), find $\Phi^2_{A,\star}$ that achieves the power with $\mathrm{df}_A = a - 1$, df_E as per design.
2. Solve for n from $\Phi^2_A = \dfrac{nb\,D^2_A}{2(a-1)\hat{\sigma}^2}$:

$$\boxed{n = \frac{2(a-1)\,\hat{\sigma}^2}{b\,D^2_A}\,\Phi^2_{A,\star}} \quad \text{(round up).}$$

Heuristics: Halve $\hat{\sigma}^2 \Rightarrow$ halve n; double $D_A \Rightarrow$ quarter n; double $b \Rightarrow$ halve n.

Introduction to Factorial Design

- Factorial Designs are most efficient methods for experiments where we study the effects of two or more factors.
- The effect of the primary factor of interest in the experiment is called **main effect**.
- If the difference in response between the levels of one factor is not the same at all levels of other factors, the **interaction** occurs.

The Battery Design Example

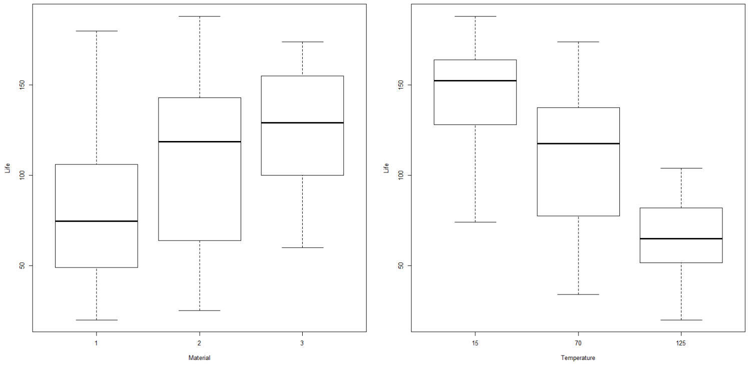
- Two factors: A = Material type; B = Temperature (quantitative variable)
- What effects do material type and temperature have on life?
- Is there a choice of material that would give uniformly long life regardless of temperature?

Life (in hours) Data for the Battery Design Example

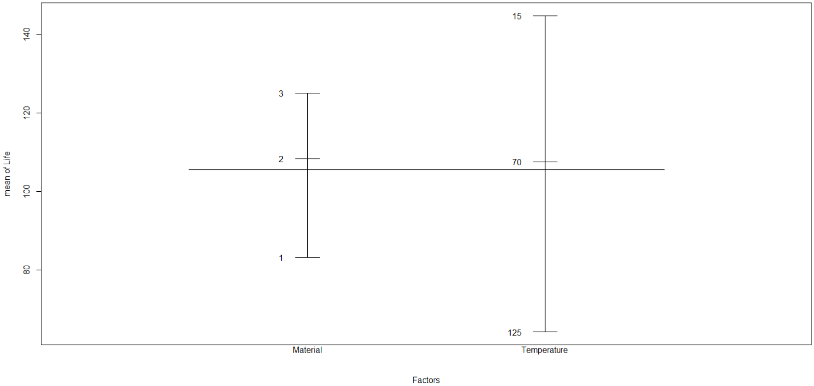
Material Type	Temperature (°F)					
	15		70		125	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

Since there are two factors at three levels, this design is sometimes called a 3^2 factorial design.

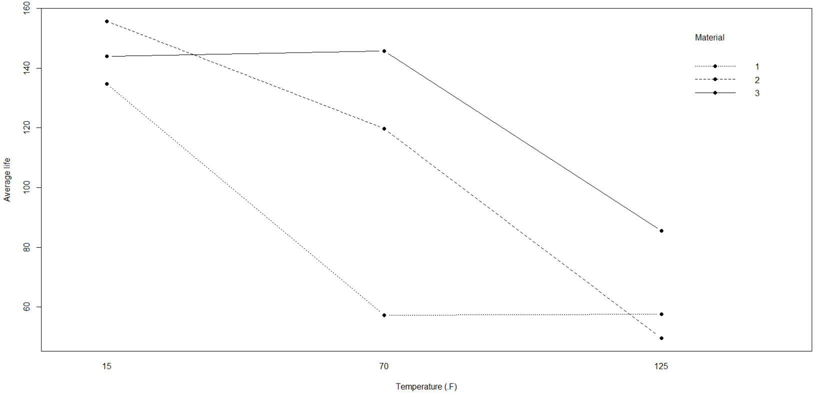
Box plot of data from the Battery Design Example



Effects plot of data from the Battery Design Example



Interactions plot of data from the Battery Design Example



ANOVA for Two-Factor Factorial Fixed-Effects Model

Effects model for the Two Factor - Fixed Factorial Experiment:

$$y_{ijk} = \mu_{ijk} + \epsilon_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ij} \quad i = 1, \ldots, a; \; j = 1, \ldots, b; \; k = 1, \ldots, n.$$

where $(\tau\beta)_{ij}$ is the effect of the interaction between τ_i and β_j.

The Analysis of Variance Table for the Two-Factor Factorial, Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
A treatments	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B treatments	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
Interaction	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n-1)$	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	SS_T	$abn-1$		

$$SS_E = SS_T - SS_{AB} - SS_A - SS_B = 77646.97 - 10683.72 - 39118.72 - 9613.78 = 18230.75$$

ANOVA for Battery Design Example

```
> battery.aov <- aov(Life~Material*Temperature,data=battery)

> summary(battery.aov)
                     Df Sum Sq Mean Sq F value   Pr(>F)
Material              2  10684    5342   7.911  0.00198 **
Temperature           2  39119   19559  28.968 1.91e-07 ***
Material:Temperature  4   9614    2403   3.560  0.01861 *
Residuals            27  18231     675
```

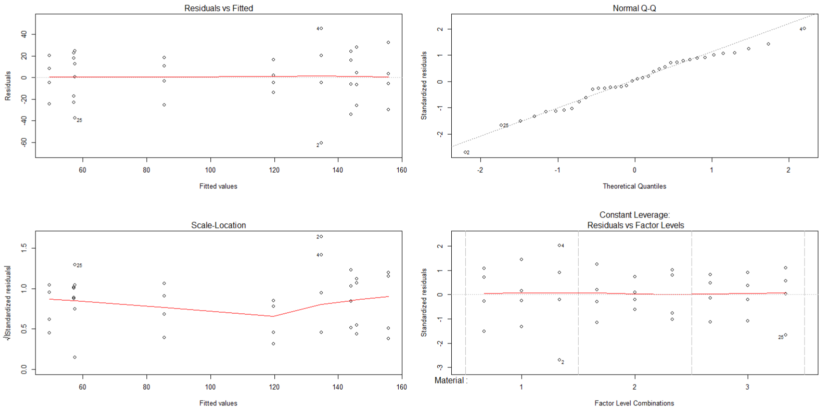
ANOVA for Battery Design Example

```
> summary(battery.aov2 <- aov(Life~Material+Temperature,
          data=battery))
            Df Sum Sq Mean Sq F value    Pr(>F)
Material     2  10684    5342   5.947  0.00651 **
Temperature  2  39119   19559  21.776 1.24e-06 ***
Residuals   31  27845     898


> anova(battery.aov2,battery.aov)
Analysis of Variance Table

Model 1: Life ~ Material + Temperature
Model 2: Life ~ Material * Temperature
  Res.Df   RSS Df Sum of Sq      F  Pr(>F)
1     31 27845
2     27 18231  4    9613.8 3.5595 0.01861 *
```

Model checking for Battery Design Example



Comparison between groups

```
> # Compute means and variance for all levels and factors
>
> with(battery, tapply(Life,list(Material,Temperature),mean))
      15     70  125
1 134.75  57.25 57.5
2 155.75 119.75 49.5
3 144.00 145.75 85.5

> with(battery, tapply(Life,list(Material,Temperature),var))
         15       70      125
1 2056.9167 556.9167 721.0000
2  656.2500 160.2500 371.0000
3  674.6667 508.2500 371.6667
```

Comparisons between the treatment group means

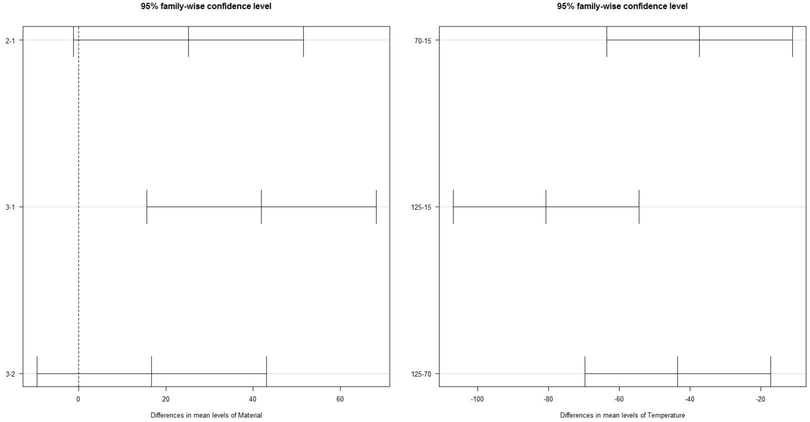
Apply Tukey HSD test to test the pairwise comparisons between the treatment group means

```
> TukeyMaterial = TukeyHSD(battery.aov,which="Material")
> Tukey multiple comparisons of means
  95% family-wise confidence level
  Material
        diff        lwr      upr     p adj
2x1 25.16667 -1.135677 51.46901 0.0627571
3x1 41.91667 15.614323 68.21901 0.0014162
3x2 16.75000 -9.552344 43.05234 0.2717815

> TukeyTemperature = TukeyHSD(battery.aov,which="Temperature")
> Tukey multiple comparisons of means
  95% family-wise confidence level
  Temperature
            diff        lwr       upr     p adj
70x15  -37.25000  -63.55234 -10.94766 0.0043788
125x15 -80.66667 -106.96901 -54.36432 0.0000001
125x70 -43.41667  -69.71901 -17.11432 0.0009787
```

Comparisons between the treatment group means

Visualization of Tukey HSD test:



Comparisons between the treatment group means

Because interaction is significant, we make the comparison at just one level of temperature, say level 2 (70Â°F).

```
battery.aov2<-aov(Life~Material,data=battery[battery$Temperature==70,]
TukeyMaterial2=TukeyHSD(battery.aov_70,which="Material")

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula=Life~Material,data=battery[battery$Temperature==70,])

$Material
diff       lwr        upr      p adj
2-1 62.5  22.59911 102.40089 0.0045670
3-1 88.5  48.59911 128.40089 0.0004209
3-2 26.0 -13.90089  65.90089 0.2177840
```

$$T_{0.05} = q_{0.05}(3,27)\sqrt{\frac{MS_E}{n}} = 3.5\sqrt{\frac{675.21}{4}} = 45.47$$

Test indicates that at the temperature level 70Â°F, the mean battery life is the same for material types 2 and 3, and that the mean battery life for material type 1 is significantly lower in comparison to both types 2 and 3.

Comparisons between the treatment group means

Since the interaction in the model is significant, the effect of Material depends on which level of Temperature is considered. Let us compute the three means at $Temperature = 70F$, appropriate 95% confidence intervals and compare the observed difference of means with the critical value.

```
mm<- with(subset(battery, Temperature==70),
          aggregate(Life,list(M=Material),mean))
  M      x    M       x    M      x
  1  57.25    2  119.75    3 145.75
>val.crit <-qtukey(.95,3,27)*sqrt(unlist(summary(battery.aov))
            [["Mean Sq4"]]/4)
[1] 45.5
```

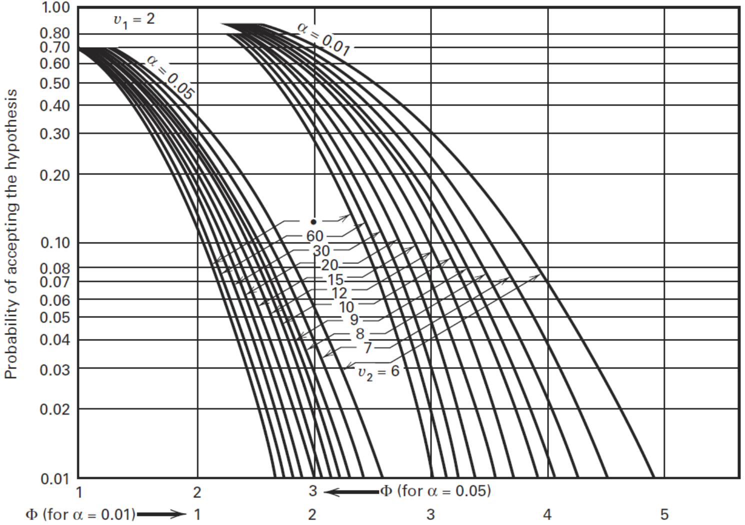
$$T_{0.05} = q_{0.05}(3,27)\sqrt{MS_E/n} = 3.5\sqrt{675.21/4} = 45.5$$

```
> diff.mm <- c(d.3.1=mm$x[3]-mm$x[1],d.3.2=mm$x[3]-mm$x[2],
              d.2.1=mm$x[2]-mm$x[1])
d.3.1 d.3.2 d.2.1
 88.5  26.0  62.5
> names(which(diff.mm > val.crit))
[1] "d.3.1" "d.2.1"
```

In conclusion, only Material type 3 vs. type 1 and Material type 2 vs. type 1 appear to be significantly different when Temperature is fixed at 70Â℉.

Sample Size Determination - Operating Characteristic Curve

Operating Characteristic Curve Parameters for the Two-Factor Factorial, Fixed Effects Model. (See charts V in the Montgomery DOE book)



Two-Treatment Factorial - What to Power

Model (balanced, *n* per cell):

$$y_{ijr} = \mu + \tau_i + \beta_j + \gamma_{ij} + \varepsilon_{ijr},$$

$$\mathrm{df}_A = a-1,\ \mathrm{df}_B = b-1,\ \mathrm{df}_{AB} = (a-1)(b-1),\ \mathrm{df}_E = ab(n-1).$$

Noncentrality parameters:

$$\lambda_A = \frac{nb}{\sigma^2}\sum_i (\mu_{i\cdot} - \bar{\mu})^2, \quad \lambda_B = \frac{na}{\sigma^2}\sum_j (\mu_{\cdot j} - \bar{\mu})^2,$$

$$\lambda_{AB} = \frac{n}{\sigma^2}\sum_{i,j} \left(\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \bar{\mu}\right)^2.$$

OC parameters:

$$\Phi_A^2 = \lambda_A/(a-1)$$
$$\Phi_B^2 = \lambda_B/(b-1)$$
$$\Phi_{AB}^2 = \lambda_{AB}/((a-1)(b-1))$$

N: Set targets (D_A, D_B, D_{AB}), choose n to satisfy the strictest β requirement.

Detectable Sizes - Worst-Case Configurations

Main effect A (pairwise $\geq D_A$):

$$\sum_i (\mu_{i\cdot} - \bar{\mu})^2 \geq D_A^2/2 \;\Rightarrow\; \boxed{\Phi_A^2 = \frac{nb\,D_A^2}{2(a-1)\sigma^2}}.$$

Main effect B (pairwise $\geq D_B$):

$$\sum_j (\mu_{\cdot j} - \bar{\mu})^2 \geq D_B^2/2 \;\Rightarrow\; \boxed{\Phi_B^2 = \frac{na\,D_B^2}{2(b-1)\sigma^2}}.$$

Interaction $A \times B$ (diff-of-diffs $\geq D_{AB}$): Embed a single 2×2 pattern with entries $\pm c$ so that the interaction contrast equals D_{AB}; this gives SS $= D_{AB}^2/4$ and

$$\boxed{\Phi_{AB}^2 = \frac{n\,D_{AB}^2}{4\,(a-1)(b-1)\sigma^2}}.$$

N: These are least favorable (hardest to detect) alternatives.

Choosing n in Factorial - Per Effect, Take the Max

From OC tables/curves, obtain $\Phi^2_\star$ for each effect at (α, df).

Planner formulas

$$n_A = \frac{2(a-1)\hat{\sigma}^2}{b\,D_A^2}\,\Phi^2_{A,\star}, \qquad n_B = \frac{2(b-1)\hat{\sigma}^2}{a\,D_B^2}\,\Phi^2_{B,\star},$$

$$n_{AB} = \frac{4\,(a-1)(b-1)\hat{\sigma}^2}{D_{AB}^2}\,\Phi^2_{AB,\star}.$$

Final choice:

$$\boxed{n = \max\{n_A, n_B, n_{AB}\}} \quad \text{(round up).}$$

N: If interaction is likely, n_{AB} often dominates. If not, size for A and B but still *test* $A \times B$ and interpret simple effects if needed.

Sample Size Determination for Battery Experiment

```
> n        = c(2, 3, 4, 5)
> a        = 3                  # number of treatment levels  A
> b        = 3                  # number of treatment levels  B
> D        = 40                 # max difference in group means
> sigma    = 25                 # standard deviation
> Fi_sq    = (n*b*(D)^2)/(2*a*(sigma^2))
> Fi       = sqrt(Fi_sq)
> errorDF  = a*b*(n-1)
> powers   = power.anova.test(groups=3, n=n,
            between.var = (n*(D^2))/((n-1)),within.var=sigma^2,
            sig.level=.05)$power
> rbind(n , Fi ,errorDF, powers)
              [,1]       [,2]      [,3]       [,4]
n       2.0000000  3.0000000  4.000000  5.0000000
Fi      1.6000000  1.9595918  2.262742  2.5298221
errorDF 9.0000000 18.0000000 27.000000 36.0000000
powers  0.6391205  0.9116119  0.979752  0.9956568
```

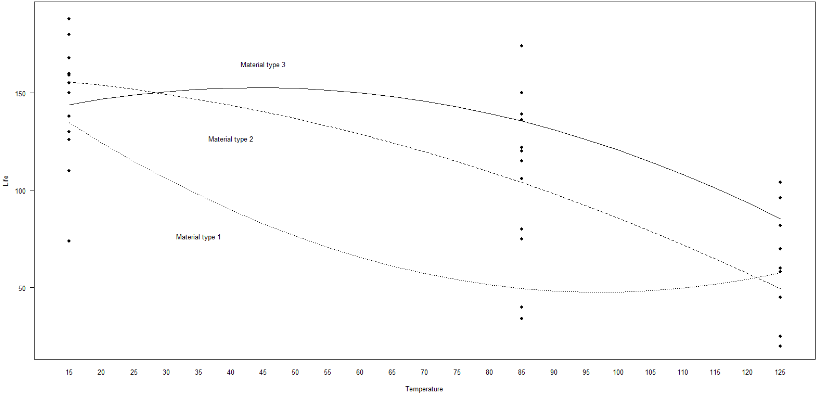
General Factorial Design - Response Curves

Let us assume that temperature is quantitative and use linear regression model to fit data.

```
>Temperature.num=as.numeric(as.character(battery$Temperature))
>battery.aov3=aov(Life~Material+Temperature.num+
               I(Temperature.num^2)+Material:Temperature.num+
               Material:I(Temperature.num^2))
>summary(battery.aov3)
                             Df  Sum Sq  Mean Sq  F value   Pr(>F)
Material                      2   10684     5342    7.911  0.00198
Temperature.num               1   39043    39043   57.823  3.53e-8
I(Temperature.num^2)          1      76       76    0.113  0.73975
Material:Temperature.num      2    2315     1158    1.714  0.19911
Material:I(Temperature.num^2) 2    7299     3649    5.405  0.01061
Residuals                    27   18231      675
```

P-values indicate that A^2 and AB are not significant, whereas the A^2B is significant. By removing A^2 and AB the model is not **hierarchical** any more.

General Factorial Design - Response Curves



P-values indicate that A^2 and AB are not significant, whereas the A^2B is significant. By removing A^2 and AB the model is not **hierarchical** any more.

Homework experiment 01

Perform the homework experiment 01 described in the Rmd/pdf/ipynb files in repo:
01NAEX/projects/01NAEX_HW01_pendulum

The report can be elaborated by group up to 5 students, written in pdf file (R Markdown or Jupyter NB are recommended) and handed till **October 18**.