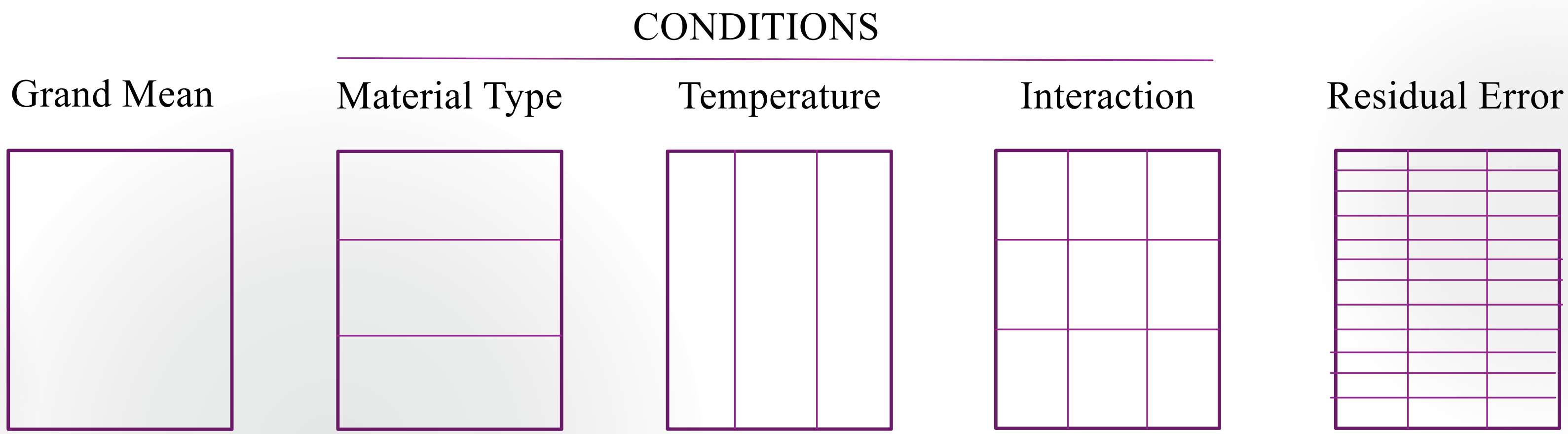
Factorial Crossing

- Each factor corresponds to a meaningful partition of the data

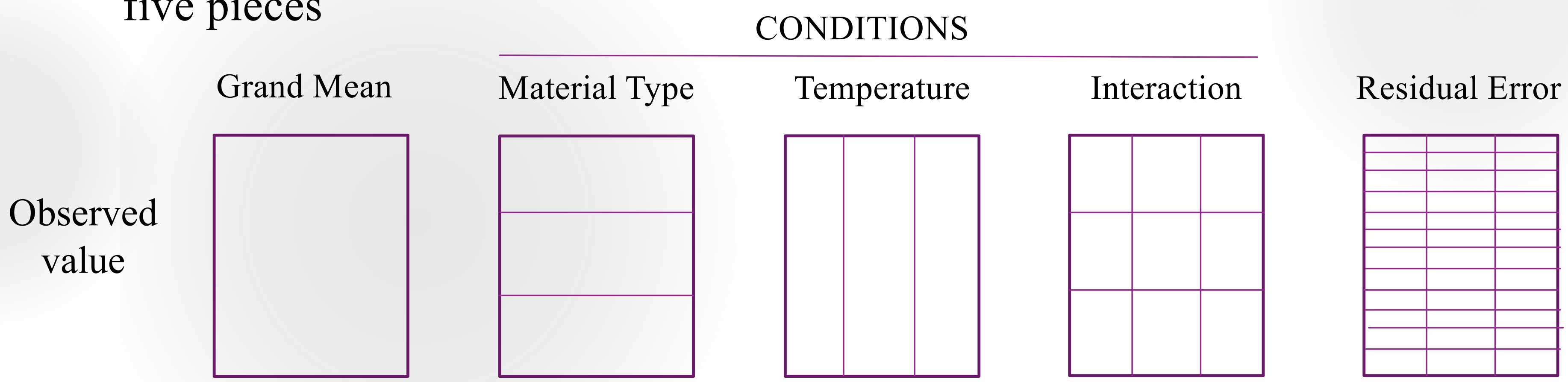


Material Type	Temperature (°F) 15		70		125	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

Factor diagram for the battery life data, a two-way factorial completely randomized design

Looking ahead

- What the factor diagram tells you?
  - Meaningful ways to group and compare observations, essentially in terms of sets of averages
  - Like the assembly line: we regard each observed value as a sum of five pieces



= Benchmark amount + Overall effect of Material Type

+ Overall effect of Temperature

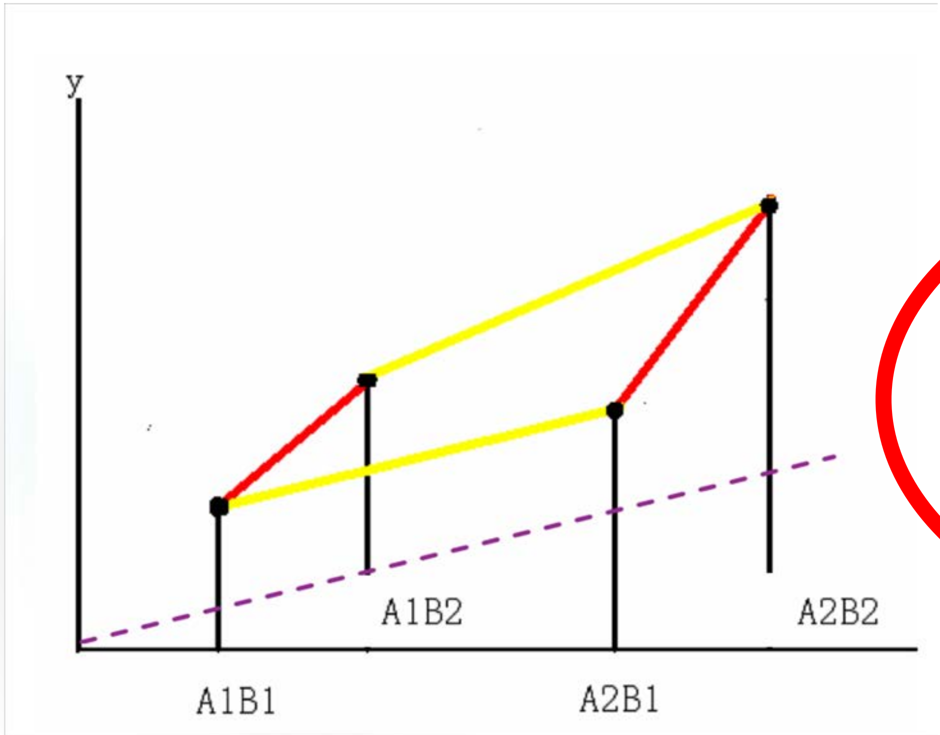
+ Interaction effect + Residual error

Another Example

	Brown eyes	Blue eyes	
Male	CELL MEAN $\bar{y}_{11.}$ Brown eyed males	CELL MEAN $\bar{y}_{12.}$ Blue eyed males	MARGINAL MEAN $\bar{y}_{1..}$ of brown eyed males and blue eyed males
Female	CELL MEAN $\bar{y}_{21.}$ Brown eyed females	CELL MEAN $\bar{y}_{22.}$ Blue eyed females	MARGINAL MEAN $\bar{y}_{2..}$ of brown eyed females and blue eyed females
	MARGINAL MEAN $\bar{y}_{.1.}$ of brown eyed males and brown eyed females	MARGINAL MEAN $\bar{y}_{.2.}$ of blue eyed males and blue eyed females	GRAND MEAN $\bar{y}_{...}$

- Interactions concern: for instance, we might be interested in whether females perform better than males depending on their eye color

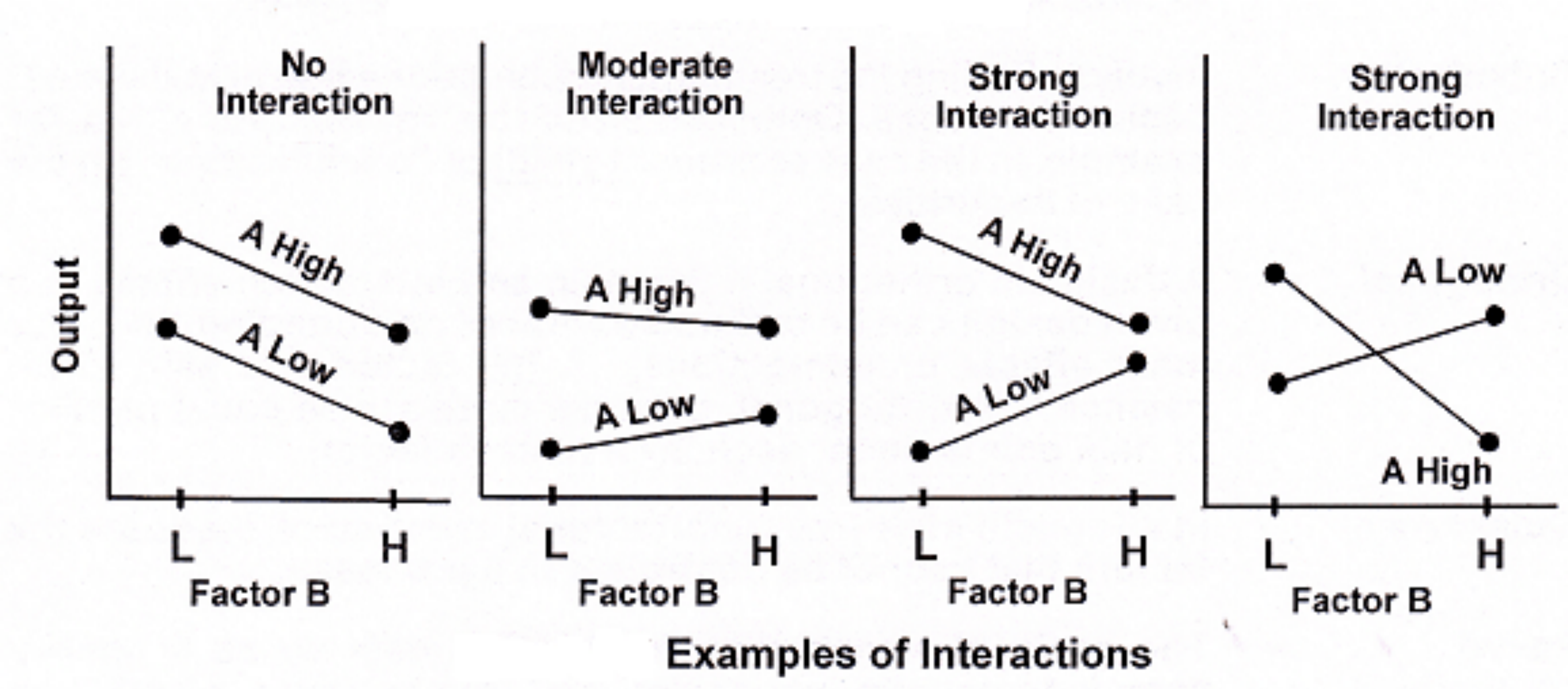
A Graphical Illustration of Interaction Effects



No interaction effect

Parallel lines means

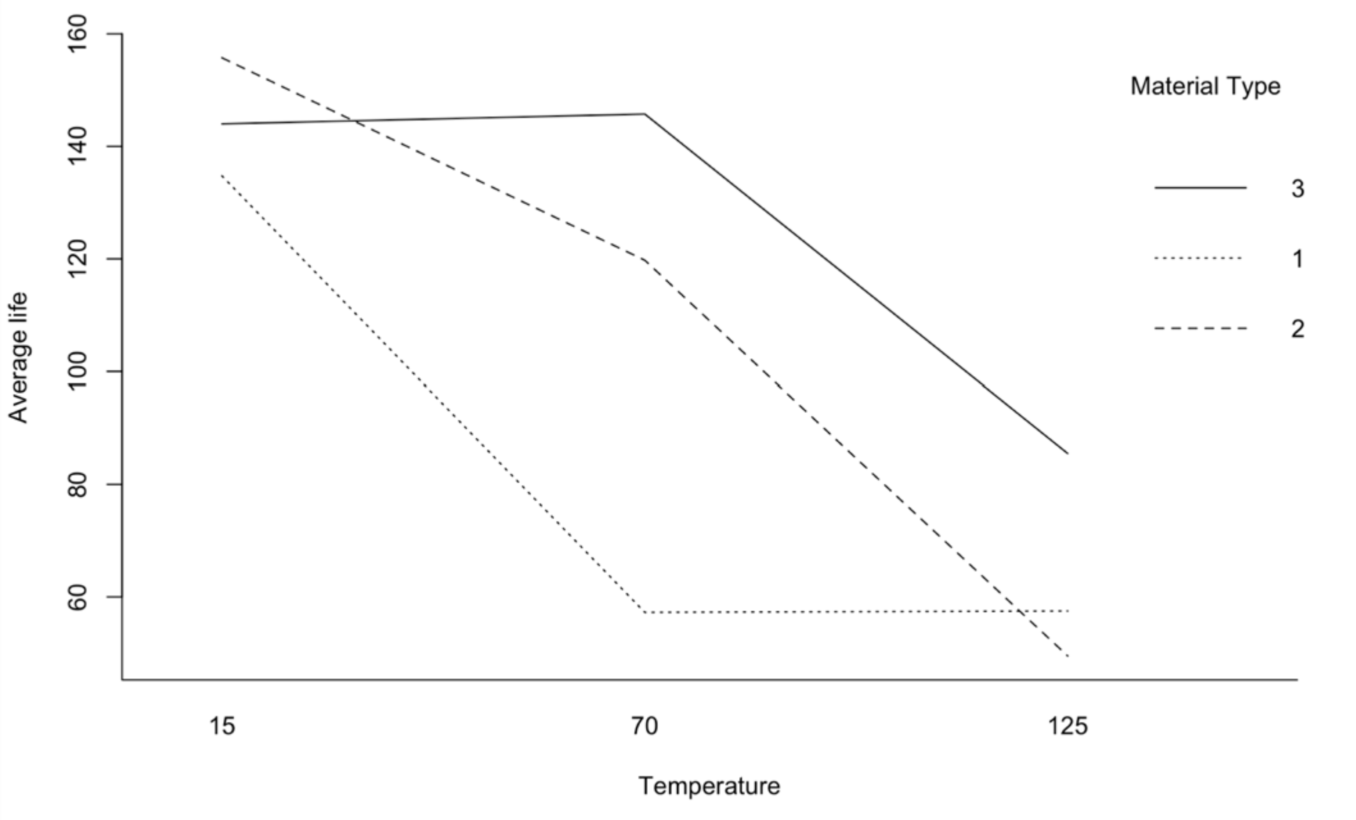
More Illustrations



No interaction effect ← → Parallel lines means

Interaction Plot for the Example

- Strong interaction!😲



Estimation of Main & Interaction Effects

$$\hat{\mu} = \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} = \bar{y}_{\cdots}$$

$$\hat{\alpha}_i = \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} - \hat{\mu} = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}$$

$$\hat{\beta}_j = \frac{1}{an}\sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk} - \hat{\mu} = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}$$

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

$$\hat{\sigma}^2 = \frac{1}{abn - ab}\sum_{i,j}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij\cdot})^2$$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

Advantages of a balanced design:
- Simple parameter estimation
- Easy variance decomposition

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

Test Main & Interaction Effects

- Test main effect of factor A:

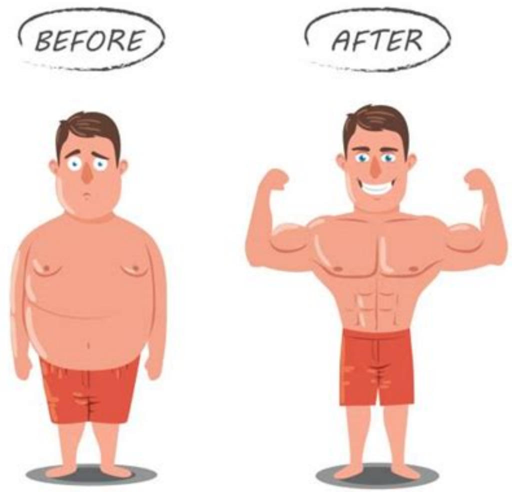
$$\alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

- Test main effect of factor B:

$$\beta_1 = \beta_2 = \cdots = \beta_b = 0$$

- Test interaction effect of factor A & B:

$$\gamma_{11} = \gamma_{12} = \cdots = \gamma_{ab} = 0$$



Two-Way Analysis of Variance (ANOV)

For $i = 1, \cdots, a;\ j = 1, \cdots, b;\ k = 1, \cdots, n;$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2 = bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 + an\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2$$

$$+ n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2$$

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

df breakdown:

$$abn - 1 = a - 1 + b - 1 + (a-1)(b-1) + ab(n-1)$$

	B1	…	Bb	
A1	$\bar{y}_{11.}$	…	$\bar{y}_{1b.}$	$\bar{y}_{1..}$
…				…
Aa	$\bar{y}_{21.}$	…	$\bar{y}_{2b.}$	$\bar{y}_{a..}$
	$\bar{y}_{.1.}$	…	$\bar{y}_{.b.}$	$\bar{y}_{...}$

A Few Facts

(1) SS_T, SS_A, SS_B and SS_E are independent;

(2) $SS_E/\sigma^2 \sim \chi^2_{ab(n-1)}$;

(3) if $\alpha_1 = \cdots = \alpha_a = 0$, then $SS_A/\sigma^2 \sim \chi^2_{a-1}$;

a因素不影响

(4) if $\beta_1 = \cdots = \beta_b = 0$, then $SS_B/\sigma^2 \sim \chi^2_{b-1}$;

(5) if $\gamma_{11} = \gamma_{12} = \cdots = \gamma_{ab} = 0$, then $SS_E/\sigma^2 \sim \chi^2_{(a-1)(b-1)}$;

没有交互影响

ANOVA Table

> summary(aov(Life ~ Type* Temp , data = Battery))

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	$abn-1$		

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Type	2	10684	5342	7.911	0.00198 **
Temp	2	39119	19559	28.968	1.91e-07 ***
Temp:Type	4	9614	2403	3.560	0.01861 *
Residuals	27	18231	675		

Experiments without Replicates

- One run for each combination of A & B

  i.e., full factorial experiment with $n = 1$
- Under the full model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

df breakdown:

$$abn - 1 = a - 1 + b - 1 + (a-1)(b-1) + ab(n-1)$$

df of SS_E is 0 - you run out of degrees of freedom!

- σ^2 is not estimable
- Thus you can't fit an interaction term and so you can't test for interaction

Example: Impurity Causes

- The impurity present in a chemical product is affected by two factors—pressure and temperature
- The data from a single replicate of a factorial experiment are shown below

	Pressure					
Temperature (°F)	25	30	35	40	45	$y_{i.}$
100	5	4	6	3	5	23
125	3	1	4	2	3	13
150	1	1	3	1	2	8
$y_{.j}$	9	6	13	6	10	$44 = y_{..}$

Additive Model & Estimation of Main Effects

Solution: reduce the full model to additive model

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

- No interaction terms
- Free parameters: 1 + (a-1) + (b-1) + 1 = a + b

$$\begin{aligned}
\hat{\mu} &= \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij} = \bar{y}_{..} \\
\hat{\alpha}_i &= \frac{1}{b}\sum_{j=1}^{b} y_{ij} - \hat{\mu} = \bar{y}_{i.} - \bar{y}_{..} \\
\hat{\beta}_j &= \frac{1}{a}\sum_{i=1}^{a} y_{ij} - \hat{\mu} = \bar{y}_{.j} - \bar{y}_{..} \\
\hat{\sigma}^2 &= \frac{1}{(a-1)(b-1)}\sum_{i,j}(y_{ij} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\mu})^2
\end{aligned}$$

ANOVA Table

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

ab -1

$(a$-1$)(b$-1$)$

清华大学统计学研究中心

Randomized Complete Block Design(RCBD)

- Elements:
  - A treatment factor A with a levels
  - A block factor B with b levels
  - Response Y

- Goals:
  - Evaluate the impact of factor A to response Y
  - Remove potential impact of block factor B
  - Assume there is no interaction between A & B

Tukey's One Degree of Freedom Test for Additivity

- Clever John Tukey didn't like this situation much and (being Tukey) he was able to figure out a way to get around the dilemma, at least partially
- The model proposed by Tukey is to formulate it as

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i\beta_j + \varepsilon_{ij}$$

- Estimation of λ
  - Multiply $\alpha_i\beta_j$ on both sides of the equation, then sum over all pairs (*i, j*):
    $\sum_{i,j}\alpha_i\beta_j y_{ij} = \mu\sum_{i,j}\alpha_i\beta_j + \sum_{i,j}\alpha_i^2\beta_j + \sum_{i,j}\alpha_i\beta_j^2 + \lambda\sum_{i,j}\alpha_i^2\beta_j^2 + \sum_{i,j}\alpha_i\beta_j\varepsilon_{ij}$
    Then

$$\lambda \approx \frac{\sum_{i,j}\alpha_i\beta_j y_{ij}}{\left(\sum_i \alpha_i^2\right)\left(\sum_j \beta_j^2\right)}$$

  - With $\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$, $\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$,

$$\hat{\lambda} = \frac{\sum_{i,j}(\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..})y_{ij}}{\left(\sum_i(\bar{y}_{i.} - \bar{y}_{..})^2\right)\left(\sum_j(\bar{y}_{.j} - \bar{y}_{..})^2\right)}$$

ANOVA Decomposition & The F Ratio

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i\beta_j + \varepsilon_{ij}$$

- Break down sum of squares:

$$= \sum_{i,j}(\lambda\alpha_i\beta_j)^2$$

$$SS_A = b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2,\ SS_B = a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2, SS_{AB} = \frac{\left(\sum_{i,j}(\bar{y}_{i.}-\bar{y}_{..})(\bar{y}_{.j}-\bar{y}_{..})y_{ij}\right)^2}{\left(\sum_i(\bar{y}_{i.}-\bar{y}_{..})^2\right)\left(\sum_j(\bar{y}_{.j}-\bar{y}_{..})^2\right)}$$

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

$$ab - 1 = (a-1) + (b-1) + 1 + \underline{(ab - a - b)}$$

- Tukey's test of additivity:

$$H_0\!: \lambda = 0 \text{ (i.e., no interaction) vs } H_1\!: \lambda \neq 0$$

- Test statistic, F ratio:

$$F_0 = \frac{SS_{AB}/1}{MS_E} \sim F_{1,ab-a-b}$$

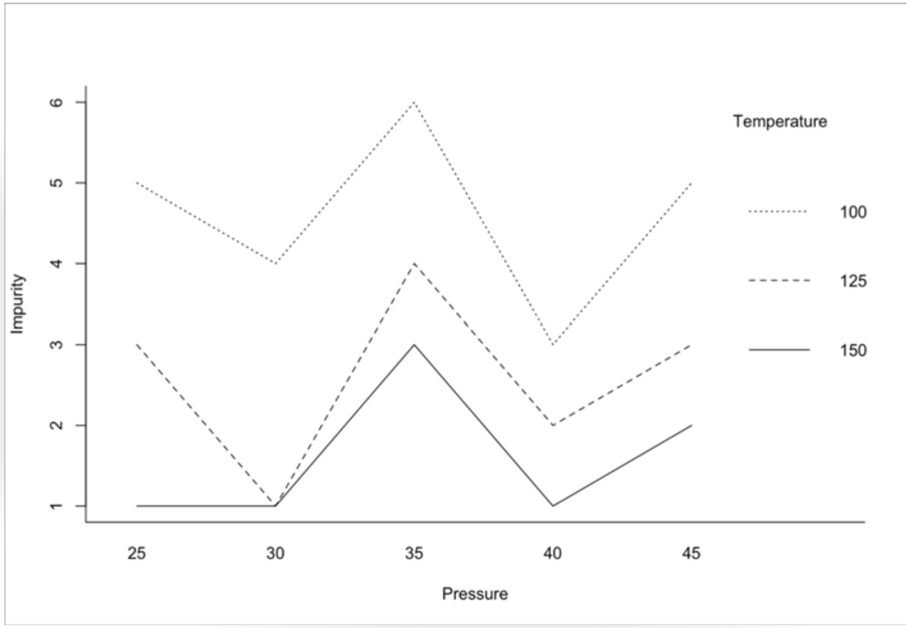
- Decision rule: Reject H_0 if

$$F_0 > F_{1-\alpha,\ 1,\ ab-a-b}$$

> 👉 It is important to remember that "$\alpha_i\beta_j$" is a covariate. The test for the "significance" of $\alpha_i\beta_j$ forms the test for interaction

Impurity Causes Check

- Interaction plot



- R implementation

```
> library(additivityTests)
> response <-matrix(impurity, nr = nlevels(temp))

> tukey.test(response)
```

Tukey test on 5% alpha-level:

Test statistic: 0.3627
Critival value: 5.591
The additivity hypothesis cannot be rejected.

Experiments with Three Factors

- **Elements:**
  - Factor *A* with *a* levels
  - Factor *B* with *b* levels
  - Factor *C* with *c* levels
  - Response *Y*
- **Potential effects:**
  - Main effects of *A*, *B* & *C*
  - Two-factor interactions *AB, AC, BC*
  - Three-factor interaction *ABC*
- **Goals:**
  - Evaluate the impact of factor *A, B* & *C* to response *Y*
  - Find the best combination of (*A,B,C*)
- **Design Methods:**
  - One factor at a time
  - Full factorial design
  - Latin square design

ANOVA Table for a General Three-Way Layout

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \delta_k + (\alpha\beta)_{ij} + (\alpha\delta)_{ik} + (\beta\delta)_{jk} + \gamma_{ijk} + \varepsilon_{ijkl}$$

Source	df	Sum of Squares
A	$I-1$	$\sum_{i=1}^{I} nJK(\hat{\alpha}_i)^2$
B	$J-1$	$\sum_{j=1}^{J} nIK(\hat{\beta}_j)^2$
C	$K-1$	$\sum_{k=1}^{K} nIJ(\hat{\delta}_k)^2$
$A \times B$	$(I-1)(J-1)$	$\sum_{i=1}^{I} \sum_{j=1}^{J} nK(\widehat{(\alpha\beta)}_{ij})^2$
$A \times C$	$(I-1)(K-1)$	$\sum_{i=1}^{I} \sum_{k=1}^{K} nJ(\widehat{(\alpha\delta)}_{ik})^2$
$B \times C$	$(J-1)(K-1)$	$\sum_{j=1}^{J} \sum_{k=1}^{K} nI(\widehat{(\beta\delta)}_{jk})^2$
$A \times B \times C$	$(I-1)(J-1)(K-1)$	$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n(\hat{\gamma}_{ijk})^2$
residual	$IJK(n-1)$	$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{n} (y_{ijkl} - \bar{y}_{ijk\cdot})^2$
total	$IJKn-1$	$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{n} (y_{ijkl} - \bar{y}_{\cdots})^2$

Example: The Soft Drink Bottling Problem

- A bottler is interested in obtaining more uniform fill heights in the bottles
- The filling machine is expected to fill each bottle to the correct target height. The bottler would like to understand the sources of the variability better and eventually reduce it
- The process engineer can control three variables during the filling process:
  - the percent carbonation (A), difficult to control, 3 levels
  - the operating pressure in the filler (B), 2 levels
  - the bottles produced per minute or the line speed (C), 2 levels
- Two replicates are run of a factorial design in these three factors, with all 24 runs taken in random order
- The response variable observed is the average deviation from the target fill height

Experiment Result

- Positive deviations are fill heights above the target, whereas negative deviations are fill heights below the target
- The circled numbers are the three-way cell totals

	Operating Pressure (B)								
	25 psi				30 psi				
	Line Speed (C)				Line Speed (C)				
Percent Carbonation (A)	**200**		**250**		**200**		**250**		$y_{i...}$
10	−3 −1	−4	−1 0	−1	−1 0	−1	1 1	2	−4
12	0 1	1	2 1	3	2 3	5	6 5	11	20
14	5 4	9	7 6	13	7 9	16	10 11	21	59
$B \times C$ Totals $y_{.jk.}$	6		15		20		34		$75 = y_{....}$
$y_{.j..}$	21				54				

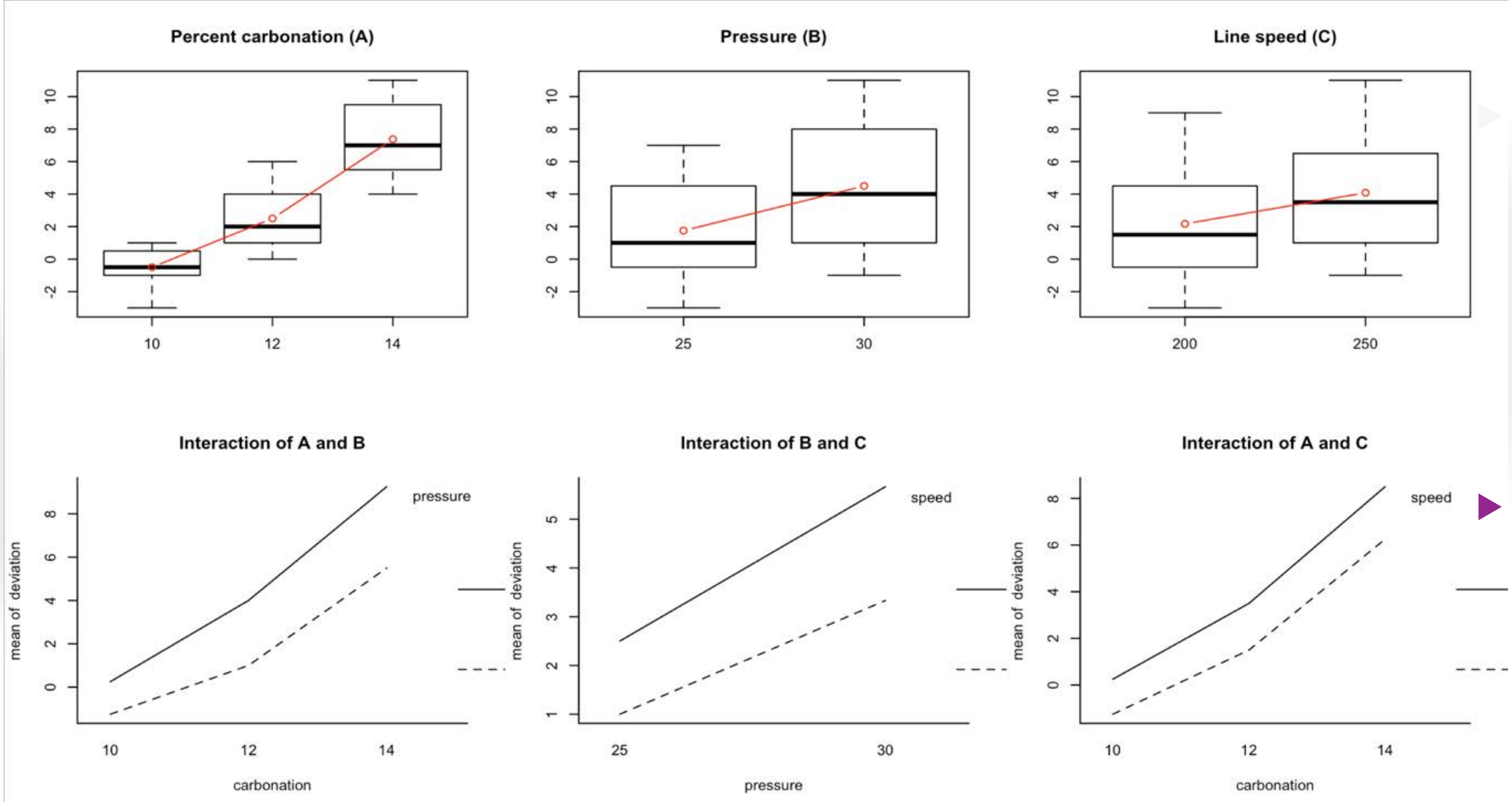
$A \times B$ Totals $y_{ij..}$

A	B = 25	30
10	−5	1
12	4	16
14	22	37

$A \times C$ Totals $y_{i.k.}$

A	C = 200	250
10	−5	1
12	6	14
14	25	34

Plots of Factor Effects & Interactions



- All three variables have positive main effects
- The interactions are fairly small

ANOVA Result

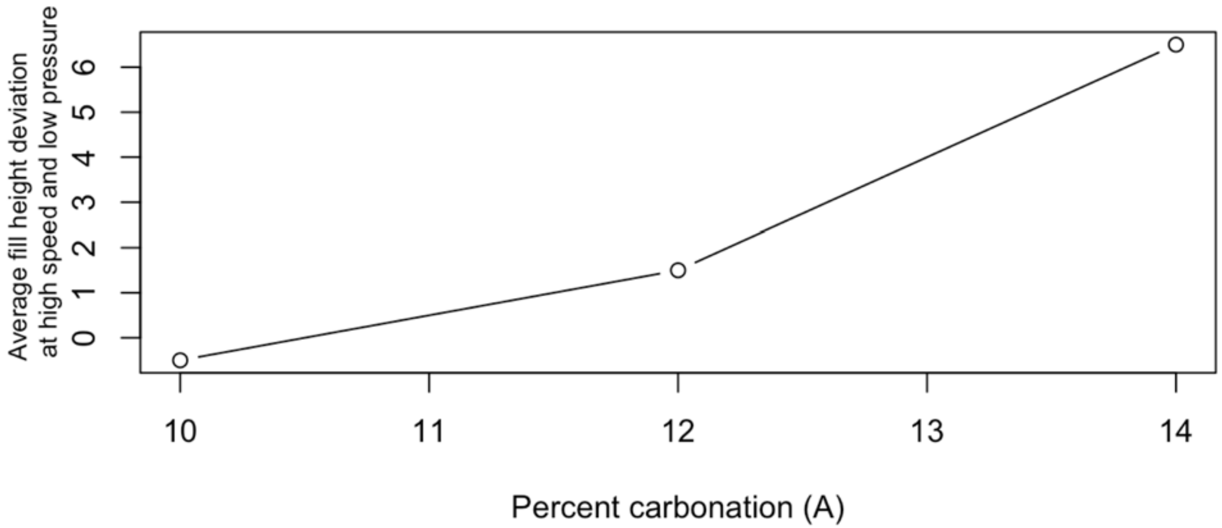
- > summary(aov(deviation ~ carbonation * pressure * speed, data=Bottle))

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
carbonation	2	252.75	126.38	178.412	1.19e-09 ***
pressure	1	45.38	45.38	64.059	3.74e-06 ***
speed	1	22.04	22.04	31.118	0.00012 ***
carbonation:pressure	2	5.25	2.63	3.706	0.05581 .
carbonation:speed	2	0.58	0.29	0.412	0.67149
pressure:speed	1	1.04	1.04	1.471	0.24859
carbonation:pressure:speed	2	1.08	0.54	0.765	0.48687
Residuals	12	8.50	0.71		

- Don’t forget analysis of the residuals

Plot for a Specific Setting

- Since high level of line speed (250 bpm) is preferred to maximize the production rate, and the percent carbonation is difficult to control, we recommend the low level of operating pressure (25 psi)
- Average fill height deviation at high speed and low pressure for different carbonation levels



- Better have a small percent carbonation, and control the variability in its levels

Blocking in a Factorial Design

- We usually treat nuisance factors as blocks
- Sometimes, it is not practical to completely randomize all of the runs in a factorial

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij}$$

- For example, the presence of a nuisance factor may require that the experiment be run in blocks
- We should assume that interaction between blocks and treatments is negligible

Split plot with 4 blocks , Factor A is 2 levels, Factor B is 5 levels

B1	B5	B2	B4	B3		Mainplot A1
B3	B2	B4	B1	B5		Mainplot A2
B1	B5	B2	B4	B3		Mainplot A2
B3	B2	B4	B1	B5		Mainplot A1
B1	B5	B2	B4	B3		Mainplot A1
B3	B2	B4	B1	B5		Mainplot A2
B1	B5	B2	B4	B3		Mainplot A2
B3	B2	B4	B1	B5		Mainplot A1

Latin Square Design

A	B	C
B	C	A
C	A	B

- A Latin square design of order *k*
  - *k* Latin letters (i.e., treatments)
  - Each appears once in each row and once in each column

- Latin square design
  - It is used to eliminate two nuisance sources of variability
  - One treatment factor + two blocking factors
  - It is an extension of RBD to accommodate two blocking factors
  - Randomization applied to assignments to rows, columns, treatments

Example: Wear Experiment

- Testing the abrasion resistance of rubber-covered fabric
- Y = loss in weight over a period of time
- One treatment factor: Material type A, B, C, D
- Two blocking factors
  - Four positions on the tester
  - Four applications(four different times for setting up the tester)

	Position			
Application	1	2	3	4
1	C	D	B	A
2	A	B	D	C
3	D	C	A	B
4	B	A	C	D

ANOVA Table for Latin Square Design

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$$

Source	Degrees of Freedom	Sum of Squares
row	$k-1$	$k\sum_{i=1}^{k}(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$
column	$k-1$	$k\sum_{j=1}^{k}(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$
treatment	$k-1$	$k\sum_{l=1}^{k}(\bar{y}_{\cdot\cdot l} - \bar{y}_{\cdot\cdot\cdot})^2$
residual	$(k-1)(k-2)$	$\sum_{(i,j,l)\in S}(y_{ijl} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot l} + 2\bar{y}_{\cdot\cdot\cdot})^2$
total	k^2-1	$\sum_{(i,j,l)\in S}(y_{ijl} - \bar{y}_{\cdot\cdot\cdot})^2$

Graeco-Latin Square Design for Four Factors

- Two Latin squares are orthogonal if each pair of letters appears once in the two squares, when superimposed
- The superimposed square is called a Graeco-Latin square

$A\alpha$	$B\beta$	$C\gamma$
$B\gamma$	$C\alpha$	$A\beta$
$C\beta$	$A\gamma$	$B\alpha$

Aα	Bγ	Cδ	Dβ
Bβ	Aδ	Dγ	Cα
Cγ	Dα	Aβ	Bδ
Dδ	Cβ	Bα	Aγ

Aα	Bδ	Cβ	Dε	Eγ
Bβ	Cε	Dγ	Eα	Aδ
Cγ	Dα	Eδ	Aβ	Bε
Dδ	Eβ	Aε	Bγ	Cα
Eε	Aγ	Bα	Cδ	Dβ

- Useful for studying four factors (1 treatment, 3 blocking factors; or 2 treatment, 2 blocking factors, etc.) allowing one more factor to be studied
- They are restricted, however, to the case in which all the factors have the same number of levels

ANOVA Table for Graeco-Latin Square Design

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \tau_l + \varepsilon_{ijkl}$$

Source	Degrees of Freedom	Sum of Squares
row	$k-1$	$k\sum_{i=1}^{k}(\bar{y}_{i\cdots} - \bar{y}_{\cdots\cdot})^2$
column	$k-1$	$k\sum_{j=1}^{k}(\bar{y}_{\cdot j\cdot\cdot} - \bar{y}_{\cdots})^2$
Latin letter	$k-1$	$k\sum_{l=1}^{k}(\bar{y}_{\cdot\cdot l\cdot} - \bar{y}_{\cdots\cdot})^2$
Greek letter	$k-1$	$k\sum_{m=1}^{k}(\bar{y}_{\cdots m} - \bar{y}_{\cdots\cdot})^2$
residual	$(k-3)(k-1)$	by subtraction
total	k^2-1	$\sum_{(i,j,l,m)\in S}(y_{ijlm} - \bar{y}_{\cdots\cdot})^2$