Chapter 6
Randomized Block Design
Two Factor ANOVA
Interaction in ANOVA

Two factor (two-way) ANOVA

Two-factor ANOVA is used when:

- Y is a quantitative response variable
- There are two categorical explanatory variables, called Factors:
  - Factor A has K levels, $k = 1, \ldots, K$
  - Factor B has J levels, $j = 1, \ldots, J$
- The combination of level k for A and level j for B has sample size n_{kj} but if all equal, just use n.
- Use N for overall sample size.

Special case: Using "BLOCKS"

Definition: A <u>block</u> is a group of similar units, or the same unit measured multiple times.

Blocks are used to reduce known sources of variability, by comparing levels of a factor <u>within</u> blocks.

Examples (explained in detail in class):

- Factor = 3 methods of reducing blood pressure; Blocks defined using initial blood pressure.
- Factor = 4 methods for enhancing memory; Blocks defined by age.
- Factor = Impairment while driving (alcohol, marijuana, no sleep, control); Blocks = individuals.

Simple Block Design, all $n_{kj} = 1$

A simple block design has two factors with:

- Exactly one data value (observation) in each combination of the factors.
- Factor A is factor of interest, called *treatment*
- Factor B, called *blocks*, used to control a known source of variability

Main interest is comparing levels of the *treatment*.

Notation: Factor A (Treatments) has K levels
Factor B (Blocks) has J levels
$\rightarrow$ $N = KJ$ data values

Example: Do Means Differ for 4 Exam Formats?

	Adam	Brenda	Cathy	Dave	Emily	Mean
Exam #1:	62	94	68	86	50	72
Exam #2:	87	95	93	97	63	87
Exam #3:	74	86	82	70	28	68
Exam #4:	77	89	73	79	47	73
Mean	75	91	79	83	47	75

Treatments: 4 different exam formats, Blocks: 5 different students

Question: Is there a difference in population means for the 4 exams?

Use students as *blocks* because we know student abilities differ. Controls for that *known* source of variability.

Two-way ANOVA: Main Effects Model
Shown here for simple block design

$$Y = \mu + \alpha_k + \beta_j + \varepsilon$$

- μ: Grand mean — Mean of all exams for all students
- α_k: Effect for k^{th} treatment — How mean for k^{th} exam differs from overall mean
- β_j: Effect for j^{th} block — How mean for j^{th} student differs from overall mean
- ε: Random error

Randomized Block—Calculations

1. Find the mean for each treatment (row means), each block (column means), and grand mean.
2. Partition the *SSTotal* into three pieces:

$$SSTotal = SSA + SSB + SSE$$

$$SSTotal = \sum (y - \bar{y})^2 = (n-1)s_Y^2$$ (As usual)

$$SSA = \sum J(\bar{y}_k - \bar{y})^2$$ Compare row means (exams)

$$SSB = \sum K(\bar{y}_j - \bar{y})^2$$ Compare column means (students)

$$SSE = SSTotal - SSA - SSB$$ (Unexplained error)

Randomized Block ANOVA Table

Source	d.f.	S.S.	M.S.	t.s.	p-value
Trts = A	K-1	$SSTr$	$SSTr/(K\text{-}1)$	MSTr/MSE	
Blocks	J-1	SSB	$SSB/(J\text{-}1)$	MSB/MSE	
Error	$(K\text{-}1)(J\text{-}1)$	SSE	$SSE/(K\text{-}1)(J\text{-}1)$		
Total	N-1	$SSTotal$			

Testing TWO hypotheses:

H_0: $\alpha_1 = \alpha_2 = \ldots = \alpha_K = 0$

H_a: Some $\alpha_k \neq 0$

(Factor A: Difference in treatment means)

H_0: $\beta_1 = \beta_2 = \ldots = \beta_J = 0$

H_a: Some $\beta_j \neq 0$

(Factor B: Difference in block means)

ANOVA Output in *R*

```
> BlockMod=aov(Grade~as.factor(Exam)+Student)
> summary(BlockMod)
                Df Sum Sq Mean Sq F value    Pr(>F)
as.factor(Exam)  3   1030  343.33  5.7222   0.01144 *
Student          4   4480 1120.00 18.6667 4.347e-05 ***
Residuals       12    720   60.00
```

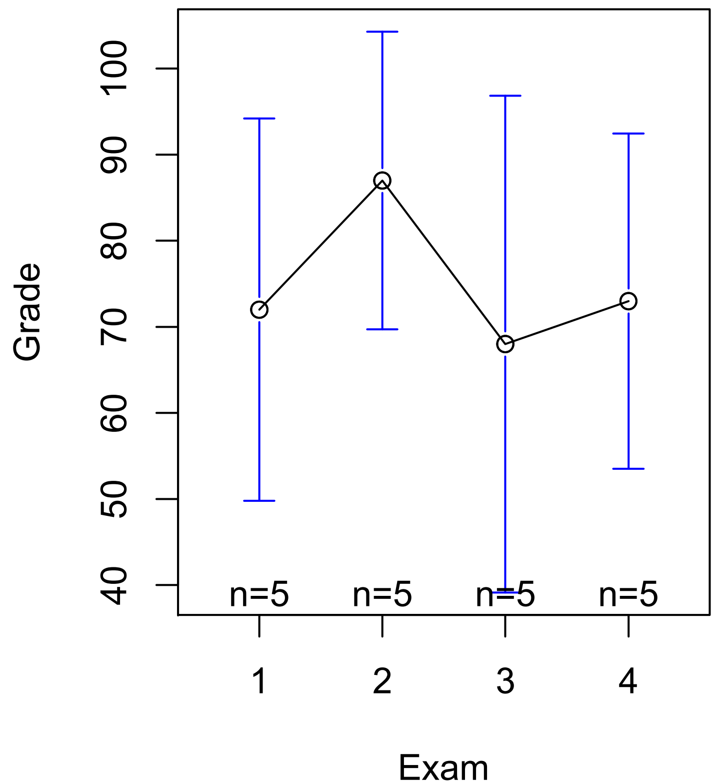
What if we ignored Blocks (Students) and treated it as a one-factor ANOVA? (See Lecture 15 – didn't take into account blocks!)

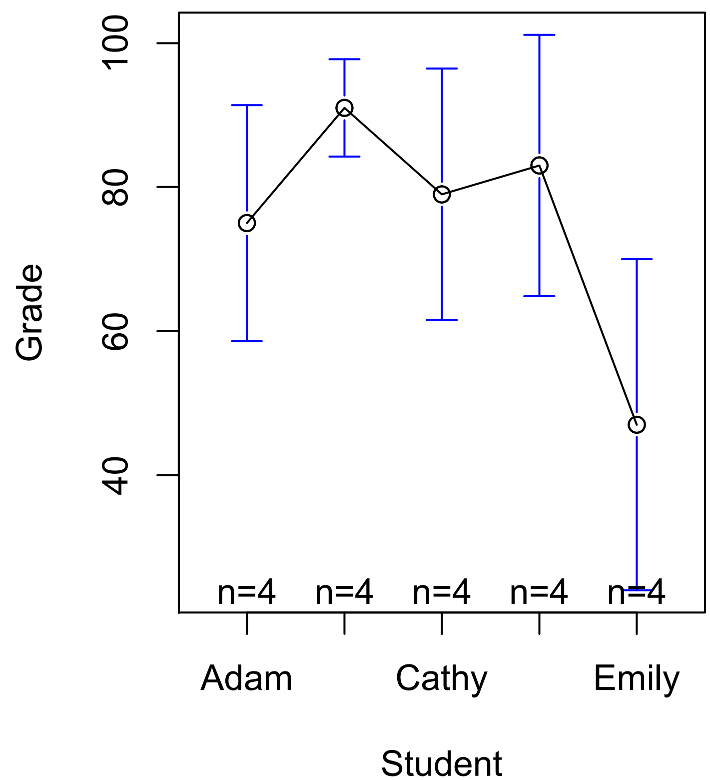
```
> model=aov(Grade~as.factor(Exam))
> summary(model)
                Df Sum Sq Mean Sq F value Pr(>F)
as.factor(Exam)  3 1030.0   343.3  1.0564  0.395
Residuals       16 5200.0   325.0
```

Ignoring "student effect," exams don't seem to differ; but including student effect, exams do differ. SS(Student) becomes part of SSE if Blocks are ignored, which *inflates* the estimate of the standard deviation.

After Installing Three Packages in *R*: gplots, gdata, gtools

```
> plotmeans(Grade~Exam)
> plotmeans(Grade~Student)
```





95% CI's for each group mean are shown in blue.

Fisher's LSD CIs After Two-Way ANOVA in a Simple Block Design

Same as one-way, but we know that

$$\frac{1}{n_i} + \frac{1}{n_j} = \begin{cases} 2/J & \text{for row means} \\ 2/K & \text{for column means} \end{cases}$$

For treatment (row) means:

$$LSD = t^* \sqrt{MSE} \sqrt{\frac{2}{J}}$$

For block (column) means:

$$LSD = t^* \sqrt{MSE} \sqrt{\frac{2}{K}}$$

Two-way ANOVA model with Interaction

$$Y = \mu + \alpha_k + \beta_j + \gamma_{kj} + \varepsilon$$

Grand mean

Main effect for Factor A

Main effect for Factor B

Interaction effect

Random error

What's an Interaction Effect?

An *interaction effect* occurs when differences in mean level effects for one factor *depend* on the level of the other factor.

Example: Y = GPA
Factor A = Year in School (FY, So, Jr, Sr)
Factor B = Major (Psych, Bio, Math)

FY is hard. $\Rightarrow \alpha_1 < 0$ (Main effect)

Bio is easy. $\Rightarrow \beta_2 > 0$ (Main effect)

Jr in Math is harder than just Jr or just Math $\Rightarrow \gamma_{33} < 0$ (Interaction effect)

Example

Fire extinguishers tested to see how quickly they put out fires.

Factor A: 3 different chemicals in the extinguishers A_1, A_2, A_3

Factor B: 2 types of fires, B_1 = wood, B_2 = gas

Y_{kj} = time to put out the fire of type B_j with chemical A_k

Questions of interest:

- Do the 3 chemicals differ in mean time required? (If so, there is a *Factor A effect.*)
- Does mean time to put out fire depend on the type of fire? (If so, there is a *Factor B effect.*)
- Do the differences in times for the 3 chemicals depend on the type of fire? (If so, there is an *interaction* between chemical type and fire type.)

Example: Putting out fires

Factor A: Chemical (A_1, A_2, A_3)
Factor B: Fire type (*wood*, *gas*)
Response: Time until fire is completely out (in seconds)

Data:	Wood (j=1)	Gas (j=2)
A1 (k=1)	**52 64**	**72 60**
A2 (k=2)	**67 55**	**78 68**
A3 (k=3)	**86 72**	**43 51**

$K = 3$
$J = 2$
$n = 2$
$N = 12$

Interpreting Interaction

Cell means plot (Interaction plot)

Data:	Wood	Gas
A1	**58.0**	**66.0**
A2	**61.0**	**73.0**
A3	**79.0**	**47.0**

Interaction Plot via *R*

Generic

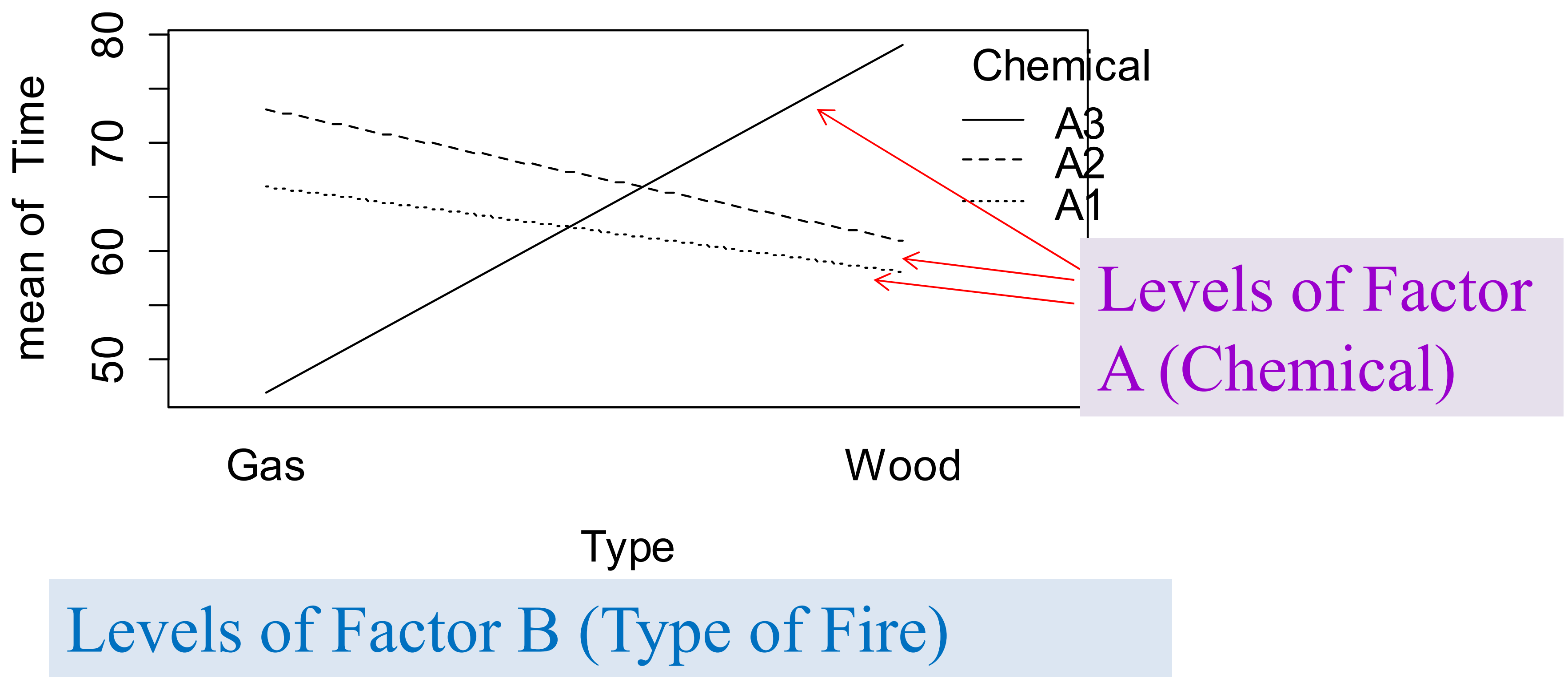
```
> interaction.plot(FactorA,FactorB,Y)
```

```
> interaction.plot(Chemical,Type,Time)
```

OR

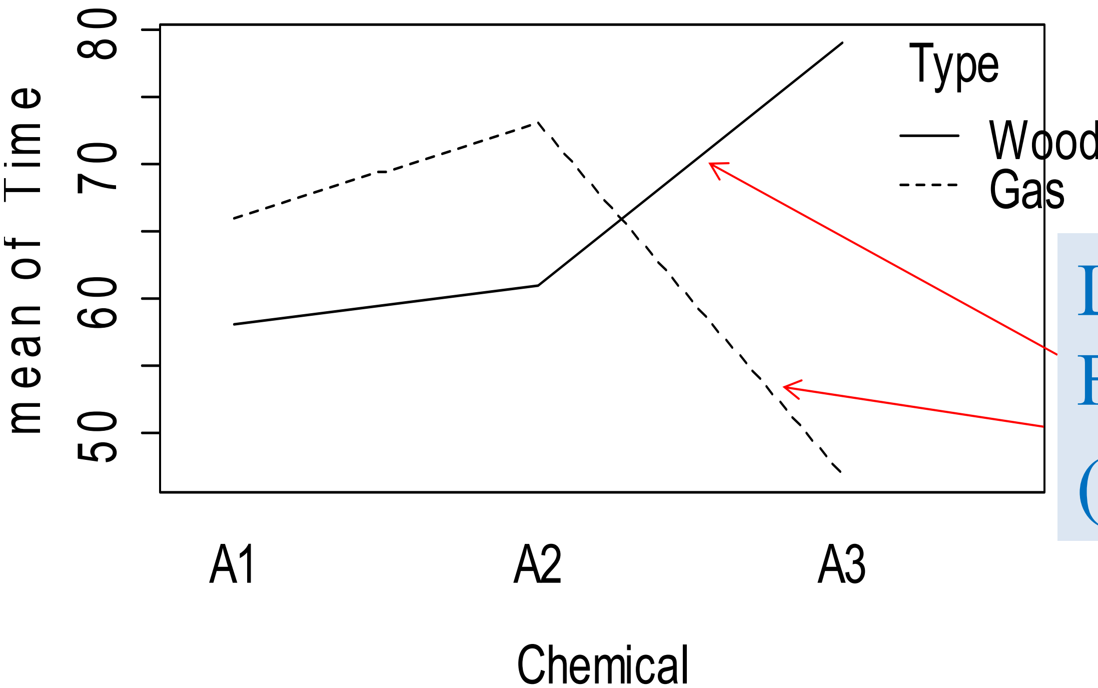
```
> interaction.plot(Type,Chemical, Time)
```

```
> interaction.plot(Type,Chemical,Time)
```



Interpretation discussed in class.

```
> interaction.plot(Chemical,Type,Time)
```



Levels of Factor B

(Type of Fire)

Levels of Factor A (Chemical)

Interpretation discussed in class.

Two-way ANOVA (with Interaction)

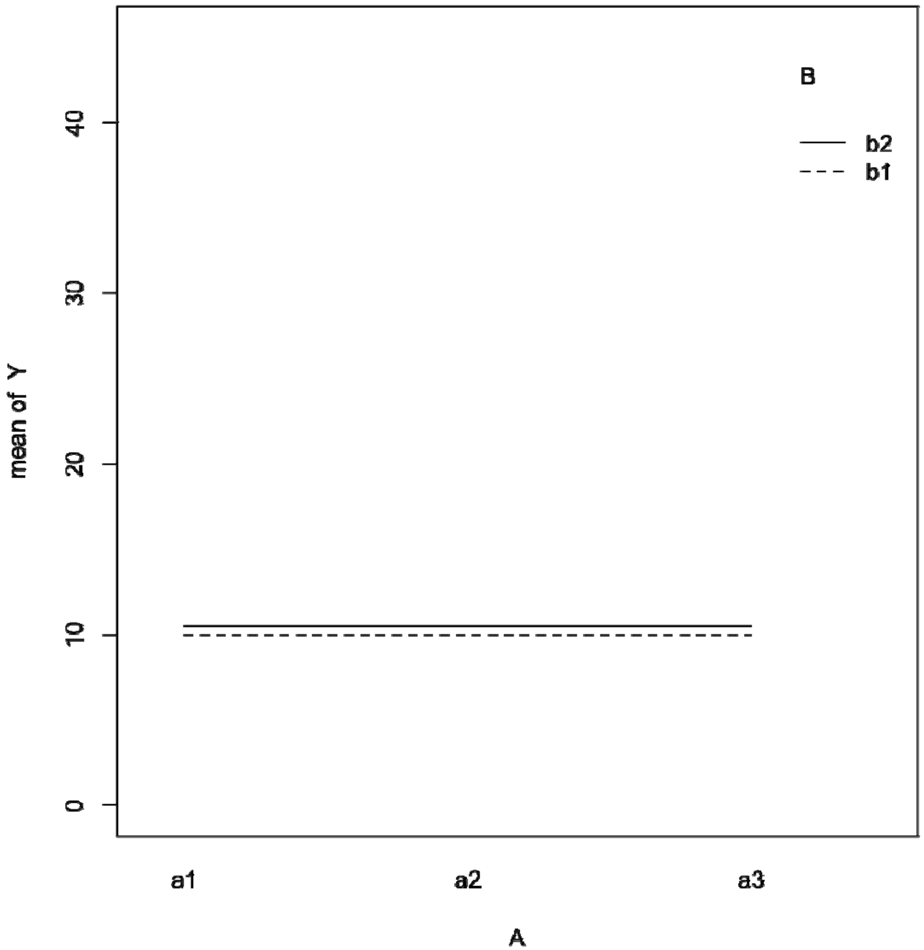
$$Y = \mu + \alpha_k + \beta_j + \gamma_{kj} + \varepsilon$$

- μ: Grand mean
- α_k: Main effect for Factor A
- β_j: Main effect for Factor B
- γ_{kj}: Interaction effect
- ε: Random error

Generic Interaction Plots: 7 Cases

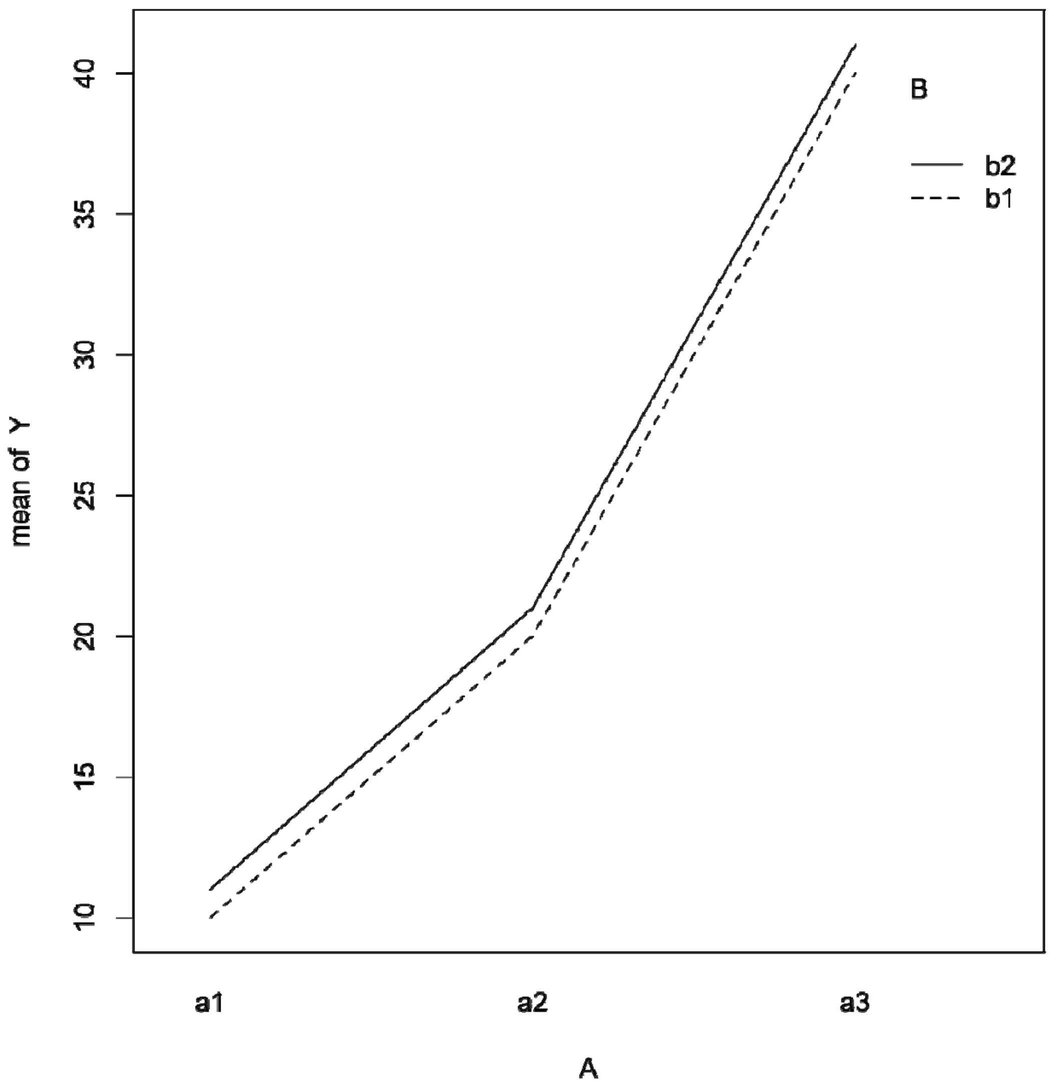
Suppose A has three levels and B has two levels.

Grand mean only (no treatment effects)



$$Y = \mu + \varepsilon$$

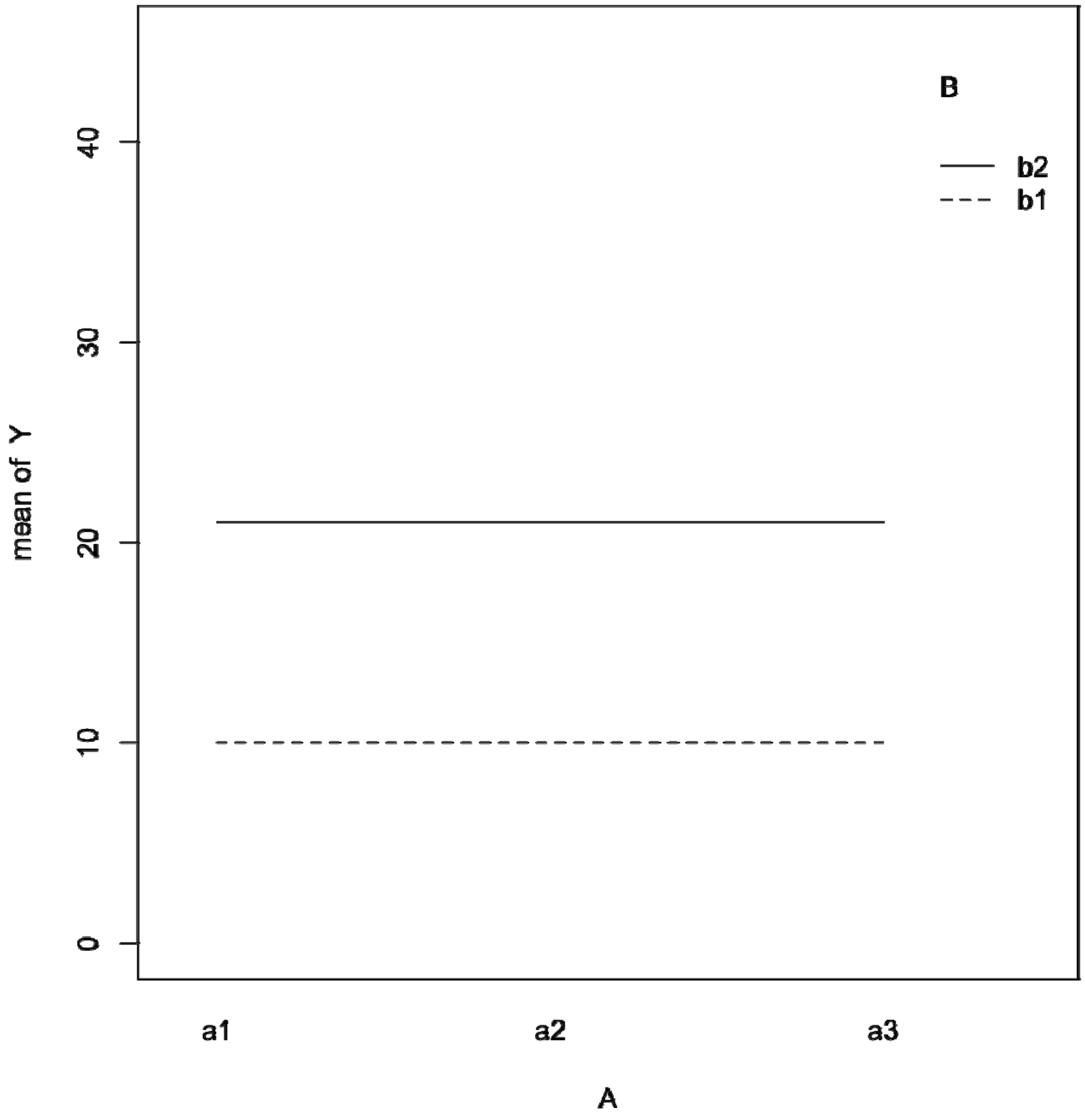
Treatment A Effect, No Treatment B Effect



$$Y = \mu + \alpha_k + \varepsilon$$

Means differ for levels of A, *not* for levels of B.

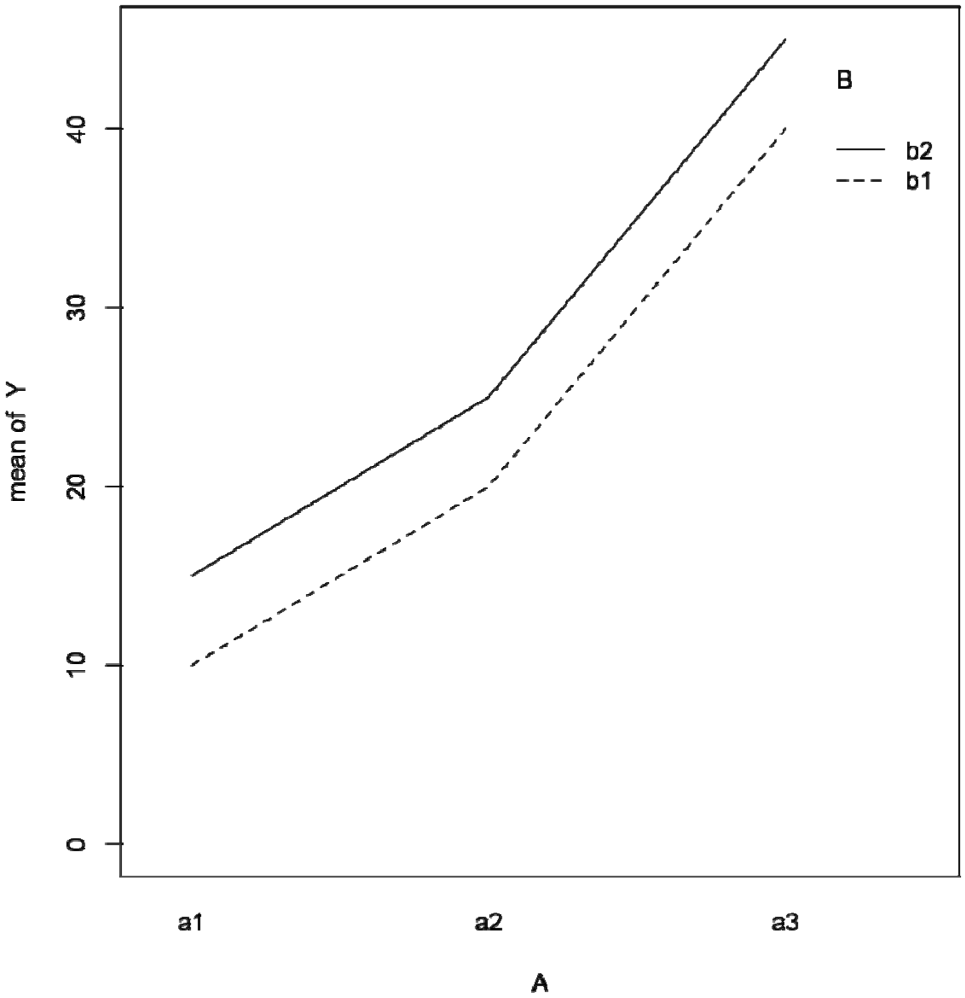
Treatment B Effect, No Treatment A Effect



$$Y = \mu + \beta_j + \varepsilon$$

Means differ for levels of B, *not* for levels of A.

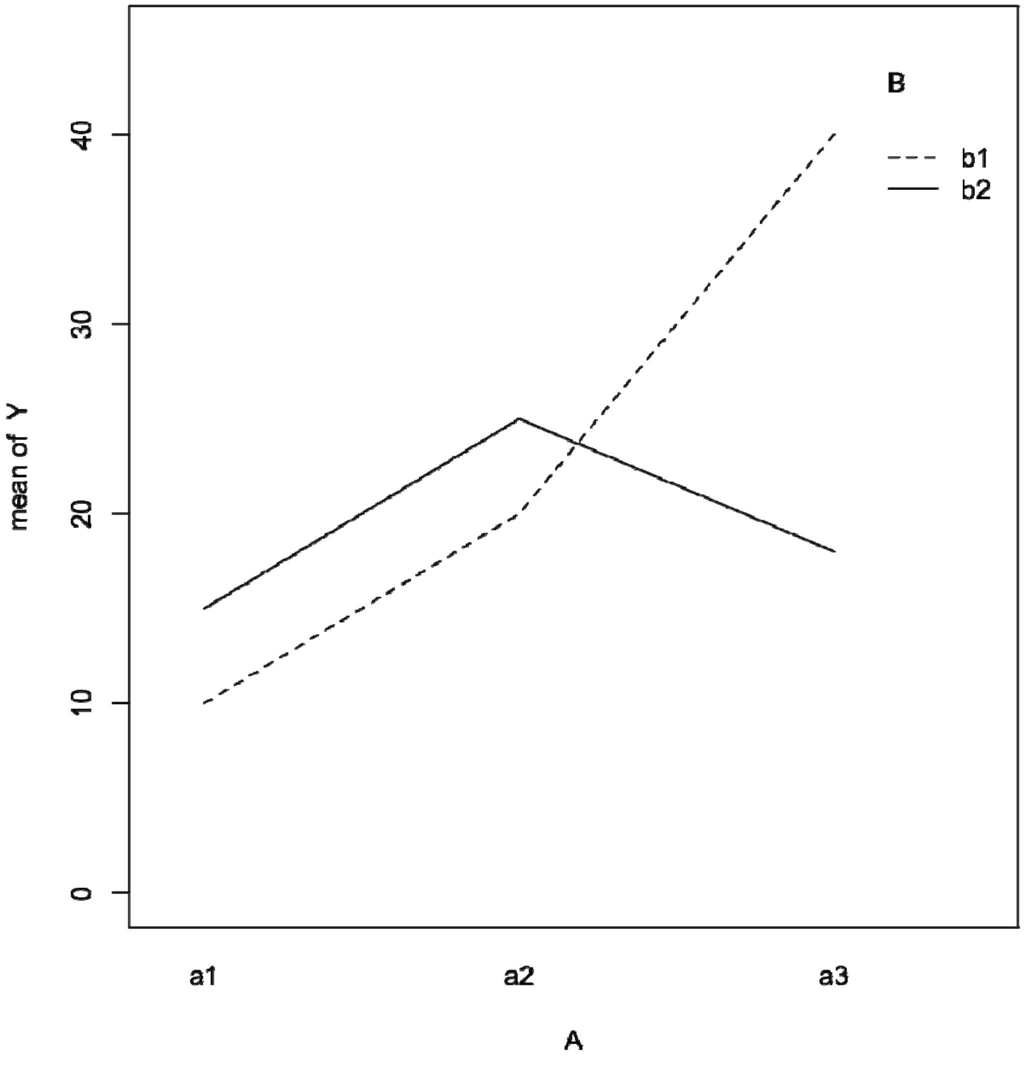
Treatments A and B Have Effects, No Interaction



$$Y = \mu + \alpha_k + \beta_j + \varepsilon$$

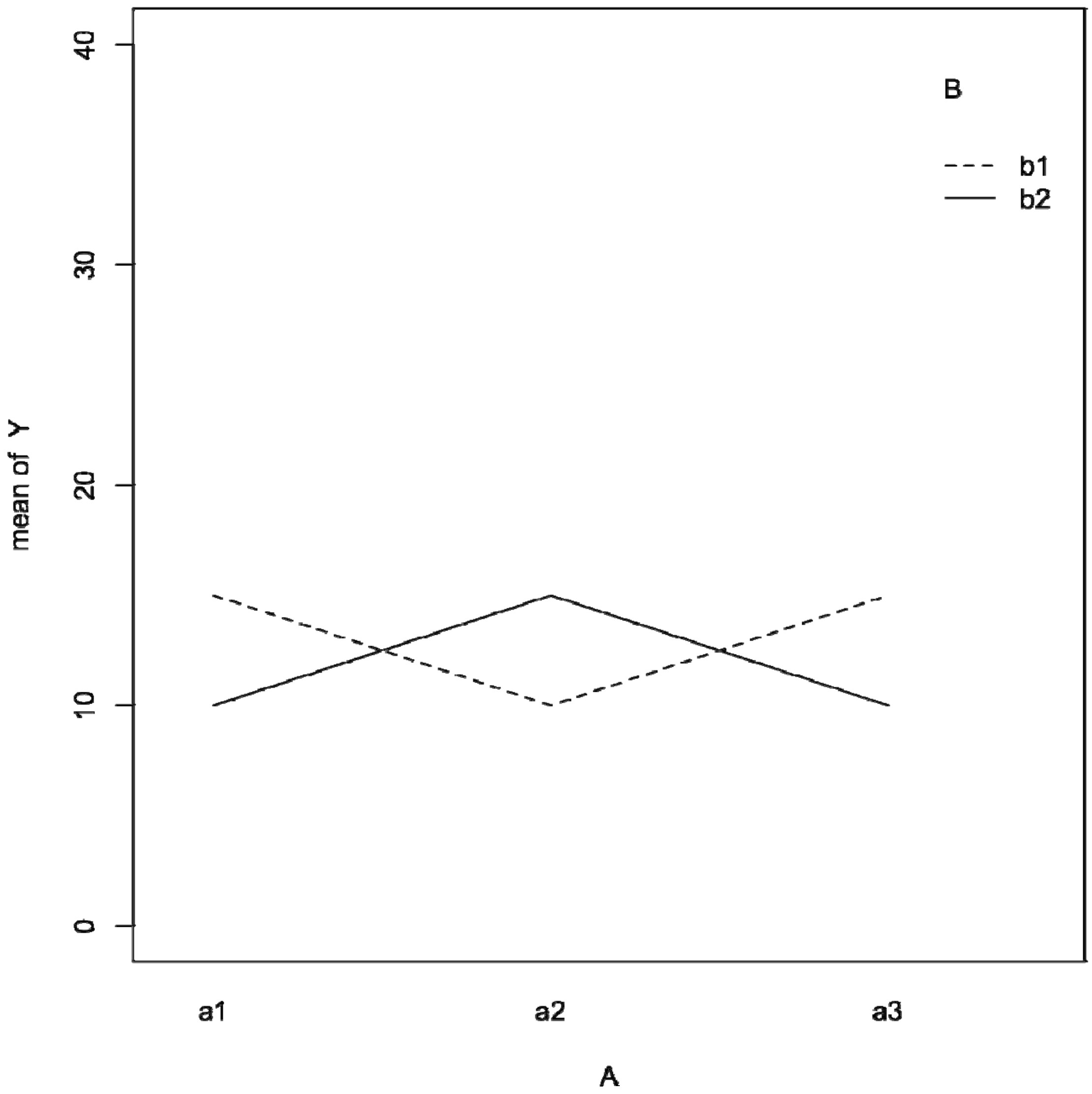
Differences between the 2 levels of Factor B don’t depend on level of Factor A (and vice versa).

A and B Effects Plus Interaction



$$Y = \mu + \alpha_k + \beta_j + \gamma_{kj} + \varepsilon$$

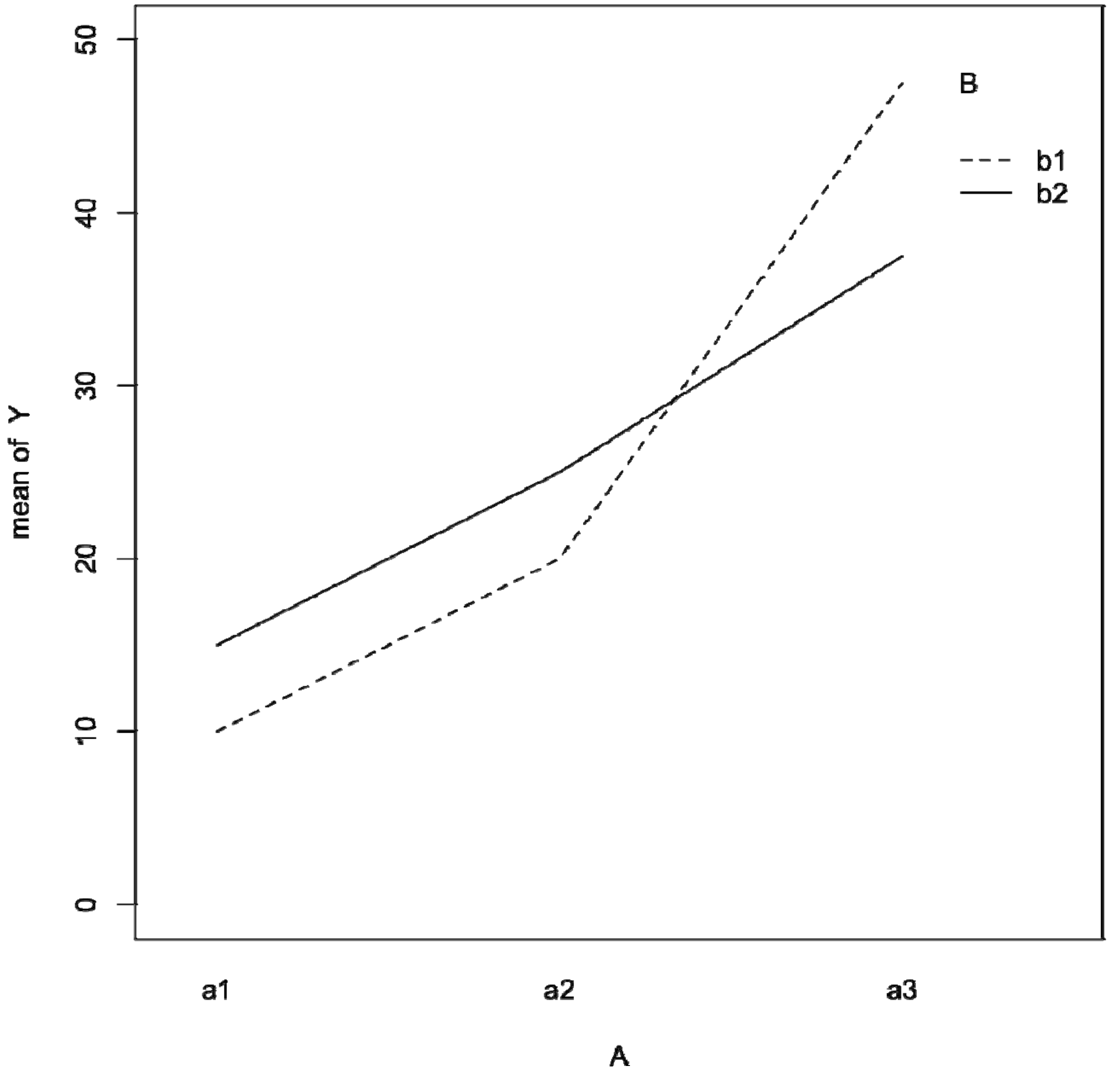
No A Effect but a B Effect Plus Interaction



$$Y = \mu + \beta_j + \gamma_{kj} + \varepsilon$$

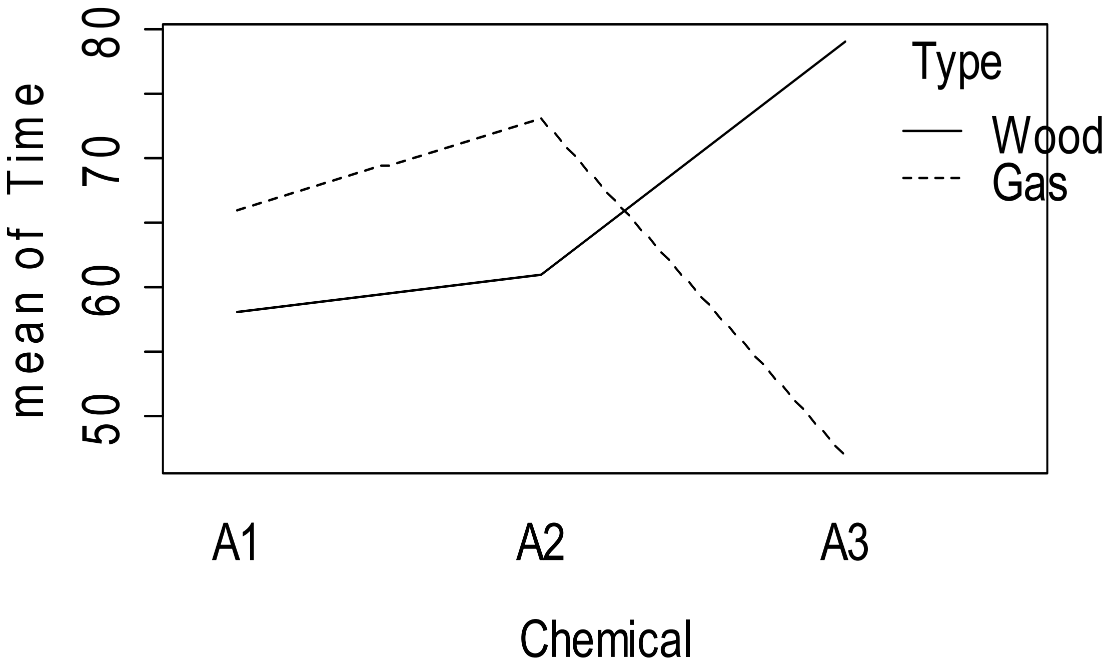
Cannot say means of A do not differ! They do, at *each* level of B.

A Effect but No B Effect Plus Interaction



$$Y = \mu + \alpha_k + \gamma_{kj} + \varepsilon$$

Cannot say means of B do not differ! They do, at *each* level of A.



Looks like a *significant* interaction

No interaction: lines would be parallel.

Interaction: Differences in A *depend* on level of B.

However, looks like almost no *overall* A effect or B effect!

Chapter 6 Section 6.3
The Gory Details!

Recall: Main Effects Model

$$Y = \mu + \alpha_k + \beta_j + \varepsilon$$

Grand Mean

Effect for k^{th} level of Factor A

Effect for j^{th} level of Factor B

Random error

Factorial Anova—Example: Putting out fires

Factor A: Chemical (A1, A2, A3)
Factor B: Fire type (*wood, gas*)
Response: Time required to put out fire (seconds)

Data:	Wood	Gas	Row mean
A1	**52 64**	**72 60**	**62**
A2	**67 55**	**78 68**	**67**
A3	**86 72**	**43 51**	**63**
Col mean	**66**	**62**	

Two-way ANOVA (with Interaction)

$$Y = \mu + \alpha_k + \beta_j + \gamma_{kj} + \varepsilon$$

Grand mean

Main effect for Factor A

Main effect for Factor B

Interaction effect

Random error

Factorial Design

Assume:

Factor A has K levels, Factor B has J levels.

To estimate an interaction effect, we need *more than one* observation for each combination of factors.

Let n_{kj} = sample size in $(k,j)^{th}$ cell.

Definition: For a balanced design, n_{kj} is constant for all cells.

$$n_{kj} = n$$

$n = 1$ in a typical randomized block design

$n > 1$ in a balanced factorial design

Fire Extinguishers

Factor A: Chemical (*A1, A2, A3*)
Factor B: Fire type (*wood, gas*)
Response: Time required to put out fire (seconds)

Data:	Wood	Gas
A1	**52 64**	**72 60**
A2	**67 55**	**78 68**
A3	**86 72**	**43 51**

$K = 3$
$J = 2$
$n = 2$
$N = 12$

Estimating Factorial Effects

$\bar{y}_{kj}$ = mean for $(k, j)^{th}$ cell

$\bar{y}_k$ = mean for k^{th} row

$\bar{y}_j$ = mean for j^{th} column

$\bar{y}$ = Grand mean

$$y = \mu + \alpha_k + \beta_j + \gamma_{kj} + \varepsilon$$

$$(y - \bar{y}) = (\bar{y}_k - \bar{y}) + (\bar{y}_j - \bar{y}) + (\bar{y}_{kj} - \bar{y}_k - \bar{y}_j + \bar{y}) + (y - \bar{y}_{kj})$$

Total = Factor A + Factor B + Interaction + Error

$$SSTotal = SSA + SSB + SSAB + SSE$$

Partitioning Variability (Balanced)

$$SSTotal = \sum (y - \bar{y})^2 = (N-1)s_Y^2$$ (As usual)

$$SSA = \sum_k Jn(\bar{y}_k - \bar{y})^2$$ (Row means)

$$SSB = \sum_j Kn(\bar{y}_j - \bar{y})^2$$ (Column means)

$$SSAB = \sum_{k,j} n(\bar{y}_{kj} - \bar{y}_k - \bar{y}_j + \bar{y})^2$$ (Cell means)

$$SSE = \sum \left(y - \bar{y}_{kj}\right)^2 = SSTotal - SSA - SSB - SSAB$$

$$SSTotal = SSA + SSB + SSAB + SSE$$ (Error)

Total = Factor A + Factor B + Interaction + Error

Decomposition: Fire extinguishers

Data:	Wood	Gas
A1	**52 64**	**72 60**
A2	**67 55**	**78 68**
A3	**86 72**	**43 51**

Cell Means:	Wood	Gas	Row mean	Trt A effect
A1	**58.0**	**66.0**	**62**	**−2**
A2	**61.0**	**73.0**	**67**	**+3**
A3	**79.0**	**47.0**	**63**	**−1**
Col mean	**66**	**62**	***64***	
Trt B effect	**+2**	**−2**		

Interaction effects	Wood	Gas	Row mean	Trt A effect
A1	**−6**	**6**	**62**	**−2**
A2	**−8**	**8**	**67**	**+3**
A3	**14**	**−14**	**63**	**−1**
Col mean	**66**	**62**	***64***	
Trt B effect	**−2**	**−2**		

$$58 - 66 - 62 + 64 = -6$$

Decomposition: Fire Extinguishers

Top left cell: **52,64** Top right cell: **72,60**

Observed Value		Grand Mean		Trt A Effect		Trt B Effect		Inter-action		Residual
52	=	64	+	-2	+	2	+	-6	+	-6
64	=	64	+	-2	+	2	+	-6	+	6
72	=	64	+	-2	+	-2	+	6	+	6
60	=	64	+	-2	+	-2	+	6	+	-6

Etc.

Two-way ANOVA Table (with Interaction)

Source	d.f.	S.S.	M.S.	t.s.	p
Factor A	$K-1$	SSA	$SSA/(K-1)$	**MSA/MSE**	
Factor B	$J-1$	SSB	$SSB/(J-1)$	**MSB/MSE**	
A × B	$(K-1)(J-1)$	$SSAB$	$SSAB$/df	**MSAB/MSE**	
Error	$KJ(n-1)$	SSE	SSE/df		
Total	$N-1$	SSY			

H_0: $\alpha_1 = \alpha_2 = \ldots = \alpha_K = 0$

H_0: $\beta_1 = \beta_2 = \ldots = \beta_J = 0$

H_0: All $\gamma_{kj} = 0$

(Looking back) If $n = 1$ then df(interaction) = 0

Recall: Randomized Block ANOVA Table

Source	d.f.	S.S.	M.S.	t.s.	p-value
Trts/A	K−1	*SSTr*	*SSTr*/(*K*−1)	MSTr/MSE	
Block	J−1	*SSB*	*SSB*/(*J*−1)	MSB/MSE	
Error	(K−1)(J−1)	*SSE*	*SSE*/(*K*−1)(*J*−1)		
Total	N−1	*SSTotal*			

Fire Example: Two-way ANOVA Table, with Interaction

Source	d.f.	S.S.	M.S.	t.s.	p
Chemical	2	56	28.0	0.42	0.672
Type	1	48	48.0	0.73	0.426
A × B	2	1184	592.0	8.97	0.016
Error	6	396	66.0		
Total	11	1684			

$H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_K = 0$

$H_0: \beta_1 = \beta_2 = \ldots = \beta_J = 0$

H_0: All $\gamma_{kj} = 0$

Two-way ANOVA with *R*

Option #1 - aov

```
> model=aov(Time~Chemical+Type+Chemical:Type)
> summary(model)
              Df Sum Sq Mean Sq F value Pr(>F)
Chemical       2     56      28   0.424 0.6725
Type           1     48      48   0.727 0.4265
Chemical:Type  2   1184     592   8.970 0.0157 *
Residuals      6    396      66
```

Two-way ANOVA with *R*

Option #2 – anova(lm), when predictors are categorical

```
> anova(lm(Time~Chemical+Type+Chemical:Type))
Analysis of Variance Table

Response: Time
              Df Sum Sq Mean Sq F value  Pr(>F)
Chemical       2     56      28  0.4242 0.67247
Type           1     48      48  0.7273 0.42649
Chemical:Type  2   1184     592  8.9697 0.01574 *
Residuals      6    396      66
```

If sample sizes are <u>not</u> equal, order matters.

New example (on website):
Y = GPA
Explanatory variables are:
Seat location (front, middle, back)
Alcohol consumption
(none, some, lots)