Part VI

Analysis of Variance: II

Chapter 19

Two–Factor Analysis of Variance—Equal Sample Sizes

We look at two–factor ANOVA where we assume there is an equal sample size for each treatment.

19.1 Multifactor Studies

We look at some examples of two–factor experiments.

Exercise 19.1 (Multifactor Studies)

1. *Experimental Study: Temperature, Noise and Mice ROC.*
   Consider the effect of air temperature *and* noise on the ROC of deer mice. Twenty–four different mice are subjected, two at a time, to four different temperatures and three different noises. For instance, the ROCs of the first two mice, subjected to 0^o F temperature and a low noise level, are 10.3 and 7.2 units, respectively.

	noise →	low	medium	high
temperature	0^o F	10.3, 7.2	9.1, 5.4	6.1, 2.1
	10^o F	1.8, 9.8	12.1, 4.2	5.1, 6.2
	20^o F	1.2, 8.1	6.5, 4.1	1.2, 2.1
	30^o F	12.4, 15.1	16.1, 17.2	18.1, 19.1

	Factor B, noise →	low	medium	high
Factor A,	0^o F	treatment 1	treatment 2	treatment 3
temperature	10^o F	treatment 4	treatment 5	treatment 6
	20^o F	treatment 7	treatment 8	treatment 9
	30^o F	treatment 10	treatment 11	treatment 12

(a) Match the statistical terms with the various parts of this example.

statistical terms	temp, noise and ROC example
(i) factor	**(i)** 0^o F and low noise
(ii) a treatment	**(ii)** deer mouse
(iii) response variable	**(iii)** ROC
(iv) observational unit	**(iv)** temperature and noise
	(v) temperature
	(vi) lighting

statistical terms	(i)	(ii)	(iii)	(iv)
ROC example				

where, notice, a *treatment* is any combination of the levels of any of the factors in an experiment.

(b) If "temperature" is one factor of this research study, then another factor of this research study is (circle one)

i. low noise.
ii. 10.3, 7.2.
iii. noise.
iv. ROC of mice.

(c) If one treatment is (low noise, 0^oF), then there are (circle one) **nine** / **ten** / **eleven** / **twelve** treatments. In other words, the number of treatments equals the product of the number of levels of each factor in a research study.

(d) **True** / **False** The control, in this case, is any one of the treatments, and could be "0^o F and low noise", say.

(e) **True** / **False** The observational unit is a single mouse and the experimental unit is the *pair* of mice receiving one of the twelve treatments.

(f) This is an experimental study because (choose one)

i. the mice decided which treatments to be tested at.
ii. the experimenters decided how to assign the treatments to each mouse.
iii. no one decided how to assign the treatments to each mouse; it was done at random.

(g) In this case, we have,

i. qualitative temperature, qualitative noise
ii. qualitative temperature, quantitative noise
iii. quantitative temperature, qualitative noise
iv. quantitative temperature, quantitative noise

(h) This is a (choose one) **single–factor** / **multifactor** study.

(i) **True** / **False** In a single–factor study, a treatment corresponds to a factor level.

(j) Possible analyses we could conduct on this multifactor study are (choose none, one or more)

i. Do temperature and noise interact in some way when influencing the ROC of mice? How do they interact?

ii. Does one of the two factors, temperature or noise, have a greater influence on the ROC of mice?

(k) *Statistic, Sample, Parameter and Population.* Match the statistical terms with the various parts of this example.

statistical terms	example
(i) statistics	**(i)** ROCs of mice in study
(ii) sample	**(ii)** average ROC per treatment for mice in study
(iii) parameters	**(iii)** ROCs of all mice
(iv) population	**(iv)** average ROC per treatment for all mice

statistical terms	(i)	(ii)	(iii)	(iv)
example				

2. *Experimental Study: Catalysts, Rice Liquor Blends and Prozac Yield.* Consider the effect of different catalysts and blends on Prozac yields from 10 milliliters of rice liquor. For instance, the yield of catalyst A with 10 milliliters of rice liquor blend 2 is 84 units of Prozac.

	Factor B, blend →	1	2	3	4	5
Factor A,	A	89	84	81	87	79
catalyst	B	88	77	87	92	81
	C	97	92	87	89	80
	D	94	79	85	84	88

(a) Match the statistical terms with the various parts of this example.

statistical terms	example
(i) factor	**(i)** average yield per 10 milliliters of rice liquor
(ii) a treatment	**(ii)** 10 milliliters of rice liquor
(iii) response variable	**(iii)** Prozac yield
(iv) observational unit	**(iv)** catalyst and blend
	(v) temperature
	(vi) catalyst

statistical terms	(i)	(ii)	(iii)	(iv)
example				

where, notice, a *treatment* is any combination of the levels of any of the factors in an experiment.

(b) If "catalyst" is one factor of this study, then another factor of this study is (circle one)

i. low noise.
ii. 87.
iii. noise.
iv. blend.

(c) There are (circle one) **nine** / **ten** / **eleven** / **twenty** treatments.

(d) This is an experimental study because (choose one)

i. the rice liquor decided which treatments to be assigned.
ii. the experimenters decided how to assign the treatments to each 10 milliliter rice liquor batches.
iii. no one decided how to assign the treatments to each batch; it was done at random.

(e) In this case, we have,

i. qualitative catalyst, qualitative blend
ii. qualitative catalyst, quantitative blend
iii. quantitative catalyst, qualitative blend
iv. quantitative catalyst, quantitative blend

(f) This is a (choose one) **single–factor** / **multifactor** study.

(g) **True** / **False**
In a two–factor study you find out about the influence of two factors as well as any interaction between the two; in a single–factor study, you find out about the influence of one factor only. Furthermore, *less* data is required in a two–factor analysis to provide the same amount of precision as two one–factor analyses.

(h) This is a (choose one) **full** / **fractional** two–factor study because every factor combination (treatment) has data associated with it. Sometimes this is not possible. For example, in an experiment with 10 levels of one factor and 15 levels of another factor (150 treatments) it might be possible to provide data for all 150 treatments. In this case, some (clever) subset or fraction of all possible treatments might be considered in this study.

(i) *Statistic, Sample, Parameter and Population.* Match the statistical terms with the various parts of this example.

statistical terms	example
(i) statistics	**(i)** Prozac yields from 20 treatments
(ii) sample	**(ii)** average Prozac yield from 20 treatments
(iii) parameters	**(iii)** Prozac yields from 20 populations
(iv) population	**(iv)** average Prozac yields from 20 populations

statistical terms	(i)	(ii)	(iii)	(iv)
example				

19.2 Meaning of ANOVA Model Elements

Exercise 19.2 (Meaning of ANOVA Model Elements)

1. *Notation for Population: Catalysts, Rice Liquor Blends and Prozac Yield.* Consider the effect of different catalysts and blends on Prozac yields from 10 milliliters of rice liquor.

	Factor B, blend →	1	2	3	4	5	factor A means
Factor A,	A	89 (μ_{11})	84	81	87	79	84 ($\mu_{1\cdot}$)
catalyst ↓	B	88	77	87	92	81	85 ($\mu_{2\cdot}$)
	C	97	92	87	89	80	89 ($\mu_{3\cdot}$)
	D	94	79	85 (μ_{43})	84	88	86 ($\mu_{4\cdot}$)
	factor B means	92 ($\mu_{\cdot 1}$)	83 ($\mu_{\cdot 2}$)	85 ($\mu_{\cdot 3}$)	88 ($\mu_{\cdot 4}$)	82 ($\mu_{\cdot 5}$)	86 ($\mu_{\cdot\cdot}$)

Instead of dealing with the statistics based on an *observed* sample, we look at the population parameters associated with the two–factor ANOVA model; that is, we *assume all of the means calculated here are, in fact, the population (not sample) averages.*

(a) treatment mean, μ_{11} = (circle one) **84** / **89** / **92**

(b) treatment mean, μ_{21} = (circle one) **84** / **88** / **92**

(c) treatment mean, μ_{45} = (circle one) **84** / **88** / **92**

(d) factor A, level 1, mean,
$\mu_{1\cdot} = \frac{1}{b}\sum_{j=1}^{b}\mu_{1j} = \frac{1}{5}(420)$ = (circle one) **84** / **89** / **92**

(e) factor A, level 2, mean,
$\mu_{2\cdot} = \frac{1}{b}\sum_{j=1}^{b}\mu_{2j} = \frac{1}{5}(425)$ = (circle one) **84** / **85** / **92**

(f) factor B, level 5, mean,
$\mu_{\cdot 5} = \frac{1}{a}\sum_{i=1}^{a}\mu_{i5} = \frac{1}{4}(328)$ = (circle one) **82** / **88** / **92**

(g) overall mean,
$\mu_{\cdot\cdot} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\mu_{ij} = \frac{1}{(4)5}(1720)$ = (circle one) **85** / **86** / **87**

(h) *main effect*, factor A, level 1,
$\alpha_1 = \mu_{1\cdot} - \mu_{\cdot\cdot} = 84 - 86 =$ (circle one) **−2 / −5 / −6**

(i) main effect, factor A, level 2,
$\alpha_2 = \mu_{2\cdot} - \mu_{\cdot\cdot} = 85 - 86 =$ (circle one) **−1 / −4 / −5**

(j) main effect, factor B, level 5,
$\beta_5 = \mu_{\cdot 5} - \mu_{\cdot\cdot} = 82 - 86 =$ (circle one) **−3 / −4 / −5**

(k) **True** / **False**
The main effects are given by

$$\alpha_i = \mu_{i\cdot} - \mu_{\cdot\cdot}, \quad \beta_j = \mu_{\cdot j} - \mu_{\cdot\cdot}$$

where $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$

(l) *interaction* effect AB, levels 15,
$(\alpha\beta)_{15} = \mu_{15} - (\mu_{\cdot\cdot} + \alpha_1 + \beta_5) = 79 - (86 - 2 - 4) =$
(circle one) **−1 / 0 / −1**

(m) interaction effect AB, levels 25,
$(\alpha\beta)_{25} = \mu_{25} - (\mu_{\cdot\cdot} + \alpha_2 + \beta_5) = 81 - (86 - 1 - 4) =$
(circle one) **−1 / 0 / −1**

(n) *Interaction effect.*
True / **False** The *interaction effect* is

$$\begin{aligned}(\alpha\beta)_{ij} &= \mu_{ij} - (\mu_{\cdot\cdot} + \alpha_i + \beta_j) \\ &= \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot}\end{aligned}$$

where $\sum_{ij} (\alpha\beta)_{ij} = 0$.

(o) *Additive factors*
The factors are *additive if*, for all i, j,

$$\begin{aligned}\mu_{ij} &= \mu_{\cdot\cdot} + \alpha_i + \beta_j \\ &= \mu_{\cdot\cdot} + (\mu_{i\cdot} - \mu_{\cdot\cdot}) + (\mu_{\cdot j} - \mu_{\cdot\cdot}) \\ &= \mu_{i\cdot} + \mu_{\cdot j} - \mu_{\cdot\cdot}\end{aligned}$$

Since, for example,

$$\begin{aligned}\mu_{15} &= 79 \\ &\neq \mu_{\cdot\cdot} + \alpha_1 + \beta_5 \\ &= 86 + (-2) + (-4) \\ &= 80\end{aligned}$$

the two factors in this experiment (choose one) **are** / **are not** additive. In fact, they *interact* with one another.

(p) *More on the interaction effect.*
True / **False** The interaction effect,

$$(\alpha\beta)_{ij} = \mu_{ij} - (\mu_{\cdot\cdot} + \alpha_i + \beta_j) = \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot}$$

is the difference between what the treatment means *actually* are, μ_{ij}, and what the treatment means *would be* if the factor were additive, $\mu_{\cdot\cdot}+\alpha_i+\beta_j$.

(q) *Still more on the interaction effect.*
True / **False**
Since the interaction effect is

$$(\alpha\beta)_{ij} = \mu_{ij} - (\mu_{\cdot\cdot} + \alpha_i + \beta_j)$$

then

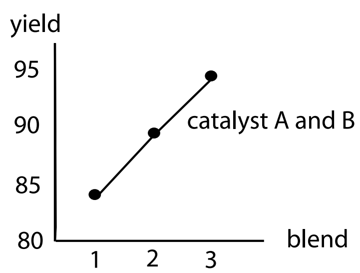
$$\mu_{ij} = \mu_{\cdot\cdot} + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

is the *factor effects* formulation of the two–factor (fixed factor level) ANOVA model I.

2. *Understanding Additive and Interaction Effects: Catalysts, Rice Liquor Blends and Prozac Yield.*
We look at the treatment mean plots for three data sets to help understand additive and interaction effects.

(a) *Additive Factor Effects: Blend Effect, But No Catalyst Effect and No Interactions*
Consider the effect of different catalysts and blends on Prozac yields from 10 milliliters of rice liquor[1].

	Factor B, blend →	1	2	3	row ave
Factor A, catalyst	A	84	89	94	$\mu_{1\cdot} = 89$
	B	84	89	94	$\mu_{2\cdot} = 89$
column ave		$\mu_{\cdot 1} = 84$	$\mu_{\cdot 2} = 89$	$\mu_{\cdot 3} = 94$	$\mu_{\cdot\cdot} = 89$



[1]Notice that this data is *different* from the previous data above. This data has two catalysts and three blends, whereas the data above had four catalysts and five blends!

Figure 19.1 (Additive Factor Effects: Blend Effect, But No Catalyst Effect and No Interactions)

There is no catalyst effect because, for example,
$\alpha_1 = \mu_{1\cdot} - \mu_{\cdot\cdot} = 89 - 89 =$ (circle one) $\mathbf{-1}$ / $\mathbf{0}$ / $\mathbf{1}$
There is a blend effect, though, because, for example,
$\beta_1 = \mu_{\cdot 1} - \mu_{\cdot\cdot} = 84 - 89 =$ (circle one) $\mathbf{-7}$ / $\mathbf{-5}$ / $\mathbf{1}$
There is no interaction effect also because, for example,
$(\alpha\beta)_{11} = \mu_{11} - (\mu_{\cdot\cdot} + \alpha_1 + \beta_1) = 84 - (89 + 0 - 5) =$
(circle one) $\mathbf{-7}$ / $\mathbf{0}$ / $\mathbf{1}$

(b) *Additive Factor Effects:*
Blend and Catalyst Effect, But No Interactions
Consider the effect of different catalysts and blends on Prozac yields from 10 milliliters of rice liquor.

	Factor B, blend →	1	2	3	row ave
Factor A,	A	84	86	93	$\mu_{1\cdot} = 263/3$
catalyst	B	80	82	89	$\mu_{2\cdot} = 251/3$
column ave		$\mu_{\cdot 1} = 82$	$\mu_{\cdot 2} = 84$	$\mu_{\cdot 3} = 91$	$\mu_{\cdot\cdot} = 257/3$

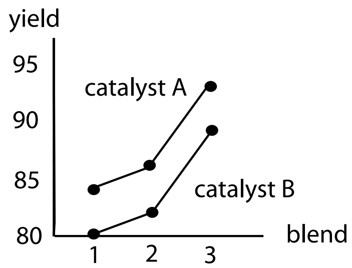


Figure 19.2 (Additive Factor Effects: Blend And Catalyst Effect, But No Interactions)

There is a catalyst effect because, for example,
$\alpha_1 = \mu_{1\cdot} - \mu_{\cdot\cdot} = \frac{263}{3} - \frac{257}{3} =$ (circle one) $\mathbf{2}$ / $\mathbf{3}$ / $\mathbf{4}$
There is also a blend effect because, for example,
$\beta_1 = \mu_{\cdot 1} - \mu_{\cdot\cdot} = 82 - \frac{257}{3} =$ (circle one) $\mathbf{-\frac{11}{3}}$ / $\mathbf{-\frac{10}{3}}$ / $\mathbf{1}$
But there is *no* interaction effect because, for example,
$(\alpha\beta)_{11} = \mu_{11} - (\mu_{\cdot\cdot} + \alpha_1 + \beta_1) = 84 - \left(\frac{257}{3} + 2 - \frac{11}{3}\right) =$
(circle one) $\mathbf{-7}$ / $\mathbf{0}$ / $\mathbf{1}$

(c) *Interacting Blend And Catalyst Factor Effects*
Consider the effect of different catalysts and blends on Prozac yields from 10 milliliters of rice liquor.

	Factor B, blend →	1	2	3	row ave
Factor A,	A	84	92	93	$\mu_{1\cdot} = 269/3$
catalyst	B	80	82	89	$\mu_{2\cdot} = 251/3$
column ave		$\mu_{\cdot 1} = 82$	$\mu_{\cdot 2} = 87$	$\mu_{\cdot 3} = 91$	$\mu_{\cdot\cdot} = 260/3$

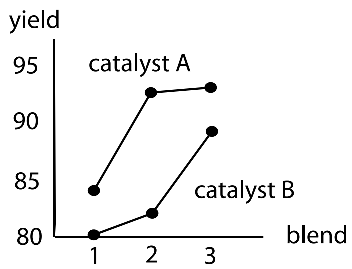


Figure 19.3 (Interacting Blend And Catalyst Factor Effects)

There is a catalyst effect because, for example,
$\alpha_1 = \mu_{1\cdot} - \mu_{\cdot\cdot} = \frac{269}{3} - \frac{260}{3} =$ (circle one) **2 / 3 / 4**
There is also a blend effect because, for example,
$\beta_1 = \mu_{\cdot 1} - \mu_{\cdot\cdot} = 82 - \frac{260}{3} =$ (circle one) $\mathbf{-\frac{11}{3}}$ / $\mathbf{-\frac{14}{3}}$ / **1**
And there is also interaction effect because, for example,
$(\alpha\beta)_{11} = \mu_{11} - (\mu_{\cdot\cdot} + \alpha_1 + \beta_1) = 84 - \left(\frac{260}{3} + 3 - \frac{14}{3}\right) =$
(circle one) **−1 / 0 / 1**

(d) *Interaction if treatment plots non–parallel.*
True / False Non–parallel treatment means plots are an indication of interaction between the factors.

3. *Transformable and Nontransformable Interactions.*
Some models with interaction can be transformed into additive models.

(a) **True / False**
The interactive multiplicative factor effects model,

$$\mu_{ij} = \mu_{\cdot\cdot}\alpha_i\beta_j$$

can be transformed, by taking (natural) logarithms, to become

$$\begin{aligned} \ln \mu_{ij} &= \ln \mu_{\cdot\cdot} + \ln \alpha_i + \ln \beta_j \\ \mu'_{ij} &= \mu'_{\cdot\cdot} + \alpha'_i + \beta'_j \end{aligned}$$

where $\mu'_{ij} = \ln \mu_{ij}$, $\mu'_{\cdot\cdot} = \ln \mu_{\cdot\cdot}$, $\alpha'_i = \ln \alpha_i$ and $\beta'_j = \ln \beta_j$, which is an additive factor effect model.

(b) **True** / **False**
The interactive factor effects model,

$$\mu_{ij} = \alpha_i + \beta_j + \sqrt{\alpha_i}\sqrt{\beta_j}$$

can be rearranged and transformed[2], to become

$$\begin{aligned} \mu_{ij} &= \left(\sqrt{\alpha_i} + \sqrt{\beta_j}\right)^2 \\ \sqrt{\mu_{ij}} &= \sqrt{\alpha_i} + \sqrt{\beta_j} \\ \mu'_{ij} &= \alpha'_i + \beta'_j \end{aligned}$$

where $\mu'_{ij} = \sqrt{\mu_{ij}}$, $\alpha'_i = \sqrt{\alpha_i}$ and $\beta'_j = \sqrt{\beta_j}$, which is an additive factor effect model.

19.3 Model I (Fixed Factor Levels) for Two–Factor Studies

The two–factor (fixed factor levels) ANOVA model I can be described in at least two different ways. One formulation is called the *cell means model*, given by

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk},$$

where

Y_{ijk} is the value of the response variable
in the kth case of the ith factor level of factor A
and the jth factor level of factor B

μ_{ij} are cell mean parameters

ε_{ijk} are independent $N(0, \sigma^2)$

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

The alternative *factor effects model* formulation is

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

where

Y_{ijk} is the value of the response variable
in the kth case of the ith factor level of factor A
and the jth factor level of factor B

[2]Take the square root of both sides in the second step.

$\mu_{..} = \frac{1}{ab}\sum_i \sum_j \mu_{ij}$ is the average of all cell mean parameters

$\alpha_i = \mu_{i\cdot} - \mu_{..}$ main effect for factor A at ith level

$\beta_j = \mu_{\cdot j} - \mu_{..}$ main effect for factor B at jth level

$(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{..}$ interaction effect
when factor A is at ith level and factor B is at jth level

ε_{ijk} are independent $N(0, \sigma^2)$

$i = 1, \ldots, r;\;\; j = 1, \ldots, n_i$

Exercise 19.3 (Notation Used In Model I (Fixed Factor Levels) for Two–Factor Studies) Consider the effect of air temperature *and* noise on the ROC of deer mice. Twenty–four different mice are subjected, two at a time, to four different temperatures and three different noises. For instance, the ROCs of the first two mice, subjected to 0^o F temperature and a low noise level, are 10.3 and 7.2 units, respectively.

	Factor B, noise →	$j = 1$, low	$j = 2$, medium	$j = 3$, high	row ave
Factor A,	$i = 1$, 0^o F	10.3, 7.2	9.1, 5.4	6.1, 2.1	$\overline{Y}_{1..} = 6.7$
temperature	$i = 2$, 10^o F	1.8, 9.8	12.1, 4.2	5.1, 6.2	$\overline{Y}_{2..} = 6.5$
	$i = 3$, 20^o F	1.2, 8.1	6.5, 4.1	1.2, 2.1	$\overline{Y}_{3..} = 3.9$
	$i = 4$, 30^o F	12.4, 15.1	16.1, 17.2	18.1, 19.1	$\overline{Y}_{4..} = 16.3$
	column ave	$\overline{Y}_{\cdot 1 \cdot} = 8.2$	$\overline{Y}_{\cdot 2 \cdot} = 9.3$	$\overline{Y}_{\cdot 3 \cdot} = 7.5$	$\overline{Y}_{...} = 8.4$

1. $Y_{ijk} = Y_{111} =$ (circle one) **10.3 / 7.2 / 1.8**
2. $Y_{112} =$ (circle one) **10.3 / 7.2 / 1.8**
3. $Y_{212} =$ (circle one) **5.4 / 7.2 / 9.8**
4. $Y_{431} =$ (circle one) **18.1 / 19.1 / 17.2**
5. treatment total, $Y_{11\cdot} = \sum_{k=1}^{n} Y_{11k} =$ (circle one) **16.5 / 17.5 / 18.5**
6. treatment total, $Y_{21\cdot} = \sum_{k=1}^{n} Y_{21k} =$ (circle one) **11.5 / 11.6 / 11.7**
7. treatment total, $Y_{43\cdot} = \sum_{k=1}^{n} Y_{43k} =$ (circle one) **28.1 / 29.1 / 37.2**
8. treatment average,
   $\overline{Y}_{11\cdot} = \frac{1}{n}\sum_{k=1}^{n} Y_{11k} = \frac{1}{n}Y_{11\cdot} = \frac{1}{2}(17.5) =$ (circle one) **6.5 / 7.5 / 8.8**
9. treatment average,
   $\overline{Y}_{21\cdot} = \frac{1}{n}\sum_{k=1}^{n} Y_{21k} = \frac{1}{n}Y_{21\cdot} = \frac{1}{2}(11.6) =$ (circle one) **5.8 / 6.6 / 7.7**

10. treatment average,
$\overline{Y}_{43\cdot} = \frac{1}{n}\sum_{k=1}^{n} Y_{43k} = \frac{1}{n}Y_{43\cdot} = \frac{1}{2}(37.2) =$ (circle one) **18.6 / 19.1 / 22.2**

11. factor A, level 1, total, $Y_{1\cdot\cdot} = \sum_{j=1}^{b}\sum_{k=1}^{n} Y_{1jk} =$ (circle one) **36.5 / 37.5 / 40.2**

12. factor A, level 2, total, $Y_{2\cdot\cdot} = \sum_{j=1}^{b}\sum_{k=1}^{n} Y_{2jk} =$ (circle one) **39.2 / 41.6 / 41.7**

13. factor B, level 2, total, $Y_{\cdot 2\cdot} = \sum_{i=1}^{a}\sum_{k=1}^{n} Y_{i2k} =$ (circle one) **68.1 / 69.1 / 74.7**

14. factor B, level 3, total, $Y_{\cdot 3\cdot} = \sum_{i=1}^{a}\sum_{k=1}^{n} Y_{i3k} =$ (circle one) **58.1 / 59.1 / 60.0**

15. overall total, $Y_{\cdot\cdot\cdot} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} =$ (circle one) **128.1 / 179.1 / 200.6**

16. factor A, level 1, average,
$\overline{Y}_{1\cdot\cdot} = \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{1jk} = \frac{1}{bn}Y_{1\cdot\cdot} = \frac{1}{3(2)}(40.2) =$
(circle one) **6.5 / 6.7 / 7.2**

17. factor B, level 2, average,
$\overline{Y}_{\cdot 2\cdot} = \frac{1}{an}\sum_{i=1}^{a}\sum_{k=1}^{n} Y_{i2k} = \frac{1}{an}Y_{\cdot 2\cdot} = \frac{1}{4(2)}(74.7) =$
(circle one) **8.1 / 9.1 / 9.3**

18. factor B, level 3, average,
$\overline{Y}_{\cdot 3\cdot} = \frac{1}{an}\sum_{i=1}^{a}\sum_{k=1}^{n} Y_{i3k} = \frac{1}{an}Y_{\cdot 3\cdot} = \frac{1}{4(2)}(60) =$
(circle one) **7.5 / 9.1 / 10.0**

19. overall average,
$\overline{Y}_{\cdot\cdot\cdot} = \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} = \frac{1}{abn}Y_{\cdot\cdot\cdot} = \frac{1}{4(3)(2)}(200.6) =$
(circle one) **8.1 / 8.4 / 9.6**

20. treatment residual, $e_{111} = Y_{111} - \bar{Y}_{11\cdot} = 10.3 - 8.8 =$ (circle one) **1.5 / 7.5 / 8.7**

21. treatment residual, $e_{211} = Y_{211} - \bar{Y}_{21\cdot} = 1.8 - 5.8 =$ (circle one) **−3 / −4 / −5**

22. fitted value, $\hat{Y}_{111} = \overline{Y}_{11\cdot} =$ (circle one) **16.5 / 17.5 / 8.8**

23. fitted value, $\hat{Y}_{112} = \overline{Y}_{11\cdot} =$ (circle one) **16.5 / 17.5 / 8.8**

24. fitted value, $\hat{Y}_{211} = \overline{Y}_{21\cdot} =$ (circle one) **11.5 / 11.6 / 5.8**

25. overall mean estimate,
$\hat{\mu}_{\cdot\cdot} = \overline{Y}_{\cdot\cdot\cdot} = \frac{1}{abn}Y_{\cdot\cdot\cdot} = \frac{1}{4(3)(2)}(200.6) =$ (circle one) **8.1 / 8.4 / 9.6**

26. factor A, level 1, effect estimate,
$\hat{\alpha}_1 = \overline{Y}_{1\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot} = 6.7 - 8.4 =$ (circle one) **−1.7 / −1.9 / −2.1**

27. factor B, level 1, effect estimate,
$\hat{\beta}_1 = \overline{Y}_{\cdot 1\cdot} - \overline{Y}_{\cdot\cdot\cdot} = 8.2 - 8.4 =$ (circle one) **−0.4 / −0.2 / 0.2**

28. interaction AB effect estimate,
$\hat{\alpha\beta}_{11} = \overline{Y}_{11\cdot} - \overline{Y}_{1\cdot\cdot} - \overline{Y}_{\cdot 2\cdot} + \overline{Y}_{\cdots} = 8.8 - 6.7 - 8.2 + 8.4 =$
(circle one) **2.1 / 2.2 / 2.3**

19.4 Analysis of Variance

The two–factor analysis of variance (ANOVA) procedure, assuming equal sample sizes, is demonstrated in this section. This ANOVA test uses the data to test whether the

- mean main effect due to factor A is zero or not,
- mean main effect due to factor B is zero or not,
- mean interaction effect due to factor AB
(after eliminating the main effect A and eliminating main effect B) is zero or not

An important part of this procedure involves the calculation of the ANOVA table,

Source	Degrees of Freedom, *df*	Sum Of Squares, *SS*	Mean Squares, *MS*
Factor A	$a-1$	$SSA = nb\sum(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots})^2$	$MSTA = \frac{SSA}{a-1}$
Factor B	$b-1$	$SSB = na\sum(\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots})^2$	$MSTB = \frac{SSB}{b-1}$
Interaction AB	$(a-1)(b-1)$	$SSAB = n\sum\sum(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots})^2$	$MSTAB = \frac{SSAB}{(a-1)(b-1)}$
Error	$ab(n-1)$	$SSE = \sum\sum\sum(Y_{ijk} - \overline{Y}_{ij\cdot})^2$	$MSE = \frac{SSE}{ab(n-1)}$
Total	$nab-1$	$SSTO = \sum\sum(Y_{ijk} - \overline{Y}_{\cdots})^2$	

19.5 Evaluation of Appropriateness of ANOVA Model

SAS program: att3-19-5-deermice-residual,twoANOVA

As explained in chapter 18, residuals are plotted against various quantities, including the response (fitted) variable, time (or other appropriate sequence variables), frequency (dot plots, histograms, stem–and–leaf plots) and percentiles (normal probability plot) in an effort to assess the validity of the assumptions of a model used to fit the data.

Exercise 19.4 (Evaluation of Appropriateness of ANOVA Model)
Consider the effect of air temperature *and* noise on the ROC of deer mice.

	Factor B, noise →	$j = 1$, low	$j = 2$, medium	$j = 3$, high	row ave
Factor A,	$i = 1$, 0^o F	10.3, 7.2	9.1, 5.4	6.1, 2.1	$\overline{Y}_{1\cdot\cdot} = 6.7$
temperature	$i = 2$, 10^o F	1.8, 9.8	12.1, 4.2	5.1, 6.2	$\overline{Y}_{2\cdot\cdot} = 6.5$
	$i = 3$, 20^o F	1.2, 8.1	6.5, 4.1	1.2, 2.1	$\overline{Y}_{3\cdot\cdot} = 3.9$
	$i = 4$, 30^o F	12.4, 15.1	16.1, 17.2	18.1, 19.1	$\overline{Y}_{4\cdot\cdot} = 16.3$
	column ave	$\overline{Y}_{\cdot 1\cdot} = 8.2$	$\overline{Y}_{\cdot 2\cdot} = 9.3$	$\overline{Y}_{\cdot 3\cdot} = 7.5$	$\overline{Y}_{\cdot\cdot\cdot} = 8.4$

1. *Treatment means plots*
   From SAS, either one of the two temperature–noise treatment means plots clearly show (choose one) **no interaction** / **interaction** between the two main factors, temperature and noise, because the plots are non–parallel. The interaction appears to be (choose one) **important** / **unimportant** because the plots are not too badly non–parallel.

2. *More treatment means plots*
   From SAS, both the temperature treatment mean plot and the noise treatment mean plot clearly shows the two main factors, temperature and noise, (choose one) **significant** / **not significant** because the plots are *not* horizontal.

3. *Residuals*
   From SAS, $e_{211} = -4$, and also
   $e_{212} =$ (choose one) $\mathbf{-4}$ / **0** / **4**

4. *Residual plot*
   From SAS, the residual plot appears to (choose one)

   (a) be randomly scattered.

   (b) have greater variability for the lower predicted values than for the upper predicted values.

   (c) have less variability for the lower predicted values than for the upper predicted values.

5. *Normal probability plot*
   The normal probability plot appears (choose one)

   (a) to be linear and so this indicates the error is normally distributed.

   (b) to be linear and so this indicates the error is *not* normally distributed.

   (c) to be *non*linear and so this indicates the error is normally distributed.

   (d) to be *non*linear and so this indicates the error is *not* normally distributed.

6. *Residual sequence plots*
   True / False
   Although only two points on each, all of the residual sequence plots appear to show random scatter and so there does not appear to be any dependence between the rate of oxygen consumption of each of the two mice in each of the treatments.

19.6 F Tests

SAS program: att3-19-6-deermice-twoANOVA

Exercise 19.5 (F Tests)
Consider the effect of air temperature *and* noise on the ROC of deer mice.

	Factor B, noise →	$j = 1$, low	$j = 2$, medium	$j = 3$, high	row ave
Factor A,	$i = 1$, 0^o F	10.3, 7.2	9.1, 5.4	6.1, 2.1	$\overline{Y}_{1\cdot\cdot} = 6.7$
temperature	$i = 2$, 10^o F	1.8, 9.8	12.1, 4.2	5.1, 6.2	$\overline{Y}_{2\cdot\cdot} = 6.5$
	$i = 3$, 20^o F	1.2, 8.1	6.5, 4.1	1.2, 2.1	$\overline{Y}_{3\cdot\cdot} = 3.9$
	$i = 4$, 30^o F	12.4, 15.1	16.1, 17.2	18.1, 19.1	$\overline{Y}_{4\cdot\cdot} = 16.3$
	column ave	$\overline{Y}_{\cdot 1\cdot} = 8.2$	$\overline{Y}_{\cdot 2\cdot} = 9.3$	$\overline{Y}_{\cdot 3\cdot} = 7.5$	$\overline{Y}_{\cdot\cdot\cdot} = 8.4$

Test if either of the two main effects or the interaction effect are significant at $\alpha = 0.05$.

1. *Preliminary work.*
   True / False
   From SAS, the ANOVA table is given by:

Source	*df*	*SS*	*MS*
Factor A (Temperature)	3	539.14	179.71
Factor B (Noise)	2	13.68	13.68
Interaction AB	6	55.70	9.28
Error	12	115.3	9.61
Total	23	723.82	

2. *Test Factor A, P–Value Versus Level of Significance.*

   (a) *Statement*
   The statement of the test is (check none, one or more):

   i. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a : \alpha_1 \neq \alpha_2, \alpha_1 = \alpha_3$.
   ii. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a : \alpha_1 \neq \alpha_3, \alpha_1 \neq \alpha_2$.

iii. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus
$H_a :$ at least one $\alpha_i \neq 0,\ i = 1, 2, 3, 4$.

iv. $H_0 : \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$ versus
$H_a :$ at least one $\mu_{i\cdot} \neq \mu_{j\cdot},\ i, j = 1, 2, 3, 4$.

(b) *Test*
Since the test statistic is $F = \frac{179.71}{9.61} = 18.70$, the p–value, with $a - 1 = 4 - 1 = 3$ and $ab(n - 1) = 4(3)(2 - 1) = 12$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 18.70)$$

which equals (circle one) **0** / **0.26** / **0.43**.
The level of significance is 0.05.

(c) *Conclusion*
Since the p–value, 0, is smaller than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the average mice ROC responses to the four temperatures are the same.

3. *Test Factor B, P–Value Versus Level of Significance.*

(a) *Statement*
The statement of the test is (check none, one or more):

i. $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ versus $H_a :\ \beta_1 \neq \beta_2, \beta_1 = \beta_3$.

ii. $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ versus $H_a :\ \beta_1 \neq \beta_3, \beta_1 \neq \beta_2$.

iii. $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ versus
$H_a :$ at least one $\beta_i \neq 0,\ i = 1, 2, 3$.

iv. $H_0 : \mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3}$ versus
$H_a :$ at least one $\mu_{i\cdot} \neq \mu_{j\cdot},\ i, j = 1, 2, 3$.

(b) *Test*
Since the test statistic is $F = \frac{6.84}{9.61} = 0.71$, the p–value, with $b-1 = 3-1 = 2$ and $ab(n - 1) = 4(3)(2 - 1) = 12$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 0.71)$$

which equals (circle one) **0.00** / **0.35** / **0.51**.
The level of significance is 0.05.

(c) *Conclusion.*
Since the p–value, 0.51, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the average mice ROC responses to the three noise levels are the same.

4. *Test Interaction Factor AB, P–Value Versus Level of Significance.*

(a) *Statement*
The statement of the test is (check none, one or more):

i. $H_0 : \alpha\beta_{11} = \alpha\beta_{12} = \alpha_4 = 0$ versus $H_a :\ \alpha_1 \neq \alpha_2, \alpha_1 = \alpha_3$.
ii. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a :\ \alpha_1 \neq \alpha_3, \alpha_1 \neq \alpha_2$.
iii. $H_0 : \alpha\beta_{11} = \alpha\beta_{12} = \cdots = \alpha\beta_{43} = 0$ versus
$H_a :$ at least one $\alpha\beta_{ij} \neq 0,\ i = 1, 2, 3, 4; j = 1, 2, 3$.
iv. $H_0 : \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$ versus
$H_a :$ at least one $\mu_{i\cdot} \neq \mu_{j\cdot},\ i, j = 1, 2, 3, 4$.

(b) *Test*
Since the test statistic is $F = \frac{9.28}{9.61} = 0.97$, the p–value, with $(a-1)(b-1) = (4-1)(3-1) = 6$ and $ab(n-1) = 4(3)(2-1) = 12$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 0.97)$$

which equals (circle one) **0.00** / **0.49** / **0.43**.
The level of significance is 0.05.

(c) *Conclusion*
Since the p–value, 0.49, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that there is no interaction effect.

5. *Kimball inequality*
The Kimball inequality is an *upper bound* on the family level of significance for the three tests (factor A, factor B, interaction AB) given above where, in each, $\alpha = 0.05$. It is determined by

$$\begin{aligned} \alpha &\leq 1 - (1-\alpha_1)(1-\alpha_2)(1-\alpha_3) \\ &= 1 - (1-0.05)(1-0.05)(1-0.05) = \end{aligned}$$

(choose one) **0.05** / **0.10** / **0.143**

19.7 Regression Approach to Two–Factor Analysis of Variance

SAS program: att3-19-7-gas-regressionANOVA

Exercise 19.6 (Regression Approach to Two–Factor Analysis of Variance)
Consider the effect of two different catalysts and three blends on the yields from 10 milliliters of gasoline.

	Factor B, blend →	1	2	3
Factor A,	A	84, 82	89, 87	94, 94
catalyst ↓	B	84, 85	89, 88	94, 93

1. *Regression model*
True / False
The factor means version of this two–factor ANOVA,

$$Y_{ijk} = \mu_{\cdot\cdot} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

can be written as a multiple regression model in the following way,

$$\begin{aligned} Y_{ijk} &= \mu_{\cdot\cdot} + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} \\ &\quad +(\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \varepsilon_{ijk} \end{aligned}$$

where[3] $i = 1, 2$; $j = 1, 2, 3$ and $k = 1, 2$
and where $\alpha_2 = -\alpha_1$ since $\sum_{i=1}^{2} \alpha_i = 0$,
and $\beta_3 = -\beta_1 - \beta_2$ since $\sum_{i=1}^{3} \beta_j = 0$
and so, as a consequence,

$$X_{ijk1} = \begin{cases} 1, & \text{if case from Factor A level 1} \\ -1, & \text{if case from Factor A level 2} \end{cases}$$

$$X_{ijk2} = \begin{cases} 1, & \text{if case from Factor B level 1} \\ -1, & \text{if case from Factor B level 3} \\ 0, & \text{otherwise,} \end{cases}$$

$$X_{ijk3} = \begin{cases} 1, & \text{if case from Factor B level 2} \\ -1, & \text{if case from Factor B level 3} \\ 0, & \text{otherwise,} \end{cases}$$

2. *More regression model*
True / False
Since

$$\begin{aligned} Y_{ijk} &= \mu_{\cdot\cdot} + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} \\ &\quad +(\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \varepsilon_{ijk} \end{aligned}$$

[3]The regression model could *not* be written as, say,

$$\begin{aligned} Y_{ijk} &= \mu_{\cdot\cdot} + \alpha_1 X_{ijk1} + \alpha_2 X_{ijk2} \\ &\quad + \beta_1 X_{ijk3} + \beta_2 X_{ijk4} + \beta_3 X_{ijk5} \\ &\quad + (\alpha\beta)_{11} X_{ijk1} X_{ijk3} + (\alpha\beta)_{12} X_{ijk1} X_{ijk4} + (\alpha\beta)_{13} X_{ijk1} X_{ijk5} \\ &\quad + (\alpha\beta)_{21} X_{ijk2} X_{ijk3} + (\alpha\beta)_{22} X_{ijk2} X_{ijk4} + (\alpha\beta)_{23} X_{ijk2} X_{ijk5} \\ &\quad + \varepsilon_{ijk} \end{aligned}$$

because, then, the resulting **X** matrix would be singular; in other words, unsolvable.

where the X_{ijkl} are defined as above, then

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

where

$$\mathbf{Y} = \begin{bmatrix} 84 \\ 82 \\ \vdots \\ 94 \\ 93 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix}, \quad \beta = \begin{bmatrix} \mu. \\ \alpha_1 \\ \beta_1 \\ \beta_2 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \end{bmatrix}$$

and so

$$\mathbf{X}\beta = \begin{bmatrix} \mu. + \alpha_1 + (\alpha\beta)_{11} \\ \mu. + \alpha_1 + (\alpha\beta)_{11} \\ \vdots \\ \mu. - \alpha_1 - \beta_1 - \beta_2 + (\alpha\beta)_{11} + (\alpha\beta)_{12} \\ \mu. - \alpha_1 - \beta_1 - \beta_2 + (\alpha\beta)_{11} + (\alpha\beta)_{12} \end{bmatrix} = \begin{bmatrix} \mu_{11} \\ \mu_{11} \\ \vdots \\ \mu_{23} \\ \mu_{23} \end{bmatrix}$$

3. *Regression*
   From SAS, the approximate regression to

$$\begin{aligned} Y_{ijk} &= \mu_{..} + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} \\ &\quad + (\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \varepsilon_{ijk} \end{aligned}$$

   is (choose one)

   (i) $\hat{Y} = 88.583 - 0.25X_1 - 4.833X_2 - 0.333X_3 - 0.5X_1X_2 - 0.2X_1X_3$

   (ii) $\hat{Y} = 88.583 - 0.25X_1 - 4.833X_2 - 0.333X_3 - 0.5X_1X_2$

   (iii) $\hat{Y} = 88.583 + 0.25X_1 - 4.833X_2 - 0.333X_3 - 0.5X_1X_2 - 0.2X_1X_3$

4. *Using the regression*
   Since

$$\hat{Y} = 88.583 - 0.25X_1 - 4.833X_2 - 0.333X_3 - 0.5X_1X_2$$

   and using the defined values of $X_{ijk1} = X_1$, $X_{ijk2} = X_2$ and $X_{ijk3} = X_3$ above,

(a) treatment mean for factor A, level 1, where $(X_1, X_2, X_3) = (1, 0, 0)$,

$$\begin{aligned}\hat{Y} &= 88.583 - 0.25X_1 - 4.833X_2 - 0.333X_3 - 0.5X_1X_2 \\ &= 88.583 - 0.25(1) - 4.833(0) - 0.333(0) - 0.5(0)(1) \\ &= \bar{Y}_{1\cdot\cdot} =\end{aligned}$$

(choose one) **86.56** / **87.78** / **88.333**

(b) treatment mean for factor B, level 2, where $(X_1, X_2, X_3) = (0, 0, 1)$,

$$\begin{aligned}\hat{Y} &= 88.583 - 0.25X_1 - 4.833X_2 - 0.333X_3 - 0.5X_1X_2 \\ &= 88.583 - 0.25(0) - 4.833(0) - 0.333(1) - 0.5(0)(0) \\ &= \bar{Y}_{\cdot 2\cdot} =\end{aligned}$$

(choose one) **88.25** / **88.77** / **89.3**

(c) treatment mean for factor B, level 3, where $(X_1, X_2, X_3) = (0, -1, -1)$,

$$\begin{aligned}\hat{Y} &= 88.583 - 0.25X_1 - 4.833X_2 - 0.333X_3 - 0.5X_1X_2 \\ &= 88.583 - 0.25(0) - 4.833(-1) - 0.333(-1) - 0.5(0)(-1) \\ &= \bar{Y}_{\cdot 3\cdot} =\end{aligned}$$

(choose one) **88.25** / **88.77** / **93.75**

(d) treatment mean for interaction, level 13, where $(X_1, X_2, X_3) = (1, -1, -1)$,

$$\begin{aligned}\hat{Y} &= 88.583 - 0.25X_1 - 4.833X_2 - 0.333X_3 - 0.5X_1X_2 \\ &= 88.583 - 0.25(1) - 4.833(-1) - 0.333(-1) - 0.5(1)(-1) \\ &= \bar{Y}_{13\cdot} =\end{aligned}$$

(choose one) **83.167** / **88.77** / **94**

5. *Comparing regression ANOVA with treatment mean ANOVA*
   The regression ANOVA is given by,

Source	*df*	*SS*	*MS*
Model	5	203.4166	40.683
X_1	1	0.75	0.75
X_2	1	200	200
X_3	1	0.666	0.666
X_1X_2	1	2	2
X_1X_3	1	0	0
Error	6	5.5	0.916
Total	11	208.9166	

whereas the treatment means ANOVA is given by,

Source	*df*	*SS*	*MS*
Model	5	203.4166	40.683
Factor A (Catalyst)	1	0.75	0.75
Factor B (Blend)	2	200.66	100.333
Interaction AB	2	2	1
Error	6	5.5	0.916
Total	11	723.82	

where, notice, the factor A (catalyst) SS is equal to the X_1 SS and the factor B (blend) SS is equal to (choose one)

(a) X_2 SS plus the X_3 SS
(b) X_2 SS
(c) X_3 SS

and the interaction SS is equal to the sum of the X_1X_2 and X_1X_3 SS.

6. *Test Interaction Factor AB, P–Value Versus Level of Significance.*

   (a) *Statement*
   The statement of the test is (check none, one or more):

      i. $H_0 : \alpha\beta_{11} = \alpha\beta_{12} = \alpha_4 = 0$ versus $H_a : \alpha_1 \neq \alpha_2, \alpha_1 = \alpha_3$.
      ii. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a : \alpha_1 \neq \alpha_3, \alpha_1 \neq \alpha_2$.
      iii. $H_0 : (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{23} = 0$ versus
      $H_a :$ at least one $(\alpha\beta)_{ij} \neq 0$, $i = 1, 2; j = 1, 2, 3$.
      iv. H_0 : additive model versus
      H_a : model with interaction.

   (b) *Test*
   Since the test statistic is $F = \frac{1}{0.916} = 1.09$, the p–value, with $(a-1)(b-1) = (2-1)(3-1) = 2$ and $ab(n-1) = 2(3)(2-1) = 6$ degrees of freedom, is given by

   $$\text{p–value} = P(F \geq 1.09)$$

   which equals (circle one) **0.00** / **0.39** / **0.43**.
   The level of significance is 0.05.

   (c) *Conclusion*
   Since the p–value, 0.39, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the model is additive.

19.8 Pooling Sums of Squares in Two–Factor Analysis of Variance

SAS program: att3-19-8-deermice-twoANOVA,poolinteract

Although somewhat controversial, if the interaction is found to be *non*significant, then it is sometimes pooled with the error term in a two–factor ANOVA.

Exercise 19.7 (Pooling Sums of Squares in Two–Factor Analysis of Variance) Consider the effect of air temperature *and* noise on the ROC of deer mice.

	Factor B, noise →	$j = 1$, low	$j = 2$, medium	$j = 3$, high	row ave
Factor A,	$i = 1$, 0^o F	10.3, 7.2	9.1, 5.4	6.1, 2.1	$\overline{Y}_{1\cdot\cdot} = 6.7$
temperature	$i = 2$, 10^o F	1.8, 9.8	12.1, 4.2	5.1, 6.2	$\overline{Y}_{2\cdot\cdot} = 6.5$
	$i = 3$, 20^o F	1.2, 8.1	6.5, 4.1	1.2, 2.1	$\overline{Y}_{3\cdot\cdot} = 3.9$
	$i = 4$, 30^o F	12.4, 15.1	16.1, 17.2	18.1, 19.1	$\overline{Y}_{4\cdot\cdot} = 16.3$
	column ave	$\overline{Y}_{\cdot 1\cdot} = 8.2$	$\overline{Y}_{\cdot 2\cdot} = 9.3$	$\overline{Y}_{\cdot 3\cdot} = 7.5$	$\overline{Y}_{\cdot\cdot\cdot} = 8.4$

From SAS, the ANOVA table is given,

Source	*df*	*SS*	*MS*
Factor A (Temperature)	3	539.14	179.71
Factor B (Noise)	2	13.68	13.68
Error	18	170.99	9.49
Total	23	723.82	

Pool the interaction AB with the error since it is not significant. Re–test if either of the two main effects are significant at $\alpha = 0.05$.

1. *Test Factor A, P–Value Versus Level of Significance.*

   (a) *Statement*
   The statement of the test is (check none, one or more):

      i. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a :\ \alpha_1 \neq \alpha_2, \alpha_1 = \alpha_3$.
      ii. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a :\ \alpha_1 \neq \alpha_3, \alpha_1 \neq \alpha_2$.
      iii. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus
      $H_a :$ at least one $\alpha_i \neq 0,\ i = 1, 2, 3, 4$.
      iv. $H_0 : \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$ versus
      $H_a :$ at least one $\mu_{i\cdot} \neq \mu_{j\cdot},\ i, j = 1, 2, 3, 4$.

   (b) *Test.*
   Since the test statistic is $F = \frac{179.71}{9.49} = 18.92$, the p–value, with $a - 1 = 4 - 1 = 3$ and $ab(n - 1) = 4(3)(2 - 1) = 12$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 18.92)$$

which equals (circle one) **0.00** / **0.35** / **0.43**.
The level of significance is 0.05.

(c) *Conclusion.*
Since the p–value, 0.00, is smaller than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the average mice ROC responses to the four temperatures are the same.

2. *Test Factor B, P–Value Versus Level of Significance.*

(a) *Statement*
The statement of the test is (check none, one or more):

i. $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ versus $H_a : \beta_1 \neq \beta_2, \beta_1 = \beta_3$.
ii. $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ versus $H_a : \beta_1 \neq \beta_3, \beta_1 \neq \beta_2$.
iii. $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ versus
$H_a :$ at least one $\beta_i \neq 0,\ i = 1, 2, 3$.
iv. $H_0 : \mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3}$ versus
$H_a :$ at least one $\mu_{i\cdot} \neq \mu_{j\cdot},\ i, j = 1, 2, 3$.

(b) *Test.*
Since the test statistic is $F = \frac{6.84}{9.49} = 0.72$, the p–value, with $b-1 = 3-1 = 2$ and $ab(n-1) = 4(3)(2-1) = 12$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 0.72)$$

which equals (circle one) **0.00** / **0.35** / **0.50**.
The level of significance is 0.05.

(c) *Conclusion.*
Since the p–value, 0.50, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the average mice ROC responses to the three noise levels are the same.

3. *One final comment*
It (choose one) **does** / **does not** make sense to pool the interaction with the error in this case, because we know from above, that although there is interaction, this interaction is not significant.

Chapter 20

Analysis of Factor Effects in Two–Factor Studies–Equal Sample Sizes

We look at analyzing the factor effects of two–factor studies in detail in this chapter. We find out that this analysis is influenced by whether the *interaction* is significant (important) or not.

20.1 Strategy for Analysis

On the one hand, if the interaction is *not* significant (*un*important), or can be made to be unimportant by using a simplifying transformation on the data, then inference is conducted separately on the two main effects using the factor level means. On the other hand, if the interaction *is* significant, and can*not* be made to be unimportant by using a simplifying transformation on the data, than inference is conducted on the *treatment* (cell) means.

20.2 Analysis of Factor Effects when Factors Do Not Interact

SAS program: att3-20-2-deermice-analysis,nointeract

In this section, we consider the case when the factors do *not* interact. The analysis focuses on comparing individual treatment main *factor* means, rather than comparing treatment *cell* means, which is necessary when factors do interact. A table of the various individual and multiple confidence intervals and tests that are often conducted in this case, is given below.

Inference	Factor	confidence interval	test	
Individual	A	$\overline{Y}_{i\cdot\cdot} \pm t(1-\alpha/2;(n-1)ab)s\left\{\overline{Y}_{i\cdot\cdot}\right\}$	$t^* = [\overline{Y}_{i\cdot\cdot} - c]/s\left\{\overline{Y}_{i\cdot\cdot}\right\}$	$s\left\{\hat{\overline{Y}}_{i\cdot\cdot}\right\} = \sqrt{\frac{MSE}{bn}}$
Factor Level	B	$\overline{Y}_{\cdot j\cdot} \pm t(1-\alpha/2;(n-1)ab)s\left\{\overline{Y}_{\cdot j\cdot}\right\}$	$t^* = [\overline{Y}_{i\cdot\cdot} - c]/s\left\{\overline{Y}_{\cdot j\cdot}\right\}$	$s\left\{\hat{\overline{Y}}_{\cdot j\cdot}\right\} = \sqrt{\frac{MSE}{an}}$
Individual	A	$\hat{L} \pm t(1-\alpha/2;(n-1)ab)s\left\{\hat{L}\right\}$	$t^* = [\hat{L} - c]/s\left\{\hat{L}\right\}$	$s\left\{\hat{L}\right\} = \sqrt{\frac{MSE}{bn}\sum c_i^2}$, $\hat{L} = \sum c_i \overline{Y}_{i\cdot\cdot}$
Contrast/Combination	B	$\hat{L} \pm t(1-\alpha/2;(n-1)ab)s\left\{\hat{L}\right\}$	$t^* = [\hat{L} - c]/s\left\{\hat{L}\right\}$	$s\left\{\hat{L}\right\} = \sqrt{\frac{MSE}{an}\sum c_i^2}$, $\hat{L} = \sum c_i \overline{Y}_{\cdot j\cdot}$
Multiple Pairwise	A	$\hat{D} \pm \frac{1}{\sqrt{2}}q(1-\alpha;a,(n-1)ab)s\left\{\hat{D}\right\}$	$q^* = [\sqrt{2}\hat{D}]/s\left\{\hat{D}\right\}$	$s\left\{\hat{D}\right\} = \sqrt{\frac{2MSE}{bn}}$, $\hat{D} = \overline{Y}_{i\cdot\cdot} - \overline{Y}_{i'\cdot\cdot}$
Tukey (many)	B	$\hat{D} \pm \frac{1}{\sqrt{2}}q(1-\alpha;b,(n-1)ab)s\left\{\hat{D}\right\}$	$q^* = [\sqrt{2}\hat{D}]/s\left\{\hat{D}\right\}$	$s\left\{\hat{D}\right\} = \sqrt{\frac{2MSE}{an}}$, $\hat{D} = \overline{Y}_{i\cdot\cdot} - \overline{Y}_{i'\cdot\cdot}$
Multiple Pairwise	A	$\hat{D} \pm t(1-\alpha/2g;(n-1)ab)s\left\{\hat{D}\right\}$	$t^* = [\hat{D}]/s\left\{\hat{D}\right\}$	$s\left\{\hat{D}\right\} = \sqrt{\frac{2MSE}{bn}}$, $\hat{D} = \overline{Y}_{i\cdot\cdot} - \overline{Y}_{i'\cdot\cdot}$
Bonferroni (few)	B	$\hat{D} \pm t(1-\alpha/2g;(n-1)ab)s\left\{\hat{D}\right\}$	$t^* = [\hat{D}]/s\left\{\hat{D}\right\}$	$s\left\{\hat{D}\right\} = \sqrt{\frac{2MSE}{an}}$, $\hat{D} = \overline{Y}_{i\cdot\cdot} - \overline{Y}_{i'\cdot\cdot}$
Multiple Contrast	A	$\hat{L} \pm \sqrt{(a-1)F(1-\alpha;a-1,(n-1)ab)}s\left\{\hat{L}\right\}$	$F^* = [\hat{L}]/[(a-1)s^2\left\{\hat{L}\right\}]$	$s\left\{\hat{L}\right\} = \sqrt{\frac{MSE}{bn}\sum c_i^2}$, $\hat{L} = \sum c_i \overline{Y}_{i\cdot\cdot}$
Scheffe (many)	B	$\hat{L} \pm \sqrt{(b-1)F(1-\alpha;b-1,(n-1)ab)}s\left\{\hat{L}\right\}$	$F^* = [\hat{L}]/[(b-1)s^2\left\{\hat{L}\right\}]$	$s\left\{\hat{L}\right\} = \sqrt{\frac{MSE}{an}\sum c_i^2}$, $\hat{L} = \sum c_i \overline{Y}_{\cdot j\cdot}$
Multiple Contrast	A	$\hat{L} \pm t(1-\alpha/2g;(n-1)ab)s\left\{\hat{L}\right\}$	$t^* = [\hat{L}]/[s\left\{\hat{L}\right\}]$	$s\left\{\hat{L}\right\} = \sqrt{\frac{MSE}{bn}\sum c_i^2}$, $\hat{L} = \sum c_i \overline{Y}_{i\cdot\cdot}$
Bonferroni (few)	B	$\hat{L} \pm t(1-\alpha/2g;(n-1)ab)s\left\{\hat{L}\right\}$	$t^* = [\hat{L}]/[s\left\{\hat{L}\right\}]$	$s\left\{\hat{L}\right\} = \sqrt{\frac{MSE}{an}\sum c_i^2}$, $\hat{L} = \sum c_i \overline{Y}_{\cdot j\cdot}$

Exercise 20.1 (Analysis of Factor Effects when Factors Do Not Interact) Consider the effect of air temperature *and* noise on the ROC of deer mice.

	Factor B, noise $\rightarrow$	$j=1$, low	$j=2$, medium	$j=3$, high	row ave
Factor A,	$i=1$, 0^o F	10.3, 7.2	9.1, 5.4	6.1, 2.1	$\overline{Y}_{1\cdot\cdot} = 6.7$
temperature	$i=2$, 10^o F	1.8, 9.8	12.1, 4.2	5.1, 6.2	$\overline{Y}_{2\cdot\cdot} = 6.5$
	$i=3$, 20^o F	1.2, 8.1	6.5, 4.1	1.2, 2.1	$\overline{Y}_{3\cdot\cdot} = 3.9$
	$i=4$, 30^o F	12.4, 15.1	16.1, 17.2	18.1, 19.1	$\overline{Y}_{4\cdot\cdot} = 16.3$
	column ave	$\overline{Y}_{\cdot 1\cdot} = 8.2$	$\overline{Y}_{\cdot 2\cdot} = 9.3$	$\overline{Y}_{\cdot 3\cdot} = 7.5$	$\overline{Y}_{\cdot\cdot\cdot} = 8.4$

Source	*df*	*SS*	*MS*
Factor A (Temperature)	3	539.14	179.71
Factor B (Noise)	2	13.68	13.68
Interaction AB	6	55.70	55.70
Error	12	115.30	9.61
Total	23	723.82	

Conduct inference on various individual and multiple linear combinations at $\alpha = 0.05$.

1. *Interaction significant?*
   Recall from above,

   (a) *Statement*
   The statement of the test is (check none, one or more):

   i. $H_0 : \alpha\beta_{11} = \alpha\beta_{12} = \alpha_4 = 0$ versus $H_a :\ \alpha_1 \neq \alpha_2, \alpha_1 = \alpha_3$.

   ii. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a :\ \alpha_1 \neq \alpha_3, \alpha_1 \neq \alpha_2$.

   iii. $H_0 : \alpha\beta_{11} = \alpha\beta_{12} = \cdots = \alpha\beta_{43} = 0$ versus
   $H_a :$ at least one $\alpha\beta_{ij} \neq 0,\ i = 1,2,3,4; j = 1,2,3$.

iv. $H_0 : \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$ versus
$H_a :$ at least one $\mu_{i\cdot} \neq \mu_{j\cdot}$, $i, j = 1, 2, 3, 4$.

(b) *Test*
Since the test statistic is $F = \frac{9.28}{9.61} = 0.97$, the p–value, with $(a-1)(b-1) = (4-1)(3-1) = 6$ and $ab(n-1) = 4(3)(2-1) = 12$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 0.97)$$

which equals (circle one) **0.00** / **0.49** / **0.43**.
The level of significance is 0.05.

(c) *Conclusion*
Since the p–value, 0.49, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that there is no interaction effect.

2. *Test, Individual Factor Level A.*
Test if the ROC subjected to factor A, level 1 (0^o F), is greater than 6.3 units at $\alpha = 0.05$.

(a) *Statement.*
The statement of the test is (check none, one or more):

i. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a :\ \alpha_1 \neq \alpha_2, \alpha_1 = \alpha_3$.
ii. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a :\ \alpha_1 \neq \alpha_3, \alpha_1 \neq \alpha_2$.
iii. $H_0 : \mu_{1\cdot} = 6.3$ versus $H_a :\ \mu_{1\cdot} > 6.3$.
iv. $H_0 : \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$ versus
$H_a :$ at least one $\mu_{i\cdot} \neq \mu_{j\cdot}$, $i, j = 1, 2, 3, 4$.

(b) *Test.*
Since

$$s\left\{\overline{Y}_{1\cdot\cdot}\right\} = \sqrt{\frac{MSE}{bn}} = \sqrt{\frac{9.61}{(3)(2)}} = 1.266$$

the test statistic is

$$t^* = [\overline{Y}_{1\cdot\cdot} - c]/s\left\{\overline{Y}_{1\cdot\cdot}\right\} = [6.7 - 6.3]/1.266 = 0.316$$

the p–value, with $(n-1)ab = (2-1)(4)(3) = 12$ degrees of freedom, is given by

$$\text{p–value} = P(t \geq 0.316)$$

which equals (circle one) **0.00** / **0.38** / **0.40**.
The level of significance is 0.05.

(c) *Conclusion.*
Since the p–value, 0.38, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the ROC, subjected to factor A, level 1 (0^o F), is equal to 6.3 units.

3. *Test, Individual Factor Level B.*
Test if the ROC subjected to factor B, level 2 (medium noise), is less than 9.5 units at $\alpha = 0.05$.

(a) *Statement.*
The statement of the test is (check none, one or more):

i. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a : \alpha_1 \neq \alpha_2, \alpha_1 = \alpha_3$.
ii. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a : \alpha_1 \neq \alpha_3, \alpha_1 \neq \alpha_2$.
iii. $H_0 : \mu_{\cdot 2} = 9.5$ versus $H_a : \mu_{\cdot 2} < 9.5$.
iv. $H_0 : \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$ versus
$H_a :$ at least one $\mu_{i\cdot} \neq \mu_{j\cdot}$, $i, j = 1, 2, 3, 4$.

(b) *Test.*
Since

$$s\left\{\overline{Y}_{\cdot 2\cdot}\right\} = \sqrt{\frac{MSE}{an}} = \sqrt{\frac{9.61}{(4)(2)}} = 1.096$$

the test statistic is

$$t^* = [\overline{Y}_{1\cdot\cdot} - c]/s\left\{\overline{Y}_{1\cdot\cdot}\right\} = [9.3 - 9.5]/1.096 = -0.182$$

the p–value, with $(n-1)ab = (2-1)(4)(3) = 12$ degrees of freedom, is given by

$$\text{p–value} = P(t \leq -0.182)$$

which equals (circle one) **0.00** / **0.36** / **0.43**.
The level of significance is 0.05.

(c) *Conclusion.*
Since the p–value, 0.43, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the ROC, subjected to factor B, level 2 (medium noise), is equal to 9.5 units.

4. *Confidence Interval, Multiple Pairwise, Factor A, Tukey.*
Since there are $a = 4$ levels of factor A (temperature), there are a total of $\frac{a(a-1)}{2} = \frac{4(3)}{2} = 6$ possible pairwise comparisons. In particular, determine a 95%

Tukey confidence interval[1] for $\mu_{3\cdot} - \mu_{4\cdot}$. Since

$$\begin{aligned} \hat{D} &= \overline{Y}_{3\cdot\cdot} - \overline{Y}_{4\cdot\cdot} \\ &= 3.9 - 16.3 \\ &= -12.4 \\ s\left\{\hat{D}\right\} &= \sqrt{\frac{2MSE}{bn}} \\ &= \sqrt{\frac{2(9.61)}{(3)(2)}} \\ &= 1.790 \\ q(1-\alpha; a, (n-1)ab) &= q(1-0.05; 4, (2-1)(4)(3)) \\ &= q(0.95; 4, 12) \\ &= 4.20 \end{aligned}$$

the confidence interval for $\mu_{3\cdot} - \mu_{4\cdot}$ is given by

$$\hat{D} \pm \frac{1}{\sqrt{2}} q(1-\alpha; a, (n-1)ab) s\left\{\hat{D}\right\} = -12.4 \pm \frac{1}{\sqrt{2}}(1.790)(4.20) =$$

(choose one) $\mathbf{(-0.88, 8.11)}$ / $\mathbf{(-2.88, 8.11)}$ / $\mathbf{(-17.72, -7.08)}$

5. *Test, Multiple Contrast, Bonferroni.*
   Let $L = -\frac{1}{2}\mu_{1\cdot} - \frac{1}{2}\mu_{2\cdot} + \mu_{3\cdot}$ and assume this is one of six ($g = 6$) contrasts. Test if contrast L is zero at $\alpha = 0.05$.

   (a) *Statement.*
   The statement of the test is (check none, one or more):

   i. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus $H_a :\ \alpha_1 \neq \alpha_2, \alpha_1 = \alpha_3$.
   ii. $H_0 : L = 0$ versus $H_a :\ L \neq 0$.
   iii. $H_0 : \mu_{\cdot 2} = 9.5$ versus
   $H_a :\ \mu_{\cdot 2} < 9.5$.
   iv. $H_0 : \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$ versus
   $H_a :$ at least one $\mu_{i\cdot} \neq \mu_{j\cdot}$, $i, j = 1, 2, 3, 4$.

[1]Since there is a small number of pairwise comparisons, it probably makes more sense to use the Bonferroni procedure, but let's calculate a Tukey multiple pairwise confidence interval anyway.

(b) *Test.*
Since

$$\begin{aligned}\hat{L} &= -\frac{1}{2}\overline{Y}_{1..} - \frac{1}{2}\overline{Y}_{2..} + \overline{Y}_{3..} \\ &= -\frac{1}{2}(6.7) - \frac{1}{2}(6.5) + 3.9 \\ &= -2.7 \\ s\left\{\hat{L}\right\} &= \sqrt{\frac{MSE}{bn}\sum c_i^2} \\ &= \sqrt{\frac{9.61}{(3)(2)}\left(\left(-\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + (1)^2\right)} \\ &= 1.55\end{aligned}$$

the test statistic is

$$|t^*| = [\hat{L}]/[s\left\{\hat{L}\right\}] = |-2.7/1.55| = 0.574$$

the p–value, with $(n-1)ab = (2-1)(4)(3) = 12$ degrees of freedom, is given by

$$\text{p–value} = P(t \geq 0.574)$$

which equals (circle one) **0.29** / **0.36** / **0.43**.
(Use 2nd DISTR $t$cdf(0.574,E99,12).)
The level of significance is

$$\frac{\alpha}{2g} = \frac{0.05}{2(6)} \approx 0.004$$

(c) *Conclusion.*
Since the p–value, 0.29, is larger than the *family* level of significance, 0.004, we (circle one) **accept** / **reject** the null hypothesis that the contrast is equal to zero.

6. *Comparing critical values for contrast CIs.*
The most *efficient* confidence interval is the one with the *smallest*[2] critical value.Using the TI–83 INVT and INVF programs appropriately, we can determine the following table of critical values.

[2]This is because, for the same sample size, the confidence interval will be the *narrowest* and so the most informative of the location of the unknown parameter. In addition to the Bonferroni and Scheffe critical values given in the table above, a modified Scheffe critical value is compared in the table below when determining the critical value for a combined factor A (three contrasts, say) and factor B (two contrasts, say) family at $\alpha 0.05$.

type	critical value
Bonferroni	$t(1-\alpha/2g;(n-1)ab) = t(1-0.05/[2(3+2)];(2-1)(4)(3)) = 3.055$
Scheffe factor A	$\sqrt{(a-1)F(1-\alpha/2;a-1;(n-1)ab)} = \sqrt{(4-1)F(1-0.05/2;4-1;(2-1)(4)(3))} = 3.663$
Scheffe factor B	$\sqrt{(b-1)F(1-\alpha/2;b-1;(n-1)ab)} = \sqrt{(3-1)F(1-0.05/2;3-1;(2-1)(4)(3))} = 3.192$
Scheffe (modified)	$\sqrt{(a+b-2)F(1-\alpha;a+b-2;(n-1)ab)} = \sqrt{(4+3-2)F(1-0.05;4+3-2;(2-1)(4)(3))} = 3.941$

The most efficient confidence interval is (choose one)

(i) Bonferroni

(ii) Scheffe factor A

(iii) Scheffe factor B

(iv) Scheffe (modified)

20.3 Analysis of Factor Effects when Interactions Important

SAS program: att3-20-3-deermice-analysis,interact

Since the interactions are important, it does not make sense to conduct inference on factor level means, or contrasts and combinations of factor level means. However, it does make sense to conduct inference on treatment *cell* means, as is given below.

Inference	confidence interval	test	
Multiple Pairwise Tukey (many)	$\hat{D} \pm \frac{1}{\sqrt{2}} q(1-\alpha;ab,(n-1)ab)s\left\{\hat{D}\right\}$	$q^* = \frac{\sqrt{2}\hat{D}}{s\{\hat{D}\}}$	$s\left\{\hat{D}\right\} = \sqrt{\frac{2MSE}{n}}$, $\hat{D} = \overline{Y}_{ij\cdot} - \overline{Y}_{i'j'\cdot}$
Multiple Pairwise Bonferroni (few)	$\hat{D} \pm t(1-\alpha/2g;(n-1)ab)s\left\{\hat{D}\right\}$	$t^* = \frac{\hat{D}}{s\{\hat{D}\}}$	$s\left\{\hat{D}\right\} = \sqrt{\frac{2MSE}{n}}$, $\hat{D} = \overline{Y}_{ij\cdot} - \overline{Y}_{i'j'\cdot}$
Multiple Contrast Scheffe (many)	$\hat{L} \pm \sqrt{(ab-1)F(1-\alpha;ab-1,(n-1)ab)}s\left\{\hat{L}\right\}$	$F^* = \frac{\hat{L}}{(ab-1)s^2\{\hat{L}\}}$	$s\left\{\hat{L}\right\} = \sqrt{\frac{MSE}{n}\sum c_{ij}^2}$, $\hat{L} = \sum\sum c_{ij}\overline{Y}_{ij\cdot}$
Multiple Contrast Bonferroni (few)	$\hat{L} \pm t(1-\alpha/2g;(n-1)ab)s\left\{\hat{L}\right\}$	$t^* = \frac{\hat{L}}{s\{\hat{L}\}}$	$s\left\{\hat{L}\right\} = \sqrt{\frac{MSE}{n}\sum c_{ij}^2}$, $\hat{L} = \sum\sum c_{ij}\overline{Y}_{ij\cdot}$

Exercise 20.2 (Analysis of Factor Effects when Interactions Important) Consider the effect of air temperature and noise on the ROC of deer mice.

	Factor B, noise →	$j = 1$, low	$j = 2$, medium	$j = 3$, high	row ave
Factor A,	$i = 1$, 0^o F	10.3, 7.2	9.1, 5.4	6.1, 2.1	$\overline{Y}_{1\cdot\cdot} = 6.7$
temperature	$i = 2$, 10^o F	1.8, 9.8	12.1, 4.2	5.1, 6.2	$\overline{Y}_{2\cdot\cdot} = 6.5$
	$i = 3$, 20^o F	1.2, 8.1	6.5, 4.1	1.2, 2.1	$\overline{Y}_{3\cdot\cdot} = 3.9$
	$i = 4$, 30^o F	12.4, 15.1	16.1, 17.2	18.1, 19.1	$\overline{Y}_{4\cdot\cdot} = 16.3$
	column ave	$\overline{Y}_{\cdot 1\cdot} = 8.2$	$\overline{Y}_{\cdot 2\cdot} = 9.3$	$\overline{Y}_{\cdot 3\cdot} = 7.5$	$\overline{Y}_{\cdot\cdot\cdot} = 8.4$

The ANOVA table is given by,

Source	*df*	*SS*	*MS*
Factor A (Temperature)	3	539.14	179.71
Factor B (Noise)	2	13.68	13.68
Interaction AB	6	55.70	55.70
Error	12	115.30	9.61
Total	23	723.82	

Also, a table of treatment *cell* means is given by

	noise 1	noise 2	noise 3
temperature 1	8.75	7.25	4.10
temperature 2	5.80	8.15	5.65
temperature 3	4.65	5.30	1.65
temperature 4	13.75	16.65	18.60

Although we know the interaction is *not* significant in this case, we will conduct various inferences as though it is, at $\alpha = 0.05$. In particular, we will look at comparing treatment *cell* means.

1. *Confidence Interval, Multiple Pairwise, Bonferroni.*
   Let $L = \mu_{23} - \mu_{13}$ and assume this is one of six ($g = 6$) contrasts under scrutiny. Calculate a 95% confidence interval for the contrast L. It makes sense to calculate a Bonferroni type of a multiple confidence interval because only a few (6) contrasts are under consideration. Since

$$\begin{aligned}
\hat{D} &= \overline{Y}_{ij\cdot} - \overline{Y}_{i'j'\cdot} \\
&= \overline{Y}_{23\cdot} - \overline{Y}_{13\cdot} \\
&= 5.65 - 4.10 \\
&= 1.55 \\
s\left\{\hat{D}\right\} &= \sqrt{\frac{2MSE}{n}} \\
&= \sqrt{\frac{2(9.61)}{2}} \\
&= 3.017 \\
t(1-\alpha/2g;(n-1)ab) &= t(1-0.05/[2(6)];(2-1)(4)(3)) \\
&= t(0.9958;12) \\
&= 3.148
\end{aligned}$$

   the confidence interval for $\mu_{23} - \mu_{13}$ is given by

$$\hat{D} \pm t(1-\alpha/2g;(n-1)ab)s\left\{\hat{D}\right\} = 1.55 \pm (3.017)(3.148) =$$

   (choose one) **(−7.95, 11.05)** / **(−2.88, 8.11)** / **(−4.88, 8.11)**

2. *Test, Multiple Contrast, Scheffe.*
   Let $L = (\mu_{23} - \mu_{13}) - (\mu_{21} - \mu_{11})$ and assume this is one of six ($g = 6$) contrasts. This is not a large number of contrasts, but perform a Scheffe test anyway. Test if contrast L is zero at $\alpha = 0.05$.

   (a) *Statement.*
   The statement of the test is (check none, one or more):

   i. $H_0 : (\mu_{23} - \mu_{13}) - (\mu_{21} - \mu_{11}) = 0$ versus
   $H_a : (\mu_{23} - \mu_{13}) - (\mu_{21} - \mu_{11}) \neq 0.$

   ii. $H_0 : L = 0$ versus $H_a : L \neq 0.$

   iii. $H_0 : \mu_{\cdot 2} = 9.5$ versus $H_a : \mu_{\cdot 2} < 9.5.$

   iv. $H_0 : \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$ versus
   $H_a :$ at least one $\mu_{i\cdot} \neq \mu_{j\cdot},\ i, j = 1, 2, 3, 4.$

   (b) *Test.*
   Since

$$\begin{aligned} \hat{L} &= \left(\overline{Y}_{23\cdot} - \overline{Y}_{13\cdot}\right) - \left(\overline{Y}_{21\cdot} - \overline{Y}_{11\cdot}\right) \\ &= (5.65 - 4.10) - (5.80 - 8.75) \\ &= 4.5 \\ s\left\{\hat{L}\right\} &= \sqrt{\frac{MSE}{n}\sum\sum c_i^2} \\ &= \sqrt{\frac{9.61}{2}(1^2 + (-1)^2 + (-1)^2 + 1^2)} \\ &= 4.38 \end{aligned}$$

   the test statistic is

$$F^* = [\hat{L}]/[(ab-1)s^2\left\{\hat{L}\right\}] = \frac{4.5}{[(4)(3)-1]4.38^2} = 0.021$$

   the p–value, with $ab - 1 = 11$ and $(n-1)ab = (2-1)(4)(3) = 12$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 0.021)$$

   which equals (circle one) **0.29** / **0.36** / **0.99**.
   (Use 2nd DISTR Fcdf(0.021,E99,11,12).)
   The level of significance is 0.05.

   (c) *Conclusion.*
   Since the p–value, 0.99, is larger than the *family* level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the contrast is equal to zero.

20.4 Analysis when One or Both Factors Quantitative

Not covered.