

R09 - Analysis of Experiments with Two Factors

Two-way ANOVA and Contrasts

STAT 5870 (Engineering)
Iowa State University

November 22, 2024

Two factors

Consider the question of the affect of variety and density on yield under various experimental designs:

- Balanced, complete design
- Unbalanced, complete
- Incomplete

We will also consider the problem of finding the density that maximizes yield.

Data

An experiment was run on tomato plants to determine the effect of

- 3 different varieties (A,B,C) and
- 4 different planting densities (10,20,30,40)

on yield.

A **balanced completely randomized** design (CRD) with **replication** was used.

- complete: each treatment (variety \times density) is represented
- balanced: each treatment has the same number of replicates
- randomized: treatment was randomly assigned to the plot
- replication: each treatment is represented more than once

This is also referred to as a **full factorial** or **fully crossed** design.

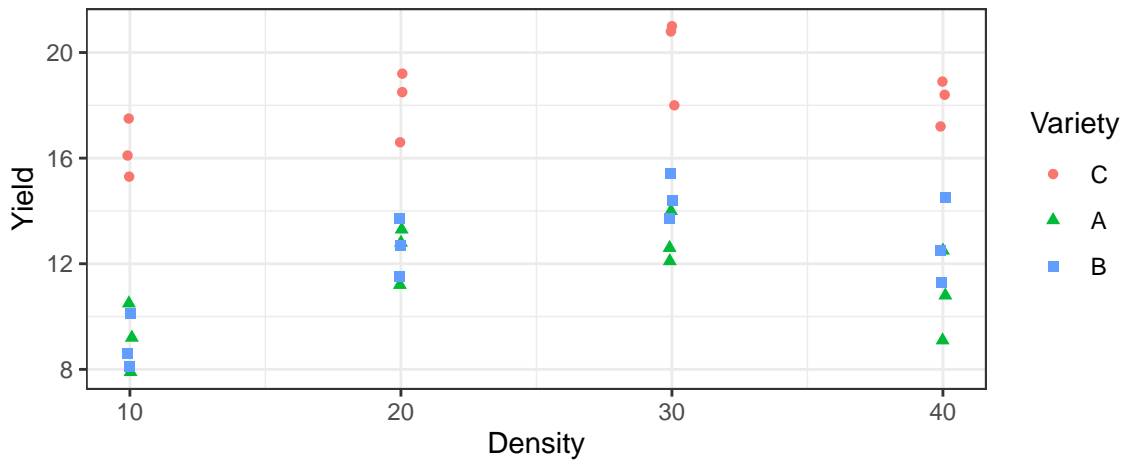
Hypotheses

- How does variety affect mean yield?
 - How is the mean yield for variety A different from B **on average**?
 - How is the mean yield for variety A different from B **at a particular value for density**?
- How does density affect mean yield?
 - How is the mean yield for density 10 different from density 20 **on average**?
 - How is the mean yield for density 10 different from density 20 **at a particular value for variety**?
- How does density affect yield differently for each variety?

For all of these questions, we want to know

- is there any effect and
- if yes, what is the magnitude and direction of the effect.

Confidence/credible intervals can answer these questions.



Summary statistics

```
# A tibble: 12 x 5
# Groups:   Variety [3]
  Variety Density    n mean    sd
  <fct>    <int> <int> <dbl> <dbl>
1 C         10     3 16.3  1.11
2 C         20     3 18.1  1.35
3 C         30     3 19.9  1.68
4 C         40     3 18.2  0.874
5 A         10     3  9.2  1.3
6 A         20     3 12.4  1.10
7 A         30     3 12.9  0.985
8 A         40     3 10.8  1.7
9 B         10     3  8.93 1.04
10 B        20     3 12.6  1.10
11 B        30     3 14.5  0.854
12 B        40     3 12.8  1.62
```

Two-way ANOVA

- Setup: Two categorical explanatory variables with I and J levels respectively
- Model:

$$Y_{ijk} \stackrel{ind}{\sim} N(\mu_{ij}, \sigma^2)$$

where Y_{ijk} is the

- k th observation at the
- i th level of variable 1 (variety) with $i = 1, \dots, I$ and the
- j th level of variable 2 (density) with $j = 1, \dots, J$.

Consider the models:

- Additive/Main effects: $\mu_{ij} = \mu + \nu_i + \delta_j$
- Cell-means: $\mu_{ij} = \mu + \nu_i + \delta_j + \gamma_{ij}$

	10	20	30	40
A	μ_{11}	μ_{12}	μ_{13}	μ_{14}
B	μ_{21}	μ_{22}	μ_{23}	μ_{24}
C	μ_{31}	μ_{32}	μ_{33}	μ_{34}

As a regression model

1. Assign a reference level for both variety (C) and density (40).
2. Let V_i and D_i be the variety and density for observation i .
3. Build indicator variables, e.g. $I(V_i = A)$ and $I(D_i = 10)$.
4. The additive/main effects model:

$$\begin{aligned}\mu_i = & \beta_0 \\ & + \beta_1 I(V_i = A) + \beta_2 I(V_i = B) \\ & + \beta_3 I(D_i = 10) + \beta_4 I(D_i = 20) + \beta_5 I(D_i = 30).\end{aligned}$$

5. The cell-means model:

$$\begin{aligned}\mu_i = & \beta_0 \\ & + \beta_1 I(V_i = A) + \beta_2 I(V_i = B) \\ & + \beta_3 I(D_i = 10) + \beta_4 I(D_i = 20) + \beta_5 I(D_i = 30) \\ & + \beta_6 I(V_i = A)I(D_i = 10) + \beta_7 I(V_i = A)I(D_i = 20) + \beta_8 I(V_i = A)I(D_i = 30) \\ & + \beta_9 I(V_i = B)I(D_i = 10) + \beta_{10} I(V_i = B)I(D_i = 20) + \beta_{11} I(V_i = B)I(D_i = 30)\end{aligned}$$

ANOVA Table

ANOVA Table - Additive/Main Effects model

Source	SS	df	MS	F
Factor A	SSA	(I-1)	$SSA/(I-1)$	MSA/MSE
Factor B	SSB	(J-1)	$SSB/(J-1)$	MSB/MSE
Error	SSE	$n-I-J+1$	$SSE/(n-I-J+1)$	
Total	SST	$n-1$		

ANOVA Table - Cell-means model

Source	SS	df	MS	
Factor A	SSA	I-1	$SSA/(I-1)$	MSA/MSE
Factor B	SSB	J-1	$SSB/(J-1)$	MSB/MSE
Interaction AB	SSAB	$(I-1)(J-1)$	$SSAB/(I-1)(J-1)$	$MSAB/MSE$
Error	SSE	$n-IJ$	$SSE/(n-IJ)$	
Total	SST	$n-1$		

```
tomato$Density = factor(tomato$Density)
m = lm(Yield~Variety+Density, tomato)
drop1(m, test="F")
```

Single term deletions

Model:

```
Yield ~ Variety + Density
      Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>      46.07 20.880
Variety  2    327.60 373.67 92.235 106.659 2.313e-14 ***
Density  3     86.69 132.76 52.980  18.816 4.690e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m = lm(Yield~Variety*Density, tomato)
drop1(m, scope = ~Variety+Density+Variety:Density, test="F")
```

Single term deletions

Model:

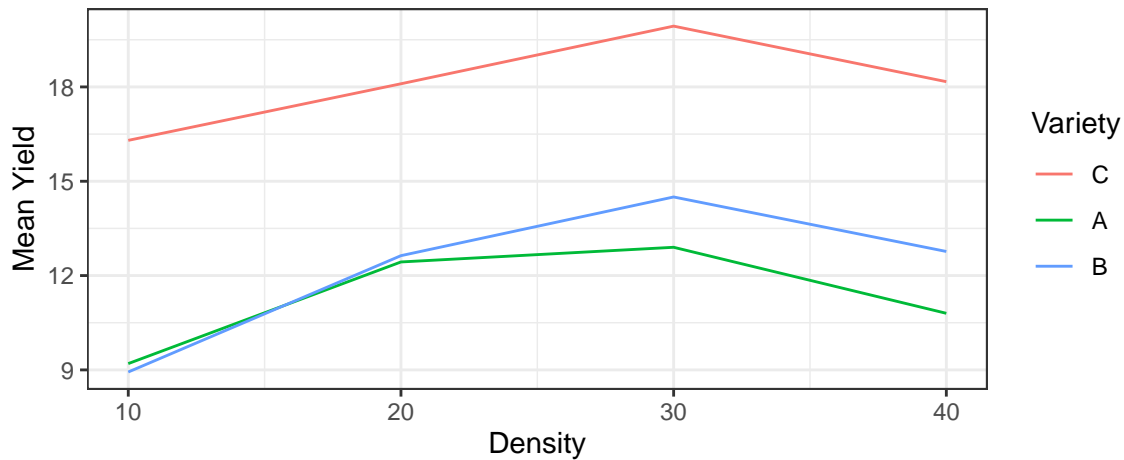
```
Yield ~ Variety * Density
      Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>      38.040 25.984
Variety  2    104.749 142.789 69.603 33.0438 1.278e-07 ***
Density  3     19.809  57.849 35.076  4.1660  0.01648 *
Variety:Density 6      8.032  46.072 20.880  0.8445  0.54836
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Additive vs cell-means

Opinions differ on whether to use an additive vs a cell-means model when the interaction is not significant. Remember that an insignificant test does not prove that there is no interaction.

	Additive	Cell-means
Interpretation	Direct	More complicated
Estimate of σ^2	Biased	Unbiased

We will continue using the cell-means model to answer the scientific questions of interest.



Two-way ANOVA in R

```
tomato$Density = factor(tomato$Density)
m = lm(Yield~Variety*Density, tomato)
anova(m)
```

Analysis of Variance Table

Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety	2	327.60	163.799	103.3430	1.608e-12 ***
Density	3	86.69	28.896	18.2306	2.212e-06 ***
Variety:Density	6	8.03	1.339	0.8445	0.5484
Residuals	24	38.04	1.585		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Variety comparison

```
library(emmeans)
emmeans(m, pairwise~Variety)
```

```
$emmeans
  Variety emmean    SE df lower.CL upper.CL
C          18.1 0.363 24     17.4     18.9
A          11.3 0.363 24     10.6     12.1
B          12.2 0.363 24     11.5     13.0
```

Results are averaged over the levels of: Density
Confidence level used: 0.95

```
$contrasts
  contrast estimate    SE df t.ratio p.value
C - A         6.792 0.514 24   13.214 <.0001
C - B         5.917 0.514 24   11.512 <.0001
A - B        -0.875 0.514 24    -1.702 0.2249
```

Results are averaged over the levels of: Density
P value adjustment: tukey method for comparing a family of 3 estimates

Density comparison

```
emmeans(m, pairwise~Density)
```

```
$emmeans
```

Density	emmean	SE	df	lower.CL	upper.CL
10	11.5	0.42	24	10.6	12.3
20	14.4	0.42	24	13.5	15.3
30	15.8	0.42	24	14.9	16.6
40	13.9	0.42	24	13.0	14.8

Results are averaged over the levels of: Variety

Confidence level used: 0.95

```
$contrasts
```

contrast	estimate	SE	df	t.ratio	p.value
Density10 - Density20	-2.911	0.593	24	-4.905	0.0003
Density10 - Density30	-4.300	0.593	24	-7.245	<.0001
Density10 - Density40	-2.433	0.593	24	-4.100	0.0022
Density20 - Density30	-1.389	0.593	24	-2.340	0.1169
Density20 - Density40	0.478	0.593	24	0.805	0.8514
Density30 - Density40	1.867	0.593	24	3.145	0.0213

Results are averaged over the levels of: Variety

P value adjustment: tukey method for comparing a family of 4 estimates

```
emmeans(m, pairwise~Variety*Density)
```

```
$emmeans
```

Variety	Density	emmean	SE	df	lower.CL	upper.CL
C	10	16.30	0.727	24	14.80	17.8
A	10	9.20	0.727	24	7.70	10.7
B	10	8.93	0.727	24	7.43	10.4
C	20	18.10	0.727	24	16.60	19.6
A	20	12.43	0.727	24	10.93	13.9
B	20	12.63	0.727	24	11.13	14.1
C	30	19.93	0.727	24	18.43	21.4
A	30	12.90	0.727	24	11.40	14.4
B	30	14.50	0.727	24	13.00	16.0
C	40	18.17	0.727	24	16.67	19.7
A	40	10.80	0.727	24	9.30	12.3
B	40	12.77	0.727	24	11.27	14.3

Confidence level used: 0.95

```
$contrasts
```

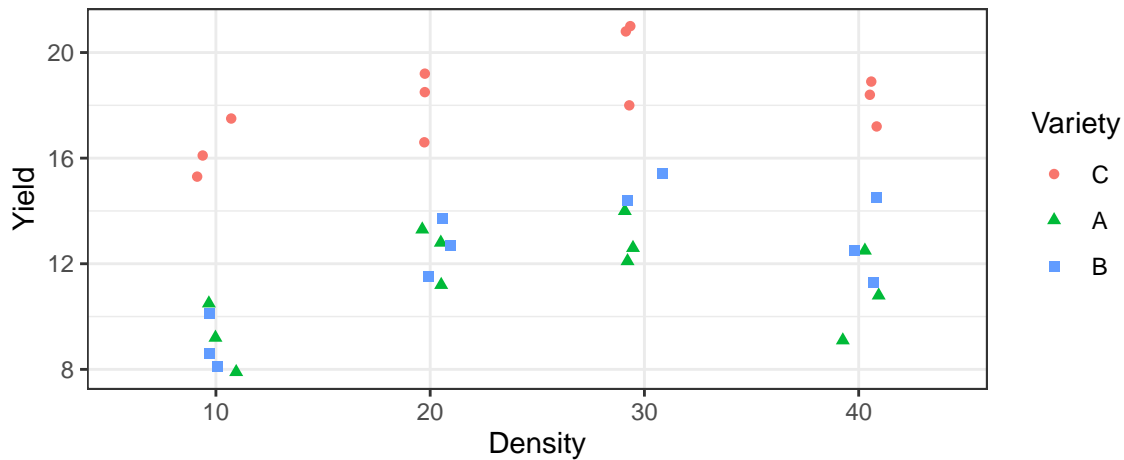
contrast	estimate	SE	df	t.ratio	p.value
C Density10 - A Density10	7.1000	1.03	24	6.907	<.0001
C Density10 - B Density10	7.3667	1.03	24	7.166	<.0001
C Density10 - C Density20	-1.8000	1.03	24	-1.751	0.8276
C Density10 - A Density20	3.8667	1.03	24	3.762	0.0356
C Density10 - B Density20	3.6667	1.03	24	3.567	0.0543
C Density10 - C Density30	-3.6333	1.03	24	-3.535	0.0582
C Density10 - A Density30	3.4000	1.03	24	3.308	0.0932
C Density10 - B Density30	1.8000	1.03	24	1.751	0.8276
C Density10 - C Density40	-1.8667	1.03	24	-1.816	0.7947
C Density10 - A Density40	5.5000	1.03	24	5.350	0.0008
C Density10 - B Density40	3.5333	1.03	24	3.437	0.0714

Summary

- Use `emmeans` to answer questions of scientific interest.
- Check model assumptions
- Consider alternative models, e.g. treating density as continuous

Unbalanced design

Suppose for some reason that a variety B, density 30 sample was contaminated. Although you started with a balanced design, the data is now unbalanced. Fortunately, we can still use the tools we have used previously.



Summary statistics

```
# A tibble: 12 x 5
# Groups:   Variety [3]
  Variety Density    n mean    sd
  <fct>    <fct>  <int> <dbl> <dbl>
1 C      10      3 16.3  1.11
2 C      20      3 18.1  1.35
3 C      30      3 19.9  1.68
4 C      40      3 18.2  0.874
5 A      10      3  9.2  1.3
6 A      20      3 12.4  1.10
7 A      30      3 12.9  0.985
8 A      40      3 10.8  1.7
9 B      10      3  8.93 1.04
10 B     20      3 12.6  1.10
11 B     30      2 14.9  0.707
12 B     40      3 12.8  1.62
```

Two-way ANOVA in R

```
m = lm(Yield~Variety*Density, tomato_unbalanced)
anova(m)
```

Analysis of Variance Table

Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety	2	329.99	164.994	102.343	3.552e-12 ***
Density	3	84.45	28.150	17.461	3.947e-06 ***
Variety:Density	6	8.80	1.467	0.910	0.5052
Residuals	23	37.08	1.612		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Variety comparison

```
emmeans(m, pairwise~Variety)
```

```
$emmeans
```

Variety	emmean	SE	df	lower.CL	upper.CL
C	18.1	0.367	23	17.4	18.9
A	11.3	0.367	23	10.6	12.1
B	12.3	0.389	23	11.5	13.1

Results are averaged over the levels of: Density

Confidence level used: 0.95

```
$contrasts
```

contrast	estimate	SE	df	t.ratio	p.value
C - A	6.792	0.518	23	13.102	<.0001
C - B	5.817	0.534	23	10.886	<.0001
A - B	-0.975	0.534	23	-1.825	0.1839

Results are averaged over the levels of: Density

P value adjustment: tukey method for comparing a family of 3 estimates

Density comparison

```
emmeans(m, pairwise~Density)
```

```
$emmeans
```

Density	emmean	SE	df	lower.CL	upper.CL
10	11.5	0.423	23	10.6	12.4
20	14.4	0.423	23	13.5	15.3
30	15.9	0.457	23	15.0	16.9
40	13.9	0.423	23	13.0	14.8

Results are averaged over the levels of: Variety

Confidence level used: 0.95

```
$contrasts
```

contrast	estimate	SE	df	t.ratio	p.value
Density10 - Density20	-2.911	0.599	23	-4.864	0.0004
Density10 - Density30	-4.433	0.623	23	-7.116	<.0001
Density10 - Density40	-2.433	0.599	23	-4.065	0.0025
Density20 - Density30	-1.522	0.623	23	-2.443	0.0967
Density20 - Density40	0.478	0.599	23	0.798	0.8545
Density30 - Density40	2.000	0.623	23	3.210	0.0189

Results are averaged over the levels of: Variety

P value adjustment: tukey method for comparing a family of 4 estimates

```
emmeans(m, pairwise~Variety*Density)
```

```
$emmeans
```

Variety	Density	emmean	SE	df	lower.CL	upper.CL
C	10	16.30	0.733	23	14.78	17.8
A	10	9.20	0.733	23	7.68	10.7
B	10	8.93	0.733	23	7.42	10.4
C	20	18.10	0.733	23	16.58	19.6
A	20	12.43	0.733	23	10.92	13.9
B	20	12.63	0.733	23	11.12	14.1
C	30	19.93	0.733	23	18.42	21.4
A	30	12.90	0.733	23	11.38	14.4
B	30	14.90	0.898	23	13.04	16.8
C	40	18.17	0.733	23	16.65	19.7
A	40	10.80	0.733	23	9.28	12.3
B	40	12.77	0.733	23	11.25	14.3

Confidence level used: 0.95

```
$contrasts
```

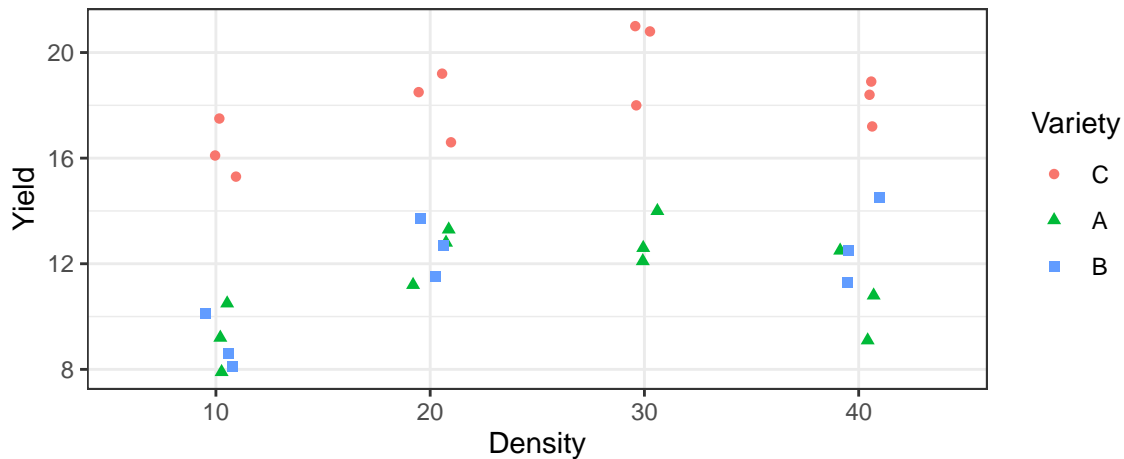
contrast	estimate	SE	df	t.ratio	p.value
C Density10 - A Density10	7.1000	1.04	23	6.849	<.0001
C Density10 - B Density10	7.3667	1.04	23	7.106	<.0001
C Density10 - C Density20	-1.8000	1.04	23	-1.736	0.8341
C Density10 - A Density20	3.8667	1.04	23	3.730	0.0396
C Density10 - B Density20	3.6667	1.04	23	3.537	0.0597
C Density10 - C Density30	-3.6333	1.04	23	-3.505	0.0638
C Density10 - A Density30	3.4000	1.04	23	3.280	0.1008
C Density10 - B Density30	1.4000	1.16	23	1.208	0.9828
C Density10 - C Density40	-1.8667	1.04	23	-1.801	0.8022
C Density10 - A Density40	5.5000	1.04	23	5.305	0.0011
C Density10 - B Density40	3.5333	1.04	23	3.408	0.0778

Unbalanced Summary

The analysis can be completed just like the balanced design using `emmeans` to answer scientific questions of interest.

Incomplete design

Suppose none of the samples from variety B, density 30 were obtained. Now the analysis becomes more complicated.



Summary statistics

```
# A tibble: 11 x 5
# Groups:   Variety [3]
  Variety Density    n  mean    sd
  <fct>    <fct>  <int> <dbl> <dbl>
1 C      10      3 16.3  1.11
2 C      20      3 18.1  1.35
3 C      30      3 19.9  1.68
4 C      40      3 18.2  0.874
5 A      10      3  9.2  1.3
6 A      20      3 12.4  1.10
7 A      30      3 12.9  0.985
8 A      40      3 10.8  1.7
9 B      10      3  8.93 1.04
10 B     20      3 12.6  1.10
11 B     40      3 12.8  1.62
```

Treat as a One-way ANOVA

When the design is incomplete, use a one-way ANOVA combined with contrasts to answer questions of interest. For example, to compare the average difference between B and C, we want to only compare at densities 10, 20, and 40.

	10	20	30	40
A	μ_{11}	μ_{12}	μ_{13}	μ_{14}
B	μ_{21}	μ_{22}		μ_{24}
C	μ_{31}	μ_{32}	μ_{33}	μ_{34}

Thus, the contrast is

$$\begin{aligned}
 \gamma &= \frac{1}{3}(\mu_{31} + \mu_{32} + \mu_{34}) - \frac{1}{3}(\mu_{21} + \mu_{22} + \mu_{24}) \\
 &= \frac{1}{3}(\mu_{31} + \mu_{32} + \mu_{34} - \mu_{21} - \mu_{22} - \mu_{24})
 \end{aligned}$$

The Regression model

The regression model here considers variety-density combination as a single explanatory variable with 11 levels: A10, A20, A30, A40, B10, B20, B40, C10, C20, C30, and C40. Let C40 be the reference level. For observation i , let

- Y_i be the yield
- V_i be the variety
- D_i be the density

The model is then $Y_i \stackrel{ind}{\sim} N(\mu_i, \sigma^2)$ and

$$\begin{aligned} \mu_i &= \beta_0 \\ &+ \beta_1 I(V_i = A, D_i = 10) + \beta_2 I(V_i = A, D_i = 20) + \beta_3 I(V_i = A, D_i = 30) + \beta_4 I(V_i = A, D_i = 40) \\ &+ \beta_5 I(V_i = B, D_i = 10) + \beta_6 I(V_i = B, D_i = 20) + \beta_7 I(V_i = B, D_i = 40) \\ &+ \beta_8 I(V_i = C, D_i = 10) + \beta_9 I(V_i = C, D_i = 20) + \beta_{10} I(V_i = C, D_i = 30) \end{aligned}$$

Two-way ANOVA in R

```
m <- lm(Yield ~ Variety*Density, data=tomato_incomplete)
anova(m)
```

Analysis of Variance Table

Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety	2	347.38	173.691	104.462	5.868e-12 ***
Density	3	66.65	22.218	13.362	3.514e-05 ***
Variety:Density	5	7.06	1.412	0.849	0.53
Residuals	22	36.58	1.663		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

How can you tell the design is not complete?

One-way ANOVA in R

```
m = lm(Yield~Variety:Density, tomato_incomplete)
anova(m)
```

Analysis of Variance Table

Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety:Density	10	421.09	42.109	25.326	8.563e-10 ***
Residuals	22	36.58	1.663		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Contrasts

```
m = lm(Yield ~ VarietyDensity, tomato_incomplete)
em <- emmeans(m, ~ VarietyDensity)
contrast(em, method = list(
#           A10 A20 A30 A40 B10 B20      B40 C10 C20 C30 C40
"C-B" = c( 0,  0,  0,  0, -1, -1,      -1,  1,  1,  0,  1)/3,
"C-A" = c(-1, -1, -1, -1,  0,  0,       0,  1,  1,  1,  1)/4,
"B-A" = c(-1, -1,  0, -1,  1,  1,       1,  0,  0,  0,  0)/3)) |>
  confint()
```

contrast	estimate	SE	df	lower.CL	upper.CL
C-B	6.078	0.608	22	4.817	7.34
C-A	6.792	0.526	22	5.700	7.88
B-A	0.633	0.608	22	-0.627	1.89

Confidence level used: 0.95

```
m = lm(Yield~Variety:Density, tomato_incomplete)
emmeans(m, pairwise~Variety:Density) # We could have used the VarietyDensity model, but this looks nicer
```

```
$emmeans
```

Variety	Density	emmean	SE	df	lower.CL	upper.CL
C	10	16.30	0.744	22	14.76	17.8
A	10	9.20	0.744	22	7.66	10.7
B	10	8.93	0.744	22	7.39	10.5
C	20	18.10	0.744	22	16.56	19.6
A	20	12.43	0.744	22	10.89	14.0
B	20	12.63	0.744	22	11.09	14.2
C	30	19.93	0.744	22	18.39	21.5
A	30	12.90	0.744	22	11.36	14.4
B	30	nonEst	NA	NA	NA	NA
C	40	18.17	0.744	22	16.62	19.7
A	40	10.80	0.744	22	9.26	12.3
B	40	12.77	0.744	22	11.22	14.3

```
Confidence level used: 0.95
```

```
$contrasts
```

contrast	estimate	SE	df	t.ratio	p.value
C Density10 - A Density10	7.1000	1.05	22	6.744	<.0001
C Density10 - B Density10	7.3667	1.05	22	6.997	<.0001
C Density10 - C Density20	-1.8000	1.05	22	-1.710	0.8157
C Density10 - A Density20	3.8667	1.05	22	3.673	0.0407
C Density10 - B Density20	3.6667	1.05	22	3.483	0.0606
C Density10 - C Density30	-3.6333	1.05	22	-3.451	0.0646
C Density10 - A Density30	3.4000	1.05	22	3.229	0.1007
C Density10 - B Density30	nonEst	NA	NA	NA	NA
C Density10 - C Density40	-1.8667	1.05	22	-1.773	0.7829
C Density10 - A Density40	5.5000	1.05	22	5.224	0.0012

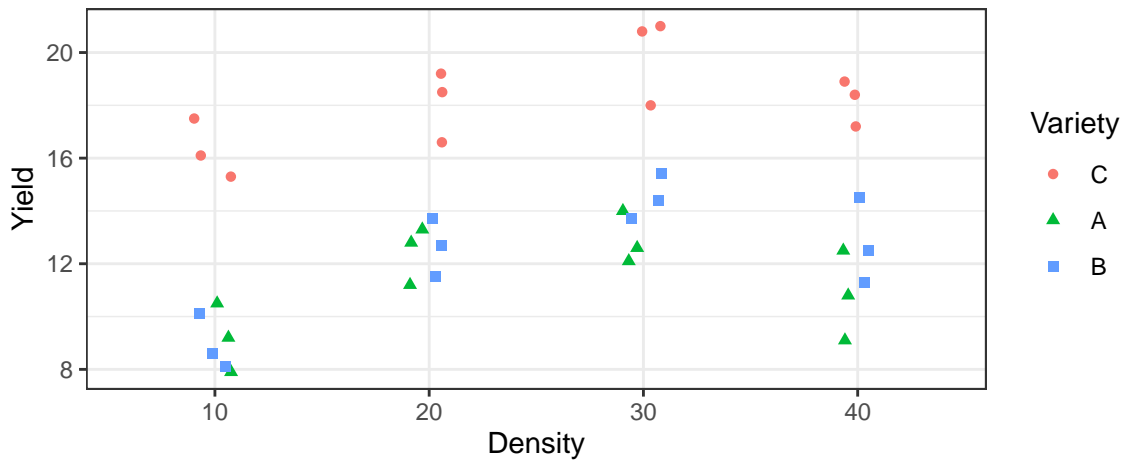
Summary

When dealing with an incomplete design, it is often easier to treat the analysis as a one-way ANOVA and use contrasts to answer scientific questions of interest.

Optimal yield

Now suppose you have the same data set, but your scientific question is different. Specifically, you are interested in choosing a variety-density combination that provides the optimal yield.

You can use the ANOVA analysis to choose from amongst the 3 varieties and one of the 4 densities, but there is no reason to believe that the optimal density will be one of those 4.



Modeling

Considering a single variety, if we assume a linear relationship between Yield (Y_i) and Density (D_i) then the maximum Yield will occur at either $-\infty$ or $+\infty$ which is unreasonable. The easiest way to have a maximum (or minimum) is to assume a quadratic relationship, e.g.

$$E[Y_i] = \mu_i = \beta_0 + \beta_1 D_i + \beta_2 D_i^2$$

Now we can incorporate Variety (V_i) in many ways. Two options are parallel curves or completely independent curves.

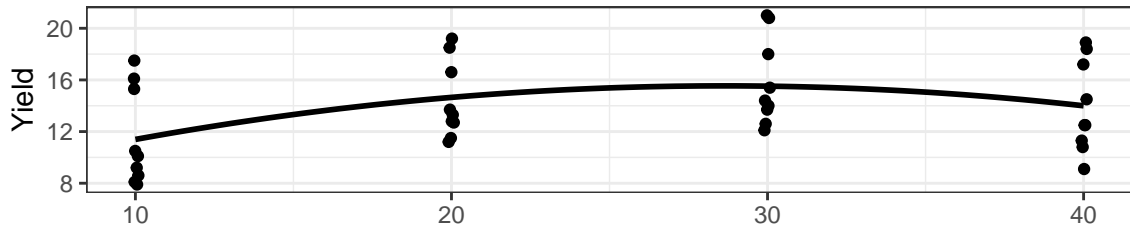
Parallel curves:

$$\mu_i = \beta_0 + \beta_1 D_i + \beta_2 D_i^2 + \beta_3 I(V_i = A) + \beta_4 I(V_i = B)$$

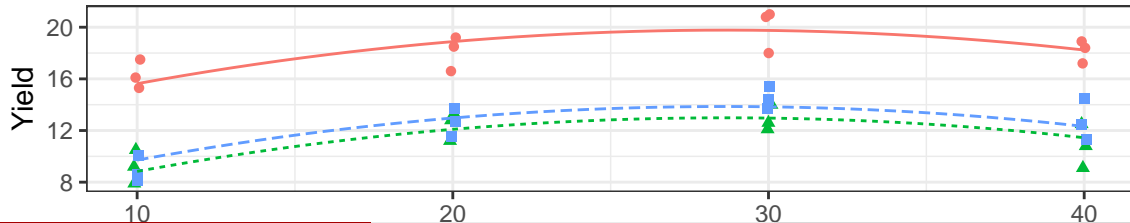
Independent curves:

$$\begin{aligned} \mu_i = & \beta_0 + \beta_1 D_i + \beta_2 D_i^2 \\ & + \beta_3 I(V_i = A) + \beta_4 I(V_i = B) \\ & + \beta_5 I(V_i = A) D_i + \beta_6 I(V_i = B) D_i \\ & + \beta_7 I(V_i = A) D_i^2 + \beta_8 I(V_i = B) D_i^2 \end{aligned}$$

No variety



Parallel curves



Finding the maximum

For a particular variety, there will be an equation like

$$E[Y_i] = \mu_i = \beta_0 + \beta_1 D_i + \beta_2 D_i^2$$

where these β_1 and β_2 need not correspond to any particular β_1 and β_2 we have discussed thus far.

If $\beta_2 < 0$, then the quadratic curve has a maximum and it occurs at $-\beta_1/2\beta_2$.

No variety

```
Call:
lm(formula = Yield ~ Density + I(Density^2), data = tomato)

Residuals:
    Min       1Q   Median       3Q      Max
-4.898 -2.721 -1.320  3.364  6.109

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.744444    3.128242   1.836  0.0753 .
Density        0.684111    0.285384   2.397  0.0223 *
I(Density^2) -0.011944    0.005618  -2.126  0.0411 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.371 on 33 degrees of freedom
Multiple R-squared:  0.1854, Adjusted R-squared:  0.136
F-statistic: 3.755 on 2 and 33 DF,  p-value: 0.03395
```

Parallel curves

```
Call:
lm(formula = Yield ~ Density + I(Density^2) + Variety, data = tomato)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.3422	-0.9039	0.1744	0.8082	2.1828

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	9.980556	1.184193	8.428	1.61e-09	***
Density	0.684111	0.104707	6.534	2.71e-07	***
I(Density^2)	-0.011944	0.002061	-5.794	2.21e-06	***
VarietyA	-6.791667	0.504942	-13.450	1.76e-14	***
VarietyB	-5.916667	0.504942	-11.718	6.39e-13	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.237 on 31 degrees of freedom

Multiple R-squared: 0.897, Adjusted R-squared: 0.8837

F-statistic: 67.48 on 4 and 31 DF, p-value: 7.469e-15

Independent curves

```
Call:
lm(formula = Yield ~ Density * Variety + I(Density^2) * Variety,
    data = tomato)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.04500	-0.82125	-0.01417	0.94000	1.71000

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	11.808333	1.968364	5.999	2.12e-06	***
Density	0.520167	0.179570	2.897	0.00739	**
VarietyA	-8.458333	2.783687	-3.039	0.00523	**
VarietyB	-9.733333	2.783687	-3.497	0.00165	**
I(Density^2)	-0.008917	0.003535	-2.522	0.01787	*
Density:VarietyA	0.199167	0.253951	0.784	0.43971	
Density:VarietyB	0.292667	0.253951	1.152	0.25924	
VarietyA:I(Density^2)	-0.004417	0.005000	-0.883	0.38482	
VarietyB:I(Density^2)	-0.004667	0.005000	-0.933	0.35889	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.225 on 27 degrees of freedom

Multiple R-squared: 0.912, Adjusted R-squared: 0.886

F-statistic: 34.99 on 8 and 27 DF, p-value: 2.678e-12

Completely randomized design (CRD)

This semester, we have assumed a completely randomized design. As an example, consider 36 plots and we are randomly assigning our variety-density combinations to the plots such that we have 3 reps of each combination. The result may look something like this

A20	A10	A20	B10	B10	A30
C10	C30	C30	C10	C20	A10
B30	B10	B20	B30	B40	B40
C40	B20	A10	C20	B30	A40
C30	B40	A30	C40	B20	C40
C10	C20	A40	A30	A20	A40

Complete randomized block design (RBD)

A randomized block design is appropriate when there is a nuisance factor that you want to control for. In our example, imagine you had 12 plots at 3 different locations and you expect these locations would have impact on yield. A randomized block design might look like this.

B10	B40
C30	A30
C40	C10
A20	B20
B30	A40
A10	C20

Block 1

C20	B40
C30	A30
C10	B10
A10	A20
B20	C40
B30	A40

Block 2

A20	B30
C10	A30
A10	C30
B20	C40
B40	A40
C20	B10

Block 3

RBD Analysis

Generally, you will want to model a randomized block design using an additive model for the treatment and blocking factor. If you have the replication, you should test for an interaction. Let's compute the degrees of freedom for the ANOVA tables for this current design considering the variety-density combination as the treatment.

V+D+B		T+B		Cell-means	
Factor	df	Factor	df	Factor	df
Variety	2				
Density	3	Treatment	11	Treatment	11
Block	2	Block	2	Block	2
				Treatment x Block	22
Error	28	Error	22	Error	0
Total	35	Total	35	Total	35

The cell-means model does not have enough degrees of freedom to estimate the interaction because there is no replication of the treatment within a block.

Why block?

Consider a simple experiment with 2 blocks each with 3 experimental units and 3 treatments (A, B, C).

Blocked		Unblocked	
A	C	A	C
B	B	B	B
A	C	C	A
Block 1	Block 2	Block 1	Block 2

Let's consider 3 possible analyses:

- Blocked experiment using an additive model for treatment and block (RBD)
- Unblocked experiment using only treatment (CRD)

Why block?

Now suppose, the true model is

$$\mu_{ij} = \mu + T_i + B_j$$

where $T_1 = T_2 = T_3$ and $B_1 = 0$ and $B_2 = \delta$.

In the Blocked experiment using an additive model for treatment and block, the expected treatment differences to all be zero.

In the Unblocked design using only treatment, the expected difference between treatments is

$$\mu_C - \mu_B = \delta \quad \text{and} \quad \mu_C - \mu_A = \delta/2.$$

In the Unblocked design using an additive model for treatment and block, we would have an unbalanced design and it would be impossible to compare B and C.

Summary

Block what you can control; randomize what you cannot.