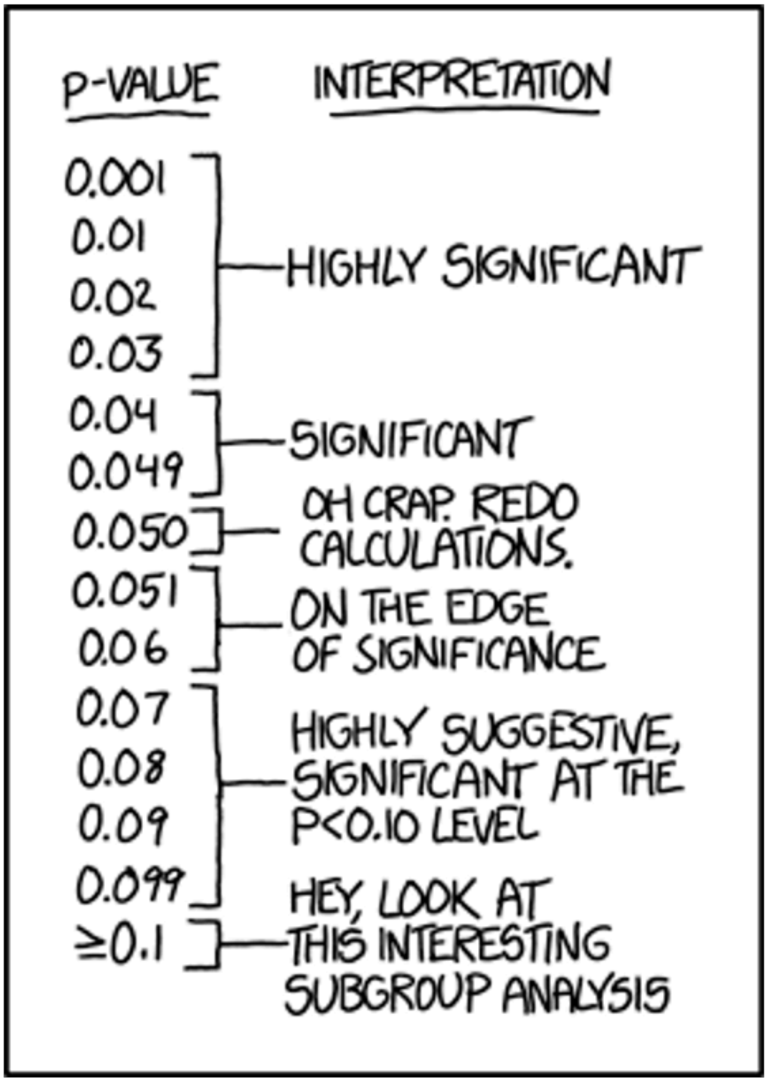
Lecture 13 - Multifactor ANOVA

Bill Perry

Lecture 13: Review

Multifactor ANOVA

- Example
- Linear model
- Analysis of variance
- Null hypotheses
- Interactions and main effects
- Unequal sample size
- Assumptions



If all else fails, use "significant at a p>0.05 level" and hope no one notices.

Lecture 13: 2 Factor or 2 Way ANOVA

Often consider more than 1 factor (independent categorical variable):

- reduce unexplained variance
- look at interactions

2-factor designs (2-way ANOVA) very common in ecology

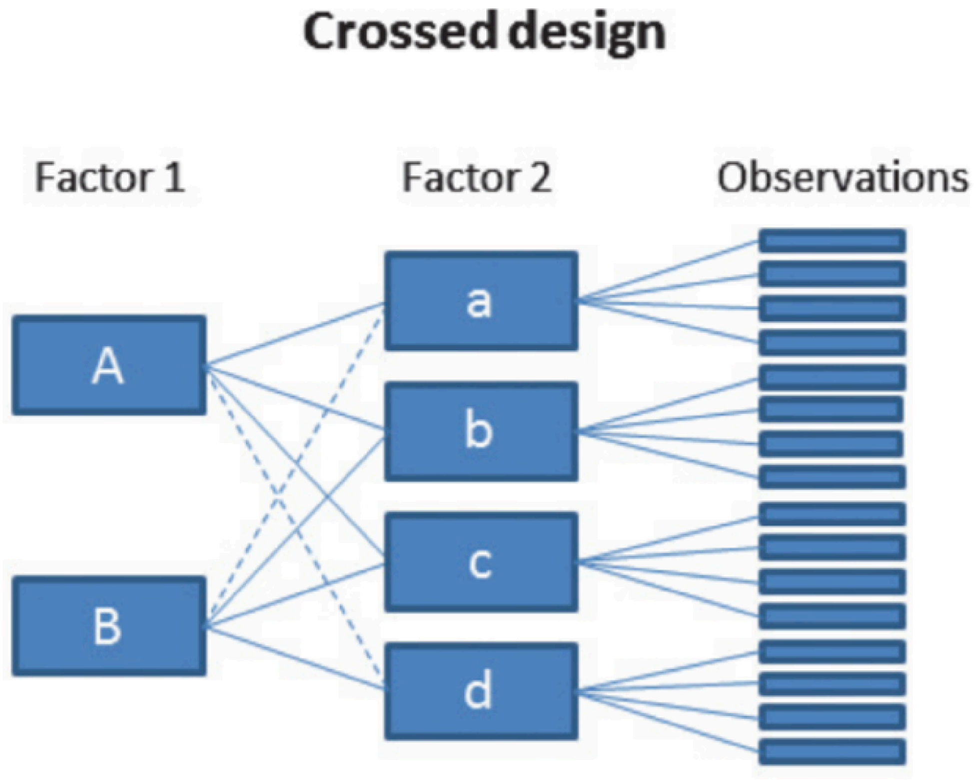
- Can have more factors (e.g., 3-way ANOVA)
  - interpretation tricky…

Most multifactor designs: nested or factorial

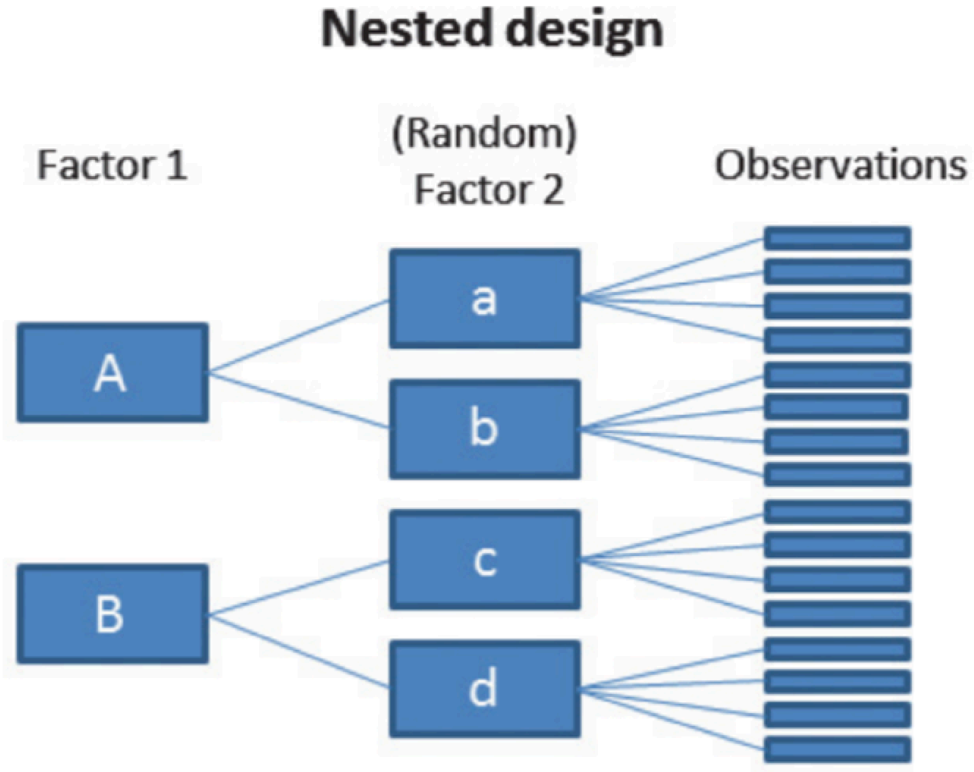
Factorial Versus Nested Designs

Consider two factors: A and B

- Factorial/crossed: every level of B in every level of A
- Nested/hierarchical: levels of B occur only in 1 level of A



Schielzth & Nakagawa 2013



Lecture 13: Nested ANOVA Overview

Nested design examples

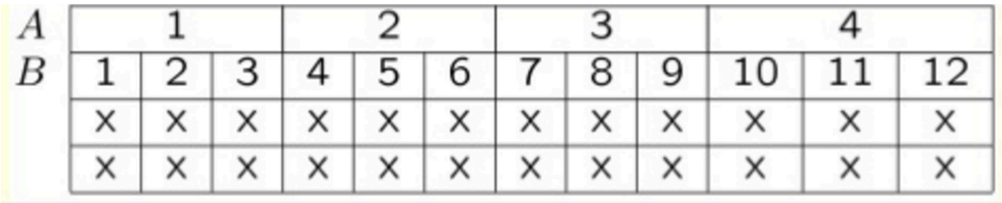
- Nested designs
- Linear model

- Analysis of variance
- Null hypotheses
- Unbalanced designs
- Assumptions

Nested Designs Overview

Nested Designs:

- Factor A usually fixed
- Factor B usually random

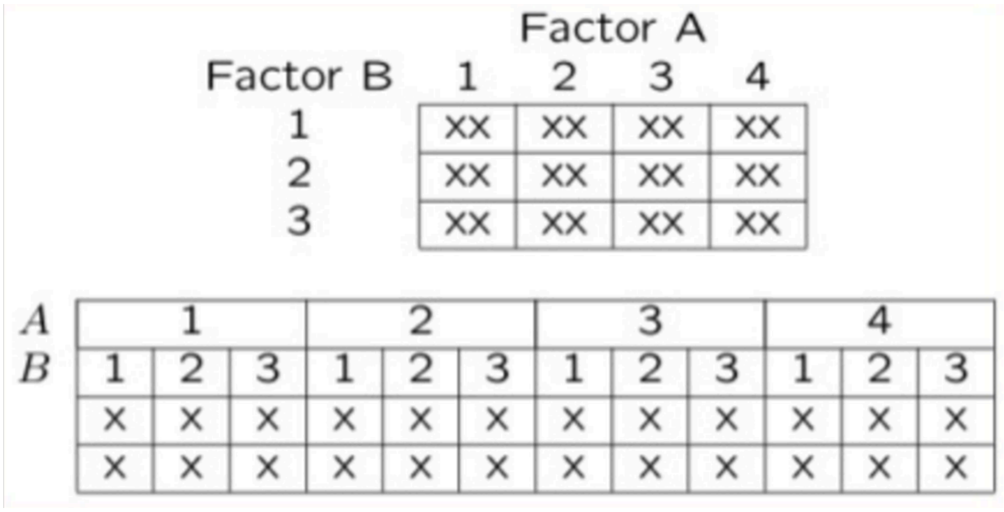


A	1			2			3			4		
B	1	2	3	4	5	6	7	8	9	10	11	12
	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×	×	×	×	×	×	×	×

Factorial Designs Overview

Factorial Designs:

- Both factors typically fixed (but not always)



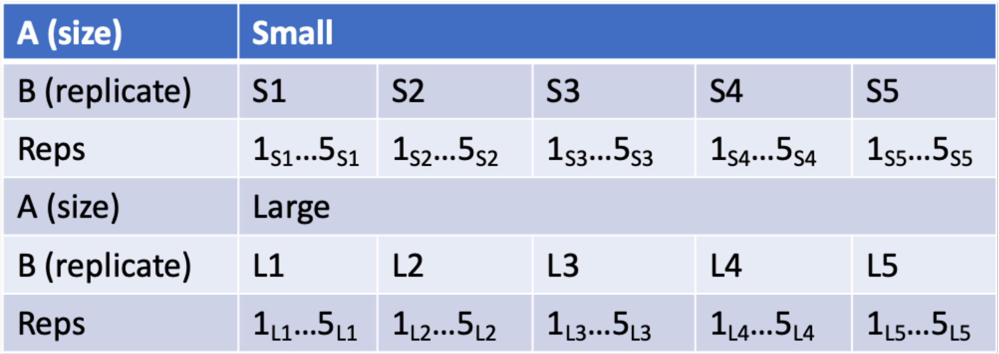
	Factor A			
Factor B	1	2	3	4
1	XX	XX	XX	XX
2	XX	XX	XX	XX
3	XX	XX	XX	XX

A	1			2			3			4		
B	1	2	3	1	2	3	1	2	3	1	2	3
	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×	×	×	×	×	×	×	×

Nested Design Example: Limpet Growth

Study on effects of enclosure size on limpet growth:

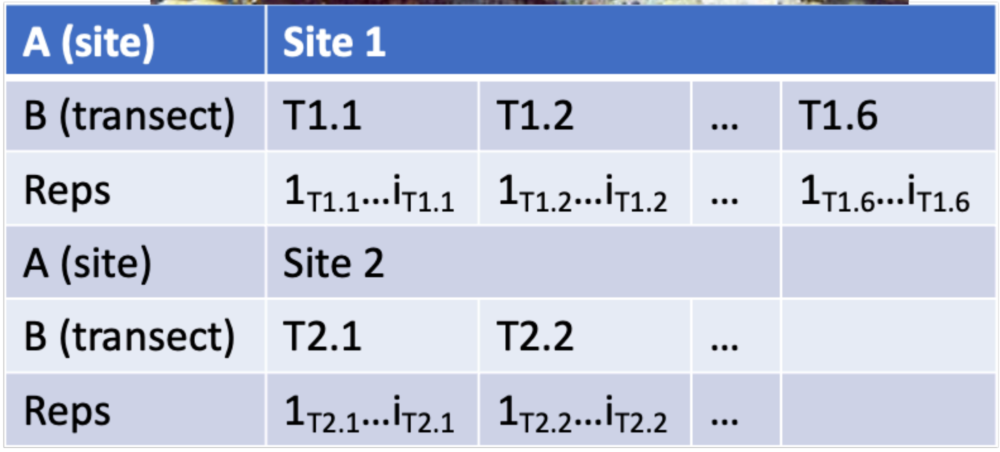
- 2 enclosure sizes (factor A)
- 5 replicate enclosures (factor B)
- 5 replicate limpets per enclosure



A (size)	Small				
B (replicate)	S1	S2	S3	S4	S5
Reps	$1_{S1}…5_{S1}$	$1_{S2}…5_{S2}$	$1_{S3}…5_{S3}$	$1_{S4}…5_{S4}$	$1_{S5}…5_{S5}$
A (size)	Large				
B (replicate)	L1	L2	L3	L4	L5
Reps	$1_{L1}…5_{L1}$	$1_{L2}…5_{L2}$	$1_{L3}…5_{L3}$	$1_{L4}…5_{L4}$	$1_{L5}…5_{L5}$

Nested Design Example: Reef Fish

Study on reef fish recruitment: 5 sites (factor A) 6 transects at each site (factor B) replicate observations along each transect

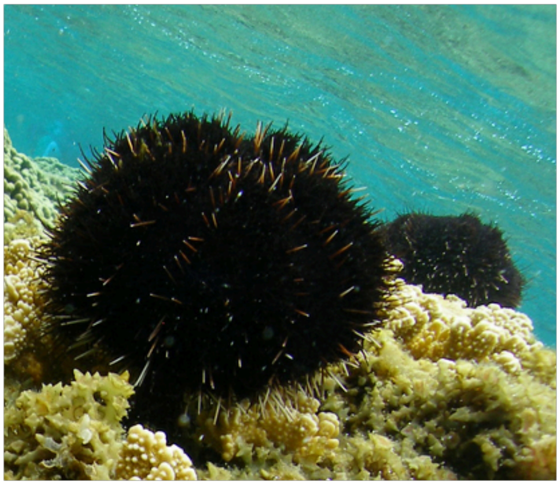


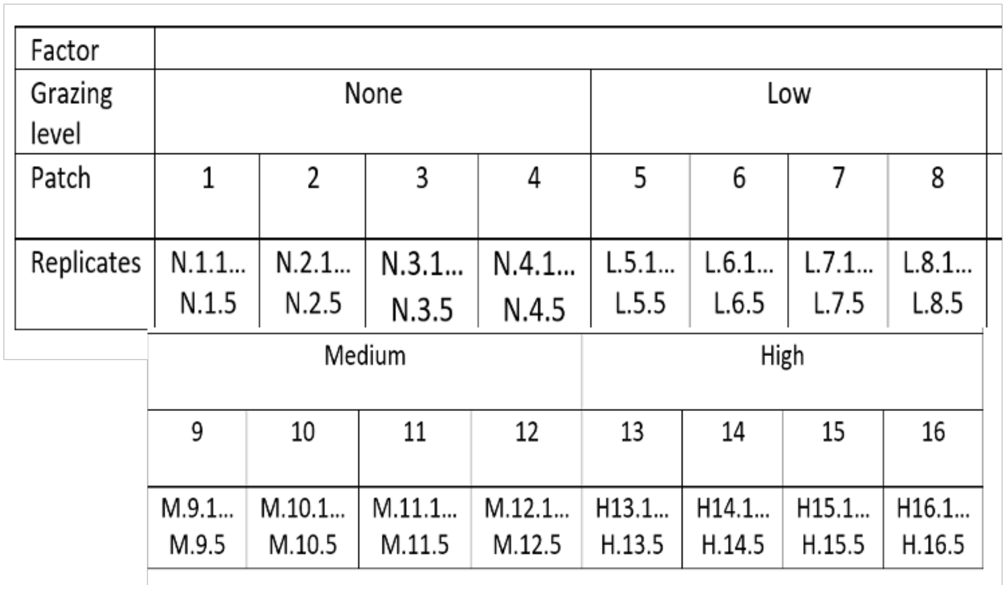
A (site)	Site 1			
B (transect)	T1.1	T1.2	…	T1.6
Reps	$1_{T1.1}…i_{T1.1}$	$1_{T1.2}…i_{T1.2}$	…	$1_{T1.6}…i_{T1.6}$
A (site)	Site 2			
B (transect)	T2.1	T2.2	…	
Reps	$1_{T2.1}…i_{T2.1}$	$1_{T2.2}…i_{T2.2}$	…	

Nested Design Example: Sea Urchin Grazing

Effects of sea urchin grazing on biomass of filamentous algae:

- 4 levels of urchin grazing: none, L, M, H
- 4 patches of rocky bottom (3-4 m2) nested in each level of grazing
- 5 replicate quadrats per patch



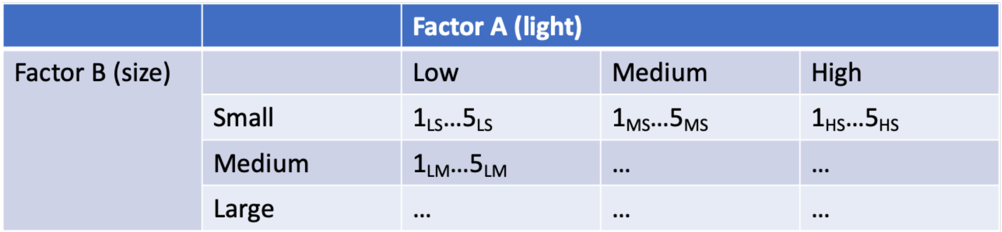


Factor																
Grazing level	None				Low				Medium				High			
Patch	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Replicates	N.1.1… N.1.5	N.2.1… N.2.5	N.3.1… N.3.5	N.4.1… N.4.5	L.5.1… L.5.5	L.6.1… L.6.5	L.7.1… L.7.5	L.8.1… L.8.5	M.9.1… M.9.5	M.10.1… M.10.5	M.11.1… M.11.5	M.12.1… M.12.5	H13.1… H.13.5	H14.1… H.14.5	H15.1… H.15.5	H16.1… H.16.5

Factorial Design Example: Seedling Growth

Effects of light level on growth of seedlings of different size:

- 3 light levels (factor A)
- 3 size classes (factor B)
- 5 replicate seeding in each cell

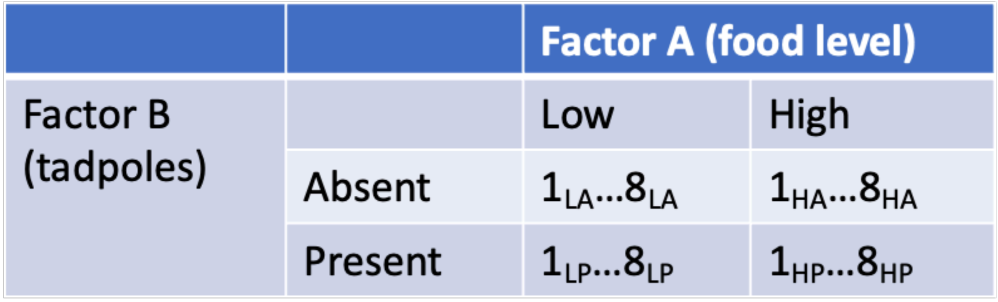


		Factor A (light)		
Factor B (size)		Low	Medium	High
	Small	$1_{LS}…5_{LS}$	$1_{MS}…5_{MS}$	$1_{HS}…5_{HS}$
	Medium	$1_{LM}…5_{LM}$	…	…
	Large	…	…	…

Factorial Design Example: Salamander Growth

Effects of food level and tadpole presence on larval salamander growth

- 2 food levels (factor A)
- presence/absence of tadpoles (factor B)
- 8 replicates in each cell



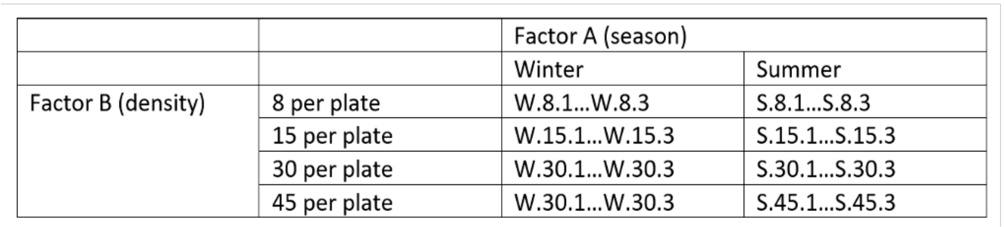
		Factor A (food level)	
Factor B (tadpoles)		Low	High
	Absent	$1_{LA}…8_{LA}$	$1_{HA}…8_{HA}$
	Present	$1_{LP}…8_{LP}$	$1_{HP}…8_{HP}$

Factorial Design Example: Limpet Fecundity

Effect of season and density on limpet fecundity.

- 2 seasons (factor A)
- 4 density treatments (factor B)

- 3 replicates in each cell

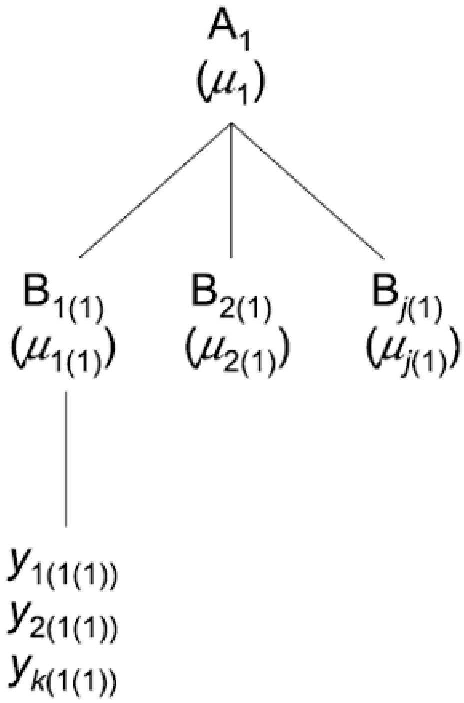


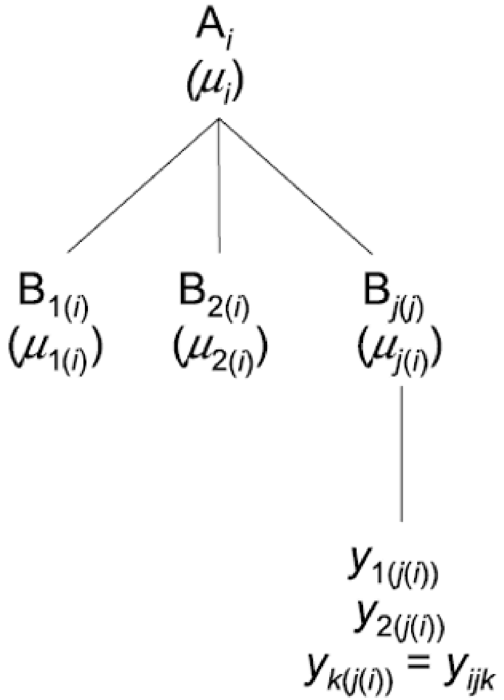
		Factor A (season)	
		Winter	Summer
Factor B (density)	8 per plate	W.8.1…W.8.3	S.8.1…S.8.3
	15 per plate	W.15.1…W.15.3	S.15.1…S.15.3
	30 per plate	W.30.1…W.30.3	S.30.1…S.30.3
	45 per plate	W.30.1…W.30.3	S.45.1…S.45.3

Nested Design: Linear Model Structure

Consider a nested design with:

- p levels of factor A (i= 1…p) (e.g., 4 grazing levels)
- q levels of factor B (j= 1…q), nested within each level of A (e.g., 4 - diff. patches per grazing level)
- n replicates (k= 1…n) in each combination of A and B (5 replicate - quadrats in each patch in each grazing level)

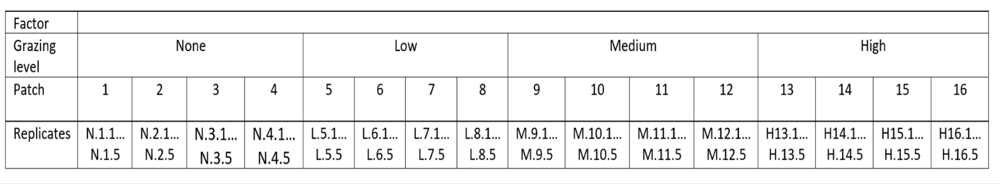




Calculating Means in Nested Design

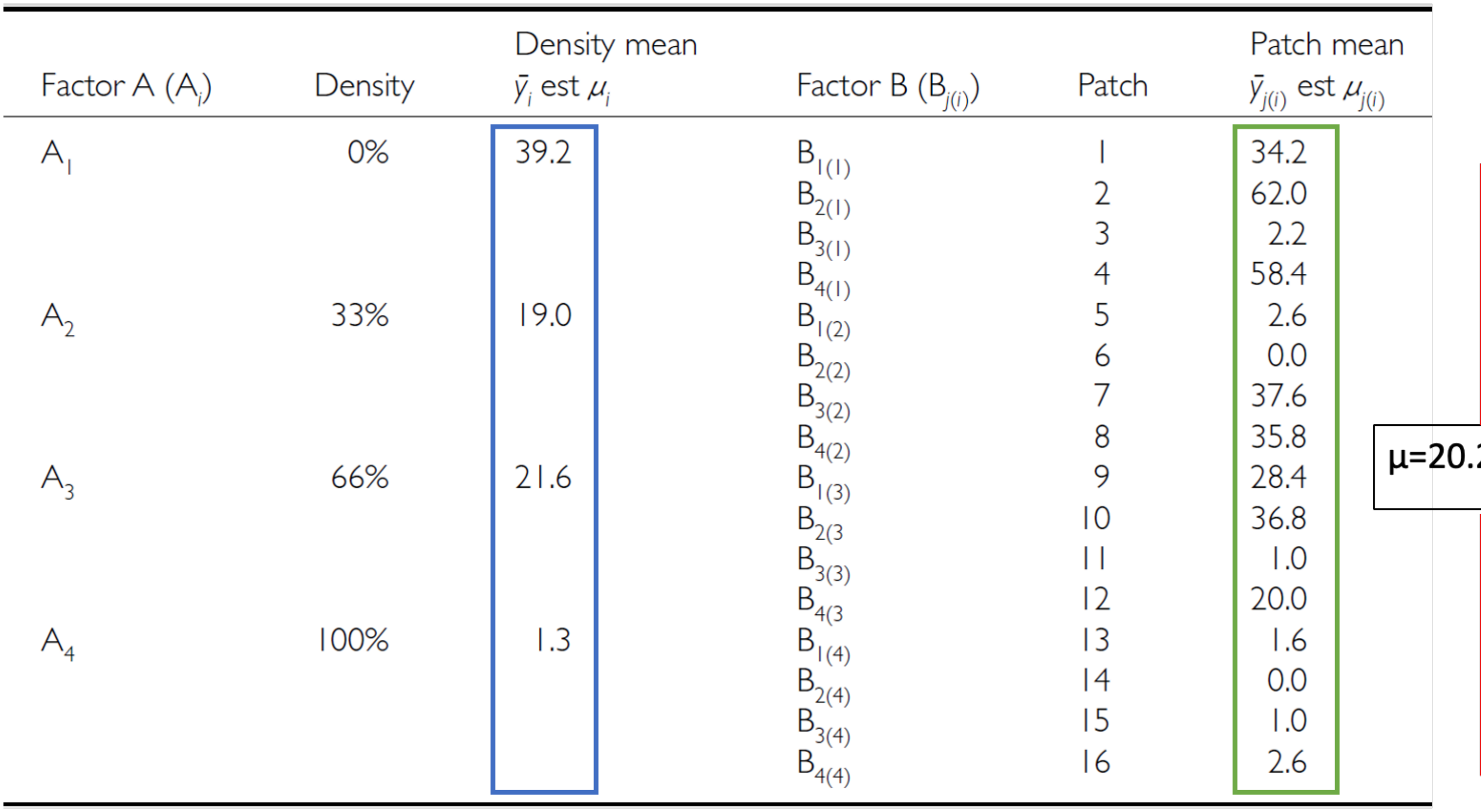
Can calculate several means:

- overall mean (across all levels of A and B)= $\bar{y}$;
- a mean for each level of A (across all levels of B in that A)= $\bar{y}i$;
- a mean for each level of B within each A= $\bar{y}j(i)$



Factor																
Grazing level	None				Low				Medium				High			
Patch	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Replicates	N.1.1… N.1.5	N.2.1… N.2.5	N.3.1… N.3.5	N.4.1… N.4.5	L.5.1… L.5.5	L.6.1… L.6.5	L.7.1… L.7.5	L.8.1… L.8.5	M.9.1… M.9.5	M.10.1… M.10.5	M.11.1… M.11.5	M.12.1… M.12.5	H13.1… H.13.5	H14.1… H.14.5	H15.1… H.15.5	H16.1… H.16.5

Nested Design Means Visualization



Factor A (A_i)	Density	Density mean $\bar{y}_i$ est μ_i	Factor B ($B_{j(i)}$)	Patch	Patch mean $\bar{y}_{j(i)}$ est $\mu_{j(i)}$
A_1	0%	39.2	$B_{1(1)}$	1	34.2
			$B_{2(1)}$	2	62.0
			$B_{3(1)}$	3	2.2
			$B_{4(1)}$	4	58.4
A_2	33%	19.0	$B_{1(2)}$	5	2.6
			$B_{2(2)}$	6	0.0
			$B_{3(2)}$	7	37.6
			$B_{4(2)}$	8	35.8
A_3	66%	21.6	$B_{1(3)}$	9	28.4
			$B_{2(3}$	10	36.8
			$B_{3(3)}$	11	1.0
			$B_{4(3}$	12	20.0
A_4	100%	1.3	$B_{1(4)}$	13	1.6
			$B_{2(4)}$	14	0.0
			$B_{3(4)}$	15	1.0
			$B_{4(4)}$	16	2.6

μ=20.

Nested Design Linear Model

The linear model for a nested design is:

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}$$

Where:

- y_{ijk} is the response variable
  - value of the k-th replicate in j-th level of B in the i-th level of A
  - (algal biomass in 3rd quadrat, in 2nd patch in low grazing treatment)
- μ is the overall mean
  - (overall average algal biomass)

Fixed Effects in Nested Model

The linear model for a nested design is:

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}$$

- α_i is the fixed effect of factor i
- (difference between average biomass in all low grazing level quadrats and overall mean)
- $\beta_{j(i)}$ is the random effect of factor j nested within factor i
- usually random variable, measuring variance among all possible levels of B within each level of A

- (variance among all possible patches that may have been used in the low grazing treatment)

Error Term in Nested Model

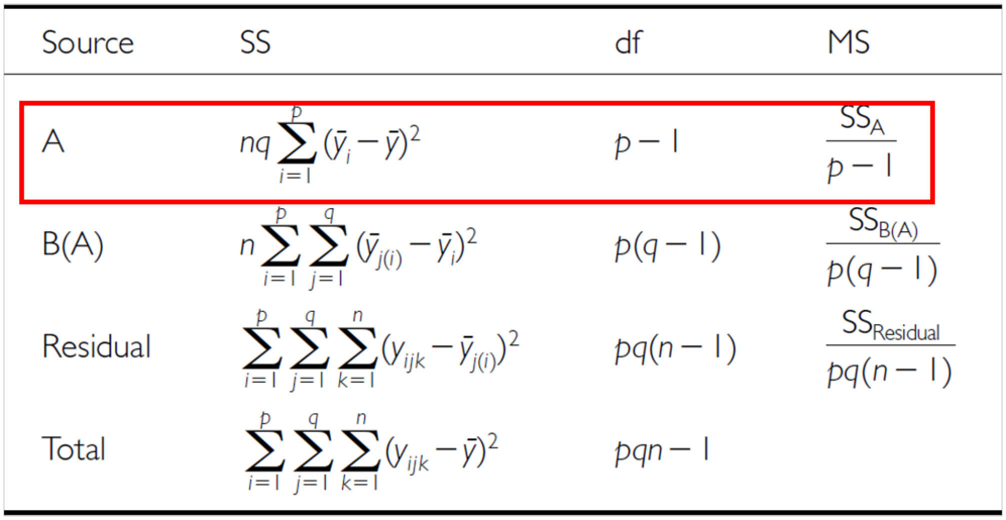
The linear model for a nested design is:

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}$$

- ε_{ijk} is the error term
- αi: is the effect of the ith level of A: µi- µ
- unexplained variance associated with the kth replicate in jth level of B in the ith level of A
- (difference bw observed algal biomass in 3rd quadrat in 2nd patch in low grazing treatment and predicted biomass - average biomass in 2nd patch in low grazing treatment)

Analysis of Variance: SSA

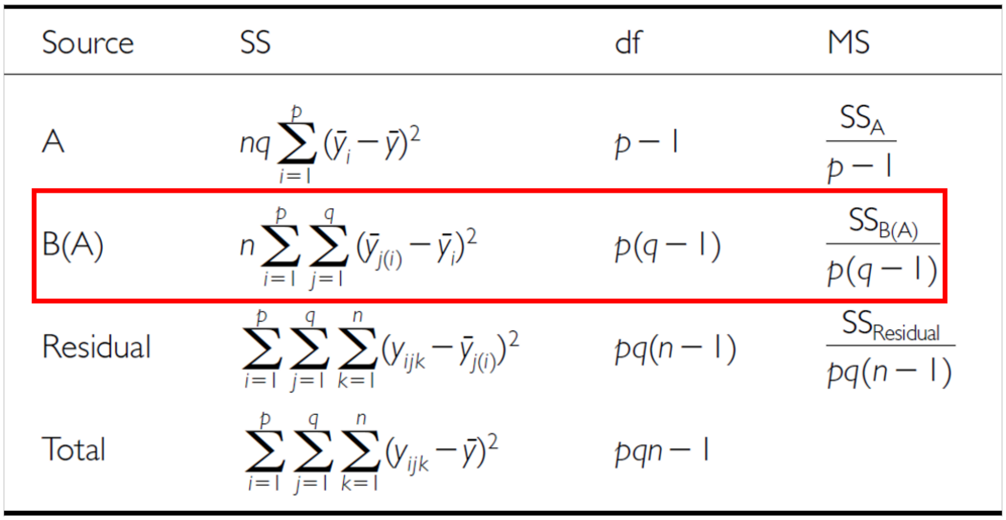
As before, partition the variance in the response variable using SS SSA is SS of differences between means in each level of A and overall mean



Source	SS	df	MS
A	$nq\sum_{i=1}^{p}(\bar{y}_i-\bar{y})^2$	$p-1$	$\frac{SS_A}{p-1}$
B(A)	$n\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{j(i)}-\bar{y}_i)^2$	$p(q-1)$	$\frac{SS_{B(A)}}{p(q-1)}$
Residual	$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{j(i)})^2$	$pq(n-1)$	$\frac{SS_{Residual}}{pq(n-1)}$
Total	$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(y_{ijk}-\bar{y})^2$	$pqn-1$	

Analysis of Variance: SSB

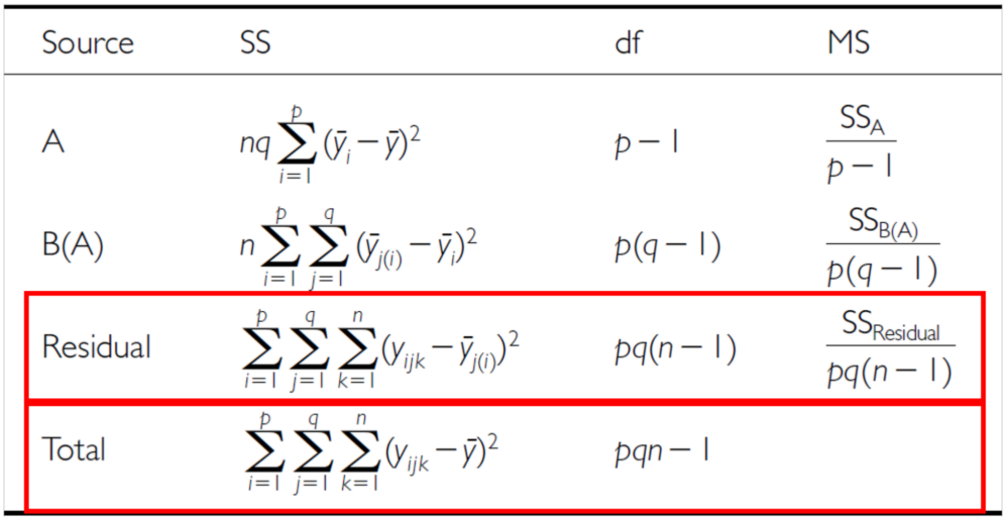
SSB is SS of difference between means in each level of B and the mean of corresponding level of A summed across levels of A



Source	SS	df	MS
A	$nq\sum_{i=1}^{p}(\bar{y}_i-\bar{y})^2$	$p-1$	$\frac{SS_A}{p-1}$
B(A)	$n\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{j(i)}-\bar{y}_i)^2$	$p(q-1)$	$\frac{SS_{B(A)}}{p(q-1)}$
Residual	$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{j(i)})^2$	$pq(n-1)$	$\frac{SS_{Residual}}{pq(n-1)}$
Total	$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(y_{ijk}-\bar{y})^2$	$pqn-1$	

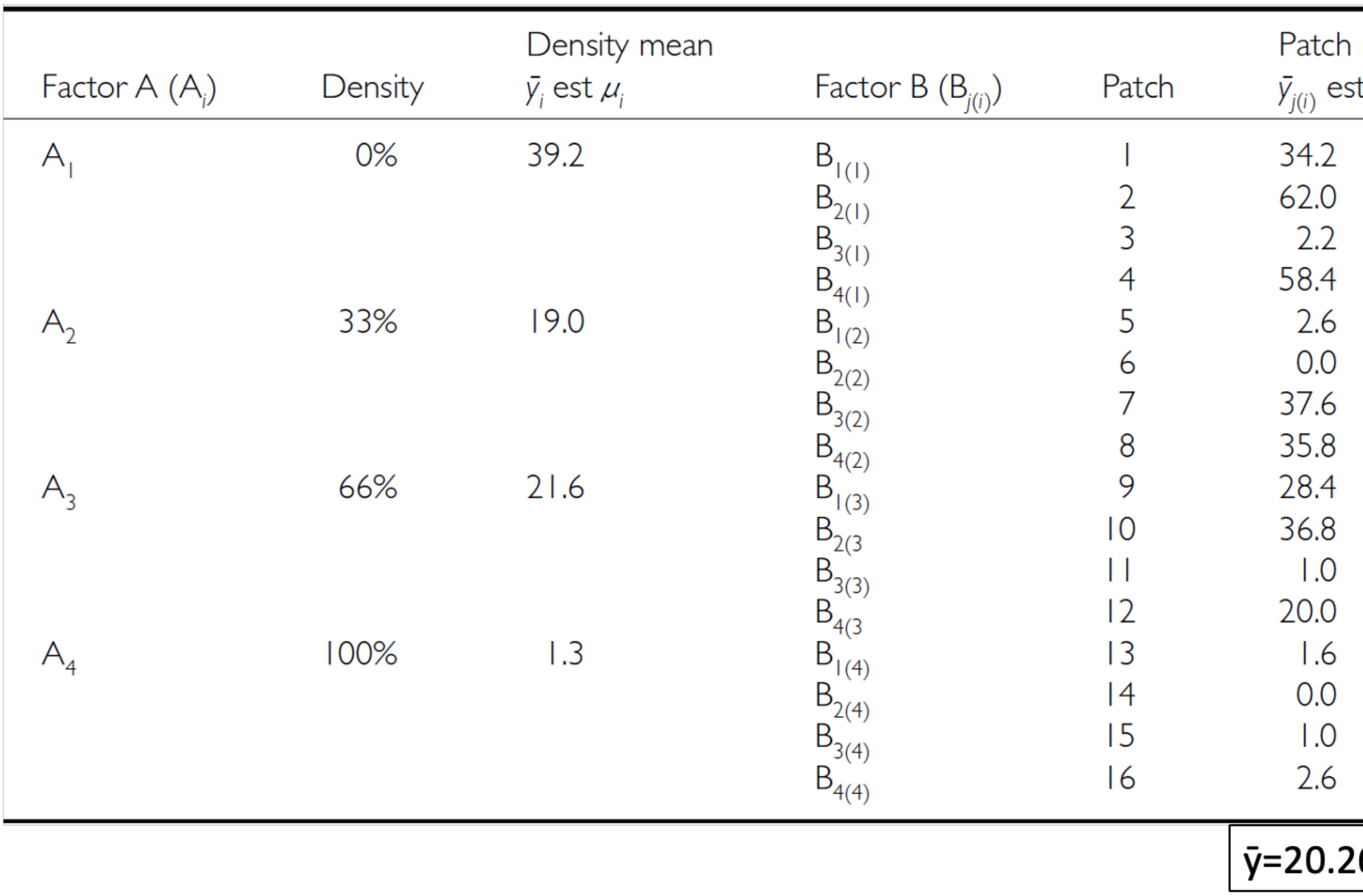
Analysis of Variance: Residual and Total

- SSresid is difference bw each observation and mean for its level of factor B, summed over all observations
- SStotal = SSA + SSB + SSresid
- SS can be turned into MS by dividing by appropriate df



Source	SS	df	MS
A	$nq\sum_{i=1}^{p}(\bar{y}_i-\bar{y})^2$	$p-1$	$\frac{SS_A}{p-1}$
B(A)	$n\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{j(i)}-\bar{y}_i)^2$	$p(q-1)$	$\frac{SS_{B(A)}}{p(q-1)}$
Residual	$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{j(i)})^2$	$pq(n-1)$	$\frac{SS_{Residual}}{pq(n-1)}$
Total	$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(y_{ijk}-\bar{y})^2$	$pqn-1$	

Analysis of Variance Table



Factor A (A_i)	Density	Density mean $\bar{y}_i$ est μ_i	Factor B ($B_{j(i)}$)	Patch	Patch $\bar{y}_{j(i)}$ est
A_1	0%	39.2	$B_{1(1)}$	1	34.2
			$B_{2(1)}$	2	62.0
			$B_{3(1)}$	3	2.2
			$B_{4(1)}$	4	58.4
A_2	33%	19.0	$B_{1(2)}$	5	2.6
			$B_{2(2)}$	6	0.0
			$B_{3(2)}$	7	37.6
			$B_{4(2)}$	8	35.8
A_3	66%	21.6	$B_{1(3)}$	9	28.4
			$B_{2(3}$	10	36.8
			$B_{3(3)}$	11	1.0
			$B_{4(3}$	12	20.0
A_4	100%	1.3	$B_{1(4)}$	13	1.6
			$B_{2(4)}$	14	0.0
			$B_{3(4)}$	15	1.0
			$B_{4(4)}$	16	2.6

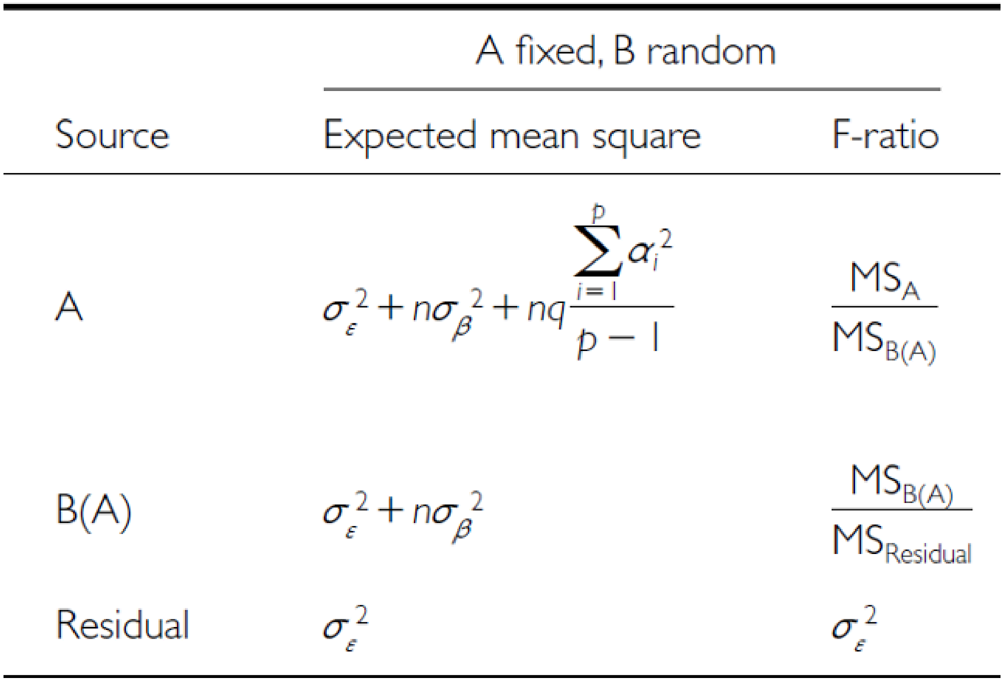
$\bar{y}$=20.2

Null Hypotheses: Factor A

Two hypotheses tested on values of MS:

1. no effects of factor A

- Assuming A is fixed:
- Ho(A): μ1= μ2= μ3=.... μi= μ
- Same as in 1-factor ANOVA, using means from B factors nested within each - level of A
- (no difference in algal biomass across all levels of grazing: μnone= - μlow= μmed= μhigh)



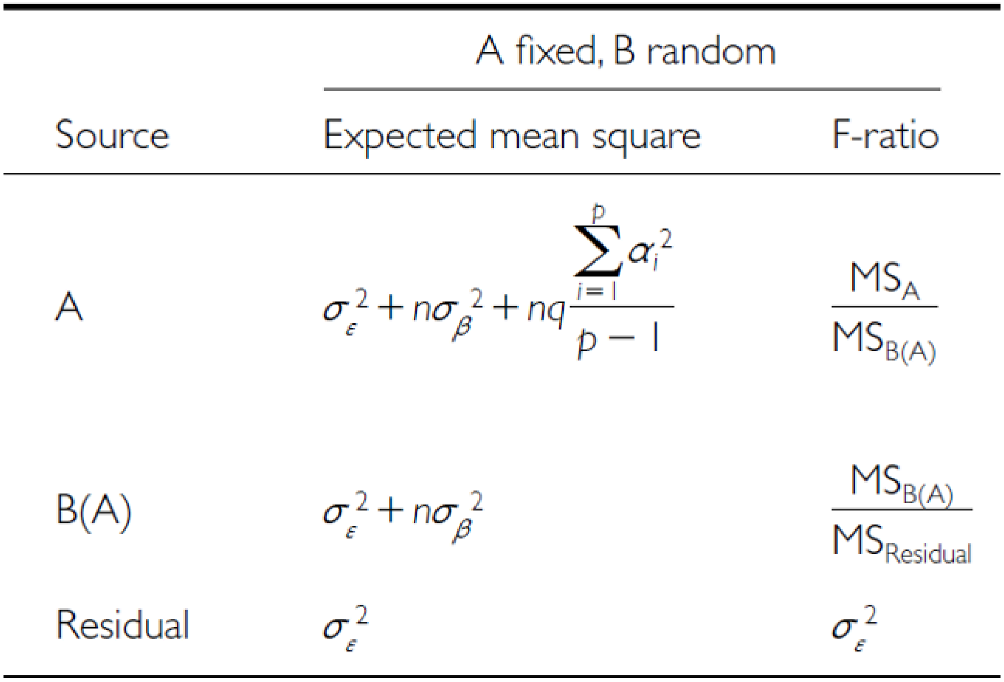
	A fixed, B random	
Source	Expected mean square	F-ratio
A	$\sigma_\varepsilon^2 + n\sigma_\beta^2 + nq\frac{\sum_{i=1}^{p}\alpha_i^2}{p-1}$	$\frac{MS_A}{MS_{B(A)}}$
B(A)	$\sigma_\varepsilon^2 + n\sigma_\beta^2$	$\frac{MS_{B(A)}}{MS_{Residual}}$
Residual	σ_ε^2	σ_ε^2

Null Hypotheses: Factor B

Two hypotheses tested on values of MS:

2. No effects of factor B nested in A

- Assuming B is random:
- Ho(B): σβ2= 0 (no variance added due to differences between all possible - levels of B)
- (no variance added due to differences between patches)

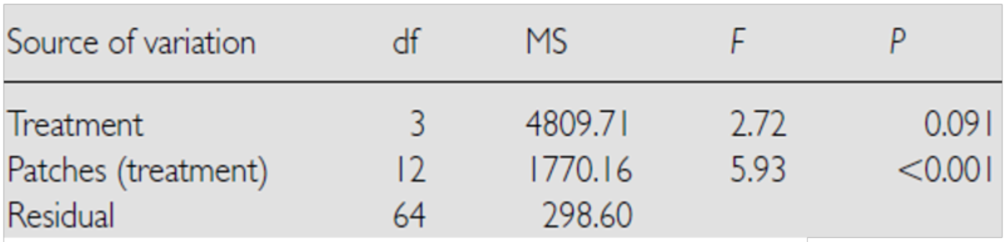


	A fixed, B random	
Source	Expected mean square	F-ratio
A	$\sigma_\varepsilon^2 + n\sigma_\beta^2 + nq\frac{\sum_{i=1}^{p}\alpha_i^2}{p-1}$	$\frac{MS_A}{MS_{B(A)}}$
B(A)	$\sigma_\varepsilon^2 + n\sigma_\beta^2$	$\frac{MS_{B(A)}}{MS_{Residual}}$
Residual	σ_ε^2	σ_ε^2

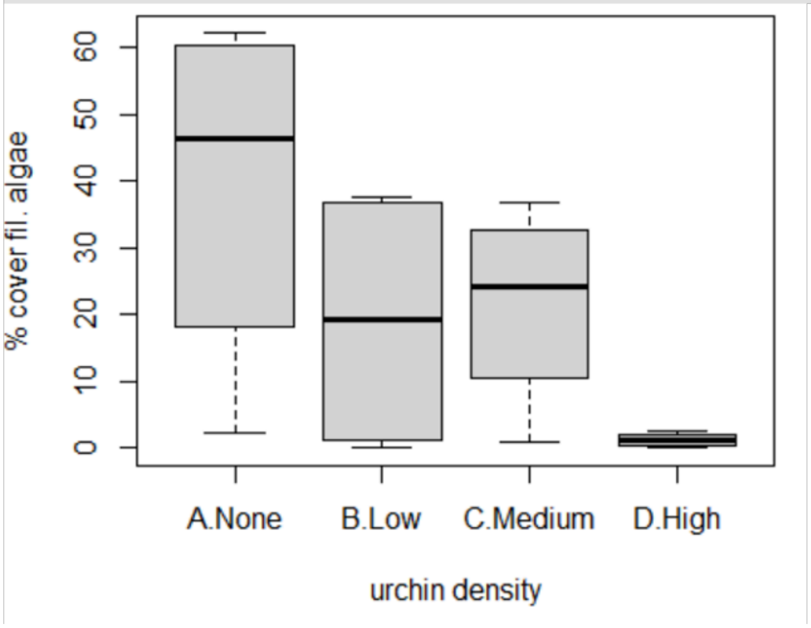
Conclusions from Analysis

Conclusions?

“significant variation between replicate patches within each treatment, but no significant difference in amount of filamentous algae between treatments”



Source of variation	df	MS	F	P
Treatment	3	4809.71	2.72	0.091
Patches (treatment)	12	1770.16	5.93	<0.001
Residual	64	298.60		

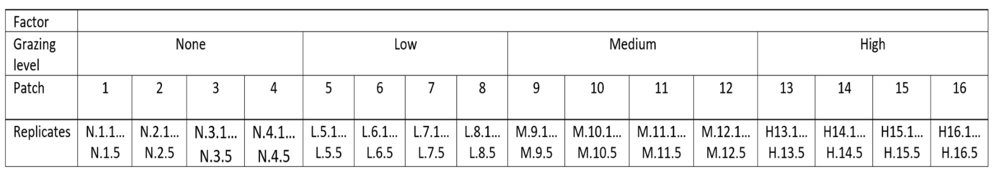


Unbalanced Nested Designs

Unequal sample sizes can be because of:

- uneven number of B levels within each A
- uneven number of replicates within each level of B

Not a problem, unless have unequal variance or large deviation from - normality



Factor																
Grazing level	None				Low				Medium				High			
Patch	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Replicates	N.1.1... N.1.5	N.2.1... N.2.5	N.3.1... N.3.5	N.4.1... N.4.5	L.5.1... L.5.5	L.6.1... L.6.5	L.7.1... L.7.5	L.8.1... L.8.5	M.9.1... M.9.5	M.10.1... M.10.5	M.11.1... M.11.5	M.12.1... M.12.5	H13.1... H.13.5	H14.1... H.14.5	H15.1... H.15.5	H16.1... H.16.5

Nested Design Assumptions

As usual, we assume

- equal variance
- normality
- independence of observations

Equal variance + normality need to be assessed at both levels:

- Since means for each level of B within each A are used for the H-test about A, need to assess whether those means meet normality and equal variance
- Examine residuals for H-test about B
- Transformations can be used