

A Small Exercise in Normal Forms

Earthworm Invasions

Study to identify the cause and extent of exotic earthworm infestations

- Many sample sites, 10 meters in diameter, over a range of conditions. The worms present were counted and identified by species. GPS location was taken along with the type of soil, amount/depth of forest litter and vegetation at each sample site is also noted.
- 5 different Earthworm species are found in varying amounts at each site
- These data were organized in an attribute data file associated with each sampling point

Maple without Worms



Maple with Worms



Exotic Earthworm Study Non-normalized Form

Site Table

Site	UTM - East	UTM - North	Veg-code	Vegetation	Soil	DESCRIPTION	Litter	Worm Species 1	Number 1	Worm Species 2	Number 2	Worm Species 3	Number 3	Worm Species 4	Number 4	Worm Species 5	Number 5
1	56565	538980	b	Blk Spruce	34	Hyck	3	A	12	B	13	C	44				
2	56580	538952	m	MxdHrdwd	34	Hyck	3	B	14	C	33						
3	56590	538950	m	MxdHrdwd	13	Johns	4	C	16	B	22						
4	56540	538990	l	LowHrdwd	86	Chero	2	A	40	B	11						
5	56567	538700	o	Oak Sav	4	Sonb	1	G	23	A	12	B	32	C	12	E	24
6	56577	538950	o	Oak Sav	34	Hyck	1	E	23	C	43	B	31				
7	56597	538990	r	Red Pine	13	Johns	2	A	22	C	12	D	23	E	12		
8	56527	538950	j	Jak Pine	86	Chero	2	A	23	D	22	E	32				
9	56537	539100	a	Alder	13	Johns	3	A	11	B	44	C	15	D	24		

Repeat columns

Normal Forms Summary

No repeat columns (create new records such that there are multiple records per entry)

Split the tables, so that all non-key attributes depend on a primary key.

Split tables further, if there are transitive functional dependencies. This results in tables with a single, primary key per table.

Un-normalized File

Site	UTM - East	UTM North	Veg-code	Vegetation	Soil	DESCRIPTION	Litter	Worm Species 1	Number 1	Worm Species 2	Number 2	Worm Species 3	Number 3	Worm Species 4	Number 4	Worm Species 5	Number 5
1	56565	538980	b	Blk Spruce	34	Hyck	3	A	12	B	13	C	44				
2	56580	538952	m	MxdHrdwd	34	Hyck	3	B	14	C	33						
3	56590	538950	m	MxdHrdwd	13	Johns	4	C	16	B	22						
4	56540	538990	l	LowHrdwd	86	Chero	2	A	40	B	11						
5	56567	538700	o	Oak Sav	4	Sonb	1	G	23	A	12	B	32	C	12	E	24
6	56577	538950	o	Oak Sav	34	Hyck	1	E	23	C	43	B	31				
7	56597	538990	r	Red Pine	13	Johns	2	A	22	C	12	D	23	E	12		
8	56527	538950	j	Jack Pine	86	Chero	2	A	23	D	22	E	32				
9	56537	539100	a	Alder	13	Johns	3	A	11	B	44	C	15	D	24		

You remove duplicate columns to create a 1st normal form

Site	East	North	?	?	?	?	?	?	?
1	56565	538980							
1	56565	538980							
1	56565	538980							
2	56580	538952							
2	56580	538952							
3	56590	538950							
3	56590	538950							
4	56540	538990							
4	56540	538990							
5	56567	538700							
5	56567	538700							
5	56567	538700							
5	56567	538700							
5	56567	538700							
6	56577	538950							
6	56577	538950							
6	56577	538950							
6	56577	538950							
7	56597	538990							
7	56597	538990							
7	56597	538990							
7	56597	538990							
7	56597	538990							
8	56527	538950							
8	56527	538950							
8	56527	538950							
8	56527	538950							
9	56537	539100							
9	56537	539100							
9	56537	539100							
9	56537	539100							

1st normal form
– no repeat columns

Site	East	North	Veg-code	Vegetation	Soil	DESCRIPTION	Litter	Worm Species	Number
1	56565	538980	b	Blk Spruce	34	Hyck	3	A	12
1	56565	538980	b	Blk Spruce	34	Hyck	3	B	13
1	56565	538980	b	Blk Spruce	34	Hyck	3	C	12
2	56580	538952	m	MxdHrdwd	34	Hyck	3	B	14
2	56580	538952	m	MxdHrdwd	34	Hyck	3	C	33
3	56590	538950	m	MxdHrdwd	13	Johns	4	C	16
3	56590	538950	m	MxdHrdwd	13	Johns	4	B	22
4	56540	538990	l	LowHrdwd	86	Chero	2	A	40
4	56540	538990	l	LowHrdwd	86	Chero	2	B	11
5	56567	538700	o	Oak Sav	4	Sonb	1	G	23
5	56567	538700	o	Oak Sav	4	Sonb	1	A	12
5	56567	538700	o	Oak Sav	4	Sonb	1	B	32
5	56567	538700	o	Oak Sav	4	Sonb	1	C	12
5	56567	538700	o	Oak Sav	4	Sonb	1	E	24
6	56577	538950	o	Oak Sav	34	Hyck	1	E	23
6	56577	538950	o	Oak Sav	34	Hyck	1	C	43
6	56577	538950	o	Oak Sav	34	Hyck	1	B	31
7	56597	538990	r	Red Pine	13	Johns	2	A	22
7	56597	538990	r	Red Pine	13	Johns	2	C	12
7	56597	538990	r	Red Pine	13	Johns	2	D	23
7	56597	538990	r	Red Pine	13	Johns	2	E	12
8	56527	538950	j	Jack Pine	86	Chero	2	A	23
8	56527	538950	j	Jack Pine	86	Chero	2	D	22
8	56527	538950	j	Jack Pine	86	Chero	2	E	32
9	56537	539100	a	Alder	13	Johns	3	A	11
9	56537	539100	a	Alder	13	Johns	3	B	44
9	56537	539100	a	Alder	13	Johns	3	C	15
9	56537	539100	a	Alder	13	Johns	3	D	24

2nd normal form
Functional Dependencies

Break up the 1st normal table. Retain the Site/Species/Count table and create a second table with eight items, using Site as the primary key and removing the dependent items from the 1st NF table

Site Table

Site	UTM - East	UTM North	?	?	?	?	?
1	56565	538980					
2	56580	538952					
3	56590	538950					
4	56540	538990					
5	56567	538700					
6	56577	538950					
7	56597	538990					
8	56527	538950					
9	56537	539100					

Site/Species/Count Table

Site	?	?
1		
1		
1		
2		
2		
3		
3		
4		
4		
5		
5		
5		
5		
5		
5		
6		
6		
6		
7		
7		
7		
7		
8		
8		
8		
9		
9		
9		
9		

2nd normal form Functional Dependencies

Break up the 1st normal table. Retain the Site/Species/Count table and create a second table with eight items, using Site as the primary key and removing the dependent items from the 1st NF table

Site Table

Site	UTM - East	UTM North	Veg-code	Vegetation	Soil	DESCRIPTION	Litter
1	56565	538980	b	Blk Spruce	34	Hyck	3
2	56580	538952	m	MxdHrdwd	34	Hyck	3
3	56590	538950	m	MxdHrdwd	13	Johns	4
4	56540	538990	l	LowHrdwd	86	Chero	2
5	56567	538700	o	Oak Sav	4	Sonb	1
6	56577	538950	o	Oak Sav	34	Hyck	1
7	56597	538990	r	Red Pine	13	Johns	2
8	56527	538950	j	Jack Pine	86	Chero	2
9	56537	539100	a	Alder	13	Johns	3

Site/Species/Count
Table

Site	Worm Specie	Number
1	a	12
1	b	23
1	c	12
2	b	242
2	c	232
2	d	33
3	b	12
3	d	121
3	e	123
4	a	23
4	b	123
4	c	55
5	d	55
5	g	55
6	e	23
6	c	43
6	b	31
7	c	22
7	a	12
7	c	22
8	a	23
8	d	22
8	e	32
9	a	11
9	b	44
9	c	15
9	d	24

2nd normal form Functional Dependencies,

Site -> East, North, Vegcode, Vegetation, soil, description, litter

So I make a second table with these eight items, with Site as the primary key and remove the dependent items from the 1st NF table

Site table

Site	UTM - East	UTM North	Veg-code	Vegetation	Soil	DESCRIPTION	Litter
1	56565	538980	b	Blk Spruce	34	Hyck	3
2	56580	538952	m	MxdHrdwd	34	Hyck	3
3	56590	538950	m	MxdHrdwd	13	Johns	4
4	56540	538990	l	LowHrdwd	86	Chero	2
5	56567	538700	o	Oak Sav	4	Sonb	1
6	56577	538950	o	Oak Sav	34	Hyck	1
7	56597	538990	r	Red Pine	13	Johns	2
8	56527	538950	j	Jack Pine	86	Chero	2
9	56537	539100	a	Alder	13	Johns	3

3rd normal form - Transitive Functional Dependencies

Veg-code -> Vegetation
Soil -> Description

Split the 2nd normal table these into several tables, and removing the transitively dependent columns from the site table.

Veg Table

VEG_CODE	?

Soil Table

Soil	?

Site/CodeTable

Site	UTM - East	UTM North	?	?	?
1	56565	538980			
2	56580	538952			
3	56590	538950			
4	56540	538990			
5	56567	538700			
6	56577	538950			
7	56597	538990			
8	56527	538950			
9	56537	539100			

Site/Species/Count Table

Site	?	?
1		
1		
1		
2		
2		
3		
3		
4		
4		
5		
5		
5		
5		
5		
6		
6		
6		
7		
7		
7		
7		
7		
8		
8		
8		
9		
9		
9		
9		

3rd normal form - Transitive Functional Dependencies

Veg-code -> Vegetation

Soil -> Description
So split these into their own tables, and remove the transitively dependent columns from the site table, above

Vegetation Table

VEG_CODE	Vegetation
b	Blk Spruce
m	MxdHrdwd
l	LowHrdwd
o	Oak Sav
r	Red Pine
j	Jack Pine
a	Alder

Site/Code Table

Site	UTM - East	UTM North	Veg-code	Soil	Litter
1	56565	538980	b	34	3
2	56580	538952	m	34	3
3	56590	538950	m	13	4
4	56540	538990	l	86	2
5	56567	538700	o	4	1
6	56577	538950	o	34	1
7	56597	538990	r	13	2
8	56527	538950	j	86	2
9	56537	539100	a	13	3

Soil Table

Soil	DESCRIPTION
34	Hyck
86	Chero
4	Sonb
13	Johns

Exotic Earthworm Study 3rd Normal Form

Site/Code Table

Site	UTM - East	UTM North	Veg-code	Soil	Litter
1	56565	538980	b	34	3
2	56580	538952	m	34	3
3	56590	538950	m	13	4
4	56540	538990	l	86	2
5	56567	538700	o	4	1
6	56577	538950	o	34	1
7	56597	538990	r	13	2
8	56527	538950	j	86	2
9	56537	539100	a	13	3

Vegetation Table

VEG_CODE	Vegetation
b	Blk Spruce
m	MxdHrdwd
l	LowHrdwd
o	Oak Sav
r	Red Pine
j	Jak Pine
a	Alder

Soil Table

Soil	DESCRIPTION
34	Hyck
86	Chero
4	Sonb
13	Johns

Site/
Species/
Count
Table

Site	Worm Specie	Num ber
1	A	12
1	B	13
1	C	12
2	B	14
2	C	33
3	C	16
3	B	22
4	A	40
4	B	11
5	G	23
5	A	12
5	B	32
5	C	12
5	G	23
5	E	24
6	E	23
6	C	43
6	B	31
7	A	22
7	C	12
7	D	23
7	E	12
8	A	23
8	D	22
8	E	32
9	A	11
9	B	44
9	C	15
9	D	24