(Crossed and nested subscripts). Let A and B be two given factors indexed by i and j, respectively. If A and B are crossed, then i and j are said to be *crossed subscripts*. This fact is denoted by writing $(i)(j)$ where i and j are separated by two pairs of parentheses. If, however, B is nested within A, then j is said to be *nested* within i, and this fact is denoted by writing $i : j$ where a colon separates i and j with j appearing to the right of the colon.

(Population structure). The *population structure* associated with a given experiment is a complete description of the nesting and nonnesting (crossed) relationships that exist among the factors considered in the experiment.

The population structure plays an important role in setting up the complete model and the corresponding ANOVA table for the experiment under investigation ….

An Example

A manufacturer wants to investigate the variation of the quality of a product with regard to type A pre-production processes and type B pre-production processes. Factor A has 4 levels and each level has 5 sub levels. Factor B has 4 levels and each level has 6 sub levels. Each sub level of each level of the A-factor is combined with each sub level of each level of the B-factor. The same no. of replications (3) is available for each sub level combination.

1. Give the population structure and write the corresponding model.
2. Give expressions for the expected mean squares assuming that the effects of A and B are fixed, while the remaining effects are random (the sub levels are chosen at random).

[(i:j)(k:l)]:m	A i = 1,…,4 C(A) j = 1, …, 5 B k = 1, …, 4 D(B) l = 1, …, 6 Rep m = 1, …, 3 gives (4)(5)(4)(6)(3) = 1440 obs		
--		DF	E(MS)
i	(i) A Fixed	4-1 = 3	$(5)(4)(6)(3)\,\Theta_A + (4)(6)(3)\sigma^2_{C(A)} + (6)(3)\sigma^2_{BC(A)} + (5)(3)\sigma^2_{AD(B)} + 3\sigma^2_{C(A)D(B)} + \sigma^2_\varepsilon$
XX j	(k) B Fixed	4-1 = 3	$(4)(5)(6)(3)\,\Theta_B + (4)(5)(3)\sigma^2_{D(B)} + (6)(3)\sigma^2_{BC(A)} + (5)(3)\sigma^2_{AD(B)} + 3\sigma^2_{C(A)D(B)} + \sigma^2_\varepsilon$
k	(ik) AB Fixed	(4-1)(4-1) = 9	$(5)(6)(3)\,\Theta_{AB} + (6)(3)\sigma^2_{BC(A)} + (5)(3)\sigma^2_{AD(B)} + 3\sigma^2_{C(A)D(B)} + \sigma^2_\varepsilon$
XX l	i(j) C(A) Random	4(5-1) = 16	$(4)(6)(3)\sigma^2_{C(A)} + (6)(3)\sigma^2_{BC(A)} + 3\sigma^2_{C(A)D(B)} + \sigma^2_\varepsilon$
XX m	k(l) D(B) Random	4(6-1)= 20	$(4)(5)(3)\sigma^2_{D(B)} + (5)(3)\sigma^2_{AD(B)} + 3\sigma^2_{C(A)D(B)} + \sigma^2_\varepsilon$
ij	i(jk) B C(A) Random	4(5-1)(4-1) = 48	$(6)(3)\sigma^2_{BC(A)} + 3\sigma^2_{C(A)D(B)} + \sigma^2_\varepsilon$
ik	k(il) A D(B) Random	4(4-1)(6-1) = 60	$(5)(3)\sigma^2_{AD(B)} + 3\sigma^2_{C(A)D(B)} + \sigma^2_\varepsilon$
XX il	ik(jl) C(A) D(B) Random	(4)(4)(5-1)(6-1) = 320	$3\sigma^2_{C(A)D(B)} + \sigma^2_\varepsilon$
XX im	ijkl(m) Error Random	(4)(5)(4)(6)(3-1) = 960	σ^2_ε
XX jk		Total= 1440 – 1 = 1439	
XX jl			
XX jm			
kl			
XX km			
XX lm			
ijk			
XX ijl			
XX ijm			
ikl			
xx ikm			
XX ilm			
XX jkl			
xx jkm			
XX jlm			
xx klm			
ijkl			
xx ijkm			
xx ijlm			
xx iklm			
xx jklm			
ijklm			

$$y_{ijklm} = \mu + \alpha_i + \beta_k + \alpha\beta_{ik} + \gamma_{ij} + \tau_{kl} + \alpha\tau_{ikl} + \beta\gamma_{ijk} + \gamma\tau_{ijkl} + \varepsilon_{ijklm}$$

Consider a company that buys raw material in batches from three different suppliers. The purity of this raw material differs considerably, which causes problems in manufacturing the finished product. … wish to determine if the differences in purity is attributable to differences between the suppliers. Four batches of raw material are selected at random from each supplier, and three determinations of purity are made on each batch. --- hence a two-stage nested design. Data is provided below … coded response …

```
options nodate nonumber ls=80 nocenter;
data nested; input purity supplier batch $ @@; cards;
1 1 1 -2 1 2 -2 1 3 1 1 4
-1 1 1 -3 1 2  0 1 3 4 1 4
0 1 1 -4 1 2  1 1 3 0 1 4
1 2 a  0 2 b -1 2 c 0 2 d
-2 2 a  4 2 b  0 2 c 3 2 d
-3 2 a  2 2 b -2 2 c 2 2 d
2 3 w -2 3 x  1 3 y 3 3 z
4 3 w  0 3 x -1 3 y 2 3 z
0 3 w  2 3 x  2 3 y 1 3 z
; run;

proc glm data = nested;
class supplier batch;
model purity = supplier batch(supplier);
random batch(supplier) / test;
run;
```

Dependent Variable: purity

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	84.9722222	7.7247475	2.93	0.0135
Error	24	63.3333333	2.6388889		
Corrected Total	35	148.3055556			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
supplier	2	15.05555556	7.52777778	2.85	0.0774
batch(supplier)	9	69.91666667	7.76851852	2.94	0.0167

Source	Type III Expected Mean Square
supplier	Var(Error) + 3 Var(batch(supplier)) + Q(supplier)
batch(supplier)	Var(Error) + 3 Var(batch(supplier))

Tests of Hypotheses for Mixed Model Analysis of Variance

Source	DF	Type III SS	Mean Square	F Value	Pr > F
supplier	2	15.055556	7.527778	0.97	0.4158
Error: MS(batch(supplier))	9	69.916667	7.768519		

Source	DF	Type III SS	Mean Square	F Value	Pr > F
batch(supplier)	9	69.916667	7.768519	2.94	0.0167
Error: MS(Error)	24	63.333333	2.638889		

We conclude that there is insufficient evidence of a supplier effect on purity, but the purity of batches of raw material from the same supplier does differ significantly … Note the practical implications of this experiment and the analysis are very important. If differences in raw material results from differences among the suppliers, we may be able to solve the problem by selecting the 'best' supplier. However, that solution is not applicable here b/c the major source of variability is the batch-to-batch purity variation within suppliers. We must attack the problem by working with the suppliers to reduce their batch-to-batch variability. … may involve modifications to the suppliers' production processes or their internal quality assurance system.