

Statistical analyses for nested experimental design

Jixiang Wu, South Dakota State University

Monday, April 03, 2017

```
## Warning: package 'knitr' was built under R version 3.3.2
```

Introduction

When two or more factors are considered in an experimental design, these treatment factors can be in a factorial way, which has been introduced in the previous chapter. However, other ways related to these treatment factors are possible too. Another common one is nested treatment designs. In this case, one factor is nested within another factor. The data analysis for such treatment designs can be conducted by either ANOVA methods or linear mixed model approaches, depending on the data structures. We will use several data sets to demonstrate the use of different methods.

Two-factor nested experimental designs

There are several models to analyze two-factor nested experimental designs based on if blocks are included. We assume that block effects are included.

Assume the above used data come from a nested genetic mating design, we can work on the nested experimental designs. When demonstrating data analysis, we frequently use the term "assume" or "assumption" or related word to use an appropriate method. In reality, you need to choose particular linear models and/or particular methods based on the objectives of a research project. It is very important to clarify in your own research work.

In this demonstration, we will use a previously used data set, which is available in the minque package. It is easy to load this data set following these R codes provided below.

```
require(minque)

## Loading required package: minque
## Loading required package: klaR
## Loading required package: MASS
## Loading required package: Matrix

data(ncii)
```

ANOVA based analysis

We may try to use the built-in function to run Unlike factorial factor design, the model system for nested factor design used in the aov will be a little different. Before we run an ANOVA analysis, we need to convert several variables in this data set.

```
dat=ncii
dat=transform(dat,Female=factor(Female),Male=factor(Male),Rep=factor(Rep))
```

I found three model expressions showed the same results. Knowing these expression sometimes is very helpful.

```
# Model expression 1
mod=aov(Yld~Male+Male/Female,data=dat)
a=summary(mod)
data.frame(a[[1]])
```

	Df	Sum.Sq	Mean.Sq	F.value	Pr..F.
Male	2	1718.764	859.38216	9.587483	0.0003401
Male:Female	12	24777.782	2064.81514	23.035596	0.0000000
Residuals	45	4033.613	89.63585	NA	NA

```
# Model expression 2
mod=aov(Yld~Male+Female:Male,data=dat)
a=summary(mod)
data.frame(a[[1]])
```

	Df	Sum.Sq	Mean.Sq	F.value	Pr..F.
Male	2	1718.764	859.38216	9.587483	0.0003401
Male:Female	12	24777.782	2064.81514	23.035596	0.0000000
Residuals	45	4033.613	89.63585	NA	NA

```
# Model expression 3
mod=aov(Yld~Male+Male:Female,data=dat)
a=summary(mod)
data.frame(a[[1]])
```

	Df	Sum.Sq	Mean.Sq	F.value	Pr..F.
Male	2	1718.764	859.38216	9.587483	0.0003401
Male:Female	12	24777.782	2064.81514	23.035596	0.0000000
Residuals	45	4033.613	89.63585	NA	NA

Remember that the above data analysis is only suitable when both female and male are fixed effects and no block effect is assumed.

If you want to include a block effect in the model, it is very easy to add that effect in the model as showed below.

```
# Model expression 1
```

```
mod=aov(Yld~Male+Male/Female+Rep,data=dat)
```

```
a=summary(mod)
```

```
data.frame(a[[1]])
```

	Df	Sum.Sq	Mean.Sq	F.value	Pr..F.
Male	2	1718.7643	859.38216	9.2088698	0.0004828
Rep	3	114.1255	38.04184	0.4076444	0.7482971
Male:Female	12	24777.7816	2064.81514	22.1259116	0.0000000
Residuals	42	3919.4876	93.32113	NA	NA

```
# Model expression 2
```

```
mod=aov(Yld~Male+Female:Male+Rep,data=dat)
```

```
a=summary(mod)
```

```
data.frame(a[[1]])
```

	Df	Sum.Sq	Mean.Sq	F.value	Pr..F.
Male	2	1718.7643	859.38216	9.2088698	0.0004828
Rep	3	114.1255	38.04184	0.4076444	0.7482971
Male:Female	12	24777.7816	2064.81514	22.1259116	0.0000000
Residuals	42	3919.4876	93.32113	NA	NA

```
# Model expression 3
```

```
mod=aov(Yld~Male+Male:Female+Rep,data=dat)
```

```
a=summary(mod)
```

```
data.frame(a[[1]])
```

	Df	Sum.Sq	Mean.Sq	F.value	Pr..F.
Male	2	1718.7643	859.38216	9.2088698	0.0004828
Rep	3	114.1255	38.04184	0.4076444	0.7482971
Male:Female	12	24777.7816	2064.81514	22.1259116	0.0000000
Residuals	42	3919.4876	93.32113	NA	NA

When female is random, based on EMS provided in this class, the F-test for factor Male is not right, by doing so, we may slightly change the model expression to obtain a right F-test for factor Male.

```
mod=aov(Yld~Male+Error(Male:Female)+Rep,data=dat)
```

```
## Warning in aov(Yld ~ Male + Error(Male:Female) + Rep, data = dat): Error()
## model is singular
```

```
summary(mod)
```

```
##
```

```
## Error: Male:Female
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Male       2   1719    859.4    0.416  0.669
## Residuals 12  24778   2064.8
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Rep        3    114    38.04    0.408  0.748
## Residuals 42   3919    93.32
```

Now you have a right F-test for factor male. All other F-tests please refer to the regular ANOVA results. Accordingly, if we want to compare the differences between any two males, we will need to use the mean square for female instead of the MSE in the original ANOVA analysis. The degrees of freedom will be that for female when a test or any pair-wise mean comparisons are performed.

Linear mixed model analysis

We may use linear mixed model analysis for this nested factor experimental design. By doing so, we will need to determine which effects are fixed or random. Again, the vertical bar "|" will be used to separate fixed effect(s) and random effects. The first case is for that both male and female are fixed.

Case 1: Both male and female are fixed

```
res=lmm.jack(Yld~Male+Male:Female|Rep,data=dat)[[1]]
data.frame(res$Var[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
V(Rep)	0.0194057	0.0613661	0.9780305	-0.1973558	0.2361671
V(e)	93.9641254	5.6661721	0.0000002	73.9496834	113.9785673

```
data.frame(res$PVar[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
V(Rep)/VP	0.0002351	0.0007435	0.9780305	-0.0023910	0.0028612
V(e)/VP	0.9997649	0.0007435	0.0000000	0.9971388	1.0023910

```
data.frame(res$FixedEffect[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
mu	86.0898489	0.4210378	0.0000000	84.6026302	87.577067
Male(6)	-0.0405724	0.5213790	0.9937588	-1.8822230	1.801078
Male(7)	5.5177798	0.3578130	0.0000004	4.2538883	6.781671
Male(8)	-5.4772074	0.3373617	0.0000002	-6.6688594	-4.285555
Male:Female(6:1)	-17.4100053	0.7450692	0.0000000	-20.0417900	-14.778221
Male:Female(7:1)	16.1111281	1.7422977	0.0000274	9.9568646	22.265391

Male:Female(8:1)	-27.9503081	0.8455030	0.0000000	-30.9368520	-24.963764
Male:Female(6:2)	-15.2984431	1.7201770	0.0000376	-21.3745705	-9.222316
Male:Female(7:2)	6.0323713	1.8531502	0.0389770	-0.5134531	12.578196
Male:Female(8:2)	4.9460252	1.4527019	0.0308352	-0.1853083	10.077359
Male:Female(6:3)	14.8386596	0.9707360	0.0000004	11.4097593	18.267560
Male:Female(7:3)	20.0430379	1.1039191	0.0000001	16.1436993	23.942377
Male:Female(8:3)	37.1223585	1.0908651	0.0000000	33.2691299	40.975587
Male:Female(6:4)	19.7284929	1.8524965	0.0000085	13.1849775	26.272008
Male:Female(7:4)	-27.1576287	1.9519811	0.0000009	-34.0525503	-20.262707
Male:Female(8:4)	-33.8026415	1.3939913	0.0000000	-38.7265932	-28.878690
Male:Female(6:5)	-1.8992764	2.1877705	0.8387200	-9.6270697	5.828517
Male:Female(7:5)	-9.5111287	1.8232955	0.0022051	-15.9514982	-3.070759
Male:Female(8:5)	14.2073585	1.5388521	0.0000277	8.7717195	19.642998

Case 2: Both male and female are random

The second case is that both male and female are random.

```
res=lmm.jack(Yld~1|Male+Male:Female+Rep,data=dat)[[1]]
data.frame(res$Var[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
V(Male)	0.0000000	0.0000000	1.0000000	0.0000000	0.0000000
V(Rep)	0.1965852	0.4555052	0.9631591	-1.412381	1.805552
V(Male:Female)	492.0468867	19.7713325	0.0000000	422.209229	561.884545
V(e)	93.0763519	7.9319043	0.0000037	65.058735	121.093969

```
data.frame(res$PVar[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
V(Male)/VP	0.0000000	0.0000000	1.0000000	0.0000000	0.0000000
V(Rep)/VP	0.0003424	0.0007928	0.9630983	-0.0024578	0.0031427
V(Male:Female)/VP	0.8404374	0.0151787	0.0000000	0.7868223	0.8940525
V(e)/VP	0.1592202	0.0154661	0.0000112	0.1045898	0.2138506

```
data.frame(res$FixedEffect[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
mu	86.12719	0.5796184	0	84.07982	88.17456

```
data.frame(res$RandomEffect[[1]])
```

	Pre	SE	PValue	X2.5.LL	X97.5.UL
Male(6)	0.0000000	0.0000000	1.0000000	0.0000000	0.0000000
Male(7)	0.0000000	0.0000000	1.0000000	0.0000000	0.0000000
Male(8)	0.0000000	0.0000000	1.0000000	0.0000000	0.0000000
Rep(1)	-0.1482982	0.3135516	0.9562620	-1.2558469	0.9592504
Rep(2)	-0.1648107	0.3809187	0.9629888	-1.5103177	1.1806964
Rep(3)	0.0843874	0.1803467	0.9571511	-0.5526457	0.7214205
Rep(4)	0.2287215	0.5086158	0.9602497	-1.5678463	2.0252893
Male:Female(6:1)	-18.4872867	1.4838209	0.0000022	-23.7285408	-13.2460325
Male:Female(7:1)	17.2282551	1.9636810	0.0000421	10.2920062	24.1645040
Male:Female(8:1)	-29.2982991	2.9900500	0.0000169	-39.8599589	-18.7366392
Male:Female(6:2)	-15.9907095	1.9515285	0.0000731	-22.8840324	-9.0973866
Male:Female(7:2)	6.7067474	1.6651579	0.0118711	0.8249623	12.5885324
Male:Female(8:2)	5.2397503	2.3881439	0.2028366	-3.1958155	13.6753161
Male:Female(6:3)	15.5824360	1.4869107	0.0000097	10.3302681	20.8346039
Male:Female(7:3)	21.2069782	1.7776041	0.0000032	14.9280031	27.4859533
Male:Female(8:3)	38.4966882	1.8263558	0.0000000	32.0455089	44.9478676
Male:Female(6:4)	20.8556309	2.6123627	0.0000900	11.6280639	30.0831979
Male:Female(7:4)	-28.2849839	1.9834755	0.0000007	-35.2911522	-21.2788155
Male:Female(8:4)	-35.8561715	2.4805515	0.0000006	-44.6181459	-27.0941970
Male:Female(6:5)	-1.9990954	2.2839311	0.8356955	-10.0665536	6.0683628
Male:Female(7:5)	-9.8162580	1.9259665	0.0025889	-16.6192890	-3.0132271
Male:Female(8:5)	14.4163180	2.2360906	0.0004741	6.5178451	22.3147908

Case 3: Male is fixed while female is random

The second case is that male is fixed while female is random.

```
res=lmm.jack(Yld~Male|Male:Female+Rep,data=dat)[[1]]
data.frame(res$Var[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
V(Rep)	0.0090026	0.0284687	0.9780305	-0.0915565	0.1095617
V(Male:Female)	494.5684316	15.6889615	0.0000000	439.1508046	549.9860585
V(e)	93.1527521	5.7332742	0.0000000	72.9012875	113.4042166

```
data.frame(res$PVar[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
V(Rep)/VP	0.0000159	0.0000501	0.9780305	-0.0001612	0.0001930
V(Male:Female)/VP	0.8414660	0.0091246	0.0000000	0.8092356	0.8736965
V(e)/VP	0.1585181	0.0091327	0.0000001	0.1262590	0.1907772

```
data.frame(res$FixedEffect[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
mu	86.1566203	0.4437518	0.0000000	84.589170	87.724071
Male(6)	-0.0875655	0.5473883	0.9900794	-2.021088	1.845957
Male(7)	6.6900818	0.7274284	0.0000286	4.120609	9.259554
Male(8)	-6.6025162	0.6648016	0.0000152	-8.950774	-4.254258

```
data.frame(res$RandomEffect[[1]])
```

	Pre	SE	PValue	X2.5.LL	X97.5.UL
Rep(1)	-0.0135217	0.0427593	0.9780305	-0.1645591	0.1375157
Rep(2)	-0.0353696	0.1118484	0.9780305	-0.4304483	0.3597092
Rep(3)	0.0184557	0.0583621	0.9780305	-0.1876949	0.2246064
Rep(4)	0.0304355	0.0962456	0.9780305	-0.3095299	0.3704010
Male:Female(6:1)	-18.3996020	1.8451663	0.0000147	-24.9172250	-11.8819789
Male:Female(7:1)	15.6203008	2.8033656	0.0013852	5.7180603	25.5225412
Male:Female(8:1)	-28.1483675	2.7057003	0.0000103	-37.7056278	-18.5911073
Male:Female(6:2)	-16.0852564	1.7666733	0.0000311	-22.3256211	-9.8448918
Male:Female(7:2)	5.0711576	2.0849847	0.1417215	-2.2935683	12.4358836
Male:Female(8:2)	6.3273876	2.2475897	0.0780198	-1.6117029	14.2664781
Male:Female(6:3)	15.5572582	1.2702584	0.0000026	11.0703642	20.0441523
Male:Female(7:3)	19.7213502	2.0402493	0.0000190	12.5146414	26.9280589
Male:Female(8:3)	40.2594507	1.4437948	0.0000000	35.1595794	45.3593219
Male:Female(6:4)	20.9083422	2.6889826	0.0001111	11.4101331	30.4065513
Male:Female(7:4)	-29.9442983	1.9964438	0.0000005	-36.9962742	-22.8923224
Male:Female(8:4)	-34.5788162	2.1852240	0.0000003	-42.2976144	-26.8600180
Male:Female(6:5)	-1.9807420	2.2712069	0.8370424	-10.0032549	6.0417709
Male:Female(7:5)	-10.4685103	3.7694202	0.0828263	-23.7831152	2.8460946
Male:Female(8:5)	16.1403454	1.6288323	0.0000154	10.3868721	21.8938187

Two-factor experimental designs with nested block effects

It is very often that some experiments are conducted in different environments. For such experiments, randomized complete blocks in each environment are independent of those in another one. That means block effects are nested within each environment.

```
library(agricolae)

data(plrv)
names(plrv)

## [1] "Genotype"      "Locality"      "Rep"           "WeightPlant"  "WeightPlot"
## [6] "Yield"

dat=plrv
dat=transform(dat,Geno=Genotype,Loc=Locality, Rep=factor(Rep))
names(dat)

## [1] "Genotype"      "Locality"      "Rep"           "WeightPlant"  "WeightPlot"
## [6] "Yield"         "Geno"          "Loc"
```

ANOVA methods

```
require(minque)
mod=aov(Yield~Geno*Loc+Loc/Rep,data=dat)
a=summary(mod)
data.frame(a[[1]])
```

	Df	Sum.Sq	Mean.Sq	F.value	Pr..F.
Geno	27	17533.281	649.38078	17.535906	0.0000000
Loc	5	122284.346	24456.86916	660.434332	0.0000000
Geno:Loc	135	23761.795	176.01330	4.753070	0.0000000
Loc:Rep	12	1141.785	95.14878	2.569402	0.0028887
Residuals	324	11998.204	37.03149	NA	NA

If we assume all effects are fixed, then the results in the above ANOVA table look ok. However, if block effect is random or any other factor is random, based on the derived EMS, some F-tests may not be right. Thus, with such a model, ANOVA analysis for F-tests may be problematic and linear mixed model approached are more preferred.

Linear mixed model approaches

Case 1: Both genotype and environment are fixed

```
require(minque)

res=lmm.jack(Yield~Loc*Geno|Loc:Rep,data=dat)[[1]]
data.frame(res$Var[[1]])
```


	Estimate	SE	PValue	X2.5.LL	X97.5.UL
V(Loc:Rep)	2.202455	0.6632316	0.035163	-0.1402571	4.545167
V(e)	36.972805	1.3270367	0.000000	32.2853549	41.660255

```
data.frame(res$PVar[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
V(Loc:Rep)/VP	0.0563057	0.0172517	0.0384545	-0.0046318	0.1172432
V(e)/VP	0.9436943	0.0172517	0.0000000	0.8827568	1.0046318

```
head(data.frame(res$FixedEffect[[1]]))
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
mu	29.7436691	2.698851	0.0000063	20.210601	39.276738
Loc(Ayac)	-3.6389076	3.287643	0.7237249	-15.251745	7.973930
Loc(LM-02)	-3.6389068	3.287642	0.7237249	-15.251743	7.973929
Loc(SR-02)	-0.3181576	1.658233	0.9882247	-6.175483	5.539168
Loc(Hyo-02)	1.5261900	3.107071	0.9529330	-9.448818	12.501199
Loc(LM-03)	3.7239352	5.095013	0.8906925	-14.273021	21.720891

```
head(data.frame(res$RandomEffect[[1]]))
```

	Pre	SE	PValue	X2.5.LL	X97.5.UL
Loc:Rep(Ayac:1)	0.1693823	0.4632113	0.9723861	-1.4668046	1.8055692
Loc:Rep(Ayac:2)	0.6057069	0.3347713	0.3472509	-0.5767954	1.7882093
Loc:Rep(Ayac:3)	-0.7750893	0.3992675	0.2909507	-2.1854092	0.6352307
Loc:Rep(LM-02:1)	2.2544395	0.4114606	0.0015614	0.8010501	3.7078289
Loc:Rep(LM-02:2)	0.6566264	0.3493362	0.3163697	-0.5773230	1.8905758
Loc:Rep(LM-02:3)	-2.9110659	0.5721572	0.0026202	-4.9320788	-0.8900530

Case 2: Both genotype and environment are random

```
res=lmm.jack(Yield~1|Geno*Loc+Rep:Loc,data=dat)[[1]]
```

```
data.frame(res$Var[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
V(Geno)	26.264434	1.9588603	0.0000012	19.3452131	33.183655
V(Loc)	288.643219	5.0426925	0.0000000	270.8310744	306.455364
V(Geno:Loc)	45.993977	1.2512883	0.0000000	41.5740909	50.413864
V(Loc:Rep)	1.964619	0.4560123	0.0078398	0.3538612	3.575377
V(e)	37.362891	1.6082036	0.0000000	31.6822835	43.043498

```
data.frame(res$PVar[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
V(Geno)/VP	0.0656085	0.0045344	0.0000006	0.0495919	0.0816251
V(Loc)/VP	0.7211740	0.0059276	0.0000000	0.7002361	0.7421119
V(Geno:Loc)/VP	0.1149207	0.0028200	0.0000000	0.1049596	0.1248817
V(Loc:Rep)/VP	0.0049131	0.0011560	0.0085357	0.0008299	0.0089964
V(e)/VP	0.0933837	0.0045735	0.0000000	0.0772287	0.1095386

```
data.frame(res$FixedEffect[[1]])
```

	Estimate	SE	PValue	X2.5.LL	X97.5.UL
mu	30.67555	0.1212494	0	30.24727	31.10384

```
head(data.frame(res$RandomEffect[[1]]))
```

	Pre	SE	PValue	X2.5.LL	X97.5.UL
Geno(102.18)	-3.6511460	0.5944079	0.0006808	-5.7507544	-1.5515377
Geno(104.22)	0.5418619	0.2408313	0.1867132	-0.3088189	1.3925427
Geno(121.31)	-0.4919683	0.4131696	0.6781046	-1.9513942	0.9674576
Geno(141.28)	7.7195038	0.6655875	0.0000041	5.3684701	10.0705375
Geno(157.26)	5.3266253	0.4751735	0.0000055	3.6481848	7.0050658
Geno(163.9)	-7.7677790	0.6176531	0.0000021	-9.9494958	-5.5860621

Conclusions

When two factors or more factors involve, experimental designs can be very complicated. In the same manner, the corresponding data analyses can be complicated too. The key for these experimental analyses is to write an appropriate linear model. ANOVA based approaches can be used for data analyses once you can provide an appropriate linear model. The ANOVA results, especially F-testing results, generated by the function of aov are based on that all effects are fixed. When at least one effect is random, some F-tests can be very complicated. Thus, linear mixed model approaches are preferred.