

General Linear Model of an Unbalanced Factorial Design

Potato Damage Data

This is an unbalanced design as we have different numbers of replicated in the $2 \times 2 \times 2 = 8$ cells of the table.

Temperature Pre-treatment * Potato variety * Acclimatization Routine Crosstabulation

Count

Acclimatization Routine			Potato variety		Total
			Variety 1	Variety 2	
Room Temp	Temperature	-4 C	5	13	18
	Pre-treatment	-8 C	5	13	18
	Total		10	26	36
Cold Room	Temperature	-4 C	12	7	19
	Pre-treatment	-8 C	13	7	20
	Total		25	14	39

Number of parameters:
k=7

Three-way Interaction Model (COMPLETE MODEL)

Dependent Variable: Damage Score: Ion Leakage

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8842.339(a)	7	1263.191	17.033	.000
Intercept	8055.406	1	8055.406	108.619	.000
potato	1892.313	1	1892.313	25.516	.000
regime	1493.822	1	1493.822	20.143	.000
temp	803.280	1	803.280	10.831	.002
potato * regime	2087.539	1	2087.539	28.148	.000
potato * temp	48.135	1	48.135	.649	.423
regime * temp	13.891	1	13.891	.187	.667
potato * regime * temp	89.198	1	89.198	1.203	.277
Error	4968.876	67	74.162		
Total	27481.316	75			
Corrected Total	13811.215	74			

a R Squared = .640 (Adjusted R Squared = .603)

It appears that the fit is moderate (R squared = 0.640), but that there is some explanatory power in the variables.

Note that we cannot interpret the quoted F statistics, as this is an unbalanced design, and therefore the stated p-values are not in general exact. However, these results do give an indication of which terms might be omitted.

REDUCED MODEL 1

Number of parameters:
g=4

Dependent Variable: Damage Score: Ion Leakage

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8717.469(a)	4	2179.367	29.950	.000
Intercept	8060.776	1	8060.776	110.774	.000
potato	1890.703	1	1890.703	25.983	.000
regime	1492.360	1	1492.360	20.509	.000
temp	1225.714	1	1225.714	16.844	.000
potato * regime	2089.928	1	2089.928	28.721	.000
Error	5093.746	70	72.768		
Total	27481.316	75			
Corrected Total	13811.215	74			

Interaction terms omitted:

Three-way interaction:
potato*regime*temp

Two-way interactions:
potato*temp
regime*temp

a R Squared = .631 (Adjusted R Squared = .610)

REDUCED MODEL 2

Number of parameters:
g=3

Dependent Variable: Damage Score: Ion Leakage

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6627.541(a)	3	2209.180	21.834	.000
Intercept	13233.292	1	13233.292	130.792	.000
potato	1502.970	1	1502.970	14.855	.000
regime	1977.340	1	1977.340	19.543	.000
temp	1255.583	1	1255.583	12.410	.001
Error	7183.674	71	101.179		
Total	27481.316	75			
Corrected Total	13811.215	74			

Remaining interaction term
potato*regime omitted

a R Squared = .480 (Adjusted R Squared = .458)

REDUCED MODEL 3

Number of parameters:
g=3

Dependent Variable: Damage Score: Ion Leakage

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7491.755(a)	3	2497.252	28.057	.000
Intercept	8119.673	1	8119.673	91.226	.000
potato	1862.829	1	1862.829	20.929	.000
regime	1467.591	1	1467.591	16.489	.000
potato * regime	2119.797	1	2119.797	23.816	.000
Error	6319.460	71	89.006		
Total	27481.316	75			
Corrected Total	13811.215	74			

Interaction replaced,
but temp main effect
removed.

a R Squared = .542 (Adjusted R Squared = .523)

Balanced two-factor predictor, one covariate linear model

Task Distraction Data

This is a balanced design, with 15 replicates in each of the 3 x 3 = 9 cells of the table. For a balanced design, the quoted p-values are more reliable as indications of significance

COMPLETE MODEL

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39671.916(a)	17	2333.642	48.240	.000
Intercept	309.723	1	309.723	6.402	.013
Group	252.027	2	126.014	2.605	.078
Task	450.584	2	225.292	4.657	.011
Distract	2790.513	1	2790.513	57.684	.000
Group * Task	172.095	4	43.024	.889	.473
Group * Distract	335.100	2	167.550	3.463	.035
Task * Distract	2535.238	2	1267.619	26.203	.000
Group * Task * Distract	142.924	4	35.731	.739	.567
Error	5660.010	117	48.376		
Total	90341.000	135			
Corrected Total	45331.926	134			

k=17

a R Squared = .875 (Adjusted R Squared = .857)

REDUCED MODEL 1

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	37704.447(a)	9	4189.383	68.656	.000
Intercept	537.895	1	537.895	8.815	.004
Group	228.483	2	114.242	1.872	.158
Task	494.293	2	247.147	4.050	.020
Distract	3575.111	1	3575.111	58.589	.000
Group * Distract	343.795	2	171.898	2.817	.064
Task * Distract	2540.469	2	1270.235	20.817	.000
Error	7627.479	125	61.020		
Total	90341.000	135			
Corrected Total	45331.926	134			

g=9

a R Squared = .832 (Adjusted R Squared = .820)

REDUCED MODEL 2

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	37360.652(a)	7	5337.236	85.034	.000
Intercept	619.535	1	619.535	9.871	.002
Group	433.379	2	216.690	3.452	.035
Task	500.794	2	250.397	3.989	.021
Distract	3796.748	1	3796.748	60.491	.000
Task * Distract	2597.561	2	1298.780	20.692	.000
Error	7971.274	127	62.766		
Total	90341.000	135			
Corrected Total	45331.926	134			

g=7

a R Squared = .824 (Adjusted R Squared = .814)

REDUCED MODEL 3

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	36927.272(a)	5	7385.454	113.357	.000
Intercept	522.634	1	522.634	8.022	.005
Task	513.356	2	256.678	3.940	.022
Distract	3565.647	1	3565.647	54.728	.000
Task * Distract	2750.062	2	1375.031	21.105	.000
Error	8404.654	129	65.152		
Total	90341.000	135			
Corrected Total	45331.926	134			

g=5

a R Squared = .815 (Adjusted R Squared = .807)

REDUCED MODEL 4

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	34177.211(a)	3	11392.404	133.791	.000
Intercept	1192.389	1	1192.389	14.003	.000
Task	23726.782	2	11863.391	139.323	.000
Distract	5515.685	1	5515.685	64.776	.000
Error	11154.715	131	85.150		
Total	90341.000	135			
Corrected Total	45331.926	134			

g=3

a R Squared = .754 (Adjusted R Squared = .748)

Task Distraction Data: Follow-Up Analysis

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39671.916(a)	17	2333.642	48.240	.000
Intercept	309.723	1	309.723	6.402	.013
Group	252.027	2	126.014	2.605	.078
Task	450.584	2	225.292	4.657	.011
Distract	2790.513	1	2790.513	57.684	.000
Group * Task	172.095	4	43.024	.889	.473
Group * Distract	335.100	2	167.550	3.463	.035
Task * Distract	2535.238	2	1267.619	26.203	.000
Group * Task * Distract	142.924	4	35.731	.739	.567
Error	5660.010	117	48.376		
Total	90341.000	135			
Corrected Total	45331.926	134			

a R Squared = .875 (Adjusted R Squared = .857)

Now we omit the three-way interaction only

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39528.992(a)	13	3040.692	63.403	.000
Intercept	305.861	1	305.861	6.378	.013
Group	316.691	2	158.345	3.302	.040
Task	481.392	2	240.696	5.019	.008
Distract	2802.472	1	2802.472	58.436	.000
Group * Task	1824.545	4	456.136	9.511	.000
Group * Distract	414.362	2	207.181	4.320	.015
Task * Distract	2643.278	2	1321.639	27.558	.000
Error	5802.934	121	47.958		
Total	90341.000	135			
Corrected Total	45331.926	134			

a R Squared = .872 (Adjusted R Squared = .858)

Here, to compare these models,

$$F = \frac{(5802.934 - 5660.010)/(17-13)}{5660.010/(135-17-1)} = 0.739$$

We compare this with the Fisher-F(17-13,135-17-1) = Fisher-F(4,117) distribution:
the 0.05 tail quantile Critical Value is 2.45.

Therefore we do not reject the simpler model as an adequate simplification: we CAN drop the three-way interaction.

Now we try to drop the least significant two-way interaction: group*distract

Dependent Variable: Errors

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39114.630(a)	11	3555.875	70.348	.000
Intercept	332.570	1	332.570	6.579	.012
Group	364.584	2	182.292	3.606	.030
Task	521.312	2	260.656	5.157	.007
Distract	2871.665	1	2871.665	56.812	.000
Group * Task	1753.978	4	438.495	8.675	.000
Task * Distract	2725.029	2	1362.514	26.955	.000
Error	6217.296	123	50.547		
Total	90341.000	135			
Corrected Total	45331.926	134			

a R Squared = .863 (Adjusted R Squared = .851)

Here, to compare these models,

$$F = \frac{(6217.296 - 5802.934)/(13-11)}{5802.934/(135-13-1)} = 4.320$$

We compare this with the Fisher-F(13-11,135-17-1) = Fisher-F(2,121) distribution:
the approximate 0.05 tail quantile Critical Value is 3.07.

Therefore we reject the simpler model as an adequate simplification.

The conclusion is that the most appropriate model in terms of ANOVA F-test selection is

group + task + distract + group.task + group.distract + task.distract

Note that there is very little difference between the R squared statistics for the models.