Emperical Methods for Marketing Research and Analytics Using R

Prof. Dr. Martin Wetzels
Maastricht University

About Me

- **Name:** Prof. Dr. Martin Wetzels
  Professor in Marketing and Supply Chain Research

- **Address:** Maastricht University
  School of Business and Economics
  Department of Marketing and SCM
  P.O. Box 616, 6200 MD Maastricht
  Tongersestraat 53, F1.05
  T: +31 43 388 3250/3839
  F: +31 43 388 4918
  E: m.wetzels@maastrichtuniversity.nl
  W: https://www.maastrichtuniversity.nl/m.wetzels

Using R for AN(C)OVA

Prof. Dr. Martin Wetzels
Maastricht University

Maastricht University

Course Outline

Session:	TOPIC:
DAY 1	
Session 1	INTRODUCING MULTIVARIATE ANALYSIS AND R
Session 2	USING R FOR BASIC ANALYSIS
DAY 2	
Session 3	USING R FOR AN(C)OVA
Session 4	USING R FOR REGRESSION ANALYSIS
DAY 3	
Session 5	USING R FOR SCALING AND FACTOR ANALYSIS
Session 6	USING R FOR SEM and PLS PATH MODELING

Positioning AN(C)OVA…

Hair et al. (2018); Malhotra (2010)

Multivariate Data Analysis Tools

Dependence Techniques

	Y	X_i
AN(C)OVA	Metric	Nominal
Regression Analysis	Metric	Metric
Conjoint Analysis	Metric	Nominal

Interdependence Techniques

Variables

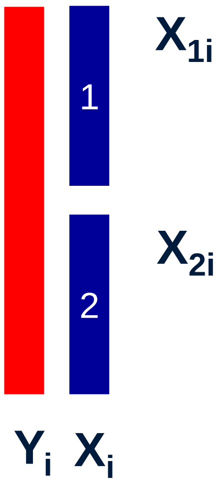
- Factor Analysis

Objects

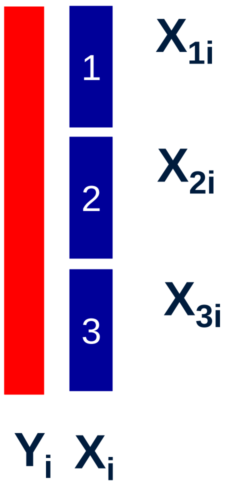
- Cluster Analysis
- Perceptual Mapping

Analysis of Variance (ANOVA): Oneway ANOVA

"Two"-Sample Independent-Samples



Oneway ANOVA



Analysis of Variance (ANOVA): Oneway ANOVA

Malhotra (2010); Pallant (2016)

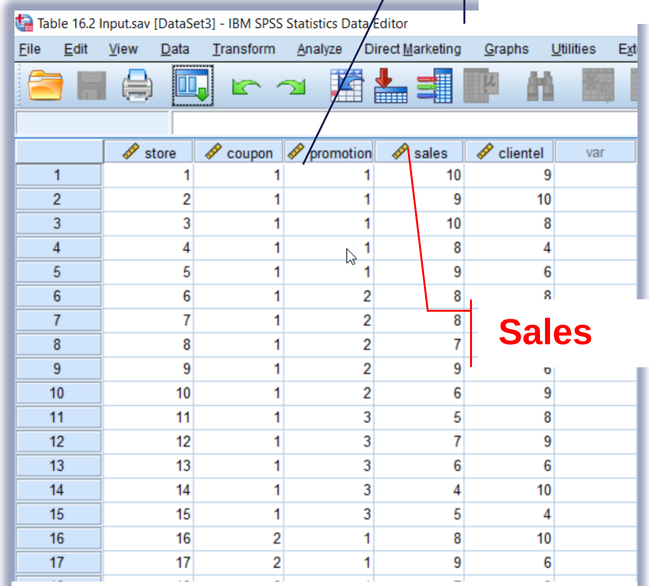
- Assumptions
  - Variables
    - **Y_i = intervally scaled**
    - **X_i = nominal (“Group”)**
  - Independent samples
  - For each “group” of X, Y has the same variance (*homogeneity of variance*) and is randomly sampled from a normal distribution
- Hypotheses
  - H_0: $M_1 = M_2 = M_3 = M_j$
- Model
  - $Y_i = a + b_1 * X_i +$ **ERROR**

Error term is normally distributed (M=0, VAR=c) and the error terms are uncorrelated

Analysis of Variance (ANOVA): Oneway ANOVA

Department Store Data (Malhotra, 2010, ...

Promotion



	store	coupon	promotion	sales	clientel	var
1	1	1	1	10	9	
2	2	1	1	9	10	
3	3	1	1	10	8	
4	4	1	1	8	4	
5	5	1	1	9	6	
6	6	1	2	8	8	
7	7	1	2	8	[illegible]	
8	8	1	2	7	[illegible]	
9	9	1	2	9	[illegible]	
10	10	1	2	6	9	
11	11	1	3	5	8	
12	12	1	3	7	9	
13	13	1	3	6	6	
14	14	1	3	4	10	
15	15	1	3	5	4	
16	16	2	1	8	10	
17	17	2	1	9	6	

Sales

Factorial design with two factors with 3 (Promotion) and 2 (Coupon) levels (= 6 cells) with 5 stores randomly assigned (n=6*5=30).

Experimental Design

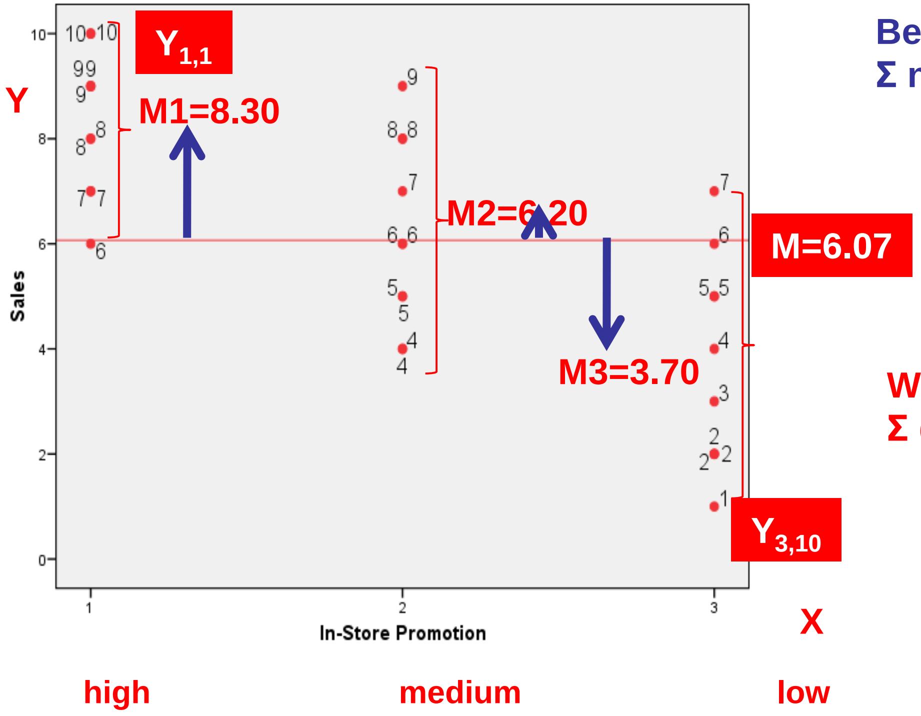
"X" Promotion
(1=high, 2=medium, 3= low)
"X" Coupon
(1=$20 coupon, 2=No coupon)
"X" Clientel
(Affluence of clientele, 1-10)
"Y" Sales
(normalized, scale 1-10)

Two procedures:

- Analyze→Compare Means →Oneway ANOVA
- Analyze→General Linear Model→Univariate

Analysis of Variance (ANOVA): Oneway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)

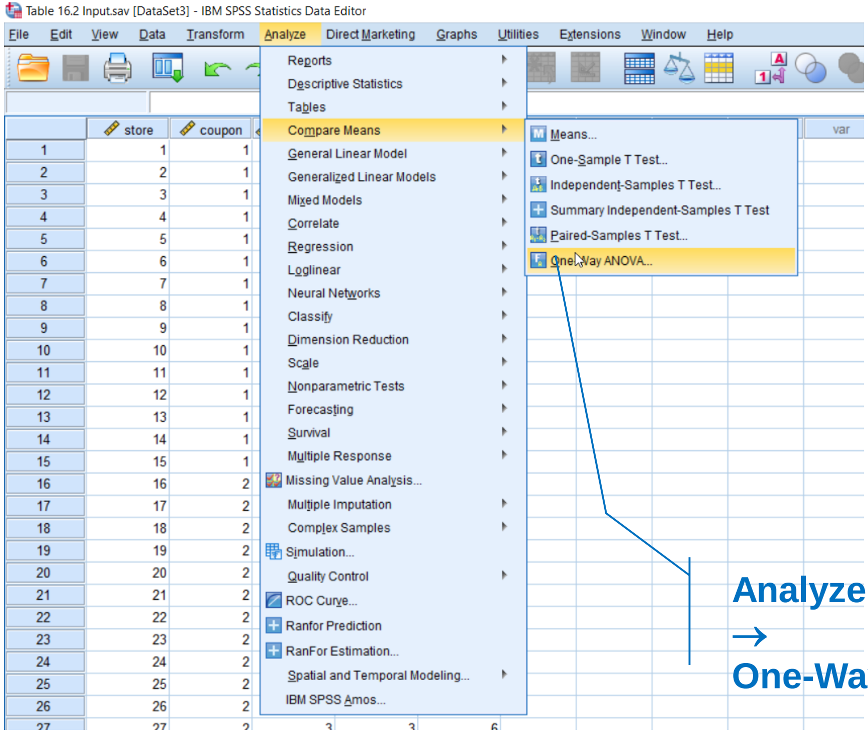


Between SS:
$\Sigma\ n_j*(M_j-M)^2$ **for all groups**

Within SS (Error):
$\Sigma\ (Y_{ij}-M_j)^2$ **for all groups**

Analysis of Variance (ANOVA): Oneway ANOVA

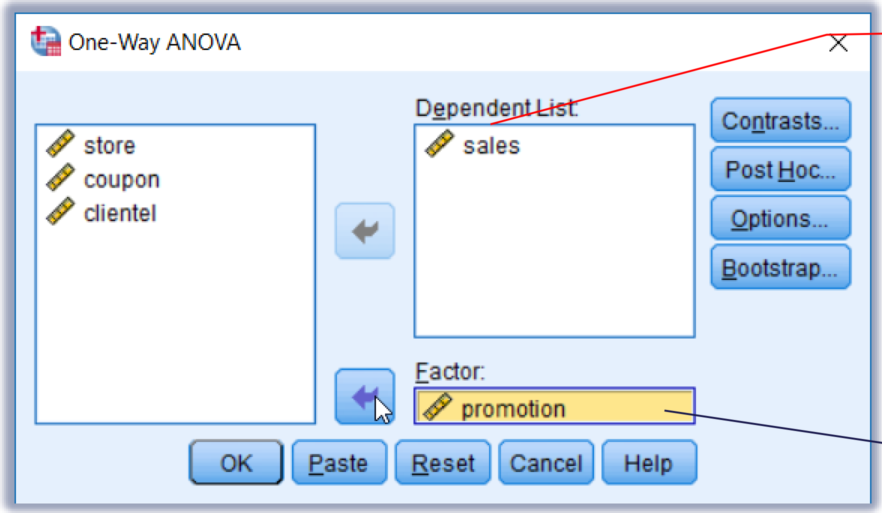
Department Store Data (Malhotra, 2010, Table 16.2)



Analyze → Compare Means → One-Way ANOVA

Analysis of Variance (ANOVA): Oneway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)



DV: SALES

IV: PROMOTION

Analysis of Variance (ANOVA): Oneway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)

Descriptives and Assumptions

Descriptives

sales Sales

	N	M Mean	SD Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 High	10	8.30	1.337	.423	7.34	9.26	6	10
2 Medium	10	6.20	1.751	.554	4.95	7.45	4	9
3 Low	10	3.70	2.003	.633	2.27	5.13	1	7
Total	30	6.07	2.532	.462	5.12	7.01	1	10

Test of Homogeneity of Variances

sales Sales

Levene Statistic	df1	df2	Sig.
1.353	2	27	.275

Homogeneity of Variance Assumption

H_0: all variances are equal

Analysis of Variance (ANOVA): Oneway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)

Explained by PROMOTION
(η^2=106.067/185.867=0.571)

ANOVA Table

df_X=c-1 = 3-1 =2
df_E=N-c=30-3 = 27

106.067/2

F(2,27)=53.033/2.956 =17.944 (p < 0.001!!!)

ANOVA

sales Sales

		Sum of Squares	df	Mean Square	F	Sig.
SS_X	Between Groups	106.067	2	53.033	17.944	.000
SS_E	Within Groups	79.800	27	2.956		
SS_Y	Total	185.867	29			

79.800/27

Error – Not Explained by PROMOTION

Analysis of Variance (ANOVA): Oneway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)

- What to do if Levene's test is significant (i.e., homogeneity of variance is violated)?
  - Nothing! ANOVA is quite robust
  - Welch or Brown-Forsythe's F test (available in Oneway ANOVA)
  - Kruskal-Wallis test (nonparametric alternative!)

Analysis of Variance (ANOVA): Oneway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)

Robust Tests of Equality of Means

sales Sales

	Statistic[a]	df1	df2	Sig.
Welch	18.090	2	17.470	.000
Brown-Forsythe	17.944	2	24.659	.000

a. Asymptotically F distributed.

Welch (1951)
$F_W(2,17.470) = 18.090$, $p<0.001$

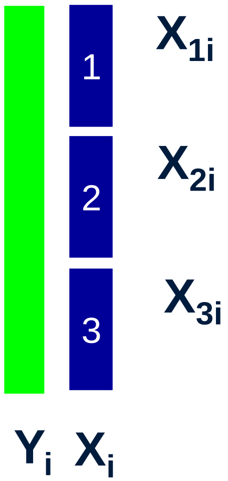
Brown and Forsyth (1974)
$F_{BF}(2,24.659) = 17.944$, $p<0.001$

Kruskal-Wallis Test

Department Store Data (Malhotra, 2010, Table 16.2)

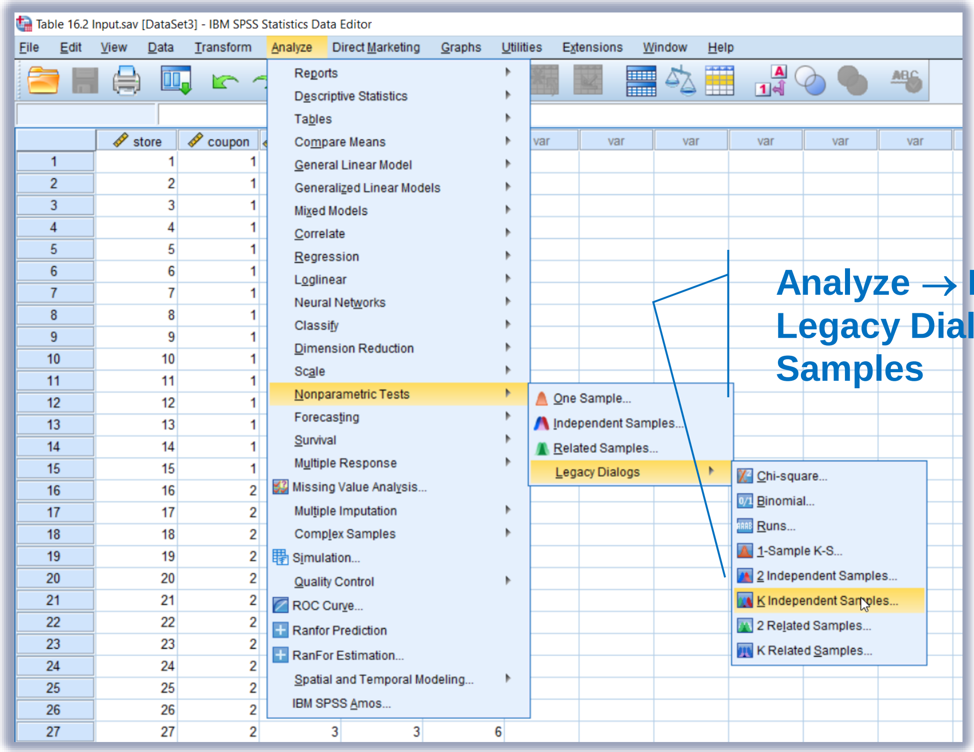
- Assumptions
  - Three or more (random) samples
  - "Dependent“ variable must be at least **ordinally** scaled and the "independent" variable must be **nominally** scaled
- Hypotheses
  - H_0: $Median_1 = Median_2 = \ldots = Median_k$
- Extension of Wilcoxon-Mann-Whitney test for more than 2 samples

Kruskal-Wallis Test



Kruskal-Wallis Test

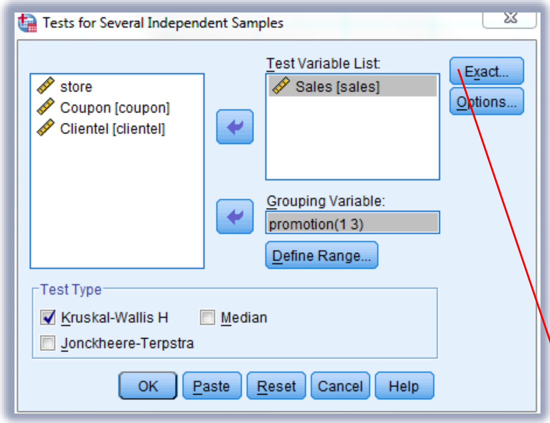
Department Store Data (Malhotra, 2010, Table 16.2)



Analyze → Nonparametric Tests → Legacy Dialogs → K Independent Samples

Kruskal-Wallis Test

Department Store Data (Malhotra, 2010, Table 16.2)



EXACT!

RANKS

Ranks

	In-Store Promotion	N	Mean Rank
Sales	High	10	23.50
	Medium	10	15.40
	Low	10	7.60
	Total	30	

Test Statistics[a,b]

	Sales
Chi-Square	16.529
df	2
Asymp. Sig.	.000
Exact Sig.	.000
Point Probability	.000

a. Kruskal Wallis Test
b. Grouping Variable: In-Store Promotion

$\chi^2(2)=16.529$ ($p_2<0.001$)

ES

$\eta^2=\chi^2/(n-1)$

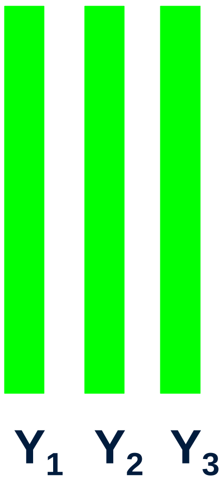
$\eta^2=16.529/29=0.570$

Friedman Test

Pallant (2016)

- Assumptions
  - Three or more (random) samples
  - Variables must be at least **ordinally** scaled
- Hypotheses
  - H_0: $Median_{Y1} = Median_{Y2} = \ldots = Median_{Yk}$
- Extension of Wilcoxon signed rank test for more than 2 samples

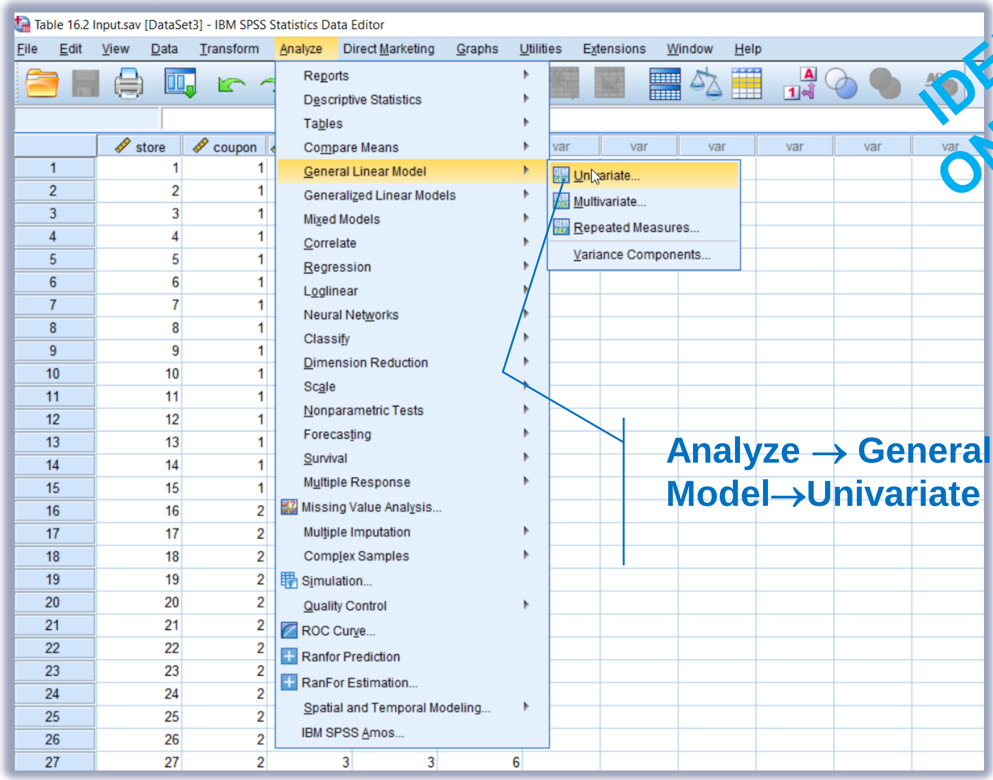
Friedman Test



GLM Univariate

Department Store Data (Malhotra, 2010, Table 16.2)

IDENTICAL TO ONEWAY-ANOVA



Analyze → General Linear Model→Univariate

GLM Univariate

Department Store Data (Malhotra, 2010, Table 16.2)

IDENTICAL TO ONEWAY-ANOVA

Descriptives and Assumptions

Descriptive Statistics

Dependent Variable: sales Sales

promotion In-Store	Mean (M)	Std. Deviation (SD)	N (N)
1 High	8.30	1.337	10
2 Medium	6.20	1.751	10
3 Low	3.70	2.003	10
Total	6.07	2.532	30

Levene's Test of Equality of Error Variances[a]

Dependent Variable: sales Sales

F	df1	df2	Sig.
1.353	2	27	.275

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+promotion

Homogeneity of Variance Assumption

H_0: all variances are equal

GLM Univariate

Department Store Data (Malhotra, 2010, Table 16.2)

IDENTICAL TO ONEWAY-ANOVA

ANOVA Table

Tests of Between-Subjects Effects

Dependent Variable: sales Sales

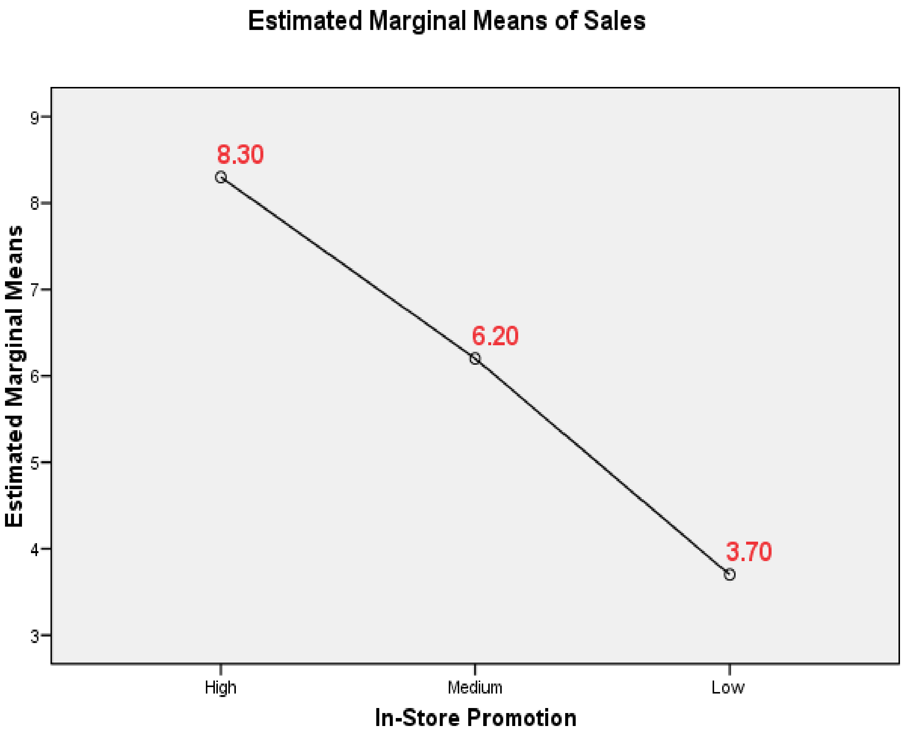
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	106.067[a]	2	53.033	17.944	.000	.571
Intercept	~~1104.133~~	1	1104.133	373.579	.000	.933
promotion	106.067	2	53.033	17.944	.000	.571
Error	79.800	27	2.956			
Total	~~1290.000~~	30				
Corrected Total	185.867	29				

a. R Squared = .571 (Adjusted R Squared = .539)

GLM Univariate

Department Store Data (Malhotra, 2010, Table 16.2)

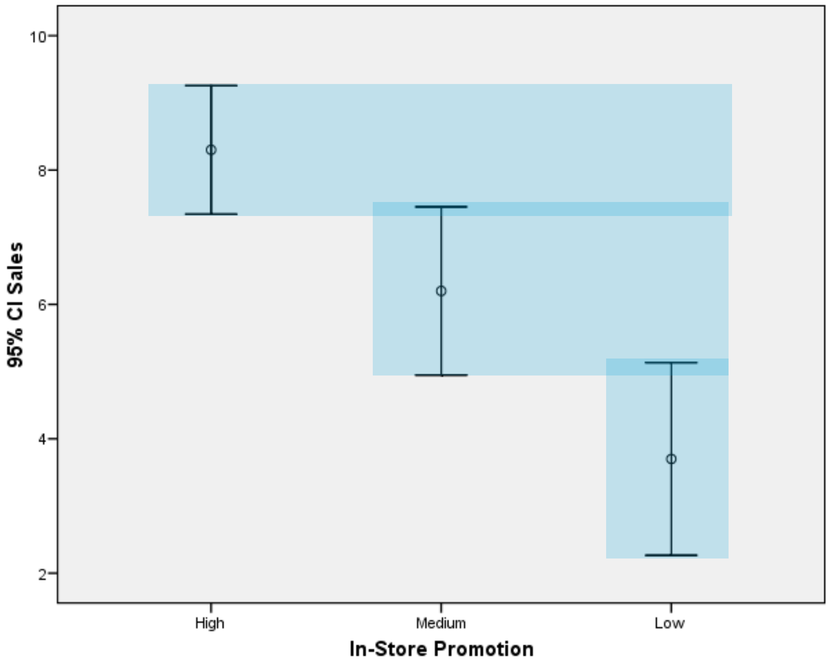
Mean Plot



GLM Univariate

Department Store Data (Malhotra, 2010, Table 16.2)

Error Bar Graph



GLM Univariate: Post Hoc Tests

Department Store Data (Malhotra, 2010, Table 16.2)

Why do we need Post Hoc Multiple Comparison tests?

Multiple tests Inflate Familywise α!
$FW\alpha = 1-(1-\alpha)^T$

EXAMPLE
T=1, α=0.05, FWα=0.05
T=2, α=0.05, FWα=0.10
T=3, α=0.05, FWα=0.14

Sig. = p value!

p value for the individual tests

Multiple Comparisons

Dependent Variable: sales Sales
Tukey HSD

(I) In-Store Promotion	(J) In-Store Promotion	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1 High	2 Medium	2.10*	.769	.029	.19	4.01
	3 Low	4.60*	.769	.000	2.69	6.51
2 Medium	1 High	-2.10*	.769	.029	-4.01	-.19
	3 Low	2.50*	.769	.008	.59	4.41
3 Low	1 High	-4.60*	.769	.000	-6.51	-2.69
	2 Medium	-2.50*	.769	.008	-4.41	-.59

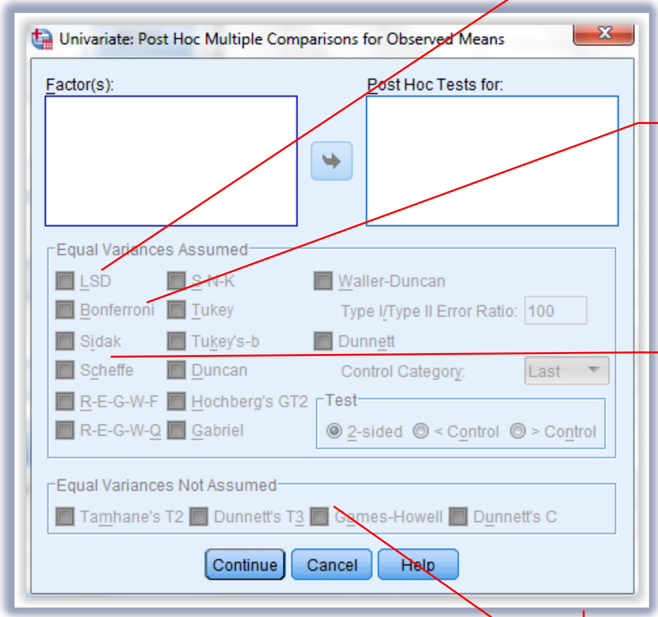
Based on observed means.

*. The mean difference is significant at the .05 level.

GLM Univariate: Post Hoc Tests

Department Store Data (Malhotra, 2010, Table 16.2)

Robust to nonnormality!



LSD

- Equal variances and sample sizes assumed
- Very liberal (t tests)!

Bonferroni, REGWQ and Tukey

- Equal variances and sample sizes assumed
- All pairwise comparisons

Scheffé

- Equal variances and sample sizes assumed
- All pairwise comparisons
- Very conservative!

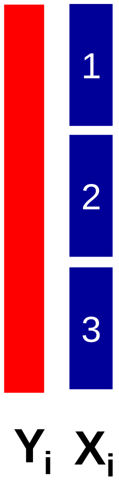
Games-Howell

- Equal variances and sample sizes **NOT** assumed
- All pairwise comparisons

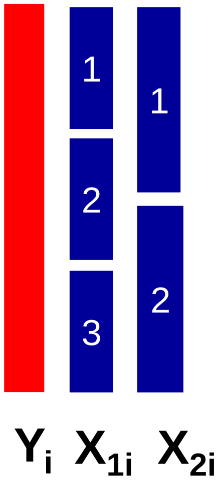
Analysis of Variance (ANOVA): Twoway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)

Oneway ANOVA



Twoway ANOVA



Analysis of Variance (ANOVA): Twoway ANOVA

Malhotra (2010); Pallant (2016)

- Assumptions
  - Variables
    - **Y_i = intervally scaled**
    - **X_i = nominal ("Group")**
  - Independent samples
  - For each "group" of X, Y has the same variance (*homogeneity of variance*) and is randomly sampled from a normal distribution
  - Variables

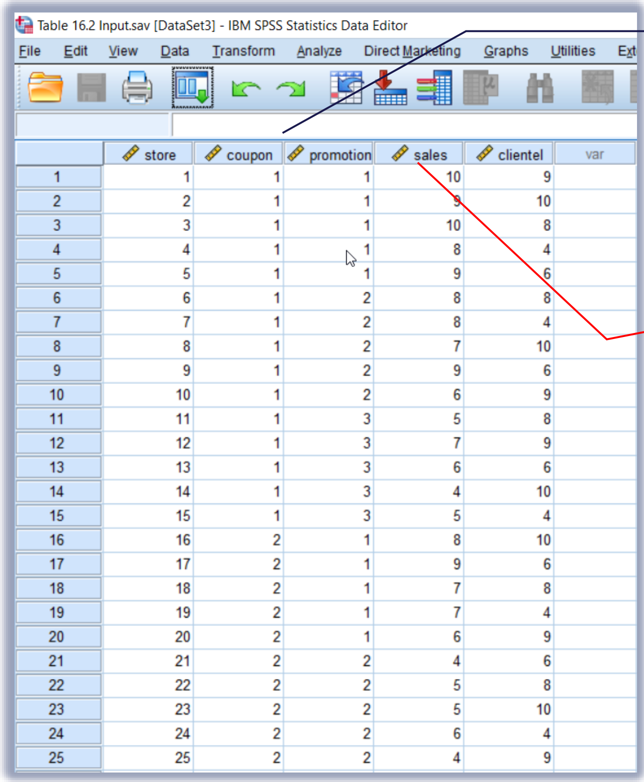
Error term is normally distributed (M=0, VAR=c) and the error terms are uncorrelated

- Model
  - $Y_i = a + b_1 * X_{1i} + + b_2 * X_{2i} + b_3 * X_{1i} * X_{2i}$ + **ERROR**

Interaction Term

Analysis of Variance (ANOVA): Twoway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)



	store	coupon	promotion	sales	clientel	var
1	1	1	1	10	9	
2	2	1	1	9	10	
3	3	1	1	10	8	
4	4	1	1	8	4	
5	5	1	1	9	6	
6	6	1	2	8	8	
7	7	1	2	8	4	
8	8	1	2	7	10	
9	9	1	2	9	6	
10	10	1	2	6	9	
11	11	1	3	5	8	
12	12	1	3	7	9	
13	13	1	3	6	6	
14	14	1	3	4	10	
15	15	1	3	5	4	
16	16	2	1	8	10	
17	17	2	1	9	6	
18	18	2	1	7	8	
19	19	2	1	7	4	
20	20	2	1	6	9	
21	21	2	2	4	6	
22	22	2	2	5	8	
23	23	2	2	5	10	
24	24	2	2	6	4	
25	25	2	2	4	9	

IVs: Coupon + Promotion

DV: Sales

Analysis of Variance (ANOVA): Twoway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)

Table 16.2 Input.sav [DataSet3] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Ext

	store	coupon	promotion	sales	clientel	var
1	1	1	1	10	9	
2	2	1	1	9	10	
3	3	1	1	10	8	
4	4	1	1	8	4	
5	5	1	1	9	6	
6	6	1	2	8	8	
7	7	1	2	8	4	
8	8	1	2	7	10	
9	9	1	2	9	6	
10	10	1	2	6	9	
11	11	1	3	5	8	
12	12	1	3	7	9	
13	13	1	3	6	6	
14	14	1	3	4	10	
15	15	1	3	5	4	
16	16	2	1	8	10	
17	17	2	1	9	6	
18	18	2	1	7	8	
19	19	2	1	7	4	
20	20	2	1	6	9	
21	21	2	2	4	6	
22	22	2	2	5	8	
23	23	2	2	5	10	
24	24	2	2	6	4	
25	25	2	2	4	9	

Hypotheses:

Promotion ***(Main Effect)***

H_0: $M_L = M_M = M_H$

Coupon ***(Main Effect)***

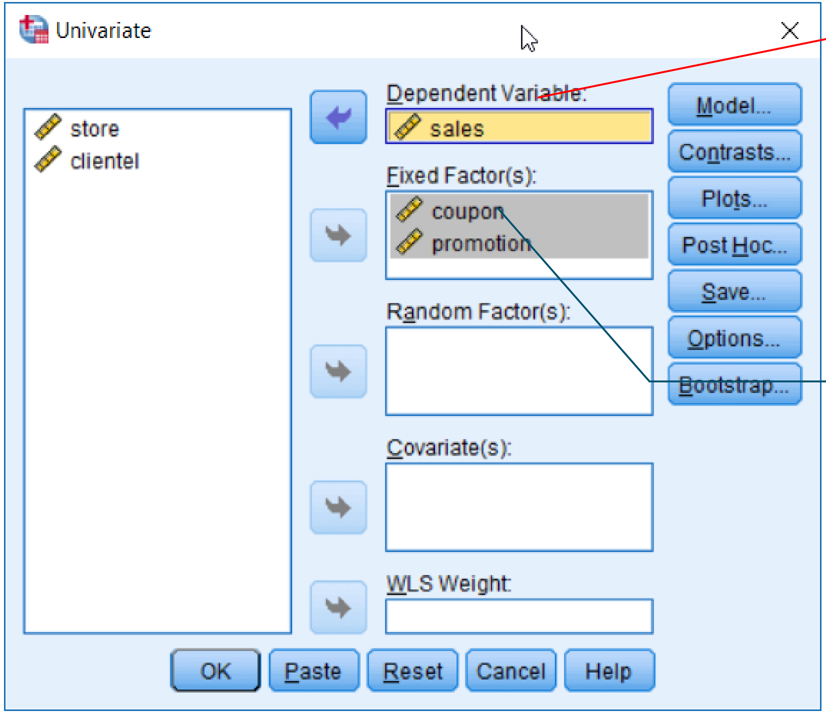
H_0: $M_{No} = M_{yes}$

Promotion * Coupon (***Interaction Effect*****)**

H_0: Promotion * Coupon=0

Analysis of Variance (ANOVA): Twoway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)



DV: Sales

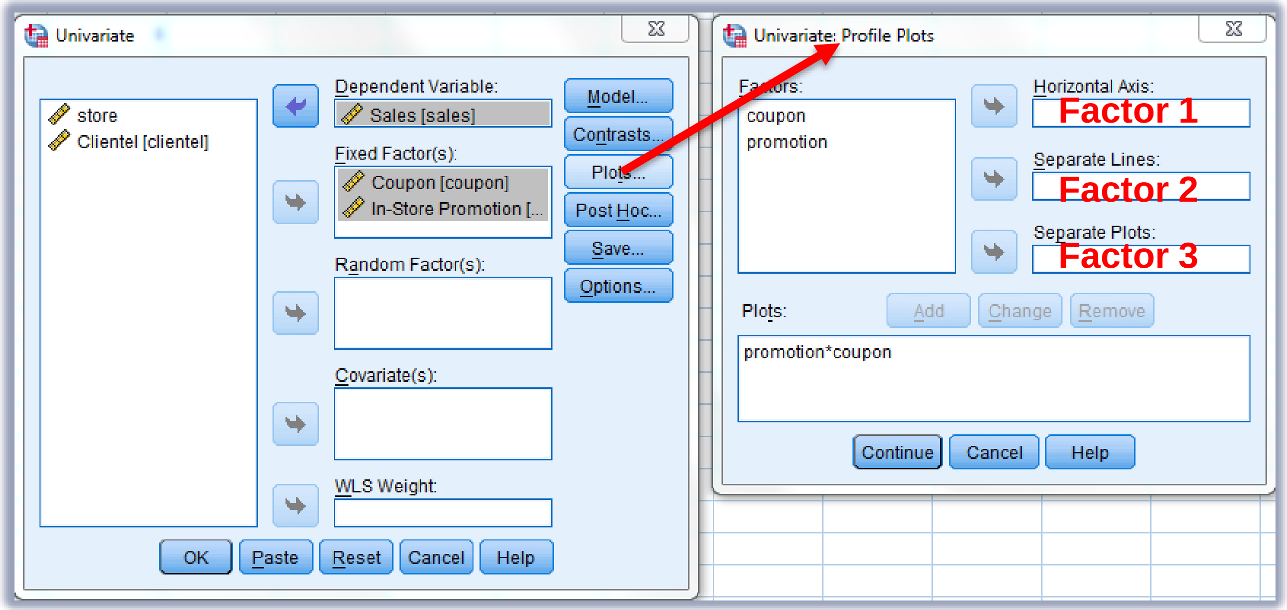
IVs: Coupon + Promotion

Maastricht University

Analysis of Variance (ANOVA): Twoway ANOVA

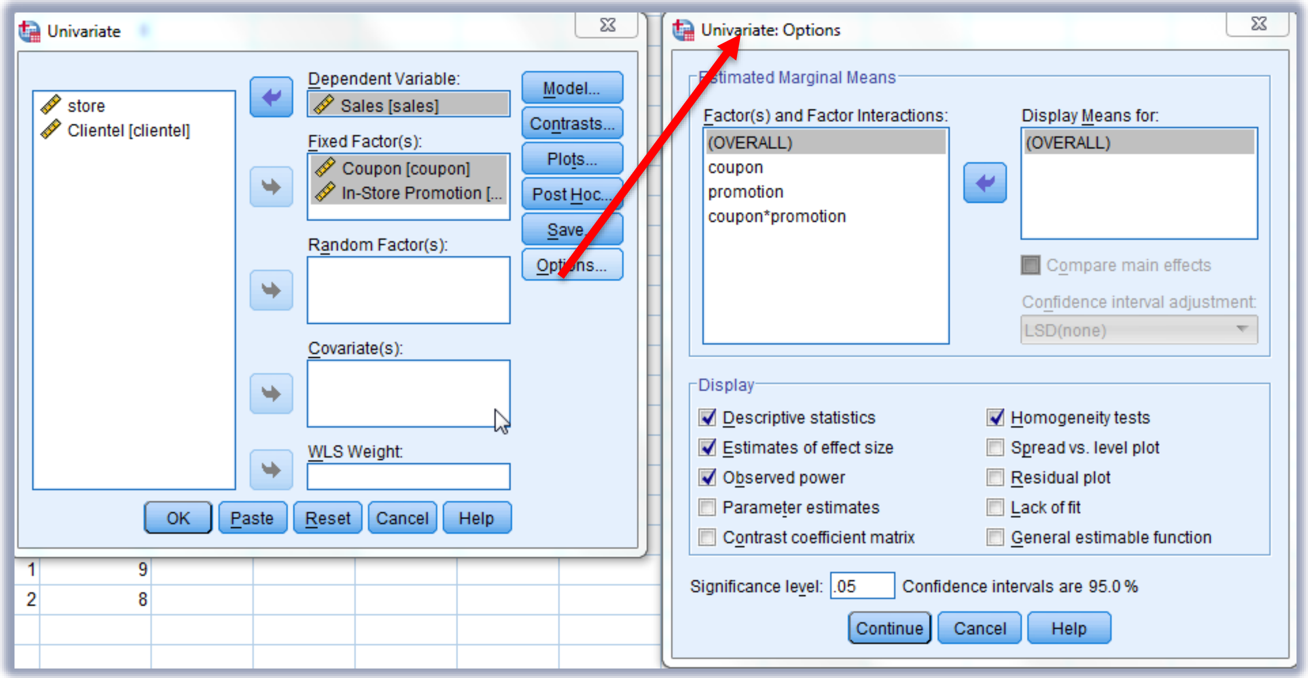
Department Store Data (Malhotra, 2010, Table 16.2)

Three-way Interaction Plot



Analysis of Variance (ANOVA): Twoway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)



Analysis of Variance (ANOVA): Twoway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)

Descriptives and Assumptions

Descriptive Statistics

Dependent Variable: sales Sales

promotion In-Store	coupon Coupon	Mean (M)	Std. Deviation (SD)	N (N)
1 High	1 Yes	9.20	.837	5
	2 No	7.40	1.140	5
	Total	8.30	1.337	10
2 Medium	1 Yes	7.60	1.140	5
	2 No	4.80	.837	5
	Total	6.20	1.751	10
3 Low	1 Yes	5.40	1.140	5
	2 No	2.00	.707	5
	Total	3.70	2.003	10
Total	1 Yes	7.40	1.882	15
	2 No	4.73	2.434	15
	Total	6.07	2.532	30

Levene's Test of Equality of Error Variances[a]

Dependent Variable: sales Sales

F	df1	df2	Sig.
.689	5	24	.637

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+promotion+coupon+promotion * coupon

Homogeneity of Variance Assumption

H_0: all variances are equal

Analysis of Variance (ANOVA): Twoway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)

ANOVA Table

Tests of Between-Subjects Effects

Dependent Variable: sales Sales

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	162.667[a]	5	32.533	33.655	.000	.875
Intercept	~~1104.133~~	1	1104.133	1142.207	.000	.979
promotion	106.067	2	53.033	54.862	.000	.821
coupon	53.333	1	53.333	55.172	.000	.697
promotion * coupon	3.267	2	1.633	1.690	.206	.123
Error	23.200	24	.967			
Total	~~1200.000~~	30				
Corrected Total	185.867	29				

a. R Squared = .875 (Adjusted R Squared = .849)

Total explained by model Promotion, Coupon, Promotion*Coupon

p value

$\eta^2 = 0.571$
$\eta^2 = 0.287$
$\eta^2 = 0.018$

(partial) η^2

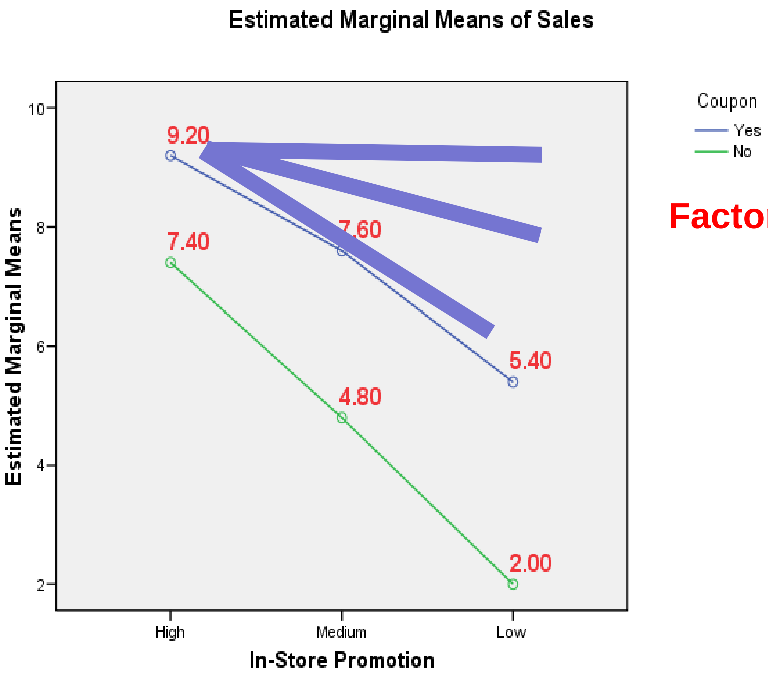
ERROR Not explained by model

Total including ERROR

Analysis of Variance (ANOVA): Twoway ANOVA

Department Store Data (Malhotra, 2010, Table 16.2)

Factor 3?



Factor 2

DV

One-way ANOVA

Editor

```
# One Way Anova (Completely Randomized Design)

attach(Table16.2)
# detach() # To detach data set ...

FIT.01<- aov(sales ~ factor(promotion))

library(lsr)
etaSquared(FIT.01, type=3, anova=TRUE)

summary(FIT.01)

# Type III === SPSS
# To get SPSS equivalent SPSS results

options(contrasts=c("contr.sum", "contr.poly"))

drop1(FIT.01,~.,test="F")

# Install.packages("car") is R Base

library(car)

Anova(FIT.01, type=3)

pairwise.t.test(sales, factor(promotion), p.adj = "bonf")

TukeyHSD(FIT.01)
```

One-way ANOVA

Console

```
> summary(FIT.01)
                  Df Sum Sq Mean Sq F value   Pr(>F)
factor(promotion)  2  106.1      53    17.9 0.000011 ***
Residuals         27   79.8       3
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # Type III === SPSS
> # To get SPSS equivalent SPSS results
>
> options(contrasts=c("contr.sum", "contr.poly"))
>
> drop1(FIT.01,~.,test="F")
Single term deletions

Model:
sales ~ factor(promotion)
                  Df Sum of Sq   RSS  AIC F value   Pr(>F)
<none>                          79.8 35.3
factor(promotion)  2       106 185.9 56.7    17.9 0.000011 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

One-way ANOVA

```
> pairwise.t.test(sales, factor(promotion), p.adj = "bonf")

 Pairwise comparisons using t tests with pooled SD

data:  sales and factor(promotion)

  1        2
2 0.033    -
3 0.000007 0.009

P value adjustment method: bonferroni
>
> TukeyHSD(FIT.01)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = sales ~ factor(promotion))

$`factor(promotion)`
    diff    lwr     upr   p adj
2-1 -2.1 -4.006 -0.1937 0.0286
3-1 -4.6 -6.506 -2.6937 0.0000
3-2 -2.5 -4.406 -0.5937 0.0084
```

Package **multcomp** for more options

One-way ANOVA

Editor

```r

# Regression Approach

FIT.LM.01<-lm(sales ~ factor(promotion))
summary(FIT.LM.01)
anova(FIT.LM.01)

library(car)
Anova(FIT.LM.01, type=3)
```

One-way ANOVA

Console

```
> # Regression Approach
>
> FIT.LM.01<-lm(sales ~ factor(promotion))
> summary(FIT.LM.01)

Call:
lm(formula = sales ~ factor(promotion))

Residuals:
   Min     1Q Median     3Q    Max
  -2.7   -1.3   -0.2    1.3    3.3

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          8.3000     0.5437  15.267 8.39e-15 ***
factor(promotion)2  -2.1000     0.7688  -2.731    0.011 *
factor(promotion)3  -4.6000     0.7688  -5.983 2.21e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.719 on 27 degrees of freedom
Multiple R-squared:  0.5707, Adjusted R-squared:  0.5389
F-statistic: 17.94 on 2 and 27 DF,  p-value: 1.104e-05
```

One-way ANOVA

Console

```
> anova(FIT.LM.01)
Analysis of Variance Table

Response: sales
                  Df Sum Sq Mean Sq F value    Pr(>F)    
factor(promotion)  2 106.07  53.033  17.944 1.104e-05 ***
Residuals         27  79.80   2.956                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> 
```

One-way ANOVA

Console

```
> library(car)
> Anova(FIT.LM.01, type=3)
Anova Table (Type III tests)

Response: sales
                  Sum Sq Df F value    Pr(>F)    
(Intercept)       688.90  1 233.087 8.388e-15 ***
factor(promotion) 106.07  2  17.944 1.104e-05 ***
Residuals          79.80 27                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> 
```

One-way ANOVA

Editor

```
# Descriptives and Diagnostics

print(model.tables(FIT.01,"means"),digits=3)

library(psych)
describeBy(sales, factor(promotion))

library(lawstat)
levene.test(sales,factor(promotion), location="mean")

library(car)
leveneTest(sales,factor(promotion), center="mean")

# bartlett.test(sales ~ factor(promotion))

```

One-way ANOVA

Console

```
> print(model.tables(FIT.01,"means"),digits=3)
Tables of means
Grand mean

6.066667

 factor(promotion)
factor(promotion)
  1   2   3
8.3 6.2 3.7
>
> library(psych)
> describeBy(sales, factor(promotion))
group: 1
  vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
1    1 10  8.3 1.34    8.5    8.38 1.48   6  10     4 -0.24     -1.4 0.42
-----------------------------------------------------------------
group: 2
  vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
1    1 10  6.2 1.75      6    6.12 2.22   4   9     5 0.17    -1.58 0.55
-----------------------------------------------------------------
group: 3
  vars  n mean sd median trimmed  mad min max range skew kurtosis   se
1    1 10  3.7  2    3.5    3.62 2.22   1   7     6 0.22    -1.57 0.63
>
```

One-way ANOVA

Console

```
> library(lawstat)
> levene.test(sales,factor(promotion), location="mean")

 classical Levene's test based on the absolute deviations from the mean (
 none not applied because the location is not set to median )

data:  sales
Test Statistic = 1.3532, p-value = 0.2754

>
> library(car)
> leveneTest(sales,factor(promotion), center="mean")
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group  2  1.3532 0.2754
      27
```

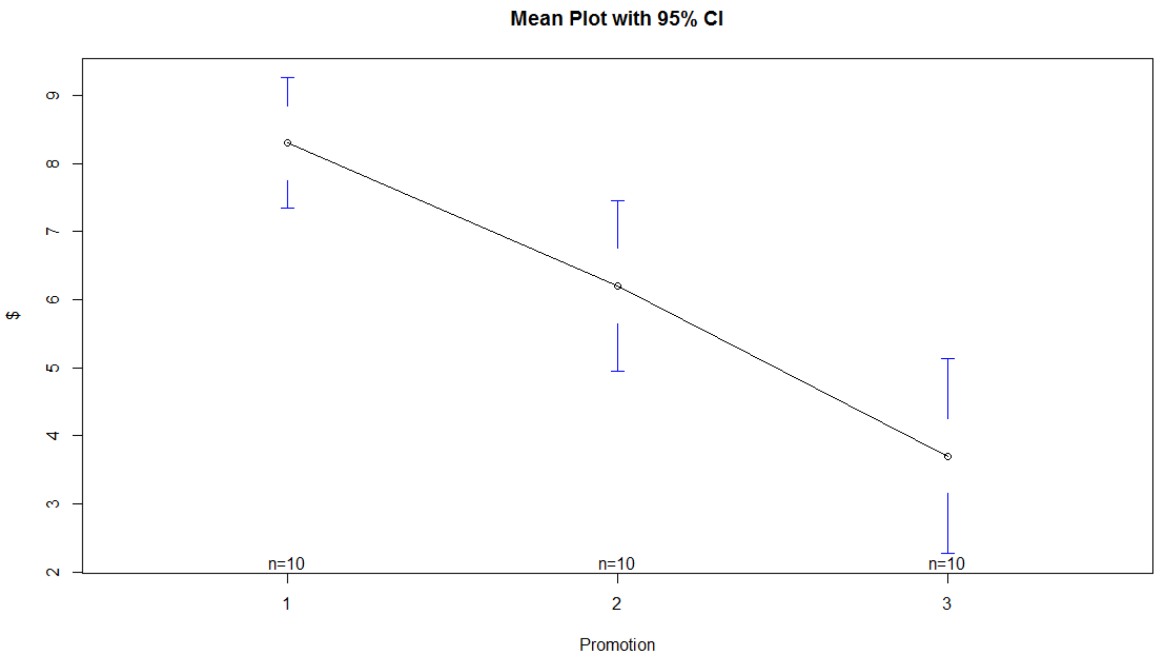
One-way ANOVA

Editor

```r

# Plots

library(gplots)

plotmeans(sales ~ factor(promotion), xlab="Promotion",
          ylab="$", main="Mean Plot with 95% CI")


```

One-way ANOVA

Console



Welch Correction

```
> # Welch Correction
>
> oneway.test(sales ~ factor(promotion))

 One-way analysis of means (not assuming equal variances)

data:  sales and factor(promotion)
F = 18.0902, num df = 2.00, denom df = 17.47, p-value = 5.541e-05
```

Kruskal-Wallis Test

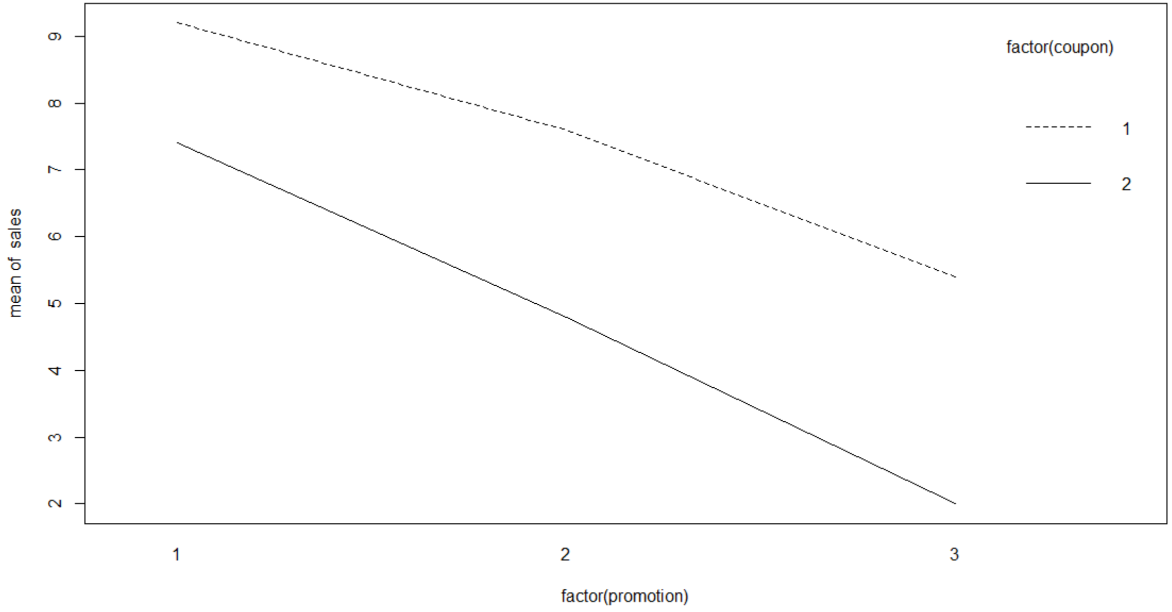
```
> # Kruskal-Wallis Test
>
> kruskal.test(sales ~ factor(promotion))

 Kruskal-Wallis rank sum test

data:  sales by factor(promotion)
Kruskal-Wallis chi-squared = 16.5292, df = 2, p-value = 0.0002575

>
> pairwise.wilcox.test(sales,factor(promotion), p.adj="bonf")

 Pairwise comparisons using Wilcoxon rank sum test

data:  sales and factor(promotion)

  1      2
2 0.0431 -
3 0.0011 0.0541

P value adjustment method: bonferroni
```

Package coin for more options

Two-way ANOVA

```
> # Two-way ANOVA
>
> FIT.02<- aov(sales ~ factor(promotion)*factor(coupon))
>
> summary(FIT.02)
                                Df Sum Sq Mean Sq F value   Pr(>F)    
factor(promotion)                2 106.07   53.03   54.86 1.12e-09 ***
factor(coupon)                   1  53.33   53.33   55.17 1.14e-07 ***
factor(promotion):factor(coupon) 2   3.27    1.63    1.69    0.206    
Residuals                       24  23.20    0.97                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> print(model.tables(FIT.02,c("means"),digits=3))
Tables of means
Grand mean

6.066667 

 factor(promotion) 
factor(promotion)
  1   2   3 
8.3 6.2 3.7 

 factor(coupon) 
factor(coupon)
    1     2 
7.400 4.733 

 factor(promotion):factor(coupon) 
                 factor(coupon)
factor(promotion) 1   2  
                1 9.2 7.4
                2 7.6 4.8
                3 5.4 2.0
```

Two-way ANOVA



ANCOVA

```
> FIT.03<- aov(sales ~ clientel + factor(promotion)*factor(coupon))
> summary(FIT.03)
                                  Df Sum Sq Mean Sq F value       Pr(>F)    
clientel                           1    0.8     0.8    0.86         0.36    
factor(promotion)                  2  106.1    53.0   54.55 0.000000019 ***
factor(coupon)                     1   53.3    53.3   54.85 0.0000001568 ***
factor(promotion):factor(coupon)   2    3.3     1.6    1.68         0.21    
Residuals                         23   22.4     1.0                         
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> library(car)
> Anova(FIT.03, type=3)
Anova Table (Type III tests)

Response: sales
                                  Sum Sq Df F value       Pr(>F)    
(Intercept)                        103.3  1  106.29 0.0000000043 ***
clientel                             0.8  1    0.86         0.36    
factor(promotion)                  106.1  2   54.55 0.0000000186 ***
factor(coupon)                      53.3  1   54.85 0.00000015684 ***
factor(promotion):factor(coupon)     3.3  2    1.68         0.21    
Residuals                           22.4 23                         
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

G*Power

http://www.gpower.hhu.de/

G*Power: Statistical Power Analyses for Windows and Mac

Kontrast

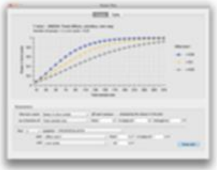
Suchbegriff

G*Power is a tool to compute statistical power analyses for many different *t* tests, *F* tests, $\chi 2$ tests, *z* tests and some exact tests. G*Power can also be used to compute effect sizes and to display graphically the results of power analyses.

Screenshots (click to enlarge)


Main Window


Main Window (Table)


Power Plot

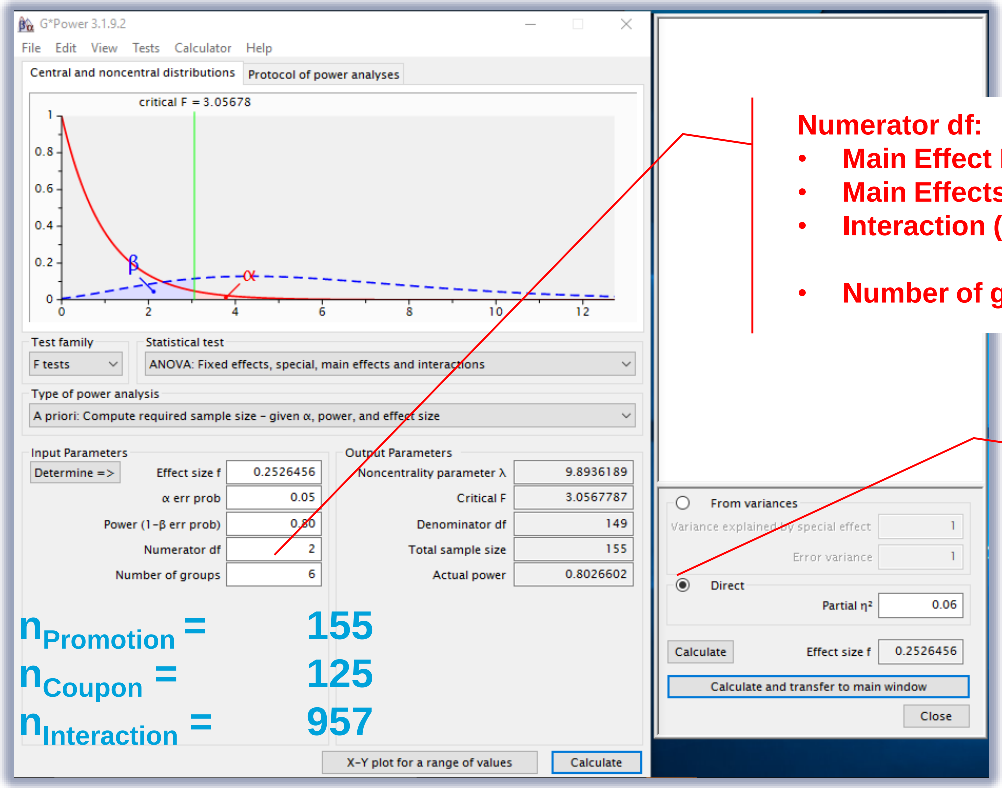

Power Plot (Table)

Register

Whenever we find a problem with G*Power we provide an update as quickly as we can. We will inform you about updates if you click here and add your e-mail address to our mailing list. We will only use your e-mail address to inform you about updates. We will not use your e-mail address for other purposes. We will not give your e-mail address to anyone else. You can withdraw your e-mail address from the mailing list at any time.

G*Power

http://www.gpower.hhu.de/



Numerator df:

- Main Effect Promotion (df=2 [3-1])
- Main Effects Coupon (df=1 [2-1])
- Interaction (df=2 [(2-1)*(3-1)])
- Number of groups: 6 (2*3)

f

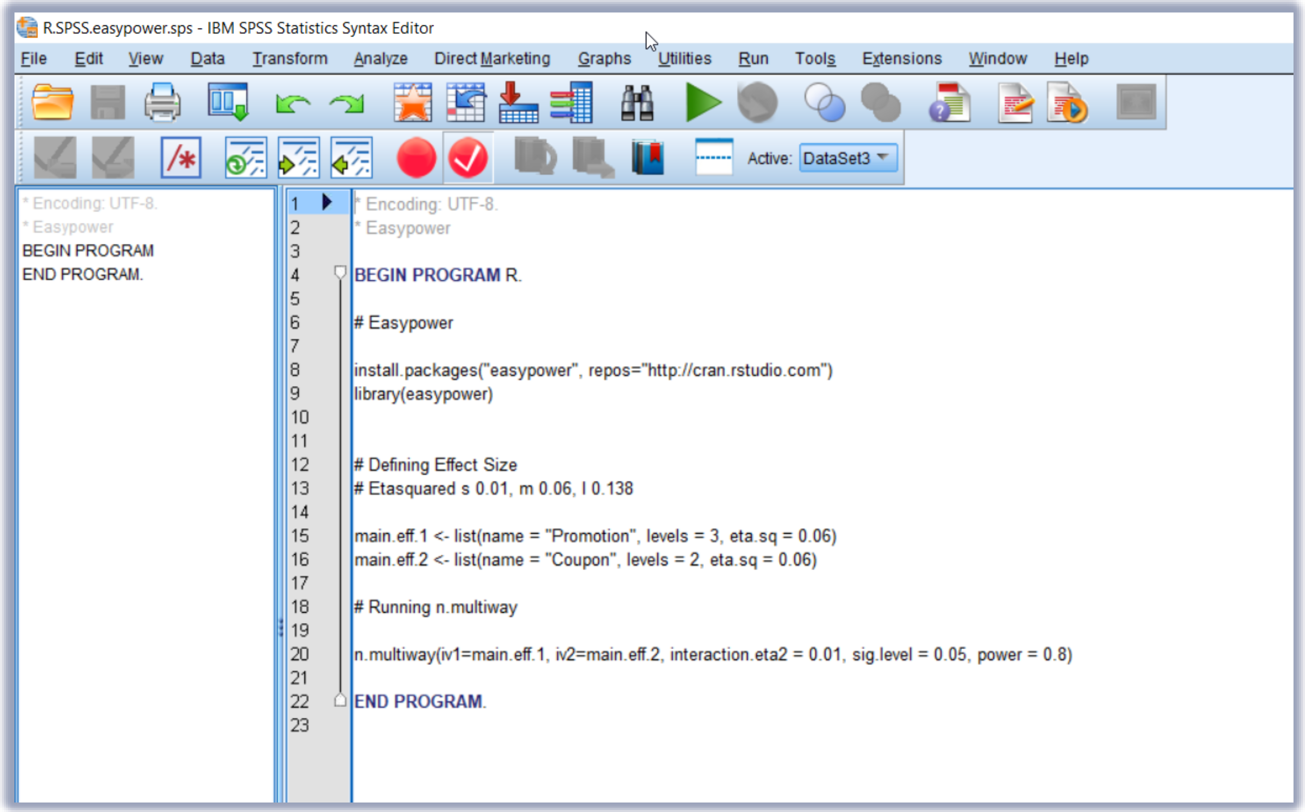
η^2=0.06

$n_{Promotion}$ = 155

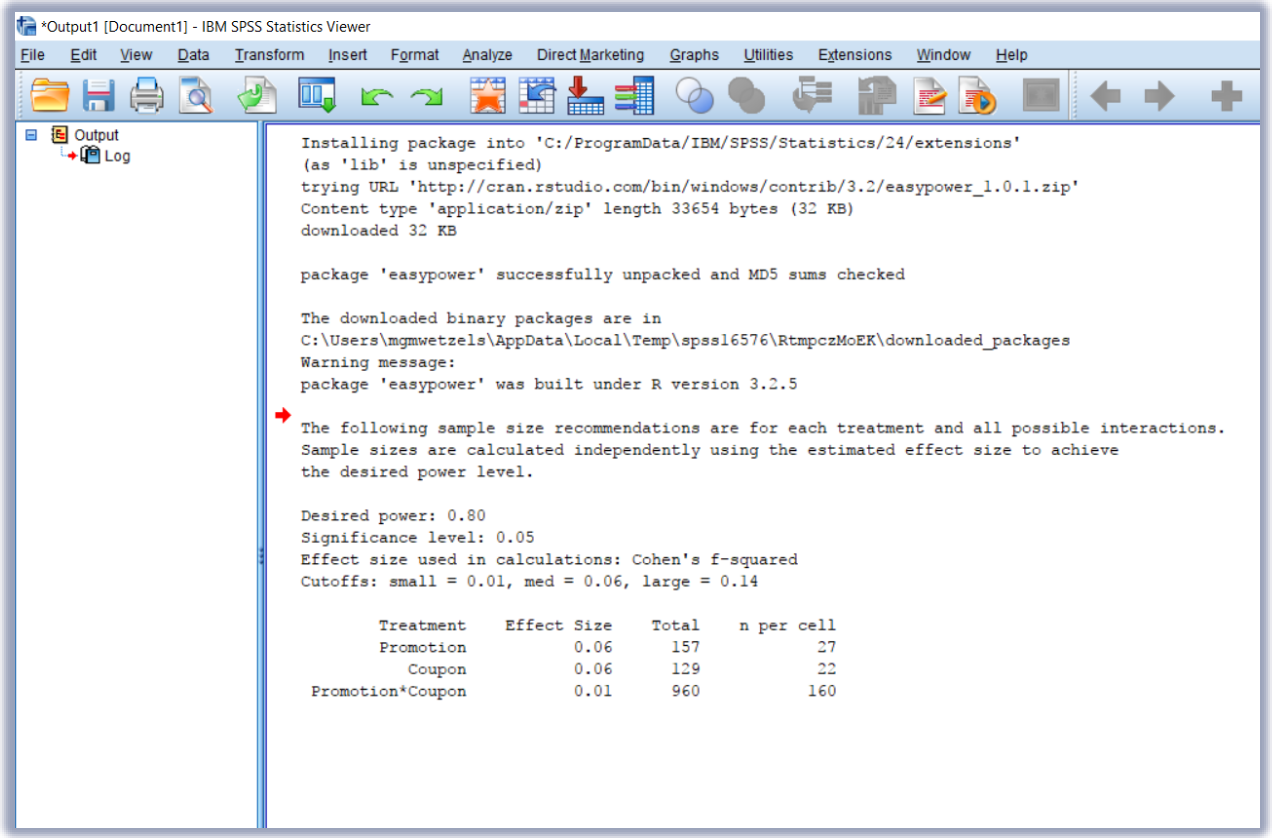
n_{Coupon} = 125

$n_{Interaction}$ = 957

SPSS and R


R.SPSS.easypower.sps - IBM SPSS Statistics Syntax Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Run Tools Extensions Window Help
Active: DataSet3
* Encoding: UTF-8.
* Easypower
BEGIN PROGRAM
END PROGRAM.
* Encoding: UTF-8.
* Easypower
BEGIN PROGRAM R.
# Easypower
install.packages("easypower", repos="http://cran.rstudio.com")
library(easypower)
# Defining Effect Size
# Etasquared s 0.01, m 0.06, l 0.138
main.eff.1 <- list(name = "Promotion", levels = 3, eta.sq = 0.06)
main.eff.2 <- list(name = "Coupon", levels = 2, eta.sq = 0.06)
# Running n.multiway
n.multiway(iv1=main.eff.1, iv2=main.eff.2, interaction.eta2 = 0.01, sig.level = 0.05, power = 0.8)
END PROGRAM.

SPSS and R


*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Extensions Window Help
Output
Log
Installing package into 'C:/ProgramData/IBM/SPSS/Statistics/24/extensions'
(as 'lib' is unspecified)
trying URL 'http://cran.rstudio.com/bin/windows/contrib/3.2/easypower_1.0.1.zip'
Content type 'application/zip' length 33654 bytes (32 KB)
downloaded 32 KB
package 'easypower' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
C:\Users\mgmwetzels\AppData\Local\Temp\spss16576\RtmpczMoEK\downloaded_packages
Warning message:
package 'easypower' was built under R version 3.2.5
The following sample size recommendations are for each treatment and all possible interactions.
Sample sizes are calculated independently using the estimated effect size to achieve
the desired power level.
Desired power: 0.80
Significance level: 0.05
Effect size used in calculations: Cohen's f-squared
Cutoffs: small = 0.01, med = 0.06, large = 0.14
Treatment Effect Size Total n per cell
Promotion 0.06 157 27
Coupon 0.06 129 22
Promotion*Coupon 0.01 960 160

References

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Cohen, J. (1992). “A Power Primer,” *Psychological Bulletin*, 112 (1), 155-159. [condensed version].

Everitt, B. S. (1992). *The Analysis of Contingency Tables*. CRC Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). “G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences,” *Behavior Research Methods, 39* (2), 175-191.

Hair, J.F., Jr., Black, W.C., Babin, B.J, and Anderson, R.E. (2018). *Multivariate Data Analysis.*Cengage.

Howell, D. C. (2012). *Statistical Methods for Psychology*. Cengage Learning.

Malhotra, N. (2010). *Marketing Research: An Applied Orientation*. Upper Saddle River: Pearson/Prentice-Hall.

References

Mundry, R. and Fischer, J. (1998). Use of Statistical Programs for Nonparametric Tests of Small Samples often Leads to Incorrect P Values: Examples from Animal Behaviour. *Animal Behaviour*, 56, 256-259.

Pallant, J. (2016). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows*. Maidenhead: Open University Press/McGraw-Hill.

Siegel, S. and Castellan, N.J. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

References

Chapman, C., & Feit, E. M. (2015). *R for Marketing Research and Analytics.* New York, NY: Springer.

Crawley, MJ (2013). *The R Book.* Chichester, UK: John Wiley and Sons.

Everitt, BS and Hothorn, T (2006). *A Handbook of Statistical Analysis Using R.* Boca Raton, FL: Chapman and Hall/CRC.

Faraday, J.A. (2015). *Linear Models with R*. Boca Raton, FL: CRC Press.

Field, A, Miles, J and Field, Z (2012). *Discovering Statistics Using R.* Los Angeles, CA: Sage Publications.

Kabacoff, RI (2011). *R in Action*. Shelter Island, NY: Manning.

Muenchen, RA (2009). *R for SAS and SPSS Users.* New York, NY; Springer Science and Business Media.

Thank you for Your Attention!