# Analysis of Variance

**Objective** of Analysis of variance (ANOVA) is to determine whether the differences on criterion or dependent variable among groups are "reliable" or statistically significant. Variable(s) that create group membership are called control or independent variable(s).

- **Application**

  There are two broad classes of situation where ANOVA may be used.

  - Experiment: In this situation, control or independent variable(s) *within* and / or *between* experimental unit is changed and a measure of difference(s) on criterion variable is used to demonstrate the cause (experimental treatment) and the effect. For example, higher number of brand advertising exposures increases the likelihood of preference towards brand.
  - Groups with "natural" belonging: In this situation, groups with natural belongings is used to demonstrate reliable differences exists between groups. For example, reliable difference between gender on attitude towards consumer's willingness to search for bargains.

- **Data Requirement**

  - Criterion variable must have interval or ratio scale measurement properties.
  - Control variable(s) may be either nominal or ordinal. If control variable is interval scale and it is desired to understand the impact of each level of control variable on the criterion variable then, control variable must be converted to ordinal scale.

  In following notes, we will present situation involving

  - Two Group Comparison,
  - Matched Comparison,
  - Multiple Group Comparison, and
  - Multiple Factorial Comparison.

The main idea behind ANOVA is to compare two measures of variability. One measure of variability is a result of differences among dependent variable values *within* each group and the second measure of variability is a result of differences *between* group means.

## Two Group Comparison

Two groups, one treatment and another control group, are compared on one dependent variable. The Test units (individuals, stores etc.) assigned randomly to two groups.

- **Experimental Situation**

| | | | | |
|---|---|---|---|---|
| Treatment Group | $R$ | $X$ | $O_1$ | $n_1$ test units. |
| Control Group | $R$ | | $O_2$ | $n_2$ test units. |

Consider an experiment where $R$ is used to denote random assignment of treatment and $X$ is used to denote treatment application to test units. Moreover, we take measurement (denoted by $O_1$ and $O_2$) on $n_1$ and $n_2$ test units for treatment and control condition respectively. That is, we obtain $y_{11}, y_{12}, \cdots y_{1n_1}$ and $y_{21}, y_{22}, \cdots y_{2n_2}$ dependent variable values for treatment and control groups.

| | **City** | |
|---|---|---|
| Store | Treatment | Control |
| 1 | 18 | 12 |
| 2 | 15 | 9 |
| 3 | 14 | 8 |
| 4 | 13 | 7 |
| 5 | 10 | |
| Total | 70 | 36 |
| Mean | 14 | 9 |

An experiment is conducted in two cities. Under treatment condition, advertisement is placed in a local newspaper for one day with a coupon and a sample of five stores are monitored for sales. In a control city, there is same advertisement without coupon and four stores are monitored for sales. The dependent variable is the number of cases sold. Table contains collected data for nine stores from two cities.

It is of interest to know whether advertisement with coupon resulted in systematically different sales. Moreover, we need to determine whether observed differences are large enough to have occurred by chance. First question is somewhat easy to answer. We have higher sales for all stores (with the exception of store 5) where we placed advertisement with coupon than the control condition. In general, we would demonstrate that on an average treatment had higher (lower) mean than the control condition. To answer our second question, we need to know something about variability in observed values and then compare that variability to some standard, for example normal distribution of variability. Suppose we wrote observations as follows:

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \end{pmatrix} \equiv \begin{pmatrix} 18 \\ 15 \\ 14 \\ 13 \\ 10 \\ \\ 12 \\ 9 \\ 8 \\ 7 \end{pmatrix} = \begin{pmatrix} 14 \\ 14 \\ 14 \\ 14 \\ 14 \\ \\ 9 \\ 9 \\ 9 \\ 9 \end{pmatrix} + \begin{pmatrix} 4 \\ 1 \\ 0 \\ -1 \\ -4 \\ \\ 3 \\ 0 \\ -1 \\ -2 \end{pmatrix}$$
$$\text{Response} = \text{Treatment} + \text{Error}$$

One purpose of this sort of decomposition is to better understand source of variation and attribute that source of variation to that "factor". In order to better understand various components of variation, we could square each entry. Following squaring, we could add rows that correspond to a particular treatment. These summed entries will be called sums of squares. Instead of nine entries in our matrices, in sums of squares matrices, we will have two entries. That is,

$$\begin{pmatrix} 1014 \\ 338 \end{pmatrix} = \begin{pmatrix} 980 \\ 324 \end{pmatrix} + \begin{pmatrix} 34 \\ 14 \end{pmatrix}$$

Note that entry 1014 is based on sums of squares for the treatment group of stores. That is, $18^2 + 15^2 + 14^2 + 13^2 + 10^2 = 1014$. Note also that variation due to treatment can be further decomposed as variation due to overall mean as well as variation around overall mean. Such decomposition for responses would be

$$\begin{pmatrix} 18 \\ 15 \\ 14 \\ 13 \\ 10 \\ \\ 12 \\ 9 \\ 8 \\ 7 \\ \text{Response} \end{pmatrix} = \begin{pmatrix} 11\frac{7}{9} \\ 11\frac{7}{9} \\ 11\frac{7}{9} \\ 11\frac{7}{9} \\ 11\frac{7}{9} \\ \\ 11\frac{7}{9} \\ 11\frac{7}{9} \\ 11\frac{7}{9} \\ 11\frac{7}{9} \\ \text{Overall} \end{pmatrix} + \begin{pmatrix} 2\frac{2}{9} \\ 2\frac{2}{9} \\ 2\frac{2}{9} \\ 2\frac{2}{9} \\ 2\frac{2}{9} \\ \\ -2\frac{7}{9} \\ -2\frac{7}{9} \\ -2\frac{7}{9} \\ -2\frac{7}{9} \\ \text{Treatment} \end{pmatrix} + \begin{pmatrix} 4 \\ 1 \\ 0 \\ -1 \\ -4 \\ \\ 3 \\ 0 \\ -1 \\ -2 \\ \text{Error} \end{pmatrix}$$

Based on this decomposition, we can also write various sums of squares. For the sample of nine observations, we may write following equality

$$\begin{pmatrix} 1352 \\ \text{Response} \\ \text{sum of squares} \end{pmatrix} = \begin{pmatrix} 1248.44 \\ \text{Overall} \\ \text{sum of squares} \end{pmatrix} + \begin{pmatrix} 55.56 \\ \text{Treatment} \\ \text{sum of squares} \end{pmatrix} + \begin{pmatrix} 48 \\ \text{Error} \\ \text{sum of squares} \end{pmatrix}$$

Note that overall sum of squares number look complicated but is equal to $\left(11\frac{7}{9}\right)^2 \times 9$. Moreover, we can construct a test based on above set of numbers. Suppose we computed two mean sum squares (one due to treatment while other one due to error) and took ratio of these two mean of squares. If the mean sum of squares due to treatment was large in comparison to the mean error sum of squares, then we should conclude that variation due to treatment is substantial compared to error variation and vice versa. Let us look at above ideas more formally.

- **Hypothesis and Test Statistic**

Suppose $\bar{y}_1$, and $\bar{y}_2$ denote the average for treatment and control stores respectively. Then, the null and alternative hypothesis can be stated as: $H_0 : \bar{y}_1 = \bar{y}_2$ and $H_A : \bar{y}_1 \neq \bar{y}_2$.
In this situation, there are two alternative ways to test this particular hypothesis. First relies on t-statistic while other one relies on F-statistic. Both approaches will give you identical results. Let us look at both of these approaches with the simple example. Suppose $\bar{y}_1 = 14$ is the average sales for treatment stores and $\bar{y}_2 = 9$ is the average for control stores. Then, t-statistic is

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})}},$$

where $\bar{y}_1 - \bar{y}_2$ is variation due to treatment and $s$ is square root of the pooled sums of squares, or *between group variation*. That is,

$$\begin{aligned} s^2 &= \frac{ss_1 + ss_2}{n_1 + n_2 - 2} \\ &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \end{aligned}$$

and $s_1^2 = \frac{1}{n_1-1}\sum_{j=1}^{n_1}(y_{1j} - \bar{y}_1)^2$, $s_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(y_{2j} - \bar{y}_2)^2$. Note that we have already computed $ss_1$ and $ss_2$ to be 34 and 14 respectively. Consequently, $s$ is equal to $\sqrt{\frac{48}{7}} = \sqrt{6.857} = 2.6186$. This would result in

$$\begin{aligned} t &= \frac{14 - 9}{2.6186 \times \sqrt{\frac{1}{5} + \frac{1}{4}}} \\ &= \frac{5}{2.6186 \times 0.6708} = 2.85 \end{aligned}$$

An alternative test based on ratio of two mean sums of squares, leads to F-statistic. It is given by

$$
\begin{aligned}
F &= \frac{\text{Mean Treatment Sum of Squares}}{\text{Mean Error Sum of Squares}} \\
&= \frac{55.555/1}{48/7} = 8.10
\end{aligned}
$$

As we will see, t-statistic is a special of F-statistic and conclusion reached by either tests are identical[1].

- **Decision Rule**

If $|t| > t_{\alpha/2, n_1+n_2-2}$, then reject the null hypothesis. That is, look at tabled $t$ value with $n_1 + n_2 - 2$ degrees of freedom with probability level of $\alpha/2$. We take half of probability level because we do not have prior knowledge about direction of difference. Suppose we decided on critical probability ($\alpha = 0.05$), then we would reject the null hypothesis at prob. $\leq 0.05$, if t-statistic is less than[2] 2.365. We would thus conclude that advertisement with coupons produced higher sales than one without and difference is statistically significant 19 times out of 20.

- **Assumptions**

1. Each test unit is independent from others.
2. Each observation within in a group has equal influence on the estimated group means.
3. Variations within group is a similar.
4. variances of both groups are equal.
5. $y_{11}, y_{12}, \cdots y_{1n_1}$ are normally distributed, and also $y_{21}, y_{22}, \cdots y_{2n_1}$ are normally distributed. Alternatively, error values across treatments are normally distributed.

- **Comments about Assumptions and Study Design Issues**

Assumption about the independence is sometimes easy to test, especially if test units can be grouped in particular categories. For example, stores located in particular city or order in which subjects were recruited. Assumption about normality is testable. It appears that $t$-test is robust to most distributions except when distribution is "flat" or multimodal. Assumption

[1] To compute $F$-statistic, we needed sum of squares, a form of difference operation, followed by squaring. On the other hand, for t-statistic, we took difference between two means. Note that F-statistic is squared quantity of $t$-statistic.

[2] This is based on *Table of critical values of t* at 7 degrees of freedom at prob. of 0.025

about the equality of variances can also be tested. It appears that as long as the ratio of variances is less than 10 and sample size is large (more than 25 per cell), $t$-test is robust.
If the whole purpose of research is to demonstrate there are differences in two groups, there are two alternative approaches. First, we could select treatments that are substantially different. This would result in a larger difference in mean estimates. We also could reduce within group variation by choosing homogeneous groups.

- **Using SAS to Conduct above analysis**

Above analysis also can be done using SAS. Following inputs were used to conduct such an analysis.

```
options ls=70 ps = 80 nocenter nodate;
data stores;
input Cond sales;
datalines;
 1 18
 1 15
 1 14
 1 13
 1 10
 2 12
 2  9
 2  8
 2  7
;;;;
proc glm ;
class cond;
model sales = cond ;
means cond;
run;
```

- **SAS produced following Output:**

```
The SAS System

General Linear Models Procedure
Class Level Information

Class    Levels    Values
COND          2    1 2

Number of observations in data set = 9

Dependent Variable: SALES

Source                DF    Sum of Squares     F Value      Pr > F
Model                  1       55.55555556        8.10      0.0248
Error                  7       48.00000000
Corrected Total        8      103.55555556

                R-Square              C.V.             SALES Mean
                0.536481          22.23352             11.7777778
```

```
Source                 DF        Type I SS     F Value      Pr > F
COND                    1      55.55555556        8.10      0.0248

General Linear Models Procedure

Level of       ------------SALES------------
COND       N       Mean              SD

1          5     14.0000000       2.91547595
2          4      9.0000000       2.16024690
```

- **Note on Using Regression Analysis to Conduct ANOVA**

As one may have expected, ANOVA is special case of regression analysis. That is, conclusions about differences and associated statistical tests would often produce comparable results. As shown in following Table, regression analysis with different coding scheme would produce different estimate for intercept or constant term and slope but $R^2$ and absolute value of t-statistics for slope are identical with all coding schemes.

| Code Used | Constant | | Slope | | |
|---|---|---|---|---|---|
| for Cond. 2 | Estimate | t-stat. | Estimate | t-stat. | $R^2$ |
| 2 | 19.00 | 7.08 | −5.00 | −2.85 | 0.537 |
| 0 | 9.00 | 6.87 | 5.00 | 2.85 | 0.537 |
| -1 | 11.50 | 13.09 | 2.50 | 2.85 | 0.537 |

Note that the first two rows have slope of same magnitude but opposite in direction because codes are opposite. Moreover, slope is equal to difference across two groups of stores. In the last instance, slope is equal to the average of difference.

- **A Comparison of Personal Computer Prices (Example of ANCOVA)**

There are two types of personal computers: Wintel[3] and Apple Macintosh (Mac). It is common perception among computer buyers and users that Mac systems are more expensive than Wintel systems. We are going to use ANOVA to test whether such perception is true and examine magnitude of price differences. To examine this question, I gathered data for 36 computer systems from *Business Week* web site `http://www.maven.businessweek.com/Maven` around January 1998. I did not include in this analysis seven computer systems that were priced below $1,000. I also excluded one Macintosh system that was priced above $4,000. Since price of computer system is related to CPU speed, I have used CPU as covariate.

- **SAS input for Price Comparison**

[3] These are computers with Windows based operating system and Intel like central processing unit (CPU).

```
options nodate nocenter ps = 80 ls = 75;
data pc;
input price pcmac cpu Name $;
if price > 4000 then price = .;
datalines;
1899 1 200 Acer Aspire 2761
3224 1 300 Compulink CLR Infinity PT2-DAX
2249 1 350 Dell Dimension XPS R350
2599 1 400 Dell Dimension XPS R400
2098 1 300 Gateway G6/300
2398 1 333 Gateway G6/333
2598 1 333 Gateway G6/333 XL
1399 1 350 HP Pavillion 8370
1799 1 400 HP Pavillion 8380
1048 1 266 IBM Aptiva E2N
1399 1 233 Micro Express MicroFLEX K6-233
1399 1 200 Micro Express MicroFLEX-MMX/200
2199 1 300 Micron Millennia 300
1999 1 350 Micron Millennia 350
2099 1 400 Micron Millennia 400
1999 1 400 Pionex Model 640036
2498 1 300 Polywell Poly 7300L5
1788 1 333 Polywell Poly 7333 Lx
2750 1 400 Polywell Poly 8400S1
1149 1 266 Quantex QP6/266 M-1x
1199 1 266 Quantex QP6/266 SB-2x
1399 1 300 Quantex QP6/300 M-2x
1799 1 300 Quantex QP6/300 SM-3X
2499 1 400 Quantex QP6/400 SM-4x SE
1599 1 200 Toshiba Infinia 7202
1899 1 200 Xi K200 Mtower
2099 1 200 Xi MX200 Mtower
2299 2 250 Apple Power Mac 6500/250
2399 2 200 Apple Power Mac 7300/200
2999 2 200 Apple Power Mac 8600/200
2499 2 266 Apple Power Mac G3/266
4899 2 300 Apple Power Mac G3/300
1199 2 160 UMAX SuperMac C500/180
1799 2 200 UMAX SuperMac C600/200
1999 2 240 UMAX SuperMac C600/240
2499 2 180 UMAX SuperMac J700/180
;;;;
proc glm;
class pcmac;
model price = pcmac cpu / solution ;
means pcmac;
lsmeans pcmac;
run;
```

- **Output from SAS is as follows**

```
The SAS System
General Linear Models Procedure
Class Level Information

Class    Levels    Values
```

```
PCMAC         2    1 2

Number of observations in data set = 36


NOTE: Due to missing values, only 35 observations can be used in this
      analysis.

General Linear Models Procedure

Dependent Variable: PRICE
                                   Sum of           Mean
Source                  DF        Squares         Square  F Value    Pr > F

Model                    2     2039876.48     1019938.24     4.10 ①    0.0259
Error                   32     7956729.80      248647.81
Corrected Total         34     9996606.29

                  R-Square           C.V.      Root MSE         PRICE Mean
                  0.204057 ②      24.65929       498.646            2022.14


Source                  DF      Type I SS   Mean Square  F Value    Pr > F

PCMAC                    1 ③     371841.32     371841.32     1.50    0.2303
CPU                      1     1668035.16    1668035.16     6.71    0.0143

Source                  DF    Type III SS   Mean Square  F Value    Pr > F

PCMAC                    1 ④    1458109.46    1458109.46     5.86    0.0213
CPU                      1     1668035.16    1668035.16     6.71    0.0143


                                  T for H0:     Pr > |T|    Std Error of
Parameter           Estimate     Parameter=0                 Estimate

INTERCEPT       1471.296776 ⑤            4.38      0.0001     335.7893906
PCMAC     1     -575.993962            -2.42      0.0213     237.8565871
CPU                3.491525             2.59      0.0143       1.3480464

General Linear Models Procedure

Level of      -----------PRICE-----------    ------------CPU------------
PCMAC      N      Mean            SD            Mean            SD

1         27    1966.03704 ⑥    539.088082      306.666667     70.0340576
2          8    2211.50000     543.631966      212.000000     36.4574116

General Linear Models Procedure
Least Squares Means

PCMAC         PRICE
             LSMEAN

1        1890.48709 ⑦
2        2466.48106
```

- **Comments About SAS Output**

① We conclude from F-statistic that we should reject the null hypothesis that prices for Wintel and Apple Macintosh systems are identical at prob $\leq 0.05$. We will look at relative contribution of CPU speed and type of system on prices below.

② $R^2$ indicate that only 20.4% of variation in price is explained by type of system and CPU speed. It is expected that remaining 79.6% variation is explained by other factors not included in the model. These might include brand name, size of hard disk, type and speed of CDROM, and other optional hardware included with the system.

③ Note that our model consisting of two variables account for 2,039,878.5 sum of squares. Of this total sum of squares about 82% variation is accounted by variable `CPU` Speed and remaining 18% by type of system (Wintel or Mac).

④ There are four different approaches to sum of squares calculation when multiple independent variables are involved. We will look at details about them when we look at multiple factorial designs.

⑤ Based on estimates we would conclude following equation:

$$\text{System Price} = 1471.3 - 576 \times \text{PCMAC} + 3.5 \times \text{CPU Speed.}$$

Note that variable PCMAC is equal to 1, if it is Wintel system and zero otherwise. From above equation, I would conclude that Apple Macintosh computers cost about $576 more than comparable Wintel computers. Note that `Proc GLM` generally will not produce these estimates. You have to request them by stating `solution` on the `model` statement of SAS input.

⑥ These means are simple averages and note that without correcting for `CPU` speed, the average price of Wintel PC is $1,966 while that of Mac is $2,211.5. Moreover typical Mac machine has `CPU` speed of 200 megahertz (mhz) while those based on Wintel standard is about 300 mhz. This would mean that Mac computers are more expensive by about $250.

⑦ Keyword `lsmeans` is used to produce means by classification variable. To obtain price estimate of $1,890.5, SAS uses equation summarized above. Specifically, to obtain Wintel based PC Price, we would start with base of $1,471.3, substract from it $576 and add 3.5× 284 (or the average `CPU` speed all computers)[4]. On the other hand, the predicted price for Mac is $2,466.5.

[4]Note that various truncation of numbers result in estimated price to be slightly different from that produced by SAS software.

Based on above analysis, we would conclude that Wintel based computers are cheaper by about $250 when corrected for `CPU` speed than Mac. When `CPU` speed is matched for both systems, Wintel based computers are cheaper by about $576. This latter difference in price is statistically significant at prob. of 0.05 or less.

- **Evaluating ANOVA Assumptions**

We stated five assumptions for ANOVA. These are

1. Each test unit is independent from others.
2. Each observation within in a group has equal influence on the estimated group means.
3. Variations within group is a similar.
4. variances of both groups are equal.
5. $y_{11}, y_{12}, \cdots y_{1n_1}$ are normally distributed, and also $y_{21}, y_{22}, \cdots y_{2n_1}$ are normally distributed. Alternatively, error values across treatments are normally distributed.

Note that these assumptions are similar to assumptions stated for regression analysis and as such most of test procedure as well as mathematical details given there are applicable in present situation as well. SAS software provides facility to save variables by using `OUTPUT` statement. For example, I could modify my previous `PROC GLM` statement as follows:

```
proc glm;
class pcmac;
model price = pcmac cpu / solution ;
means pcmac;
lsmeans pcmac;
 output out=predpc
  predicted=prdprc
  residual=resprc student=stprc dffits=dfprc covratio=covprc;
run;

proc print data=predpc1;
var Name price prdprc resprc stprc dfprc covprc;
run;
```

Note that keyword `predicted` is used to save predicted values of dependent variable, and other keywords are same as those we summarized in regression analysis.

- **SAS Output as result of `PROC PRINT`**

| OBS | NAME | PRICE | PRDPRC | RESPRC | STPRC | DFPRC | COVPRC |
|---|---|---|---|---|---|---|---|
| 1 | Acer | 1899 | 1593.61 | 305.39 | 0.65294 | 0.23913 | 1.20089 |
| 2 | Compulin | 3224 | 1942.76 | 1281.24 | 2.61883 | 0.57289 | 0.55416 |
| 3 | Dell | 2249 | 2117.34 | 131.66 | 0.27101 | 0.06175 | 1.15079 |
| 4 | Dell | 2599 | 2291.91 | 307.09 | 0.64941 | 0.21531 | 1.17537 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 5 | Gateway | 2098 | 1942.76 | 155.24 | 0.31731 | 0.06162 | 1.13187 |
| 6 | Gateway | 2398 | 2057.98 | 340.02 | 0.69671 | 0.14487 | 1.09681 |
| 7 | Gateway | 2598 | 2057.98 | 540.02 | 1.10652 | 0.23283 | 1.02145 |
| 8 | HP | 1399 | 2117.34 | -718.34 | -1.47859 | -0.34866 | 0.93711 |
| 9 | HP | 1799 | 2291.91 | -492.91 | -1.04238 | -0.34930 | 1.10269 |
| 10 | IBM | 1048 | 1824.05 | -776.05 | -1.59601 | -0.37216 | 0.90192 |
| 11 | Micro | 1399 | 1708.83 | -309.83 | -0.64663 | -0.18465 | 1.14521 |
| 12 | Micro | 1399 | 1593.61 | -194.61 | -0.41608 | -0.15177 | 1.23001 |
| 13 | Micron | 2199 | 1942.76 | 256.24 | 0.52375 | 0.10200 | 1.11349 |
| 14 | Micron | 1999 | 2117.34 | -118.34 | -0.24358 | -0.05549 | 1.15232 |
| 15 | Micron | 2099 | 2291.91 | -192.91 | -0.40796 | -0.13472 | 1.20411 |
| 16 | Pionex | 1999 | 2291.91 | -292.91 | -0.61943 | -0.20525 | 1.17963 |
| 17 | Polywell | 2498 | 1942.76 | 555.24 | 1.13490 | 0.22463 | 1.01013 |
| 18 | Polywell | 1788 | 2057.98 | -269.98 | -0.55320 | -0.11470 | 1.11565 |
| 19 | Polywell | 2750 | 2291.91 | 458.09 | 0.96873 | 0.32385 | 1.11862 |
| 20 | Quantex | 1149 | 1824.05 | -675.05 | -1.38829 | -0.32037 | 0.96008 |
| 21 | Quantex | 1199 | 1824.05 | -625.05 | -1.28546 | -0.29530 | 0.98665 |
| 22 | Quantex | 1399 | 1942.76 | -543.76 | -1.11143 | -0.21980 | 1.01534 |
| 23 | Quantex | 1799 | 1942.76 | -143.76 | -0.29384 | -0.05705 | 1.13340 |
| 24 | Quantex | 2499 | 2291.91 | 207.09 | 0.43793 | 0.14467 | 1.20124 |
| 25 | Toshiba | 1599 | 1593.61 | 5.39 | 0.01153 | 0.00419 | 1.25018 |
| 26 | Xi | 1899 | 1593.61 | 305.39 | 0.65294 | 0.23913 | 1.20089 |
| 27 | Xi | 2099 | 1593.61 | 505.39 | 1.08054 | 0.40046 | 1.11828 |
| 28 | Apple | 2299 | 2344.18 | -45.18 | -0.09745 | -0.03799 | 1.27128 |
| 29 | Apple | 2399 | 2169.60 | 229.40 | 0.49210 | 0.18465 | 1.23022 |
| 30 | Apple | 2999 | 2169.60 | 829.40 | 1.77921 | 0.70063 | 0.92079 |
| 31 | Apple | 2499 | 2400.04 | 98.96 | 0.21479 | 0.08758 | 1.28288 |
| 32 | Apple | . | 2518.75 | . | . | . | . |
| 33 | UMAX | 1199 | 2029.94 | -830.94 | -1.80191 | -0.76976 | 0.93300 |
| 34 | UMAX | 1799 | 2169.60 | -370.60 | -0.79501 | -0.30015 | 1.18546 |
| 35 | UMAX | 1999 | 2309.26 | -310.26 | -0.66736 | -0.25652 | 1.21325 |
| 36 | UMAX | 2499 | 2099.77 | 399.23 | 0.85959 | 0.33451 | 1.18209 |

To test that each observation is independent from all others, we could follow to different approaches. First we could sort observations by values of dependent variable (as we did in regression analysis) and compute autocorrelation or Durbin-Watson's statistic. A significant autocorrelation will indicate that there is a departure from this assumption. An alternative is to compute the average of residuals for a known independent variable which is not included in the analysis. As a possibility, consider variable brand name for our example. If brand name and unexplained variability in dependent variable (`RESPRC`) is not related, then the average residuals will be about zero for most of brands[5].

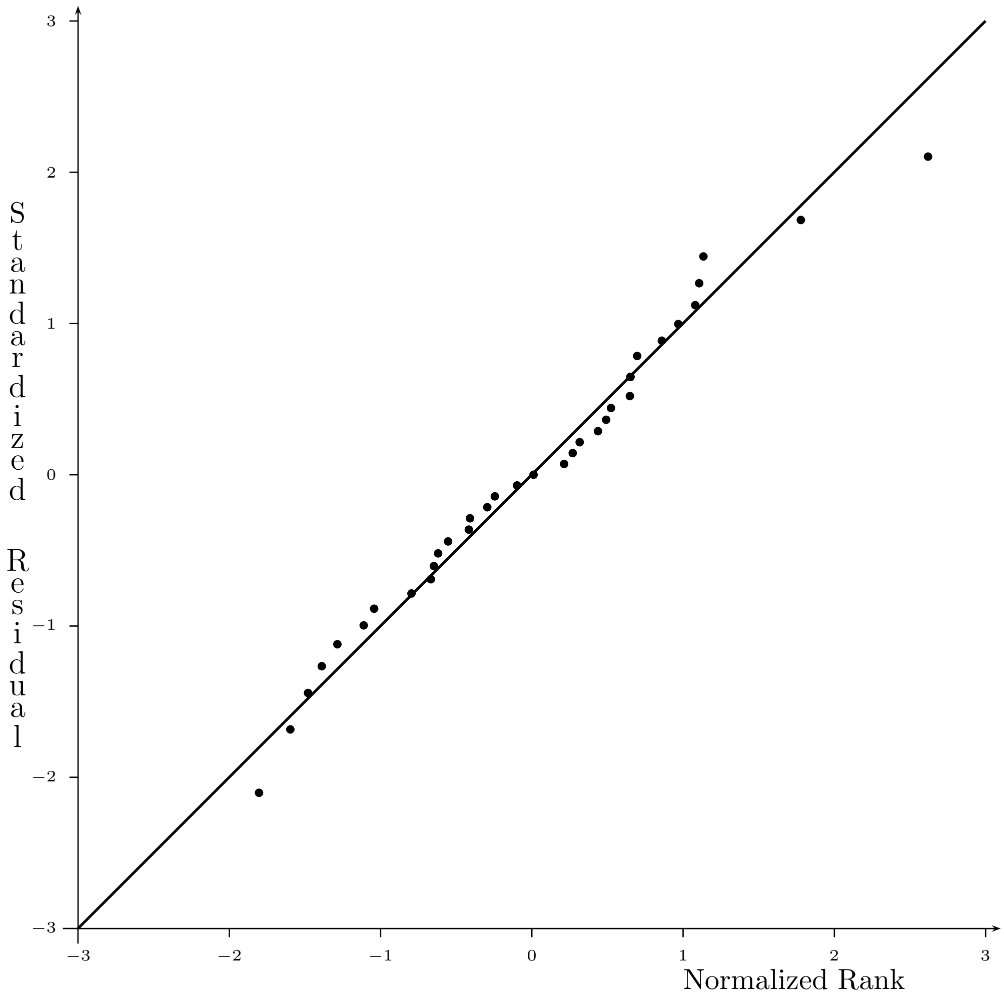
A summary table indicates that ignoring brand name in our analysis is likely to influence our conclusion whether Wintel or Macintosh computers are on an average cheaper. An indication that each observation may not be independent from others. From the table, I would conclude that prices across different observations are systematically related. Consequently, one may want to modify model that includes brand name as another classification variable.

[5] We also have practical implication of these residuals. A large negative value of residual would indicate that predicted price is higher than the observed. An indication that product might be a good bargain! See for example, IBM. On the other hand, a large positive residual might suggest that product is overpriced as is the case for Compulink

**The Average Residuals by Brand Name**

| Wintel Based | | | Macintosh Based | | |
|---|---|---|---|---|---|
| Brand | Residual | n | Brand | Residual | n |
| Dell | 219.4 | 2 | Apple | 278.1 | 4 |
| Gateway | 345.1 | 3 | Umax | −278.1 | 4 |
| HP | −605.6 | 2 | | | |
| Micro Express | −252.2 | 2 | | | |
| Micron | −18.3 | 3 | | | |
| Polywell | 247.8 | 3 | | | |
| Quantex | 356.1 | 5 | | | |
| Xi | 405.4 | 2 | | | |
| All Others | 104.6 | 5 | | | |

To test the assumption that residuals are normally distributed or dependent variable values are normally distributed for each group, we would plot Q-Q plot and examine relevant statistics that examine the null hypothesis that variable is normally distributed. First we examine, Q-Q plot and conclude that standardized residuals (`STPRC`) are normally distributed. Our examination of Skewness and Kurtosis as well as Shapiro-Wilk's Ominibus normality test lead us to conclude that we can accept the null hypothesis that residuals are normally distributed.




```
Normality Test for variable stprc N=35

G1=0.261        SQRTB1=0.250        z =    0.689 prob = 0.4906
G2=-0.037       B2=2.801            z =    0.177 prob = 0.8596

      K**2 = Chisquare (2 df) =    0.506  prob = 0.7763
              Shapiro-Wilk Test =   0.976 prob = 0.7076
```

To examine whether dependent variable variances for two group of equal, we would need to look at the standard deviations. On page 10, we noted that the standard deviation of price for Wintel based computers is 539 while those of Mac is 544. Since these are very close, we will not investigate formal procedure to test whether these standard deviations are equal.

## Matched Comparison

Two observations are made on the same test unit, once before the treatment and another one after the treatment. The main idea behind such experiment is to treat each test unit as a control of itself. Otherwise, one could treat this test procedure same as comparing two groups.

- **Experimental Situation**

  $O_1 \quad X \quad O_2 \quad n$ test units.

Note that measurment is made on $n$ test units at two different times. That is, we obtain $y_{11}, y_{12}, \cdots y_{1n}$ and $y_{21}, y_{22}, \cdots y_{2n}$ dependent variable values as pre- and post-test for the same unit. Our interest lies in finding whether differences are different from zero and statistically significant. Accompanying box indicates nature of creation of dependent variable.

$$
\begin{aligned}
y_{d_1} &= y_{11} - y_{21} \\
y_{d_2} &= y_{12} - y_{22} \\
&. \\
&. \\
y_{d_n} &= y_{1n} - y_{2n}
\end{aligned}
$$

- **Illustrative Example**

This example is from Stevens (1996)[6] which contains multiple groups. For our purpose, we will look at one group here. The example deals with a short instructional slide presentation about behaviour of individual in a group situation and in specifics the roles or character types that members assume. The presentation was in a lecture format of 5 minutes with 18 slides. Each role was identified by an animal. Each animal was shown in two forms: in a cartoon sketch and a realistic picture. Of the 179 subjects who participated in the study, about half saw slides in black and white and the other half saw the slides in colour.

| subject number | Test Score | | |
|---|---|---|---|
| | Time 1 | Time 2 | Difference |
| 29 | 6 | 2 | 4 |
| 57 | 7 | 6 | 1 |
| 59 | 8 | 7 | 1 |
| 60 | 9 | 8 | 1 |
| 61 | 9 | 4 | 5 |
| 63 | 7 | 4 | 3 |
| 83 | 9 | 6 | 3 |
| 85 | 1 | 0 | 1 |
| 86 | 0 | 0 | 0 |
| 87 | 4 | 1 | 3 |
| Mean | 6.0 | 3.8 | 2.2 |
| Std. Dev. | 3.3 | 2.9 | 1.6 |

After seeing the slides, the subjects took a test on the material. The slides were presented in random order, and the subject wrote down the character type represented by that slide. Here we will analyze data for one hospital and the number of cartoon characters correctly identified. Each score could range from 0 to 9, since there were 9 characters. Four weeks later the subjects were retested. 10 subjects who participated in both parts are used below for analysis.
Note that subjects on an average recalled 6 items in the first session while on the second session about 4, a drop in about 2 items recalled. In addition, note that while the standard deviation of number of items recalled in both sessions is about 3, the standard deviation of difference is

[6]Stevens, James (1996) *Applied Multivariate Statistics for the Social Sciences*, 3rd Edition, Lawrence Erlbaum Associates:New Jersy, page 607.

1.6, about 50% decrease in standard deviation.

- **Hypothesis and Test Statistic**

  $H_0 : \bar{y}_d = 0$ and $H_A : \bar{y}_d \neq 0$. This would be tested using $t$-statistic and it is equal to

$$
\begin{aligned}
t &= \frac{\bar{y}_d}{s_d\sqrt{\frac{1}{n}}}, \\
&= \frac{2.2}{1.6 \times \sqrt{.1}}, \\
&= 4.35
\end{aligned}
$$

$s_d^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_d - \bar{y}_d)^2$ and $\bar{y}_d$ is the average difference.

- **Decision Rule**

If $|t| > t_{\alpha/2,n-1}$, then reject the null hypothesis. For our example, since tabled $t$ statistic is 2.26 for $\alpha/2 = 0.025$ and 9 degrees of freedom, we would reject the null hypothesis that the difference score is equal to zero.

- **Assumptions**

  1. Each test unit is independent from others.
  2. Each observation within in a group has equal influence on the estimated group means.
  3. $y_{d1}, y_{d2}, \cdots y_{dn}$ are normally distributed.

- **Comments about Assumptions and Design Issues**

Assumptions about independence and normality can be tested using suggested procedures outlined for two group comparison. As we noted above, one interesting feature about this procedure to test difference is the fact that we generally, should see reduced estimate of variance for difference score. There is statistical reason for this to occur. Consider two random variables, say $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$. Then, it can be shown that variance of $\boldsymbol{y_1} - \boldsymbol{y_2}$ will always be lower than the average variance of both variables, as long as both variables are correlated. Moreover, note that *Matched Comparison* method has intuitive appeal, easy to compute (most of our calculation were performed by calculator) and easy to explain to non-statistician. Consequently, such method is popular among experimental researcher.
There are two weaknesses of this method. First, by definition, we are suggesting that the absolute levels are not important. For example, time 1 mean may have important information about pre-test. Second problem with this method is the fact that we have to collect information from same test unit at two points in time and test unit may not be interested in participating

in the second phase. This usually leads to far lower response rate than if all of information is collected at one occasion.

## Multiple Group Comparison

Suppose that more than two treatments are to be compared with the control group on one dependent variable. Test units (individuals, stores etc.) assigned randomly to their respective groups. As an example, suppose there are three treatments and one control group. Note that we take measurment on $n_1,\ n_2 \cdots n_k$ test units. That is, we obtain

| Test | Experimental Details | | | Test Units |
|---|---|---|---|---|
| 1 | $R$ | $X_1$ | $O_1$ | $n_1.$ |
| 2 | $R$ | $X_1$ | $O_2$ | $n_2.$ |
| 3 | $R$ | $X_3$ | $O_3$ | $n_3.$ |
| Control | $R$ | | $O_4$ | $n_4.$ |

| | Test1 | Test2 | Test3 | Control |
|---|---|---|---|---|
| | $y_{11}$ | $y_{21}$ | $y_{31}$ | $y_{41}$ |
| | $y_{12}$ | $y_{22}$ | $y_{32}$ | $y_{42}$ |
| | . | . | . | . |
| | . | . | . | $y_{4n_4}$ |
| | . | $y_{2n_2}$ | | |
| | $y_{1n_1}$ | | . | |
| | | | $y_{3n_3}$ | |
| mean | $\bar{y}_1$ | $\bar{y}_2$ | $\bar{y}_3$ | $\bar{y}_4$ |
| $s_k^2$ | $\sum_{j=1}^{n_1}(y_{1j}-\bar{y}_1)^2$ | $\sum_{j=1}^{n_2}(y_{2j}-\bar{y}_2)^2$ | $\sum_{j=1}^{n_3}(y_{3j}-\bar{y}_3)^2$ | $\sum_{j=1}^{n_4}(y_{4j}-\bar{y}_4)^2$ |

That is, we took measurement on $n = n_1 + n_2 + n_3 + n_4$ test units. The last two rows in above table are computed averages and within treatment variation.

- **Illustrative Example**

**City and Coupon Discount**

| Store | 15% | 10% | 5% | Control |
|---|---|---|---|---|
| 1 | 25 | 18 | 17 | 12 |
| 2 | 23 | 15 | 14 | 9 |
| 3 | 19 | 14 | 11 | 8 |
| 4 | 17 | 13 | 10 | 7 |
| 5 | 16 | 10 | 8 | |
| 6 | 14 | | | |
| Total | 114 | 70 | 60 | 36 |
| Mean | 19 | 14 | 12 | 9 |
| $n_i$ | 6 | 5 | 5 | 4 |

Let us revisit our coupon experiment from page 2. Suppose we also had 11 more stores in two more cities (6 stores with 15% off coupons, 5 stores with 5% off coupons). Adjoining table provides data for 20 stores. Our results indicate that on an average as coupon incentive percent increases, so does store sales for the brand. To determine the effect of individual treatment or test condition, we need to decompose response from the overall mean (equal to 14) and compute treatment sums of squares, as well as error sums of squares. Four treatment sums of squares will be equal to $6 \times (19-14)^2$, $5 \times (14-14)^2$, $5 \times (12-14)^2$ and $4 \times (9-14)^2$ respectively; or 150, 0, 20 and 100. To compute error sums of squares we need to compute error values and then compute sums of squares. Following table provides error and error squared calculations for each observation.

**City and Coupon Discount**

| Store | 15% | | 10% | | 5% | | Control | |
|---|---|---|---|---|---|---|---|---|
| | Error | Square | Error | Square | Error | Square | Error | Square |
| 1 | 6 | 36 | 4 | 16 | 5 | 25 | 3 | 9 |
| 2 | 4 | 16 | 1 | 1 | 2 | 4 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | −1 | 1 | −1 | 1 |
| 4 | −2 | 4 | −1 | 1 | −2 | 4 | −2 | 4 |
| 5 | −3 | 9 | −4 | 16 | −4 | 16 | | |
| 6 | −5 | 25 | | | | | | |
| Total | 0 | 90 | 0 | 34 | 0 | 50 | 0 | 14 |

Note that the overall sum of squares is equal to $20 \times 14^2$ or 3920. This would lead to following identity,

$$\begin{pmatrix} 4378 \\ \text{Response} \\ \text{sum of squares} \end{pmatrix} = \begin{pmatrix} 3920 \\ \text{Overall} \\ \text{sum of squares} \end{pmatrix} + \begin{pmatrix} 270 \\ \text{Treatment} \\ \text{sum of squares} \end{pmatrix} + \begin{pmatrix} 188 \\ \text{Error} \\ \text{sum of squares.} \end{pmatrix}$$

- **Hypothesis Tested and Statistic**

$H_0 : \bar{y}_1 = \bar{y}_2 = \bar{y}_3 = \bar{y}_4$ and $H_A$ : at least two means are different from each other. To test this hypothesis, we could use following statistic.

$$F = \frac{s_T^2/(k-1)}{(s_1^2 + s_2^2 + s_3^2 + s_4^2)/(n-k)},$$

$s_T^2$ is the sums of squares due to treatments weighted by respective sample size, or $s_T^2 = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2$ and $\bar{y}$ is the overall mean, or $\bar{y} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$. Note that holding error sum of squares constant, as treatment means are closer to the overall mean, treatment sum of squares will be small and as a result $F$-statistic will be small. On the other hand, holding treatment sum of squares constant, having smaller error values lead to smaller error sum of squares and as a result $F$-statistic will be larger. For our example,

$$\begin{aligned} F &= \frac{\text{Mean Treatment Sum of Squares}}{\text{Mean Error Sum of Squares}} \\ &= \frac{270/3}{188/16} \\ &= 7.66 \end{aligned}$$

Note that $s_T^2$ is often called between group variation and $s_1^2, s_2^2 \cdots s_k^2$ are called within group variation.

- **Decision Rule**

If $F > F_{(k-1),(n-k)}(1-\alpha)$, then reject the null hypothesis. That is, look at tabled $F$ distribution value with $k-1$, and $n-1$ degrees of freedom with probability level of $\alpha$. Note also that F-statistic is always positive number. For our example, Upper percentage point of the $F$ distribution for 3 (column of 3) and 16 (row of 16) at $\alpha = 0.05$ is 3.25 and thus we would reject the null hypothesis that all treatment means are equal. Note also that four treatments account for $\frac{270}{458}$ or about 59% total of variation. In addition, treatment of 15% price discount and no price discount account for major portion of this variation. In other words, these two treatment conditions are likely to be far more distinct compared to other two treatments.

- **Assumptions**

1. Each test unit is independent from others.
2. Each test unit has equal influence on group means.
3. Variations within group is a similar.
4. Observed dependent variable values for each group are normally distributed, and also $y_{21}, y_{22}, \cdots y_{2n_1}$ are normally distributed. Alternatively, error values across treatments are normally distributed.
5. variances of all groups are equal.

- **Additional Comments**

With more than two groups, interest often lies in determining similarities in groups. That is, whether means for one or more groups are equal. Consider the present example. There are several comparisons we could propose. A partial list is provided below.

| Comparison | First Group | Compared to | Comments |
|---|---|---|---|
| 1-a | 1 | 2 and 3 | These comparisons involve |
| 1-b | 2 | 1 and 3 | one treatment being |
| 1-c | 3 | 1 and 2 | compared to two other |
| 2-a | 1 | 4 | Here control is compared to |
| 2-b | 2 | 4 | each of three treatments |
| 2-c | 3 | 4 | |
| 3-a | 4 | 1 and 2 | Here comparison involves |
| 3-b | 4 | 1 and 3 | control to two treatments |
| 3-c | 4 | 2 and 3 | taken two at a time |

Each of these comparisons can be stated as a hypothesis. These are called *contrasts.* Consider comparison (1-a). We can state the null hypothesis as $H_0 : 2\bar{y}_1 = \bar{y}_2 + \bar{y}_3$ and the alternative hypothesis would be $H_A : 2\bar{y}_1 \neq \bar{y}_2 + \bar{y}_3$. Note that we can also write out null hypothesis as $2\bar{y}_1 - \bar{y}_2 - \bar{y}_3 = 0$. This, of course, is like our $t-$test or two group comparison. There is a general approach to stating these comparisons which are based on constrast coefficients. That is to write our contrast as

$$c_1\bar{y}_1 + c_2\bar{y}_2 + c_3\bar{y}_3 + c_4\bar{y}_4 = 0,$$

where $c_1, c_2 \cdots$ are chosen to reflect nature of comparison. More compactly, this can be written as $\sum_{j=1}^{k} c_j\bar{y}_j = 0$. For (1-a) comparison, contrast coefficient would be $c_1 = 2$, $c_2 = -1$, $c_3 = -1$ and $c_4 = 0$. Suppose we denote one such constrast as $\hat{C}$. We could state the null hypothesis as $H_0 : \hat{C} = \sum_{j=1}^{k} c_j\bar{y}_j = 0$ with the condition that $\sum_{j=1}^{k} c_j = 0$. To test this null hypothesis, we would need variance of such a linear combination. Note that estimated variance of $\hat{C}$ is equal to

$$\text{Var}(\hat{C}) = (\text{Mean Error Sum of Squares}) \sum_{j=1}^{k} \frac{c_j^2}{n_j}.$$

Note that computation of such variances is relatively easy because most of our contrast coefficients involve 2 and 1 and squaring such quantities is easy. Moreover, we already have computed Error sum of squares and does not require additional computations. Note that for (1-a) comparison,

$$\hat{C} = 2 \times 19 - 14 - 12 = 12$$

and estimated variance is

$$\begin{aligned} \text{Var}(\hat{C}) &= \frac{188}{16}\left[\frac{2^2}{6} + \frac{1^2}{5} + \frac{1^2}{5}\right] \\ &= \frac{188}{16}\left[\frac{4}{6} + \frac{1}{5} + \frac{1}{5}\right] = 12.533. \end{aligned}$$

This would result in $t$-statistic of $\frac{12}{\sqrt{12.533}} = 3.39$. Since *Tabled t-distribution value* at $\alpha = 0.05$ for 16 degrees of freedom is 2.12 we would reject the null hypothesis. Consequently, we conclude that the average sales for 15% off coupon had significantly more sales than condition when coupon amounts were 10% or 5%. Similar computation may be performed to compare remaining contrasts. Both SAS and SPSS allows one to specify and test various planned as well as unplanned comparisons.

- **Realistic Example: Which Snack and How Many Calories?**

Below we will examine calorie content of four group of snacks. Note that all of data come from *Nutrition Action Healthletter*[7] and used here for illustrative purpose. SAS input and output follows.

[7] "Sugar, How Bittersweet It Is", Nov. 1998, p. 3–7

```
options nodate nocenter ps = 80 ls = 75;
data snacks;
input calories sugar tot_fat snack product $;
/*     Calories
       Sugar    is in grams
       tot_fat  is in grams
       snack = 1 is for Cookies
             = 3 is for Cakes, Donuts
             = 4 is for Candy
             = 6 is for Ice Cream and Frozen Desserts
        Data came from Nutrition Action Healthletter, Nov. 1998, p 6-7.   */
if snack = 1 then snacks = "Cookies";
if snack = 3 then snacks = "Cakes";
if snack = 4 then snacks = "Candy";
if snack = 6 then snacks = "Ice_crm";
datalines;
180 16  9 3 Little Debbie Chocolate Cupcakes (1-45g)
230 16 10 3 Tim Hortons Honey Dip Donut (1)
250 16 12 3 Tim Hortons Old Fashion Glazed Donut (1)
130 17  0 3 Entenmann's Light Fat Free Golden Loaf (1/8 - 48g)
130 17  0 3 Entenmann's Light Fat Free Marble Loaf (1/8 - 50g)
160 17  3 3 Entenmann's Low Fat Cinnamon Buns (1 - 61g)
120 18  0 3 Entenmann's 97% Fat Free Chocolate Loaf (1/8 - 53g)
210 18  9 3 Entenmann's All Butter Loaf (1/6 - 57g)
230 18 10 3 Tim Hortons Chocolate Dip Donut (1)
310 18 13 3 Tim Hortons Dutchie (1)
210 19  6 3 Entenmann's 50% Less Fat Crumb Delight (1/9 - 57g)
370 22 22 3 Tim Hortons Chocolate glaze Donut (1)
440 22 34 3 Tim Hortons French Cruller (1)
270 24 13 3 Little Debbie Fudge Brownie (1 - 61g)
260 30 12 3 Little Debbie Swiss Cake Rolls (2 - 61g)
190 33  0 3 Entenman's Light Fat Free Fudge Iced Chocolate Cake (1/6 - 85g)
330 35 14 3 Entenman's Lousiana Crunch Cake (1/9 - 82g)
340 35 18 3 Entenman's Marshmallow Iced Devil's Food Cake (1/6 - 85g)
310 36 14 3 Entenman's Chocolate Fudge Cake (1/6 - 85g)

150 16  5 4 Hershey Sweet Escapes Caramel Chewy Bars (1 - 39g)
190 22  8 4 Hershey Sweet Escapes Chocolate Crisp Bars (1 - 39g)
240 22 14 4 Hershey Kisses (9 - 44g)
280 23 17 4 Reese Peanut Butter Cups (3 - 51g)
160 26  3 4 York Peppermint Patties (3 - 41g)
280 27 14 4 Twix Caramel Cookie Bars (1 package - 57g)
160 28  1 4 Hershey Twizzlers, Strawberry (45g)
270 29 11 4 Butterfinger (1 - 60g)
280 29 14 4 Snickers (1 - 62g)
240 31 10 4 M&M Plain (1 package - 48g)
180 37  3 4 Junior Mints (1 box - 45g)
260 40  8 4 3Musketeers (1 - 60g)

240  24 14 6 Baskin-Robbins Vanilla (1 Scoop - 113g)
260  27 16 6 Baskin-Robbins Mint Chocolate Chip (1 scoop - 113g)
260  29 16 6 Baskin-Robbins Chocolate (1 Scoop - 113g)
270  33 14 6 Baskin-Robbins Pralines'n Cream (1 Scoop - 113g)
330  38  9 6 Dairy Queen Vanilla Cone, Medium (198g)
540  42 36 6 Haagen-Dazs Chocolate, Coffee or Vanilla (1 cup)
490  43 24 6 Dairy Queen Dipped Cone, Medium (220g)
600  48 40 6 Haagen-Dazs Chocolate Chocolate Chip (1 cup)
410  49 12 6 Dairy Queen Vanilla Cone, Large (253g)
400  61 10 6 Dairy Queen Sundae with Chocolate Topping medium (234g)
520  61 18 6 Dairy Queen Chocolate Sandwich Cookie Blizzard Small (276g)
770 113 20 6 Dairy Queen Chocolate Shake, Medium (539g)
```

```
170  8  3 1 Honey Maid 40% less Fat Graham Wafers (6-40g)
180 10  5 1 Christie Teddy Graham Wafers (32-40g)
180 10  5 1 Honey Maid Graham Wafers (6-40g)
180 10  6 1 Christie Teddy Graham, Chocolate (32-40g)
170 12  3 1 Snackwell's Chocolate Creme Sandwich (3-40g)
140 12  5 1 Christie Oreo 25% Less Fat (2-30g)
200 12  9 1 Christie Chips Ahoy! (3-40g)
170 13  3 1 Snackwell's Vanilla Creme Sandwich (3-40g)
200 14  8 1 Christie Oreo Mini (13-40g)
140 15  8 1 Pepperidge Farm Chesapeake Pecan (1-26g)
150 17  7 1 Christie Oreo (2-30g)
;;;;
proc glm;
class snacks;
model calories = snacks ;
means snacks ;
run;

General Linear Models Procedure
Class Level Information

Class    Levels    Values

SNACKS        4    Cakes Candy Cookies Ice_crm

Number of observations in data set = 54

Dependent Variable: CALORIES
                                 Sum of          Mean
Source                 DF       Squares        Square  F Value    Pr > F
Model                   3    428405.192    142801.731    15.14    0.0001
Error                  50    471737.400      9434.748
Corrected Total        53    900142.593

                 R-Square          C.V.      Root MSE      CALORIES Mean
                 0.475930      36.60267       97.1326            265.370


Source                 DF      Type I SS   Mean Square  F Value    Pr > F
SNACKS                  3    428405.192    142801.731    15.14    0.0001

Level of      ----------CALORIES---------
SNACKS     N      Mean            SD

Cakes     19    245.789474      87.769635
Candy     12    224.166667      52.303022
Cookies   11    170.909091      20.714510
Ice_crm   12    424.166667     164.784064
```

Our analysis here indicates that cookies have on an average lower calories (about 171 per serving) than all other snacks. Candies, and cakes have about 224 and 246 calories respectively. Not surprisingly, ice creams are loaded with calories. Note also that the standard deviation for cookies compared to ice creams, there is eight times more variability in ice creams than variability in cookies. This is an indication that we may be violating assumption that variability across subgroups is about same.

## Multiple Factorial Comparison

Consider a situation where interest is in determining impact of more than one experimental factor. One example might be price level (−5%, 0% and +5%) and whether brand was promoted via feature advertising (two levels). One would be interested in comparing changes to test units for both factors as well as interaction between two factors. Note that interaction have both statistical as well as substative meaning.

- **Experimental Situation**

Consider an experiment with two factors. First factor has $L$-levels and the second factor has $K$-levels. In order to understand interactions, we would need at least $L \times K$ treatment groups. If $L$ is 3 and $K$ is 2, then we would need 6 treatment groups. Note that we take measurment on $n_{11}$, $n_{12}$, $\cdots n_{32}$ test units for six different treatments respectively. That is, we obtain $y_{11}, y_{12}, \cdots y_{1n_{11}}$, $y_{21}, y_{22}, \cdots y_{2n_{32}}$ and so on dependent variable values.

| Group | | | | Test Units | Factor 1 | Factor 2 |
|---|---|---|---|---|---|---|
| 1 | $R$ | $X_{11}$ | $O_1$ | $n_{11}$ | 1 | 1 |
| 2 | $R$ | $X_{12}$ | $O_2$ | $n_{12}$ | 1 | 2 |
| 3 | $R$ | $X_{21}$ | $O_3$ | $n_{21}$ | 2 | 1 |
| 4 | $R$ | $X_{22}$ | $O_4$ | $n_{22}$ | 2 | 2 |
| 5 | $R$ | $X_{31}$ | $O_5$ | $n_{31}$ | 3 | 1 |
| 6 | $R$ | $X_{32}$ | $O_6$ | $n_{32}$ | 3 | 2 |

- **Illustrative Example - Balanced Design**

| | Price | | |
|---|---|---|---|
| | −5% | 0% | +5% |
| In-Store Display | 28 | 24 | 11 |
| | 26 | 23 | 10 |
| | 24 | 21 | 9 |
| | 22 | 16 | 9 |
| | 20 | 16 | 6 |
| No In-Store Display | 20 | 18 | 12 |
| | 17 | 15 | 10 |
| | 16 | 14 | 7 |
| | 15 | 12 | 6 |
| | 12 | 11 | 5 |

In an experiment that involved manipulation of price level (5% price reduction, no price change and 5% price increase) and presence or absence of store display. Data collection consisted of monitoring store sales, in terms of number of cases sold for one week for 24 stores. Five stores were used per experimental condition and data were collected for one week.

- **Hypothesis Tested**

There are three hypotheses pertaining to price, display and the effect of price conditional on display. Unlike one factor multiple levels, here we have multiple tests planned. Suppose price is labeled as factor 1 and in-store display is labeled as factor 2, then we can write our hypotheses formally as follows:

1. We would expect that higher price reduction results in higher sales. This would be stated as $H_0 : \bar{y}_{1\cdot} = \bar{y}_{2\cdot} = \bar{y}_{3\cdot}$ and $H_A$ : At least one mean is different from other two.
2. We would expect that putting brand on display results in higher sales. This would be writen as $H_0 : \bar{y}_{\cdot 1} = \bar{y}_{\cdot 2}$ and $H_A : \bar{y}_{\cdot 1} \neq \bar{y}_{\cdot 2}$.
3. When brand is on display and there is a price reduction, customers who normally do not buy particular brand would be tempted to buy product and thus resulting in higher sales. On the other hand, when brand is not on display and price is increased no one likely to be interested in brand (except those who normally would buy brand) and this would result in lower sales. Formally, we would stat this as $H_0$ :

$$\bar{y}_{11} + \bar{y}_{12} + \bar{y}_{21} + \bar{y}_{22} + \bar{y}_{31} + \bar{y}_{32} = \bar{y}_{1\cdot} + \bar{y}_{2\cdot} + \bar{y}_{3\cdot} + \bar{y}_{\cdot 1} + \bar{y}_{\cdot 2}$$

   and $H_A$ : These means are not equal.

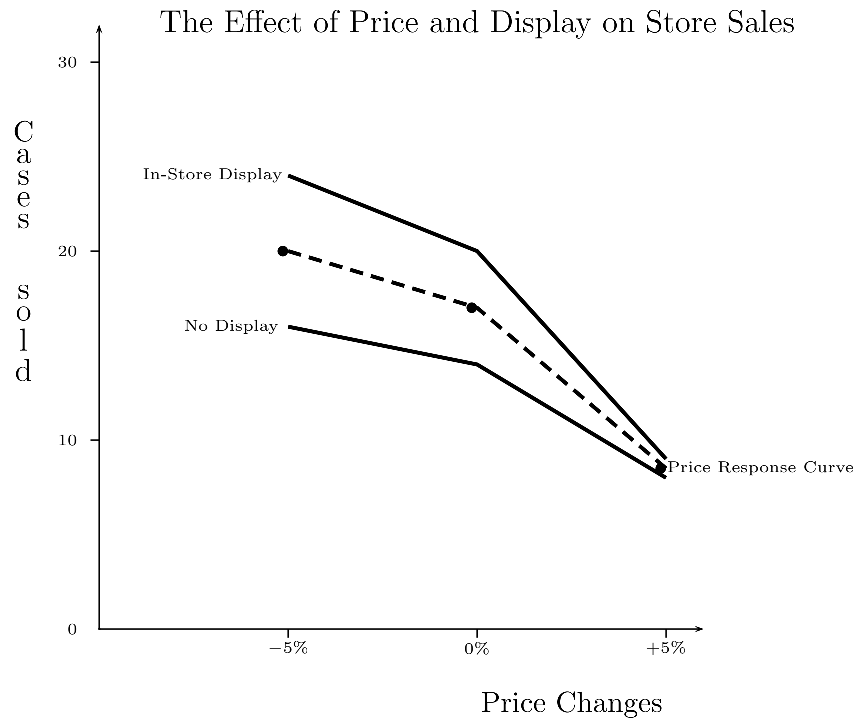
Note that $\bar{y}_{1\cdot}$, $\bar{y}_{2\cdot}$ and $\bar{y}_{3\cdot}$ denote means for factor 1 and level 1, level 2 and level 3 respectively. Similarly, $\bar{y}_{\cdot 1}$ and $\bar{y}_{\cdot 2}$ denote means for factor 2 and level 1 and level 2 respectively.

- **Test Statistic**

**Means by Conditions**

| | Price | | | |
|---|---|---|---|---|
| Display | −5% | 0% | +5% | |
| Yes | 24 | 20 | 9 | 17.67 |
| No | 16 | 14 | 8 | 12.67 |
| | 20.00 | 17.00 | 8.50 | 15.17 |

One useful starting point to understand these effects is to construct a table that indicates various averages and if possible their standard deviations. It is also useful to display graphically nature of main effects and source of interaction. Adjoining table provide averages by various experimental conditions. Following graph provides pictorial summary of most of the averages. Note that there is likely to be main effects of price and display. A closer examination of graph reveals that there might be interaction between price and display as well. That is, price increase with or without display bears no advantage in terms of sales.




In order to understand relative variability within and across test conditions, we need to decompose observed response to assign variability to various factors that may have contributed it. To accomplish that let us look at decomposition. Suppose we denote observed value as $y_{lki}$ where $l$ indicates level associated with factor A, $k$ indicates level associated with factor B and $i$ individual response such that $i = 1, 2, \cdots n$ where $n$ is sample size or $n = \sum_{l=1}^{L} \sum_{k=1}^{K} n_{lk}$. Each observation can be decomposed as the overall mean, the effect due to price, display, interaction between price and display and error. Following table displays this decomposition. Note that to determine the effect of price at $-5\%$ reduction we need to subtract $15\frac{1}{6}$ from 20.

$$
\underset{\text{Sales}}{\begin{pmatrix} 28\\26\\24\\22\\20\\ \\24\\23\\21\\16\\16\\ \\11\\10\\9\\9\\6\\ \\20\\17\\16\\15\\12\\ \\18\\15\\14\\12\\11\\ \\12\\10\\7\\6\\5 \end{pmatrix}}
=
\underset{\text{Overall}}{\begin{pmatrix} 15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\ \\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\ \\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\ \\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\ \\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\ \\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6}\\15\frac{1}{6} \end{pmatrix}}
+
\underset{\text{Price}}{\begin{pmatrix} 4\frac{5}{6}\\4\frac{5}{6}\\4\frac{5}{6}\\4\frac{5}{6}\\4\frac{5}{6}\\ \\1\frac{5}{6}\\1\frac{5}{6}\\1\frac{5}{6}\\1\frac{5}{6}\\1\frac{5}{6}\\ \\-6\frac{2}{3}\\-6\frac{2}{3}\\-6\frac{2}{3}\\-6\frac{2}{3}\\-6\frac{2}{3}\\ \\4\frac{5}{6}\\4\frac{5}{6}\\4\frac{5}{6}\\4\frac{5}{6}\\4\frac{5}{6}\\ \\1\frac{5}{6}\\1\frac{5}{6}\\1\frac{5}{6}\\1\frac{5}{6}\\1\frac{5}{6}\\ \\-6\frac{2}{3}\\-6\frac{2}{3}\\-6\frac{2}{3}\\-6\frac{2}{3}\\-6\frac{2}{3} \end{pmatrix}}
+
\underset{\text{Display}}{\begin{pmatrix} 2\frac{1}{2}\\2\frac{1}{2}\\2\frac{1}{2}\\2\frac{1}{2}\\2\frac{1}{2}\\ \\2\frac{1}{2}\\2\frac{1}{2}\\2\frac{1}{2}\\2\frac{1}{2}\\2\frac{1}{2}\\ \\2\frac{1}{2}\\2\frac{1}{2}\\2\frac{1}{2}\\2\frac{1}{2}\\2\frac{1}{2}\\ \\-2\frac{1}{2}\\-2\frac{1}{2}\\-2\frac{1}{2}\\-2\frac{1}{2}\\-2\frac{1}{2}\\ \\-2\frac{1}{2}\\-2\frac{1}{2}\\-2\frac{1}{2}\\-2\frac{1}{2}\\-2\frac{1}{2}\\ \\-2\frac{1}{2}\\-2\frac{1}{2}\\-2\frac{1}{2}\\-2\frac{1}{2}\\-2\frac{1}{2} \end{pmatrix}}
+
\underset{\text{Interaction}}{\begin{pmatrix} 1\frac{1}{2}\\1\frac{1}{2}\\1\frac{1}{2}\\1\frac{1}{2}\\1\frac{1}{2}\\ \\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\ \\-2\\-2\\-2\\-2\\-2\\ \\-1\frac{1}{2}\\-1\frac{1}{2}\\-1\frac{1}{2}\\-1\frac{1}{2}\\-1\frac{1}{2}\\ \\-\frac{1}{2}\\-\frac{1}{2}\\-\frac{1}{2}\\-\frac{1}{2}\\-\frac{1}{2}\\ \\2\\2\\2\\2\\2 \end{pmatrix}}
+
\underset{\text{Error}}{\begin{pmatrix} 4\\2\\0\\-2\\-4\\ \\4\\3\\1\\-4\\-4\\ \\2\\1\\0\\0\\-3\\ \\4\\1\\0\\-1\\-4\\ \\4\\1\\0\\-2\\-3\\ \\4\\2\\-1\\-2\\-3 \end{pmatrix}}
$$

A summary of variability is provided below.

$$
\begin{pmatrix} y_{lki}^2 \\ \text{Sales} \\ 8075 \end{pmatrix}
=
\begin{pmatrix} 30 \times \bar{y}_{\cdots}^2 \\ \text{Overall} \\ \text{Mean} \\ 6900.83 \end{pmatrix}
+
\begin{pmatrix} (\bar{y}_{l\cdot\cdot} - \bar{y}_{\cdots})^2 \\ \text{Price} \\ \text{Effect} \\ 711.667 \end{pmatrix}
+
\begin{pmatrix} (\bar{y}_{\cdot k\cdot} - \bar{y}_{\cdots})^2 \\ \text{Display} \\ \text{Effect} \\ 187.5 \end{pmatrix}
+
$$

$$
\begin{pmatrix} (\bar{y}_{lk\cdot} - \bar{y}_{l\cdot\cdot} - \bar{y}_{\cdot k\cdot} + \bar{y}_{\cdots})^2 \\ \text{Price} \times \text{Display} \\ \text{Effect} \\ 65 \end{pmatrix}
+
\begin{pmatrix} (y_{lki} - \bar{y}_{lk\cdot})^2. \\ \text{Error} \\ 210 \end{pmatrix}
$$

| Source of variation | Denoted by | Sum of squares | Degrees of freedom |
|---|---|---|---|
| Factor 1 | $ss_{\text{fac}_1}$ | $\sum_{l=1}^{L} n_l(\bar{y}_{l\cdot} - \bar{y})^2$ | $L-1$ |
| Factor 2 | $ss_{\text{fac}_2}$ | $\sum_{j=1}^{K} n_k(\bar{y}_{\cdot k} - \bar{y})^2$ | $K-1$ |
| Interaction | $ss_{\text{int}}$ | $\sum_{l=1}^{L}\sum_{k=1}^{K}(\bar{y}_{lk} - \bar{y}_{l\cdot} - \bar{y}_{\cdot k} + \bar{y})^2$ | $(L-1)(K-1)$ |
| Residual (Error) | $ss_{\text{res}}$ | $\sum_{i=1}^{n}(y_{lki} - \bar{y}_{lk})^2$ | $(n-LK)$ |

Note that $n$ is total number of test units used in the study which is equal to $\sum_{l=1}^{L}\sum_{k=1}^{K} n_{lk}$. The $F$-ratios pertaining to various mean squares can be computed to test null hypothesis. To test the effect of factor 1, we would look at ratio of $ss_{\text{fac}_1}/(L-1)$ and $ss_{\text{res}}/(n-LK)$ and our decision rule would be same as multiple group comparison. We will use SAS output to study various tests below.

- **SAS Output to Test various Hypotheses**

```
Dependent Variable: SALES
                                   Sum of           Mean
Source                  DF        Squares         Square  F Value    Pr > F
Model                    5     964.166667     192.833333    22.04    0.0001
Error                   24     210.000000       8.750000
Corrected Total         29    1174.166667

                  R-Square           C.V.       Root MSE         SALES Mean
                  0.821150       19.50356        2.95804            15.1667

Source                  DF      Type I SS    Mean Square  F Value    Pr > F

PRICE                    2     711.666667     355.833333    40.67    0.0001
DISPLAY                  1     187.500000     187.500000    21.43    0.0001
PRICE*DISPLAY            2      65.000000      32.500000     3.71    0.0393

General Linear Models Procedure

Level of        ------------SALES------------
PRICE       N       Mean              SD

1          10     20.0000000       5.09901951
2          10     17.0000000       4.44722135
3          10      8.5000000       2.36877840

Level of        ------------SALES------------
DISPLAY     N       Mean              SD

2          15     17.6666667       7.14809332
1          15     12.6666667       4.40238031

Level of   Level of        ------------SALES------------
PRICE      DISPLAY    N       Mean              SD

1          2          5     24.0000000       3.16227766
1          1          5     16.0000000       2.91547595
2          2          5     20.0000000       3.80788655
2          1          5     14.0000000       2.73861279
3          2          5      9.0000000       1.87082869
```

```
3          1          5      8.0000000       2.91547595
```

From above analysis, we would conclude that we would reject that price and display had no influence on sales. Moreover, we would also that display and price reduction jointly influence sales as well. This is tested as interaction. Note that our model consisting of five parameters accounts for 964.17 sum of squares of that 711.67 or 73.81% total of variation is accounted by price factor. Moreover, display accounts for 187.5 sum of squares or 19.45% of total variation and price × display interaction account for 65 sum of squares or 6.74% of total variation explained by the model. These measures often are used assign relative importance of factors.

- **Illustrative Example - Unbalanced Case**

| | Price | | |
|---|---|---|---|
| | −5% | 0% | +5% |
| In-Store Display | 24<br>23<br>21<br>20 | 22<br>18 | 21<br>18<br>17<br>15<br>13 |
| No In-Store Display | 20<br>19<br>17<br>16 | 10<br>9<br>9<br>8<br>7 | 11<br>8<br>7<br>6 |

In many situations, either by design (2 × 2, and control condition) or due to lack of true control in the experimental variables, planned experiment results in "empty cells" or unequal observations per cell. Such situations are called unbalanced experiments. As a consequence of such situation, our experimental comparisons are not statistically independent. This generally, influences statistical power of experimental treatments. Following example is similar to the one we discussed above, except number of observations per cell are not equal.

Consider an experiment that involved manipulation of price level (5% price reduction, no price change and 5% price increase) and presence or absence of store display. Data collection consisted of monitoring store sales, in terms of number of cases sold for one week for 24 stores. It was intended that there would be five stores per experimental condition, but number last minute cancellations as well as not implementing experimental treatment correctly resulted in varying observations per condition.

**Means by Conditions**

| | Price | | | |
|---|---|---|---|---|
| Display | −5% | 0% | +5% | |
| Yes | 22 | 20 | 16.8 | 19.27 |
| No | 18 | 8.6 | 8 | 11.30 |
| | 20.00 | 11.86 | 12.89 | 14.96 |

- **Hypothesis Tested**

There are three hypotheses pertaining to price, display and the effect of price conditional on display. All three hypotheses and test statistics are listed on page 23 where balanced case is discussed.

We start by constructing a table that indicates various averages and if possible their standard deviations. A summary table on previous page provides averages by various experimental conditions. Following graph provides pictorial summary of most of the averages. Note that there is likely to be main effects of price and display. A closer examination of graph reveals that there might be interaction between price and display as well.


- **SAS Input to Test various Hypotheses**

```
options nodate nocenter ps = 80 ls = 75;
data AnoEx4;
input display price sales;
if display = 2 then display = -1;
price1 = 0;
price2 = 0;
if price = 1 then price1 = 1;
if price = 2 then price2 = 1;
if price =3 then price1 = -1;
if price = 3 then price2 = -1;
prc1dis = price1*display;
prc2dis = price2*display;
datalines;
1 1 24
1 1 23
1 1 21
1 1 20
1 2 22
1 2 18
1 3 21
1 3 18
1 3 17
1 3 15
1 3 13
2 1 20
2 1 19
2 1 17
```

```
2 1 16
2 2 10
2 2  9
2 2  9
2 2  8
2 2  7
2 3 11
2 3  8
2 3  7
2 3  6
;;;;
proc glm;
class price display;
model sales = price display price*display / ss1 ss2 ss3 ss4 e e1 e2 e3 e4;
means  price display price*display;
run;
```

- **SAS Output from PROC GLM**

```
Dependent Variable: SALES
                                 Sum of          Mean
Source                 DF       Squares        Square  F Value    Pr > F

Model                   5    698.958333    139.791667    29.96    0.0001
Error                  18     84.000000      4.666667
Corrected Total        23    782.958333

                 R-Square          C.V.      Root MSE         SALES Mean

                 0.892715      14.44176       2.16025            14.9583

Source                 DF     Type I SS   Mean Square  F Value    Pr > F

PRICE                   2    309.212302    154.606151    33.13    0.0001
DISPLAY                 1    340.139290    340.139290    72.89    0.0001
PRICE*DISPLAY           2     49.606742     24.803371     5.32    0.0153

Source                 DF    Type II SS   Mean Square  F Value    Pr > F

PRICE                   2    271.344307    135.672154    29.07    0.0001
DISPLAY                 1    340.139290    340.139290    72.89    0.0001
PRICE*DISPLAY           2     49.606742     24.803371     5.32    0.0153

Source                 DF   Type III SS   Mean Square  F Value    Pr > F

PRICE                   2    255.539326    127.769663    27.38    0.0001
DISPLAY                 1    354.933333    354.933333    76.06    0.0001
PRICE*DISPLAY           2     49.606742     24.803371     5.32    0.0153

Source                 DF    Type IV SS   Mean Square  F Value    Pr > F

PRICE                   2    255.539326    127.769663    27.38    0.0001
DISPLAY                 1    354.933333    354.933333    76.06    0.0001
PRICE*DISPLAY           2     49.606742     24.803371     5.32    0.0153

General Linear Models Procedure
```

```
Level of      ------------SALES------------
PRICE      N       Mean            SD

1          8     20.0000000      2.72554058
2          7     11.8571429      5.75698334
3          9     12.8888889      5.27836254

Level of       ------------SALES------------
DISPLAY    N       Mean            SD

 1         11     19.2727273      3.40854541
-1         13     11.3076923      4.90551756

Level of   Level of       ------------SALES------------
PRICE      DISPLAY    N       Mean            SD

1          1          4     22.0000000      1.82574186
1         -1          4     18.0000000      1.82574186
2          1          2     20.0000000      2.82842712
2         -1          5      8.6000000      1.14017543
3          1          5     16.8000000      3.03315018
3         -1          4      8.0000000      2.16024690
```

- **Assumptions**

1. Each test unit is independent from others.
2. Each test unit has equal influence on group means.
3. Variations within group is a similar.
4. Observed dependent variable values for each group are normally distributed. Alternatively, error values across treatments are normally distributed.
5. variances of all groups are equal.

- **Additional Comments and Design Issues**

When there are an equal number of test units in each cell in a factorial design, then sum of squares for different factors (main effects and interactions) are uncorrelated. That means, when testing for a particular main effect, we should not be concerned about what other factors might be affecting dependent variables. This certainly is an ideal situation and most research designs that contain factorial designs strive towards equal cell sizes. In some situations, either by design or by chance, we do not have equal cell sizes. SAS and SPSS allow three or four alternative approaches and sum of squares computed from these methods may lead to differing conclusions about hypotheses tested. Four methods used in SAS are

1. *Sequential sums of squares* are the incremental improvement in error sum of squares as each effect is added to the model (often called **Type I tests**).

2. *Incremental sum of squares* indicate particular factor contribution to the model, assuming all other factors are already included in the model (called **Type II tests**).

3. *Partial sum of squares* indicate marginal contribution of a factor with no other other factor included in the model ( called **Type III and IV tests**).

4. When there are missing cells in design, Type III and IV tests differ, in all other instances they will produce identical results. Type III tests have independence property while Type IV tests have balancing property. In other words, Type III tests rely on cell means without considering cell sample sizes. On the other hand, Type IV tests weigh cell means proportional to cell variance.

A descriptive comparison of these four alternative forms of sum of squares is provided below

| | Sum of Squares Computation | | | |
|---|---|---|---|---|
| Property | Type I | Type II | Type III | Type IV |
| Add upto total Model effect | Yes | May not | May not | May not |
| Order importance | Yes | May not | May not | May not |
| Unequal cell $n$'s | Results vary | | Do not vary | |
| Useful for | Balanced designs, Polynomial Model | Compare subset, Main Effects | Empty cells by design or choice | |

- **Concluding Comments**

In above we have discussed some of the simplest designs and tools to analyze them. There are many other experimental designs such as fractional factorial, incomplete block and latin square designs to name few that require careful examination of sources of variation and its implication on dependent variables. Moreover, although one would like to include and vary many factors simultaneously in one experiment, many researcher believe that studying factors in isolation and controlling others is one of the best approach to understand impact of factor on a dependent variable.