Bryan L. Mack

Week 2 R Training #1 – Factorial Design

# Factorial Design

In a **factorial design**, there are more than one factors under consideration in the experiment. The test subjects are assigned to treatment levels of every factor combinations at random.

#### Example

A fast food franchise is test marketing 3 new menu items in both East and West Coasts of continental United States. To find out if they the same popularity, 12 franchisee restaurants from each Coast are randomly chosen for participation in the study. In accordance with the factorial design, within the 12 restaurants from East Coast, 4 are randomly chosen to test market the first new menu item, another 4 for the second menu item, and the remaining 4 for the last menu item. The 12 restaurants from the West Coast are arranged likewise.

#### Problem

Suppose the following tables represent the sales figures of the 3 new menu items after a week of test marketing. Each row in the upper table represents the sales figures of 3 different East Coast restaurants. The lower half represents West Coast restaurants. At .05 level of significance, test whether the [mean](http://www.r-tutor.com/node/35) sales volume for the new menu items are all equal. Decide also whether the mean sales volume of the two coastal regions differs.

East Coast:   
==========   
   Item1 Item2 Item3   
E1    25    39    36   
E2    36    42    24   
E3    31    39    28   
E4    26    35    29   
   
West Coast:   
==========   
   Item1 Item2 Item3   
W1    51    43    42   
W2    47    39    36   
W3    47    53    32   
W4    52    46    33

#### Solution

The solution consists of the following steps:

1. Save the sales figure into a file named "fastfood-3.csv" in [CSV format](http://www.r-tutor.com/node/69) as follows.

Item1,Item2,Item3   
E1,25,39,36   
E2,36,42,24   
E3,31,39,28   
E4,26,35,29   
W1,51,43,42   
W2,47,39,36   
W3,47,53,32   
W4,52,46,33

Mack - done

1. Load the data into a [data frame](http://www.r-tutor.com/node/10) named df3 with the read.csv function.

> df3 = read.csv("fastfood-3.csv")

Mack – done, note that this file has to be in your current working directory. In R, find this by calling the function getwd()

1. Concatenate the data rows in df3 into a single vector r .

> r = c(t(as.matrix(df3))) # response data   
> r   
 [1] 25 39 36 36 42 ...

done

1. Assign new variables for the treatment levels and number of observations.

> f1 = c("Item1", "Item2", "Item3") # 1st factor levels   
> f2 = c("East", "West")            # 2nd factor levels   
> k1 = length(f1)          # number of 1st factors   
> k2 = length(f2)          # number of 2nd factors   
> n = 4                    # observations per treatment

done

1. Create a vector that corresponds to the 1*th* treatment level of the response data r in step 3 element-by-element with the gl function.

> tm1 = gl(k1, 1, n\*k1\*k2, factor(f1))   
> tm1   
 [1] Item1 Item2 Item3 Item1 Item2 ...

Done – but not sure I understand it

1. Similarly, create a vector that corresponds to the 2*nd* treatment level of the response data r in step 3.

> tm2 = gl(k2, n\*k1, n\*k1\*k2, factor(f2))   
> tm2   
 [1] East East East East East ...

Done – but not sure I understand it

1. Apply the function aov to a formula that describes the response r by the two treatment factors tm1 and tm2 with interaction.

> av = aov(r ~ tm1 \* tm2)  # include interaction

Done – but not sure I understand it

1. Print out the ANOVA table with summary function.

> summary(av)   
            Df Sum Sq Mean Sq F value  Pr(>F)   
tm1          2    385     193    9.55  0.0015 \*\*   
tm2          1    715     715   35.48 1.2e-05 \*\*\*   
tm1:tm2      2    234     117    5.81  0.0113 \*   
Residuals   18    363      20

Done

#### Answer

Since the p-value of 0.0015 for the menu items is less than the .05 significance level, we reject the null hypothesis that the mean sales volume of the new menu items are all equal. Moreover, the p-value of 1.2e-05 for the east-west coasts comparison is also less than the .05 significance level. It shows there is a difference in overall sales volume between the coasts. Finally, the last p-value of 0.0113 (< 0.05) indicates that there is a possible interaction between the menu item and coast location factors, i.e., customers from different coastal regions have different tastes.

Not sure I get this

#### Exercise

Create the response data in step 3 above along *vertical*columns instead of horizontal rows. Adjust the factor levels in steps 5 and 6 accordingly.

Didn’t do this yet