Stats 424 DOE Project
Shane McIntyre

Background:

As a kid in middle school my school would have an annual competition. We would all line up on one side of the gymnasium and throw our paper airplanes. Whosever plane flew the furthest was declared the winner for the class. Everyone had to build their plane themselves. The one rule: the plane had to be made out of paper. Being a 10-year-old kid, I just folded the paper into a simple plane shape. I remember the excitement I had as the teacher told everyone to throw their plane. I threw mine, watched it fly forward, and after just a couple feet my plane proceeded to take a sudden and sharp nose dive into the ground. My plane maybe flew 10 feet, an abysmal attempt.

Design:

From my failures as a child, I will design an experiment to create the perfect paper airplane. My experiment will consist of a 2^4 factorial experiment measuring the factors of the type of paper (printer vs. card stock), type of wing style (rectangular wing vs. triangular wing), weight of plane (put pen in base vs. no pen), and the length of the plane (short vs. long).

Factor	Variable	High (+)	Low (-)
Paper type	A	Card Stock	Printer
Wing style	B	Rectangle	Triangle
Weight	C	W/ pen	W/out pen
Length	D	Long ()	Short ()

I was curious as to what factors would lead to the creation of the most efficient paper airplane. I wanted to see if the type of paper used, such as thin paper like printer paper or thick paper like card stock, would affect how far the plane would go. Similarly, the style of the wing may have an important effect on the plane. Typically, most paper airplanes are folded with rectangular shaped wings or triangular shaped wings. Next, I wanted to see the effect weight had, and if adding a pen inside the folding would affect the distance flown. Lastly, I wanted to see how the length of the plane would affect the distance flown and tested a standard length of paper versus paper with an inch cut off.

Data Collection:

Since my experiment is 2^4 factorial design, I had a total of 16 combinations. For each of the 16 different combinations of paper airplane I chose to do 3 trials. I found a room that would be long enough to properly throw the plane. For throwing the plane, I held the paper airplane next to my ear in the same location for each throw, and let my arm fall forward from gravity with trying to limit the amount of additional force of pushing with my arm. To begin, I measured out 12 feet (144 inches). Typically, a paper airplane can be thrown further than 144 inches, but for the purpose of accuracy in my experiment I wanted to limit the throwing motion of my arms by letting them just fall, which limited to plane distance to 144 inches. After the plane landed, I would measure the furthermost point of the plane. These measurements ranged anywhere from 40 inches to 140 inches. I chose to put the measurements in inches because numbers worked better mathematically if left in inches rather than feet. There is a lot of error that can go into this. First, since the plane is thrown by a human, there is a lot of error in the fact that no two throwing motions will be exactly the same. Second, my experiment took place in a carpeted room, which means when the plane hits the ground it wouldn't slide on the ground that far. However, if this experiment was done in a room with a wooden floor, a paper airplane would be able to slide further on the floor once it landed and the end distance could be much further. Some other factors that can change the accuracy is if there's a draft in the room that could add tailwind to the plane, or if each plane was folded to the same level of care. Any small part of paper that could be sticking up could affect air resistance. With any experiment there will always be error, especially when a lot of human use is involved. Improvements to the experiment will be discussed in the conclusion.

Data Collection Table:

paper	wing	weight	length	y1	y2	y3	y_tot
-	-	-	-	85	80	86	83.66667
+	-	-	-	96	112	93	100.33333
-	+	-	-	141	139	137	139.00000
+	+	-	-	33	58	55	48.66667
-	-	+	-	73	85	87	81.66667
+	-	+	-	94	84	90	89.33333
-	+	+	-	139	130	126	131.66667
+	+	+	-	138	116	109	121.00000
-	-	-	+	113	120	114	115.66667
+	-	-	+	84	101	117	100.66667
-	+	-	+	67	58	58	61.00000
+	+	-	+	32	47	66	48.33333
-	-	+	+	100	119	88	102.33333
+	-	+	+	114	109	102	108.33333
-	+	+	+	90	64	96	83.33333
+	+	+	+	104	112	111	109.00000

The four factors and the interactions were assigned through random factorial design. Y1,Y2, and Y3 are the results from the distance flown in inches for each trial, with y_tot being the mean of the 3 trials.

Modeling the Data:

In general, a factorial design is represented by a regression model:

$$z_i = \beta_0 + \sum_{j=1}^{7} \beta_j x_{ij} + \varepsilon_i$$

The first beta value is the intercept of the model, and each succeeding beta is the coefficient estimate for each main effect and interaction effect value. The relationship between the beta coefficient values and the main/interaction effects is the beta coefficient is equal to ½ of the main/interaction effect value.

$$\hat{\beta}_j = \frac{1}{1-(-1)}(\bar{z}(x_{ij}=+1) - \bar{z}(x_{ij}=-1))$$
$$= \tfrac{1}{2}(\text{factorial effect of variable } x_j)$$

The regression output is as follows:

```
                                Estimate Std. Error t value Pr(>|t|)    
(Intercept)                      95.2500     1.5537  61.304  < 2e-16 ***
paper_1                           4.5417     1.5537   2.923 0.006315 ** 
wing_1                            2.5000     1.5537   1.609 0.117431    
weight_1                         -8.0833     1.5537  -5.203 1.10e-05 ***
length_1                          4.1667     1.5537   2.682 0.011487 *  
paper_1:wing_1                   -6.4583     1.5537  -4.157 0.000225 ***
paper_1:weight_1                  8.1250     1.5537   5.229 1.02e-05 ***
wing_1:weight_1                  10.4167     1.5537   6.704 1.44e-07 ***
paper_1:length_1                  5.0417     1.5537   3.245 0.002752 ** 
wing_1:length_1                 -13.1667     1.5537  -8.474 1.10e-09 ***
weight_1:length_1                 1.5833     1.5537   1.019 0.315819    
paper_1:wing_1:weight_1          -6.6250     1.5537  -4.264 0.000166 ***
paper_1:wing_1:length_1          -9.2083     1.5537  -5.927 1.34e-06 ***
paper_1:weight_1:length_1         0.7083     1.5537   0.456 0.651545    
wing_1:weight_1:length_1         -0.6667     1.5537  -0.429 0.670743    
paper_1:wing_1:weight_1:length_1 -4.4583     1.5537  -2.869 0.007227 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

From the output, the interaction effects that are significant are paper*wing, paper*weight, wing*weight, paper*length, wing*length, paper*wing*weight, paper*wing*length, and paper*wing*weight*length. That’s 5 2-factor interaction, 2 3-factor interaction, and 1 4-factor interaction. I used p-value < .1 since my experiment has a lot of human errors and I wanted to test how all the factors effected the distance.

While the main effect wing is not significant by itself, it is still included in my model since weight is important to specific interactions.

Important factors:
Paper = +
Wing = +
Weight = -
Length = +

From the 4 main effects, and the significant interaction effects my model reduces to:

Distance = 95.25 + 4.5417*paper + 2.5*wing + -8.0833*weight + 4.1667*length + -6.4583*paper:wing + 8.125*paper:weight + 10.4167*wing:weight + 5.0417*paper:length + -13.1667*wing:length + -6.6250*paper:wing:weight + -9.2083*paper:wing:length + -4.4583*paper:wing:weight:length

Main effects (+,-,+,+) that maximize distance with interaction effects, you get Distance = 136.0417 inches

The importance of interaction effects is further displayed in an ANOVA table:

```
Response: yvec
                              Df Sum Sq Mean Sq F value    Pr(>F)    
paper_                         1  990.1   990.1  8.5444 0.0063151 ** 
wing_                          1  300.0   300.0  2.5890 0.1174312    
weight_                        1 3136.3  3136.3 27.0665 1.102e-05 ***
length_                        1  833.3   833.3  7.1917 0.0114865 *  
paper_:wing_                   1 2002.1  2002.1 17.2780 0.0002251 ***
paper_:weight_                 1 3168.7  3168.7 27.3463 1.020e-05 ***
wing_:weight_                  1 5208.3  5208.3 44.9479 1.436e-07 ***
paper_:length_                 1 1220.1  1220.1 10.5293 0.0027522 ** 
wing_:length_                  1 8321.3  8321.3 71.8130 1.104e-09 ***
weight_:length_                1  120.3   120.3  1.0385 0.3158194    
paper_:wing_:weight_           1 2106.7  2106.7 18.1812 0.0001659 ***
paper_:wing_:length_           1 4070.1  4070.1 35.1248 1.342e-06 ***
paper_:weight_:length_         1   24.1    24.1  0.2078 0.6515454    
wing_:weight_:length_          1   21.3    21.3  0.1841 0.6707427    
paper_:wing_:weight_:length_   1  954.1   954.1  8.2337 0.0072271 ** 
Residuals                     32 3708.0   115.9                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Inference:

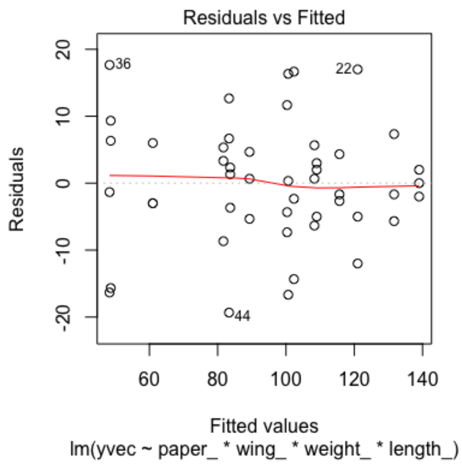
From the raw data, without doing any statistical analysis the plane that went the furthest for me on average was plane #3 (-,+,-,-) with an average distance of 139 inches. This plane was made out of printer paper, rectangular wings, no pen, and was the shorter version in length. The purpose of my experiment was to find the plane that went the furthest. When plane #3 (-,+,-,-) was put into my model it got a projected distance of 108.12, which is a good distance. However, once all 16 plane combinations were entered into my model, the combination that got the best distance was (+,-,+,+) which has a plane made out of card stock, triangular wings, a pen inside, and the longer version. When this was put through my model it got a projected distance of 136 inches.

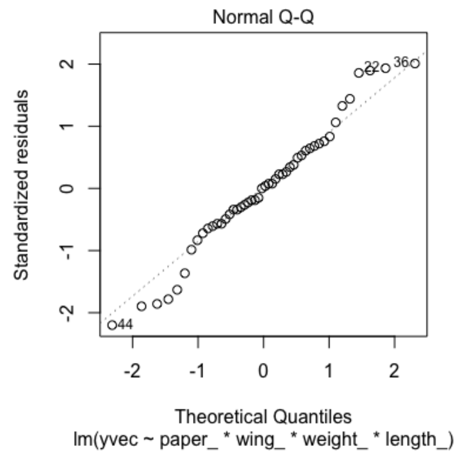
The statistical analysis tells you how to make a good paper airplane. Weight has a lot to do with how long in the air it will glide for. From my experiment, I recommend making the paper airplane out of a sturdier type of paper. Adding some weight can increase distance, but if you add too much the plane could just nose dive straight into the ground. For longer planes, you should use triangular wings.

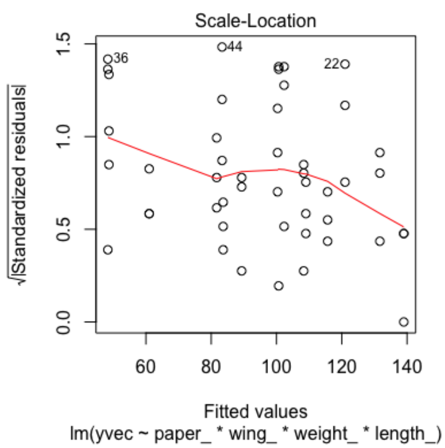
If this experiment were to be recreated, some improvements I would focus on would be a consistent way to throw the planes. Some form of machinery that could consistently throw each plane the same exact way would help take out error from a human throwing the plane. Also, for the weight added, trying different weights to find the most optimal weight for distance traveled could make for a better plane.

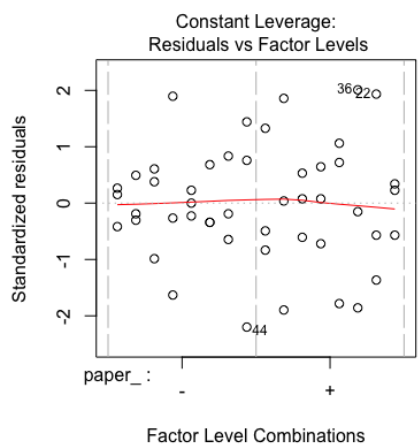
My final conclusion for the best paper airplane is a plane made out of sturdier paper (card stock), with triangular wings, an added weight (pen in the folding), and a standard length of paper used for the plane creation.

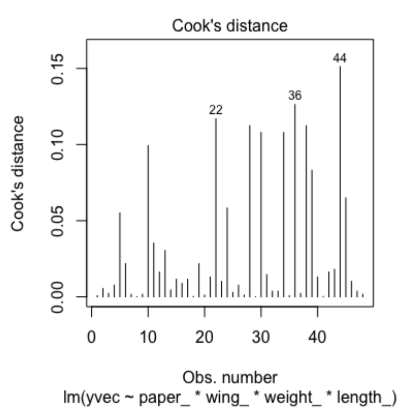
R analysis:


Residuals vs Fitted
Residuals
Fitted values
lm(yvec ~ paper_ * wing_ * weight_ * length_)
36
22
44
20
10
0
-10
-20
60
80
100
120
140


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(yvec ~ paper_ * wing_ * weight_ * length_)
22
36
44
2
1
0
-1
-2

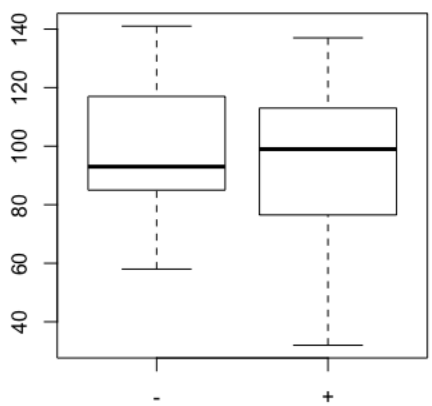

Scale-Location
√|Standardized residuals|
Fitted values
lm(yvec ~ paper_ * wing_ * weight_ * length_)
36
44
22
1.5
1.0
0.5
0.0
60
80
100
120
140


Constant Leverage:
Residuals vs Factor Levels
Standardized residuals
Factor Level Combinations
paper_ :
-
+
36
22
44
2
1
0
-1
-2

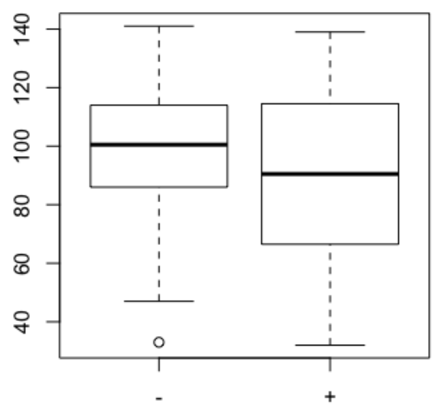

Cook's distance
Cook's distance
Obs. number
lm(yvec ~ paper_ * wing_ * weight_ * length_)
22
36
44
0.15
0.10
0.05
0.00
0
10
20
30
40

From the above residual plots and QQ plots, normality and linearity are held decently well. From the Cook's distance plot, you can tell there were many trials that had a lot of error. This error most likely comes from human error of how the plane was thrown. A more consistent throwing method would most likely fix the Cook's distance plot.

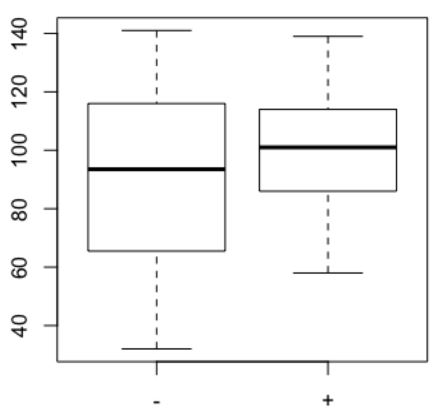
Main Effect Box Plot Paper vs Distance:



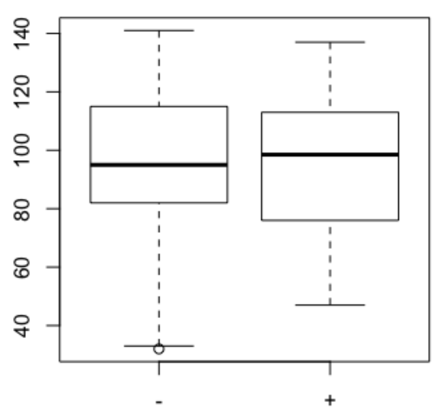
Main Effect Box Plot Wing vs Distance:



Main Effect Box Plot Weight vs Distance:

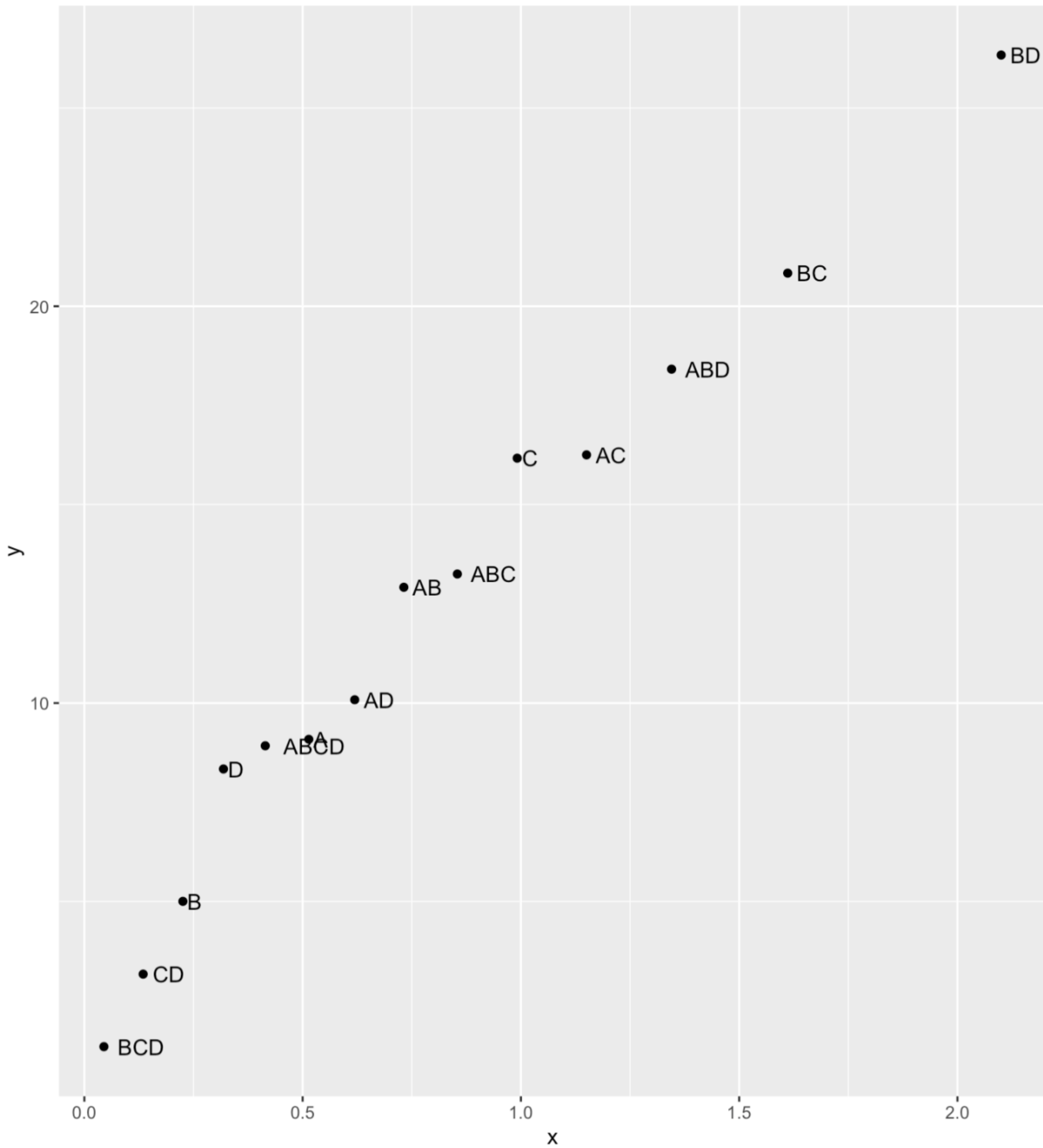


Main Effect Box Plot Length vs Distance:



These main effect box plots prove again that the best combination in the order of paper, wing, weight, and length is (+,-,+,+). This gets you a predicted distance of over 136 inches.

Half-Normal Plot for Factorial Effects:



Paper = A
Wing = B
Weight = C
Length = D

From the Half-Normal plot you can see how each factorial effect affects the model. For example, BD (Wing:Length) appears to have the most significant effect on the model, and from the Regression output the p-value for Wing:Length is $1.1 \times e^{-9}$, which proves it is in fact the most significant factorial effect. The higher the Y value on the half-normal plot, the more influence on Distance flow the factorial effect had.

Appendix:

Notes from Stats 424 and Stats 333

Compressed R script:

```
Library(ggplot2)
y1 = c(85,96,141,33,73,94,139,138,113,84,67,32,100,114,90,104)
y2 = c(80,112,139,58,85,84,130,116,120,101,58,47,119,109,64,112)
y3 = c(86,93,137,55,87,90,126,109,114,117,58,66,88,102,96,111)
paper = rep(c('-','+'),8)
wing = rep(c('-','-','+','+'),4)
weight = rep(c('-','-','-','-','+','+','+','+'),2)
length = c(rep('-',8),rep('+',8))

y_tot = c()
for (i in 1:16){
  y_tot[i] = (y1[i]+y2[i]+y3[i])/3
}

paper_ = rep(as.factor(paper), rep(3,16))
wing_ = rep(as.factor(wing), rep(3,16))
weight_ = rep(as.factor(weight), rep(3,16))
length_ = rep(as.factor(length), rep(3,16))

ydat = data.frame(y1,y2,y3)
ydat

data = data.frame(paper,wing,weight,length,y1,y2,y3,y_tot)
data

yvec = c(t(ydat))

model = lm(yvec~paper_*wing_*weight_*length_)
summary(model)
anova(model)
plot(model)
#plot(MainEffect, c(y1,y2,y3))


FourWayFE = function(Resp){
  A = rep(c(-1,1),8)
  B = rep(c(-1,-1,1,1),4)
  C = rep(c(-1,-1,-1,-1,1,1,1,1),2)
  D = c(rep(-1,8),rep(1,8))
  FE = c(sum(A * Resp)/8,
       sum(B * Resp)/8,
       sum(C * Resp)/8,
       sum(D * Resp)/8,
       sum(A * B * Resp)/8,
       sum(A * C * Resp)/8,
       sum(A * D * Resp)/8,
       sum(B * C * Resp)/8,
       sum(B * D * Resp)/8,
       sum(C * D * Resp)/8,
       sum(A * B * C * Resp)/8,
       sum(A * B * D * Resp)/8,
       sum(B * C * D * Resp)/8,
```

```
        sum(A * B * C * D * Resp)/8)
  FEID = c("A","B","C","D","AB","AC","AD","BC","BD","CD","ABC","ABD","BCD","ABCD")
  return(list(FE,FEID))
}

HNplot = function(FactEff,FactId)
{
  #Input = Vector
  n = length(FactEff)
  ind = order(abs(FactEff))
  y = abs(FactEff[ind])
  x = qnorm(0.5+0.5*(1:n-0.5)/n)
  Names = FactId[ind]
  dat = data.frame(x,y,names = FactId)
  ggplot(dat,aes(x,y))+geom_point()+geom_text(aes(label = Names),hjust=-.3)
}

FE = FourWayFE(y_tot)
FE[1]
FE[2]

HNplot(FE[[1]],FE[[2]])
```