STA305/1004-Class 23

March 28, 2016

Today's Class

2^2 design

Run	1	2	12	Y
1	$-$	$-$	$+$	Y_1
2	$+$	$-$	$-$	Y_2
3	$-$	$+$	$-$	Y_3
4	$+$	$+$	$+$	Y_4

- Assessing significance in unreplicated factorial designs
  - Normal plots
  - half-Normal plots
  - Lenth's method (wednesday)

main effect of 1

if no replication $\frac{Y_2+Y_4-(Y_1+Y_3)}{2}$

Not possible to est. the standard error of factorial effects

Factorial Assignment

HW #4

- Read the sample report.
- You are supposed to design an experiment using a factorial design.
- This means I want you to collect the data. So finding data (e.g., on the web) is not appropriate.
- What are the controllable input variables (factors) in your experiment? What is the response variable?
- Example: How does coffee consumption and hours of sleep affect running speed?

Coffee = $\begin{cases} 1 \text{ /day} & (-) \\ >1\text{/day} & (+) \end{cases}$

hours sleep = $\begin{cases} \geq 8 \text{ hours} & (+) \\ < 8 \text{ hours} & (-) \end{cases}$

2^2

Coffee	Sleep	Running Speed (s)
−	−	
+	−	
−	+	
+	+	

Running Speed: Run 100 m on specific track

Sleep effect $= \dfrac{(y_3+y_4)-(y_1+y_2)}{2} = 0$

Coffee effect $= \dfrac{(y_2+y_4)-(y_1+y_3)}{2} = 0$

Quantile-Quantile Plots

plot quantiles of one set of numbers vs. other set of numbers

if straight line then same dist.

- Quantile-quantile (Q-Q) plots are useful for comparing distribution functions.
- If X is a continuous random variable with strictly increasing distribution function $F(x)$ then the pth quantile of the distribution is the value of x_p such that,

$$F(x_p) = p$$

or

$$x_p = F^{-1}(p).$$

- In a Q-Q plot, the quantiles of one distribution are plotted against another distribution.
- Q-Q plots can be used to investigate if a set of numbers follows a certain distribution.

Quantile-Quantile Plots

- Suppose that we have observations independent observations $X_1, X_2, ..., X_n$ from a uniform distribution on $[0, 1]$ or Unif[0,1].
- The ordered sample values (also called the order statistics) are the values $X_{(j)}$ such that

$$\underbrace{X_{(1)}}_{\text{min}} < X_{(2)} < \cdots < \underbrace{X_{(n)}}_{\text{max}}$$

.

- It can be shown that

$$E\left(X_{(j)}\right) = \frac{j}{n+1}.$$

$$E(X_{(1)}) = \frac{1}{n+1}$$
$$E(X_{(n)}) = \frac{n}{n+1}$$

- This suggests that if we plot

$$X_{(j)} \quad \text{vs.} \quad \frac{j}{n+1}$$

then if the underlying distribution is Unif[0,1] then the plot should be rooughly linear.

Quantile-Quantile Plots

- A continuous random variable with strictly increasing CDF F_X can be transformed to a Unif[0,1] by defining a new random variable $Y = F_X(X)$. (PIT)
- Suppose that it's hypothesized that X follows a certain distribution function with CDF F.
- Given a sample $X_1, X_2, ..., X_n$ plot

unif[0,1]

$$F(X_{(k)}) \quad \text{vs.} \quad \frac{k}{n+1}$$

or equivalently

observed quantiles

$$X_{(k)} \quad \text{vs.} \quad F^{-1}\left(\frac{k}{n+1}\right)$$

- $X_{(k)}$ can be thought of as empirical quantiles and $F^{-1}\left(\frac{k}{n+1}\right)$ as the hypothesized quantiles.
- The quantile assigned to $X_{(k)}$ is not unique.
- Instead of assigning it $\frac{k}{n+1}$ it is often assigned $\frac{k-0.5}{n}$. In practice it makes little difference which definition is used.

easier to interpret this definition

prob. integral transformation

If a random var. X has cdf F_X, then $Y = F_X(X)$ has unif[0,1]

$F_Y(y) = P(Y \le y) = P(F_X(X) \le y)$

$= P(X \le F^{-1}(y))$

$= F(F^{-1}(y))$

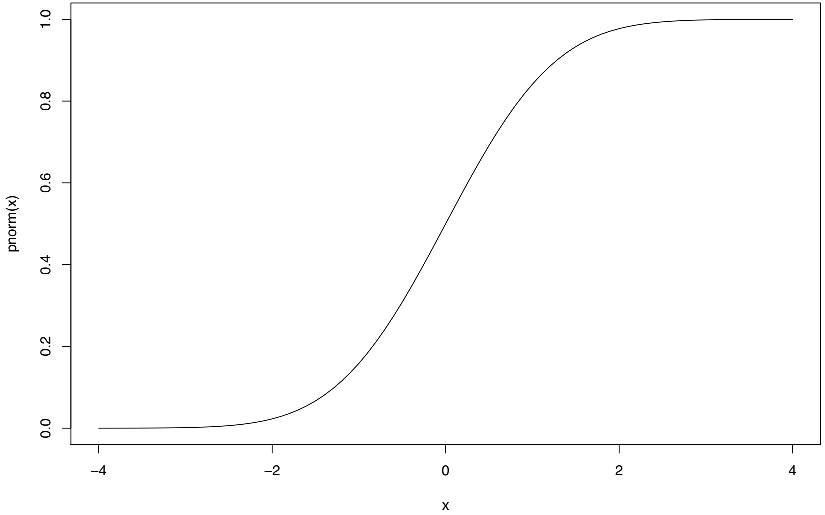
$= y$

This is the CDF of Y which is unif[0,1]

Normal Quantile-Quantile Plots

The cumulative distribution function (CDF) of the normal has an S-shape.

```
x <- seq(-4,4,by=0.1)
plot(x,pnorm(x),type="l")
```



Normal Quantile-Quantile Plots

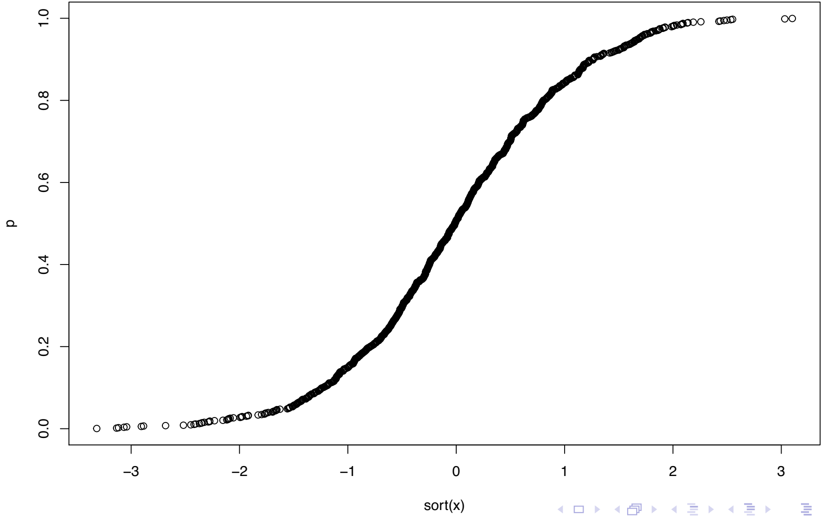
The normality of a set of data can be assessed by the following method.

- Let $r_{(1)} < \ldots < r_{(N)}$ denote the ordered values of $r_1, \ldots, r_N$.
- A test of normality for a set of data is to plot the ordered values $r_{(i)}$ of the data versus $p_i = (i - 0.5)/N$.
- If the plot has the same S-shape as the normal CDF then this is evidence that the data come from a normal distribution.

Normal Quantile-Quantile Plots

- A plot of $r_{(i)}$ vs. $p_i = (i - 0.5)/N, i = 1, \dots, N$ for a random sample of 1000 simulated from a $N(0, 1)$.

```
N <- 1000;x <- rnorm(N);p <- ((1:N)-0.5)/N
plot(sort(x),p)
```



Normal Quantile-Quantile Plots

- It can be shown that $\Phi(r_i)$ has a uniform distribution on $[0, 1]$.
- This implies that $E(\Phi(r_{(i)})) = i/(N+1)$ (this is the expected value of the *jth* order statistic from a uniform distribution over $[0, 1]$.
- This implies that the N points $(p_i, \Phi(r_{(i)}))$ should fall on a straight line.
- Now apply the Φ^{-1} transformation to the horizontal and vertical scales. The N points

  $$\left(\Phi^{-1}(p_i), r_{(i)}\right),$$

  form the normal probability plot of $r_1, ..., r_N$.
- If $r_1, ..., r_N$ are generated from a normal distribution then a plot of the points $\left(\Phi^{-1}(p_i), r_{(i)}\right)$, $i = 1, ..., N$ should be a straight line.

Normal Quantile-Quantile Plots

In R qnorm() is Φ^{-1}.

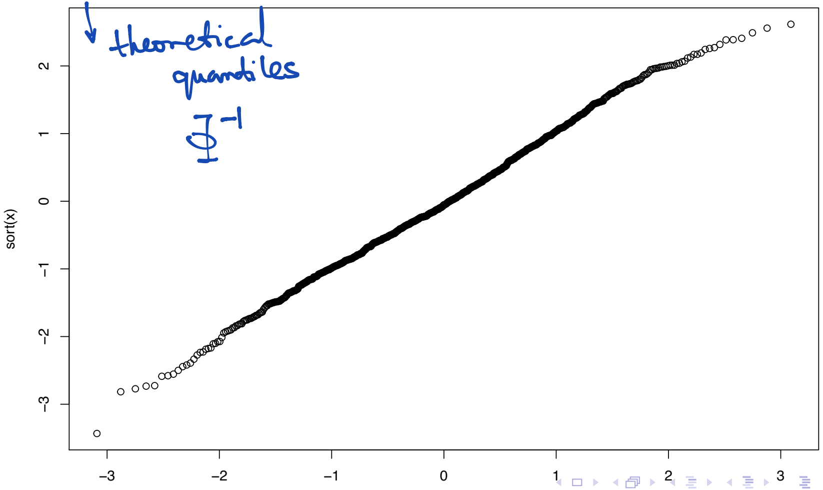
```
set.seed(2503)
N <- 1000
x <- rnorm(N)
p <- (1:N)/(N+1)
plot(qnorm(p),sort(x))
```

setting random number generator

$N(0,1)$

$\frac{i}{N+1}$

theoretical quantiles Φ^{-1}



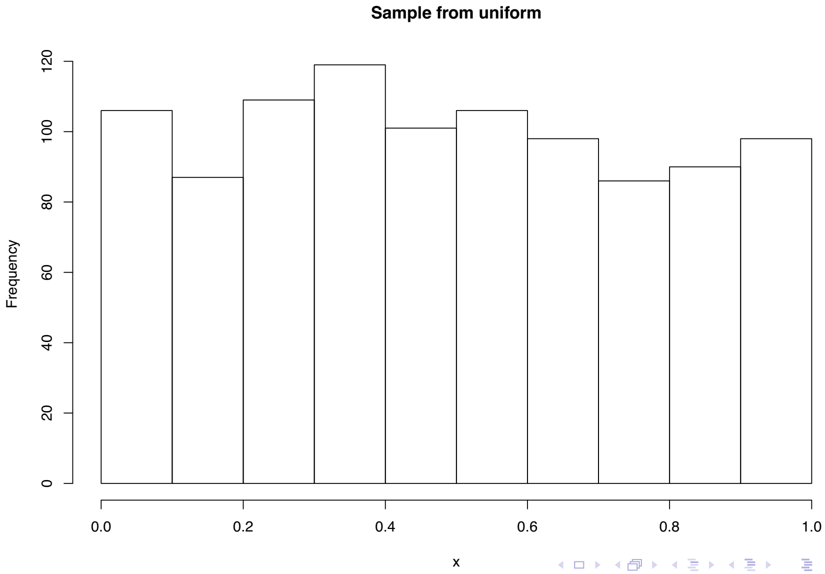
Normal Quantile-Quantile Plots

A marked (systematic) deviation of the plot from the straight line would indicate that:

1. The normality assumption does not hold.
2. The variance is not constant.

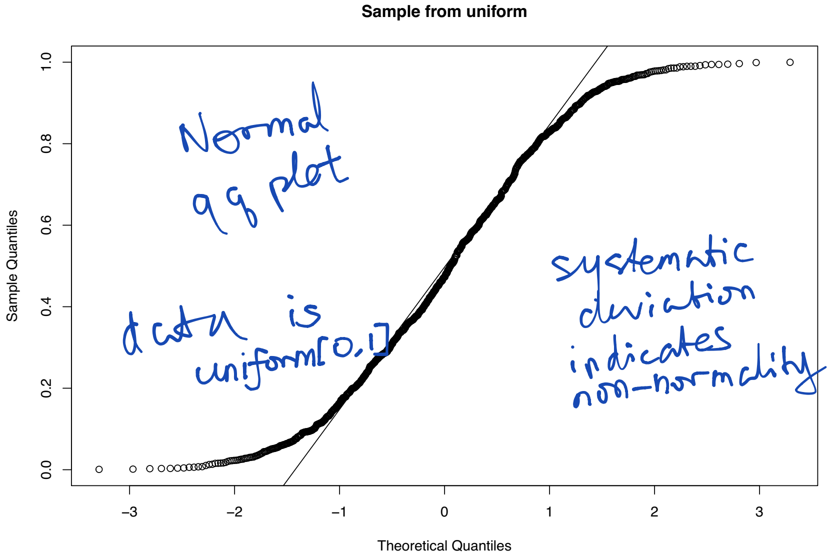
Normal Quantile-Quantile Plots

```
x <- runif(1000)
hist(x,main = "Sample from uniform")
```



Normal Quantile-Quantile Plots

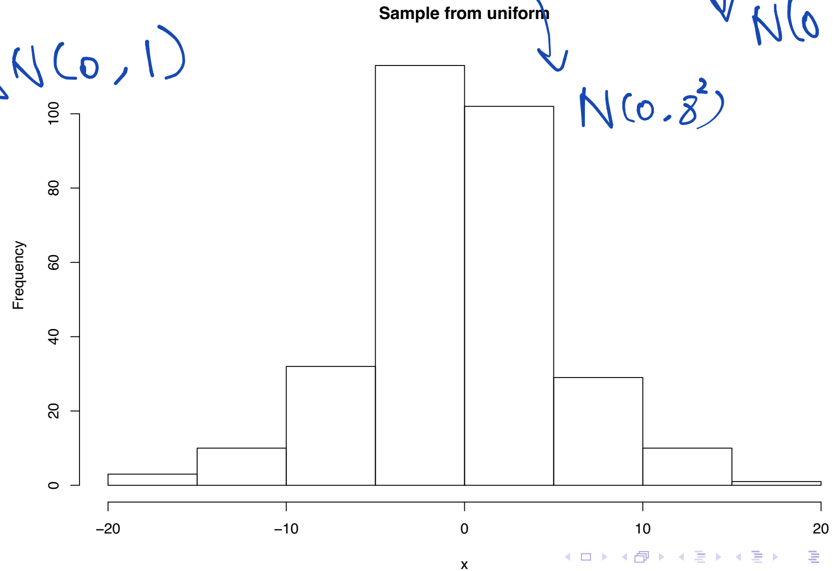
```
qqnorm(x,main = "Sample from uniform");qqline(x)
```



Normal Quantile-Quantile Plots

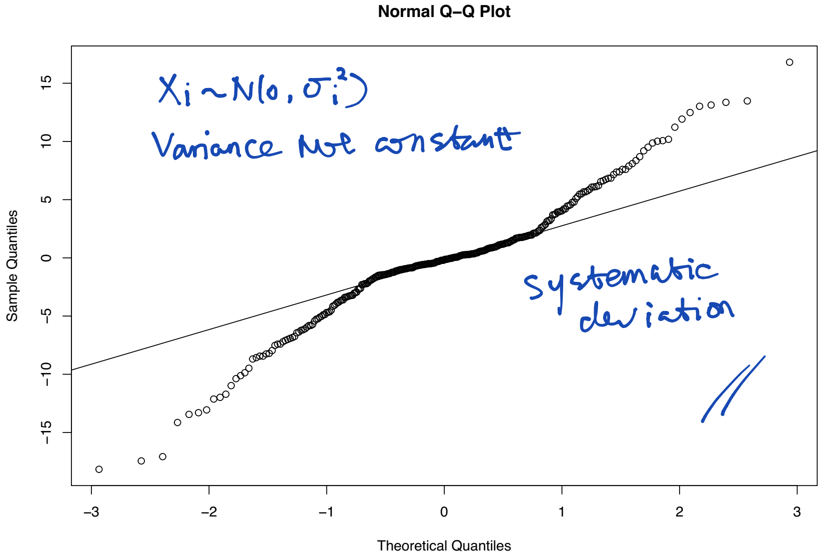
```
x1 <- rnorm(100,mean = 0,sd = 1);x2 <- rnorm(100,mean = 0,sd = 5)
x3 <- rnorm(100,mean = 0,sd = 8); x <- c(x1,x2,x3)
hist(x,main = "Sample from uniform")
```

$N(0,1)$ $N(0,8^2)$ $N(0,5^2)$



Normal Quantile-Quantile Plots

```
qqnorm(x);qqline(x)
```



Normal plots in factorial experiments

main effects & interactions

- A major application is in factorial designs where the $r(i)$ are replaced by ordered factorial effects.
- Let $\hat{\theta}_{(1)} < \hat{\theta}_{(2)} < \cdots < \hat{\theta}_{(N)}$ be N ordered factorial estimates.
- If we plot

$$\hat{\theta}_{(i)} \text{ vs. } \Phi^{-1}(p_i).\ i = 1, \ldots, N.$$

then factorial effects $\hat{\theta}_i$ that are close to 0 will fall along a straight line. Therefore, points that fall off the straight line will be declared significant.

N(0,1) cdf

Normal plots in factorial experiments

factorial effect = 0

The rationale is as follows:

1. Assume that the estimated effects $\hat{\theta}_i$ are $N(\theta, \sigma)$ (estimated effects involve averaging of N observations and CLT ensures averages are nearly normal for N as small as 8).
2. If $H_0 : \theta_i = 0$, $i = 1, ..., N$ is true then all the estimated effects will be zero.
3. The resulting normal probability plot of the estimated effects will be a straight line.
4. Therefore, the normal probability plot is testing whether all of the estimated effects have the same distribution (i.e. same means).

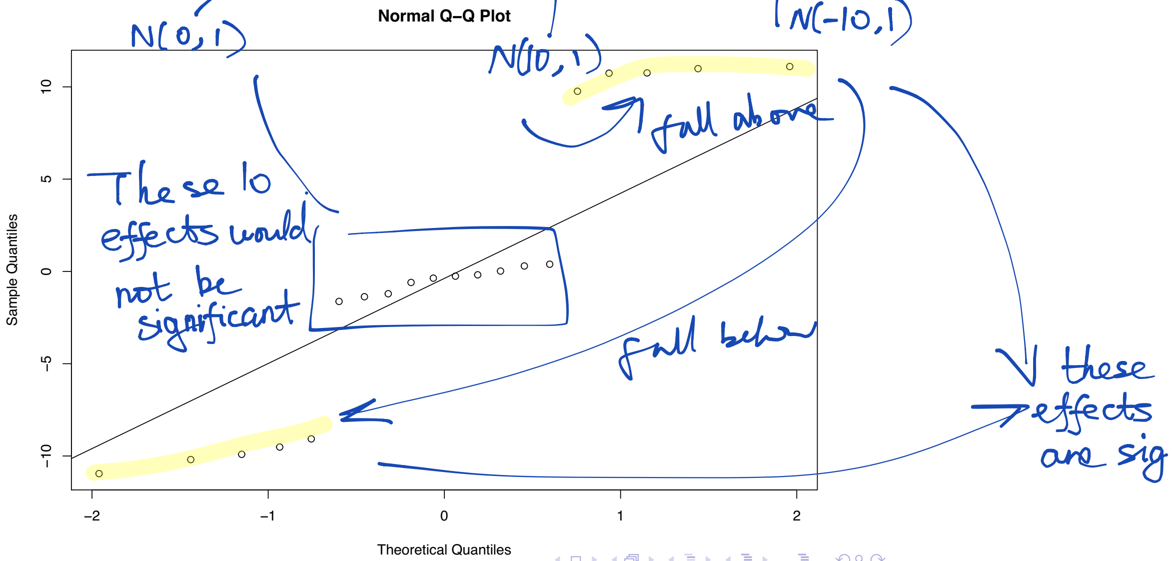
- When some of the effects are nonzero the corresponding estimated effects will tend to be larger and fall off the straight line.

– when smaller than zero will fall below the line

Normal plots in factorial experiments

Positive effects fall above the line and negative effects fall below the line.

```
set.seed(10);x1 <- rnorm(10,0,1); x2 <- rnorm(5,10,1);x3 <- rnorm(5,-10,1)
x <- c(x1,x2,x3);qqnorm(x);qqline(x)
```



Example - 2^4 design for studying a chemical reaction

x1	x2	x3	x4	conversion
-1	-1	-1	-1	70
1	-1	-1	-1	60
-1	1	-1	-1	89
1	1	-1	-1	81
-1	-1	1	-1	69
1	-1	1	-1	62
-1	1	1	-1	88
1	1	1	-1	81
-1	-1	-1	1	60
1	-1	-1	1	49
-1	1	-1	1	88
1	1	-1	1	82
-1	-1	1	1	60
1	-1	1	1	52
-1	1	1	1	86
1	1	1	1	79

The design is not replicated so it's not possible to estimate the standard errors of the factorial effects.

Example - 2^4 design for studying a chemical reaction

```
fact1 <- lm(conversion~x1*x2*x3*x4,data=tab0510a)
round(2*fact1$coefficients,2)
```

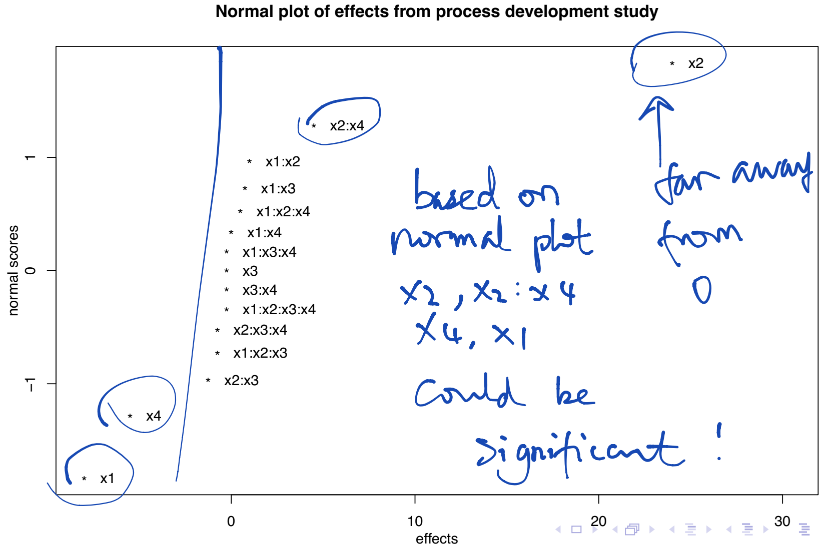
multi x2

```
(Intercept)          x1          x2          x3          x4       x1:x2
     144.50       -8.00       24.00       -0.25       -5.50        1.00
      x1:x3       x2:x3       x1:x4       x2:x4       x3:x4    x1:x2:x3
       0.75       -1.25        0.00        4.50       -0.25       -0.75
   x1:x2:x4    x1:x3:x4    x2:x3:x4 x1:x2:x3:x4
       0.50       -0.25       -0.75       -0.25
```

Example - 2^4 design for studying a chemical reaction

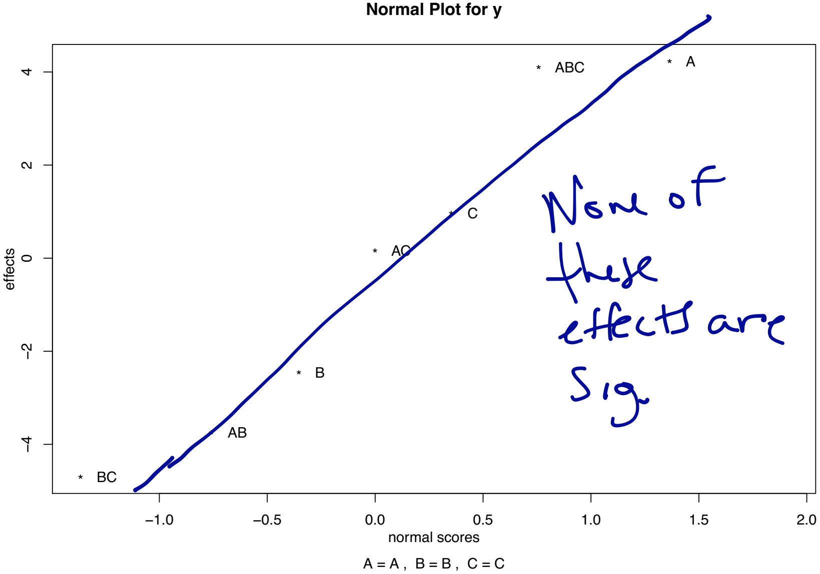
A normal plot of the factorial effects is obtained by using the function `DanielPlot()` in the `FrF2` library.

```r
library(FrF2)
DanielPlot(fact1,half=FALSE,autolab=F, main="Normal plot of effects from process development study")
```

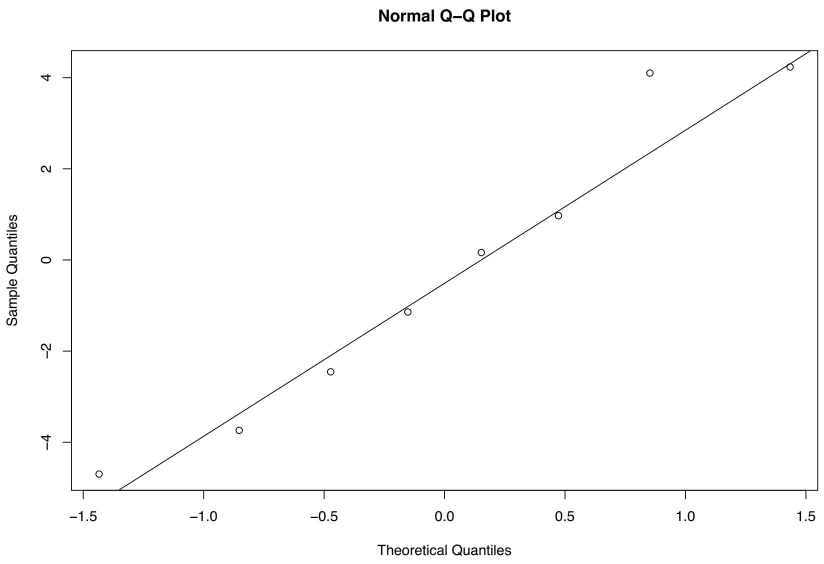


Question

Which effects are not explained by chance according to the normal plot?



Question


Normal Q–Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

Half-Normal Plots

- A related graphical method is called the half-normal probability plot.
- Let
$$\left|\hat{\theta}\right|_{(1)} < \left|\hat{\theta}\right|_{(2)} < \cdots < \left|\hat{\theta}\right|_{(N)} .$$
denote the ordered values of the unsigned factorial effect estimates.
- Plot them against the coordinates based on the half-normal distribution - the absolute value of a normal random variable has a half-normal distribution.
- The half-normal probability plot consists of the points

$$\left|\hat{\theta}\right|_{(i)} \text{ vs. } \Phi^{-1}(0.5 + 0.5[i - 0.5]/N).\ i = 1, ..., N.$$

Half-Normal Plots

- An advantage of this plot is that all the large estimated effects appear in the upper right hand corner and fall above the line.
- The half-normal plot for the effects in the process development example is can be obtained with `DanielPlot()` with the option `half=TRUE`.

Half-Normal Plots

```
library(FrF2)
DanielPlot(fact1,half=TRUE,autolab=F,
           main="Normal plot of effects from process development study")
```