STAT5303HW1

P1.2

The number of experimental units is 4.Because an experimental unit should be able to receive any treatment. In this experiment , each organ as an experimental unit is randomly assigned a different treatment. The twelve is the number of measurement units

P1.3

The good thing is that the school district can choose the most suitable set of standards according to its real situation and desire. The bad thing is that all school districts may choose the same one of the two sets of standards, and no school district chooses the other. In addition, this experiment takes 10 years, which is too long.

E2.5

```r
library(cfcdae)
```

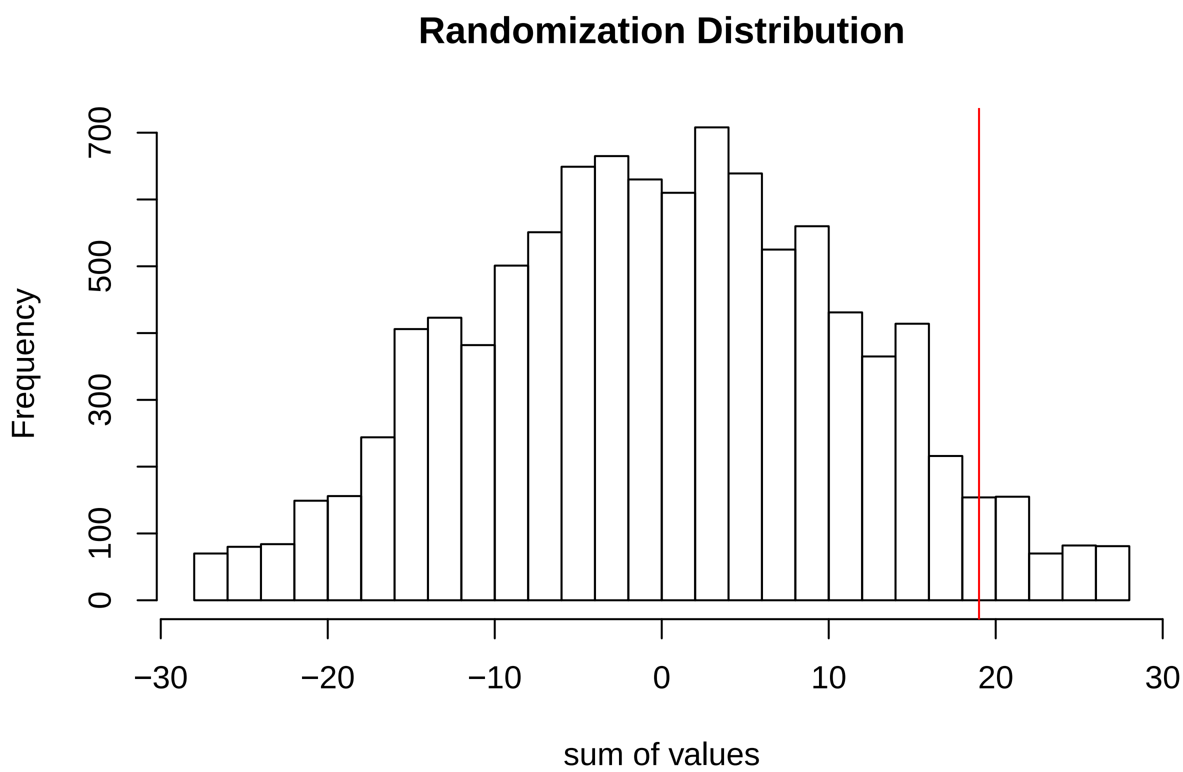
```
## Warning: package 'cfcdae' was built under R version 4.0.2
```

```
## Registered S3 method overwritten by 'DoE.base':
##   method           from
##   factorize.factor conf.design
```

```r
data("CalfCounts")
head(CalfCounts)
```

```
##   differences
## 1          -1
## 2           6
## 3           4
## 4           6
## 5           2
## 6          -3
```

```r
permsign.test(CalfCounts[,"differences"],plot=TRUE)
```




```
## $x0
## [1] 19
##
## $lower.p
## [1] 0.9612
##
## $upper.p
## [1] 0.0542
##
## $twosided.p
## [1] 0.1084
```

```
## As the alternative hypothesis is the drug improves fertility,p-value is 0.05.
```

P2.3

The standard p-value cutoff used in your field of study in order to declare some result "significant" is $\alpha = 0.05$.I think this cutoff is not always appropriate for the kinds of experiments conducted in my field.Like,if we want to reduce the Type II error, choosing a higher values of α is better bacause higher values of α will make it easier to reject the null hypothesis.However,it also makes Type I error laregr.

P2.5

Since 95% is not the probability,the statement on this result shoul be that we are 95% confident that the mean response lies within the interval form 1.73 to 2.11.

E3.1

```r
library(cfcdae)
data("RatLiverWeight")
head(RatLiverWeight)
```

```
##   weight diet
```

```
## 1   3.52    1
## 2   3.36    1
## 3   3.57    1
## 4   4.19    1
## 5   3.88    1
## 6   3.76    1
```

```
attach(RatLiverWeight)
```

(a)

```
separate.means<-lm(weight~diet)
single.mean<-lm(weight~1,data = RatLiverWeight)
summary(single.mean)##the overall mean is 3.71828
```

```
##
## Call:
## lm(formula = weight ~ 1, data = RatLiverWeight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36828 -0.19828 -0.00828  0.15172  0.59172
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.71828    0.04456   83.44   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.24 on 28 degrees of freedom
```

```
model.effects(separate.means,"diet") ##for treatment effect
```

```
##           1           2           3           4
##  0.03407738 -0.13163690 -0.11330357  0.21086310
```

(b)

```
anova(separate.means)
```

```
## Analysis of Variance Table
##
## Response: weight
##           Df  Sum Sq  Mean Sq F value  Pr(>F)
## diet       3 0.57821 0.192736  4.6581 0.01016 *
## Residuals 25 1.03440 0.041376
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because the p-value 0.01016 is samller than 0.05, the difference between the four diets is statistically significant.

E3.5

The null hypothesis that delay after exposure does not affect leaflet angle

The alternative hypothesis that delay after exposure does affect leaflet angle

```
library(cfcdae)
data("Albizia")
head(Albizia)
```

```
##   delay.z angle delay
## 1      30   140    30
## 2      30   138    30
## 3      30   140    30
## 4      30   138    30
## 5      30   142    30
## 6      45   140    45
```

```
attach(Albizia)
mod<-lm(angle~delay.z)
summary(mod)
```

```
##
## Call:
## lm(formula = angle ~ delay.z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.867  -4.567  -1.867   3.133  18.133
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 157.6667     7.3177  21.546 1.48e-11 ***
## delay.z      -0.5733     0.1569  -3.654  0.00292 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.443 on 13 degrees of freedom
## Multiple R-squared:  0.5067, Adjusted R-squared:  0.4687
## F-statistic: 13.35 on 1 and 13 DF,  p-value: 0.002915
```

Because the p-value of angle 0.00292 is samller than 0.05, delay.z has a statistically significant effect on leaflet angle.So we reject the null hypothesis and get the evidence that delay after exposure does affect leaflet angle.

P3.1

```
library(cfcdae)
data("PacemakerPins")
head(PacemakerPins)
```

```
##   operator substrate strength
## 1        1         1     5.60
## 2        1         1     6.80
## 3        1         1     8.32
## 4        1         1     8.70
## 5        1         2     7.64
## 6        1         2     7.44
```

```
mod_1<-aov(strength~operator,data=PacemakerPins)
summary(mod_1)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## operator     3  15.19   5.063   4.243 0.0102 *
## Residuals   44  52.51   1.193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because the p-value 0.0102 is samller than 0.05, there is statistically significant difference between the four operators. So,it shows that the operators produce different mean shear strengths.

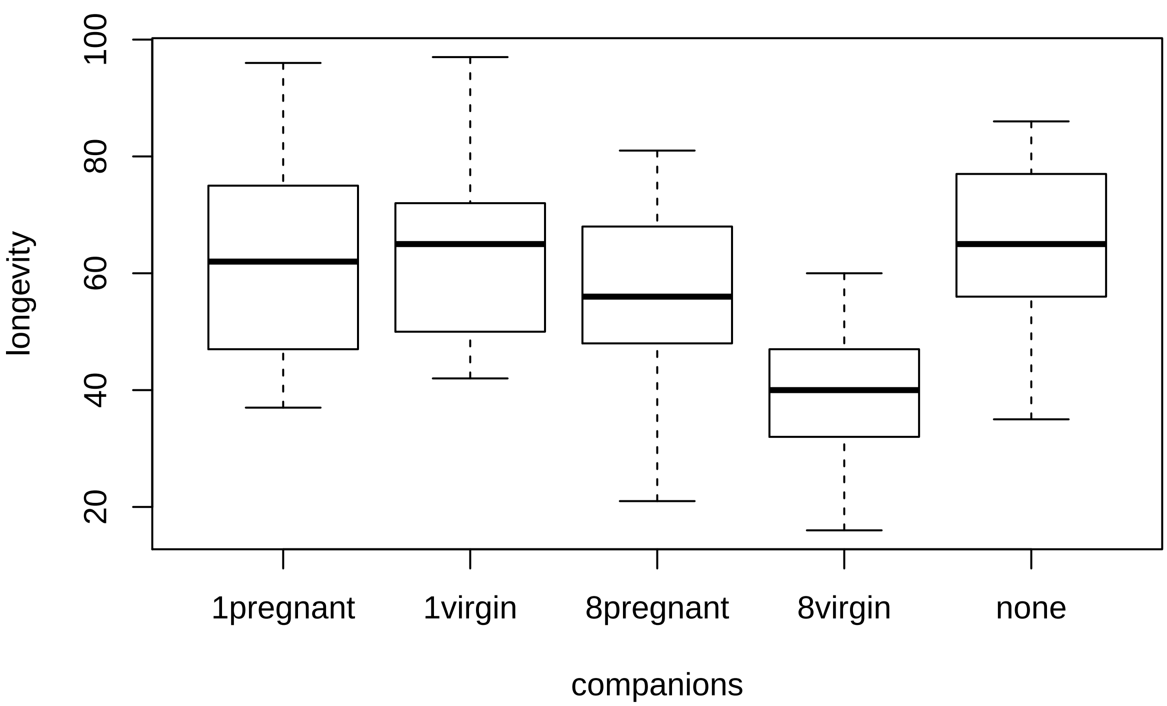
P3.2

```r
library(cfcdae)
data("FruitFlyLifespan")
head(FruitFlyLifespan)
```

```
##   companions longevity
## 1       none        35
## 2       none        37
## 3       none        49
## 4       none        46
## 5       none        63
## 6       none        39
```

```r
attach(FruitFlyLifespan)
```

```r
boxplot(longevity~companions)
```




```r
mod_2<-lm(longevity~companions,data = FruitFlyLifespan)
summary(mod_2);model.effects(mod_2,"companions")
```

```
##
## Call:
## lm(formula = longevity ~ companions, data = FruitFlyLifespan)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.76  -8.76   0.20  11.20  32.44
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   57.440      1.325  43.368  < 2e-16 ***
```

```
## companions1    6.120      2.649   2.310  0.02258 *
## companions2    7.360      2.649   2.778  0.00634 **
## companions3   -0.680      2.649  -0.257  0.79785
## companions4  -18.720      2.649  -7.067 1.13e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.81 on 120 degrees of freedom
## Multiple R-squared:  0.3121, Adjusted R-squared:  0.2892
## F-statistic: 13.61 on 4 and 120 DF,  p-value: 3.516e-09
```

```
## 1pregnant   1virgin 8pregnant   8virgin      none
##      6.12      7.36     -0.68    -18.72      5.92
```

```
anova(mod_2)
```

```
## Analysis of Variance Table
##
## Response: longevity
##             Df Sum Sq Mean Sq F value    Pr(>F)
## companions   4  11939 2984.82  13.612 3.516e-09 ***
## Residuals  120  26314  219.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

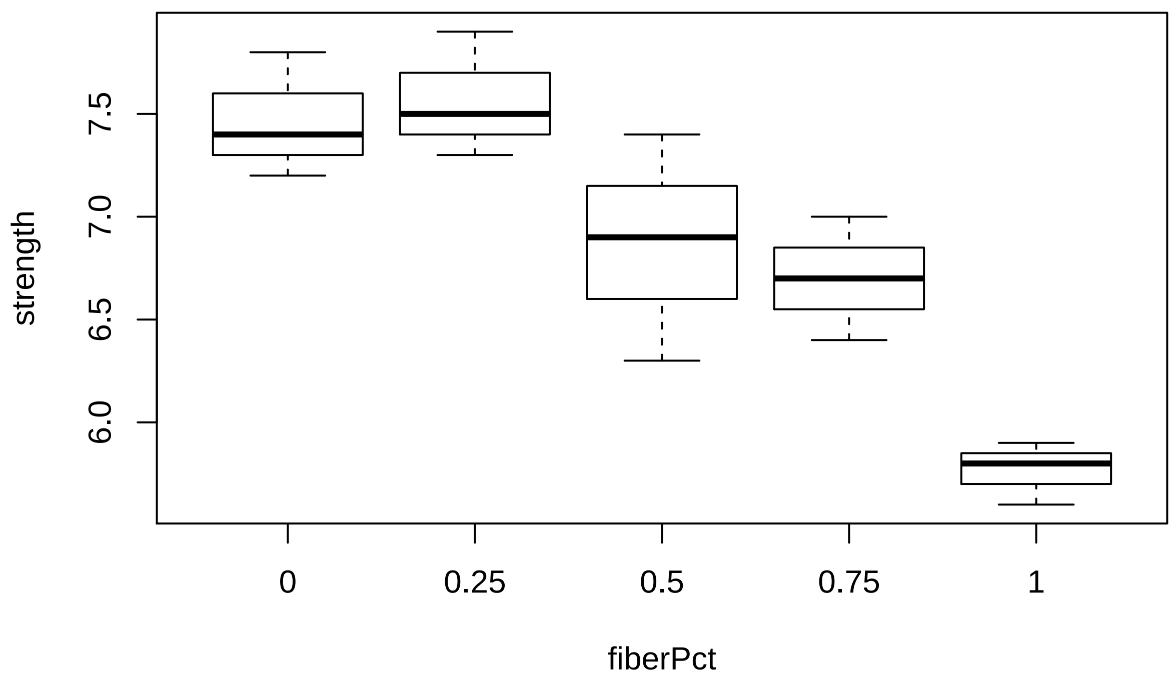
The boxplot shows that there is no significant diffrence of longevity within the levels of "1pregnant","1virgin","8pregnant"and"none".But apparently the level of "8virgin" has significant diffrence with other levels on longevity.Besides,by the LSE model,the p-value indicates that except "8pregnant",all the other levels has significant diffrence on longevity. The p-value in anova table also shows that there is significant diffrence between these level.Therefore,we reject the null hypothesis that reproductive activity does not affect longevity.

P3.7

```
data("ConcreteStrength")
head(ConcreteStrength)
```

```
##   fiberPct.z fiberPct strength
## 1       0.00        0      7.8
## 2       0.00        0      7.4
## 3       0.00        0      7.2
## 4       0.25     0.25      7.9
## 5       0.25     0.25      7.5
## 6       0.25     0.25      7.3
```

```
attach(ConcreteStrength)
boxplot(strength~fiberPct)
```




```
model<-lm(strength~fiberPct);
anova(model)
```

```
## Analysis of Variance Table
##
## Response: strength
##           Df Sum Sq Mean Sq F value    Pr(>F)
## fiberPct   4 6.2627 1.56567  12.975 0.0005713 ***
## Residuals 10 1.2067 0.12067
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model);model.effects(model,"fiberPct")
```

```
##
## Call:
## lm(formula = strength ~ fiberPct)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5667 -0.2167  0.0000  0.2167  0.5333
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.873333   0.089691  76.634  3.5e-15 ***
## fiberPct1    0.593333   0.179382   3.308  0.00791 **
## fiberPct2    0.693333   0.179382   3.865  0.00313 **
## fiberPct3   -0.006667   0.179382  -0.037  0.97109
## fiberPct4   -0.173333   0.179382  -0.966  0.35669
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3474 on 10 degrees of freedom
## Multiple R-squared:  0.8385, Adjusted R-squared:  0.7738
```

```
## F-statistic: 12.98 on 4 and 10 DF,  p-value: 0.0005713
```

```
##            0         0.25          0.5         0.75            1
##  0.593333333  0.693333333 -0.006666667 -0.173333333 -1.106666667
```

Form the anava table and boxplot, we can see that there is a statistical significance between these five levels of fiber content.Also, form the boxplot, we can estinate that the higher fiber content,the lower mean of compressive strength of concrete. Besides,from the lm model,the p-values of (Intercept),fiberPct1 and fiberPct2 are smaller than 0.05, which implies that these levels have signficant effect on the compressive strength of concrete.