HW5

Nikhil Gopal

10/16/2020

Do treatments have an effect on the distribution of post minus pre scores?

```
setwd("C:/Users/d/Google Drive/Notability/Applied Linear Regression Analysis")

data <- read.csv("deadseaminerals.csv")



#create a column for the difference between post and pre scores
data$difference <- data$post_y - data$pre_x

#create a linear model on this data and use treatment group as a factor
linmod_trt_group <- lm(data$difference ~ as.factor(data$trt_Grp))

#run an ANOVA test on the model
anova(linmod_trt_group)
```

```
## Analysis of Variance Table
##
## Response: data$difference
##                          Df Sum Sq Mean Sq F value   Pr(>F)
## as.factor(data$trt_Grp)  2  33666 16833.1  26.792 6.27e-09 ***
## Residuals                57  35813   628.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#generate a p value (should be same as ANOVA?)
p_value_f_dist <- pf(26.792, 2,57,lower.tail = FALSE)

summary(linmod_trt_group)
```

```
##
## Call:
## lm(formula = data$difference ~ as.factor(data$trt_Grp))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -58.500 -12.852   0.078  18.065  43.690
##
## Coefficients:
```

```
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -51.799      5.605  -9.242 6.23e-13 ***
## as.factor(data$trt_Grp)2  -25.269      7.927  -3.188 0.002328 **
## as.factor(data$trt_Grp)3   32.599      7.927   4.113 0.000127 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.07 on 57 degrees of freedom
## Multiple R-squared:  0.4846, Adjusted R-squared:  0.4665
## F-statistic: 26.79 on 2 and 57 DF,  p-value: 6.27e-09
```

```
p_value_f_dist
```

```
## [1] 6.269068e-09
```

To determine weather treatments have an effect on the distribution of post minus pre scores, a one-way ANOVA was conducted with 57/2 degrees of freedom (3 groups, n = 60). The observed sum of squared means was 33666 and 16833.1, and our F value was 26.792, giving us a p value of less than 0.001. Thus, we can conclude that these treatments have a statistically significant effect on distribution of post - pre scores.

Does gel treatment have an effect on distribution of post minus pre scores?

```
#gel

linmod_gel <- lm(data$difference ~ as.factor(data$gel))
anova(linmod_gel)
```

```
## Analysis of Variance Table
##
## Response: data$difference
##                      Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(data$gel)  1    179   179.1  0.1499 0.7001
## Residuals           58  69300  1194.8
```

```
p_value_f_dist_gel <- pf(26.792, 2,57,lower.tail = FALSE)

summary(linmod_gel)
```

```
##
## Call:
## lm(formula = data$difference ~ as.factor(data$gel))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -85.965 -28.224   6.562  26.955  45.345
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -48.134      5.465  -8.807 2.76e-12 ***
## as.factor(data$gel)1   -3.665      9.466  -0.387      0.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 34.57 on 58 degrees of freedom
## Multiple R-squared:  0.002578,   Adjusted R-squared:  -0.01462
## F-statistic: 0.1499 on 1 and 58 DF,  p-value: 0.7001
```

```
p_value_f_dist_gel
```

```
## [1] 6.269068e-09
```

The exact same analysis was used here as above. However, our p value was 0.7001, which is much to high to conclude statistical significance. Thus, we cannot say that the gel treatment has a statistically significant effect.

Does gel + dead sea concentrate treatment have an effect on distribution of post minus pre scores?

```
linmod_gel_DS <- lm(data$difference ~ as.factor(data$gelDS))
anova(linmod_gel_DS)
```

```
## Analysis of Variance Table
##
## Response: data$difference
##                        Df Sum Sq Mean Sq F value    Pr(>F)
## as.factor(data$gelDS)  1  23039 23039.2  28.774 1.481e-06 ***
## Residuals             58  46440   800.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
p_value_f_dist_gelDS <- pf(26.792, 2,57,lower.tail = FALSE)

summary(linmod_gel_DS)
```

```
##
## Call:
## lm(formula = data$difference ~ as.factor(data$gelDS))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -74.800 -18.355   6.405  21.590  35.058
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -35.500      4.474  -7.935 7.88e-11 ***
## as.factor(data$gelDS)1   -41.568      7.749  -5.364 1.48e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.3 on 58 degrees of freedom
## Multiple R-squared:  0.3316, Adjusted R-squared:  0.3201
## F-statistic: 28.77 on 1 and 58 DF,  p-value: 1.481e-06
```

In this case, we can say that the gel treatment plus the dead sea concentrate has a significant effect on the distribution. The p value was less than 0.001, meaning there is a statistically significant effect.