MA50259_Lab1_Submission_by_Han.R

cne

2024-02-16

```r
library(tidyverse)
library(ggplot2)



replicates_r <- 10
treatment_levels_t <- 6
number_of_experimental_units_n <- replicates_r * treatment_levels_t
overall_response_average_mu <- 500
treatment_effects_tau <- c(-20, 50, 0, -30, -10, 100)
mean_response <- overall_response_average_mu + treatment_effects_tau %>% rep(each = replicates_r)
sd <- 10
levels <- c("level 1","level 2","level 3","level 4", "level 5","level 6")
factor_f <- factor(rep(levels, each = replicates_r))



set.seed(5678)
fac <- sample(factor_f, size = number_of_experimental_units_n, replace = FALSE)
dataframe <- tibble(units=1:number_of_experimental_units_n, treatment=fac)
dataframe_rearranged_by_treatment_level <- arrange(dataframe, treatment)
dataframe_with_original_factor <- tibble(units=1:number_of_experimental_units_n, treatment=factor_f)
response_y <- rnorm(n = number_of_experimental_units_n, mean = mean_response, sd = sd) # note the units
dataframe_with_original_factor$individual_response <- response_y


ggplot(dataframe_with_original_factor, aes(x = treatment, y = individual_response)) + geom_point()
```
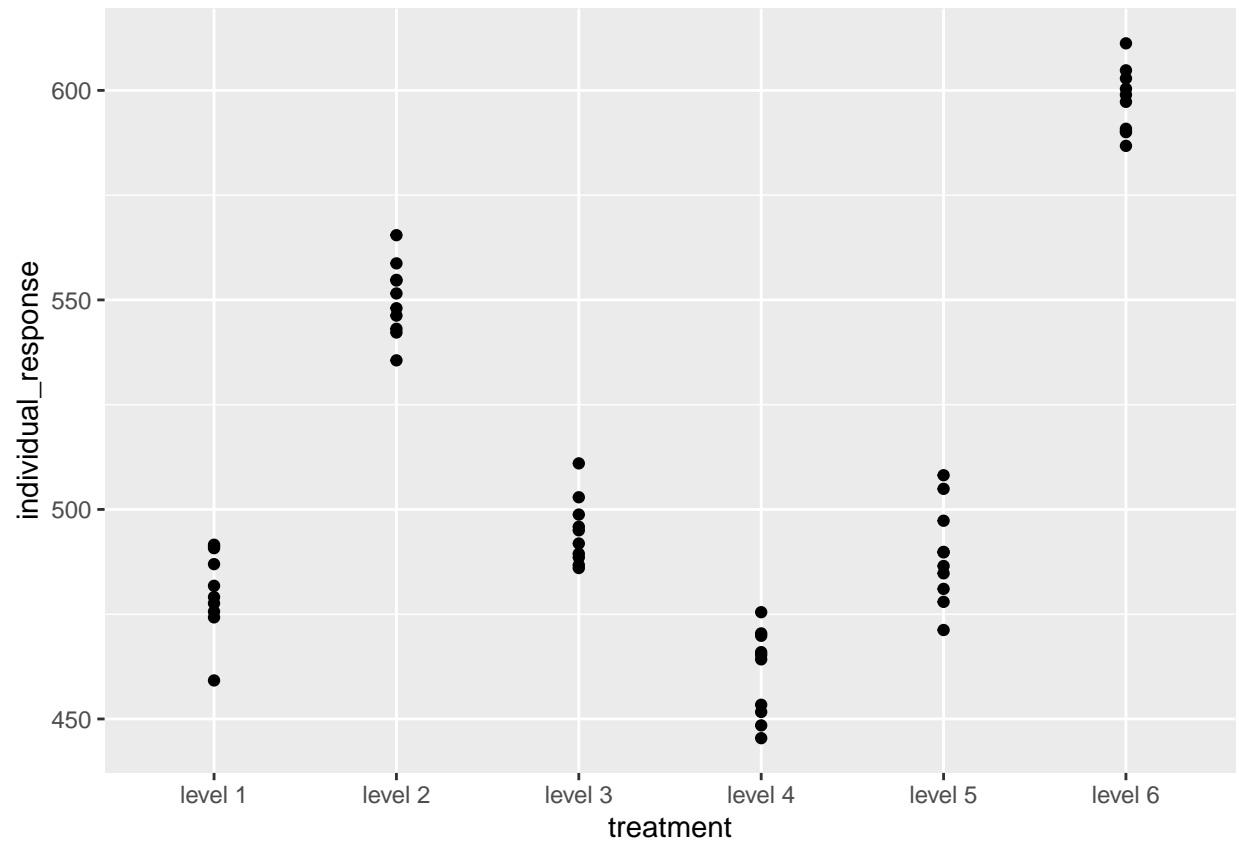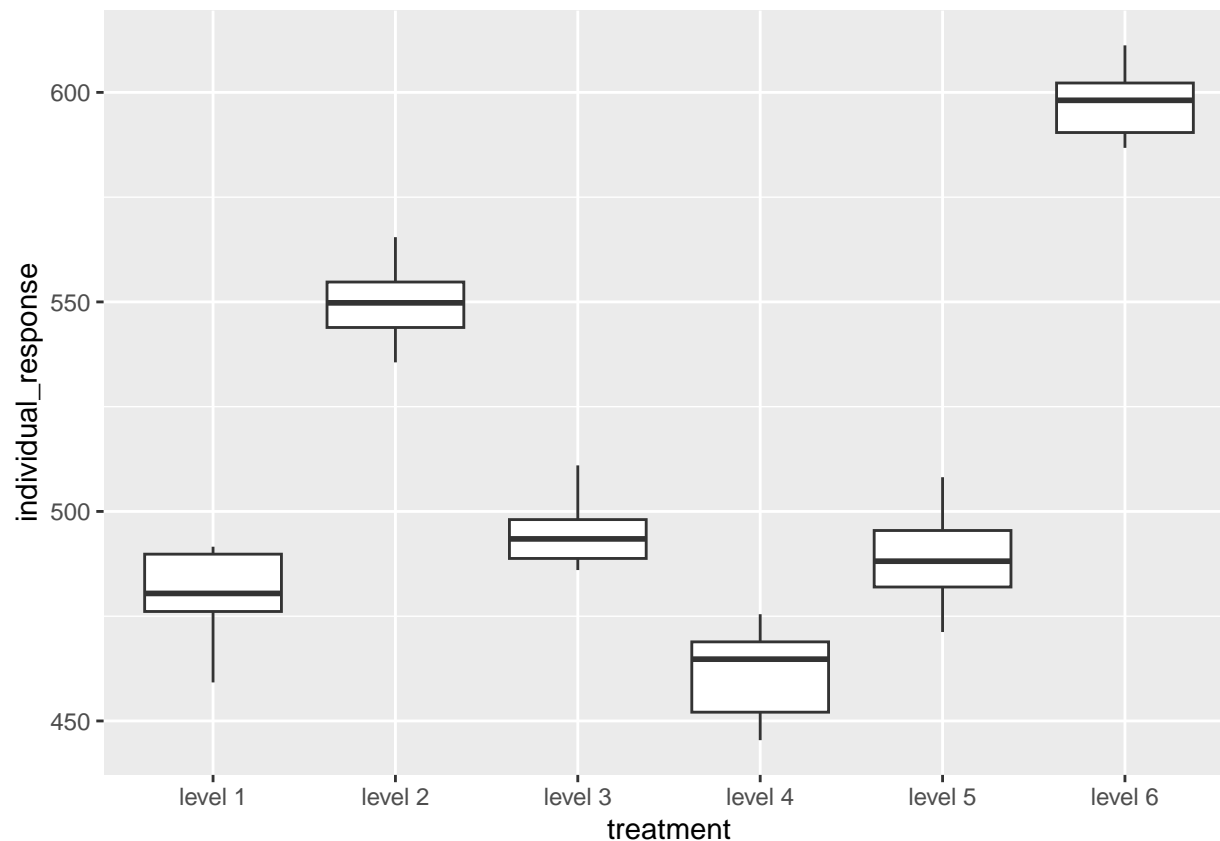
```r
ggplot(dataframe_with_original_factor, aes(x = treatment, y = individual_response)) + geom_boxplot()
```

```r
linear_model.dataframe_with_original_factor <- lm(individual_response ~ treatment, data = dataframe_with
summary(model.dataframe_with_original_factor)
```

```
##
## Call:
## lm(formula = individual_response ~ treatment, data = dataframe_with_original_factor)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.6029  -6.9514   0.5019   6.4716  19.0177
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       480.800      3.015 159.463  < 2e-16 ***
## treatmentlevel 2   69.233      4.264  16.237  < 2e-16 ***
## treatmentlevel 3   13.818      4.264   3.241  0.00204 **
## treatmentlevel 4  -19.788      4.264  -4.641 2.25e-05 ***
## treatmentlevel 5    8.349      4.264   1.958  0.05540 .
## treatmentlevel 6  116.542      4.264  27.332  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.535 on 54 degrees of freedom
## Multiple R-squared:  0.964,  Adjusted R-squared:  0.9606
## F-statistic: 288.8 on 5 and 54 DF,  p-value: < 2.2e-16
```

```
linear_model_coefficients <- coef(linear_model.dataframe_with_original_factor)
c(overall_response_average_mu, treatment_effects_tau[-1]) # why do we ignore tau_1?
```

```
## [1] 500  50   0 -30 -10 100
```

```
new_taus <- c(0, coefficients[-1]) # why do we make tau_1 = 0?
new_means <- coefficients[1] + new_taus
new_means
```

```
##                 treatmentlevel 2 treatmentlevel 3 treatmentlevel 4 treatmentlevel 5
##         480.7995          550.0329         494.6177         461.0119         489.1487
## treatmentlevel 6
##         597.3417
```

```
overall_response_average_mu + treatment_effects_tau
```

```
## [1] 480 550 500 470 490 600
```

```
by_group <- group_by(dataframe_with_original_factor, treatment)
by_group
```

```
## # A tibble: 60 x 3
## # Groups:   treatment [6]
##    units treatment individual_response
##    <int> <fct>                   <dbl>
## 1     1 level 1                   476.
## 2     2 level 1                   487.
## 3     3 level 1                   459.
## 4     4 level 1                   478.
## 5     5 level 1                   474.
## 6     6 level 1                   479.
## 7     7 level 1                   492.
## 8     8 level 1                   491.
## 9     9 level 1                   491.
## 10   10 level 1                   482.
## # i 50 more rows
```

```
summaries.dataframe_with_original_factor <- summarize(by_group, mean = mean(individual_response), sd = 
glimpse(summaries.dataframe_with_original_factor)
```

```
## Rows: 6
## Columns: 3
## $ treatment <fct> level 1, level 2, level 3, level 4, level 5, level 6
## $ mean      <dbl> 480.7995, 550.0329, 494.6177, 461.0119, 489.1487, 597.3417
## $ sd        <dbl> 10.061297, 8.790670, 7.911896, 10.406406, 11.626105, 7.803252
```