Estudo de Impacto de Tratamentos no Peso das Plantas

Problema de Negócio

- Uma empresa agrícola está enfrentando dificuldades em sua produão e deseja reavaliar a eficácia de seus tratamentos. Para realizar essa avaliação, a empresa reuniu dados de seus tratamentos, juntamente com um grupo de controle, usando um delineamento inteiramente casualizado (DIC). Agora, busca uma avaliação profissional para testar e analisar os resultados obtidos com esse experimento.
- Quais são os objetivos do projeto?
 - Avaliar a eficácia dos tratamentos.
 - Auxiliar na otimização do processo de produção.
 - Fornecer uma base sólida para a tomada de decisões futuras sobre a gestão dos tratamentos agrícolas.
- Metodologia:
 - Verificação dos dados.
 - Realização de uma ANOVA para determinar se existe diferença estatística significativa entre os tratamentos.
 - Execução do teste de dunnett para análise direta de cada tratamento em relação ao grupo de controle.

Importando Bibliotecas

```
pacman::p_load("dplyr", "DescTools", "car", "rstatix", "ggplot2", "qqplotr", "multcomp")
```

Entendendo os Dados

Dicionário dos Dados

- weight: O peso da planta em gramas(g) após submissão ao tratamento.
- group: O grupo de tratamento ao qual a planta pertence.
- ctrl: Grupo de controle.
- trt1: Primeiro grupo de tratamento.
- trt2: Segundo grupo de tratamento.

Importando e vizualizando as 6 primeiras linhas da base de dados

```
data(PlantGrowth); head(PlantGrowth)
```

```
## weight group
## 1 4.17 ctrl
## 2 5.58 ctrl
## 3 5.18 ctrl
## 4 6.11 ctrl
## 5 4.50 ctrl
## 6 4.61 ctrl
```

Os dados estão organizados em duas colunas, a primeira contendo nossa varíavel de interesse "weight", e a segunda composta pelos grupos do experimento.

```
PlantGrowth %>% group_by(group) %>% get_summary_stats(weight)
```

```
## # A tibble: 3 x 14
##
     group variable
                                                            iqr
                                  max median
                                                       q3
                                                                  mad
                        n
                            min
                                                 q1
     <fct> <fct>
                    <dbl> <dbl> <dbl>
                                       <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <
## 1 ctrl weight
                       10
                          4.17
                                 6.11
                                        5.16
                                               4.55
                                                    5.29 0.743 0.719
                                                                       5.03 0.583
## 2 trt1 weight
                          3.59
                                 6.03
                                         4.55
                                               4.21
                                                     4.87 0.662 0.534 4.66 0.794
## 3 trt2 weight
                       10 4.92 6.31
                                         5.44
                                              5.27 5.74 0.467 0.363 5.53 0.443
## # i 2 more variables: se <dbl>, ci <dbl>
```

De acordo com o sumário estatístico vemos que o experimento é balanceado, são 10 observações para cada um dos três grupos da base de dados. Além disso, nota-se certa semelhança nas medidas estatísticas de cada grupo, condição esta que vamos investigar mais a fundo ao longo da análise.

Para a realização da ANOVA os seguintes pressupostos devem ser atendidos:

- Homogêniedade das variâncias dos dados;
- Normalidade dos erros:
- Independência dos erros;
- Aditividade dos fatores.

Verificação dos Pressupostos

Para verificar a homogeniedade dos dados utilizarei o teste de levene em torno da mediana para ter um resultado mais exato considerando que nossa amostra é relativamente pequena.

Hipóteses

```
{
m H0}={
m As} variâncias dos grupos são semelhantes. 
 {
m H1}={
m Ao} menos duas variâncias diferem entre si.
```

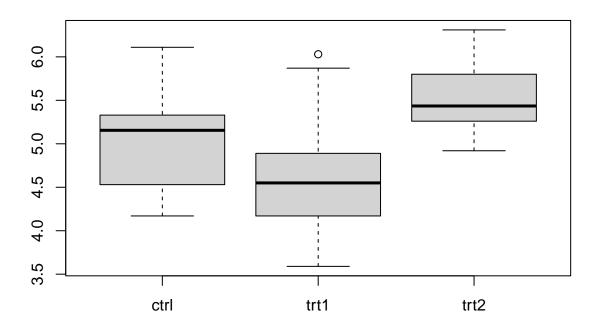
```
leveneTest(weight ~ group, PlantGrowth, center = median)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
## Df F value Pr(>F)
## group 2 1.1192 0.3412
## 27
```

O P-valor foi maior que 0.05 então aceita-se H0 ao nível de significância de 5%, portanto, há evidências de que nossos grupos possuem variâncias semlhantes.

Para vizualizarmos isto gráficamente vamos fazer um boxplot.

Boxplot do Peso por Grupo



Conforme observado nos resultados do teste, as variâncias dos dados são semelhantes. No entanto, o boxplot revelou a presença de um outlier no tratamento 1. Vamos analisar mais detalhadamente as observações desse tratamento para decidir a melhor forma de lidar com essa anomalia.

PlantGrowth[PlantGrowth\$group == 'trt1',]

```
##
      weight group
## 11
        4.81
              trt1
              trt1
##
        4.17
        4.41
##
   13
               trt1
        3.59
## 14
               trt1
        5.87
## 15
               trt1
## 16
        3.83
              trt1
## 17
        6.03
              trt1
        4.89
## 18
              trt1
## 19
        4.32
              trt1
## 20
        4.69
```

Analisando a base de dados, percebe-se que o valor não é tão discrepante e possui uma alta chance de ser legítimo. Naturalmente, algumas plantas podem apresentar um desenvolvimento menor ou maior que o comum. Portanto, ponderando as consequências de perder essa informação, optarei por manter o outlier na nossa análise.

Agora vamos montar o modelo da ANOVA.

```
modelo <- lm(weight ~ group, data = PlantGrowth)
resultado_anova <- anova(modelo)</pre>
```

Antes de olhar para o resultado da ANOVA vamos usar o teste de Shapiro-Wilk para verificar o pressuposto da normalidade dos resíduos.

Hipóteses

H0 = Os resíduos seguem distribuição normal.

H1 = Os resíduos não seguem distribuição normal.

```
shapiro.test(residuals(modelo))
```

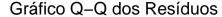
```
##
## Shapiro-Wilk normality test
##
## data: residuals(modelo)
## W = 0.96607, p-value = 0.4379
```

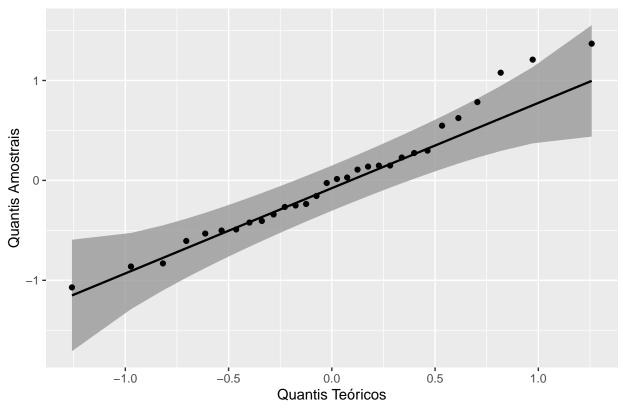
O P-valor foi maior que 0,05 então aceita-se H0 ao nível de significância de 5%, portanto, há evidências de que os resíduos seguem uma distribuição normal.

Para vizualizar melhor isto vamos fazer um plot dos quantis dos resíduos.

```
PlantGrowth$residuos <- residuals(modelo)

ggplot(PlantGrowth, aes(sample = residuos)) +
    stat_qq_band() +
    stat_qq_line() +
    stat_qq_point() +
    labs(title = "Gráfico Q-Q dos Resíduos", x = "Quantis Teóricos", y = "Quantis Amostrais") +
    theme(plot.title = element_text(hjust = 0.5))</pre>
```





Anova

Agora que sabemos que os pressupostos da normalidade dos resíduos e homogêniedade das variâncias dos grupos foram atendidos, e, supondo que os erros são independentes e os fatores aditivos, vamos vizualizar o resultado da nossa ANOVA

Hipóteses

H0 = As médias dos tratamentos são semelhantes.

H1 = Ao menos duas médias dos tratamentos diferem entre si.

resultado_anova

O P-valor foi menor que 0.05 então rejeita-se H0 ao nível de 5% de significância, portanto, há evidências de que ao menos duas médias dos tratamentos diferem entre si.

Teste de Dunnett

Após realizar a ANOVA e identificar a existência de diferenças significativas entre os grupos, aplicaremos o teste de Dunnett, ideal para compararação de múltiplos grupos de tratamento com um grupo de controle.

Hipóteses

 $\operatorname{H0:}\,$ Não há diferença significativa entre as médias do grupo de controle e do tratamento.

H1: Há uma diferença significativa entre as médias do grupo de controle e do tratamento.

```
dunnett <- glht(modelo, linfct = mcp(group = "Dunnett"))
summary(dunnett)</pre>
```

```
##
##
     Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Dunnett Contrasts
##
##
## Fit: lm(formula = weight ~ group, data = PlantGrowth)
##
## Linear Hypotheses:
##
                    Estimate Std. Error t value Pr(>|t|)
## trt1 - ctrl == 0 -0.3710
                                 0.2788
                                         -1.331
                                                    0.323
## trt2 - ctrl == 0
                      0.4940
                                 0.2788
                                           1.772
                                                    0.153
## (Adjusted p values reported -- single-step method)
```

O P-valor foi maior que 0,05 para os dois tratamentos então aceitamos H0 ao nível de 5% de significância, portanto, há evidências de que não existem diferenças entre as médias do grupo de controle e dos tratamentos.

Conclusão

Após a aplicação da ANOVA e do Teste de Dunnett, observamos que nenhum dos tratamentos difere significativamente do grupo de controle. Isso implica que os dois tratamentos analisados não têm um efeito estatisticamente relevante na produção. Com base nisso, recomendamos a interrupção desses métodos na produção das plantas e a busca por técnicas potencialmente mais eficientes.