Delineamento de Experimentos

Efeito de agrotóxicos na produção de laranjas - Análise

Jonatha Azevedo/Leonardo Filgueira

24 de abril de 2018

## Descrição do experimento

Todo o plantio foi realizado numa área 2.197,97 e divida em 3 estufas de mesmo tamanho (11,6 X 63.16) onde o solo, temperatura,umidade e irrigação foram tratados e controlados sob as mesmas condições. Cada semente foi plantada com uma distância de 2 metros uma da outra e as covas com abertura de 40 x 40 x 40 centímetros totalizando 100 sementes em cada estufa. Todas as sementes usadas foram do mesmo fabricante e plantadas no mesmo dia e hora. Em cada estufa, a mesma condição foi aplicada, sendo que, na primeira estufa não foi aplicado nenhum agrotóxico; na segunda estufa foi aplicada metade da dose de agrotóxicos atualmente utilizada e na terceira estufa foi aplicada a dose atualmente utilizada. Ao final do experimento será comparado o peso total de laranjas em cada estufa.

### Descrição das variáveis medidas no objeto de estudo

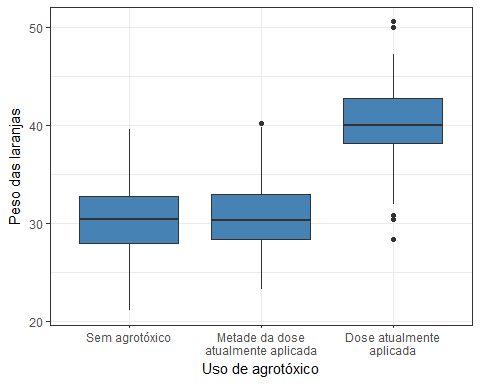
* **Variável resposta**: Peso total das laranjas (em Kg) por pé.
* **Fator**: Grupo de utilização de agrotóxico
* **Níveis**:
  + 1 = Sem agrotóxico;
  + 2 = Metade da dose atualmente aplicada;
  + 3 = Dose atualmente aplicada.

## Análise

### Análise descritiva

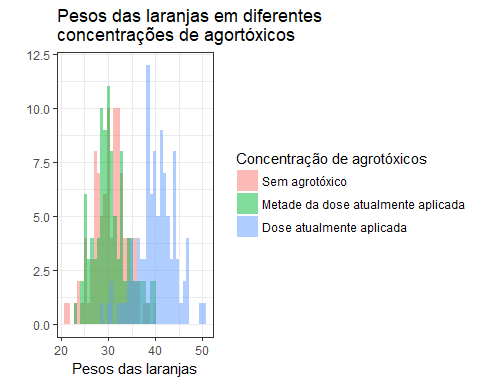
Primeiramente, será feita uma análise descritiva dos dados. Abaixo está um box-plot do peso das laranjas considerando as diferentes concentrações de agrotóxicos listadas [acima](#desc).

base %>%   
 mutate(ind = factor(ind,   
 labels = c("Sem agrotóxico", "Metade da dose atualmente aplicada",  
 "Dose atualmente aplicada"))) %>%   
 ggplot(aes(x = ind, y = pesos)) +   
 geom\_boxplot(fill = "steelblue") +   
 theme\_bw() +   
 labs(x = "Uso de agrotóxico", y = "Peso das laranjas") +   
 scale\_x\_discrete(labels = function(x) str\_wrap(x, width = 20))



Ao analisar o boxplot, pode-se perceber que a distribuição do peso das laranjas é diferente quando a dose de agrotóxicos normalmente utilizada é aplicada. Nota-se que com esta concenctração de agrotóxicos, o peso das laranjas é maior. A seguir, observa-se um histograma dos pesos das laranjas, onde é dividida a distribuição, também pela concentração de agrotóxicos. Ao observar o histograma, confirma-se o que foi notado no box-plot: a distribuição de pesos das laranjas é diferente na situação em que a dose de agrotóxicos utilizada é a concentração normalmente aplicada.

base %>%   
 mutate(ind = factor(ind, labels = c("Sem agrotóxico", "Metade da dose atualmente aplicada", "Dose atualmente aplicada"))) %>%   
 ggplot(aes(x = pesos, fill = ind)) +   
 geom\_histogram(binwidth=.7, alpha=0.5, position="identity") +   
 theme\_bw() +   
 labs(x = "Pesos das laranjas", y = "",  
 title = "Pesos das laranjas em diferentes \nconcentrações de agortóxicos") +   
 guides(fill=guide\_legend(title="Concentração de agrotóxicos"))



### Inferência

A seguir serão realizados testes de hipóteses para confirmar o que já fora notado na [análise exploratória](#expl). Serão realizadas uma série de testes como objetivo de detectar diferença na distribuição do peso das laranjas de acordo com a concentração de agrotóxicos. Para os testes aplicados em seguida, será utilizado um nível de significância estatística de 5%.

#### Teste de normalidade (Shapiro-Wilk)

O teste de Shapiro-Wilk, teste para normalidade, considera as seguintes hipóteses:

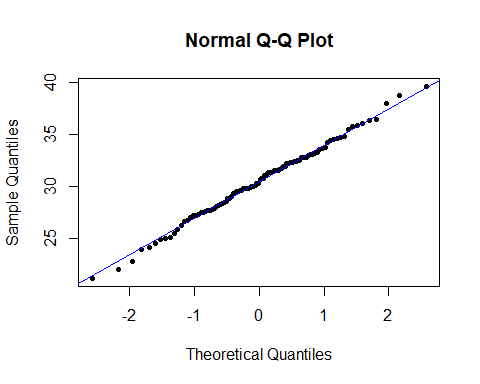
Será verificada a hipótese de normalidade para cada um dos níveis do fator.

* Primeiro para o tratamento **sem agrotóxicos** (*nível 1*):

base %>%   
 filter(ind == 1) %>%   
 pull(pesos) %>%   
 shapiro.test()

##   
## Shapiro-Wilk normality test  
##   
## data: .  
## W = 0.9956, p-value = 0.9876

Com o p-valor de 0.988 não é possível rejeitar a hipótese de normalidade dos dados para o tratamento do nível 1. Também ao verificar o qqplot pode-se supor que é razoável supor que estes dados podem vir de uma população com distribuição normal.

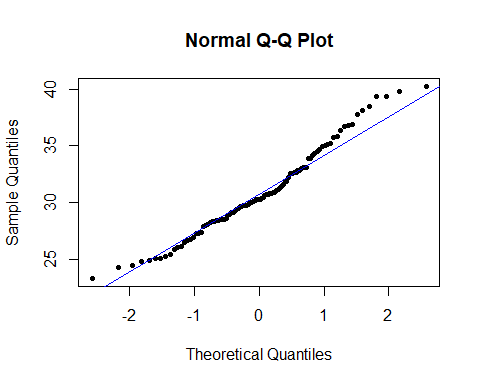


* Em seguida para o tratamento com **metade da dose atualmente aplicada** (*nível 2*):

base %>%   
 filter(ind == 2) %>%   
 pull(pesos) %>%   
 shapiro.test()

##   
## Shapiro-Wilk normality test  
##   
## data: .  
## W = 0.97431, p-value = 0.04763

Com o p-valor de 0.048 rejeita-se a hipótese de normalidade dos dados para o tratamento do nível 2. Ao observar o qqplot pode se notar que os quantis da amostra estão distantes dos quantis de uma distribuição normal.



Como a hipótese de normalidade foi rejeitada para o segundo nível, não é necessário verificar se, para o terceiro nível, a hipótese de normalidade será ou não rejeitada. Com essa configuração, ao nível de significância estatística de 5%, a hipótese de normalidade da ANOVA foi violada. Logo, será necessária uma abordagem não paramétrica para o experimento.

#### Teste de Kruskal-Wallis

Verificamos se os níveis representados, possuem a mesma distribuição (hipótese nula). O teste de Kruskal é uma extensão do teste de Wilcoxon, que será também utilizado na análise. O teste de Kruskal-Wallis é o análogo ao teste F utilizado na ANOVA com 1 fator. Enquanto a análise de variância depende da hipótese de que todas as populações em confronto são independentes e normalmente distribuídas, o teste de Kruskal-Wallis não coloca nenhuma restrição sobre a comparação. Suponha que os dados provenham de k amostras aleatórias independentes com tamanhos amostrais sendo o número total de elementos considerados em todas as amostras.

Considerando as seguintes hipóteses:

Onde é o efeito do nível sobre a variável resposta.

kruskal.test(pesos~ind, data = base)

##   
## Kruskal-Wallis rank sum test  
##   
## data: pesos by ind  
## Kruskal-Wallis chi-squared = 161.3, df = 2, p-value < 2.2e-16

com o p-valor muito abaixo do nível de significância, rejeita-se a hipótese nula (distribuições iguais) para o teste de Kruskal. Em seguida, realizam-se as comparações múltiplas com o teste de Wilcoxon, a fim de verificar para qual(is) pares de grupos há diferença de distribuições.

#### Comparações múltiplas

A seguir, as hipóteses a serem testadas serão:

Também será necessário ajustar o nível de significância, a fim de que a probabilidade de cometer o erro do tipo I nas comparações múltiplas não se descontrole. Então .

Comparando o **nível 1** (sem agrotóxico) e o **nível 2** (metade da dose aplicada):

wilcox.test(base$pesos[base$ind == 1], base$pesos[base$ind == 2])

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: base$pesos[base$ind == 1] and base$pesos[base$ind == 2]  
## W = 4861, p-value = 0.7351  
## alternative hypothesis: true location shift is not equal to 0

Nota-se que as distribuições não diferem.

Comparando o **nível 1** (sem agrotóxico) e o **nível 3** (dose total aplicada):

wilcox.test(base$pesos[base$ind == 1], base$pesos[base$ind == 3])

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: base$pesos[base$ind == 1] and base$pesos[base$ind == 3]  
## W = 426, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0

Neste teste, conclui-se que as distribuições diferem.

Comparando o **nível 2** (metade da dose aplicada) e o **nível 3** (dose total aplicada):

wilcox.test(base$pesos[base$ind == 2], base$pesos[base$ind == 3])

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: base$pesos[base$ind == 2] and base$pesos[base$ind == 3]  
## W = 582, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0

Conclui-se que as distribuições diferem.

## Conclusão

Como havia sido já observado na análise descritiva, existe diferença significativa entre os grupos 1 e 3 e entre 2 e 3. Logo, pode-se dizer que há diferença significativa do peso das laranjas com dose atualmente aplicada de agrotóxicos, em relação às laranjas sem agrotóxicos e com a metade da dose atualmente aplicada. Embora o não uso de agrotóxicos possa deixar de trazer malefícios à saúde, esta prática resulta num menor peso das frutas produzidas.