

5.1. ANOVA – Apêndice 1 – Exemplo

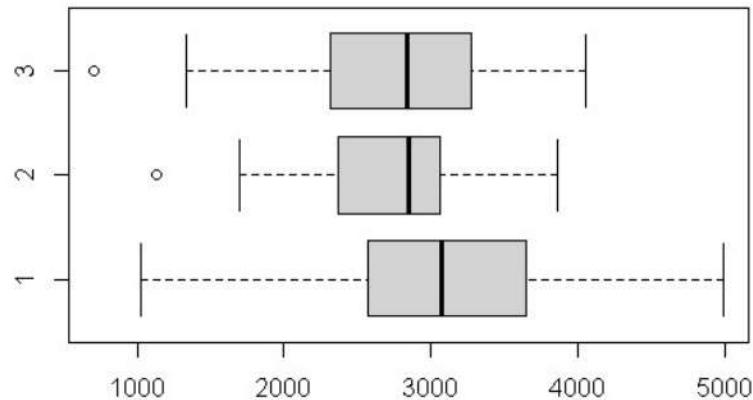
Para verificar que fatores podem levar uma criança a nascer com peso abaixo do normal (o baixo peso ao nascer põe em risco a saúde e o crescimento da criança), foi feito um estudo a partir de uma amostra de 189 nascimentos registrados no *Baystate Medical Center*, Springfield (EUA), in 1986. Um dos fatores considerados foi a raça da mãe; as mães foram classificadas em três raças, *brancas*, *negras* ou *outras*. Os dados estão disponíveis em R no arquivo *lowbw.txt*, do pacote *aplore3*.

Uma ANOVA foi feita para testar se existe diferença significativa entre os pesos ao nascer das crianças de mães das três raças. Abaixo são mostrados os comandos usados, e as saídas do R.

1) Leitura dos dados e gráfico dos pesos ao nascer, por raça

```
x=read.table("lowbw.txt", header=T)
attach(x)

boxplot(bwt[race==1],bwt[race==2],bwt[race==3],horizontal=T)
```



Vemos que as medianas dos pesos das crianças das raças 2 e 3 (negras e outras) parecem aproximadamente iguais entre si, mas diferentes da mediana da raça 1 (brancas).

2) Testes dos pressupostos: normalidade e igualdade das variâncias

```
# testes de normalidade
shapiro.test(bwt[race==1])
  Shapiro-Wilk normality test
  data: bwt[race == 1]
  W = 0.98728, p-value = 0.4864
shapiro.test(bwt[race==2])
  Shapiro-Wilk normality test
  data: bwt[race == 2]
  W = 0.97696, p-value = 0.8038
shapiro.test(bwt[race==3])
  Shapiro-Wilk normality test
  data: bwt[race == 3]
  W = 0.97547, p-value = 0.2071
```

```
# teste da igualdade das variâncias
library(car)
leveneTest(bwt, race, center=mean)

Levene's Test for Homogeneity of Variance (center = mean)
  Df F value Pr(>F)
group  2  0.4368 0.6467
      186
```

Os três testes de Shapiro tiveram resultados não-significativos (os três valores p acima de 0,05), o que significa que o pressuposto de normalidade dos pesos das crianças nas três populações não foi rejeitado. O teste de Levene também teve resultado não-significativo ($p = 0,6467$), o que indica que o pressuposto da igualdade das variâncias dos pesos das crianças nas três populações não foi rejeitado. A ANOVA portanto pode ser usada para analisar estes dados.

3) Análise da variância

```
# a variável que separa os tratamentos deve ser tipo _factor_
# (no arquivo original, esta' como _integer_)
race = as.factor(race)

# anova
anova.bwt = aov(bwt~race)
summary(anova.bwt)

  Df  Sum Sq Mean Sq F value  Pr(>F)
race      2 5070608 2535304  4.972 0.00788 ***
Residuals 186 94846445 509927
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A ANOVA teve resultado significativo ($p = 0,0078$), o que quer dizer que a hipótese de as crianças das três raças terem o mesmo peso médio ao nascer foi rejeitada. É preciso agora fazer um teste *post-hoc*, para verificar qual das três raças tem pesos médios significativamente diferente das outras.

```
# testes post-hoc
TukeyHSD(anova.bwt)

  Tukey multiple comparisons of means
  95% family-wise confidence level
  Fit: aov(formula = bwt ~ race)

  $race
  diff      lwr      upr      p adj
2-1 -384.04728 -757.0458 -11.04871 0.0419620
3-1 -299.72466 -568.3026 -31.14673 0.0245289
3-2   84.32262 -305.4999 474.14515 0.8660552
```

O teste *post-hoc* de Tukey indica que a raça 1 (brancas) tem média de peso ao nascer que difere significativamente das médias da raça 2 ($p = 0,042$) e da raça 3 ($p = 0,025$). As médias das raças 2 e 3, por outro lado, não diferem significativamente entre si ($p=0,86$).