Discussion 9

Berman & Rummerfield

12/1/2017

Generations of athletes have been cautioned that cigarette smoking hinders performance. One measure of the truth of that warning is the effect of smoking on heart rate. In one study examining that impact, six each of non-smokers, light smokers, moderate smokers, and heavy smokers undertook sustained physical exercise. Their heart rates were measured after resting for three minutes. The results appear in the following table:

Non-Smokers	Light Smokers	Moderate Smokers	Heavy Smokers
69	55	66	91
52	60	81	72
71	78	70	81
58	58	77	67
59	62	57	95
65	66	79	84

```
exercise = data.frame(HR = c(69,52,71,58,59,65,
                             55,60,78,58,62,66,
                             66,81,70,77,57,79,
                             91,72,81,67,95,84),
                      Smoke = factor(c(rep("NS",6), rep("LS",6),
                                       rep("MS",6),rep("HS",6)),
                                     levels=c("NS","LS","MS","HS")))
```

Write the population model using group mean and factor effects and write the corresponding null and alternative hypothesis for an ANOVA in symbols and in context of the problem.

Next, lets take a look at the means and check if the conditions for a one-way ANOVA are satisfied.

```
means <- tapply(exercise$HR, exercise$Smoke, mean)
s <- tapply(exercise$HR, exercise$Smoke, sd)
```

```
##       NS       LS       MS       HS
## 62.33333 63.16667 71.66667 81.66667
```

```
##        NS        LS        MS        HS
##  7.257180  8.158840  9.157875 10.764138
```

```
library(ggplot2)
ggplot(exercise) + geom_boxplot(aes(y = HR, x = Smoke, fill = Smoke)) +
  labs(title = "Heart Rate by Smoking Status", x = "Smoking Status",
       y = "Heart Rate") + theme_bw()
```

Parameter Estimates

Let's calculate the estimates of the parameters in the cell means model.

$\hat{\mu}_{\text{NS}} =$ ______________ $\hat{\mu}_{\text{LS}} =$ ______________

$\hat{\mu}_{\text{MS}} =$ ______________ $\hat{\mu}_{\text{HS}} =$ ______________

Let's calculate the estimates of the parameters in the factor effects model.

$\hat{\mu} =$ ______________ $\hat{\alpha}_{\text{NS}} =$ ______________

$\hat{\alpha}_{\text{LS}} =$ ______________ $\hat{\alpha}_{\text{MS}} =$ ______________

$\hat{\alpha}_{\text{HS}} =$ ______________

ANOVA table

Next, we are going to make an ANOVA table by hand.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Smoke					
Residuals					

```r
ybar_k <- means
ybar = mean(exercise$HR)
SSgrp <- sum(6*(ybar_k - ybar)^2)
SSE <- sum(5*s^2)
SSTot <- SSgrp + SSE
MSgrp <- SSgrp/ (4-1)
MSE <- SSE/(24-4)
F_stat <- MSgrp / MSE
p_val <- pf( (SSgrp/3) / (SSE/20), 3, 20, lower.tail = FALSE)
```

Doing the work in R

Now compare your results to the ANOVA table produced in R. Recall, there are many ways to get the F statistic and p-value.

```
mod1 <- aov(HR ~ Smoke, data=exercise)
summary(mod1) # 1

mod2 <- lm(HR ~ Smoke, data=exercise)
anova(mod2) # 2
summary(mod2) # 3

exercise <- cbind(exercise, model.matrix(~ Smoke - 1, data = exercise))
mod3 <- lm(HR ~ SmokeLS + SmokeMS + SmokeHS, data = exercise)
summary(mod3) # 4
anova(mod3) # 5
```

Finally, check all pairwise relationships to get a more detailed look at which mean(s) were different from one another and by how much.

```
TukeyHSD(mod1)
plot(TukeyHSD(mod1, ordered=TRUE, conf.level=0.95), las=1)
```