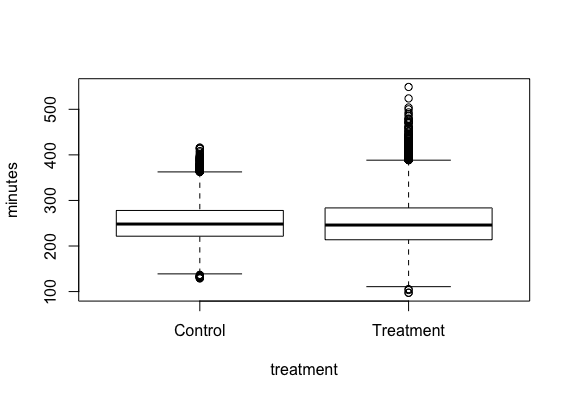
1. Merged data based on unique ID
2. Inspected data for missing data (imputed NA when this was the case)
3. Converted seconds into minutes, for easier reading of data
4. Plotted distribution of running time for both control group and treatment group.

At first sight, there is not much difference between the treatment group and the control group. Interestingly, there are more outliers toward longer running time for the treatment group.



1. The effect of wearing RunPro 3000 shoes can be modelled with a linear regression model: yi=β0+β1xi1+ϵi

Modelled effect of treatment on running time (model 1):

Response: minutes

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

treatment 1 408 407.83 0.1749 0.6758

Residuals 32390 75521463 2331.63

* 1. No significant effect of RunPro3000 on running time.

1. Modelled effect of treatment on running time, taking into account effects of age, nationality and sex (model 2):

Response: minutes

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

treatment 1 513 513 0.2537 0.6145

ageClass 11 1604294 145845 72.1202 <2e-16 \*\*\*

sex 1 6843115 6843115 3383.9163 <2e-16 \*\*\*

nationality 91 1630670 17919 8.8612 <2e-16 \*\*\*

Residuals 32134 64982888 2022

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

* 1. All other variables do have a significant effect, but treatment effect is still insignificant.

1. Modelled effect of treatment on running time, conditioned on gender (interaction model 3):

Response: minutes

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

treatment 1 513 513 0.2560 0.6129

sex 1 6438448 6438448 3213.1702 <2e-16 \*\*\*

ageClass 11 2008961 182633 91.1447 <2e-16 \*\*\*

nationality 91 1630670 17919 8.9429 <2e-16 \*\*\*

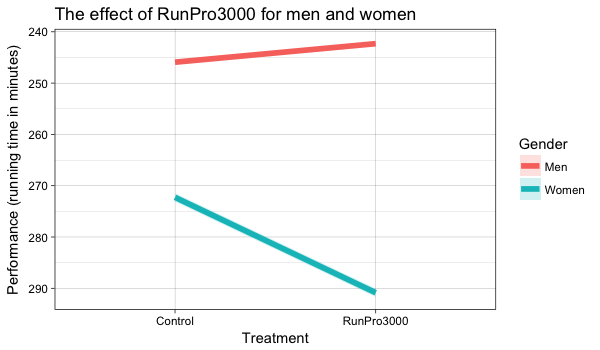
treatment:sex 1 595803 595803 297.3415 <2e-16 \*\*\*

Residuals 32133 64387084 2004

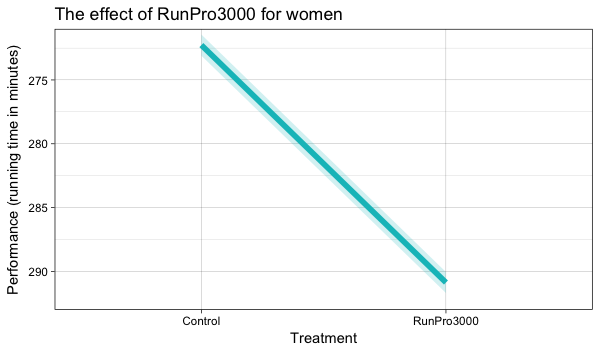
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

* 1. Interacted with gender, treatment does have a significant effect on running time.







* 1. Plotting the predicted effect of RunPro3000 for either men or women shows that men with RunPro3000 run approximately 4 minutes faster than without them. On the other hand, women run approximately 18 minutes slower with RunPro3000.

1. Plotting predicted effects age class (whether or not they were wearing RunPro3000 (interaction model 4). There are differences between age groups when it comes to their running time. Younger runners are faster than older runners, although there is not much difference between 20 and 40 years old (they are all predicted to finish the marathon in about 250 minutes. There is no significant difference in running time for any of the age groups as a result of the RunPro3000.

Response: minutes

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

treatment 1 513 513 0.2537 0.6145

ageClass 11 1604294 145845 72.1227 <2e-16 \*\*\*

sex 1 6843115 6843115 3384.0311 <2e-16 \*\*\*

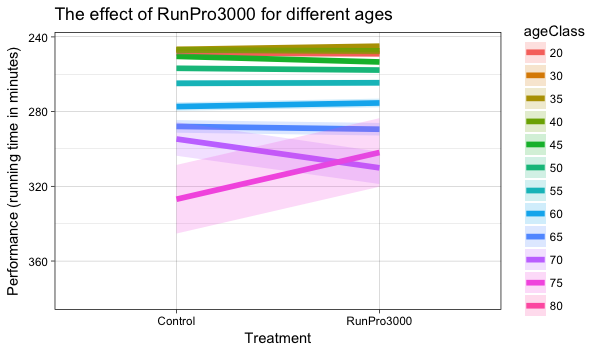
nationality 91 1630670 17919 8.8615 <2e-16 \*\*\*

treatment:ageClass 10 22426 2243 1.1090 0.3506

Residuals 32124 64960461 2022

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



1. Plotting predicted effects nationality (whether or not they were wearing RunPro3000 (interaction model 5). There are no significant differences in running time for any of the nationalities as a result of the RunPro3000. When you are an Ethiopian you are expected to finish the marathon the quickest no matter the type of shoes (in around 150 minutes).

Response: minutes

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

treatment 1 513 513 0.2537 0.6145

nationality 91 1842388 20246 10.0118 <2e-16 \*\*\*

sex 1 6315787 6315787 3123.1916 <2e-16 \*\*\*

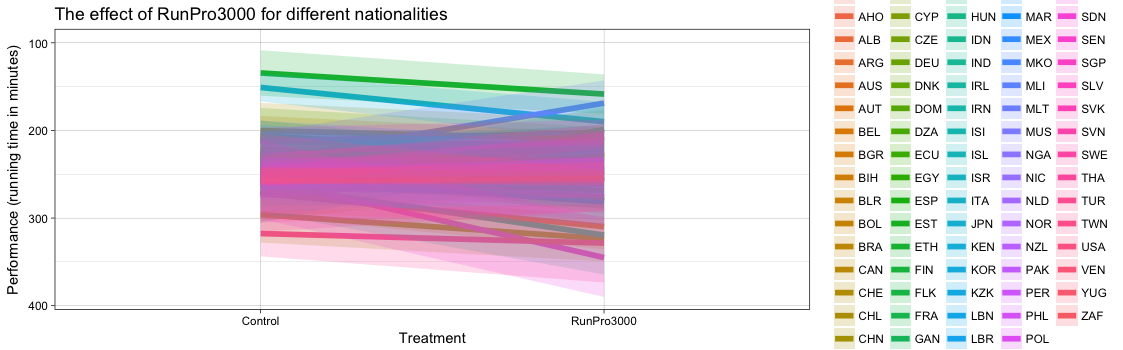
ageClass 11 1919903 174537 86.3093 <2e-16 \*\*\*

treatment:nationality 59 120112 2036 1.0067 0.4611

Residuals 32075 64862776 2022

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



###########################################################

###################### Set Directory ######################ˇ

getwd()

setwd('/Users/reinischrama/Desktop/Private/ProPulsion/Challenge')

###########################################################

######################## Load Data ########################ˇ

runner = read.csv("runner\_attributes.csv")

marathon = read.csv("marathon\_times.csv")

###########################################################

####################### Merge Data ########################ˇ

run.mar.data <- merge(runner, marathon,by="unique\_id")

run.mar.data

summary(run.mar.data)

###########################################################

###################### Inspect Data #######################

summary(run.mar.data)

summary(run.mar.data$netTime)

summary(run.mar.data$nationality)

summary(run.mar.data$ageClass)

run.mar.data$netTime[run.mar.data$netTime<0] <- NA

run.mar.data$ageClass[run.mar.data$ageClass=="U20"] <- NA

run.mar.data$minutes <- run.mar.data$netTime/60

run.mar.data$hours <- run.mar.data$minutes/60

sub.treatment.time <- subset(run.mar.data, select = c("treatment", "minutes"))

summary(sub.treatment.time)

###########################################################

###################### Analyse Data #######################

# plot time and treatment

plot(sub.treatment.time)

###################### Simple Model #######################

# treatment as categorical variable

str(run.mar.data$treatment)

run.mar.data$treatment <- factor(run.mar.data$treatment)

# model treatment and runners net time

model.1 <- lm(minutes ~ treatment,

data=run.mar.data)

# results

coef(summary(model.1))

anova(model.1)

################## Model with controls ###################

# model treatment, time and controls

model.2 <- lm(minutes ~ treatment + ageClass

+ sex + nationality,

data=run.mar.data)

# results

coef(summary(model.2))

anova(model.2)

################### Interaction Model ####################

# model treatment\*sex, time and controls

model.3 <- lm(minutes ~ treatment\*sex + ageClass

+ nationality,

data=run.mar.data)

# results

coef(summary(model.3))

anova(model.3)

###########################################################

###################### Visualize Data #####################

################# Plot Interaction Model ##################

#install.packages("effects")

#library("effects")

# Interaction effet

interaction <- effect('treatment\*sex', model.3,

se=TRUE)

# Data Frame

interaction.df<-as.data.frame(interaction)

# Labels

interaction.df$treatment.f <- factor(interaction.df$treatment,

level=c("Control", "Treatment"),

labels=c("Control", "RunPro3000"))

interaction.df$Gender <- factor(interaction.df$sex,

level=c("M", "W"),

labels=c("Men", "Women"))

# Interaction Plot

plot.interaction<-ggplot(data=interaction.df, aes(x=treatment.f, y=fit, group=Gender))+

geom\_line(size=2, aes(color=Gender))+

geom\_ribbon(aes(ymin=fit-se, ymax=fit+se,fill=Gender),alpha=.2)+

scale\_y\_reverse("Performance (running time in minutes)") +

scale\_x\_discrete("Treatment") +

ggtitle("The effect of RunPro3000 for men and women")+

theme\_linedraw()

plot.interaction

# The bad news...

# Zoom in plot for Women

plot.interaction<-ggplot(data=interaction.df, aes(x=treatment.f, y=fit, group=Gender))+

geom\_line(size=2, aes(color=Gender))+

geom\_ribbon(aes(ymin=fit-se, ymax=fit+se,fill=Gender),alpha=.2)+

scale\_y\_reverse("Performance (running time in minutes)") +

scale\_x\_discrete("Treatment") +

ggtitle("The effect of RunPro3000 for women")+

theme\_linedraw() +

theme(legend.position="none") +

coord\_cartesian(ylim=c(292,272))

plot.interaction

# The good news...

# Zoom in plot for Men

plot.interaction<-ggplot(data=interaction.df, aes(x=treatment.f, y=fit, group=Gender))+

geom\_line(size=2, aes(color=Gender))+

geom\_ribbon(aes(ymin=fit-se, ymax=fit+se,fill=Gender),alpha=.2)+

scale\_y\_reverse("Performance (running time in minutes)") +

scale\_x\_discrete("Treatment") +

ggtitle("The effect of RunPro3000 for men")+

theme\_linedraw() +

theme(legend.position="none") +

coord\_cartesian(ylim=c(247,242))

plot.interaction

########## Predicting effect of all attributes ###########

######################### Age ############################

# model treatment\*age, time and controls

model.4 <- lm(minutes ~ treatment\*ageClass + sex

+ nationality,

data=run.mar.data)

# results

coef(summary(model.4))

anova(model.4)

# Interaction effet

interaction\_age <- effect('treatment\*ageClass', model.4,

se=TRUE)

# Data Frame

interaction\_age.df<-as.data.frame(interaction\_age)

# Labels

interaction\_age.df$treatment.f <- factor(interaction\_age.df$treatment,

level=c("Control", "Treatment"),

labels=c("Control", "RunPro3000"))

# Interaction Plot

plot.interaction\_age<-ggplot(data=interaction\_age.df, aes(x=treatment.f, y=fit, group=ageClass))+

geom\_line(size=2, aes(color=ageClass))+

geom\_ribbon(aes(ymin=fit-se, ymax=fit+se,fill=ageClass),alpha=.2)+

scale\_y\_reverse("Performance (running time in minutes)") +

scale\_x\_discrete("Treatment") +

ggtitle("The effect of RunPro3000 for different ages")+

theme\_linedraw()

plot.interaction\_age

##################### Nationality ##########################

interaction\_nat\_new.df <- na.omit(interaction\_nat.df)

# model treatment\*nationality, time and controls

model.5 <- lm(minutes ~ treatment\*nationality + sex

+ ageClass,

data=run.mar.data)

# results

coef(summary(model.5))

anova(model.5)

# Interaction effet

interaction\_nat <- effect('treatment\*nationality', model.5,

se=TRUE)

# Data Frame

interaction\_nat.df<-as.data.frame(interaction\_nat)

interaction\_nat\_new.df <- na.omit(interaction\_nat.df)

# Labels

interaction\_nat\_new.df$treatment.f <- factor(interaction\_nat\_new.df$treatment,

level=c("Control", "Treatment"),

labels=c("Control", "RunPro3000"))

# Interaction Plot

plot.interaction\_nat<-ggplot(data=interaction\_nat\_new.df, aes(x=treatment.f, y=fit, group=nationality))+

geom\_line(size=2, aes(color=nationality))+

geom\_ribbon(aes(ymin=fit-se, ymax=fit+se,fill=nationality),alpha=.2)+

scale\_y\_reverse("Performance (running time in minutes)") +

scale\_x\_discrete("Treatment") +

ggtitle("The effect of RunPro3000 for different nationalities")+

theme\_linedraw()

plot.interaction\_nat