Block 4 Week 4: Personal Training & Marathon Report

1 Introduction

The data available at https://uoepsy.github.io/data/marathonDat1.csv and https://uoepsy.github.io/data/marathonDat2.csv were provided by a Personal Trainer who specializes in workouts that involve running. Their first data set contained information on the base fitness levels (`BaseFitness`; levels = Low, Moderate, High) of 90 clients; whilst the second contained information regarding the training methods (`Group`; levels = New Method, Typical, Solo) they use with their clients and whether they could complete a marathon (`Marathon`; levels = None, Half, Full).

1.1 Research Questions

- RQ1: Is the proportion of participants in each fitness category consistent with the expected proportions?
- RQ2: Is marathon completion associated with the type of training completed?

2 Analysis

2.1 Research Question 1

To investigate whether the proportion of participants in each fitness category (Low, Moderate, High) was consistent with the expected proportions, we performed a χ^2 Goodness of Fit test. The Goodness of Fit test was not significant ($\chi^2(2, 90) = 1.87, p = .393$). Since we failed to reject the null hypothesis, we concluded that our sample was equal in proportion across levels of fitness (see Figure 1).

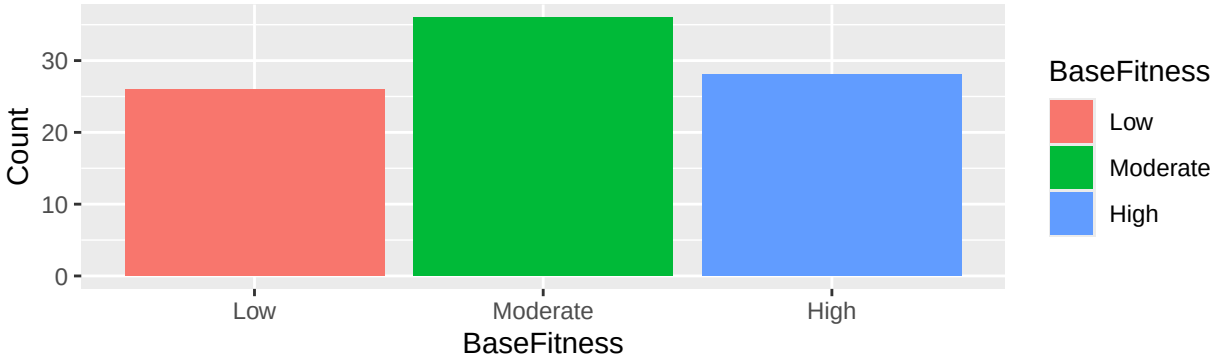

Figure 1: Fitness Level Proportions

2.2 Research Question 2

To investigate whether marathon completion (None, Half, Full) was associated with the type of training completed (New Method, Typical Method, Solo Method), we performed a χ^2 Test of Independence. The χ^2 Test of Independence was significant ($\chi^2(4, 90) = 10.03, p = .040$), which suggested that marathon completion was dependent on the type of training completed (see Figure 2). The size of the effect was found to be small-medium $V = 0.18[0.00, 1.00]$. Specifically, more participants than expected ran a full marathon after training with the new method, and fewer participants than expected could not run a marathon after training with the new method. Additionally, fewer participants than expected completed a full marathon when training solo.

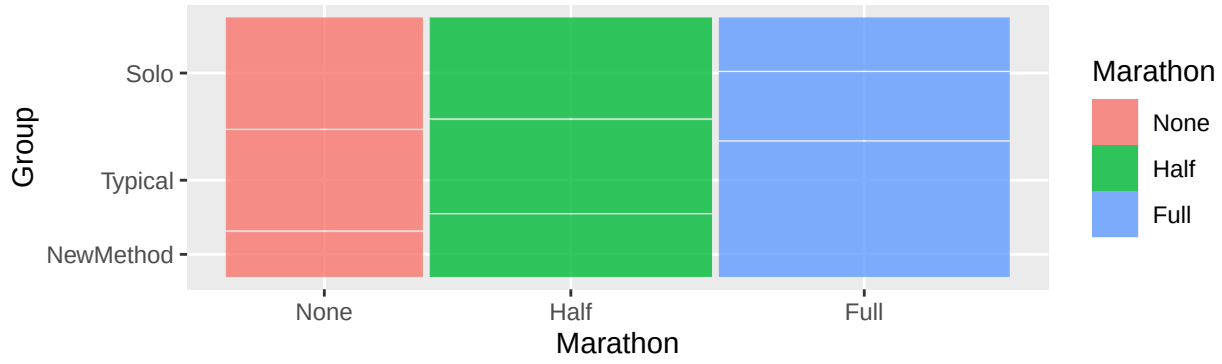

Figure 2: Association between Marathon Completion and Training Method

3 Appendix

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)
######## LOAD LIBRARIES & DATA ########
library(tidyverse)
library(patchwork)
library(kableExtra)
library(psych)
library(effectsize)

dat1 <- read_csv("https://uoepsy.github.io/data/marathonDat1.csv")
dat2 <- read_csv("https://uoepsy.github.io/data/marathonDat2.csv")

###################### RQ1 ########################

###### DATA CHECKS ######

str(dat1)
dat1$BaseFitness <- as_factor(dat1$BaseFitness)
summary(dat1)
levels(dat1$BaseFitness)

######  DESCRIBE & VISUALISE ######

dat1
```

```r
gof_plot <- ggplot(data = dat1, aes(x = BaseFitness, y = Count, fill = BaseFitness)) +
  geom_col()
gof_plot

######  GOF TEST ######

#Option 1
obs <- c(26, 36, 28) #counts of observed cases
exp <- c(1/3, 1/3, 1/3) #equal props
GOFtest1 <- chisq.test(x = obs, p = exp)
GOFtest1
GOFtest1$statistic

#Option 2
GOFtest2 <- chisq.test(x = dat1$Count) # p by default tests for equal
GOFtest2

### How to extract specific parts of output:

GOFtest2$expected #data requirement: check that our expected count is at least 5 in each category
GOFtest2$statistic #check our X2 stat
GOFtest2$residuals #use to examine residuals - non-sig result, so no need to do this

#what if you had a more specific idea of what the proportions would look like as a result of having a v

levels(dat1$BaseFitness) # need to match order of levels
GOFtest3 <- chisq.test(x = dat1$Count, p = c(0.25, 0.50, 0.25))
GOFtest3


###################### RQ2 ########################

###### DATA CHECKS ######
summary(dat2)

dat2$Group <- as_factor(dat2$Group)
dat2$Marathon <- as_factor(dat2$Marathon)

######  DESCRIBE & VISUALISE ######

table(dat2$Group, dat2$Marathon)

library(ggmosaic)
toi_plot <- ggplot(data = dat2) +
  geom_mosaic(aes(x = product(Group, Marathon), fill = Marathon))
toi_plot

######  TOI TEST ######

TOI <- chisq.test(dat2$Group, dat2$Marathon)
TOI
TOI$expected #data requirement: check that our expected count is at least 5 in each category
TOI$statistic #check our X2 stat
```

```r
TOI$residuals #examine residuals

######  EFFECT SIZE ######
cramers_v(TOI)

#not bias corrected
cramers_v(TOI, adjust = FALSE)


gof_plot
toi_plot
```