

# Step1 Team Project Multivariate Analysis

Adrian White, Cesar Conejo, Xavier Bryant

12/6/2020

## Introduction data set

We have selected the CRASH-2 data set provided by Vanderbilt School of Biostatistics for our project. It describes the outcome of a randomized controlled trial and economic valuation of the effects of tranexamic acid on death, vascular occlusive events, and transfusion requirement in bleeding trauma patients. Tranexamic acid reduces bleeding in trauma patients undergoing surgery but is an expensive treatment option. The trial's objective was to assess the effects and cost-effectiveness of an early administration of this medication.

Participants of the study were adults with, or at risk of, significant bleeding within 8 hours of injury. Sample randomization was determined by the allocation of an eight-digit sequence randomly generated by a computer. Patients and staff were masked to the treatment allocation of the tranexamic acid.

We have adjusted the original data set to remove some variables that were not relevant to our investigation. We have removed variables regarding the exact surgical procedures administered to patients, various IDs, and details on the patient outcome. We removed the health outcome columns because of complications regarding missing data, where the boolean structure of the columns relating to specific outcomes, like stroke or pulmonary embolism, left a large number of cases with missing values. Instead, we added a boolean variable for a general outcome of survival to assess the efficacy of the procedure, rather than looking at particular health outcomes in post-surgery for living patients.

We will be using variables regarding the sex, age, and injury of the patient as well as certain biometrics, like blood pressure, respiratory and heart rates, details on surgical blood transfusion, and a boolean variable on the survival of the patient. Our selection provides us with a balance of continuous and categorical variables, many of which are boolean, with minimal complications due to missing data.

## Summary variables in the data set

The variables in this dataset are the following:

1. entryid: (Numerical) Unique Numbers for Entry Forms
2. sex: (Boolean) The sex of the patient (Male/Female)
3. age : (Numerical) Age of the patient(Years)
4. injurytime: (Numerical) Hours since injury (Hours)
5. injurytype: (Categorical) Type of injury {Blunt, Penetrating, Blunt and Penetrating}
6. sbp: (Numerical) Systolic Blood Pressure (mmHg)
7. rr: (Numerical) Respiratory Rate (rate per minute)
8. cc: (Numerical) Central Capillary Refill Time (seconds)
9. hr: (Numerical) Heart Rate (rate per minute)
10. ndaysicu: (Numerical) Number of days in ICU (days)

11. Death: (Boolean) Indicator if the patient survived after the procedure

A summary of the data type is the following:

variable	type_variable	sub_type_variable
entryid	Quantitative	Continuous
sex	Qualitative	Nominal
age	Quantitative	Continuous
injurytime	Quantitative	Continuous
injurytype	Qualitative	Nominal
sbp	Quantitative	Continuous
rr	Quantitative	Continuous
cc	Quantitative	Continuous
hr	Quantitative	Continuous
ndaysicu	Quantitative	Discrete
death	Qualitative	Nominal

## Visual Analysis

## Sample Estimators

## Principal Component Analysis

```
load('rda/clinical_trial_complete.rda')
X <- clinical_trial_complete[, -c(1,11,13,14,15,16,18)]
X <- X[, -c(3,6,7,9,10)]
X_quan <- X[, -c(1,3,6)]

# Categorical variables
Y <- X[, c(1,3,6)]
X <- X_quan

# Remove the file
rm(clinical_trial_complete)
rm(X_quan)

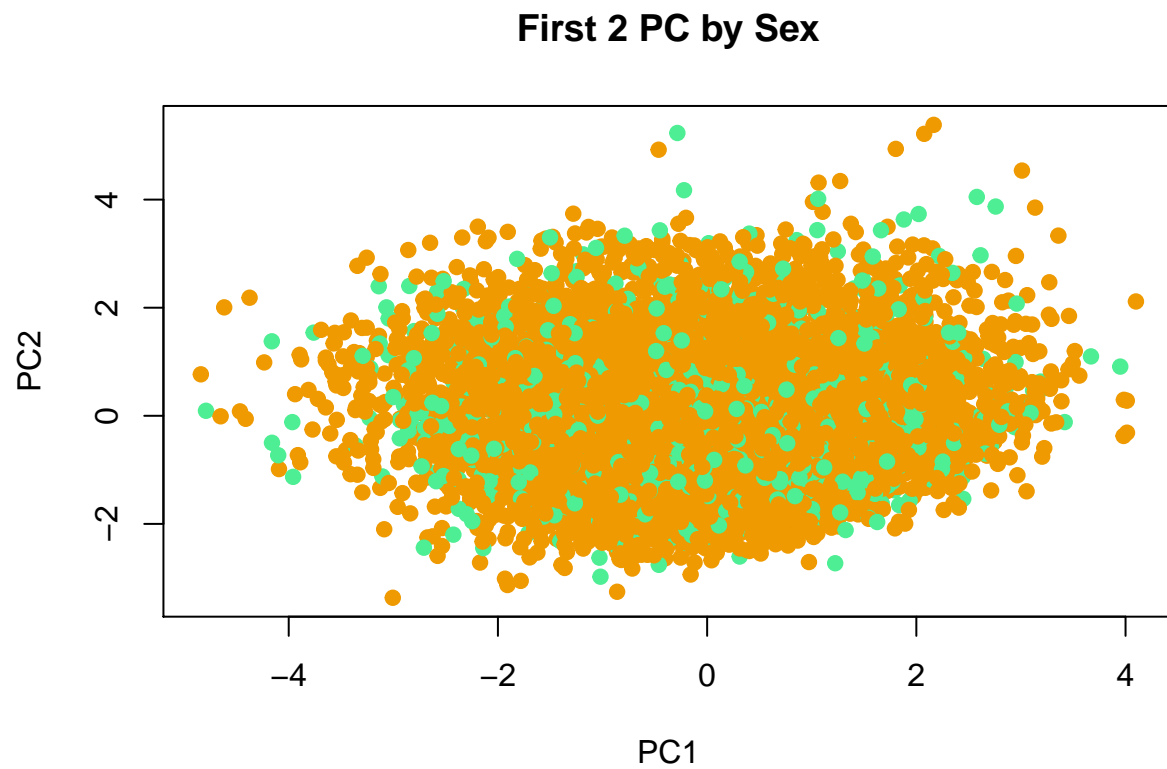
# Define colors for plots
color_1 <- "deepskyblue2"
color_2 <- "seagreen2"
color_3 <- "orange2"
color_4 <- "darkorchid4"

n <- nrow(X)
p <- ncol(X)

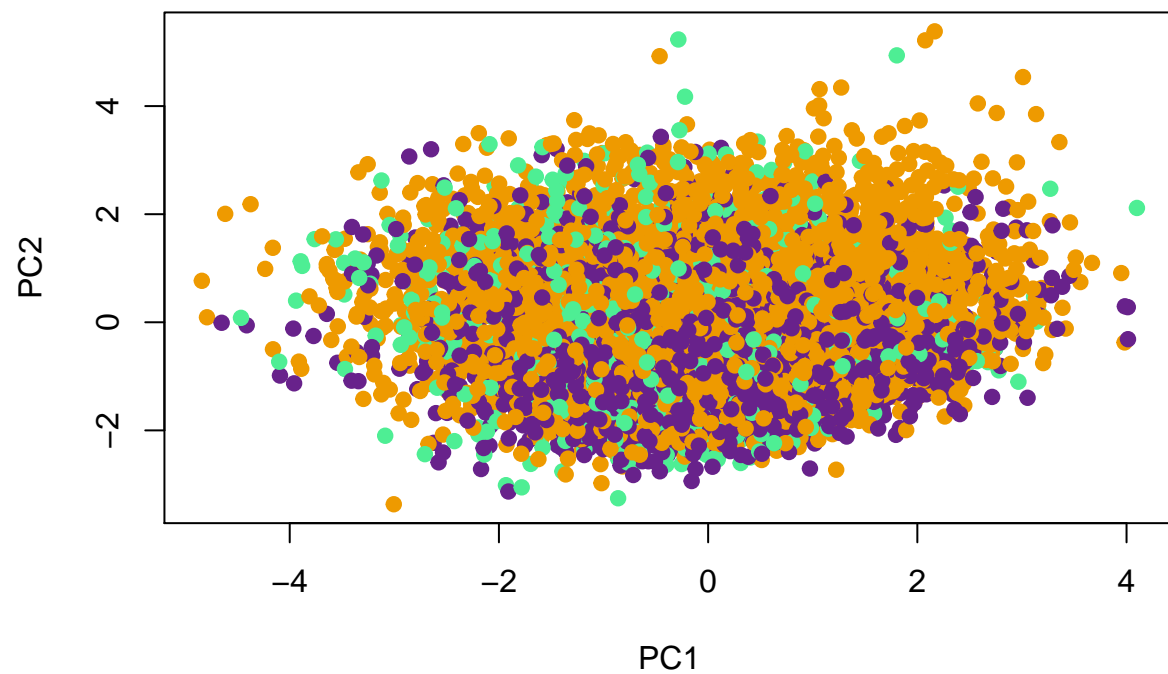
X_pcs <- prcomp(X, scale=TRUE)

##### SEX
colors_sex <- c(color_2, color_3)[1*(Y[,1]=="male")+1]
```

```
plot(X_pcs$x[,1:2],pch=19,col=colors_sex, main = "First 2 PC by Sex")
```



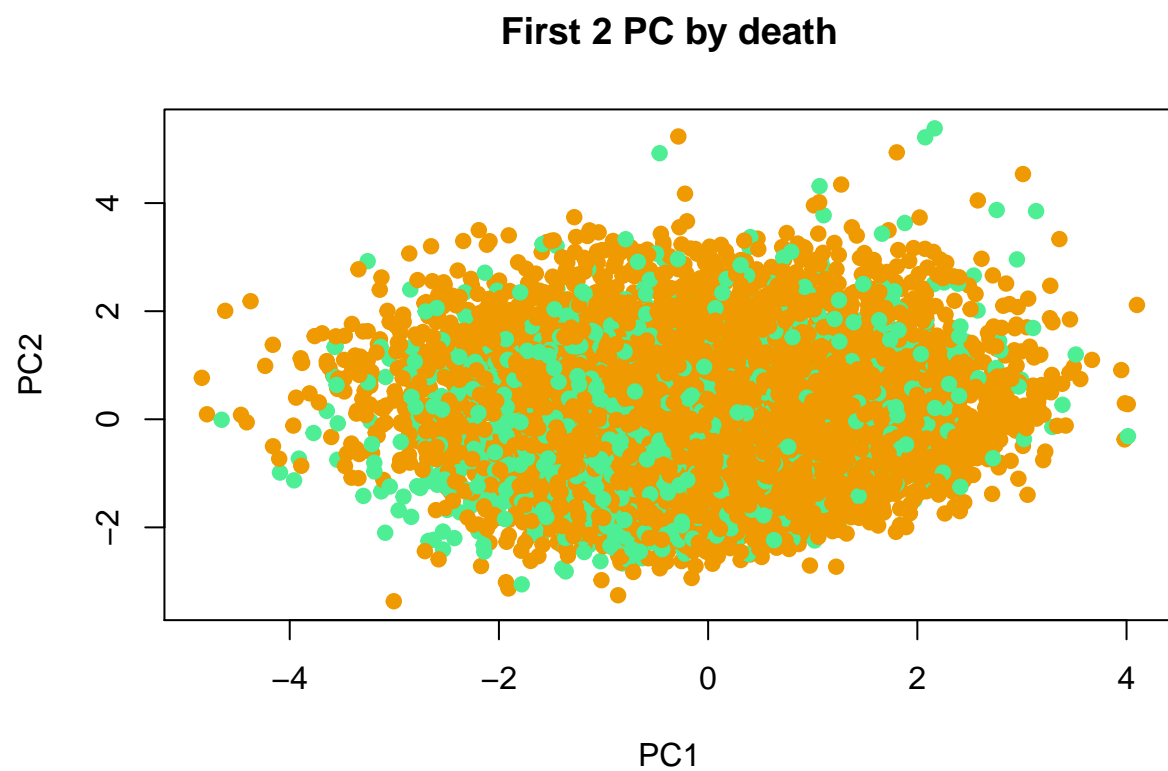
```
##### PAIN
colors_pain <- c(color_2,color_3,color_4)[1 + ifelse(Y[,2]== "blunt", 1,
                                                    ifelse(Y[,2]=="penetrating",2,
                                                            0))]
plot(X_pcs$x[,1:2],pch=19,col=colors_pain)
```



```
##### death 1 (death)
```

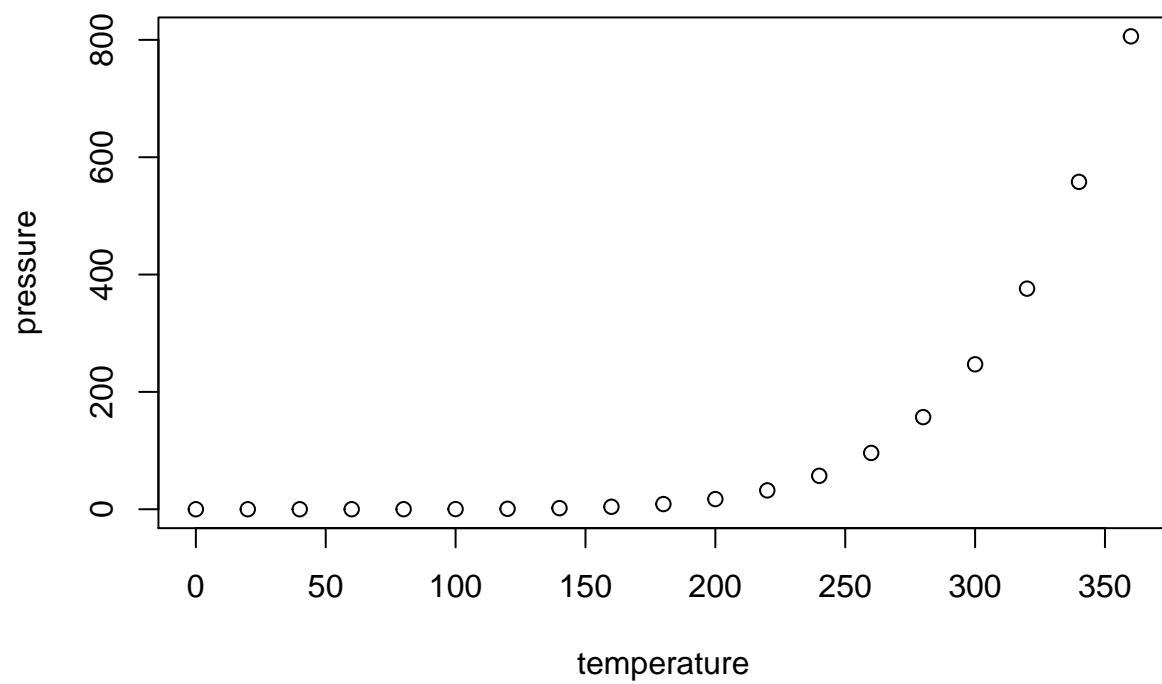
```
colors_death <- c(color_2,color_3)[1*(Y[,3]==0)+1]
```

```
plot(X_pcs$x[,1:2],pch=19,col=colors_death, main = "First 2 PC by death")
```



### Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.