ASSSIGNMENT

ANKUSH SHARMA

**Aggregate all parameters to a patient-level (using the aggregate function) by building means of all continuous parameters**

> head(df\_agg)

PATIENT\_ID max\_days NLR SBP DBP MALE ALBUMIN DIED

1 1 362 7.016651 140.5503 84.95289 0 5.035644 0.00000000

2 2 270 6.360841 139.6204 85.05419 1 4.264230 0.00862069

3 3 187 6.525698 139.9281 85.00840 1 5.383948 0.00000000

4 4 363 5.822375 140.3275 84.91455 0 4.510595 0.00000000

5 5 327 5.432448 139.9982 85.08366 1 4.222784 0.00000000

6 6 363 4.996217 139.9578 85.20283 1 4.875493 0.00000000

**Count the number of deaths**

14

**Was there a difference between SBP and DBP between those that died and those that didn’t?**

df = 97

a = .05

t-SBP = -0.396396 < t from table, so we do not have enough evidence to claim that there is a difference between SBP of those who died and those who didn’t.

t-DBP = -0.6371365 < t from table, so we do not have enough evidence to claim that there is a difference between DBP of those who died and those who didn’t.

**Did Albumin differ between men and women?**

a = .05

z-alb = -0.4471543

2 \* p(z < -0.4471543) = 2 \* .3264 = .653 > a, so we do not have enough evidence to

claim that Albumin differs between men and women.

**Did NLR differ between the highest and the lowest tertile of SBP?**

a = .05

z-NLR = 1.633019

2 \* p(z < 1.633019) = 2 \* .9484 = 1.8968 > a, we do not have enough evidence to claim that NLR differs between the highest and lowest tertiles of SBP.

|  |
| --- |
| rm(list=ls())  pathname<-"C:/Users/Raffi Wiesen/Desktop/BTM6000 Spring 2020/Class 9/"  df<-read.csv(paste(pathname,"Dataset Participation Lab 6.csv",sep=""))  # number of deaths  sum(df$DIED)  14  # max number of days a patient 1 was around  max(df$DAYS\_FROM\_FDD[df$PATIENT\_ID==1])  362  # Aggregate all parameters to a patient-level (using the aggregate function) by building means of all continuous parameters  # data frame of unique patient entries - one entry for each patient  df\_agg <- data.frame("PATIENT\_ID" = unique(df$PATIENT\_ID))  # aggregates by list of ID  df\_temp\_max <- aggregate(df$DAYS\_FROM\_FDD, by = list(df$PATIENT\_ID), FUN =  + max)  colnames(df\_temp\_max) <- c("PATIENT\_ID", "max\_days")  df\_temp\_NLR <- aggregate(df$NLR, by = list(df$PATIENT\_ID), FUN = mean, na.rm =  + TRUE)  colnames(df\_temp\_NLR) <- c("PATIENT\_ID", "NLR")  df\_temp\_SBP <- aggregate(df$SBP, by = list(df$PATIENT\_ID), FUN = mean, na.rm =  + TRUE)  colnames(df\_temp\_SBP) <- c("PATIENT\_ID", "SBP")  df\_temp\_DBP <- aggregate(df$DBP, by = list(df$PATIENT\_ID), FUN = mean, na.rm =  + TRUE)  colnames(df\_temp\_DBP) <- c("PATIENT\_ID", "DBP")  df\_temp\_MALE <- aggregate(df$MALE, by = list(df$PATIENT\_ID), FUN = mean,  + na.rm = TRUE)  colnames(df\_temp\_MALE) <- c("PATIENT\_ID", "MALE")  df\_temp\_ALBUMIN <- aggregate(df$ALBUMIN, by = list(df$PATIENT\_ID), FUN =  + mean, na.rm = TRUE)  colnames(df\_temp\_ALBUMIN) <- c("PATIENT\_ID", "ALBUMIN")  df\_temp\_DIED <- aggregate(df$DIED, by = list(df$PATIENT\_ID), FUN = mean, na.rm  + = TRUE)  colnames(df\_temp\_DIED) <- c("PATIENT\_ID", "DIED")  library(dplyr)  # merge tables by shared characteristic  df\_agg <- left\_join(df\_agg, df\_temp\_max, by = "PATIENT\_ID")  df\_agg <- left\_join(df\_agg, df\_temp\_NLR, by = "PATIENT\_ID")  df\_agg <- left\_join(df\_agg, df\_temp\_SBP, by = "PATIENT\_ID")  df\_agg <- left\_join(df\_agg, df\_temp\_DBP, by = "PATIENT\_ID")  df\_agg <- left\_join(df\_agg, df\_temp\_MALE, by = "PATIENT\_ID")  df\_agg <- left\_join(df\_agg, df\_temp\_ALBUMIN, by = "PATIENT\_ID")  df\_agg <- left\_join(df\_agg, df\_temp\_DIED, by = "PATIENT\_ID")  head(df\_agg)    PATIENT\_ID max\_days NLR SBP DBP MALE ALBUMIN DIED  1 1 362 7.016651 140.5503 84.95289 0 5.035644 0.00000000  2 2 270 6.360841 139.6204 85.05419 1 4.264230 0.00862069  3 3 187 6.525698 139.9281 85.00840 1 5.383948 0.00000000  4 4 363 5.822375 140.3275 84.91455 0 4.510595 0.00000000  5 5 327 5.432448 139.9982 85.08366 1 4.222784 0.00000000  6 6 363 4.996217 139.9578 85.20283 1 4.875493 0.00000000  # SBP tertiles for NLR  df\_agg$tertiles\_SBP <- factor(ntile(df\_agg$SBP, 3), labels = c("tert 1", "tert 2", "tert 3"))  df\_agg$tertiles\_SBP <- relevel(df\_agg$tertiles\_SBP, "tert 1")  summary(df\_agg$tertiles\_SBP)  tert 1 tert 2 tert 3  33 33 33  mean(df\_agg$NLR)  5.405617  mean(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 1"])  5.273648  mean(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 2"])  5.229242  mean(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 3"])  5.713961  sd(df\_agg$NLR)  1.289131  sd(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 1"])  0.9420905  sd(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 2"])  1.389552  sd(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 3"])  1.45704  # put into table  Demographics <- data.frame("Parameter" = unique(df\_agg$tertiles\_SBP), "NLR\_mean"  + = NA, "NLR\_sd" = NA)  Demographics[1,2] <- mean(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 1"])  Demographics[1,3] <- sd(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 1"])  Demographics[2,2] <- mean(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 2"])  Demographics[2,3] <- sd(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 2"])  Demographics[3,2] <- mean(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 3"])  Demographics[3,3] <- sd(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 3"])  # Was there a difference between SBP and DBP between those that died and those that didn’t?  mean\_SBPdied <- mean(df\_agg$SBP[df\_agg$DIED != "0"])  mean\_SBPdidnt <- mean(df\_agg$SBP[df\_agg$DIED == "0"])  sd\_SBPdied <- sd(df\_agg$SBP[df\_agg$DIED != "0"])  sd\_SBPdidnt <- sd(df\_agg$SBP[df\_agg$DIED == "0"])  mean\_DBPdied <- mean(df\_agg$DBP[df\_agg$DIED != "0"])  mean\_DBPdidnt <- mean(df\_agg$DBP[df\_agg$DIED == "0"])  sd\_DBPdied <- sd(df\_agg$DBP[df\_agg$DIED != "0"])  sd\_DBPdidnt <- sd(df\_agg$DBP[df\_agg$DIED == "0"])  n\_died <- nrow(subset(df\_agg, DIED != "0"))  n\_didnt <- nrow(subset(df\_agg, DIED == "0"))  tSBP <- (mean\_SBPdied - mean\_SBPdidnt)/ sqrt((sd\_SBPdied/ n\_died) + (sd\_SBPdidnt/  + n\_didnt))  tDBP <- (mean\_DBPdied - mean\_DBPdidnt)/ sqrt((sd\_DBPdied/ n\_died) +  + (sd\_DBPdidnt/ n\_didnt))  # Did Albumin differ between men and women?  mean\_alb\_men <- mean(df\_agg$ALBUMIN[df\_agg$MALE == "1"])  mean\_alb\_women <- mean(df\_agg$ALBUMIN[df\_agg$MALE == "0"])  sd\_alb\_men <- sd(df\_agg$ALBUMIN[df\_agg$MALE == "1"])  sd\_alb\_women <- sd(df\_agg$ALBUMIN[df\_agg$MALE == "0"])  n\_men <- nrow(subset(df\_agg, MALE =="1"))  n\_women <- nrow(subset(df\_agg, MALE =="0"))  z\_alb <- (mean\_alb\_men - mean\_alb\_women)/ (sqrt((sd\_alb\_men/n\_men) +  + (sd\_alb\_women/n\_women)))  # Did NLR differ between the highest and the lowest tertile of SBP?  mean\_NLR\_1 <- mean(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 1"])  mean\_NLR\_3 <- mean(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 3"])  sd\_NLR\_1 <- sd(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 1"])  sd\_NLR\_3 <- sd(df\_agg$NLR[df\_agg$tertiles\_SBP == "tert 3"])  n1 <- nrow(subset(df\_agg, tertiles\_SBP == "tert 1"))  n3 <- nrow(subset(df\_agg, tertiles\_SBP == "tert 3"))  zNLR <- (mean\_NLR\_3 - mean\_NLR\_1)/ (sqrt((sd\_NLR\_3/n3) + (sd\_NLR\_1/n1))) |
|  |
| |  | | --- | | > | |