* inspect the [Lab Week 6 - NHANES subset.csvPreview the document](https://yu.instructure.com/courses/39395/files/1245823/download?wrap=1)(n=1000) by computing descriptive statistics characterizing the cohort.
* Categorize the cohort in CKD stages (see criteria in the table below)

|  |  |
| --- | --- |
| Chronic kidney disease (CKD) stage | Criterion |
| Stage 1 | eGFR ≥ 90 mL/min/1.73m2 and ACR ≥ 30 mg/g |
| Stage 2 | eGFR ≥ 60 to 89 mL/min/1.73m2 |
| Stage 3 | eGFR ≥ 30 to 59 mL/min/1.73m2 |
| Stage 4 | eGFR ≥ 15 to 29 mL/min/1.73m2 |
| Stage 5 | eGFR < 15 mL/min/1.73m2 |

* Characterize subjects in the respective CKD stages (merge CKD stage 3 to 5).
* Create contingency tables categorizing subjects with CKD stage 3 or worse against elevated creatinine (reference range for Women 0.5 to 1.1 mg/dL and 0.6 to 1.2 for men).
* Calculate sensitivity, specificity, PPV and NPV of elevated creatinine to diagnose CKD stage 3 or worse
* Construct a ROC curve

# Clean the environment and load file

rm(list = ls())

pathname<-"C:/BME/Sem 2/Biostat/Class 6/"

df<-read.csv(paste(pathname,"Lab Week 6 - NHANES subset.csv",sep=""))

ls(df)

A screenshot of a cell phone

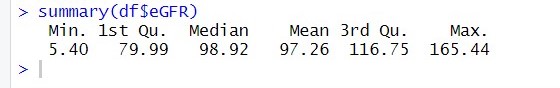
Description automatically generated

hist(df$eGFR)

A screenshot of a cell phone

Description automatically generated

summary(df$eGFR)



df[1,]

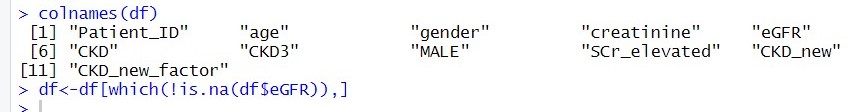
df[,1]



colnames(df[1])

colnames(df)

df<-df[which(!is.na(df$eGFR)),]



df$CKD\_new<-NA

df$CKD\_new<-ifelse(df$eGFR<60,1,0)

summary(df$CKD\_new)

df$CKD\_new\_factor<-factor(df$CKD\_new,labels = c("no CKD","CKD"))

summary(df$CKD\_new\_factor)

A screenshot of a cell phone

Description automatically generated

Create contingency tables categorizing subjects with CKD stage 3 or worse against elevated creatinine (reference range for Women 0.5 to 1.1 mg/dL and 0.6 to 1.2 for men)

ls(df)

summary(df$SCr\_elevated)

df$SCr\_elevated<-factor(df$SCr\_elevated,labels = c("elevated","not elevated"))

ctable<-table(df$SCr\_elevated,df$CKD\_new\_factor)

A screenshot of a cell phone

Description automatically generated

Calculate sensitivity, specificity, PPV and NPV of elevated creatinine to diagnose CKD stage 3 or worse

Sensitivty <-ctable[4]/(ctable[4]+ctable[3])

specifcity <- ctable[1]/(ctable[1]+ctable[2])

Construct a ROC curve

library(pROC)

roc(df$CKD3,df$creatinine,plot=TRUE)

A screenshot of a cell phone

Description automatically generated