# Assessment03

STAT414

2024-12-18

```r
library(EnvStats)
```

```
##
## Attaching package: 'EnvStats'

## The following objects are masked from 'package:stats':
##
##     predict, predict.lm
```

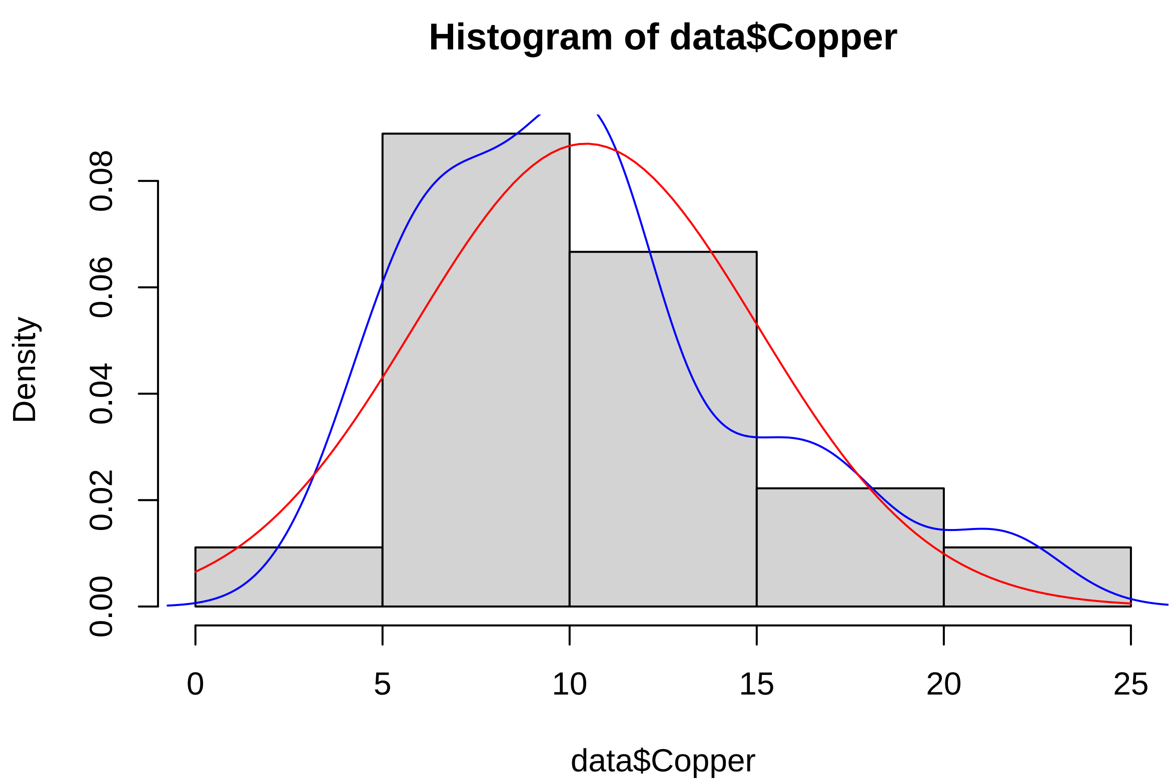
```r
# 1. Consider the copper data provided in the package EnvStats in the data frame EPA.92c.copper1.df.
data <- EPA.92c.copper1.df
data
```

```
##    Copper Month Well  Well.type
## 1     4.2     1    1 Background
## 2     5.8     2    1 Background
## 3    11.3     3    1 Background
## 4     7.0     4    1 Background
## 5     7.3     5    1 Background
## 6     8.2     6    1 Background
## 7     5.2     1    2 Background
## 8     6.4     2    2 Background
## 9    11.2     3    2 Background
## 10   11.5     4    2 Background
## 11   10.1     5    2 Background
## 12    9.7     6    2 Background
## 13    9.4     1    3 Compliance
## 14   10.9     2    3 Compliance
## 15   14.5     3    3 Compliance
## 16   16.1     4    3 Compliance
## 17   21.5     5    3 Compliance
## 18   17.6     6    3 Compliance
```

```r
# (a) Create a normal probability plot by combining the data from both background
# wells to determine whether these data may be adequately modeled with a normal
# distribution.
hist(data$Copper, probability=TRUE)
lines(density(data$Copper), add=TRUE, col="blue")
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "add" is not a graphical
## parameter
```

```r
curve(dnorm(x,mean=mean(data$Copper), sd=sd(data$Copper)),add=TRUE,col="red")
```




```r
skewness(data$Copper)
```

```
## [1] 0.9159713
```

```r
kurtosis(data$Copper)
```

```
## [1] 0.5889153
```

```r
cat("The histogram and superimpose normal curve show that the data could possibly be modeled
    with normal distribution, and the skewness and kurtosis which are both relatively close
    to 0 show that the data can be modeled with normal distribution.")
```

```
## The histogram and superimpose normal curve show that the data could possibly be modeled
##     with normal distribution, and the skewness and kurtosis which are both relatively close
##     to 0 show that the data can be modeled with normal distribution.
```

```r
# (b) Assume there is no substantial spatial or temporal variability in the copper concentrations.
# Also, assume the data at the compliance well will be tested monthly, and use the
# background well data to create a one-sided upper 95% prediction interval for the
# next k = 6 future observations, and compare the data from the compliance well to
```

```r
# the upper prediction limit. Is there any evidence of contamination based on this method?

compliance_data <- data[data$Well.type=="Compliance",]
background_data <- data[data$Well.type=="Background",]

k = 6
alpha = 1 - (1-.05)^(1/k)
conf = .95
n <- length(background_data)

xbar = mean(background_data$Copper)
s <- sd(background_data$Copper)

pub <- xbar + ( qt(.95, df=n-1) * s * sqrt(1/n + 1/k))
plb <- xbar + ( qt(.05, df=n-1) * s * sqrt(1/n + 1/k))

cat("(", plb, ",", pub, ")")
```

```
## ( 4.282985 , 12.03368 )
```

```r
cat("The upper 95% prediction limit of the background well is a concentration of",
    pub, "
    while the mean concentration of the compliance well is",
    mean(compliance_data$Copper), "The concentration of the complaince
    well is higher than the upper prediction limit for the background wells, therefore
    it is appropriate to say that it is likely that the compliance wells are
    contaminated.")
```

```
## The upper 95% prediction limit of the background well is a concentration of 12.03368
##     while the mean concentration of the compliance well is 15 The concentration of the complaince
##     well is higher than the upper prediction limit for the background wells, therefore
##     it is appropriate to say that it is likely that the compliance wells are
##     contaminated.
```

```r
# (c) Compute a one-sided upper 95% beta-content tolerance interval with associated
#confidence level 95% based on the background data, and compare the data from the
#compliance well with the upper tolerance limit.
beta <- .95
p1 <- .5 - beta/2
p2 <- .5 + beta/2

clb <- qnorm(p1, mean=mean(background_data$Copper), sd=sd(background_data$Copper))
cub <- qnorm(p2, mean=mean(background_data$Copper), sd=sd(background_data$Copper))
clb
```

```
## [1] 3.158271
```

```r
cub
```

```
## [1] 13.1584
```

```r
cat("The upper 95% beta-content tolerance bound of the background well is a
    concentration of", cub, "while the mean concentration of the compliance well is",
    mean(compliance_data$Copper), "
    .The concentration of the complaince well is
    greater and than the upper tolerance bound for the background wells, and outside
    of the interval, therefore it is appropriate to say that the compliance wells are
    contaminated.")
```

```
## The upper 95% beta-content tolerance bound of the background well is a
##     concentration of 13.1584 while the mean concentration of the compliance well is 15
##     .The concentration of the complaince well is
##     greater and than the upper tolerance bound for the background wells, and outside
##     of the interval, therefore it is appropriate to say that the compliance wells are
##     contaminated.
```

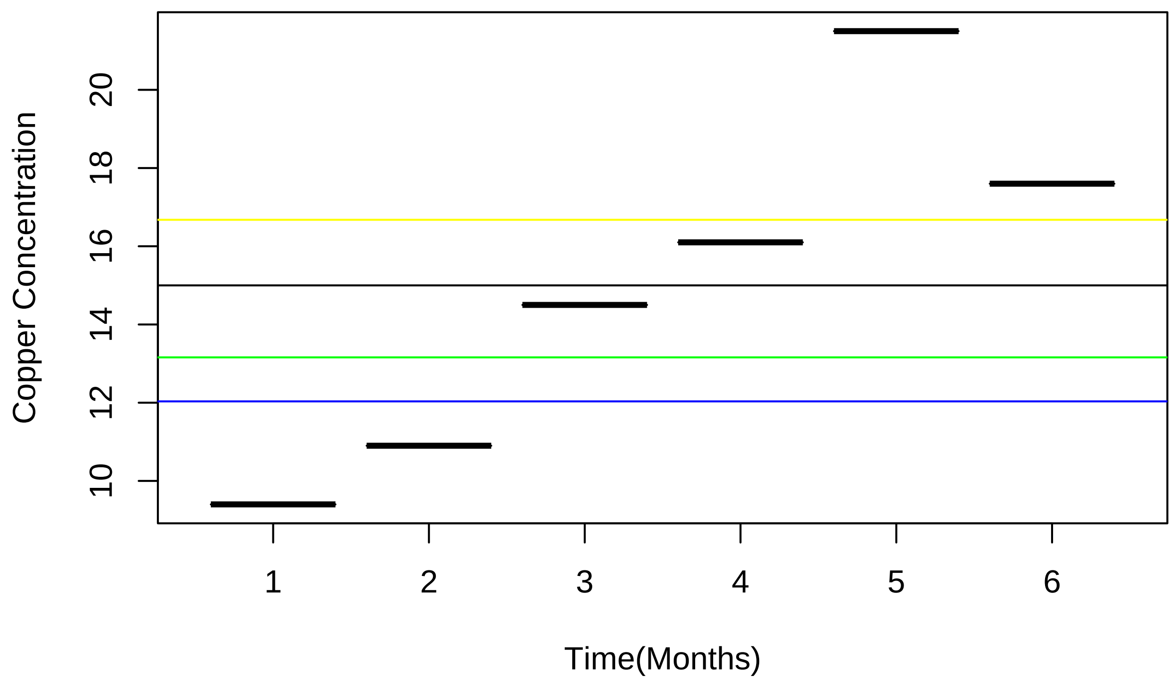
```r
# 2. Continuing with the copper concentration data analysis from # above:
# (a) Plot the compliance well data vs. time, and add a horizontal line at the upper
# prediction limit to the plot. Add a horizontal line at the upper simultaneous
# prediction limit to the plot you created in part c above. Add a horizontal line at
# the upper tolerance limit to the plot you created in part c above.

simultaneous <- predIntNorm(background_data$Copper, n.mean = 1, k = 6,
                            method = "Bonferroni",
                            pi.type = "two-sided", conf.level = 0.95)
simultaneous
```

```
##
## Results of Distribution Parameter Estimation
## --------------------------------------------
##
## Assumed Distribution:            Normal
##
## Estimated Parameter(s):          mean = 8.158333
##                                  sd   = 2.551099
##
## Estimation Method:               mvue
##
## Data:                            background_data$Copper
##
## Sample Size:                     12
##
## Prediction Interval Method:      Bonferroni
##
## Prediction Interval Type:        two-sided
##
## Confidence Level:                95%
##
## Number of Future Observations:   6
##
## Prediction Interval:             LPL = -0.3600913
##                                  UPL = 16.6767580
```

```r
sub <- 16.6767580

plot(compliance_data$Month, compliance_data$Copper, type="l",
     ylab="Copper Concentration", xlab="Time(Months)",
     col="red")
abline(h=pub, col= "blue")
abline(h=cub, col="green")
abline(h=sub, col="yellow")
abline(h=mean(compliance_data$Copper))
```




```r
# (b) Explain the difference between the upper prediction limit, the upper simultaneous
# prediction limit, and the upper tolerance limit. What does each limit assume?

cat("The upper prediction limit says that for any future observation from the data
    of a population, the future observation will have a 100(1-alpha)% probability
    to be less than upper prediction limit and greater than the lower prediction
    limit.

    The upper simultaneous prediction limit is for the upper bound of the simultaneous
    intervals in ANOVA. It says that the future observation will have a 100(1-alpha)%
    probability to be less than upper prediction limit and greater than the lower prediction limit.
    For the mean value of the copper concentration of the compliance wells, this is the
    only limit where the compliance concentrations are within the interval for, and lower
    than the upper simultaneous prediciton limit.
```

```
    The upper tolerance limit says that we are 100(1-alpha)% confident that the
    proportion of the population covered by the tolerance interval is Beta*100%
    100(1-alpha)% of all the Beta tolerance intervals generated contains at least
    Beta*100% of the population.")
```

```
## The upper prediction limit says that for any future observation from the data
##     of a population, the future observation will have a 100(1-alpha)% probability
##     to be less than upper prediction limit and greater than the lower prediction
##     limit.
##
##     The upper simultaneous prediction limit is for the upper bound of the simultaneous
##     intervals in ANOVA. It says that the future observation will have a 100(1-alpha)%
##     probability to be less than upper prediction limit and greater than the lower prediction limit.
##     For the mean value of the copper concentration of the compliance wells, this is the
##     only limit where the compliance concentrations are within the interval for, and lower
##     than the upper simultaneous prediciton limit.
##
##     The upper tolerance limit says that we are 100(1-alpha)% confident that the
##     proportion of the population covered by the tolerance interval is Beta*100%
##     100(1-alpha)% of all the Beta tolerance intervals generated contains at least
##     Beta*100% of the population.
```