

Produit final fonctionnel

Prédiction du pourcentage (%) de gras

Nom, prénom: Lavergne, Bruno

No étudiant:22101859

Trimestre d'inscription: Hiver 2024

Date d'envoi : 08-04-2024

Description du “Dataset”

Body Fat Prediction :

Les données proviennent du Dr. A. Garth Fisher, FACSM, Human Performance Research Center, Brigham Young University, Provo, Utah.

Les données sont regroupées en 15 variables, notamment :

1. La densité (déterminée par la pesée sous l'eau)
2. Pourcentage de graisse corporelle (à partir de l'équation de Siri (1956))
3. Âge (années)
4. Poids (livres)
5. Tailles (pouces)
6. Circonférence du cou (cm)
7. Circonférence de la poitrine (cm)
8. Circonférence de l'abdomaine (cm)
9. Circonférence de taille (cm)
10. Circonférence de la cuisse (cm)
11. Circonférence du genou (cm)
12. Circonférence de la cheville (cm)
13. Circonférence du biceps (cm)
14. Circonférence de l'avant-bras (cm)
15. Circonférence du poignet (cm)

Nettoyage des données

```
library(readr)
bodyfat <- read_csv("C:/Users/Bruno/Documents/Sciences_de_donnees/SCI1402/Body_Fat_Prediction/bodyfat.csv")
```

Importer les données provenant de Kaggle.

```
## Rows: 252 Columns: 15
## -- Column specification -----
## Delimiter: ","
## dbl (15): Density, BodyFat, Age, Weight, Height, Neck, Chest, Abdomen, Hip, ...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
summary(bodyfat)
```

Aperçu du jeu de données

```
##      Density      BodyFat      Age      Weight
## Min.   :0.995   Min.   : 0.00   Min.   :22.00   Min.   :118.5
## 1st Qu.:1.041   1st Qu.:12.47   1st Qu.:35.75   1st Qu.:159.0
## Median :1.055   Median :19.20   Median :43.00   Median :176.5
## Mean   :1.056   Mean   :19.15   Mean   :44.88   Mean   :178.9
## 3rd Qu.:1.070   3rd Qu.:25.30   3rd Qu.:54.00   3rd Qu.:197.0
## Max.   :1.109   Max.   :47.50   Max.   :81.00   Max.   :363.1
##      Height      Neck      Chest      Abdomen
## Min.   :29.50   Min.   :31.10   Min.   : 79.30   Min.   : 69.40
## 1st Qu.:68.25   1st Qu.:36.40   1st Qu.: 94.35   1st Qu.: 84.58
## Median :70.00   Median :38.00   Median : 99.65   Median : 90.95
## Mean   :70.15   Mean   :37.99   Mean   :100.82   Mean   : 92.56
## 3rd Qu.:72.25   3rd Qu.:39.42   3rd Qu.:105.38   3rd Qu.: 99.33
## Max.   :77.75   Max.   :51.20   Max.   :136.20   Max.   :148.10
##      Hip      Thigh      Knee      Ankle      Biceps
## Min.   : 85.0   Min.   :47.20   Min.   :33.00   Min.   :19.1   Min.   :24.80
## 1st Qu.: 95.5   1st Qu.:56.00   1st Qu.:36.98   1st Qu.:22.0   1st Qu.:30.20
## Median : 99.3   Median :59.00   Median :38.50   Median :22.8   Median :32.05
## Mean   : 99.9   Mean   :59.41   Mean   :38.59   Mean   :23.1   Mean   :32.27
## 3rd Qu.:103.5   3rd Qu.:62.35   3rd Qu.:39.92   3rd Qu.:24.0   3rd Qu.:34.33
## Max.   :147.7   Max.   :87.30   Max.   :49.10   Max.   :33.9   Max.   :45.00
##      Forearm      Wrist
## Min.   :21.00   Min.   :15.80
## 1st Qu.:27.30   1st Qu.:17.60
## Median :28.70   Median :18.30
## Mean   :28.66   Mean   :18.23
## 3rd Qu.:30.00   3rd Qu.:18.80
## Max.   :34.90   Max.   :21.40
```

```
any(is.na(bodyfat))
```

Vérifier s'il manque des valeurs dans le tableau.

```
## [1] FALSE
```

Supprimer les valeurs avec un “bodyfat” sous 1%.

Supprimer les données avec un âge supérieur a 60 ans

L âge maximal pour un membre des Forces Armées Canadiennes est de 60 ans.

```
bodyfat
```

```
## # A tibble: 252 x 15
##   Density BodyFat Age Weight Height Neck Chest Abdomen Hip Thigh Knee
##   <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  1.07    12.3  23  154.  67.8  36.2  93.1  85.2  94.5  59  37.3
## 2  1.09     6.1  22  173.  72.2  38.5  93.6  83  98.7  58.7  37.3
## 3  1.04    25.3  22  154  66.2  34  95.8  87.9  99.2  59.6  38.9
## 4  1.08    10.4  26  185.  72.2  37.4 102.  86.4 101.  60.1  37.3
## 5  1.03    28.7  24  184.  71.2  34.4  97.3 100  102.  63.2  42.2
## 6  1.05    20.9  24  210.  74.8  39  104.  94.4 108.  66  42
## 7  1.05    19.2  26  181  69.8  36.4 105.  90.7 100.  58.4  38.3
## 8  1.07    12.4  25  176  72.5  37.8  99.6  88.5  97.1  60  39.4
## 9  1.09     4.1  25  191  74  38.1 101.  82.5  99.9  62.9  38.3
## 10 1.07    11.7  23  198.  73.5  42.1  99.6  88.6 104.  63.1  41.7
## # ... with 242 more rows, and 4 more variables: Ankle <dbl>, Biceps <dbl>,
## # Forearm <dbl>, Wrist <dbl>
```

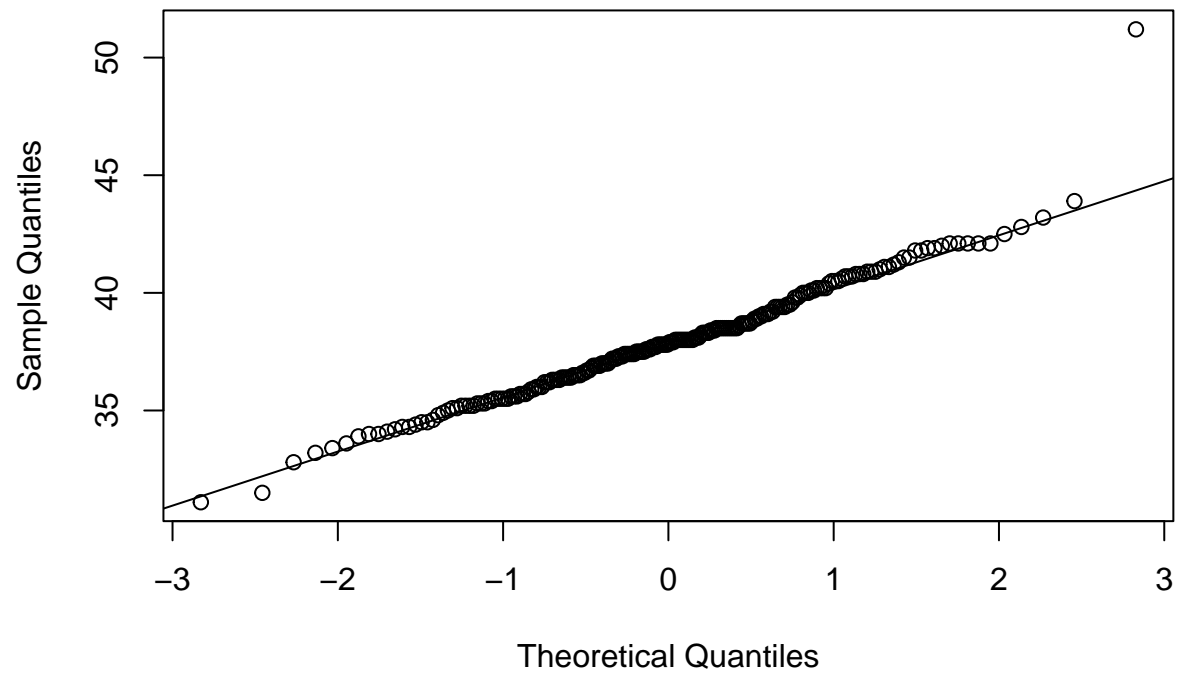
```
bodyfat <- subset(bodyfat, BodyFat >=1)
bodyfat <- subset(bodyfat, Age < 60)
```

Test de normalité (Shapiro-Wilk) des données avec les divers mensurations. Vérifier si les données représentent une courbe normale pour faire des prédictions.

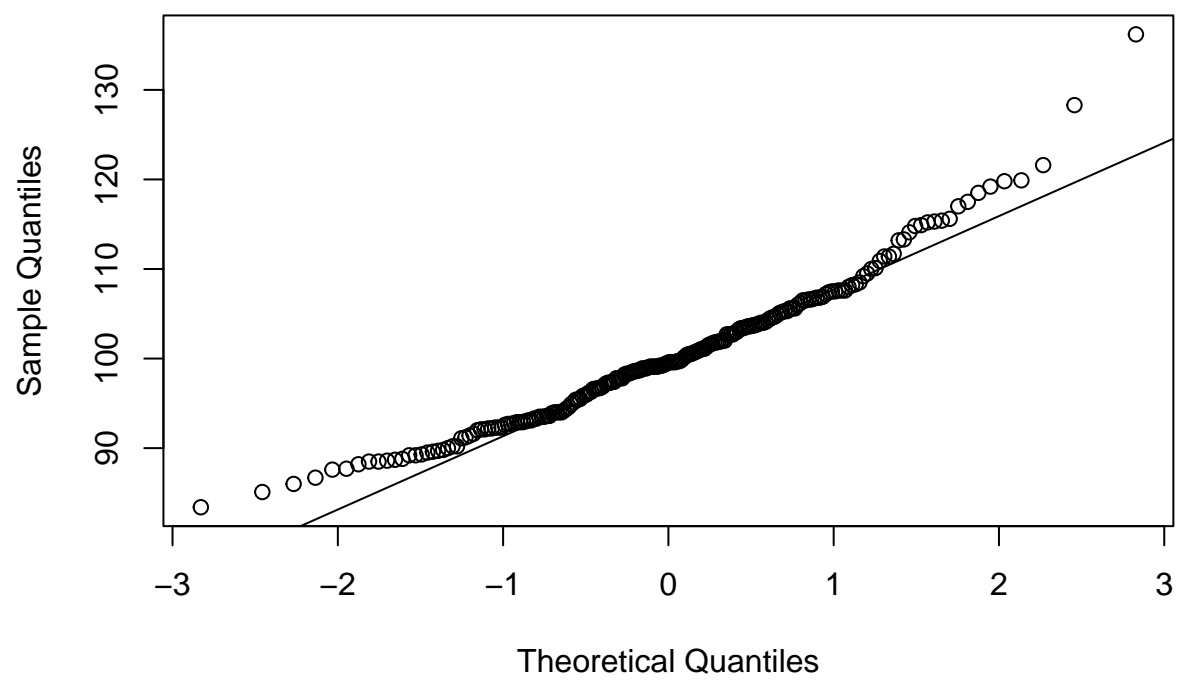
```
#Le graphique quantile-quantile permet de visualiser la normalité des résidus.
normality_tests <- list()

par(mfrow=c(1, 1))
for (col in names(bodyfat)[-c(1:5)]) {
  qqnorm(bodyfat[[col]], main = paste("QQ plot de", col))
  qqline(bodyfat[[col]])
}
```

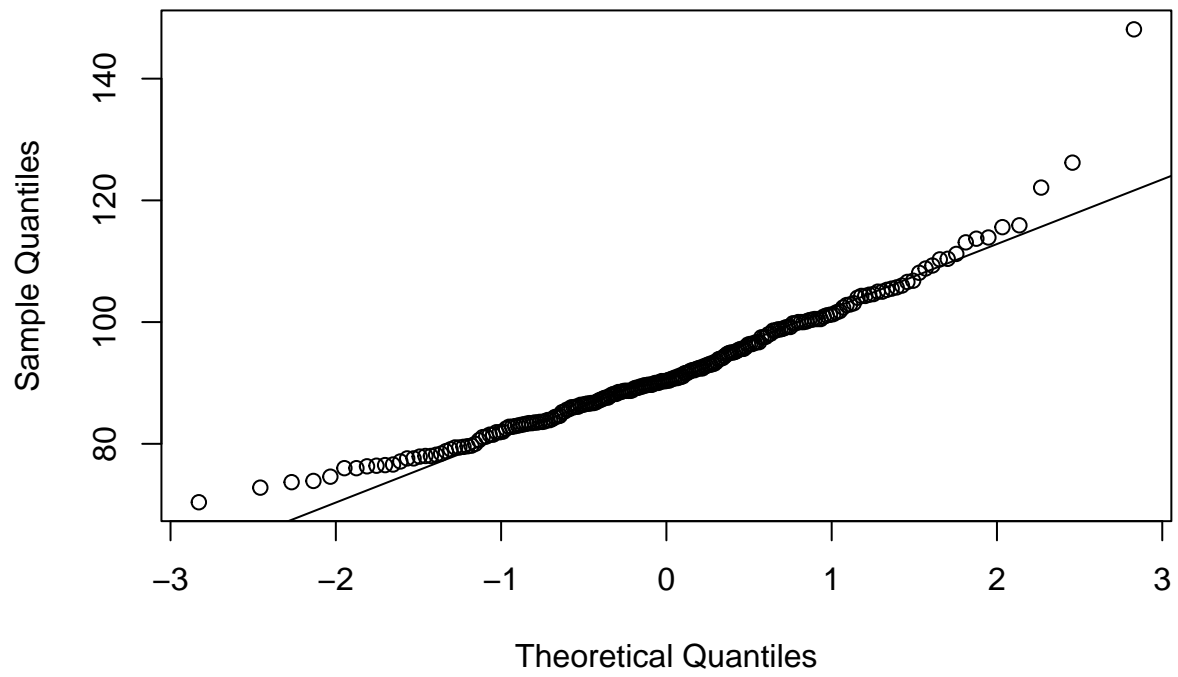
QQ plot de Neck



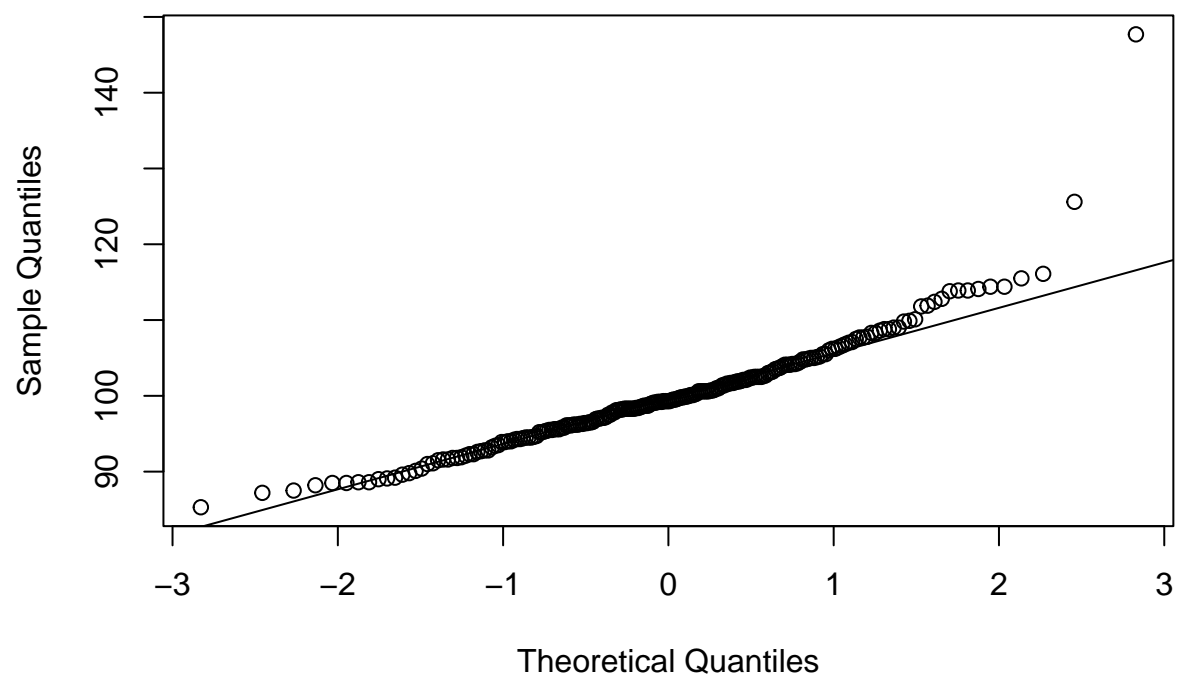
QQ plot de Chest



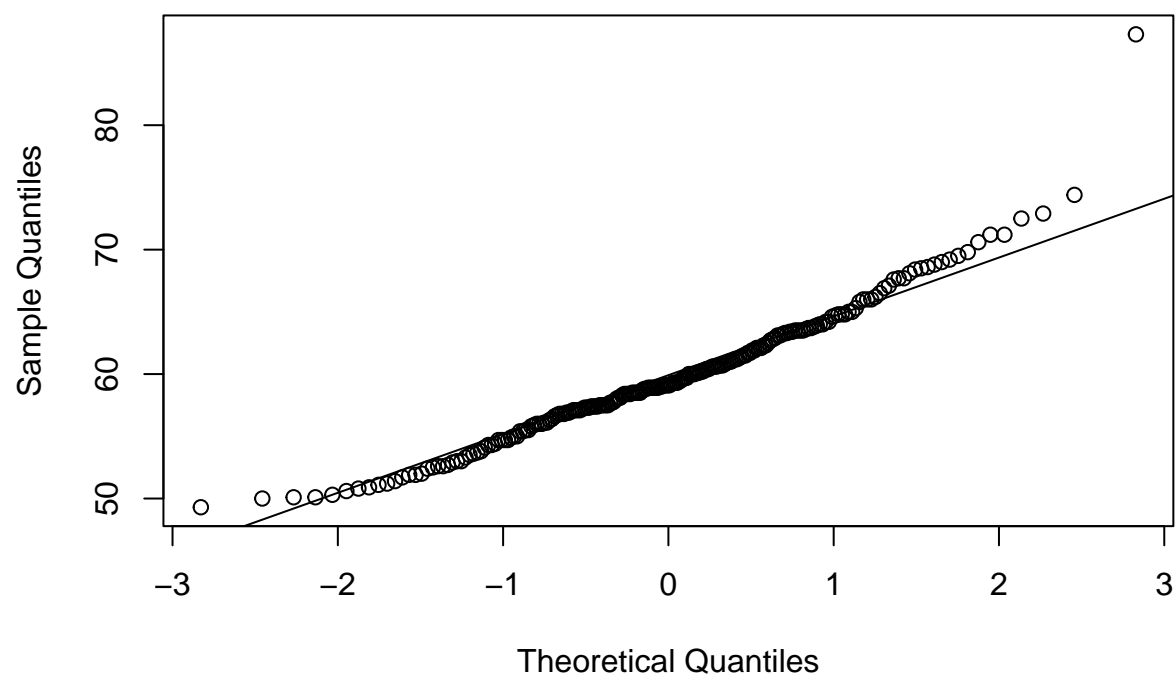
QQ plot de Abdomen



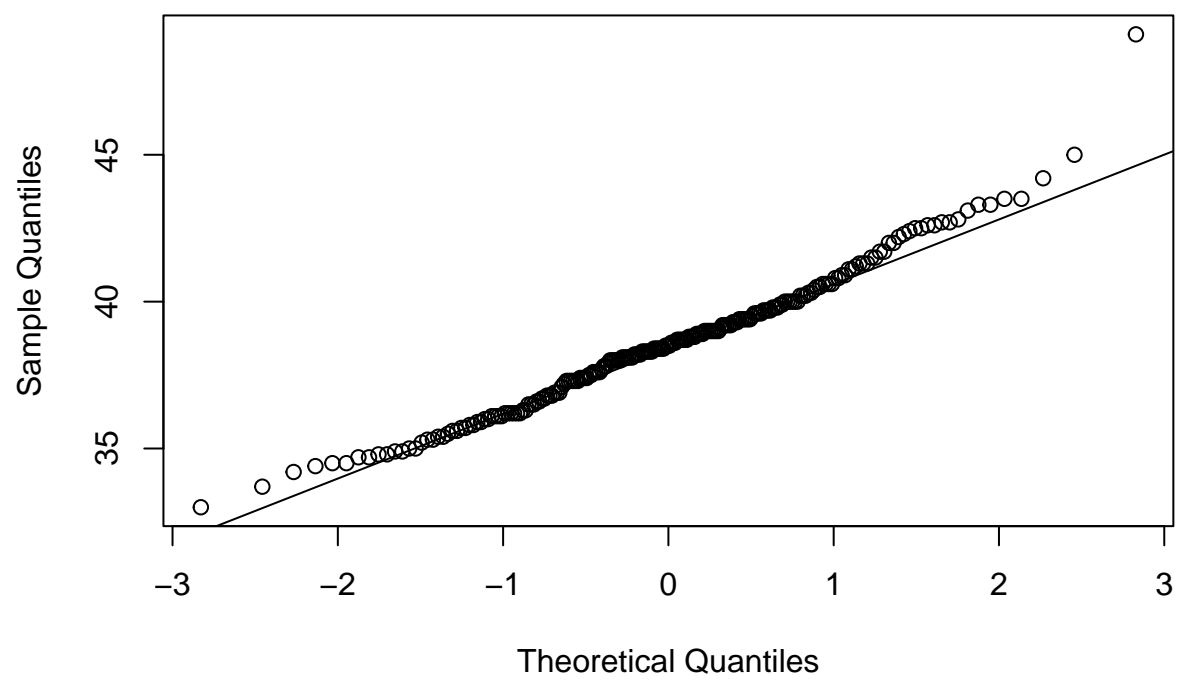
QQ plot de Hip



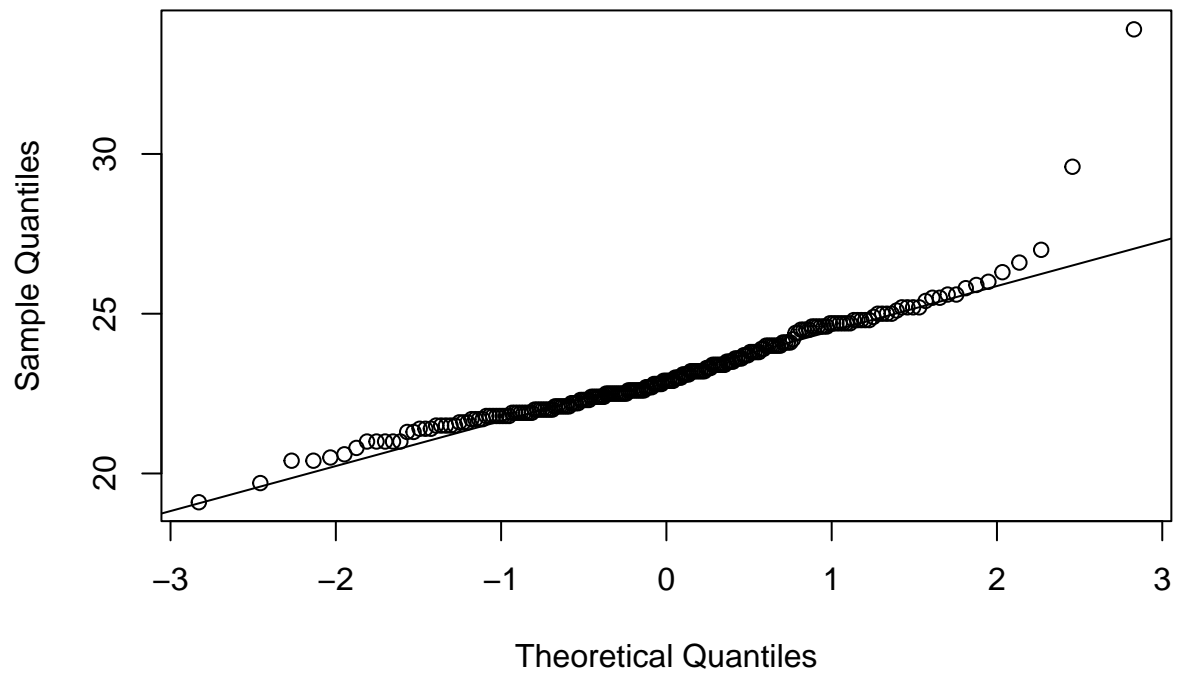
QQ plot de Thigh



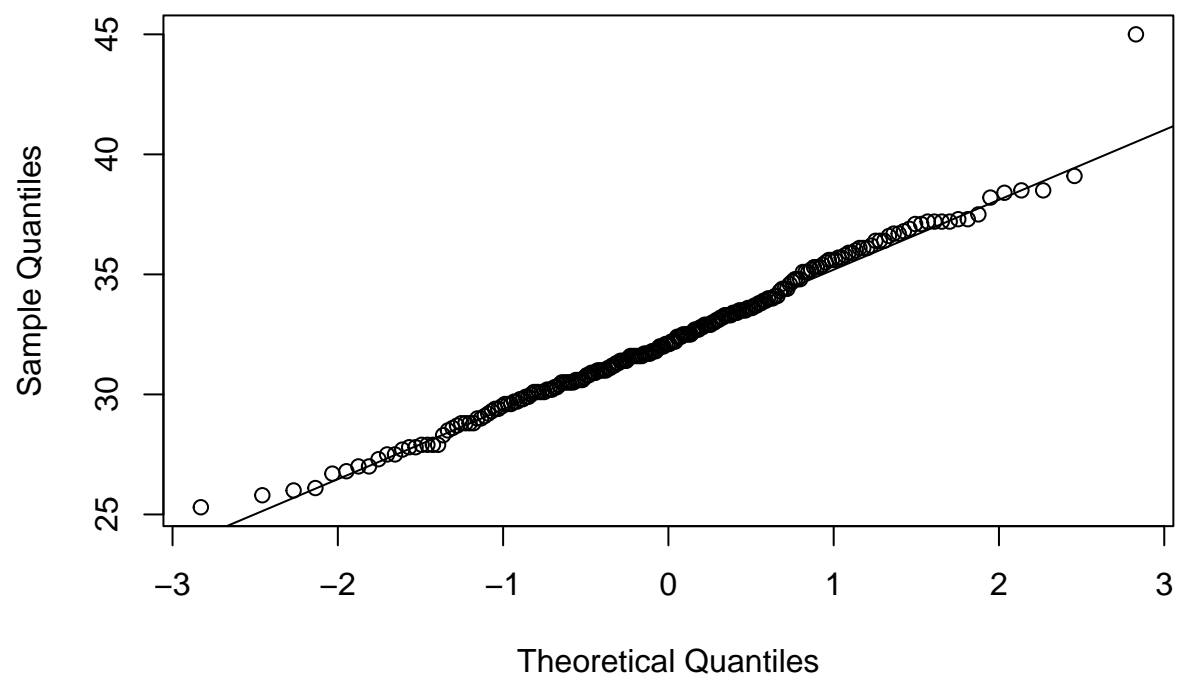
QQ plot de Knee



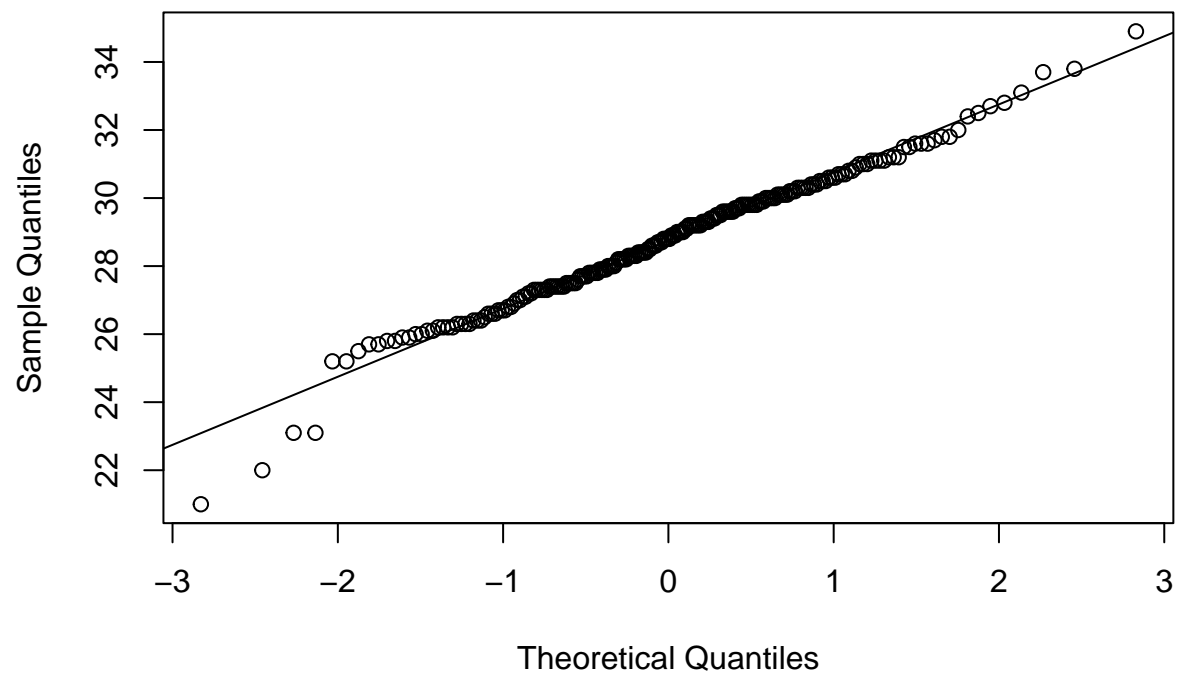
QQ plot de Ankle



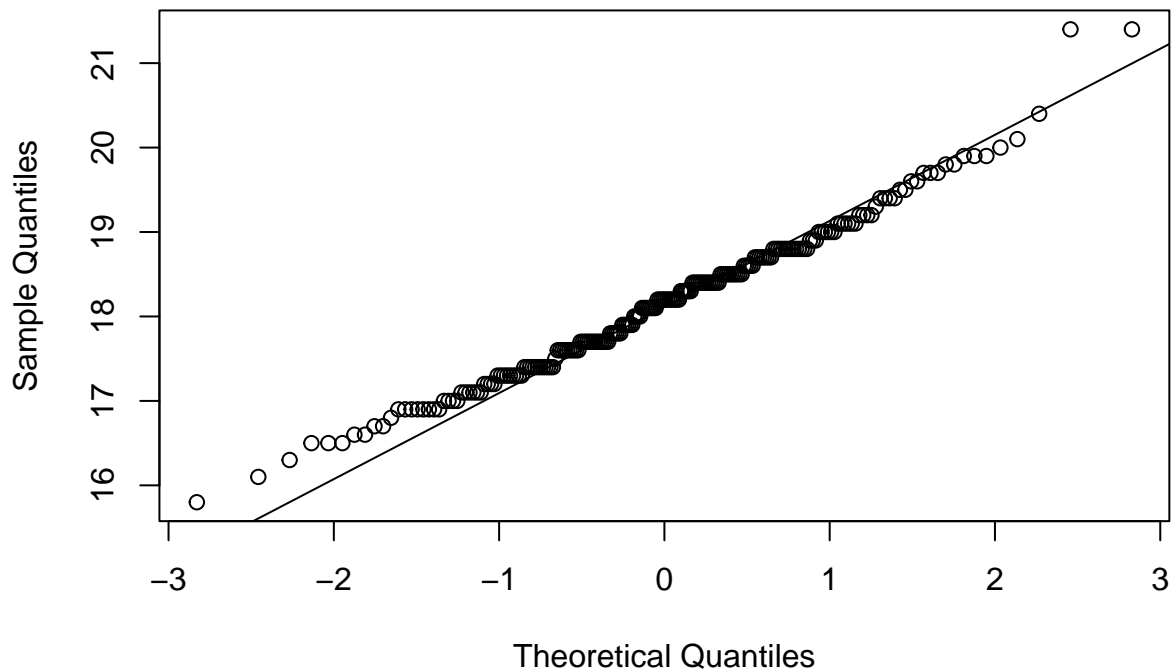
QQ plot de Biceps



QQ plot de Forearm



QQ plot de Wrist



```
#Test de Shapiro-Wilk pour chaque circonférence
for (col in names(bodyfat)[-c(1:5)]) { # Exclure les 5 premières colonnes
normality_tests[[col]] <- shapiro.test(bodyfat[[col]])
cat("Test de normalité de Shapiro-Wilk pour", col, "\n")
print(normality_tests[[col]])
}
```

```
## Test de normalité de Shapiro-Wilk pour Neck :
##
##  Shapiro-Wilk normality test
##
## data:  bodyfat[[col]]
## W = 0.96631, p-value = 5.575e-05
##
## Test de normalité de Shapiro-Wilk pour Chest :
##
##  Shapiro-Wilk normality test
##
## data:  bodyfat[[col]]
## W = 0.96237, p-value = 1.903e-05
##
## Test de normalité de Shapiro-Wilk pour Abdomen :
##
##  Shapiro-Wilk normality test
##
## data:  bodyfat[[col]]
```

```

## W = 0.95009, p-value = 9.146e-07
##
## Test de normalité de Shapiro-Wilk pour Hip :
##
## Shapiro-Wilk normality test
##
## data: bodyfat[[col]]
## W = 0.9099, p-value = 4.256e-10
##
## Test de normalité de Shapiro-Wilk pour Thigh :
##
## Shapiro-Wilk normality test
##
## data: bodyfat[[col]]
## W = 0.96097, p-value = 1.314e-05
##
## Test de normalité de Shapiro-Wilk pour Knee :
##
## Shapiro-Wilk normality test
##
## data: bodyfat[[col]]
## W = 0.97895, p-value = 0.002735
##
## Test de normalité de Shapiro-Wilk pour Ankle :
##
## Shapiro-Wilk normality test
##
## data: bodyfat[[col]]
## W = 0.89526, p-value = 4.506e-11
##
## Test de normalité de Shapiro-Wilk pour Biceps :
##
## Shapiro-Wilk normality test
##
## data: bodyfat[[col]]
## W = 0.98646, p-value = 0.03918
##
## Test de normalité de Shapiro-Wilk pour Forearm :
##
## Shapiro-Wilk normality test
##
## data: bodyfat[[col]]
## W = 0.98373, p-value = 0.01448
##
## Test de normalité de Shapiro-Wilk pour Wrist :
##
## Shapiro-Wilk normality test
##
## data: bodyfat[[col]]
## W = 0.98633, p-value = 0.03736

```

Les données ne représentent pas une courbe normale selon le test de Shapiro-Wilk et les graphiques des quantiles-quantiles démontrent qu'il y a présence de données aberrantes.

Nous allons utiliser les scores z pour éliminer les données qui sont trop à l'écart et ainsi normaliser l'ensemble

des données.

```
#Supprimer les données de la ligne 39 qui représentent des données abberantes.
bodyfat2 <- bodyfat[-39,]
bodyfat[39,]
```

```
## # A tibble: 1 x 15
##   Density BodyFat   Age Weight Height  Neck Chest Abdomen  Hip Thigh  Knee
##   <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1    1.02    35.2   46   363.   72.2  51.2  136.    148.  148.   87.3  49.1
## # ... with 4 more variables: Ankle <dbl>, Biceps <dbl>, Forearm <dbl>,
## #   Wrist <dbl>
```

```
#Score z de la colonne chest
z_scores_chest <- scale(bodyfat2$Chest)
#Définir un seuil pour le score Z
seuil_z_chest <- 2.2
#Remplacer les valeurs aberrantes par NA pour chest
bodyfat2$Chest[abs(z_scores_chest) > seuil_z_chest] <- NA
#Le nombre de données supprimées pour Chest
nb_na_chest <- sum(is.na(bodyfat2$Chest))
cat(nb_na_chest, "Données supprimées")
```

6 Données supprimées

```
#score z pour Abdomen
z_scores_abdomen <- scale(bodyfat2$Abdomen)
#Définir un seuil pour le score Z
seuil_z_abdomen <- 2.39
#Remplacer les valeurs aberrantes par NA pour abdomen
bodyfat2$Abdomen[abs(z_scores_abdomen) > seuil_z_abdomen] <- NA
#Le nombre de données supprimées pour Abdomen
nb_na_abdomen <- sum(is.na(bodyfat2$Abdomen))
cat(nb_na_abdomen, "Données supprimées")
```

4 Données supprimées

```
#score pour Hip
z_scores_hip <- scale(bodyfat2$Hip)
#Définir un seuil pour le score Z
seuil_z_hip <- 2.39
#Remplacer les valeurs aberrantes par NA pour hip
bodyfat2$Hip[abs(z_scores_hip) > seuil_z_hip] <- NA
#Le nombre de données supprimées pour Hip
nb_na_hip <- sum(is.na(bodyfat2$Hip))
cat(nb_na_hip, "Données supprimées")
```

3 Données supprimées

```

#score pour Thigh
z_scores_thigh <- scale(bodyfat2$Thigh)
#Définir un seuil pour le score Z
seuil_z_thigh <- 2.59
#Remplacer les valeurs aberrantes par NA pour thigh
bodyfat2$Thigh[abs(z_scores_thigh) > seuil_z_thigh] <- NA
#Le nombre de données supprimées pour Thigh
nb_na_thigh <- sum(is.na(bodyfat2$Thigh))
cat(nb_na_thigh, "Données supprimées")

```

2 Données supprimées

```

#score pour Forearm
z_scores_forearm <- scale(bodyfat2$Forearm)
#Définir un seuil pour le score Z
seuil_z_forearm <- 3
#Remplacer les valeurs aberrantes par NA pour forearm
bodyfat2$Forearm[abs(z_scores_forearm) > seuil_z_forearm] <- NA
#Le nombre de données supprimées pour Forearm
nb_na_thigh <- sum(is.na(bodyfat2$Thigh))
cat(nb_na_thigh, "Données supprimées")

```

2 Données supprimées

Pour les données de “Ankle”, nous devons faire une transformation logarithmique afin de rendre les données normales.

```

#transformation logarithmique pour ankle
bodyfat2$Ankle_log <- log(bodyfat2$Ankle)

```

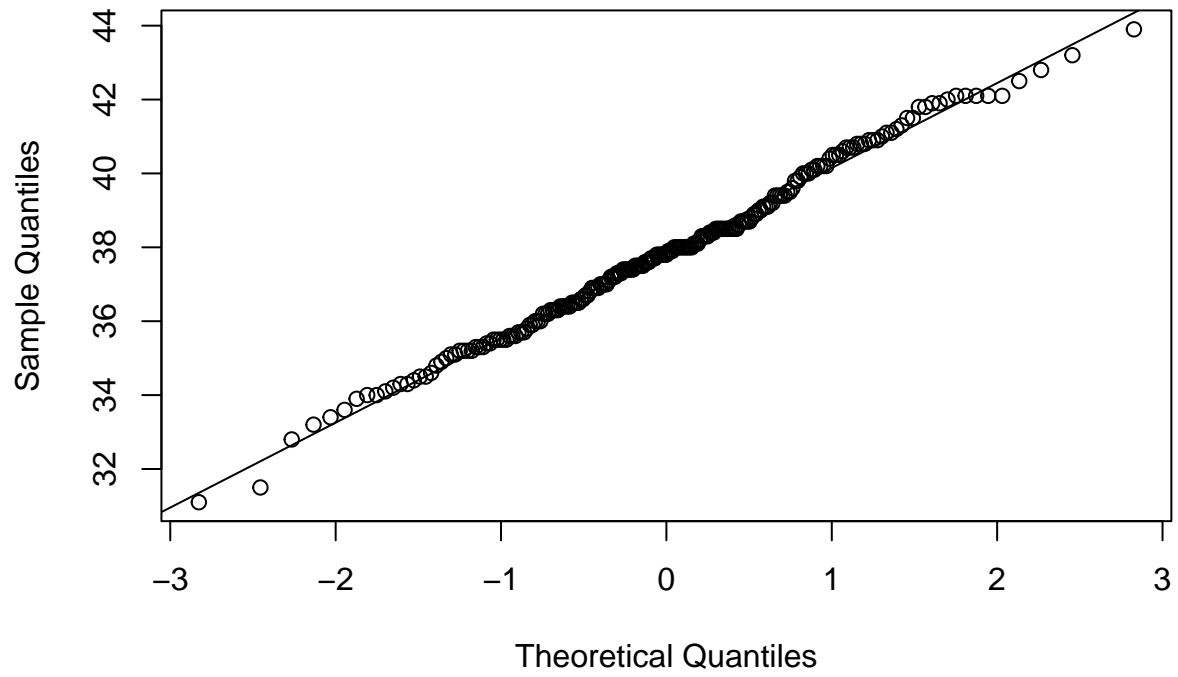
Refaisont le test de Shapiro-Wilk et les graphiques quantiles-quantiles pour observer les résultats après l’élimination de données abberantes.

```

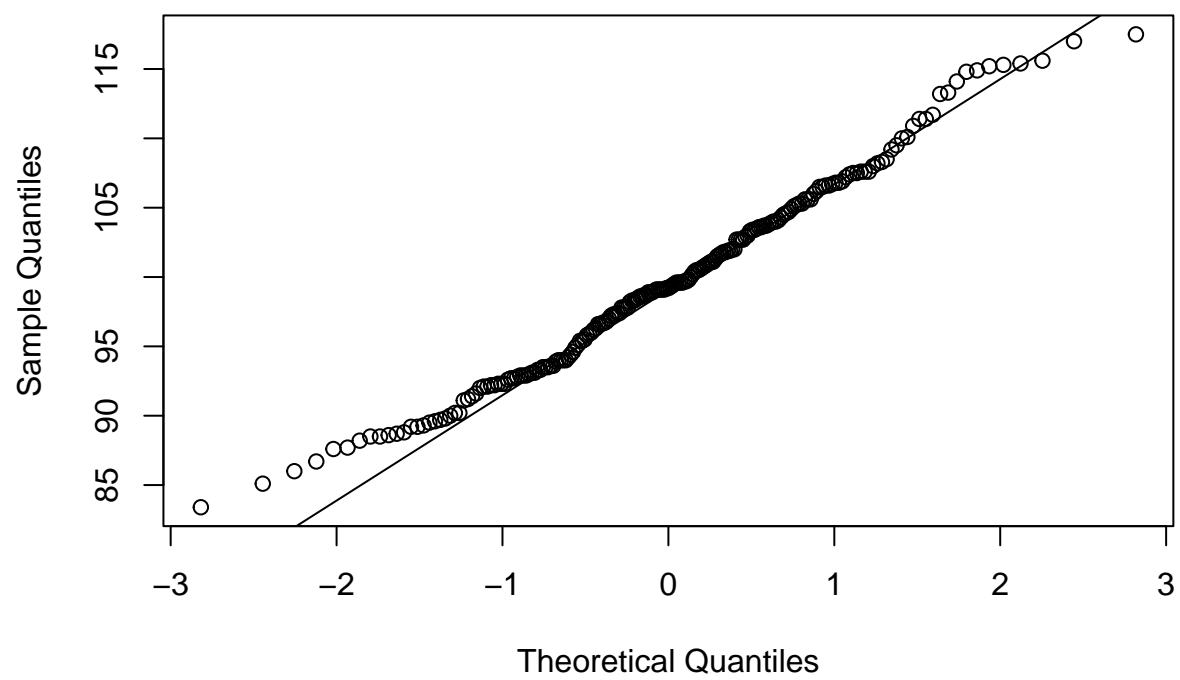
par(mfrow=c(1, 1))
for (col in names(bodyfat2)[-c(1:5)]) {
  qqnorm(bodyfat2[[col]], main = paste("QQ plot de", col))
  qqline(bodyfat2[[col]])
}

```

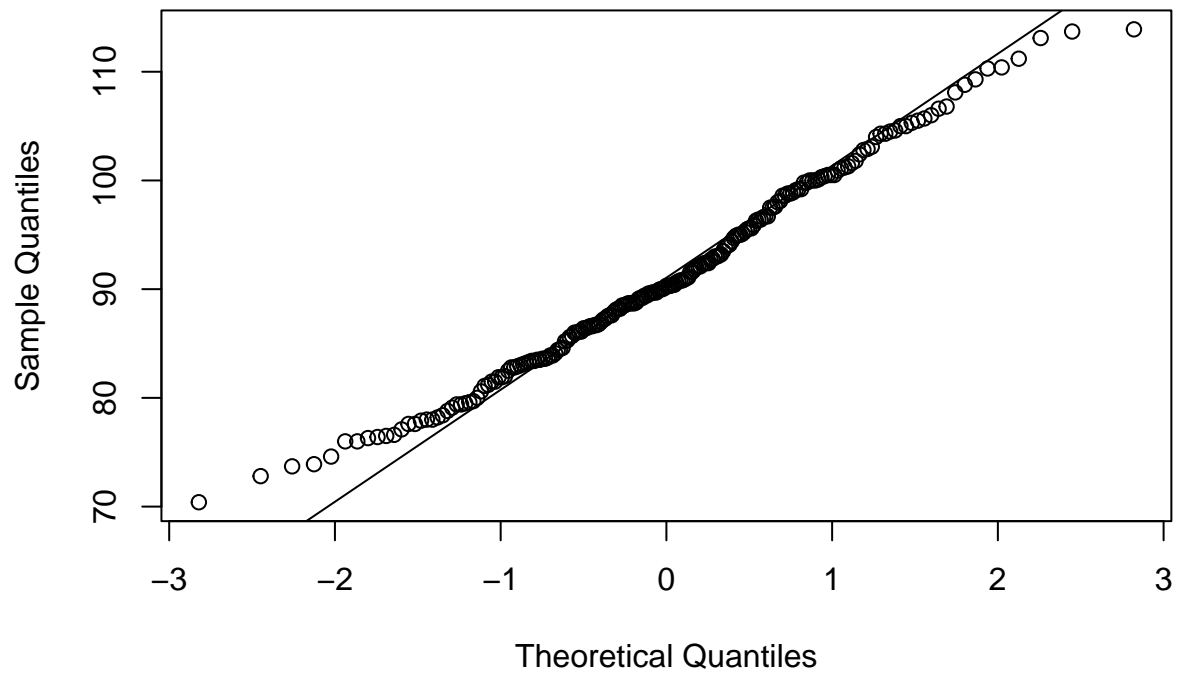

QQ plot de Neck



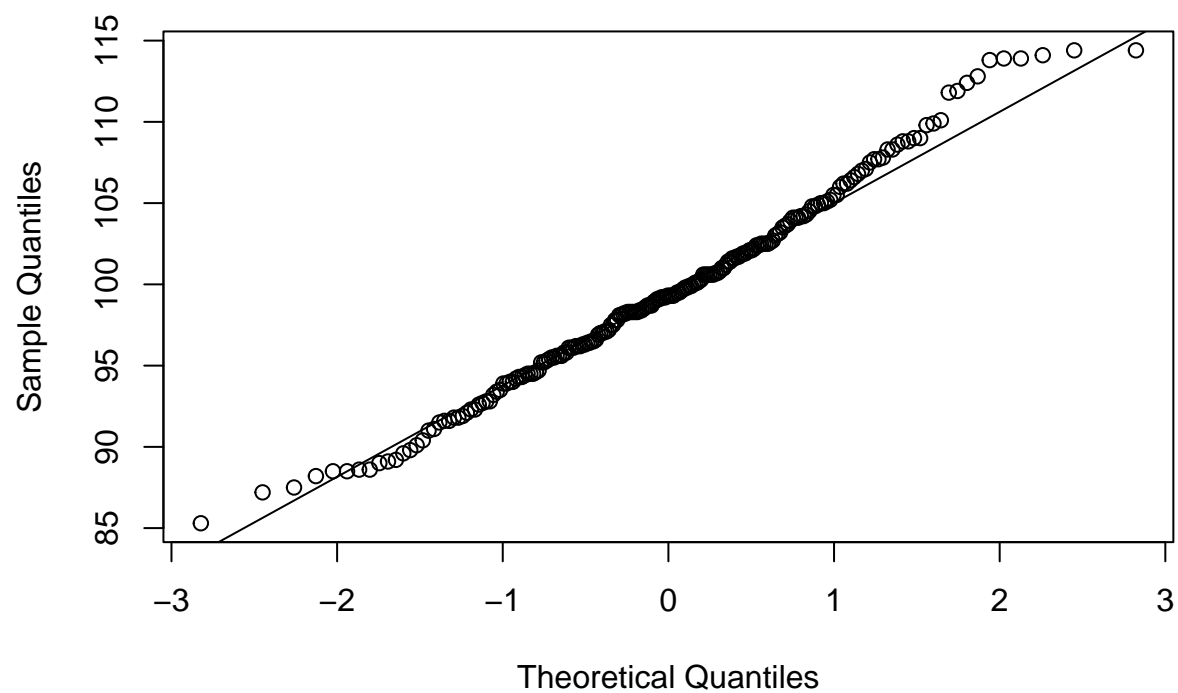
QQ plot de Chest



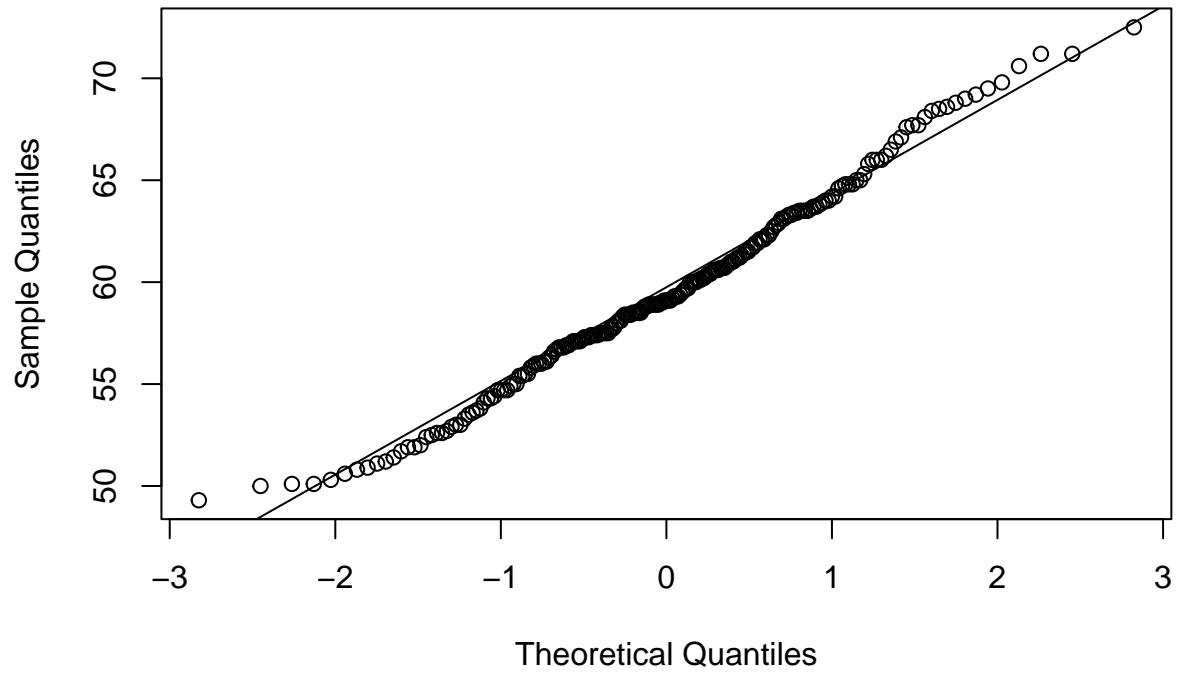
QQ plot de Abdomen



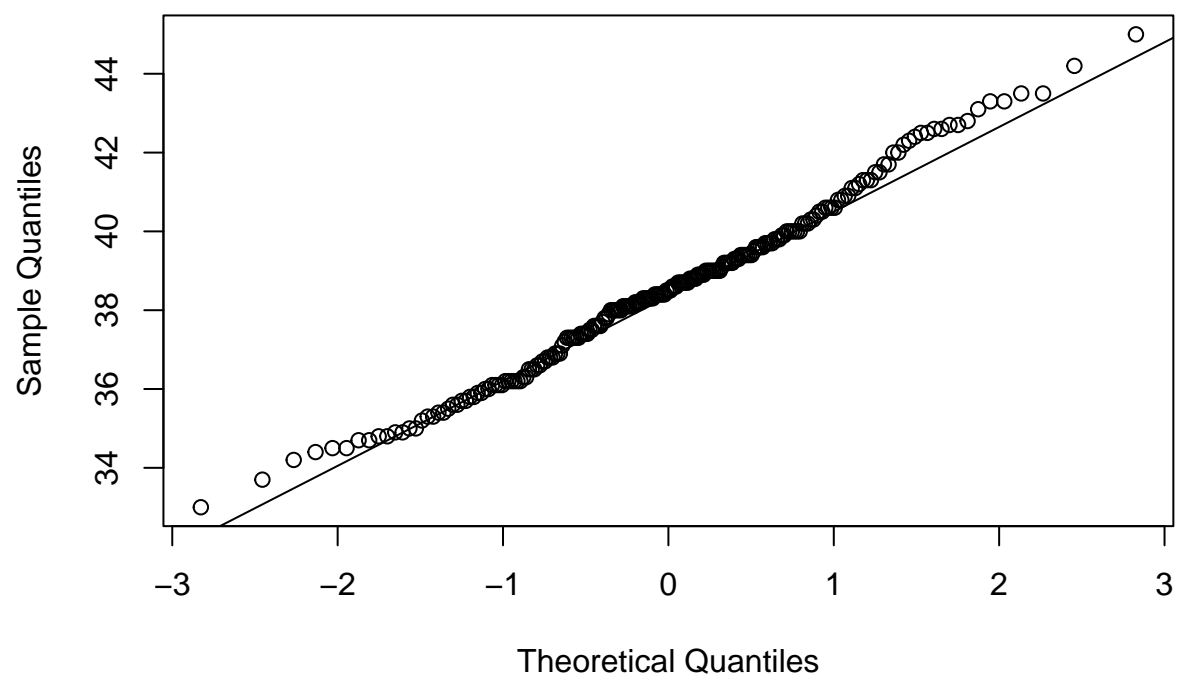
QQ plot de Hip



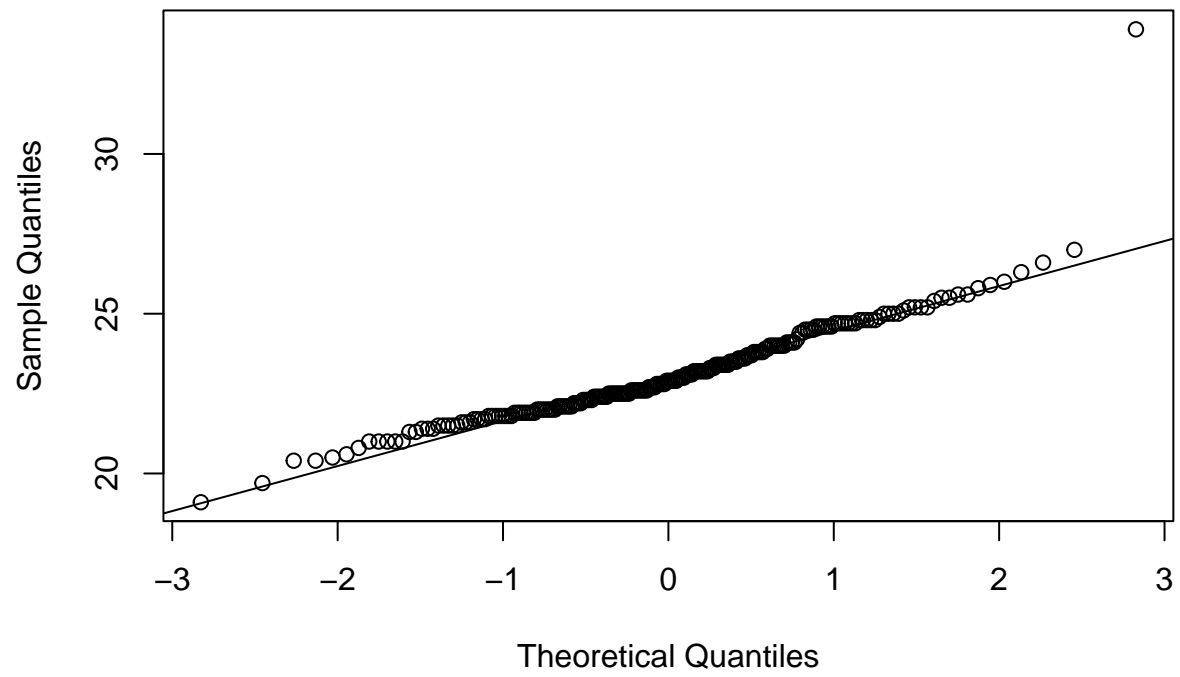
QQ plot de Thigh



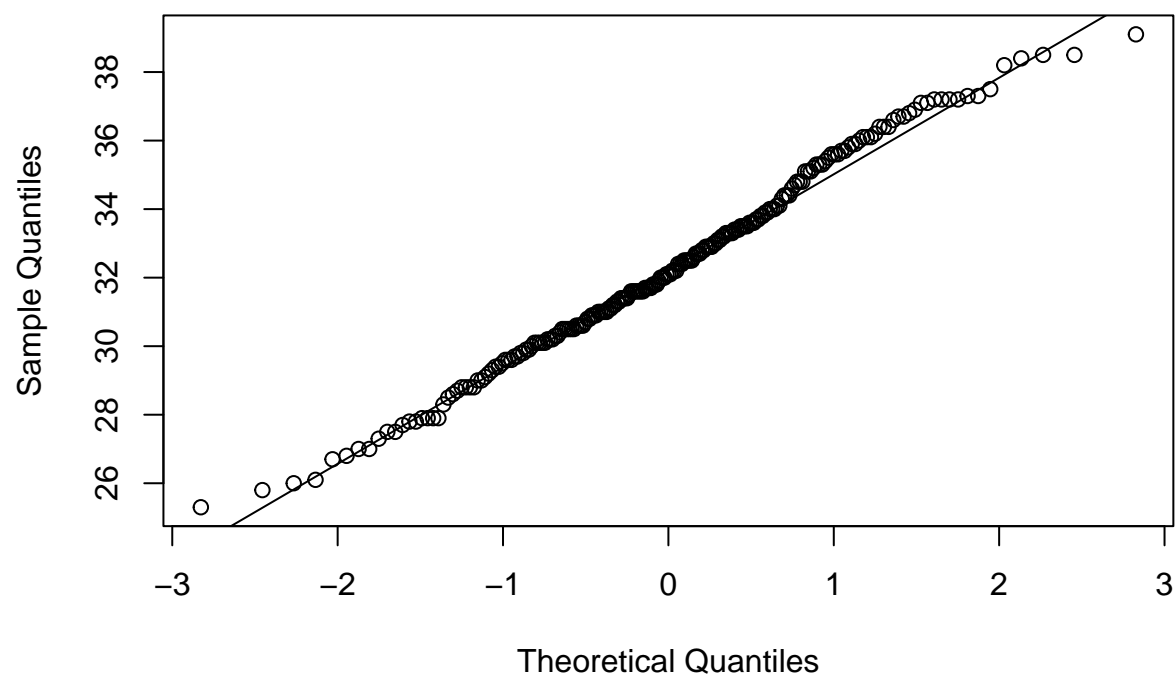
QQ plot de Knee



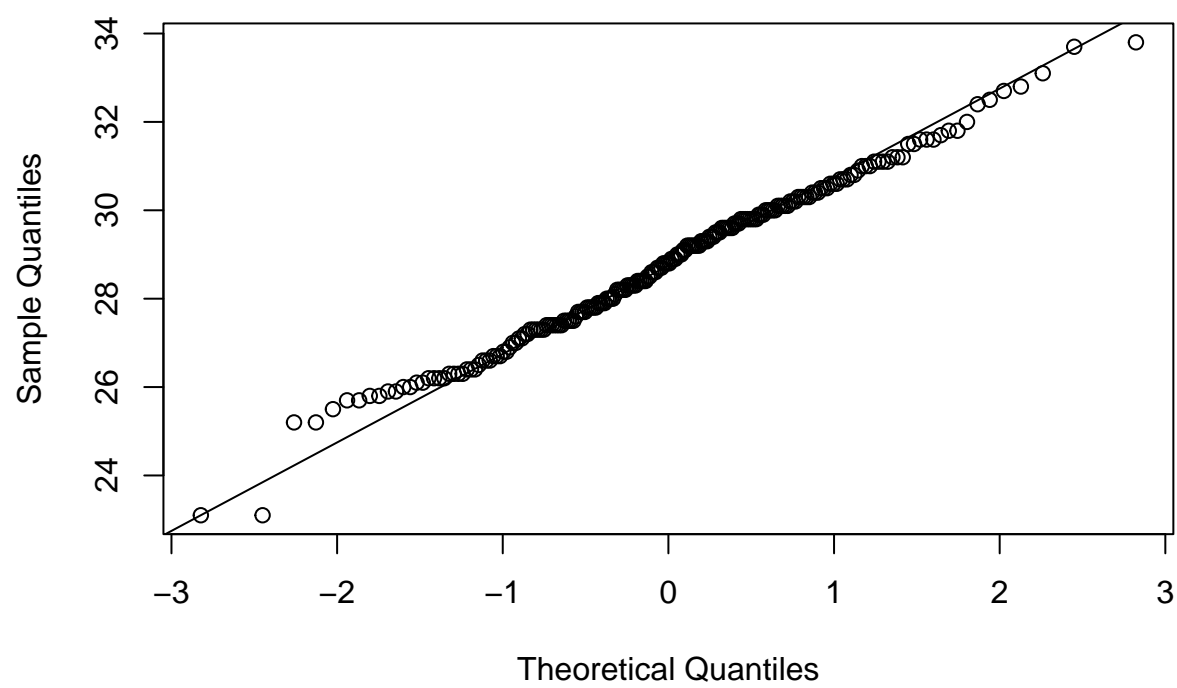
QQ plot de Ankle



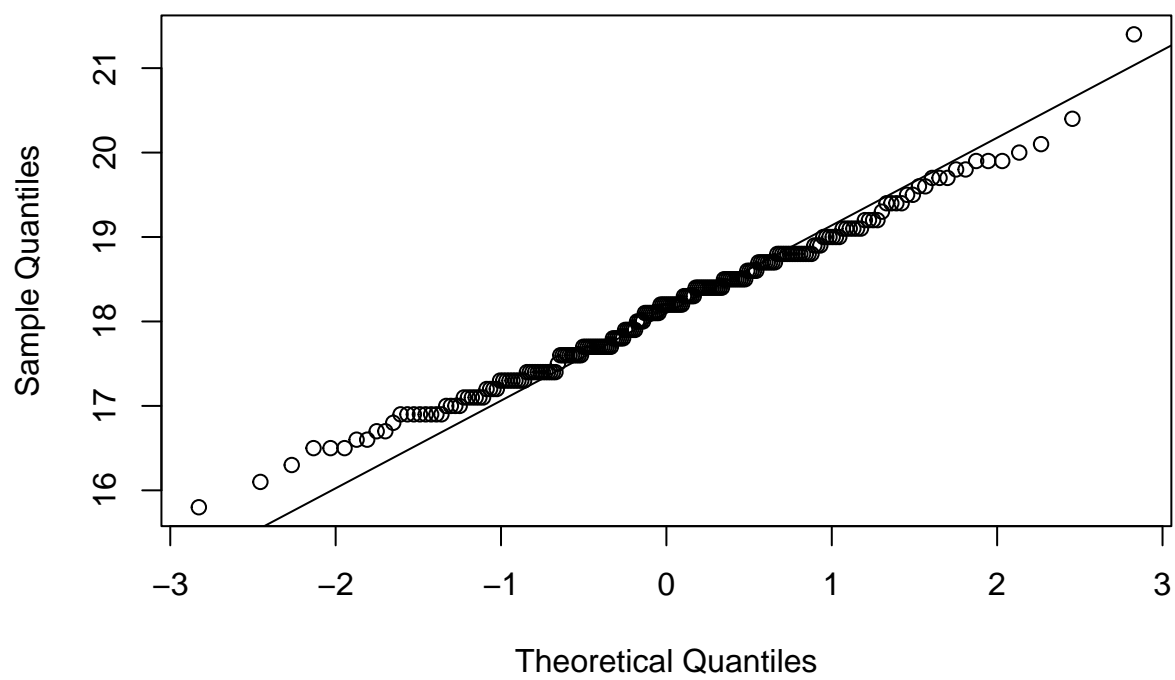
QQ plot de Biceps



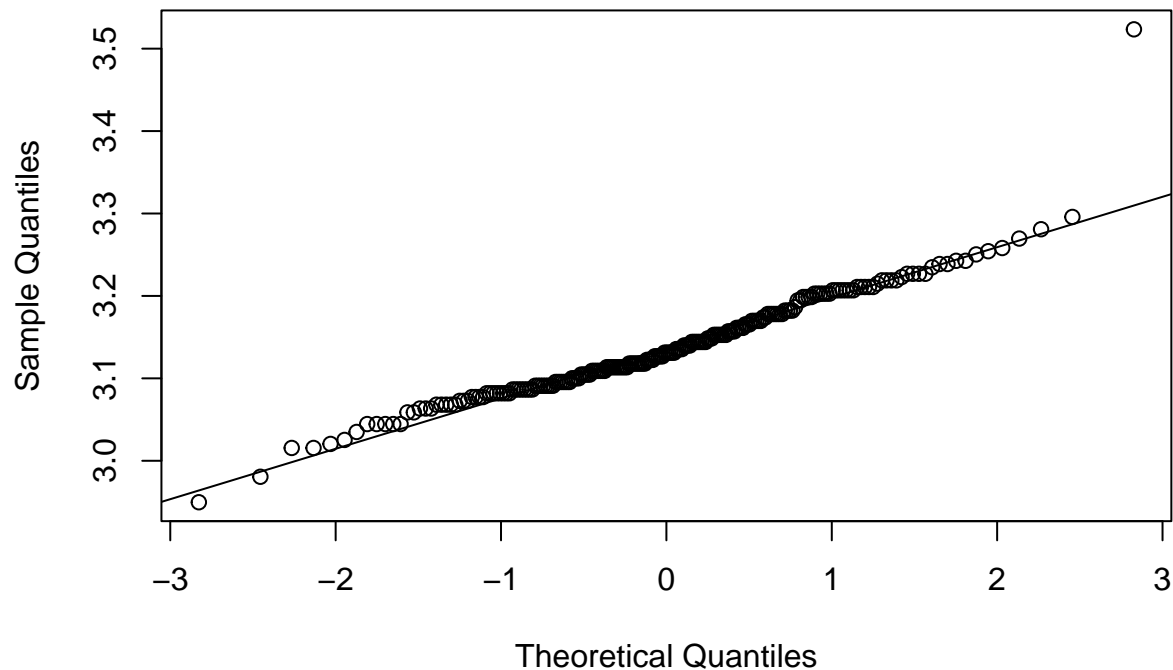
QQ plot de Forearm



QQ plot de Wrist



QQ plot de Ankle_log



```
for (col in names(bodyfat2)[-c(1:5)]) { # Exclure les 5 premières colonnes
  normality_tests[[col]] <- shapiro.test(bodyfat2[[col]])
  cat("Test de normalité de Shapiro-Wilk pour", col, ":\n")
  print(normality_tests[[col]])
}
```

```
## Test de normalité de Shapiro-Wilk pour Neck :
##
##  Shapiro-Wilk normality test
##
## data:  bodyfat2[[col]]
## W = 0.99433, p-value = 0.6019
##
## Test de normalité de Shapiro-Wilk pour Chest :
##
##  Shapiro-Wilk normality test
##
## data:  bodyfat2[[col]]
## W = 0.98756, p-value = 0.06758
##
## Test de normalité de Shapiro-Wilk pour Abdomen :
##
##  Shapiro-Wilk normality test
##
## data:  bodyfat2[[col]]
## W = 0.98871, p-value = 0.09854
```

```

##
## Test de normalité de Shapiro-Wilk pour Hip :
##
## Shapiro-Wilk normality test
##
## data: bodyfat2[[col]]
## W = 0.98873, p-value = 0.09725
##
## Test de normalité de Shapiro-Wilk pour Thigh :
##
## Shapiro-Wilk normality test
##
## data: bodyfat2[[col]]
## W = 0.989, p-value = 0.1054
##
## Test de normalité de Shapiro-Wilk pour Knee :
##
## Shapiro-Wilk normality test
##
## data: bodyfat2[[col]]
## W = 0.9913, p-value = 0.2331
##
## Test de normalité de Shapiro-Wilk pour Ankle :
##
## Shapiro-Wilk normality test
##
## data: bodyfat2[[col]]
## W = 0.90172, p-value = 1.269e-10
##
## Test de normalité de Shapiro-Wilk pour Biceps :
##
## Shapiro-Wilk normality test
##
## data: bodyfat2[[col]]
## W = 0.99166, p-value = 0.2637
##
## Test de normalité de Shapiro-Wilk pour Forearm :
##
## Shapiro-Wilk normality test
##
## data: bodyfat2[[col]]
## W = 0.99266, p-value = 0.3787
##
## Test de normalité de Shapiro-Wilk pour Wrist :
##
## Shapiro-Wilk normality test
##
## data: bodyfat2[[col]]
## W = 0.99208, p-value = 0.3049
##
## Test de normalité de Shapiro-Wilk pour Ankle_log :
##
## Shapiro-Wilk normality test
##

```

```
## data: bodyfat2[[col]]  
## W = 0.94265, p-value = 1.874e-07
```

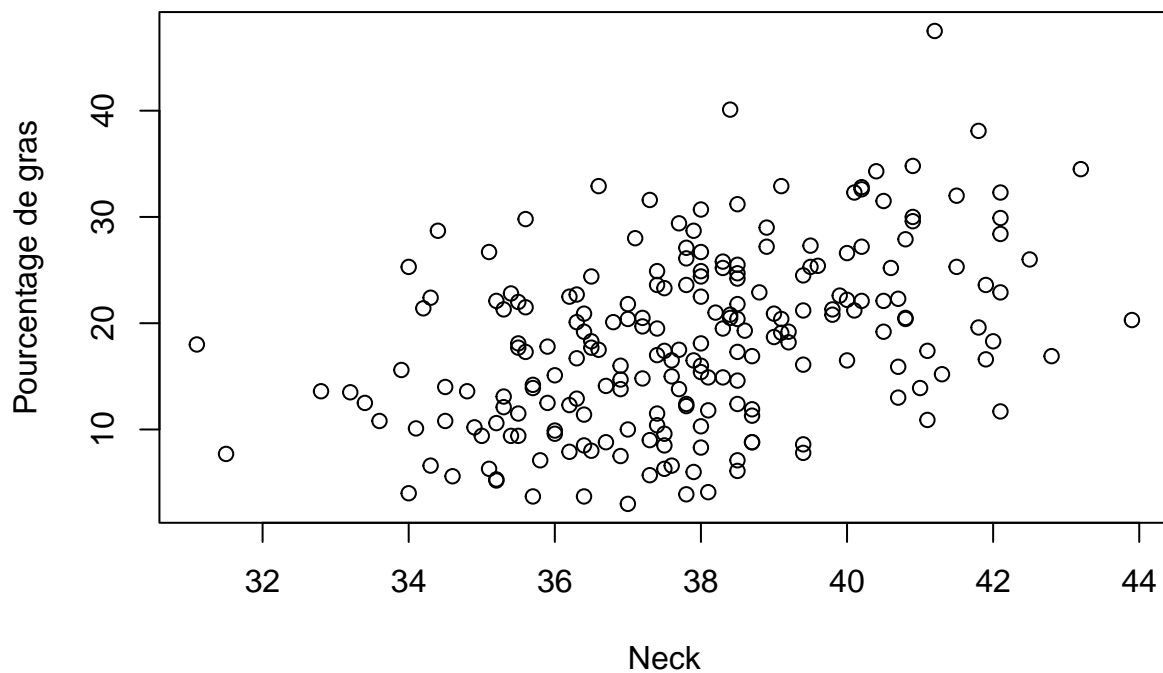
Tous les données sont maintenant normalisées avec un seuil de 0.05% à l'exception de "Ankle".

Explorations des données

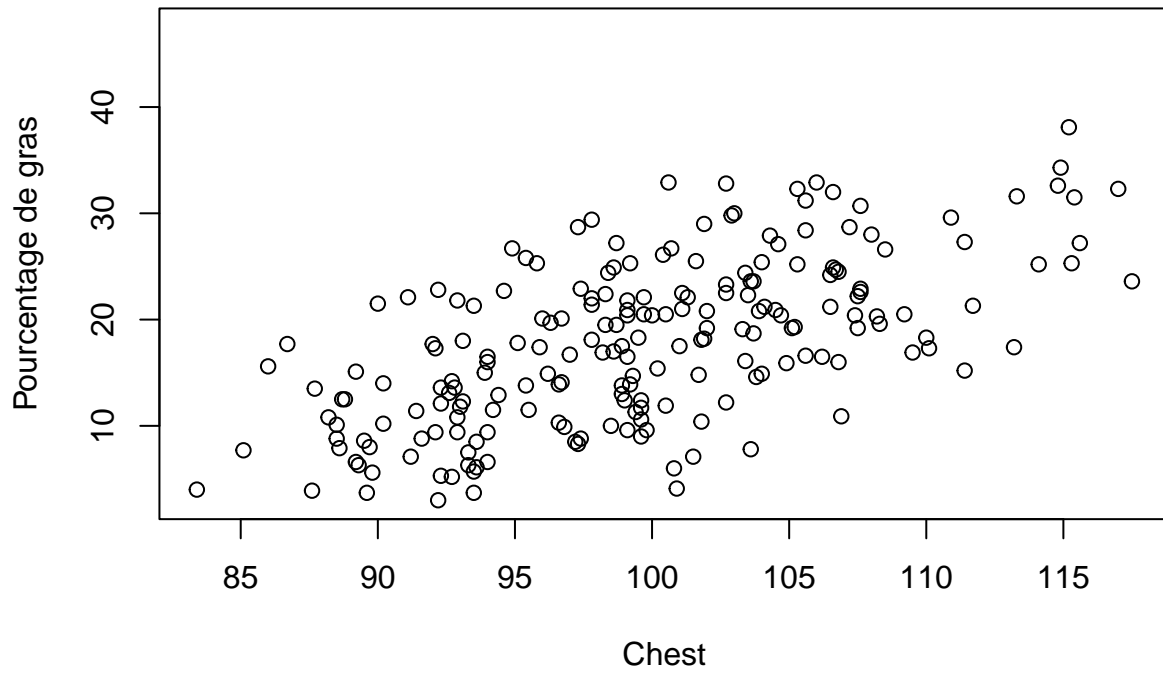
Visualisation des relations entre les divers mensurations et le pourcentage (%) de gras.
L'objectif ici est voir si nous pouvons observer une relation linéaire entre deux variables afin de poursuivre l'analyse vers un test de corrélation, puis bâtir un modèle prédictif avec la régression.

```
par(mfrow=c(1, 1))  
  
col_percentage_gras <- "BodyFat"  
  
for (col in names(bodyfat2)[-c(1:5)]) {  
  if (col != col_percentage_gras) {  
    plot(bodyfat2[[col]], bodyfat2[[col_percentage_gras]],  
        main = paste("Relation entre", col, "et le % de gras"),  
        xlab = col, ylab = "Pourcentage de gras")  
  }  
}
```

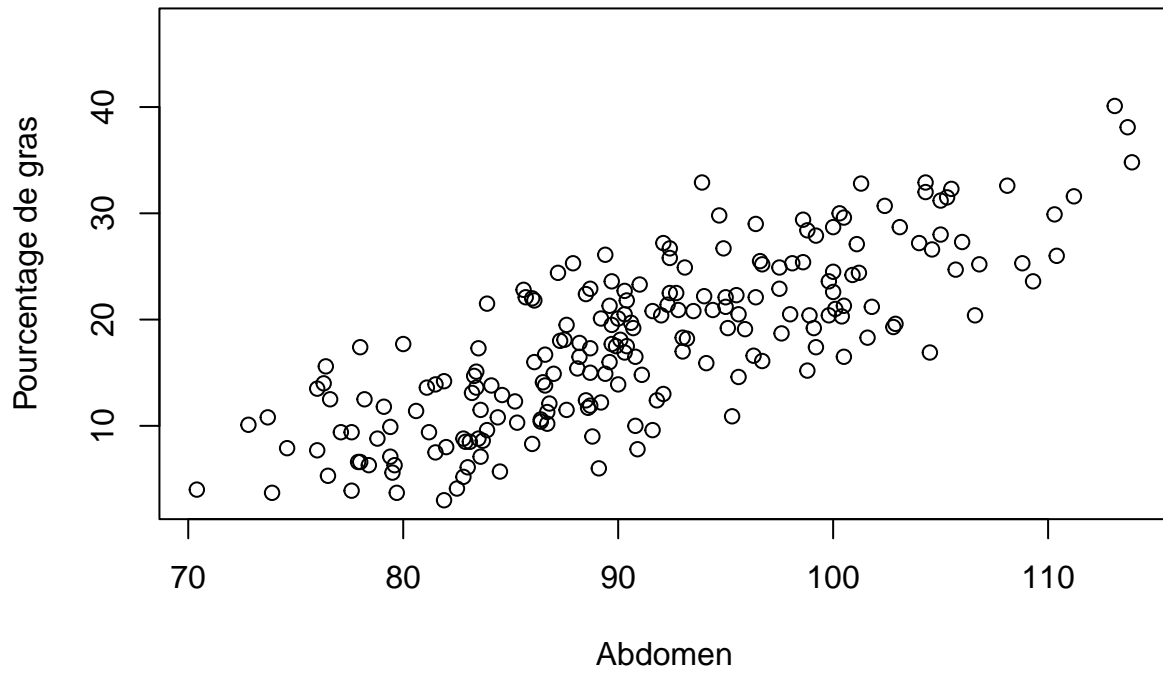
Relation entre Neck et le % de gras



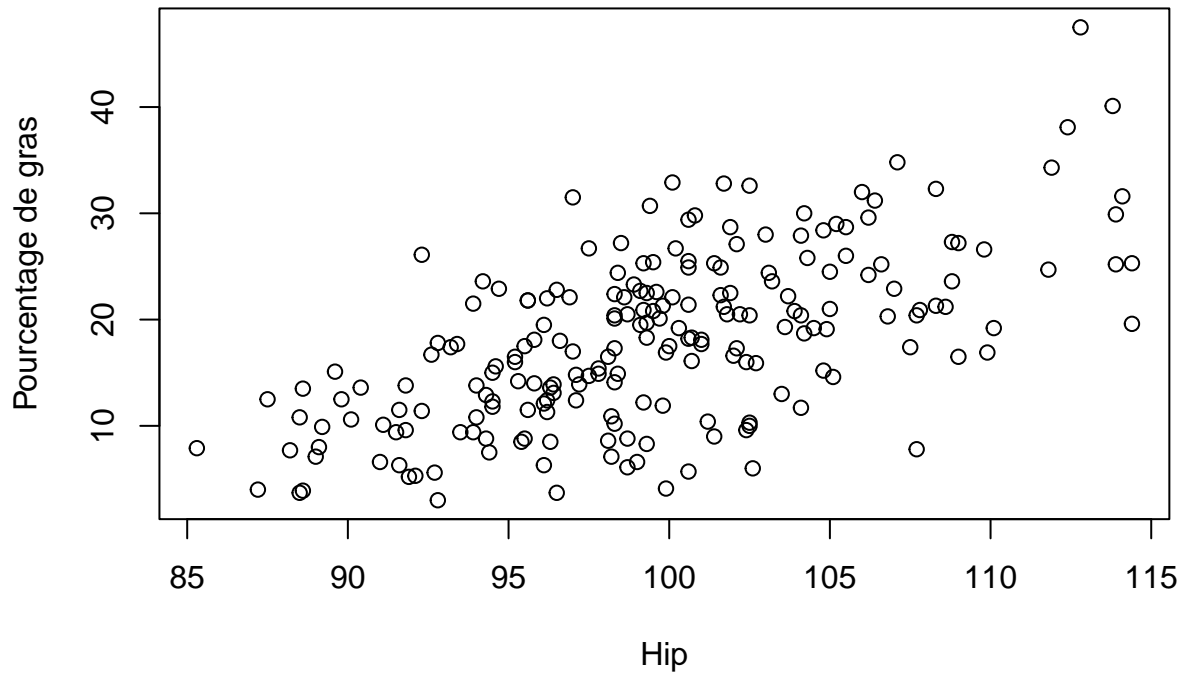
Relation entre Chest et le % de gras



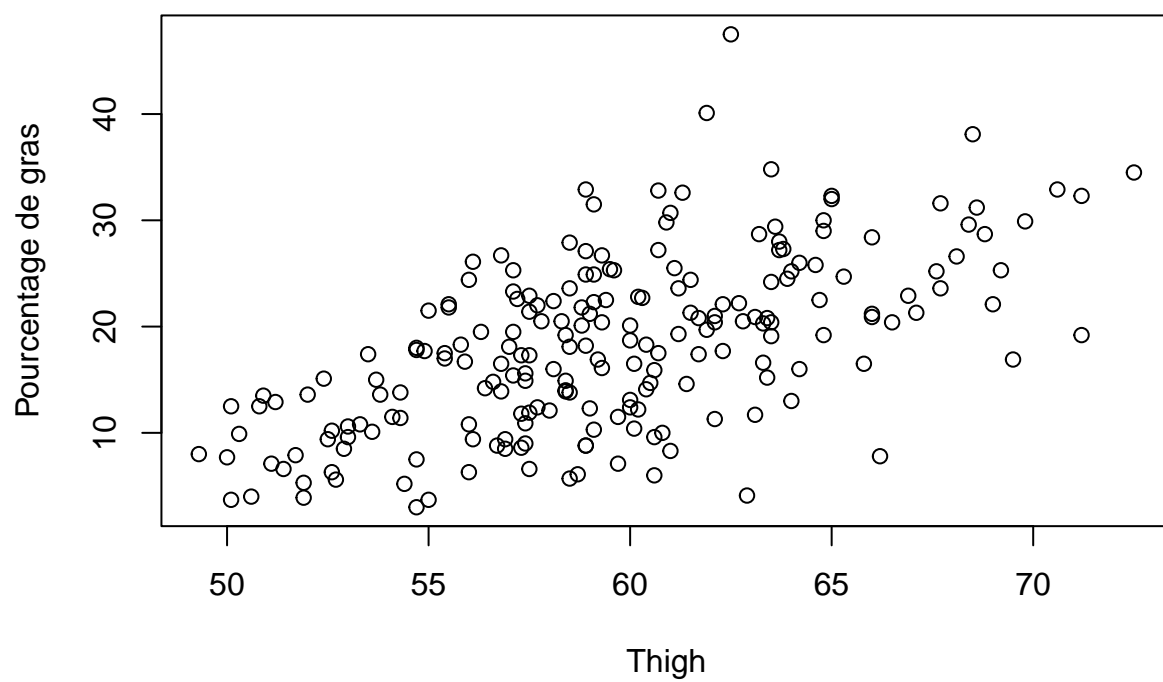
Relation entre Abdomen et le % de gras



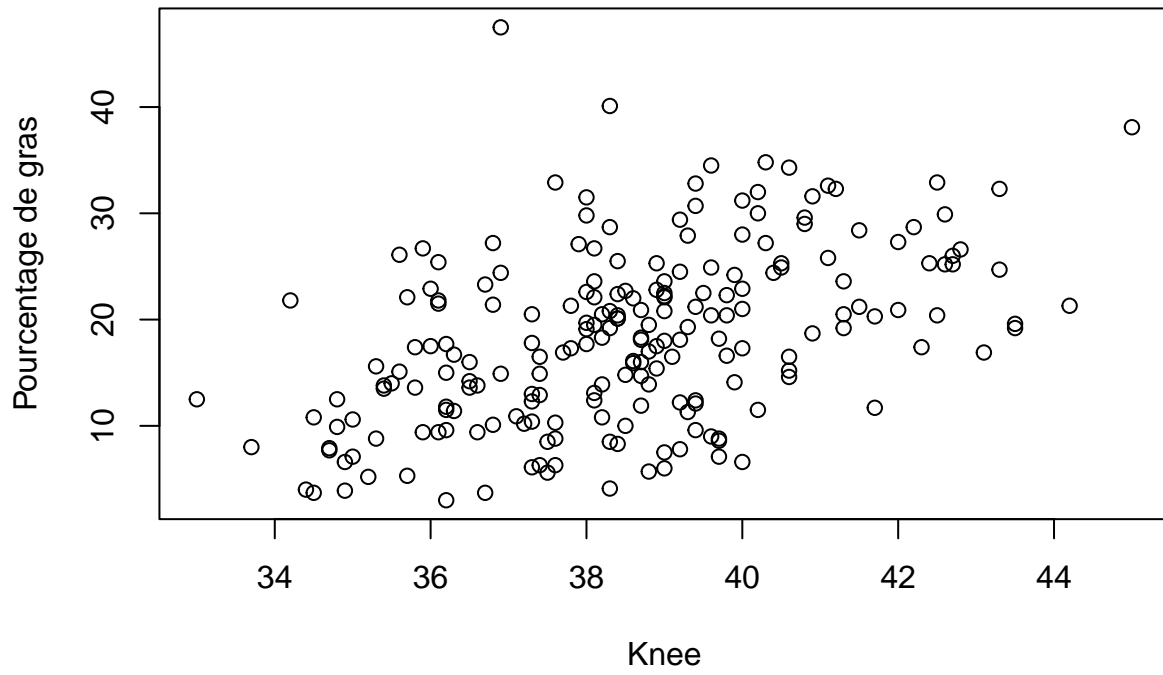
Relation entre Hip et le % de gras



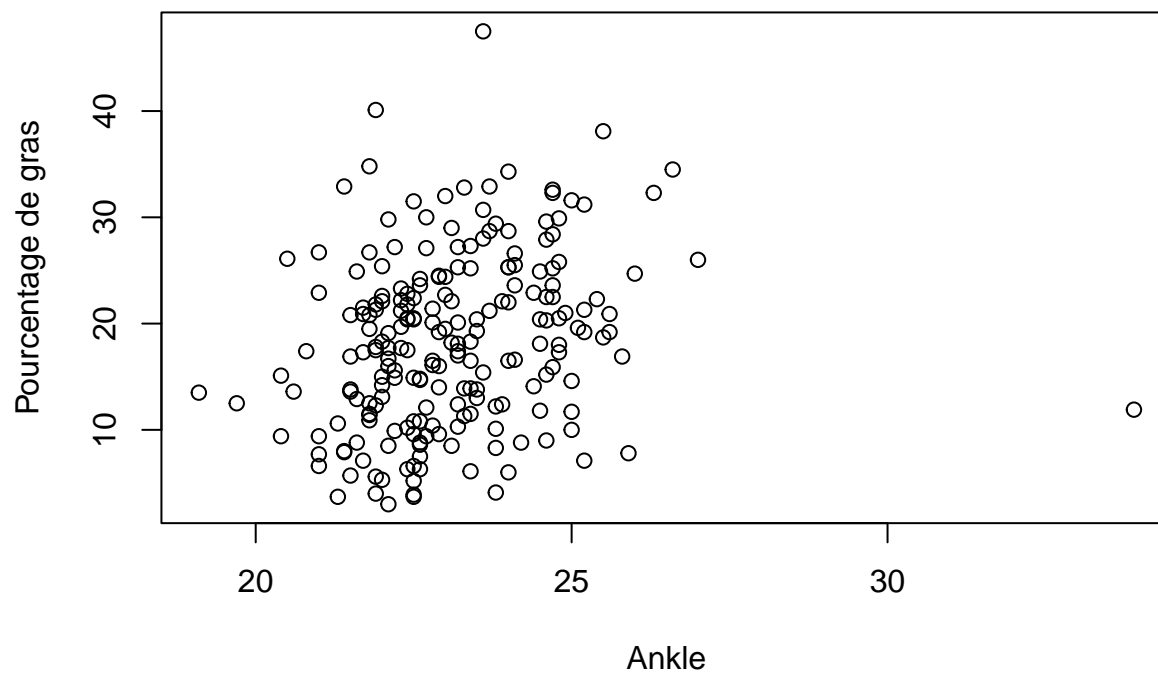
Relation entre Thigh et le % de gras



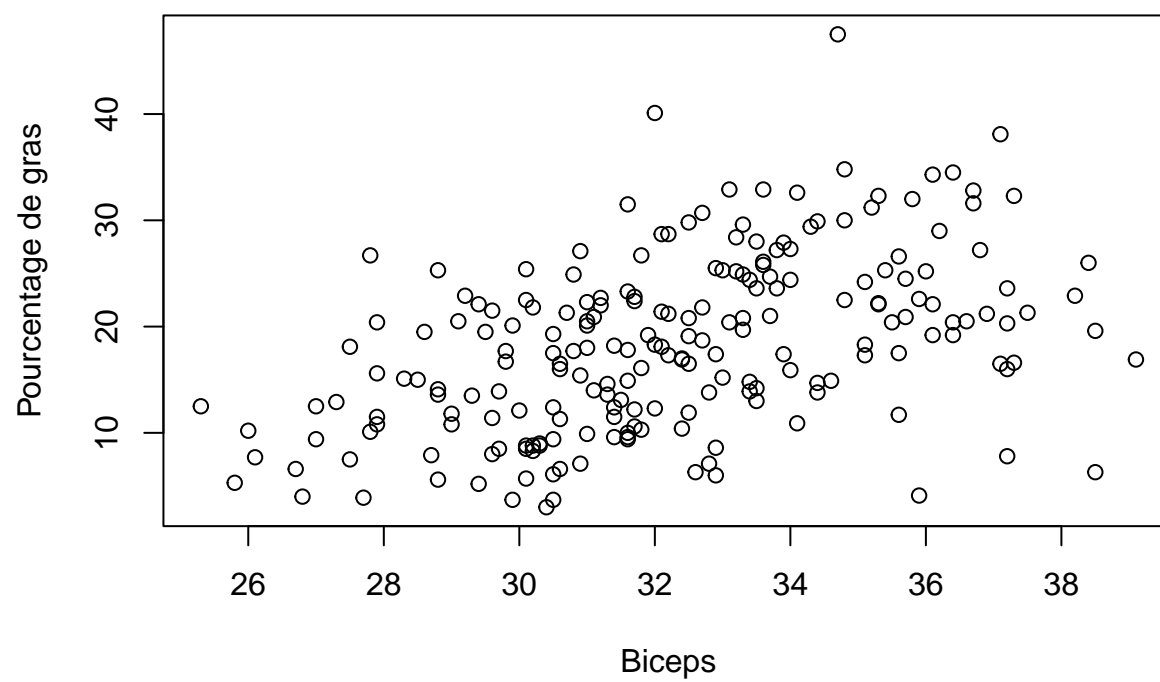
Relation entre Knee et le % de gras



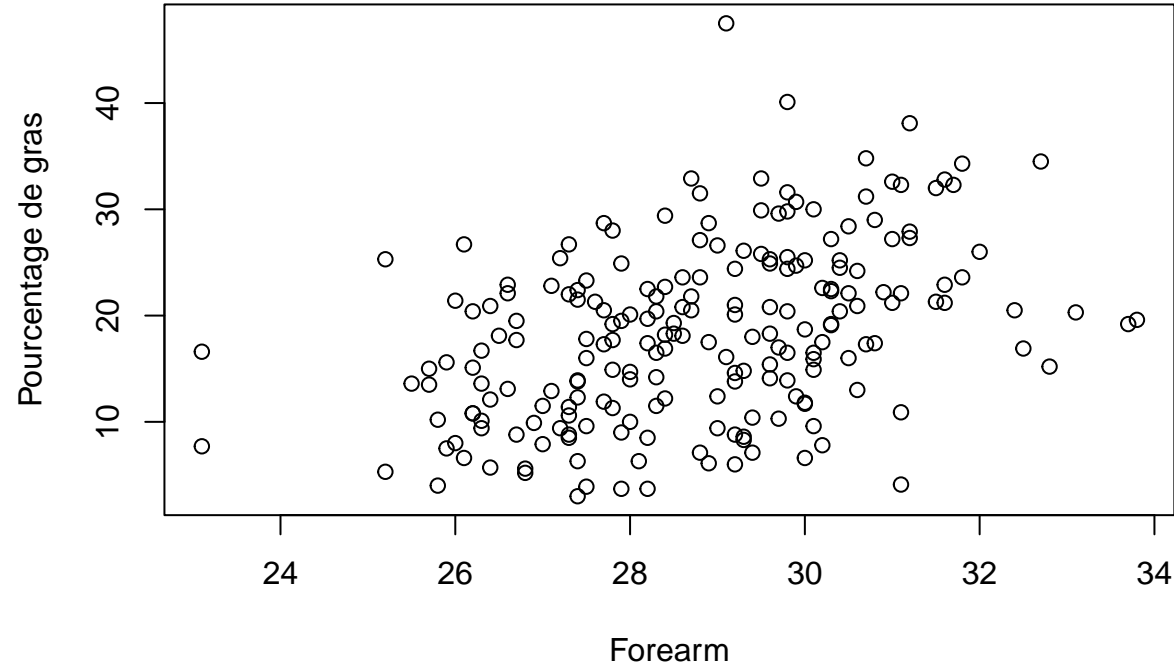
Relation entre Ankle et le % de gras



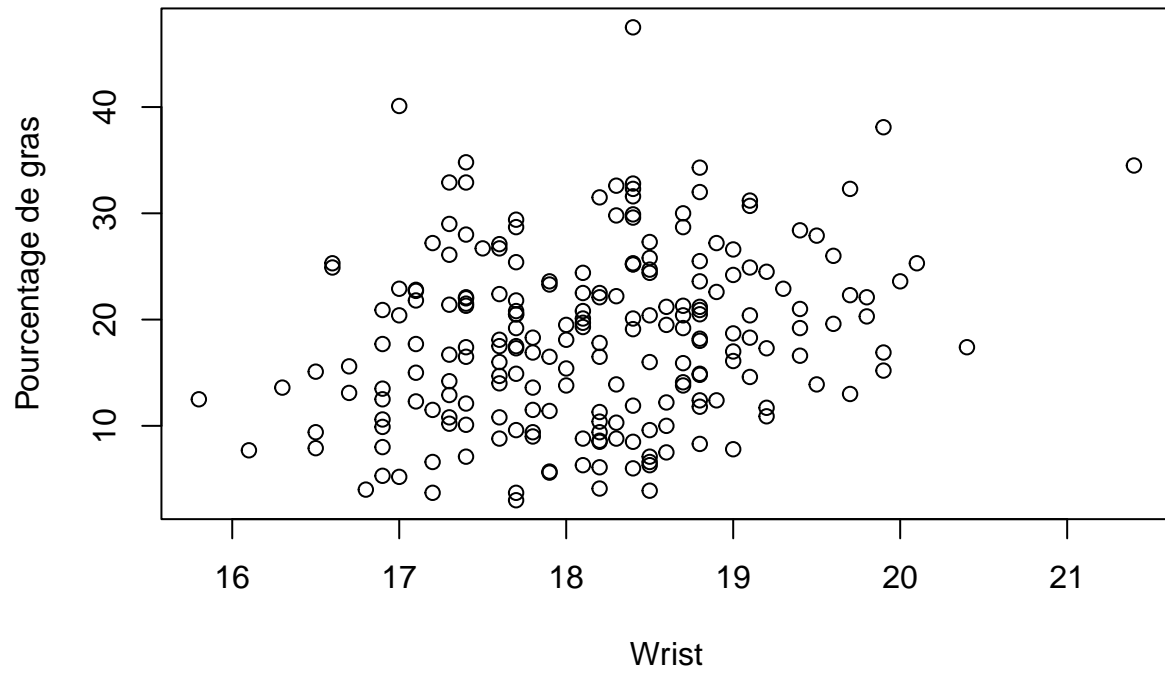
Relation entre Biceps et le % de gras



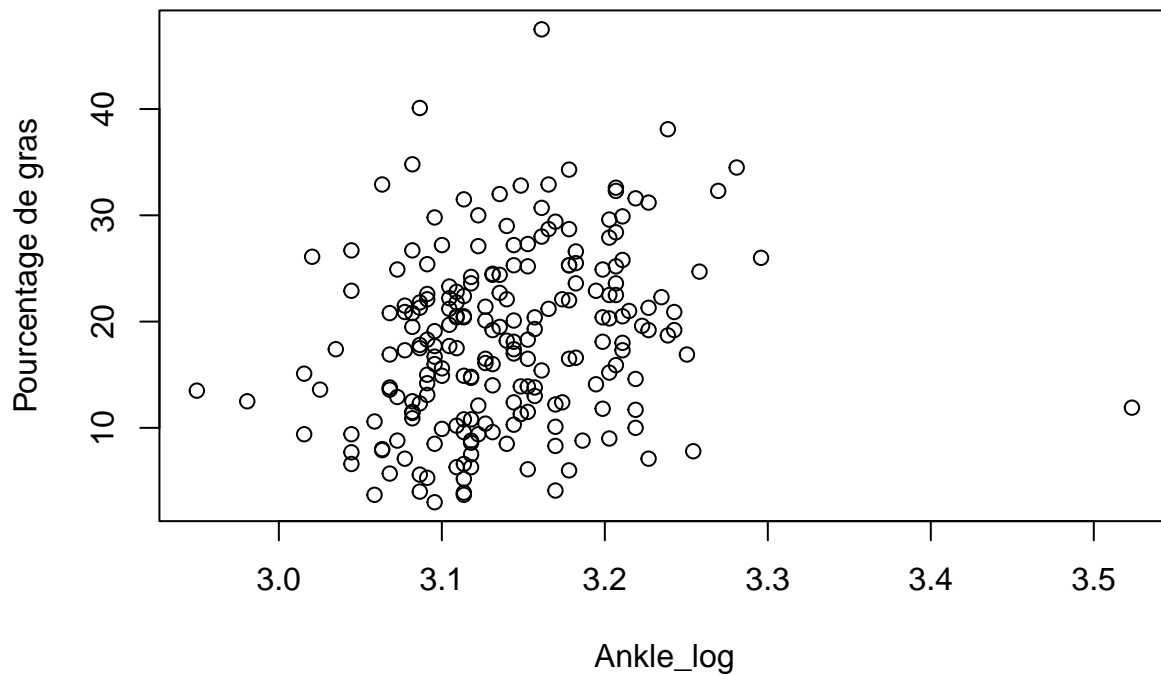
Relation entre Forearm et le % de gras



Relation entre Wrist et le % de gras



Relation entre Ankle_log et le % de gras



Analyse

Pour la majorité des données à l'exception de "Ankle", nous observons une relation linéaire positive. Test de corrélation pour chaque circonférence:

```
#correlation neck et % gras
bodyfat_Neck <- bodyfat2[complete.cases(bodyfat2$BodyFat, bodyfat2$Neck), c("BodyFat", "Neck")]
cor_neck <- cor(bodyfat2$Neck, bodyfat2$BodyFat)
cor_neck
```

```
## [1] 0.4288981
```

```
#correlation Chest et % de gras
bodyfat_Chest <- bodyfat2[complete.cases(bodyfat2$BodyFat, bodyfat2$Chest), c("BodyFat", "Chest")]
cor_chest <- cor(bodyfat_Chest$Chest, bodyfat_Chest$BodyFat)
cor_chest
```

```
## [1] 0.6271569
```

```
#Correlation Abdomen et % de gras
bodyfat_Abdomen <- bodyfat2[complete.cases(bodyfat2$BodyFat, bodyfat2$Abdomen), c("BodyFat", "Abdomen")]
cor_abdomen <- cor(bodyfat_Abdomen$Abdomen, bodyfat_Abdomen$BodyFat)
cor_abdomen
```

```
## [1] 0.7928256
```

```
#Correlation hip et % de gras
```

```
bodyfat_Hip <- bodyfat2[complete.cases(bodyfat2$BodyFat, bodyfat2$Hip), c("BodyFat", "Hip")]  
cor_hip <- cor(bodyfat_Hip$Hip, bodyfat_Hip$BodyFat)  
cor_hip
```

```
## [1] 0.6213323
```

```
#Correlation thigh et % de gras
```

```
bodyfat_Thigh <- bodyfat2[complete.cases(bodyfat2$BodyFat, bodyfat2$Thigh), c("BodyFat", "Thigh")]  
cor_thigh <- cor(bodyfat_Thigh$Thigh, bodyfat_Thigh$BodyFat)  
cor_thigh
```

```
## [1] 0.6002684
```

```
#Correlation knee et % de gras
```

```
bodyfat_Knee <- bodyfat2[complete.cases(bodyfat2$BodyFat, bodyfat2$Knee), c("BodyFat", "Knee")]  
cor_knee <- cor(bodyfat_Knee$Knee, bodyfat_Knee$BodyFat)  
cor_knee
```

```
## [1] 0.4697214
```

```
#correlation Ankle et % de gras
```

```
bodyfat_Ankle <- bodyfat2[complete.cases(bodyfat2$BodyFat, bodyfat2$Ankle), c("BodyFat", "Ankle")]  
cor_ankle <- cor(bodyfat_Ankle$Ankle, bodyfat_Ankle$BodyFat)  
cor_ankle
```

```
## [1] 0.2293663
```

```
#Correlation biceps et % de gras
```

```
bodyfat_Biceps <- bodyfat2[complete.cases(bodyfat2$BodyFat, bodyfat2$Biceps), c("BodyFat", "Biceps")]  
cor_biceps <- cor(bodyfat_Biceps$Biceps, bodyfat_Biceps$BodyFat)  
cor_biceps
```

```
## [1] 0.465773
```

```
#Correlation Forearm et % de gras
```

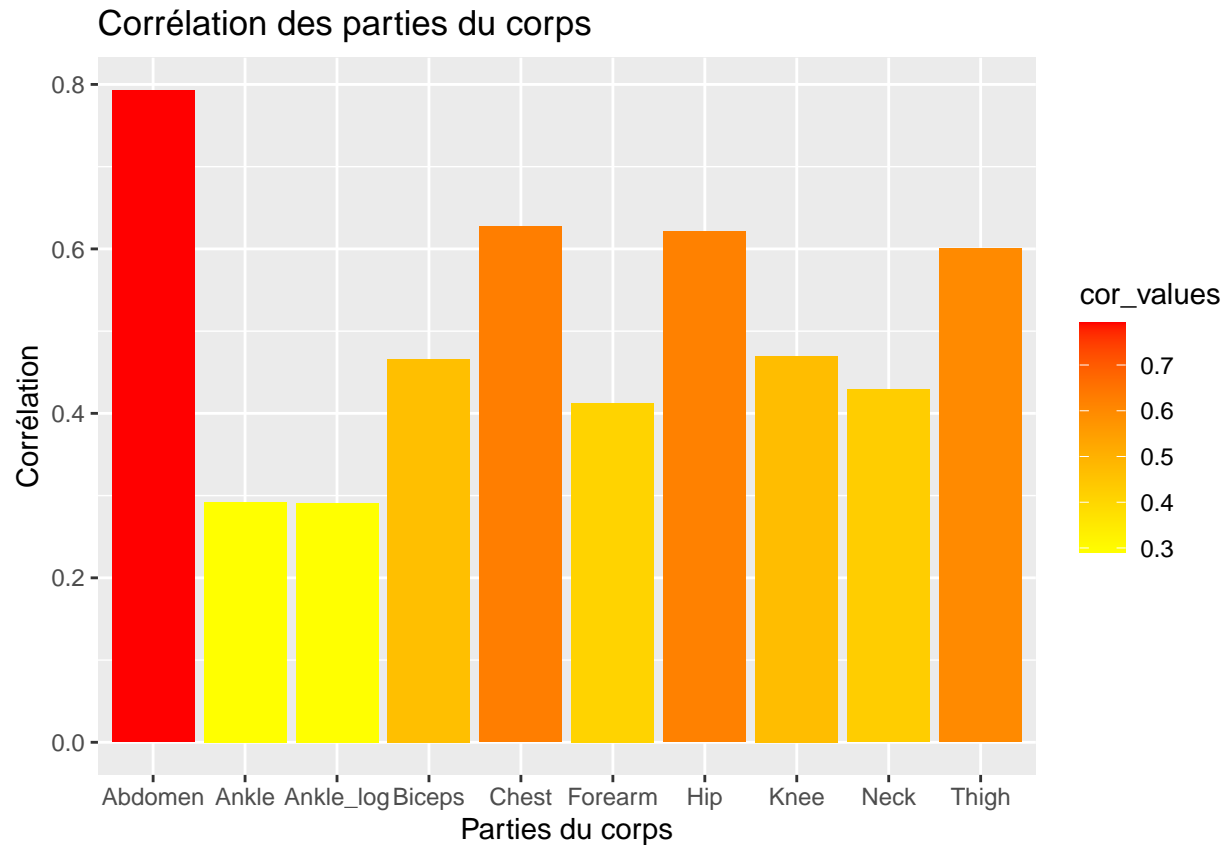
```
bodyfat_Forearm <- bodyfat2[complete.cases(bodyfat2$BodyFat, bodyfat2$Forearm), c("BodyFat", "Forearm")]  
cor_forearm <- cor(bodyfat_Forearm$Forearm, bodyfat_Forearm$BodyFat)  
cor_forearm
```

```
## [1] 0.4125673
```

```
#Correlation wrist et % de gras
```

```
bodyfat_Wrist <- bodyfat2[complete.cases(bodyfat2$BodyFat, bodyfat2$Wrist), c("BodyFat", "Wrist")]  
cor_wrist <- cor(bodyfat_Wrist$Wrist, bodyfat_Wrist$BodyFat)  
cor_wrist
```

```
## [1] 0.2359589
```

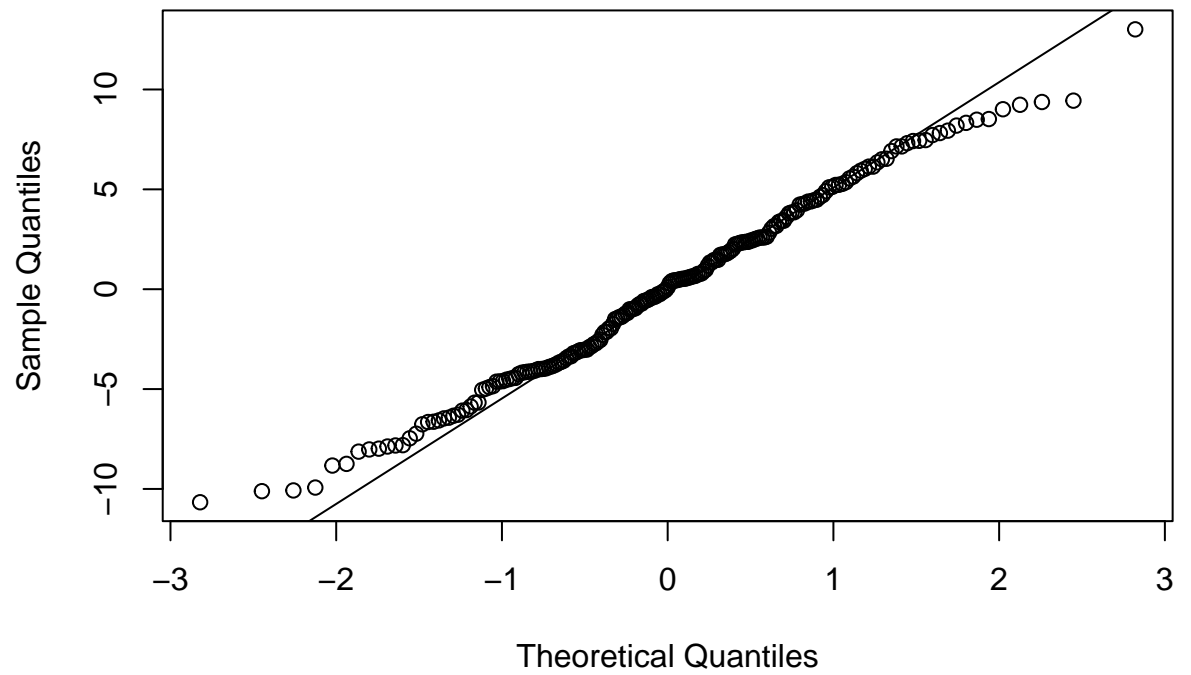



Les parties du corps les plus corrélées sont l'abdomen, chest, hip et thigh.

Nous allons faire un test de régression pour bâtir un modèle prédictif pour la meilleur corrélation, soit l'abdomen.

```
lm_abdomen <- lm(BodyFat ~ Abdomen, data = bodyfat2)
qqnorm(residuals(lm_abdomen))
qqline(residuals(lm_abdomen))
```

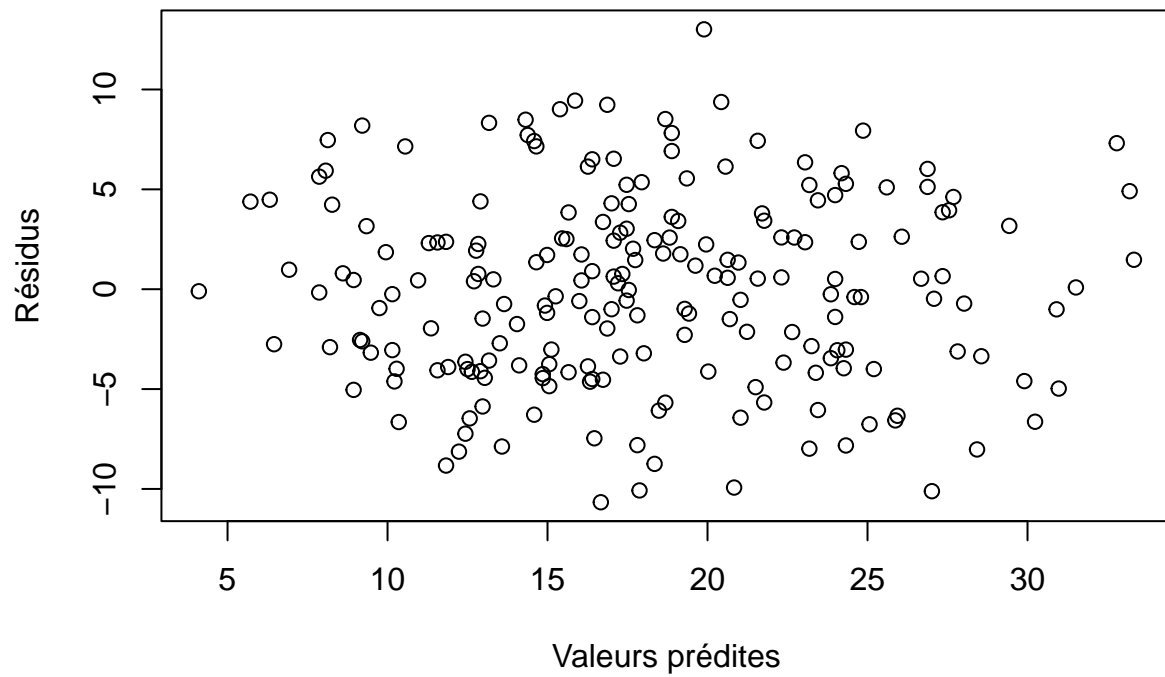
Normal Q-Q Plot



Abdomen

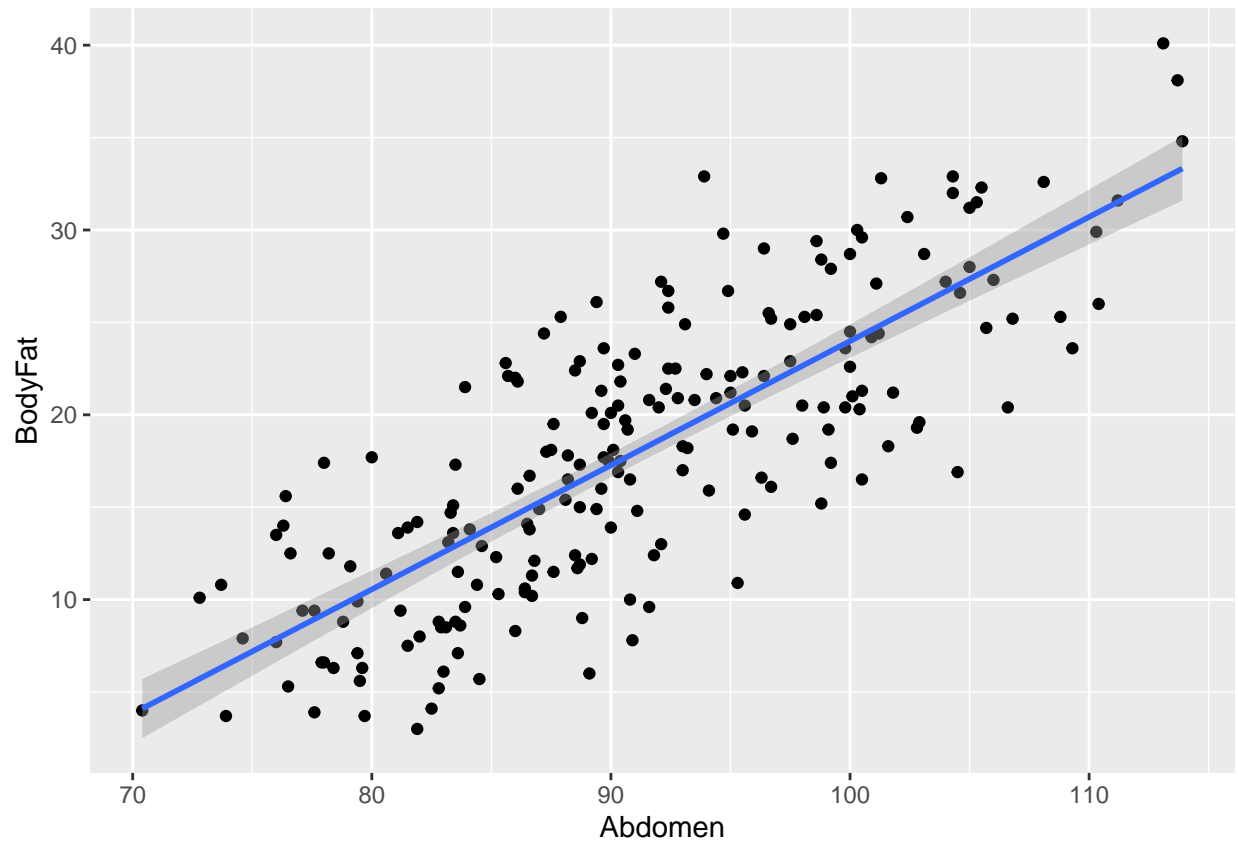
```
plot(fitted(lm_abdomen), residuals(lm_abdomen),  
     ylab = "Résidus", xlab = "Valeurs prédites",  
     main = "Homogénéité des variances")
```

Homogénéité des variances



```
library(ggplot2)
ggplot(bodyfat_Abdomen, aes(x = Abdomen, y = BodyFat)) + geom_point() + geom_smooth(method=lm)

## 'geom_smooth()' using formula = 'y ~ x'
```



```
summary(lm_abdomen)
```

```
##
## Call:
## lm(formula = BodyFat ~ Abdomen, data = bodyfat2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6658  -3.7534   0.0873   3.3670  13.0096
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -43.19208    3.28646  -13.14  <2e-16 ***
## Abdomen      0.67181    0.03589   18.72  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.735 on 207 degrees of freedom
## (4 observations effacées parce que manquantes)
## Multiple R-squared:  0.6286, Adjusted R-squared:  0.6268
## F-statistic: 350.3 on 1 and 207 DF, p-value: < 2.2e-16
```

Relation entre la circonférence de l'abdomen et le pourcentage de graisse corporelle : Les résultats de la régression linéaire indiquent une relation significative entre la circonférence de l'abdomen et le pourcentage de graisse corporelle. L'estimation de la pente (0.67) suggère qu'une **augmentation de 1**

cm dans la circonférence de l'abdomen est associée à une augmentation moyenne de 0.67% du pourcentage de graisse corporelle.

Ajustement du modèle : Le coefficient de détermination ajusté (Adjusted R-squared) de 0.6268 indique que **62.68% de la variance du pourcentage de graisse corporelle peut être expliquée par la circonférence de l'abdomen dans ce modèle.** Cela signifie que la circonférence de l'abdomen est une variable importante dans la prédiction du pourcentage de graisse corporelle, mais d'autres variables peuvent également contribuer à expliquer la variance restante.

Importance de la mesure de la circonférence de l'abdomen : Les résultats soutiennent l'hypothèse selon laquelle la circonférence de l'abdomen est fortement corrélée au pourcentage de graisse corporelle. Cela suggère que la mesure de la circonférence de l'abdomen peut être une méthode utile pour estimer le pourcentage de graisse corporelle, ce qui peut être important dans un contexte militaire afin d'estimer le % de gras chez les membres des Forces Armées Canadiennes.

Modèle prédictif du pourcentage de gras en fonction de la circonférence de l'abdomen: Exemple de pourcentage de gras pour le sujet de la ligne 8 du data frame bodyfat2 ayant 88,5cm de circonférence de l'abdoemen.

```
bodyfat2[8,]
```

```
## # A tibble: 1 x 16
##   Density BodyFat   Age Weight Height  Neck Chest Abdomen  Hip Thigh  Knee
##   <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1    1.07    12.4   25    176   72.5  37.8  99.6    88.5  97.1   60  39.4
## # ... with 5 more variables: Ankle <dbl>, Biceps <dbl>, Forearm <dbl>,
## #   Wrist <dbl>, Ankle_log <dbl>
```

```
pred_abdomen <- predict(lm_abdomen, newdata = bodyfat2[8,], se.fit = TRUE)
inf95_abdomen <- pred_abdomen$fit + qt(p = 0.025, df = lm_abdomen$df.residual) * pred_abdomen$se.fit
sup95_abdomen <- pred_abdomen$fit - qt(p = 0.025, df = lm_abdomen$df.residual) * pred_abdomen$se.fit
c(inf95_abdomen, sup95_abdomen)
```

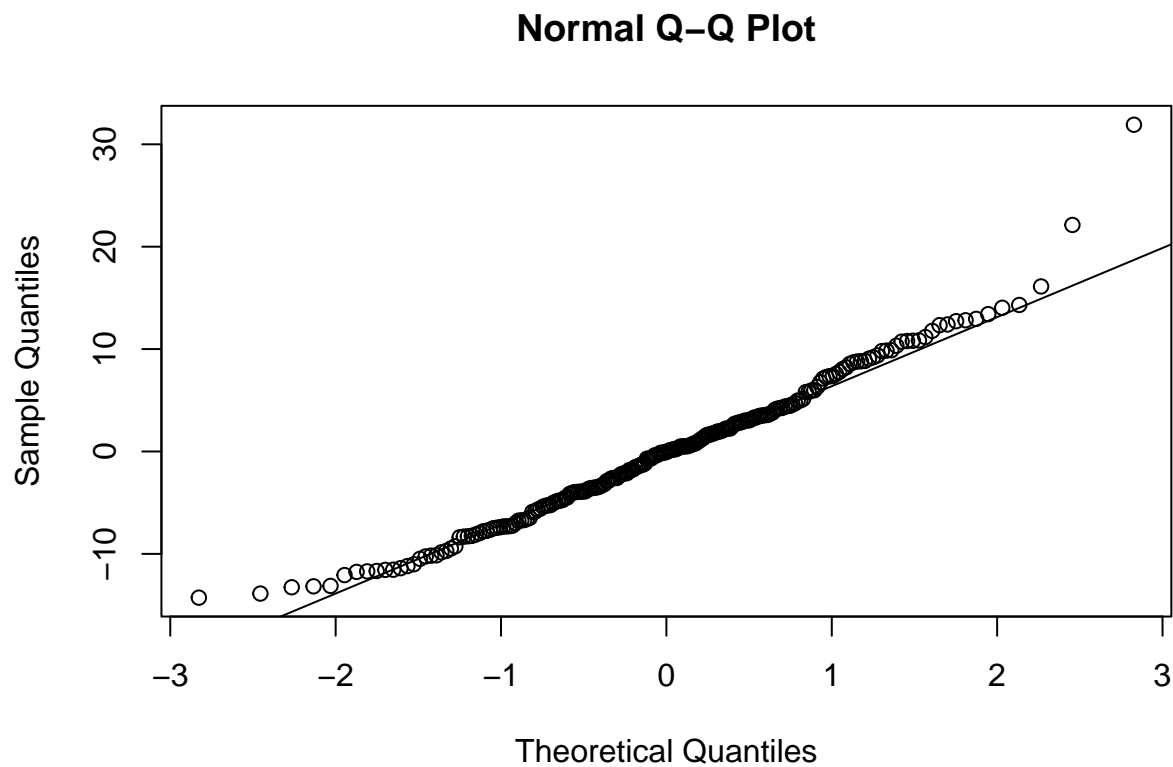
```
##           1           1
## 15.59118 16.93419
```

La différence entre le résultat de l'intervalle de confiance (15.59% - 16.93%) et celui du réel pourcentage de gras du sujet 8 (12.4%) peut provenir du fait que le modèle a été normalisé et qu'il est modifié pour une tranche d'âge maximal de 60 ans afin de refléter la population militaire.

```
lm_knee <- lm(BodyFat ~ Knee, data = bodyfat2)
qqnorm(residuals(lm_knee))
qqline(residuals(lm_knee))
```

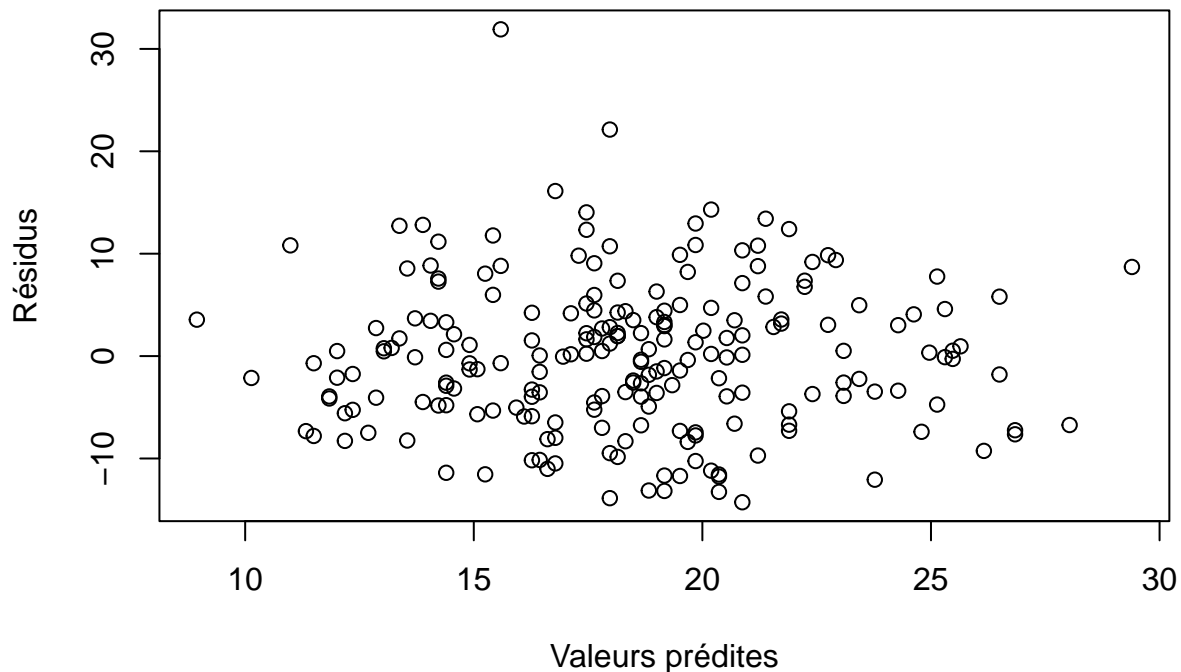
L'objectif de l'évaluation était d'explorer la faisabilité d'utiliser une mesure alternative à la circonférence de l'abdomen des membres dans les Forces Armées Canadiennes. Dans cette optique, nous avons examiné la circonférence du genou comme une mesure potentielle, en raison

de son côté pratique et de son caractère moins intrusif par rapport à d'autres mensurations:



```
plot(fitted(lm_knee), residuals(lm_knee),  
     ylab = "Résidus", xlab = "Valeurs prédites",  
     main = "Homogénéité des variances")
```

Homogénéité des variances



```
summary(lm_knee)
```

```
##
## Call:
## lm(formula = BodyFat ~ Knee, data = bodyfat2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.274  -4.928  -0.054   4.176  31.910
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -47.3034     8.5119  -5.557 8.21e-08 ***
## Knee         1.7044     0.2205   7.729 4.36e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.211 on 211 degrees of freedom
## Multiple R-squared:  0.2206, Adjusted R-squared:  0.2169
## F-statistic: 59.73 on 1 and 211 DF,  p-value: 4.355e-13
```

Les résultats de l'analyse de régression linéaire indiquent une relation moins marquée que celle observée pour la circonférence de l'abdomen, ce qui en fait un modèle moins fiable.

On peut observer un coefficient de pente positif de 1.7044, suggérant qu'une **augmentation de 1 cm dans la circonférence du genou est associée à une augmentation moyenne de 1.7044% du pourcentage**

de graisse corporelle.

Ici, il faut comprendre que la circonférence du genou est naturellement beaucoup plus petite que la circonférence de l'abdomen. C'est pourquoi nous pouvons voir qu'une augmentation de 1cm de circonférence du genou fait augmenter le pourcentage de gras plus rapidement (1.70%) que 1cm de circonférence de l'abdomen (0.67%). De plus, la circonférence du genou peut être affectée par la masse musculaire, en particulier chez les individus ayant des muscles quadriceps bien développés.

Ce modèle montre un coefficient de détermination ajusté (Adjusted R-squared) de 0.2169, indiquant que **21.69% de la variance du pourcentage de graisse corporelle peut être expliquée par la circonférence du genou**. Bien que cette relation soit significative, il faut reconnaître que d'autres facteurs peuvent également influencer le pourcentage de graisse corporelle. Ainsi, nous examinerons si un modèle de régression multiple, intégrant plusieurs variables, permet d'aboutir à une prédiction plus précise.

```
#Combinaison des meilleurs corrélations pour obtenir un meilleur modèle
```

```
lm_Abd_Chest <- lm(BodyFat ~ Abdomen + Chest, data = bodyfat2)
```

```
lm_Abd_Hip <- lm(BodyFat ~ Abdomen + Hip, data = bodyfat2)
```

```
lm_Abd_Thigh <- lm(BodyFat ~ Abdomen + Thigh, data = bodyfat2)
```

```
lm_multiple <- lm(BodyFat ~ Abdomen + Thigh + Chest + Hip, data = bodyfat2)
```

```
summary(lm_Abd_Chest)
```

```
##
## Call:
## lm(formula = BodyFat ~ Abdomen + Chest, data = bodyfat2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.9146 -3.3550  0.0694  3.4163 12.7122
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -31.44507    4.89187   -6.428 9.13e-10 ***
## Abdomen       0.87752    0.07573   11.587 < 2e-16 ***
## Chest       -0.30582    0.09578   -3.193  0.00163 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.635 on 202 degrees of freedom
## (8 observations effacées parce que manquantes)
## Multiple R-squared:  0.6227, Adjusted R-squared:  0.619
## F-statistic: 166.7 on 2 and 202 DF, p-value: < 2.2e-16
```

```
summary(lm_Abd_Hip)
```

```
##
## Call:
## lm(formula = BodyFat ~ Abdomen + Hip, data = bodyfat2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.1731 -3.3350 -0.2283  3.1670 12.7142
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -25.55943    5.95597  -4.291 2.74e-05 ***
## Abdomen      0.86590    0.06764  12.801 < 2e-16 ***
## Hip         -0.35527    0.10336  -3.437 0.000711 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.608 on 205 degrees of freedom
## (5 observations effacées parce que manquantes)
## Multiple R-squared:  0.6453, Adjusted R-squared:  0.6418
## F-statistic: 186.5 on 2 and 205 DF, p-value: < 2.2e-16
```

```
summary(lm_Abd_Thigh)
```

```
##
## Call:
## lm(formula = BodyFat ~ Abdomen + Thigh, data = bodyfat2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.4722  -3.6133   0.1302   3.2815  12.7912
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -41.08644    4.15203  -9.896 <2e-16 ***
## Abdomen      0.71940    0.05706  12.609 <2e-16 ***
## Thigh       -0.10792    0.11053  -0.976    0.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.726 on 205 degrees of freedom
## (5 observations effacées parce que manquantes)
## Multiple R-squared:  0.6335, Adjusted R-squared:  0.6299
## F-statistic: 177.1 on 2 and 205 DF, p-value: < 2.2e-16
```

```
summary(lm_multipe)
```

```
##
## Call:
## lm(formula = BodyFat ~ Abdomen + Thigh + Chest + Hip, data = bodyfat2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##  -9.7647  -3.2134  -0.2223   3.3883  12.8167
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15.60484    6.72899  -2.319 0.021413 *
## Abdomen      0.99315    0.08741  11.362 < 2e-16 ***
## Thigh        0.30305    0.15270   1.985 0.048576 *
## Chest       -0.23737    0.09551  -2.485 0.013776 *
## Hip         -0.51487    0.14868  -3.463 0.000655 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.508 on 198 degrees of freedom
## (10 observations effacées parce que manquantes)
## Multiple R-squared:  0.6428, Adjusted R-squared:  0.6356
## F-statistic: 89.1 on 4 and 198 DF,  p-value: < 2.2e-16
```

Dans ce cas, les diverses combinaisons entre les variables les plus corrélées au pourcentage de graisse dans une analyse univariée ne fournissent pas un modèle amélioré. En effet, l'association de chaque variable avec la circonférence de l'abdomen révèle une diminution du pourcentage de graisse corporelle avec l'augmentation de leurs circonférences, ce qui peut entraîner de la multicollinéarité.

Explorons la possibilité de combiner des variables moins corrélées à la mesure de l'abdomen.

```
#Combinaison des plus faibles corrélations avec la mesure de l'abdomen.
lm_Abd_Ankle <- lm(BodyFat ~ Abdomen + Ankle_log, data = bodyfat2)
lm_Abd_Biceps <- lm(BodyFat ~ Abdomen + Biceps, data = bodyfat2)
lm_Abd_Forearm <- lm(BodyFat ~ Abdomen + Forearm, data = bodyfat2)
lm_Abd_Neck <- lm(BodyFat ~ Abdomen + Neck, data = bodyfat2)

lm_multiple_cor <- lm(BodyFat ~ Abdomen + Neck + Forearm + Biceps + Ankle_log, data = bodyfat2)

summary(lm_Abd_Ankle)
```

```
##
## Call:
## lm(formula = BodyFat ~ Abdomen + Ankle_log, data = bodyfat2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.5658  -3.2965  -0.2789   3.5521  11.0366
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  25.23509   16.25333   1.553   0.122
## Abdomen       0.74816    0.03879  19.286 < 2e-16 ***
## Ankle_log    -24.03358    5.59996  -4.292 2.73e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.547 on 206 degrees of freedom
## (4 observations effacées parce que manquantes)
## Multiple R-squared:  0.6591, Adjusted R-squared:  0.6557
## F-statistic: 199.1 on 2 and 206 DF,  p-value: < 2.2e-16
```

```
summary(lm_Abd_Biceps)
```

```
##
## Call:
## lm(formula = BodyFat ~ Abdomen + Biceps, data = bodyfat2)
##
## Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -10.3163 -3.5689  0.1218   3.0422  13.1112
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -38.79180    3.84797 -10.081  <2e-16 ***
## Abdomen      0.73677    0.04669  15.780  <2e-16 ***
## Biceps      -0.32032    0.14904  -2.149   0.0328 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.694 on 206 degrees of freedom
## (4 observations effacées parce que manquantes)
## Multiple R-squared:  0.6367, Adjusted R-squared:  0.6332
## F-statistic: 180.5 on 2 and 206 DF,  p-value: < 2.2e-16
```

```
summary(lm_Abd_Forearm)
```

```
##
## Call:
## lm(formula = BodyFat ~ Abdomen + Forearm, data = bodyfat2)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -10.1910 -3.6425  0.1323   3.2714  13.2504
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -33.65729    5.09122  -6.611 3.31e-10 ***
## Abdomen      0.75848    0.04556  16.648  < 2e-16 ***
## Forearm     -0.60728    0.22271  -2.727  0.00696 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.67 on 203 degrees of freedom
## (7 observations effacées parce que manquantes)
## Multiple R-squared:  0.6424, Adjusted R-squared:  0.6389
## F-statistic: 182.3 on 2 and 203 DF,  p-value: < 2.2e-16
```

```
summary(lm_Abd_Neck)
```

```
##
## Call:
## lm(formula = BodyFat ~ Abdomen + Neck, data = bodyfat2)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -10.217  -3.461  -0.010   3.111  13.844
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.87956    5.20799  -4.009 8.52e-05 ***
## Abdomen      0.84116    0.04638  18.137  < 2e-16 ***
```

```
## Neck          -0.99870      0.18768  -5.321 2.68e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.45 on 206 degrees of freedom
## (4 observations effacées parce que manquantes)
## Multiple R-squared:  0.6735, Adjusted R-squared:  0.6703
## F-statistic: 212.4 on 2 and 206 DF, p-value: < 2.2e-16
```

```
summary(lm_multiple_cor)
```

```
##
## Call:
## lm(formula = BodyFat ~ Abdomen + Neck + Forearm + Biceps + Ankle_log,
##     data = bodyfat2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.5114 -3.1230 -0.2201  2.9993 12.1389
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.32758   16.07185   1.700  0.09062 .
## Abdomen       0.86646    0.04916  17.626 < 2e-16 ***
## Neck         -0.94601    0.22669  -4.173 4.48e-05 ***
## Forearm       0.15414    0.29142   0.529  0.59744
## Biceps        0.07154    0.17940   0.399  0.69047
## Ankle_log    -18.88525    5.80353  -3.254  0.00134 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.379 on 200 degrees of freedom
## (7 observations effacées parce que manquantes)
## Multiple R-squared:  0.6902, Adjusted R-squared:  0.6824
## F-statistic: 89.11 on 5 and 200 DF, p-value: < 2.2e-16
```

Nous observons exactement le même phénomène qu’avec la combinaison de variables fortement corrélées.

Conclusion

N’étant pas parvenus à élaborer un meilleur modèle en combinant différentes circonférences, et bien que la circonférence du genou puisse être envisagée comme une mesure alternative pour estimer le pourcentage de graisse corporelle dans le contexte des Forces Armées Canadiennes, la circonférence de l’abdomen demeure la mesure la plus fiable. Ceci s’explique par son coefficient de détermination supérieur (62.68% contre 21.69%) et par son impact potentiellement moins important de la masse musculaire du quadriceps chez les sujets bien entraînés.

De plus, étant donné que la graisse située près des organes vitaux est plus dommageable que celle autour du genou, il est plus important d’obtenir la donnée concernant la circonférence de l’abdomen.