SOUTENANCE PROJET 10 PAUL BÉTOUS

OPENCLASSROOMS

Détectez des faux billets avec R ou Python

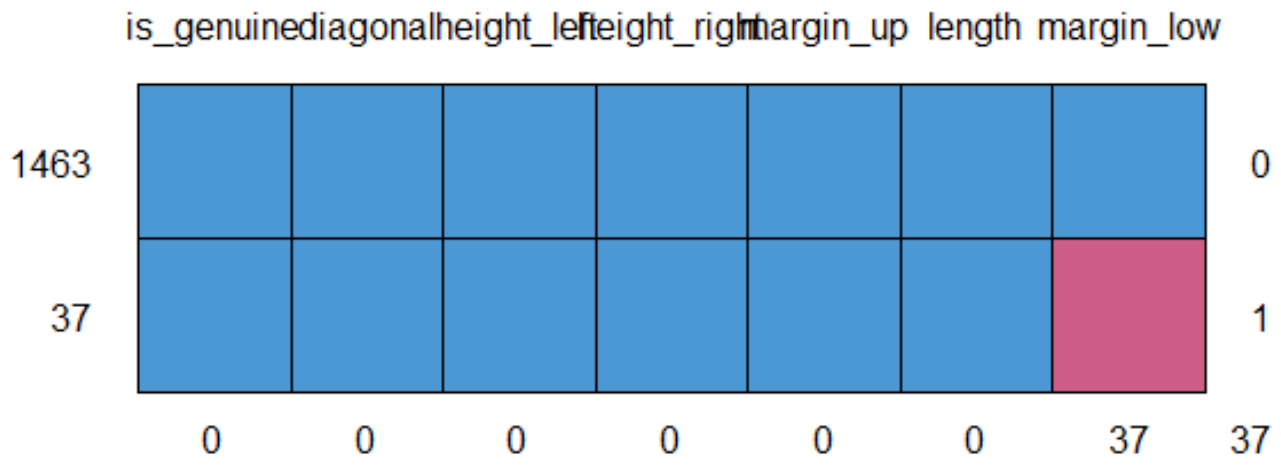
SOMMAIRE

Introduction

- *Distribution des billets*
- *Algorithmes prédictifs*

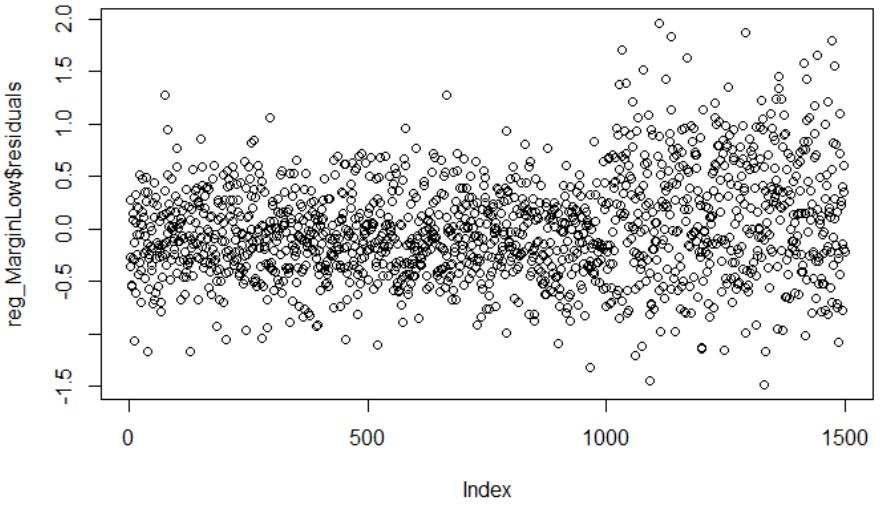
Conclusion

VALEURS NULLES



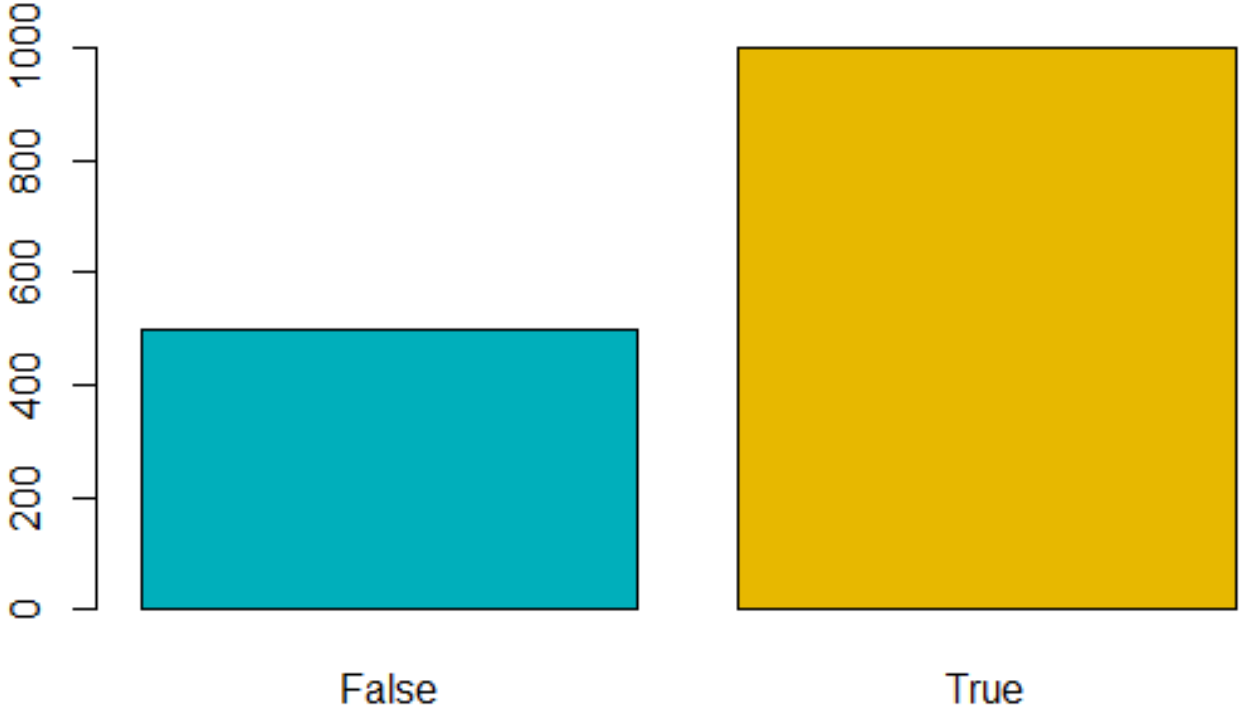
Info Régression Linéaire

- *L'ensemble des variables sont corrélées à la variable « Margin Low »*
- *Test de Shapiro Wilk → les résidus ne suivent pas une loi normale*
- *Test de Breush Pagan → Homoscédasticité des résidus*
- *Le R^2 multiple est égal à 0,48*

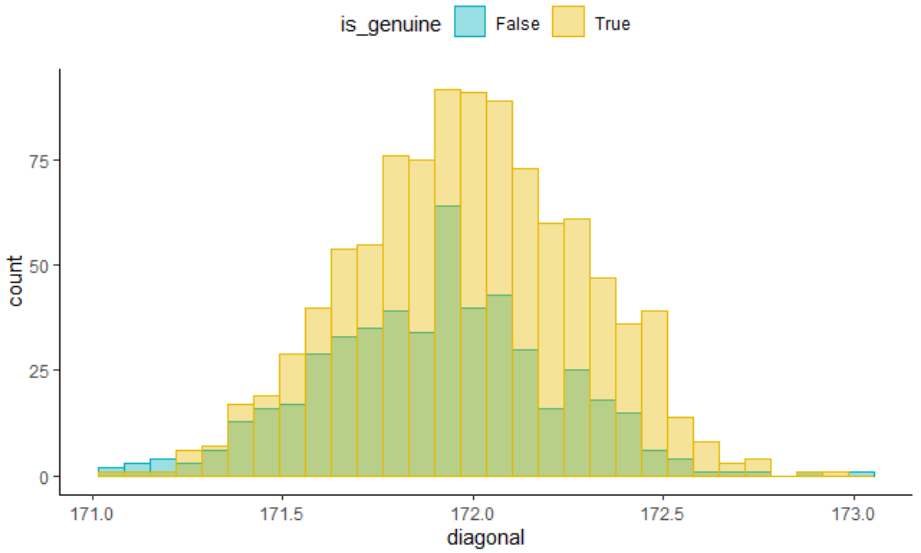


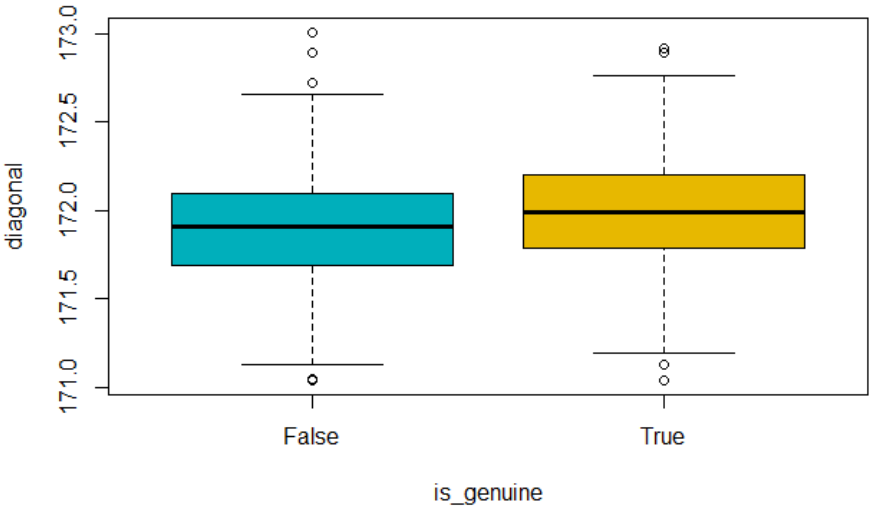
DISTRIBUTION DES BILLETS

RÉPARTITION



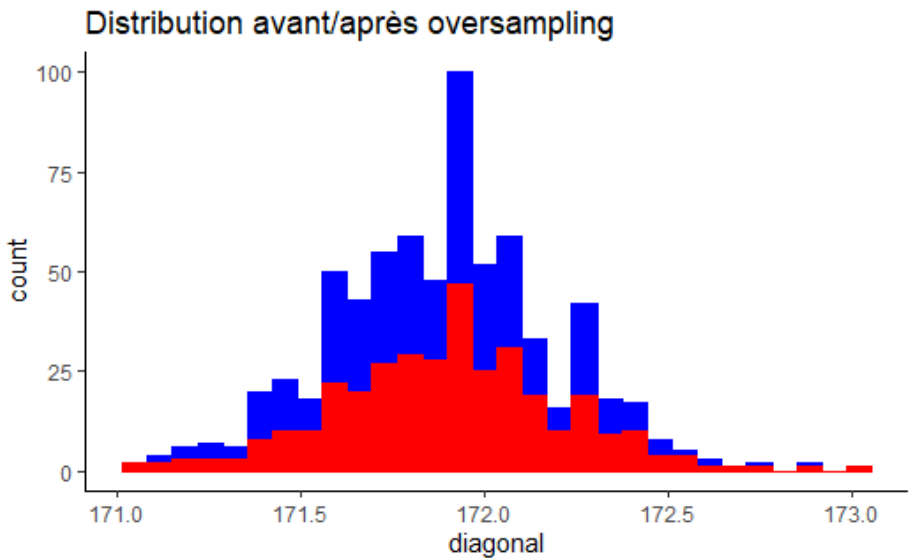
VARIABLE DIAGONAL

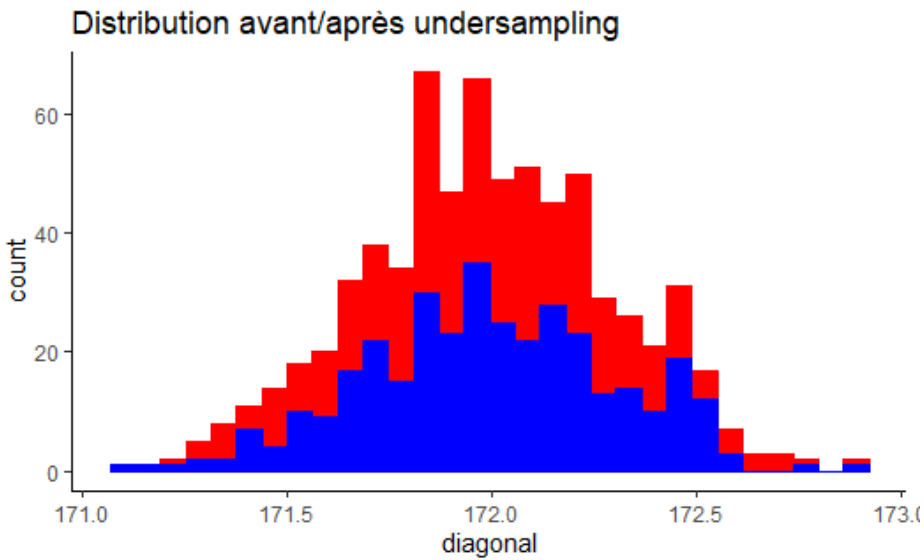




MOY. (mm) = 171,9 / 172
ÉCART-TYPE (mm) = 0,31 / 0,3

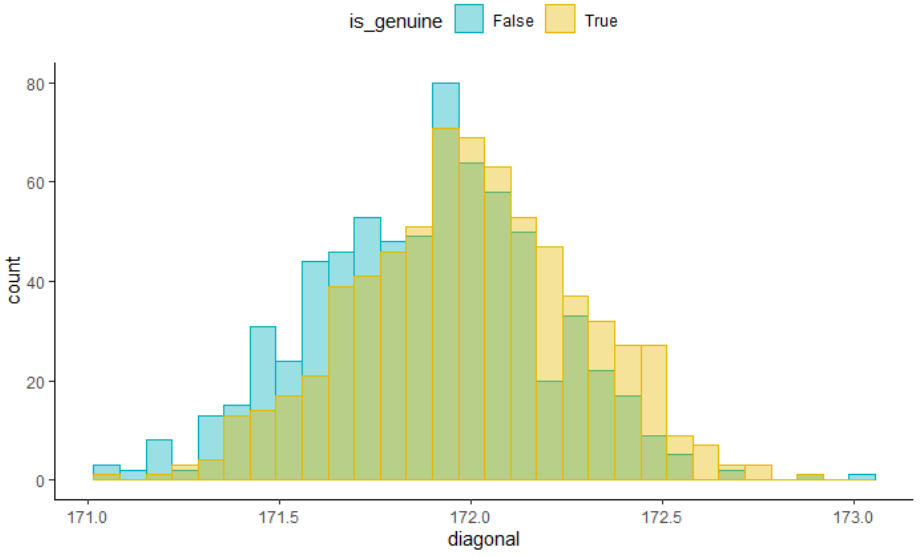
VARIABLE DIAGONAL avec RESAMPLING

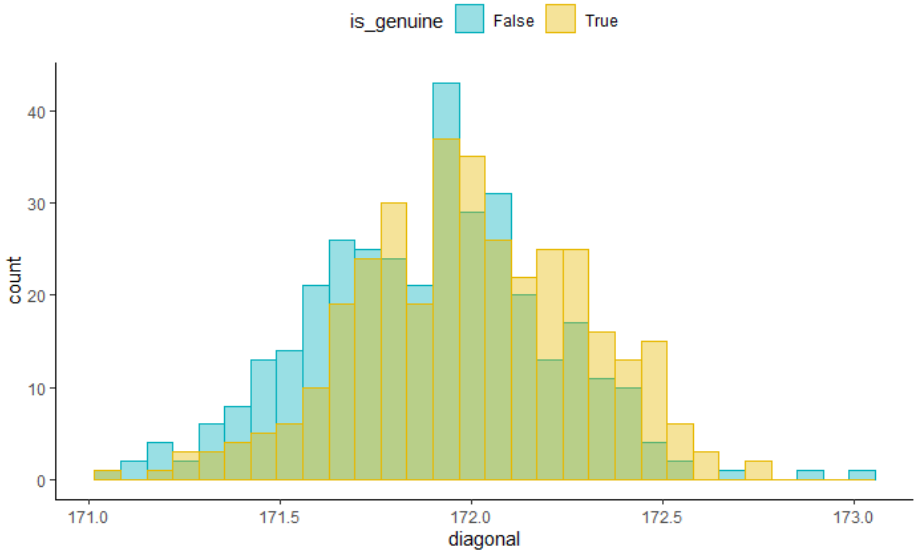




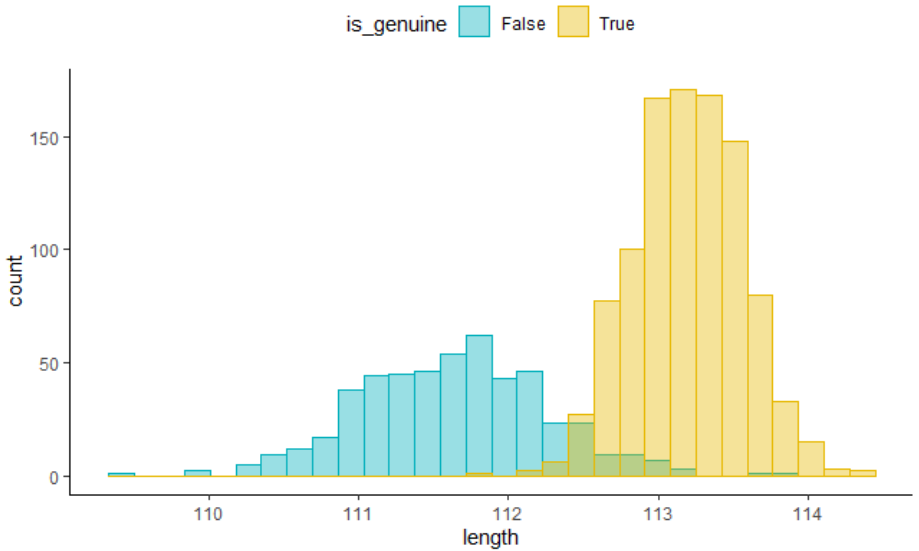
MOY. (mm) = 171,9 (Faux) / 171,9 (Vrai)
ÉCART-TYPE (mm) = 0,31 / 0,31

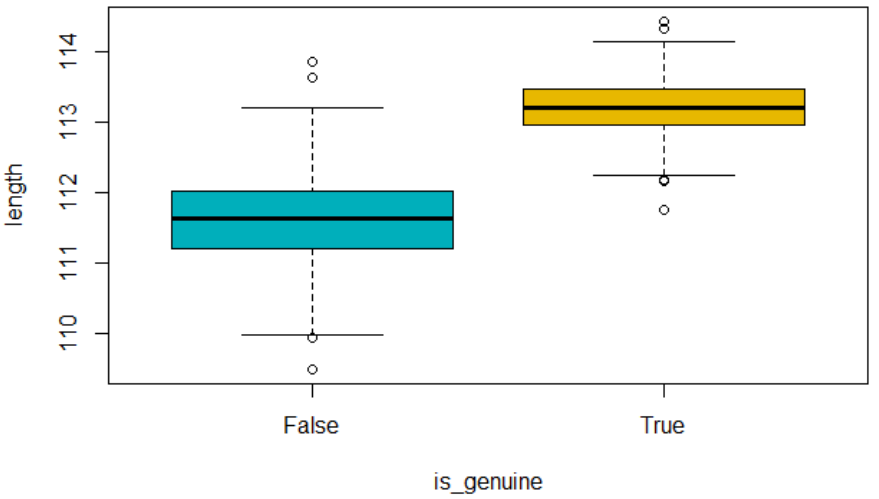
VARIABLE DIAGONAL avec RESAMPLING





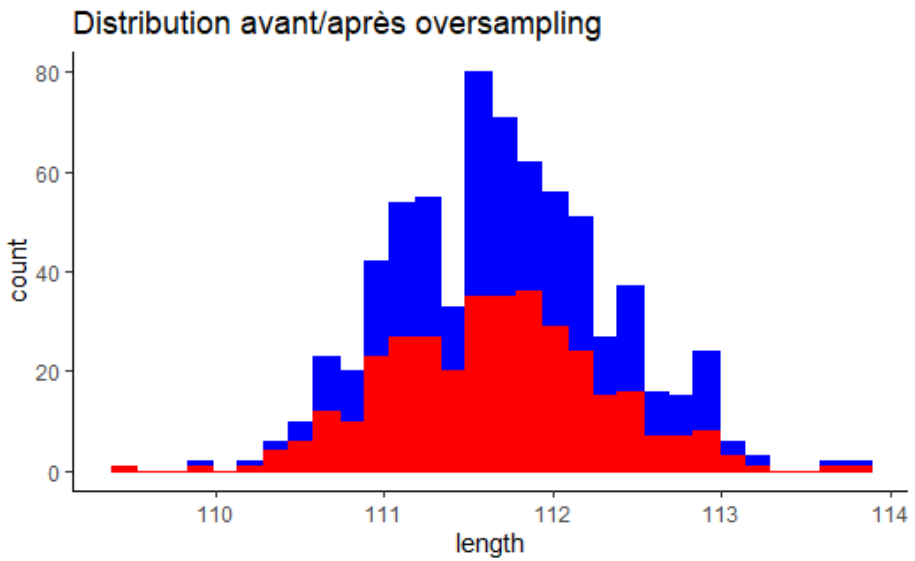
VARIABLE LONGUEUR

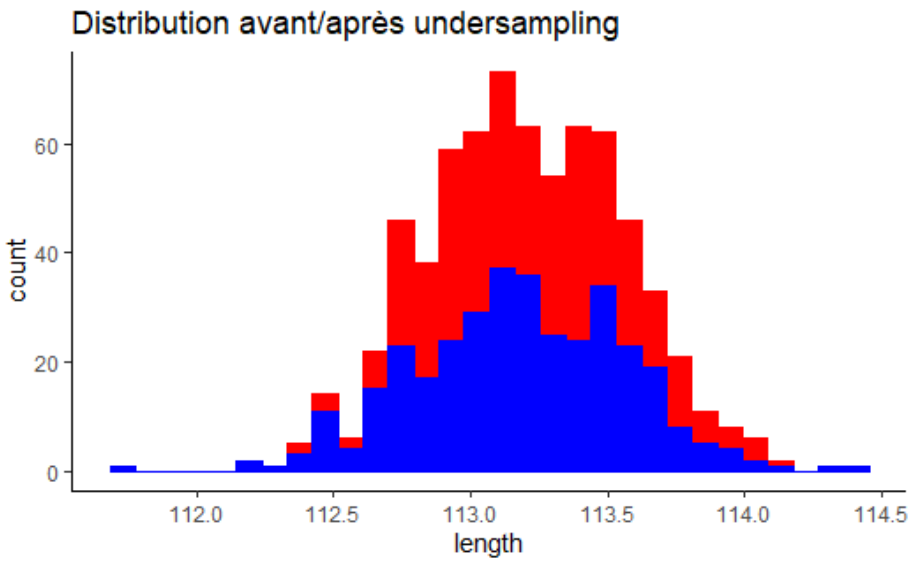




MOY. (mm) = 111,6 / 113,2
ÉCART-TYPE (mm) = 0,62 / 0,36

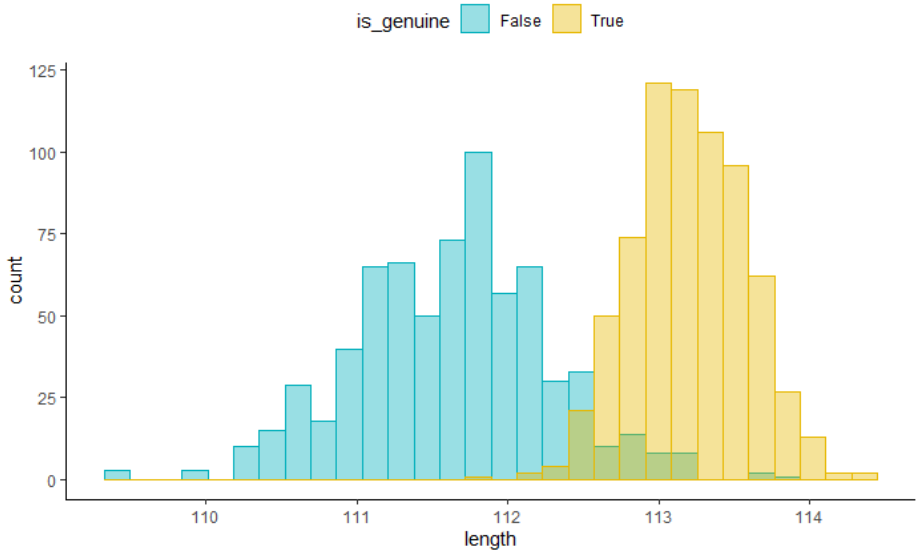
VARIABLE LONGUEUR avec RESAMPLING

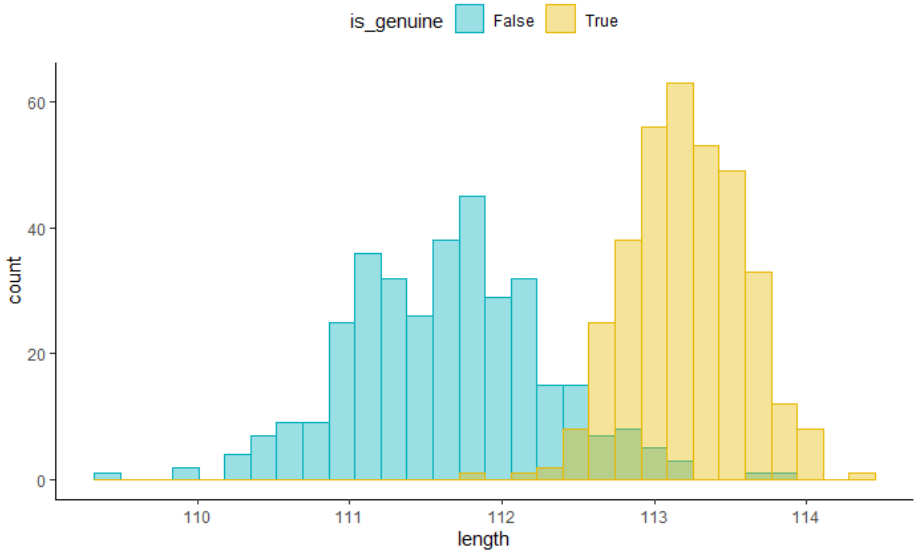




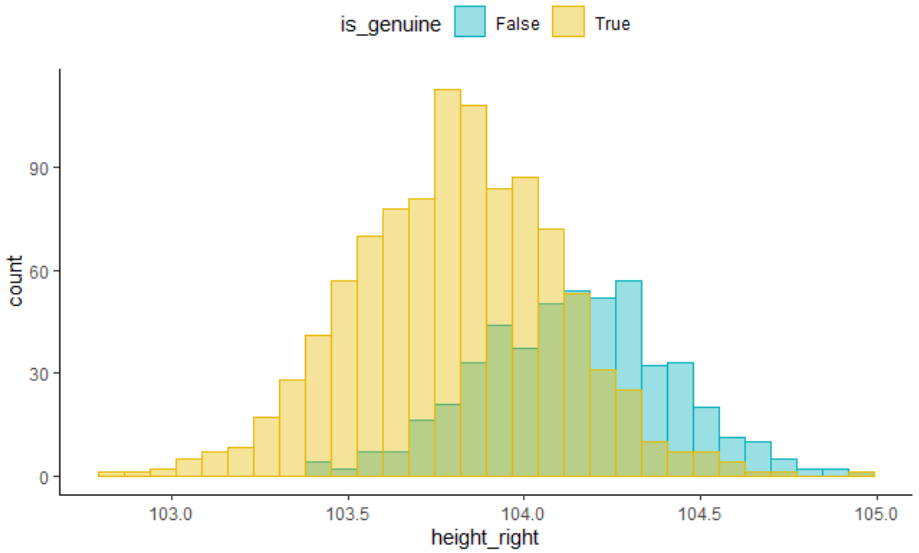
MOY. (mm) = 112,5 (Faux) / 112,4 (Vrai)
ÉCART-TYPE (mm) = 0,93 / 0,94

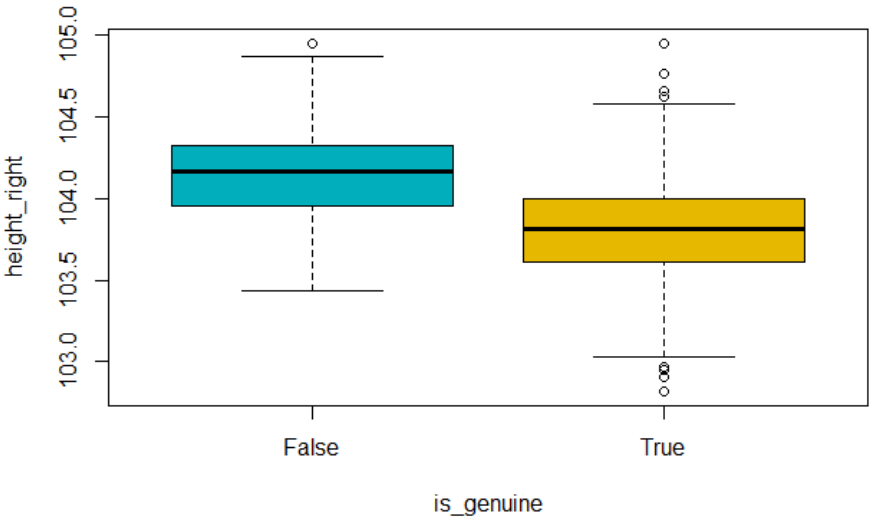
VARIABLE LONGUEUR avec RESAMPLING





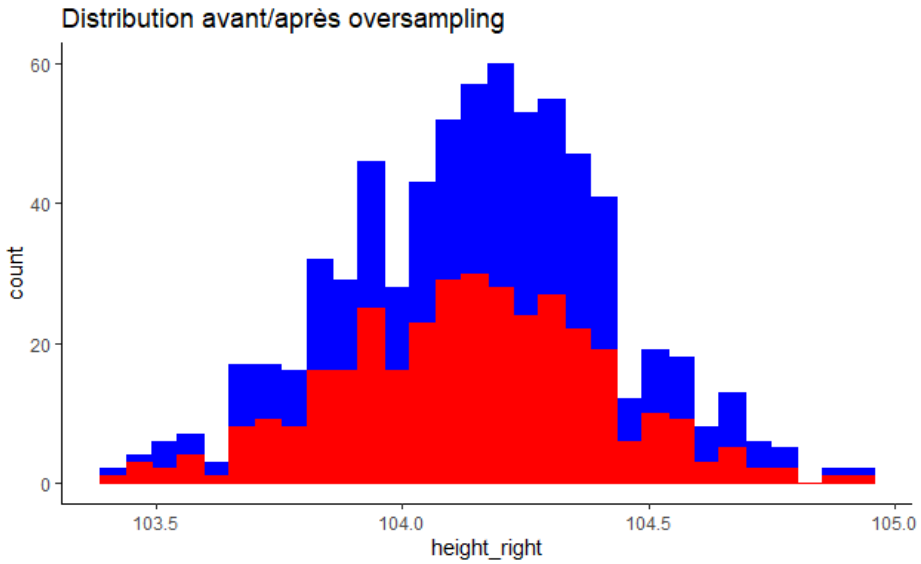
VARIABLE LARGEUR DROITE

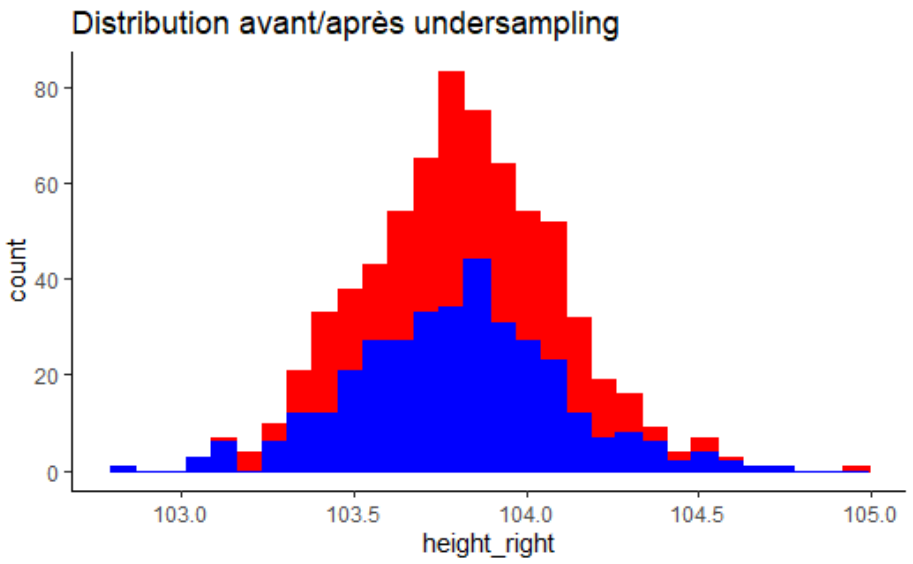




MOY. (mm) = 104,1 / 103,8
ÉCART-TYPE (mm) = 0,27 / 0,29

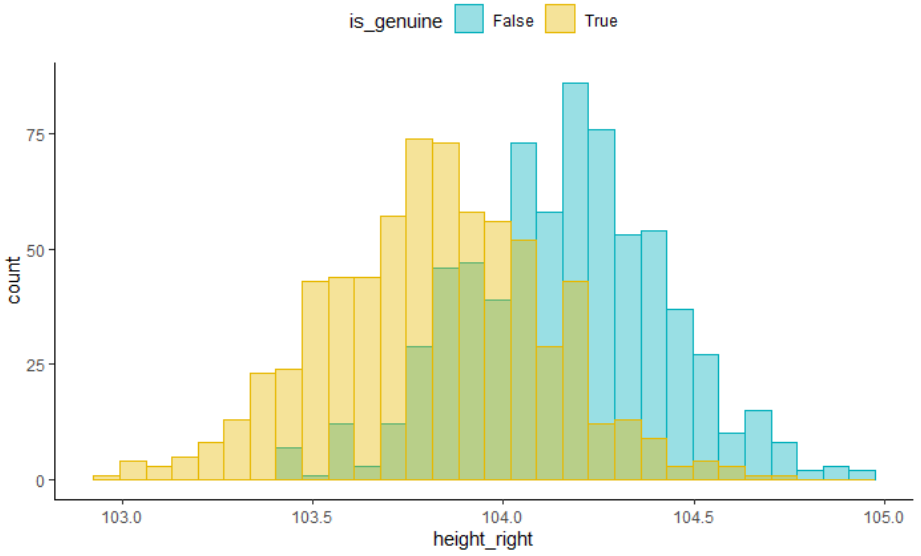
VARIABLE LARGEUR DROITE avec RESAMPLING

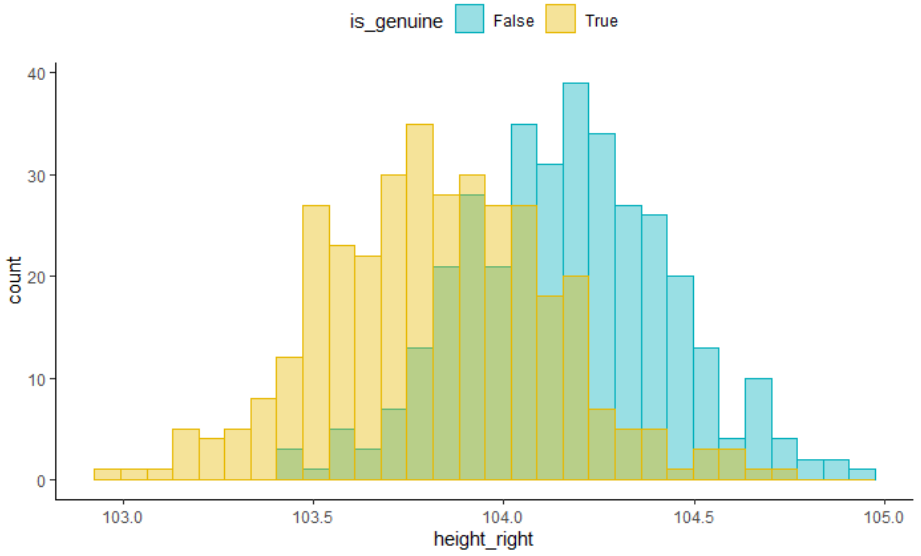




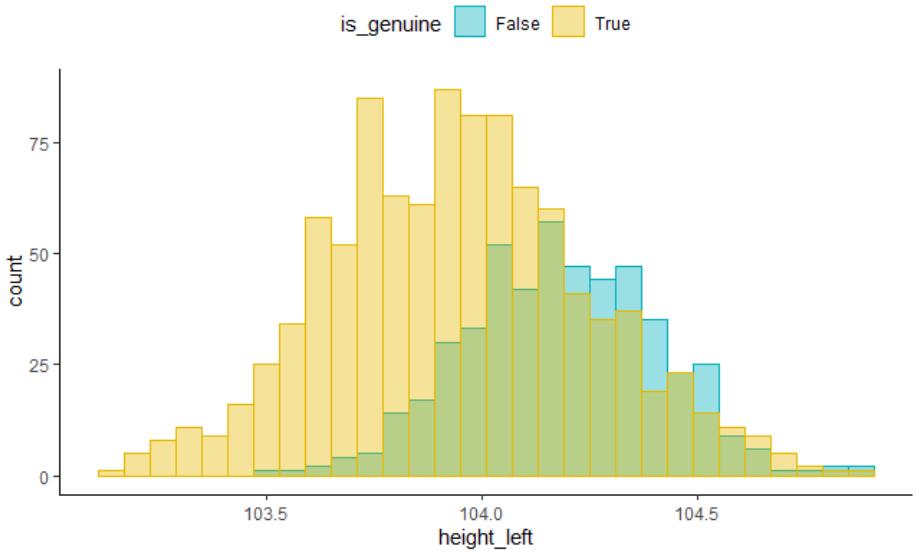
MOY. (mm) = 104 (Faux) / 104 (Vrai)
ÉCART-TYPE (mm) = 0,33 / 0,33

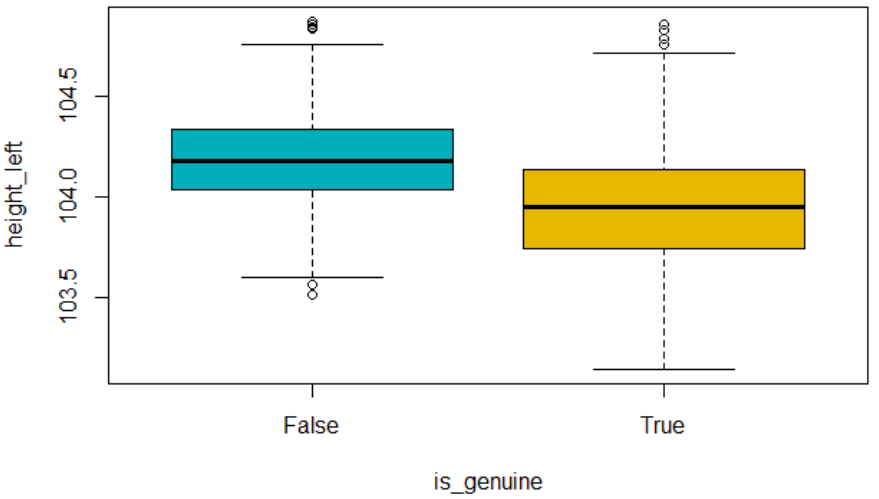
VARIABLE LARGEUR DROITE avec RESAMPLING





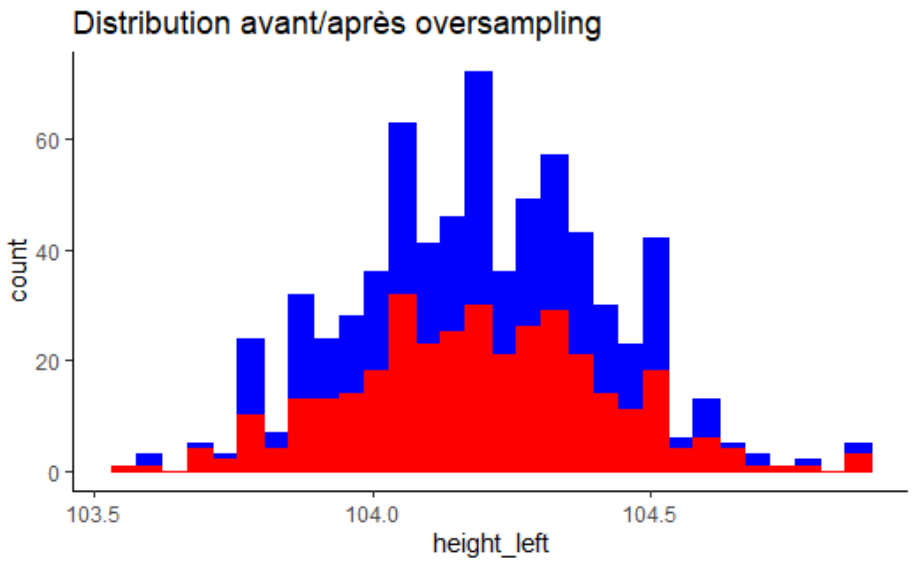
VARIABLE LARGEUR GAUCHE

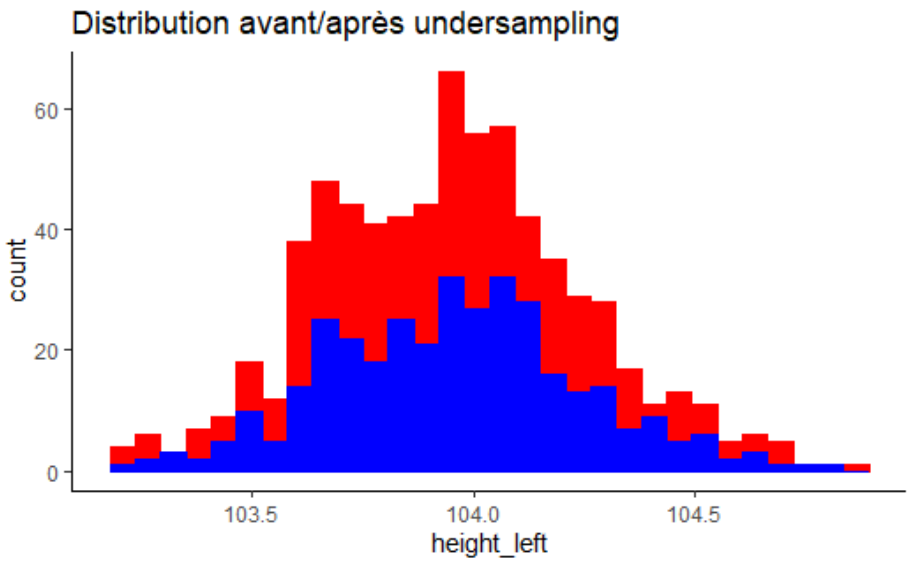




MOY. (mm) = 104,2 / 103,9
ÉCART-TYPE (mm) = 0,22 / 0,3

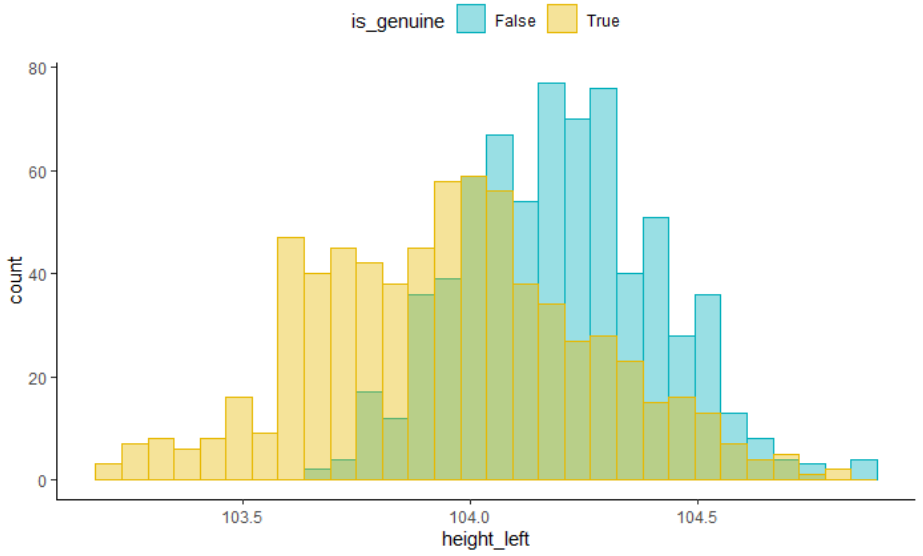
VARIABLE LARGEUR GAUCHE avec RESAMPLING

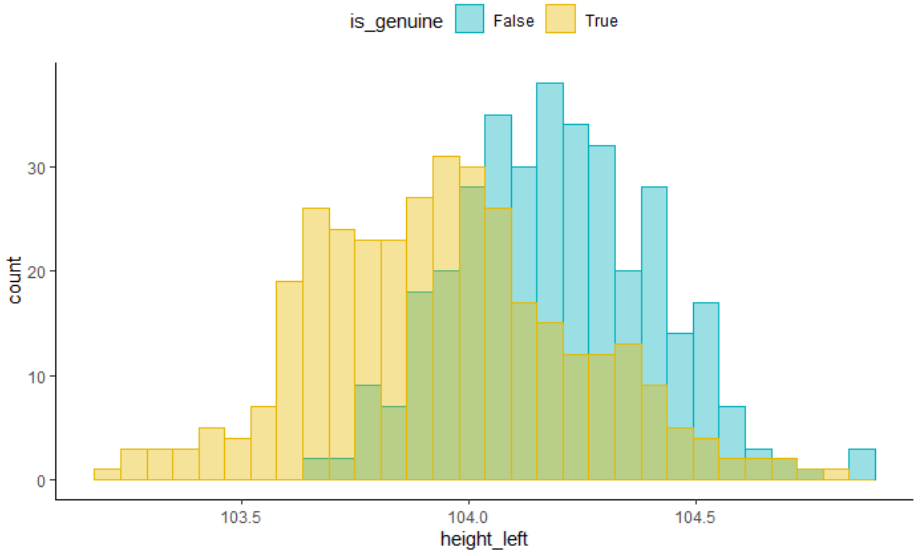




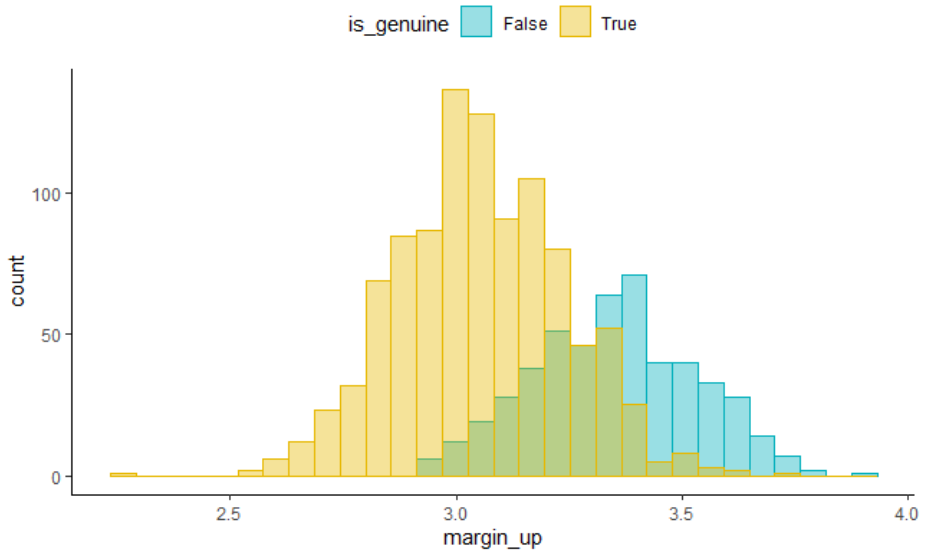
MOY. (mm) = 104,1 (Faux) / 104,1 (Vrai)
ÉCART-TYPE (mm) = 0,3 / 0,3

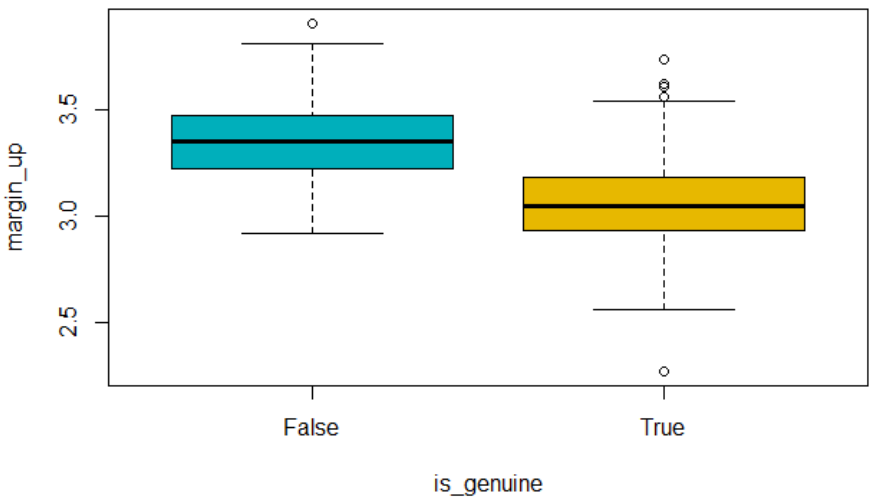
VARIABLE LARGEUR GAUCHE avec RESAMPLING





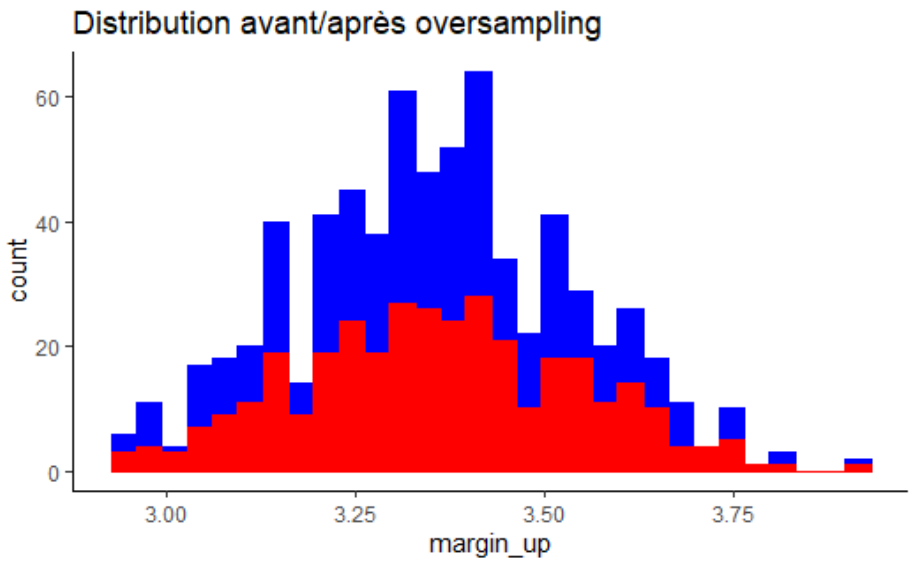
VARIABLE MARGE HAUTE

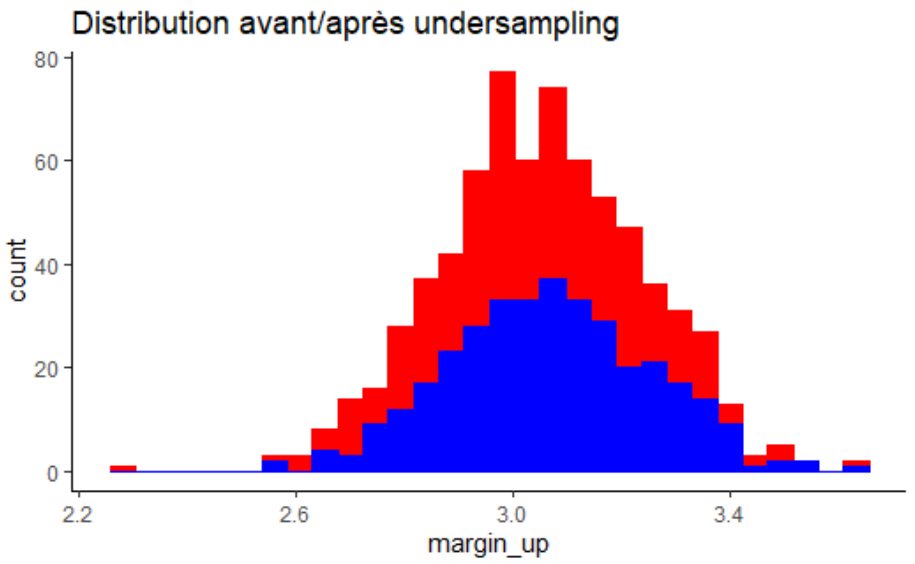




MOY. (mm) = 3,35 / 3,052
ÉCART-TYPE (mm) = 0,18 / 0,19

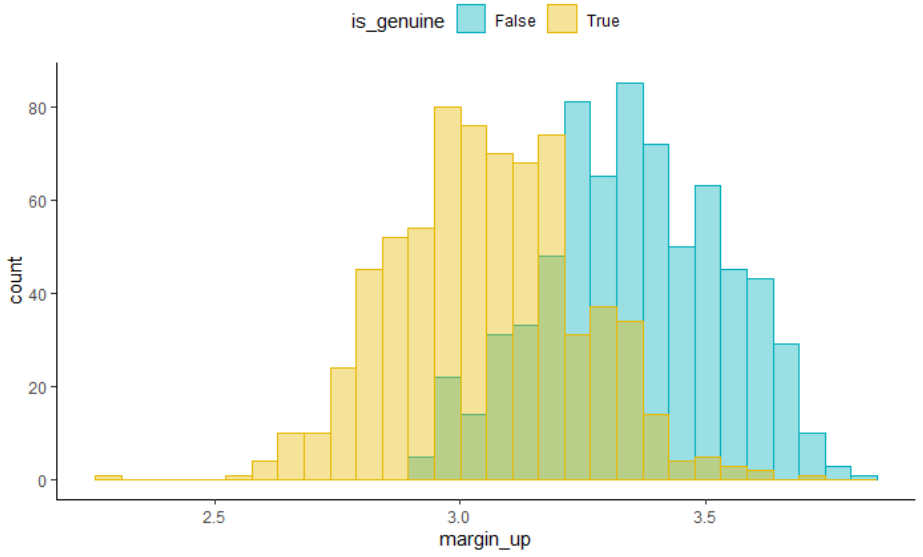
VARIABLE MARGE HAUTE avec RESAMPLING

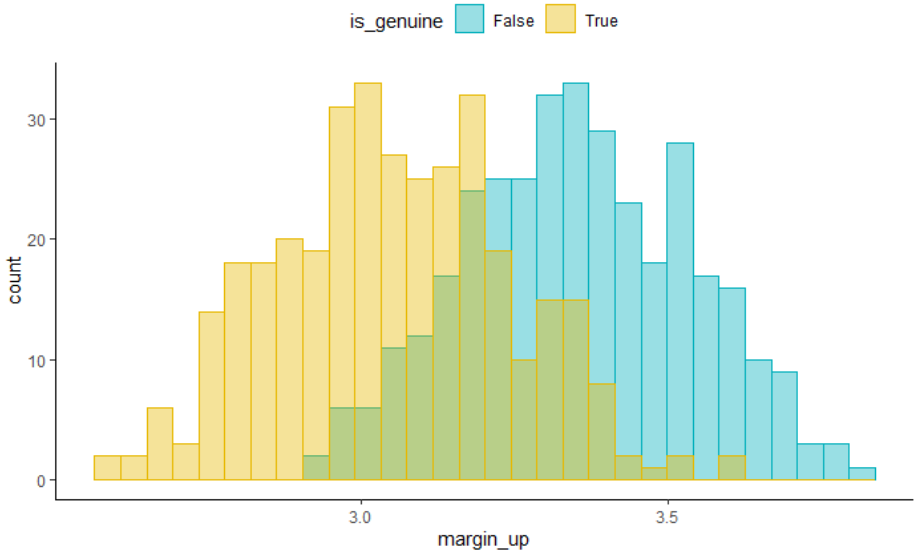




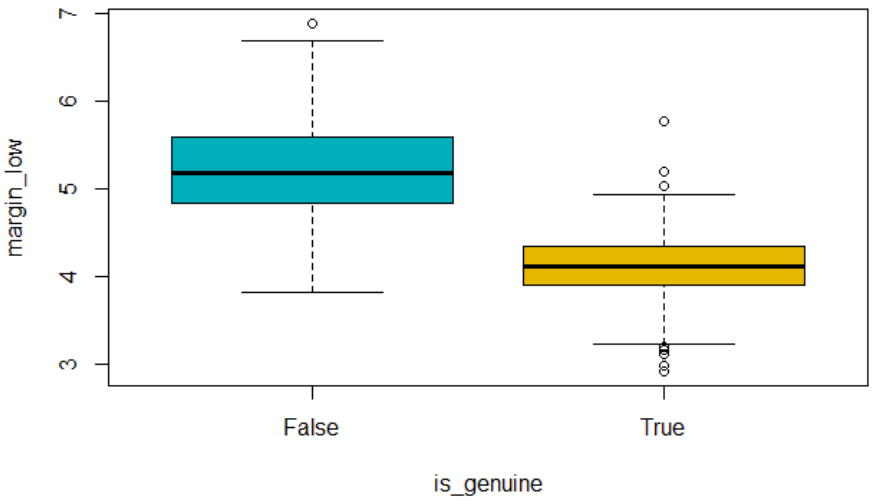
MOY. (mm) = 3,205 (Faux) / 3,209 (Vrai)
ÉCART-TYPE (mm) = 0,24 / 0,24

VARIABLE MARGE HAUTE avec RESAMPLING



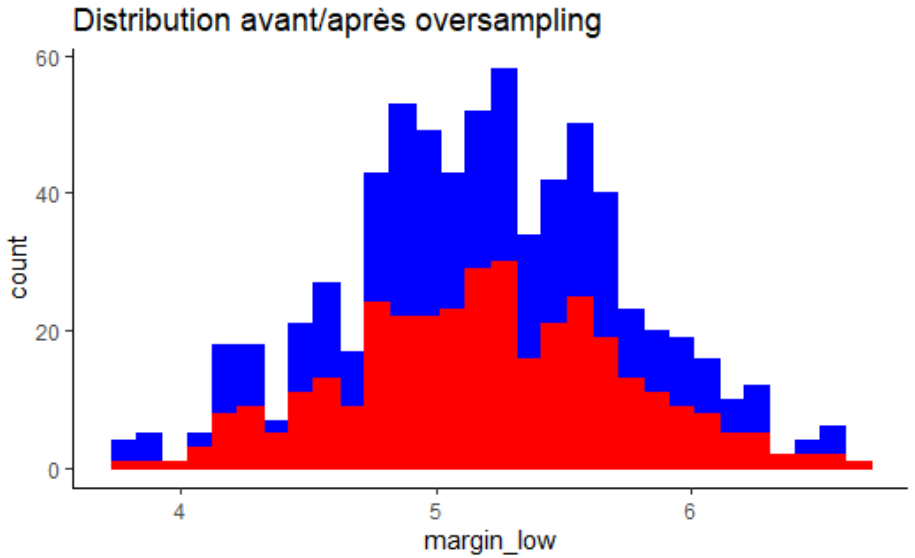


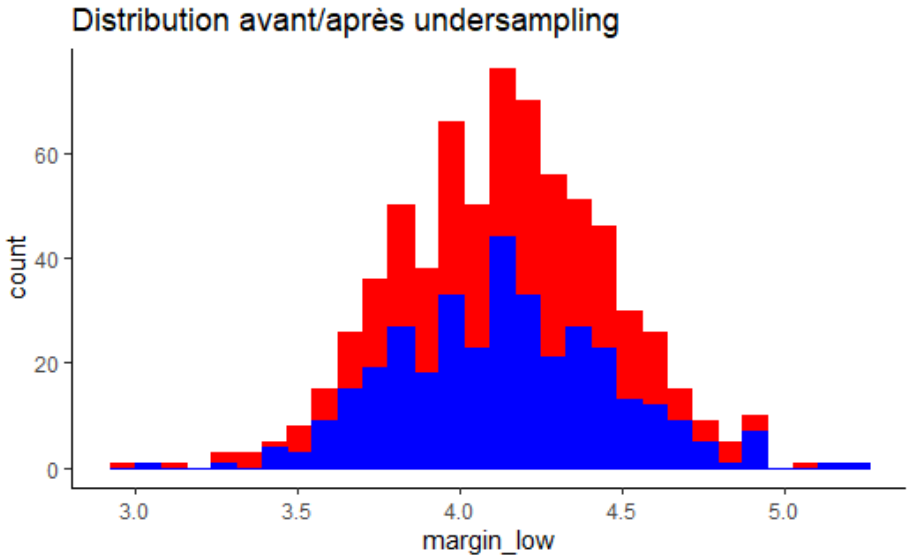
VARIABLE MARGE BASSE



MOY. (mm) = 5,216 / 4,118
ÉCART-TYPE (mm) = 0,55 / 0,33

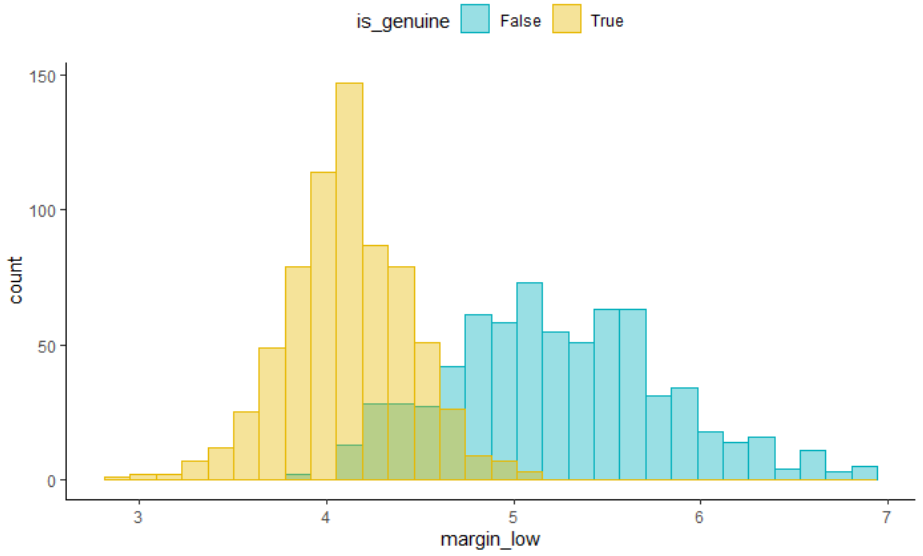
VARIABLE MARGE BASSE avec RESAMPLING

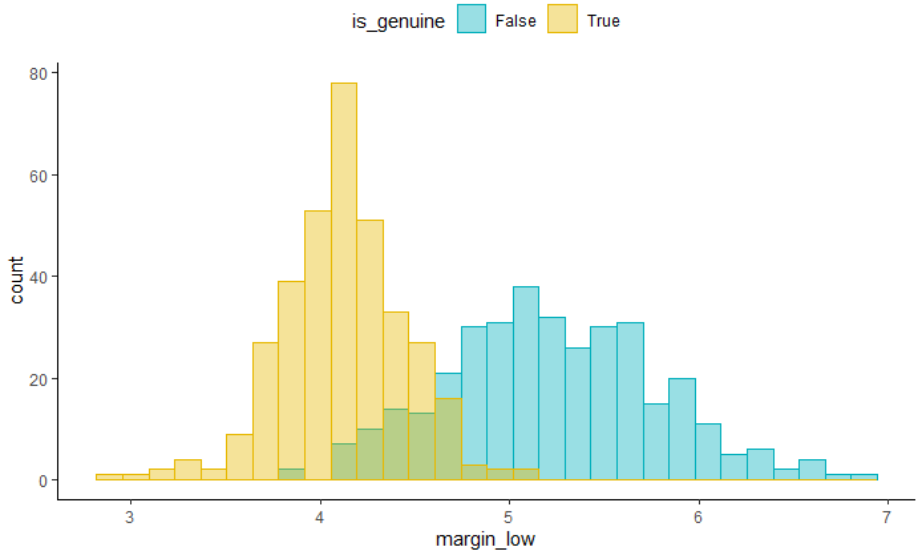




MOY. (mm) = 4,651 (Faux) / 4,650 (Vrai)
ÉCART-TYPE (mm) = 0,72 / 0,71

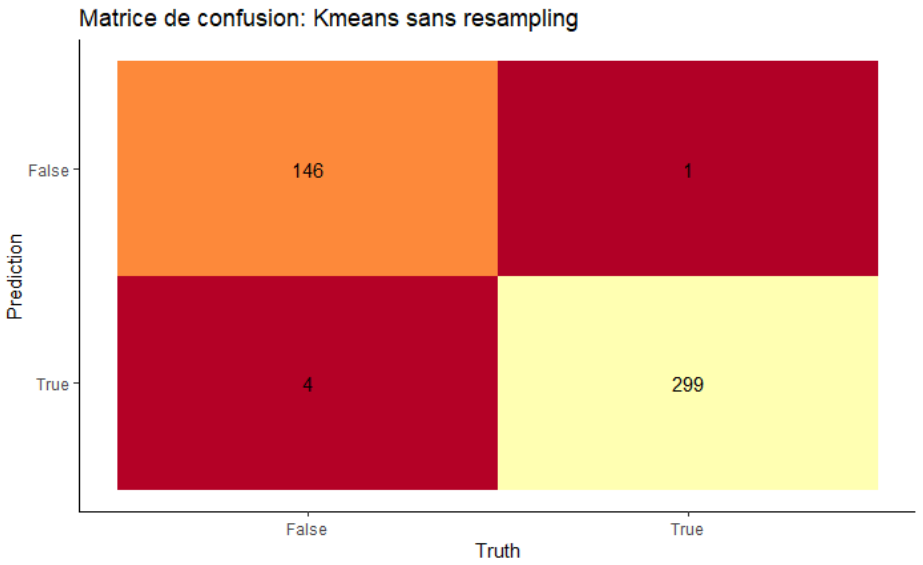
VARIABLE MARGE BASSE avec RESAMPLING

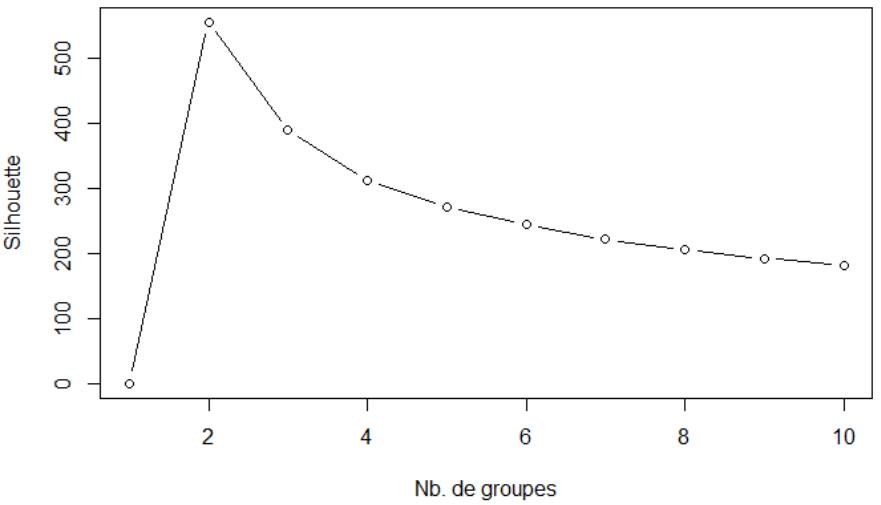




ALGORITHMES PRÉDICTIFS

SANS RESAMPLING

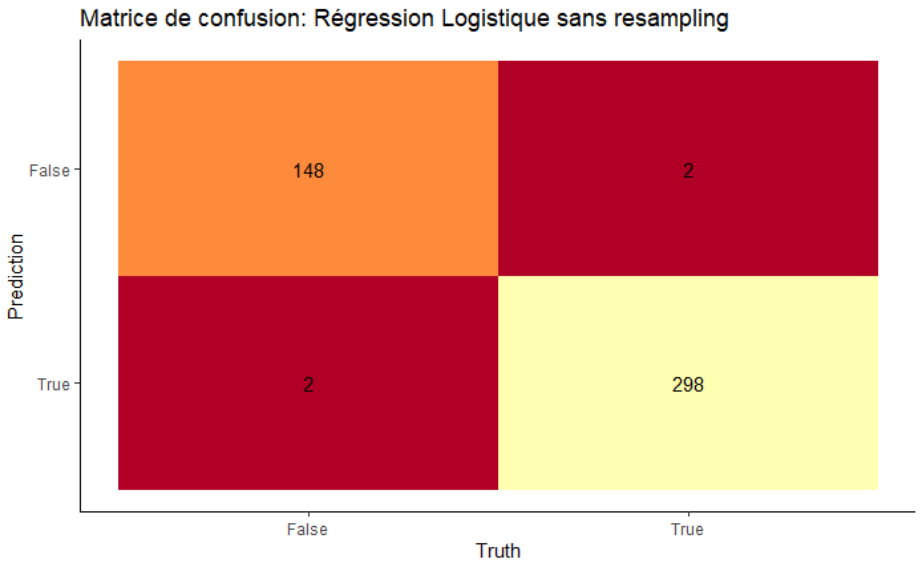


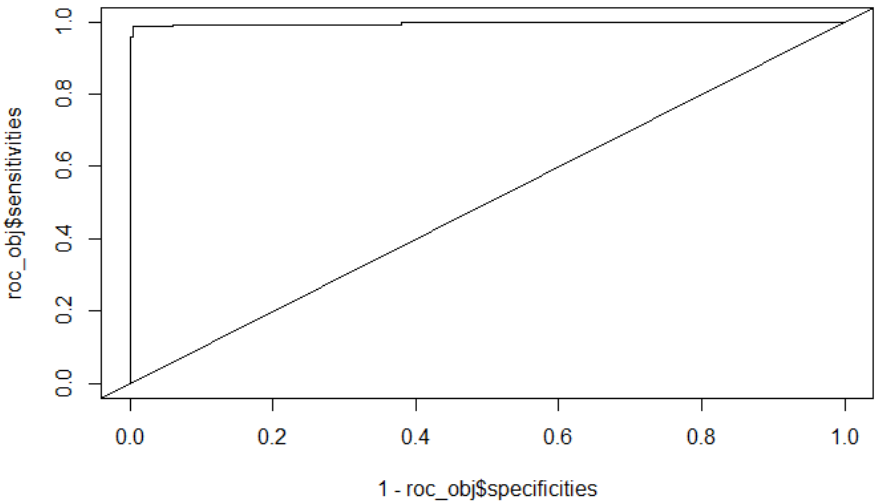


Accuracy = 0.99 / Recall = 0,97 / F-score = 0,98 / Precision = 0,99

SANS RESAMPLING

SANS RESAMPLING

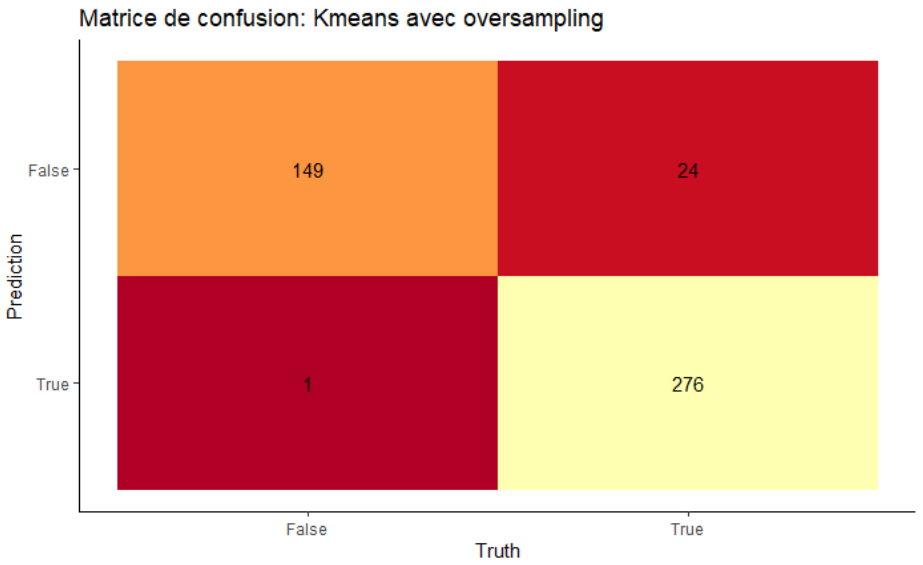


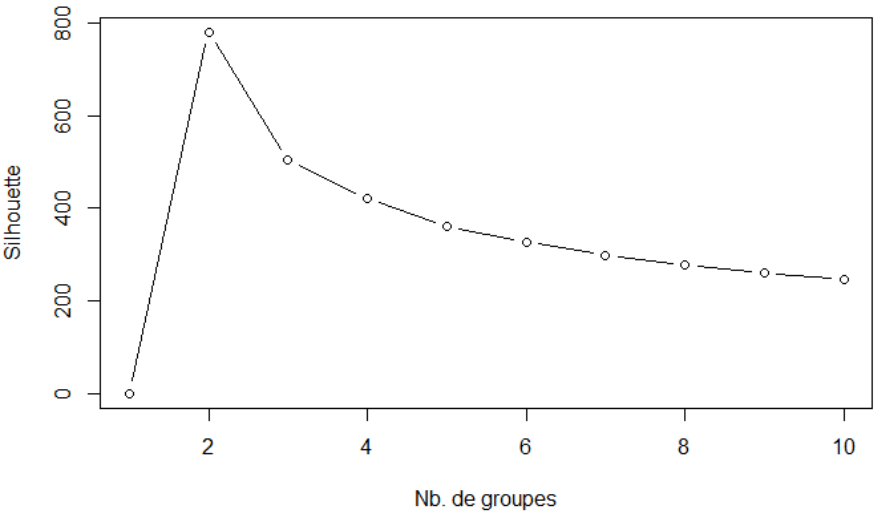


Accuracy = 0.9911 / Recall = 0,9867 /
F-score = 0,9867 / Precision = 0,9867

AUC = 0.997
Les billets sont considérés comme faux à partir d'un score de 0,76

OVERSAMPLING

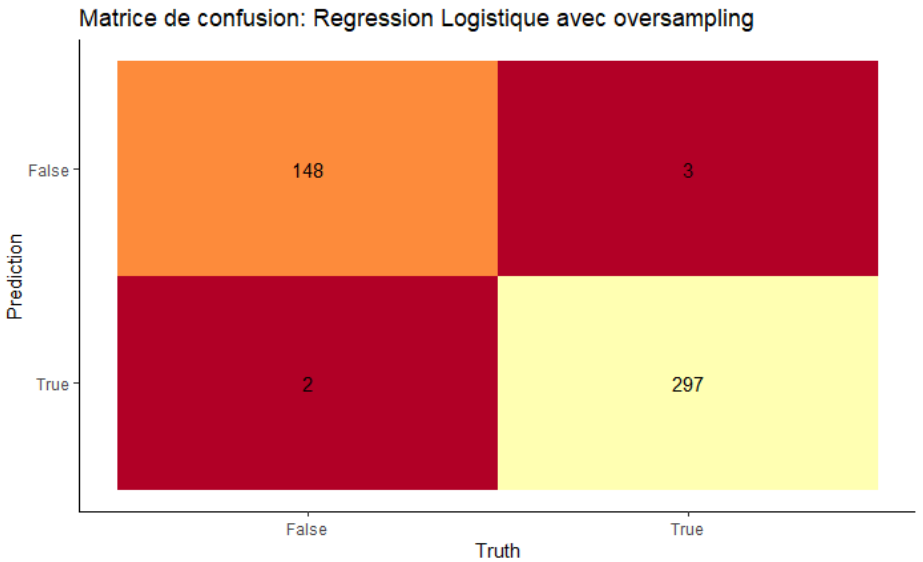


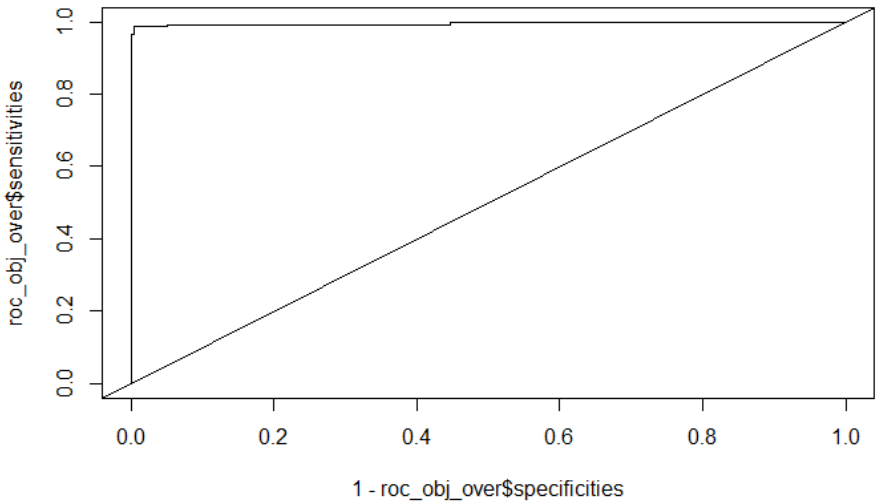


Accuracy = 0.94 / Recall = 0,99 / F-score = 0,92 / Precision = 0,86

OVERSAMPLING

OVERSAMPLING

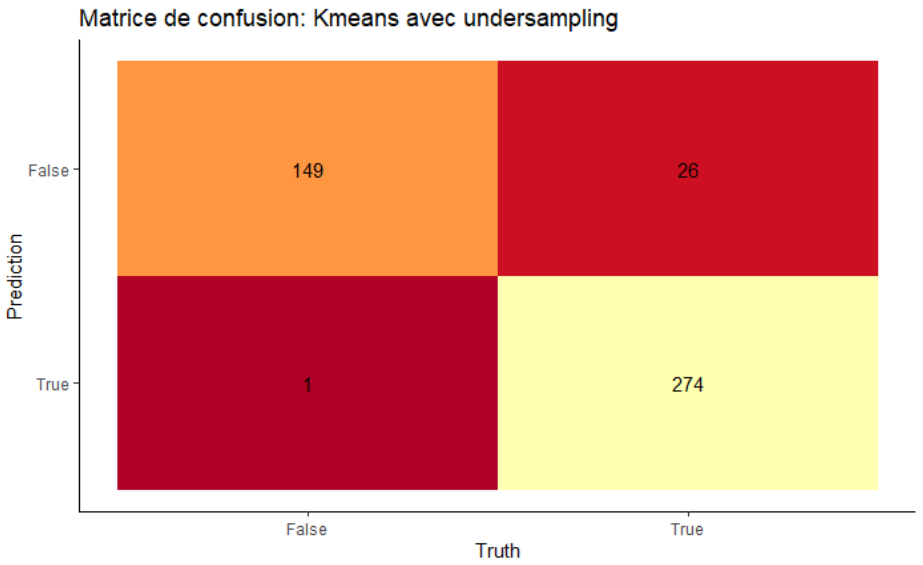


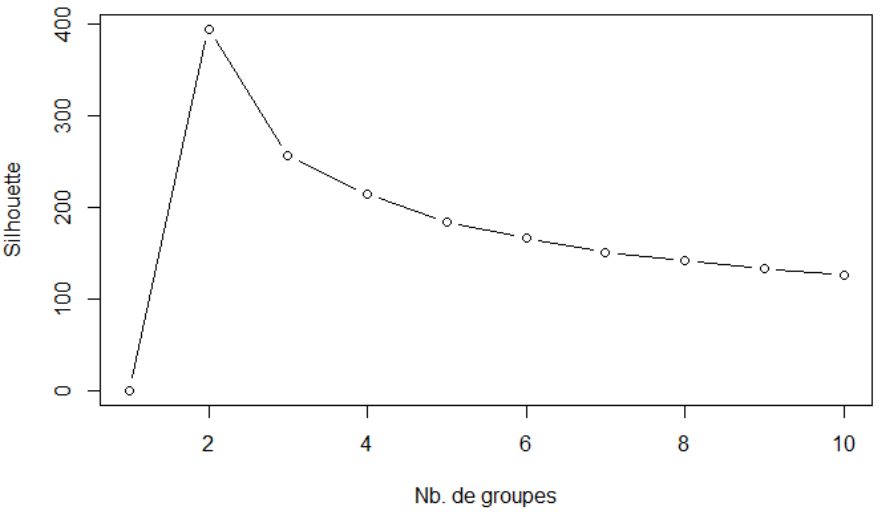


Accuracy = 0.99 / Recall = 0,99 / F-score = 0,98 / Precision = 98

AUC = 0.996
Les billets sont considérés comme faux à partir d'un score de 0,57

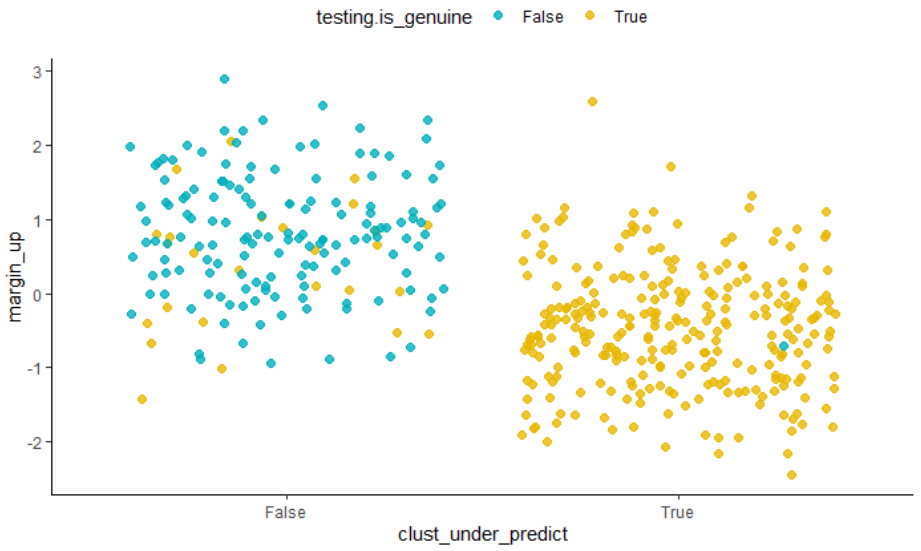
UNDERSAMPLING



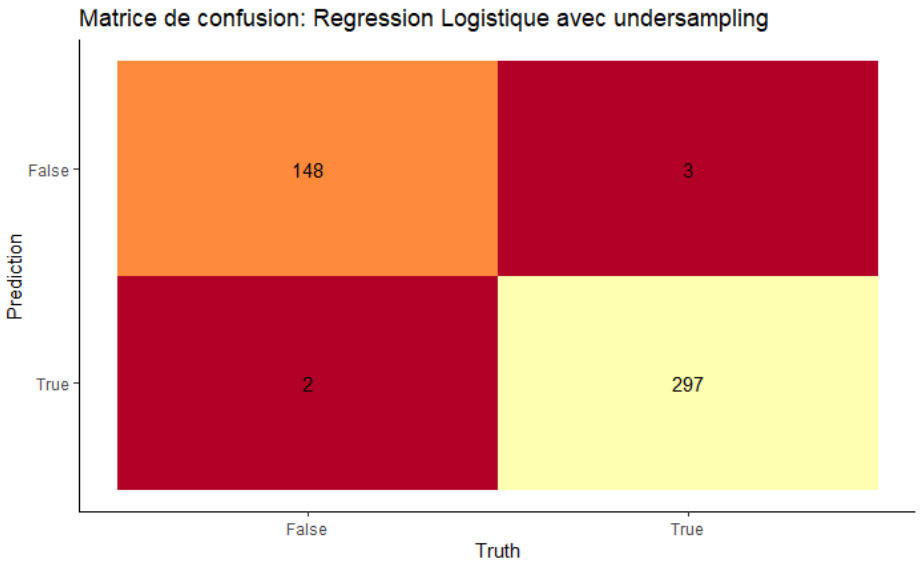


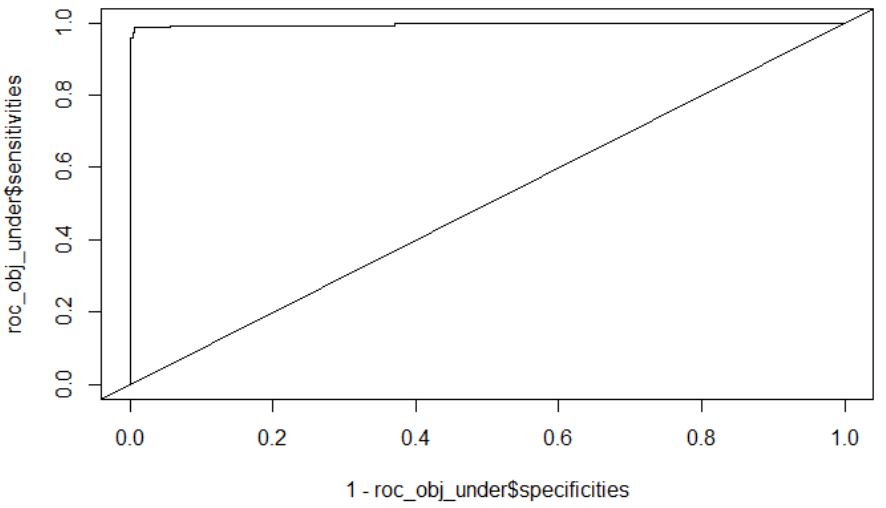
Accuracy = 0.94 / Recall = 0,99 / F-score = 0,92 / Precision = 0,85

UNDERSAMPLING



UNDERSAMPLING





Accuracy = 0.99 / Recall = 0,99 / F-score = 0,98 / Precision = 0,98

AUC = 0.997
Les billets sont considérés comme faux à partir d'un score de 0,77

CONCLUSION

- *Distribution des billets : des faux bien différents*
- *Algorithmes prédictifs : avantage à la régression logistique*