First find the subset of the data with the same size of one of the folds in five-fold cross validation that has the lowest Shapiro – Wilk test p-value and is closest to a normal distribution:

Results are the following 35 subject IDs =

|  |  |  |
| --- | --- | --- |
| **11** | 11 | 16.57 |
| **14** | 14 | 16.20 |
| **25** | 25 | 15.27 |
| **56** | 56 | 20.53 |
| **61** | 61 | 13.57 |
| **70** | 70 | 14.67 |
| **72** | 72 | 14.67 |
| **74** | 74 | 14.70 |
| **77** | 77 | 14.43 |
| **78** | 78 | 14.47 |
| **80** | 80 | 14.47 |
| **81** | 81 | 14.43 |
| **85** | 85 | 14.50 |
| **87** | 87 | 15.00 |
| **90** | 90 | 14.77 |
| **91** | 91 | 14.77 |
| **98** | 98 | 14.97 |
| **102** | 102 | 16.07 |
| **106** | 106 | 19.97 |
| **108** | 108 | 15.50 |
| **109** | 109 | 13.17 |
| **124** | 124 | 14.07 |
| **125** | 125 | 15.03 |
| **139** | 139 | 14.33 |
| **143** | 143 | 13.77 |
| **145** | 145 | 14.20 |
| **148** | 148 | 14.33 |
| **149** | 149 | 15.03 |
| **150** | 150 | 14.50 |
| **151** | 151 | 14.50 |
| **152** | 152 | 14.50 |
| **160** | 160 | 14.63 |
| **161** | 161 | 14.33 |
| **168** | 168 | 15.87 |
| **171** | 171 | 19.97 |

The QQ-plot, the Shapiro-Wilk test, and the histogram of such a holdout test shows that the age distribution of this hold-out test is close to normal.

A picture containing text, diagram, screenshot, plot

Description automatically generatedA picture containing text, line, screenshot, diagram

Description automatically generated

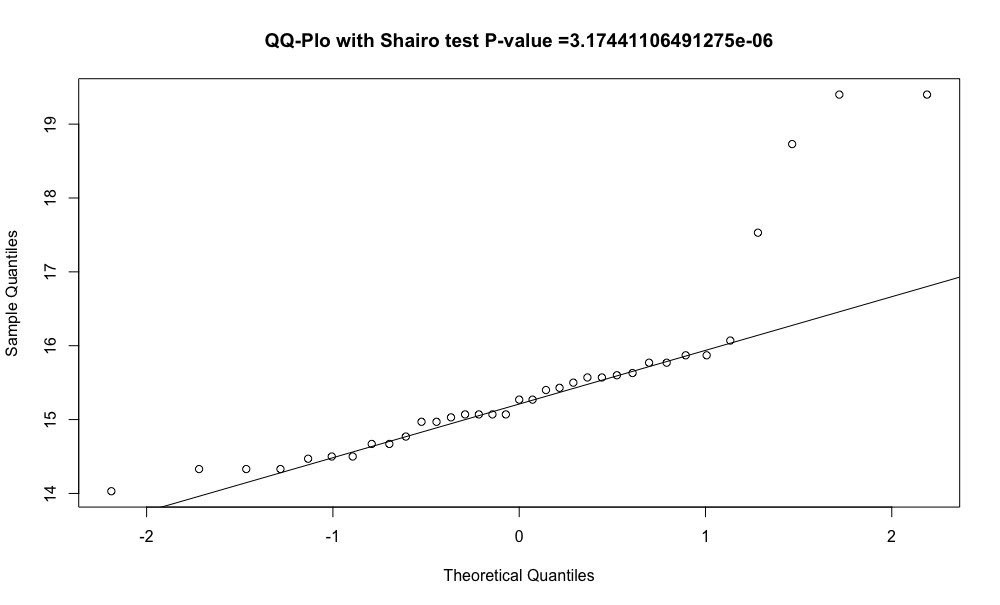
Now from the remaining data we find the best validation set with a similar procedure

|  |  |  |
| --- | --- | --- |
| **9** | 9 | 15.57 |
| **10** | 10 | 15.57 |
| **20** | 20 | 17.53 |
| **26** | 26 | 15.27 |
| **29** | 29 | 15.27 |
| **34** | 34 | 19.40 |
| **36** | 36 | 18.73 |
| **38** | 38 | 19.40 |
| **65** | 65 | 15.40 |
| **66** | 66 | 14.77 |
| **69** | 69 | 14.03 |
| **71** | 71 | 14.67 |
| **73** | 73 | 14.67 |
| **79** | 79 | 14.47 |
| **82** | 82 | 14.50 |
| **84** | 84 | 14.50 |
| **88** | 88 | 15.03 |
| **92** | 92 | 14.97 |
| **96** | 96 | 14.97 |
| **100** | 100 | 15.50 |
| **104** | 104 | 15.87 |
| **114** | 114 | 15.87 |
| **115** | 115 | 15.60 |
| **122** | 122 | 16.07 |
| **131** | 131 | 15.63 |
| **132** | 132 | 15.77 |
| **133** | 133 | 15.77 |
| **135** | 135 | 15.07 |
| **136** | 136 | 15.07 |
| **137** | 137 | 15.07 |
| **138** | 138 | 15.07 |
| **146** | 146 | 15.43 |
| **159** | 159 | 14.33 |
| **163** | 163 | 14.33 |
| **167** | 167 | 14.3 |

The QQ-plot, the Shapiro-Wilk test, and the histogram of such a validation test shows that the age distribution of this validation test is close to normal.

A picture containing text, diagram, screenshot, rectangle

Description automatically generated



They both seem to be sampled from a normal distribution. Because normal distribution has two parameters (mu and sigma) we can test if the hold-out test and validation have different means and variances by a t-test and F-test because we tested and they are significantly sampled from a normal distributions

Welch Two Sample t-test

data: sample\_validation$AGE\_MONTHS and sample\_holdout$AGE\_MONTHS

t = 0.92661, df = 64.051, p-value = 0.3576

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.3877293 1.0585864

sample estimates:

mean of x mean of y

15.52857 15.19314

Therefore we fail to reject that the means are different!

F test to compare two variances

data: sample\_validation$AGE\_MONTHS and sample\_holdout$AGE\_MONTHS

F = 0.60217, num df = 34, denom df = 34, p-value = 0.1442

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.3039522 1.1929618

sample estimates:

ratio of variances

0.6021656

Therefore we fail to reject the variances are different.

And the following non-paramteric Kolmogorov Smirnov test that does not assume the normality of the two sets fails to reject that these two data are from the same distribution.

Exact two-sample Kolmogorov-Smirnov test

data: sample\_validation$AGE\_MONTHS and sample\_holdout$AGE\_MONTHS

D = 0.37143, p-value = 0.01121

alternative hypothesis: two-sided