**Lab #1 – Regression Analysis**

> install.packages("MASS")

Installing package into ‘C:/Users/Dor Meir/Documents/R/win-library/3.5’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/MASS\_7.3-51.3.zip'

Content type 'application/zip' length 1171555 bytes (1.1 MB)

downloaded 1.1 MB

package ‘MASS’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\Dor Meir\AppData\Local\Temp\Rtmpodn1Cj\downloaded\_packages

Warning message:

In dir.create(tempPath, recursive = TRUE) :

cannot create dir 'C:\Users\Dor Meir\Google Drive\????? ???????? ???????', reason 'Invalid argument'

> library(MASS)

> dat<-Boston

> ?Boston

The Boston data frame has 506 rows and 14 columns.

> summary(Boston)

crim zn indus chas

Min. : 0.00632 Min. : 0.00 Min. : 0.46 Min. :0.00000

1st Qu.: 0.08204 1st Qu.: 0.00 1st Qu.: 5.19 1st Qu.:0.00000

Median : 0.25651 Median : 0.00 Median : 9.69 Median :0.00000

Mean : 3.61352 Mean : 11.36 Mean :11.14 Mean :0.06917

3rd Qu.: 3.67708 3rd Qu.: 12.50 3rd Qu.:18.10 3rd Qu.:0.00000

Max. :88.97620 Max. :100.00 Max. :27.74 Max. :1.00000

nox rm age dis

Min. :0.3850 Min. :3.561 Min. : 2.90 Min. : 1.130

1st Qu.:0.4490 1st Qu.:5.886 1st Qu.: 45.02 1st Qu.: 2.100

Median :0.5380 Median :6.208 Median : 77.50 Median : 3.207

Mean :0.5547 Mean :6.285 Mean : 68.57 Mean : 3.795

3rd Qu.:0.6240 3rd Qu.:6.623 3rd Qu.: 94.08 3rd Qu.: 5.188

Max. :0.8710 Max. :8.780 Max. :100.00 Max. :12.127

rad tax ptratio black

Min. : 1.000 Min. :187.0 Min. :12.60 Min. : 0.32

1st Qu.: 4.000 1st Qu.:279.0 1st Qu.:17.40 1st Qu.:375.38

Median : 5.000 Median :330.0 Median :19.05 Median :391.44

Mean : 9.549 Mean :408.2 Mean :18.46 Mean :356.67

3rd Qu.:24.000 3rd Qu.:666.0 3rd Qu.:20.20 3rd Qu.:396.23

Max. :24.000 Max. :711.0 Max. :22.00 Max. :396.90

lstat medv

Min. : 1.73 Min. : 5.00

1st Qu.: 6.95 1st Qu.:17.02

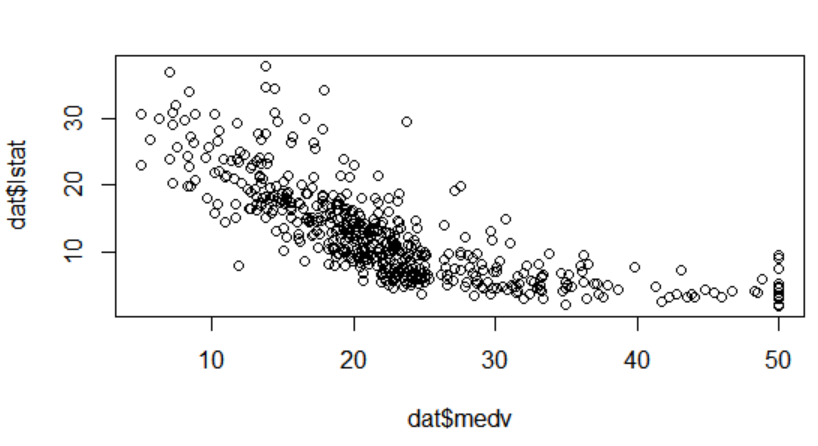
Median :11.36 Median :21.20

Mean :12.65 Mean :22.53

3rd Qu.:16.95 3rd Qu.:25.00

Max. :37.97 Max. :50.00

> plot(dat$medv,dat$lstat)

a. The relationship between the variables is negative, possibly squared, meaning the highest median house prices (medv) correlated with low percent of lower status of population (lstat), and as mdev is decreasing, lstat is increasing.

> fit<-lm(dat$medv~dat$lstat)

> summary(fit)

Call:

lm(formula = dat$medv ~ dat$lstat)

Residuals:

Min 1Q Median 3Q Max

-15.168 -3.990 -1.318 2.034 24.500

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 34.55384 0.56263 61.41 <2e-16 \*\*\*

dat$lstat -0.95005 0.03873 -24.53 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

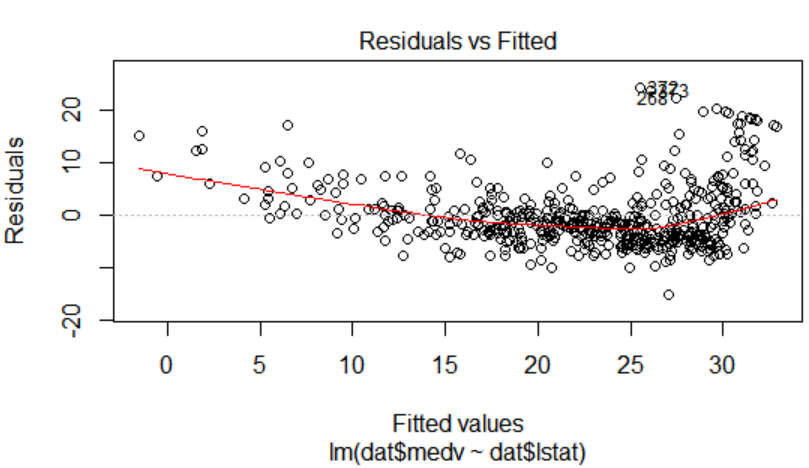
Residual standard error: 6.216 on 504 degrees of freedom

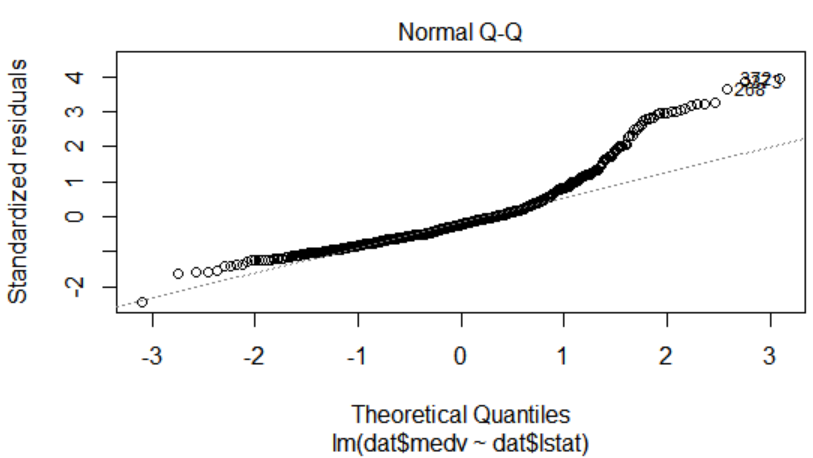
Multiple R-squared: 0.5441, Adjusted R-squared: 0.5432

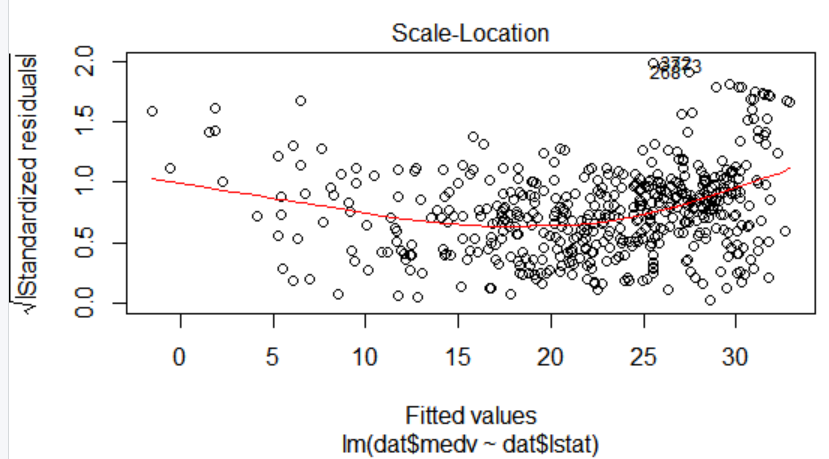
F-statistic: 601.6 on 1 and 504 DF, p-value: < 2.2e-16

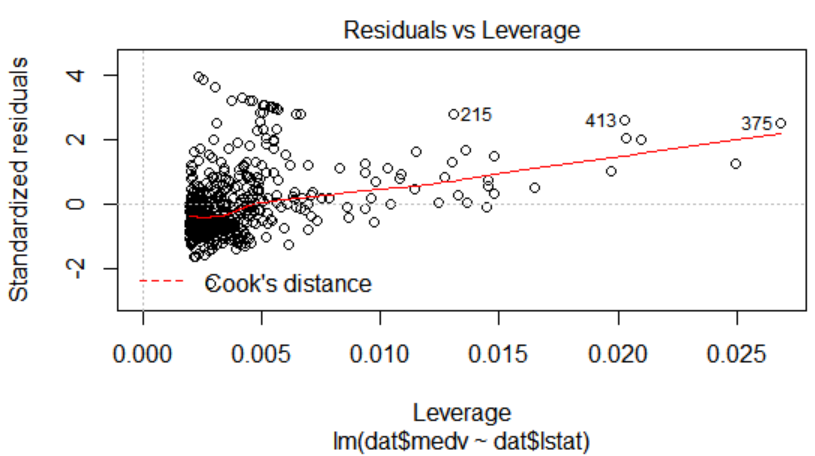
c. The lstat regression coefficient is -0.95005, which means a one percent increase of lower status of population is correlated with 950.05$ decrease in the median value of owner-occupied home, and this coefficient is statistically significant at the level of 0.1%. The intercept coefficient is 34,554$ of the median value of owner-occupied home, and it is also statistically significant at the level of 0.1%. The Multiple and Adjusted R-squared are about 54.5%, which means the model explains 54.5% of the variance in the median value of owner-occupied home. The reported Residuals appear to be not strongly symmetrical, meaning the model predicts certain point that fall far away from the actual observed points.

> plot(fit)









> fit$fitted.values

1 2 3 4 5 6 7

29.8225951 25.8703898 30.7251420 31.7606958 29.4900778 29.6040837 22.7447274

8 9 10 11 12 13 14

16.3603958 6.1188637 18.3079969 15.1253316 21.9466860 19.6285655 26.7064332

15 16 17 18 19 20 21

24.8063345 26.5069229 28.3025161 20.6166169 23.4477639 23.8372842 14.5838035

22 23 24 25 26 27 28

21.4146583 16.7689170 15.6668597 19.0680364 18.8685260 20.4836100 18.1369880

29 30 31 32 33 34 35

22.3932092 23.1722496 13.0827255 22.1651973 8.2279733 17.1204352 15.2298370

36 37 38 39 40 41 42

25.3573631 23.7137778 26.2219080 24.9298409 30.4496277 32.6727432 29.9556020

43 44 45 46 47 48 49

29.0340541 27.4854737 25.4808696 24.8538370 21.1106425 16.6929130 5.2828203

50 51 52 53 54 55 56

19.1630413 21.7756771 25.5948755 29.5375803 26.5449248 20.4931104 29.9841035

57 58 59 60 61 62 63

29.0720561 30.8011459 28.0365023 25.7943858 22.0606919 20.8351282 28.1600087

64 65 66 67 68 69 70

25.5283720 26.9059436 30.1171104 24.8253355 26.8584411 22.1176948 26.2029071

71 72 73 74 75 76 77

28.1695092 25.1673533 29.3095684 27.3904688 28.1125063 26.0603997 23.1817501

78 79 80 81 82 83 84

24.7968340 22.8302319 25.9083918 29.5280798 27.6944845 28.1695092 27.4189702

85 86 87 88 89 90 91

25.4143661 28.3500186 22.3362062 26.5354243 29.3285694 29.1385596 26.1839061

92 93 94 95 96 97 98

26.7634362 26.8014382 28.6540344 24.4928182 28.2360127 23.7802812 30.5541331

99 100 101 102 103 104 105

31.1621647 28.6730354 25.6043760 27.2669623 24.4548162 21.7851776 22.8397323

106 107 108 109 110 111 112

18.9065280 16.8259199 21.1676455 22.8967353 19.7805734 22.2031993 24.9013394

113 114 115 116 117 118 119

19.1535409 18.3174974 24.6258251 19.5810631 23.1152467 24.7683325 19.9515823

120 121 122 123 124 125 126

21.6236692 20.9016317 20.9966366 17.5194560 10.4130868 17.8519732 20.4836100

127 128 129 130 131 132 133

8.6554955 18.2224925 19.9325813 17.1299357 22.5832190 22.9062358 23.9892921

134 135 136 137 138 139 140

20.2745991 18.1084866 18.4410038 18.4980068 20.6926208 14.2987887 17.0159298

141 142 143 144 145 146 147

11.6006485 1.8626426 9.0735172 9.4535370 6.7268953 8.1424688 18.7355191

148 149 150 151 152 153 154

6.4988835 7.6484432 14.1752822 21.1581450 21.9371855 23.0392427 19.5525616

155 156 157 158 159 160 161

20.1890947 20.2840996 19.2200443 30.1931143 28.4450235 27.5329762 29.3285694

162 163 164 165 166 167 168

32.9102555 32.7297461 31.3996770 23.4952664 25.2338567 31.0386583 23.0202417

169 170 171 172 173 174 175

24.0082931 23.7992822 20.8446287 23.1247472 20.5976159 25.9653947 25.3953651

176 177 178 179 180 181 182

29.4900778 24.9488419 28.5780304 27.9794994 29.7655921 27.3714678 25.5758745

183 184 185 186 187 188 189

29.9746030 29.1575606 21.2721509 22.0606919 30.3261213 28.2075112 30.2216158

190 191 192 193 194 195 196

29.4330749 29.7085892 30.0981094 31.8271992 29.7750926 30.3926247 31.7321943

197 198 199 200 201 202 203

30.6776395 26.3739159 28.2645142 30.2216158 30.3261213 27.4949742 31.5991874

204 205 206 207 208 209 210

30.9341528 31.8176987 24.2268044 24.1317995 17.3959496 20.6261174 12.6172013

211 212 213 214 215 216 217

18.1464885 11.7716574 19.3245497 25.6423779 6.4798825 25.5568735 21.7186741

218 219 220 221 222 223 224

25.3478626 17.5289565 24.5783227 25.3288617 14.1657817 25.1198508 27.3334658

225 226 227 228 229 230 231

30.6206366 30.1551124 31.5801864 28.5115270 30.8296474 30.9816553 23.4857659

232 233 234 235 236 237 238

29.5660818 32.2072190 30.8011459 26.9059436 24.2173039 25.4903700 30.0601074

239 240 241 242 243 244 245

28.5115270 27.5519771 23.7422792 22.7732289 23.8942871 29.6230847 22.6782240

246 247 248 249 250 251 252

17.0159298 25.8513888 24.9108399 25.5093710 28.3215171 28.9485497 31.1431637

253 254 255 256 257 258 259

31.2001667 31.1906662 28.3120166 25.7658844 31.5991874 29.6895882 27.1529564

260 261 262 263 264 265 266

27.9985003 25.4428676 27.6564826 28.9390492 23.8657856 26.8584411 24.6258251

267 268 269 270 271 272 273

20.5026109 27.4854737 31.5516849 21.5856672 22.2031993 28.2930156 27.2099594

274 275 276 277 278 279 280

28.3025161 31.2001667 31.7226938 28.8060423 30.6016356 27.7229860 29.9461015

281 282 283 284 285 286 287

30.9816553 30.1931143 31.6941923 31.5516849 27.0959535 26.7349347 22.2697027

288 289 290 291 292 293 294

27.7704885 27.3334658 25.5188715 31.3901765 31.1716652 30.0886089 26.4024174

295 296 297 298 299 300 301

24.6733276 28.5970314 27.5329762 19.5050591 29.8320956 30.0506069 28.7870413

302 303 304 305 306 307 308

25.5283720 26.3169130 29.9366010 27.9699989 26.0699002 28.4070216 27.3999692

309 310 311 312 313 314 315

30.2406168 25.0818488 22.5452170 28.8725457 23.4192625 27.0484510 25.7373829

316 317 318 319 320 321 322

23.6282733 17.1394362 19.4100542 24.7113296 22.4597126 27.7134855 28.0270018

323 324 325 326 327 328 329

27.2384609 23.4002615 28.7395388 29.7275902 28.7110374 22.4027096 25.0818488

330 331 332 333 334 335 336

27.5804786 25.9178923 22.7447274 27.1149544 29.1575606 28.1410077 26.9439456

337 338 339 340 341 342 343

25.2433572 24.5213197 26.4689209 25.3003602 25.7278824 29.3380699 26.3359140

344 345 346 347 348 349 350

27.7324865 30.1741134 24.5498212 22.5167156 28.5115270 28.8630453 28.9580502

351 352 353 354 355 356 357

28.8725457 29.3380699 27.1529564 30.2786188 26.9059436 29.2620660 17.8329723

358 359 360 361 362 363 364

21.9466860 23.6472743 22.5167156 27.1529564 21.0726405 24.8728380 20.6451183

365 366 367 368 369 370 371

29.5280798 27.7894895 21.2531499 21.8896830 31.4566800 31.0101568 31.7416948

372 373 374 375 376 377 378

25.4998705 26.1174026 1.5206248 -1.5195331 21.7851776 12.4746939 14.3747926

379 380 381 382 383 384 385

12.0471717 13.8617660 18.2034915 14.5268005 12.1326761 11.2206288 5.4538292

386 387 388 389 390 391 392

5.2828203 7.6864452 4.1617621 5.4633297 14.7453119 18.2984964 16.7309150

393 394 395 396 397 398 399

10.1565735 20.1415922 19.0205339 18.2889959 16.1513849 15.6288578 5.4918311

400 401 402 403 404 405 406

6.0808617 9.1210197 15.2488380 15.2583385 15.7713652 8.5414896 12.7217067

407 408 409 410 411 412 413

12.3796890 23.0297422 9.4725379 15.7618647 24.9488419 14.3937936 1.9006446

414 415 416 417 418 419 420

15.4768499 -0.5789842 6.9549072 10.0520680 9.2445261 14.9638232 12.9497186

421 422 423 424 425 426 427

20.2840996 19.6380660 21.1581450 12.4271914 18.2509940 11.3821371 19.6475665

428 429 430 431 432 433 434

20.7591243 14.1087788 11.6766524 17.7949703 15.8473691 23.1247472 19.1440404

435 436 437 438 439 440 441

20.1415922 12.4461924 17.4054500 9.4250355 2.2331619 12.8167117 13.5482497

442 443 444 445 446 447 448

16.0088775 18.7925221 16.6454106 11.9521668 11.7716574 17.6524629 18.9350295

449 450 451 452 453 454 455

17.3294461 16.2083879 17.9849801 17.7094658 18.1464885 18.6500147 16.7784175

456 457 458 459 460 461 462

17.3294461 16.4934027 18.4600048 19.1345399 20.5881154 18.9540305 20.6356178

463 464 465 466 467 468 469

21.2626504 24.7778330 21.9941884 21.1296435 18.2604945 14.2987887 17.3294461

470 471 472 473 474 475 476

20.5311124 19.0775369 22.3267057 20.9111322 23.4762654 17.3199456 11.6576515

477 478 479 480 481 482 483

16.8069190 10.8881115 17.4244510 22.0986939 24.3503108 27.2004589 27.8939949

484 485 486 487 488 489 490

24.6543266 21.8801825 24.5023187 20.3221016 23.6757758 17.3959496 11.7811579

491 492 493 494 495 496 497

6.3563761 17.3864491 21.8706820 23.1437481 21.6426702 17.8329723 14.4697975

498 499 500 501 502 503 504

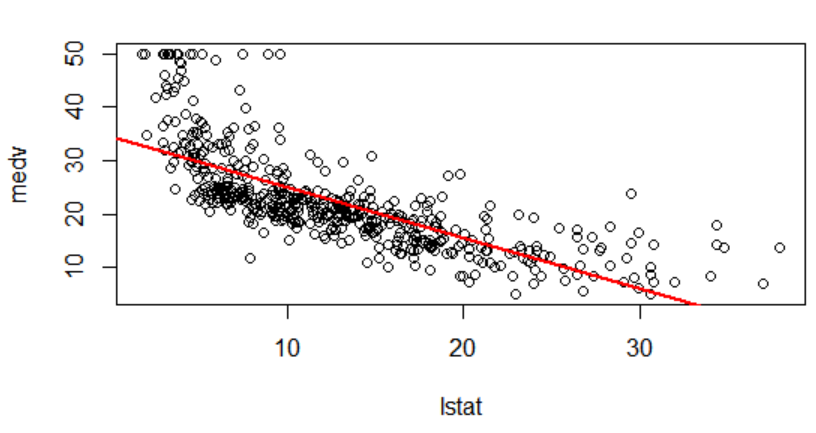
21.1581450 22.2792032 20.2080956 20.9396336 25.3668636 25.9273927 29.1955625

505 506

28.3975211 27.0674520

> plot(medv~lstat)

> abline(fit, col="red", lwd=2)



f. This regression does not seem to give a good fit, since it the observations with lstat values under 5 and over 35 are not fitted well by the regression. It seems as if the Data Generating Proccess was quadratic, and so this fitted linear model has high bias (but low variance).

**Lab #2 – KNN Analysis**

> install.packages("ISLR")

Installing package into ‘C:/Users/Dor Meir/Documents/R/win-library/3.5’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/ISLR\_1.2.zip'

Content type 'application/zip' length 2923754 bytes (2.8 MB)

downloaded 2.8 MB

package ‘ISLR’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\Dor Meir\AppData\Local\Temp\RtmpYXpUiI\downloaded\_packages

Warning message:

In dir.create(tempPath, recursive = TRUE) :

cannot create dir 'C:\Users\Dor Meir\Google Drive\????? ???????? ???????', reason 'Invalid argument'

> install.packages("class")

Installing package into ‘C:/Users/Dor Meir/Documents/R/win-library/3.5’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/class\_7.3-15.zip'

Content type 'application/zip' length 106333 bytes (103 KB)

downloaded 103 KB

package ‘class’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\Dor Meir\AppData\Local\Temp\RtmpYXpUiI\downloaded\_packages

> library("ISLR","class")

?Default

A data frame with 10000 observations on the following 4 variables.

> summary(Default)

default student balance income

No :9667 No :7056 Min. : 0.0 Min. : 772

Yes: 333 Yes:2944 1st Qu.: 481.7 1st Qu.:21340

Median : 823.6 Median :34553

Mean : 835.4 Mean :33517

3rd Qu.:1166.3 3rd Qu.:43808

Max. :2654.3 Max. :73554

> dat <-Default[c("balance","income","default")]

> dat$balance <- as.vector(scale(dat$balance))

> dat$income <- as.vector(scale(dat$income))

> index <- sample(x=1:nrow(dat), size=.3\*nrow(dat))

> test <- dat[index,]

> train <- dat[-index,]

> test\_pred <- test[c("income","balance")]

> train\_pred <- train[c("income","balance")]

> test\_default <- test$default

> train\_default <- train$default

> knn.1 <- knn(train = train\_pred, test = test\_pred, cl = train\_default, k = 1)

> knn.5 <- knn(train = train\_pred, test = test\_pred, cl = train\_default, k = 5)

> knn.20 <- knn(train = train\_pred, test = test\_pred, cl = train\_default, k = 20)

> knn.70 <- knn(train = train\_pred, test = test\_pred, cl = train\_default, k = 70)

> table(knn.1, test\_default)

test\_default

knn.1 No Yes

No 2839 64

Yes 64 33

> table(knn.5, test\_default)

test\_default

knn.5 No Yes

No 2888 65

Yes 15 32

> table(knn.20, test\_default)

test\_default

knn.20 No Yes

No 2892 67

Yes 11 30

> table(knn.70, test\_default)

test\_default

knn.70 No Yes

No 2896 78

Yes 7 19

> prop.table(table(knn.1, test\_default),2)

test\_default

knn.1 No Yes

No 0.97795384 0.65979381

Yes 0.02204616 0.34020619

> prop.table(table(knn.5, test\_default),2)

test\_default

knn.5 No Yes

No 0.994832931 0.670103093

Yes 0.005167069 0.329896907

> prop.table(table(knn.20, test\_default),2)

test\_default

knn.20 No Yes

No 0.996210816 0.690721649

Yes 0.003789184 0.309278351

> prop.table(table(knn.70, test\_default),2)

test\_default

knn.70 No Yes

No 0.997588701 0.804123711

Yes 0.002411299 0.195876289