

---

# Student Performance Analysis

— Richard Hou, Ahnaf Ayub,  
Shihao Tong, Siqi Wang —

---

---

---

# Problem Context & Exploratory Data Analysis

---

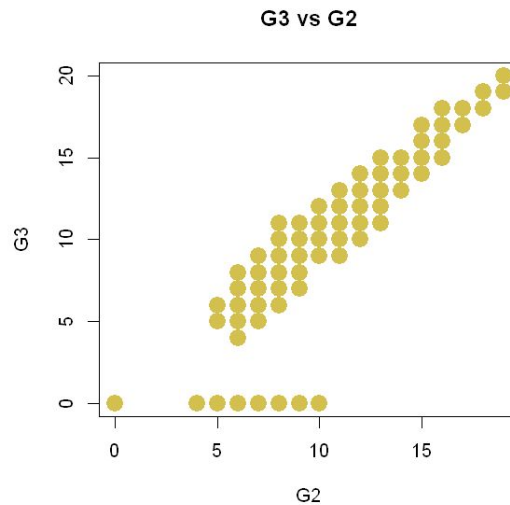
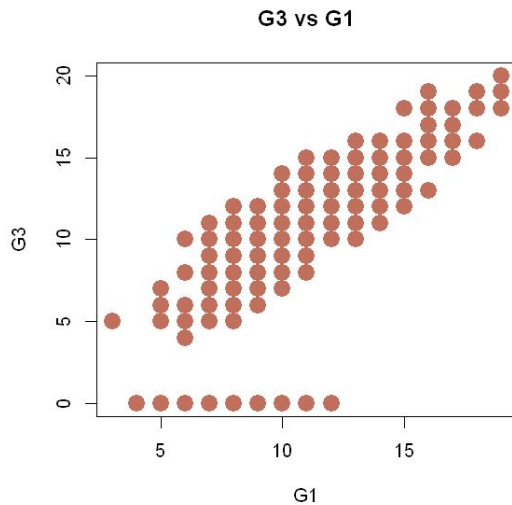
---

# Problem Context

- Observational Study- Surveys and Questionnaires were used
- Null Hypothesis: All Regression Coefficients are Zero

# Problem Context

Can we predict a student's **Final Course Grade** with Linear Regression?



# The Data

Response Variable: G3 (ie. Final Grade)

Explanatory Variables:

- First/Second period grades (G1/G2)
- Age of Student
- Sex(F/M)
- Time spent studying(Weekly)
- Daily commute time

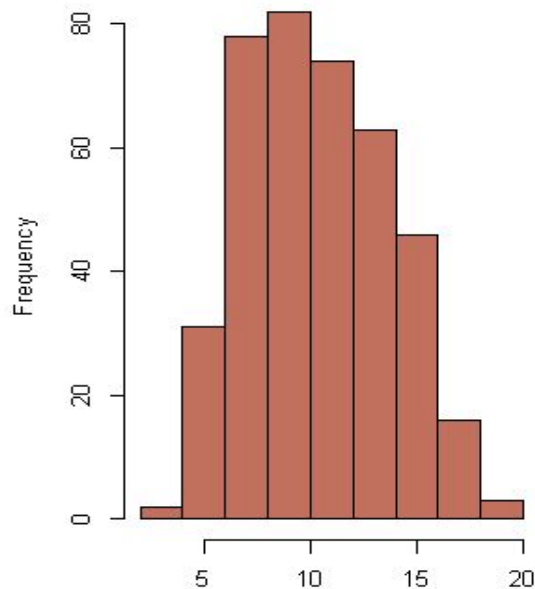
... (and 26 others)

## Attribute Information:

# Attributes for both student-mat.csv (Math course) and student-1 school - student's school (binary: 'GP' - Gabriel Pereira or 'MS'  
2 sex - student's sex (binary: 'F' - female or 'M' - male)  
3 age - student's age (numeric: from 15 to 22)  
4 address - student's home address type (binary: 'U' - urban or 'R' - rural)  
5 famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)  
6 Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - alone)  
7 Medu - mother's education (numeric: 0 - none, 1 - primary education, 2 - secondary education, 3 - tertiary education)  
8 Fedu - father's education (numeric: 0 - none, 1 - primary education, 2 - secondary education, 3 - tertiary education)  
9 Mjob - mother's job (nominal: 'teacher', 'health' care related, 'other')  
10 Fjob - father's job (nominal: 'teacher', 'health' care related, 'other')  
11 reason - reason to choose this school (nominal: close to 'home', 'school', 'other')  
12 guardian - student's guardian (nominal: 'mother', 'father' or 'other')  
13 traveltime - home to school travel time (numeric: 1 - <15 min, 2 - 15 to 30 min, 3 - 30 to 45 min, 4 - 45 to 60 min, 5 - 60 to 90 min, 6 - 90 to 120 min, 7 - 120 to 150 min, 8 - 150 to 180 min, 9 - 180 to 210 min, 10 - 210 to 240 min, 11 - 240 to 270 min, 12 - 270 to 300 min, 13 - 300 to 330 min, 14 - 330 to 360 min, 15 - 360 to 390 min, 16 - 390 to 420 min, 17 - 420 to 450 min, 18 - 450 to 480 min, 19 - 480 to 510 min, 20 - 510 to 540 min, 21 - 540 to 570 min, 22 - 570 to 600 min, 23 - 600 to 630 min, 24 - 630 to 660 min, 25 - 660 to 690 min, 26 - 690 to 720 min, 27 - 720 to 750 min, 28 - 750 to 780 min, 29 - 780 to 810 min, 30 - 810 to 840 min, 31 - 840 to 870 min, 32 - 870 to 900 min, 33 - 900 to 930 min, 34 - 930 to 960 min, 35 - 960 to 990 min, 36 - 990 to 1020 min, 37 - 1020 to 1050 min, 38 - 1050 to 1080 min, 39 - 1080 to 1110 min, 40 - 1110 to 1140 min, 41 - 1140 to 1170 min, 42 - 1170 to 1200 min, 43 - 1200 to 1230 min, 44 - 1230 to 1260 min, 45 - 1260 to 1290 min, 46 - 1290 to 1320 min, 47 - 1320 to 1350 min, 48 - 1350 to 1380 min, 49 - 1380 to 1410 min, 50 - 1410 to 1440 min, 51 - 1440 to 1470 min, 52 - 1470 to 1500 min, 53 - 1500 to 1530 min, 54 - 1530 to 1560 min, 55 - 1560 to 1590 min, 56 - 1590 to 1620 min, 57 - 1620 to 1650 min, 58 - 1650 to 1680 min, 59 - 1680 to 1710 min, 60 - 1710 to 1740 min, 61 - 1740 to 1770 min, 62 - 1770 to 1800 min, 63 - 1800 to 1830 min, 64 - 1830 to 1860 min, 65 - 1860 to 1890 min, 66 - 1890 to 1920 min, 67 - 1920 to 1950 min, 68 - 1950 to 1980 min, 69 - 1980 to 2010 min, 70 - 2010 to 2040 min, 71 - 2040 to 2070 min, 72 - 2070 to 2100 min, 73 - 2100 to 2130 min, 74 - 2130 to 2160 min, 75 - 2160 to 2190 min, 76 - 2190 to 2220 min, 77 - 2220 to 2250 min, 78 - 2250 to 2280 min, 79 - 2280 to 2310 min, 80 - 2310 to 2340 min, 81 - 2340 to 2370 min, 82 - 2370 to 2400 min, 83 - 2400 to 2430 min, 84 - 2430 to 2460 min, 85 - 2460 to 2490 min, 86 - 2490 to 2520 min, 87 - 2520 to 2550 min, 88 - 2550 to 2580 min, 89 - 2580 to 2610 min, 90 - 2610 to 2640 min, 91 - 2640 to 2670 min, 92 - 2670 to 2700 min, 93 - 2700 to 2730 min, 94 - 2730 to 2760 min, 95 - 2760 to 2790 min, 96 - 2790 to 2820 min, 97 - 2820 to 2850 min, 98 - 2850 to 2880 min, 99 - 2880 to 2910 min, 100 - 2910 to 2940 min, 101 - 2940 to 2970 min, 102 - 2970 to 3000 min, 103 - 3000 to 3030 min, 104 - 3030 to 3060 min, 105 - 3060 to 3090 min, 106 - 3090 to 3120 min, 107 - 3120 to 3150 min, 108 - 3150 to 3180 min, 109 - 3180 to 3210 min, 110 - 3210 to 3240 min, 111 - 3240 to 3270 min, 112 - 3270 to 3300 min, 113 - 3300 to 3330 min, 114 - 3330 to 3360 min, 115 - 3360 to 3390 min, 116 - 3390 to 3420 min, 117 - 3420 to 3450 min, 118 - 3450 to 3480 min, 119 - 3480 to 3510 min, 120 - 3510 to 3540 min, 121 - 3540 to 3570 min, 122 - 3570 to 3600 min, 123 - 3600 to 3630 min, 124 - 3630 to 3660 min, 125 - 3660 to 3690 min, 126 - 3690 to 3720 min, 127 - 3720 to 3750 min, 128 - 3750 to 3780 min, 129 - 3780 to 3810 min, 130 - 3810 to 3840 min, 131 - 3840 to 3870 min, 132 - 3870 to 3900 min, 133 - 3900 to 3930 min, 134 - 3930 to 3960 min, 135 - 3960 to 3990 min, 136 - 3990 to 4020 min, 137 - 4020 to 4050 min, 138 - 4050 to 4080 min, 139 - 4080 to 4110 min, 140 - 4110 to 4140 min, 141 - 4140 to 4170 min, 142 - 4170 to 4200 min, 143 - 4200 to 4230 min, 144 - 4230 to 4260 min, 145 - 4260 to 4290 min, 146 - 4290 to 4320 min, 147 - 4320 to 4350 min, 148 - 4350 to 4380 min, 149 - 4380 to 4410 min, 150 - 4410 to 4440 min, 151 - 4440 to 4470 min, 152 - 4470 to 4500 min, 153 - 4500 to 4530 min, 154 - 4530 to 4560 min, 155 - 4560 to 4590 min, 156 - 4590 to 4620 min, 157 - 4620 to 4650 min, 158 - 4650 to 4680 min, 159 - 4680 to 4710 min, 160 - 4710 to 4740 min, 161 - 4740 to 4770 min, 162 - 4770 to 4800 min, 163 - 4800 to 4830 min, 164 - 4830 to 4860 min, 165 - 4860 to 4890 min, 166 - 4890 to 4920 min, 167 - 4920 to 4950 min, 168 - 4950 to 4980 min, 169 - 4980 to 5010 min, 170 - 5010 to 5040 min, 171 - 5040 to 5070 min, 172 - 5070 to 5100 min, 173 - 5100 to 5130 min, 174 - 5130 to 5160 min, 175 - 5160 to 5190 min, 176 - 5190 to 5220 min, 177 - 5220 to 5250 min, 178 - 5250 to 5280 min, 179 - 5280 to 5310 min, 180 - 5310 to 5340 min, 181 - 5340 to 5370 min, 182 - 5370 to 5400 min, 183 - 5400 to 5430 min, 184 - 5430 to 5460 min, 185 - 5460 to 5490 min, 186 - 5490 to 5520 min, 187 - 5520 to 5550 min, 188 - 5550 to 5580 min, 189 - 5580 to 5610 min, 190 - 5610 to 5640 min, 191 - 5640 to 5670 min, 192 - 5670 to 5700 min, 193 - 5700 to 5730 min, 194 - 5730 to 5760 min, 195 - 5760 to 5790 min, 196 - 5790 to 5820 min, 197 - 5820 to 5850 min, 198 - 5850 to 5880 min, 199 - 5880 to 5910 min, 200 - 5910 to 5940 min, 201 - 5940 to 5970 min, 202 - 5970 to 6000 min, 203 - 6000 to 6030 min, 204 - 6030 to 6060 min, 205 - 6060 to 6090 min, 206 - 6090 to 6120 min, 207 - 6120 to 6150 min, 208 - 6150 to 6180 min, 209 - 6180 to 6210 min, 210 - 6210 to 6240 min, 211 - 6240 to 6270 min, 212 - 6270 to 6300 min, 213 - 6300 to 6330 min, 214 - 6330 to 6360 min, 215 - 6360 to 6390 min, 216 - 6390 to 6420 min, 217 - 6420 to 6450 min, 218 - 6450 to 6480 min, 219 - 6480 to 6510 min, 220 - 6510 to 6540 min, 221 - 6540 to 6570 min, 222 - 6570 to 6600 min, 223 - 6600 to 6630 min, 224 - 6630 to 6660 min, 225 - 6660 to 6690 min, 226 - 6690 to 6720 min, 227 - 6720 to 6750 min, 228 - 6750 to 6780 min, 229 - 6780 to 6810 min, 230 - 6810 to 6840 min, 231 - 6840 to 6870 min, 232 - 6870 to 6900 min, 233 - 6900 to 6930 min, 234 - 6930 to 6960 min, 235 - 6960 to 6990 min, 236 - 6990 to 7020 min, 237 - 7020 to 7050 min, 238 - 7050 to 7080 min, 239 - 7080 to 7110 min, 240 - 7110 to 7140 min, 241 - 7140 to 7170 min, 242 - 7170 to 7200 min, 243 - 7200 to 7230 min, 244 - 7230 to 7260 min, 245 - 7260 to 7290 min, 246 - 7290 to 7320 min, 247 - 7320 to 7350 min, 248 - 7350 to 7380 min, 249 - 7380 to 7410 min, 250 - 7410 to 7440 min, 251 - 7440 to 7470 min, 252 - 7470 to 7500 min, 253 - 7500 to 7530 min, 254 - 7530 to 7560 min, 255 - 7560 to 7590 min, 256 - 7590 to 7620 min, 257 - 7620 to 7650 min, 258 - 7650 to 7680 min, 259 - 7680 to 7710 min, 260 - 7710 to 7740 min, 261 - 7740 to 7770 min, 262 - 7770 to 7800 min, 263 - 7800 to 7830 min, 264 - 7830 to 7860 min, 265 - 7860 to 7890 min, 266 - 7890 to 7920 min, 267 - 7920 to 7950 min, 268 - 7950 to 7980 min, 269 - 7980 to 8010 min, 270 - 8010 to 8040 min, 271 - 8040 to 8070 min, 272 - 8070 to 8100 min, 273 - 8100 to 8130 min, 274 - 8130 to 8160 min, 275 - 8160 to 8190 min, 276 - 8190 to 8220 min, 277 - 8220 to 8250 min, 278 - 8250 to 8280 min, 279 - 8280 to 8310 min, 280 - 8310 to 8340 min, 281 - 8340 to 8370 min, 282 - 8370 to 8400 min, 283 - 8400 to 8430 min, 284 - 8430 to 8460 min, 285 - 8460 to 8490 min, 286 - 8490 to 8520 min, 287 - 8520 to 8550 min, 288 - 8550 to 8580 min, 289 - 8580 to 8610 min, 290 - 8610 to 8640 min, 291 - 8640 to 8670 min, 292 - 8670 to 8700 min, 293 - 8700 to 8730 min, 294 - 8730 to 8760 min, 295 - 8760 to 8790 min, 296 - 8790 to 8820 min, 297 - 8820 to 8850 min, 298 - 8850 to 8880 min, 299 - 8880 to 8910 min, 300 - 8910 to 8940 min, 301 - 8940 to 8970 min, 302 - 8970 to 9000 min, 303 - 9000 to 9030 min, 304 - 9030 to 9060 min, 305 - 9060 to 9090 min, 306 - 9090 to 9120 min, 307 - 9120 to 9150 min, 308 - 9150 to 9180 min, 309 - 9180 to 9210 min, 310 - 9210 to 9240 min, 311 - 9240 to 9270 min, 312 - 9270 to 9300 min, 313 - 9300 to 9330 min, 314 - 9330 to 9360 min, 315 - 9360 to 9390 min, 316 - 9390 to 9420 min, 317 - 9420 to 9450 min, 318 - 9450 to 9480 min, 319 - 9480 to 9510 min, 320 - 9510 to 9540 min, 321 - 9540 to 9570 min, 322 - 9570 to 9600 min, 323 - 9600 to 9630 min, 324 - 9630 to 9660 min, 325 - 9660 to 9690 min, 326 - 9690 to 9720 min, 327 - 9720 to 9750 min, 328 - 9750 to 9780 min, 329 - 9780 to 9810 min, 330 - 9810 to 9840 min, 331 - 9840 to 9870 min, 332 - 9870 to 9900 min, 333 - 9900 to 9930 min, 334 - 9930 to 9960 min, 335 - 9960 to 9990 min, 336 - 9990 to 10020 min, 337 - 10020 to 10050 min, 338 - 10050 to 10080 min, 339 - 10080 to 10110 min, 340 - 10110 to 10140 min, 341 - 10140 to 10170 min, 342 - 10170 to 10200 min, 343 - 10200 to 10230 min, 344 - 10230 to 10260 min, 345 - 10260 to 10290 min, 346 - 10290 to 10320 min, 347 - 10320 to 10350 min, 348 - 10350 to 10380 min, 349 - 10380 to 10410 min, 350 - 10410 to 10440 min, 351 - 10440 to 10470 min, 352 - 10470 to 10500 min, 353 - 10500 to 10530 min, 354 - 10530 to 10560 min, 355 - 10560 to 10590 min, 356 - 10590 to 10620 min, 357 - 10620 to 10650 min, 358 - 10650 to 10680 min, 359 - 10680 to 10710 min, 360 - 10710 to 10740 min, 361 - 10740 to 10770 min, 362 - 10770 to 10800 min, 363 - 10800 to 10830 min, 364 - 10830 to 10860 min, 365 - 10860 to 10890 min, 366 - 10890 to 10920 min, 367 - 10920 to 10950 min, 368 - 10950 to 10980 min, 369 - 10980 to 11010 min, 370 - 11010 to 11040 min, 371 - 11040 to 11070 min, 372 - 11070 to 11100 min, 373 - 11100 to 11130 min, 374 - 11130 to 11160 min, 375 - 11160 to 11190 min, 376 - 11190 to 11220 min, 377 - 11220 to 11250 min, 378 - 11250 to 11280 min, 379 - 11280 to 11310 min, 380 - 11310 to 11340 min, 381 - 11340 to 11370 min, 382 - 11370 to 11400 min, 383 - 11400 to 11430 min, 384 - 11430 to 11460 min, 385 - 11460 to 11490 min, 386 - 11490 to 11520 min, 387 - 11520 to 11550 min, 388 - 11550 to 11580 min, 389 - 11580 to 11610 min, 390 - 11610 to 11640 min, 391 - 11640 to 11670 min, 392 - 11670 to 11700 min, 393 - 11700 to 11730 min, 394 - 11730 to 11760 min, 395 - 11760 to 11790 min, 396 - 11790 to 11820 min, 397 - 11820 to 11850 min, 398 - 11850 to 11880 min, 399 - 11880 to 11910 min, 400 - 11910 to 11940 min, 401 - 11940 to 11970 min, 402 - 11970 to 12000 min, 403 - 12000 to 12030 min, 404 - 12030 to 12060 min, 405 - 12060 to 12090 min, 406 - 12090 to 12120 min, 407 - 12120 to 12150 min, 408 - 12150 to 12180 min, 409 - 12180 to 12210 min, 410 - 12210 to 12240 min, 411 - 12240 to 12270 min, 412 - 12270 to 12300 min, 413 - 12300 to 12330 min, 414 - 12330 to 12360 min, 415 - 12360 to 12390 min, 416 - 12390 to 12420 min, 417 - 12420 to 12450 min, 418 - 12450 to 12480 min, 419 - 12480 to 12510 min, 420 - 12510 to 12540 min, 421 - 12540 to 12570 min, 422 - 12570 to 12600 min, 423 - 12600 to 12630 min, 424 - 12630 to 12660 min, 425 - 12660 to 12690 min, 426 - 12690 to 12720 min, 427 - 12720 to 12750 min, 428 - 12750 to 12780 min, 429 - 12780 to 12810 min, 430 - 12810 to 12840 min, 431 - 12840 to 12870 min, 432 - 12870 to 12900 min, 433 - 12900 to 12930 min, 434 - 12930 to 12960 min, 435 - 12960 to 12990 min, 436 - 12990 to 13020 min, 437 - 13020 to 13050 min, 438 - 13050 to 13080 min, 439 - 13080 to 13110 min, 440 - 13110 to 13140 min, 441 - 13140 to 13170 min, 442 - 13170 to 13200 min, 443 - 13200 to 13230 min, 444 - 13230 to 13260 min, 445 - 13260 to 13290 min, 446 - 13290 to 13320 min, 447 - 13320 to 13350 min, 448 - 13350 to 13380 min, 449 - 13380 to 13410 min, 450 - 13410 to 13440 min, 451 - 13440 to 13470 min, 452 - 13470 to 13500 min, 453 - 13500 to 13530 min, 454 - 13530 to 13560 min, 455 - 13560 to 13590 min, 456 - 13590 to 13620 min, 457 - 13620 to 13650 min, 458 - 13650 to 13680 min, 459 - 13680 to 13710 min, 460 - 13710 to 13740 min, 461 - 13740 to 13770 min, 462 - 13770 to 13800 min, 463 - 13800 to 13830 min, 464 - 13830 to 13860 min, 465 - 13860 to 13890 min, 466 - 13890 to 13920 min, 467 - 13920 to 13950 min, 468 - 13950 to 13980 min, 469 - 13980 to 14010 min, 470 - 14010 to 14040 min, 471 - 14040 to 14070 min, 472 - 14070 to 14100 min, 473 - 14100 to 14130 min, 474 - 14130 to 14160 min, 475 - 14160 to 14190 min, 476 - 14190 to 14220 min, 477 - 14220 to 14250 min, 478 - 14250 to 14280 min, 479 - 14280 to 14310 min, 480 - 14310 to 14340 min, 481 - 14340 to 14370 min, 482 - 14370 to 14400 min, 483 - 14400 to 14430 min, 484 - 14430 to 14460 min, 485 - 14460 to 14490 min, 486 - 14490 to 14520 min, 487 - 14520 to 14550 min, 488 - 14550 to 14580 min, 489 - 14580 to 14610 min, 490 - 14610 to 14640 min, 491 - 14640 to 14670 min, 492 - 14670 to 14700 min, 493 - 14700 to 14730 min, 494 - 14730 to 14760 min, 495 - 14760 to 14790 min, 496 - 14790 to 14820 min, 497 - 14820 to 14850 min, 498 - 14850 to 14880 min, 499 - 14880 to 14910 min, 500 - 14910 to 14940 min, 501 - 14940 to 14970 min, 502 - 14970 to 15000 min, 503 - 15000 to 15030 min, 504 - 15030 to 15060 min, 505 - 15060 to 15090 min, 506 - 15090 to 15120 min, 507 - 15120 to 15150 min, 508 - 15150 to 15180 min, 509 - 15180 to 15210 min, 510 - 15210 to 15240 min, 511 - 15240 to 15270 min, 512 - 15270 to 15300 min, 513 - 15300 to 15330 min, 514 - 15330 to 15360 min, 515 - 15360 to 15390 min, 516 - 15390 to 15420 min, 517 - 15420 to 15450 min, 518 - 15450 to 15480 min, 519 - 15480 to 15510 min, 520 - 15510 to 15540 min, 521 - 15540 to 15570 min, 522 - 15570 to 15600 min, 523 - 15600 to 15630 min, 524 - 15630 to 15660 min, 525 - 15660 to 15690 min, 526 - 15690 to 15720 min, 527 - 15720 to 15750 min, 528 - 15750 to 15780 min, 529 - 15780 to 15810 min, 530 - 15810 to 15840 min, 531 - 15840 to 15870 min, 532 - 15870 to 15900 min, 533 - 15900 to 15930 min, 534 - 15930 to 15960 min, 535 - 15960 to 15990 min, 536 - 15990 to 16020 min, 537 - 16020 to 16050 min, 538 - 16050 to 16080 min, 539 - 16080 to 16110 min, 540 - 16110 to 16140 min, 541 - 16140 to 16170 min, 542 - 16170 to 16200 min, 543 - 16200 to 16230 min, 544 - 16230 to 16260 min, 545 - 16260 to 16290 min, 546 - 16290 to 16320 min, 547 - 16320 to 16350 min, 548 - 16350 to 16380 min, 549 - 16380 to 16410 min, 550 - 16410 to 16440 min, 551 - 16440 to 16470 min, 552 - 16470 to 16500 min, 553 - 16500 to 16530 min, 554 - 16530 to 16560 min, 555 - 16560 to 16590 min, 556 - 16590 to 16620 min, 557 - 16620 to 16650 min, 558 - 16650 to 16680 min, 559 - 16680 to 16710 min, 560 - 16710 to 16740 min, 561 - 16740 to 16770 min, 562 - 16770 to 16800 min, 563 - 16800 to 16830 min, 564 - 16830 to 16860 min, 565 - 16860 to 16890 min, 566 - 16890 to 16920 min, 567 - 16920 to 16950 min, 568 - 16950 to 16980 min, 569 - 16980 to 17010 min, 570 - 17010 to 17040 min, 571 - 17040 to 17070 min, 572 - 17070 to 17100 min, 573 - 17100 to 17130 min, 574 - 17130 to 17160 min, 575 - 17160 to 17190 min, 576 - 17190 to 17220 min, 577 - 17220 to 17250 min, 578 - 17250 to 17280 min, 579 - 17280 to 17310 min, 580 - 17310 to 17340 min, 581 - 17340 to 17370 min, 582 - 17370 to 17400 min, 583 - 17400 to 17430 min, 584 - 17430 to 17460 min, 585 - 17460 to 17490 min, 586 - 17490 to 17520 min, 587 - 17520 to 17550 min, 588 - 17550 to 17580 min, 589 - 17580 to 17610 min, 590 - 17610 to 17640 min, 591 - 17640 to 17670 min, 592 - 17670 to 17700 min, 593 - 17700 to 17730 min, 594 - 17730 to 17760 min, 595 - 17760 to 17790 min, 596 - 17790 to 17820 min, 597 - 17820 to 17850 min, 598 - 17850 to 17880 min, 599 - 17880 to 17910 min, 600 - 17910 to 17940 min, 601 - 17940 to 17970 min, 602 - 17970 to 18000 min, 603 - 18000 to 18030 min, 604 - 18030 to 18060 min, 605 - 18060 to 18090 min, 606 - 18090 to 18120 min, 607 - 18120 to 18150 min, 608 - 18150 to 18180 min, 609 - 18180 to 18210 min, 610 - 18210 to 18240 min, 611 - 18240 to 18270 min, 612 - 18270 to 18300 min, 613 - 18300 to 18330 min, 614 - 18330 to 18360 min, 615 - 18360 to 18390 min, 616 - 18390 to 18420 min, 617 - 18420 to 18450 min, 618 - 18450 to 18480 min, 619 - 18480 to 18510 min, 620 - 18510 to 18540 min, 621 - 18540 to 18570 min, 622 - 18570 to 18600 min, 623 - 18600 to 18630 min, 624 - 18630 to 18660 min, 625 - 18660 to 18690 min, 626 - 18690 to 18720 min, 627 - 18720 to 18750 min, 628 - 18750 to 18780 min, 629 - 18780 to 18810 min, 6

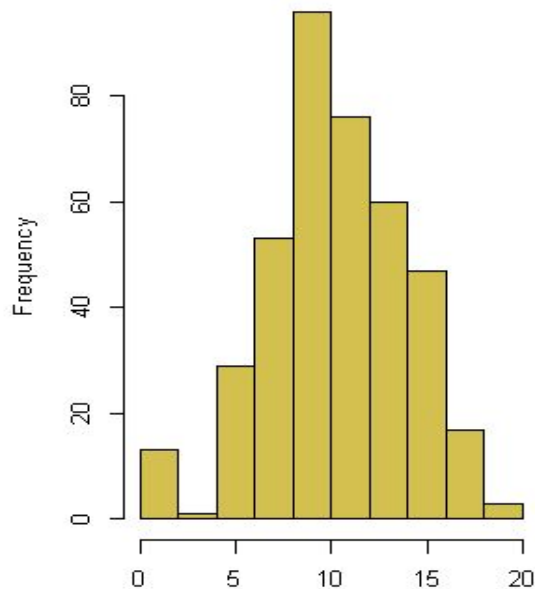
# Underlying distributions

First Period Grade(G1)



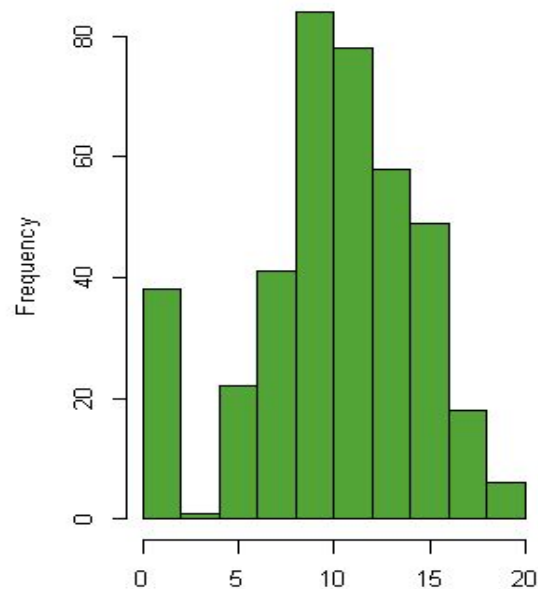
G1

Second Period Grade(G2)



G2

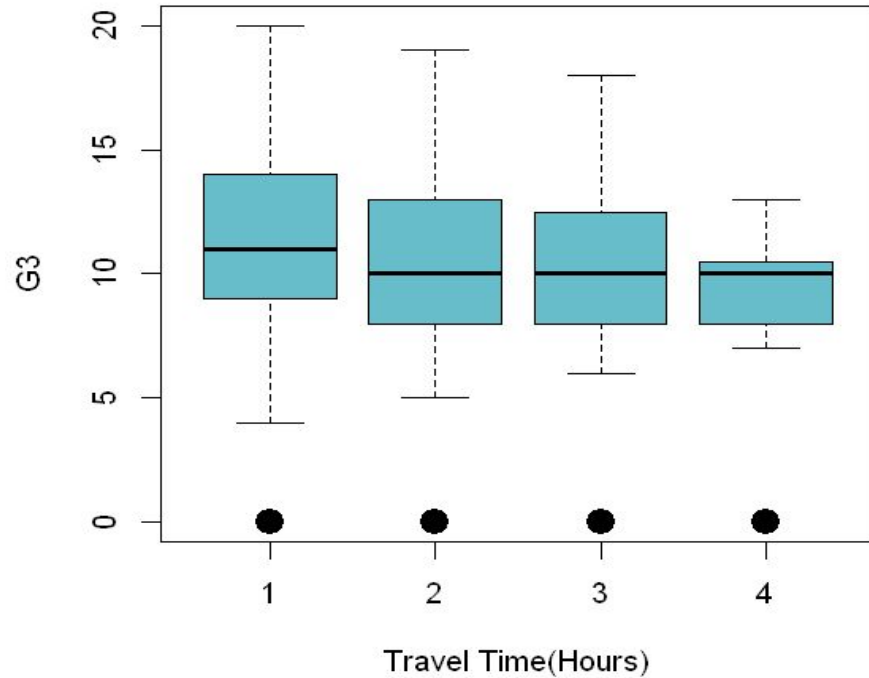
Final Grade(G3)



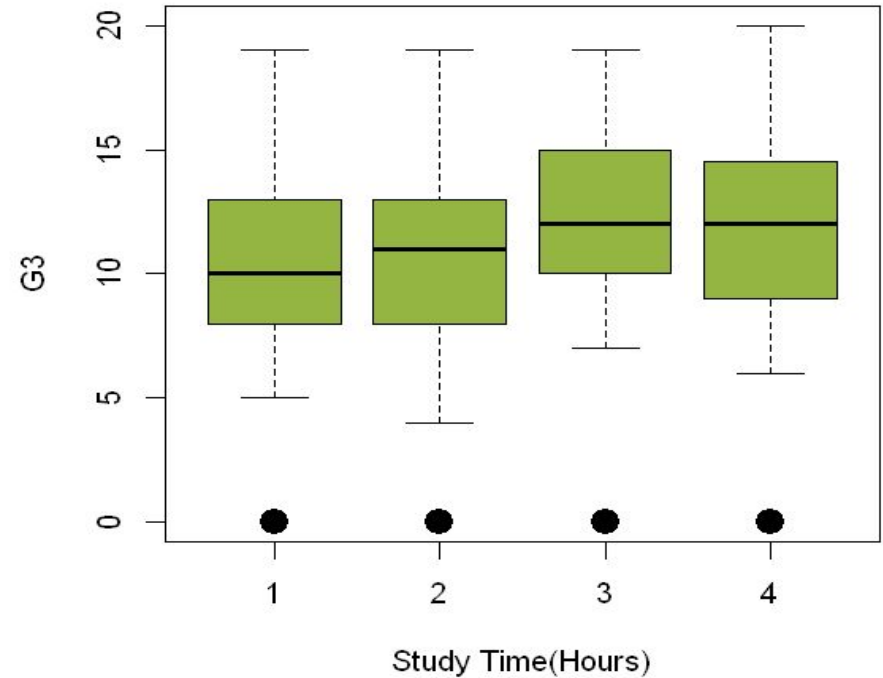
G3

# Travel-time and Study-time

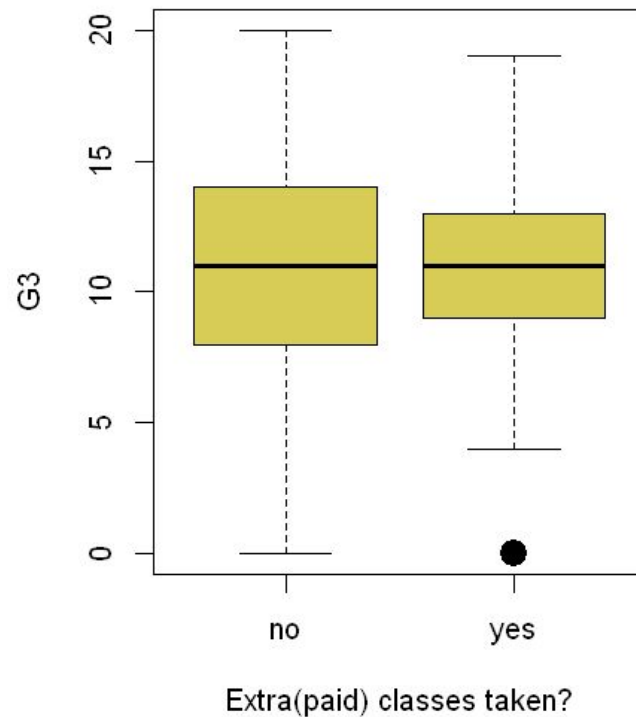
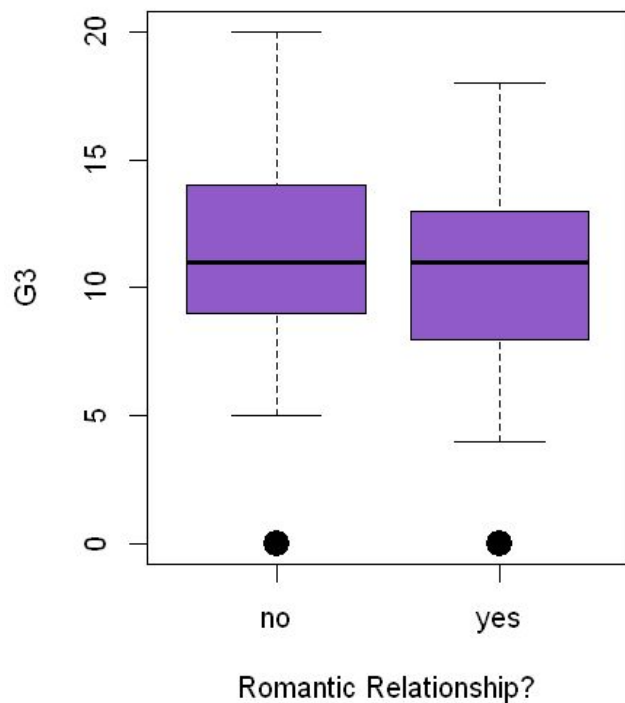
G3 vs Travel-time



G3 vs Study-time



# More features...





---

---

# Models and Methodology

---

---

## Models and Methodology

### Backward Elimination

- ❑ All the independent variables are entered into the equation first and each one is deleted one at a time if they do not contribute to the regression equation.
- ❑ Calculate the AIC(Akaike information criterion) and find the lowest score.

# AIC

- ❑ AIC: an estimator of out-of-sample prediction error and estimates the quality of each model, relative to each of the other models
- ❑ AIC provides a means for **model selection**.

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

Navigation icons: back, forward, search, etc.

# Backward Elimination

Code:

## Backward Elimination

```
reg_ob <- lm(G3~., data=df)
```

```
step(reg_ob, direction="backward")
```

**#Find the final set of features**

```
reg_2 <- lm( G3 ~ school + age + activities + romantic + famrel + Walc +
```

```
absences + G1 + G2, data = df)
```

```
summary(reg_2)
```

Call:

```
lm(formula = G3 ~ school + age + activities + romantic + famrel +  
    Walc + absences + G1 + G2, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-8.8416	-0.4534	0.2645	1.0247	4.0315

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.65012	1.45806	0.446	0.655933
schoolMS	0.45495	0.32293	1.409	0.159693
age	-0.26017	0.08405	-3.095	0.002109 **
activitiesyes	-0.32049	0.18989	-1.688	0.092278 .
romanticyes	-0.32668	0.20622	-1.584	0.113989
famrel	0.38361	0.10668	3.596	0.000365 ***
Walc	0.11979	0.07521	1.593	0.112050
absences	0.04714	0.01232	3.826	0.000152 ***
G1	0.18058	0.05517	3.273	0.001159 **
G2	0.96153	0.04904	19.607	< 2e-16 ***

---

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

Residual standard error: 1.864 on 385 degrees of freedom

Multiple R-squared: 0.8382, Adjusted R-squared: 0.8344

F-statistic: 221.6 on 9 and 385 DF, p-value: < 2.2e-16

# Models and Methodology

## Forward Selection

Start with a NULL model with no variable and going forward to choose the one with lowest AIC.

```
basemodel <- lm(G3~NULL, data = d1)

model.forward <- step(basemodel, direction = "forward", trace = F, scope = ~ school + sex + age +
  Fedu + Mjob + Fjob + reason + guardian + traveltime + studytime +
  failures + schoolsup + famsup + paid + activities + nursery +
  higher + internet + romantic + famrel + freetime + goout +
  Dalc + Walc + health + absences + G1 + G2)

summary(model.forward)
```

# Forward Selection

```
Call:
lm(formula = G3 ~ G2 + famrel + absences + G1 + age + activities +
    Walc + romantic + school, data = d1)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-8.8416	-0.4534	0.2645	1.0247	4.0315

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.65012	1.45806	0.446	0.655933
G2	0.96153	0.04904	19.607	< 2e-16 ***
famrel	0.38361	0.10668	3.596	0.000365 ***
absences	0.04714	0.01232	3.826	0.000152 ***
G1	0.18058	0.05517	3.273	0.001159 **
age	-0.26017	0.08405	-3.095	0.002109 **
activitiesyes	-0.32049	0.18989	-1.688	0.092278 .
Walc	0.11979	0.07521	1.593	0.112050
romanticyes	-0.32668	0.20622	-1.584	0.113989
schoolMS	0.45495	0.32293	1.409	0.159693

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.864 on 385 degrees of freedom

Multiple R-squared: 0.8382, Adjusted R-squared: 0.8344

F-statistic: 221.6 on 9 and 385 DF, p-value: < 2.2e-16

Adjusted R-square: 0.8344

---

---

# Interpretation of Results & Conclusion

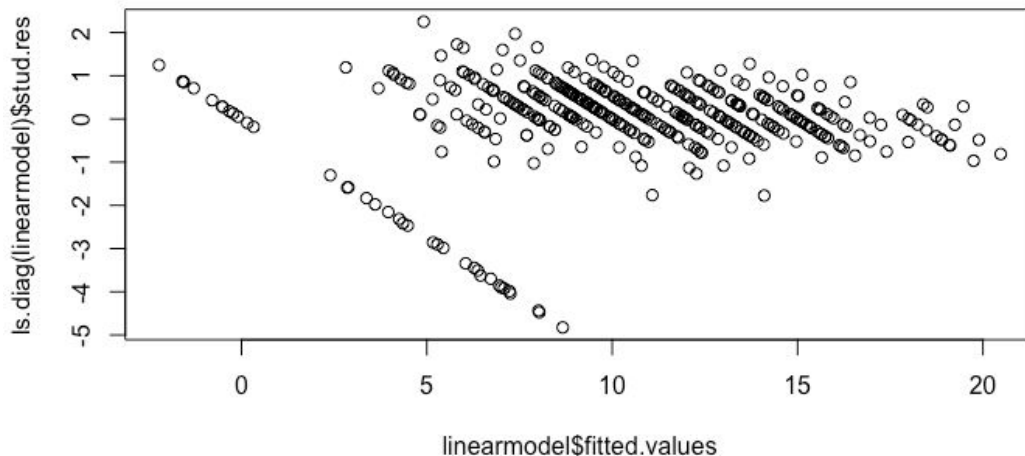
---

---

# Interpretation of Results

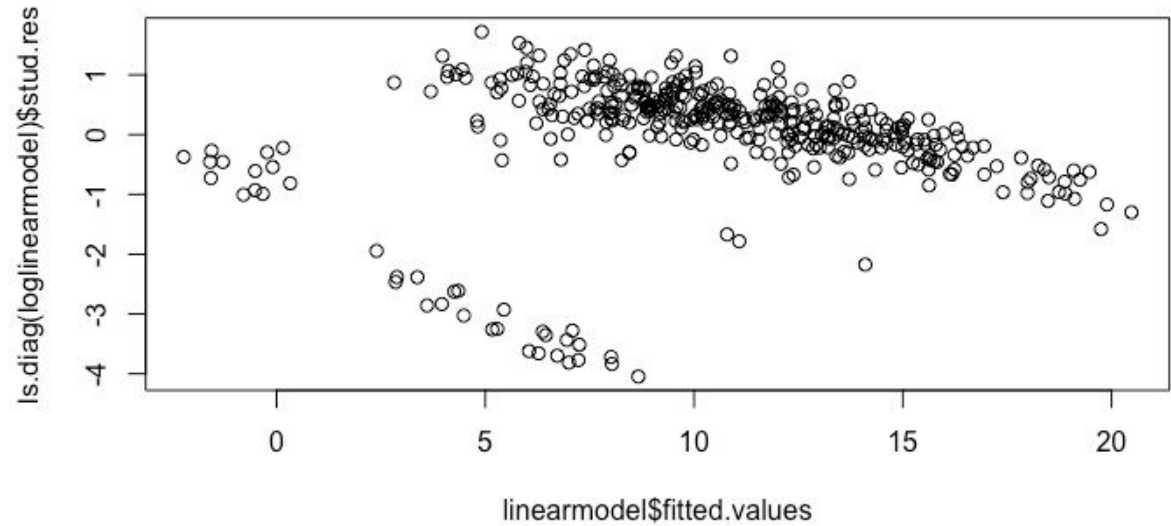
$$\hat{G3} = 0.65012 + 0.45495 * schoolMS - 0.26017 * age - 0.32049 * activity - 0.32668 * romantic + 0.38361 * famrel + 0.11979 * Walc + 0.04714 * absences + 0.18058 * G1 + 0.96153 * G2$$

$$adjR^2 = 0.8344$$

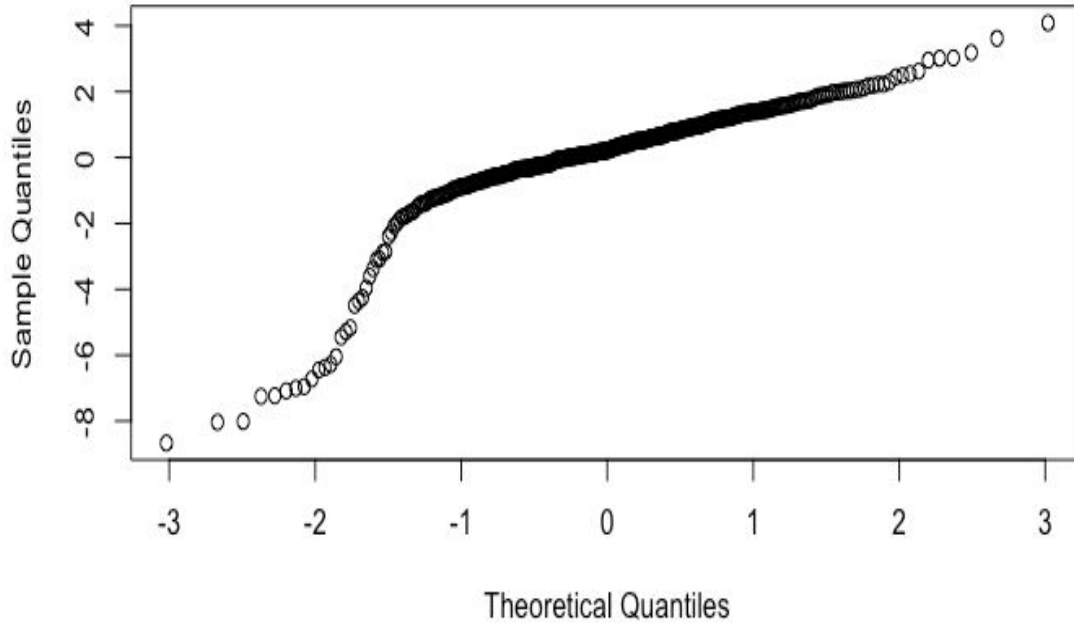




Log-linear model:  
 $\text{adj}R^2 = 0.6099$



Normal Q-Q Plot



The effect of a p-unit increase in explanatory variable  $X_j$  is to multiply the mean of  $y$  by  $\exp(p\beta_j)$ .

# Conclusion

We started with 32 features, went down to ... 9 features

We only did this for Mathematics grades

A similar approach can be taken for Portuguese data

# Reference

**Citation:** P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUTURE BUSINESS TECHNOLOGY Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7.