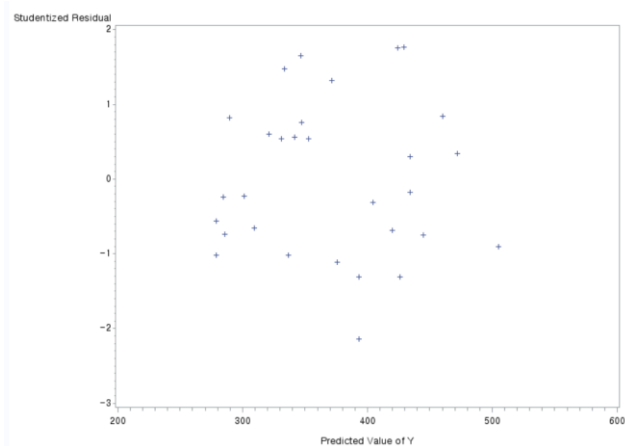


오유진(2015325) 회귀분석론 과제2

6.3

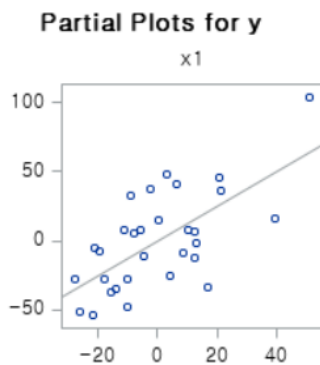
6.3.1

헬스클럽자료 중 Y를 설명변수, X1,X2,X3,X4를 설명변수라고 두고 스튜던트화잔차와 적합값의 잔차산점도는 아래와 같다.



이는 회귀모형이 적절하게 적합되었을 때 보여지는 잔차산점도라고 볼 수 있기 때문에 정규분포를 따른다는 귀무가설에 문제가 없다는 것을 알 수 있고 추가적인 변환이 필요없다. 비선형의 관계가 매우 약하다.

6.3.2



양의 상관관계가 있긴하지만 미약하다는 것을 알 수있고 두 잔차의 선형관계가 뚜렷이 나타나 있지않기 때문에 X1의 추가적인 설명력이 존재를 나타낸다. 그리고 맨 오른쪽 상단에 이상점으로 보이는 관측값이 있다.

6.3.3

| OBS | no | x1 | x2 | x3 | x4 | y | ei | ri | di | hi | ti |
|-----|----|-----|----|-----|----|-----|----------|----------|---------|---------|----------|
| 1 | 1 | 217 | 67 | 260 | 91 | 481 | 20.8626 | 0.83988 | 0.04688 | 0.24941 | 0.83477 |
| 2 | 2 | 141 | 52 | 190 | 66 | 292 | -17.4516 | -0.65115 | 0.01225 | 0.12622 | -0.64348 |
| 3 | 3 | 152 | 58 | 203 | 68 | 338 | 16.5149 | 0.60394 | 0.00725 | 0.09037 | 0.59610 |
| 4 | 4 | 153 | 56 | 183 | 70 | 357 | 15.2904 | 0.55560 | 0.00527 | 0.07867 | 0.54777 |
| 5 | 5 | 180 | 66 | 170 | 77 | 396 | -8.5700 | -0.31339 | 0.00195 | 0.09034 | -0.30767 |
| 6 | 6 | 193 | 71 | 178 | 82 | 429 | -4.8981 | -0.18113 | 0.00081 | 0.11043 | -0.17759 |
| 7 | 7 | 162 | 65 | 160 | 74 | 345 | -30.6190 | -1.11601 | 0.02294 | 0.08432 | -1.12176 |
| 8 | 8 | 180 | 80 | 170 | 84 | 469 | 44.4618 | 1.75652 | 0.17464 | 0.22059 | 1.83819 |
| 9 | 9 | 205 | 77 | 188 | 83 | 425 | -19.8113 | -0.74806 | 0.01926 | 0.14680 | -0.74129 |
| 10 | 10 | 168 | 74 | 170 | 79 | 358 | -34.9628 | -1.30629 | 0.05035 | 0.12857 | -1.32595 |
| 11 | 11 | 232 | 65 | 220 | 72 | 393 | -33.2460 | -1.30752 | 0.09283 | 0.21353 | -1.32729 |
| 12 | 12 | 146 | 68 | 158 | 68 | 346 | 14.6468 | 0.53827 | 0.00639 | 0.09929 | 0.53048 |
| 13 | 13 | 173 | 51 | 243 | 56 | 279 | -5.7448 | -0.23907 | 0.00484 | 0.29756 | -0.23450 |
| 14 | 14 | 155 | 64 | 198 | 59 | 311 | 21.4653 | 0.81764 | 0.02577 | 0.16160 | 0.81205 |
| 15 | 15 | 212 | 66 | 220 | 77 | 401 | -18.8832 | -0.68310 | 0.00707 | 0.07043 | -0.67563 |
| 16 | 16 | 138 | 70 | 180 | 62 | 267 | -18.6332 | -0.74132 | 0.03310 | 0.23147 | -0.73446 |
| 17 | 17 | 147 | 54 | 150 | 75 | 404 | 32.6650 | 1.31391 | 0.11396 | 0.24815 | 1.33426 |
| 18 | 18 | 197 | 76 | 228 | 88 | 442 | 7.4904 | 0.29708 | 0.00517 | 0.22665 | 0.29159 |
| 19 | 19 | 165 | 59 | 188 | 70 | 368 | 15.1795 | 0.54221 | 0.00287 | 0.04660 | 0.53441 |
| 20 | 20 | 125 | 58 | 160 | 66 | 295 | -6.1689 | -0.22858 | 0.00134 | 0.11399 | -0.22420 |
| 21 | 21 | 161 | 52 | 190 | 69 | 391 | 44.4867 | 1.65338 | 0.07408 | 0.11933 | 1.71654 |
| 22 | 22 | 132 | 62 | 163 | 59 | 264 | -15.1054 | -0.56216 | 0.00876 | 0.12171 | -0.55432 |
| 23 | 23 | 257 | 64 | 313 | 96 | 487 | -18.1672 | -0.90827 | 0.17402 | 0.51332 | -0.90498 |
| 24 | 24 | 236 | 72 | 225 | 84 | 481 | 9.0489 | 0.34096 | 0.00389 | 0.14319 | 0.33485 |
| 25 | 25 | 149 | 57 | 173 | 68 | 374 | 40.6420 | 1.47158 | 0.03368 | 0.07215 | 1.50867 |
| 26 | 26 | 161 | 57 | 173 | 65 | 309 | -27.8606 | -1.01556 | 0.01903 | 0.08448 | -1.01623 |
| 27 | 27 | 198 | 59 | 220 | 62 | 367 | 19.7324 | 0.75405 | 0.02279 | 0.16697 | 0.74736 |
| 28 | 28 | 245 | 70 | 218 | 69 | 469 | 39.5998 | 1.76478 | 0.39407 | 0.38750 | 1.84806 |
| 29 | 29 | 141 | 63 | 193 | 60 | 252 | -26.7412 | -1.01968 | 0.04061 | 0.16337 | -1.02052 |
| 30 | 30 | 177 | 53 | 183 | 75 | 338 | -55.2235 | -2.14405 | 0.21987 | 0.19299 | -2.32538 |

위 그림과 위 표를 보더라도 30번째 관측값에서 가장 큰 t_i (외적 스튜던트화잔차)의 절댓값이 가장 크므로 30번째 관측값에 대해 이상점검정을 할 필요가 있다.

위 경우는 $p'=5$, $n=30$ 이므로 부록 [표 5B]에서 유의수준 0.01의 검정에 대한 기각값은 4.18이므로 $|t_{30}| < 4.18$ 이기 때문에 30번째 관측값은 이상점으로 판단할 수 없다.

6.3.4

6.3.3의 표에 d_i 가 cook의 D통계량을 말한다. Cook의 D통계량 같은 경우 28번째 관측값이 가장 큰 값으로 0.39407, 두번째로 30번째 관측값이 0.21987로 큰 값을 가진다.

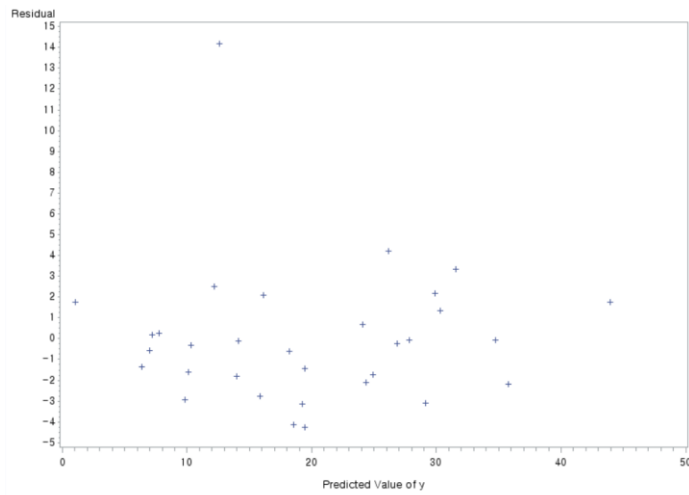
| Output Statistics | | | | | | | | | | |
|-------------------|----------|----------|------------|-----------|---------|-----------|---------|---------|---------|---------|
| Obs | Residual | RStudent | Hat Diag H | Cov Ratio | DFFITS | DFBETAS | | | | |
| | | | | | | Intercept | x1 | x2 | x3 | x4 |
| 1 | 20.8626 | 0.8348 | 0.2494 | 1.4161 | 0.4812 | -0.3106 | -0.2006 | -0.0011 | 0.2802 | 0.2785 |
| 2 | -17.4516 | -0.6435 | 0.1262 | 1.2885 | -0.2446 | -0.1129 | 0.0953 | 0.1233 | -0.0550 | -0.0603 |
| 3 | 16.5149 | 0.5961 | 0.0904 | 1.2528 | 0.1879 | 0.0110 | -0.1197 | -0.0084 | 0.1125 | 0.0201 |
| 4 | 15.2904 | 0.5478 | 0.0787 | 1.2510 | 0.1601 | 0.0812 | -0.0313 | -0.0924 | -0.0130 | 0.0603 |
| 5 | -8.5700 | -0.3077 | 0.0903 | 1.3220 | -0.0970 | -0.0242 | -0.0401 | 0.0210 | 0.0703 | -0.0267 |
| 6 | -4.8981 | -0.1776 | 0.1104 | 1.3697 | -0.0626 | 0.0077 | -0.0197 | -0.0035 | 0.0354 | -0.0207 |
| 7 | -30.6190 | -1.1218 | 0.0843 | 1.0374 | -0.3404 | -0.1046 | -0.0532 | 0.0427 | 0.2147 | -0.0964 |
| 8 | 44.4618 | 1.8382 | 0.2206 | 0.8144 | 0.9779 | -0.5051 | -0.1867 | 0.6034 | -0.1233 | 0.2333 |
| 9 | -19.8113 | -0.7413 | 0.1468 | 1.2836 | -0.3075 | 0.1297 | -0.0719 | -0.1449 | 0.0966 | -0.0204 |
| 10 | -34.9628 | -1.3259 | 0.1286 | 0.9883 | -0.5093 | 0.2015 | 0.1408 | -0.2824 | 0.0557 | -0.1291 |
| 11 | -33.2460 | -1.3273 | 0.2135 | 1.0943 | -0.6916 | -0.1541 | -0.5859 | 0.0981 | 0.2627 | 0.3486 |
| 12 | 14.6468 | 0.5305 | 0.0993 | 1.2845 | 0.1761 | 0.0096 | -0.0432 | 0.0896 | -0.0262 | -0.0272 |
| 13 | -5.7448 | -0.2345 | 0.2976 | 1.7261 | -0.1526 | -0.0363 | 0.0120 | 0.0146 | -0.0826 | 0.0801 |
| 14 | 21.4653 | 0.8120 | 0.1616 | 1.2774 | 0.3565 | 0.0014 | -0.0963 | 0.2000 | 0.1675 | -0.2468 |
| 15 | -18.8832 | -0.6756 | 0.0704 | 1.2007 | -0.1860 | 0.0235 | -0.0936 | 0.0089 | 0.0148 | 0.0287 |
| 16 | -18.6332 | -0.7345 | 0.2315 | 1.4280 | -0.4031 | 0.0989 | 0.2166 | -0.3167 | -0.1892 | 0.1687 |
| 17 | 32.6650 | 1.3343 | 0.2482 | 1.1406 | 0.7665 | 0.4306 | 0.0987 | -0.5497 | -0.4427 | 0.4381 |
| 18 | 7.4904 | 0.2916 | 0.2267 | 1.5581 | 0.1579 | -0.1335 | -0.0915 | 0.0872 | 0.0885 | 0.0629 |
| 19 | 15.1795 | 0.5344 | 0.0466 | 1.2125 | 0.1182 | 0.0603 | 0.0013 | -0.0533 | -0.0177 | 0.0156 |
| 20 | -6.1689 | -0.2242 | 0.1140 | 1.3698 | -0.0804 | -0.0303 | 0.0417 | 0.0098 | -0.0019 | -0.0201 |
| 21 | 44.4867 | 1.7165 | 0.1193 | 0.7805 | 0.6319 | 0.4123 | 0.0609 | -0.5082 | -0.1218 | 0.1707 |
| 22 | -15.1054 | -0.5543 | 0.1217 | 1.3103 | -0.2064 | -0.0672 | 0.0591 | -0.0774 | -0.0192 | 0.0992 |
| 23 | -18.1672 | -0.9050 | 0.5133 | 2.1308 | -0.9294 | 0.5057 | 0.1988 | 0.1358 | -0.5425 | -0.3881 |
| 24 | 9.0489 | 0.3349 | 0.1432 | 1.3984 | 0.1369 | -0.0444 | 0.0771 | 0.0122 | -0.0312 | -0.0043 |
| 25 | 40.6420 | 1.5087 | 0.0721 | 0.8403 | 0.4207 | 0.2643 | -0.0382 | -0.2064 | -0.1050 | 0.0918 |
| 26 | -27.8606 | -1.0162 | 0.0845 | 1.0852 | -0.3087 | -0.2540 | -0.1318 | 0.1545 | 0.1622 | 0.0671 |
| 27 | 19.7324 | 0.7474 | 0.1670 | 1.3123 | 0.3346 | 0.1212 | 0.1713 | -0.0261 | 0.0053 | -0.2495 |
| 28 | 39.5998 | 1.8481 | 0.3875 | 1.0295 | 1.4699 | 0.2151 | 1.2032 | 0.1444 | -0.5380 | -0.9746 |
| 29 | -26.7412 | -1.0205 | 0.1634 | 1.1854 | -0.4510 | 0.0221 | 0.2416 | -0.2439 | -0.2628 | 0.2117 |
| 30 | -55.2235 | -2.3254 | 0.1930 | 0.5502 | -1.1372 | -0.6831 | -0.4391 | 0.9938 | 0.6054 | -0.4707 |

DFFITS 값은 28번째, 30번째 순서대로 절댓값이 크고 기준값이 $2\sqrt{(p'/n)}=2\sqrt{(1/6)}=0.8165$ 보다 크므로 영향력이 크다고 할 수 있다. 그리고 28,30번째 관측값의 DFBETAS의 값들은 모두 매우 크므로 28번째 관측값은 영향력이 크다고 볼 수 있다.

따라서 28번째, 30번째 관측값이 영향력이 크다.

6.5

6.5.1



잔차산점도는 위와 같은데 11번째 관측값이 가장 잔차가 크고, 32번째 관측값이 y의 적합값이 가장 크다.

| OBS | no | x1 | x2 | x3 | x4 | y | ei | ri | di | hi | ti |
|-----|----|------|-----|-----|-----|------|---------|----------|---------|---------|----------|
| 1 | 1 | 38.4 | 6.1 | 220 | 235 | 6.9 | -2.9372 | -0.87605 | 0.02259 | 0.12828 | -0.87215 |
| 2 | 2 | 40.3 | 4.8 | 231 | 307 | 14.4 | -4.1384 | -1.17330 | 0.01007 | 0.03527 | -1.18189 |
| 3 | 3 | 40.0 | 6.1 | 217 | 212 | 7.4 | 0.1813 | 0.05451 | 0.00010 | 0.14229 | 0.05349 |
| 4 | 4 | 31.8 | 0.2 | 316 | 365 | 8.5 | -1.6013 | -0.49261 | 0.01070 | 0.18065 | -0.48559 |
| 5 | 5 | 40.8 | 3.5 | 210 | 218 | 8.0 | 0.2777 | 0.09103 | 0.00064 | 0.27851 | 0.08934 |
| 6 | 6 | 41.3 | 1.8 | 267 | 235 | 2.8 | 1.7673 | 0.54699 | 0.01408 | 0.19047 | 0.53977 |
| 7 | 7 | 38.1 | 1.2 | 274 | 285 | 5.0 | -1.3525 | -0.40020 | 0.00414 | 0.11438 | -0.39389 |
| 8 | 8 | 50.8 | 8.6 | 190 | 205 | 12.2 | -1.8145 | -0.59465 | 0.02723 | 0.27797 | -0.58739 |
| 9 | 9 | 32.2 | 5.2 | 236 | 267 | 10.0 | -0.3353 | -0.10420 | 0.00053 | 0.19717 | -0.10227 |
| 10 | 10 | 38.4 | 6.1 | 220 | 300 | 15.2 | -4.2700 | -1.23408 | 0.02350 | 0.07161 | -1.24668 |
| 11 | 11 | 40.3 | 4.8 | 231 | 267 | 26.8 | 14.1895 | 4.06942 | 0.20093 | 0.05720 | 6.42205 |
| 12 | 12 | 32.2 | 2.4 | 284 | 351 | 14.0 | -0.1040 | -0.03072 | 0.00002 | 0.11054 | -0.03015 |
| 13 | 13 | 31.8 | 0.2 | 316 | 379 | 14.7 | 2.5240 | 0.77389 | 0.02543 | 0.17514 | 0.76799 |
| 14 | 14 | 41.3 | 1.8 | 267 | 275 | 6.4 | -0.5605 | -0.16793 | 0.00089 | 0.13607 | -0.16488 |
| 15 | 15 | 38.1 | 1.2 | 274 | 365 | 17.6 | -0.6082 | -0.17763 | 0.00063 | 0.09095 | -0.17441 |
| 16 | 16 | 50.8 | 8.6 | 190 | 275 | 22.3 | -2.0883 | -0.65494 | 0.02303 | 0.21164 | -0.64786 |
| 17 | 17 | 32.2 | 5.2 | 236 | 360 | 24.8 | 0.6825 | 0.20746 | 0.00165 | 0.16088 | 0.20375 |
| 18 | 18 | 38.4 | 6.1 | 220 | 365 | 26.0 | -3.1028 | -0.90313 | 0.01510 | 0.08470 | -0.89994 |
| 19 | 19 | 40.3 | 4.8 | 231 | 395 | 34.9 | 3.3203 | 0.96398 | 0.01617 | 0.08002 | 0.96267 |
| 20 | 20 | 40.0 | 6.1 | 217 | 272 | 18.2 | 2.0895 | 0.60337 | 0.00548 | 0.07002 | 0.59612 |
| 21 | 21 | 32.2 | 2.4 | 284 | 424 | 23.2 | -1.7224 | -0.51312 | 0.00761 | 0.12628 | -0.50601 |
| 22 | 22 | 31.8 | 0.2 | 316 | 428 | 18.0 | -1.4376 | -0.44247 | 0.00867 | 0.18136 | -0.43578 |
| 23 | 23 | 40.8 | 3.5 | 210 | 273 | 13.1 | -2.7731 | -0.89268 | 0.05359 | 0.25164 | -0.88921 |
| 24 | 24 | 41.3 | 1.8 | 267 | 358 | 16.1 | -3.1608 | -0.93167 | 0.02090 | 0.10745 | -0.92932 |
| 25 | 25 | 38.1 | 1.2 | 274 | 444 | 32.1 | 2.1843 | 0.66826 | 0.01849 | 0.17151 | 0.66126 |
| 26 | 26 | 50.8 | 8.6 | 190 | 345 | 34.7 | -0.0620 | -0.01963 | 0.00002 | 0.22620 | -0.01927 |
| 27 | 27 | 32.2 | 5.2 | 236 | 402 | 31.7 | 1.3582 | 0.42058 | 0.00837 | 0.19130 | 0.41408 |
| 28 | 28 | 38.4 | 6.1 | 220 | 410 | 33.6 | -2.1716 | -0.65007 | 0.01315 | 0.13463 | -0.64297 |
| 29 | 29 | 40.0 | 6.1 | 217 | 340 | 30.4 | 4.2121 | 1.20978 | 0.01867 | 0.05996 | 1.22071 |
| 30 | 30 | 40.8 | 3.5 | 210 | 347 | 26.6 | -0.2397 | -0.07945 | 0.00053 | 0.29429 | -0.07798 |
| 31 | 31 | 41.3 | 1.8 | 267 | 416 | 27.8 | -0.0562 | -0.01703 | 0.00001 | 0.15496 | -0.01671 |
| 32 | 32 | 50.8 | 8.6 | 190 | 407 | 45.7 | 1.7498 | 0.58518 | 0.03029 | 0.30666 | 0.57792 |

11번째 관측값은 자료영역의 가운데에 놓여있어서 h_{ii} 는 0.05720으로 매우 작지만 r_i 가 4.06942로 매우 크므로 cook의 D의 통계량(d_i)는 0.20093으로 크게 나와서 영향력을 크다는 것을 알 수 있다. 이는 6.2의 결과와 d_i 값이 정확하게 같진 않지만, h_i 가 매우 작고, r_i 가 매우 크고 d_i 가 큰 값을 갖는 다는 것과 같다. 32번째 관측값은 d_i 값이 0.03029로 작게 나와서 영향력이 크다고 볼 수는 없다.

6.5.2

x_4 가 잘못 기재될 가능성이 높다. 설명변수 x_1, x_2, x_3 이 같은 값을 갖는다면 x_4 값이 클수록 y 값도 큰 값을 보여줘야한다. 하지만 11번째 관측값은 2번째, 19번째 관측값에 비해 x_4 의 값은 작지만 y 값은 크므로, 설명변수 x_4 가 잘못 기재될 가능성이 높다.