

LSM

May 10, 2018

Referência: https://en.wikipedia.org/wiki/Ordinary_least_squares

Em estatística, mínimos quadrados ordinários (ordinary least squares - OLS) ou mínimos quadrados lineares é um método para estimar os parâmetros desconhecidos em um modelo de regressão. Este método escolhe os parâmetros de uma função linear de um conjunto de variáveis pela minimização da soma da diferença dos quadrados entre o valor observado das variáveis dependentes no conjunto de dados e o valor predito pela função linear. OLS é usado em campos diversos, como economia, ciências políticas, psicologia e engenharia.

Modelo Linear

Suponha que os dados consistem de n observações $\{y_i, x_i\}_{i=1}^n$. Cada observação i inclui uma resposta escalar y_i e um vetor de valores de p preditores (regressores) x_{ij} para $j = 1..p$. No modelo de regressão linear, a variável de resposta y_i é uma função linear dos regressores:

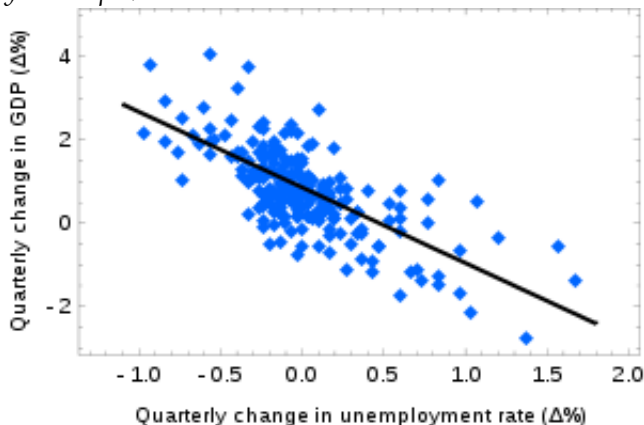
$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i,$$

ou na forma vetorial,

$$y_i = x_i^T \beta + \epsilon_i,$$

onde β é um vetor coluna de dimensão p com os parâmetros desconhecidos, ϵ_i são os escalares não observados representando os erros (variável aleatória) que é o responsável pela discrepância entre a resposta verdadeira os valores preditos $x_i \beta$, o mesmo problema pode ser expresso na seguinte forma matricial:

$$y = X^T \beta + \epsilon.$$

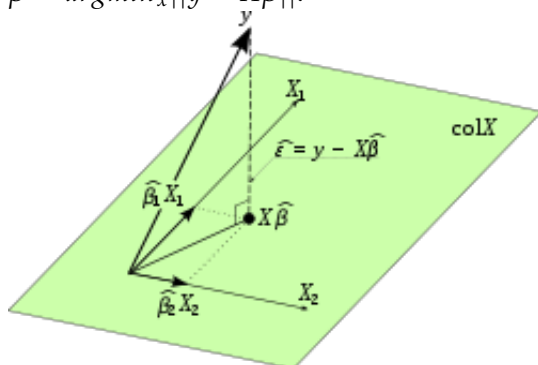


Lei de Okun em macroeconomia diz que o crescimento do PIB de uma economia deveria depender linearmente de mudanças na taxa de desemprego. Aqui o OLS é usado para construir um regressor linear descrevendo esta lei.

Abordagem geométrica

Para o matemático, OLS é uma solução aproximada para um sistema superdeterminado de equações lineares $X\beta \approx y$. Supondo que o sistema não tem solução exata, estamos olhando para uma solução que prove a menor discrepância. Em outras palavras, estamos procurando soluções que satisfaça:

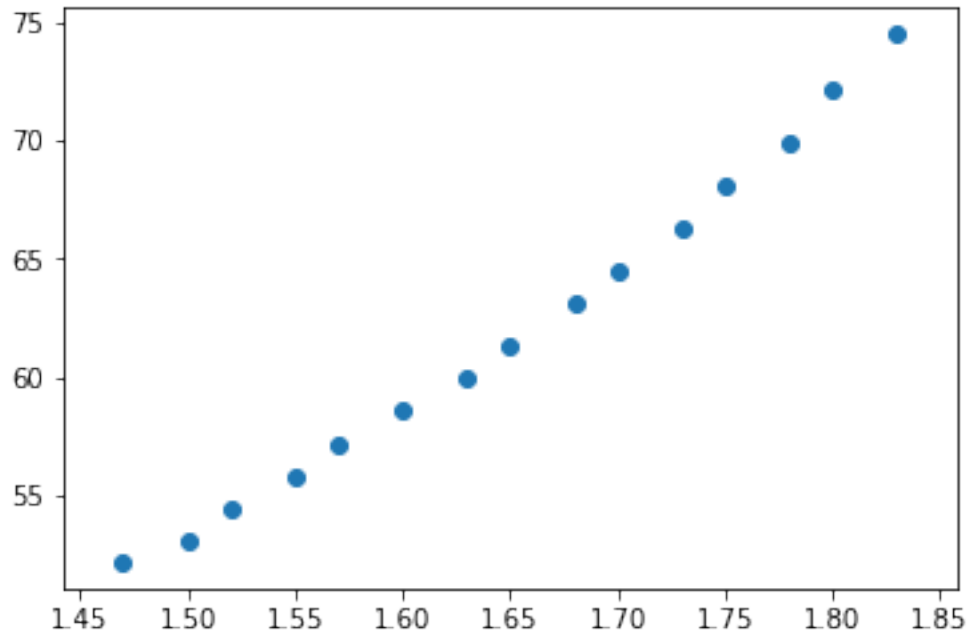
$$\hat{\beta} = \operatorname{argmin}_x ||y - X\beta||.$$



A estimação do OLS pode ser vista como uma projeção em um espaço linear coberto pelos regressores. OLS estimation can be viewed as a projection onto the linear space spanned by the regressors. (Aqui cada X_1 e X_2 refere-se a uma coluna da matriz de dados.

```
In [11]: import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets, linear_model
from sklearn.metrics import mean_squared_error, r2_score
%matplotlib inline

height = np.array( [1.47, 1.50, 1.52, 1.55, 1.57, 1.60, 1.63, 1.65, 1.68, 1.70, 1.73, 1.76, 1.78, 1.80, 1.82, 1.84, 1.86, 1.88, 1.90, 1.92, 1.94, 1.96, 1.98, 2.00])
weight = np.array( [52.21, 53.12, 54.48, 55.84, 57.20, 58.57, 59.93, 61.29, 63.11, 64.47, 66.26, 67.50, 68.75, 69.95, 71.28, 72.62, 73.86, 75.18, 76.43, 77.71, 78.97, 80.24, 81.50, 82.74, 84.09, 85.31, 86.59, 87.86, 89.15, 90.42, 91.70, 92.98, 94.26, 95.54, 96.82, 98.10, 99.38, 100.66, 101.94, 103.22, 104.50, 105.78, 107.06, 108.34, 109.62, 110.90, 112.18, 113.46, 114.74, 116.02, 117.30, 118.58, 119.86, 121.14, 122.42, 123.70, 124.98, 126.26, 127.54, 128.82, 130.10, 131.38, 132.66, 133.94, 135.22, 136.50, 137.78, 139.06, 140.34, 141.62, 142.90, 144.18, 145.46, 146.74, 148.02, 149.30, 150.58, 151.86, 153.14, 154.42, 155.70, 156.98, 158.26, 159.54, 160.82, 162.10, 163.38, 164.66, 165.94, 167.22, 168.50, 169.78, 171.06, 172.34, 173.62, 174.90, 176.18, 177.46, 178.74, 180.02, 181.30, 182.58, 183.86, 185.14, 186.42, 187.70, 188.98, 190.26, 191.54, 192.82, 194.10, 195.38, 196.66, 197.94, 199.22, 200.50, 201.78, 203.06, 204.34, 205.62, 206.90, 208.18, 209.46, 210.74, 212.02, 213.30, 214.58, 215.86, 217.14, 218.42, 219.70, 220.98, 222.26, 223.54, 224.82, 226.10, 227.38, 228.66, 229.94, 231.22, 232.50, 233.78, 235.06, 236.34, 237.62, 238.90, 240.18, 241.46, 242.74, 244.02, 245.30, 246.58, 247.86, 249.14, 250.42, 251.70, 252.98, 254.26, 255.54, 256.82, 258.10, 259.38, 260.66, 261.94, 263.22, 264.50, 265.78, 267.06, 268.34, 269.62, 270.90, 272.18, 273.46, 274.74, 276.02, 277.30, 278.58, 279.86, 281.14, 282.42, 283.70, 284.98, 286.26, 287.54, 288.82, 290.10, 291.38, 292.66, 293.94, 295.22, 296.50, 297.78, 299.06, 300.34, 301.62, 302.90, 304.18, 305.46, 306.74, 308.02, 309.30, 310.58, 311.86, 313.14, 314.42, 315.70, 316.98, 318.26, 319.54, 320.82, 322.10, 323.38, 324.66, 325.94, 327.22, 328.50, 329.78, 331.06, 332.34, 333.62, 334.90, 336.18, 337.46, 338.74, 340.02, 341.30, 342.58, 343.86, 345.14, 346.42, 347.70, 348.98, 350.26, 351.54, 352.82, 354.10, 355.38, 356.66, 357.94, 359.22, 360.50, 361.78, 363.06, 364.34, 365.62, 366.90, 368.18, 369.46, 370.74, 372.02, 373.30, 374.58, 375.86, 377.14, 378.42, 379.70, 380.98, 382.26, 383.54, 384.82, 386.10, 387.38, 388.66, 389.94, 391.22, 392.50, 393.78, 395.06, 396.34, 397.62, 398.90, 400.18, 401.46, 402.74, 404.02, 405.30, 406.58, 407.86, 409.14, 410.42, 411.70, 412.98, 414.26, 415.54, 416.82, 418.10, 419.38, 420.66, 421.94, 423.22, 424.50, 425.78, 427.06, 428.34, 429.62, 430.90, 432.18, 433.46, 434.74, 436.02, 437.30, 438.58, 439.86, 441.14, 442.42, 443.70, 444.98, 446.26, 447.54, 448.82, 450.10, 451.38, 452.66, 453.94, 455.22, 456.50, 457.78, 459.06, 460.34, 461.62, 462.90, 464.18, 465.46, 466.74, 468.02, 469.30, 470.58, 471.86, 473.14, 474.42, 475.70, 476.98, 478.26, 479.54, 480.82, 482.10, 483.38, 484.66, 485.94, 487.22, 488.50, 489.78, 491.06, 492.34, 493.62, 494.90, 496.18, 497.46, 498.74, 500.02, 501.30, 502.58, 503.86, 505.14, 506.42, 507.70, 508.98, 510.26, 511.54, 512.82, 514.10, 515.38, 516.66, 517.94, 519.22, 520.50, 521.78, 523.06, 524.34, 525.62, 526.90, 528.18, 529.46, 530.74, 532.02, 533.30, 534.58, 535.86, 537.14, 538.42, 539.70, 540.98, 542.26, 543.54, 544.82, 546.10, 547.38, 548.66, 549.94, 551.22, 552.50, 553.78, 555.06, 556.34, 557.62, 558.90, 560.18, 561.46, 562.74, 564.02, 565.30, 566.58, 567.86, 569.14, 570.42, 571.70, 572.98, 574.26, 575.54, 576.82, 578.10, 579.38, 580.66, 581.94, 583.22, 584.50, 585.78, 587.06, 588.34, 589.62, 590.90, 592.18, 593.46, 594.74, 596.02, 597.30, 598.58, 599.86, 601.14, 602.42, 603.70, 604.98, 606.26, 607.54, 608.82, 610.10, 611.38, 612.66, 613.94, 615.22, 616.50, 617.78, 619.06, 620.34, 621.62, 622.90, 624.18, 625.46, 626.74, 628.02, 629.30, 630.58, 631.86, 633.14, 634.42, 635.70, 636.98, 638.26, 639.54, 640.82, 642.10, 643.38, 644.66, 645.94, 647.22, 648.50, 649.78, 651.06, 652.34, 653.62, 654.90, 656.18, 657.46, 658.74, 660.02, 661.30, 662.58, 663.86, 665.14, 666.42, 667.70, 668.98, 670.26, 671.54, 672.82, 674.10, 675.38, 676.66, 677.94, 679.22, 680.50, 681.78, 683.06, 684.34, 685.62, 686.90, 688.18, 689.46, 690.74, 692.02, 693.30, 694.58, 695.86, 697.14, 698.42, 699.70, 700.98, 702.26, 703.54, 704.82, 706.10, 707.38, 708.66, 709.94, 711.22, 712.50, 713.78, 715.06, 716.34, 717.62, 718.90, 720.18, 721.46, 722.74, 724.02, 725.30, 726.58, 727.86, 729.14, 730.42, 731.70, 732.98, 734.26, 735.54, 736.82, 738.10, 739.38, 740.66, 741.94, 743.22, 744.50, 745.78, 747.06, 748.34, 749.62, 750.90, 752.18, 753.46, 754.74, 756.02, 757.30, 758.58, 759.86, 761.14, 762.42, 763.70, 764.98, 766.26, 767.54, 768.82, 770.10, 771.38, 772.66, 773.94, 775.22, 776.50, 777.78, 779.06, 780.34, 781.62, 782.90, 784.18, 785.46, 786.74, 788.02, 789.30, 790.58, 791.86, 793.14, 794.42, 795.70, 796.98, 798.26, 799.54, 800.82, 802.10, 803.38, 804.66, 805.94, 807.22, 808.50, 809.78, 811.06, 812.34, 813.62, 814.90, 816.18, 817.46, 818.74, 819.98, 821.26, 822.54, 823.82, 825.10, 826.38, 827.66, 828.94, 830.22, 831.50, 832.78, 834.06, 835.34, 836.62, 837.90, 839.18, 840.46, 841.74, 843.02, 844.30, 845.58, 846.86, 848.14, 849.42, 850.70, 851.98, 853.26, 854.54, 855.82, 857.10, 858.38, 859.66, 860.94, 862.22, 863.50, 864.78, 866.06, 867.34, 868.62, 869.90, 871.18, 872.46, 873.74, 875.02, 876.30, 877.58, 878.86, 880.14, 881.42, 882.70, 883.98, 885.26, 886.54, 887.82, 889.10, 890.38, 891.66, 892.94, 894.22, 895.50, 896.78, 898.06, 899.34, 900.62, 901.90, 903.18, 904.46, 905.74, 907.02, 908.30, 909.58, 910.86, 912.14, 913.42, 914.70, 915.98, 917.26, 918.54, 919.82, 921.10, 922.38, 923.66, 924.94, 926.22, 927.50, 928.78, 930.06, 931.34, 932.62, 933.90, 935.18, 936.46, 937.74, 939.02, 940.30, 941.58, 942.86, 944.14, 945.42, 946.70, 947.98, 949.26, 950.54, 951.82, 953.10, 954.38, 955.66, 956.94, 958.22, 959.50, 960.78, 962.06, 963.34, 964.62, 965.90, 967.18, 968.46, 969.74, 971.02, 972.30, 973.58, 974.86, 976.14, 977.42, 978.70, 979.98, 981.26, 982.54, 983.82, 985.10, 986.38, 987.66, 988.94, 990.22, 991.50, 992.78, 994.06, 995.34, 996.62, 997.90, 999.18, 1000.46, 1001.74, 1003.02, 1004.30, 1005.58, 1006.86, 1008.14, 1009.42, 1010.70, 1011.98, 1013.26, 1014.54, 1015.82, 1017.10, 1018.38, 1019.66, 1020.94, 1022.22, 1023.50, 1024.78, 1026.06, 1027.34, 1028.62, 1029.90, 1031.18, 1032.46, 1033.74, 1035.02, 1036.30, 1037.58, 1038.86, 1040.14, 1041.42, 1042.70, 1043.98, 1045.26, 1046.54, 1047.82, 1049.10, 1050.38, 1051.66, 1052.94, 1054.22, 1055.50, 1056.78, 1058.06, 1059.34, 1060.62, 1061.90, 1063.18, 1064.46, 1065.74, 1067.02, 1068.30, 1069.58, 1070.86, 1072.14, 1073.42, 1074.70, 1075.98, 1077.26, 1078.54, 1079.82, 1081.10, 1082.38, 1083.66, 1084.94, 1086.22, 1087.50, 1088.78, 1090.06, 1091.34, 1092.62, 1093.90, 1095.18, 1096.46, 1097.74, 1099.02, 1100.30, 1101.58, 1102.86, 1104.14, 1105.42, 1106.70, 1107.98, 1109.26, 1110.54, 1111.82, 1113.10, 1114.38, 1115.66, 1116.94, 1118.22, 1119.50, 1120.78, 1122.06, 1123.34, 1124.62, 1125.90, 1127.18, 1128.46, 1129.74, 1131.02, 1132.30, 1133.58, 1134.86, 1136.14, 1137.42, 1138.70, 1139.98, 1141.26, 1142.54, 1143.82, 1145.10, 1146.38, 1147.66, 1148.94, 1150.22, 1151.50, 1152.78, 1154.06, 1155.34, 1156.62, 1157.90, 1159.18, 1160.46, 1161.74, 1163.02, 1164.30, 1165.58, 1166.86, 1168.14, 1169.42, 1170.70, 1171.98, 1173.26, 1174.54, 1175.82, 1177.10, 1178.38, 1179.66, 1180.94, 1182.22, 1183.50, 1184.78, 1186.06, 1187.34, 1188.62, 1189.90, 1191.18, 1192.46, 1193.74, 1195.02, 1196.30, 1197.58, 1198.86, 1200.14, 1201.42, 1202.70, 1203.98, 1205.26, 1206.54, 1207.82, 1209.10, 1210.38, 1211.66, 1212.94, 1214.22, 1215.50, 1216.78, 1218.06, 1219.34, 1220.62, 1221.90, 1223.18, 1224.46, 1225.74, 1227.02, 1228.30, 1229.58, 1230.86, 1232.14, 1233.42, 1234.70, 1235.98, 1237.26, 1238.54, 1239.82, 1241.10, 1242.38, 1243.66, 1244.94, 1246.22, 1247.50, 1248.78, 1250.06, 1251.34, 1252.62, 1253.90, 1255.18, 1256.46, 1257.74, 1259.02, 1260.30, 1261.58, 1262.86, 1264.14, 1265.42, 1266.70, 1267.98, 1269.26, 1270.54, 1271.82, 1273.10, 1274.38, 1275.66, 1276.94, 1278.22, 1279.50, 1280.78, 1282.06, 1283.34, 1284.62, 1285.90, 1287.18, 1288.46, 1289.74, 1291.02, 1292.30, 1293.58, 1294.86, 1296.14, 1297.42, 1298.70, 1300.02, 1301.30, 1302.58, 1303.86, 1305.14, 1306.42, 1307.70, 1308.98, 1310.26, 1311.54, 1312.82, 1314.10, 1315.38, 1316.66, 1317.94, 1319.22, 1320.50, 1321.78, 1323.06, 1324.34, 1325.62, 1326.90, 1328.18, 1329.46, 1330.74, 1332.02, 1333.30, 1334.58, 1335.86, 1337.14, 1338.42, 1339.70, 1340.98, 1342.26, 1343.54, 1344.82, 1346.10, 1347.38, 1348.66, 1349.94, 1351.22, 1352.50, 1353.78, 1355.06, 1356.34, 1357.62, 1358.90, 1360.18, 1361.46, 1362.74, 1364.02, 1365.30, 1366.58, 1367.86, 1369.14, 1370.42, 1371.70, 1372.98, 1374.26, 1375.54, 1376.82, 1378.10, 1379.38, 1380.66, 1381.94, 1383.22, 1384.50, 1385.78, 1387.06, 1388.34, 1389.62, 1390.90, 1392.18, 1393.46, 1394.74, 1396.02, 1397.30, 1398.58, 1399.86, 1401.14, 1402.42, 1403.70, 1404.98, 1406.26, 1407.54, 1408.82, 1410.10, 1411.38, 1412.66, 1413.94, 1415.22, 1416.50, 1417.78, 1419.06, 1420.34, 1421.62, 1422.90, 1424.18, 1425.46, 1426.74, 1428.02, 1429.30, 1430.58, 1431.86, 1433.14, 1434.42, 1435.70, 1436.98, 1438.26, 1439.54, 1440.82, 1442.10, 1443.38, 1444.66, 1445.94, 1447.22, 1448.50, 1449.78, 1451.06, 1452.34, 1453.62, 1454.90, 1456.18, 1457.46, 1458.74, 1460.02, 1461.30, 1462.58, 1463.86, 1465.14, 1466.42, 1467.70, 1468.98, 1470.26, 1471.54, 1472.82, 1474.10, 1475.38, 1476.66, 1477.94, 1479.22, 1480.50, 1481.78, 1483.06, 1484.34, 1485.62, 1486.90, 1488.18, 1489.46, 1490.74, 1492.02, 1493.30, 1494.58, 1495.86, 1497.14, 1498.42, 1499.70, 1500.98, 1502.26, 1503.54, 1504.82, 1506.10, 1507.38, 1508.66, 1509.94, 1511.22, 1512.50, 1513.78, 1515.06, 1516.34, 1517.62, 1518.90, 1520.18, 1521.46, 1522.74, 1524.02, 1525.30, 1526.58, 1527.86, 1529.14, 1530.42, 1531.70, 1532.98, 1534.26, 1535.54, 1536.82, 1538.10, 1539.38, 1540.66, 1541.94, 1543.22, 1544.50, 1545.78, 1547.06, 1548.34, 1549.62, 1550.90, 1552.18, 1553.46, 1554.74, 1556.02, 1557.30, 1558.58, 1559.86, 1561.14, 1562.42, 1563.70, 1564.98, 1566.26, 1567.54, 1568.82, 1570.10, 1571.38, 1572.66, 1573.94, 1575.22, 1576.50, 1577.78, 1579.06, 1580.34, 1581.62, 1582.90, 1584.18, 1585.46, 1586.74, 1588.02, 1589.30, 1590.58, 1591.86, 1593.14, 1594.42, 1595.70, 1596.98, 1598.26, 1599.54, 1600.82, 1602.10, 1603.38, 1604.66, 1605.94, 1607.22, 1608.50, 1609.78, 1611.06, 1612.34, 1613.62, 1614.90, 1616.18, 1617.46, 1618.74, 1620.02, 1621.30, 1622.58, 1623.86, 1625.14, 1626.42, 1627.70, 1628.98, 1630.26, 1631.54, 1632.82, 1634.10, 1635.38, 1636.66, 1637.94, 1639.22, 1640.50, 1641.78, 1643.06, 1644.34, 1645.62, 1646.90, 1648.18, 1649.46, 1650.74, 1652.02, 1653.30, 1654.58, 1655.86, 1657.14, 1658.42, 1659.70, 1660.98, 1662.26, 1663.54, 1664.82, 1666.10, 1667.38, 1668.66, 1669.94, 1671.22, 1672.50, 1673.78, 1675.06, 1676.34, 1677.62, 1678.90, 1680.18, 1681.46, 1682.74, 1684.02, 1685.30, 1686.58, 1687.86, 1689.14, 1690.42, 1691.70, 1692.98, 1694.26, 1695.54, 1696.82, 1698.10, 1699.38, 1700.66, 1701.94, 1703.22, 1704.50, 1705.78, 1707.06, 1708.34, 1709.62, 1710.90, 1712.18, 1713.46, 1714.74, 1716.02, 1717.30, 1718.58, 1719.86, 1721.14, 1722.42, 1723.70, 1724.98, 1726.26, 1727.54, 1728.82, 1730.10, 1731.38, 1732.66, 1733.94, 1735.22, 1736.50, 1737.78, 1739.06, 1740.34, 1741.62, 1742.90, 1744.18, 1745.46, 1746.74, 1748.02, 1749.30, 1750.58, 1751.86, 1753.14, 1754.42, 1755.70, 1756.98, 1758.26, 1759.54, 1760.82, 1762.10, 1763.38, 1764.66, 1765.94, 1767.22, 1768.50, 1769.78, 1771.06, 1772.34, 1773.62, 1774.90, 1776.18, 1777.46, 1778.74, 1780.02, 1781.30, 1782.58, 1783.86, 1785.14, 1786.42, 1787.70, 1788.98, 1790.26, 1791.54, 1792.82, 1794.10, 1795.38, 1796.66, 1797.94, 1799.22, 1800.50, 1801.78, 1803.06, 1804.34, 1805.62, 1806.90, 1808.18, 1809.46, 1810.74, 1812.02, 1
```



```
In [12]: # montando a matriz de dados usada no sistema linear
```

```
A = np.ones((len(height), 2))
```

```
A[:, 0] = height
```

```
print(A)
```

```
[[1.47 1. ]
 [1.5  1. ]
 [1.52 1. ]
 [1.55 1. ]
 [1.57 1. ]
 [1.6  1. ]
 [1.63 1. ]
 [1.65 1. ]
 [1.68 1. ]
 [1.7  1. ]
 [1.73 1. ]
 [1.75 1. ]
 [1.78 1. ]
 [1.8  1. ]
 [1.83 1. ]]
```

```
In [13]: # calculando a pseudo-inversa da matriz A
```

```
p_A = np.linalg.pinv(A)
```

```
print(p_A)
```

```

print(p_A.shape)
print(weight.shape)

[[-0.98890673 -0.82469712 -0.71522405 -0.55101445 -0.44154138 -0.27733178
  -0.11312217 -0.0036491  0.1605605  0.27003357  0.43424318  0.54371625
   0.70792585  0.81739892  0.98160852]
 [ 1.69902204  1.42796672  1.24726317  0.97620785  0.79550431  0.52444899
   0.25339367  0.07269012 -0.1983652  -0.37906875 -0.65012407 -0.83082762
  -1.10188294 -1.28258648 -1.5536418 ]]
(2, 15)
(15,)

```

agora resolvo o sistema multiplicando $\beta = Y * X^{-1}$

```

In [14]: beta = np.matmul(p_A, weight)
print(beta)
#print(np.matmul(weight, p_A))

[ 61.27218654 -39.06195592]

```

```

In [15]: # plotando resultado junto com os dados

```

```

plt.plot(height, weight, 'o', label='Dados originais', markersize=10)
plt.plot(height, beta[0]*height + beta[1], 'r', label='Modelo ajustado')
plt.legend()
plt.show()

```

