Simple Linear Regression

Kesav Adithya Venkidusamy

Bellevue university - Master of Science in Data Science

Course Name: DSC520-T301 Statistics for Data Science (2221-1)

Assignment: Week 8.2 Assignment

Instructor: Dr Richard Bushart

Due Date: 10/31/2021

Assignment 06

```
# Assignment: ASSIGNMENT 6
# Name: Venkidusamy, Kesav Adithya
# Date: 2021-10-23

> ## Set the working directory to the root of your DSC 520 directory
> setwd("E:/Personal/Bellevue University/Course/github/dsc520")
> ## Load the `data/r4ds/heights.csv` to
> heights_df <- read.csv("data/r4ds/heights.csv")
> summary(heights_df)
      earn            height          sex                ed             age
 Min.   :   200   Min.   :57.50   Length:1192        Min.   : 3.0   Min.   :18.00
 1st Qu.: 10000   1st Qu.:64.01   Class :character   1st Qu.:12.0   1st Qu.:29.00
 Median : 20000   Median :66.45   Mode  :character   Median :13.0   Median :38.00
 Mean   : 23155   Mean   :66.92                      Mean   :13.5   Mean   :41.38
 3rd Qu.: 30000   3rd Qu.:69.85                      3rd Qu.:16.0   3rd Qu.:51.00
 Max.   :200000   Max.   :77.05                      Max.   :18.0   Max.   :91.00
    race
 Length:1192
 Class :character
 Mode  :character

> ## Load the ggplot2 library
> library(ggplot2)
> ## Fit a linear model using the `age` variable as the predictor and `earn` as the outcome
> age_lm <-  lm(earn ~ age, data = heights_df)
```

```
> ## View the summary of your model using `summary()`
> summary(age_lm)

Call:
lm(formula = earn ~ age, data = heights_df)

Residuals:
   Min     1Q Median     3Q    Max
-25098 -12622  -3667   6883 177579

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 19041.53    1571.26  12.119  < 2e-16 ***
age            99.41      35.46   2.804  0.00514 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19420 on 1190 degrees of freedom
Multiple R-squared:  0.006561,          Adjusted R-squared:  0.005727
F-statistic:  7.86 on 1 and 1190 DF,  p-value: 0.005137

> ## Creating predictions using `predict()`
> age_predict_df <- data.frame(earn = predict(age_lm, heights_df), age=heights_df$age)
> ## Plot the predictions against the original data
> ggplot(data = heights_df, aes(y = earn, x = age)) +
+   geom_point(color='blue') +
+   geom_line(color='red',data = age_predict_df, aes(y=earn, x=age))
```

```
> mean_earn <- mean(heights_df$earn)
> mean_earn
[1] 23154.77
> ## Corrected Sum of Squares Total
> sst <- sum((mean_earn - heights_df$earn)^2)
> sst
[1] 451591883937
> ## Corrected Sum of Squares for Model
> ssm <- sum((mean_earn - age_predict_df$earn)^2)
> ssm
[1] 2963111900
> ## Residuals
> residuals <- heights_df$earn - age_predict_df$earn
> residuals
  [1]  26485.21417  35192.93914   8075.70651  21912.54868  28081.64879 -12626.07618
  [7]   5087.59108   8385.80839 -19129.04732   5373.92382 -18972.90126   7578.67765
 [13]  -9725.48195 -12111.22046   -520.72812  -7807.06086  18075.70651  20584.61994
 [19] -17508.84355  30479.27188 -19111.22046  -7129.04732 -15728.45309  16882.83725
 [25]  10485.21417 -12520.72812   1994.12760  27181.05456  18678.08342  -6526.67041
 [31]   2678.08342  52081.64879   4876.89496  -9626.07618 -19295.17629   7876.89496
 [37]  -3707.65509   5373.92382 -22502.90126   2976.30074    882.83725  10075.70651
 [43] -12023.69926 -10129.04732  -3522.51081  -1924.29349 -17330.83001 -11608.24932
 [49]   -502.90126  -2117.16275  10087.59108 -19502.90126  -4824.88772  12777.48919
 [55] -16830.83001 -10508.84355 -14707.65509   3075.70651  -5725.48195  -5824.88772
 [61]  -6815.97429 -21801.11858 -20370.62623   -918.35121 -24110.03201   4274.51805
 [67] -18629.04732  -5918.35121 -13228.45309 -13801.11858 -13713.59738 -15701.71280
 [73]  -7918.35121 -20518.35121   7777.48919   7678.08342 -12818.94544 -16626.07618
```

[79] -13304.08972 -5105.27818 -2620.13389 -1327.85887 -19829.64155 -21522.51081

[85] -1725.48195 -11228.45309 3701.85257 15391.75068 -2888.63977 -16894.58206

[91] 34695.91028 -7602.30703 -6281.91658 -7023.69926 -11327.85887 9684.02571

[97] -4023.69926 -7719.53966 -228.45309 -2315.97429 14280.46034 14584.61994

[103] -6315.97429 9976.30074 2692.93914 -6915.38006 2479.27188 3684.02571

[109] -16824.88772 47181.05456 -18017.75698 -16427.26464 -1626.07618 -17129.04732

[115] 37479.27188 -17123.10504 7181.05456 47479.27188 26882.83725 21081.64879

[121] 6684.02571 -13017.75698 -216.56852 19795.31605 -7123.10504 -22099.33589

[127] -5123.10504 7777.48919 5476.30074 -20212.52435 -6608.24932 19584.61994

[133] 8584.61994 -14608.24932 36485.21417 -3011.81469 -18828.45309 -20824.88772

[139] 5274.51805 -15224.88772 -7315.97429 68373.92382 -15216.56852 -2222.51081

[145] -5830.83001 -13005.87240 385.80839 -2918.35121 -4614.19161 -13912.40892

[151] 17578.67765 -403.49549 3572.73536 2777.48919 -6105.27818 20186.99685

[157] -8912.40892 -5620.13389 -608.24932 280.46034 -13397.55320 41385.80839

[163] -18304.08972 14695.91028 -8315.97429 -1526.67041 -4011.81469 -3216.56852

[169] 1087.59108 26280.46034 -502.90126 -13707.65509 -6818.94544 11684.02571

[175] 101485.21417 -713.59738 -4824.88772 3882.83725 47777.48919 11286.40262

[181] -15204.68395 12081.64879 -7818.94544 -10918.35121 -17801.11858 -16011.81469

[187] 10988.18531 20888.77954 -6228.45309 -11695.77052 -111.22046 2777.48919

[193] 590.56222 -1824.88772 2572.73536 20385.80839 46181.05456 38584.61994

[199] 9777.48919 -2321.91658 9479.27188 988.18531 146485.21417 12678.08342

[205] 17181.05456 10976.30074 -3228.45309 6684.02571 3876.89496 -16129.04732

[211] -6298.14743 -5023.69926 10175.11228 -6228.45309 25491.15645 -14918.35121

[217] 15888.77954 16882.83725 10328.67765 -5900.52435 13081.64879 -19403.49549

[223] -10421.32235 34695.91028 16584.61994 20181.05456 22578.67765 -16918.35121

[229] -19192.79937 -2918.35121 -4327.85887 -20222.51081 43175.11228 26882.83725

[235] -11818.94544 12976.30074 4678.08342 -21605.27818 19280.46034 -1924.29349

[241] -6924.29349 -11129.04732 -20795.17629 -4327.85887 9497.09874 6175.11228

[247] -7321.91658 -1427.26464 -1924.29349 -12716.56852 -8228.45309 -14210.62623

[253] -19719.53966 2075.70651 -5520.72812 -12813.00315 8982.24302 -4626.07618

[259] -19520.72812 -19029.04732 -19229.32235 -4626.07618 8175.11228 -7222.51081

[265] 783.43148 -15701.71280 25689.96799 25689.96799 -3912.40892 5590.56222

[271] -1514.78583 5572.73536 -18818.94544 -4126.07618 -5324.88772 5081.64879

[277] 27988.18531 -6924.29349 -3222.51081 5672.14113 -9807.06086 -7321.91658

[283] -10008.84355 1976.30074 -4620.13389 -21801.11858 -20707.65509 -21920.13389

[289] -4087.45132 4081.64879 -10520.72812 -1713.59738 7379.86611 9894.72182

[295] -3216.56852 9777.48919 -16123.10504 -10918.35121 -12123.10504 -19830.83001

[301] -14099.33589 -6620.13389 2678.08342 3274.51805 9976.30074 11684.02571

[307] -15824.88772 56584.61994 -1526.67041 -21818.94544 5181.05456 4672.14113

[313] 5572.73536 -5117.16275 -22906.46663 -5602.30703 -17813.00315 -304.08972

[319] -9924.29349 -12707.65509 3403.63525 -17228.45309 3280.46034 479.27188

[325] -1321.91658 -204.68395 -8496.95897 -19415.38006 2671.54691 11982.24302

[331] 3976.30074 2379.86611 1774.51805 -9228.45309 -6029.64155 -18327.85887

[337] -19701.71280 -11900.52435 -7123.10504 151186.99685 -20526.67041 -15304.08972

[343] 17807.20063 -12093.39360 -3117.16275 12976.30074 18175.11228 8801.25834

[349] 25590.56222 75292.34491 13175.11228 1479.27188 11684.02571 11602.44680

[355] 5789.37377 -17918.35121 125181.05456 7777.48919 -15623.10504 -19719.53966

[361] -2900.52435 16882.83725 7081.64879 -8719.53966 -7421.32235 3379.86611

[367] -13526.67041 2274.51805 -20577.71280 -11526.67041 -10222.51081 27578.67765

[373] 1373.92382 17578.67765 -15626.07618 -17023.69926 87280.46034 18777.48919

[379] -3795.17629 -327.85887 -17924.29349 2379.86611 6385.80839 -10111.22046

[385] -16520.72812 -9614.19161 21274.51805 2678.08342 14397.69297 41982.24302

[391] -6222.51081 -18099.33589 -2918.35121 -12198.74166 18602.44680 -18198.74166

[397] -9824.88772 -14123.10504 -6222.51081 -20298.14743 -3222.51081 -1918.35121

[403] 19385.80839 13870.95268 -18496.95897 -2117.16275 -16223.69926 -6614.19161

[409] 87.59108 -21099.33589 -10327.85887 -11725.48195 17379.86611 16286.40262

[415] 2075.70651 2373.92382 -2321.91658 35988.18531 -12713.59738 -18017.75698

[421] 4976.30074 65882.83725 18175.11228 -6626.07618 -20029.64155 -3005.87240

[427] -11813.00315 175.11228 15391.75068 16584.61994 10689.96799 -14824.88772

[433] -8719.53966 -9608.24932 -16421.32235 -12117.16275 -12123.10504 -3117.16275

[439] -19298.14743 14385.80839 4192.93914 -3415.38006 -16707.65509 11888.77954

[445] -21915.38006 20075.70651 3771.54691 -8029.64155 5888.77954 -14327.85887

[451] 7186.99685 -8216.56852 26087.59108 -16029.64155 -20230.83001 -10111.22046

[457] 17479.27188 1379.86611 572.73536 -2620.13389 -2706.94544 -13701.71280

[463] 17982.24302 -11011.81469 37876.89496 2473.32959 15379.86611 -18725.48195

[469] -9023.69926 -16129.04732 -2719.53966 175.11228 3491.15645 -1117.16275

[475] 6584.61994 7777.48919 7876.89496 6099.47565 2684.02571 -421.32235

[481] 2678.08342 -10427.26464 -11310.03201 -13216.56852 -11514.78583 -18403.49549

[487] 34280.46034 13280.46034 7075.70651 -12930.23578 -17830.83001 -3117.16275

[493] 8292.34491 -10801.11858 -6321.91658 -21795.17629 -15315.97429 9578.67765

[499] -6927.26464 -8924.29349 -14526.67041 24497.09874 -10520.72812 -6105.27818

[505] -15508.84355 -2520.72812 32479.27188 -13017.75698 -7930.23578 -2123.10504

[511] 8373.92382 -10421.32235 3678.08342 -9225.48195 -13924.29349 -7222.51081

[517] 43473.32959 3870.95268 -13129.04732 -6129.04732 -11017.75698 12280.46034

[523] 17876.89496 7771.54691 -7123.10504 -2620.13389 -22204.68395 -19930.23578

[529] -13327.85887 -15830.83001 3274.51805 -4707.65509 7578.67765 -10520.72812

[535] -12222.51081 -4930.23578 23075.70651 15590.56222 -2093.39360 -5210.62623

[541] -4423.69926 -18701.71280 7777.48919 -5713.59738 1286.40262 -19204.68395

[547] 46783.43148 -7421.32235 2678.08342 13175.11228 -3427.26464 5976.30074

[553] -6129.04732 175.11228 12976.30074 -5614.19161 12280.46034 -4729.04732

[559] -719.53966 5192.93914 -4725.48195 2379.86611 -14111.22046 17280.46034

[565] -12804.08972 11485.21417 81087.59108 76982.24302 -3111.22046 -14608.24932

[571] 11373.92382   4473.32959 33204.82371 -5514.78583 -10105.27818 -16029.64155

[577] 36783.43148 -13900.52435   5181.05456 18684.02571 -20321.91658 -2321.91658

[583]   4982.24302 -15204.68395   8876.89496 -5719.53966 -6918.35121 12181.05456

[589] -18626.07618 -6918.35121 11783.43148   4678.08342   3075.70651 16882.83725

[595] 21684.02571 -7327.85887   2280.46034 16783.43148 11479.27188 17578.67765

[601] 13175.11228 -12111.22046   6684.02571   -17.75698 -20023.69926 -1526.67041

[607] -15204.68395 36684.02571 -11117.16275   -105.27818 -4298.14743 -13526.67041

[613] -19129.04732 -3026.07618 -1029.64155 77280.46034 26982.24302   3783.43148

[619] -23807.06086 -3029.64155 -4327.85887 -16397.55320 -8520.72812 -18402.30703

[625] 10081.64879   578.67765 -13117.16275   572.73536 -7930.23578 10876.89496

[631] 13379.86611 -12222.51081 -19005.87240 -11315.97429 -13713.59738 -2514.78583

[637]   3777.48919 -2011.81469 -1409.43778 -1427.26464 -12818.94544 -11099.33589

[643] -14427.26464 -7818.94544   3578.67765 -12123.10504 -18029.64155   3882.83725

[649]   6473.32959 -6725.48195 32976.30074 -1626.07618 -14915.38006   7976.30074

[655] -2123.10504 -4496.95897   3373.92382 -18228.45309 -7912.40892 -12918.35121

[661] -6824.88772   4385.80839   8979.27188 70789.37377 14982.24302   6882.83725

[667] 12683.43148 -9228.45309 -11526.67041 21584.61994 15391.75068   -225.48195

[673] -11204.68395 13590.56222 -19830.83001 -12298.14743 -7123.10504   783.43148

[679] -16827.85887 -7304.08972 -7029.64155 -11824.88772 -12722.51081 -14602.30703

[685] -8701.71280 -15824.88772 -22307.06086 16882.83725   2081.64879   2994.12760

[691]   9485.21417 -21403.49549 -17923.69926 -2192.79937 10175.11228 -11807.06086

[697] -13725.48195 -5818.94544 -11695.77052 -4222.51081 -9427.26464 -16912.40892

[703] -20028.45309 -7129.04732 -6017.75698   6783.43148   689.96799   -321.91658

[709] -6129.04732 -1087.45132 -1626.07618 -13421.32235 -5023.69926 -21814.19161

[715] -5614.19161   1783.43148 -3117.16275 51982.24302 -19517.75698   9900.66411

[721]   8473.32959 -1795.17629 -20626.07618 34397.69297   8403.63525 -13321.91658

[727] 7608.38908 -13807.06086 -12210.62623 6684.02571 -5707.65509 -16105.27818

[733] 8075.70651 -13222.51081 -7023.69926 -14123.10504 -16496.95897 10280.46034

[739] -9824.88772 -4719.53966 4894.72182 -9722.51081 -12620.13389 -17599.33589

[745] -2614.19161 -5695.77052 -526.67041 -19210.62623 -620.13389 8373.92382

[751] 98192.93914 -15514.78583 -19526.67041 -17604.68395 -11526.67041 11186.99685

[757] -9707.65509 -17228.45309 -5427.26464 -22121.91658 1473.32959 -5725.48195

[763] -17912.40892 -23014.19161 5578.67765 -21602.30703 -7520.72812 -6526.67041

[769] 2976.30074 -6626.07618 16584.61994 5888.77954 -16924.29349 12578.67765

[775] -6123.10504 4982.24302 -3216.56852 -1023.69926 -6216.56852 -11.81469

[781] -1496.95897 16385.80839 -12930.23578 -4514.78583 19099.47565 -17321.91658

[787] 982.24302 -15930.23578 10578.67765 2572.73536 -15327.85887 65491.15645

[793] 33286.40262 -10609.43778 7280.46034 -8017.75698 -11496.95897 9075.70651

[799] 9982.24302 4894.72182 -10818.94544 -16228.45309 -3906.46663 4385.80839

[805] -10123.10504 -4614.19161 -16520.72812 -19705.87240 -6824.88772 -6614.19161

[811] -22095.17629 -16002.90126 -19528.45309 -16723.10504 -12123.10504 -2292.20515

[817] 11684.02571 4584.61994 18982.24302 20192.93914 30888.77954 13976.30074

[823] 2876.89496 -2824.88772 13988.18531 -7719.53966 -22991.01669 74274.51805

[829] -12713.59738 -13421.32235 -16427.26464 25888.77954 -17415.38006 78771.54691

[835] -7114.19161 8982.24302 6882.83725 -13216.56852 -5930.23578 -20792.20515

[841] 695.91028 -7906.46663 -2906.46663 -13117.16275 25192.93914 -2520.72812

[847] -21723.10504 -8906.46663 -15602.30703 -10807.06086 -16029.64155 17578.67765

[853] -19508.84355 1286.40262 -12111.22046 35578.67765 -3626.07618 -19830.83001

[859] -2713.59738 -14204.68395 -9620.13389 -7789.23400 2479.27188 -22715.97429

[865] 3373.92382 14175.11228 -7321.91658 -20611.22046 -3416.32235 37590.56222

[871] 1783.43148 25789.37377 14894.72182 4373.92382 -1813.00315 -1918.35121

[877] -19394.90126 4596.50451 -20795.17629 5075.70651 -19216.56852 -21782.74166

[883] -22689.82823 21982.24302 -5514.78583 -15228.45309 -7427.26464 13572.73536

```
 [889]  -6924.29349  -9924.29349 -22713.59738   9678.08342 -19430.23578 -18327.85887
 [895]  11783.43148  46186.99685   -824.88772  29169.16999  -6129.04732 -13520.72812
 [901] -11526.67041   6286.40262  -5129.04732  -4111.22046  75590.56222 -13315.97429
 [907]   9695.91028 -15123.10504 -13725.48195 -13011.81469   1286.40262   1485.21417
 [913]   6186.99685  25578.67765  -6614.19161   4572.73536 -15105.27818  35689.96799
 [919]     75.70651 -21092.20515   1777.48919  -3017.75698 -15023.69926   6186.99685
 [925]   1175.11228  -3017.75698 -18924.29349   6882.83725  -9327.85887  17181.05456
 [931]  -6327.85887  12485.21417 -14310.03201 -10813.00315 -11725.48195   7280.46034
 [937]    777.48919  -5801.11858  -8192.79937  -5602.30703   2695.91028 -13993.98783
 [943] -14496.95897  -1327.85887   2976.30074    789.37377  -5017.75698 -19801.11858
 [949]  -5327.85887  22181.05456   -900.52435 -10818.94544 -18830.83001 -10830.83001
 [955]  -8620.13389  -7123.10504 -12918.35121 -24203.49549 -10204.68395 -15403.49549
 [961] -18830.83001 -18321.91658 -12222.51081 -19827.85887  -3626.07618 -17427.26464
 [967]   6373.92382 -11228.45309 -18228.45309  -9129.04732  -4514.78583  13081.64879
 [973] -17906.46663  12578.67765  16684.02571  -6526.67041 -15315.97429    491.15645
 [979]   5181.05456   -520.72812     72.73536 -18327.85887   1175.11228  -8298.14743
 [985]  -6099.33589  -7614.19161  -6626.07618  30777.48919   4274.51805 -17824.88772
 [991] -14906.46663  18888.77954  -5321.91658 -18818.94544  26186.99685  30888.77954
 [997]   4882.83725 -18813.00315  -5912.40892  -4204.68395
 [ reached getOption("max.print") -- omitted 192 entries ]
> ## Sum of Squares for Error
> sse <- sum(residuals^2)
> sse
[1] 448628772037
> ## R Squared R^2 = SSM\SST
> r_squared <- ssm/sst
> ## Number of observations
```

```
> n <- nrow(heights_df)
> ## Number of regression parameters
> p <- 2
> ## Corrected Degrees of Freedom for Model (p-1)
> dfm <- p-1
> dfm
[1] 1
> ## Degrees of Freedom for Error (n-p)
> dfe <- n-p
> dfe
[1] 1190
> ## Corrected Degrees of Freedom Total:   DFT = n - 1
> dft <- n-1
> dft
[1] 1191
> ## Mean of Squares for Model:   MSM = SSM / DFM
> msm <- ssm/dfm
> msm
[1] 2963111900
> ## Mean of Squares for Error:   MSE = SSE / DFE
> mse <- sse/dfe
> mse
[1] 376998968
> ## Mean of Squares Total:   MST = SST / DFT
> mst <- sst/dft
> mst
[1] 379170348
```

```
> ## F Statistic F = MSM/MSE
> f_score <- msm/mse
> f_score
[1] 7.859735
> ## Adjusted R Squared R2 = 1 - (1 - R2)(n - 1) / (n - p)
> adjusted_r_squared <- 1 - (1 - r_squared)*(n-1) / (n - p)
> adjusted_r_squared
[1] 0.005726659
> ## Calculate the p-value from the F distribution
> p_value <- pf(f_score, dfm, dft, lower.tail=F)
> p_value
[1] 0.005136826
```