Hanford Washington Problem

The Data

In[1]:=
```
index = {2.5, 2.6, 3.4, 1.3, 1.6, 3.8, 11.6, 6.4, 8.3};
```

In[2]:=
```
deaths = {147, 130, 130, 114, 138, 162, 208, 178, 210};
```

In[3]:=
```
list1 = Transpose[{index, deaths}];
```

In[4]:=
```
Text[Grid[Prepend[list1, {"index", "deaths"}],
  Alignment → Center, Dividers → {2 → True, 2 → True}, Spacings → {1, 1}]]
```

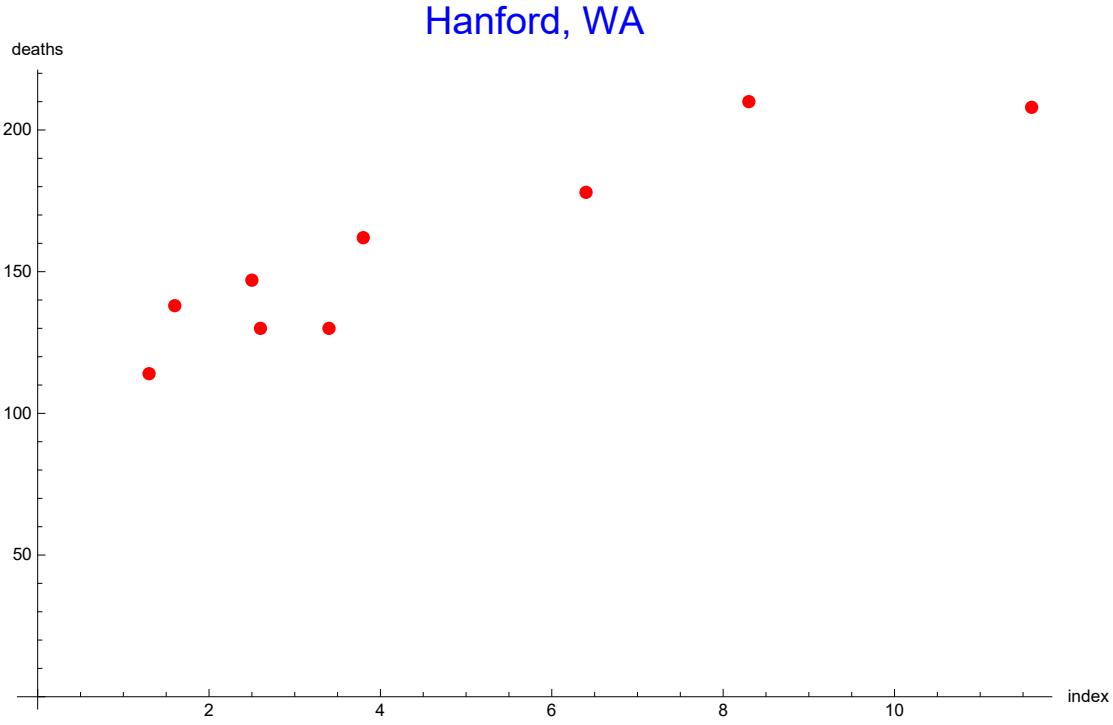
Out[4]=

index	deaths
2.5	147
2.6	130
3.4	130
1.3	114
1.6	138
3.8	162
11.6	208
6.4	178
8.3	210

In[5]:=

```
lp1 = ListPlot[list1, AxesLabel → {"index", "deaths"},
  ImageSize → Large, PlotStyle → {Red, Point : 10 },
  AxesOrigin → {0, 0},
  PlotLabel → Style ["Hanford, WA", 20, Blue]]
```

Out[5]=



Two-Point Line

In[6]:=
```
myTwoPointLine[x_] :=
 (list1[[9]][[2]] - list1[[4]][[2]])/(list1[[9]][[1]] - list1[[4]][[1]]) (x - list1[[9]][[1]]) + list1[[9]][[2]]
```

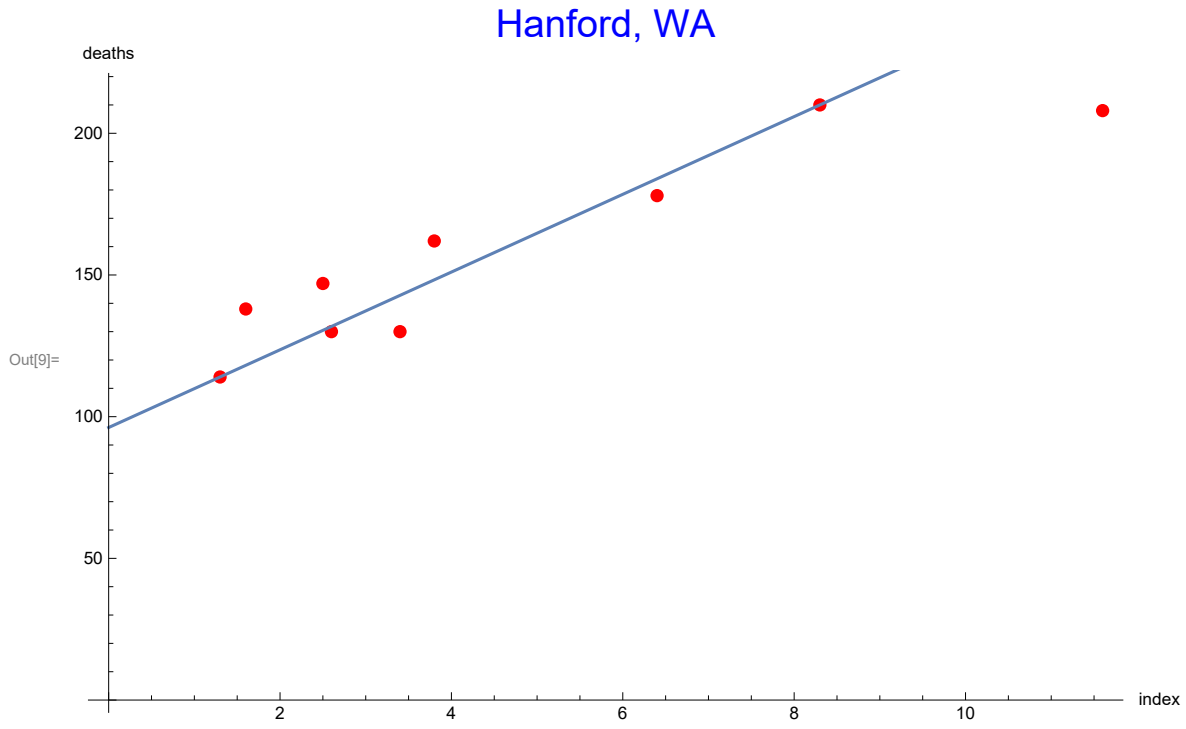
In[7]:=
```
210 + 13.714285714285714` (-8.3` + x)
```

Out[7]=
```
210 + 13.7143 (-8.3 + x)
```

In[8]:=
```
p1 = Plot[myTwoPointLine[x], {x, 0, 12}];
```

In[9]:=
```
Show[lp1, p1]
```

Out[9]=



Two-Point Line Residuals

In[10]:=
```
twoPointLineIndexResiduals = {(list1[[1]][[2]] - myTwoPointLine[list1[[1]][[1]]]),
  list1[[2]][[2]] - myTwoPointLine[list1[[2]][[1]]],
  (list1[[3]][[2]] - myTwoPointLine[list1[[3]][[1]]]),
  list1[[4]][[2]] - myTwoPointLine[list1[[4]][[1]]],
  (list1[[5]][[2]] - myTwoPointLine[list1[[5]][[1]]]),
  (list1[[6]][[2]] - myTwoPointLine[list1[[6]][[1]]]),
  (list1[[7]][[2]] - myTwoPointLine[list1[[7]][[1]]]),
  (list1[[8]][[2]] - myTwoPointLine[list1[[8]][[1]]]),
  (list1[[9]][[2]] - myTwoPointLine[list1[[9]][[1]]])}
```

Out[10]=
```
{16.5429, -1.82857, -12.8, 0., 19.8857, 13.7143, -47.2571, -5.94286, 0.}
```

In[11]:=
```
list2 = Transpose[{index, twoPointLineIndexResiduals}];
```

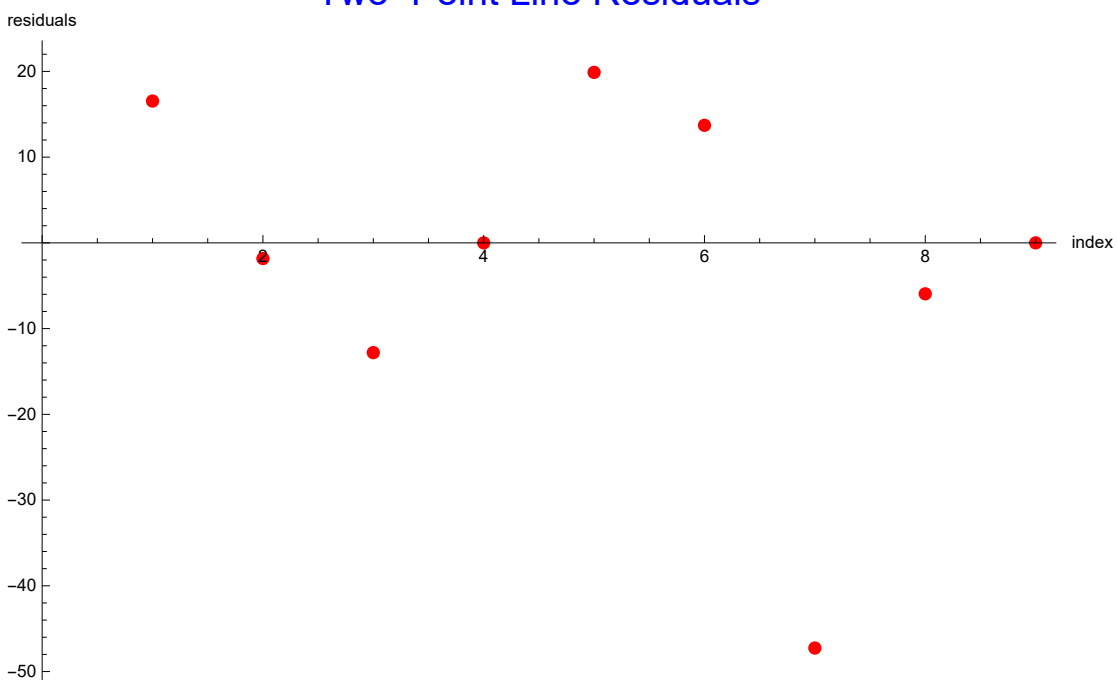
In[12]:=
```
Text[Grid[Prepend[list2, {"index", "residuals"}],
  Alignment → Center, Dividers → {2 → True, 2 → True}, Spacings → {1, 1}]]
```

Out[12]=

index	residuals
2.5	16.5429
2.6	−1.82857
3.4	−12.8
1.3	0.
1.6	19.8857
3.8	13.7143
11.6	−47.2571
6.4	−5.94286
8.3	0.

In[93]:=
```
residualsPlot1 = ListPlot[twoPointLineIndexResiduals,
  AxesLabel → {"index", "residuals"},
  ImageSize → Large, PlotStyle → {Red, Point : 10 },
  AxesOrigin → {0, 0},
  PlotLabel → Style ["Two-Point Line Residuals", 20, Blue]]
```

Out[93]=



Two-Point Line Sum of Residuals Squared

In[39]:=
```
twoPointLineIndexResidualsSquared =
 twoPointLineIndexResiduals * twoPointLineIndexResiduals
```

Out[39]= {273.666, 3.34367, 163.84, 0., 395.442, 188.082, 2233.24, 35.3176, 0.}

In[42]:=
```
sumOfResidualsSquared1 = Total[twoPointLineIndexResidualsSquared]
```

Out[42]= 3292.93

Median-Median Line

Summary Points

In[16]:=
```
sortedlist1 = Sort[list1]
```

Out[16]=
```
{{1.3, 114}, {1.6, 138}, {2.5, 147}, {2.6, 130},
 {3.4, 130}, {3.8, 162}, {6.4, 178}, {8.3, 210}, {11.6, 208}}
```

First, since there are 9 points, I will split them into three groups of three so the points with the lowest three x values are in a group, the ones with the highest three x values are in a group and the ones with the middle three x values are in a group.

In[17]:=
```
list1min = {sortedlist1[[1]], sortedlist1[[2]], sortedlist1[[3]]}
```

Out[17]=
```
{{1.3, 114}, {1.6, 138}, {2.5, 147}}
```

In[18]:=
```
list1mid = {sortedlist1[[4]], sortedlist1[[5]], sortedlist1[[6]]}
```

Out[18]=
```
{{2.6, 130}, {3.4, 130}, {3.8, 162}}
```

In[19]:=
```
list1max = {sortedlist1[[7]], sortedlist1[[8]], sortedlist1[[9]]}
```

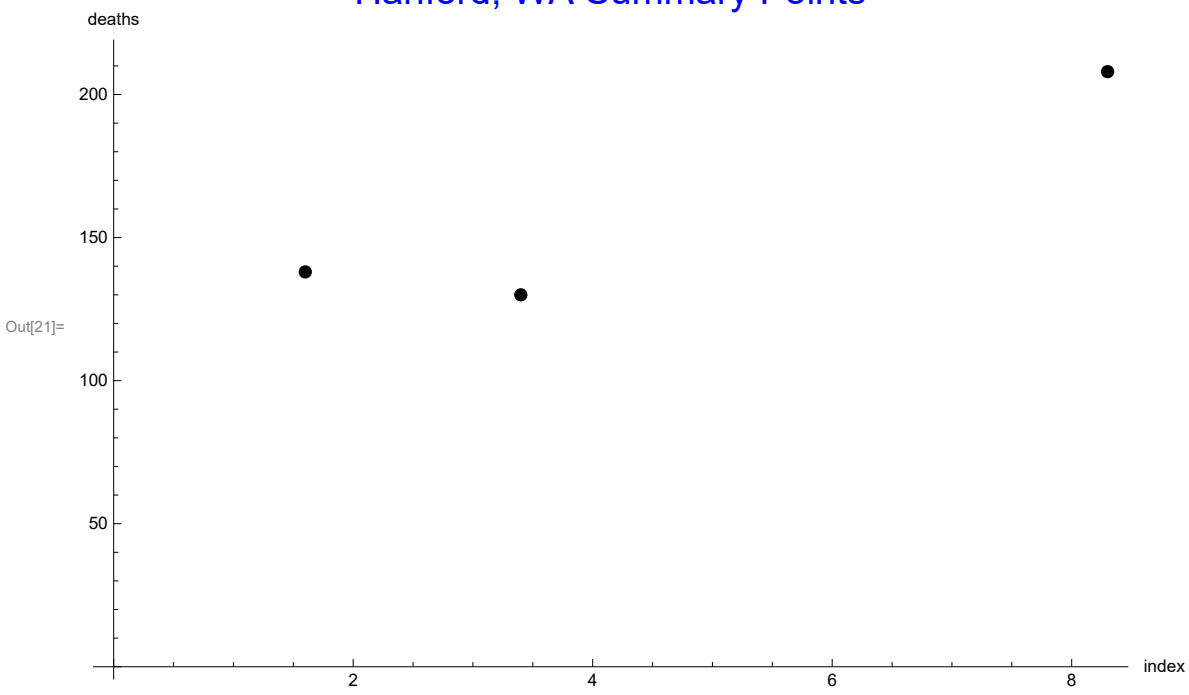
Out[19]=
```
{{6.4, 178}, {8.3, 210}, {11.6, 208}}
```

In[20]:=
```
summaryPoints = {{list1min[[2]][[1]], list1min[[2]][[2]]},
  {list1mid[[2]][[1]], list1mid[[2]][[2]]}, {list1max[[2]][[1]], list1max[[3]][[2]]}}
```

Out[20]=
```
{{1.6, 138}, {3.4, 130}, {8.3, 208}}
```

In[21]:=
```
summaryPointsPlot =
 ListPlot[summaryPoints , AxesLabel → {"index", "deaths"},
  ImageSize → Large, PlotStyle → {Black, Point : 60},
  AxesOrigin → {0, 0},
  PlotLabel → Style ["Hanford, WA Summary Points", 20, Blue]]
```

Out[21]=



Methods for Obtaining the Median-Median Line

Method 1: Point-Slope Form Derivation

In[22]:=
```
myPreMedianMedianLine[x_] :=
 (summaryPoints[[3]][[2]] - summaryPoints[[1]][[2]])/(summaryPoints[[3]][[1]] - summaryPoints[[1]][[1]]) (x - summaryPoints[[3]][[1]]) +
  summaryPoints[[3]][[2]]
```

In[23]:=
```
208 + 10.44776119402985` (-8.3` + x)
```

Out[23]=
$$208 + 10.4478\,(-8.3 + x)$$

In[89]:=
```
myMedianMedianLine[x] = myPreMedianMedianLine[x] -
   ((1/3) * (myPreMedianMedianLine[summaryPoints[[2]][[1]]] - summaryPoints[[2]][[2]]))
```

Out[89]=
$$199.065 + 10.4478\,(-8.3 + x)$$

Method 2: Slope-Intercept Form Derivation

In[125]:=
```
myMedianMedianLineSlope = (summaryPoints[[3]][[2]] - summaryPoints[[1]][[2]])/(summaryPoints[[3]][[1]] - summaryPoints[[1]][[1]])
```

Out[125]=
$$10.4478$$

In[126]:=
```
Solve[summaryPoints[[3]][[2]] - ((1/3) *
      (myPreMedianMedianLine[summaryPoints[[2]][[1]]] - summaryPoints[[2]][[2]])) ==
  myMedianMedianLineSlope * summaryPoints[[3]][[1]] + myMedianMedianLineIntercept]
```

Out[126]=
$$\{\{\text{myMedianMedianLineIntercept} \to 112.348\}\}$$

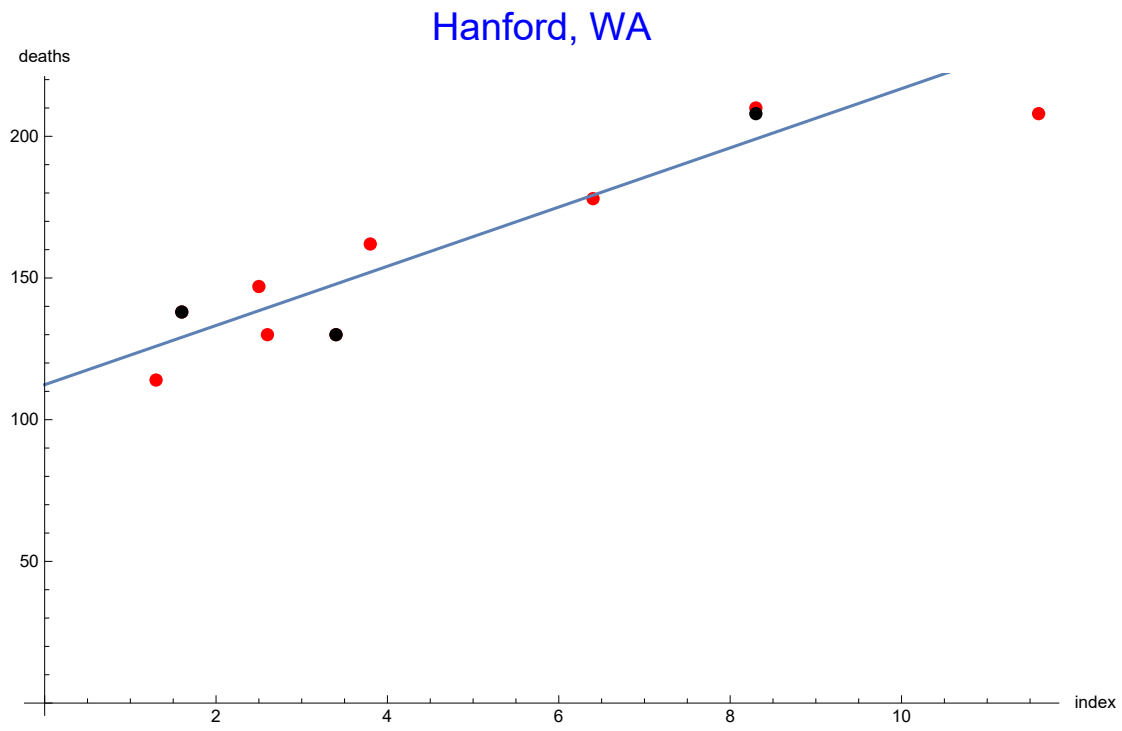
In[127]:=
```
myMedianMedianLine[x] = myMedianMedianLineSlope * x + 112.34825870646766`
```

Out[127]=
$$112.348 + 10.4478\,x$$

In[25]:=
```
p2 = Plot[myMedianMedianLine[x], {x, 0, 12}];
```

In[26]:=
```
Show[lp1, p2, summaryPointsPlot]
```

Out[26]=



Median-Median Line Residuals

In[30]:=
```
myMedianMedianLineIndexResiduals =
 {(list1[[1]][[2]] - myMedianMedianLine[list1[[1]][[1]]]),
  list1[[2]][[2]] - myMedianMedianLine[list1[[2]][[1]]],
  (list1[[3]][[2]] - myMedianMedianLine[list1[[3]][[1]]]),
  list1[[4]][[2]] - myMedianMedianLine[list1[[4]][[1]]],
  (list1[[5]][[2]] - myMedianMedianLine[list1[[5]][[1]]]),
  (list1[[6]][[2]] - myMedianMedianLine[list1[[6]][[1]]]),
  (list1[[7]][[2]] - myMedianMedianLine[list1[[7]][[1]]]),
  (list1[[8]][[2]] - myMedianMedianLine[list1[[8]][[1]]]),
  (list1[[9]][[2]] - myMedianMedianLine[list1[[9]][[1]]])}
```

Out[30]=
```
{8.53234, -9.51244, -17.8706, -11.9303, 8.93532, 9.95025, -25.5423, -1.21393, 10.9353}
```

In[31]:=
```
list3 = Transpose[{index, myMedianMedianLineIndexResiduals}];
```

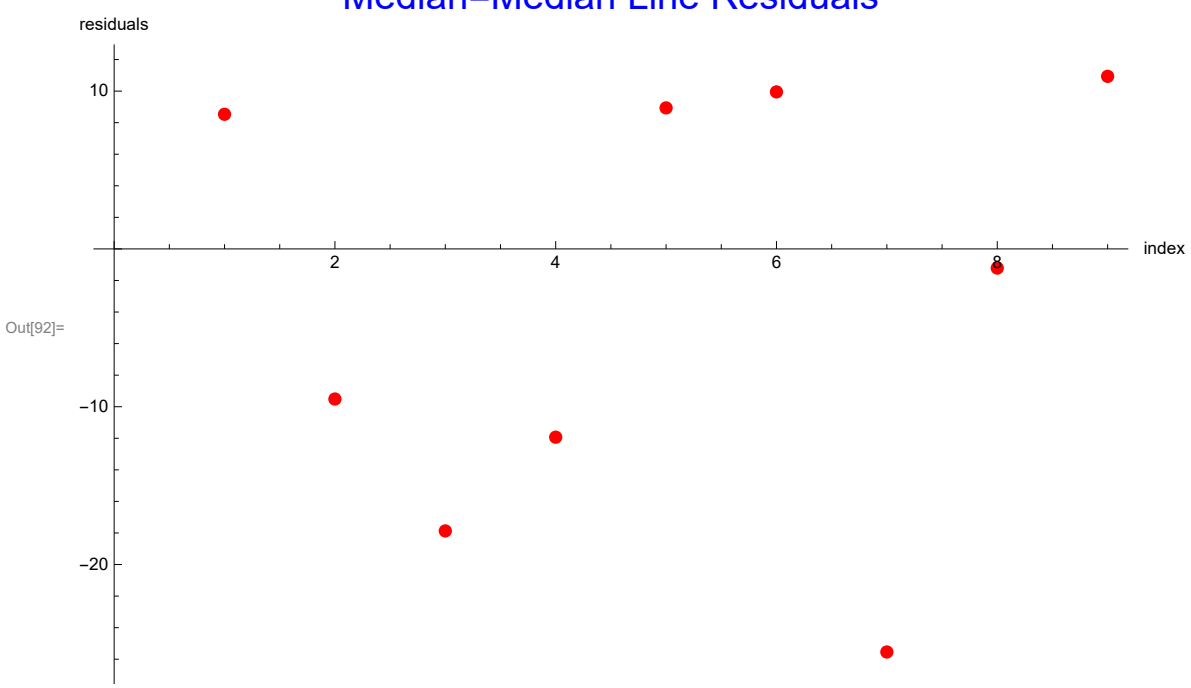
In[33]:=
```
Text[Grid[Prepend[list3, {"index", "residuals"}],
  Alignment → Center, Dividers → {2 → True, 2 → True}, Spacings → {1, 1}]]
```

Out[33]=

index	residuals
2.5	8.53234
2.6	−9.51244
3.4	−17.8706
1.3	−11.9303
1.6	8.93532
3.8	9.95025
11.6	−25.5423
6.4	−1.21393
8.3	10.9353

In[92]:=
```
residualsPlot2 = ListPlot[ myMedianMedianLineIndexResiduals ,
  AxesLabel → {"index", "residuals"},
  ImageSize → Large, PlotStyle → {Red, Point : 10 },
  AxesOrigin → {0, 0},
  PlotLabel → Style ["Median-Median Line Residuals", 20, Blue]]
```

Out[92]=



Median-Median Line Sum of Residuals Squared

In[41]:=
```
myMedianMedianLineIndexResidualsSquared =
 myMedianMedianLineIndexResiduals * myMedianMedianLineIndexResiduals
```

Out[41]= {72.8008, 90.4865, 319.36, 142.333, 79.84, 99.0075, 652.409, 1.47363, 119.581}

In[43]:=
```
sumOfResidualsSquared2 = Total[myMedianMedianLineIndexResidualsSquared]
```

Out[43]= 1577.29

Least Squares Regression Line

In[44]:=
```
indexSquared = index * index
```
Out[44]= {6.25, 6.76, 11.56, 1.69, 2.56, 14.44, 134.56, 40.96, 68.89}

In[45]:=
```
indexDeathsProduct = index * deaths
```
Out[45]= {367.5, 338., 442., 148.2, 220.8, 615.6, 2412.8, 1139.2, 1743.}

In[52]:=
```
varB = Total[indexSquared ]
```
Out[52]= 287.67

In[54]:=
```
varC = Total[index]
```
Out[54]= 41.5

In[55]:=
```
varD = Total[indexDeathsProduct]
```
Out[55]= 7427.1

In[56]:=
```
varE = Total[deaths]
```
Out[56]= 1417

In[102]:=
```
n = 9
```
Out[102]= 9

In[103]:=
$$m = \frac{(varD * n) - (varE * varC)}{(varB * n) - varC^2}$$
Out[103]= 9.27386

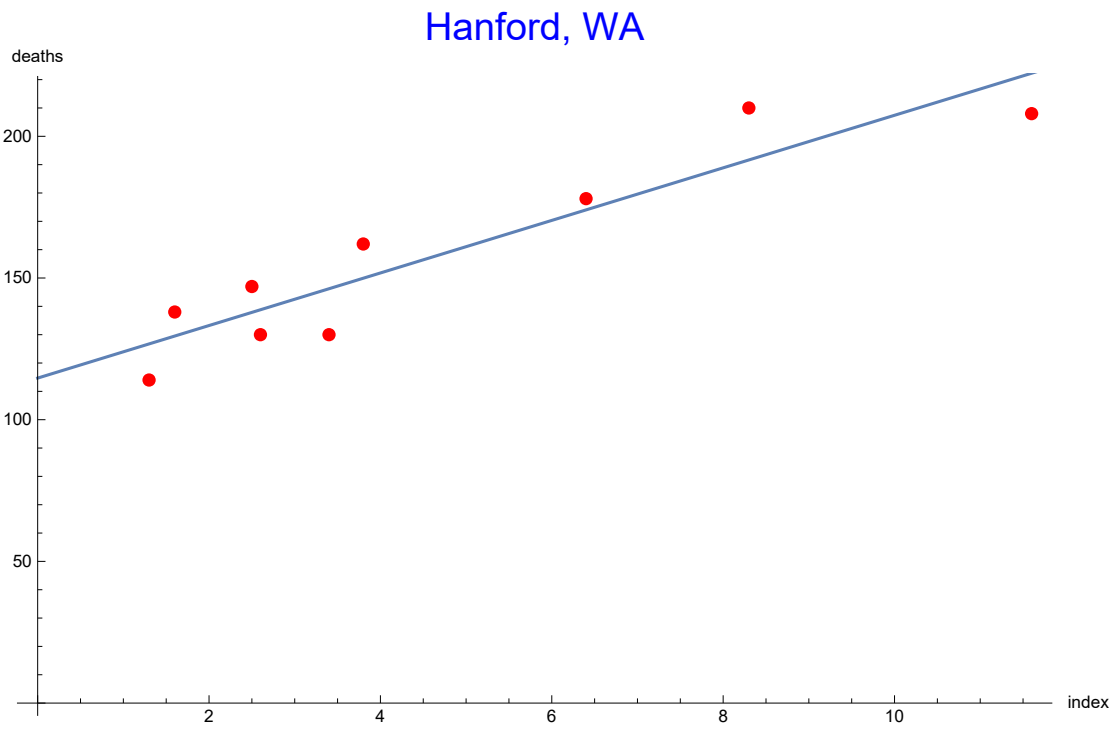
In[104]:=
$$b = \frac{varE - (m * varC)}{n}$$
Out[104]= 114.682

In[105]:=
```
leastSquaresRegessionLine[x] = m * x + b
```
Out[105]= 114.682 + 9.27386 x

In[106]:=
```
p3 = Plot[leastSquaresRegessionLine[x], {x, 0, 12}];
```

In[107]:=
```
Show[lp1, p3]
```

Out[107]=



Least Squares Regression Line Residuals

In[108]:=
```
leastSquaresRegressionLineIndexResiduals =
 {(list1[[1]][[2]] - leastSquaresRegessionLine[list1[[1]][[1]]]),
  list1[[2]][[2]] - leastSquaresRegessionLine[list1[[2]][[1]]],
  (list1[[3]][[2]] - leastSquaresRegessionLine[list1[[3]][[1]]]),
  list1[[4]][[2]] - leastSquaresRegessionLine[list1[[4]][[1]]],
  (list1[[5]][[2]] - leastSquaresRegessionLine[list1[[5]][[1]]]),
  (list1[[6]][[2]] - leastSquaresRegessionLine[list1[[6]][[1]]]),
  (list1[[7]][[2]] - leastSquaresRegessionLine[list1[[7]][[1]]]),
  (list1[[8]][[2]] - leastSquaresRegessionLine[list1[[8]][[1]]]),
  (list1[[9]][[2]] - leastSquaresRegessionLine[list1[[9]][[1]]])}
```

Out[108]=
```
{9.13371, -8.79367, -16.2128, -12.7376, 8.48019, 12.0777, -14.2585, 3.96564, 18.3453}
```

In[109]:=
```
list4 = Transpose[{index, leastSquaresRegressionLineIndexResiduals }];
```

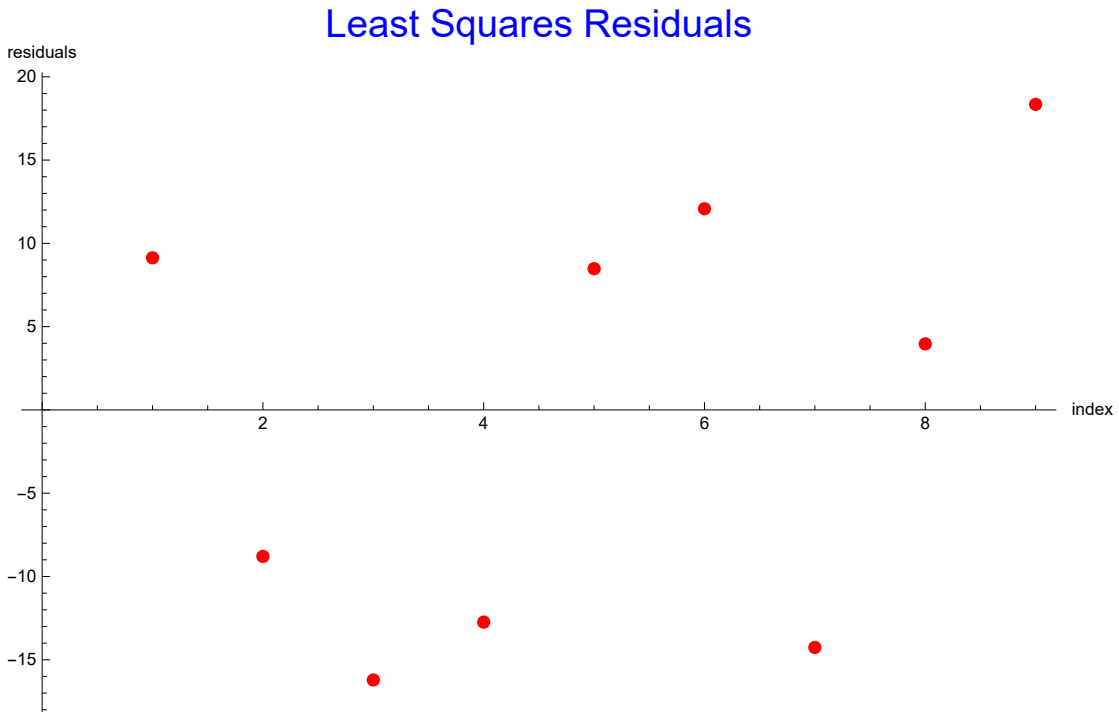
In[110]:=
```
Text[Grid[Prepend[list4, {"index", "residuals"}],
  Alignment → Center, Dividers → {2 → True, 2 → True}, Spacings → {1, 1}]]
```

Out[110]=

index	residuals
2.5	9.13371
2.6	−8.79367
3.4	−16.2128
1.3	−12.7376
1.6	8.48019
3.8	12.0777
11.6	−14.2585
6.4	3.96564
8.3	18.3453

In[123]:=
```
residualsPlot3 = ListPlot[ leastSquaresRegressionLineIndexResiduals ,
  AxesLabel → {"index", "residuals"},
  ImageSize → Large, PlotStyle → {Red, Point : 10 },
  AxesOrigin → {0, 0},
  PlotLabel → Style ["Least Squares Residuals", 20, Blue]]
```

Out[123]=



Least Squares Regression Line Sum of Residuals Squared

In[112]:=
```
leastSquaresRegressionLineIndexResidualsSquared =
 leastSquaresRegressionLineIndexResiduals * leastSquaresRegressionLineIndexResiduals
```

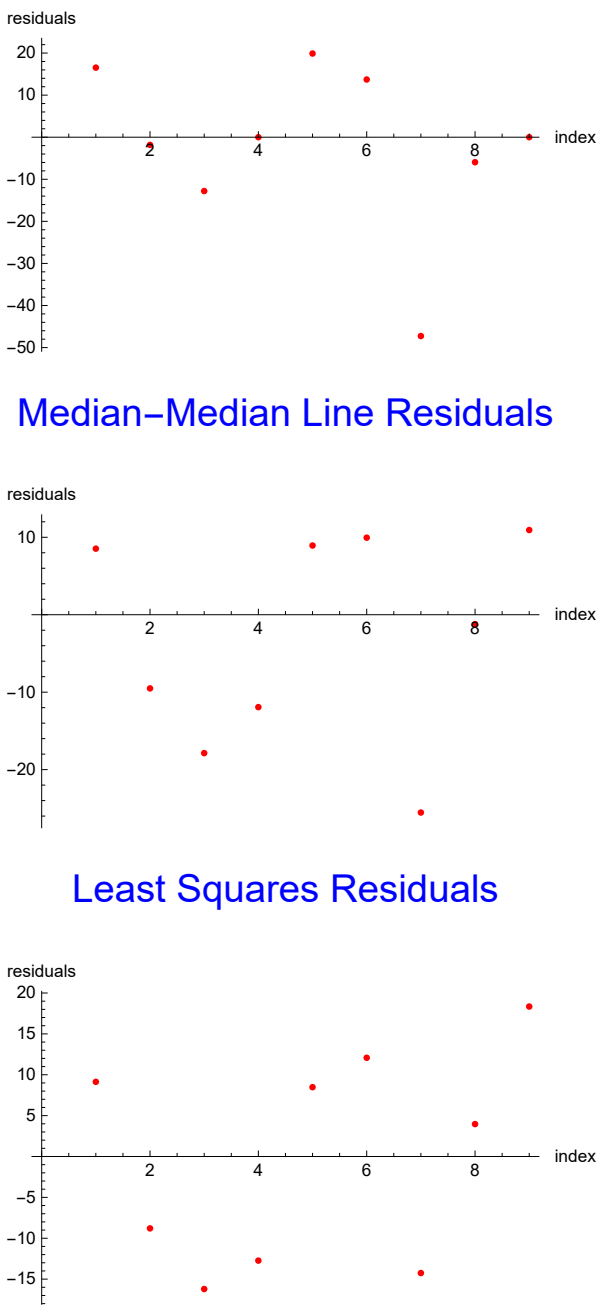
Out[112]= {83.4247, 77.3287, 262.854, 162.248, 71.9136, 145.871, 203.303, 15.7263, 336.55}

In[113]:=
```
sumOfResidualsSquared3 = Total[leastSquaresRegressionLineIndexResidualsSquared]
```

Out[113]= 1359.22

In[124]:=
```
GraphicsGrid[{{residualsPlot1}, {residualsPlot2}, {residualsPlot3}},
 PlotLabel → "Residuals Comparison", ImageSize → 400]
```

Out[124]=



Conclusion

I fit three different lines (a two point line, a median-median line, and a least squares regression line) to the same set of data as shown in previous sections. I used the residuals to compare how well each line fits the data. Since, as shown in the residuals plot above, the residuals appear to be randomly scattered above and below 0, each line is a relatively good fit for the data. Furthermore, the total error of each line can be compared using the sum of the squares of the residuals for each line. Since residuals can be positive or negative, a sum of residuals could be 0 even though the total error is not 0. Absolute values of residuals are always positive but the squaring the residuals is preferred for statistical purposes because the squaring function is easier to work with than the absolute value function, especially when using calculus. Since the least squares regression line has the smallest sum of squared residuals, as it should by definition, it is the best linear fit for the given data.