In[4]:=
```
(* Adam Beck *)
(* Problem 1*)


(* hosp-heart.nb data *)

(* {M,V} M = one year mortality rate,
percentage of patiuents that died within one year of the
 transplant operation,
V = average annual number of transplants at that center during the same 4 years *)
heart = {{17.9, 27}, {23.1, 4}, {40, 3}, {6.5, 35}, {14.9, 17}, {12.5, 4}, {15.7, 45},
   {9.8, 28}, {24, 6}, {5.0, 10}, {15.4, 13}, {4.8, 7}, {0, 1}, {19.1, 47}, {4.5, 6},
   {15, 56}, {12.5, 4}, {33.9, 8}, {10.7, 9}, {13, 14}, {28.3, 12}, {57.1, 2}, {6.3, 4},
   {10, 3}, {8.3, 12}, {17.5, 10}, {20, 3}, {29.3, 10}, {21.4, 7}, {27.3, 8}, {13.6, 6},
   {21.8, 30}, {36.4, 3}, {18.2, 11}, {33.3, 2}, {20, 4}, {38.5, 7}, {20.8, 18}, {12.2, 19},
   {22.2, 18}, {29, 8}, {0, 9}, {5.7, 9}, {50, 2}, {21.7, 15}, {66.7, 4}, {29.4, 17},
   {12.1, 27}, {10.7, 14}, {6.3, 4}, {16.2, 9}, {21.1, 5}, {17.4, 33}, {23.9, 17},
   {42.9, 2}, {40, 2}, {6.7, 15}, {44.4, 3}, {18.7, 34}, {14.7, 24}, {7.4, 7}, {12.6, 24},
   {9.7, 26}, {44.4, 2}, {16.7, 6}, {15.8, 14}, {83.3, 2}, {10.9, 22}, {13.3, 5},
   {11.1, 5}, {75, 2}, {19, 20}, {14, 13}, {60, 1}, {21.2, 8}, {9.7, 8}, {50, 2}, {25, 14},
   {18.6, 15}, {0.0, 1}, {35.3, 9}, {23.5, 85}, {15.6, 11}, {37.5, 2}, {14.3, 28},
   {14.3, 4}, {16.7, 6}, {20.0, 15}, {13.0, 17}, {9.6, 26}, {66.7, 3}, {30.8, 3},
   {14.0, 13}, {27.5, 10}, {37.5, 8}, {18.9, 13}, {0.0, 4}, {12.2, 44}, {57.1, 4},
   {21.4, 35}, {23.4, 16}, {10.9, 12}, {15.6, 8}, {16.7, 2}, {13.9, 9}, {18.2, 11},
   {11.5, 26}, {18.4, 13}, {16.7, 3}, {20.4, 14}, {40.0, 5}, {20.7, 56}, {19.6, 13},
   {13.5, 9}, {29.9, 36}, {8.4, 21}, {28.4, 24}, {7.7, 23}, {19.3, 29}, {0.0, 1},
   {22.2, 20}, {30.0, 5}, {7.0, 11}, {23.8, 7}, {18.8, 29}, {14.5, 16}, {17.0, 16},
   {20.0, 15}, {6.7, 15}, {11.4, 20}, {100.0, 1}, {31.4, 9}, {17.6, 26}, {19.6, 14}};
```

In[5]:=
```
(* Split this M and V data into separate
 lists via Transpose[] in order to parse through *)
heartTranspose = Transpose[heart];
MData = heartTranspose[[1]];
VData = heartTranspose[[2]];
```

In[8]:=
```
(* Define mean, median, quantile, and variance functions *)
```

In[9]:=
```
mean[x_] := Sum[x[[i]], {i, 1, Length[x]}] / Length[x];
(* Sum elements, divide by length *)
```

In[10]:=
```
median[x_] := (s = Sort[x]; s[[IntegerPart[.5 * Length[s]]]]);
(* Sort list, take element at index 1/2*length *)
```

In[11]:=
```
quantile[x_, alpha_] := (s = Sort[x]; s[[IntegerPart[alpha * Length[s]]]])
 (* Sort list, take element at index .alpha*length *)
```

In[12]:=
```
variance[x_] := (m = mean[x]; Sum[(x[[i]] - m)^2, {i, 1, Length[x]}] / Length[x]);
(* difference of every element from mean, squared, times 1/length *)
```

In[13]:=
```
(* Find the mean, median, q1 and q3, and variance *)
hospMeanM = mean[MData]
```

Out[13]= 21.9045

In[14]:=
```
hospMeanV = N[mean[VData]]
```
Out[14]= 13.8657

In[15]:=
```
hospMedianM = median[MData]
```
Out[15]= 18.2

In[16]:=
```
hospMedianV = median[VData]
```
Out[16]= 10

In[17]:=
```
hospQ1M = quantile[MData, .25]
```
Out[17]= 12.2

In[18]:=
```
hospQ1V = quantile[VData, .25]
```
Out[18]= 4

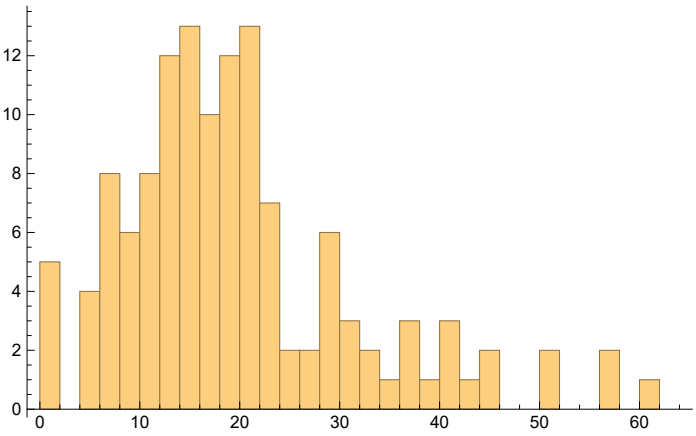
In[19]:=
```
hospQ2M = quantile[MData, .75]
```
Out[19]= 25

In[20]:=
```
hospQ2V = quantile[VData, .75]
```
Out[20]= 17

In[21]:=
```
hospVarianceM = variance[MData]
```
Out[21]= 268.634

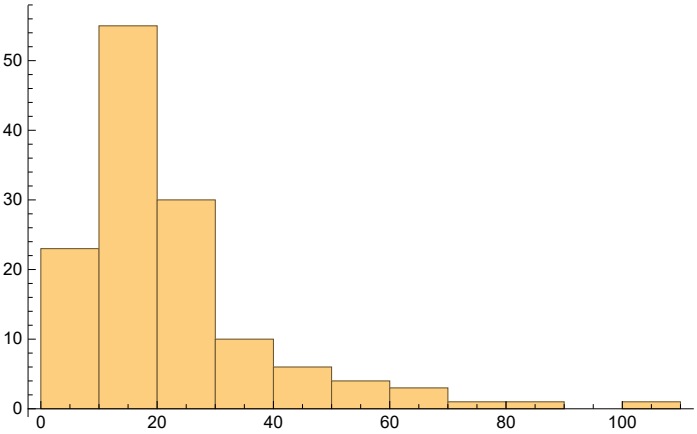
In[22]:=
```
hospVarianceV = N[variance[VData]]
```
Out[22]= 166.46

In[23]:=
```
(* Histograms using two difference bin sizes *)
(* I will use a bin size Length/2 for a very large bin,
and Length/10 for a smaller bin *)

(* MData large bin *)
Histogram[MData, IntegerPart[Length[MData] / 2]]
```
Out[23]=



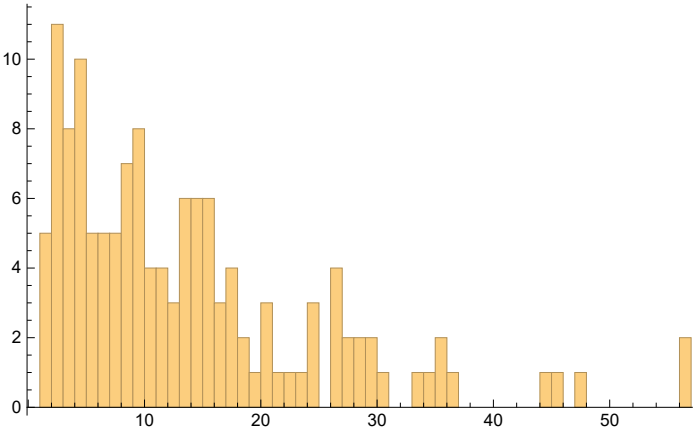
In[24]:=
```
(* MData small bin *)
Histogram[MData, IntegerPart[Length[MData] / 10]]
```

Out[24]=



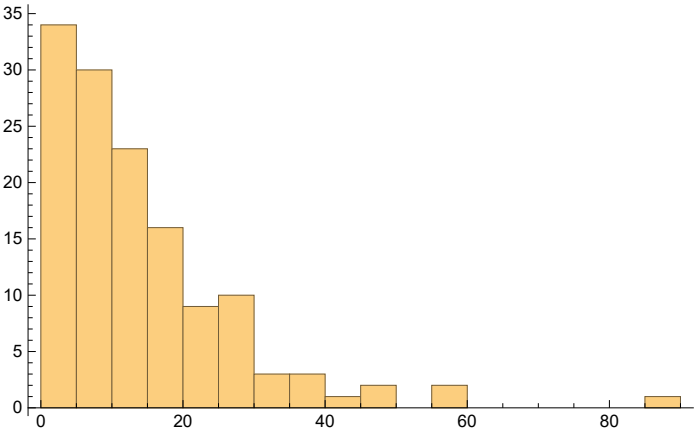
In[25]:=
```
(* VData, large bin *)
Histogram[VData, IntegerPart[Length[VData] / 2]]
```

Out[25]=



In[26]:=
```
(* VData, small bin *)
Histogram[VData, IntegerPart[Length[VData] / 10]]
```

Out[26]=

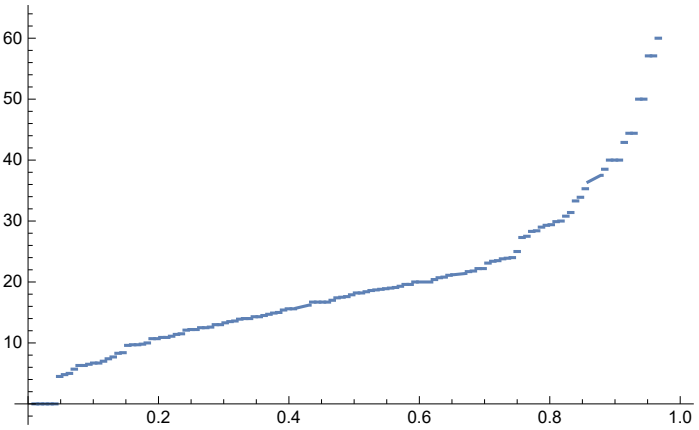


In[27]:=
```
(* Product plots of quantile functions, moment functions, and CDFs *)

(* Define functions for moments and CDF *)

(* Sum elements raised to the kth power, divide by length *)
moment[x_, k_] := N[Sum[x[[i]]^k, {i, 1, Length[x]}] / Length[x]];
```

In[28]:=
```
cdf[x_, xi_] := N[Sum[If[x[[i]] <= xi, 1, 0], {i, 1, Length[x]}]];
(* Count that an element is less than or equal to a given element *)
```

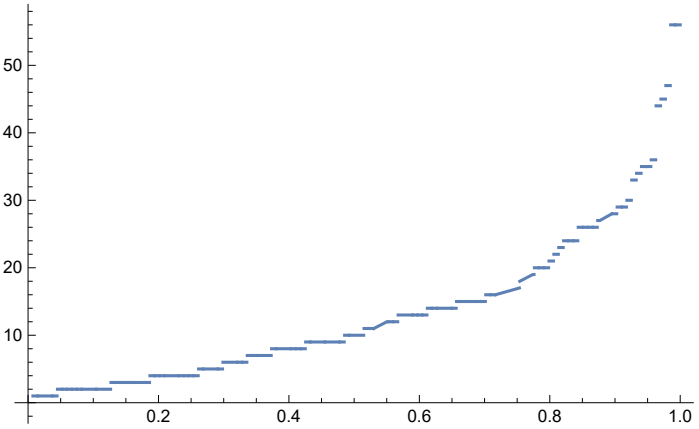
In[29]:=
```
(* Plot the quantile functions *)
(* MData *)
Plot[quantile[MData, i], {i, 0, 1}]
```

Out[29]=



In[30]:=
```
(* VData *)
Plot[quantile[VData, i], {i, 0, 1}]
```

Out[30]=

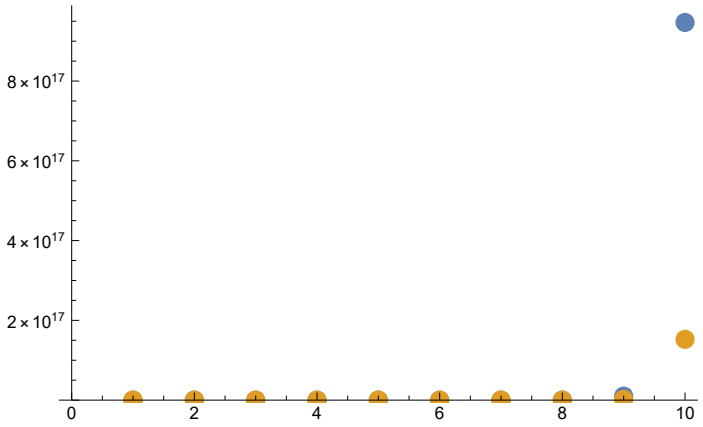


In[31]:=
```
(* Plot moment functions *)
(* Get the first 10 moments for MData and VData in a Table *)
momMData  = Table[moment[MData, i], {i, 1, 10}];
```

In[32]:=
```
momVData = Table[moment[VData, i], {i, 1, 10}];
```

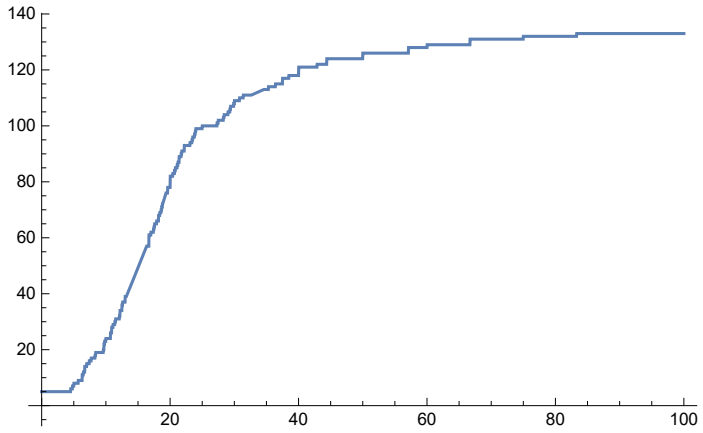
In[33]:=
```
(* Plot the two moments *)
ListPlot[{momMData, momVData}, PlotStyle → PointSize[.03], PlotRange → {0, 9.9 * 10^17}]
(* Blue is MData, Orange is VData*)
```

Out[33]=



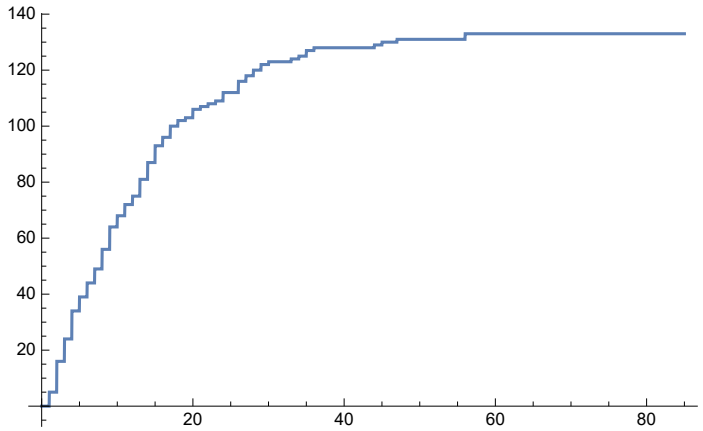
In[34]:=
```
(* Plot CDF functions *)
(* MData *)
Plot[cdf[MData, i], {i, 0, Max[MData]}]
```

Out[34]=



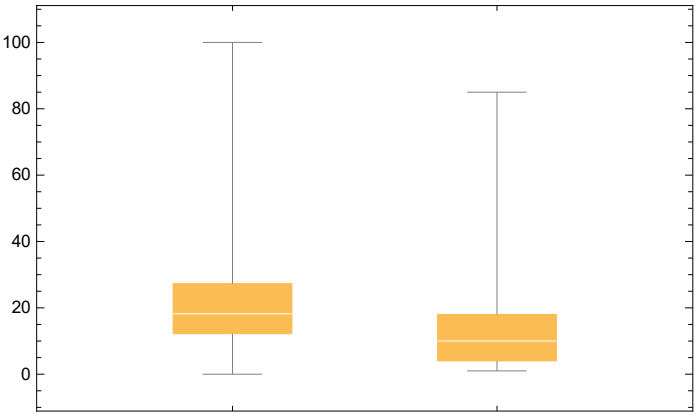
In[35]:=
```
(* VData *)
Plot[cdf[VData, i], {i, 0, Max[VData]}]
```

Out[35]=



```
(* Although the instructions do not say to compare
 any box and whisker and QQ plots for this hospital data
 against other sets of data, I will product them anyways. QQ
 will be MData(x axis) against VDaya (y axis) *)

(* Box an whisker plots *)
(* A box and whisker plot takes a min, q1, q2 (median), q3, and max *)

(* Box and whisker, MData and VData, outliers not shown *)
BoxWhiskerChart[{MData, VData}]
```

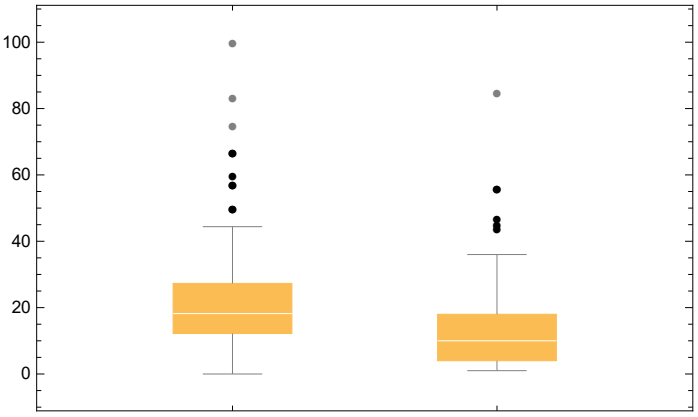
Out[46]=



In[47]:=

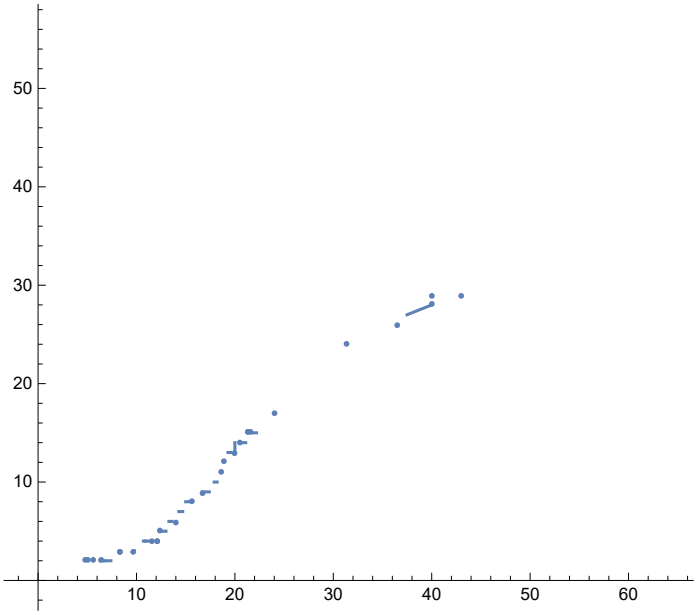
```
(* Box and whisker, MData and VData, outliers shown *)
BoxWhiskerChart[{MData, VData}, "Outliers"]
```

Out[47]=



In[54]:=
```
(* Parametric Plots (QQ Plots), MData on x axis, VData on y axis *)
ParametricPlot[{quantile[MData, i], quantile[VData, i]}, {i, 0, 1}]
```

Out[54]=




```
(* If the two sets come from a population with the same distribution,
the points should fall approximately
 along a 45 degree reference line. As we can see,
the 2 batches do not appear to have come from populations with
 a common distribution, as they do not fit along a 45 degree line *)
```