Berkson’s Paradox

Berkson’s Paradox

Removing one quadrant (where both variables are negative) for uncorrelated variables produces a fake negative correlation

In[26]:= $f[x\_, y\_] := \frac{1}{3}\begin{cases} 1 & (x == 1 \&\& y == 1) \,||\, (x == 1 \&\& y == 0) \,||\, (x == 0 \&\& y == 1) \\ 0 & \text{True} \end{cases}$

In[30]:= $\sum_{y=0}^{1}\sum_{x=0}^{1} f[x, y]$

Out[30]= 1

In[39]:= $\mu_X = \sum_{y=0}^{1}\sum_{x=0}^{1} x\, f[x, y];\ \mu_Y = \sum_{y=0}^{1}\sum_{x=0}^{1} y\, f[x, y];$

Correlation

In[43]:= $\frac{\sum_{y=0}^{1}\sum_{x=0}^{1} (x - \mu_X)\,(y - \mu_Y)\, f[x, y]}{\sqrt{\sum_{y=0}^{1}\sum_{x=0}^{1} (x - \mu_X)^2 f[x, y]}\ \sqrt{\sum_{y=0}^{1}\sum_{x=0}^{1} (y - \mu_Y)^2 f[x, y]}}$

Out[43]= $-\frac{1}{2}$

Monte Carlo Intuition

In[10]:=
```
X = 2 RandomVariate[BernoulliDistribution[1/2], 10^4] - 1;
Y = 2 RandomVariate[BernoulliDistribution[1/2], 10^4] - 1;
Z = Transpose[{X, Y}];
```

In[17]:=
```
Correlation[Z][[2]][[1]] // N
```

Out[17]= 0.0291565

In[16]:=
```
Correlation[Select[Z, #[[1]] + #[[2]] > -2 &]][[2]][[1]] // N
```

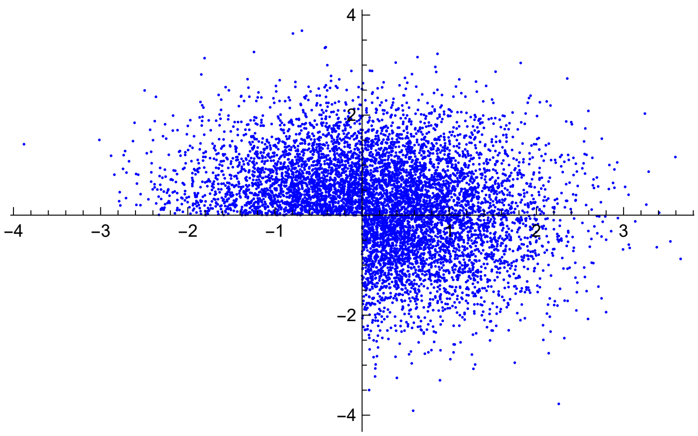
Out[16]= -0.482023

Remove one quadrant.
The quadrant where both variables are negative.
Do this for uncorrelated variables.
This gives a fake negative correlation.

```
In[◦]:= X = RandomVariate[NormalDistribution[], 10^4];
        Y = RandomVariate[NormalDistribution[], 10^4];
        Z = Transpose[{X, Y}];
        Correlation[Z][[2]][[1]]
Out[◦]=
        -0.00311412

In[◦]:= Z2 = Select[Z, Sign[#[[1]]] + Sign[#[[2]]] > -2 &];
        Correlation[Z2][[2]][[1]]
Out[◦]=
        -0.311458

In[◦]:= ListPlot[Z2, PlotStyle → Blue]
Out[◦]=
```



With Bernoulli Distribution

Remove one quadrant.
The quadrant where both variables are negative.
Do this for uncorrelated variables.
This gives a fake negative correlation.

```
In[◦]:= X = 2 RandomVariate[BernoulliDistribution[1 / 2], 10^4] - 1;
        Y = 2 RandomVariate[BernoulliDistribution[1 / 2], 10^4] - 1;
        Z = Transpose[{X, Y}];
        Correlation[Z] // N
Out[◦]=
        {{1., -0.00838715}, {-0.00838715, 1.}}

In[◦]:= Z2 = Select[Z, Sign[#[[1]]] + Sign[#[[2]]] > -2 &];
        Correlation[Z2] // N
Out[◦]=
        {{1., -0.498562}, {-0.498562, 1.}}
```