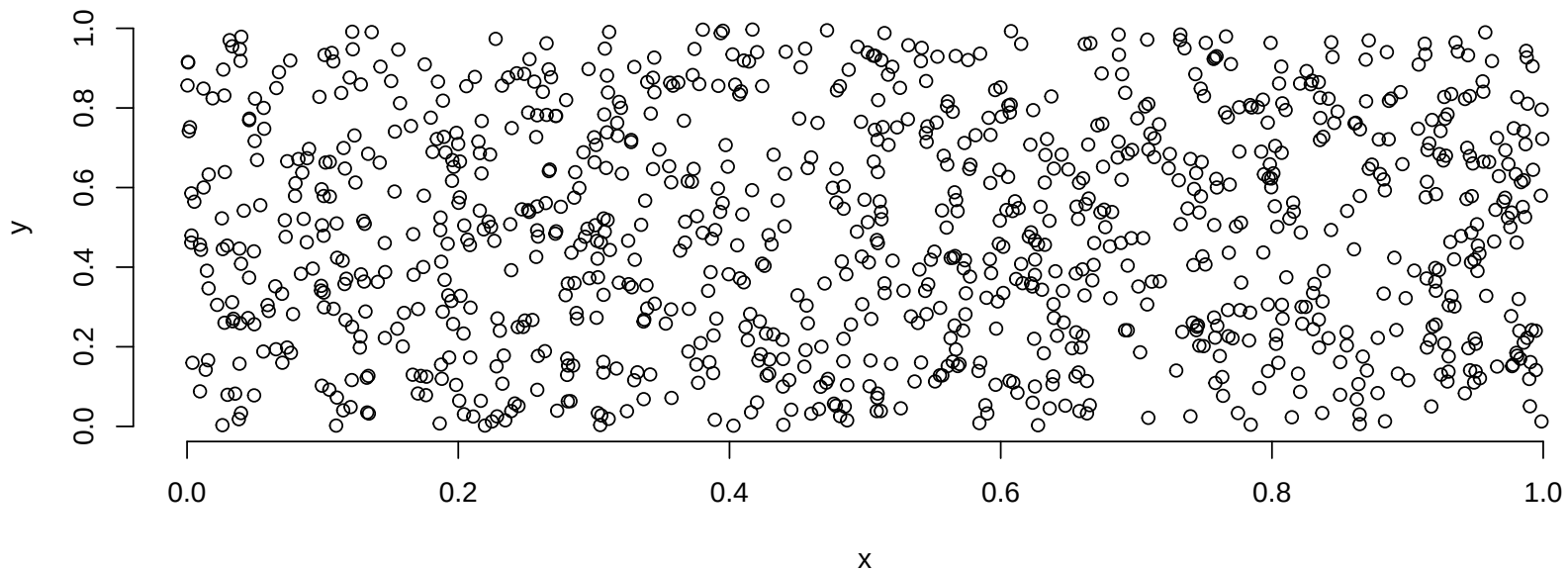
Sampling Error

Ryan Miller

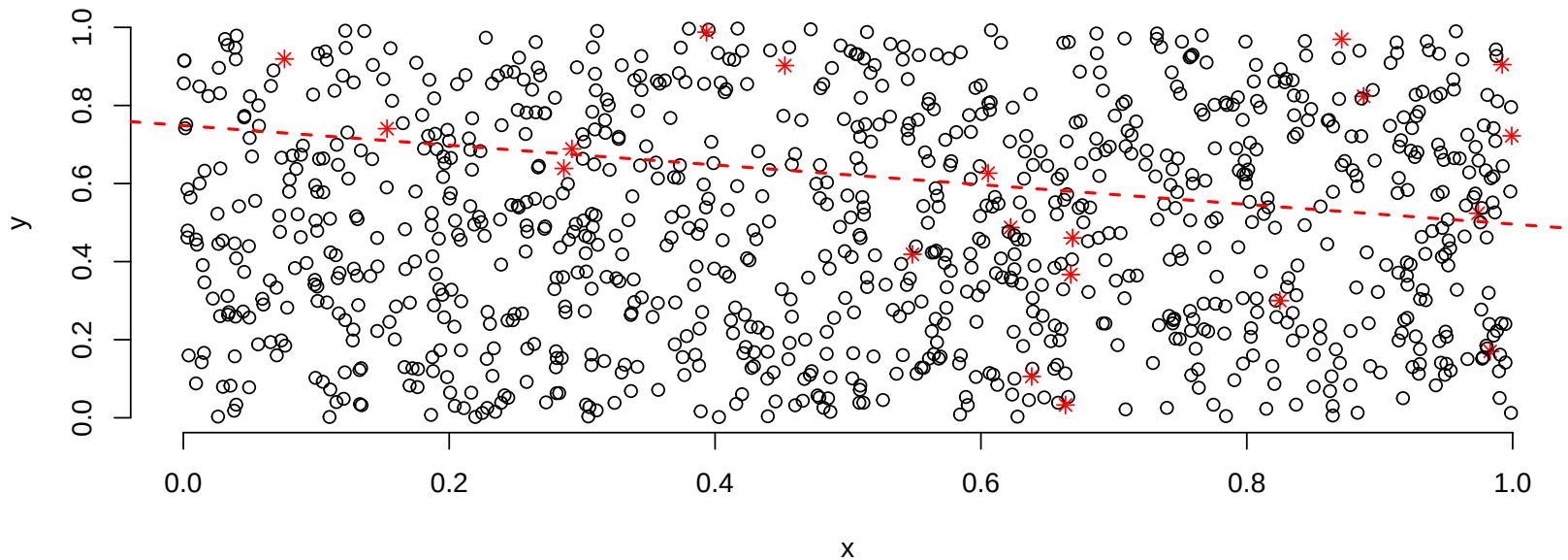
Sampling variability

The scatterplot below depicts a *population* ($N = 1000$) where the variables X and Y are *not related* (ie: $\rho = 0$):



Sampling variability

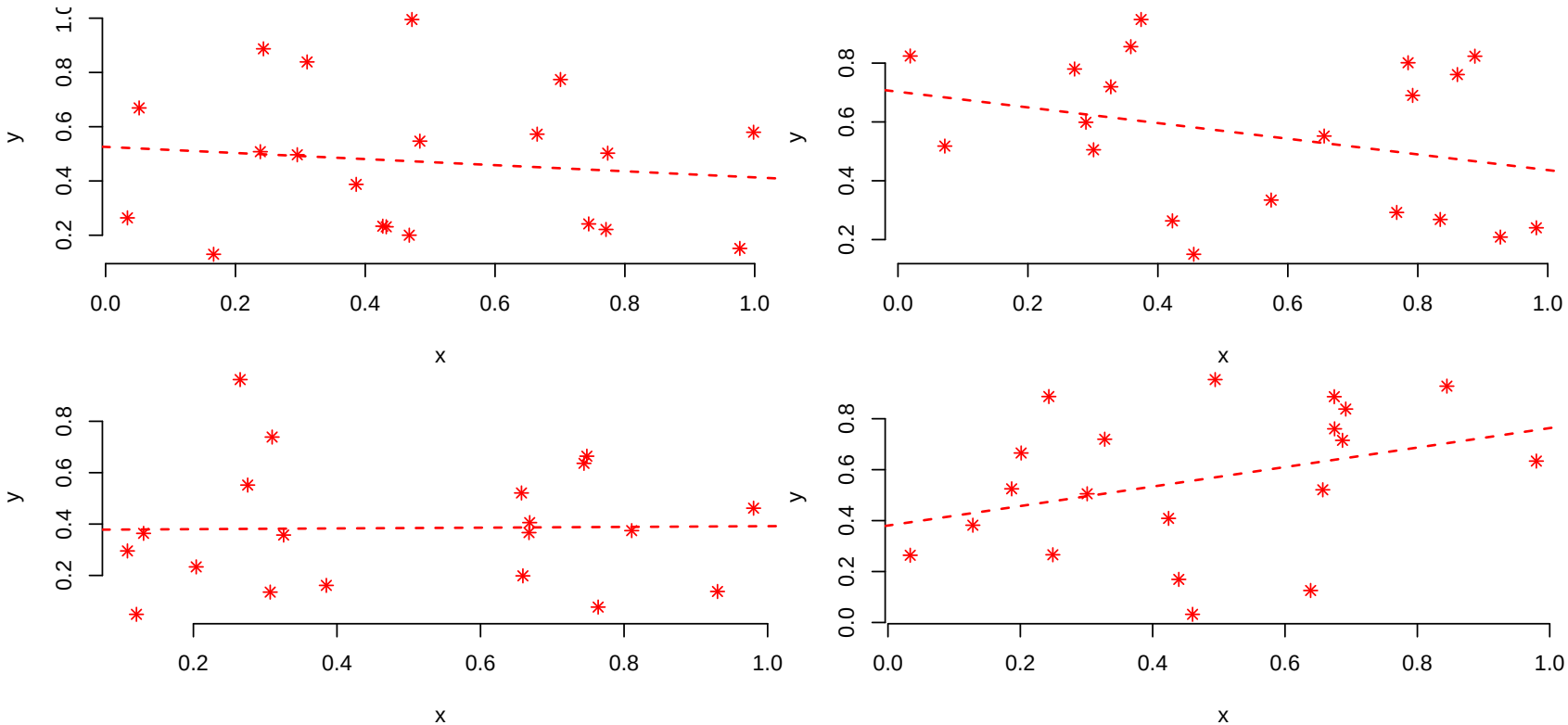
Here is a *random sample* ($n = 20$) from this population (sampled cases are colored in red), the sample correlation is $r = -0.245$:



So, the sample data suggest a *weak negative correlation* despite these variables having *no correlation* in the population

Sampling variability

Shown below are another four random samples (each $n = 20$):



Across these samples, the observed sample correlations range from $r = -0.31$ (top right) to $r = 0.35$ (bottom right)

Sampling distributions

The distribution of *all possible estimates* that could be observed when sampling is known as the **sampling distribution**: