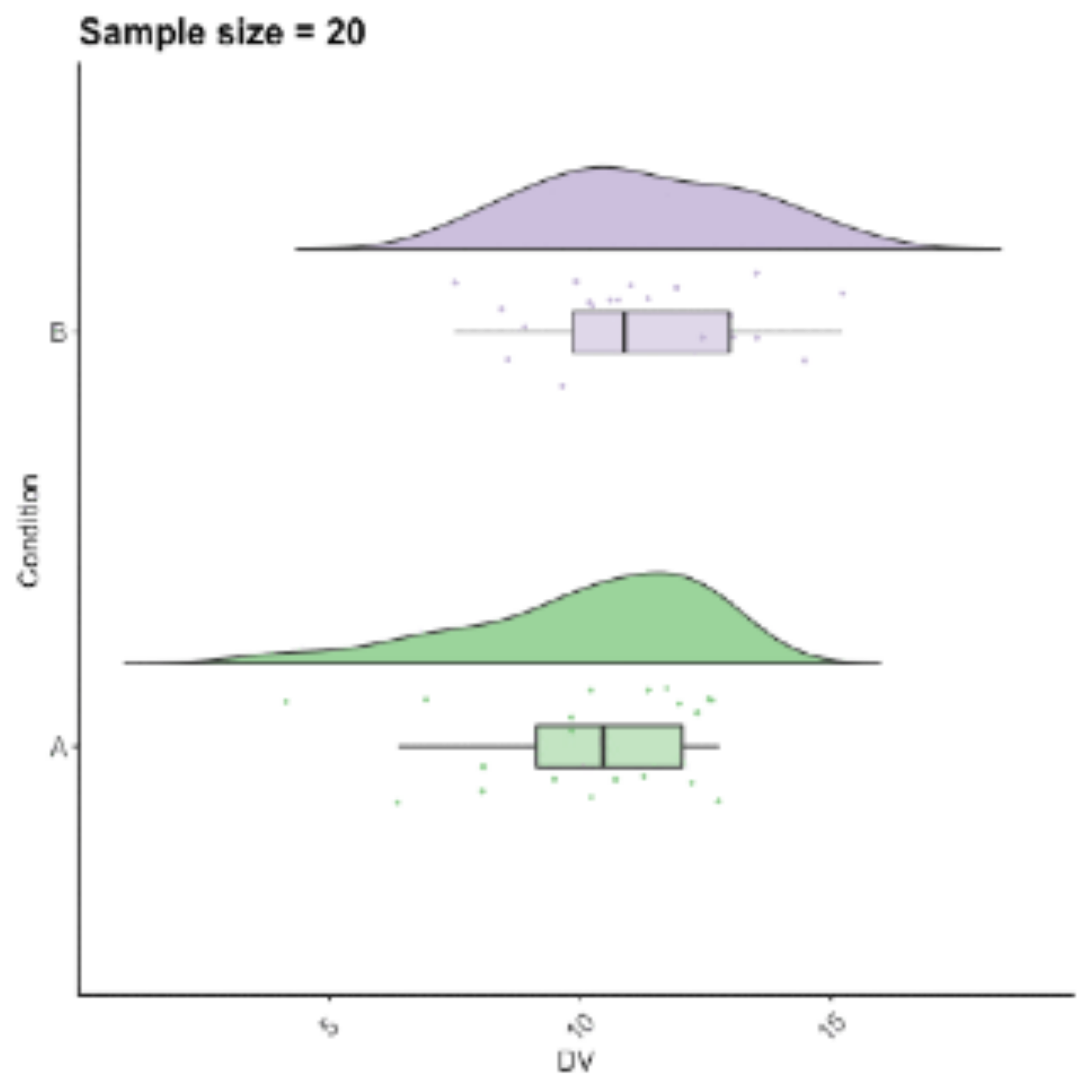
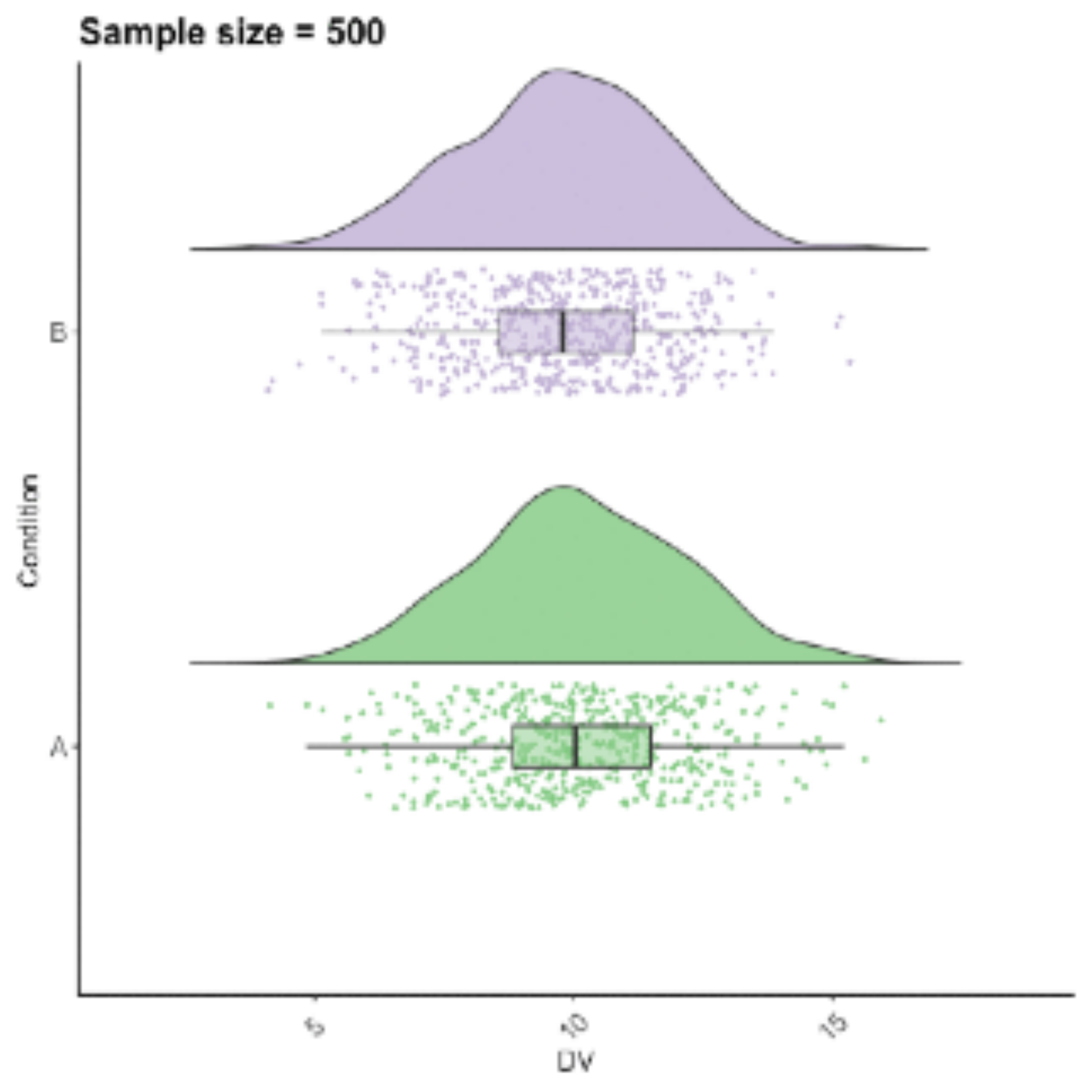
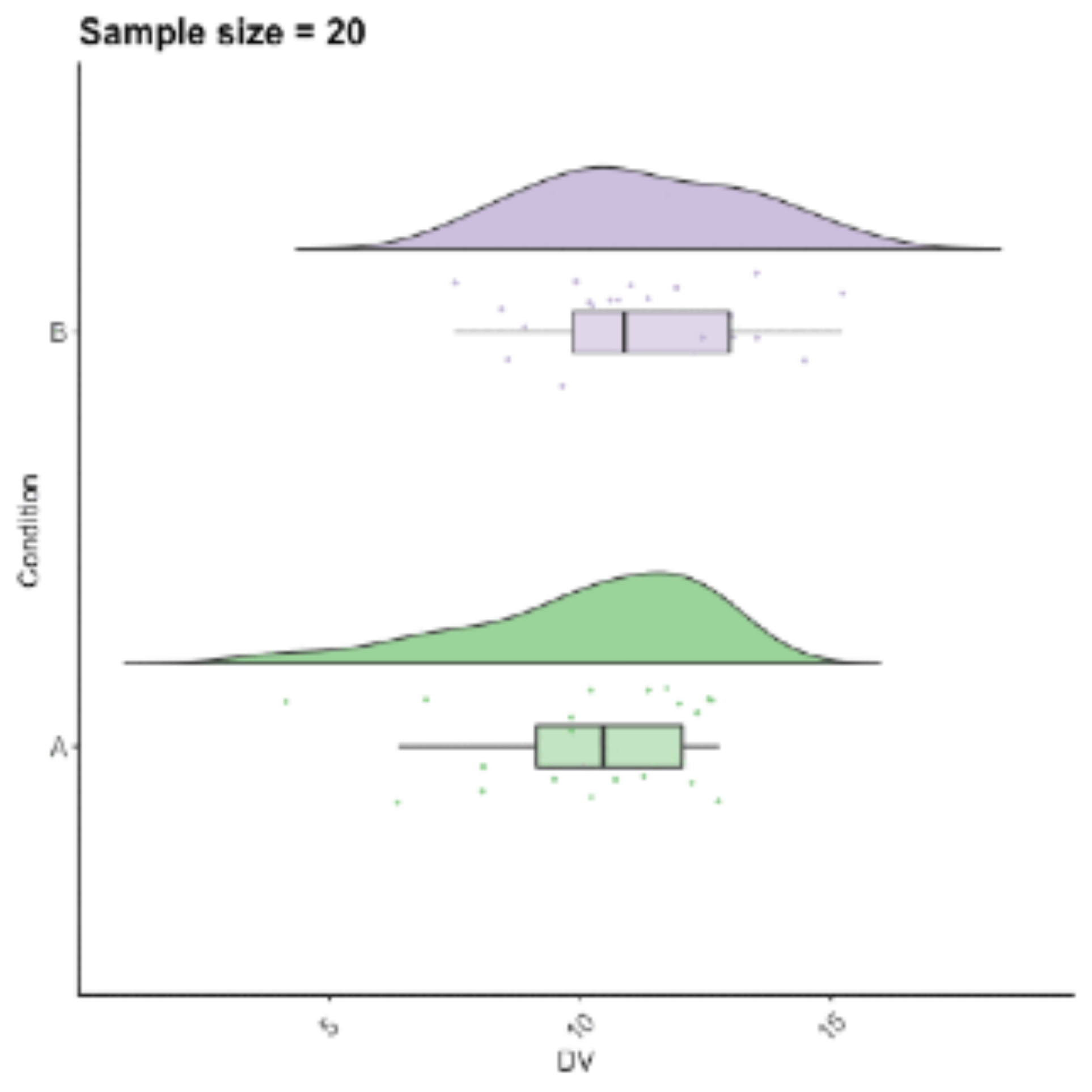
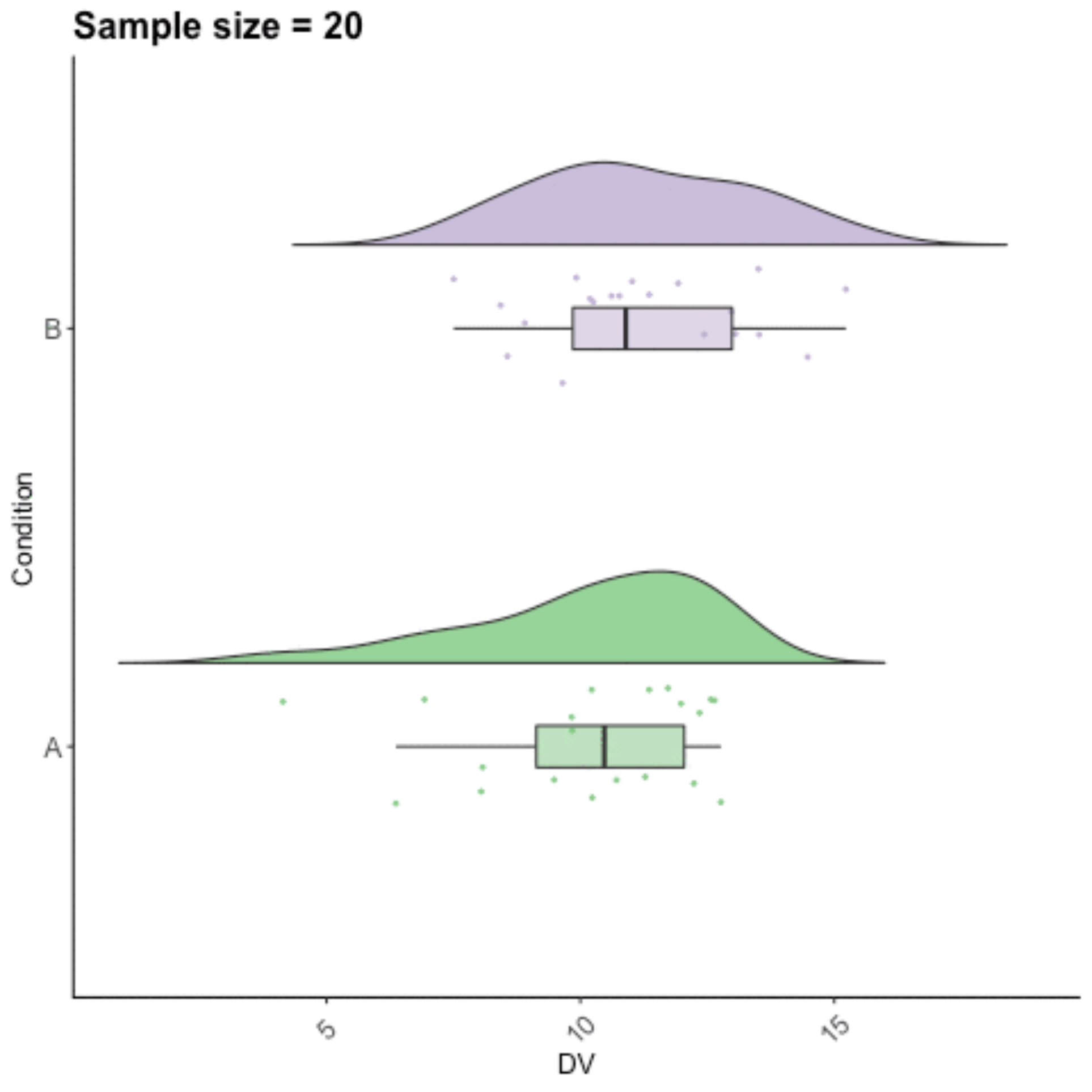
The Problem of Sampling Bias

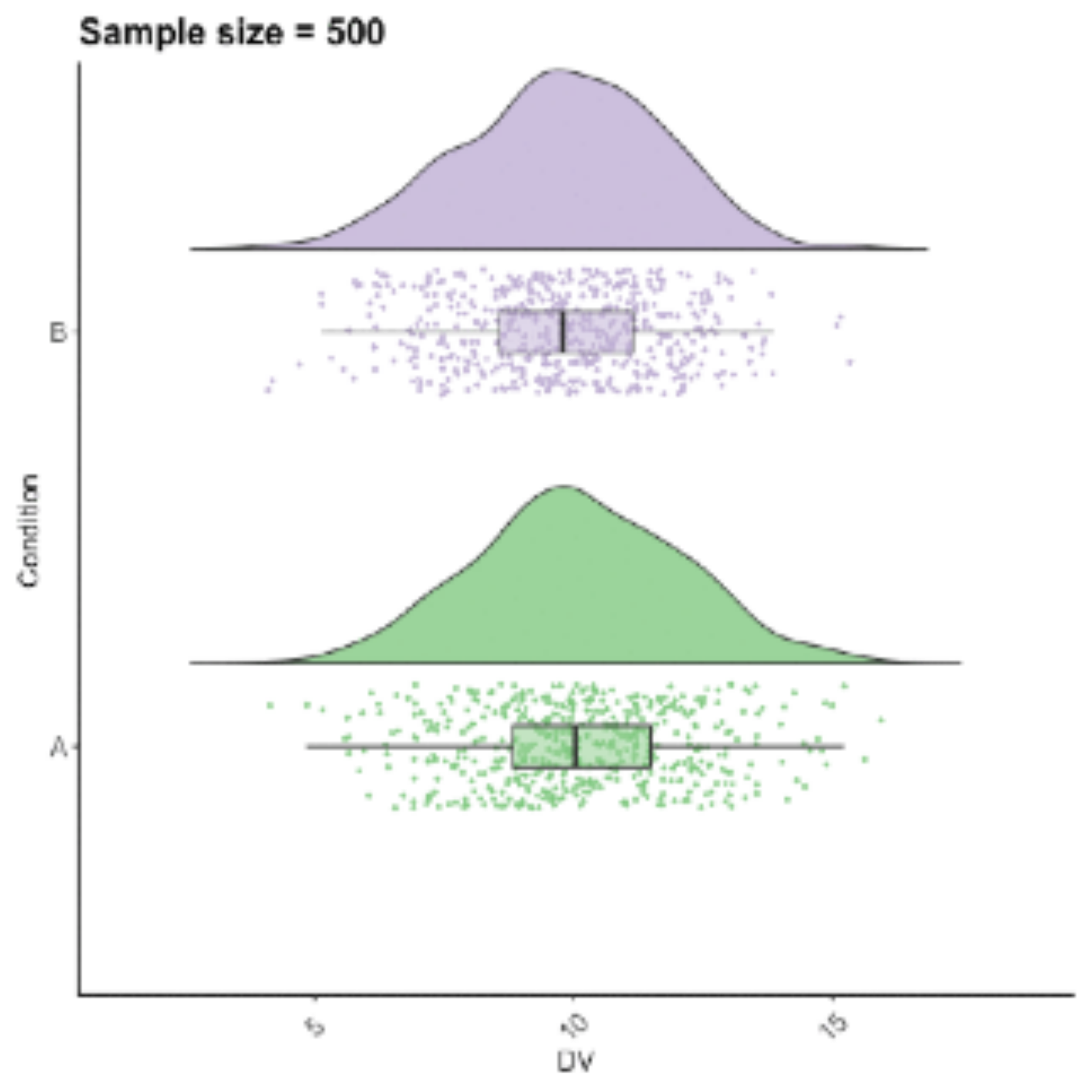
Samples for conditions A and B are drawn from the same population. Due to sampling error, with small samples (e.g., N=20) we might sometimes conclude there is a difference between A and B where there isn't one (as you can see with the N=500 samples).

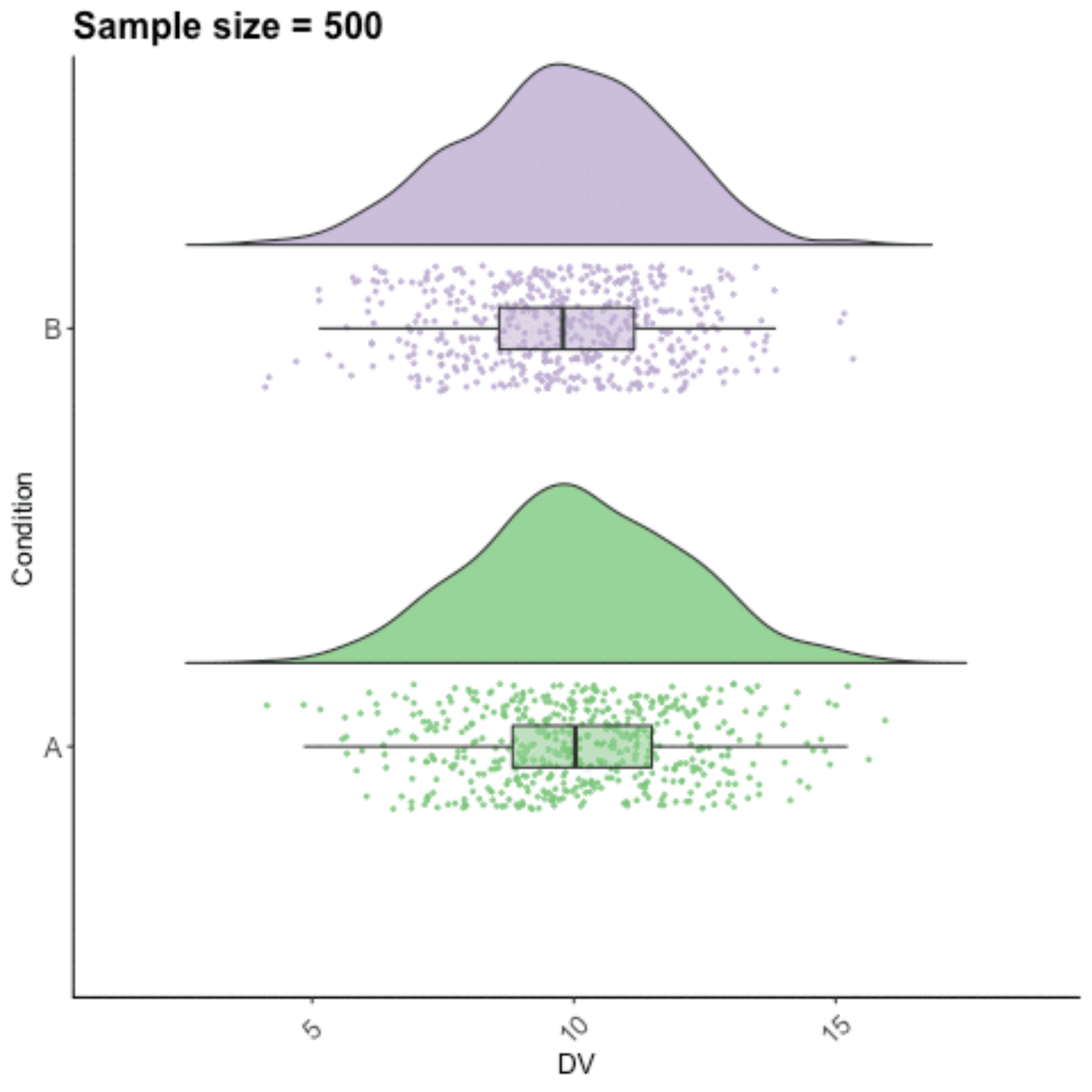

Sample size = 20
B
A
Condition
5
10
15
DV


Sample size = 500
B
A
Condition
5
10
15
DV

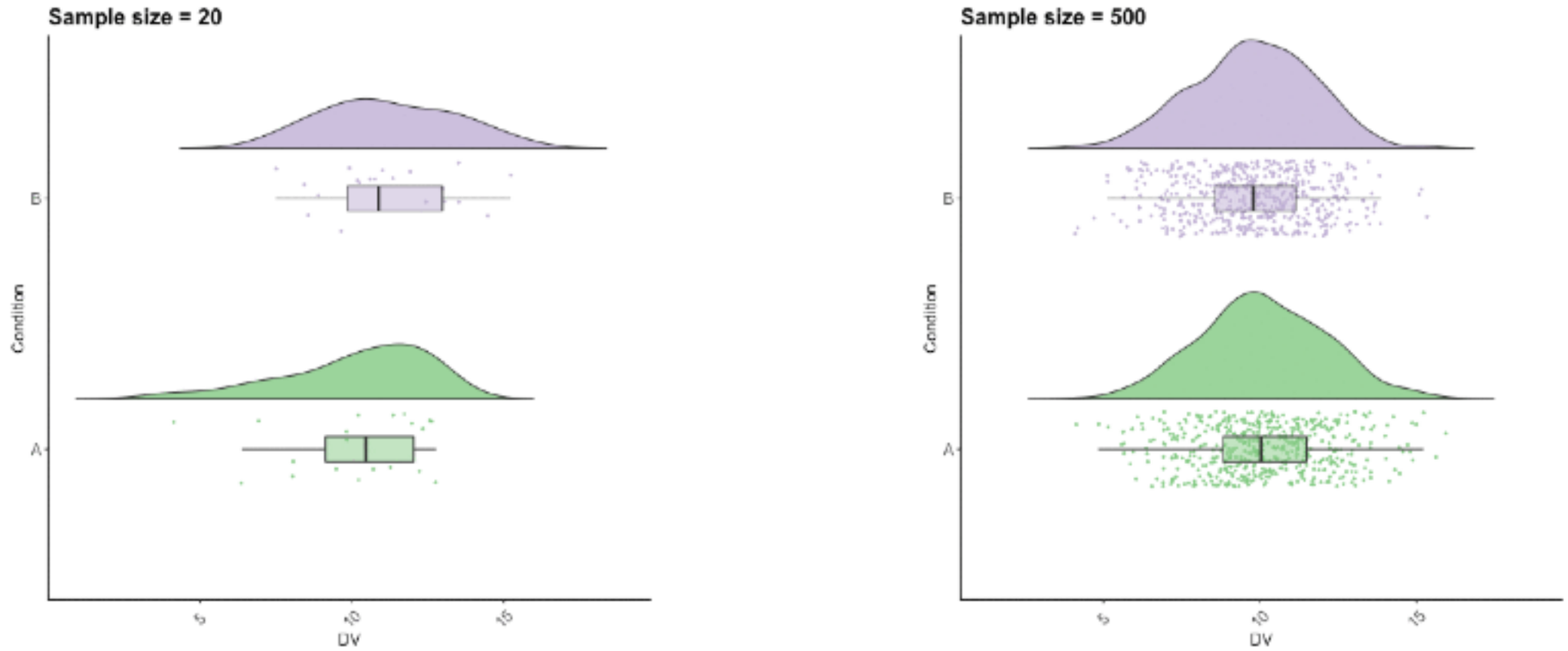

Sample size = 20
B
A
Condition
5
10
15
DV


Sample size = 20
B
A
Condition
5
10
15
DV

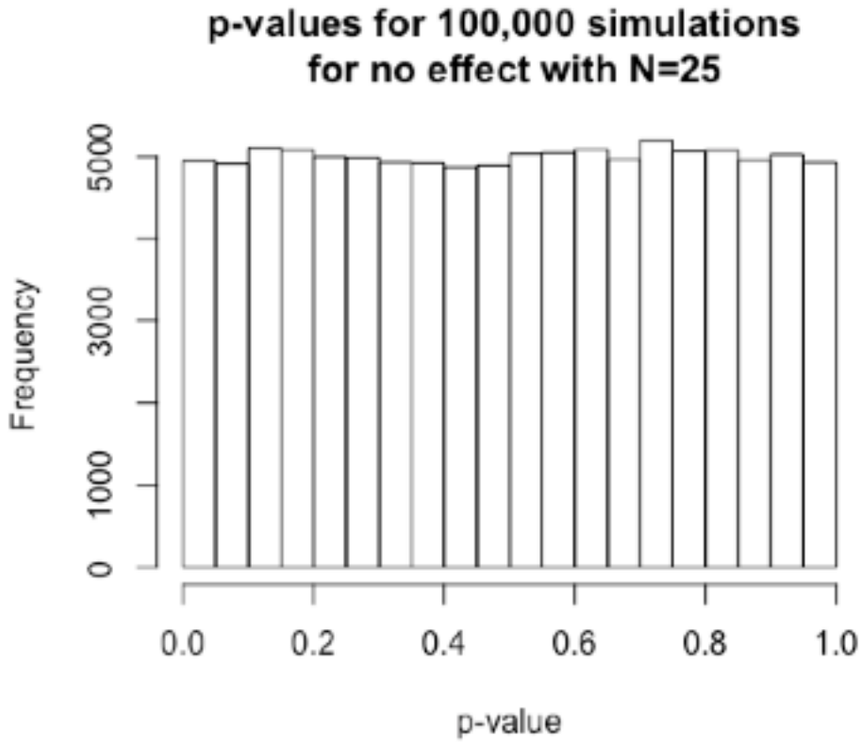

Sample size = 500
B
A
Condition
5
10
15
DV

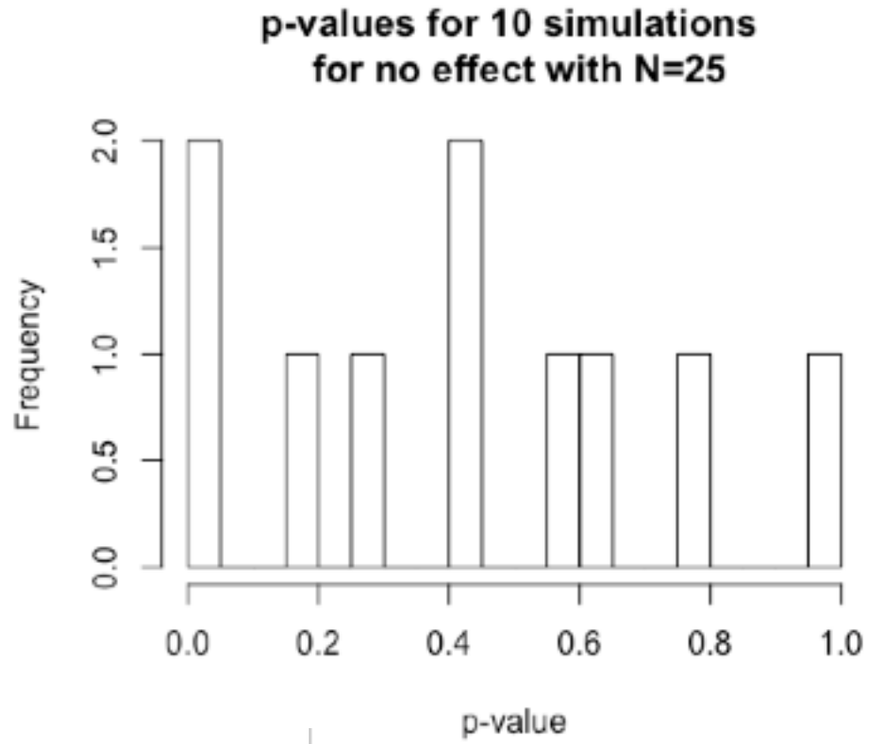

Sample size = 500
Condition
B
A
5
10
15
DV

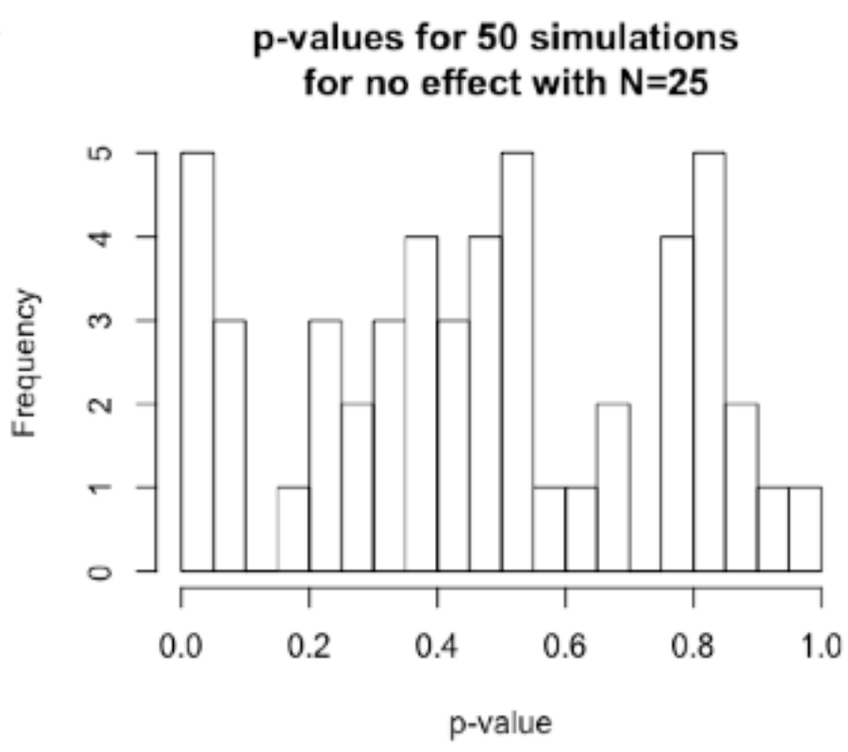
The Problem of Sampling Bias



Samples for conditions A and B are drawn from the same population. Due to sampling error, with small samples (e.g., N=20) we might sometimes conclude there is a difference between A and B where there isn't one (as you can see with the N=500 samples).


p-values for 100,000 simulations
for no effect with N=25
Frequency
0
1000
3000
5000
0.0
0.2
0.4
0.6
0.8
1.0
p-value


p-values for 10 simulations
for no effect with N=25
Frequency
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
p-value


p-values for 50 simulations
for no effect with N=25
Frequency
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
p-value