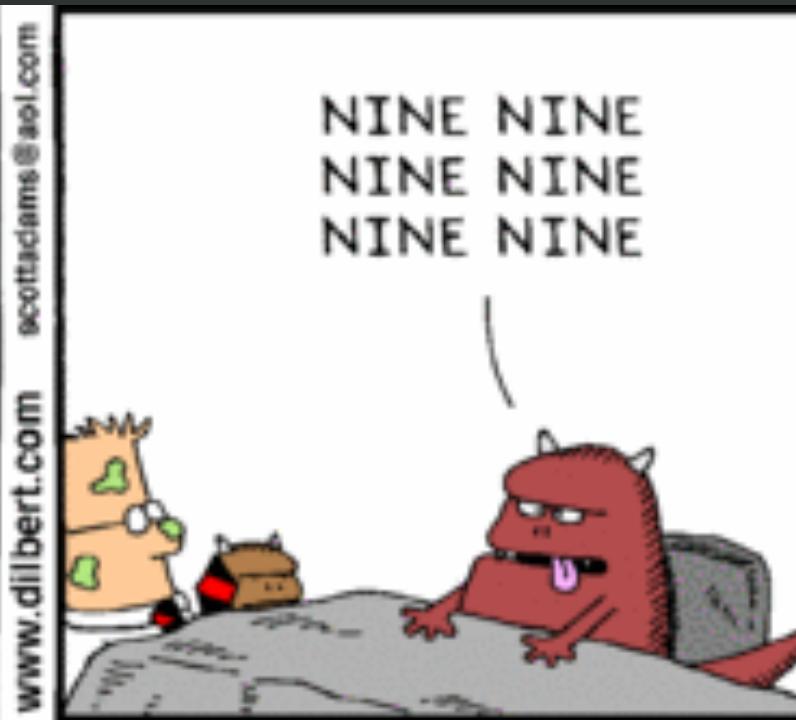


# BEYOND BASIC A/B TESTING



János Divényi  
Emarsys  
@divenyijanos  
divenyijanos.github.io







main  
90%



control  
10%



1 week

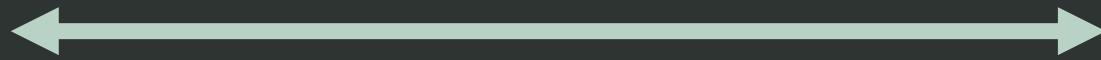
**comparison period**



main  
90%



control  
10%



1 week

**comparison period**

for 2 months

**~450k people**

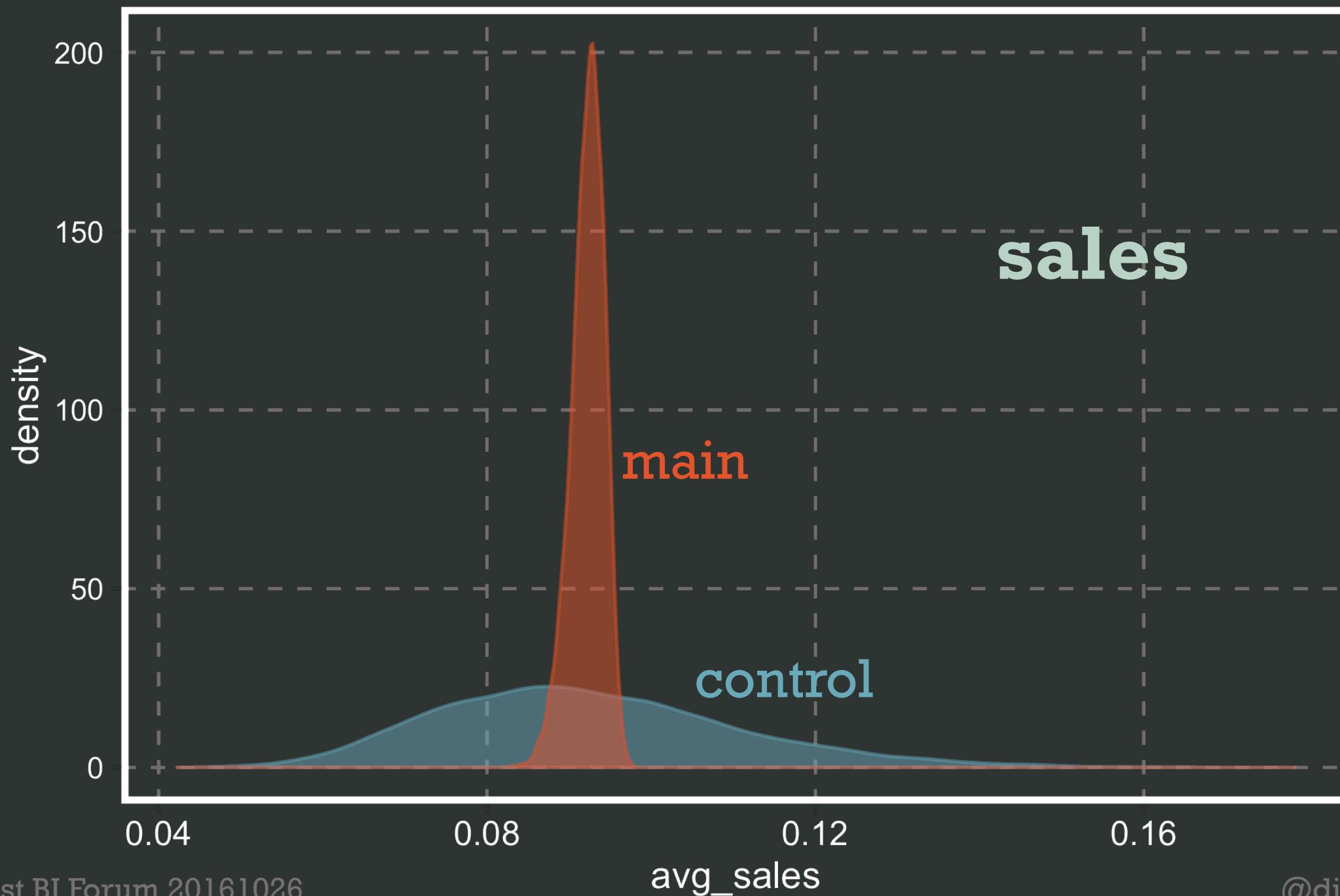
variation	average sales	improvement	lift	chance to beat baseline
main	0.0919	⬇️ -0.0037	-3.87%	10.0%
control	0.0956			

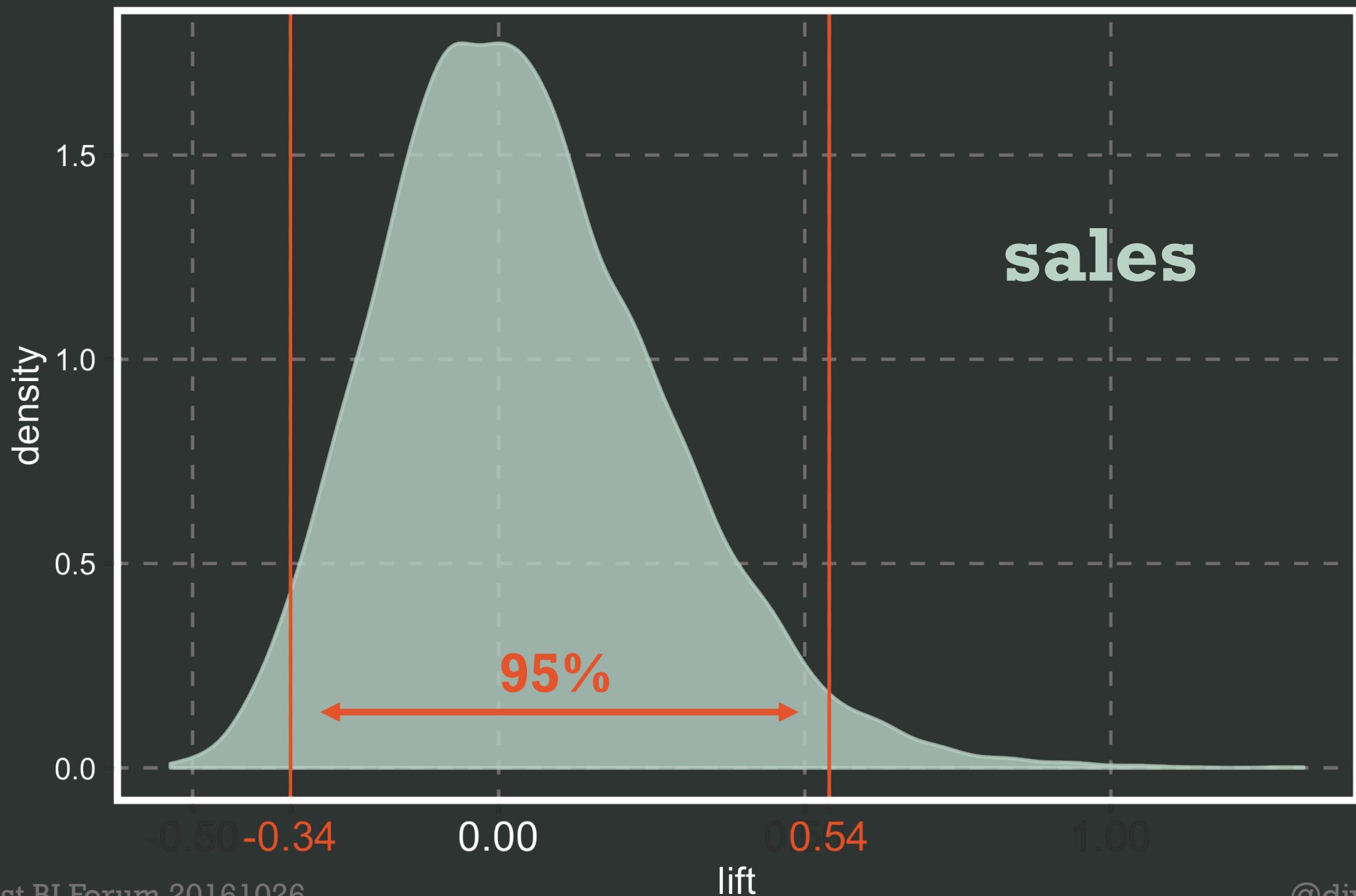
```
groupAA <- function(sales_dt, proportion = c(0.9, 0.1)) {  
  sales_dt[, .(  
    sales_amount,  
    simulated_AA_program = sample(  
      c('main', 'control'),  
      size = .N,  
      replace = TRUE,  
      prob = proportion  
    )  
  )]  
}
```

```
compareAA <- function(sales_dt, proportion = c(0.9, 0.1)) {  
  groupAA(sales_dt, proportion) %>%  
  .[  
    ,  
    .(avg_sales = mean(sales_amount)),  
    by = simulated_AA_program  
  ]  
}
```

```
compareAA <- function(sales_dt, proportion = c(0.9, 0.1)) {  
  groupAA(sales_dt, proportion) %>%  
  .[  
    ,  
    .(avg_sales = mean(sales_amount)),  
    by = simulated_AA_program  
  ]  
}
```

```
AA_results <- lapply(  
  1:10000,  
  function(i) compareAA(sales_dt)  
) %>% rbindlist()
```



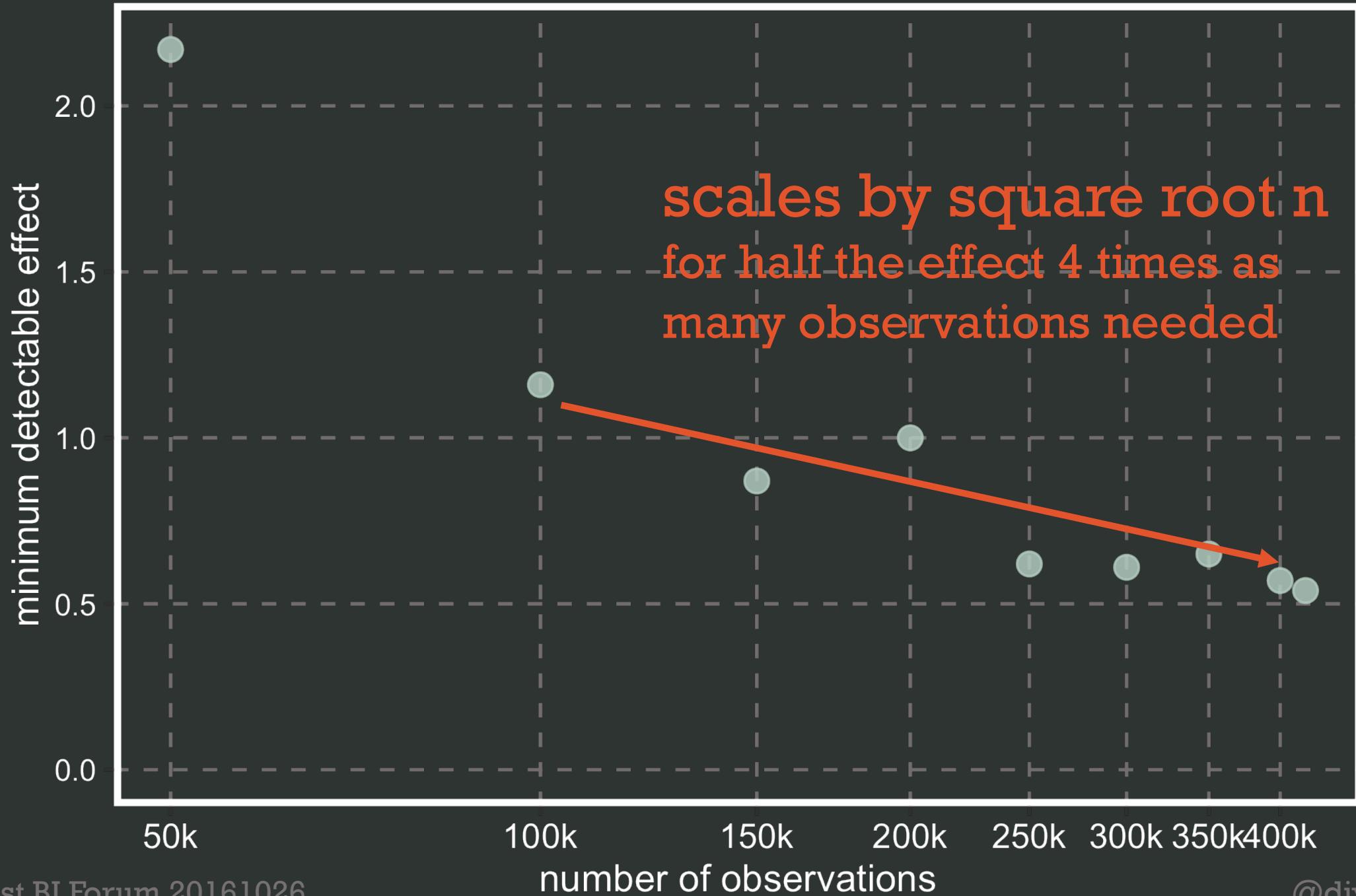


With a 90/10 split of ~450k customers  
we can detect  
an effect of at least 50%  
in average sales.

**number** of observations

**variation** of outcome

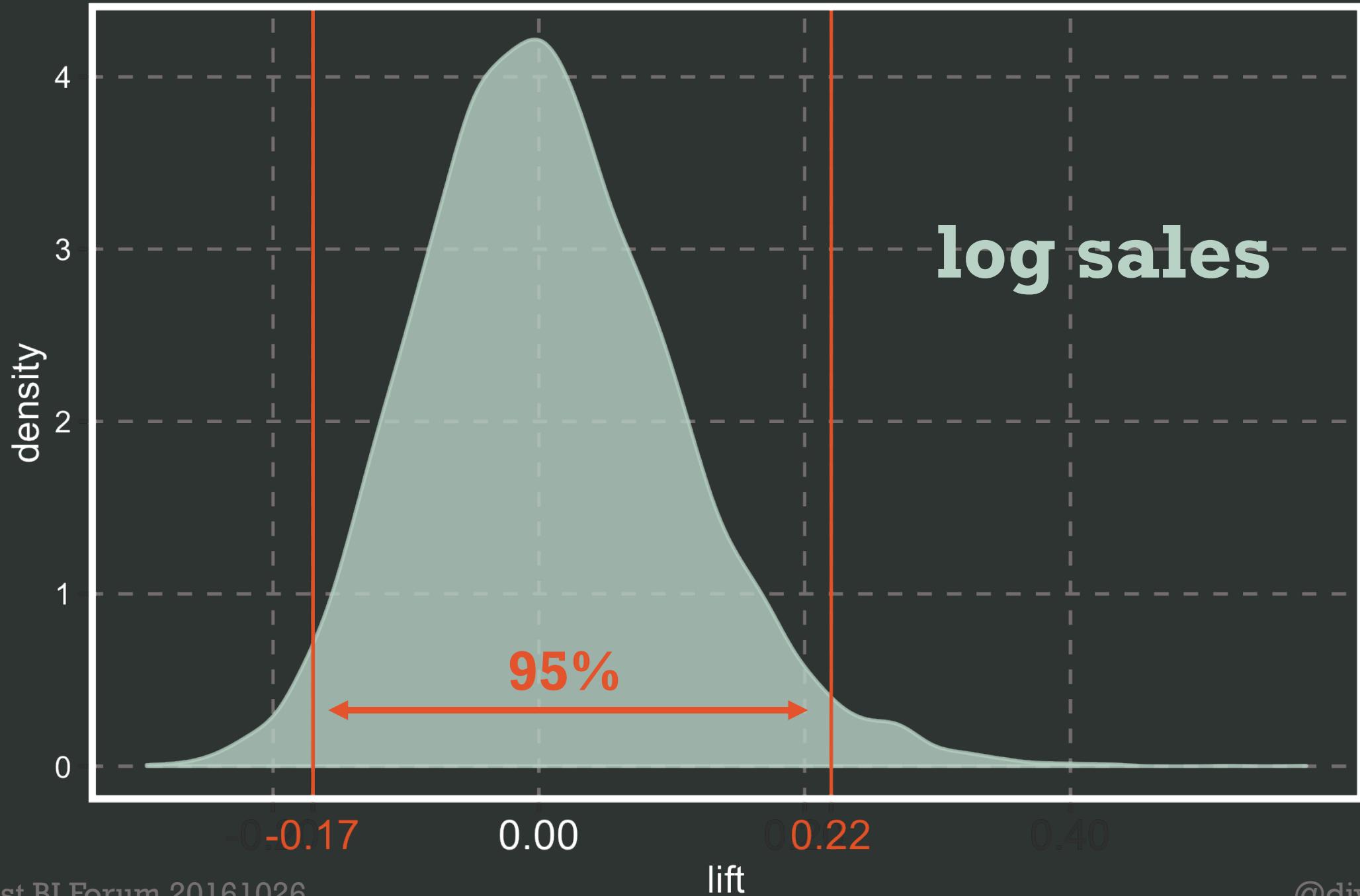
**size** of expected effect

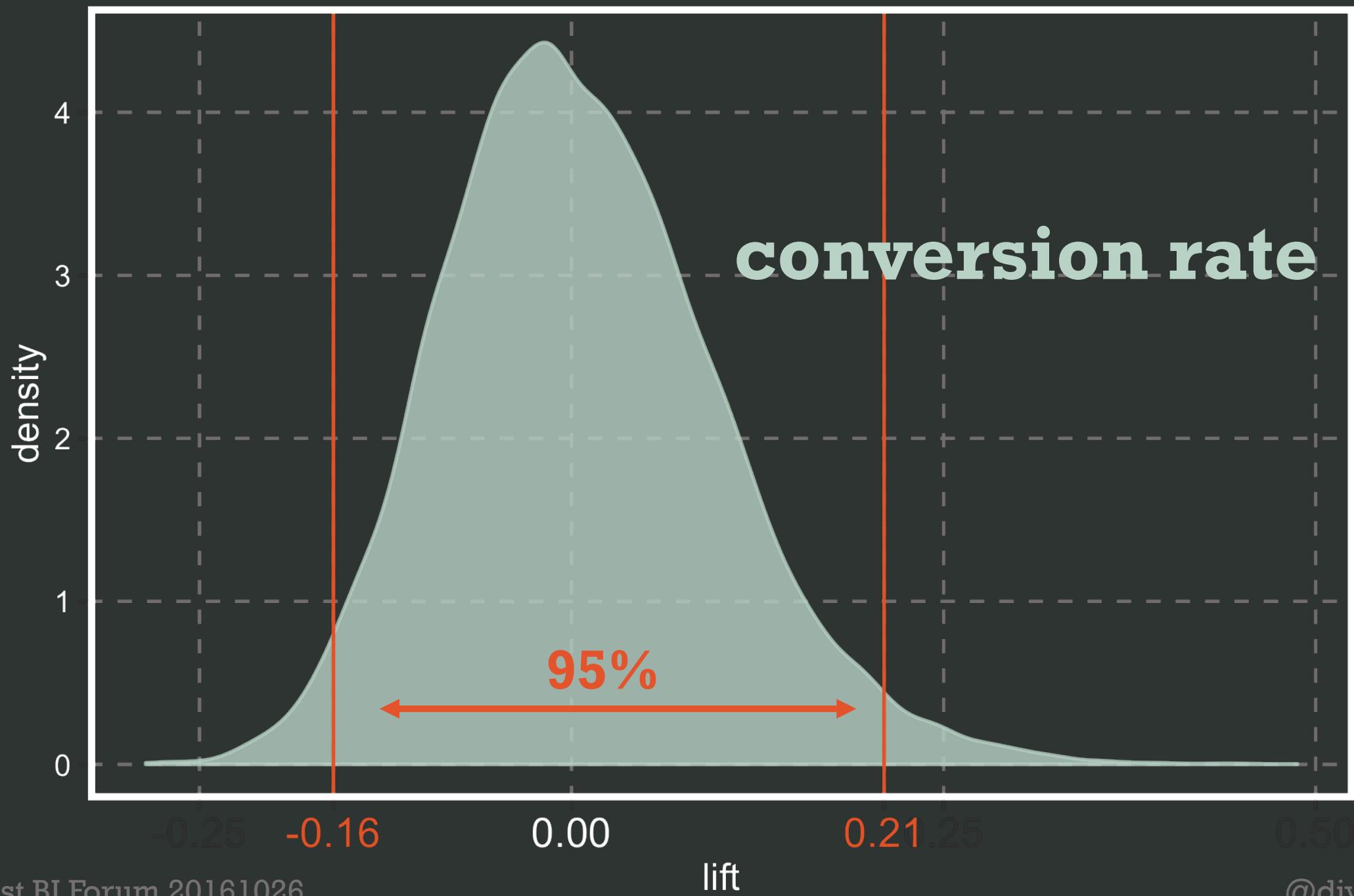


**number** of observations

**variation** of outcome

**size** of expected effect





<b>variation</b>	<b>average sales</b>	<b>improvement</b>	<b>lift</b>	<b>chance to beat baseline</b>
main	0.0919	⬇️ -0.0037	-3.87%	10.0%
control	0.0956			

	<b>avg log sales</b>				
main	0.0091	⬆️	0.0023	34.40%	99.7%
control	0.0068				

<b>variation</b>	<b>average sales</b>	<b>improvement</b>	<b>lift</b>	<b>chance to beat baseline</b>
main	0.0919	⬇️ -0.0037	-3.87%	10.0%
control	0.0956			
<b>avg log sales</b>				
main	0.0091	⬆️ 0.0023	34.40%	99.7%
control	0.0068			
<b>conversion</b>				
main	0.32%	⬆️ 0.09%p	41.26%	100.0%
control	0.23%			

**Randomized Block Design**  
**(aka Stratified Sampling)**  
ensures that less variable groups  
are being compared

number of observations

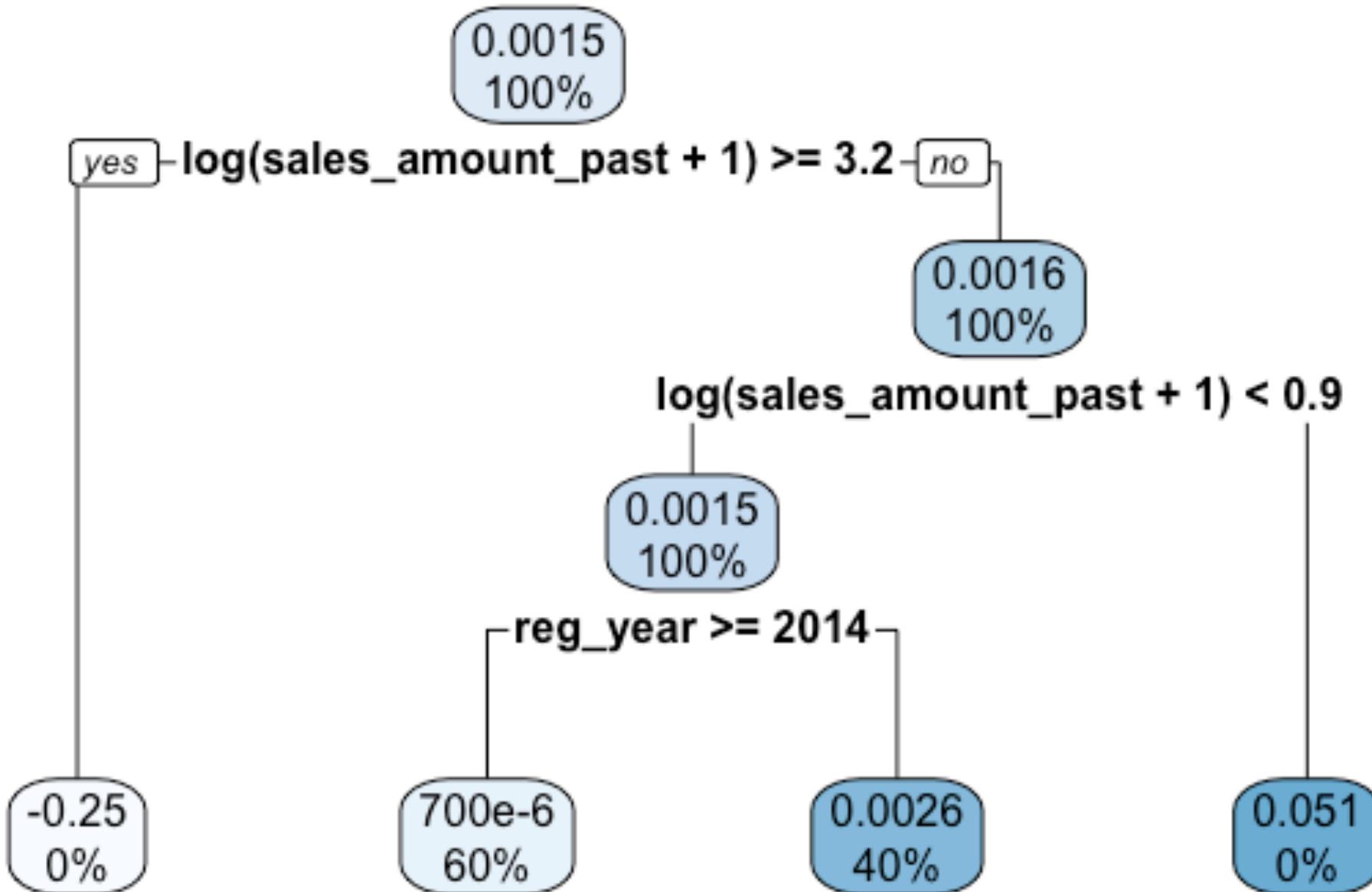
variation of outcome

size of expected effect

# CAUSAL TREE

to look for effect within segments  
while avoiding multiple testing

# CAUSAL TREE



# BAYESIAN A/B TESTING

$P(\text{data} \mid \text{versions are the same})$

# BAYESIAN A/B TESTING

$P(\text{data} \mid \text{versions are the same})$

$P(\text{versions are the same} \mid \text{data})$