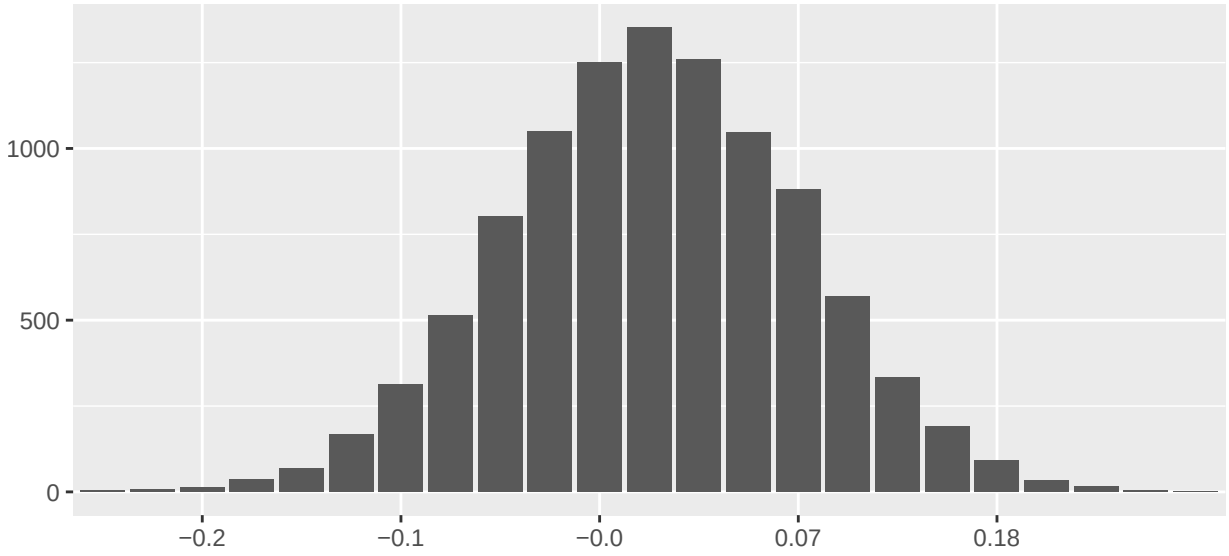
Quiz: Inference for Categorial Data

A study found that the use of bed nets was associated with a lower prevalence of malarial infections in the Gambia. Specifically, they considered a simple random sample of 154 people in the city of Banjul. They sent an inspector to each sampled resident and asked subjects whether they slept under a bed net and whether or not they had a malarial infection at any point in the past. The results of the study are below.

```
## # A tibble: 4 x 3
##   treatment malaria     n
##   <fct>     <fct>   <int>
## 1 net       no         36
## 2 net       yes        34
## 3 no_net    no         21
## 4 no_net    yes        63
```

1. Suggest a graphic that could be used to visualize the association between malarial infection and the use of a bed net (no need to sketch it).
2. Compute the proportion of people using a bed net that suffered from malaria and the proportion of people that didn't use a net that suffered from malaria.
3. Write out the hypotheses that reflect the research question of an association between these two variables, phrased in terms of the difference between two proportions.
4. The null distribution on the next page represents the collection of test statistics that result from 10,000 permuted data sets under H_0. Please label the axes and draw a vertical line representing the observed test statistic.



5. What is your decision regarding the viability of the null hypothesis (do your best to estimate a two-tailed p-value)?
6. Here we used permutation to build up the null distribution, but it is more common to use a mathematical approximation. What distribution would you use for this approximation and what conditions would you want to check before using it to compute a p-value?
7. Provide a critique of the following interpretation of the study, which appeared in a newspaper article: "This study provides strong evidence that the cheapest way to combat Malaria in West Africa is to distribute bed nets."