Simulations 2.1.1

Lauren Quesada

11/8/2021

2.1.1

```
###Load packages

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(broom)

###Define function

LFunky211<- function(beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle) {

  ###Define variables

  set.seed(4747)
  sig<- 0.05


    #If there is no intercept in the fitted regression, d.o.f. = (n-1)
    #otherwise, df=n-k-1; k=the number of variables
  dof <- n_obs-1

  simpp_val<-c()

  simp_t<-c()

  perm_pval<-c()
```

```r
###Begin loop

for(i in 1:n_data) {               # Start of the Monte Carlo loop

###Define Variables

x1<- rnorm(n_obs)     # Artificial x series, created just once

x2<- x1 + rnorm(n_obs)     ## Did we figure out why these need to be defined in the loop?

y<- beta0 + beta1*x1 + beta2*x2 + beta3*x1*x2 + rnorm(n_obs)   # The DGP includes an intercept

###Fit a linear regression model

simpfit<- lm(y ~ x1*x2 )

###Record slope and p-value

#observed p.value for each dataset (n_data)
simpp_val<- c(simpp_val,
              simpfit %>% tidy() %>% filter(term =="x1") %>% select(p.value) %>% pull())

#observed slope for each dataset (n_data)
simp_t<- c(simp_t,
              simpfit %>% tidy() %>% filter(term =="x1") %>% select(statistic) %>% pull())

#obs value for most recent dataset
obsval<- simpfit %>% tidy() %>% filter(term =="x1") %>% select(statistic) %>% pull()

#reassign vector so it doesn't add onto the concatenation each i in 1:n_data loop
simpp_valperm<-c()

  ###Begin another loop

  for(j in 1: n_shuffle) {
# A mis-specified model is estimated (unless Beta0 = 0)


  ###Fit linear regression model to PERMUTED data

  simpfitperm<- lm(y ~ sample(x1))

  ###Record slope
  simpp_valperm<- c(simpp_valperm,
                    simpfitperm %>% tidy() %>% filter(term == "sample(x1)") %>% select(statistic) %>%

  }

#p-value for each of my 200 datasets as compared to my permuted distribution of p-values
perm_pval<-c(perm_pval,mean(simpp_valperm>obsval))

}  #End of the Monte Carlo Loop
```

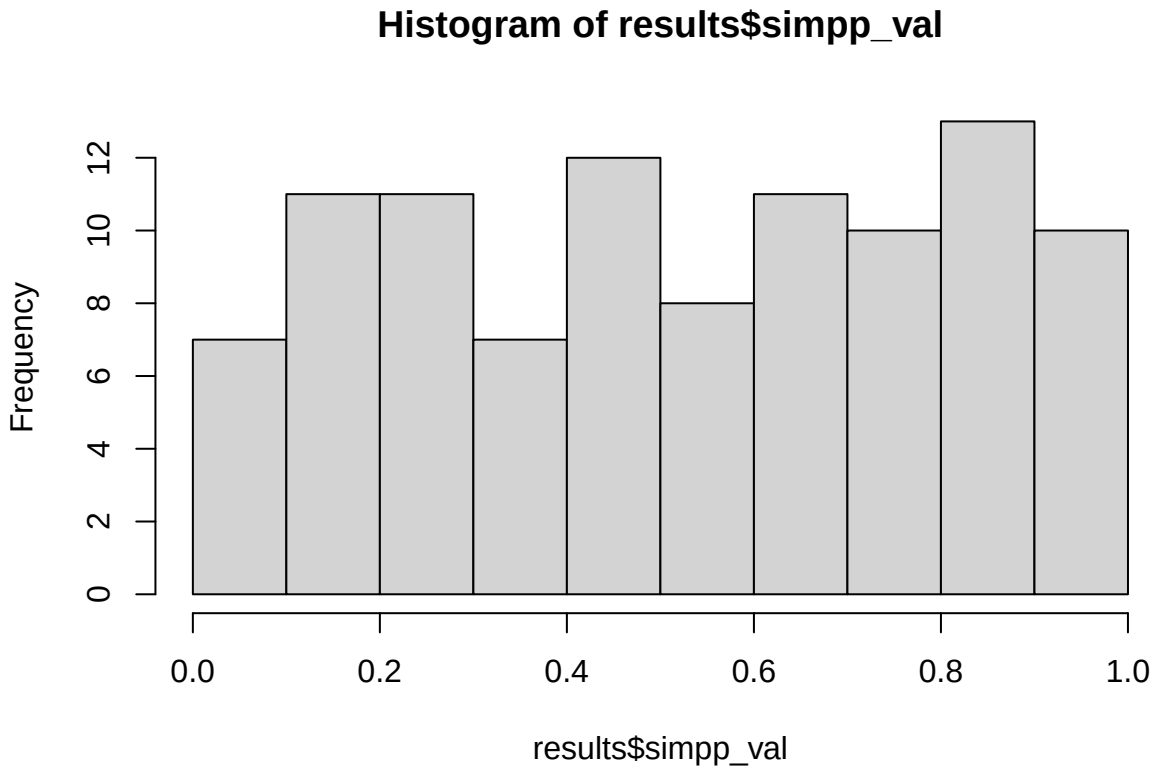
```
#coefficients, t-test p.values, permutation p.values
data.frame(simp_t, simpp_val, perm_pval)




}

#results %>% ggplot() + geom_hist(aes(x = simppval))
```

For Beta_3

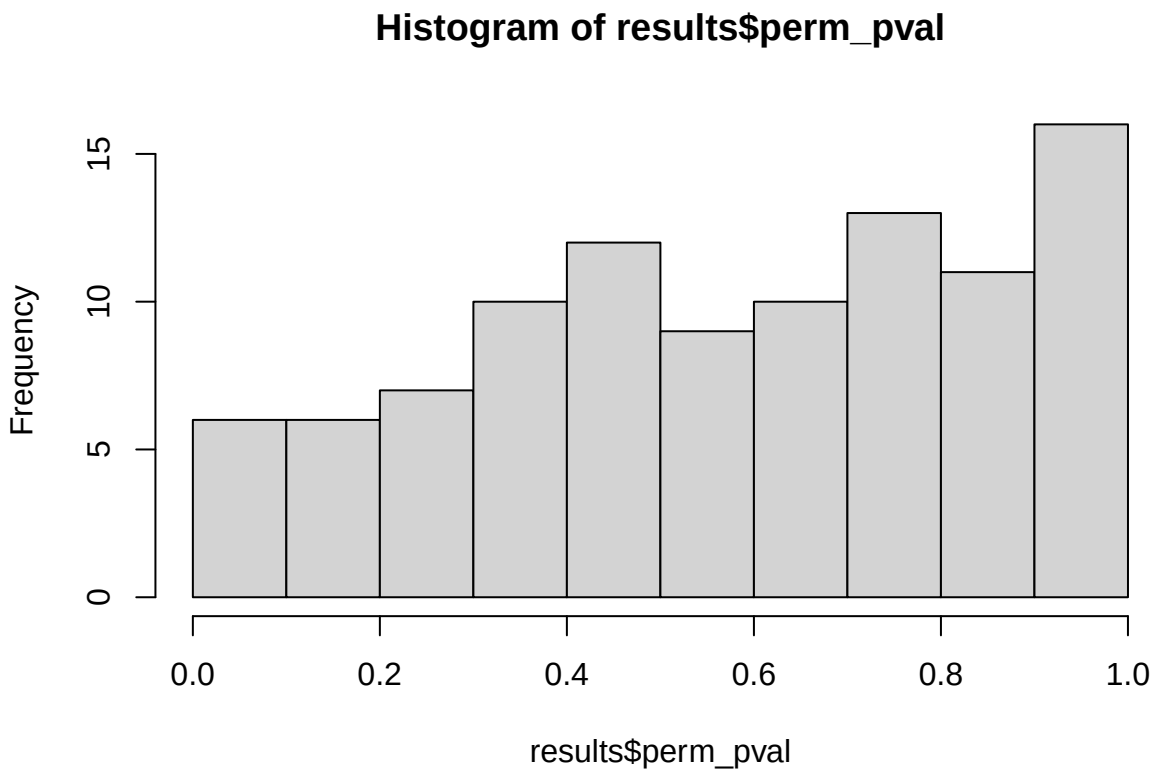
For n=15

$\beta_3 = 0.1$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.1,15,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

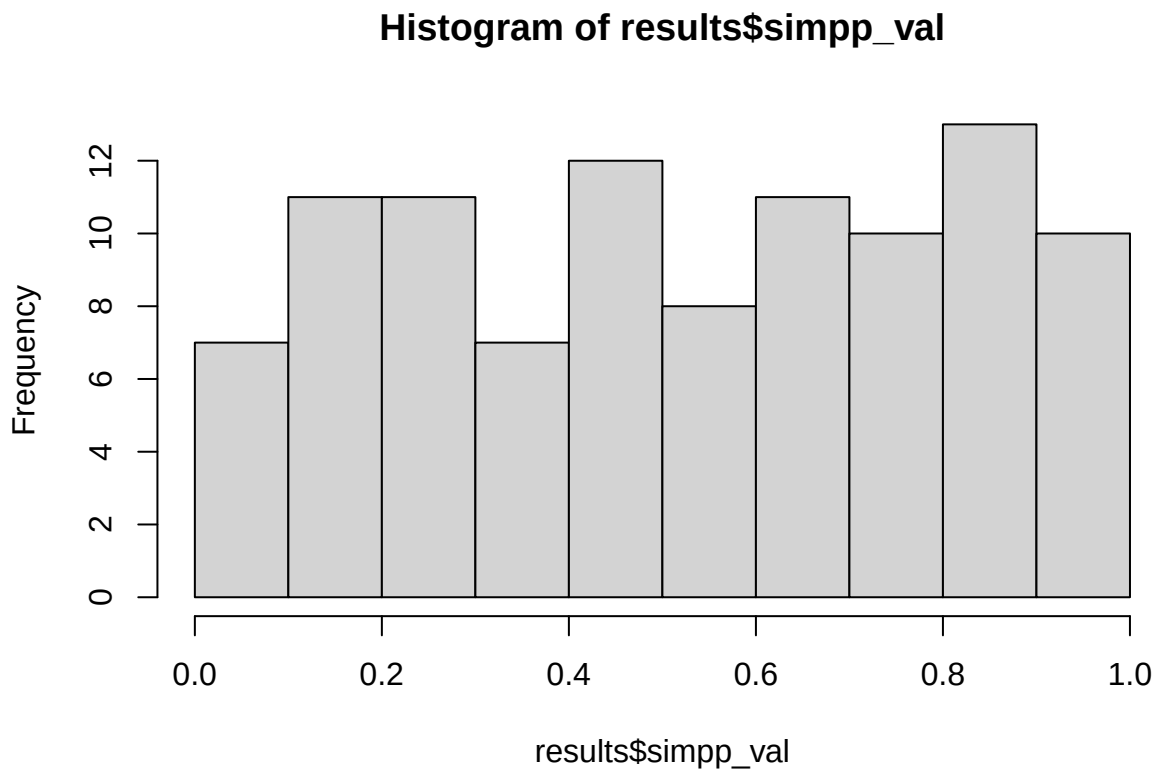
```
## [1] 0.03
```

```
mean(results$perm_pval < 0.05)
```

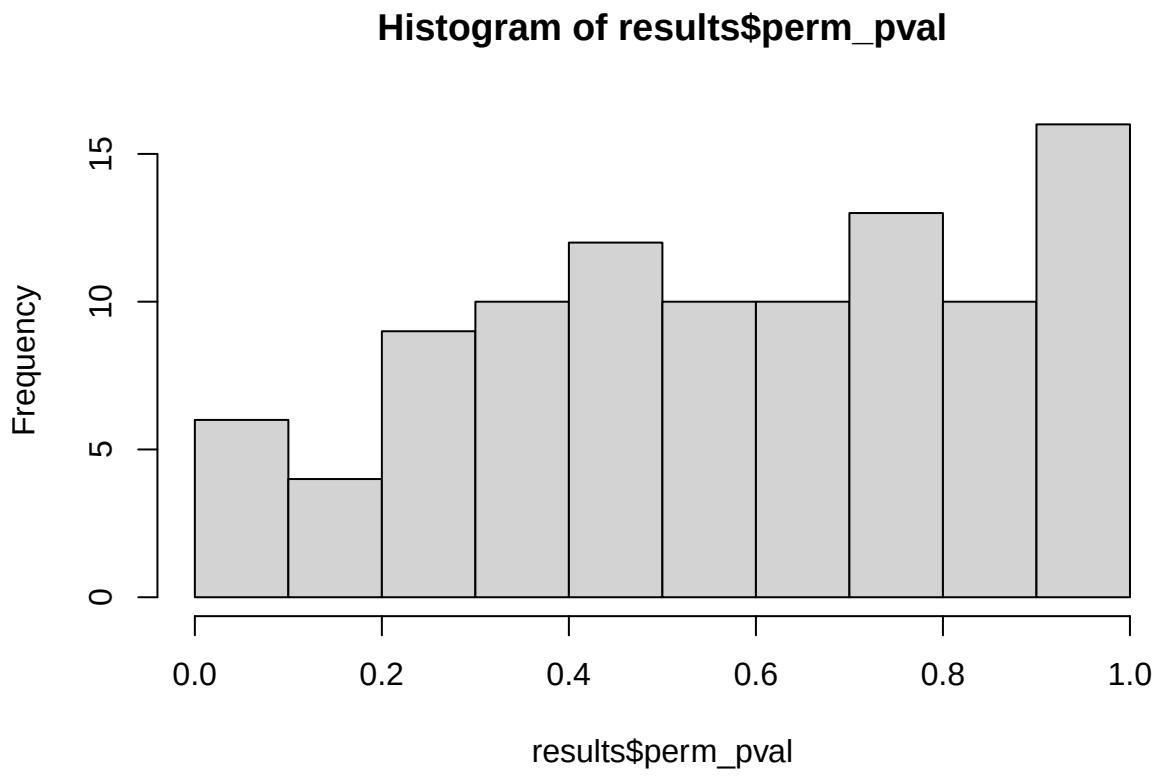
```
## [1] 0.02
```

$\beta_3 = 0.2$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.2,15,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

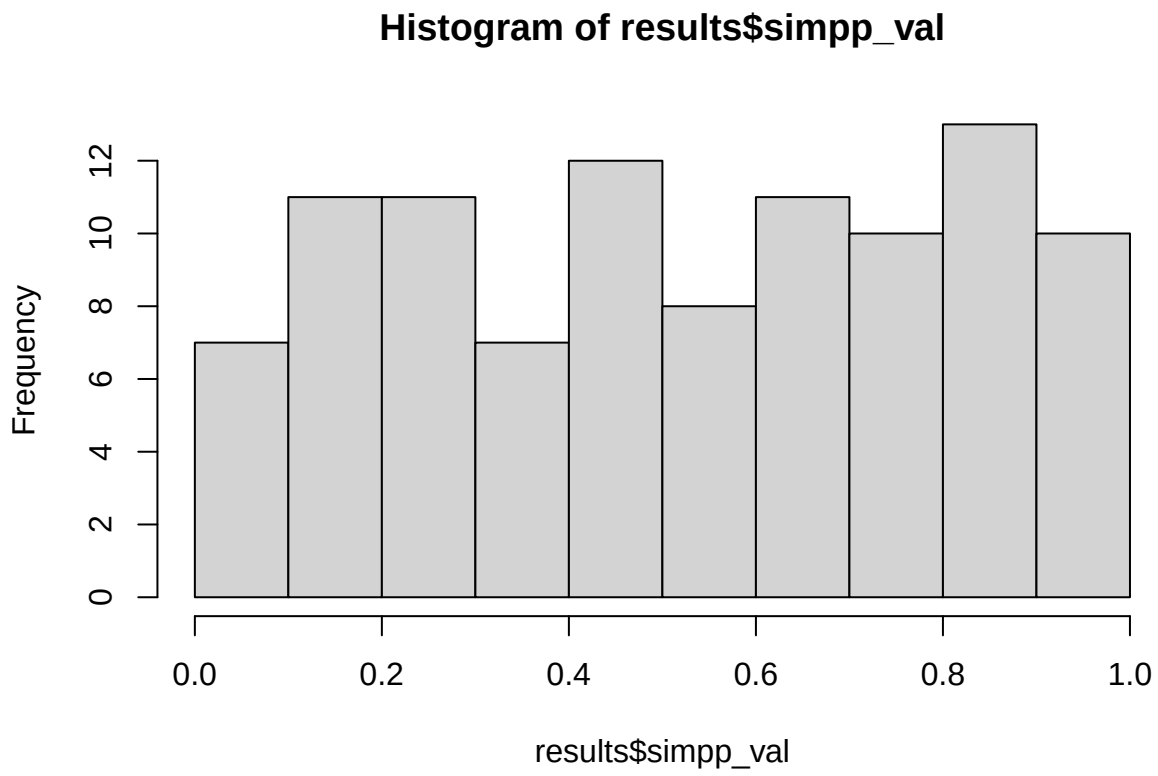
```
## [1] 0.03
```

```
mean(results$perm_pval < 0.05)
```

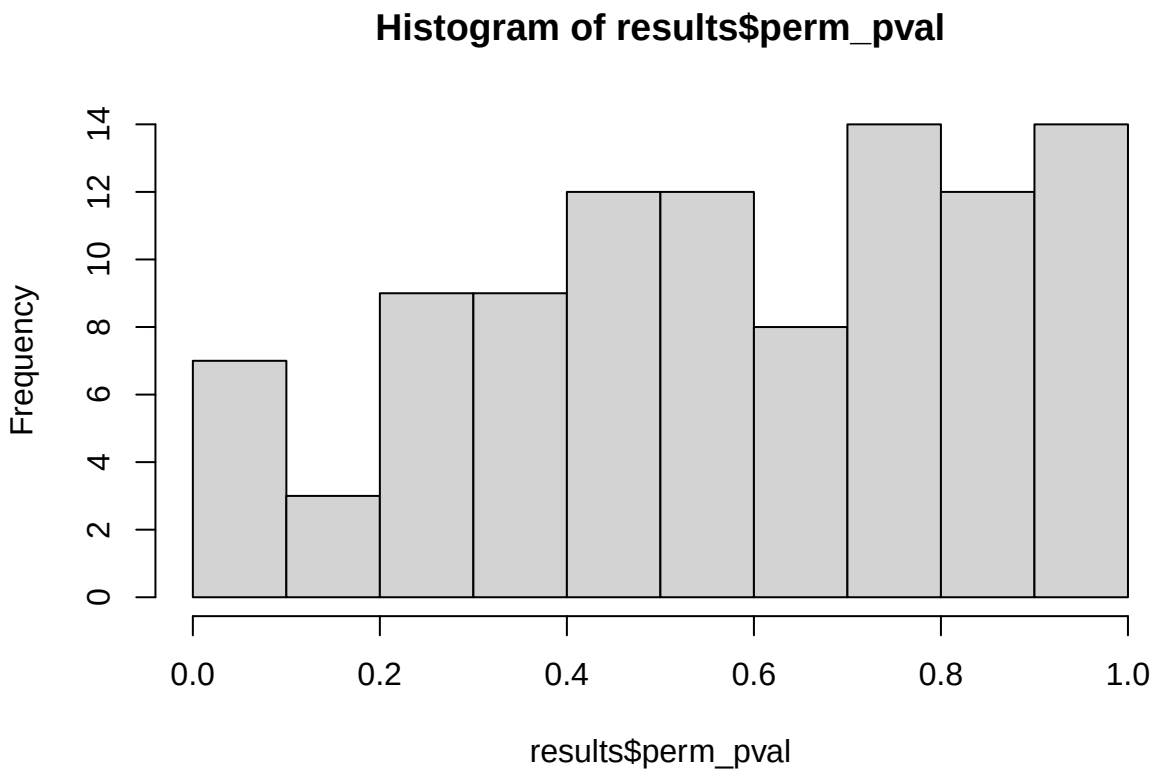
```
## [1] 0.02
```

$\beta_3 = 0.3$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.3,15,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

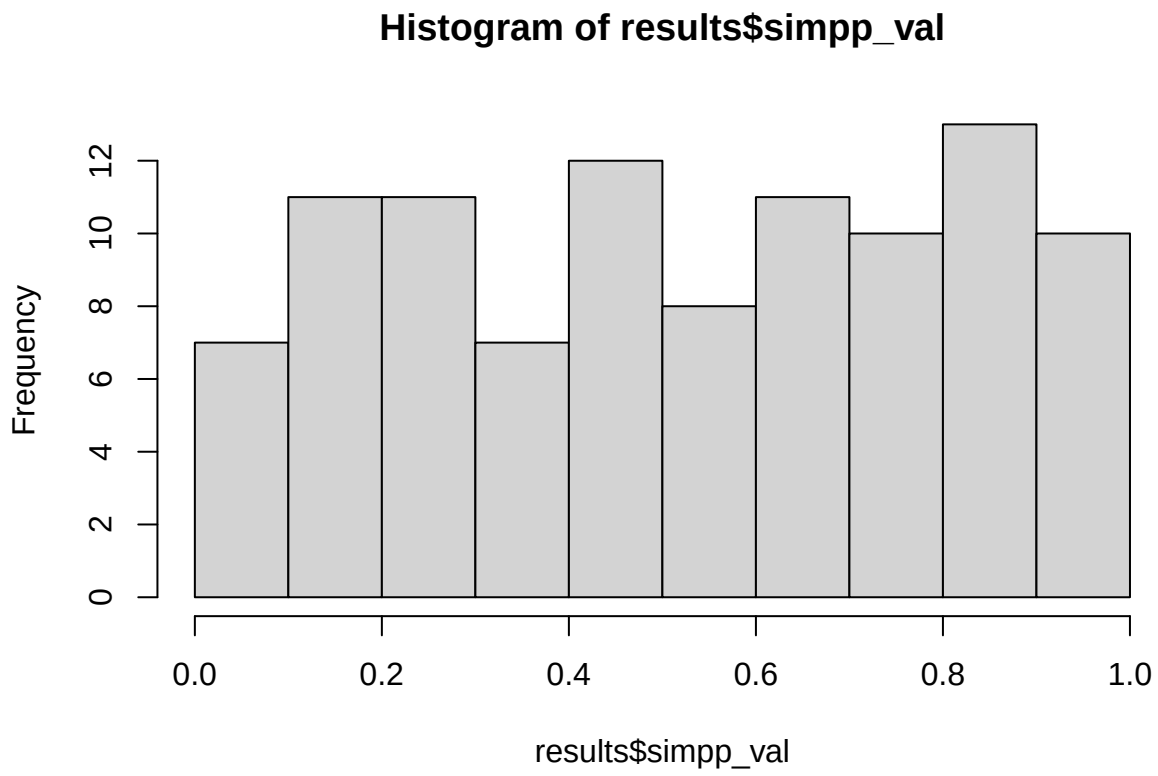
```
## [1] 0.03
```

```
mean(results$perm_pval < 0.05)
```

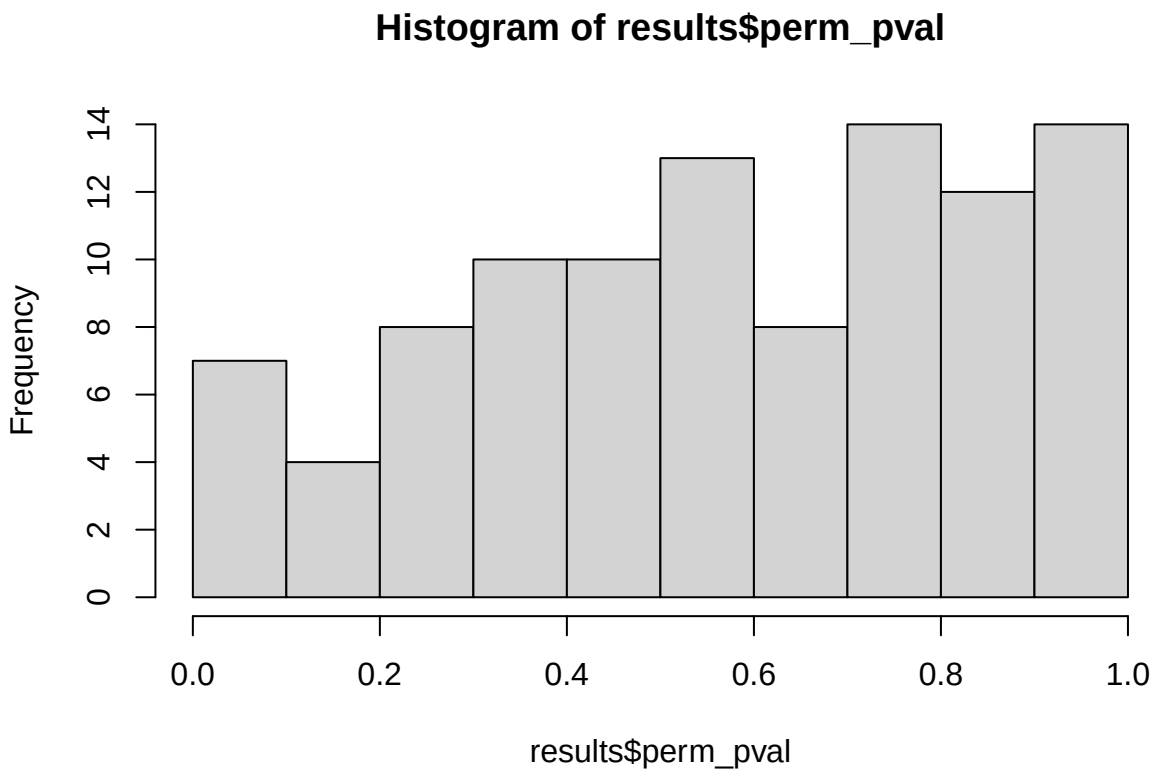
```
## [1] 0.02
```

$\beta_3 = 0.4$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.4,15,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

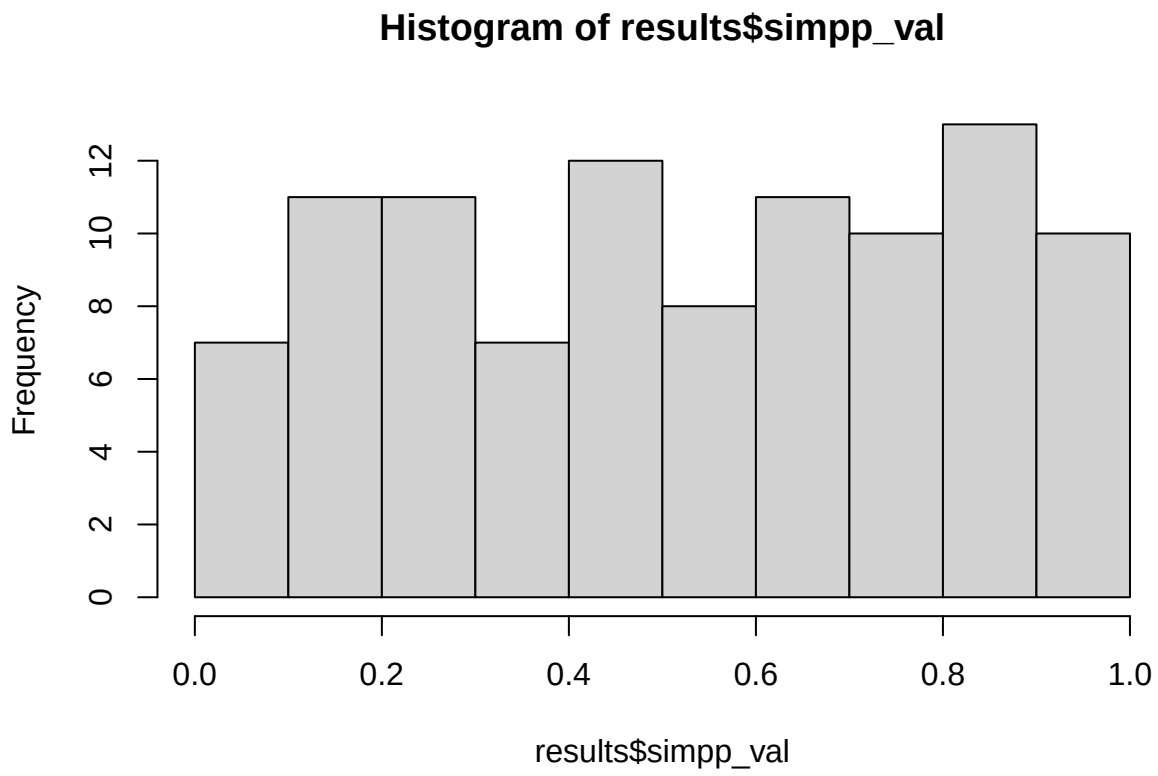
```
## [1] 0.03
```

```
mean(results$perm_pval < 0.05)
```

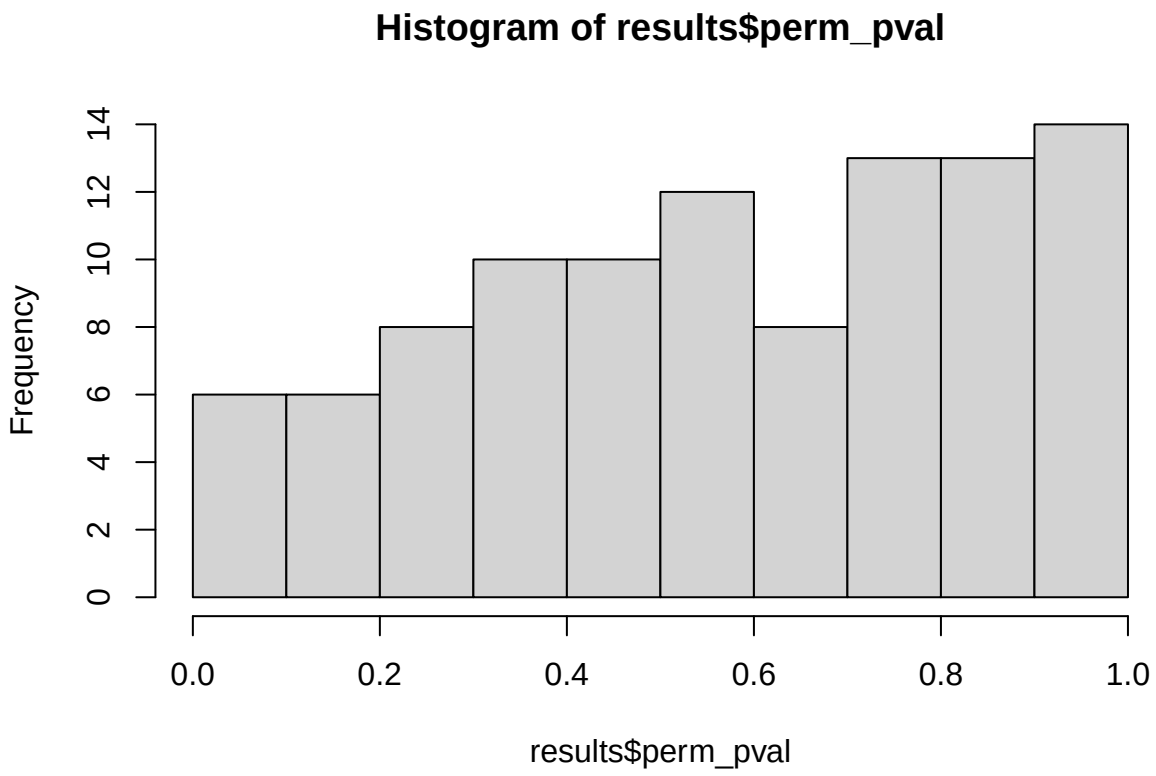
```
## [1] 0.03
```

$\beta_3 = 0.5$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.5,15,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

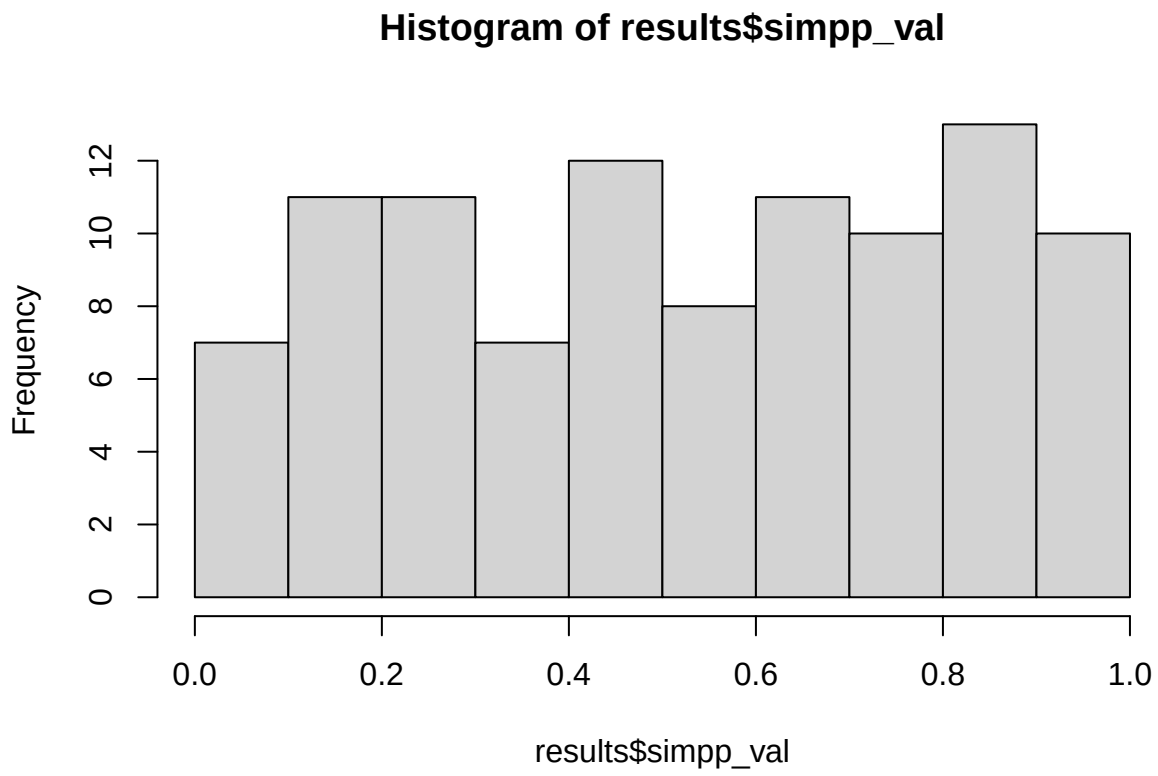
```
## [1] 0.03
```

```
mean(results$perm_pval < 0.05)
```

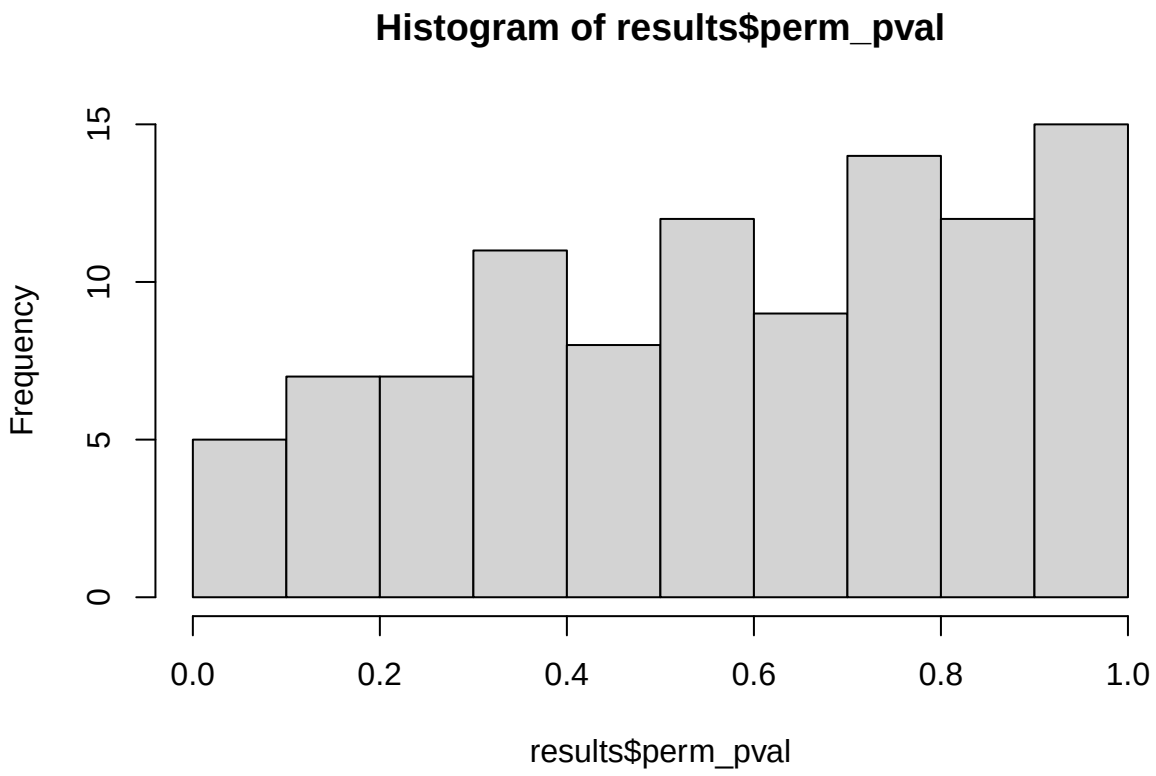
```
## [1] 0.02
```

$\beta_3 = 0.6$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.6,15,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

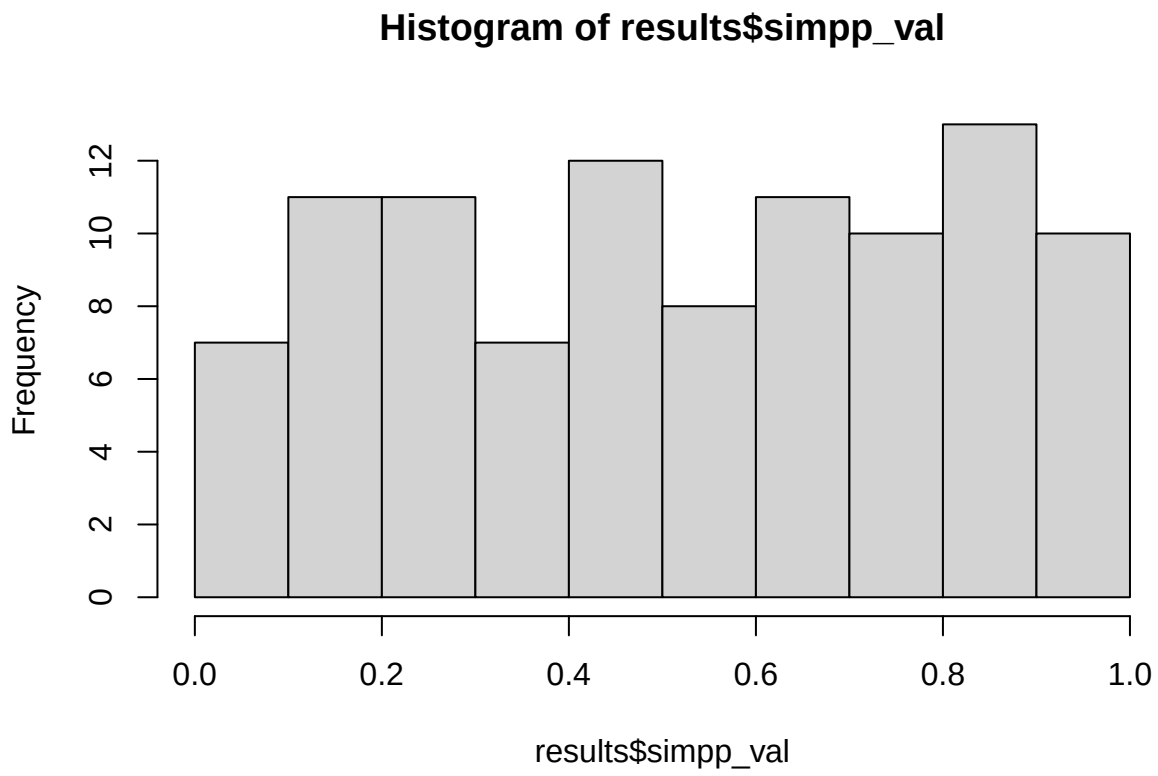
```
## [1] 0.03
```

```
mean(results$perm_pval < 0.05)
```

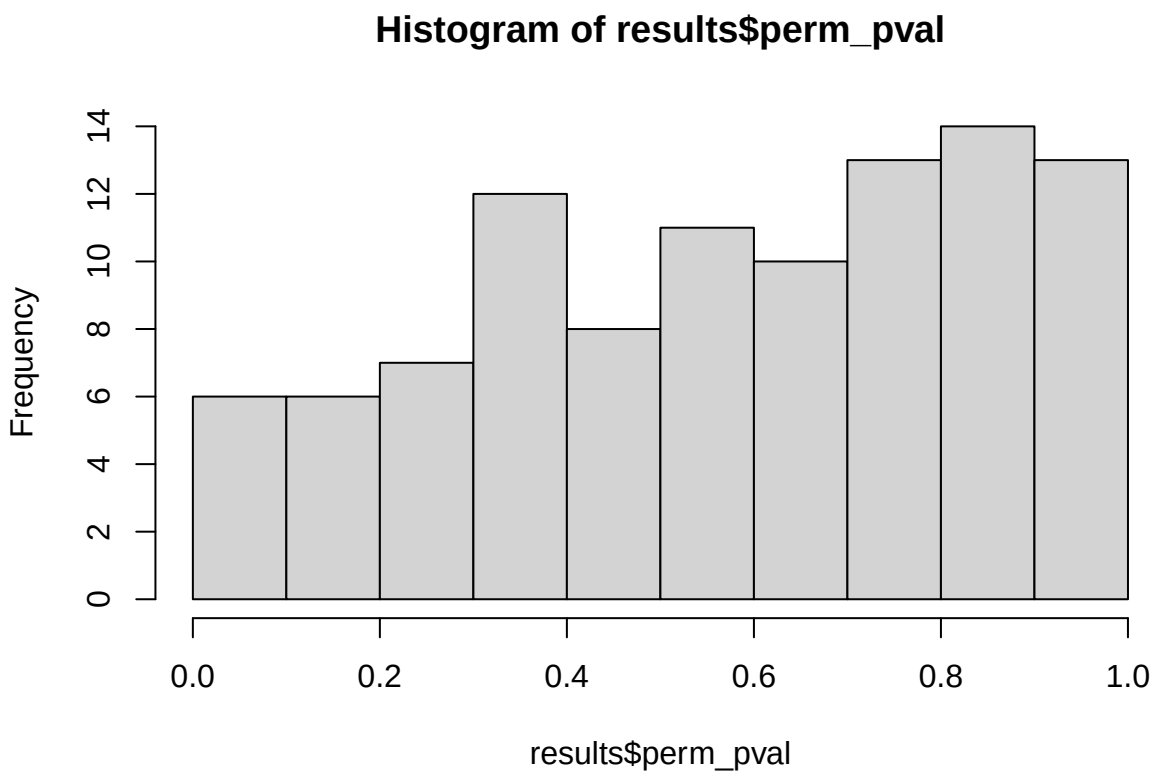
```
## [1] 0.02
```

$\beta_3 = 0.7$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.7,15,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

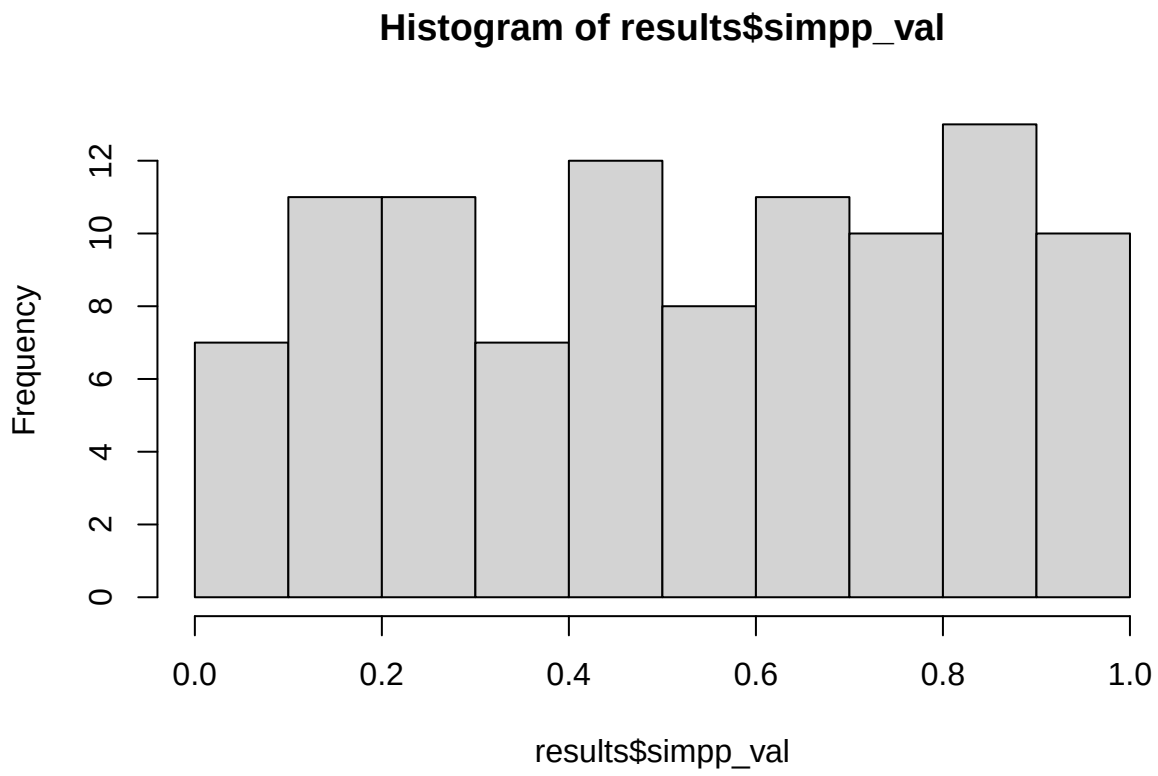
```
## [1] 0.03
```

```
mean(results$perm_pval < 0.05)
```

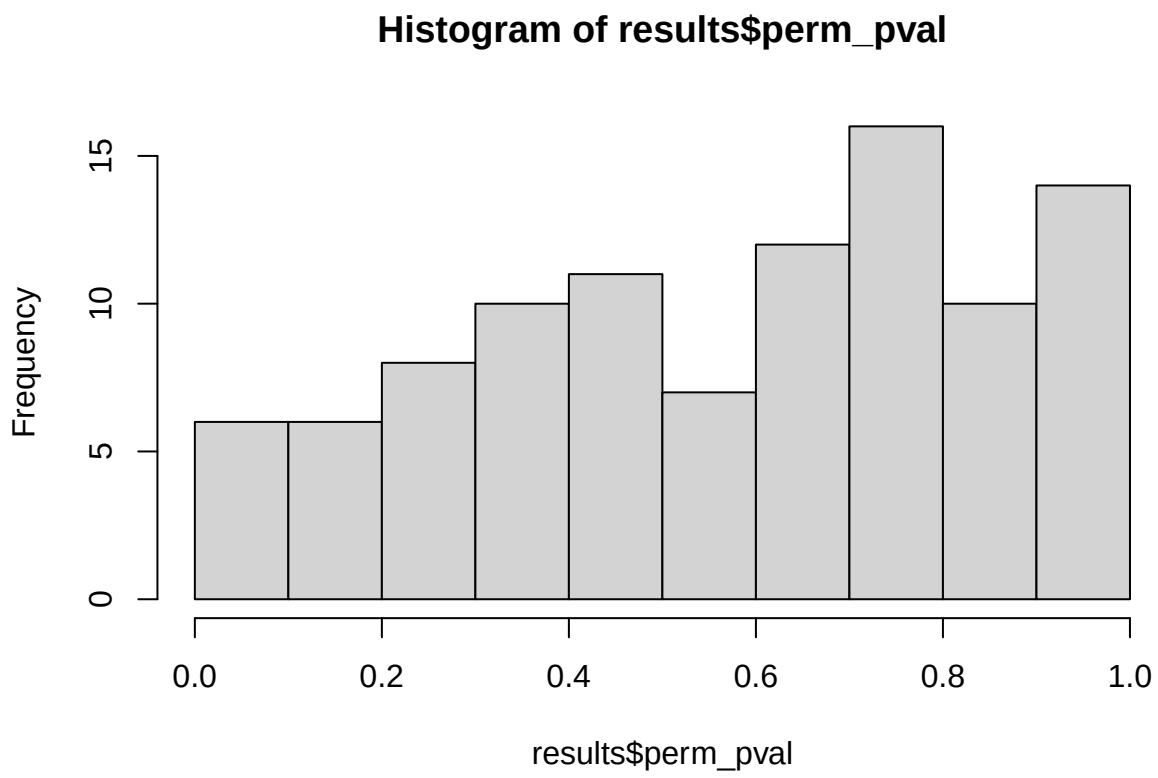
```
## [1] 0.03
```

$\beta_3 = 0.8$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.8,15,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

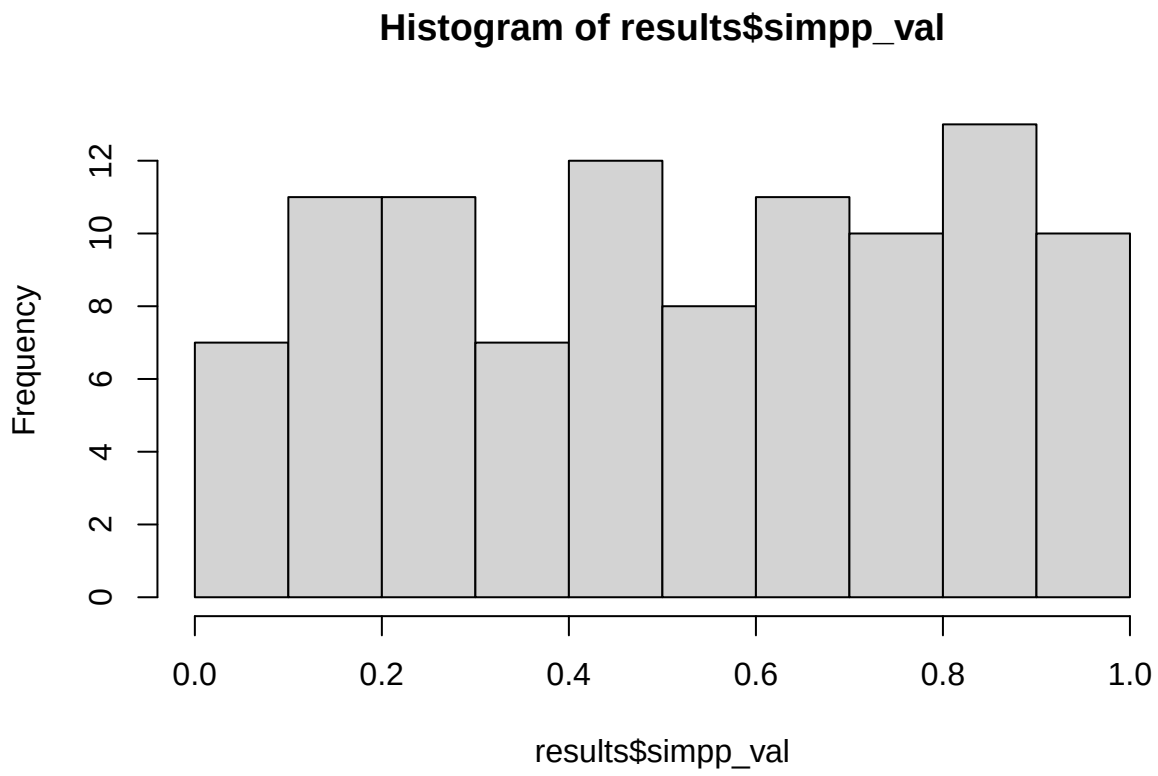
```
## [1] 0.03
```

```
mean(results$perm_pval < 0.05)
```

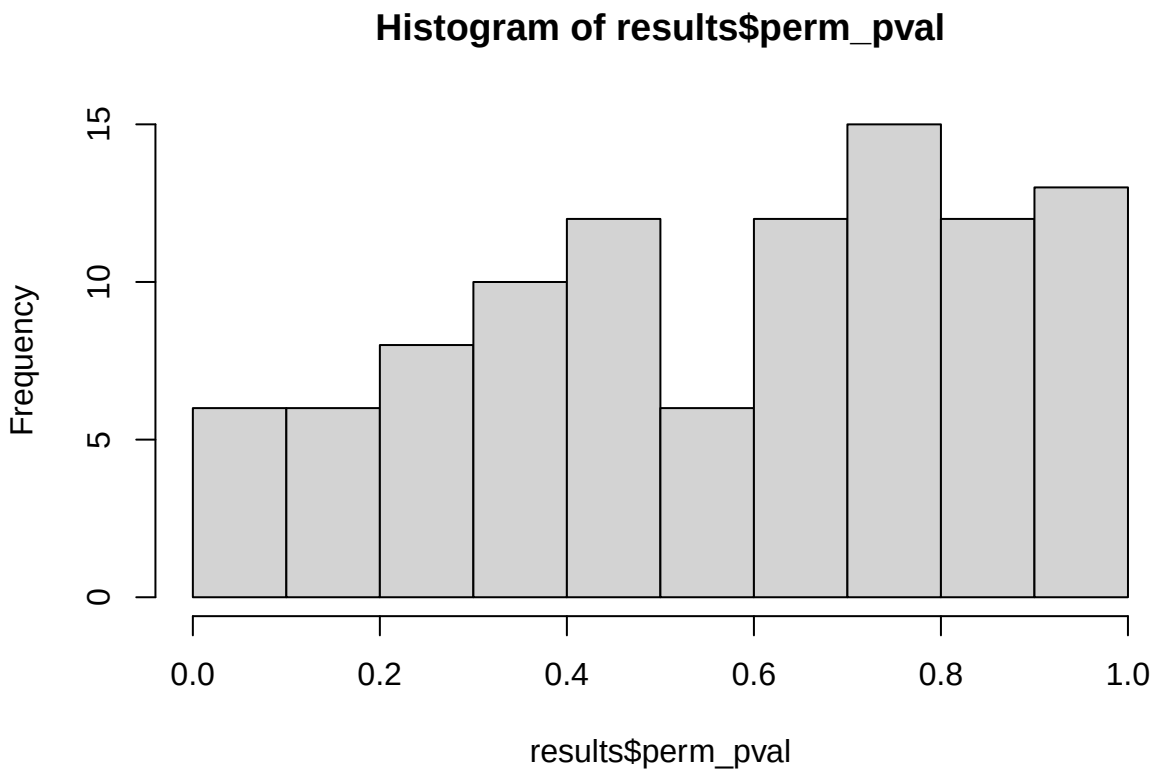
```
## [1] 0.02
```

$\beta_3 = 0.9$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.9,15,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

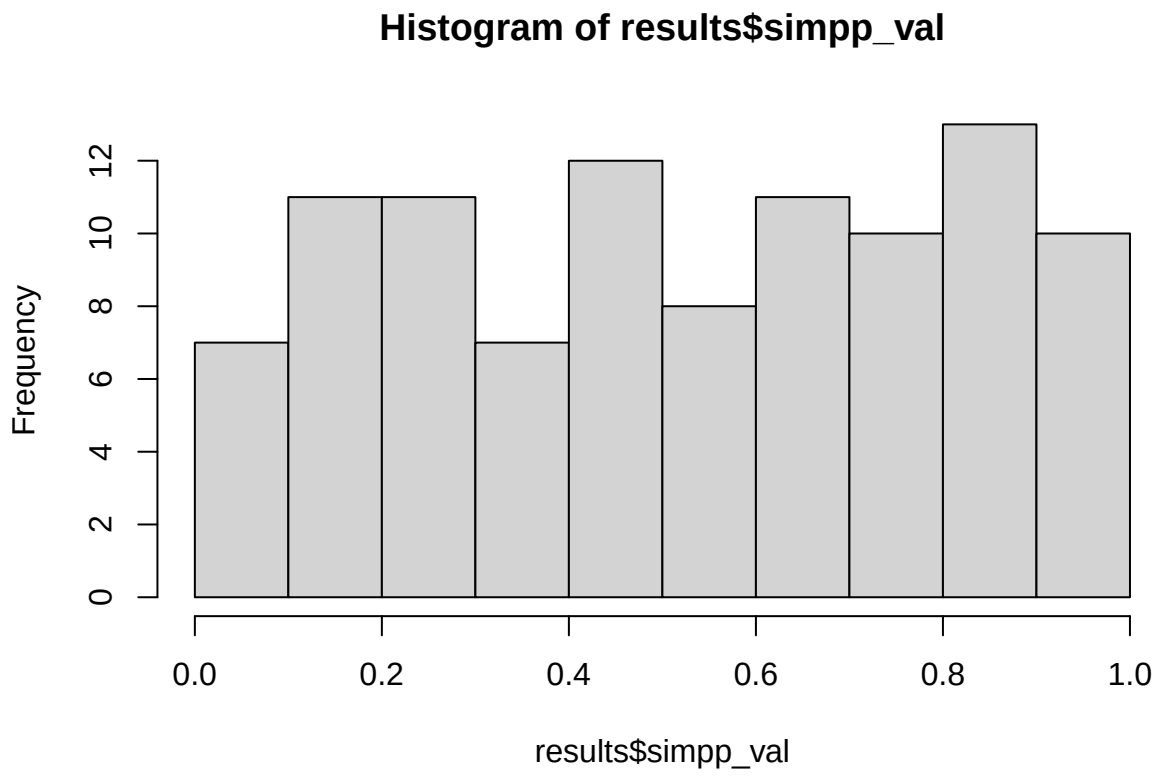
```
## [1] 0.03
```

```
mean(results$perm_pval < 0.05)
```

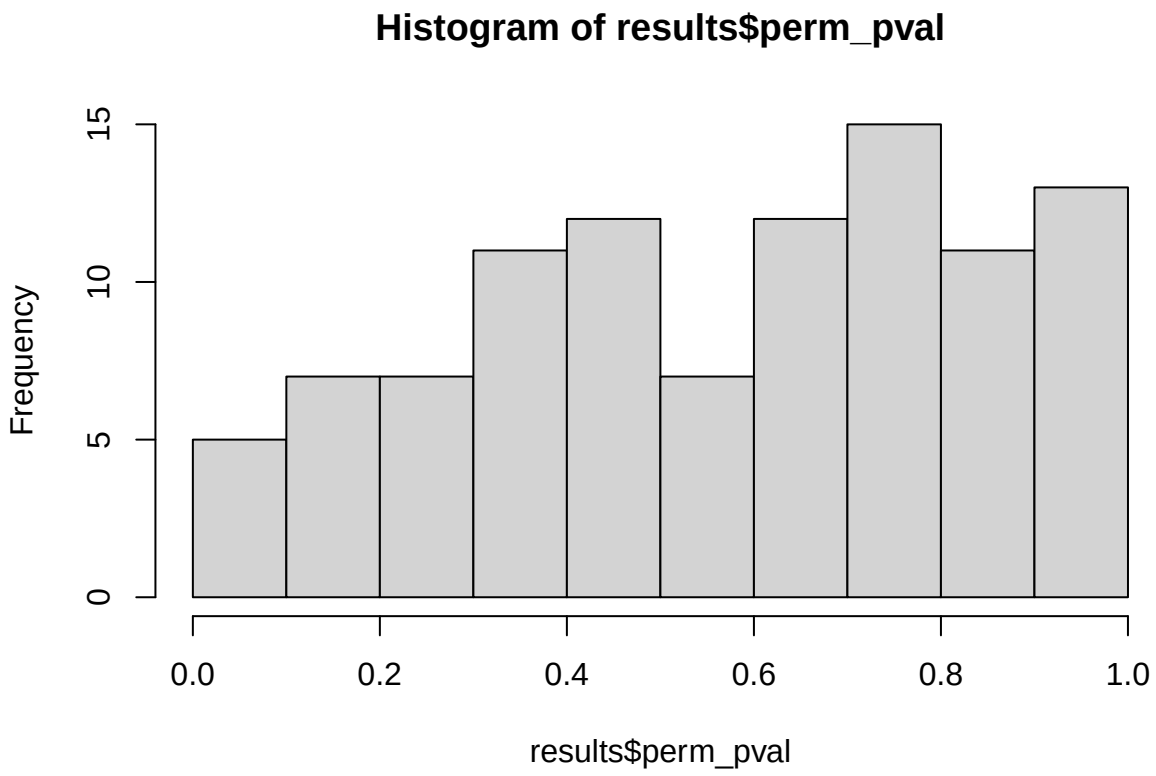
```
## [1] 0.02
```

$\beta_3 = 1.0$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,1.0,15,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

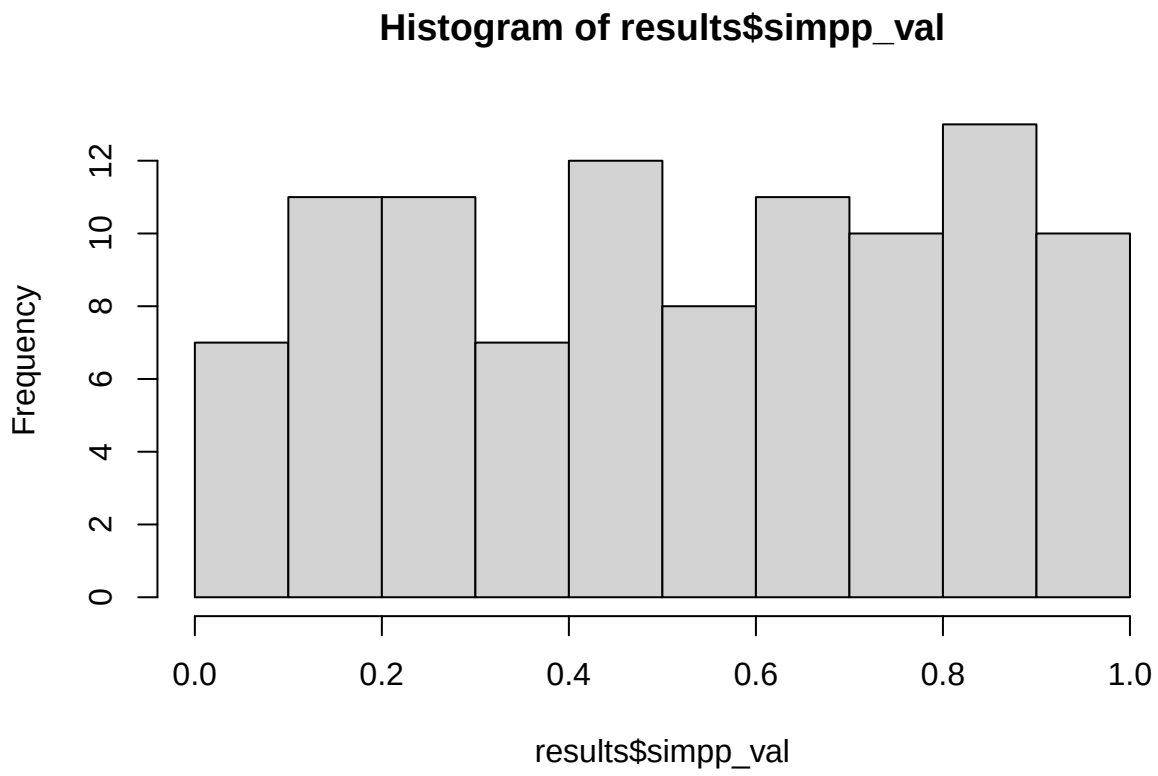
```
## [1] 0.03
```

```
mean(results$perm_pval < 0.05)
```

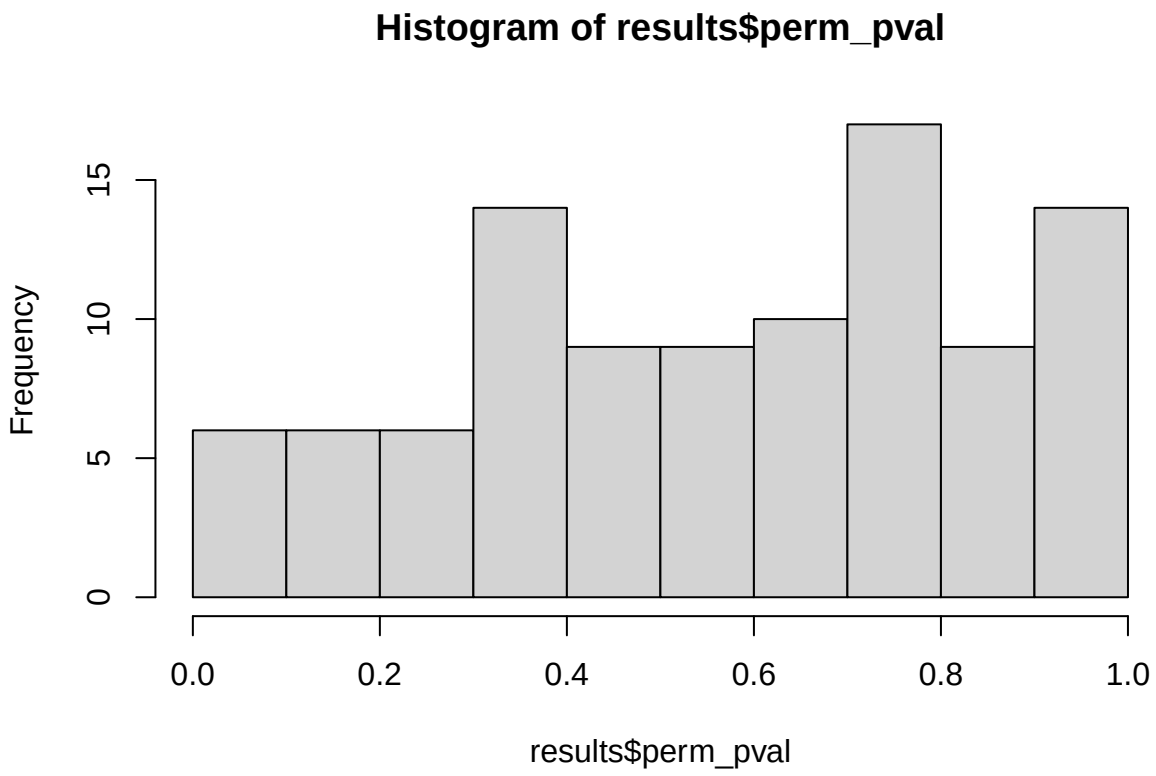
```
## [1] 0.02
```

$\beta_3 = 1.1$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,1.1,15,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

```
## [1] 0.03
```

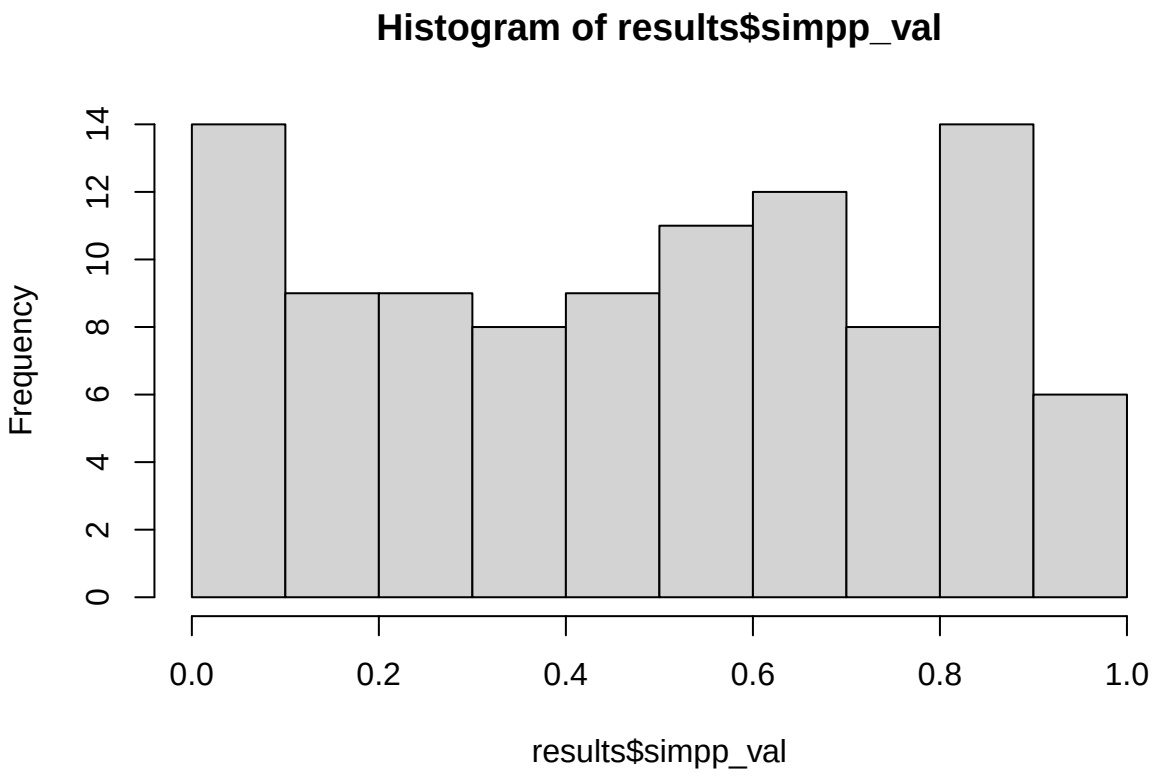
```
mean(results$perm_pval < 0.05)
```

```
## [1] 0.02
```

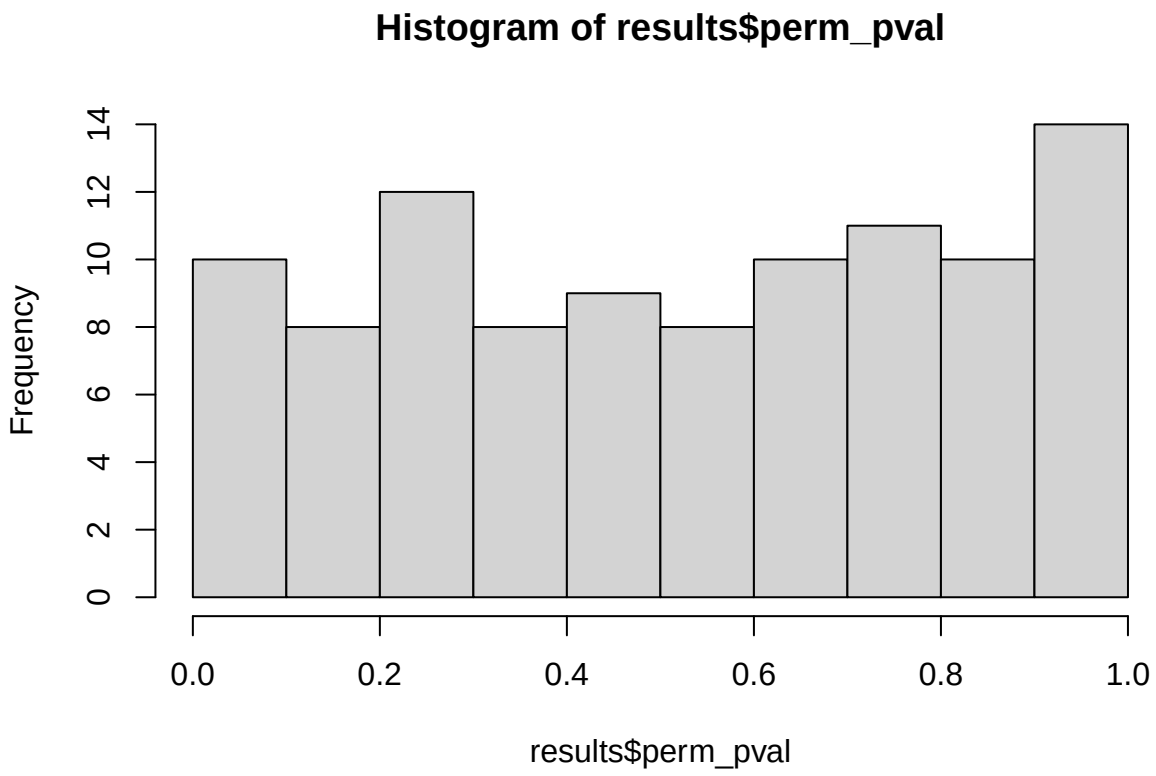
For n=30

$\beta_3 = 0.1$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.1,30,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

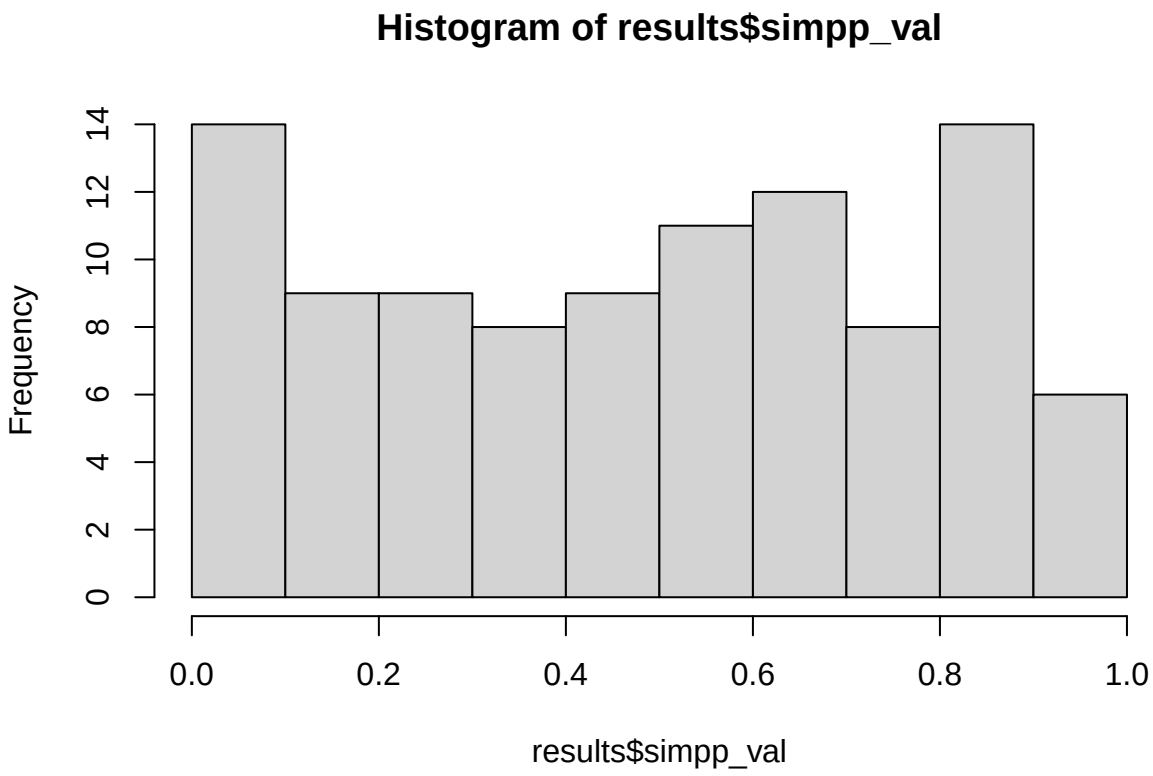
```
## [1] 0.06
```

```
mean(results$perm_pval < 0.05)
```

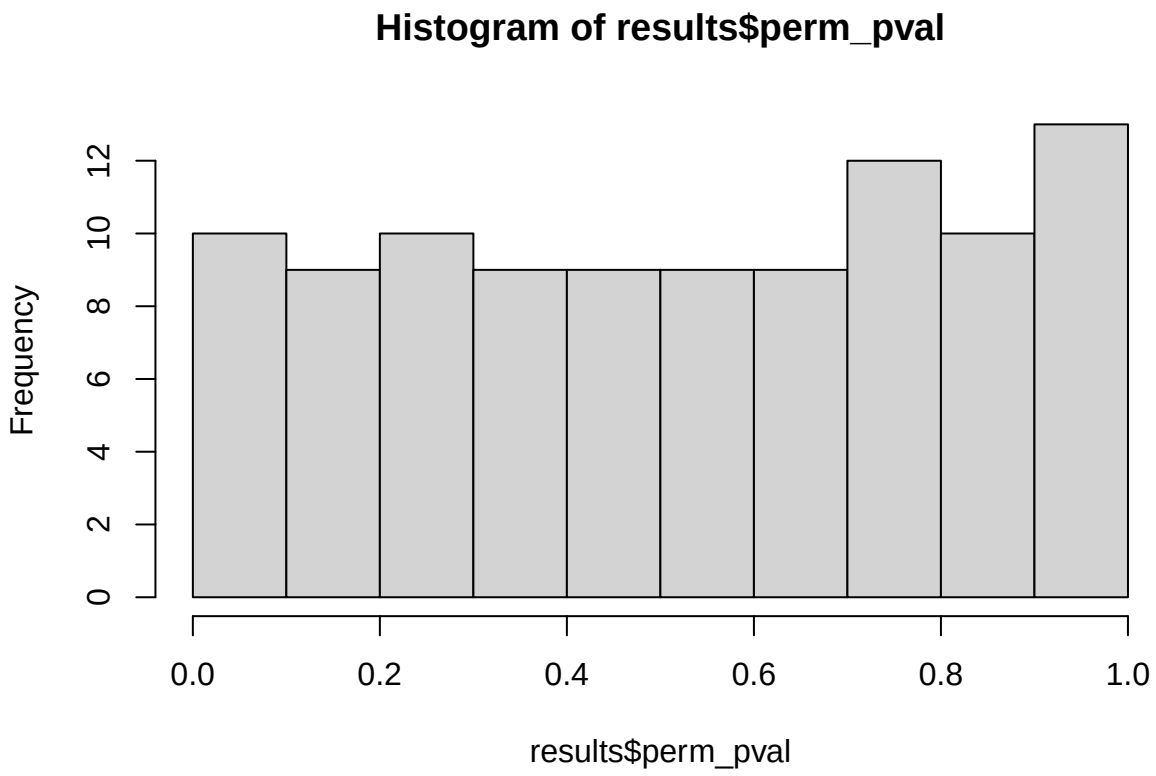
```
## [1] 0.05
```

$\beta_3 = 0.2$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.2,30,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

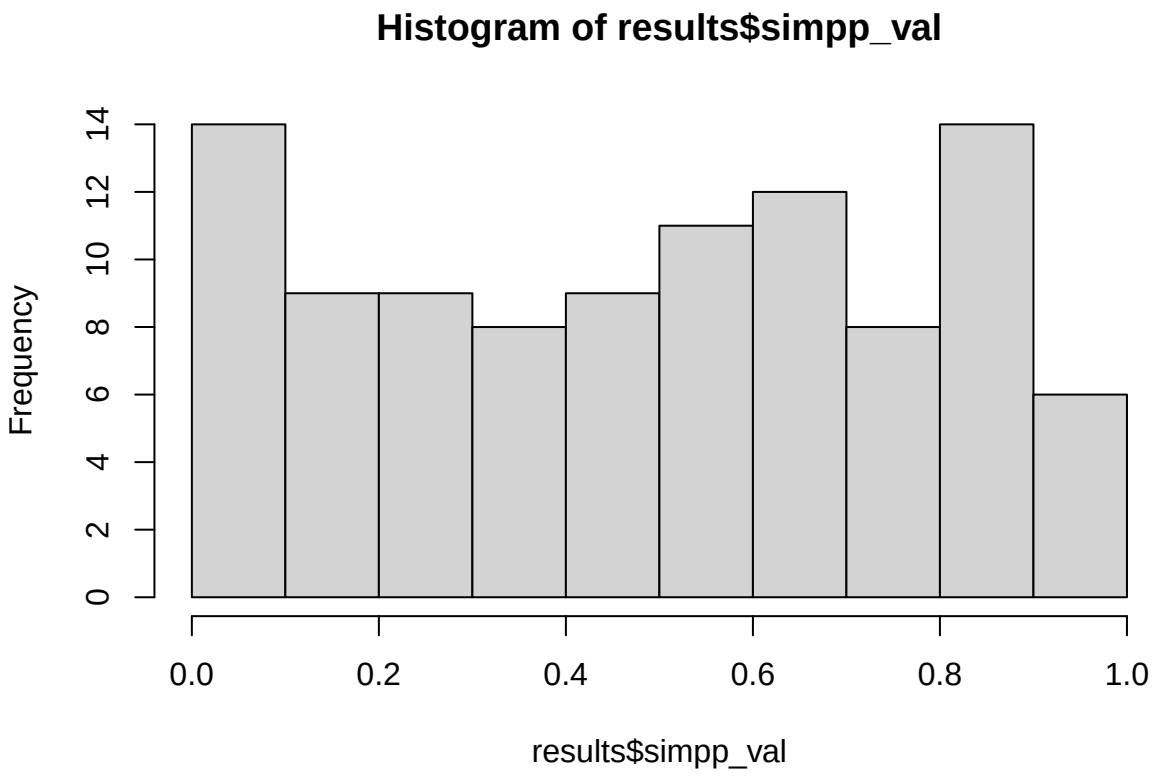
```
## [1] 0.06
```

```
mean(results$perm_pval < 0.05)
```

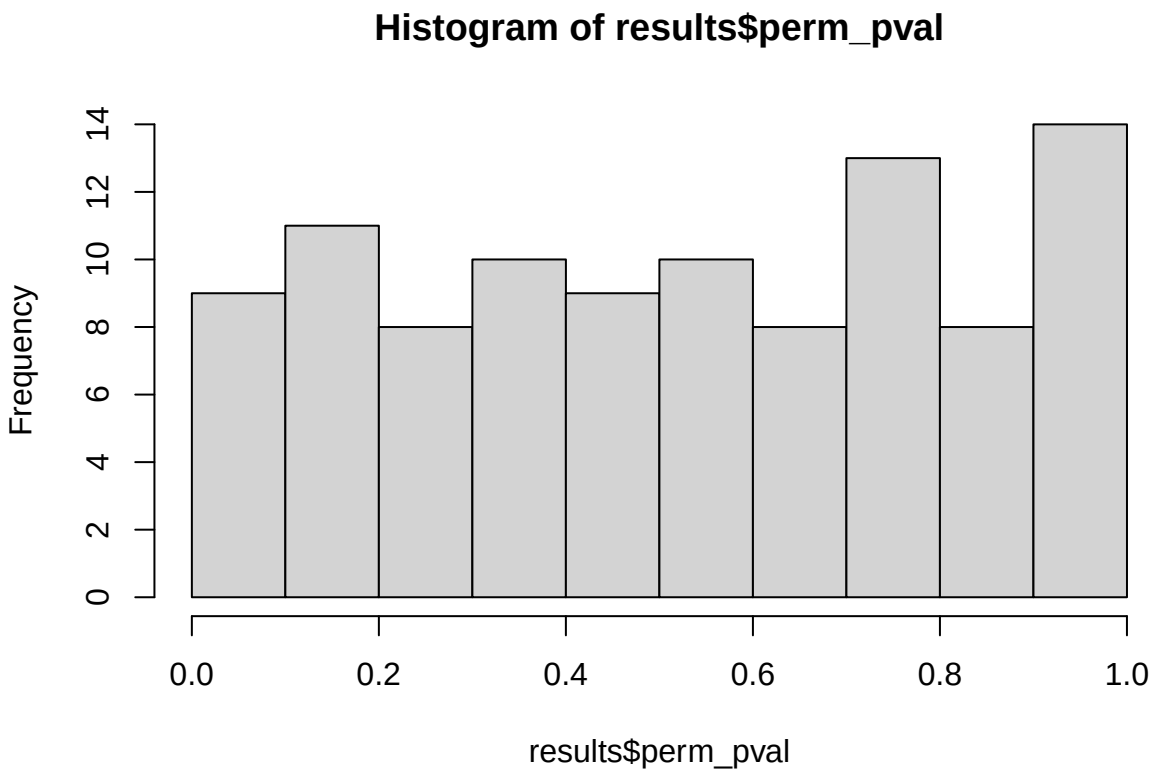
```
## [1] 0.06
```

$\beta_3 = 0.3$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.3,30,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

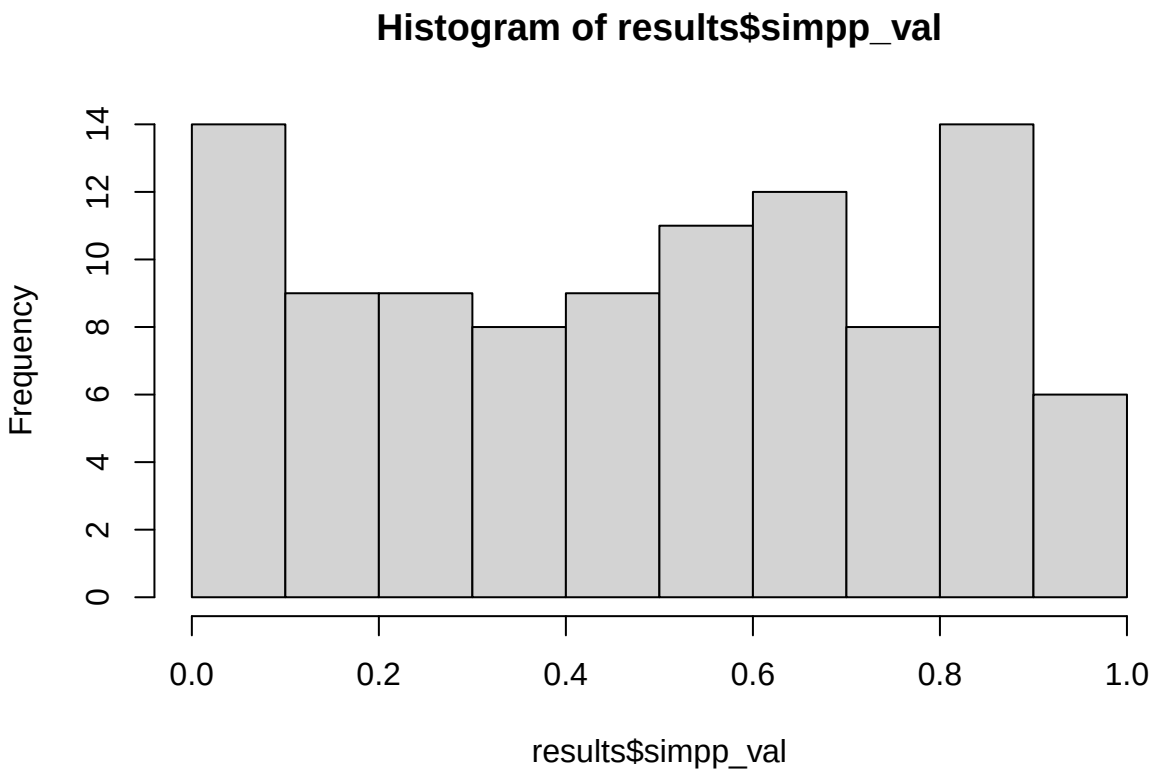
```
## [1] 0.06
```

```
mean(results$perm_pval < 0.05)
```

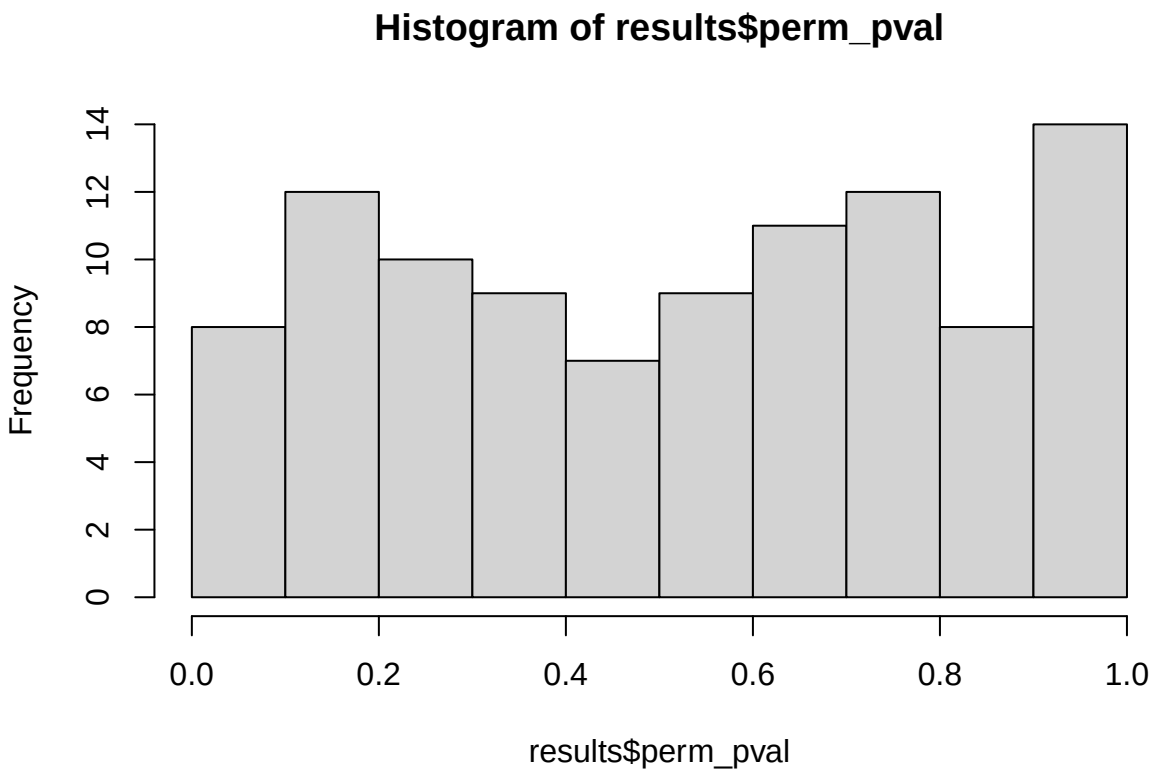
```
## [1] 0.06
```

$\beta_3 = 0.4$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.4,30,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

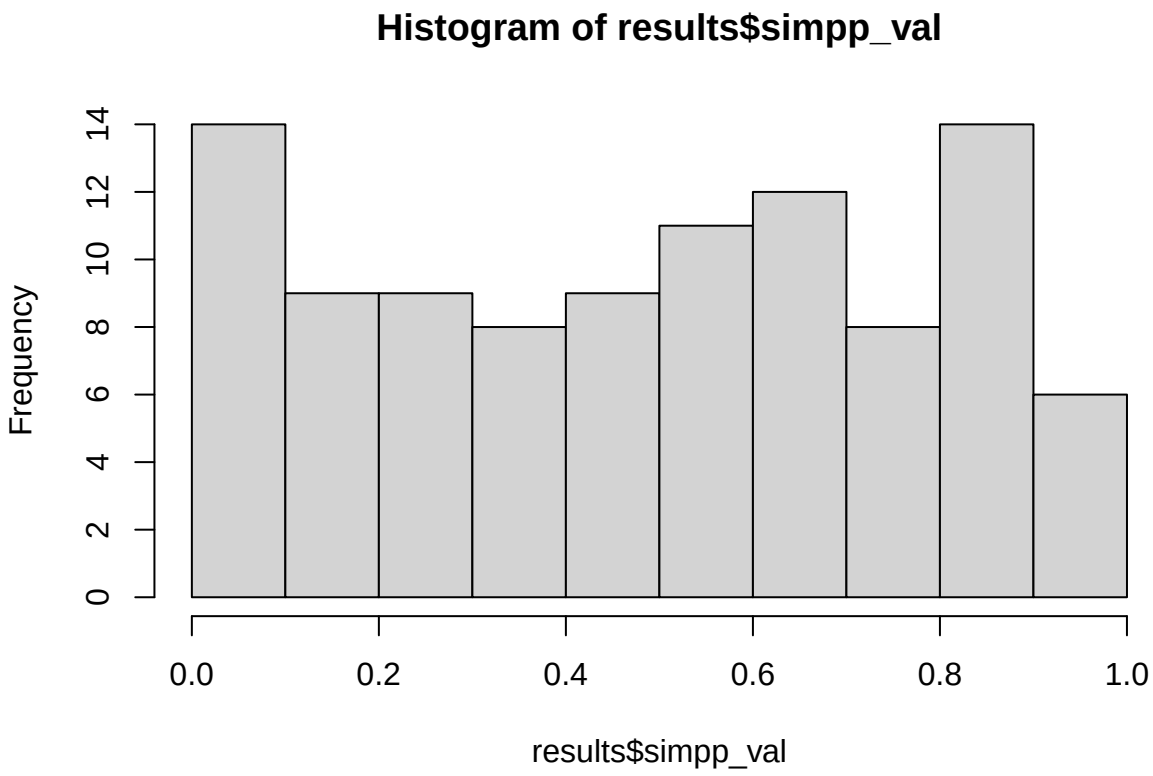
```
## [1] 0.06
```

```
mean(results$perm_pval < 0.05)
```

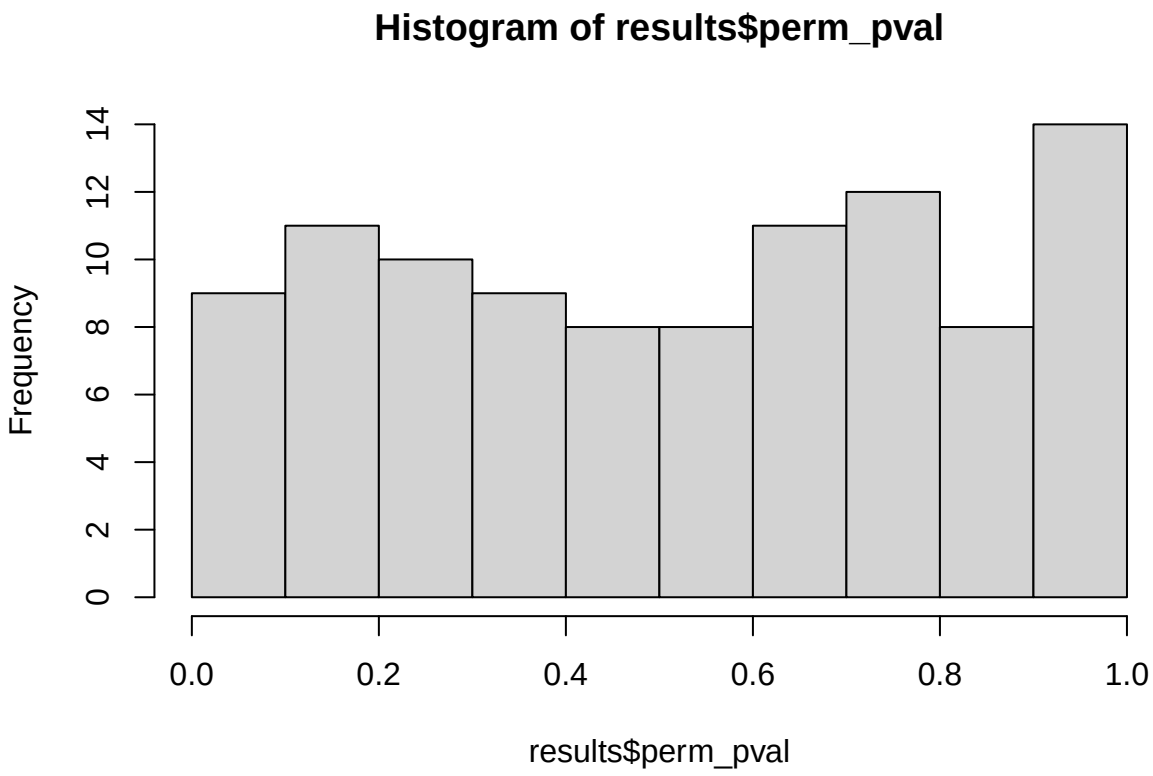
```
## [1] 0.06
```

$\beta_3 = 0.5$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.5,30,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

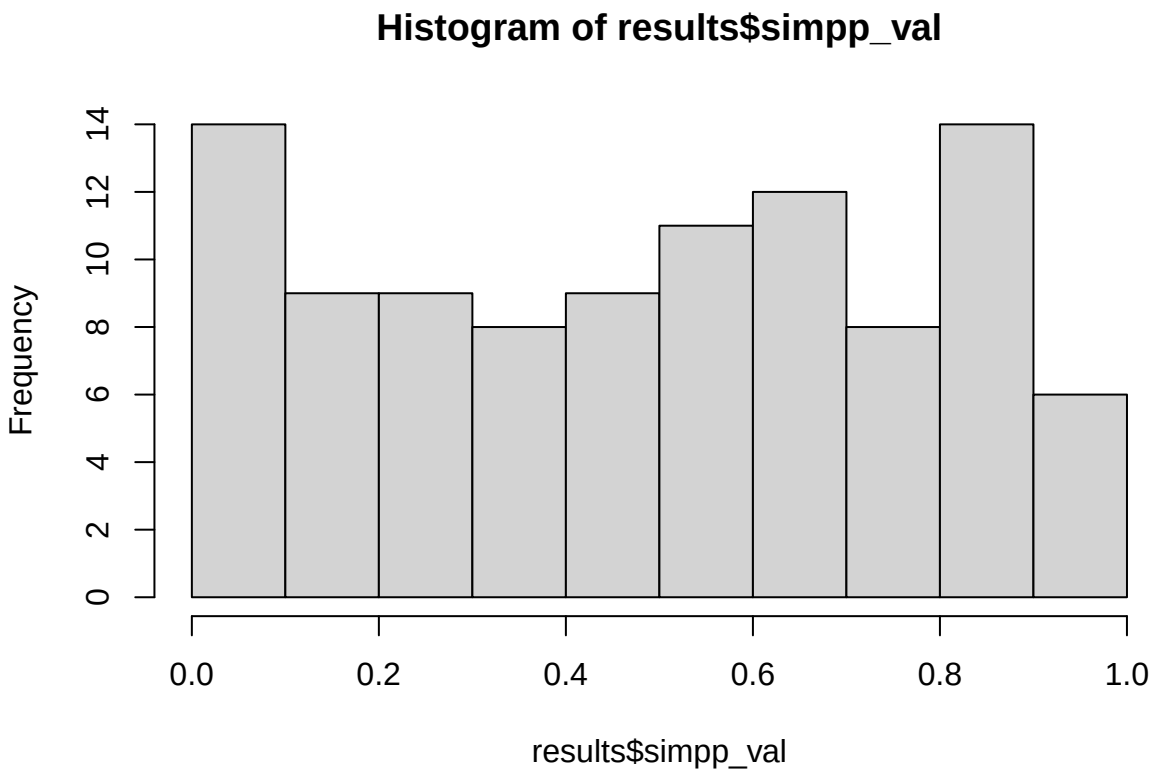
```
## [1] 0.06
```

```
mean(results$perm_pval < 0.05)
```

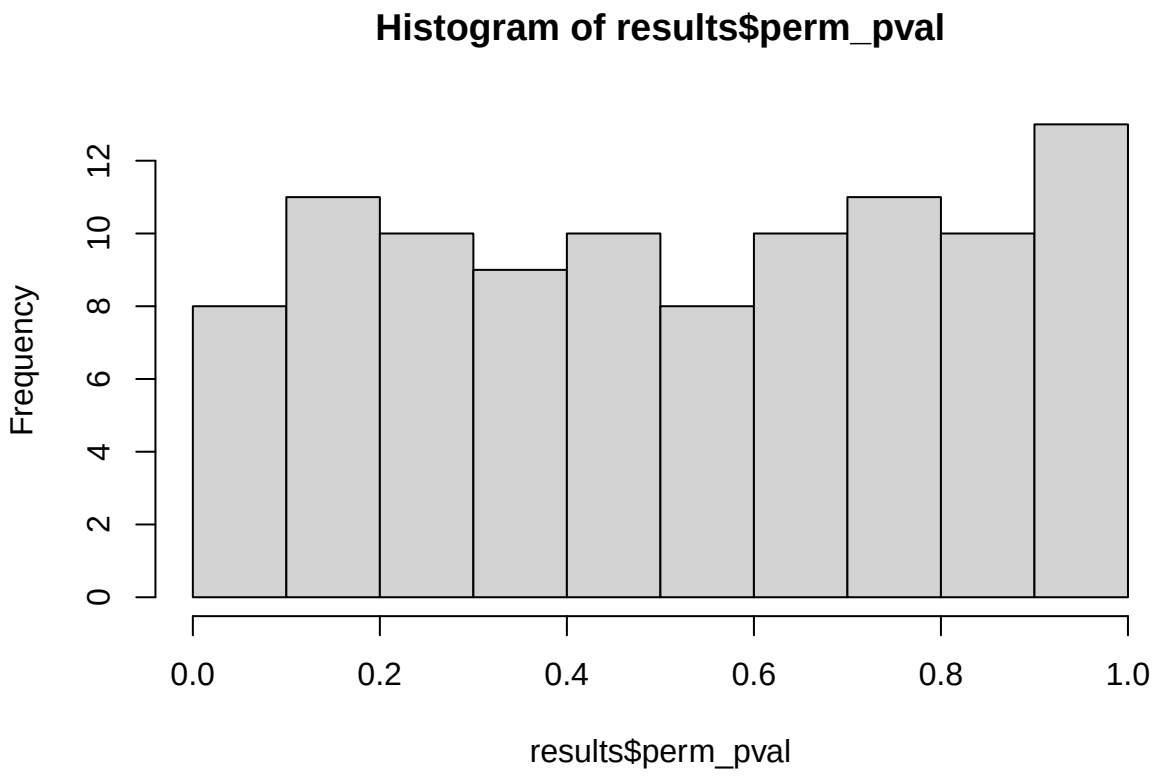
```
## [1] 0.06
```

$\beta_3 = 0.6$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.6,30,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

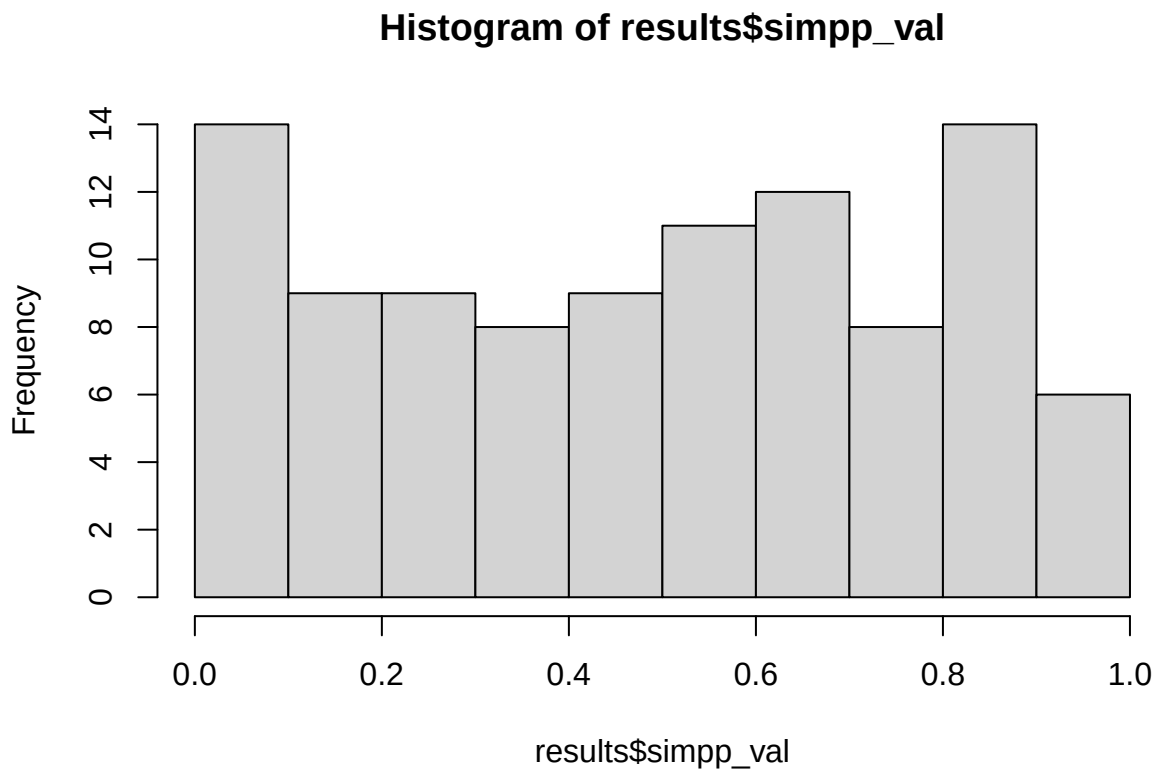
```
## [1] 0.06
```

```
mean(results$perm_pval < 0.05)
```

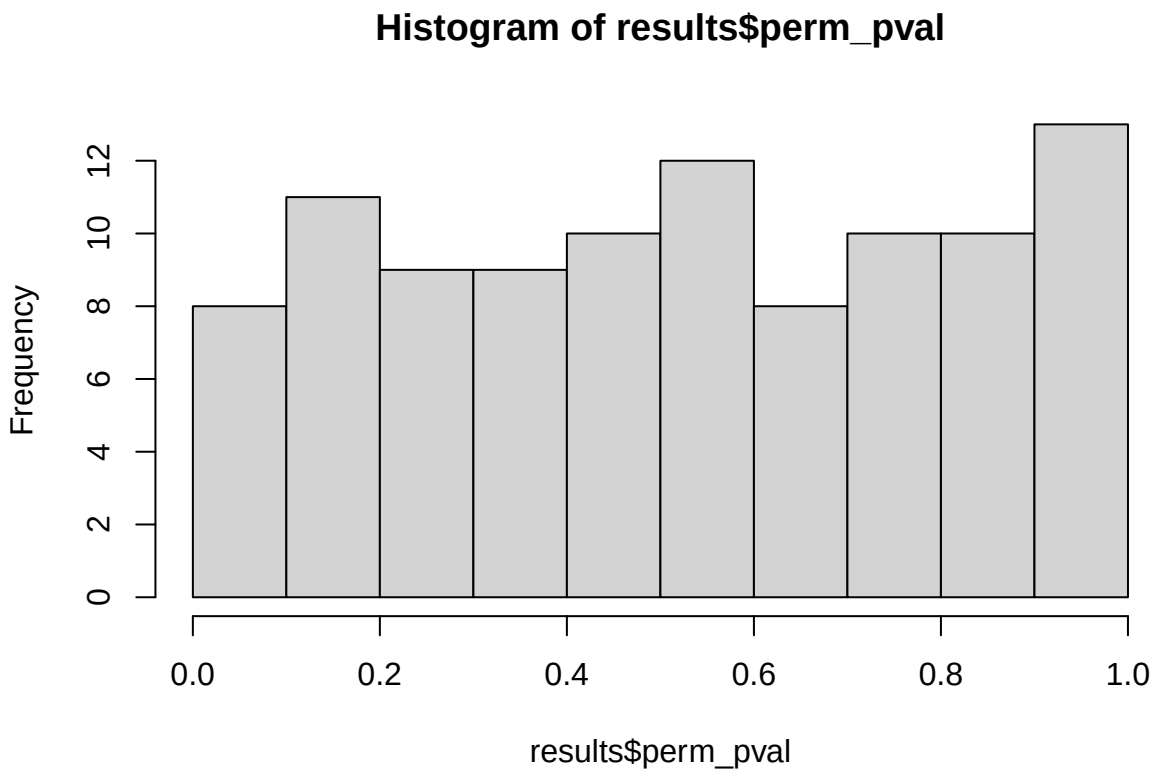
```
## [1] 0.06
```

$\beta_3 = 0.7$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.7,30,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

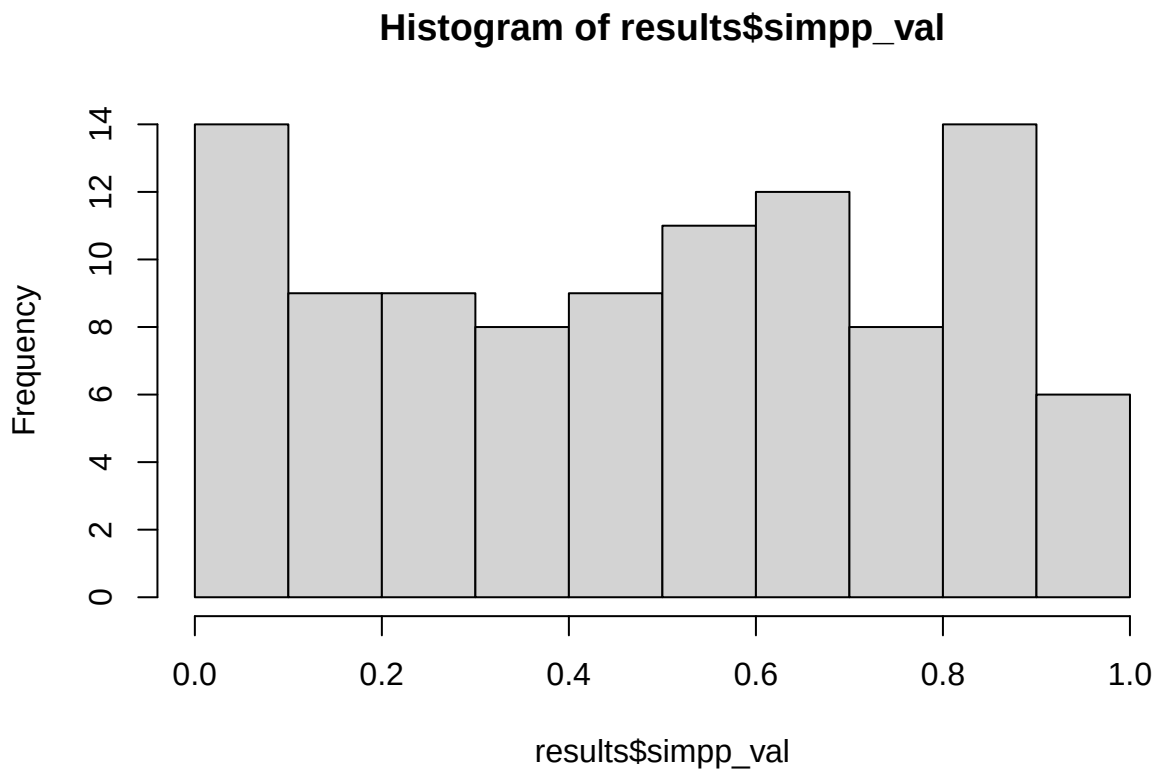
```
## [1] 0.06
```

```
mean(results$perm_pval < 0.05)
```

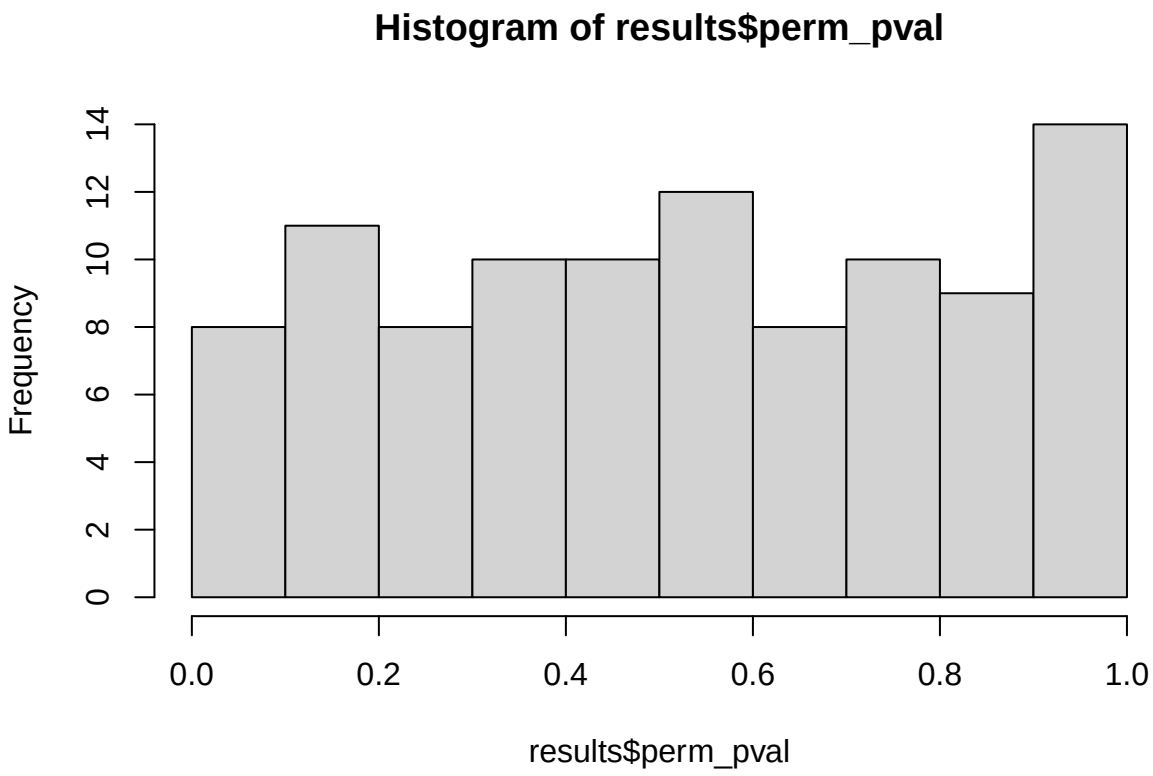
```
## [1] 0.06
```

$\beta_3 = 0.8$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.8,30,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

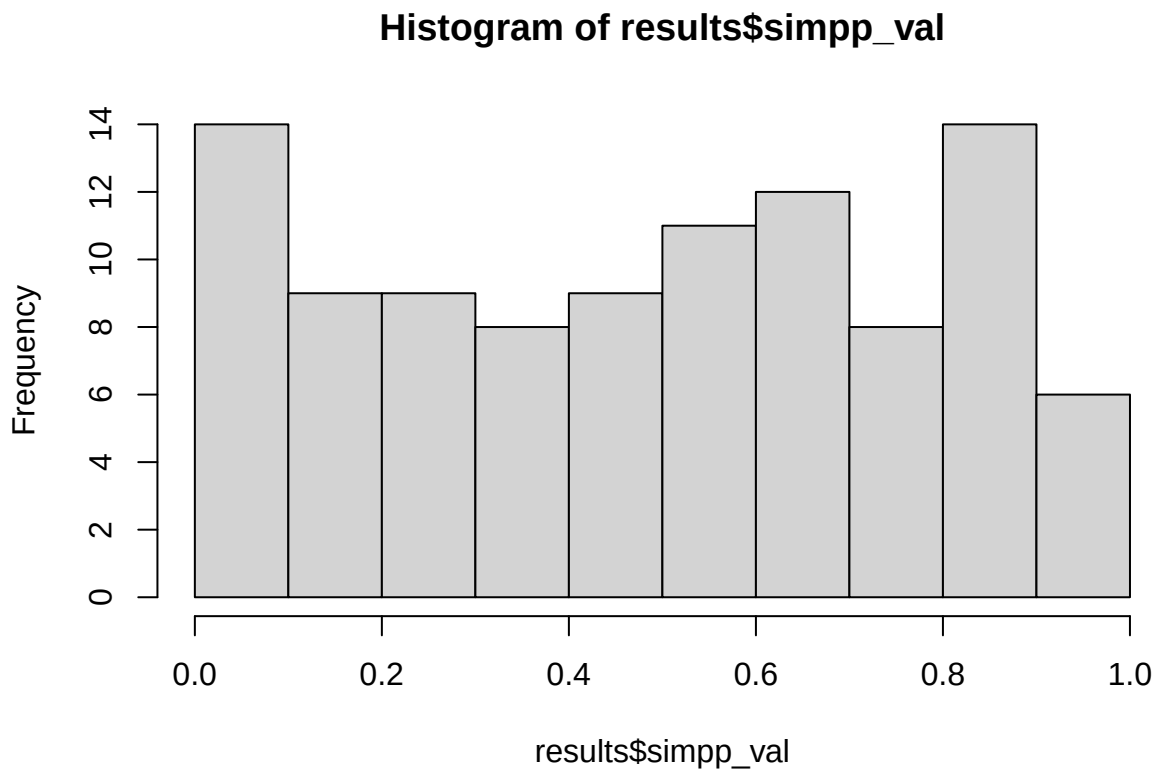
```
## [1] 0.06
```

```
mean(results$perm_pval < 0.05)
```

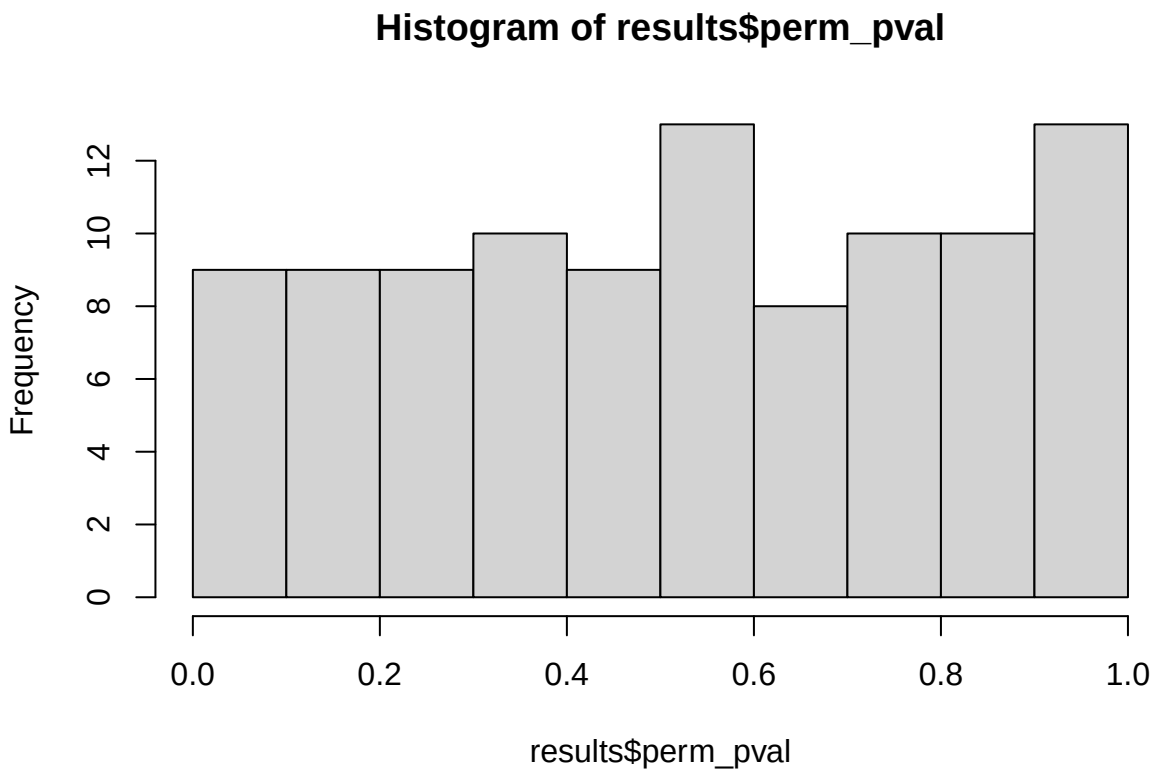
```
## [1] 0.06
```

$\beta_3 = 0.9$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.9,30,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

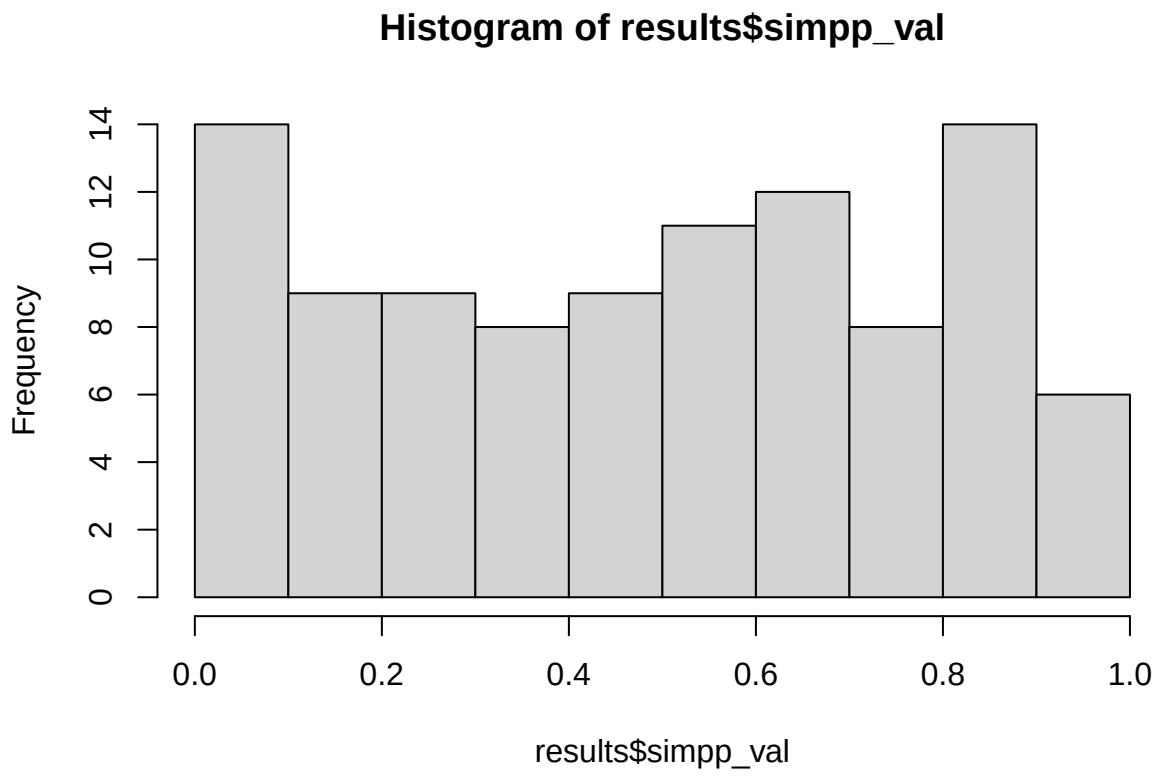
```
## [1] 0.06
```

```
mean(results$perm_pval < 0.05)
```

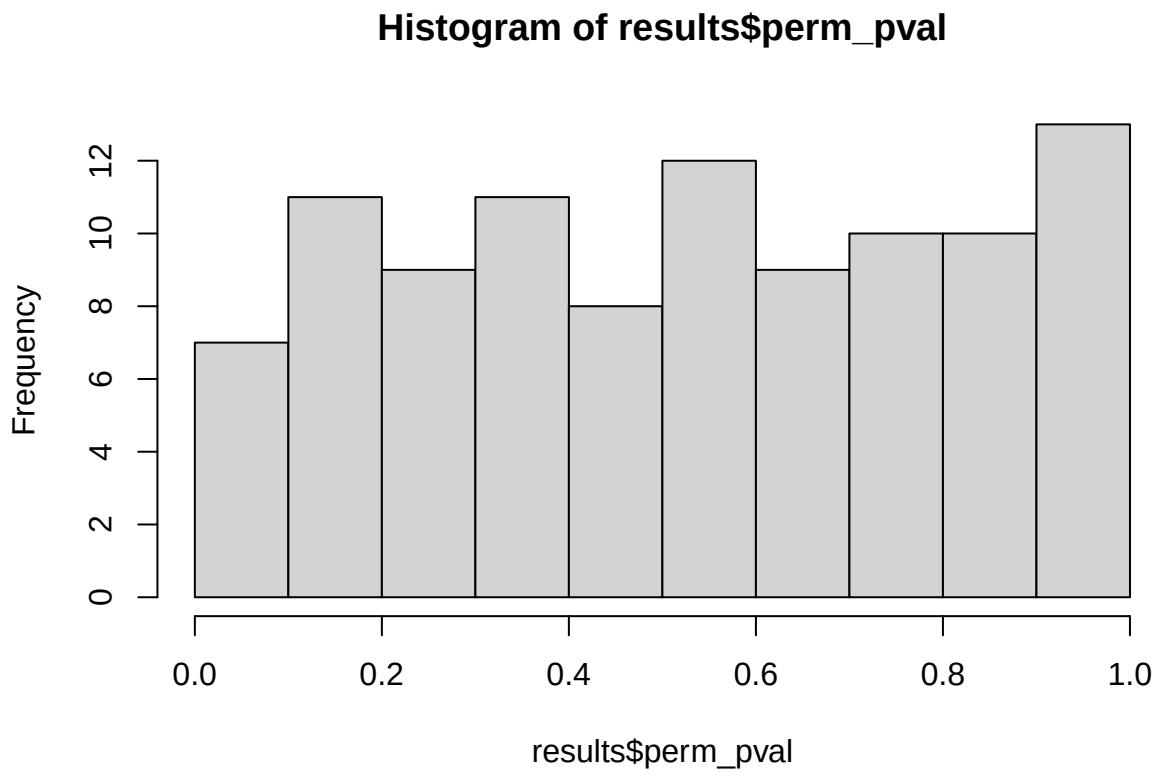
```
## [1] 0.06
```

$\beta_3 = 1.0$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,1.0,30,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

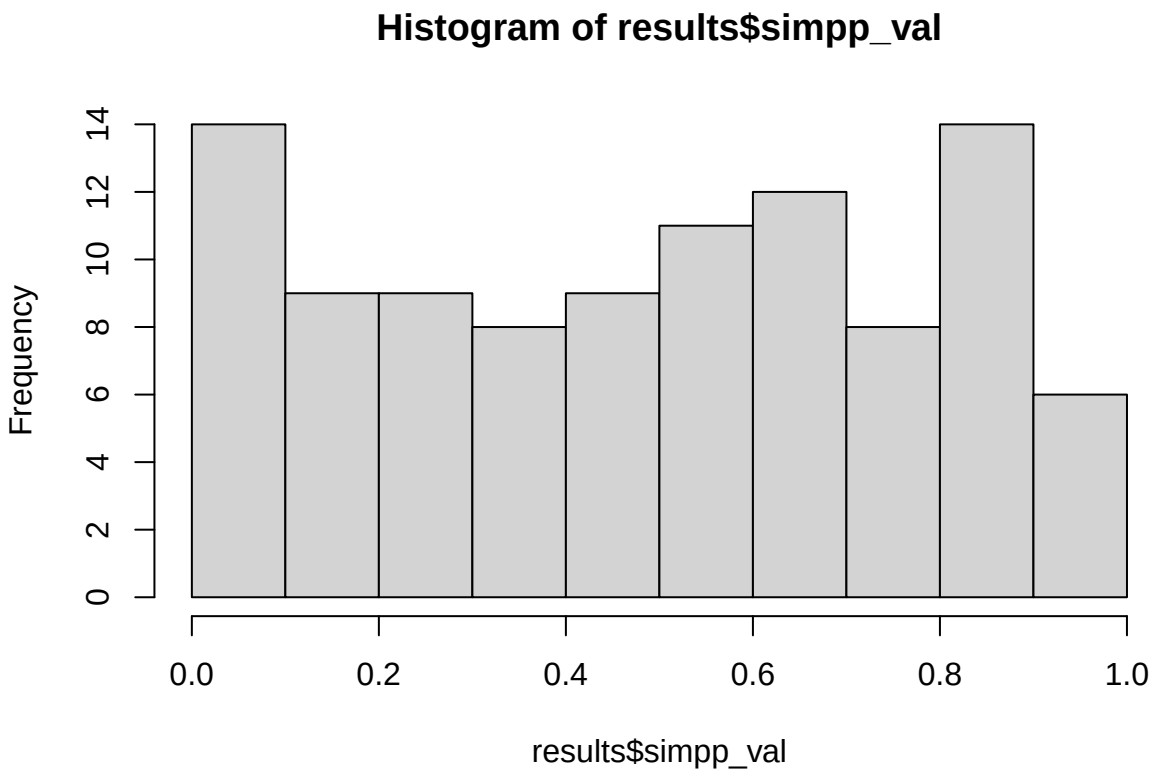
```
## [1] 0.06
```

```
mean(results$perm_pval < 0.05)
```

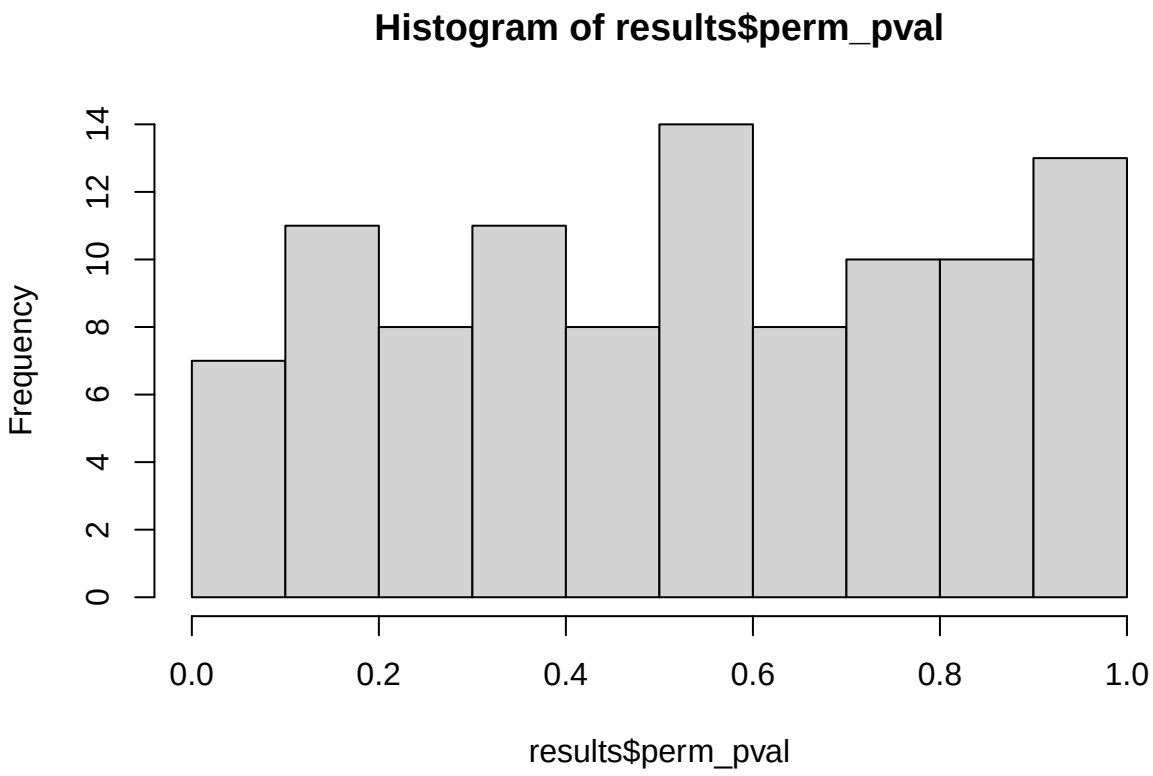
```
## [1] 0.06
```

$\beta_3 = 1.1$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,1.1,30,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

```
## [1] 0.06
```

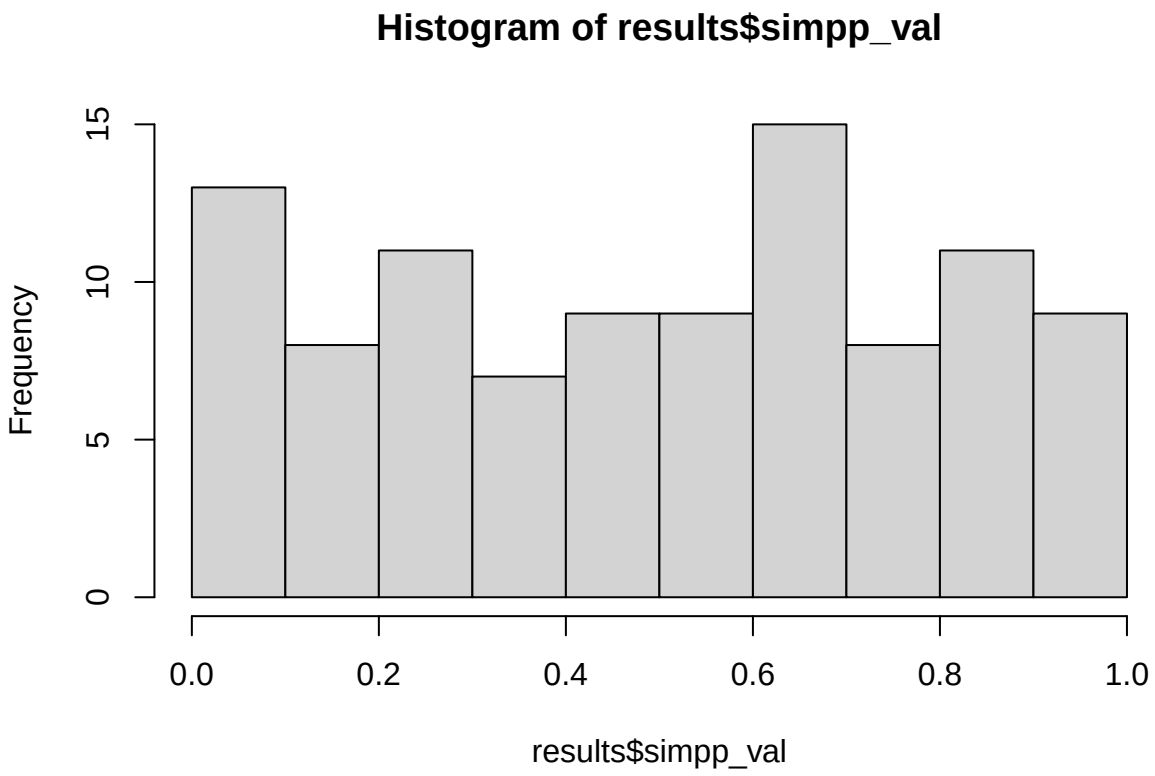
```
mean(results$perm_pval < 0.05)
```

```
## [1] 0.06
```

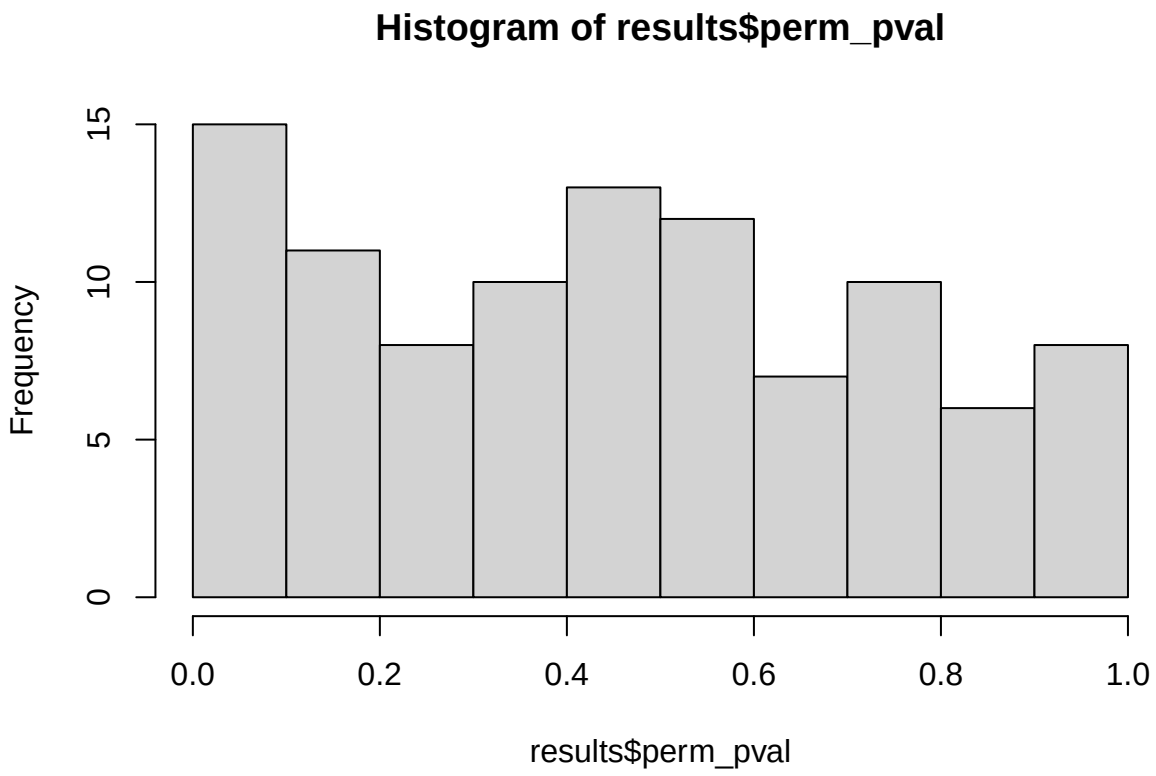
For n=60

$\beta_3 = 0.1$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.1,60,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

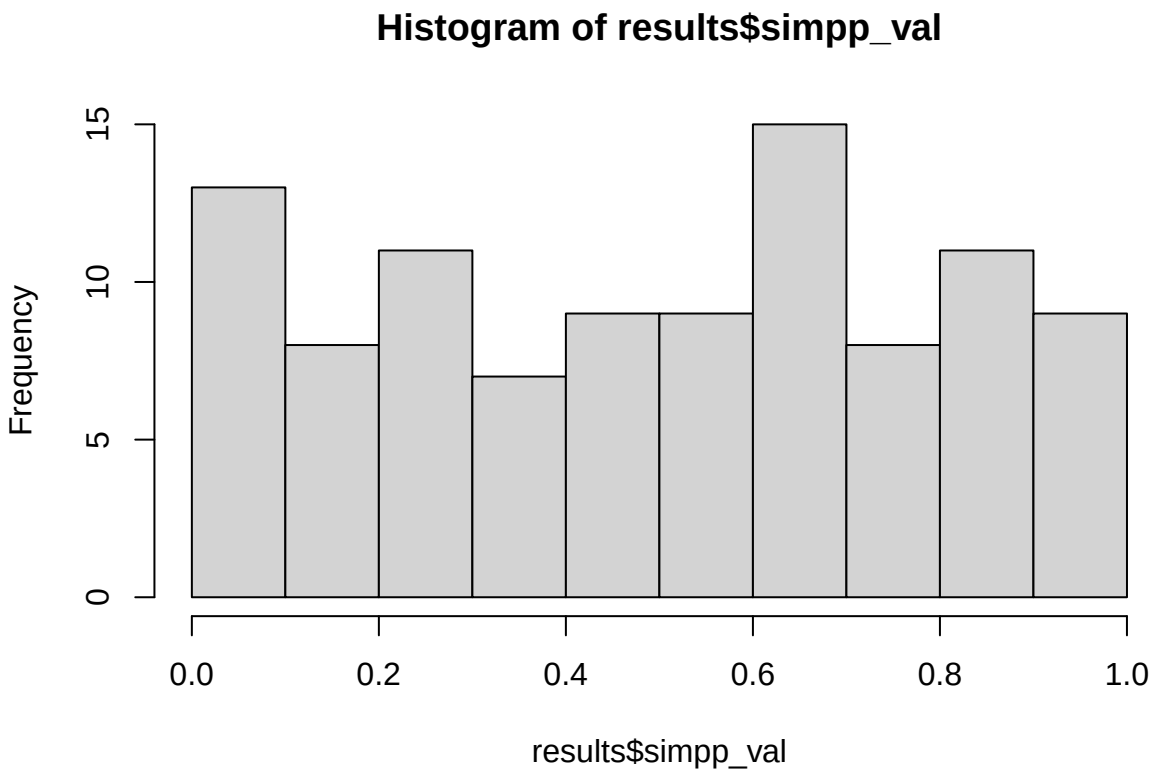
```
## [1] 0.07
```

```
mean(results$perm_pval < 0.05)
```

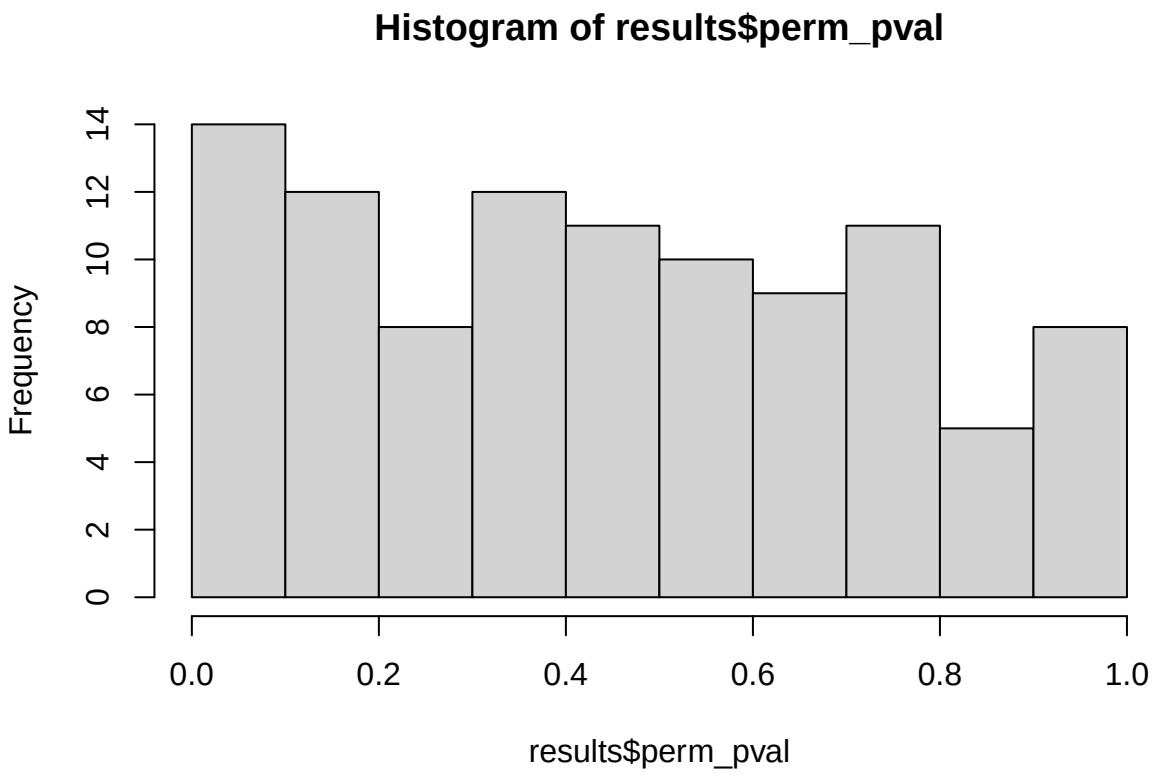
```
## [1] 0.07
```

$\beta_3 = 0.2$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.2,60,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

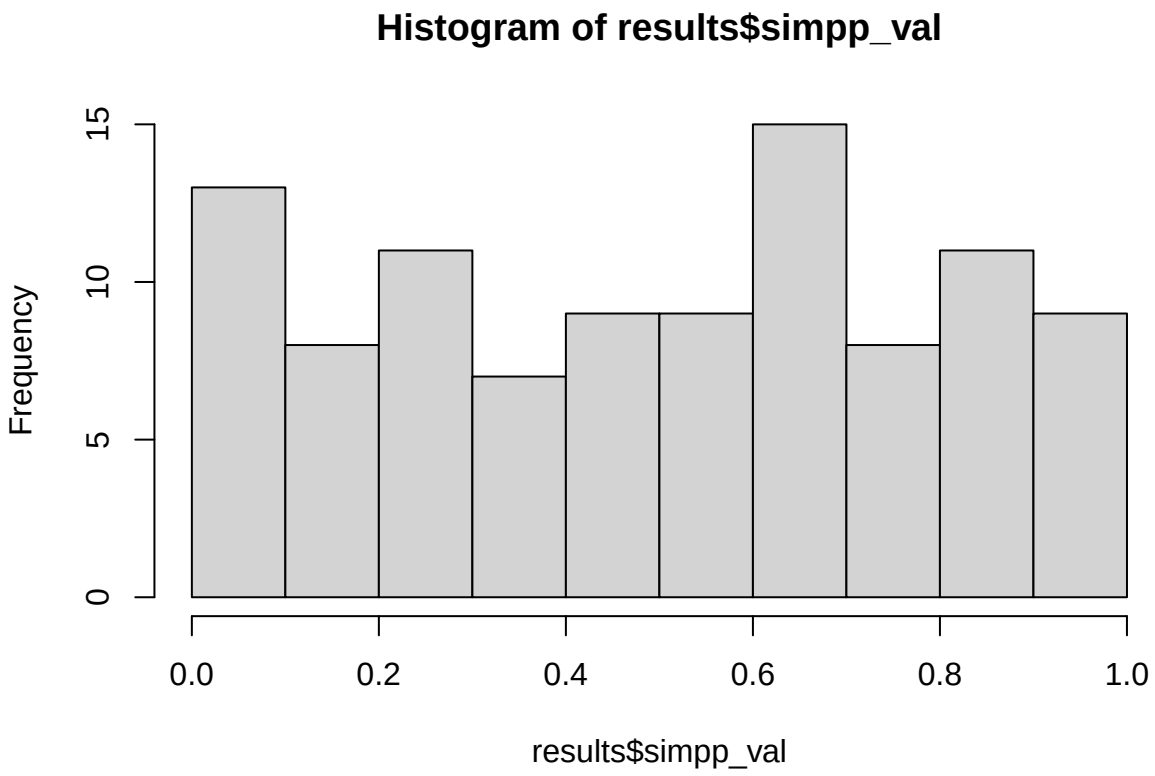
```
## [1] 0.07
```

```
mean(results$perm_pval < 0.05)
```

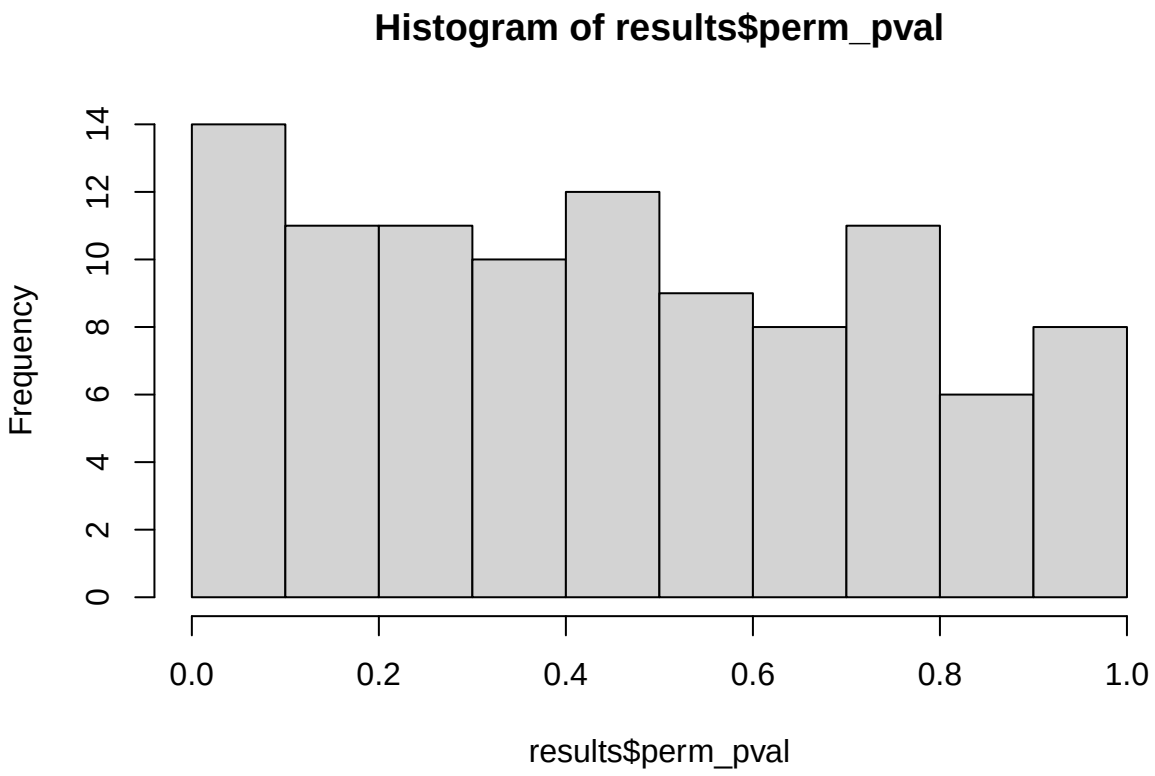
```
## [1] 0.07
```

$\beta_3 = 0.3$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.3,60,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

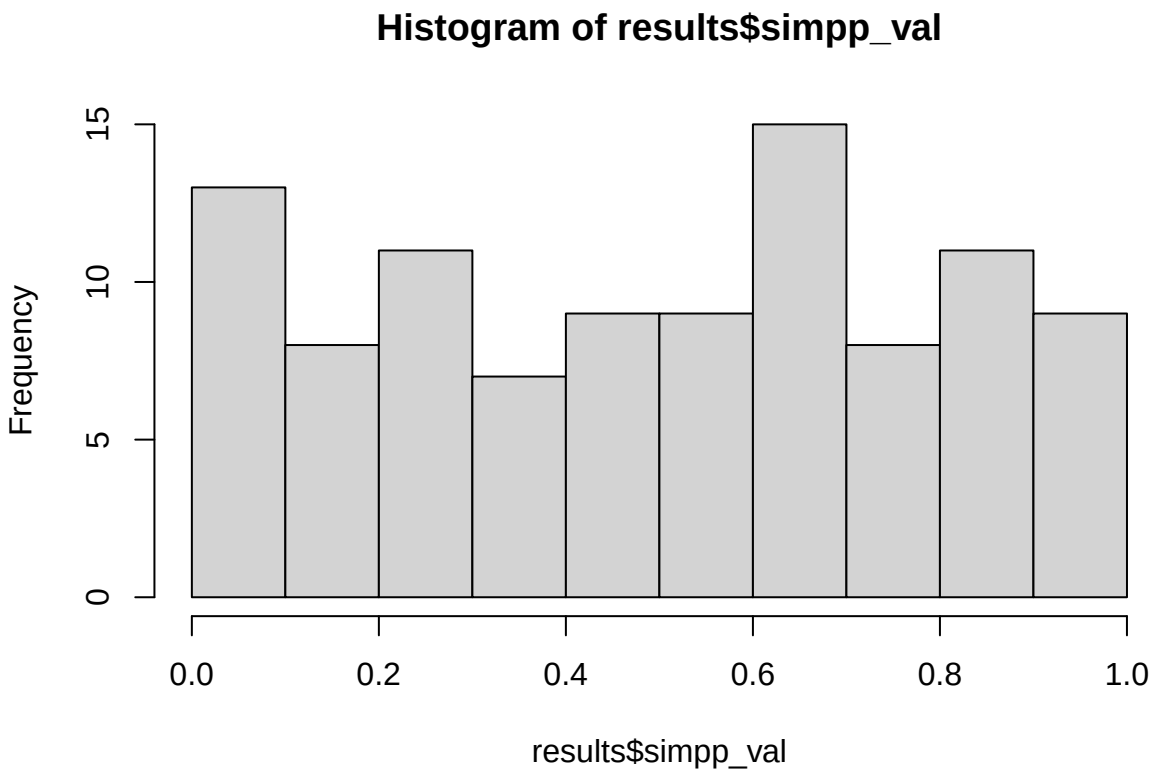
```
## [1] 0.07
```

```
mean(results$perm_pval < 0.05)
```

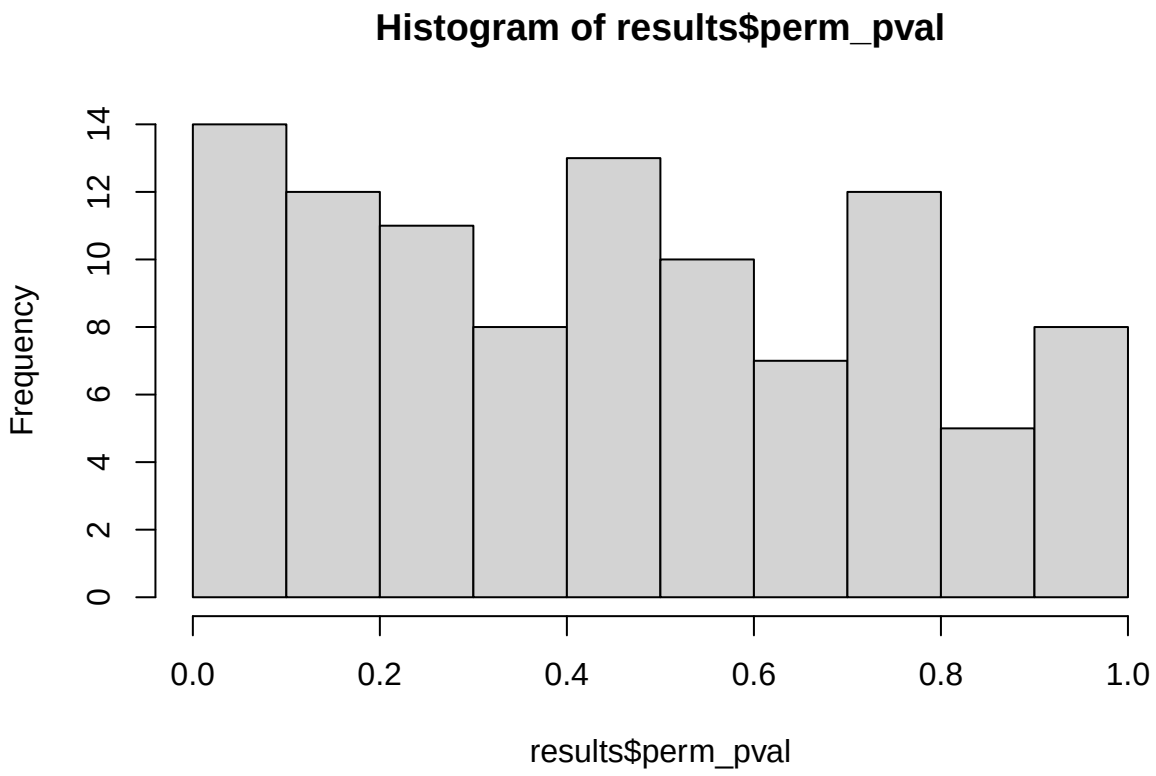
```
## [1] 0.06
```

$\beta_3 = 0.4$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.4,60,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

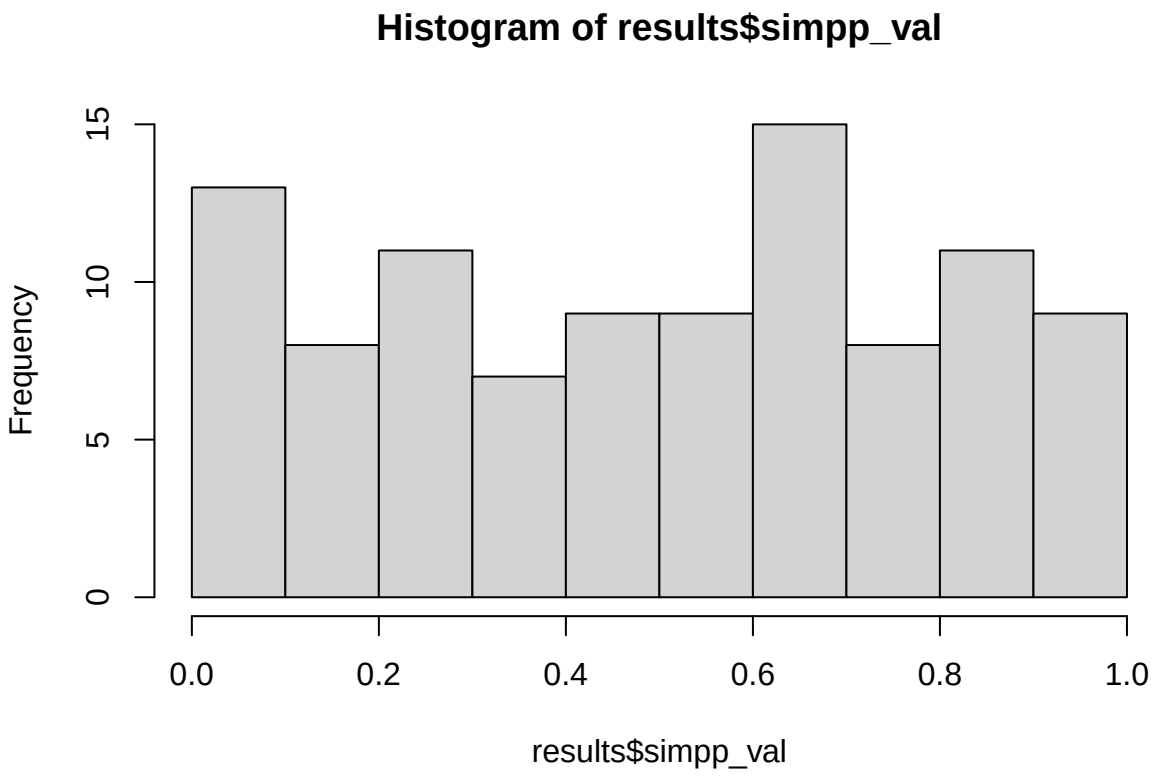
```
## [1] 0.07
```

```
mean(results$perm_pval < 0.05)
```

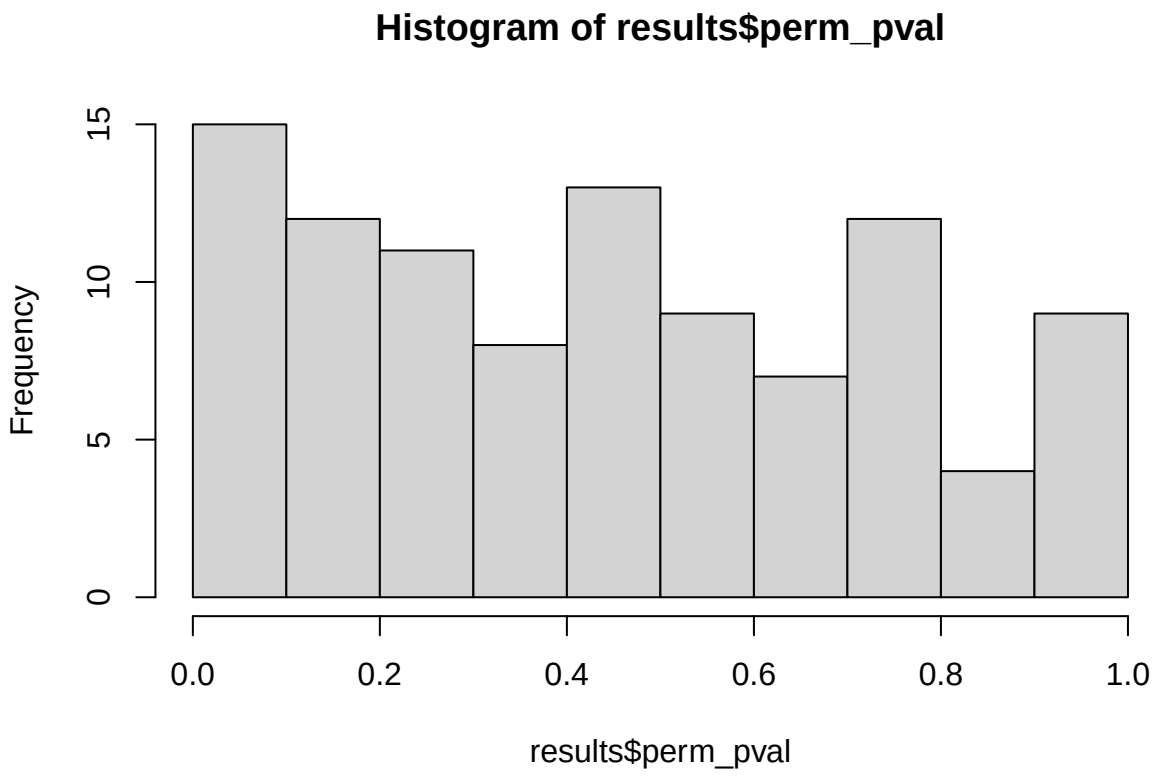
```
## [1] 0.07
```

$\beta_3 = 0.5$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.5,60,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

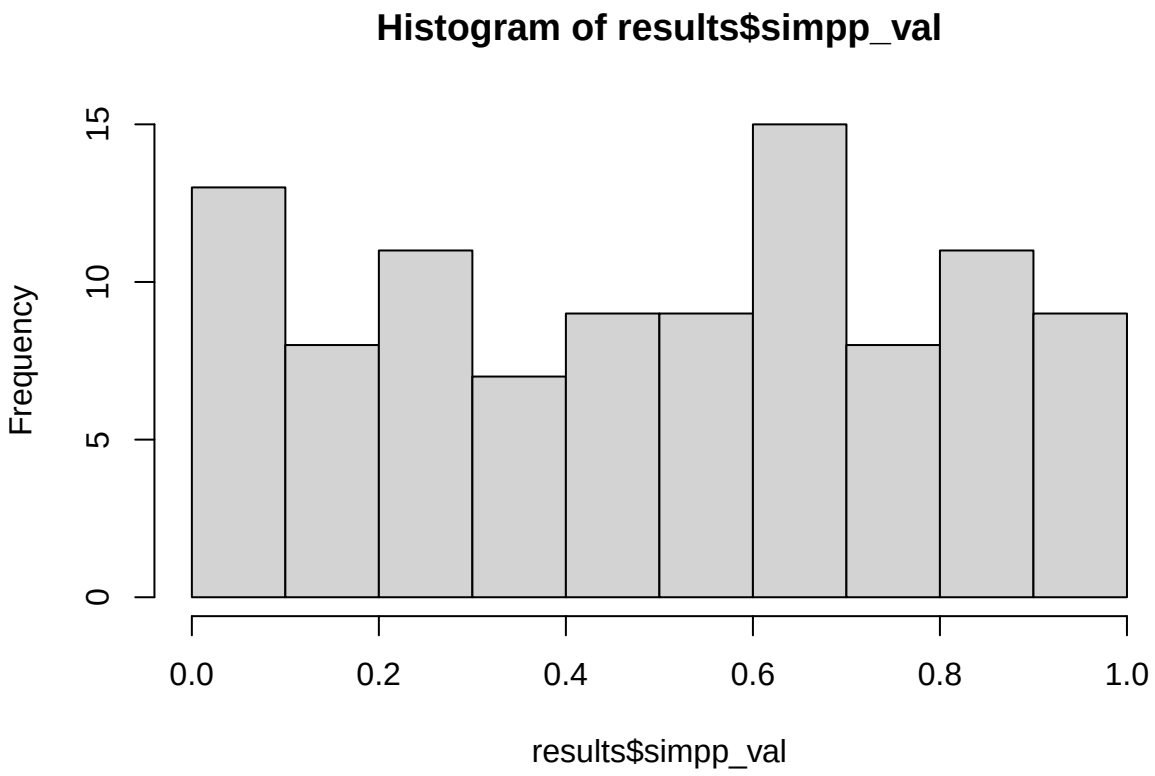
```
## [1] 0.07
```

```
mean(results$perm_pval < 0.05)
```

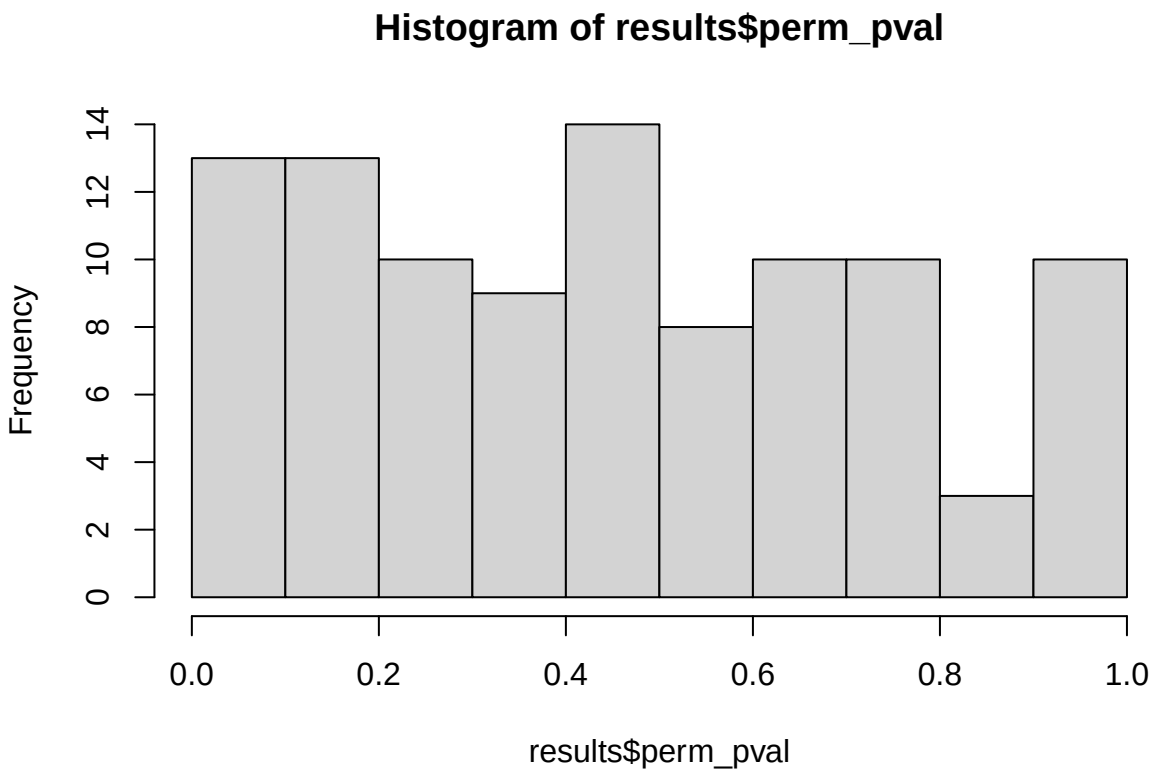
```
## [1] 0.06
```

$\beta_3 = 0.6$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.6,60,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

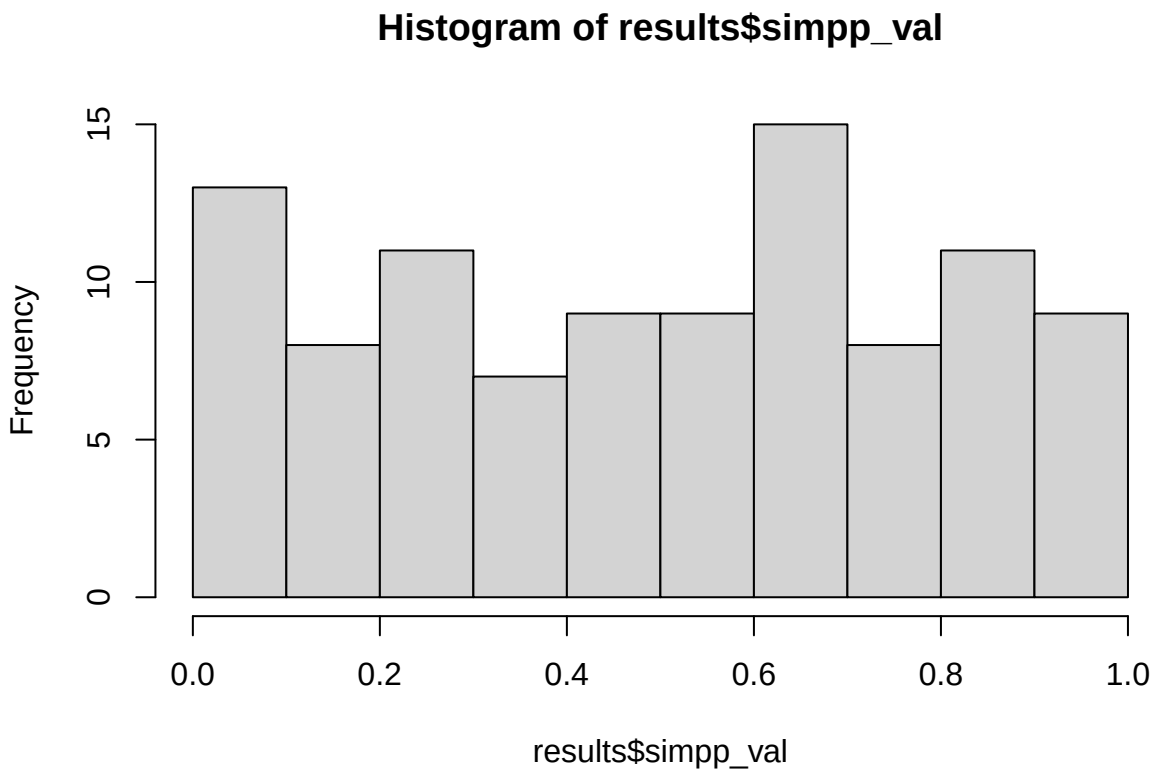
```
## [1] 0.07
```

```
mean(results$perm_pval < 0.05)
```

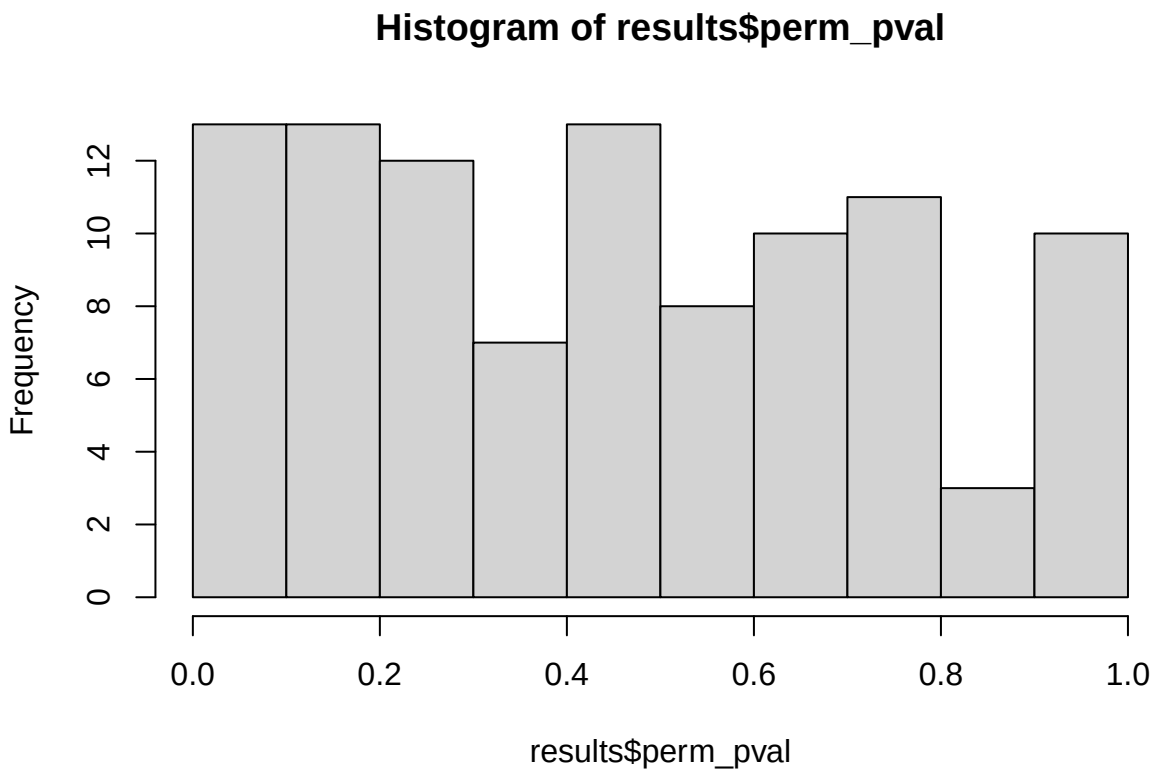
```
## [1] 0.06
```

$\beta_3 = 0.7$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.7,60,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

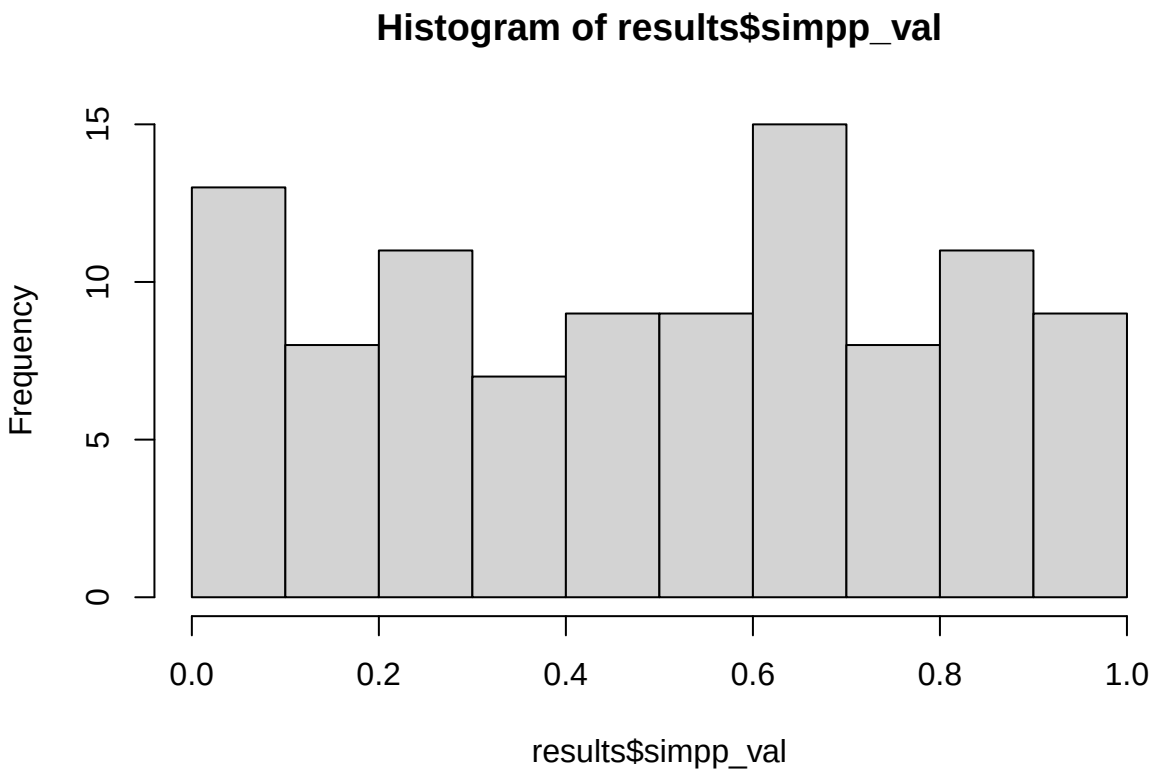
```
## [1] 0.07
```

```
mean(results$perm_pval < 0.05)
```

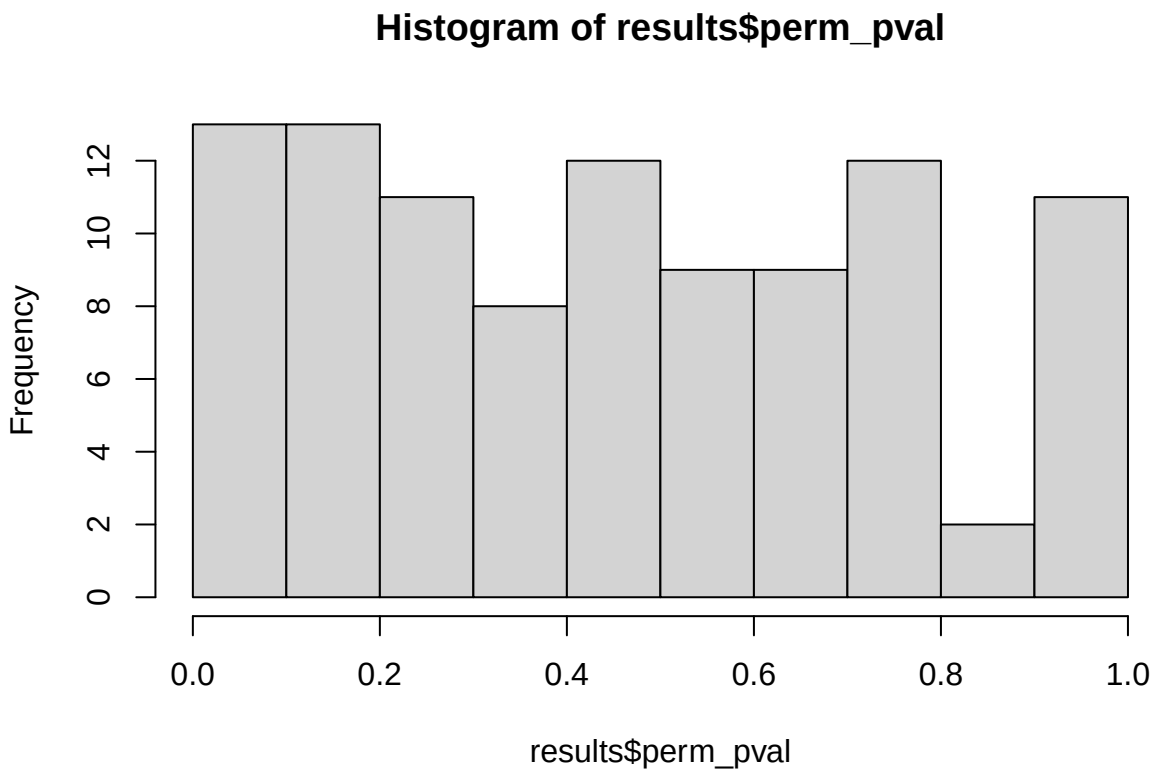
```
## [1] 0.06
```

$\beta_3 = 0.8$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.8,60,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

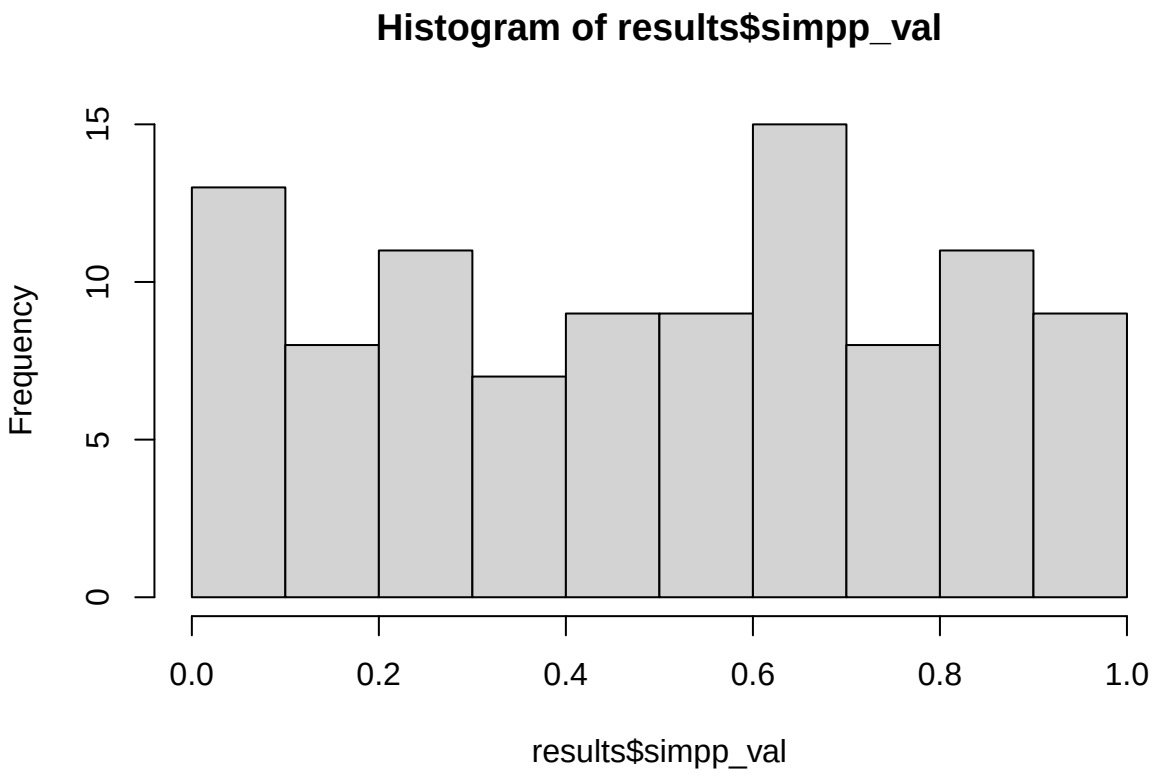
```
## [1] 0.07
```

```
mean(results$perm_pval < 0.05)
```

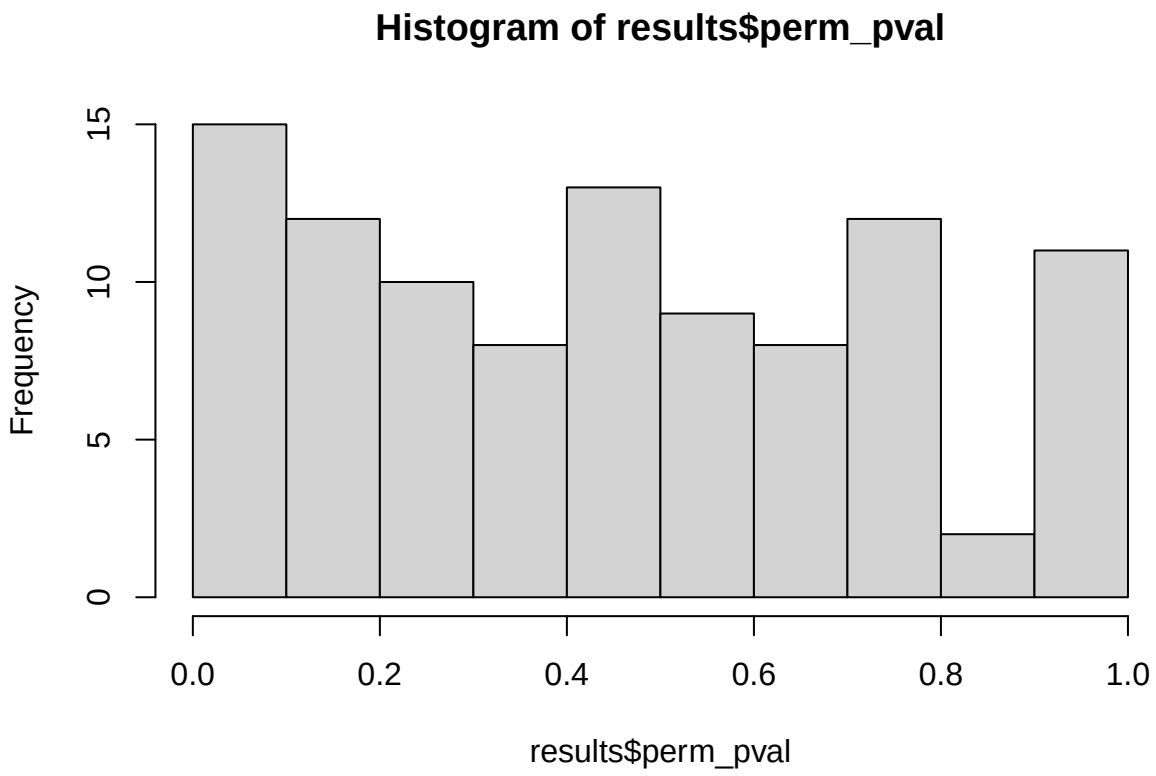
```
## [1] 0.05
```

$\beta_3 = 0.9$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,0.9,60,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

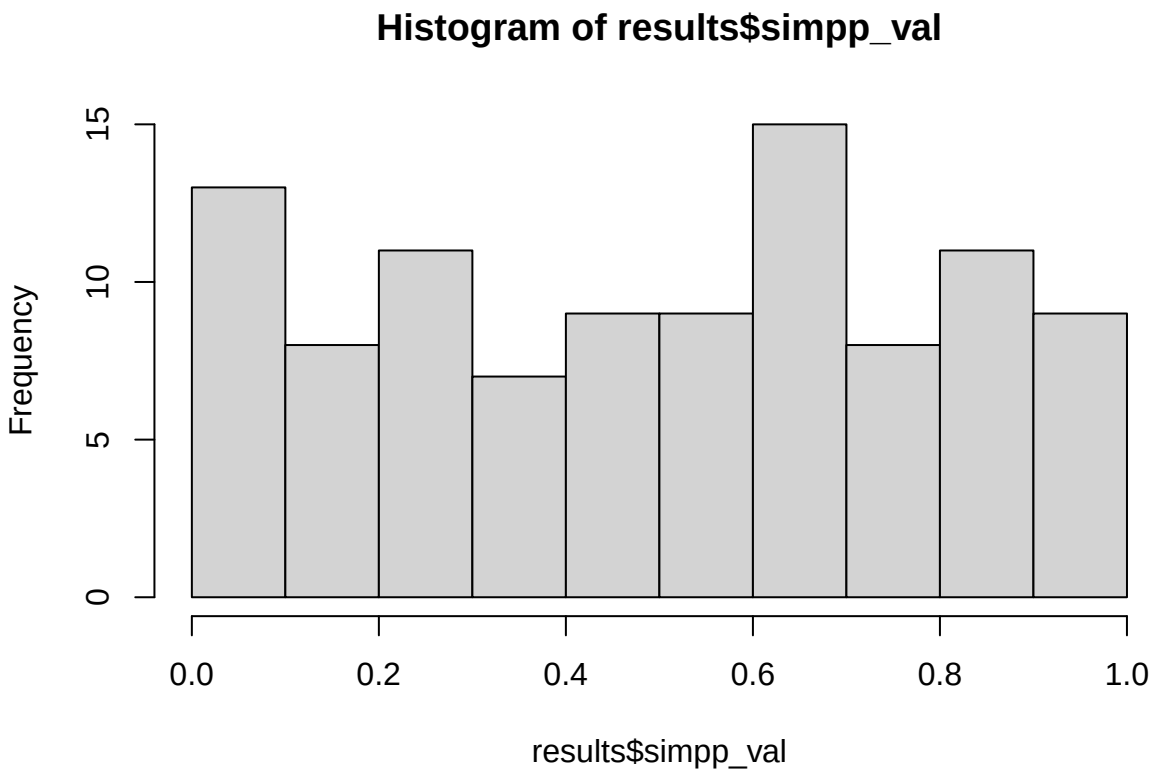
```
## [1] 0.07
```

```
mean(results$perm_pval < 0.05)
```

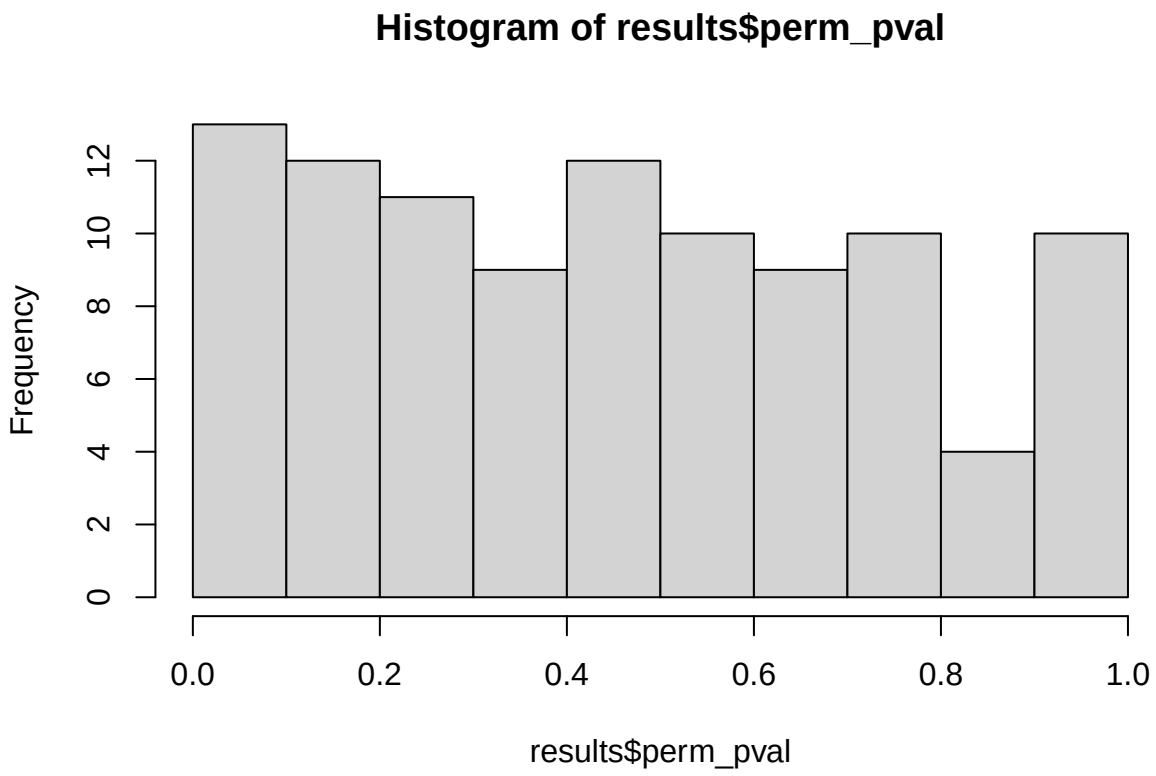
```
## [1] 0.06
```

$\beta_3 = 1.0$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,1.0,60,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

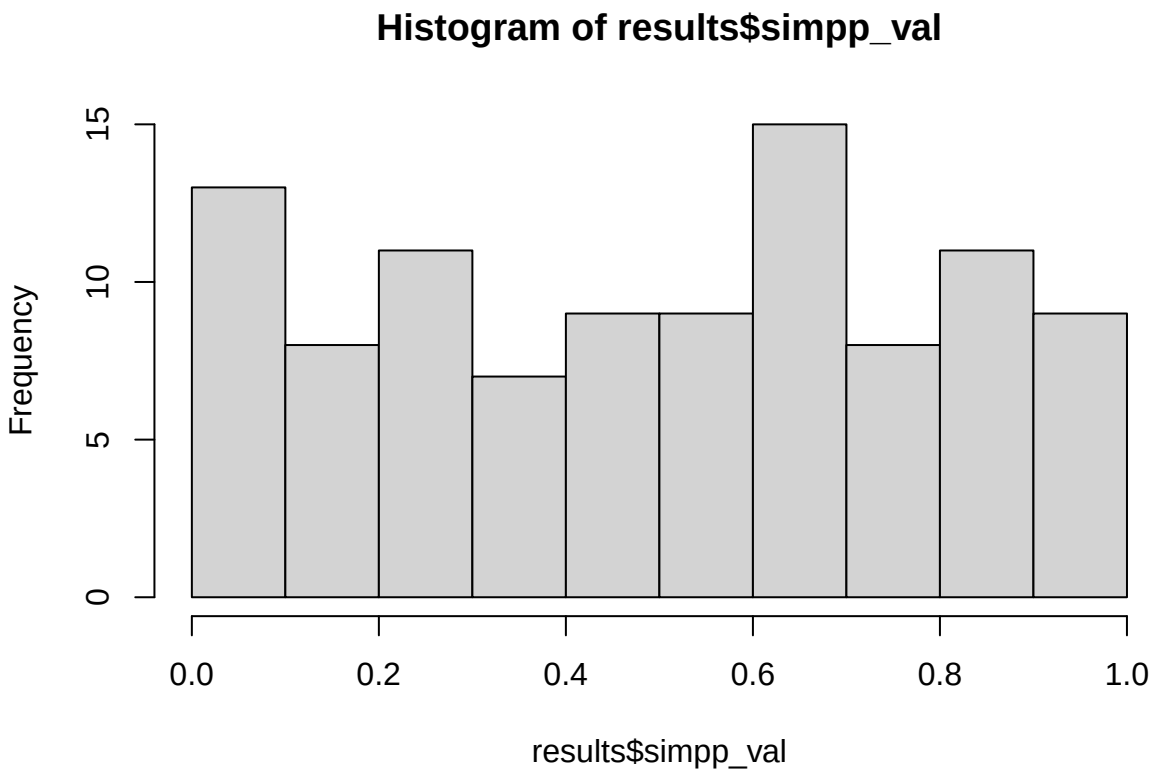
```
## [1] 0.07
```

```
mean(results$perm_pval < 0.05)
```

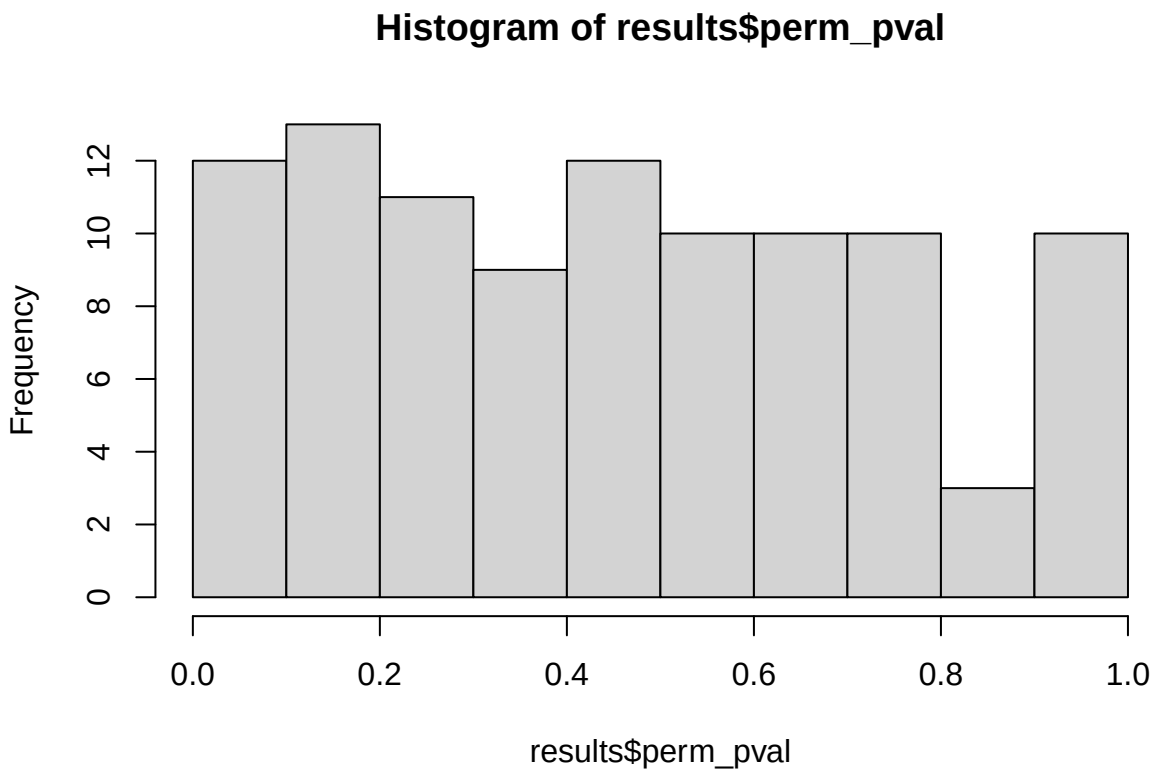
```
## [1] 0.06
```

$\beta_3 = 1.1$

```
#beta0,beta1,beta2,beta3, n_obs,n_data,n_shuffle
results<-LFunky211(0,0,0,1.1,60,100,200)
hist(results$simpp_val)
```




```
hist(results$perm_pval)
```




```
mean(results$simpp_val < 0.05)
```

```
## [1] 0.07
```

```
mean(results$perm_pval < 0.05)
```

```
## [1] 0.06
```