

## Introductions

- Name
- Year
- Make sure to sign in on the google form (linked here)

## Hypothesis testing on real data

This section will deal with a data set of country-level statistics from this source with an explanation of the data encoding found here.

```
# Read in the data
countries <- read.csv("data/countries.csv", check.names = F)
```

1. Perform a formal  $t$ -test for the difference in GDP per capita (encoded as `mad_gdppc`) by democracy status (encoded as `br_dem`) and provide a 95% confidence interval for the difference. Comment on the assumptions of the test.

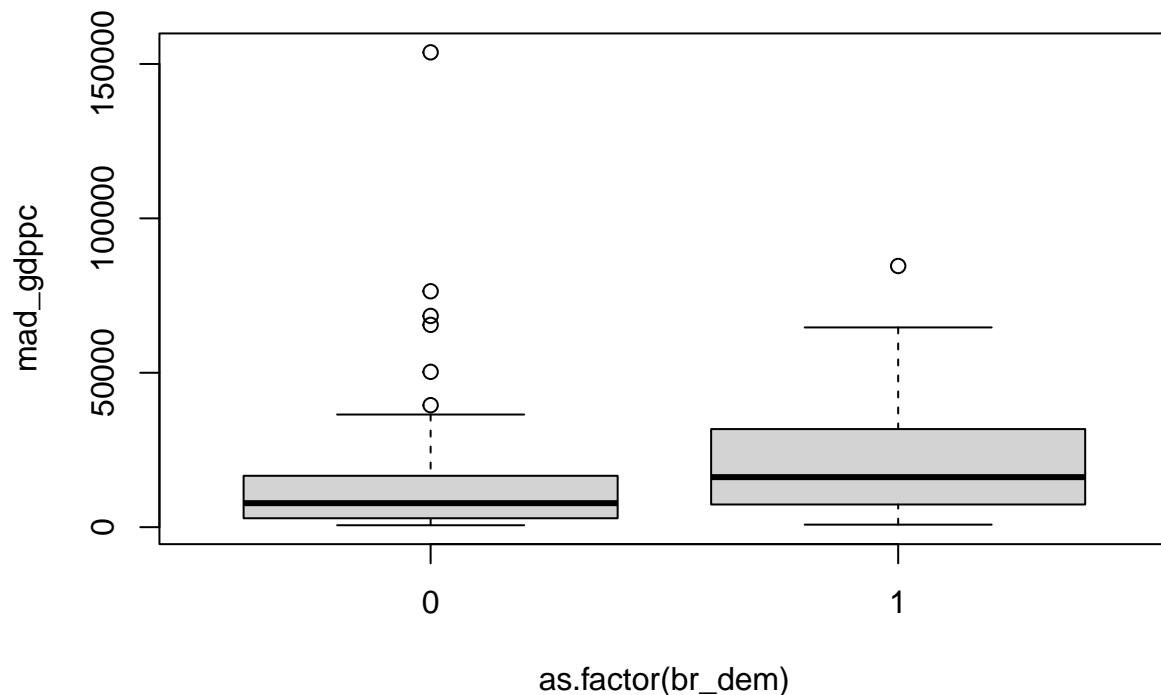
Let  $\mu_0$  be the mean GDP of countries not labeled as democracies and  $\mu_1$  be the mean GDP of countries labeled as democracies. We are testing  $H_0: \mu_0 = \mu_1$  vs  $H_a$ : the two means are different. We get a  $t$ -statistic of  $-1.95$  with  $113.99$  degrees of freedom, which corresponds to a p-value of  $0.053 > 0.05$ , so we fail to reject the null that there is a difference in GDPs by democracy status. The confidence interval for the differences in means is  $(-89, 13273)$ , which includes  $0$ , consistent with the  $t$ -test.

The assumptions are that the observations are independent both between and within groups and that the data are approximately normal. Independence probably does not hold because GDP is very likely dependent on trade and diplomacy between countries. The data are also right skewed, but there are many countries in both groups, so the central limit theorem suggests this is a reasonable assumption.

```
# T test
t.test(mad_gdppc~as.factor(br_dem), countries)
```

```
##
## Welch Two Sample t-test
##
## data: mad_gdppc by as.factor(br_dem)
## t = -1.9546, df = 113.99, p-value = 0.05308
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -13272.98486 88.92301
## sample estimates:
## mean in group 0 mean in group 1
## 14866.85 21458.88
```

```
# Distribution of GDP by democracy status
boxplot(mad_gdppc~as.factor(br_dem), countries)
```



```
# Countries in each group
table(as.factor(countries$br_dem))
```

```
##
##    0    1
##  74 118
```

2. Perform a formal analysis of variance for the difference in GDP per capita (encoded as `mad_gdppc`) by type of democracy (encoded as `gol_inst`). Comment on the assumptions of the test.

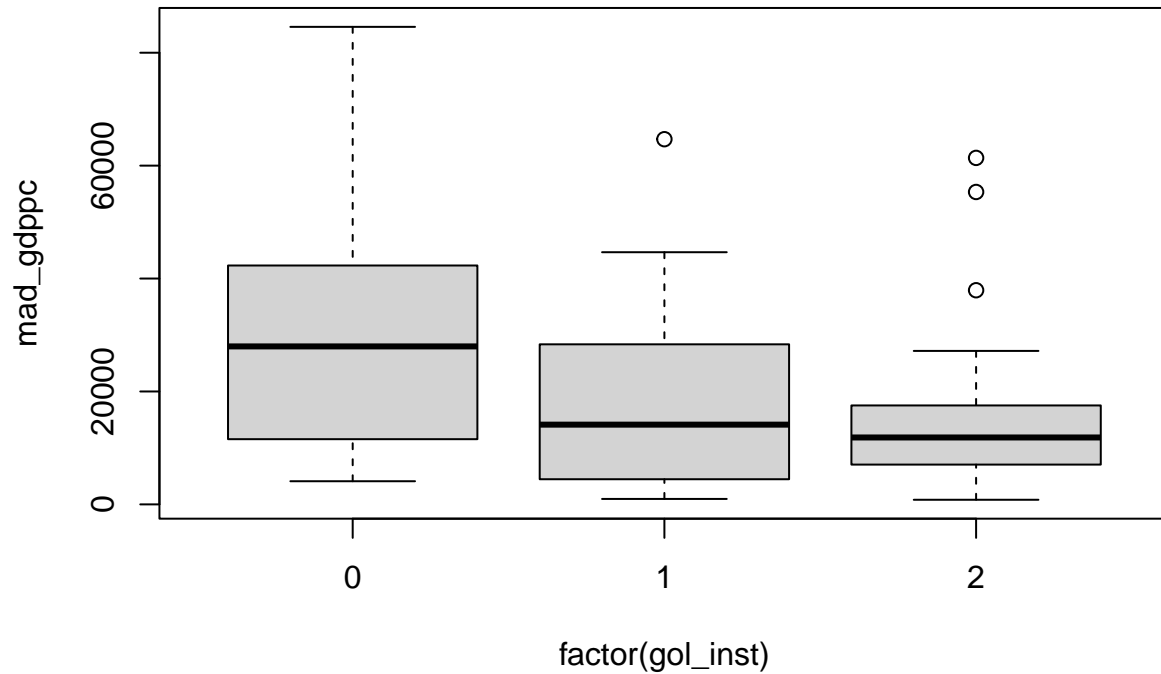
Let  $\mu_0$  be the mean GDP of countries with parliamentary democracies,  $\mu_1$  be the mean GDP of countries with semi-presidential democracies and  $\mu_2$  be the mean GDP of countries with presidential democracies. We want to test  $H_0: \mu_0 = \mu_1 = \mu_2$  vs  $H_a$ : the means are not all equal. We get an  $F$ -statistic of 5.919 for 2 and 91 degrees of freedom for a p-value of 0.004, suggesting that the mean GDPs of different types of democracies are not equal. The GDPs are slightly right skewed, but the sample size is moderate for each, so normality is a reasonable assumption. The variances are relatively similar (none are more than twice the others), so equal variance seems reasonable. Independence probably does not hold because the countries likely trade with each other, causing their GDPs to be correlated both within and between groups.

```
summary(aov(mad_gdppc~factor(gol_inst), countries))
```

```
##              Df    Sum Sq  Mean Sq F value  Pr(>F)
## factor(gol_inst)  2 3.227e+09 1.614e+09   5.919 0.00383 **
## Residuals       91 2.481e+10 2.726e+08
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 100 observations deleted due to missingness
```

```
boxplot(mad_gdppc~factor(gol_inst), countries)
```



```
by(countries$mad_gdppc, factor(countries$gol_inst), var, na.rm=T)/10^6
```

```
## factor(countries$gol_inst): 0
## [1] 340.34
## -----
## factor(countries$gol_inst): 1
## [1] 277.1644
## -----
## factor(countries$gol_inst): 2
## [1] 198.0793
```

```
table(countries$gol_inst[!is.na(countries$mad_gdppc)])
```

```
##
##  0  1  2
## 32 31 31
```

3. Formally test whether there is a difference in the proportion of countries that are democracies (**br\_dem**) by whether they are part of the British commonwealth (**br\_cw**). Comment on the assumptions of the test.

Let  $p_1$  be the proportion of democracies of British commonwealth countries and  $p_2$  be the proportion of democracies of non-British commonwealth countries. We are testing  $H_0 : p_1 = p_2$  vs  $H_a : p_1 \neq p_2$ . We get a  $z$ -statistic of  $\sqrt{4.06} = 2.02$ , which yields a  $p$ -value of 0.044, so we reject the null and conclude that the proportion of countries that are democracies is different between British commonwealth and non-British commonwealth countries.

All the cell counts are above the regularity condition ( $> 10$ ), but independence within and between groups is probably not satisfied because diplomacy and world news shared between the countries might affect whether a country is democratic.

```
prop.test(table(factor(countries$br_dem),factor(countries$br_cw)), correct = F)
```

```
##
## 2-sample test for equality of proportions without continuity
## correction
##
## data:  table(factor(countries$br_dem), factor(countries$br_cw))
## X-squared = 4.0642, df = 1, p-value = 0.0438
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.01008202 0.25560740
## sample estimates:
##      prop 1      prop 2
## 0.8108108 0.6779661
```

```
table(factor(countries$br_dem),factor(countries$br_cw))
```

```
##
##      0  1
## 0 60 14
## 1 80 38
```

## Manipulating new distributions

Let  $T_n \sim t_n$ . Find the following:

1. Distribution of  $T_n^2$

Let  $T_n = Z/\sqrt{V_n/v}$  where  $Z$  is a standard normal random variable and  $V_n$  is a  $\chi_n^2$  random variable.

$$T_n^2 = Z^2/(V_n/v) = (Z^2/1)/(V_n/v) = F_{1,n}$$

2. Distribution of  $T^{-2}$

$$T_n^{-2} = F_{1,n}^{-1} = F_{n,1}$$

## Bonferroni

Let  $H_1, H_2, \dots, H_n$  be a series of hypotheses, and let  $p_1, p_2, \dots, p_n$  be their corresponding p-values. Let  $m_0$  be the number of true null hypotheses, and let  $\alpha$  be the false discovery rate.

1. Show that the Bonferroni correction limits the family wise error rate to  $\alpha$ .

The family-wise error rate is equal to

$$P(\cup_{i=1}^{m_0} (p_i \leq \alpha/m)) \leq \sum_{i=1}^{m_0} P(p_i \leq \alpha/m) \leq m_0 \alpha/m \leq \alpha$$

2. Find the exact family-wise error rate if  $P(p_i \leq \alpha/m | p_j \leq \alpha/m) = \rho P(p_j \leq \alpha/m)$ ,  $P(p_i \leq \alpha/m | p_j \leq \alpha/m, p_k \leq \alpha/m) = \rho P(p_j \leq \alpha/m, p_k \leq \alpha/m)$ , etc.

$$\begin{aligned} P(\cup_{i=1}^{m_0} (p_i \leq \alpha/m)) &= \sum_{i=1}^{m_0} P(p_i \leq \alpha/m) - \sum_{i < j}^{m_0} P(p_j \leq \alpha/m | p_i \leq \alpha/m) P(p_j \leq \alpha/m) + \dots \\ &= m_0(\alpha/m) - \binom{m_0}{2} \rho(\alpha/m) + \binom{m_0}{3} \rho^2(\alpha/m) - \dots (-1)^{m_0+1} \rho^{m_0-1}(\alpha/m) \end{aligned}$$

3. Write a chunk of code to calculate the family-wise error rate if  $m_0 = m = 40$  (what does this mean in context?),  $\rho = 0.05$  (what does this mean in context?), and  $\alpha = 0.05$ . Comment on the rate when the errors are perfectly correlated, when the errors are mutually exclusive, and when the errors are independent.

If  $m_0 = m$ , all 40 hypotheses are actually false. If  $\rho = 0.05$ , the probability of falsely rejecting the null given that another test falsely rejected the null is 40 times what it would be alone. If  $\rho = 1$  and the observations are perfectly correlated, the family-wise error rate is the same as the probability of an individual error. If  $\rho = \alpha/m$  and there is no correlation, the family-wise error rate is slightly less than 0.05, and if it's 0, the family-wise error rate is at 0.05, the bound.

```
# This code only works for m < ~50 because of rounding errors with big choose functions,
i = 1
m_0 = 40
m = m_0
prob = 0
alpha = 0.05
rho = 1
while(i <= m_0) {
  prob = prob + choose(m_0,i) * (-1)^(i+1)*rho^(i-1)*alpha/m
  i = i + 1
}

# Bonferroni correction
print(alpha/m)
```

```
## [1] 0.00125
```

```
# Actual FWER
print(prob)
```

```
## [1] 0.001250002
```

## Simulations

1. Let  $X_1, X_2, \dots, X_n \sim \mathcal{N}(10, 3^2)$ . Then, let  $X_{i,1} = X_i + \epsilon_{i,1}$  and  $X_{i,2} = X_i + \beta + \epsilon_{i,2}$  with  $\epsilon_{i,j} \sim \mathcal{N}(0, 3^2)$ . What is the best way to test whether  $\beta = 0$ ? Let  $n = 10$ ,  $\beta = 1$ , and the number of simulations be 1000.

Paired  $t$ -test.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

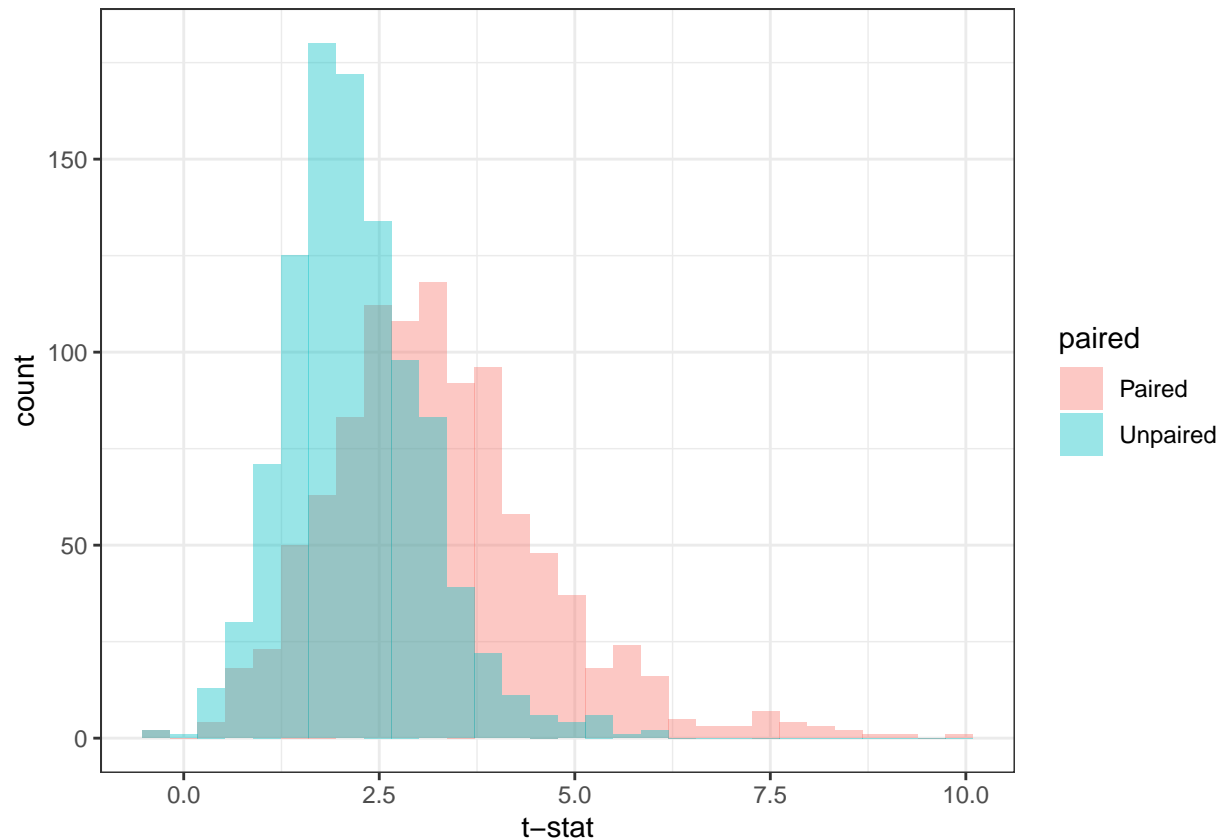
```
n = 10
beta = 1
nsim = 1000

paired_vec = vector(length = nsim)
unpaired_vec = vector(length = nsim)
for (i in 1:nsim) {
  x = rnorm(n, 10, 3)
  x_1 = x + rnorm(n, 5, 3)
  x_2 = x + beta + rnorm(n, 0, 3)
  paired_vec[i] = t.test(x_1-x_2)$statistic
  unpaired_vec[i] = t.test(x_1, x_2)$statistic
}

df = data.frame("t-stat" = c(paired_vec, unpaired_vec),
               "paired" = c(rep("Paired", nsim), rep("Unpaired", nsim)), check.names = F)

ggplot(df, aes(x=`t-stat`, fill=paired)) +
  geom_histogram(alpha=0.4, position="identity") +
  theme_bw()
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



2. Let  $\beta_1 = 1, \beta_2 = 2, \dots, \beta_{n_1} = n_1$ . Let  $X_{i,j} = \beta_i + \epsilon_{i,j}$  for  $j = 1$  to  $n_2$  with  $\epsilon_{i,j} \sim \mathcal{N}(0, 5^2)$ . What is the best way to test whether  $\beta_1 = \beta_2 = \dots = \beta_{n_1}$ ? If we break each group up into  $n_3$  subgroups, what happens to our power? (Think of this as accidentally choosing too many categories.) Let  $n_1 = 10$ ,  $n_2 = 30$ ,  $n_3 = 6$ , and the number of simulations be 1000.

```
nsim = 1000
n_1 = 10
n_2 = 30
n_3 = 6
betas = 1:n_1

f_correct = vector(length = nsim)
f_incorrect = vector(length = nsim)

for (i in 1:nsim) {
  df = data.frame(matrix(nrow = 0, ncol = 0))
  for (beta in betas) {
    x = beta + rnorm(n_2, 0, 5)
    df = rbind(df, cbind(x, beta))
  }
  df$beta <- as.factor(df$beta)
  f_correct[i] = summary(aov(x~beta, df))[[1]][1, 'F value']
  df$beta = as.numeric(as.character(df$beta))
  df$beta = (df$beta - 1) * n_3 + rep(rep(0:(n_3-1), n_2/n_3), n_1)
  df$beta <- as.factor(df$beta)
```

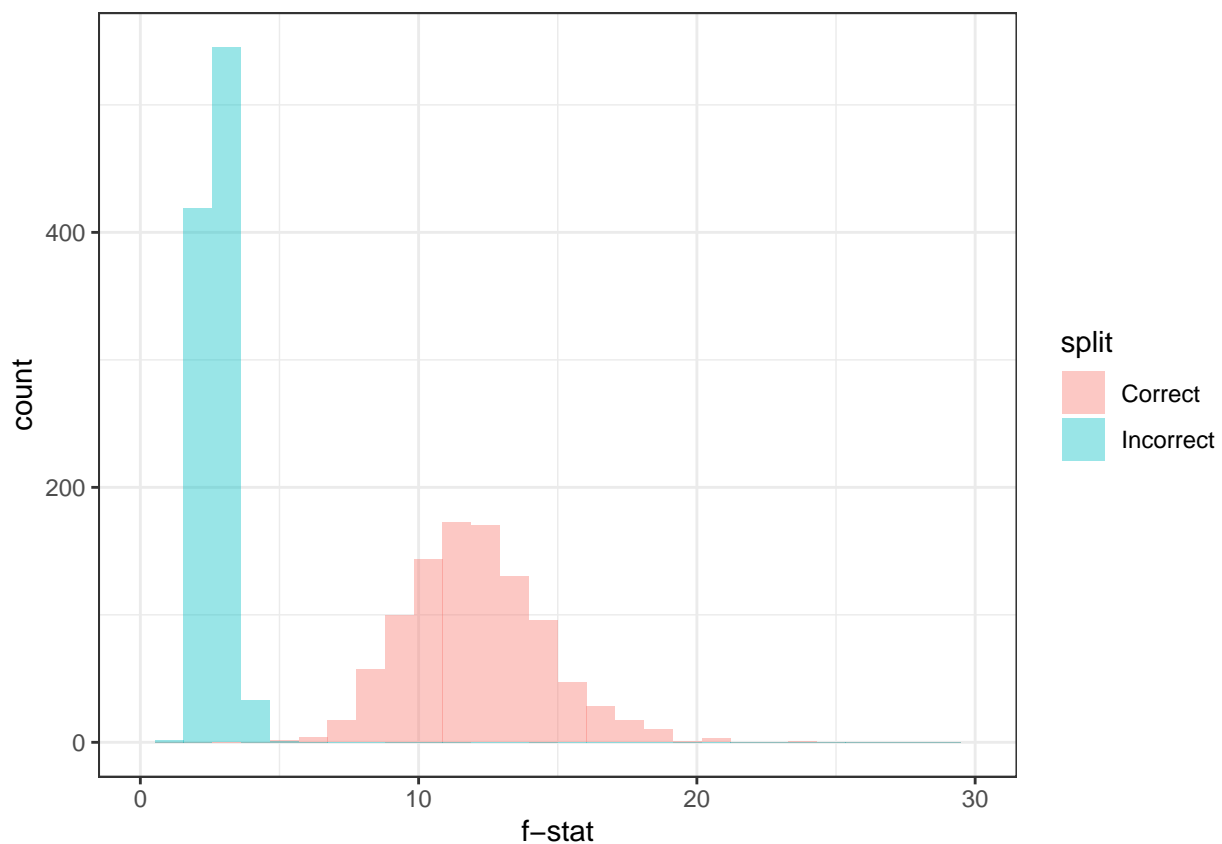
```
f_incorrect[i] = summary(aov(x~beta, df))[[1]][1,'F value']
}

df = data.frame("f-stat" = c(f_correct, f_incorrect),
               "split" = c(rep("Correct", nsim), rep("Incorrect", nsim)), check.names = F)

ggplot(df, aes(x=`f-stat`, fill=split)) +
  geom_histogram(alpha=0.4, position="identity") +
  theme_bw() + xlim(0, 30)
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

```
## Warning: Removed 4 rows containing missing values (geom_bar).
```



3. Let  $X_i \sim \mathcal{N}(0, 1)$  for  $i$  from 1 to  $n$ . Show that the p-values of the test  $H_0 : \mu = 0$ ,  $H_a : \mu \neq 0$  are uniformly distributed. Let  $X_i \sim \text{Exp}(1) - 1$  for  $i$  from 1 to  $n$ . Show that the p-values of the test  $H_0 : \mu = 0$ ,  $H_a : \mu \neq 0$  are not uniformly distributed for small  $n$ . Does the skew of the distribution make sense? Compare this to a large  $n$  ( $\sim 100$ ).

```
par(mfrow=c(1,2))
set.seed(139)
n = 5
nsim = 100000
```



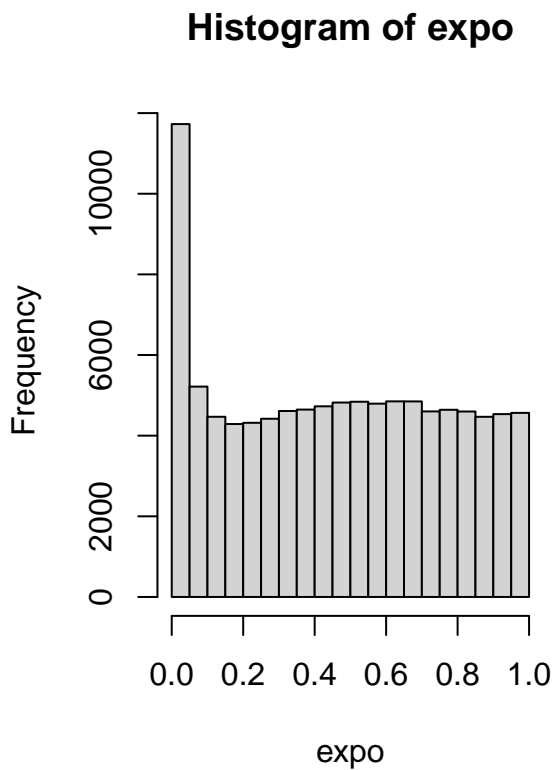
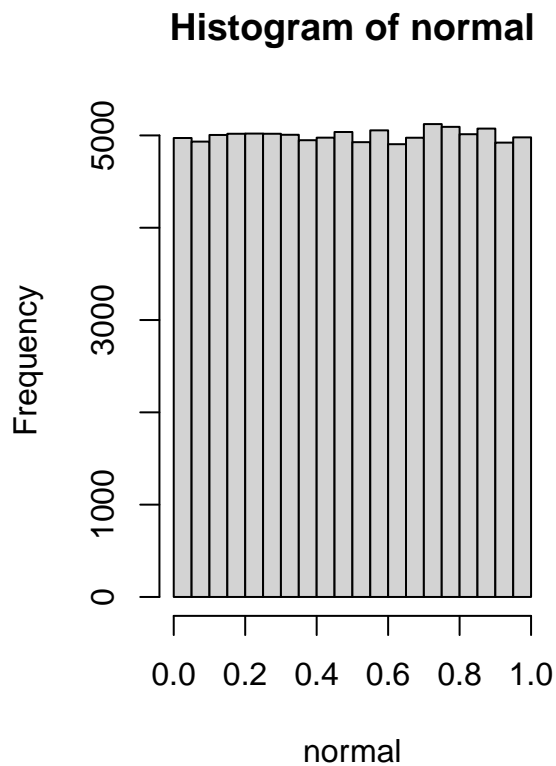
```
normal = vector(length = nsim)
for (i in 1:nsim) {
  x = rnorm(n, 0, 1)
  normal[i] = t.test(x)$p.value
}

hist(normal)
mean(normal<0.05)
```

```
## [1] 0.04972
```

```
expo = vector(length = nsim)
for (i in 1:nsim) {
  x = rexp(n, 1) - 1
  expo[i] = t.test(x)$p.value
}

hist(expo)
```



```
mean(expo<0.05)
```

```
## [1] 0.11728
```