Nonparametric Analysis of US Dairy Production and Consumption

Permutation Tests for GAM

Teo Bucci* Filippo Cipriani† Gabriele Corbo‡ Andrea Puricelli§

2023-02-17

Contents

1 Load libraries and data

```
library(pbapply)
library(mgcv)
```

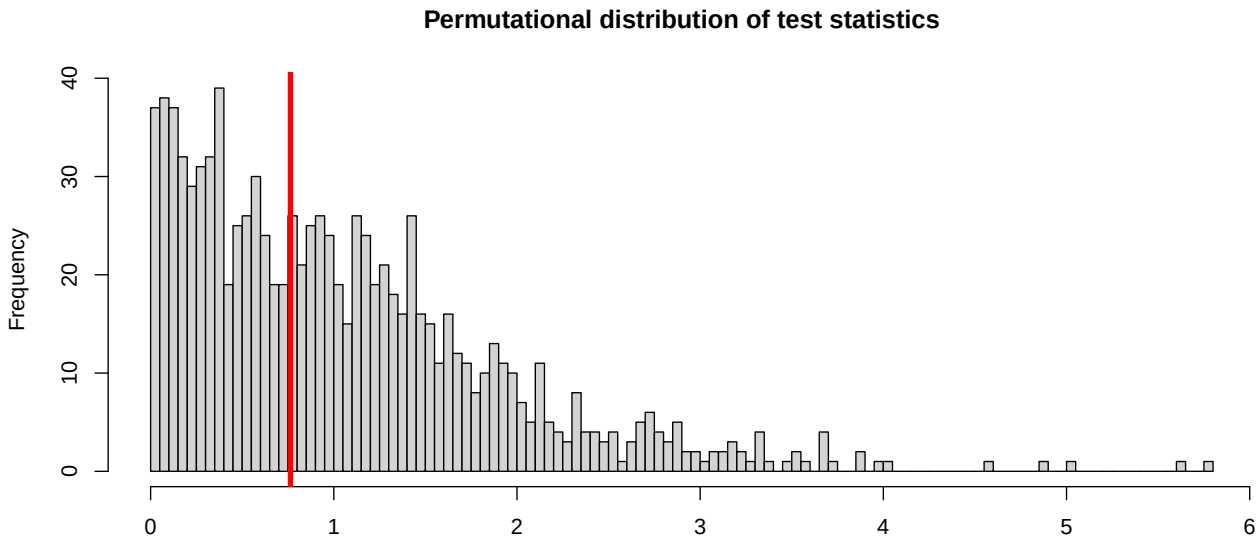
```
data_path = file.path('data_updated_2021')
output_path = file.path('output')
data_infl =
    read.table(
        file.path(data_path, 'production_facts_inflated.csv'),
        header = T,
        sep = ';'
    )
set.seed(1)
B = 1000
n = nrow(data_infl)
```

*teo.bucci@mail.polimi.it
†filippo.cipriani@mail.polimi.it
‡gabriele.corbo@mail.polimi.it
§andrea3.puricelli@mail.polimi.it

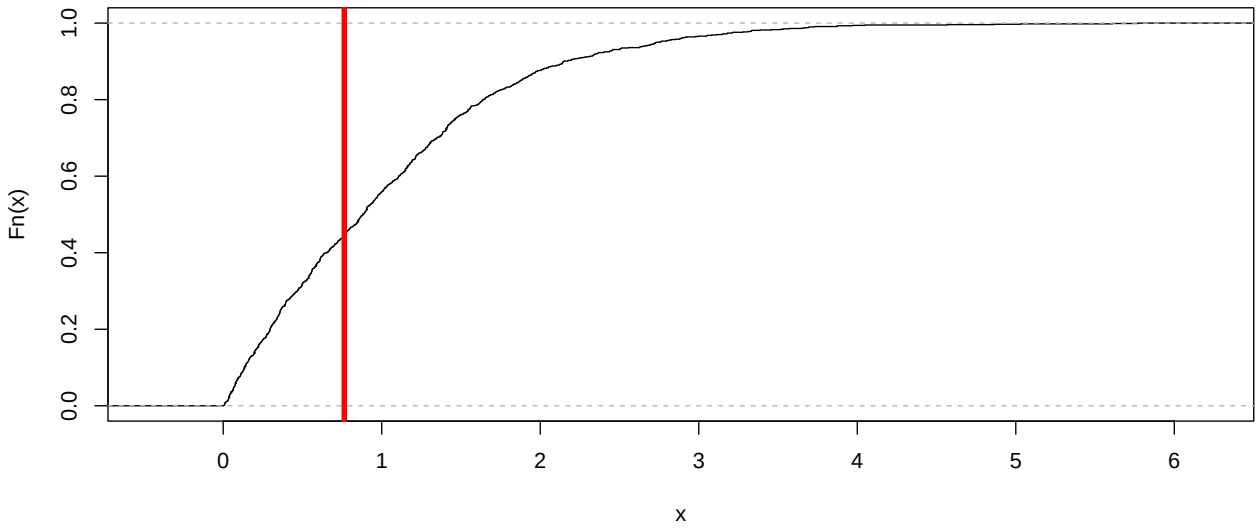
2 H_0: alfalfa_hay_price $= 0$ VS H_1: alfalfa_hay_price $\neq 0$

```
model_gam = gam(
    avg_price_milk ~ s(avg_milk_cow_number, bs = 'cr')
    + s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs
    + alfalfa_hay_price
    + s(slaughter_cow_price, bs = 'cr'),
    data = data_infl
)
T0 = abs(summary(model_gam)$p.table[5, 3])
gam.H0 = gam(
    avg_price_milk ~ s(avg_milk_cow_number, bs = 'cr')
    + s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs
    + s(slaughter_cow_price, bs = 'cr'),
    data = data_infl
)
res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
    gam.perm = gam(
        Y.perm.H0 ~ s(avg_milk_cow_number, bs = 'cr')
        + s(milk_per_cow, bs = 'cr')
        + s(dairy_ration, bs = 'cr')
        + milk_feed_price_ratio
        + milk_cow_cost_per_animal
        + milk_volume_to_buy_cow_in_lbs
        + alfalfa_hay_price
        + s(slaughter_cow_price, bs = 'cr'),
        data = data_infl
    )
    return(abs(summary(gam.perm)$p.table[5, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(sort(T_H0)[-1000],
     breaks = 100,
     main = 'Permutational distribution of test statistics',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```




```
plot(ecdf(sort(T_H0)[-1000]), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```




```
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0.555
```

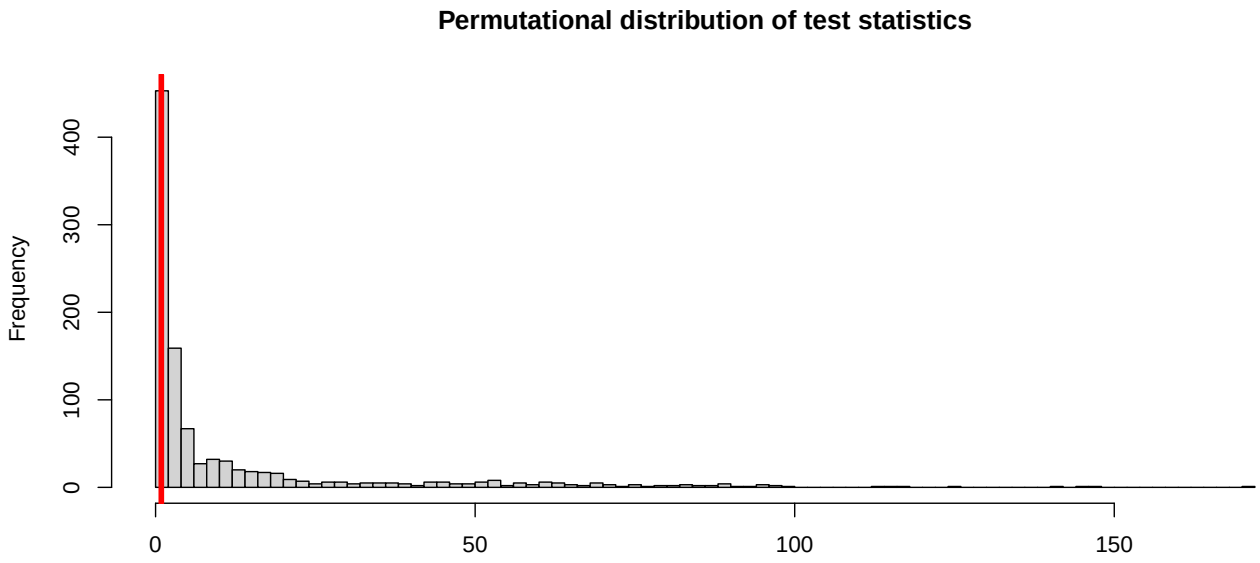
We cannot reject H_0 and proceed removing the covariate `alfalfa_hay_price` from the model.

3 H_0: `avg_milk_cow_number` $= 0$ VS H_1: `avg_milk_cow_number` $\neq 0$

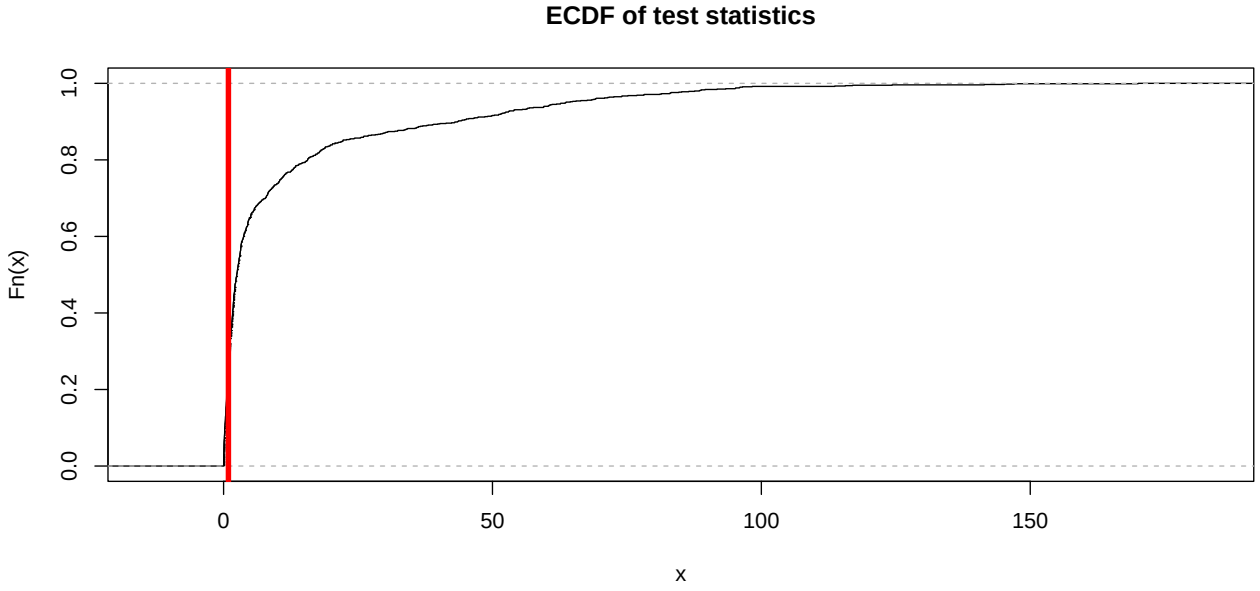
```
model_gam = gam(
    avg_price_milk ~ s(avg_milk_cow_number, bs = 'cr')
    + s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
    + milk_feed_price_ratio
```

```
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs
    + s(slaughter_cow_price, bs = 'cr'),
    data = data_infl
)
T0 = abs(summary(model_gam)$s.table[1, 3])
gam.H0 = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs
    + s(slaughter_cow_price, bs = 'cr'),
    data = data_infl
)
res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
    gam.perm = gam(
        Y.perm.H0 ~ s(avg_milk_cow_number, bs = 'cr')
        + s(milk_per_cow, bs = 'cr')
        + s(dairy_ration, bs = 'cr')
        + milk_feed_price_ratio
        + milk_cow_cost_per_animal
        + milk_volume_to_buy_cow_in_lbs
        + s(slaughter_cow_price, bs = 'cr'),
        data = data_infl
    )
    return(abs(summary(gam.perm)$s.table[1, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(T_H0,
     breaks = 100,
     main = 'Permutational distribution of test statistics',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```




```
plot(ecdf(T_H0), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```




```
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0.746
```

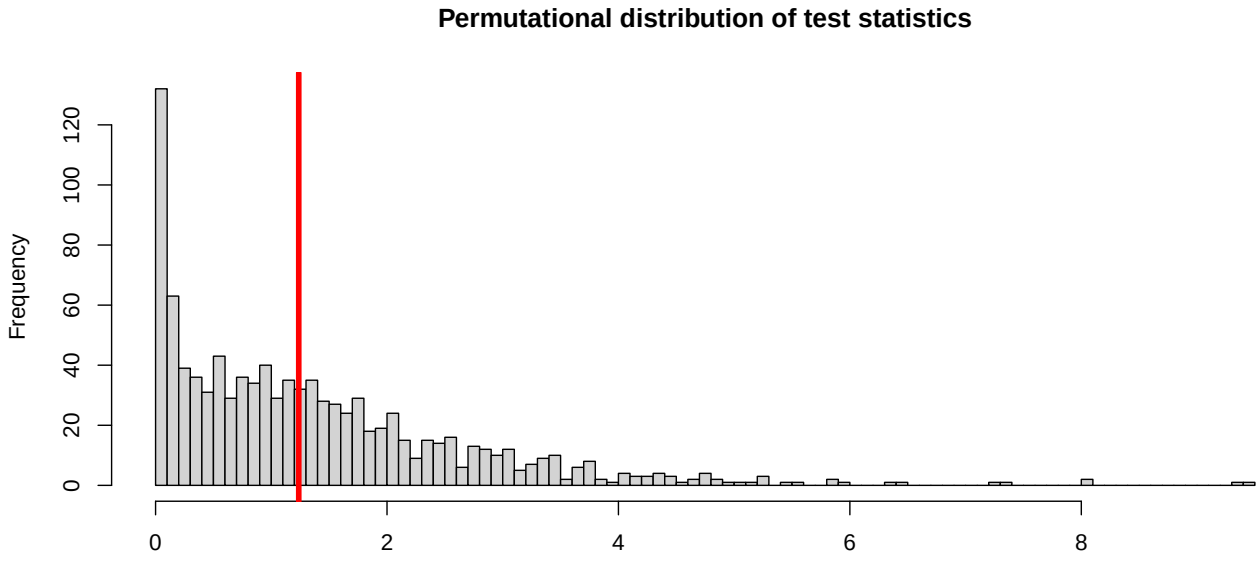
We cannot reject H_0 and proceed removing the covariate `avg_milk_cow_number` from the model.

4 H_0: `slaughter_cow_price` $= 0$ VS H_1: `slaughter_cow_price` $\neq 0$

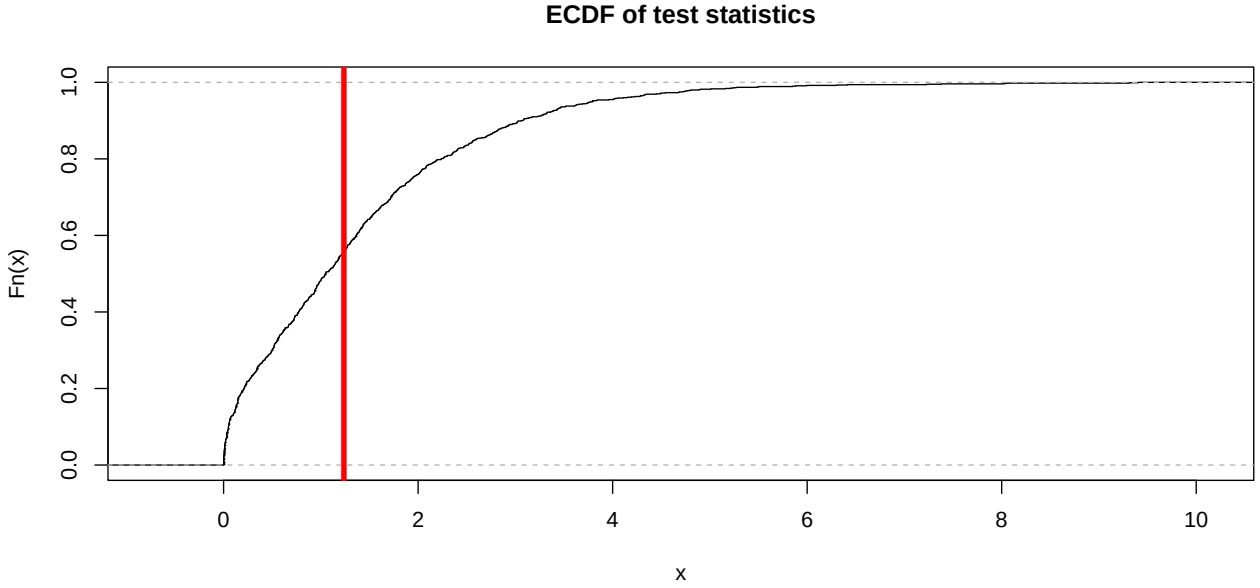
```
model_gam = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
```

```
    + milk_volume_to_buy_cow_in_lbs
    + s(slaughter_cow_price, bs = 'cr'),
    data = data_infl
)
T0 = abs(summary(model_gam)$s.table[3, 3])
gam.H0 = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs
    ,
    data = data_infl
)
res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
    gam.perm = gam(
        Y.perm.H0 ~ s(milk_per_cow, bs = 'cr')
        + s(dairy_ration, bs = 'cr')
        + milk_feed_price_ratio
        + milk_cow_cost_per_animal
        + milk_volume_to_buy_cow_in_lbs
        + s(slaughter_cow_price, bs = 'cr'),
        data = data_infl
    )
    return(abs(summary(gam.perm)$s.table[3, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(T_H0,
     breaks = 100,
     main = 'Permutational distribution of test statistics',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```




```
plot(ecdf(T_H0), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```




```
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0.443
```

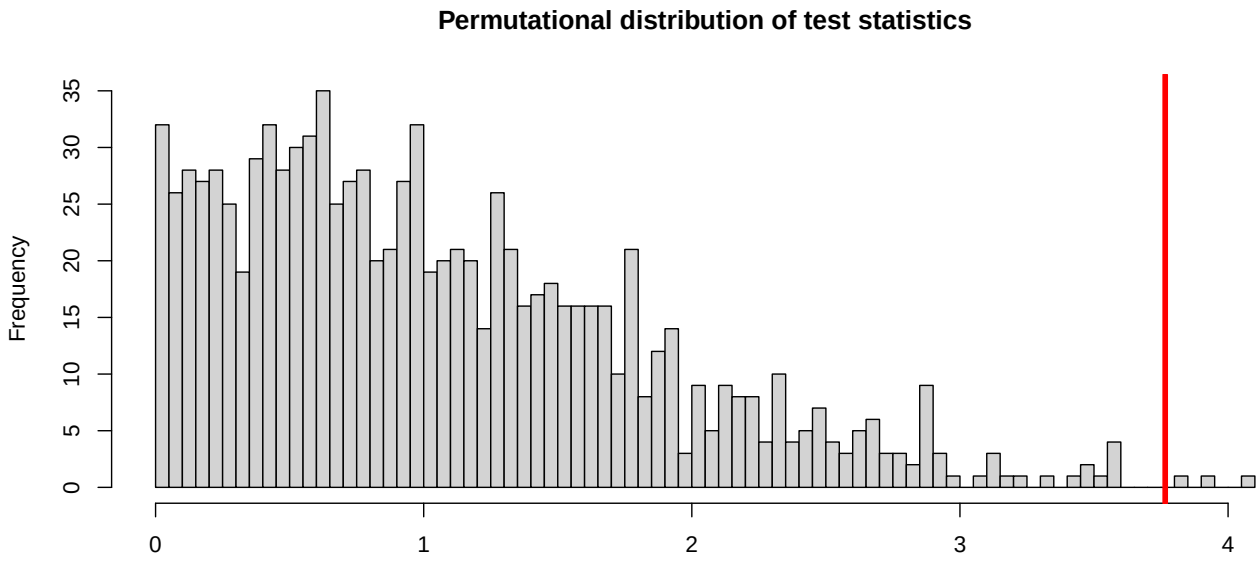
We cannot reject H_0 and proceed removing the covariate `slaughter_cow_price` from the model.

5 H_0: `milk_volume_to_buy_cow_in_lbs` $= 0$ VS H_1: `milk_volume_to_buy_cow_in` $\neq 0$

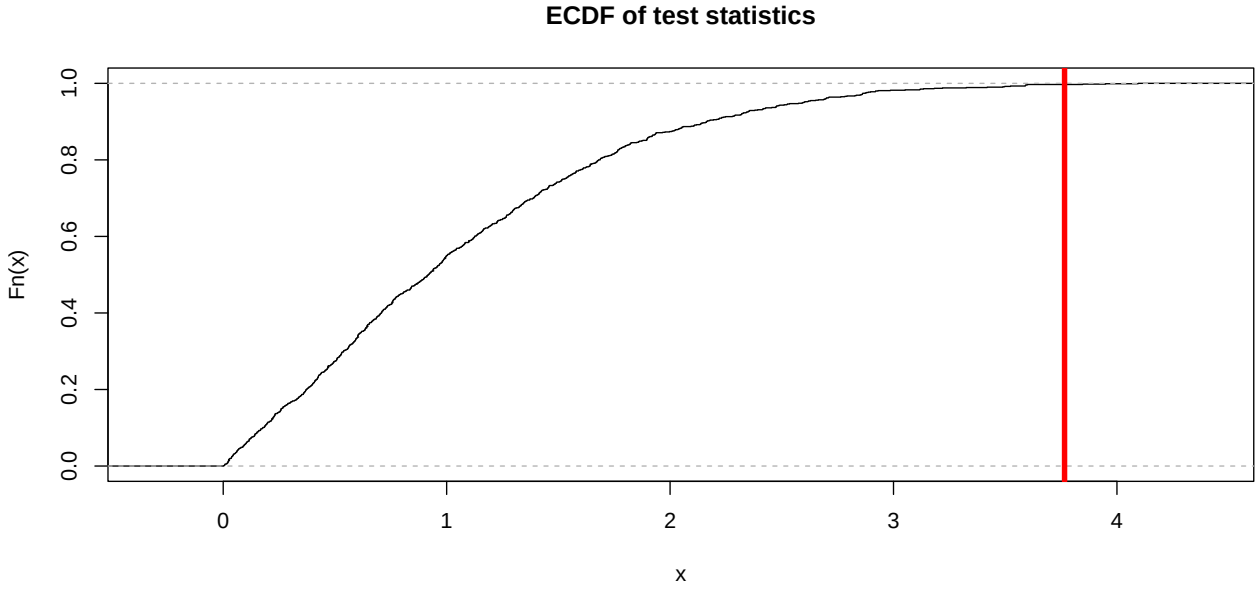
```
model_gam = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
```

```
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs,
    data = data_infl
)
T0 = abs(summary(model_gam)$p.table[4, 3])
gam.H0 = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal,
    data = data_infl
)
res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
    gam.perm = gam(
        Y.perm.H0 ~ s(milk_per_cow, bs = 'cr')
        + s(dairy_ration, bs = 'cr')
        + milk_feed_price_ratio
        + milk_cow_cost_per_animal
        + milk_volume_to_buy_cow_in_lbs,
        data = data_infl
    )
    return(abs(summary(gam.perm)$p.table[4, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(T_H0,
     breaks = 100,
     main = 'Permutational distribution of test statistics',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```




```
plot(ecdf(T_H0), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```




```
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0.003
```

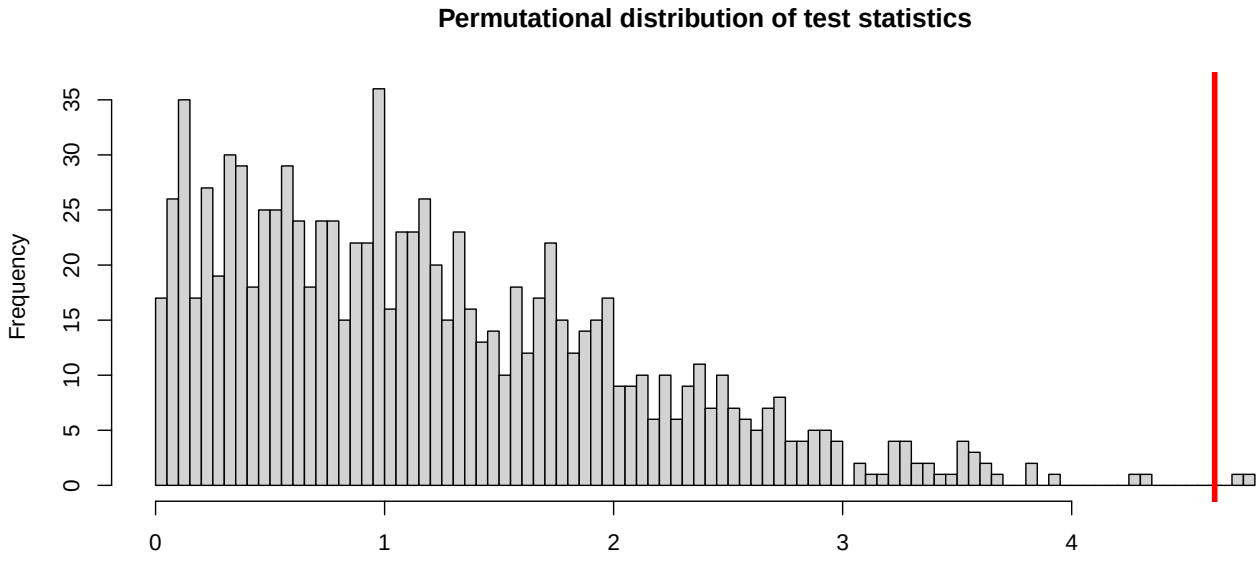
We reject H_0, maintaining the covariate `milk_volume_to_buy_cow_in_lbs`.

6 H_0: `milk_cow_cost_per_animal` $= 0$ VS H_1: `milk_cow_cost_per_animal` $\neq 0$

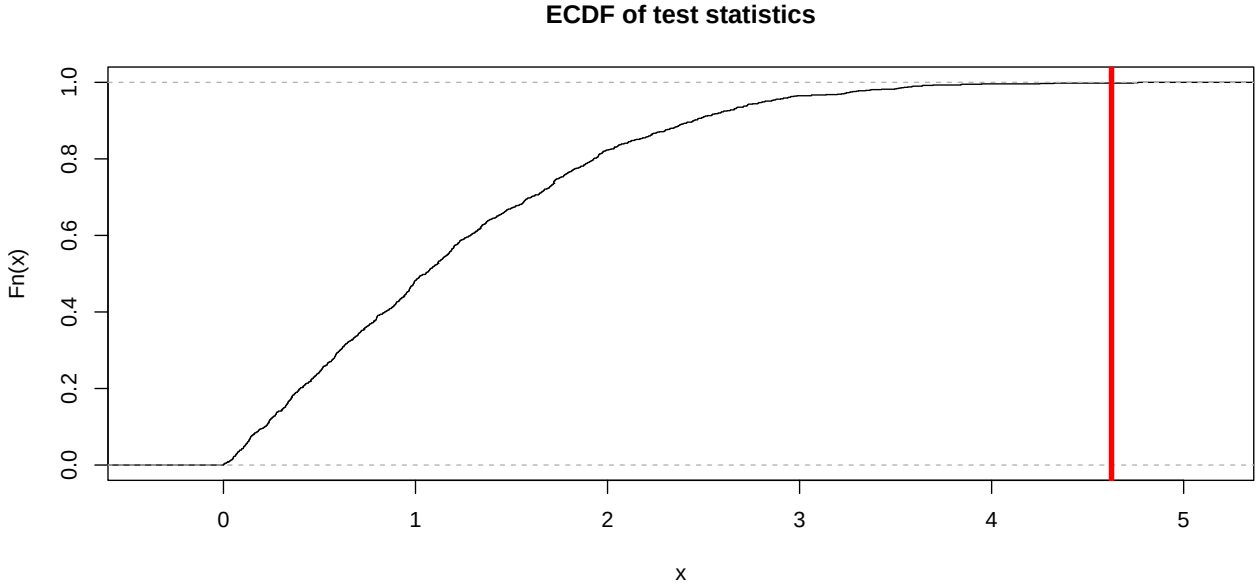
```
model_gam = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
```

```
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs,
    data = data_infl
)
T0 = abs(summary(model_gam)$p.table[3, 3])
gam.H0 = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
    + milk_feed_price_ratio
    + milk_volume_to_buy_cow_in_lbs,
    data = data_infl
)
res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
    gam.perm = gam(
        Y.perm.H0 ~ s(milk_per_cow, bs = 'cr')
        + s(dairy_ration, bs = 'cr')
        + milk_feed_price_ratio
        + milk_cow_cost_per_animal
        + milk_volume_to_buy_cow_in_lbs,
        data = data_infl
    )
    return(abs(summary(gam.perm)$p.table[3, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(T_H0,
     breaks = 100,
     main = 'Permutational distribution of test statistics',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```




```
plot(ecdf(T_H0), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```




```
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0.002
```

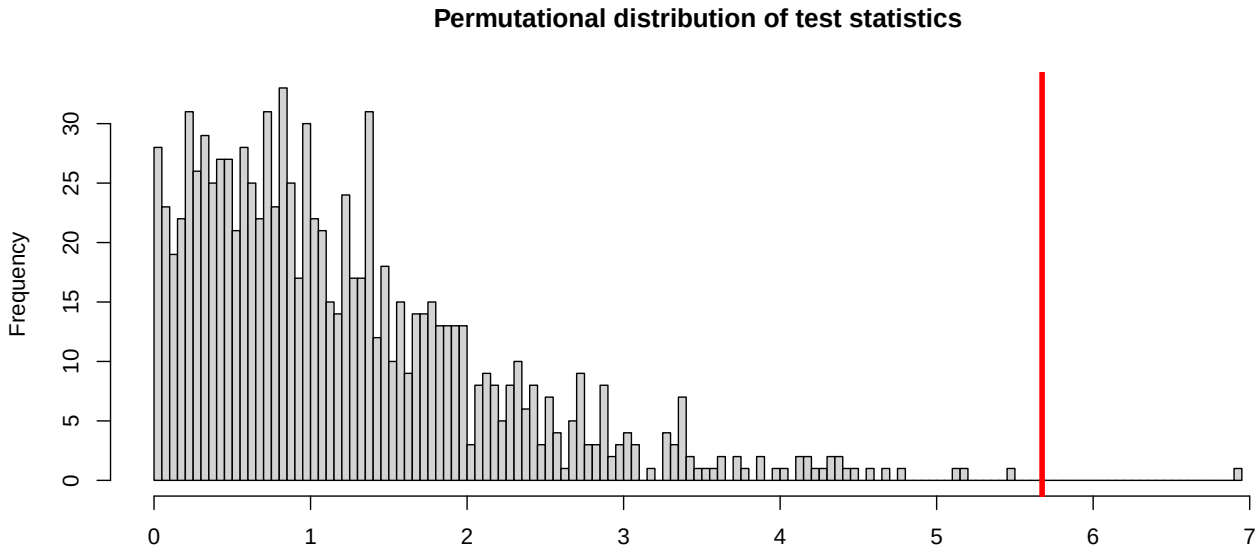
We reject H_0, maintaining the covariate `milk_cow_cost_per_animal`.

7 H_0: `milk_feed_price_ratio` $= 0$ VS H_1: `milk_feed_price_ratio` $\neq 0$

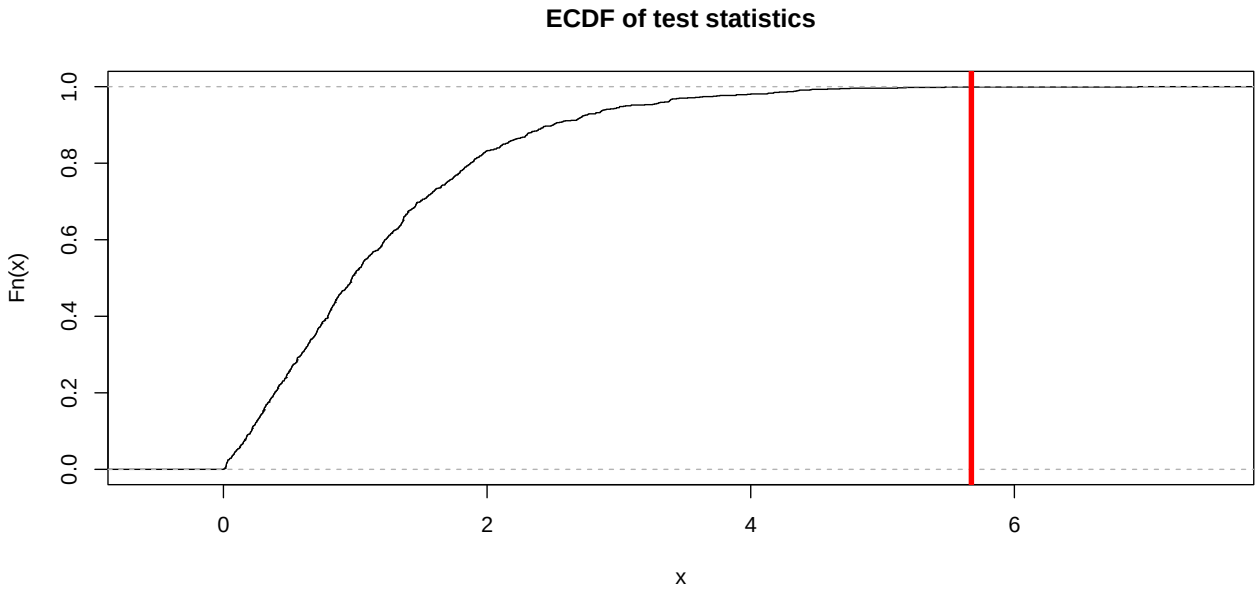
```
model_gam = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
```

```
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs,
    data = data_infl
)
T0 = abs(summary(model_gam)$p.table[2, 3])
gam.H0 = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs,
    data = data_infl
)
res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
    gam.perm = gam(
        Y.perm.H0 ~ s(milk_per_cow, bs = 'cr')
        + s(dairy_ration, bs = 'cr')
        + milk_feed_price_ratio
        + milk_cow_cost_per_animal
        + milk_volume_to_buy_cow_in_lbs,
        data = data_infl
    )
    return(abs(summary(gam.perm)$p.table[2, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(T_H0,
     breaks = 100,
     main = 'Permutational distribution of test statistics',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```




```
plot(ecdf(T_H0), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```




```
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0.001
```

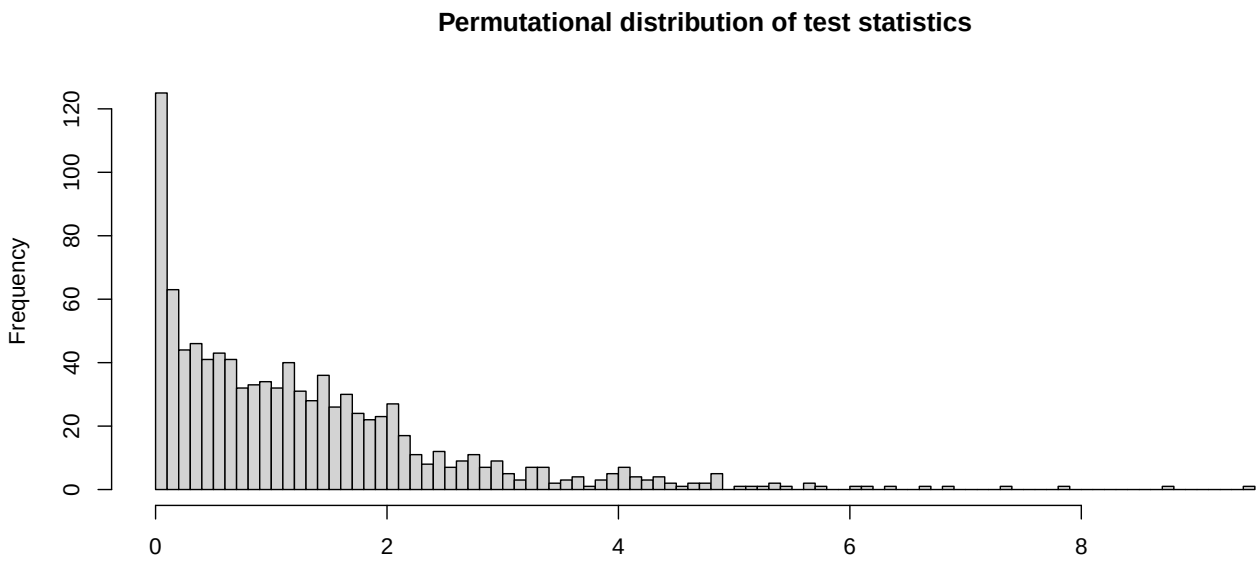
We reject H_0, maintaining the covariate `milk_feed_price_ratio`.

8 H_0: `milk_per_cow` $= 0$ VS H_1: `milk_per_cow` $\neq 0$

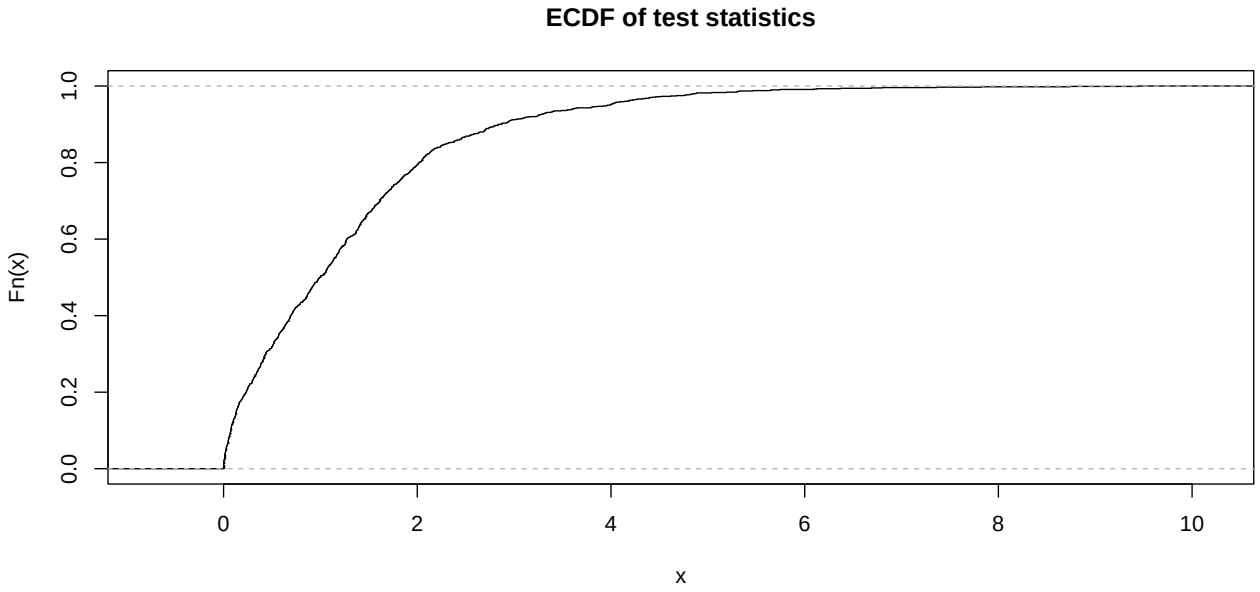
```
model_gam = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
```

```
    + milk_volume_to_buy_cow_in_lbs,
    data = data_infl
)
T0 = abs(summary(model_gam)$s.table[1, 3])
gam.H0 = gam(
    avg_price_milk ~ s(dairy_ration, bs = 'cr')
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs,
    data = data_infl
)
res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
    gam.perm = gam(
        Y.perm.H0 ~ s(milk_per_cow, bs = 'cr')
        + s(dairy_ration, bs = 'cr')
        + milk_feed_price_ratio
        + milk_cow_cost_per_animal
        + milk_volume_to_buy_cow_in_lbs,
        data = data_infl
    )
    return(abs(summary(gam.perm)$s.table[1, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(T_H0,
     breaks = 100,
     main = 'Permutational distribution of test statistics',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```



```
plot(ecdf(T_H0), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```




```
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0
```

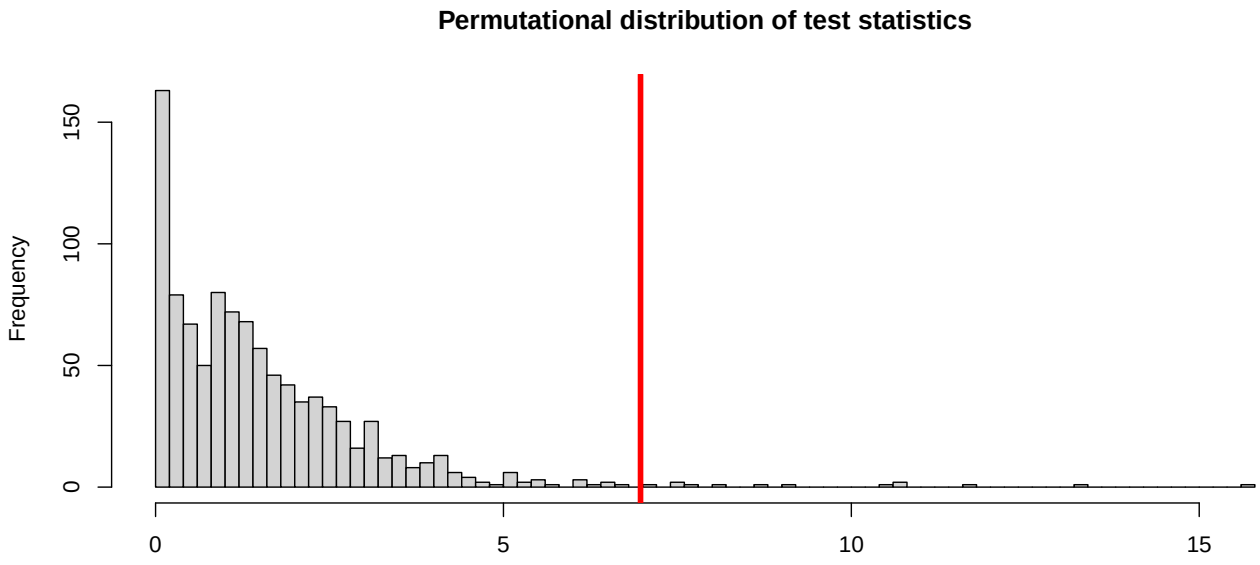
We reject H_0, maintaining the covariate `milk_per_cow`.

9 H_0: `dairy_ration` $= 0$ VS H_1: `dairy_ration` $\neq 0$

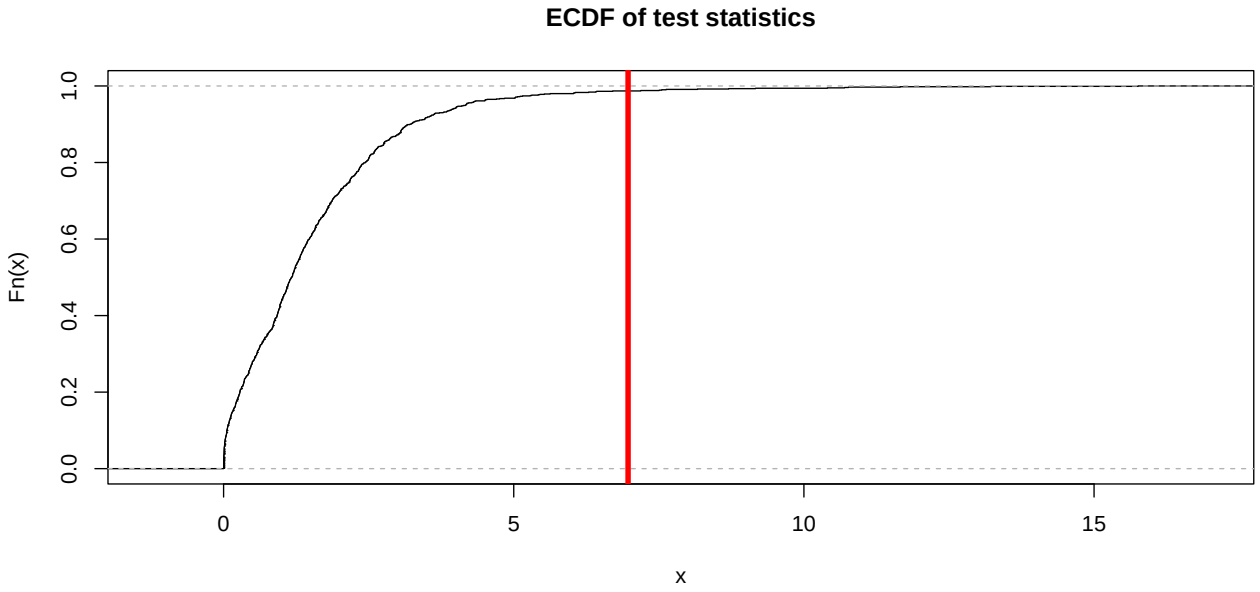
```
model_gam = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + s(dairy_ration, bs = 'cr')
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs,
    data = data_infl
)
T0 = abs(summary(model_gam)$s.table[2, 3])
gam.H0 = gam(
    avg_price_milk ~ s(milk_per_cow, bs = 'cr')
    + milk_feed_price_ratio
    + milk_cow_cost_per_animal
    + milk_volume_to_buy_cow_in_lbs,
    data = data_infl
)
res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
```

```
    gam.perm = gam(
        Y.perm.H0 ~ s(milk_per_cow, bs = 'cr')
        + s(dairy_ration, bs = 'cr')
        + milk_feed_price_ratio
        + milk_cow_cost_per_animal
        + milk_volume_to_buy_cow_in_lbs,
        data = data_infl
    )
    return(abs(summary(gam.perm)$s.table[2, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(T_H0,
     breaks = 100,
     main = 'Permutational distribution of test statistics',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```




```
plot(ecdf(T_H0), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```




```
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0.013
```

We reject H_0, maintaining the covariate `dairy_ration`.