

# 1 Gestió de les dades

Llegim les dades `pesosIndividuales.xlsx`

## 1.1 Tractament de les dades faltants

Si concluïm en que la distribució de les dades faltants es pot considerar aleatòria i que n'hi ha poques podem omitir aquestes dades i treure-les de la base de dades sabent que no esbiaixarà l'anàlisi. Encara que cal tenir en compte que hi haurà una petita pèrdua d'informació.

## 1.2 Tractament de les variables

Posem les variables `Box` i `Treat` com a factors.

Creem les següents variables d'interès amb l'objectiu de controlar l'efecte individual i el pes al naixement:

- Guany de pes:

$$BW_{41-28} = BW_{41} - BW_{28}$$

- Index de l'increment de pes:

$$Index_{41-28} = 100 * \frac{BW_{41}}{BW_{28}}$$

- Taxa de l'increment de pes:

$$Taxa_{41-28} = 100 * \frac{BW_{41} - BW_{28}}{BW_{28}}$$

Caldrà tenir en compte que els valors petits d'aquestes variables poden provindre d'individus amb pesos inicialment grans (o també petits).

Creem una base de dades sense aquells individus amb increments de pes negatius.

```
pesIndDifNoNegativa <- pesInd[-pesInd$difBW41_28<0,]
```

## 2 Comparació gràfica

Usem el package `patchwork` per a fer layouts de `ggplot2`.

Amb les llibreries `ggplot2`, `devtools` i `easyGgplot2` podrem obtenir ggplots d'una manera considerablement senzilla.

Creem una funció per a simplificar l'obtenció dels histogrames per a comparar tractaments:

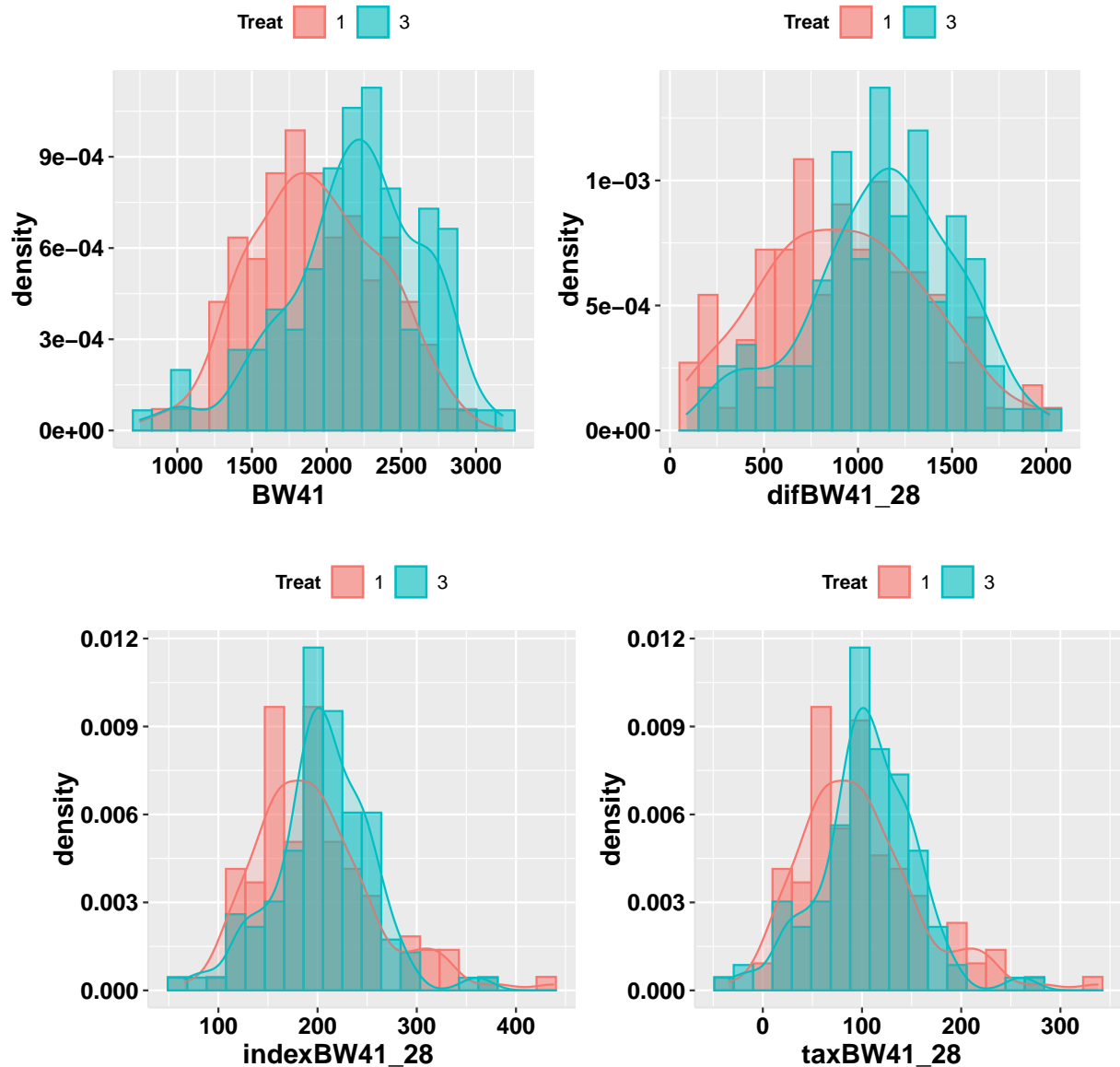
```
histComparatiu(BD, Tract1, Tract2, TipusGraf=c(1,2), NombreBins, var)
```

Els arguments de la funció són:

- BD la base de dades.

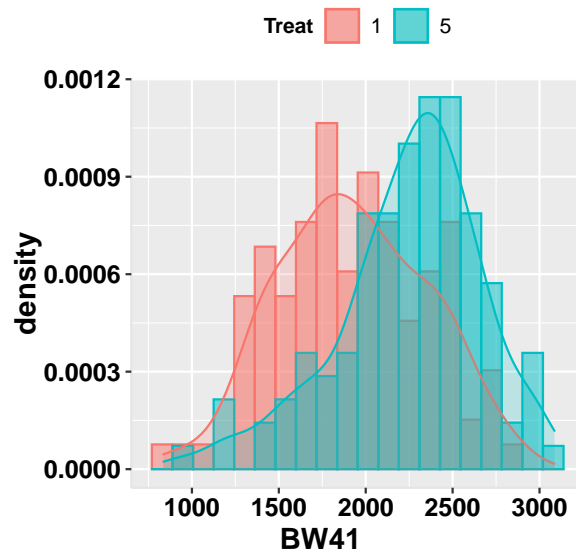
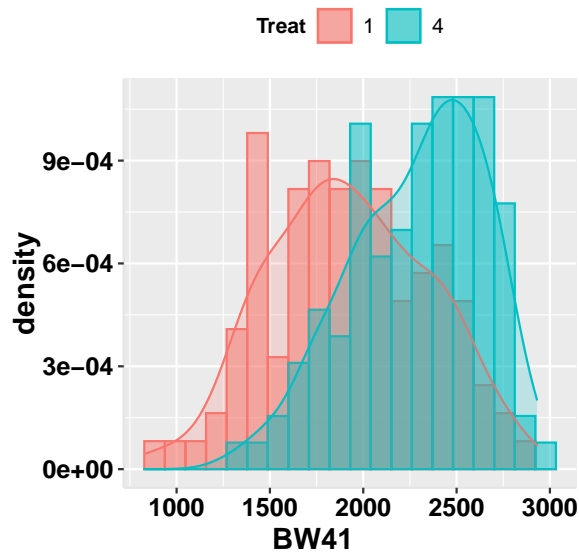
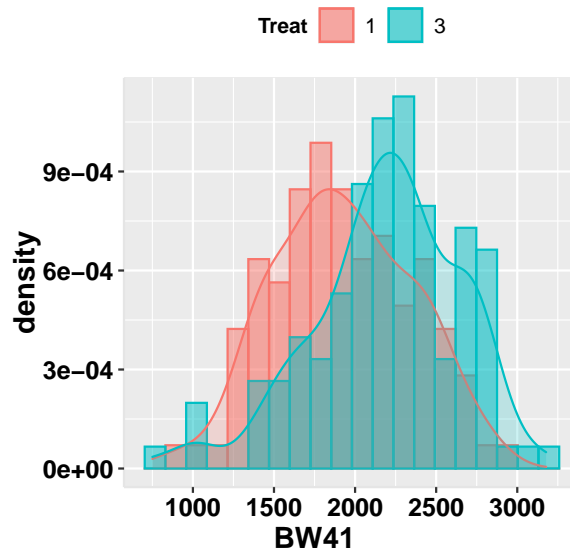
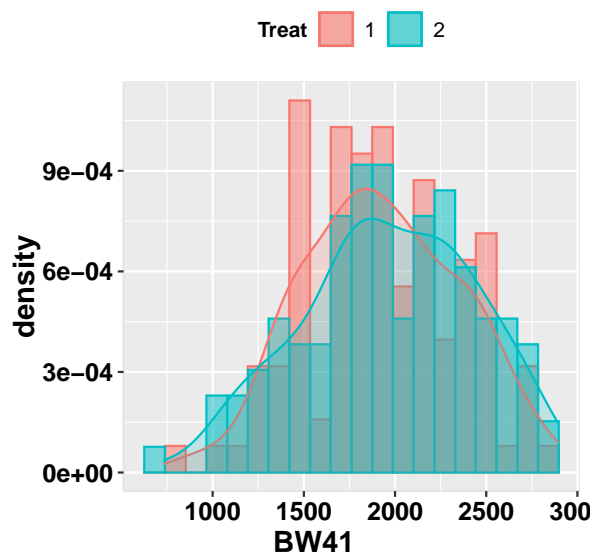
- Trac1 i Tract2 són els tractaments a comparar.
- TipusGraf si és 1 representem els histogrames solapant-se al mateix eix de les y, si val 2 representarem els histogrames un a sobre de l'altre pero en diferents eix y.
- NombreBins és el nombre de "caixes" en que estarà dividit l'histograma de cadascún dels tractaments, el valor predeterminat és 10.
- var és la variable sobre la que es faràn els histogrames.

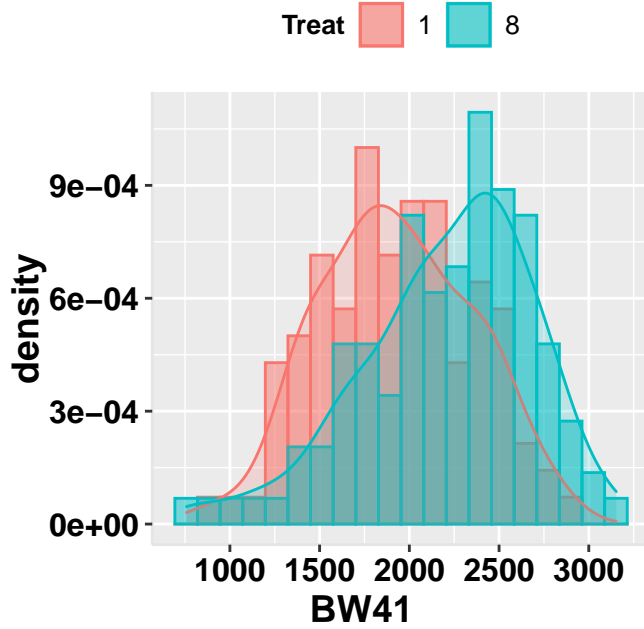
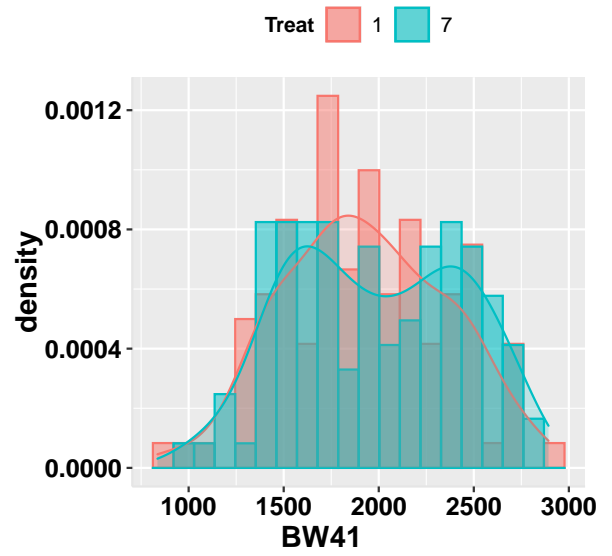
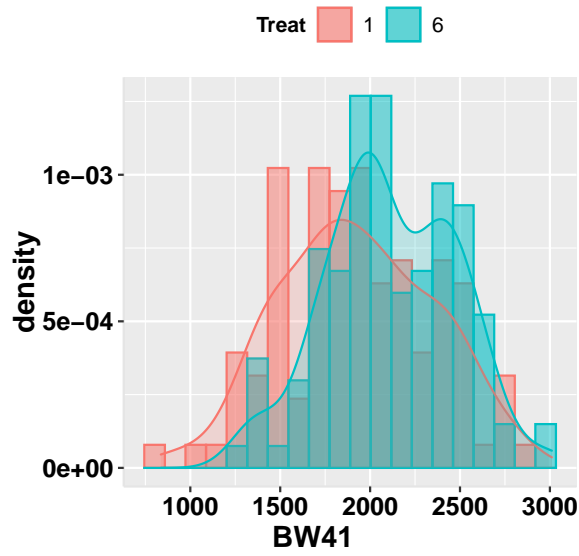
## 2.1 Histogrames de les noves variables:



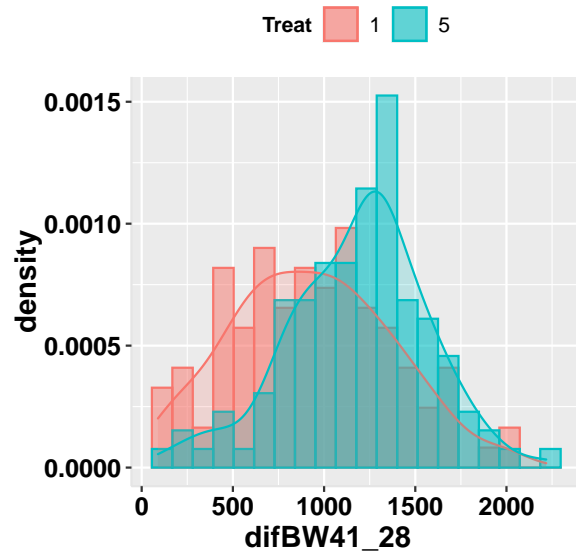
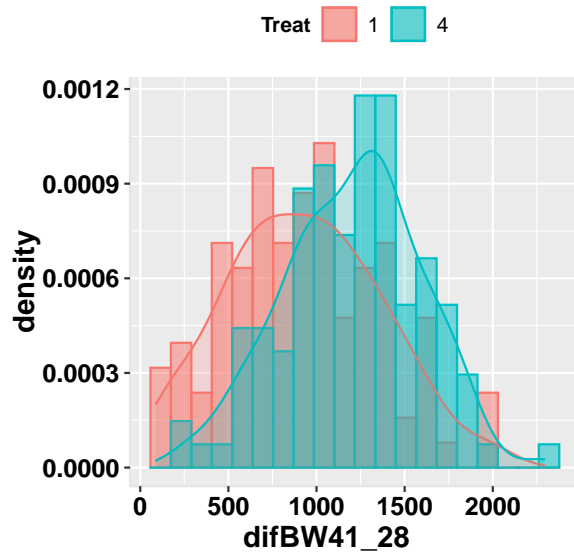
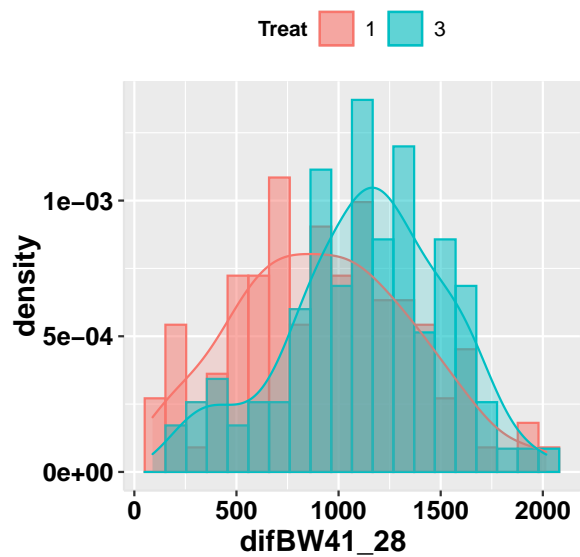
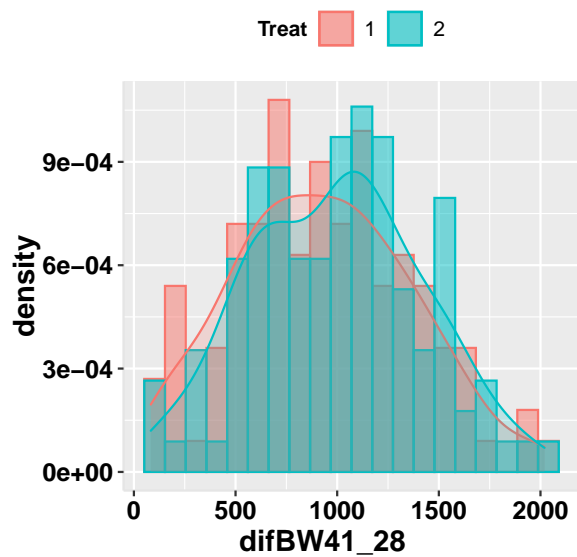
## 2.2 Comparació de tots els tractaments amb el control

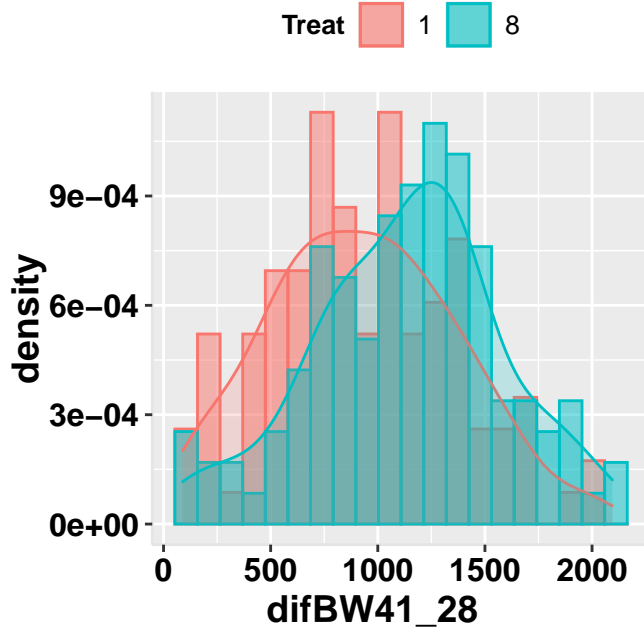
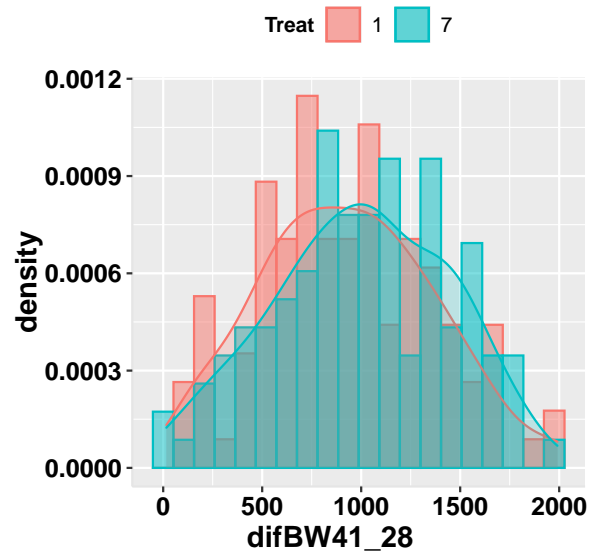
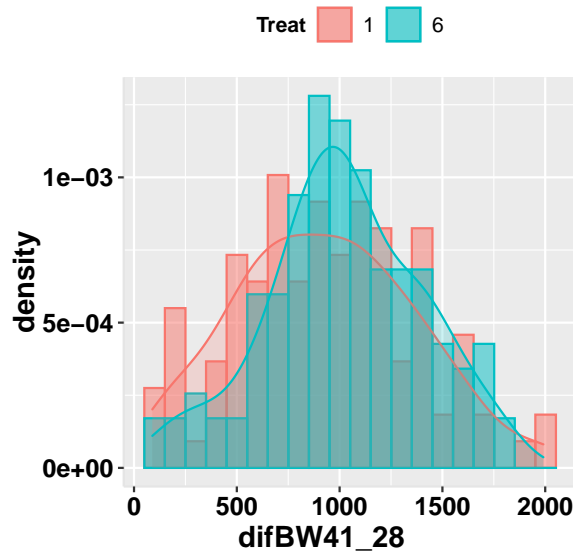
### 2.2.1 Variable BW41





### 2.2.2 Variable difBW41<sub>28</sub>





### 3 Two-Sample Rank Test to detect a shift in a proportion of the "treated" population

Test bi-mostrat per detectar un canvi positiu en una proporció de la població (tractament) comparada a una altra (control).

```
quantileTest(x, y, alternative = "greater", target.quantile = 0.5, target.r  
= NULL, exact.p = TRUE)
```

- **x**: Vector numèric d'observacions del grup tractament.
- **y**: Vector numèric d'observacions del grup control.
- **alternative**: Tipus d'hipòtesi alternativa.
  - "greater": La cua dreta del grup tractament desplaçada cap a la dreta de la cua dreta del grup control.
  - "less": La cua esquerra del grup tractament desplaçada cap a la esquerra de la cua esquerra del grup control.
- **target.quantile**: Quantil utilitzat com a punt de tall inferior per a la prova. A causa de la naturalesa discreta dels quantils empírics, el límit superior dels possibles quantils empírics sovint difereix del valor de target.quantile.

$H_1$  : La porció  $\epsilon$  de la distribució per al grup de tractament (la distribució de  $X$ ) es desplaça cap a la dreta de la distribució per al grup de referència (la distribució de  $Y$ ).

#### 3.1 Resultats del test comparant Tractament vs. control

Apliquem el test al quantil 0.2, 0.3, 0.4 i 0.75. De manera que per exemple per al quantil 0.2 un 20% de la població és menor o igual al valor del quantil. I a la variable diferència dels pesos individuals a la setmana 28 i a la setmana 41.

Veiem un exemple de la sortida de RStudio i una taula resumint els resultats per a a tots els tractaments respecte el control.

<b>Treat</b>	$q_{20}$	<b>P-value</b>	$q_{30}$	<b>P-value</b>	$q_{40}$	<b>P-value</b>	$q_{75}$	<b>P-value</b>
1 2	503 573.8	0.32064	667 693.8	0.32084	772 820.8	0.26627	1221 1266.5	0.26858
1 3	503 819	0.00325	667 936.2	8e-05	772 1060.6	0.00064	1221 1381.5	0.05803
1 4	503 883.6	4e-05	667 1011.6	0	772 1125.4	2e-05	1221 1435	0.00217
1 5	503 865.8	9e-05	667 987.2	2e-05	772 1132	3e-05	1221 1419.75	0.03003
1 6	503 745.2	0.00555	667 849.4	0.00318	772 932.8	0.02565	1221 1300	0.24579
1 7	503 609	0.29542	667 779.7	0.11261	772 865.6	0.1908	1221 1357.25	0.06462
1 8	503 766	0.01499	667 885.5	0.00143	772 1045	0.0033	1221 1405	0.03461

**Taula 1:** Taula d'anàlisi de quantiles per a la variable guany de pes  $BW_{41_{28}}$



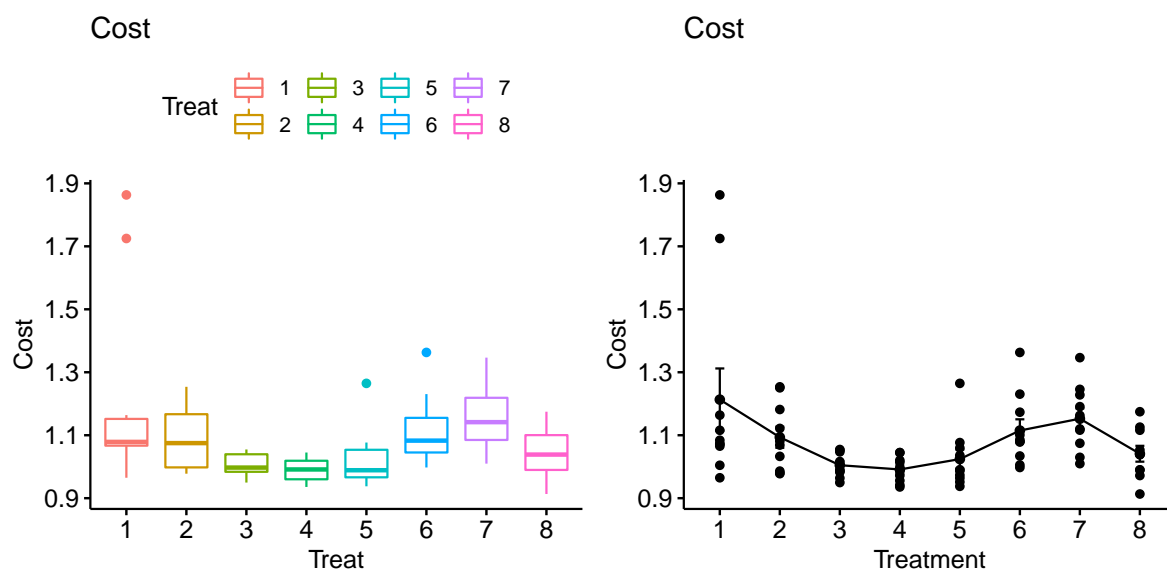
## 4 Costos

Usarem la variable `Coste_per` PV del fitxer `CalculoIndices.xlsx` modificat.

La variable s'ha calculat de la següent forma:

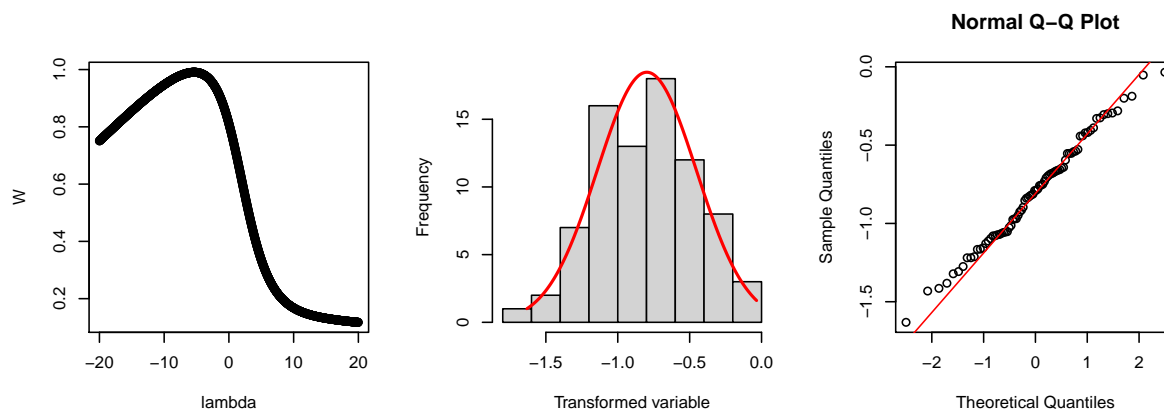
$$\text{Cost/kg PV (€/kg)} = (\text{cost pinso [€/kg]} * \text{consum pinso [kg]}) / \text{guany pes [kg]}$$

### 4.1 Exploració de la variable Cost

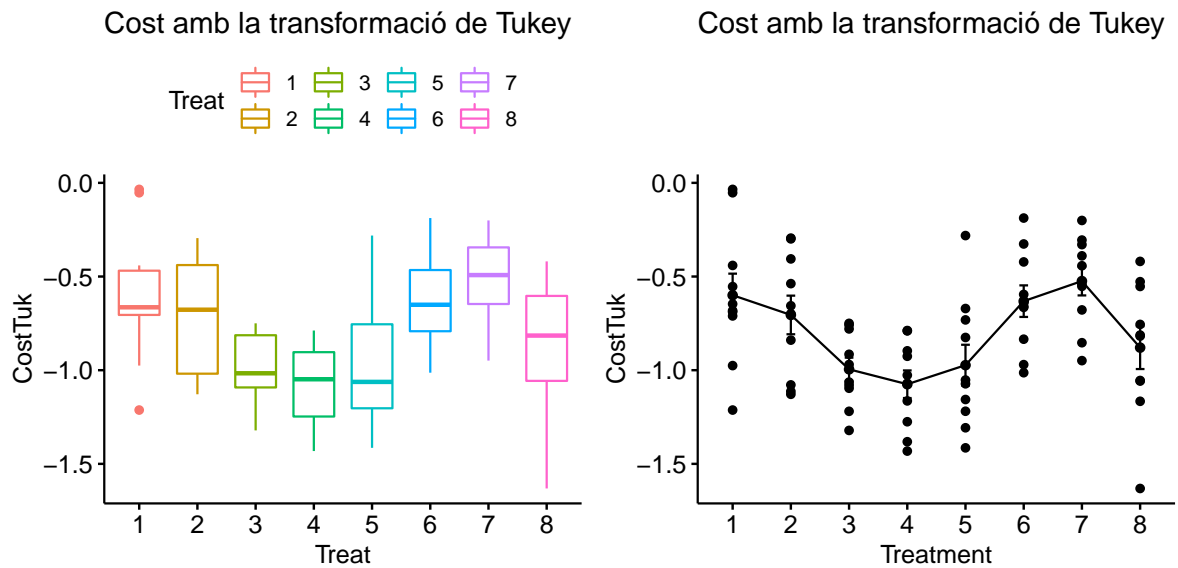


Per tal de millorar la normalitat del costos apliquem una transformació de Tukey de la llibreria `rcompanion`:

```
##  
##      lambda      W Shapiro.p.value  
## 585    -5.4 0.9906          0.8288  
##  
## if (lambda > 0){TRANS = x ^ lambda}  
## if (lambda == 0){TRANS = log(x)}  
## if (lambda < 0){TRANS = -1 * x ^ lambda}
```



El programa selecciona una lambda de  $-5.4$  de manera que com  $\lambda < 0$  la transformació aplicada als costos serà  $costTuk = -1 * Cost^\lambda$



## 4.2 Anàlisi dels resultats

Amb les dades normalitzades ajustem un model ANOVA i realitzem els següents tests:

```
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Treat      7  3.035   0.4336    4.881 0.000152 ***
## Residuals  72  6.397   0.0888
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**Tukey multiple pairwise-comparisons:** Crea un conjunt d'interval de confiança per les diferències de les mitjanes per als diferents tractaments.

```
##          diff      lwr      upr    p adj
## 2-1 -0.10534095 -0.5214742  0.31079227 0.99313670
## 3-1 -0.39553259 -0.8116658  0.02060063 0.07431879
## 4-1 -0.47522197 -0.8913552 -0.05908875 0.01427889
## 5-1 -0.37332879 -0.7894620  0.04280442 0.11097158
## 6-1 -0.03179923 -0.4479324  0.38433399 0.99999757
## 7-1  0.07552814 -0.3406051  0.49166136 0.99915669
## 8-1 -0.27980476 -0.6959380  0.13632846 0.42544395
```

Si observem el p-valor ajustat per comparacions múltiples observem que els tractaments més significatius respecte del control són els tractaments 4,3 i 5, en ordre de significació.

**Pairwise t-test:** Calcula comparacions per parelles entre els diferents tractaments, primer sense i després amb correccions per a proves múltiples. La primera taula conté el p-value usuals del t-test i la de sota conté el p-value ajustat pel mètode de Benjamin & Hockberg (BH), un p-valor que ens indica els False Discovery Rate (FDR). Se solen considerar significatives totes les diferències amb un adj-p-values menor que 0.1.

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data:  costos$CostTuk and costos$Treat
##
##      1      2      3      4      5      6      7
## 2 0.43197 -      -      -      -      -      -
## 3 0.00408 0.03276 -      -      -      -      -
## 4 0.00065 0.00703 0.55183 -      -      -      -
## 5 0.00654 0.04813 0.86817 0.44713 -      -      -
## 6 0.81213 0.58286 0.00798 0.00139 0.01250 -      -
## 7 0.57274 0.17906 0.00072 9.6e-05 0.00122 0.42338 -
## 8 0.03932 0.19476 0.38818 0.14700 0.48518 0.06690 0.00948
##
## P value adjustment method: none
##
## Pairwise comparisons using t tests with pooled SD
##
## data:  costos$CostTuk and costos$Treat
##
##      1      2      3      4      5      6      7
## 2 0.5691 -      -      -      -      -      -
## 3 0.0190 0.0764 -      -      -      -      -
## 4 0.0067 0.0246 0.6277 -      -      -      -
## 5 0.0246 0.0963 0.8682 0.5691 -      -      -
## 6 0.8422 0.6277 0.0248 0.0078 0.0318 -      -
## 7 0.6277 0.2949 0.0067 0.0027 0.0078 0.5691 -
## 8 0.0847 0.3030 0.5691 0.2572 0.5907 0.1249 0.0265
##
## P value adjustment method: BH
```

Observem les comparacions per parelles de tractaments sense cap tipus de correcció (p-valors més petits) i amb la correcció de Benjamin & Hochberg que controla el rati de falsos positius a l'hora de fer múltiples comparacions.

Cal destacar que les diferències en els p-valors per a les comparacions dos a dos es deuen a que el t.test i les comparacions de Tukey utilitzen metodologies diferents per a obtenir el p-valor.

Els resultats de la correcció de Tukey i de BH no coincideixen perquè la correcció de Tukey és més conservadora. Ens inclinem per presentar aquesta segona opció: p-value sense ajustar i ajustat per BH.

## Referències

- [1] <https://www.rdocumentation.org/packages/ggplot2/versions/3.3.2>
- [2] <https://www.rdocumentation.org/packages/devtools/versions/2.3.2>
- [3] <https://github.com/kassambara/easyGgplot2>
- [4] <https://www.rdocumentation.org/packages/patchwork/versions/1.1.0>
- [5] <https://www.rdocumentation.org/packages/EnvStats/versions/2.4.0>
- [6] <https://www.rdocumentation.org/packages/EnvStats/versions/2.3.1/topics/quantileTest>  
<https://www.jstor.org/stable/2532001?seq=1&cid=pdf-reference>
- [7] <https://www.rdocumentation.org/packages/xtable/versions/1.8-4>
- [8] <https://www.rdocumentation.org/packages/rcompanion/versions/2.3.26>
- [9] <https://www.rdocumentation.org/packages/ggpubr/versions/0.4.0>
- [10] <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/aov>
- [11] <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/TukeyHSD>
- [12] <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/pairwise.t.test>
- [13] <https://www.rdocumentation.org/packages/lawstat/versions/3.2/topics/levene.test>