

# Example report with iris data

*Gilles San Martin*

*27 November 2015*

## Contents

<b>Data used</b>	<b>2</b>
<b>Graphical representation of the data</b>	<b>3</b>
Scatterplot matrix . . . . .	3
Boxplot . . . . .	3
<b>Anova</b>	<b>3</b>
<b>Maths</b>	<b>4</b>
<b>References</b>	<b>5</b>
<b>Session Info</b>	<b>6</b>

---

## Data used

Dataset for 3 *Iris* species (*Iris virginica*, *Iris setosa*, *Iris versicolor*) with 4 morphological variables measured :

- Sepal length
- Sepal width
- Petal length
- Petal width

```
data(iris)
d <- iris
```

Random sample :

```
pander(d[sample(1:nrow(d), 5),])
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
<b>49</b>	5.3	3.7	1.5	0.2	setosa
<b>59</b>	6.6	2.9	4.6	1.3	versicolor
<b>100</b>	5.7	2.8	4.1	1.3	versicolor
<b>7</b>	4.6	3.4	1.4	0.3	setosa
<b>18</b>	5.1	3.5	1.4	0.3	setosa

Summary :

```
summary(d)
```

```
##   Sepal.Length   Sepal.Width   Petal.Length   Petal.Width   Species
##   Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa   :50
##   1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   versicolor:50
##   Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50
##   Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##   3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##   Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
```

```
str(d)
```

```
## 'data.frame':   150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

You can use inline code (i.e. calculations inside the text for which only the output is displayed in the final document.  
For example :

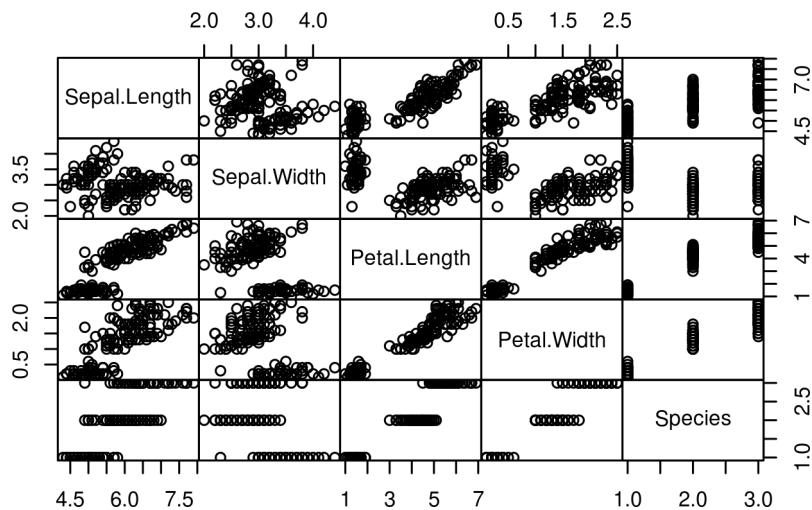
The mean sepal length for all species is 5.8433333 cm.

## Graphical representation of the data

### Scatterplot matrix

```
# dev.new(width = 12/2.54, height = 8/2.54)
```

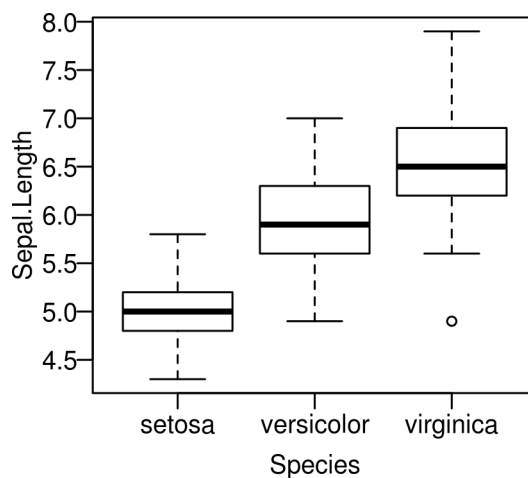
```
pairs(d, gap = 0)
```



### Boxplot

```
# dev.new(width = 7.5/2.54, height = 7/2.54)
```

```
par(mar = c(3,3,2,1), mgp = c(1.8, 0.5, 0), las = 1, cex = 0.8)  
plot(Sepal.Length ~ Species, data=d)
```



### Anova

```
m <- lm(Sepal.Length ~ Species, data = d)
Anova(m)
```

```
## Anova Table (Type II tests)
##
## Response: Sepal.Length
##           Sum Sq Df F value    Pr(>F)
## Species    63.212  2  119.26 < 2.2e-16 ***
## Residuals  38.956 147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

You can also print the results of the Anova as a table

```
pander(Anova(m))
```

Table 2: Anova Table (Type II tests)

	Sum Sq	Df	F value	Pr(>F)
<b>Species</b>	63.21	2	119.3	1.67e-31
<b>Residuals</b>	38.96	147	NA	NA

## Maths

The arithmetic mean is :  $\frac{1}{n} \sum_{i=1}^n x_i$

The arithmetic mean is :

$$\frac{1}{n} \sum_{i=1}^n x_i$$

## References

### R packages :

We used lme4 (Bates et al) for mixed models, mulcomp (Hothorn et al) for Post-Hoc tests, ggplot2 (Wickham) for most graphics, car (Fox & Weisberg) for Type II tests in the two stage analysis only (not with the mixed models !) and of course R.

Bates D, Mächler M, Bolker B and Walker S (2015). “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, 67(1), pp. 1-48. .

Fox J and Weisberg S (2011). *An R Companion to Applied Regression*, Second edition. Sage, Thousand Oaks CA. .

Hothorn T, Bretz F and Westfall P (2008). “Simultaneous Inference in General Parametric Models.” *Biometrical Journal*, 50(3), pp. 346-363.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. .

Wickham H (2009). *ggplot2: elegant graphics for data analysis*. Springer New York. ISBN 978-0-387-98140-6, .

### Other References :

Bretz, F., Hothorn, T., Westfall, P., Westfall, P.H., 2010. Multiple Comparisons Using R. CRC Press.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference. Springer.

Fox, J., 2002. An R and S Plus Companion to Applied Regression, 1st ed. Sage Publications, Inc.

Gelman, A., Hill, J., 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

Pinheiro, J.C., Bates, D.M., 2000. Mixed-Effects Models in S and S-PLUS. Springer, New-York.

Venables, W.N., Ripley, B.D., 2003. Modern Applied Statistics with S, 4th ed. Springer.

Zuur, A.F., Ieno, E.N., Smith, G.M., 2007. Analysing ecological data. Springer.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Mixed Effects Models and Extensions in Ecology With R. Springer-Verlag New York Inc.

## Session Info

```
## R version 3.2.2 (2015-08-14)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu precise (12.04.5 LTS)
##
## locale:
##  [1] LC_CTYPE=en_GB.UTF-8      LC_NUMERIC=C              LC_TIME=en_GB.UTF-8
##  [4] LC_COLLATE=en_GB.UTF-8   LC_MONETARY=en_GB.UTF-8  LC_MESSAGES=fr_BE.UTF-8
##  [7] LC_PAPER=fr_BE.UTF-8     LC_NAME=C                LC_ADDRESS=C
## [10] LC_TELEPHONE=C           LC_MEASUREMENT=fr_BE.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] car_2.0-19    lme4_1.1-10  Matrix_1.2-2 ggplot2_1.0.1 pander_0.5.1 knitr_1.10.5
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.11.3      splines_3.2.2    MASS_7.3-44      munsell_0.4      colorspace_1.2-2 lattice_0.20-3
##  [7] minqa_1.2.3      stringr_0.6.2    plyr_1.8          tools_3.2.2      nnet_7.3-11      grid_3.2.2
## [13] gtable_0.1.2     nlme_3.1-122     htmltools_0.2.6  yaml_2.1.13      digest_0.6.3     nloptr_1.0.4
## [19] reshape2_1.2.2   formatR_1.0      codetools_0.2-14 evaluate_0.7      rmarkdown_0.6.1  scales_0.3.0
## [25] proto_0.3-10
```