Introduction to Statistics for Astronomers and Physicists

Section 0: Introduction, Lecture Outline, Course Format

Angus H Wright

2021-03-19

Setup & Course Outline

This course will be taught in 4 parts, each spanning from 2-4 weeks

###Section 1: Data Description, Analysis, and Modelling (Weeks 1-2)

Topics include:

- Types of data
- Point & interval estimation
- Correlation & covariance
- Fundamentals of data exploration/mining
- Introduction to data visualisation

###Section 2: Probability & Decision Making (Weeks 3-5)

Topics include:

- Decision theory
- fundamentals of probability
- statistical distributions and their origins

###Section 3: Bayesian Statistics (Weeks 6-8

Topics include:

- Frequentist & Bayesian statistics
- Bayes theory
- prior specification
- hypothesis testing,

###Section 4: Parameter Simulation, Optimisation, and Analysis (Weeks 9-12)

Topics include:

-
-

Course Philosophy

This course is designed to be a practical introduction to statistics for astronomers and physicists who are starting their research careers and have had little (or perhaps no) previous education in statistics and statistical data analysis. The course (and these lecture notes) are not designed to be a statistics reference text. Rather the material presented here is designed to guide students on a suitable path towards robust data analysis and research.

The course will present many aspects of data analysis that are widely relevant to modern astronomy and physics. We will borrow heavily from standard statistical problems and thought experiments in an effort to convey important points, and elucidate common statistical and logical fallacies. Problems will almost always be explored using a mixture of tools simultaneously: plain English, math, computer code, graphs, and more.

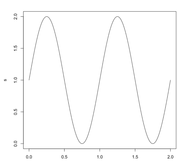
Rmarkdown

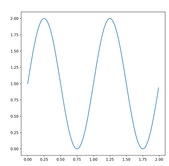
Slides and lecture notes for this course are prepared in **Rmarkdown**, and provided to you after the lectures.

The utility of **Rmarkdown** is that it allows running execution of code chunks alongside markdown-style text, in a wide array of languages. This allows us to present examples in multiple languages easily within one document. For example, if I want to plot a function, I can do so in:

```
#in R
t=seq(0,2,by=0.01)
s=1+sin(2*pi*t)
plot(t,s,type='l')
```

```
#or in python
import numpy as np
import matplotlib.pyplot as plt
t=np.arange(0.,2.,0.01)
s=1+np.sin(2*np.pi*t)
plt.plot(t,s)
plt.show()
```

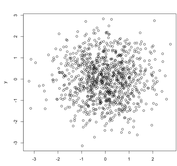


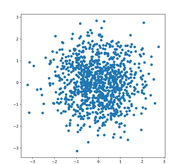


Information generated and stored within blocks is persistent, and code-blocks with different engines can also cross-communicate. This means that, for example, we can:

```
#Create some data in R
#E.g. random draws from N~N(0,1)
x=rnorm(1e5)
y=rnorm(1e5)
plot(x,y)
```

```
#and access it directly in python
plt.scatter(r.x,r.y)
plt.show()
```

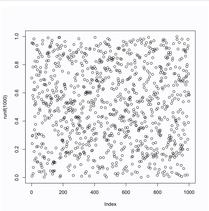




In practice, the vast majority of examples in this course will be programmed in R. However, as you can see from the above, if you can understand one, then you can probably understand both. In this lecture we will go through some examples of R and python code, so that you have an introduction to the important parts (and can follow along without much trouble).

Lecture 0: Introduction to Statistics

```
for (i in 1:10) plot(runif(1000), ylim = c(0, 1)) # for example
```



```
x = 5  # radius of a circle
```

For a circle with the radius 5, its area is 78.5398163.

Why do we need statistics?

```
print(iris)
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 1            5.1         3.5          1.4         0.2     setosa
## 2            4.9         3.0          1.4         0.2     setosa
## 3            4.7         3.2          1.3         0.2     setosa
## 4            4.6         3.1          1.5         0.2     setosa
## 5            5.0         3.6          1.4         0.2     setosa
## 6            5.4         3.9          1.7         0.4     setosa
## 7            4.6         3.4          1.4         0.3     setosa
## 8            5.0         3.4          1.5         0.2     setosa
## 9            4.4         2.9          1.4         0.2     setosa
## 10           4.9         3.1          1.5         0.1     setosa
## 11           5.4         3.7          1.5         0.2     setosa
## 12           4.8         3.4          1.6         0.2     setosa
## 13           4.8         3.0          1.4         0.1     setosa
## 14           4.3         3.0          1.1         0.1     setosa
## 15           5.8         4.0          1.2         0.2     setosa
```

```
print(iris)
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 1            5.1         3.5          1.4         0.2     setosa
## 2            4.9         3.0          1.4         0.2     setosa
## 3            4.7         3.2          1.3         0.2     setosa
## 4            4.6         3.1          1.5         0.2     setosa
## 5            5.0         3.6          1.4         0.2     setosa
## 6            5.4         3.9          1.7         0.4     setosa
## 7            4.6         3.4          1.4         0.3     setosa
## 8            5.0         3.4          1.5         0.2     setosa
## 9            4.4         2.9          1.4         0.2     setosa
## 10           4.9         3.1          1.5         0.1     setosa
## 11           5.4         3.7          1.5         0.2     setosa
## 12           4.8         3.4          1.6         0.2     setosa
## 13           4.8         3.0          1.4         0.1     setosa
## 14           4.3         3.0          1.1         0.1     setosa
## 15           5.8         4.0          1.2         0.2     setosa
```

Slide with Bullets

- Bullet 1
- Bullet 2
- Bullet 3

```
cars
```

```
##    speed dist
## 1      4    2
## 2      4   10
## 3      7    4
## 4      7   22
## 5      8   16
## 6      9   10
## 7     10   18
## 8     10   26
## 9     10   34
## 10    11   17
## 11    11   28
## 12    12   14
## 13    12   20
## 14    12   24
## 15    12   28
```

- Bullet 1
- Bullet 2
- Bullet 3

```
cars
```

```
##    speed dist
## 1      4    2
## 2      4   10
## 3      7    4
## 4      7   22
## 5      8   16
## 6      9   10
## 7     10   18
## 8     10   26
## 9     10   34
## 10    11   17
## 11    11   28
## 12    12   14
## 13    12   20
## 14    12   24
## 15    12   28
```

Slide with R Output

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Slide with Plot

```
plot(pressure)
```