

R learning notes

Rob Hayward

April 14, 2014

1 Power of tests

This comes from [Error Statistics Philosophy](#). The statistical significance of a test does not mean much without some information on the power of the test. For example, if the smoke alarm is very sensitive, the alarm going off is not a strong indication of a fire; if the alarm is lacking sensitivity, the alarm is a very good signal.

$H_0 : \mu \leq 1$, against $H_1 : \mu > 0$;

Let $\sigma = 1$, $n = 25$, so, $(\sigma/\sqrt{n}) = 0.2$

2 Using dates and time

This comes from [Bonnie Dixon](#). This is an overview of using dates and times.

```
dt1 <- as.Date("2014-02-15")
dt1

## [1] "2014-02-15"

dt2 <- as.Date("04/20/2011", format = "%m/%d/%Y")
dt2

## [1] "2011-04-20"

dt1 - dt2

## Time difference of 1032 days

dt2 + 10

## [1] "2011-04-30"
```

Create a vector of dates and find the difference between them.

```
three.dates <- as.Date(c("2010-07-22", "2011-04-20", "2012-06-10"))
three.dates

## [1] "2010-07-22" "2011-04-20" "2012-06-10"

diff(three.dates)

## Time differences in days
## [1] 272 417
```

Create a sequence of days

```
six.weeks <- seq(dt1, length = 6, by = "week")
six.weeks

## [1] "2014-02-15" "2014-02-22" "2014-03-01" "2014-03-08" "2014-03-15"
## [6] "2014-03-22"

six.weeks <- seq(dt1, length = 6, by = 14)
six.weeks

## [1] "2014-02-15" "2014-03-01" "2014-03-15" "2014-03-29" "2014-04-12"
## [6] "2014-04-26"

six.weeks <- seq(dt1, length = 6, by = "2 weeks")
six.weeks

## [1] "2014-02-15" "2014-03-01" "2014-03-15" "2014-03-29" "2014-04-12"
## [6] "2014-04-26"
```

```
unclass(dt1)

## [1] 16116

dt1

## [1] "2014-02-15"
```

2.1 POSIXct

This is for the use of times.

```
tm1 <- as.POSIXct("2009-07-24 23:55:26")
tm1

## [1] "2009-07-24 23:55:26 BST"

tm2 <- as.POSIXct("25072013 08:32:07", format = "%d%m%Y %H:%M:%S")
tm2

## [1] "2013-07-25 08:32:07 BST"
```

Specify the time zone

```
tm3 <- as.POSIXct("2010-12-01 11:42:03", tz = "GMT")
tm3

## [1] "2010-12-01 11:42:03 GMT"
```

Some calculations with times.

```
tm3 > tm2

## [1] FALSE

tm1 + 30

## [1] "2009-07-24 23:55:56 BST"

tm1 - 20

## [1] "2009-07-24 23:55:06 BST"

tm1 - tm2

## Time difference of -1461 days

Sys.time()

## [1] "2014-04-14 13:12:29 BST"

difftime(tm1, as.POSIXct("1970-01-01 00:00:00", tz = "UTC", units = "secs"))

## Time difference of 14450 days
```

2.2 POSIXlt

The 'ct' stands for *calendar time* while 'lt' stands for *local time*.

```
tm1.lt <- as.POSIXlt("2013-07-24 23:55:26")
tm1.lt

## [1] "2013-07-24 23:55:26"

unclass(tm1.lt)

## $sec
## [1] 26
##
## $min
## [1] 55
##
## $hour
## [1] 23
##
## $mday
## [1] 24
##
## $mon
## [1] 6
##
## $year
## [1] 113
##
## $yday
## [1] 204
##
## $isdst
## [1] 1

unlist(tm1.lt)

##      sec      min      hour      mday      mon      year      wday      yday      isdst
##      26      55      23      24      6      113      3      204      1
```

The components of the time object can be extracted.

```
tm1.lt$sec  
## [1] 26  
tm1.lt$yday  
## [1] 3
```

Truncate or round off the time.

```
trunc(tm1.lt, "days")  
## [1] "2013-07-24"  
trunc(tm1.lt, "mins")  
## [1] "2013-07-24 23:55:00"
```

There is information on the lubridate package in the document referenced.

3 Regular Expressions

This is from the following site [Robin Lovelace](#). The basic regex in R is grep.

```
x <- c("16_24cat", "25_34cat", "35_44catch", "45_54Cat", "55-104fat")  
grep(pattern = "cat", x = x)  
## [1] 1 2 3
```

Note the way that any case of the pattern is identified but it must be the correct case.

```
x <- c("16_24cat", "25_34cat", "35_44catch", "45_54Cat", "55-104fat")  
grep(pattern = "cat", x = x, ignore.case = T)  
## [1] 1 2 3 4
```

The function grepl is the same as grep but the output is a logical value.
The function strsplit will split a string. In this case around the underscore.

```
strsplit(x, split = "_")

## [[1]]
## [1] "16"      "24cat"
##
## [[2]]
## [1] "25"      "34cat"
##
## [[3]]
## [1] "35"      "44catch"
##
## [[4]]
## [1] "45"      "54Cat"
##
## [[5]]
## [1] "55-104fat"
```

To return the second element,

```
library(stringr)
str_split_fixed(x, "_", 2)[, 2]

## [1] "24cat" "34cat" "44catch" "54Cat" ""
```

For finding and replacing, use `sub` for the first case and `gsub` for all cases.

```
gsub(pattern = "cat$", replacement = "fat", x = x, ignore.case = T)

## [1] "16_24fat" "25_34fat" "35_44catch" "45_54fat" "55-104fat"
```

To insert an additional letter in all cases where there are two numbers followed by a letter, but only when the first number is 3 or less,

```
gsub("([1-3][1-9][a-z])", "m\\1", x, perl = T, )

## [1] "16_m24cat" "25_m34cat" "35_44catch" "45_54Cat" "55-104fat"
```