

# Summary statistics

Joao Marreiros

2021-04-22 14:06:44

---

## Brief description of the script

This R markdown document computes the output data of the resulting CSV file from the computing ISO 25178-2 parameters in ConfoMap. These data is part of the manuscript: *Dubreuil et al. A 'family of wear': Exploring use-wear patterns on ad hoc burnishing tools*

It computes the following statistics:

- n (sample size = `length`): number of measurements
- smallest value (`min`)
- largest value (`max`)
- mean
- median
- standard deviation (`sd`)

This R project and respective scripts follow the procedures described by Marwick et al. 2017.

The authors would like to thank Ivan Calandra and Lisa Schunk for their help and contribution on several chunks of code included here in the script (pieces of code are also adapted from Calandra et al. 2019, Pedergrana et al. 2020a, 2020b).

To compile this markdown document do not delete or move files from their original folders.

For any questions, comments and inputs, please contact:

Joao Marreiros, marreiros@rgzm.de

```
dir_in <- "analysis/derived_data/"
dir_out <- "analysis/summary_stats/"
```

Raw data must be located in `~/analysis/derived_data/`.

Formatted data will be saved in `~/analysis/summary_stats/`. The knit directory for this script is the project directory. —

## Load packages

```
library(openxlsx)
library(R.utils)
library(tools)
library(doBy)
```

## Get names, path and information of all files

```
data_file <- list.files(dir_in, pattern = "\\\\.Rbin$", full.names = TRUE)
md5_in <- md5sum(data_file)
info_in <- data.frame(file = basename(names(md5_in)), checksum = md5_in,
                      row.names = NULL)
```

The checksum (MD5 hashes) of the imported files are:

	file	checksum
1	MSTRdata.Rbin	422ee34326fad0785833a2fb4d402bab

## Load data into R object

```
imp_data <- loadObject(data_file)
str(imp_data)
```

```
'data.frame':  30 obs. of  53 variables:
 $ Sample.ID      : chr  "Kremasti4" "Kremasti4" "Kremasti4" "Kremasti4" ...
 $ Microscope     : chr  "LSM" "LSM" "LSM" "LSM" ...
 $ Objective      : chr  "50x" "50x" "50x" "50x" ...
 $ PolishType     : chr  "natural" "natural" "natural" "natural" ...
 $ Surface        : chr  "a" "b" "c" "d" ...
 $ Topo           : chr  "Topo" "Topo" "Topo" "Topo" ...
 $ Acquisition.Date : chr  "2021/04/15" "2021/04/15" "2021/04/15" "2021/04/15" ...
 $ Analysis.Date  : chr  "14:06:02" "14:08:39" "14:11:12" "14:13:40" ...
 $ Analysis.Time   : chr  "4/15/2021 10:24:41 AM" "4/15/2021 10:51:24 AM" "4/15/2021 11:45:47 AM" ...
 $ Axis.length.X   : num  255 255 255 255 255 ...
 $ Axis.size.X     : num  3000 3000 3000 3000 3000 3000 3000 3000 3000 3000 ...
 $ Axis.spacing.X  : num  85.2 85.2 85.2 85.2 85.2 ...
 $ Axis.length.Y   : num  255 255 255 255 255 ...
 $ Axis.size.Y     : num  3000 3000 3000 3000 3000 3000 3000 3000 3000 3000 ...
 $ Axis.spacing.Y  : num  85.2 85.2 85.2 85.2 85.2 ...
 $ Axis.length.Z   : num  40.7 49.9 92.6 31.8 29.3 ...
 $ Axis.size.Z     : num  65532 65532 65531 65532 65531 ...
 $ Axis.spacing.Z  : num  0.621 0.761 1.413 0.485 0.447 ...
 $ NM.points.ratio.Z : num  0 0 0 0 0 0 0 0 0 0 ...
 $ Sq             : num  1.58 4.09 1.47 2.24 1.77 ...
 $ Ssk            : num  -0.61 -0.391 -0.274 -0.049 -0.929 ...
```

```

$ Sku          : num  4.88 2.53 6.23 3.44 5.87 ...
$ Sp           : num  5.47 10.44 5.15 9.02 7.91 ...
$ Sv           : num  6.86 12.48 8.2 7.06 8.52 ...
$ Sz           : num  12.3 22.9 13.4 16.1 16.4 ...
$ Sa           : num  1.13 3.4 1.08 1.77 1.25 ...
$ Smr          : num  0.484 0.239 0.604 0.207 0.126 ...
$ Smc          : num  1.69 4.56 1.72 2.87 1.66 ...
$ Sxp          : num  3.99 9.02 2.67 4.65 4.7 ...
$ Sal          : num  19.5 32.4 23.5 30.6 20.7 ...
$ Str          : num  0.48 NA NA 0.614 0.813 ...
$ Std          : num  42.2 93.2 33 25.3 62 ...
$ Sdq          : num  0.383 0.658 0.521 0.403 0.403 ...
$ Sdr          : num  6 15.94 8.53 6.73 6.56 ...
$ Vm           : num  0.0895 0.1529 0.1057 0.0994 0.0945 ...
$ Vv           : num  1.78 4.72 1.82 2.97 1.75 ...
$ Vmp          : num  0.0895 0.1529 0.1057 0.0994 0.0945 ...
$ Vmc          : num  1.13 4.14 1.09 1.84 1.28 ...
$ Vvc          : num  1.5 4.24 1.65 2.68 1.41 ...
$ Vvv          : num  0.287 0.478 0.17 0.292 0.343 ...
$ Maximum.depth.of.furrows: num  7.22 10.88 9.43 6.76 8.68 ...
$ Mean.depth.of.furrows   : num  1.57 3.13 1.47 1.9 1.54 ...
$ Mean.density.of.furrows : num  3750 3056 4011 3480 3423 ...
$ First.direction        : num  89.9772 90.014 45.0229 0.0123 44.9941 ...
$ Second.direction       : num  45 135 180 26.5 63.5 ...
$ Third.direction        : num  180 45 33.7 90 90 ...
$ Texture.isotropy       : num  74 82.7 77.8 90.3 92.3 ...
$ epLsar                 : num  NA NA NA NA NA NA NA NA NA NA ...
$ NewEplsar              : num  NA NA NA NA NA NA NA NA NA NA ...
$ Asfc                   : num  9.93 25.92 17.49 11.21 10.47 ...
$ Smfc                   : num  6281985 10723090 4628049 7318909 11574299 ...
$ HAsfc9                 : num  0.539 0.39 1.927 0.603 0.546 ...
$ HAsfc81                : num  0.87 0.638 2.369 0.728 0.848 ...

```

The imported file is: “~/analysis/derived\_data//MSTRdata.Rbin”

---

## Define numeric variables

```
num.var <- 21:length(imp_data)
```

The following variables will be used:

---

## Compute summary statistics

Create function to compute the statistics at once

```
nminmaxmeanmedsd <- function(x){  
  y <- x[!is.na(x)]  
  n_test <- length(y)  
  min_test <- min(y)  
  max_test <- max(y)  
  mean_test <- mean(y)  
  med_test <- median(y)  
  sd_test <- sd(y)  
  out <- c(n_test, min_test, max_test, mean_test, med_test, sd_test)  
  names(out) <- c("n", "min", "max", "mean", "median", "sd")  
  return(out)  
}
```

Compute the summary statistics in groups

```
s_it <- summaryBy(~ Sample.ID + Impression.Time,  
  data = imp_data[c("Sample.ID", "PolishType", names(imp_data)[num.var])],  
  FUN = nminmaxmeanmedsd)
```

Warning in min(y): no non-missing arguments to min; returning Inf

Warning in max(y): no non-missing arguments to max; returning -Inf

Warning in min(y): no non-missing arguments to min; returning Inf

Warning in max(y): no non-missing arguments to max; returning -Inf

Warning in min(y): no non-missing arguments to min; returning Inf

Warning in max(y): no non-missing arguments to max; returning -Inf

Warning in min(y): no non-missing arguments to min; returning Inf

Warning in max(y): no non-missing arguments to max; returning -Inf

```
str(s_it)
```

```
'data.frame':  2 obs. of  199 variables:  
 $ Sample.ID           : chr  "Kremasti4" "MudPlaster2015MEGII"  
 $ Ssk.n                : num  15 15  
 $ Ssk.min              : num  -3.1 -1.7  
 $ Ssk.max              : num  -0.049 -0.205  
 $ Ssk.mean             : num  -1.145 -0.605
```

\$ Ssk.median	: num	-0.994 -0.426
\$ Ssk.sd	: num	0.8 0.445
\$ Sku.n	: num	15 15
\$ Sku.min	: num	2.53 2.62
\$ Sku.max	: num	21.2 7.62
\$ Sku.mean	: num	8.62 4.06
\$ Sku.median	: num	7.07 3.42
\$ Sku.sd	: num	5.57 1.61
\$ Sp.n	: num	15 15
\$ Sp.min	: num	0.715 1.59
\$ Sp.max	: num	10.4 10.7
\$ Sp.mean	: num	4.45 4.1
\$ Sp.median	: num	3.61 2.31
\$ Sp.sd	: num	2.93 2.95
\$ Sv.n	: num	15 15
\$ Sv.min	: num	1.55 2.29
\$ Sv.max	: num	12.5 12
\$ Sv.mean	: num	6.4 5.53
\$ Sv.median	: num	7.06 5.13
\$ Sv.sd	: num	2.82 3.41
\$ Sz.n	: num	15 15
\$ Sz.min	: num	2.26 4.11
\$ Sz.max	: num	22.9 22.7
\$ Sz.mean	: num	10.86 9.63
\$ Sz.median	: num	11.1 7
\$ Sz.sd	: num	5.46 6.25
\$ Sa.n	: num	15 15
\$ Sa.min	: num	0.195 0.445
\$ Sa.max	: num	3.4 3.29
\$ Sa.mean	: num	0.94 1.17
\$ Sa.median	: num	0.704 0.669
\$ Sa.sd	: num	0.804 0.908
\$ Smr.n	: num	15 15
\$ Smr.min	: num	0.126 0.183
\$ Smr.max	: num	88 26.7
\$ Smr.mean	: num	8.36 5.14
\$ Smr.median	: num	0.604 1.069
\$ Smr.sd	: num	22.65 7.63
\$ Smc.n	: num	15 15
\$ Smc.min	: num	0.309 0.658
\$ Smc.max	: num	4.56 4.8
\$ Smc.mean	: num	1.38 1.77
\$ Smc.median	: num	0.968 0.905
\$ Smc.sd	: num	1.1 1.38
\$ Sxp.n	: num	15 15
\$ Sxp.min	: num	0.571 1.408
\$ Sxp.max	: num	9.02 9.46
\$ Sxp.mean	: num	3.16 3.47
\$ Sxp.median	: num	2.88 2.52
\$ Sxp.sd	: num	2.14 2.56
\$ Sal.n	: num	15 15
\$ Sal.min	: num	16.2 14.9
\$ Sal.max	: num	32.4 29.2
\$ Sal.mean	: num	21.7 22.2

```

$ Sal.median      : num  20 20.9
$ Sal.sd          : num  5.03 4.44
$ Str.n           : num  13 11
$ Str.min         : num  0.24 0.179
$ Str.max         : num  0.813 0.847
$ Str.mean        : num  0.62 0.511
$ Str.median      : num  0.69 0.551
$ Str.sd          : num  0.17 0.237
$ Std.n           : num  15 15
$ Std.min         : num  25.25 3.25
$ Std.max         : num  148 177
$ Std.mean        : num  60.8 71
$ Std.median      : num  42.2 37.8
$ Std.sd          : num  36.3 64.3
$ Sdq.n           : num  15 15
$ Sdq.min         : num  0.113 0.147
$ Sdq.max         : num  0.658 0.688
$ Sdq.mean        : num  0.31 0.309
$ Sdq.median      : num  0.27 0.178
$ Sdq.sd          : num  0.148 0.204
$ Sdr.n           : num  15 15
$ Sdr.min         : num  0.601 1.04
$ Sdr.max         : num  15.9 16.6
$ Sdr.mean        : num  4.55 5.21
$ Sdr.median      : num  2.85 1.5
$ Sdr.sd          : num  3.95 5.87
$ Vm.n           : num  15 15
$ Vm.min          : num  0.00964 0.02037
$ Vm.max          : num  0.153 0.177
$ Vm.mean         : num  0.0666 0.0571
$ Vm.median       : num  0.059 0.0268
$ Vm.sd           : num  0.0407 0.0491
$ Vv.n           : num  15 15
$ Vv.min          : num  0.319 0.681
$ Vv.max          : num  4.72 4.98
$ Vv.mean         : num  1.44 1.83
$ Vv.median       : num  1.08 0.925
$ Vv.sd           : num  1.13 1.43
$ Vmp.n           : num  15 15
$ Vmp.min         : num  0.00964 0.02037
[1] list output truncated

```

---

## Save data

### Format name of output file

```
file_out <- "MSTRstats"
```

The file will be saved as “~/analysis/summary\_stats/MSTRstats.[ext]”.

## Write to XLSX

```
write.xlsx(list(Sample_ImpTime = s_it), file = paste0(dir_out, file_out, ".xlsx"))
```

## Save R object

```
saveObject(s_it, file = paste0(dir_out, file_out, ".Rbin"))
```

## Show files information

```
files_out <- c(paste0(dir_out, file_out, ".xlsx"), paste0(dir_out, file_out, ".Rbin"))
md5_out <- md5sum(files_out)
info_out <- data.frame(files = basename(names(md5_out)), checksum = md5_out,
                      row.names = NULL)
```

The checksum (MD5 hashes) of the exported files are:

	files	checksum
1	MSTRstats.xlsx	558977db73799a8e5516381b68aa26e9
2	MSTRstats.Rbin	af24588711e837443642be34799e6c0e

---

## sessionInfo() and RStudio version

```
sessionInfo()
```

R version 4.0.4 (2021-02-15)

Platform: x86\_64-apple-darwin17.0 (64-bit)

Running under: macOS Catalina 10.15.7

Matrix products: default

BLAS: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib

LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib

locale:

[1] en\_US.UTF-8/en\_US.UTF-8/en\_US.UTF-8/C/en\_US.UTF-8/en\_US.UTF-8

attached base packages:

[1] tools stats graphics grDevices utils datasets methods

[8] base

other attached packages:

[1] doBy\_4.6.9 R.utils\_2.10.1 R.oo\_1.24.0 R.methodsS3\_1.8.1

[5] openxlsx\_4.2.3

loaded via a namespace (and not attached):

[1] zip_2.1.1	Rcpp_1.0.6	bslib_0.2.4	compiler_4.0.4
[5] pillar_1.6.0	jquerylib_0.1.3	digest_0.6.27	lattice_0.20-41
[9] gtable_0.3.0	jsonlite_1.7.2	evaluate_0.14	lifecycle_1.0.0
[13] tibble_3.1.1	pkgconfig_2.0.3	rlang_0.4.10	Matrix_1.3-2
[17] DBI_1.1.1	yaml_2.2.1	xfun_0.22	dplyr_1.0.5
[21] stringr_1.4.0	knitr_1.32	generics_0.1.0	vctrs_0.3.7
[25] sass_0.3.1	grid_4.0.4	rprojroot_2.0.2	tidyselect_1.1.0
[29] glue_1.4.2	R6_2.5.0	fansi_0.4.2	rmarkdown_2.7
[33] tidyr_1.1.3	ggplot2_3.3.3	purrr_0.3.4	magrittr_2.0.1
[37] backports_1.2.1	MASS_7.3-53.1	scales_1.1.1	htmltools_0.5.1.1
[41] ellipsis_0.3.1	assertthat_0.2.1	colorspace_2.0-0	Deriv_4.1.3
[45] utf8_1.2.1	stringi_1.5.3	munsell_0.5.0	broom_0.7.6
[49] crayon_1.4.1			

---

END OF SCRIPT