

TRGN 527: Applied Data Science and Bioinformatics

UNIT I. Introduction and Basic Data Science

Week 2 - Lecture 1

Enrique I. Velazquez Villarreal, M.D., Ph.D., M.P.H., M.S. | Assistant Professor

Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Leader of the USC Bioinformatics Core – *USC CaRE2 Health Equity Center*

David W. Craig, Ph.D. | Professor and Vice Chair

Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Co-Director, Institute of Translational Genomics

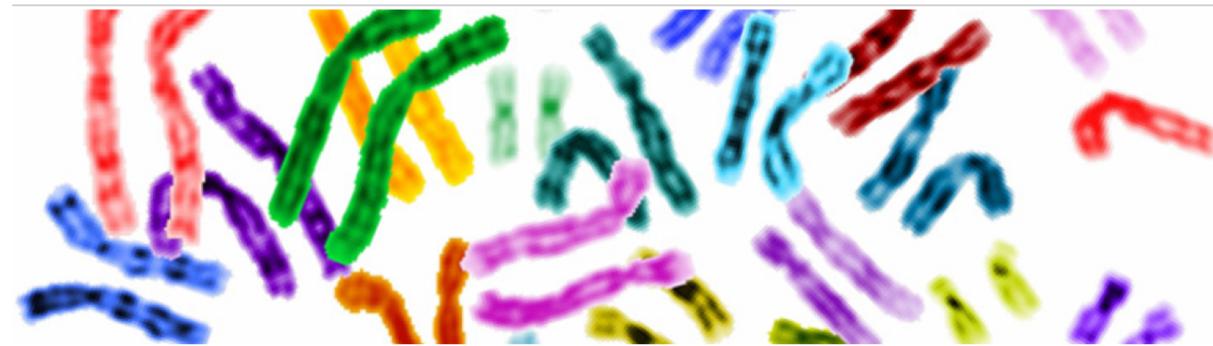
Topics

- General Statistics.



General Statistics - Introduction

- R applications includes statistics or models or graphics.
- In this week we will describe general statistics.
- Definition:
 - Statistics deals with data collection, analysis, interpretation and presentation.
 - In applying statistics in genomics we usually begin with a statistical population or a statistical model process to be studied.
 - Populations can be diverse topics such as "specific genomic regions from people in a clinical trial".



TCGA Data

Visualize copy number variations in the data portal! Details in the User's Guide

portal.gdc.cancer.gov

NATIONAL CANCER INSTITUTE GDC Data Portal Home Projects Exploration Analysis Repository Quick Search Manage Sets Login Cart GDC Apps Dismiss x

TCGA-LUSC / TCGA-39-5040

Add all files to the cart

FILES 61 ANNOTATIONS 0

Summary

Case UUID	eca4bfa2-5ba5-4708-959a-d901c1b61edd
Case ID	TCGA-39-5040
Project	TCGA-LUSC
Project Name	Lung Squamous Cell Carcinoma
Disease Type	Squamous Cell Neoplasms
Program	TCGA
Primary Site	Bronchus and lung
Images	2 (4)

File Counts by Data Category

Data Category	Files (n=61)
Sequencing Reads	6
Transcriptome Profiling	10
Simple Nucleotide Variation	16
Copy Number Variation	5
DNA Methylation	1
Clinical	7
Biospecimen	16

File Counts by Experimental Strategy

Experimental Strategy	Files (n=61)
Diagnostic Slide	1
Tissue Slide	3
WXS	18
RNA-Seq	8
miRNA-Seq	6
Genotyping Array	5
Methylation Array	1

Clinical

Demographic	Diagnoses / Treatments (1)	Family Histories (0)	Exposures (1)
UUID	ff920966-0541-5f06-a330-f257dca31c1c		
Ethnicity	not hispanic or latino		
Gender	male		
Race	white		
Year Of Birth	1949		
Year Of Death	2009		

Clinical Supplement File

Filename	Data format	Size	Action
nationwidechildrens.org_clinical_patient_lusc.txt	BCR Biobab	708106	
nationwidechildrens.org_clinical_pte_lusc.txt	BCR Biobab	8645	
nationwidechildrens.org_clinical_follow_up_v1.0_lusc.txt	BCR Biobab	241707	
nationwidechildrens.org_clinical_onmf_v4.0_lusc.txt	BCR Biobab	36505	
nationwidechildrens.org_clinical_drug_lusc.txt	BCR Biobab	146393	
nationwidechildrens.org_clinical_TCGA-39-5040.xml	BCR XML	83529	
nationwidechildrens.org_clinical_radiation_lusc.txt	BCR Biobab	19389	

Biospecimen

Search

Expand All

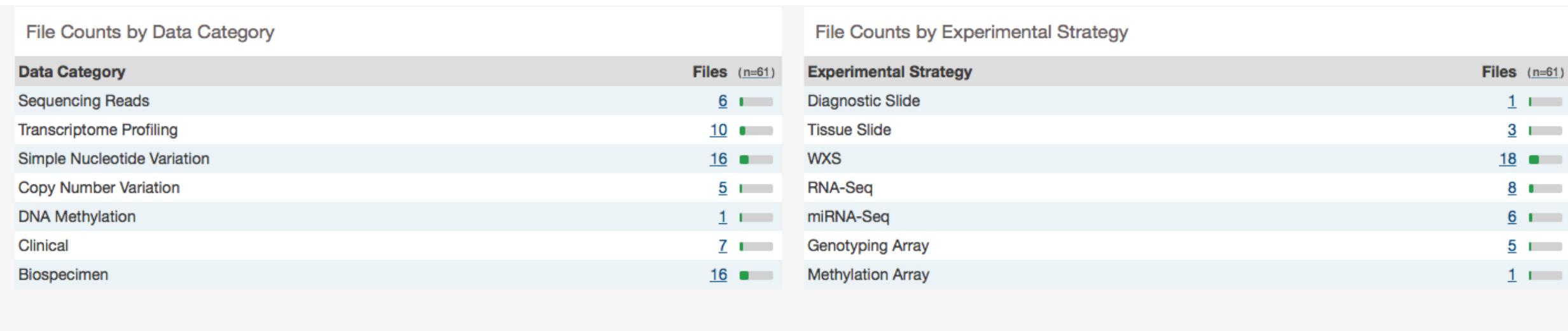
Samples	TCGA-39-5040-01A
Portions	TCGA-39-5040-11A
Protocols	
Sample ID	TCGA-39-5040-01A
Sample UUID	f7803904-c356-42b7-9e42-a2cfed2f89c
Pathology Report Uuid	6be6712-0d27-4032-9f72-3c9ff13901223
Sample Type	Primary Tumor
Sample Type Id	01
Tissue Type	Not Reported

https://portal.gdc.cancer.gov/cases/eca4bfa2-5ba5-4708-959a-d901c1b61edd?ssmsTable_offset=650

TCGA Data

Summary		 Add all files to the cart
Case UUID	eca4bfa2-5ba5-4708-959a-d901c1b61edd	
Case ID	TCGA-39-5040	
Project	TCGA-LUSC	
Project Name	Lung Squamous Cell Carcinoma	
Disease Type	Squamous Cell Neoplasms	
Program	TCGA	
Primary Site	Bronchus and lung	
Images	 (4) 	 FILES 61

TCGA Data



TCGA Data

Clinical				 Download
Demographic	Diagnoses / Treatments (1)	Family Histories (0)	Exposures (1)	
UUID	ff920986-0541-5f06-a330-f257dea31c1c			
Ethnicity	not hispanic or latino			
Gender	male			
Race	white			
Year Of Birth	1949			
Year Of Death	2009			

TCGA Data

Clinical Supplement File

Filename	Data format	Size	Action
 nationwidechildrens.org_clinical_patient_lusc.txt	BCR Biotab	708106	
 nationwidechildrens.org_clinical_nte_lusc.txt	BCR Biotab	8645	
 nationwidechildrens.org_clinical_follow_up_v1.0_lusc.txt	BCR Biotab	241707	
 nationwidechildrens.org_clinical_omf_v4.0_lusc.txt	BCR Biotab	38505	
 nationwidechildrens.org_clinical_drug_lusc.txt	BCR Biotab	146393	
 nationwidechildrens.org_clinical.TCGA-39-5040.xml	BCR XML	83529	
 nationwidechildrens.org_clinical_radiation_lusc.txt	BCR Biotab	19389	

TCGA Data

Biospecimen	
<input type="button" value="Search"/> <input type="button" value="Expand All"/>	
<input checked="" type="checkbox"/> Samples	
<input checked="" type="checkbox"/> TCGA-39-5040-01A	<input type="button" value="Download"/>
<input type="checkbox"/> Portions	
<input checked="" type="checkbox"/> TCGA-39-5040-11A	
<input type="checkbox"/> Portions	
<input checked="" type="checkbox"/> TCGA-39-5040-01Z	
<input type="checkbox"/> Portions	
Sample ID	TCGA-39-5040-01A
Sample UUID	f7803904-c356-42b7-9e42-a2cfded2f89c
Pathology Report Uuid	6bef6712-0d27-4032-9f72-3c9f13901223
Sample Type	Primary Tumor
Sample Type Id	01
Tissue Type	Not Reported
Tumor Code	--
Tumor Code Id	--
Oct Embedded	--
Shortest Dimension	0.4
Intermediate Dimension	1.5
Longest Dimension	1.5
Is Ffpe	false
Tumor Descriptor	--
Current Weight	--
Initial Weight	--
Composition	--
Time Between Clamping And Freezing	--
Time Between Excision And Freezing	--
Days To Sample Procurement	--
Freezing Method	--
Preservation Method	--
Days To Collection	--
Portions	2

TCGA Data

Biospecimen Supplement File

Filename	Data format	Size	Action
nationwidechildrens.org_biospecimen_diagnostic_slides_lusc.txt	BCR Biotab	54898	
TCGA-39-5040-01A-02-BS2.45b0881f-cac9-4963-b404-32ee901aac19.svs	SVS	129902711	
nationwidechildrens.org_biospecimen_shipment_portion_lusc.txt	BCR Biotab	50647	
nationwidechildrens.org_biospecimen_protocol_lusc.txt	BCR Biotab	426686	
TCGA-39-5040-01Z-00-DX1.F314DFF3-AEBE-4171-8F04-5D2A61FDE208.svs	SVS	637389638	
nationwidechildrens.org_biospecimen_sample_lusc.txt	BCR Biotab	479815	
nationwidechildrens.org_ssfs_tumor_samples_lusc.txt	BCR Biotab	167637	
nationwidechildrens.org_biospecimen_analyte_lusc.txt	BCR Biotab	928056	
nationwidechildrens.org_biospecimen_slide_lusc.txt	BCR Biotab	295718	
TCGA-39-5040-11A-01-TS1.3952b5ac-0a12-47d6-a518-c4da23e0e4f0.svs	SVS	145607949	
nationwidechildrens.org_ssfs.TCGA-39-5040.xml	BCR SSF XML	15782	
nationwidechildrens.org_biospecimen_aliquot_lusc.txt	BCR Biotab	1330456	
TCGA-39-5040-01A-01-BS1.6480b0fe-a8b6-46c1-ac00-4052bf427e75.svs	SVS	303811723	
nationwidechildrens.org_ssfs_normal_controls_lusc.txt	BCR Biotab	106992	
nationwidechildrens.org_biospecimen.TCGA-39-5040.xml	BCR XML	73078	
nationwidechildrens.org_biospecimen_portion_lusc.txt	BCR Biotab	180325	

TCGA Data

Biospecimen Supplement File

Filename	Data format	Size	Action
nationwidechildrens.org_biospecimen_diagnostic_slides_lusc.txt	BCR Biotab	54898	
TCGA-39-5040-01A-02-BS2.45b0881f-cac9-4963-b404-32ee901aac19.svs	SVS	129902711	
nationwidechildrens.org_biospecimen_shipment_portion_lusc.txt	BCR Biotab	50647	
nationwidechildrens.org_biospecimen_protocol_lusc.txt	BCR Biotab	426686	
TCGA-39-5040-01Z-00-DX1.F314DFF3-AE8E-4171-8F04-5D2A61FDE208.svs	SVS	637389638	
nationwidechildrens.org_biospecimen_sample_lusc.txt	BCR Biotab	479815	
nationwidechildrens.org_ssfs_tumor_samples_lusc.txt	BCR Biotab	167637	
nationwidechildrens.org_biospecimen_analyte_lusc.txt	BCR Biotab	928056	
nationwidechildrens.org_biospecimen_slide_lusc.txt	BCR Biotab	295718	
TCGA-39-5040-11A-01-TS1.3952b5ac-0a12-47d6-a518-c4da23e0e4f0.svs	SVS	145607949	
nationwidechildrens.org_ssfs.TCGA-39-5040.xml	BCR SSF XML	15782	
nationwidechildrens.org_biospecimen_aliquot_lusc.txt	BCR Biotab	1330456	
TCGA-39-5040-01A-01-BS1.6480b0fe-a8b6-46c1-ac00-4052bf427e75.svs	SVS	303811723	
nationwidechildrens.org_ssfs_normal_controls_lusc.txt	BCR Biotab	106992	
nationwidechildrens.org_biospecimen.TCGA-39-5040.xml	BCR XML	73078	
nationwidechildrens.org_biospecimen_portion_lusc.txt	BCR Biotab	180325	

TCGA Data

Showing 651 - 655 of 655 somatic mutations

☰ JSON TSV Save/Edit Mutation Set

DNA Change	Type	Consequences	# Affected Cases in TCGA-LUSC	# Affected Cases Across the GDC	Impact
chr5:g.140828418C>A	Substitution	Synonymous PCDHA6 I109I	1 / 495  0.20%	1 / 10,202 	Lo -- --
chr1:g.230747482delTTCCA	Deletion	5 Prime UTR CAPN9	1 / 495  0.20%	1 / 10,202 	MR -- --
chr6:g.29723458delG	Deletion	5 Prime UTR HLA-F	1 / 495  0.20%	1 / 10,202 	MR -- --
chr9:g.35802508A>G	Substitution	Synonymous NPR2 R572R	1 / 495  0.20%	1 / 10,202 	Lo -- --
chr22:g.42126667C>A	Substitution	Synonymous CYP2D6 S467S	1 / 495  0.20%	1 / 10,202 	Lo -- --

Show 10 ▾ entries

« < 61 62 63 64 65 66 > »

Computing basic statistics in R

- Mean
 - The sum of the values and dividing with the number of values in a data set.
- Median
 - Middle most value in a data set.
- Mode
 - Is the value that has highest number of occurrences in a data set. It can have both numeric and character data.
- Standard Deviation
 - Quantity calculated to indicate the extent of deviation for a group as a whole.
- Variance
 - Quantity equal to the square of the standard deviation.

Computing basic statistics in R: Standard Deviation

- Formula:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

- 1) Calculate the mean.
- 2) For each number: subtract the Mean and square the result.
- 3) Calculate the mean of those squared differences.
- 4) Calculate the square root of the previous step.

Computing basic statistics in R: Standard Deviation

- Sample: 9, 2, 5, 4, 12, 7, 8, 11, 9, 3, 7, 4, 12, 5, 4, 10, 9, 6, 9, 4
- Calculate the mean:

$$\underline{9 + 2 + 5 + 4 + 12 + 7 + 8 + 11 + 9 + 3 + 7 + 4 + 12 + 5 + 4 + 10 + 9 + 6 + 9 + 4}$$

20

$$\frac{140}{20} = 7$$

Mean = 7

Computing basic statistics in R: Standard Deviation

$$(x_i - \mu)^2$$

$$(9 - 7)^2 = (2)^2 = \mathbf{4}$$

$$(2 - 7)^2 = (-5)^2 = \mathbf{25}$$

$$(5 - 7)^2 = (-2)^2 = \mathbf{4}$$

$$(4 - 7)^2 = (-3)^2 = \mathbf{9}$$

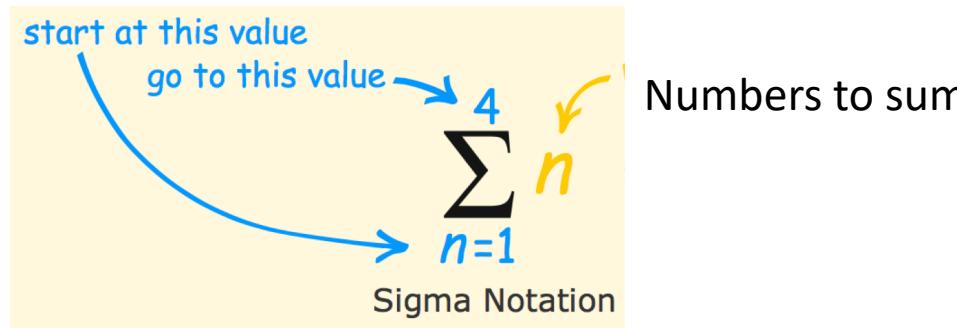
$$(12 - 7)^2 = (5)^2 = \mathbf{25}$$

$$(7 - 7)^2 = (0)^2 = \mathbf{0}$$

$$(8 - 7)^2 = (1)^2 = \mathbf{1}$$

4, 25, 4, 9, 25, 0, 1, 16, 4, 16, 0, 9, 25, 4, 9, 9, 4, 1, 4, 9

Computing basic statistics in R: Standard Deviation



$$\sum_{i=1}^N (x_i - \mu)^2$$

Which means: Sum all values from $(x_1-7)^2$ to $(x_N-7)^2$

We already calculated $(x_1-7)^2=4$ etc. in the previous step, so just sum them up:

$$= 4+25+4+9+25+0+1+16+4+16+0+9+25+4+9+9+4+1+4+9 = \mathbf{178}$$

Computing basic statistics in R: Standard Deviation

$$\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

Mean of squared differences = $(1/20) \times 178 = \mathbf{8.9}$

←This is called VARIANCE

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

$$\sigma = \sqrt{8.9} = \mathbf{2.983...}$$

←This is called Standard Deviation

Computing basic statistics in R

- Problem:
 - Want to calculate basic statistics: mean, median, standard deviation, variance and mode.
- Solution:
 - Use one of the following functions, assuming that x and y are vectors:
 - `mean(x)`
 - `median(x)`
 - `sd(x)`
 - `var(x)`
 - About calculating mode:
 - First create a function:
 - # Creating a function named "getmode"
 - `getmode <- function(x) { uniqx <- unique(x) uniqx[which.max(tabulate(match(x, uniqx)))]}`
 - # Creating a vector with numbers
 - `x <- c(1,1,2,3,4,5,6,7,8,20,22,35)`
 - # Calculating the mode
 - `calculating_mode <- getmode(x)calculating_mode`

Computing basic statistics in R: Exercise

- With the following numbers:
 - 1,1,2,3,4,5,6,7,8,20,22,35
- Calculate the mean, median, standard deviation, variance and mode.

```
8 # Computing basic statistics in R
9 # With the following numbers: 1,1,2,3,4,5,6,7,8,20,22,35
10 # Calculate the mean, median, standard deviation, variance and mode.
11 ````{R}
12 #Create a vector with numbers
13 x <- c(1,1,2,3,4,5,6,7,8,20,22,35)
14 ````

15
16 #Calculating the mean
17 ````{R}
18 mean(x)
19 ````

20
21 #Calculating the median
22 ````{R}
23 median(x)
24 ````

25
26 #Calculating the Standard Deviation
27 ````{R}
28 sd(x)
29 ````

30
31 #Calculating the variance
32 ````{R}
33 var(x)
34 ````

35
36 #Calculating the mode
37 ````{R}
38 # Creating a function named "getmode"
39 getmode <- function(x) {
40   uniqx <- unique(x)
41   uniqx[which.max(tabulate(match(x, uniqx)))]}

42
43 # Creating a vector with numbers
44 x <- c(1,1,2,3,4,5,6,7,8,20,22,35)
45 # Calculating the mode
46
47 calculating_mode <- getmode(x)
48 calculating_mode
```

Computing basic statistics in R: Exercise

- Using the file Rmarkdown_Week_2_Lecture_1.

Rmarkdown_Week_2_Lecture_1

EIVV

12/27/2018

Computing basic statistics in R

With the following numbers: 1,1,2,3,4,5,6,7,8,20,22,35

Calculate the mean, median, standard deviation, variance and mode.

```
#Create a vector with numbers  
x <- c(1,1,2,3,4,5,6,7,8,20,22,35)
```

Calculating the mean

```
mean(x)
```

```
## [1] 9.5
```

Calculating the median

```
median(x)
```

```
## [1] 5.5
```

Calculating the Standard Deviation

```
sd(x)
```

```
## [1] 10.57871
```

Calculating the variance

```
var(x)
```

```
## [1] 111.9091
```

Calculating the mode

```
# Creating a function named "getmode"  
getmode <- function(x) {  
  uniqx <- unique(x)  
  uniqx[which.max(tabulate(match(x, uniqx)))]}
```

```
# Creating a vector with numbers  
x <- c(1,1,2,3,4,5,6,7,8,20,22,35)  
# Calculating the mode
```

```
calculating_mode <- getmode(x)  
calculating_mode
```

```
## [1] 1
```

Computing basic statistics in R using TCGA data

- Upload the file "trgn599.clinical.tsv":
- Calculate the mean, median, standard deviation, variance and mode of the following variables:
 - "weight"
 - "height"
 - "days_to_birth"

```
7 ## Uploading the file "trgn599.clinical.tsv":
8 ````{R}
9 clinical_data <- read.table("trgn599.clinical.tsv", header = TRUE, sep = "\t")
10 ````

11
12
13 ## Calculate the mean, median, standard deviation, variance and mode of the following variables: "weight", "height",
14 ## "days_to_birth"
15 ## Using Weight
16 ````{R}
17 #Create a vector using the variable "weight"
18 x <- clinical_data$weight
19 ````

20
21 #Calculating the mean
22 ````{R}
23 mean(x, na.rm = TRUE)
24 ````

[1] 2.673331

25
26 #Calculating the median
27 ````{R}
28 median(x, na.rm = TRUE)
29 ````

[1] 2.465753

30
31 #Calculating the Standard Deviation
32 ````{R}
33 sd(x, na.rm = TRUE)
34 ````

[1] 1.730405

35
36 #Calculating the variance
37 ````{R}
38 var(x, na.rm = TRUE)
39 ````
```

Computing basic statistics in R using TCGA data

```
41 #Calculating the mode
42 ````{R}
43 # Creating a function named "getmode"
44 getmode <- function(x) {
45   uniqx <- unique(x)
46   uniqx[which.max(tabulate(match(x, uniqx)))]}
47
48 #Create a vector using the variable "weight"
49 x <- clinical_data$weight
50 x <- x[!is.na(x)]
51
52 # Calculating the mode
53 calculating_mode <- getmode(x)
54 calculating_mode
55 ````

56
57
58 ## Using Height
59 ````{R}
60 #Create a vector using the variable "height"
61 x <- clinical_data$height
62 ````

63
64 #Calculating the mean
65 ````{R}
66 mean(x, na.rm = TRUE)
67 ````

[1] 37.57714

68
69 #Calculating the median
70 ````{R}
71 median(x, na.rm = TRUE)
72 ````

[1] 40

73
74 #Calculating the Standard Deviation
75 ````{R}
76 sd(x, na.rm = TRUE)
77 ````

[1] 13.2857
```

```
79 #Calculating the variance
80 ````{R}
81 var(x, na.rm = TRUE)
82 ````

[1] 176.5098

83
84 #Calculating the mode
85 ````{R}
86 # Creating a function named "getmode"
87 getmode <- function(x) {
88   uniqx <- unique(x)
89   uniqx[which.max(tabulate(match(x, uniqx)))]}
90
91 #Create a vector using the variable "height"
92 x <- clinical_data$height
93 x <- x[!is.na(x)]
94
95 # Calculating the mode
96 calculating_mode <- getmode(x)
97 calculating_mode
98 ````

99
100
101 ## Using "days_to_birth"
102 ````{R}
103 #Create a vector using the variable "days_to_birth"
104 x <- clinical_data$days_to_birth
105 ````

106
107 #Calculating the mean
108 ````{R}
109 mean(x, na.rm = TRUE)
110 ````

[1] -24648.4

111
112 #Calculating the median
113 ````{R}
114 median(x, na.rm = TRUE)
115 ````

[1] -25102
```

```
117 #Calculating the Standard Deviation
118 ````{R}
119 sd(x, na.rm = TRUE)
120 ````

[1] 3598.204

121
122 #Calculating the variance
123 ````{R}
124 var(x, na.rm = TRUE)
125 ````

[1] 12947070

126
127 #Calculating the mode
128 ````{R}
129 # Creating a function named "getmode"
130 getmode <- function(x) {
131   uniqx <- unique(x)
132   uniqx[which.max(tabulate(match(x, uniqx)))]}
133
134 #Create a vector using the variable "days_to_birth"
135 x <- clinical_data$days_to_birth
136 x <- x[!is.na(x)]
137
138 # Calculating the mode
139 calculating_mode <- getmode(x)
140 calculating_mode
141 ````

[1] -18856
```

Computing basic statistics in R: Exercise

- Using the file Rmarkdown_Week_2_Lecture_1_TCGA.

Rmarkdown_Week_2_Lecture_1_TCGA

EIV

Uploading the file “trgn599.clinical.tsv”:

```
clinical_data <- read.table("trgn599.clinical.tsv", header = TRUE, sep = "\t")
```

Calculate the mean, median, standard deviation, variance and mode of the following variables: “weight”, “height”, “days_to_birth”

Using Weight

```
#Create a vector using the variable "weight"  
x <- clinical_data$weight
```

Calculating the mean

```
mean(x, na.rm = TRUE)
```

```
## [1] 2.673331
```

Calculating the median

```
median(x, na.rm = TRUE)
```

```
## [1] 2.465753
```

Calculating the Standard Deviation

```
sd(x, na.rm = TRUE)
```

```
## [1] 1.730405
```

Computing basic statistics in R: Exercise

- Using the file Rmarkdown_Week_2_Lecture_1_TCGA.

Calculating the variance

```
var(x, na.rm = TRUE)  
  
## [1] 2.994302
```

Calculating the mode

```
# Creating a function named "getmode"  
getmode <- function(x) {  
  uniqx <- unique(x)  
  uniqx[which.max(tabulate(match(x, uniqx)))]}  
  
#Create a vector using the variable "weight"  
x <- clinical_data$weight  
x <- x[!is.na(x)]  
  
# Calculating the mode  
calculating_mode <- getmode(x)  
calculating_mode  
  
## [1] 2.739726
```

Using Height

```
#Create a vector using the variable "height"  
x <- clinical_data$height
```

Calculating the mean

```
mean(x, na.rm = TRUE)  
  
## [1] 37.57714
```

Calculating the median

```
median(x, na.rm = TRUE)  
  
## [1] 40
```

Computing basic statistics in R: Exercise

- Using the file Rmarkdown_Week_2_Lecture_1_TCGA.

Calculating the Standard Deviation

```
sd(x, na.rm = TRUE)  
  
## [1] 13.2857
```

Calculating the variance

```
var(x, na.rm = TRUE)  
  
## [1] 176.5098
```

Calculating the mode

```
# Creating a function named "getmode"  
getmode <- function(x) {  
  uniqx <- unique(x)  
  uniqx[which.max(tabulate(match(x, uniqx)))]}  
  
#Create a vector using the variable "height"  
x <- clinical_data$height  
x <- x[!is.na(x)]  
  
# Calculating the mode  
calculating_mode <- getmode(x)  
calculating_mode
```



```
## [1] 50
```

Using “days_to_birth”

```
#Create a vector using the variable "days_to_birth"  
x <- clinical_data$days_to_birth
```

Calculating the mean

```
mean(x, na.rm = TRUE)  
  
## [1] -24648.4
```

Computing basic statistics in R: Exercise

- Using the file Rmarkdown_Week_2_Lecture_1_TCGA.

Calculating the median

```
median(x, na.rm = TRUE)
```

```
## [1] -25102
```

Calculating the Standard Deviation

```
sd(x, na.rm = TRUE)
```

```
## [1] 3598.204
```

Calculating the variance

```
var(x, na.rm = TRUE)
```

```
## [1] 12947070
```

Calculating the mode

```
# Creating a function named "getmode"
getmode <- function(x) {
  uniqx <- unique(x)
  uniqx[which.max(tabulate(match(x, uniqx)))]}

#Create a vector using the variable "days_to_birth"
x <- clinical_data$days_to_birth
x <- x[!is.na(x)]

# Calculating the mode
calculating_mode <- getmode(x)
calculating_mode
```

```
## [1] -18856
```