TRGN 527: Applied Data Science and Bioinformatics

UNIT I. Introduction and Basic Data Science

Week 2 – Assignment 2

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

Assignment 2

- Instructions: Please answer the following questions, each one worth 5 points (Total: 100 points).
- Load the file "trgn599.clinical.tsv" in your R-Studio.
- In order to answer the following questions you should create a vector of each variable and remove the "NA" values.
- Question 1
  - What is the mean of the variable "age_at_diagnosis"?
    - A) 24648.4
    - B) 30000.1
    - C) 25102
    - D) 27531.1

- Question 2
  - What is the median of the variable "age_at_diagnosis" ?
    - A) 24648
    - B) 25102
    - C) 30000
    - D) 27531

- Question 3
  - What is the standard deviation of the variable "age_at_diagnosis" ?
    - A) 3598.204
    - B) 30000.1
    - C) 25102
    - D) 27531.1

- Question 4
  - What is the variance of the variable "age_at_diagnosis" ?
    - A) 13000001
    - B) 12947070
    - C) 2510200
    - D) 2753101.1

Assignment 2

- Question 5
  - What is the mode of the variable "age_at_diagnosis" ?
    - A) 18856
    - B) 3000.1
    - C) 25102
    - D) 27531.1

- Question 6
  - What is the average of the variable "age_at_diagnosis" ?
    - A) 30001.1
    - B) 24648.4
    - C) 18856.4
    - D) 45672.2

Assignment 2

- Question 7
  - What is the mean of the variable "days_to_death" ?
    - A) 832.1404
    - B) 730.1334
    - C) 25102
    - D) 27531.1

- Question 8
  - What is the median of the variable "days_to_death" ?
    - A) 555
    - B) 578
    - C) 246
    - D) 978

- Question 9
  - What is the standard deviation of the variable "days_to_death" ?
    - A) 814.9373
    - B) 832.1404
    - C) 25102
    - D) 27531.1

- Question 10
  - What is the variance of the variable "days_to_death" ?
    - A) 894122.7
    - B) 664122.7
    - C) 1064123.8
    - D) 906412.7

Assignment 2

- Question 11
  - What is the mode of the variable "days_to_death" ?
    - A) 340
    - B) 240
    - C) 540
    - D) 390

- Question 12
  - What is the average of the variable "days_to_death" ?
    - A) 840.1404
    - B) 832.1404
    - C) 932.1450
    - D) 950.1450

- Question 13
  - What is the mean of the variable "days_to_last_follow_up" ?
    - A) 553.8
    - B) 653.8
    - C) 753.8
    - D) 853.8

- Question 14
  - What is the median of the variable "days_to_last_follow_up" ?
    - A) 276
    - B) 274
    - C) 275
    - D) 277

- Question 15
  - What is the standard deviation of the variable "days_to_last_follow_up" ?
    - A) 670.0331
    - B) 680.0331
    - C) 690.0332
    - D) 770.0332

- Question 16
  - What is the variance of the variable "days_to_last_follow_up" ?
    - A) 558944.3
    - B) 448944.3
    - C) 4998944.2
    - D) 418944.2

- Question 17
  - What is the mode of the variable "days_to_last_follow_up" ?
    - A) 0
    - B) 1
    - C) 2
    - D) 3

- Question 18
  - What is the average of the variable "days_to_last_follow_up" ?
    - A) 557.9
    - B) 553.8
    - C) 554.9
    - D) 555.8

- Load the file "trgn599.rna.tsv" in your R-Studio.

- Question 19

  - What is the mean expression level of the gene "ENSG00000167578.15" across all samples?

    - A) 4.031352
    - B) 5.031352
    - C) 7.001253
    - D) 9.001253

- Question 20

  - What is the median expression level of the gene "ENSG00000167578.15" across all samples?

    - A) 4.624
    - B) 3.624
    - C) 5.625
    - D) 6.625