Biostatistics 203A: Introduction to Data Management and Statistical Computing

Lab Assignment 1: Submission Template

Fall 2023

Name: Sebastian cesar Trevejo Martinez
Student ID Number: 005359531

Exercise 1

	N	Mean	Standard Deviation	Median	Minimum	Maximum
Undergraduate Enrollment	231	14946.62	10569.66	12949.00	1001.00	54513.00
In-State Tuition	133	10895.71	3038.57	10622.00	4965.00	18687.00

Exercise 2

Undergraduate Enrollment	Frequency	Percent
< 5,000	40	17.32
5,000 to 9,999	59	25.54
10,000 to 14,999	31	13.42
15,000 to 24,999	59	25.54
25,000 to 34,9999	32	13.85
35,000 or more	10	4.33

Exercise 3

	Rank 1-50		Rank 51-100		Rank > 100	
	N	**%***	***N***	**%***	***N***	**%***
Undergraduate Enrollment						
< 5,000	7	13.21	7	14.29	26	20.16
5,000 to 9,999	26	49.06	8	16.33	25	19.38
10,000 to 14,999	3	5.66	6	12.24	22	17.05
15,000 to 24,999	6	11.32	13	26.53	40	31.01
25,000 to 34,9999	9	16.98	9	18.37	14	10.85
35,000 or more	2	3.77	6	12.24	2	1.55

*Note: Percentages should reflect the column percentage, meaning that the denominator for each cell is the column total

---

Exercise 4

	Statistics for Variable Rank					
	N	Mean	Standard Deviation	Median	Minimum	Maximum
Undergraduate Enrollment						
< 5,000	40	127.98	66.99	146	7	220
5,000 to 9,999	59	87.54	71.43	82	1	220
10,000 to 14,999	31	149.23	63.03	159	15	220
15,000 to 24,999	59	128.86	55.37	135	23	220
25,000 to 34,9999	32	93.44	55.38	83	20	220
35,000 or more	10	82.7	40.51	72	50	176

Exercise 5

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
5	in_state	Num	8	FEEFMT.	COMMA10.	Annual In-State Tuition
2	location	Char	50			Location
1	name	Char	50			Name
6	rank	Num	8			Rank
3	tuition_and_fees	Num	8	FEEFMT.	COMMA10.	Tuition and Fees
4	undergrad_enrollment	Num	8		COMMA10.	Undergraduate Enrollment

---

Exercise 6

```
**** exercise 6;

libname myfmts "~/my_files/format";
libname mydat "~/my_files/data";
options fmtsearch=(myfmts);

proc format library=myfmts;
value $gendftm "M" = "male"
"F" = "female";
value yesnofmt 1 = "No"
2 = "Yes";

data lung_cancer;
infile "~/my_shared_file_links/u5338439/survey_lung_cancer.csv" dsd firstobs=2;
input gender$
age
smoking
yellow_fingers
anxiety
peer_pressure
chronic_disease
fatigue
allergy
wheezing
alcohol
coughing
shortness_of_breath
swallowing_difficulty
chest_pain
lung_cancer $;

format gender $gendfmt.
age
smoking
yellow_fingers
anxiety
peer_pressure
chronic_disease
fatigue
allergy
wheezing
alcohol
coughing
shortness_of_breath
swallowing_difficulty
chest_pain yesnofmt.;
```

run;

Exercise 7

	Lung Cancer		**No Lung Cancer**	
	N	**%[1]**	***N***	**%[2]**
Risk Factors				
Smoking	155	57%	19	49%
Anxiety	142	53%	12	31%
Peer Pressure	145	54%	10	26%
Alcohol	165	61%	7	18%
Percentages should reflect the percentage of all Lung Cancer[1]/No Lung Cancer[2] observations that had value “Yes” for the risk factor listed in the corresponding row.				
