Tony Lim
BIOSTAT 203A LAB 1A
Professor Hilary Aralis
14 Oct 2019

Lab 2

Exercise 1

```
proc means data=rankings N mean std median min max;
 var undergrad_enrollment;
 var in_state;
run;
```

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Median	Minimum	Maximum
undergrad_enrollment	Undergraduate Enrollment	231	14946.62	10569.66	12949.00	1001.00	54513.00
in_state	In-State Tuition	133	10895.71	3038.57	10622.00	4965.00	18687.00

Exercise 2

```
proc format;
 value enrollfmt low-5000 = "< 5,000"
          5000-<10000 = "5,000 to 9,999"
          10000-<15000 = "10,000 to 14,999"
          15000-<25000 = "15,000 to 24,999"
          25000-<35000 = "25,000 to 34,999"
          35000-high = "35,000 or more";
run;

proc freq data=rankings;
 format undergrad_enrollment enrollfmt.;
 tables undergrad_enrollment/nocum;
run;
```

The FREQ Procedure

Undergraduate Enrollment		
undergrad_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	91	39.39
35,000 or more	10	4.33

Exercise 3

```
proc format;
 value rankfmt 1-50 = "Rank 1-50"
          51-<101 = "Rank 51-100"
          101-high = "Rank > 100";
run;

proc freq data=rankings;
 format undergrad_enrollment enrollfmt.;
 format rank rankfmt.;
 tables undergrad_enrollment*rank/nopercent norow;
run;
```

	Rank 1-50		**Rank 51-100**		**Rank > 100**	
	N	**%***	***N***	**%***	***N***	**%***
Undergraduate Enrollment	53	100.00	49	100.00	129	100.00
< 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 34,999	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 call is the column total

Exercise 4

```
proc means data = rankings N mean std median min max;
 format undergrad_enrollment enrollfmt.;
 class undergrad_enrollment;
 var rank;
run;
```

The MEANS Procedure

Analysis Variable : rank Rank						
Undergraduate Enrollment	N Obs	Mean	Std Dev	Median	Minimum	Maximum
< 5,000	40	127.9750000	66.9934890	146.0000000	7.0000000	220.0000000
5,000 to 9,999	59	87.5423729	71.4305358	82.0000000	1.0000000	220.0000000
10,000 to 14,999	31	149.2258065	63.0284114	159.0000000	15.0000000	220.0000000
15,000 to 24,999	59	128.8644068	55.3685014	135.0000000	23.0000000	220.0000000
25,000 to 34,999	32	93.4375000	55.3790094	83.0000000	20.0000000	220.0000000
35,000 or more	10	82.7000000	40.5135094	72.0000000	50.0000000	176.0000000

Exercise 5

```
proc format;
 value feefmt low-<20000 = "< $20,000"
          20000-<30000 = "$20,000 to $29,999"
          30000-<40000 = "$30,000 to $39,999"
          40000-<50000 = "$40,000 to $49,999"
          50000-high = "$50,000 or more";
run;

proc format;
 value instfmt low-<6000 = "< $6,000"
          6000-<8000 = "$6,000 to $8,000"
          8000-<10000 = "$8,000 to $10,000"
          10000-<12000 = "$10,000 to $12,000"
          12000-<14000 = "$12,000 to $14,000"
          14000-<16000 = "14,000 to $16,000"
          16000-high = "$16,000 or more";
run;

proc contents data=rankings;
run;
```

Alphabetic List of Variables and Attributes

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

Exercise 6

```
proc format;
 value ynfmt 2 = "Yes"
             1 = "No";
run;

data lung;
 infile "/folders/myfolders/Lab_2/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 Smoking Yellow_fingers Anxiety Peer_pressure Chronic_disease Fatigue Allergy
Wheezing Alcohol Coughing Shortness_of_Breath Swallowing_Difficulty Chest_pain ynfmt.;
run;
```

Exercise 7

```
proc freq data=lung;
 tables (smoking anxiety peer_pressure alcohol)*lung_cancer;
run;
```

	Lung Cancer		**No Lung Cancer**	
	N	**%[1]**	***N***	**%[2]**
Risk Factors				
Smoking	155	57.41	19	48.72
Anxiety	142	52.59	12	30.77
Peer Pressure	145	53.70	10	25.64
Alcohol	165	61.11	7	17.95
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.				