Homework 2 STAT: 2010 Stats Methods and Computing

Yubing Li - 00808366

1. Input data:

**data** wieners;

input type $ calories sodium;

datalines ;

Beef 186 495

…

…

…

Poultry 144 545

;

**run**;

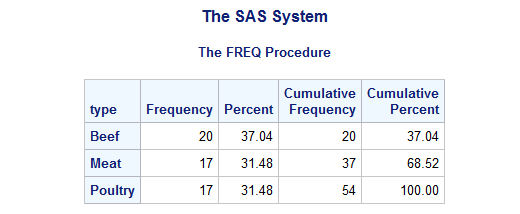
(a) Types of hotdogs qualitative nominal variables.

(b) Table of frequencies and percents is an appropriate table to describe nominal data.

**proc** **freq** data=wieners;

tables type;

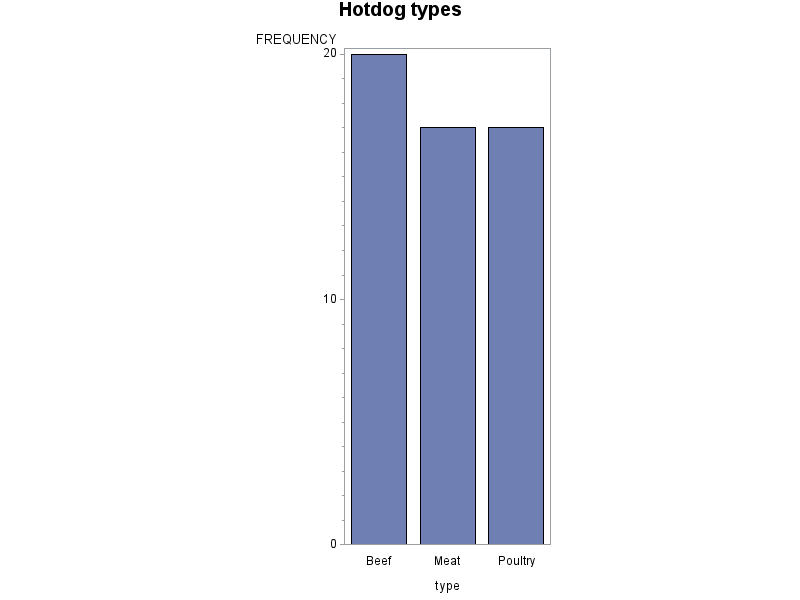
**run**;



(c) The proportion of poultry hotdogs among all types of hotdogs is 31.48%.

(d) The bar charts and the pie charts are appropriate to display qualitative variables.

(e) Bar chart:



**proc** **gchart** data=wieners ;

vbar type ;

title 'Hotdo types';

**run**;

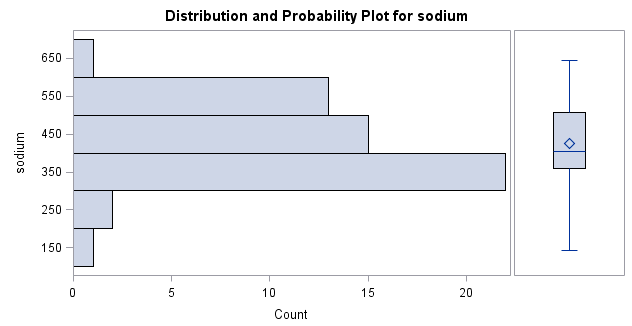
(f) Sodium is quantitative continuous variable.

**proc** **univariate** plot data=wieners;

var sodium;

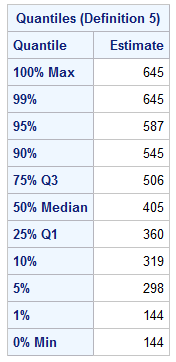
**run**;

(g) (Horizontal) histogram and boxplot of sodium:



i. The distribution of sodium is roughly symmetric

ii. There are outliers. (For example: 144, 645 seem to lie far from the main group of points)



(h) 5-number summary:

Minimum: 144

1st quantile: 360

Median: 405

3rd quantile: 506

Maximum: 645

(i) IQR=506-360=146

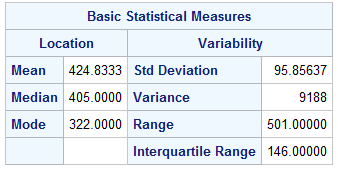
|  |
| --- |
| IQR×1.5 = 219 |
| Q1-1.5IQR = 360-219 = 141 |
| Q3+1.5IQR = 506+219=725 |

Adjacent values:

(j) 726 is the smallest value that qualify as an outlier in high direction. There’s no high outliers in this set of data.

(k) 140 is the highest value that qualify as an outlier in low direction. There’s no low outliers in this set of data.

(l) It is reasonable to compute the mean and the standard deviation of sodium because it has an approximately symmetric distribution.



Mean: 424.83

Std. Deviation: 95.86

(m) The unit for the mean is milliequivalents.

The unit for the std. deviation is milliequivalents.

Both of them have the same units as the original measurements of sodium.

3. Input data and analyze data:

**data** salaries;

input salary;

datalines ;

8000000

…

500000

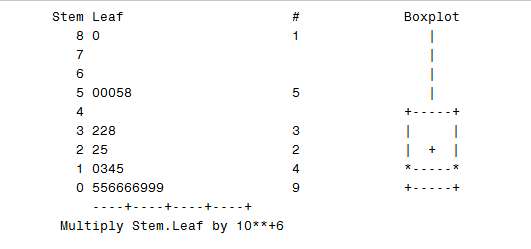
;

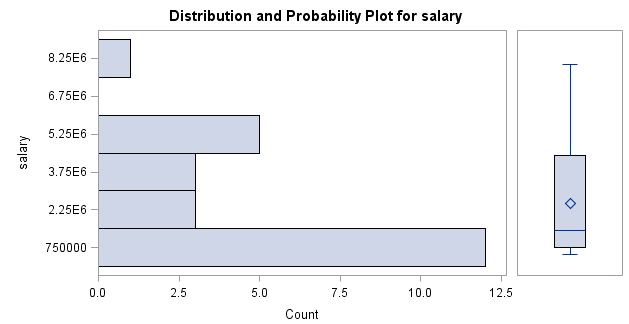
**run**;

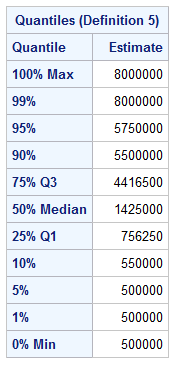
**proc** **univariate** plot data=salaries;

var salary;

**run**;







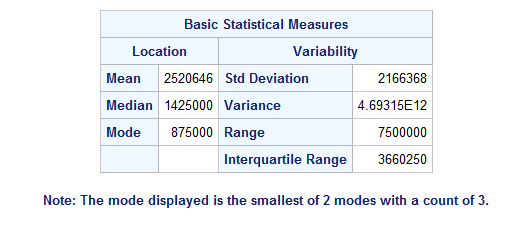
Min: 500,000

Q1: 756,250

Median: 1,425,000

Q3: 4,416,500

Max: 8,000,000



Mean: 2520646

Median: 1425000

Brief discreption:

The slemplot and the boxplot show that the distibution is not symmetric.

The 5-number summary are 500,000; 756,250; 1,425,000; 4,416,500; 8,000,000.

The stemplot shows that the distribution has larger porpotions for lower salaries and it’s skewed to the right. The mean is greater than the median.

There’s an high outlier exists, which is 8,000,000.

There’re two modes, one is on the lower end and one is around 5,000,000.