|  |  |  |
| --- | --- | --- |
| **Variable** | **Central Tendency** | **Dispersion** |
| Age | 33 | Interquartile range i.e 15.5 |
| Amount of loan | 2400.5 | 2,841.5 |
| Duration of loan | 18 | 12 |
| Instalment rate as a percentage of income | 3 | 2 |

Q3

title "Histogram plot of Age, Credit Amount, Duration in Months and Instalment Rate";

**proc** **univariate**

data = LOAN\_RISK;

histogram/normal (mu= est sigma= est); /\*normally distributed?\*/

var

age

amount

duration

instalment

;

**run**;

title;

ods noproctitle;

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Minimum | P25 | Median | P75 | Maximum | Skewness | Kurtosis |
| *age* | 19 | 26.5 | 33 | 42 | 74 | The variable is positively skewed with most observations on the left (huge gap between the Max value and P75). This is demonstrated by skewness of 1.06252405 | The variable has a low peak with thick tail as kurtosis of 0.6630448 |
| *amount of loan* | 276 | 1389.5 | 2400.5 | 4231.0 | 15945.0 | The variable is positively skewed with most of the observations on the left (huge gap between the Max value and P75). The skewness value is 1.83135718 | The variable has a high peak with thin tail. This is demonstrated by kurtosis value greater than 3 at 3.74716507 |
| *duration of loan* | 4 | 12 | 18 | 24 | 60 | The variable is positively skewed with most of the observations on the left (huge gap between the Max value and P75). This is demonstrated by skewness value of 1.11621765 | The variable has a low peak with thick tail. This is demonstrated by kurtosis value of 0.90288889 |
| *instalment rate as a percentage of income* | 1 | 2 | 3 | 4 | 4 | The variable is negatively skewed though the distribution is near normal (little gap between Max value and P75). This is also demonstrated by skewness value of -0.4598333 | The variable has a low peak with thick tail. This is demonstrated by kurtosis value far less than 3. |

Q4

ods graphics on;

title "Q-Q Plot of Age, Credit Amount, Duration in Months and Instalment Rate";

**proc** **univariate**

data = LOAN\_RISK normaltest;

qqplot/normal (mu= est sigma= est);

var

age

amount

duration

instalment

;

**run**;

title;

ods noproctitle;

ods graphics off;

ods graphics on;

title "P-P Plot of Age, Credit Amount, Duration in Months and Instalment Rate";

**proc** **univariate**

data = LOAN\_RISK normaltest;

var

age

amount

duration

instalment

;

ppplot;

**run**;

title;

ods noproctitle;

ods graphics off;

|  |  |
| --- | --- |
| Variable | Descriptive statistics |
| *age* | The descriptive statistics of age shows that the variable is not normally distributed as the p value is < 0.05. There is gap between the mean of 35.39 and median at 33.0 with a large variance of 130.49. The skewness and kurtosis are 1.06252405 and 0.6630448 respectively. The variable appears skewed as the tail end are not on the normal line of the Q-Q and p-p plots |
| *amount of loan* | The descriptive statistics of age shows that the variable is not normally distributed as the p value is < 0.05. There is large gap between the mean of 3359.44, median of 2400.50 and mode of 1393.0. The skewness and kurtosis are 1.83 and 3.75 respectively. The variable appears highly skewed with significant deviation from the normal line of the Q-Q and p-p plots. |
| *duration of loan* | The descriptive statistics of age shows that the variable is not normally distributed as the p value is < 0.05. There is large gap between the mean of 20.59, median of 18.0 and mode of 12.0. The skewness and kurtosis are 1.12 and 0.9 respectively. The points of the variable deviate from the normal line on the q-q and p-p plots |
| *instalment rate as a percentage of income* | The descriptive statistics of age shows that the variable is not normally distributed as the p value is < 0.05. There is gap between the mean of 2.94, median of 3 and mode of 4. The skewness and kurtosis are -0.46 and -1.25 respectively. The points of the variable deviate from the normal line on the q-q and p-p plots |

Q5

ods select extremevalues;

**proc** **univariate**

data = LOAN\_RISK nextrval= **5**; /\*selects 5 extreme values\*/

var

age

amount

duration

instalment

;

**run**;

ods select all;

/\*Q5b\*/

/\* requirement: to output into csv file, 10 extreme observations for Age, duration, credit amount and instalment rate\*/

ods select extremeobs;

ods csvall file= 'C:\Users\ayoyi\OneDrive - De Montfort University\Analytics programming\Week 7\age\_extreme.csv';

**proc** **univariate**

data = LOAN\_RISK nextrobs= **10**; /\*selects 10 extreme observations for age\*/

var age;

ID customer;

**run**;

ods select extremeobs;

ods csvall file= 'C:\Users\ayoyi\OneDrive - De Montfort University\Analytics programming\Week 7\amount\_extreme.csv';

**proc** **univariate**

data = LOAN\_RISK nextrobs= **10**; /\*selects 10 extreme observations for credit amount\*/

var amount;

ID customer;

**run**;

ods select extremeobs;

ods csvall file= 'C:\Users\ayoyi\OneDrive - De Montfort University\Analytics programming\Week 7\duration\_extreme.csv';

**proc** **univariate**

data = LOAN\_RISK nextrobs= **10**; /\*selects 10 extreme observations for duration\*/

var duration;

ID customer;

**run**;

ods select extremeobs;

ods csvall file= 'C:\Users\ayoyi\OneDrive - De Montfort University\Analytics programming\Week 7\instalmentextreme.csv';

**proc** **univariate**

data = LOAN\_RISK nextrobs= **10**; /\*selects 10 extreme observations for instalment\*/

var instalment;

ID customer;

**run**;

ods csv close;

ods select all;

/\*requirement: to import the csv files back to sas\*/

filename pwd 'C:\Users\ayoyi\OneDrive - De Montfort University\Analytics programming\Week 7';

**Data** AGE\_EXTREME;

infile pwd (age\_extreme.csv)dsd firstobs=**5**;

input

value

customer

obs

@@

;

**run**;

**Data** AMOUNT\_EXTREME;

infile pwd (amount\_extreme.csv )dsd firstobs=**5**;

input

value

customer

obs

@@

;

**run**;

**Data** DURATION\_EXTREME;

infile pwd (duration\_extreme.csv )dsd firstobs=**5**;

input

value

customer

obs

@@

;

**run**;

**Data** INSTALMENT\_EXTREME;

infile pwd (instalmentextreme.csv)dsd firstobs=**5**;

input

value

customer

obs

@@

;

**run**;

/\*sorts highest extremes for age\*/

**proc** **sql**;

create table highestextreme\_age as

select \*

from age\_extreme

where value > **21**;

**quit**;

/\*sorts lowest extremes for age\*/

**proc** **sql**;

create table lowestextreme\_age as

select \*

from age\_extreme

where value <= **21**;

**quit**;

/\*sorts highest extremes for credit amount\*/

**proc** **sql**;

create table highestextreme\_amount as

select \*

from amount\_extreme

where value > **454**;

**quit**;

/\*sorts lowest extremes for credit amount\*/

**proc** **sql**;

create table lowestextreme\_amount as

select \*

from amount\_extreme

where value <= **454**;

**quit**;

/\*sorts highest extremes for duration in months\*/

**proc** **sql**;

create table highestextreme\_duration as

select \*

from duration\_extreme

where value > **6**

;

**quit**;

/\*sorts lowest extremes for duration in months\*/

**proc** **sql**;

create table lowestextreme\_duration as

select \*

from duration\_extreme

where value <= **6**

;

**quit**;

/\*sorts highest extremes for instalment rate\*/

**proc** **sql**;

create table highestextreme\_instalment as

select \*

from instalmentextreme

where value > **1**;

**quit**;

/\*sorts lowest extremes for instalment rate\*/

**proc** **sql**;

create table lowestextreme\_instalment as

select \*

from instalmentextreme

where value <= **1**;

**quit**;

/\*prints highest extreme values for age in csv file\*/

ods csv file= 'highestextreme\_age.csv';

**proc** **print**

data = highestextreme\_age;

var value;

id customer;

**run**;

ods csv close;

/\*prints lowest extreme values for age in csv file\*/

ods csv file= 'lowestextreme\_age.csv';

**proc** **print**

data = lowestextreme\_age;

var value;

id customer;

**run**;

ods csv close;

/\*prints highest extreme values for credit amount in csv file\*/

ods csv file= 'highestextreme\_amount.csv';

**proc** **print**

data = highestextreme\_amount;

var value;

id customer;

**run**;

ods csv close;

/\*prints lowest extreme values for credit amount in csv file\*/

ods csv file= 'lowestextreme\_amount.csv';

**proc** **print**

data = lowestextreme\_amount;

var value;

id customer;

**run**;

ods csv close;

/\*prints highest extreme values for duration in csv file\*/

ods csv file= 'highestextreme\_duration.csv';

**proc** **print**

data = highestextreme\_duration;

var value;

id customer;

**run**;

ods csv close;

/\*prints lowest extreme values for duration in csv file\*/

ods csv file= 'lowestextreme\_duration.csv';

**proc** **print**

data = lowestextreme\_duration;

var value;

id customer;

**run**;

ods csv close;

/\*prints highest extreme values for instalment rate in csv file\*/

ods csv file= 'highestextreme\_instalment.csv';

**proc** **print**

data = highestextreme\_instalment;

var value;

id customer;

**run**;

ods csv close;

/\*prints lowest extreme values for instalment rate in csv file\*/

ods csv file= 'lowestextreme\_instalment.csv';

**proc** **print**

data = lowestextreme\_instalment;

var value;

id customer;

**run**;

ods csv close;

Q6

/\*to transform data to normal\*/

**data** LOAN\_RISK\_TRANSFORM;

set LOAN\_RISK;

trans\_age = log(age);

trans\_amount = log(amount);

trans\_duration = log(duration);

trans\_instalment = log(instalment)

;

**run**;

title "Plot of transformed variables";

**proc** **univariate**

data = LOAN\_RISK\_TRANSFORM normaltest;

var

trans\_age

trans\_amount

trans\_duration

trans\_instalment

;

qqplot/normal(mu=est sigma=est);

**run**;

title;

ods noproctitle;

Q7

title 'Frequency table of Customer ID of Customer Data';

**proc** **freq**

data = customer ;

table customer/missing;

**run**;

title 'Frequency table of Customer ID of Assessment Data';

**proc** **freq**

data = assessment;

table customer/missing;

**run**;

title 'Frequency table of Customer ID of Rating Data';

**proc** **freq**

data = rating ;

table customer/missing;

**run**;

title;

ods select all;