```sas
/*GROUP 3
Ugonna Okengwu
Kaveen Dias
Sikirat Ajike Durojaiye*/

libname mylib '/home/u63417899/BAN110ZBB';

data customer;
set mylib.customer_all;
run;

proc contents data=mylib.customer_all;
run;

 /*Question 1*/
title 'List of the numberical attributes and their descriptive statistics';
PROC MEANS DATA=mylib.customer_all N NMISS MEAN STD MIN MAX MAXDEC = 2;
  VAR age pdays day campaign previous balance;
RUN;

/*Q1 Answer:- The Range of 'pdays' is -1 to 854.
-1 means customer was not previously contacted by campaign.*/

/*Q2 Answer:- The Range of values for 'day' is from 1 to 31 and
this variable refer as last contact day of the month*/

/*Q3 Answer:- Yes, 20. is the missing values in variable "age"*/

 /*Question 2*/
*2.1*;
title 'PROC UNIVARIATE on AGE variable';
proc univariate data=mylib.Customer_all;
    var age;
    histogram / normal;
run;

*2.2*;
title 'Print the customer_id whose age is missing';
data _null_;
    file print;
    set mylib.Customer_all;
    if missing(AGE) then
    put 'Age of this ' customer_id= 'is missing';
run;

*2.3*;
proc stdize data=mylib.Customer_all
    out=mylib.Customer_all_Imputed /*new data name Customer_Imputed */
    reponly /*replace the value*/
    method=mean; /*use method by mean*/
    var AGE;
run;

*2.4*;
proc means data=mylib.customer_all_imputed nmiss mean;
    title 'Mean and Missing values of variable age in Imputed dataset';
    var age;
run;
/*Answer:- No missing value in the variable "AGE".*/

proc means data=mylib.Customer_all nmiss mean;
    title 'Mean and Missing values of variable age in Original dataset';
    var age;
run;

proc datasets library=mylib;
    delete Customer;
    title 'Modified library by delete original dataset';
run;

*2.5*;
/*Rename modified dataset 'Customer_Imputed' to Original dataset name 'Customer'*/
proc datasets library=mylib;
    change Customer_Imputed=Customer_all;
    title 'Modified library by the renamed dataset';
run;
```

```sas
title 'Influence of age on balance';
proc sgplot data=mylib.customer_all;
    reg x=age
    y=balance / lineattrs=(color=Green thickness=4);
run;


*question 3*;
*3.1*;
data mylib.customer_all;
    set mylib.customer_all;
    if 18<= age <=35 then age_cat='young_adult';
    else if 36<= age <=55 then age_cat='middle_age';
        else if age =>56 then age_cat='old';
run;


proc freq data=mylib.customer_all;
    table age_cat;
    title 'Frequency table of age_cat variable';
run;


*3.2*;
proc sgplot data=mylib.customer_all;
    vbar age_cat / response=balance group=balance groupdisplay=cluster stat=mean;
    yaxis grid;
run;


*Our group opinion: it is more evident that the age group "old" has more balances in their accounts compared to other groups*

*Question 4 Examine the variable campaign*;
*4.1*;
proc univariate data=mylib.Customer_all;
    var campaign;
    histogram / normal;
run;


*4.2*;
data mylib.customer_all;
set mylib.Customer_all;
campaign_cat = put(campaign, 8.);
drop campaign;
run;


proc contents data=mylib.customer_all;
run;


*Question 4 Examine the variable "balance"*;
*4.1*;
title "Running PROC UNIVARIATE on balance";
proc univariate data=mylib.customer_all;
    id customer_id;
    var balance;
    histogram /normal;
run;


*4.2*;
proc sgplot data=mylib.customer_all;
    vbar y / response=balance group=balance groupdisplay=cluster stat=mean;
    yaxis grid;
run;


title 'distribution of balance by y';
proc univariate data=mylib.customer_all noprint;
class y;
histogram balance;
run;


*we think that the 1st graph clearly illustrate that there is a tendancy of cutomers with high balances purchase CD*;

*Question 5*;
*5.1*;
proc univariate data=mylib.customer_all;
    id customer_id;
    var pdays;
    histogram /normal;
run;
```

```sas
*5.2*;
data mylib.customer_all;
set mylib.customer_all;
if pdays >= 1 then customer_contacted = "yes";
if pdays =-1 then customer_contacted = "No";
run;

proc print data=mylib.customer_all (obs=20) ;
run;

*5.3*;
proc print data=mylib.customer_all (firstobs=1 obs=5);
where pdays > 0;
run;

*5.4*;
data mylib.Customer_all_droppdays;
set mylib.customer_all (drop=pdays);
run;

*5.5*;
proc means data=mylib.customer_all_droppdays;
run;

*Question 6*;
proc sort data=mylib.customer_all_droppdays(keep=balance where=(balance is not missing))out=Tmp;
by balance;
run;

data _null_;
    if 0 then set  nobs=Number_of_Obs; *2;
    High = Number_of_Obs - 9;
    call symputx('High_Cutoff',High); *3;
    stop; *4;
run;

title "Ten Highest and Lowest Values for HR";
data _null_;
    set Tmp(obs=10)                      /* 10 lowest values */
        Tmp(firstobs=&High_Cutoff); *5; /* 10 highest values */
    file print; *6;
    if _n_ le 10 then do; *7;
       if _n_ = 1 then put / "Ten Lowest Values"; *8;
       put "balance = " balance @15;
    end;
    else if _n_ ge 11 then do; *9;
       if _n_ = 11 then put / "10 Highest Values";
       put "balance = " balance @15;
    end;
run;
```

The CONTENTS Procedure

Data Set Name	MYLIB.CUSTOMER_ALL	Observations	10578
Member Type	DATA	Variables	18
Engine	V9	Indexes	0
Created	07/10/2023 23:23:05	Observation Length	136
Last Modified	07/10/2023 23:23:05	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	12
First Data Page	1
Max Obs per Page	962
Obs in First Data Page	931
Number of Data Set Repairs	0
Filename	/home/u63417899/BAN110ZBB/customer_all.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	8335149
Access Permission	rw-r--r--
Owner Name	u63417899
File Size	2MB
File Size (bytes)	1703936

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
15	AGE	Num	8	F4.		AGE
14	Education	Char	9	$CHAR9.		Education
17	JOB	Char	14	$CHAR14.		JOB
18	age_cat	Char	11			
11	balance	Num	8	BEST12.	BEST32.	
5	campaign	Num	8	BEST.		campaign
2	contact	Char	9	$9.	$9.	contact
1	customer_id	Num	8	BEST12.	BEST32.	
3	day	Num	8	BEST.		day
10	default	Char	3	$3.	$3.	
12	housing	Char	3	$3.	$3.	
13	loan	Char	3	$3.	$3.	
16	marital	Char	8	$CHAR8.		marital
4	month	Char	3	$3.	$3.	month
6	pdays	Num	8	BEST.		pdays
8	poutcome	Char	7	$7.	$7.	poutcome
7	previous	Num	8	BEST.		previous
9	y	Char	3	$3.	$3.	y

List of the numberical attributes and their descriptive statistics

The MEANS Procedure

Variable	Label	N	N Miss	Mean	Std Dev	Minimum	Maximum
AGE	AGE	10558	20	41.26	12.15	18.00	146.00
pdays	pdays	10578	0	51.95	109.35	-1.00	854.00
day	day	10578	0	15.48	8.41	1.00	31.00
campaign	campaign	10578	0	2.47	2.62	1.00	50.00
previous	previous	10578	0	0.85	3.47	0.00	275.00
balance		10578	0	1548.53	3130.57	-3058.00	81204.00

PROC UNIVARIATE on AGE variable

The UNIVARIATE Procedure
Variable: AGE (AGE)

Moments			
N	10558	Sum Weights	10558
Mean	41.2641599	Sum Observations	435667
Std Deviation	12.1483452	Variance	147.582292
Skewness	1.00818411	Kurtosis	2.05204285
Uncorrected SS	19535459	Corrected SS	1558026.26
Coeff Variation	29.4404279	Std Error Mean	0.11822962

Basic Statistical Measures

Location		Variability	
Mean	41.26416	Std Deviation	12.14835
Median	39.00000	Variance	147.58229
Mode	31.00000	Range	128.00000
		Interquartile Range	17.00000

Tests for Location: Mu0=0

Test		Statistic	p Value	
Student's t	t	349.0171	Pr > |t|	<.0001
Sign	M	5279	Pr >= |M|	<.0001
Signed Rank	S	27870481	Pr >= |S|	<.0001

Quantiles (Definition 5)

Level	Quantile
100% Max	146
99%	77
95%	62
90%	58
75% Q3	49
50% Median	39
25% Q1	32
10%	28
5%	26
1%	22
0% Min	18

Extreme Observations

Lowest		Highest	
Value	Obs	Value	Obs
18	10273	95	8550
18	9679	130	3490
18	9240	139	1152
18	8880	144	1531
18	8455	146	402

Missing Values

Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	20	0.19	100.00

PROC UNIVARIATE on AGE variable
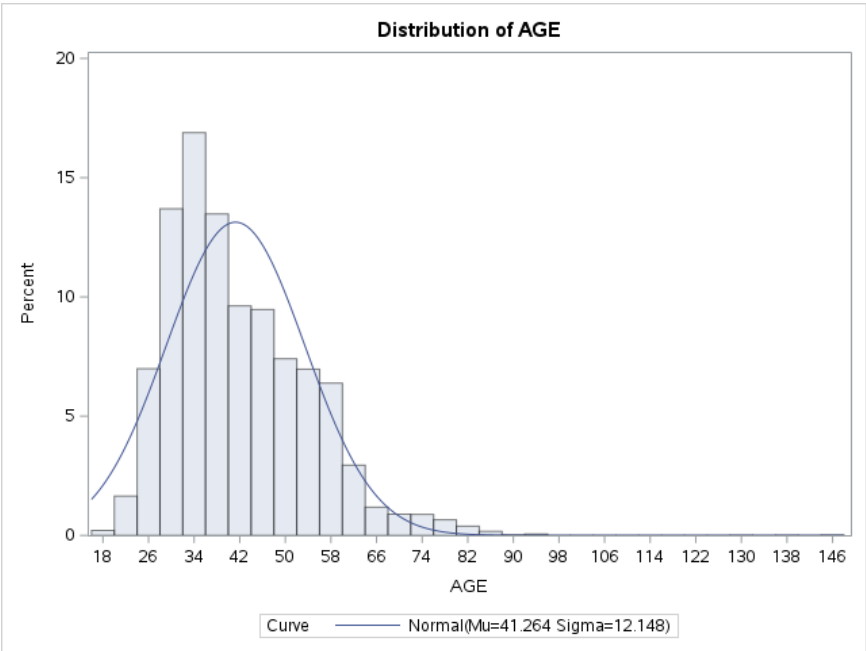
The UNIVARIATE Procedure

PROC UNIVARIATE on AGE variable

The UNIVARIATE Procedure
Fitted Normal Distribution for AGE (AGE)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	41.26416
Std Dev	Sigma	12.14835

Goodness-of-Fit Tests for Normal Distribution			
Test	Statistic		p Value
Kolmogorov-Smirnov	D	0.106237	Pr > D	<0.010
Cramer-von Mises	W-Sq	25.931403	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	150.997380	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	22.0000	13.0029
5.0	26.0000	21.2819
10.0	28.0000	25.6954
25.0	32.0000	33.0702
50.0	39.0000	41.2642
75.0	49.0000	49.4581
90.0	58.0000	56.8329
95.0	62.0000	61.2464
99.0	77.0000	69.5254

Print the customer_id whose age is missing

```
Age of this customer_id=100898 is missing
Age of this customer_id=103782 is missing
Age of this customer_id=104872 is missing
Age of this customer_id=108581 is missing
Age of this customer_id=112972 is missing
Age of this customer_id=113317 is missing
Age of this customer_id=114933 is missing
Age of this customer_id=115167 is missing
Age of this customer_id=117338 is missing
Age of this customer_id=122821 is missing
Age of this customer_id=127452 is missing
Age of this customer_id=128062 is missing
Age of this customer_id=128123 is missing
Age of this customer_id=131745 is missing
Age of this customer_id=134418 is missing
Age of this customer_id=134663 is missing
Age of this customer_id=135384 is missing
Age of this customer_id=135695 is missing
Age of this customer_id=143464 is missing
Age of this customer_id=143512 is missing
```

Mean and Missing values of variable age in Imputed dataset

The MEANS Procedure

Analysis Variable : AGE AGE	
N Miss	Mean
0	41.2641599

Mean and Missing values of variable age in Original dataset

The MEANS Procedure

Analysis Variable : AGE AGE	
N Miss	Mean
20	41.2641599

Mean and Missing values of variable age in Original dataset

Directory	
Libref	MYLIB
Engine	V9
Physical Name	/home/u63417899/BAN110ZBB
Filename	/home/u63417899/BAN110ZBB
Inode Number	23398586
Access Permission	rwxr-xr-x
Owner Name	u63417899
File Size	4KB
File Size (bytes)	4096

#	Name		Member Type	File Size	Last Modified

#	Name	Member Type	File Size	Last Modified
1	CUSTOMER_ALL	DATA	2MB	07/11/2023 03:23:05
2	CUSTOMER_ALL (2)	DATA	1MB	07/08/2023 18:58:06
3	CUSTOMER_ALL_IMPUTED	DATA	2MB	07/12/2023 01:13:40
4	CUSTOMER_BANKING_INFO	DATA	512KB	05/31/2023 02:53:31
5	CUSTOMER_BANKING_INFO_PROMO	DATA	896KB	05/31/2023 02:53:31
6	CUSTOMER_BANKING_INFO_PROMOCV	DATA	896KB	05/31/2023 02:53:32
7	CUSTOMER_DEMOGRAPHIC	DATA	146KB	05/30/2023 21:30:22
8	CUSTOMER_DEMOGRAPHICS	DATA	146KB	05/31/2023 02:24:01
9	CUSTOMER_IMPUTED	DATA	1MB	07/11/2023 03:04:45
10	PATIENTS	DATA	256KB	05/16/2023 21:28:36

Modified library by delete original dataset

Directory	
Libref	MYLIB
Engine	V9
Physical Name	/home/u63417899/BAN110ZBB
Filename	/home/u63417899/BAN110ZBB
Inode Number	23398586
Access Permission	rwxr-xr-x
Owner Name	u63417899
File Size	4KB
File Size (bytes)	4096

#	Name	Member Type	File Size	Last Modified
1	CUSTOMER_ALL	DATA	2MB	07/11/2023 03:23:05
2	CUSTOMER_ALL (2)	DATA	1MB	07/08/2023 18:58:06
3	CUSTOMER_ALL_IMPUTED	DATA	2MB	07/12/2023 01:13:40
4	CUSTOMER_BANKING_INFO	DATA	512KB	05/31/2023 02:53:31
5	CUSTOMER_BANKING_INFO_PROMO	DATA	896KB	05/31/2023 02:53:31
6	CUSTOMER_BANKING_INFO_PROMOCV	DATA	896KB	05/31/2023 02:53:32
7	CUSTOMER_DEMOGRAPHIC	DATA	146KB	05/30/2023 21:30:22
8	CUSTOMER_DEMOGRAPHICS	DATA	146KB	05/31/2023 02:24:01
9	CUSTOMER_IMPUTED	DATA	1MB	07/11/2023 03:04:45
10	PATIENTS	DATA	256KB	05/16/2023 21:28:36

Influence of age on balance

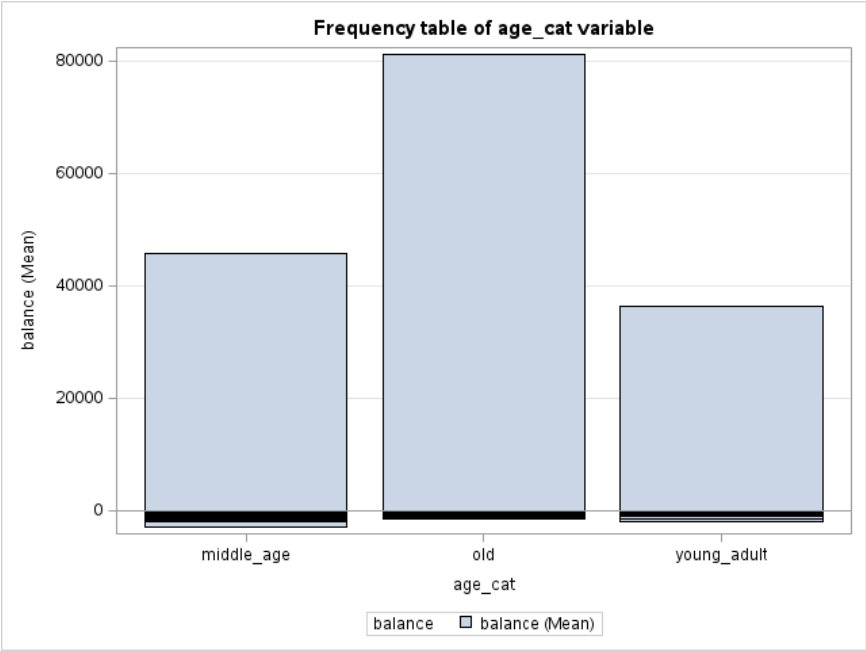**Frequency table of age_cat variable**

The FREQ Procedure

age_cat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
middle_age	4957	46.95	4957	46.95
old	1438	13.62	6395	60.57
young_adult	4163	39.43	10558	100.00

age_cat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Frequency Missing = 20				

Frequency table of age_cat variable

Frequency table of age_cat variable

The UNIVARIATE Procedure
Variable: campaign (campaign)

Moments			
N	10578	Sum Weights	10578
Mean	2.47475893	Sum Observations	26178
Std Deviation	2.61517814	Variance	6.83915672
Skewness	5.0976061	Kurtosis	44.6295296
Uncorrected SS	137122	Corrected SS	72337.7606
Coeff Variation	105.674056	Std Error Mean	0.02542726

Basic Statistical Measures			
Location		Variability	
Mean	2.474759	Std Deviation	2.61518
Median	2.000000	Variance	6.83916
Mode	1.000000	Range	49.00000
		Interquartile Range	2.00000

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	97.32702	Pr > |t|	<.0001
Sign	M	5289	Pr >= |M|	<.0001
Signed Rank	S	27976166	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	50
99%	14
95%	7
90%	5
75% Q3	3
50% Median	2
25% Q1	1
10%	1
5%	1
1%	1
0% Min	1

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	10574	31	3114

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	10573	32	543
1	10572	37	1643
1	10571	43	2276
1	10569	50	3214

Frequency table of age_cat variable
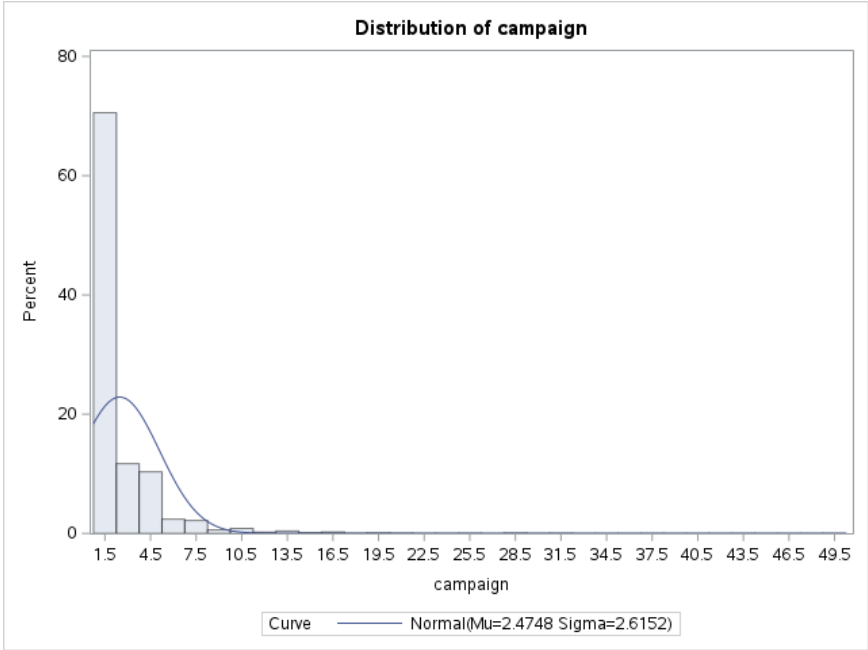
The UNIVARIATE Procedure

Distribution of campaign

Frequency table of age_cat variable

The UNIVARIATE Procedure
Fitted Normal Distribution for campaign (campaign)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	2.474759
Std Dev	Sigma	2.615178

Goodness-of-Fit Tests for Normal Distribution				
Test		Statistic	p Value	
Kolmogorov-Smirnov	D	0.28640	Pr > D	<0.010
Cramer-von Mises	W-Sq	228.76726	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	1214.01226	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	1.00000	-3.60906
5.0	1.00000	-1.82683
10.0	1.00000	-0.87673
25.0	1.00000	0.71085
50.0	2.00000	2.47476
75.0	3.00000	4.23867
90.0	5.00000	5.82624
95.0	7.00000	6.77634
99.0	14.00000	8.55857

Frequency table of age_cat variable

The CONTENTS Procedure

Data Set Name	MYLIB.CUSTOMER_ALL	Observations	10578
Member Type	DATA	Variables	18
Engine	V9	Indexes	0
Created	07/11/2023 21:13:43	Observation Length	136

Last Modified	07/11/2023 21:13:43	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	12
First Data Page	1
Max Obs per Page	962
Obs in First Data Page	931
Number of Data Set Repairs	0
Filename	/home/u63417899/BAN110ZBB/customer_all.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	8335149
Access Permission	rw-r--r--
Owner Name	u63417899
File Size	2MB
File Size (bytes)	1703936

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
14	AGE	Num	8	F4.		AGE
13	Education	Char	9	$CHAR9.		Education
16	JOB	Char	14	$CHAR14.		JOB
17	age_cat	Char	11			
10	balance	Num	8	BEST12.	BEST32.	
18	campaign_cat	Char	8			
2	contact	Char	9	$9.	$9.	contact
1	customer_id	Num	8	BEST12.	BEST32.	
3	day	Num	8	BEST.		day
9	default	Char	3	$3.	$3.	
11	housing	Char	3	$3.	$3.	
12	loan	Char	3	$3.	$3.	
15	marital	Char	8	$CHAR8.		marital
4	month	Char	3	$3.	$3.	month
5	pdays	Num	8	BEST.		pdays
7	poutcome	Char	7	$7.	$7.	poutcome
6	previous	Num	8	BEST.		previous
8	y	Char	3	$3.	$3.	y

Running PROC UNIVARIATE on balance

The UNIVARIATE Procedure
Variable: balance

Moments			
N	10578	Sum Weights	10578
Mean	1548.52978	Sum Observations	16380348
Std Deviation	3130.5653	Variance	9800439.07
Skewness	7.71681305	Kurtosis	119.649924
Uncorrected SS	1.29025E11	Corrected SS	1.03659E11
Coeff Variation	202.163713	Std Error Mean	30.4383415

Basic Statistical Measures			
Location		Variability	
Mean	1548.530	Std Deviation	3131
Median	566.000	Variance	9800439
Mode	0.000	Range	84262
		Interquartile Range	1640

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	50.87432	Pr > |t|	<.0001
Sign	M	4221.5	Pr >= |M|	<.0001
Signed Rank	S	22496590	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	81204
99%	13118

Quantiles (Definition 5)	
Level	Quantile
95%	6158
90%	3994
75% Q3	1765
50% Median	566
25% Q1	125
10%	0
5%	-76
1%	-542
0% Min	-3058

Extreme Observations					
Lowest			Highest		
Value	customer_id	Obs	Value	customer_id	Obs
-3058	132814	6058	45789	115970	2726
-1980	120418	3530	52587	140864	8152
-1944	135493	6753	52587	143154	9321
-1781	119684	3396	81204	142659	9037
-1668	115067	2532	81204	143494	9526

Running PROC UNIVARIATE on balance
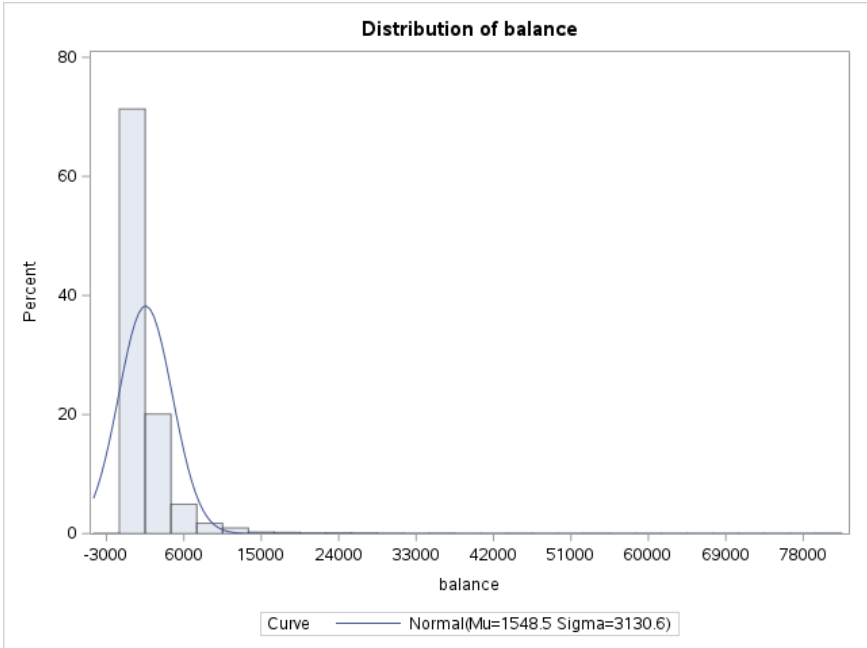
The UNIVARIATE Procedure

Distribution of balance

Running PROC UNIVARIATE on balance

The UNIVARIATE Procedure
Fitted Normal Distribution for balance

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	1548.53
Std Dev	Sigma	3130.565

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.25335	Pr > D	<0.010
Cramer-von Mises	W-Sq	260.44216	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	1347.95164	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	-542.0	-5734.254
5.0	-76.0	-3600.792
10.0	0.0	-2463.451
25.0	125.0	-563.004
50.0	566.0	1548.530
75.0	1765.0	3660.064

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
90.0	3994.0	5560.511
95.0	6158.0	6697.851
99.0	13118.0	8831.314
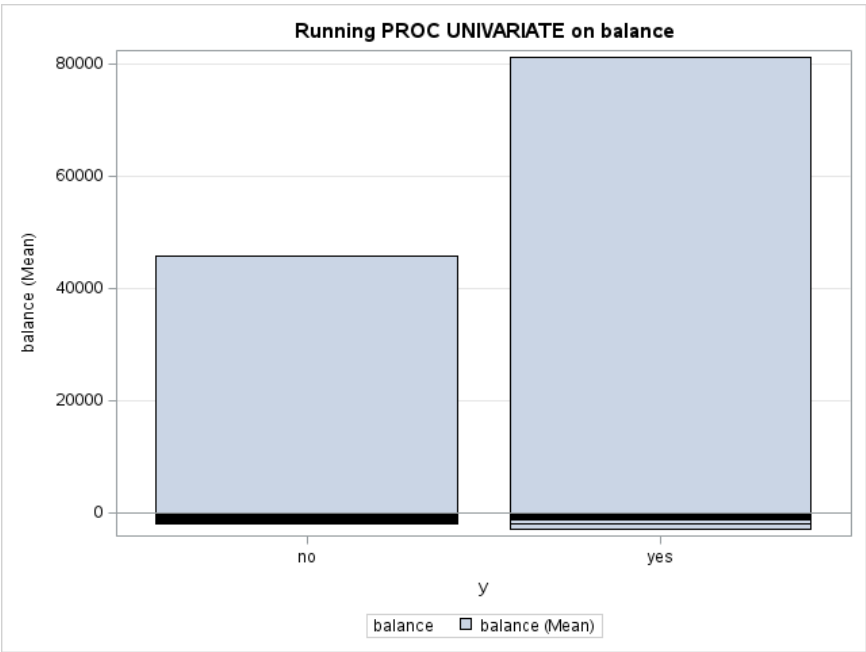
Running PROC UNIVARIATE on balance

distribution of balance by y

The UNIVARIATE Procedure

Distribution of balance
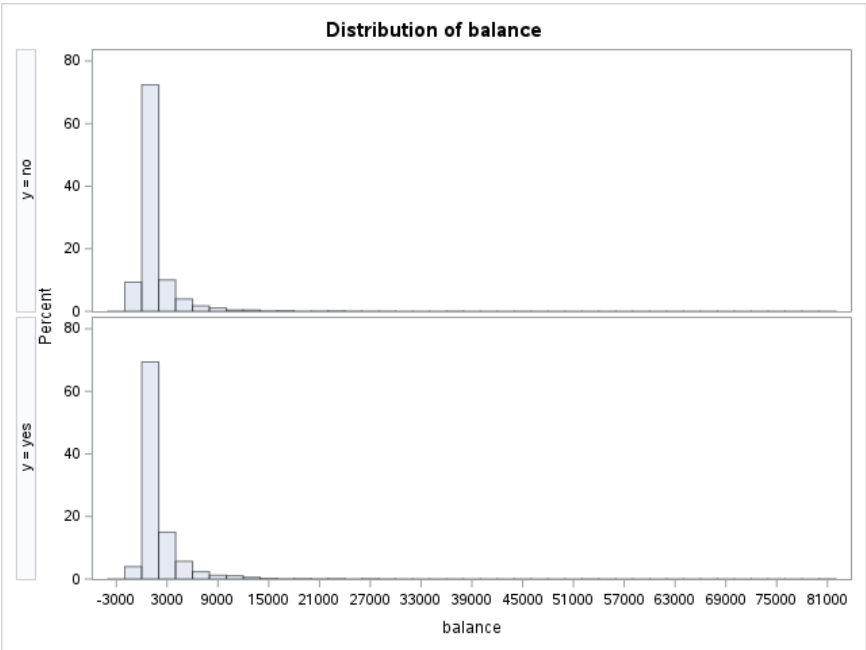
distribution of balance by y

The UNIVARIATE Procedure
Variable: pdays (pdays)

Moments			
N	10578	Sum Weights	10578
Mean	51.9548119	Sum Observations	549578
Std Deviation	109.347112	Variance	11956.791

Moments			
Skewness	2.41099367	Kurtosis	6.46379989
Uncorrected SS	155020200	Corrected SS	126466978
Coeff Variation	210.465804	Std Error Mean	1.06317691

Basic Statistical Measures			
Location		Variability	
Mean	51.95481	Std Deviation	109.34711
Median	-1.00000	Variance	11957
Mode	-1.00000	Range	855.00000
		Interquartile Range	50.00000

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	48.86751	Pr > |t|	<.0001
Sign	M	-2577	Pr >= |M|	<.0001
Signed Rank	S	-2988344	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	854
99%	430
95%	329
90%	192
75% Q3	49
50% Median	-1
25% Q1	-1
10%	-1
5%	-1
1%	-1
0% Min	-1

Extreme Observations					
Lowest			Highest		
Value	customer_id	Obs	Value	customer_id	Obs
-1	145308	10577	805	145221	10526
-1	145307	10576	828	144631	10217
-1	145306	10575	831	144799	10299
-1	145304	10573	842	144959	10396
-1	145303	10572	854	144930	10380

distribution of balance by y

The UNIVARIATE Procedure

Distribution of pdays
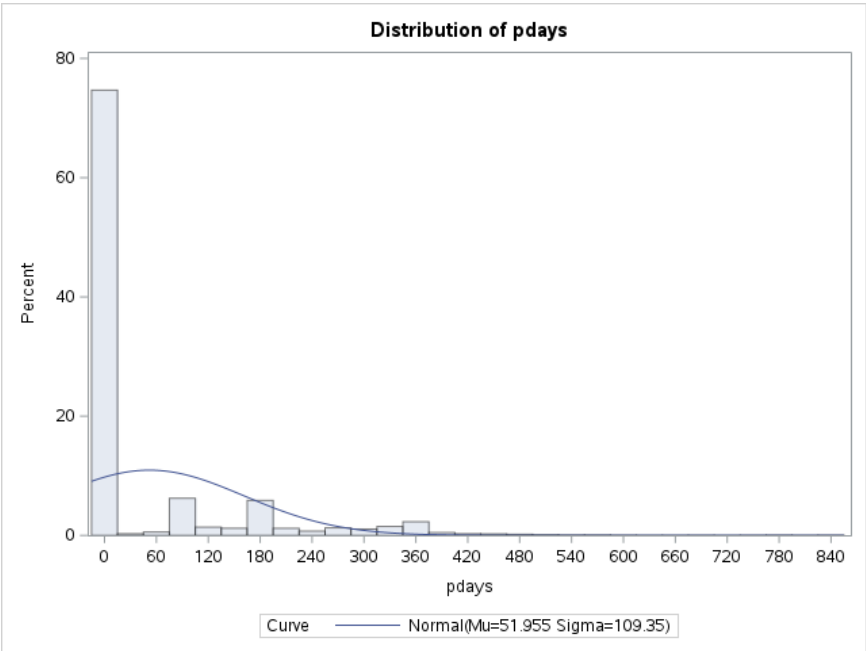
distribution of balance by y

The UNIVARIATE Procedure
Fitted Normal Distribution for pdays (pdays)

Parameters for Normal Distribution

Parameter	Symbol	Estimate
Mean	Mu	51.95481
Std Dev	Sigma	109.3471

Goodness-of-Fit Tests for Normal Distribution

Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.42953	Pr > D	<0.010
Cramer-von Mises	W-Sq	416.67946	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	2042.95965	Pr > A-Sq	<0.005

Quantiles for Normal Distribution

	Quantile	
Percent	Observed	Estimated
1.0	-1.00000	-202.4246
5.0	-1.00000	-127.9052
10.0	-1.00000	-88.1792
25.0	-1.00000	-21.7987
50.0	-1.00000	51.9548
75.0	49.00000	125.7083
90.0	192.00000	192.0888
95.0	329.00000	231.8148
99.0	430.00000	306.3342

distribution of balance by y

Obs	customer_id	contact	day	month	pdays	previous	poutcome	y	default	balance	housing	loan	Education	AGE	marital	JOB	age_cat	campaign_cat	customer_contacted
1	100103	unknown	5	may	-1	0	unknown	no	no	2	yes	yes	secondary	33	married	entrepreneur	young_adult	1	No
2	100106	unknown	5	may	-1	0	unknown	no	no	231	yes	no	tertiary	35	married	management	young_adult	1	No
3	100118	unknown	5	may	-1	0	unknown	no	no	52	yes	no	primary	57	married	blue-collar	old	1	No
4	100119	unknown	5	may	-1	0	unknown	no	no	60	yes	no	primary	60	married	retired	old	1	No
5	100121	unknown	5	may	-1	0	unknown	no	no	723	yes	yes	secondary	28	married	blue-collar	young_adult	1	No
6	100126	unknown	5	may	-1	0	unknown	no	no	-372	yes	no	secondary	44	married	admin.	middle_age	1	No
7	100130	unknown	5	may	-1	0	unknown	no	no	265	yes	yes	secondary	36	single	technician	middle_age	1	No
8	100141	unknown	5	may	-1	0	unknown	no	no	2586	yes	no	secondary	44	divorced	services	middle_age	1	No
9	100161	unknown	5	may	-1	0	unknown	no	no	0	yes	no	tertiary	32	married	admin.	young_adult	1	No
10	100168	unknown	5	may	-1	0	unknown	no	no	59	yes	no	tertiary	59	divorced	management	old	1	No
11	100172	unknown	5	may	-1	0	unknown	no	no	25	yes	yes	secondary	31	married	services	young_adult	1	No
12	100184	unknown	5	may	-1	0	unknown	yes	no	2343	yes	no	secondary	59	married	admin.	old	1	No
13	100187	unknown	5	may	-1	0	unknown	yes	no	45	no	no	secondary	56	married	admin.	old	1	No
14	100188	unknown	5	may	-1	0	unknown	yes	no	1270	yes	no	secondary	41	married	technician	middle_age	1	No
15	100189	unknown	5	may	-1	0	unknown	no	no	16	yes	yes	secondary	46	divorced	management	middle_age	2	No
16	100190	unknown	5	may	-1	0	unknown	no	no	486	yes	no	secondary	57	married	retired	old	2	No
17	100204	unknown	5	may	-1	0	unknown	no	no	179	yes	no	primary	59	married	blue-collar	old	3	No
18	100208	unknown	5	may	-1	0	unknown	no	no	61	no	yes	secondary	34	married	admin.	young_adult	1	No
19	100209	unknown	5	may	-1	0	unknown	no	no	30	yes	no	secondary	59	single	retired	old	1	No
20	100222	unknown	5	may	-1	0	unknown	no	no	26	yes	no	secondary	51	married	blue-collar	middle_age	1	No

distribution of balance by y

Obs	customer_id	contact	day	month	pdays	previous	poutcome	y	default	balance	housing	loan	Education	AGE	marital	JOB	age_cat	campaign_cat	customer_contacted
4210	124163	telephone	21	oct	166	1	other	yes	no	-247	yes	yes	secondary	42	single	admin.	middle_age	1	yes
4211	124165	telephone	21	oct	91	4	failure	yes	no	3444	yes	no	secondary	33	married	services	young_adult	1	yes
4218	124178	telephone	23	oct	143	3	failure	yes	no	0	yes	no	tertiary	36	married	management	middle_age	1	yes
4221	124181	unknown	23	oct	147	2	success	yes	no	589	yes	no	secondary	56	married	technician	old	1	yes
4262	124237	unknown	6	nov	101	11	other	no	no	1770	yes	no	tertiary	34	married	management	young_adult	1	yes

distribution of balance by y

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
customer_id		10578	127278.17	13660.22	100103.00	145309.00
day	day	10578	15.4758934	8.4137946	1.0000000	31.0000000
previous	previous	10578	0.8525241	3.4721156	0	275.0000000
balance		10578	1548.53	3130.57	-3058.00	81204.00
AGE	AGE	10558	41.2641599	12.1483452	18.0000000	146.0000000

Ten Highest and Lowest Values for HR

```
Ten Lowest Values
balance = -3058
```

```
balance = -1980
balance = -1944
balance = -1781
balance = -1668
balance = -1598
balance = -1455
balance = -1379
balance = -1350
balance = -1336
10 Highest Values
balance = 34646
balance = 36252
balance = 37378
balance = 37378
balance = 45248
balance = 45789
balance = 52587
balance = 52587
balance = 81204
balance = 81204
```