# SAS ASSIGNMENT 14

# (Summary Reports 1)

/\* Question 1 \*/

libname STA5066 "/courses/d649d56dba27fe300/STA5066";

proc freq data=STA5066.orders nlevels;

where Order\_Type = 1;

tables Customer\_ID Employee\_ID;

title "Unique Customers and Salespersons for Retail Sales";

run;

proc freq data=STA5066.orders nlevels;

where Order\_Type ne 1;

tables Customer\_ID;

title "Unique Customers for Catalog and Internet";

run;

/\* Question 2 \*/

libname STA5066 "/courses/d649d56dba27fe300/STA5066";

proc format;

value ordertypes 1='Retail' 2='Catalog' 3='Internet';

run;

proc freq data=STA5066.orders;

tables Order\_Date;

format Order\_Date YEAR4.;

title "Number of Orders in Each Year";

run;

proc freq data=STA5066.orders;

tables Order\_Type / nocum nopercent;

format Order\_Type ordertypes.;

title "Number of Orders by Order Type";

run;

proc freq data=STA5066.orders;

tables Order\_Date\*Order\_Type / nopercent;

format Order\_Date YEAR4. Order\_Type ordertypes.;

title "Number of Orders by Year and Order Type";

run;

/\* Question 3 \*/

libname sta5066 "/courses/d649d56dba27fe300/STA5066";

proc freq data=sta5066.order\_fact noprint;

tables Product\_ID / out=work.freqcount;

run;

data work.temp\_product\_list;

set sta5066.product\_list;

run;

proc sort data=work.freqcount;

by Product\_ID;

run;

proc sort data=work.temp\_product\_list;

by Product\_ID;

run;

data work.merged\_data;

merge work.freqcount(in=a) work.temp\_product\_list(in=b keep=Product\_ID Product\_Name);

by Product\_ID;

if a and b;

run;

proc sort data=work.merged\_data;

by descending Count;

run;

title "Top 10 Most Frequently Ordered Products";

proc print data=work.merged\_data(obs=10);

var Product\_ID Product\_Name Count;

run;

/\* Question 4 \*/

libname STA5066 "/courses/d649d56dba27fe300/STA5066";

proc format;

value ordertypes

1 = 'Retail'

2 = 'Catalog'

3 = 'Internet';

run;

title "Revenue (in U.S. Dollars) Earned from All Orders";

proc means data=sta5066.order\_fact noprint;

class Order\_Date Order\_Type;

var Total\_Retail\_Price;

format Order\_Date year4. Order\_Type ordertypes.;

output out=sales\_summary sum=Revenue;

run;

proc print data=sales\_summary noobs;

var Order\_Date Order\_Type Revenue;

format Order\_Date year4. Order\_Type ordertypes.;

run;

/\* Question 5 \*/

libname STA5066 "/courses/d649d56dba27fe300/STA5066";

title "Number of Missing and Non-Missing Date Values";

proc means data=STA5066.staff n nmiss;

class Gender;

var Birth\_Date Emp\_Hire\_Date Emp\_Term\_Date;

run;

/\* Question 6 \*/

libname STA5066 "/courses/d649d56dba27fe300/STA5066";

proc means data=STA5066.order\_fact sum noprint;

class Product\_ID;

var Total\_Retail\_Price;

output out=total\_revenue (drop=TYPE FREQ) sum=Total\_Revenue;

run;

data merged\_data;

merge total\_revenue (in=a) STA5066.product\_list (in=b);

by Product\_ID;

if a and b; /\* Keep only records where both datasets have matching Product\_ID \*/

run;

proc sort data=merged\_data;

by descending Total\_Revenue;

run;

title "Top 10 Products with the Highest Revenue";

proc print data=merged\_data (obs=10);

var Product\_ID Product\_Name Total\_Revenue;

run;

/\* Question 7 \*/

libname STA5066 "/courses/d649d56dba27fe300/STA5066";

data work.AnalysisTmp;

set STA5066.Analysis;

keep seqn dmaracer dmarethn dmaethnr hssex hsageir;

run;

title "Frequency of Demographic Variables";

proc freq data=work.AnalysisTmp;

tables dmaracer dmarethn hssex / nocum nopercent;

run;

title "Frequency for Female Participants under Age 50";

proc freq data=work.AnalysisTmp;

where hssex = 2 and hsageir < 50;

tables dmaracer dmarethn hssex / nocum nopercent;

run;

proc format;

value agef

low -< 45 = "<45"

45 - 59 = "45-59"

60 - high = "60+";

run;

title "Cross-tabulation of Demographic Variables by Age Groups";

proc freq data=work.AnalysisTmp;

tables dmaracer\*hsageir dmarethn\*hsageir hssex\*hsageir / nocum nopercent;

format hsageir agef.;

run;

/\* Question 8 \*/

libname STA5066 "/courses/d649d56dba27fe300/STA5066";

proc univariate data=sashelp.heart;

var cholesterol;

histogram cholesterol / normal;

inset mean="Mean" (5.1) std="Std Dev" (5.1) n="N" / position=ne;

run;

proc univariate data=sashelp.heart;

var cholesterol;

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

inset mean="Mean" (5.1) std="Std Dev" (5.1) n="N" / position=top;

run;