STAT 5193 SAS CHEAT SHEET
Complete Summary: Goad, 2020. SAS Programming for Elementary Statistics: Getting Started. ISBN-13: 987-1138589025, ISBN-10: 1138589020

Note: Chapters towards the end of book are not compiled in this cheat sheet.

PRACTICE ASSIGNMENTS
PRACTICE ASSIGNMENT 0:

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
TITLE; TITLE2;
FOOTNOTE; FOOTNOTE2;
DATA one;
INPUT Name $ Fine @@;
DATALINES;
Lynn 50 Evan 75 Thomas 24 Welsey 44 Marie 30
;
PROC PRINT DATA = one;
TITLE 'Objective 1.1: Outstanding Parking Fine';
PROC MEANS DATA = one;
VAR fine;
RUN;
QUIT;
```

PRACTICE ASSIGNMENT 1:

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;'; /* Display Manager
*/
* Note: SORT statement needed with BY statement but not
needed with CLASS statement.;
DATA prac2;
INPUT program $ gender $ time @@;
DATALINES;
R M 55.7 B M 64.3 B M 48.9 T F 67.2 T F 65.6 T F 60.5 R
M 72.0 R M 52.7 B F 83.5 B F 61.8 B F 66.4 R F 51.2
R F 49.6 T M 78.1 B M 54.2 T F 74.1 T M 71.2 B M 56.9 B
F 68.4 R M 50.9 R F 45.7 T M 69.9 B M 77.5 R F 53.4
;
PROC PRINT DATA = prac2;
TITLE "Practice 2, Data, Bijesh Mishra";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;

PROC UNIVARIATE DATA = prac2;
VAR time;
TITLE " Practice 2, Problem 1, Bijesh Mishra";
TITLE2 " Default summary statistics for time.";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;

PROC UNIVARIATE DATA = prac2;
CLASS program gender;
TITLE " Practice 2, Problem 1, Bijesh Mishra";
TITLE2 " Default summary statistics for entire data.";
FOOTNOTE "STAT 5193: SAS & R";
OUTPUT OUT = checkA MEAN = timeMean
MEDIAN = TimeMedian
VAR = TimeVariance STD = TimeSTD
MAX = TimeMax MIN = TimeMin;
RUN; QUIT;

PROC PRINT DATA = checkA;
TITLE "Practice 2, Problem 4i (Ext. 1), Bijesh Mishra";
TITLE2 "Print mean, median and variance for time
modifing 1.";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;

PROC SORT DATA = prac2;
BY gender;
ODS SELECT BASICINTERVALS;
RUN; QUIT;

/* Generates basic interval table only. */
PROC UNIVARIATE DATA = prac2 CIBASIC
(ALPHA = 0.02);
CLASS gender;
VAR time;
TITLE " Practice 2, Problem 2, Bijesh Mishra";
TITLE2 " Include 98% CI for mean for each gender class
using CLASS statement. Is sorting necessary?";
TITLE3 " Note: Sort before calculating CI.";
FOOTNOTE "STAT 5193: SAS & R";
OUTPUT OUT = checkB MEAN = timeMean
MEDIAN = TimeMedian
VAR = TimeVariance STD = TimeSTD
MAX = TimeMax MIN = TimeMin;
RUN; QUIT;

PROC PRINT DATA = checkB;
TITLE "Practice 2, Problem 4ii (Ext. 2), Bijesh Mishra";
TITLE2 "Print mean, median and variance for time
modifing 2.";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;

PROC SORT DATA = prac2;
BY gender;
ODS SELECT HISTOGRAM;
RUN; QUIT;

/* Generates histograms only */
```

```
PROC UNIVARIATE DATA = prac2;
HISTOGRAM Time / NORMAL(PERCENTS = 20 40 60 80
MIDPERCENTS) ODSTITLE = TITLE;
VAR time;
BY gender;
INSET N NORMAL (KSDPVAL) / POS = NE
FORMAT = 6.3;
TITLE " Practice 2, Problem 3, Bijesh Mishra";
TITLE2 " For each gender, include histogram for the time
variable, overlay normal curve.";
TITLE3 "Note: Sorting might be necessary if sorting is
not done previously.";
FOOTNOTE "STAT 5193: SAS & R";
OUTPUT OUT = checkC MEAN = timeMean
MEDIAN = TimeMedian
VAR = TimeVariance STD = TimeSTD
MAX = TimeMax MIN = TimeMin;
RUN; QUIT;

PROC PRINT DATA = checkB;
TITLE "Practice 2, Problem 4iii (Ext. 3), Bijesh
Mishra";
TITLE2 "Print mean, median and variance for time
modifing 3.";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;
```

PRACTICE ASSIGNMENT 2:

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;'; /* Display Manager
*/
* Note: SORT statement needed with BY statement but not
needed with CLASS statement;
DATA prac2;
INPUT program $ gender $ time @@;
DATALINES;
R M 55.7 B M 64.3 B M 48.9 T F 67.2 T F 65.6 T F 60.5 R
M 72.0 R M 52.7 B F 83.5 B F 61.8 B F 66.4 R F 51.2 R F
49.6 T M 78.1 B M 54.2 T F 74.1 T M 71.2 B M 56.9 B F
68.4 R M 50.9 R F 45.7 T M 69.9 B M 77.5 R F 53.4
;
PROC PRINT DATA = prac2;
TITLE "Practice 2, Data, Bijesh Mishra";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;

PROC UNIVARIATE DATA = prac2;
VAR time;
TITLE " Practice 2, Problem 1, Bijesh Mishra";
TITLE2 " Default summary statistics for time.";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;

PROC UNIVARIATE DATA = prac2;
CLASS program gender;
TITLE " Practice 2, Problem 1, Bijesh Mishra";
TITLE2 " Default summary statistics for entire data.";
FOOTNOTE "STAT 5193: SAS & R";
OUTPUT OUT = checkA MEAN = timeMean
MEDIAN = TimeMedian
VAR = TimeVariance STD = TimeSTD
MAX = TimeMax MIN = TimeMin;
RUN; QUIT;

PROC PRINT DATA = checkA;
TITLE "Practice 2, Problem 4i (Ext. 1), Bijesh Mishra";
TITLE2 "Print mean, median and variance for time
modifing 1.";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;

PROC SORT DATA = prac2;
BY gender;
ODS SELECT BASICINTERVALS;
RUN; QUIT;

/* Generates basic interval table only. */
PROC UNIVARIATE DATA = prac2 CIBASIC
(ALPHA = 0.02);
CLASS gender;
VAR time;
TITLE " Practice 2, Problem 2, Bijesh Mishra";
TITLE2 " Include 98% CI for mean for each gender class
using CLASS statement. Is sorting necessary?";
TITLE3 " Note: Sort before calculating CI.";
FOOTNOTE "STAT 5193: SAS & R";
OUTPUT OUT = checkB MEAN = timeMean
MEDIAN = TimeMedian
VAR = TimeVariance STD = TimeSTD
MAX = TimeMax MIN = TimeMin;
RUN; QUIT;

PROC PRINT DATA=checkB;
TITLE "Practice 2, Problem 4ii (Ext. 2), Bijesh Mishra";
TITLE2 "Print mean, median and variance for time
modifing 2.";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;

PROC SORT DATA = prac2;
BY gender;
```

```
ODS SELECT HISTOGRAM;
RUN; QUIT;

/* Generates histograms only */
PROC UNIVARIATE DATA = prac2;
HISTOGRAM Time / NORMAL(PERCENTS = 20 40 60 80
MIDPERCENTS) ODSTITLE = TITLE;
VAR time;
BY gender;
INSET N NORMAL (KSDPVAL) / POS = NE
FORMAT = 6.3;
TITLE " Practice 2, Problem 3, Bijesh Mishra";
TITLE2 " For each gender, include histogram for the time
variable, overlay normal curve.";
TITLE3 "Note: Sorting might be necessary if sorting is
not done previously.";
FOOTNOTE "STAT 5193: SAS & R";
OUTPUT OUT = checkC MEAN = timeMean
MEDIAN = TimeMedian
VAR = TimeVariance STD = TimeSTD
MAX = TimeMax MIN = TimeMin;
RUN; QUIT;

PROC PRINT DATA = checkB;
TITLE "Practice 2, Problem 4iii (Ext. 3), Bijesh
Mishra";
TITLE2 "Print mean, median and variance for time
modifing 3.";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;
```

PRACTICE ASSIGNMENT 3: NOT DONE.

PRACTICE ASSIGNMENT 4:

```
*** TEMPHUMIDWIND21.SAS *** ;
DM 'ODSRESULTS; CLEAR; LOG; CLEAR';
data meso;
input Month $ DAY MAXTEMP MINTEMP MAXHUMID MINHUMID
WINDDIR $ SPEEDAVG;
datalines;
 Jan        1        36          32          99           72
NNW          8.3
 Jan        2        45          27          98           61
SW           2.8
Apr       29        78          52          97           33
NNE          8.7
 Apr       30        81          48          98           23
SE           2.6
;
RUN; QUIT;

DM 'ODSRESULTS; CLEAR; LOG; CLEAR';
* input Month $  DAY    MAXTEMP    MINTEMP    MAXHUMID
MINHUMID    WINDDIR $   SPEEDAVG ;
PROC SORT DATA = meso; BY month;
PROC GCHART DATA = meso;
VBAR maxtemp / MIDPOINTS = 5 15 25 35 45 55 65 75 85;
BY month;
TITLE 'Practice 4, Q1(a)';
RUN; QUIT;

DM 'ODSRESULTS; CLEAR; LOG; CLEAR';
PROC GCHART DATA = meso;
VBAR maxtemp / MIDPOINTS = 5 15 25 35 45 55 65 75 85
AXIS = 0 TO 15;
BY month;
TITLE 'Practice 4, Q1(b)';
RUN; QUIT;

DM 'ODSRESULTS; CLEAR; LOG; CLEAR';
DATA  meso; SET meso;
davgtemp = MEAN (maxtemp, mintemp);
avghumid = MEAN(maxhumid, minhumid);
rh = avghumid/MEAN(avghumid)*100;
meandew = davgtemp - ((100 - rh)/5);
PROC GCHART DATA = meso;
HBAR month / TYPE = MEAN SUMVAR = meandew; * How to get
mean dew point temperature?;
BY month;
TITLE 'Practice 4, Q1(c)';
RUN; QUIT;
PROC MEANS DATA = meso;
CLASS month;
VAR meandew;
TITLE 'Practice 4, Q1(c) Verification';
RUN; QUIT;

DM 'ODSRESULTS; CLEAR; LOG; CLEAR';
PROC TTEST DATA = meso ALPH = 0.01 CIBASIC;
VAR davgtemp;
WHERE month = Jan;
TITLE 'Practice 4, Q2(a)';
RUN; QUIT;
```

PRACTICE ASSIGNMENT 5:

```
/* Practice 5 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
TITLE; FOOTNOTE;
TITLE "Assignment 4, Problem 1, Bijesh Mishra";
/* Data Year 2020 */
```

```
DATA y20;
INPUT Team $ 1-25 FG 26-30 FGA 31-35
P3 36-40 FT 41-45;
DATALINES;
Dallas Mavericks          3124 6772 1136 1392
…
Charlotte Hornets         2425 5586 785  1052
;

/* Data Year 2021 */
DATA y21;
INPUT Team $ 1-25 FG 26-30 FGA 31-35
P3 36-40 FT 41-45;
DATALINES;
Milwaukee Bucks           3221 6610 1038 1169
…
Cleveland Cavaliers       2778 6175 720  1200
;

/* Create Year and Conference Variables */
DATA y20;
SET y20;
year = 2020;
IF team = "Dallas Mavericks"
THEN conf = "west";
IF team = "Milwaukee Bucks"
THEN conf = "east";
IF team = "Portland Trail Blazers"
THEN conf = "west";
IF team = "Houston Rockets"
THEN conf = "west";
IF team = "Los Angeles Clippers"
THEN conf = "west";
IF team = "New Orleans Pelicans"
THEN conf = "west";
IF team = "Phoneix Suns"
THEN conf = "west";
IF team = "Washington Wizards"
THEN conf = "east";
IF team = "Memphis Grizzlies"
THEN conf = "west";
IF team = "Boston Celtics"
THEN conf = "east";
IF team = "Miami Heat"
THEN conf = "east";
IF team = "Denver Nuggets"
THEN conf = "west";
IF team = "Toronto Raptors"
THEN conf = "east";
IF team = "San Antonio Spurs"
THEN conf = "west";
IF team = "Philadelphia 76ers"
THEN conf = "east";
IF team = "Los Angeles Lakers"
THEN conf = "west";
IF team = "Brooklyn Nets"
THEN conf = "east";
IF team = "Utah Jazz"
THEN conf = "west";
IF team = "Indiana Pacers"
THEN conf = "east";
IF team = "Oklahoma City Thunder"
THEN conf = "west";
IF team = "Sacremento Kings"
THEN conf = "west";
IF team = "Orlando Magic"
THEN conf = "east";
IF team = "Atlanta Hawks"
THEN conf = "east";
IF team = "Minnesota Timberwolves"
THEN conf = "west";
IF team = "Detroit Pistons"
THEN conf = "east";
IF team = "New York Knicks"
THEN conf = "east";
IF team = "Cleveland Cavaliers"
THEN conf = "east";
IF team = "Chicago Bulls"
THEN conf = "east";
IF team = "Golden State Warriors"
THEN conf = "west";
IF team = "Charlotte Hornets"
THEN conf = "east";

/* Create Year and Conference Variables */
DATA y21;
SET y21;
year = 2021;
IF team = "Dallas Mavericks"
THEN conf = "west";
IF team = "Milwaukee Bucks"
THEN conf = "east";
IF team = "Portland Trail Blazers"
THEN conf = "west";
IF team = "Houston Rockets"
THEN conf = "west";
IF team = "Los Angeles Clippers"
THEN conf = "west";
IF team = "New Orleans Pelicans"
THEN conf = "west";
```

```
IF team = "Phoneix Suns"
THEN conf = "west";
IF team = "Washington Wizards"
THEN conf = "east";
IF team = "Memphis Grizzlies"
THEN conf = "west";
IF team = "Boston Celtics"
THEN conf = "east";
IF team = "Miami Heat"
THEN conf = "east";
IF team = "Denver Nuggets"
THEN conf = "west";
IF team = "Toronto Raptors"
THEN conf = "east";
IF team = "San Antonio Spurs"
THEN conf = "west";
IF team = "Philadelphia 76ers"
THEN conf = "east";
IF team = "Los Angeles Lakers"
THEN conf = "west";
IF team = "Brooklyn Nets"
THEN conf = "east";
IF team = "Utah Jazz"
THEN conf = "west";
IF team = "Indiana Pacers"
THEN conf = "east";
IF team = "Oklahoma City Thunder"
THEN conf = "west";
IF team = "Sacremento Kings"
THEN conf = "west";
IF team = "Orlando Magic"
THEN conf = "east";
IF team = "Atlanta Hawks"
THEN conf = "east";
IF team = "Minnesota Timberwolves"
THEN conf = "west";
IF team = "Detroit Pistons"
THEN conf = "east";
IF team = "New York Knicks"
THEN conf = "east";
IF team = "Cleveland Cavaliers"
THEN conf = "east";
IF team = "Chicago Bulls"
THEN conf = "east";
IF team = "Golden State Warriors"
THEN conf = "west";
IF team = "Charlotte Hornets"
THEN conf = "east";

/* Merge Dataset, create Points */
```

```
TITLE 'Practice 5, Q1(a)';
DATA combo; SET y20 y21;
points = sum((fg*2), ft, p3);
RUN; QUIT;
PROC PRINT DATA = combo; RUN; QUIT;
DM 'ODSRESULTS; CLEAR; ';
```

PRACTICE ASSIGNMENT 6: NOT DONE.
PRACTICE ASSIGNMENT 7: NOT DONE.
PRACTICE ASSIGNMENT 8: NOT DONE.
PRACTICE ASSIGNMENT 9: NOT DONE.
PRACTICE ASSIGNMENT 10: NOT DONE.
PRACTICE ASSIGNMENT 11: NOT DONE.
PRACTICE ASSIGNMENT 12: NOT DONE.
PRACTICE ASSIGNMENT 13: NOT DONE.
PRACTICE ASSIGNMENT 14: NOT DONE.

ASSIGNMENTS:
ASSIGNMENT 0:

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
TITLE; TITLE2;
FOOTNOTE; FOOTNOTE2;
DATA one;
INPUT Name $ Fine;
DATALINES;
Lynn 50
Evan 75
Thomas 24
Welsey 44
Marie 30
;
PROC PRINT DATA = one;
TITLE 'Objective 1.1: Outstanding Parking Fine';
PROC MEANS DATA = one;
VAR fine;
RUN;
QUIT;
```

ASSIGNMENT 1:

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';

TITLE "Assignment 1, Problem 1, Bijesh Mishra";
TITLE2 "Q: Enter data. ";
FOOTNOTE "STAT 5193: SAS & R";
DATA assign1;
```

```
INPUT Group Dosage $ Response @@;
DATALINES;
1 low 6.9 1 low 8.3 1 medium 23.5 1 medium 19.2
1 high 21.0 1 high 24.0 2 low 10.7 2 low 5.3
2 medium 8.9 2 medium 11.2 2 high 15.6 2 high 18.3
;
PROC PRINT DATA = assign1;
RUN; QUIT;

PROC SORT DATA = assign1;
BY group DESCENDING response;
PROC PRINT DATA = assign1 NOOBS;
VAR group response dosage;
TITLE "Assignment 1, Problem 2, Bijesh Mishra";
TITLE2 "Q: Sort data: descending response, ascending
group, noobs,";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;

PROC PRINT DATA = assign1;
VAR group response;
BY group;
TITLE "Assignment 1, Problem 3, Bijesh Mishra";
TITLE2 "Q: Use asorted data, print by group, supress
dose level.";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;

PROC SORT DATA = assign1;
BY dosage group;
PROC PRINT DATA = assign1;
VAR Dosage group response;
TITLE "Assignment 1, Problem 4, Bijesh Mishra";
TITLE2 "Q: Sort data by dosage level and group.";
FOOTNOTE "STAT 5193: SAS & R";
RUN; QUIT;
```

ASSIGNMENT 2:

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
TITLE; FOOTNOTE;
TITLE "Assignment 2, Problem 1, Bijesh Mishra";
DATA y20;
INPUT Team $ 1-25 FG 26-30 FGA 31-35
P3 36-40 FT 41-45;
DATALINES;
Dallas Mavericks          3124 6772 1136 1392
…
Charlotte Hornets         2425 5586 785  1052
;
TITLE "Assignment 2, Problem 1, Bijesh Mishra";
DATA y21;
INPUT Team $ 1-25 FG 26-30 FGA 31-35 P3 36-40 FT 41-45;
DATALINES;
Milwaukee Bucks           3221 6610 1038 1169
…
Cleveland Cavaliers       2778 6175 720  1200
;
TITLE "Assignment 2, Problem 2, Bijesh Mishra";
PROC SORT DATA = y20;
BY DESCENDING ft;
PROC print DATA = y20;
RUN; QUIT;

TITLE "Assignment 2, Problem 2, Bijesh Mishra";
PROC SORT DATA = y21;
BY DESCENDING ft;
PROC print DATA = y21;
RUN; QUIT;

DATA bbcombo;
INPUT Team $ 1-25 FG 26-30 FGA 31-35
P3 36-40 FT 41-45;
DATALINES;
Dallas Mavericks          3124 6772 1136 1392
…
Cleveland Cavaliers       2778 6175 720  1200
;
TITLE "Assignment 2, Problem 3, Bijesh Mishra";
PROC SORT DATA = bbcombo;
BY fg;
PROC PRINT DATA = bbcombo NOOBS;
VAR Team fg;
RUN; QUIT;

TITLE "Assignment 2, Problem 4, Bijesh Mishra";
PROC UNIVARIATE DATA = bbcombo CIBASIC (ALPHA =
0.08)NORMAL;
VAR ft fg;
HISTOGRAM fg/NORMAL;
OUTPUT OUT = summary MEAN = meanFT meanFG MEDIAN =
medianFT medianFG;
PROC PRINT DATA = summary NOOBS;
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
PROC SORT DATA = bbcombo;
BY team;
PROC PRINT DATA = bbcombo NOOBS;
RUN; QUIT;
```

ASSIGNMENT 3:

```
*** TEMPHUMIDWIND21.SAS *** ;
data mesonet21 ;
input Month $  DAY    MAXTEMP    MINTEMP    MAXHUMID
MINHUMID    WINDDIR $   SPEEDAVG ;
datalines;
 Jan       1       36         32         99          72
NNW          8.3
…
Apr      30       81         48         98          23
SE           2.6
;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE "Assignment 3, Problem 1, Bijesh Mishra";

PROC MEANS DATA = mesonet21;
VAR MAXTEMP MINTEMP MAXHUMID MINHUMID SPEEDAVG;
OUTPUT OUT = MesoSum;
PROC PRINT DATA = mesoSum;
RUN; QUIT;

TITLE "Assignment 3, Problem 2, Bijesh Mishra";
PROC SORT DATA = mesonet21;
BY month;
PROC MEANS DATA = mesonet21 MEAN RANGE CV STD;
BY month;
VAR MAXHUMID;
OUTPUT OUT = humid MEAN = HDMaxMean
CV = HDMaxCV;
PROC PRINT DATA = humid;
RUN; QUIT;

TITLE "Assignment 3, Problem 3, Bijesh Mishra";
PROC MEANS DATA = mesonet21 MIN MAX MEAN;
CLASS month;
VAR speedavg;
OUTPUT OUT = speed MIN = SpeedMin
MAX = SpeedMax;
PROC PRINT DATA = speed;
RUN; QUIT;

TITLE "Assignment 3, Problem 4, Bijesh Mishra";
DATA prob4; SET mesonet21;
avgdaytemp1 = MEAN(maxtemp, mintemp); * First Way (4.1);
avgdaytemp2 = (maxtemp + mintemp)/2; * Second Way (4.2);
PROC PRINT DATA = prob4;
WHERE Month = "Mar";
RUN; QUIT;
```

ASSIGNMENT 4:

```
/* Assignment 4 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
TITLE; FOOTNOTE;
TITLE "Assignment 4, Problem 1, Bijesh Mishra";
/* Data Year 2020 */
DATA y20;
INPUT Team $ 1-25 FG 26-30 FGA 31-35 P3 36-40 FT 41-45;
DATALINES;
Dallas Mavericks          3124 6772 1136 1392
…
Charlotte Hornets         2425 5586 785  1052
;

/* Data Year 2021 */
DATA y21;
INPUT Team $ 1-25 FG 26-30 FGA 31-35 P3 36-40 FT 41-45;
DATALINES;
Milwaukee Bucks           3221 6610 1038 1169
…
Cleveland Cavaliers       2778 6175 720  1200
;

/* Create Year and Conference Variables */
DATA y20;
SET y20;
year = 2020;
IF team = "Dallas Mavericks"
THEN conf = "west";
IF team = "Milwaukee Bucks"
THEN conf = "east";
IF team = "Portland Trail Blazers"
THEN conf = "west";
IF team = "Houston Rockets"
THEN conf = "west";
IF team = "Los Angeles Clippers"
THEN conf = "west";
IF team = "New Orleans Pelicans"
THEN conf = "west";
IF team = "Phoneix Suns"
THEN conf = "west";
IF team = "Washington Wizards"
THEN conf = "east";
IF team = "Memphis Grizzlies"
THEN conf = "west";
IF team = "Boston Celtics"
THEN conf = "east";
IF team = "Miami Heat"
THEN conf = "east";
```

```
IF team = "Denver Nuggets"
THEN conf = "west";
IF team = "Toronto Raptors"
THEN conf = "east";
IF team = "San Antonio Spurs"
THEN conf = "west";
IF team = "Philadelphia 76ers"
THEN conf = "east";
IF team = "Los Angeles Lakers"
THEN conf = "west";
IF team = "Brooklyn Nets"
THEN conf = "east";
IF team = "Utah Jazz"
THEN conf = "west";
IF team = "Indiana Pacers"
THEN conf = "east";
IF team = "Oklahoma City Thunder"
THEN conf = "west";
IF team = "Sacremento Kings"
THEN conf = "west";
IF team = "Orlando Magic"
THEN conf = "east";
IF team = "Atlanta Hawks"
THEN conf = "east";
IF team = "Minnesota Timberwolves"
THEN conf = "west";
IF team = "Detroit Pistons"
THEN conf = "east";
IF team = "New York Knicks"
THEN conf = "east";
IF team = "Cleveland Cavaliers"
THEN conf = "east";
IF team = "Chicago Bulls"
THEN conf = "east";
IF team = "Golden State Warriors"
THEN conf = "west";
IF team = "Charlotte Hornets"
THEN conf = "east";

/* Create Year and Conference Variables */
DATA y21;
SET y21;
year = 2021;
IF team = "Dallas Mavericks"
THEN conf = "west";
IF team = "Milwaukee Bucks"
THEN conf = "east";
IF team = "Portland Trail Blazers"
THEN conf = "west";
IF team = "Houston Rockets"
THEN conf = "west";
IF team = "Los Angeles Clippers"
THEN conf = "west";
IF team = "New Orleans Pelicans"
THEN conf = "west";
IF team = "Phoneix Suns"
THEN conf = "west";
IF team = "Washington Wizards"
THEN conf = "east";
IF team = "Memphis Grizzlies"
THEN conf = "west";
IF team = "Boston Celtics"
THEN conf = "east";
IF team = "Miami Heat"
THEN conf = "east";
IF team = "Denver Nuggets"
THEN conf = "west";
IF team = "Toronto Raptors"
THEN conf = "east";
IF team = "San Antonio Spurs"
THEN conf = "west";
IF team = "Philadelphia 76ers"
THEN conf = "east";
IF team = "Los Angeles Lakers"
THEN conf = "west";
IF team = "Brooklyn Nets"
THEN conf = "east";
IF team = "Utah Jazz"
THEN conf = "west";
IF team = "Indiana Pacers"
THEN conf = "east";
IF team = "Oklahoma City Thunder"
THEN conf = "west";
IF team = "Sacremento Kings"
THEN conf = "west";
IF team = "Orlando Magic"
THEN conf = "east";
IF team = "Atlanta Hawks"
THEN conf = "east";
IF team = "Minnesota Timberwolves"
THEN conf = "west";
IF team = "Detroit Pistons"
THEN conf = "east";
IF team = "New York Knicks"
THEN conf = "east";
IF team = "Cleveland Cavaliers"
THEN conf = "east";
IF team = "Chicago Bulls"
THEN conf = "east";
IF team = "Golden State Warriors"
```

```
THEN conf = "west";
IF team = "Charlotte Hornets"
THEN conf = "east";

/* Merge Dataset */
DATA combo;
SET Y20 Y21;
RUN; QUIT;

/* Assignment 4, Problem 1(A) */
PROC SORT DATA = combo; BY year;
PROC GCHART DATA = combo;
TITLE "Assignment 4, Problem 1(A), Bijesh Mishra";
VBAR fg/ AXIS = 0 TO 9 MIDPOINTS = (2400 2500 2600 2700
2800 2900 3000 3100 3200);
BY year;
RUN; QUIT;

/* Assignment 4, Problem 1(B) */
TITLE "Assignment 4, Problem 1(B), Bijesh Mishra";
PROC GCHART DATA = combo;
VBAR p3/ MIDPOINTS = (700 800 900 1000 1100 1200) GROUP
= Year AXIS = (1 TO 12);
RUN; QUIT;
/* Assignment 4, Problem 1(C) */
TITLE "Assignment 4, Problem 1(C), Bijesh Mishra";
PROC GCHART DATA = combo;
VBAR year / SUMVAR = P3 TYPE = MEAN MIDPOINTS = (2020
2021);
RUN; QUIT;

/* Assignment 4, Problem 2(A) */
TITLE "Assignment 4, Problem 2(A), Bijesh Mishra";
PROC TTEST DATA = combo ALPHA = 0.04
PLOTS = (HISTOGRAM INTERVAL BOXPLOT);
CLASS conf;
VAR ft;
BY year;
RUN; QUIT;

/* Assignment 4, Problem 2(B) */
TITLE "Assignment 4, Problem 2(B), Bijesh Mishra";
PROC TTEST DATA = combo ALPHA = 0.01
PLOTS (ONLY) = (BOXPLOT INTERVAL PROFILES);
PAIRED fg*ft;
WHERE year = 2020 and conf = 'west';
RUN; QUIT;
```

ASSIGNMENT 5:

```
/* Assignment 5 Problem 1*/
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE "Bijesh Mishra, Assignment 5, Problem 1 ";
data stn;
input Stations $ RspMin @@ ;
datalines;
s1 15 s2 13 s3 17 s4 10 s1 17 s2 9  s3 21 s4 8
s1 22 s2 14 s3 15 s4 11 s1 19 s2 11 s3 23 s4 11
s1 25 s2 15 s3 18 s4 12
;

TITLE "Bijesh Mishra, Assignment 5, Problem 1(a) ";
/* Kurskal-Wallis or Wilcoxon Rank Sum Test*/
PROC NPAR1WAY DATA = stn WILCOXON
ALPHA = 0.05 PLOTS = NONE;
CLASS Stations;
VAR rspmin;
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 5, Problem 1(b) ";
PROC GLM DATA = stn ALPHA = 0.05 PLOTS (ONLY) =
(BOXPLOT);
CLASS stations;
MODEL rspmin = stations ;
MEANS stations / CLM LSD PLOTS = NONE;
RUN; QUIT;

      /* Assignment 5 Problem 2*/
      *** TEMPHUMIDWIND21.SAS *** ;
      DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
      TITLE "Bijesh Mishra, Assignment 5, Problem 2 ";
      data meso ;
      input Month $  DAY    MAXTEMP    MINTEMP
MAXHUMID    MINHUMID    WINDDIR $   SPEEDAVG ;
      datalines;
       Jan       1        36         32         99
72         NNW         8.3
…
       Apr      30        81         48         98
23         SE          2.6
      ;
      RUN; QUIT;
      * input Month $  DAY    MAXTEMP    MINTEMP
MAXHUMID    MINHUMID    WINDDIR $   SPEEDAVG ;
PROC RANK DATA = meso OUT = mesoA
TIES = LOW;
VAR mintemp;
RANKS rmintemp;
TITLE "Bijesh Mishra, Assignment 5, Problem 2";
TITLE2 ' Minimum Temperature ';
RUN; QUIT;
```

```
/* PROC SORT DATA = mesoA; BY mintemp; */
PROC PRINT DATA = mesoA;
VAR month day mintemp rmintemp;
WHERE rmintemp < 11 & mintemp ne .;
RUN; QUIT;

PROC RANK DATA = meso OUT = mesoB
TIES = LOW;
VAR maxhumid;
RANKS rmaxhumid;
TITLE "Bijesh Mishra, Assignment 5, Problem 2";
TITLE2 ' Maximum Humidity ';
RUN; QUIT;

/* PROC SORT DATA = mesoB; BY maxhumid; */
PROC PRINT DATA = mesoB;
VAR  month day maxhumid rmaxhumid;
WHERE rmaxhumid < 11 & maxhumid ne .;
RUN; QUIT;

PROC RANK DATA = meso OUT = mesoC
TIES = HIGH;
VAR mintemp;
RANKS rmintemp;
TITLE "Bijesh Mishra, Assignment 5, Problem 2";
TITLE2 ' Minimum Temperature ';
RUN; QUIT;

/* PROC SORT DATA = mesoC; BY mintemp; */
PROC PRINT DATA = mesoC;
VAR month day mintemp rmintemp;
WHERE rmintemp < 11 & mintemp ne .;
RUN; QUIT;

PROC RANK DATA = meso OUT = mesoD
TIES = HIGH;
VAR maxhumid;
RANKS rmaxhumid;
TITLE "Bijesh Mishra, Assignment 5, Problem 2";
TITLE2 ' Maximum Humidity ';
RUN; QUIT;

/* PROC SORT DATA = mesoD; BY maxhumid; */
PROC PRINT DATA = mesoD;
VAR  month day maxhumid rmaxhumid;
WHERE rmaxhumid < 15 & maxhumid ne .;
RUN; QUIT;
```

ASSIGNMENT 6 & 7:

```
/* Assign 6 & 7 */
DM ' LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE; FOOTNOTE;
DATA arch;
INFILE 'C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\Data\architecture.txt' DLM =""
FIRSTOBS = 2;
INPUT SUBJ BLDG SAT BTY FNC INT DIG CST FSH;
LABEL SUBJ = " Individuals "
        BLDG = " Building Structures "
      SAT  = " Overall "
      BTY  = " Beauty "
      FNC  = " Function "
      INT  = " Intimacy "
      DIG  = " Dignity "
      CST  = " Cost "
      FSH  = " Fashion ";
TITLE "Bijesh Mishra, Assignment 6 & 7, Problem 1";
PROC SORT DATA = arch; BY BLDG;
PROC FREQ DATA = arch;
TABLES CST / PLOTS = FREQPLOT NOCOL;
BY BLDG;
TITLE "Bijesh Mishra, Assignment 6 & 7, Problem 2";
RUN; QUIT;

PROC FREQ DATA = arch;
TABLES BLDG*CST / PLOTS = FREQPLOT NOPERCENT NOCOL CHISQ
ALPHA = 0.05;
TITLE "Bijesh Mishra, Assignment 6 & 7, Problem 3";
RUN; QUIT;

/* Assignment 6 & 7 Q4 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
TITLE; FOOTNOTE;
TITLE "Bijesh Mishra, Assignment 6 & 7, Problem 4";

Data Used: Previous Basketball Data with Year and
Conference information.

/* Merge Dataset */
DATA combo;
SET Y20 Y21;
RUN; QUIT;
* PROC SORT DATA = combo;
* BY TEAM;
* PROC PRINT DATA = combo;
* RUN; * QUIT;
* INPUT Team $ 1-25 FG 26-30 FGA 31-35 P3 36-40 FT 41-45
conf;
```

```
TITLE "Bijesh Mishra, Assignment 6 & 7, Problem 4 (a)";
PROC SORT DATA = combo; BY TEAM;
PROC REPORT DATA = combo;
COLUMN Team fg p3 ft;
BY TEAM;
DEFINE Team/ GROUP;
DEFINE fg / 'Field Goals';
DEFINE p3 / '3 Point Field Goals';
DEFINE ft / 'Free Throws';
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 6 & 7, Problem 4 (b)";
PROC REPORT DATA = combo;
COLUMN Team Year fg p3 ft;
DEFINE Team / GROUP;
DEFINE YEAR / GROUP;
DEFINE fg / 'Field Goals';
DEFINE p3 / '3 Point Field Goals';
DEFINE ft / 'Free Throws';
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 6 & 7, Problem 4 (c)";
PROC REPORT DATA = combo;
COLUMN  Year Team fg p3 ft;
DEFINE Team / GROUP;
DEFINE YEAR / GROUP;
DEFINE fg / 'Field Goals';
DEFINE p3 / '3 Point Field Goals';
DEFINE ft / 'Free Throws';
RUN; QUIT;
```

ASSIGNMENT 8:

```
      /* Assignment 8 Problem 1*/
      *** TEMPHUMIDWIND21.SAS *** ;
      DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
      TITLE "Bijesh Mishra, Assignment 8, Problem 1 ";
      data meso ;
      input Month $  DAY    MAXTEMP    MINTEMP
MAXHUMID    MINHUMID    WINDDIR $   SPEEDAVG ;
      datalines;
       Jan       1       36         32         99
72         NNW         8.3
       Apr      30       81         48         98
23         SE          2.6
      ;
      RUN; QUIT;
PROC CORR DATA = meso PLOTS = MATRIX;
VAR mintemp maxtemp ;
WITH minhumid maxhumid;
BY month;
WHERE month EQ "Feb" | month EQ "Mar";
TITLE "Bijesh Mishra, Assignment 8, Problem 1 ";
RUN; QUIT;

/* Assignment 8 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
TITLE; FOOTNOTE;
TITLE "Assignment 8, Problem 2, Bijesh Mishra";
/* Data Year 2020 */
DATA y20;
INPUT Team $ 1-25 FG 26-30 FGA 31-35 P3 36-40 FT 41-45;
Year = 2020;
DATALINES;
Dallas Mavericks          3124 6772 1136 1392
Charlotte Hornets         2425 5586 785  1052
;

/* Data Year 2021 */
DATA y21;
INPUT Team $ 1-25 FG 26-30 FGA 31-35 P3 36-40 FT 41-45;
Year = 2021;
DATALINES;
Milwaukee Bucks           3221 6610 1038 1169
Cleveland Cavaliers       2778 6175 720  1200
;

DATA hw8;
INPUT fg Year @@;
DATALINES;
7000 2021 7100 2021
;
RUN; QUIT;

/* Merge Dataset */
DATA combo; SET HW8 y20 y21;
Points = 2*(fg - p3) + 3*p3 +ft;
RUN; QUIT;

/* Assignment 8, Problem 2 */
TITLE "Assignment 8, Problem 2, Bijesh Mishra";
PROC REG DATA = combo SIMPLE;
MODEL Points = fga / CLI CLM CLB ALPHA = 0.02;
WHERE year = 2021;
ID fga;
Q4B: TEST fga = 0;
RUN; QUIT;
```

ASSIGNMENT 9:

```
      /* Assignment 9*/
```

```
      *** TEMPHUMIDWIND21.SAS *** ;
      DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
      TITLE "Bijesh Mishra, Assignment 9, Problem 1 ";
      data meso ;
      input Month $  DAY    MAXTEMP    MINTEMP
MAXHUMID    MINHUMID    WINDDIR $   SPEEDAVG ;
      datalines;
       Jan       1       36         32         99
72          NNW          8.3
       Apr      30       81         48         98
23          SE           2.6
      ;
      RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 9, Problem 1(a) ";
LIBNAME ClassHw " C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Assignments\Assignment 9";
DATA CLASSHW.MESONET2021;
SET meso;
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 9, Problem 1(b) ";
PROC SORT DATA = ClassHw.MesoNet2021;
BY month;
PROC REPORT DATA = ClassHw.MesoNet2021;
COLUMN month maxtemp maxhumid;
BY month;
DEFINE maxtemp / "Maximum Temperature" MEAN;
DEFINE maxhumid / "Maximum Humidity" MEAN;
DEFINE Month / GROUP;
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 9, Problem 2(a) ";
DATA CLASSHW.ARCH;
FILENAME arch "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Data\architecture.txt";
INFILE Arch FIRSTOBS = 2;
INPUT SUBJ BLDG SAT BTY FNC INT DIG CST FSH;
LABEL SUBJ = " Individuals "
      BLDG = " Building Structures "
      SAT = " Overall "
      BTY = " Beauty "
      FNC = " Function "
      INT = " Intimacy "
      DIG = " Dignity "
      CST = " Cost "
      FSH = " Fashion ";
RUN; QUIT;
* PROC PRINT DATA = CLASSHW.ARCH (OBS = 10) NOOBS LABEL;
* RUN; * QUIT;

TITLE "Bijesh Mishra, Assignment 9, Problem 2(b) ";
PROC GCHART DATA = ClassHw.Arch;
PIE int / TYPE = PERCENT;
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 9, Problem 3 ";
DATA CIINT;
Mean = 94.22;
StdErr = 6.90;
ssize = 18;
t = TINV(0.97, 17);
t1 = TINV((1 - 0.025), 1000); /* = 1.96; Gives
Probability */
cilower = mean - t*StdErr;
ciupper = mean + t*StdErr;
f = 1- PROBF (4.47, 4, 14); /* Gives P-value Or
Significance Level */
LABEL Mean    = "Mean"
      cilower = "CI Lower Limit"
        stderr = "Standard Error"
        ssize = "Sample Size"
        ciupper = "CI Upper Limit"
        t = "T-Critical"
        t1 = "TINV (0.025, 1000)"
        f = "F-Dist. P-value (Q3b)";
;
PROC PRINT DATA = CIINT LABEL NOOBS;
VAR Mean StdErr ssize t cilower ciupper f;
RUN; QUIT;
```

ASSIGNMENT 10:

```
/* Assignment 10 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE "Bijesh Mishra, Assignment 10, Problem 1 ";
/*
PROC IMPORT OUT = WORK.Jan2021
DATAFILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Data\Assign10Data.xlsx"
DBMS = EXCEL REPLACE ;
SHEET = "Jan2021";
GETNAMES = YES;
RUN; QUIT;
*/

PROC IMPORT OUT = WORK.Jan2021
DATAFILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Data\Jan2021.csv"
DBMS = CSV REPLACE;
DATA WORK.Jan2021;
```

```
SET WORK.Jan2021;
Month = "Jan";
PROC IMPORT OUT = WORK.Feb2021
DATAFILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Data\Feb2021.csv"
DBMS = CSV REPLACE;
DATA WORK.Feb2021;
SET WORK.Feb2021;
Month = "Feb";
PROC IMPORT OUT = WORK.March2021
DATAFILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Data\March2021.csv"
DBMS = CSV REPLACE;
DATA WORK.March2021;
SET WORK.March2021;
Month = "Mar";
PROC IMPORT OUT = WORK.April2021
DATAFILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Data\April2021.csv"
DBMS = CSV REPLACE;
DATA WORK.April2021;
SET WORK.April2021;
Month = "Apr";
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 10, Problem 2 ";
LIBNAME CLASS "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Data";
DATA CLASS.MESO2021_2;
SET WORK.Jan2021 WORK.Feb2021 WORK.March2021
WORK.April2021;
LABEL Month = "Month"
      DAY = "Day"
      RAIN_IN = "Rain (Inch/Day)"
        STNPRESSUR = " Hg Pressure (Inch)"
        SOLAR_MJ_M2 = " Solar Radiation (Mj/MSq)"
        SODSOIL = " Avg Soil Temp.(F)/Day"
        BARESOIL = " Avg BareSoil Temp.(F)/Day";
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 10, Problem 3 ";
*** TEMPHUMIDWIND21.SAS *** ;
data mesonet21 ;
INPUT Month $  DAY    MAXTEMP    MINTEMP    MAXHUMID
MINHUMID    WINDDIR $   SPEEDAVG ;
LABEL Month = "Month"
      DAY   = "Day"
        MAXTEMP = "Maximum Temperature"
        MINTEMP = "Minimum Temperature"
        MAXHUMID = "Maximum Humidity"
        MINHUMID = "Minimum Humidity"
        WINDDIR = "Wind Direction"
      SPEEDAVG = "Wind Average Speed";
datalines;
 Jan        1        36          32          99           72
NNW          8.3
Apr       30        81          48          98           23
SE           2.6
;

TITLE "Bijesh Mishra, Assignment 10, Problem 4";
PROC SORT DATA = CLASS.MESO2021_2; BY Month;
PROC SORT DATA = work.mesonet21; BY Month;
DATA CLASS.MESONET1_4;
MERGE CLASS.MESO2021_2 work.mesonet21;
BY Month Day;

PROC PRINT data = CLASS.MESONET1_4 LABEL;
VAR MaxTemp MinTemp MaxHumid MinHumid SpeedAvg Rain_In
StnPressur Solar_Mj_M2 SodSoil BareSoil;
ID Month Day;
WHERE Month = "Mar";
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 10, Problem 5 (a)";
PROC MEANS DATA = CLASS.MESONET1_4;
CLASS Month;
VAR MaxTemp MinTemp MaxHumid MinHumid SpeedAvg Rain_In
StnPressur Solar_Mj_M2 SodSoil BareSoil;
OUTPUT OUT = work.sumstats MEAN = MeanMaxTemp
MeanMinTemp MeanMaxHumid MeanMinHumid MeanSpeedAvg

MeanRain_In MeanStnPressur MeanSolar_Mj_M2 MeanSodSoil
MeanBareSoil;

TITLE "Bijesh Mishra, Assignment 10, Problem 5 (b)";
/*
PROC EXPORT DATA = work.sumstats
OUTFILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Data\sumstats_ASSIGN10DATA.XLSX"
DBMS = EXCEL;
SHEET =  "WORK.SUMSTAT";
RUN; QUIT;
*/

PROC EXPORT DATA = work.sumstats
OUTFILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Assignments\Assignment
10\sumstats_ASSIGN10DATA.csv"
DBMS = csv REPLACE;
```

```
RUN; QUIT;

/* Export Permanent dataset to Assignment 11 */
LIBNAME Assgn11 " C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Assignments\Assignment 11";
DATA Assgn11.MESONET1_4;
SET CLASS.MESONET1_4;
RUN; QUIT;
```

ASSIGNMENT 11:

```
/* Assignment 11 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
/* Assignment 11, Problem 1 */
/* Export Permanent dataset to Assignment 11 */
LIBNAME Assgn11 " C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Assignments\Assignment 11";
TITLE "Bijesh Mishra, Assignment 11, Problem 1 ";
DATA Assgn11.A11;
SET Assgn11.MESONET1_4;
ARRAY b{4} MaxTemp MinTemp SODSOIL BARESOIL;
ARRAY c{4} MaxTempC MinTempC SODSOILC BARESOILC;
Do i = 1 TO 4;
c{i} = (b {i} - 32)* (5/9);
END;
DROP MaxTemp MinTemp SODSOIL BARESOIL;
LABEL Month = "Month"
      DAY   = "Day"
        MAXTEMP = "Maximum Temperature (F)"
        MINTEMP = "Minimum Temperature (F)"
        MAXHUMID = "Maximum Humidity"
        MINHUMID = "Minimum Humidity"
        WINDDIR = "Wind Direction"
      SPEEDAVG = "Wind Average Speed"
      RAIN_IN = "Rain (Inch/Day)"
        STNPRESSUR = " Hg Pressure (Inch)"
        SOLAR_MJ_M2 = " Solar Radiation (Mj/MSq)"
        SODSOIL = " Avg Soil Temp.(F)/Day"
        BARESOIL = " Avg BareSoil Temp.(F)/Day"
        MAXTEMPC = "Maximum Temperature (C)"
        MINTEMPC = "Minimum Temperature (C)"
        SODSOILC = " Avg Soil Temp.(C)/Day"
        BARESOILC = " Avg BareSoil Temp.(C)/Day";
RUN; QUIT;
PROC PRINT DATA = Assgn11.A11 LABEL NOOBS;
ID month Day;
RUN; QUIT;

/* Assignment 11, Problem 2 */
TITLE "Bijesh Mishra, Assignment 11, Problem 2";
DATA A11Q2;
seed = 1234;
DO i = 1 TO 15; * Size 15;
DO J = 1 TO 10; * Sample 10;
CALL RANUNI (seed, x);
y = x*10 + 10;
OUTPUT;
END; END;
RUN; QUIT;

PROC MEANS DATA = A11Q2;
CLASS J; * Sample;
VAR y;
RUN; QUIT;

/* Assignment 11, Problem 3 */
TITLE "Bijesh Mishra, Assignment 11, Problem 3";
DATA A11Q3;
DO I = 1 TO 100;
X = 150 + SQRT(8)* RANNOR(1234); * X ~ N(150,8);
OUTPUT;
END;
RUN; QUIT;

PROC GCHART DATA = A11Q3;
VBAR X / LEVELS = 10 MIDPOINTS = 140 142 144 146 148 150
152 154 156 158;
RUN; QUIT;
```

ASSIGNMENT 12:

```
/* Assignment 12 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
* ODS LISTING;
*PROC REGISTRY LIST STARTAT = "COLORNAMES";
* RUN;
TITLE "Bijesh Mishra, Assignment 11 ";
LIBNAME A12 "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Data";
*DATA MESO1_4;
*SET A12.MESONET1_4;
* Obs DAY RAIN_IN STNPRESSUR SOLAR_MJ_m2 SODSOIL
BARESOIL Month MAXTEMP MINTEMP MAXHUMID MINHUMID WINDDIR
SPEEDAVG;
PROC SGPLOT DATA = A12.MESONET1_4;
TITLE "Bijesh Mishra, Assignment 11, Problem 1(a) ";
SCATTER Y = MaxTemp X = SodSoil / GROUP = Month;
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 11, Problem 1(b) ";
PROC SGPLOT DATA = A12.MESONET1_4;
```

```
SCATTER Y = MaxTemp X = SodSoil / GROUP = Month
MARKERCHAR = Month;
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 11, Problem 1(c) ";
PROC SGPANEL DATA = A12.MESONET1_4;
PANELBY Month / LAYOUT = ROWLATTICE;
SCATTER Y = MaxTemp X = SodSoil / GROUP = Month
MARKERCHAR = Month;
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 11, Problem 2 ";
PROC SGPANEL DATA = A12.MESONET1_4;
PANELBY Month;
WHERE Month = "Jan" | Month = "Feb";
SERIES Y = MaxTemp X = Day/ LINEATTRS = (COLOR = Red);
SCATTER Y = MaxTemp X = Day/ MARKERATTRS = (COLOR = Red
SYMBOL = DiamondFilled);
SERIES Y = SodSoil X = Day/ LINEATTRS = (COLOR = Black);
SCATTER Y = SodSoil X = Day / MARKERATTRS = (COLOR =
Black SYMBOL = SquareFilled);
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 11, Problem 3 ";
*PROC MEANS DATA = A12.MESONET1_4 MIN MAX SUM MEAN;
*CLASS MONTH;
*VAR RAIN_IN;
PROC SGPLOT DATA = A12.MESONET1_4;
HBAR Month / RESPONSE = RAIN_IN STAT = SUM TRANSPARENCY
= 0.50 FILLATTRS = (COLOR = Blue);
RUN; QUIT;

TITLE "Bijesh Mishra, Assignment 11, Problem 4 ";
PROC SGPANEL DATA = A12.MESONET1_4;
PANELBY Month;
HISTOGRAM SODSOIL/ FILLATTRS = (COLOR = Green) SCALE =
Percent;
HISTOGRAM BARESOIL /FILLATTRS = (COLOR =
Orange)TRANSPARENCY = 0.50 SCALE = Percent;
RUN; QUIT;
```

ASSIGNMENT 13:

```
/* Assignment 13 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
GOPTIONS RESET = ALL;
TITLE "Bijesh Mishra, Assignment 13 ";
DATA meso ;
INPUT Month $  DAY    MAXTEMP    MINTEMP    MAXHUMID
MINHUMID    WINDDIR $   SPEEDAVG ;
LABEL Month = "Month"
      DAY   = "Day"
        MAXTEMP = "Maximum Temperature"
        MINTEMP = "Minimum Temperature"
        MAXHUMID = "Maximum Humidity"
        MINHUMID = "Minimum Humidity"
        WINDDIR = "Wind Direction"
      SPEEDAVG = "Wind Average Speed";
CARDS; *DATALINES;
       Jan       1        36         32         99
72         NNW          8.3
       Apr      30        81         48         98
23         SE           2.6
;
LIBNAME A13 " C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Assignments\Assignment 13";
DATA work.BBC; * NBA Basketball data as BBC ;
SET A13.combo;
RUN; QUIT;
* input Month $  DAY    MAXTEMP    MINTEMP    MAXHUMID
MINHUMID    WINDDIR $   SPEEDAVG ;
TITLE "Bijesh Mishra, Assignment 13 Q1";
DATA meso;
set meso;
DOY = _N_;
LABEL DOY = "Day of the Year";
*PROC PRINT DATA = meso (OBS = 5)LABEL;
*ID doy month day;
*RUN; *QUIT;

GOPTIONS RESET = ALL;
TITLE "Bijesh Mishra, Assignment 13 Q2";
AXIS1 LABEL = (A = 90); * Rotate Axix by 90 Degree;
PROC GPLOT DATA = meso;
PLOT (MaxTemp MinTemp) * DOY /VAXIS = AXIS1;
SYMBOL INTERPOL = JOIN;
RUN; QUIT;

GOPTIONS RESET = ALL; * Reset All Goptions;
TITLE "Bijesh Mishra, Assignment 13 Q3";
AXIS2 LABEL = ("Temperature"); * Can I also rotate 90
Degree? ;
SYMBOL1 INTERPOL = join COLOR = Green L = 1 W = 1.9;
SYMBOL2 I = JOIN CV = Red LINE = 3 WIDTH = 1.9;
LEGEND1 POSITION = (BOTTOM LEFT OUTSIDE) LABEL = NONE
DOWN = 2 MODE = SHARE;
LEGEND2 POSITION = (BOTTOM LEFT OUTSIDE) LABEL = NONE;
PROC GPLOT DATA = meso;
PLOT (MaxTemp MinTemp) * DOY / OVERLAY LEGEND = LEGEND1
VAXIS = AXIS2 ;
RUN; QUIT;
```

```
GOPTIONS RESET = ALL; * Reset All Goptions;
TITLE "Bijesh Mishra, Assignment 13 Q4(a)";
PROC GCHART DATA = meso;
HBAR Month/ TYPE = MEAN SUMVAR = MaxTemp;
PATTERN COLOR = GREEN VALUE = EMPTY;
RUN; QUIT;

GOPTIONS RESET = ALL; * Reset All Goptions;
TITLE "Bijesh Mishra, Assignment 13 Q4(b)";
PROC GCHART DATA = meso;
HBAR Month/ TYPE = MEAN SUMVAR = MaxTemp NOSTAT AUTOREF;
PATTERN COLOR = CYAN VALUE = SOLID;
RUN; QUIT;

GOPTIONS RESET = ALL; * Reset All Goptions;
TITLE "Bijesh Mishra, Assignment 13 Q4(c)";
PROC GCHART DATA = meso;
HBAR Month/ TYPE = MEAN SUMVAR = MaxTemp NOSTAT AUTOREF
CLIPREF;
PATTERN COLOR = CYAN VALUE = SOLID;
RUN; QUIT;

GOPTIONS RESET = ALL; * Reset All Goptions;
TITLE "Bijesh Mishra, Assignment 13 Q5(a)";
PROC GCHART DATA = bbc;
VBAR Team/TYPE = MEAN SUMVAR = fg GROUP = Year PATTERNID
= MIDPOINT;
WHERE Team = "Oklahoma City Thuhnder" OR
        Team = "Miami Heat" OR
      Team = "Dallas Mavericks" OR
        Team = "Denver Nuggets";
PATTERN1 COLOR = Violet VALUE = SOLID;
PATTERN2 COLOR = Brown VALUE = SOLID;
PATTERN3 COLOR = Yellow VALUE = SOLID;
RUN; QUIT;

GOPTIONS RESET = ALL; * Reset All Goptions;
TITLE "Bijesh Mishra, Assignment 13 Q5(b)";
PROC GCHART DATA = bbc;
VBAR Year/ DISCRETE TYPE = MEAN SUMVAR = fg GROUP = Team
PATTERNID = GROUP;
WHERE Team = "Oklahoma City Thuhnder" OR
        Team = "Miami Heat" OR
      Team = "Dallas Mavericks" OR
        Team = "Denver Nuggets";
PATTERN4 COLOR = Yellow VALUE = SOLID;
PATTERN5 COLOR = Red VALUE = SOLID;
PATTERN7 COLOR = Green VALUE = SOLID;
RUN; QUIT;
```

```
GOPTIONS RESET = ALL; * Reset All Goptions;
TITLE "Bijesh Mishra, Assignment 13 Q5(c)";
PROC GCHART DATA = bbc;
VBAR Year/ DISCRETE TYPE = MEAN SUMVAR = fg GROUP = Team
PATTERNID = MIDPOINT;
WHERE Team = "Oklahoma City Thuhnder" OR
        Team = "Miami Heat" OR
      Team = "Dallas Mavericks" OR
        Team = "Denver Nuggets";
PATTERN7 COLOR = Black VALUE = SOLID;
PATTERN8 COLOR = Blue VALUE = SOLID;
RUN; QUIT;
```

ASSIGNMENT 14:

Graduate Project:

```
/* Gradaute Project MishraODS1 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
ODS TRACE ON / LISTING;
ODS PDF FILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Graduate
Project\MishraODS1.pdf";
TITLE "Bijesh Mishra, Graduate Project Q1";
DATA one;
INPUT x y @@;
DATALINES;
3.1 5.5 2.3 4.8 3.0 4.7 1.9 3.9
2.5 4.5 3.7 6.2 3.4 6.0 2.6 5.2
2.8 4.7 1.6 4.3 2.0 4.9 2.9 5.4
2.3 5.0 3.2 6.3 1.8 4.6 1.4 4.3
2.0 5.0 3.8 5.9 2.2 4.1 1.5 4.7
;
PROC REG DATA = one;
MODEL Y = X/ ALPHA = 0.02 CLB CLM;
ID X;
ODS PDF SELECT ParameterEstimates OutputStatistics
FitPlot;
ODS TRACE OFF;
RUN;
ODS PDF CLOSE;
QUIT;
*FILE "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Graduate Project\MishraODS1.sas"
;

/* Gradaute Project MishraODS2 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
ODS TRACE ON / LISTING;
ODS TRACE OFF;
```

```
TITLE "Bijesh Mishra, Graduate Project Q2";
LIBNAME ODSHW2 "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Graduate Project";
ODS HTML FILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Graduate
Project\MishraODS2b.HTML";
* INPUT Iteration Group Sample Response;
TITLE "Bijesh Mishra, Graduate Project Q2(A) and Q2(B)";
ODS EXCLUDE ALL; * Supress all output from GLM
Procedure;
ODS OUTPUT  OverallANOVA = odshw2.Anova; * Anova table
as Temp. DataSet (Q2(a));
PROC GLM DATA = odshw2.odshw2;
CLASS Group;
MODEL Response = Group Sample;
MEANS Group; * Get Means;
BY Iteration;
ODS HTML SELECT Means; * Print Means in HTML file *;
RUN;
ODS HTML CLOSE; * Close HTML;
QUIT;

TITLE "Bijesh Mishra, Graduate Project Q2(C) Bonus
Question";
ODS SELECT NONE; * Supress all output;
ODS HTML FILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Graduate
Project\MishraODS2c.HTML";
DATA odshw2.Anova;
SET odshw2.Anova;
I = ProbF LE 0.05; * Indicator Variable;
PROC FREQ DATA = odshw2.Anova;
TABLES I;
ODS HTML SELECT OneWayFreqs;
RUN;
DATA odshw2.Anova;
SET odshw2.Anova;
IF ProbF LE 0.05 THEN II = 1;
ELSE II = 0; * IF THEN ELSE;
RUN;
PROC FREQ DATA = odshw2.Anova;
TABLES II;
ODS HTML SELECT OneWayFreqs;
RUN;
ODS HTML CLOSE; * Close HTML;
QUIT;


/* Gradaute Project MishraODS3 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
ODS TRACE ON / LISTING;
*ODS TRACE OFF;
TITLE "Bijesh Mishra, Graduate Project Q3";
/* Data Year 2020 */
DATA y20;
INPUT Team $ 1-25 FG 26-30 FGA 31-35 P3 36-40 FT 41-45;
DATALINES;
Dallas Mavericks          3124 6772 1136 1392
Charlotte Hornets         2425 5586 785  1052
;

/* Data Year 2021 */
DATA y21;
INPUT Team $ 1-25 FG 26-30 FGA 31-35 P3 36-40 FT 41-45;
DATALINES;
Milwaukee Bucks           3221 6610 1038 1169
Cleveland Cavaliers       2778 6175 720  1200
;

/* Create Year and Conference Variables */
DATA y20;
SET y20;
year = 2020;
IF team = "Dallas Mavericks" THEN conf = "west";
IF team = "Milwaukee Bucks" THEN conf = "east";
IF team = "Portland Trail Blazers" THEN conf = "west";
IF team = "Houston Rockets"   THEN conf = "west";

IF team = "Los Angeles Clippers" THEN conf = "west";
IF team = "New Orleans Pelicans" THEN conf = "west";
IF team = "Phoneix Suns" THEN conf = "west";
IF team = "Washington Wizards" THEN conf = "east";
IF team = "Memphis Grizzlies" THEN conf = "west";
IF team = "Boston Celtics" THEN conf = "east";
IF team = "Miami Heat" THEN conf = "east";
IF team = "Denver Nuggets" THEN conf = "west";
IF team = "Toronto Raptors"   THEN conf = "east";

IF team = "San Antonio Spurs" THEN conf = "west";
IF team = "Philadelphia 76ers" THEN conf = "east";

IF team = "Los Angeles Lakers" THEN conf = "west";
IF team = "Brooklyn Nets" THEN conf = "east";
IF team = "Utah Jazz" THEN conf = "west";
IF team = "Indiana Pacers" THEN conf = "east";
IF team = "Oklahoma City Thunder" THEN conf = "west";
IF team = "Sacremento Kings" THEN conf = "west";

IF team = "Orlando Magic" THEN conf = "east";
```

```
IF team = "Atlanta Hawks" THEN conf = "east";

IF team = "Minnesota Timberwolves" THEN conf = "west";

IF team = "Detroit Pistons"   THEN conf = "east";

IF team = "New York Knicks"   THEN conf = "east";

IF team = "Cleveland Cavaliers" THEN conf = "east";
IF team = "Chicago Bulls" THEN conf = "east";
IF team = "Golden State Warriors" THEN conf = "west";
IF team = "Charlotte Hornets" THEN conf = "east";

/* Create Year and Conference Variables */
DATA y21;
SET y21;
year = 2021;
IF team = "Dallas Mavericks" THEN conf = "west";
IF team = "Milwaukee Bucks" THEN conf = "east";
IF team = "Portland Trail Blazers" THEN conf = "west";
IF team = "Houston Rockets"   THEN conf = "west";

IF team = "Los Angeles Clippers" THEN conf = "west";
IF team = "New Orleans Pelicans" THEN conf = "west";
IF team = "Phoneix Suns" THEN conf = "west";
IF team = "Washington Wizards" THEN conf = "east";
IF team = "Memphis Grizzlies" THEN conf = "west";
IF team = "Boston Celtics" THEN conf = "east";
IF team = "Miami Heat" THEN conf = "east";
IF team = "Denver Nuggets" THEN conf = "west";
IF team = "Toronto Raptors"   THEN conf = "east";

IF team = "San Antonio Spurs" THEN conf = "west";
IF team = "Philadelphia 76ers" THEN conf = "east";

IF team = "Los Angeles Lakers" THEN conf = "west";
IF team = "Brooklyn Nets" THEN conf = "east";
IF team = "Utah Jazz" THEN conf = "west";
IF team = "Indiana Pacers" THEN conf = "east";
IF team = "Oklahoma City Thunder" THEN conf = "west";
IF team = "Sacremento Kings" THEN conf = "west";

IF team = "Orlando Magic" THEN conf = "east";

IF team = "Atlanta Hawks" THEN conf = "east";

IF team = "Minnesota Timberwolves" THEN conf = "west";

IF team = "Detroit Pistons"   THEN conf = "east";

IF team = "New York Knicks"   THEN conf = "east";

IF team = "Cleveland Cavaliers" THEN conf = "east";
IF team = "Chicago Bulls" THEN conf = "east";
IF team = "Golden State Warriors" THEN conf = "west";
IF team = "Charlotte Hornets" THEN conf = "east";

/* Merge Dataset */
DATA combo;
SET Y20 Y21;
RUN; QUIT;

ODS TRACE ON / LISTING;
*ODS TRACE OFF;
ODS SELECT NONE;
TITLE "Bijesh Mishra, Graduate Project Q3";
PROC REPORT DATA = combo;
COLUMN Team fg p3 ft;
DEFINE Team / GROUP;
DEFINE fg / 'Field Goals';
DEFINE p3 / '3 Point Field Goals';
DEFINE ft / 'Free Throws';
ODS OUTPUT Report = work.Ods3; * Report as Temp.
DataSet;
RUN; QUIT;

/*
PROC EXPORT DATA = Ods3
OUTFILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Graduate
Project\MishraODS3.XLSX"
DBMS = EXCEL;
SHEET = "Ods3";
RUN; QUIT;
*/

PROC EXPORT DATA = Ods3
OUTFILE = "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\SAS\Graduate Project\MishraODS3.csv"
DBMS = csv REPLACE;
RUN; QUIT;
```

Book Chapter Codes:

Chapter 1: Introduction

```
/* Goad, 2021. SAS programming */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; OUTPUT; CLEAR;';
TITLE; FOOTNOTE;

DATA one;
INPUT Name $ Fine @@;
DATALINES;
Lynn 50 Evan 70 Thomas 24 Wesley 44 Marie 30
;
PROC PRINT DATA = one;
TITLE 'Objective 1.1; Outstanding Parking Fines';
PROC MEANS DATA = one;
VAR fine;
RUN;
QUIT;

FILE C:\Users\bmishra\Dropbox\OSU\PhD\Fall 2021\STAT5193
SAS R\Course Book\Chapter1.sas
```

Chapter 2: Data Step Information 1

```
/* Goad, 2021. SAS programming */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; OUTPUT; CLEAR;';
TITLE; FOOTNOTE;

DATA one;
INPUT Name $ Fine;
DATALINES;
Lynn 50 Evan 70 Thomas 24 Wesley 44 Marie 30
;
PROC SORT DATA = one;
BY name;
PROC PRINT DATA = one NOOBS DOUBLE UNIFORM LABEL N;
TITLE 'Objective 1.1; Outstanding Parking Fines';
ID name;
VAR name fine;
BY name;
RUN; QUIT;

PROC MEANS DATA = one;
VAR fine;
RUN;
QUIT;
FILE C:\Users\bmishra\Dropbox\OSU\PhD\Fall 2021\STAT5193
SAS R\Course Book\Chapter2.sas
```

Chapter 3: Summarizing Data Basics

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE; TITLE3; FOOTNOTE;

DATA instruction;
INPUT program $ score @@;
DATALINES;
A 71 A 82 A 88 A 64 A 59 A 78 A 72 A 81 A 83 A 66 A 83 A
91 A 79 A 70 B 65 B 88 B 92 B 76 B 87 B 89 B 85 B 90 B
81 B 91 B 78 B 81 B 86 B 82 B 73 B 79
;
PROC UNIVARIATE DATA = instruction;
VAR score;
TITLE 'Objective 3.1: Defalut Information';
RUN; QUIT;

PROC UNIVARIATE DATA = instruction CIBASIC ALPHA = 0.01
NORMAL;
CLASS program; /* Categorical variable and undorted
data. */
VAR score;
HISTOGRAM score/NORMAL;
TITLE ' Objective 3.2';
TITLE3 'Using a CLASS statement';
FOOTNOTE ' Objective 3.2 can be achieved by using
1) NORMAL option on PROC UNIVARIATE only,
2) NORMAL option in HISTOGRAM only
3) Include both NORMAL options.';
RUN; QUIT;

/* Recovering Statistics in Data Set */
PROC SORT DATA = instruction;
BY program;
TITLE; TITLE3; FOOTNOTE;
PROC UNIVARIATE DATA = instruction MU0 = 75;
BY program;
VAR score;
HISTOGRAM score/NORMAL (MU = 75);
OUTPUT OUT = three MEAN = mnscore N = nscore
STDMEAN = semscore NOBS = nobs MIN = min
RANGE = range VAR = var STD = std CV = cv;
TITLE 'Objective 3.3';
TITLE3 'Using a BY statement';
TITLE; TITLE3; FOOTNOTE;
PROC PRINT DATA = three;
TITLE3 'Output Data Set Identifying Each Program';
RUN; QUIT;

/* Test for Location Syntax Options */
PROC UNIVARIATE DATA = data MU0 = (15 27);
var a b c;
FOOTNOTE 'Ho: MU = 15, 27 and 0 for variables a, b & c
respectively';
```

```
/* The MEANS Procedure */
TITLE; TITLE3; FOOTNOTE;
PROC MEANS DATA = instruction MIN MAX MEAN STD N CV;
CLASS program;
VAR score;
TITLE ' Objective 3.4, 3.5 & 3.6';
OUTPUT OUT = six MIN = score_min
MAX = score_max RANGE = score_range;
TITLE; TITLE3; FOOTNOTE;
PROC PRINT DATA = six;
RUN; QUIT;

PROC SORT DATA = instruction;
BY program;
TITLE; TITLE3; FOOTNOTE;
PROC MEANS DATA = instruction
MIN MAX MEAN STD N CV STDERR T PRT CLM ALPHA = 0.01;
BY program;
VAR score;
TITLE ' Objective 3.7';
OUTPUT OUT = seven MIN = score_min MAX = score_max RANGE
= score_range;
PROC PRINT DATA = seven;
RUN; QUIT;

TITLE; TITLE3; FOOTNOTE;
PROC MEANS DATA = instruction MEAN STDERR T PRT CLM
ALPHA = 0.01;
CLASS program;
VAR score;
TITLE ' Objective 3.8';
RUN; QUIT;

/* The MEANS Procedure */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE; TITLE3; FOOTNOTE;
DATA instruction;
INPUT program $ score @@;
DATALINES;
A 71 A 82 A 88 A 64 A 59 A 78 A 72 A 81 A 83 A 66 A 83 A
91 A 79 A 70 B 65 B 88 B 92 B 76 B 87 B 89 B 85 B 90 B
81 B 91 B 78 B 81 B 86 B 82 B 73 B 79
;
PROC MEANS DATA = instruction MIN MAX MEAN STD N CV;
CLASS program;
VAR score;
TITLE ' Objective 3.4, 3.5 & 3.6';
OUTPUT OUT = six MIN = score_min
MAX = score_max RANGE = score_range;
TITLE; TITLE3; FOOTNOTE;
PROC PRINT DATA = six;
RUN; QUIT;

PROC SORT DATA = instruction;
BY program;
TITLE; TITLE3; FOOTNOTE;
PROC MEANS DATA = instruction
MIN MAX MEAN STD N CV STDERR T PRT CLM ALPHA = 0.01;
BY program;
VAR score;
TITLE ' Objective 3.7';
OUTPUT OUT = seven MIN = score_min MAX = score_max RANGE
= score_range;
PROC PRINT DATA = seven;
RUN; QUIT;

TITLE; TITLE3; FOOTNOTE;
PROC MEANS DATA = instruction MEAN STDERR T PRT CLM
ALPHA = 0.01;
CLASS program;
VAR score;
TITLE ' Objective 3.8';
RUN; QUIT;
```

Chapter 4: Data Step Information 2

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE; FOOTNOTE;
DATA fish;
INPUT ID Location $ Length Weight Age Gender $ @@;
Length_in = Length/25.4;
Gender2 = LOWCASE(Gender); * Lower Case: Obj. 4.4;
Gender3 = UPCASE(Gender); * Upper Case;
Location = PROPCASE(Location); * Proper Case: Obj. 4.5;
State = "OK";

IF length < 60 THEN Size = 'Small';
ELSE size = 'Large';
IF age LE 2.0 THEN Group = 1;
IF 2.0 < age < 3.0 THEN Group = 2;
IF age GE 3.0 THEN Group = 3;
IF location = "payne" THEN Lake = 'CB';

DATALINES;
23 payne 75 24 2.5 f 41 payne 68 16 2 m 17 payne 57 12
1.5 F 33 payne 45 14 0.5 m 18 payne 71 20 3 F 77 payne
60 19 2.5 f
;
PROC PRINT  DATA = fish;
TITLE 'Objective 4.1';
RUN; QUIT;
```

```
PROC SORT DATA = fish; BY gender;
PROC PRINT DATA = fish; BY gender;
TITLE 'Objective 4.3 & 4.4';
RUN; QUIT;

PROC PRINT  DATA = fish;
WHERE weight LE 15;
TITLE 'Objective 4.6';
TITLE2 ' Observations with Weight <= 15';
RUN; QUIT;

PROC PRINT  DATA = fish;
WHERE gender = 'm';
VAR length length_in weight;
TITLE 'Objective 4.6';
TITLE2 ' Summary Statistics for Males';
RUN; QUIT;

PROC PRINT DATA = fish NOOBS;
VAR location state length_in weight gender;
TITLE 'Objective 4.5';
RUN; QUIT;

DATA Grades;
INPUT Name $ ID Exam1 Exam2 Exam3 Q1 Q2 Q3 Q4 Q5;
ExamTotal1 = SUM(exam1, exam2, exam3);
ExamTotal2 = exam1 + exam2 + exam3;
ExamsN = N(exam1, exam2, exam3); * Number of exams
completed;
ExamAvg1 = MEAN(exam1, exam2, exam3);
ExamAvg2 = (exam1 + exam2 + exam3)/ExamsN;
MinQuiz = MIN(q1, q2, q3, q4, q5);
Best4QuizTotal = SUM (q1, q2, q3, q4, q5) - MIN(q1, q2,
q3, q4, q5);
Best4QuizTotal1 = SUM (q1, q2, q3, q4, q5) - MinQuiz;
DATALINES;
Bill 123000000 85 88 84 20 22 16 . 21 Helen 234000000 96
90 89 16 25 20 18 22 Steven 345000000 80 92 82 19 24 19
20 21 Caria 456000000 65 78 74 18 20 23 20 24 Dana
567000000 97 94 . 22 17 24 18 20 Lisa 789000000 81 88 92
15 20 22 18 19
;
PROC PRINT DATA = Grades NOOBS;
VAR name exam1 exam2 exam3 examtotal1 examtotal2 examsn
examavg1 examavg2;
TITLE 'Objective 4.2';
TITLE3 'Exam Info Only';
RUN; QUIT;

PROC PRINT DATA = Grades NOOBS;
VAR name q1 q2 q3 q4 q5 best4quiztotal;
TITLE 'Objective 4.2';
TITLE3 'Quiz Info Only';
RUN; QUIT;

DATA males_fish; * Creates New Dataset males_fish;
SET fish;
IF gender = 'f' THEN DELETE;
TITLE 'Objective 4.7';
PROC PRINT DATA = males_fish;
RUN; QUIT;

DATA  fish; * Update Old Dataset;
SET fish;
Species = "darter";
TITLE 'Objective 4.8';
PROC PRINT DATA = FISH;
RUN; QUIT;

/* Combine/Merge Two Datasets */
DATA fish_noble;
INPUT ID Lake $ Gender $ Weight Length Age @@;
Length_in = length/25.4;
Location = "Noble";
State = "OK";
DATALINES;
83 PRY f 20 61 2 72 MCM m 24 80 3 30 MCM m 19 69 1.5 46
pry f 18 50 2.5 78 MCM f 19 54 2
;
PROC PRINT DATA = fish_noble;
TITLE 'Objective 4.9';
RUN; QUIT;
DATA combine;
SET fish fish_noble;
PROC PRINT DATA = combine;
TITLE 'Objective 4.9';
PROC MEANS DATA = combine;
CLASS location;
RUN; QUIT;
```

Chapter 5: Beginning Charts

```
/* Chapter 5 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE; TITLE3; FOOTNOTE;
DATA instruction;
INPUT program $ score @@;
DATALINES;
```

```
A 71 A 82 A 88 A 64 A 59 A 78 A 72 A 81 A 83 A 66 A 83 A
91 A 79 A 70 B 65 B 88 B 92 B 76 B 87 B 89 B 85 B 90 B
81 B 91 B 78 B 81 B 86 B 82 B 73 B 79
;
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC GCHART DATA = instruction;
VBAR program ; /* /MIDPOINTS = 1 2*/
TITLE 'Objective 5.1';
TITLE3 'Vertical Bar Chart for Program';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC SORT DATA = instruction; BY program;
PROC GCHART DATA = instruction; BY program;
VBAR score / SPACE = 0;
TITLE 'Objective 5.2';
TITLE3 'Histogram of the Scores for Each Program' ;
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC GCHART DATA = instruction;
BY program;
VBAR score / SPACE = 0 AXIS = 1 TO 6 MIDPOINTS = 60 65
70 75 80 85 90;
TITLE 'Objective 5.3';
TITLE2 'Histogram for the Scores for Each Program';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC GCHART DATA = instruction;
VBAR score / SPACE = 0 GROUP = program AXIS = 1 TO 6
                  MIDPOINTS = 60 65 70 75 80 85 90;
TITLE 'Objective 5.4';
TITLE2 'Histogram for the Scores for Each Program';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC GCHART DATA = instruction;
HBAR score;
TITLE 'Objective 5.5: Default Horizantal Bar Chart for
Score';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC GCHART DATA = instruction;
HBAR score / SPACE = 0 LEVELS = 4 NOSTATS AUTOREF
CLIPREF;
TITLE 'Objective 5.6: Histogram';
TITLE3 ' Number of Levels Specified - All Frequency
Information is supressed';
TITLE4 'Reference lines are included in the
background.';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC GCHART DATA = instruction;
HBAR program; / TYPE = MEAN SUMVAR = score MEAN;
TITLE 'Objective 5.7';
TITLE3 ' The mean of each program';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC GCHART DATA = instruction;
VBAR3D program / TYPE = MEAN SUMVAR = score MEAN;
TITLE 'Objective 5.8: Mean of each program in a 3-D
horizantal bar chart';
* FOOTNOTE ' SUMVAR and TYPE tell us which variable and
statistics to use
as reference variable in the chart. See practice 4,
Q1(c)';
HBAR3D program / TYPE = MEAN SUMVAR = score MEAN;
TITLE 'Objective 5.8: Mean of each program in a 3-D
vertical bar chart';
BLOCK program / TYPE = MEAN SUMVAR = score NOHEADING;
TITLE 'Objective 5.8: Mean of each program in a block
chart';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC GCHART DATA = instruction;
PIE program / TYPE = PERCENT ;
TITLE 'Objective 5.9: Pie Chart';
TITLE2 'for the percent of subjects in each of the
programs.';
RUN; QUIT;
```

Chapter 6: One and Two Population Hypothesis Tests about the Means

```
/* Chapter 6 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE; TITLE3; FOOTNOTE;
DATA instruction;
INPUT program $ score @@;
DATALINES;
A 71 A 82 A 88 A 64 A 59 A 78 A 72 A 81 A 83 A 66 A 83 A
91 A 79 A 70 B 65 B 88 B 92 B 76 B 87 B 89 B 85 B 90 B
81 B 91 B 78 B 81 B 86 B 82 B 73 B 79
;

PROC PRINT DATA = instruction;
RUN;
```

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC TTEST DATA = instruction H0 = 75;
VAR score;
TITLE 'Objective 6.1';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC TTEST DATA = instruction H0 = 75 ALPHA = 0.02 PLOTS
= NONE;
VAR score;
TITLE 'Objective 6.2';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC SORT DATA = instruction; BY program;
PROC TTEST DATA = instruction PLOTS = NONE H0 = 75 ALPHA
= 0.02 CIBASIC SIDES = U;
BY program;
VAR score;
TITLE 'Objective 6.3';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC TTEST DATA = instruction SIDES = 2 ALPHA = 0.05 H0
= 75;
VAR score;
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC UNIVARIATE DATA = instruction ALPHA = 0.05 CIBASIC
MU0 = 75;
VAR score;
RUN; QUIT;

data improvement;
input subject before after @@;
DATALINES;
1 138 324 2 284 520 3 234 318 4 132 220 5 183 232
;
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC TTEST DATA = improvement SIDES = L CI = NONE ALPHA
= 0.01;
PAIRED before*after;
TITLE 'Objective 6.4';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC TTEST DATA = improvement ALPH = 0.01
                PLOTS (ONLY) = (HISTOGRAM BOXPLOT);
PAIRED before*after;
TITLE 'Objective 6.5 (1)';
RUN; QUIT;

PROC TTEST DATA = improvement ALPH = 0.01
                PLOTS (ONLY) = (BOXPLOT INTERVAL) SIDED =
L;
PAIRED before*after;
TITLE 'Objective 6.5 (2)';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC TTEST DATA = instruction;
CLASS program;
VAR score;
TITLE 'Objective 6.6';
TITLE3 't-test for the Difference Between Two
Independent Means';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
PROC TTEST DATA = instruction SIDED = L;
CLASS program;
VAR score;
TITLE 'Objective 6.7 (1)';
TITLE3 't-test (Lower Sided) for the Difference Between
Two Independent Means';
RUN; QUIT;

PROC TTEST DATA = instruction SIDED = U;
CLASS program;
VAR score;
TITLE 'Objective 6.7 (2)';
TITLE3 't-test (Upper Sided) for the Difference Between
Two Independent Means';
RUN; QUIT;
```

Chapter 7: One-way ANOVA METHODS, Non-parametric Methods and Ranking Data

```
/* Chapter 7 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE; TITLE3; FOOTNOTE;

DATA meat;
INPUT Group $ pH CookYield @@;
DATALINES;
C 6.14 20.9 C 5.98 22.1 C 6.30 21.8 C 6.25 20.3 C 6.07
21.2 T1 5.98 22.4 T1 6.32 23.8 T1 5.89 23.0 T1 6.08 24.
```

```
5 T1 6.11 22.8 T2 6.18 23.4 T2 6.22 20.8 T2 6.03 22.6 T2
5.97 24.8 T2 5.93 25.1
;
PROC PRINT DATA = meat; RUN; QUIT;
DM 'ODSRESULTS; CLEAR;'; TITLE; TITLE3; FOOTNOTE;

/* PROC GLM for ANOVA */
PROC GLM DATA = meat PLOTS = NONE; /* PLOTS (ONLY) =
(RESIDUALS DIAGNOSTICS); */
CLASS group;
MODEL ph cookyield = group; * Two dependent variables;
MEANS group;
TITLE 'Objective 7.1 and 7.2 ANOVA';
RUN; QUIT;

PROC GLM DATA = meat  PLOTS (ONLY) = (RESIDUALS
DIAGNOSTICS);
CLASS group;
MODEL cookyield = group;
MEANS group / CLM LSD; * Confidence Limit and Least
Significant Difference;
TITLE "Objective 7.3 ANOVA, CI's and Residuals";
RUN; QUIT;

/* Non Parametric Test: NPAR1WAY */
PROC NPAR1WAY DATA = meat WILCOXON;
WHERE group NE "T2"; * T2 group excluded;
CLASS group;
VAR ph;
TITLE "Objective 74. Wilcoxon Score";
RUN; QUIT;

PROC NPAR1WAY DATA = meat WILCOXON ANOVA PLOTS = NONE;
CLASS Group;
VAR CookYield;
TITLE 'Objective 7.5';
RUN; QUIT;

/* Non Parametric Test: RANK */
DATA six;
INPUT X Y Z @@;
DATALINES;
89 25 41  47 33 37 73 27 37 66 25 29 50 42 37
;
PROC PRINT DATA = six;
TITLE "Data Six";
RUN; QUIT;
PROC RANK DATA = six OUT = new6;
VAR x y z;
RANKS RX RY RZ; * Order matters, Must be VAR = RANKS;
PROC PRINT DATA = new6;
VAR rx x ry y rz z; * Reorder variables to print ;
TITLE 'Objective 7.6';
RUN; QUIT;

TITLE 'Objective 7.7';
PROC RANK DATA = six OUT = high TIES = HIGH;
VAR x y z;
RANKS rx ry rz; * Order matters, Must be VAR = RANKS;
PROC PRINT DATA = high ;
VAR rx x ry y rz z; * Reorder variables to print ;
TITLE 'Objective 7.7 (TIES = HIGH)';
RUN; QUIT;

PROC RANK DATA = six OUT = highdes TIES = HIGH;
VAR x y z;
RANKS rx ry rz; * Order matters, Must be VAR = RANKS;
PROC PRINT DATA = highdes ;
VAR rx x ry y rz z; * Reorder variables to print ;
TITLE 'Objective 7.7 (TIES = HIGH DESCENDING)';
RUN; QUIT;

PROC RANK DATA = six OUT = low TIES = HIGH;
VAR x y z;
RANKS rx ry rz; * Order matters, Must be VAR = RANKS;
PROC PRINT DATA = low ;
VAR rx x ry y rz z; * Reorder variables to print ;
TITLE 'Objective 7.7 (TIES = LOW)';
RUN; QUIT;

PROC RANK DATA = six OUT = lowdes TIES = HIGH;
VAR x y z;
RANKS rx ry rz; * Order matters, Must be VAR = RANKS;
PROC PRINT DATA = lowdes;
VAR rx x ry y rz z; * Reorder variables to print ;
TITLE 'Objective 7.7 (TIES = LOW DESCENDING)';
RUN; QUIT;

PROC RANK DATA = six OUT = SixA;
VAR x;
RANKS rx;
TITLE 'Objective 7.8, Option 1';
RUN; QUIT;

PROC RANK DATA = six OUT = SixB;
VAR y;
RANKS ry;
PROC SORT DATA = SixA; BY x y;
PROC SORT DATA = SixB; BY x y;
TITLE 'Objective 7.8, Option 1';
```

```
RUN; QUIT;
/* Sorting the data sets by X (or RX) before merging is
all that is
necessary since X has all unique values. If there were
duplicate values
for X one would want to include more than one variable
in the BY
statements of the SORT procedures. See Section 20.2 for
information
on annotating programs with block, comments such as
this. */
DATA eight;
MERGE SixA sixB; BY x y;
PROC PRINT DATA = eight;
TITLE "Objective 7.8 - Option 1";
RUN; QUIT;

PROC RANK DATA = SIX OUT = new6;
VAR x;
VAR rx;
PROC RANK DATA = new6 OUT = eight DESCENDING;
VAR y;
RANKS ry;

PROC SORT DATA = eight; BY x; * or SORT BY rx;
PROC PRINT DATA = eight;
VAR X rx y ry z; * Optional Statement to order
variables;
TITLE "Objective 7.8 - Option 2.";
RUN; QUIT;
```

Chapter 8: Data Step information 3: Reading Data Files and Labeling Variables:

```
/* Chapter 8 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
/* Objective 8.1 */
FILENAME t1 'C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\Course Book\tornado1.dat';
FILENAME t2 'C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\Course Book\tornado2.dat';
FILENAME t3 'C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\Course Book\tornado3.dat';
DATA one;
INFILE t1;
INPUT Year Number Damage;
PROC PRINT DATA = one;
TITLE 'Objective 8.1';

DATA two;
INFILE t2 FIRSTOBS = 2;
INPUT Year Number Damage;
PROC PRINT DATA = two;
TITLE 'Objective 8.1';

DATA three;
INFILE t3;
INPUT Year 1-4 Number 5-6 Damage 7-9;
PROC PRINT DATA = three;
TITLE 'Objective 8.1';
RUN; QUIT;

/* Objective 8.2 */
DATA four;
INFILE 'C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\Course Book\tornado4.dat' DLM =
'09'X DSD;
INPUT Year Number Damage;
PROC PRINT DATA = four;
TITLE 'Objective 8.2';
RUN; QUIT;

/* Objective 8.3 */
DATA four;
INFILE 'C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\Course Book\tornado4.dat' DLM =
'09'X DSD;
INPUT Year Number Damage;
LABEL Number = "Number of Tornadoes"
      Damage = "Damage Assessment, X$10,0000";
PROC PRINT DATA = four;
TITLE 'Objective 8.2';
TITLE2 'Print with no LABEL options';
PROC MEANS DATA = four;
VAR number damage;
TITLE2;
RUN; QUIT;

DATA five;
SET four;
RENAME Year = Year1 Number = Number1 Damage = Damage1;
KEEP Year1 Number1;
PROC PRINT DATA = five;
TITLE 'RENAME & KEEP Demonstration';
RUN; QUIT;

DATA six;
SET four;
DROP Year;
PROC PRINT DATA = four;
TITLE 'Drop Demonstration';
```

```
RUN; QUIT;
```

Chapter 9: Frequency Analysis

```
/* Chapter 9 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
DATA rtw;
INPUT EmpClass $ Opinion $ Y @@;
LABEL EmpClass = "Employment Classification"
        Opinion = "Opinion on Right to Work"
        Y = "Observed Frequency";
DATALINES;
I F 20 I DNF 24 I U 16 B F 40 B DNF 51 B U 9 U F 20 U
DNF 15 U U 7
;
* The following FREQ procedure will count the number of
occurrences;
* of the levels of the variables CLASS and OPINION *;
PROC FREQ DATA = rtw;
TABLES Opinion;
TITLE "Objective 9.1 - No WEIGHT Statement";
* The WEIGHT statement is necessary in order to get
correct frequencies;
* when the counts for each level of a variable are
included in the data;
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
PROC FREQ DATA = rtw ORDER = DATA;
TABLES Opinion / CHISQ;
WEIGHT Y;
TITLE "Objective 9.1 - with a WEIGHT Statement";
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
PROC FREQ DATA = rtw ORDER = DATA;
TABLES EmpClass / TESTP = (45 50 5);
WEIGHT y;
TITLE 'Objective 9.2 - Goodness of Fit Test for Employee
Classification';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
PROC FREQ DATA = rtw;
TABLES EmpClass * Opinion / LIST;
WEIGHT y;
TITLE 'Objective 9.3 - LIST Option';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
PROC FREQ DATA = rtw ORDER = DATA;
TABLES EmpClass*Opinion / CHISQ;
WEIGHT y;
TITLE 'Objective 9.3 - Ordered Data in a Two-way Table';
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
PROC FREQ DATA = rtw ORDER = DATA;
tables EmpClass*Opinion/ CHISQ PLOTS = FREQPLOT;
WEIGHT Y;
TITLE "Objective 9.4" ;
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
PROC FREQ DATA = rtw ORDER = DATA;
TABLES EmpClass*Opinion/ CHISQ PLOTS = FREQPLOT (TYPE =
DOTPLOT);
WEIGHT Y;
TITLE "Objective 9.5";
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
PROC FREQ DATA = rtw ORDER = DATA;
TABLES EmpClass*Opinion/ CHISQ NOROW NOCOL NOPERCENT
EXPECTED;
*The options on the TABLES statement can be in any
order;
WEIGHT y;
TITLE "Objective 9.6";
RUN; QUIT;

DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
PROC FREQ DATA = rtw ORDER = DATA;
TABLES Opinion / CHISQ; *Objective 9.1;
TABLES EmpClass / TESTP = (45 50 5); *Objective 9.2;
tables EmpClass * Opinion / LIST PLOTS = NONE;
*Objective 9.6;
WEIGHT y;
TITLE "Multiple Tables";
RUN; QUIT;
```

Chapter 10: Summarizing a Data Table in a Formal Report

```
/* Chapter 10 */
DM ' LOG; CLEAR; ODSRESULTS; CLEAR;';
DATA gradebook;
INPUT Student $ 9. ID Group Hw1 Hw2 Ex1 Hw3 Hw4 Ex2 @@;
HwTotal = SUM (HW1, HW2, HW3, HW4);
ExTotal = SUM (Ex1, Ex2);
CourseTL = SUM (HwTotal, ExTotal);
LABEL HwTotal = "Homework Points Total"
      ExTotal = "Exam Score Total"
```

```
        Course TL = " Point Total for the Course";
DATALINES;
Total      . . 75 110 100 50 25 100 Dave       101 1 71 88
93 46 23 88 Lynn       381 2 64 96 95 48 25 . Michael 987
2 68 75 97 35 12 60 Leslie    579 3 55 75 81 . 17 82
Andrew    239 1 70 79 77 38 23 77 Elizabeth 128 3 67 103
94 42 20 92
;
PROC PRINT DATA = gradebook LABEL;
TITLE "Objective 10.1";
RUN; QUIT;

DATA gradebook;
SET gradebook;                          * max points =
460;
IF CourseTL GE 414 THEN Grade = "A";    * 460*0.90 = 414
Lowest A;
ELSE IF 368 LE CourseTL LE 413 THEN Grade = "B";    *
460*0.80 = 368 Lowest B;
ELSE IF 322 LE CourseTL LE 367 THEN Grade = "B";    *
460*0.70 = 322 Lowest C;
ELSE IF 276 LE CourseTL LE 321 THEN Grade = "D";    *
460*0.60 = 276 Lowest D;
ELSE Grade = "F";
IF Student = "Total" THEN Grade = " ";

DM 'ODSRESULTS; CLEAR;';
PROC PRINT DATA = gradebook LABEL NOOBS;
TITLE "Objective 10.1";
TITLE2 ' PRINT Procedure with LABEL Options';
TITLE3 "Including new Variabels";
RUN; QUIT;

DM 'ODSRESULTS; CLEAR;';
PROC FREQ DATA = gradebook NLEVELS;
WHERE Student NE "Total";
TABLES Grade;
TITLE "Objective 10.2";
TITLE2 "FREQ Procedure with NLEVELS Options";
TITLE3 "One-way Table";
RUN; QUIT;

DM 'ODSRESULTS; CLEAR;';
PROC FREQ DATA = gradebook NLEVELS;
WHERE Student NE "Total";
TABLES Group*Grade/ NOPERCENT NOROW NOCOL;
TITLE "Objective 10.3";
TITLE2 "Two-way Table with NLEVELS Option";
RUN; QUIT;

DM ' ODSRESULTS; CLEAR;';
PROC FREQ DATA = gradebook;
WHERE Student NE "Total";
TABLES Group*Grade/ CROSSLIST;
TITLE "Objective 10.4";
TITLE2 "FREQ Procedure: Two-way Table with CROSSLIST
Option";
RUN; QUIT;

PROC FREQ DATA = gradebook;
WHERE Student NE "Total";
TABLES Group*Grade/ LIST;
TITLE "Objective 10.4";
TITLE2 "FREQ Procedure: Two-way Table with LIST Option";
RUN; QUIT;

DM ' ODSRESULTS; CLEAR;';
PROC MEANS DATA = gradebook MEAN MIN MAX N;
WHERE Student NE "Total";
VAR HwTotal ExTotal CourseTL;
TITLE "Objective 10.5";
TITLE2 "MEANS Procedure with no CLASS Statement ";
RUN; QUIT;

PROC MEANS DATA = gradebook MEAN MIN MAX N;
WHERE Student NE "Total";
CLASS group;
VAR HwTotal ExTotal CourseTL;
TITLE "Objective 10.6";
TITLE2 "Group is identified as CLASS Variable";
RUN; QUIT;

/* The REPORT Procedure */
DM ' ODSRESULTS; CLEAR;';
PROC REPORT DATA = gradebook;
WHERE Student NE "Total";
COLUMN Student HWTotal ExTotal CourseTL Grade;
RBREAK AFTER / SUMMARIZE; * summarize at the end;
TITLE " Objective 10.7";
TITLE2 " REPORT Procedure with a RBREAK Statement";
RUN; QUIT;

* Does same thing as above;
PROC PRINT DATA = gradebook NOOBS LABEL;
WHERE Student NE "Total";
VAR Student HWTotal ExTotal CourseTL Grade;
SUM HWTotal ExTotal CourseTL;
TITLE " Objective 10.7";
TITLE2 " PRINT Procedure with a SUM Statement";
RUN; QUIT;
```

```
DM ' ODSRESULTS; CLEAR;';
PROC SORT DATA  = gradebook; BY Grade DESCENDING
CourseTL;
PROC REPORT DATA = gradebook;
WHERE Student NE "Total";
BY Grade;
COLUMN Student hwtotal extotal CourseTL grade;
RBREAK AFTER / SUMMARIZE;
TITLE " Objective 10.8";
TITLE2 "REPORT Procedure: Ordered Grade List";
RUN; QUIT;

* Does same thing as above;
PROC PRINT DATA = gradebook NOOBS LABEL;
WHERE Student NE "Total";
BY Grade;
VAR Student hwtotal extotal CourseTL grade;
SUM hwtotal extotal CourseTL;
TITLE " Objective 10.8";
TITLE2 "PRINT Procedure: Ordered Grade Lists";
RUN; QUIT;

DM ' ODSRESULTS; CLEAR;';
PROC REPORT DATA = gradebook;
WHERE Student NE "Total";
COLUMN Student hwtotal extotal CourseTL grade;
DEFINE Grade / GROUP;
DEFINE CourseTL / ORDER DESCENDING;
TITLE " Objective 10.9";
TITLE2 "REPORT Procedure: One Ordered Grade List";
TITLE3 " Option 1";
RUN; QUIT;

* Different arrangement of variabels;
PROC REPORT DATA = gradebook;
WHERE Student NE "Total";
COLUMN Student Grade hwtotal extotal CourseTL;
DEFINE Grade / GROUP;
DEFINE CourseTL / ORDER DESCENDING;
TITLE " Objective 10.9";
TITLE2 "REPORT Procedure: One Ordered Grade List";
TITLE3 " Option 2";
RUN; QUIT;

DM 'ODSRESULTS; CLEAR;';
PROC REPORT DATA = gradebook;
WHERE Student NE "Total";
COLUMN Student hw1 hw2 hwtotal;
DEFINE hw1 / 'Homework 1';
DEFINE hw2 / 'Homework 2';
TITLE " Objective 10.10";
TITLE2 "Report of Homework Scores";
RUN; QUIT;

* Ordered hwtotal in above code;
PROC REPORT DATA = gradebook;
WHERE Student NE "Total";
COLUMN Student hw1 hw2 hwtotal;
DEFINE hw1 / 'Homework 1';
DEFINE hw2 / 'Homework 2';
DEFINE hwtotal / 'Ordered Homework Total' ORDER;
TITLE " Objective 10.11";
TITLE2 "Report of Homework Scores";
RUN; QUIT;

DM 'ODSRESULTS; CLEAR;';
PROC REPORT DATA = gradebook;
WHERE Student NE "Total";
COLUMN Group Student id CourseTL grade;
DEFINE group / GROUP 'Project Group';
DEFINE id / ID 'ID #';
DEFINE hwtotal / 'Ordered Homework Total' ORDER;
TITLE " Objective 10.12";
RUN; QUIT;

PROC REPORT DATA = gradebook;
WHERE Student NE "Total";
COLUMN Group CourseTL;
DEFINE group / GROUP 'Project Group';
DEFINE CourseTL / 'Project group course average' MEAN ;
TITLE " Objective 10.13";
RUN; QUIT;
```

Chapter 11: Regression and Correlation Anaysis

```
/* Chapter 11 */
/* The REG Procedure */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
DATA beef;
INFILE "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\Data\BookData\beef data set Table
11_2.sas";
INPUT DMI ADG CWT BackFat REA @@;
PROC REG DATA = beef;
MODEL cwt = dmi;
TITLE "Objective 11.1";
RUN; QUIT;

PROC REG DATA = beef PLOTS = NONE;
MODEL cwt = dmi/ CLB ALPHA = 0.01;
```

```
slope9_2: TEST dmi = 9.2;
Intercept500: TEST INTERCEPT = 500;
TITLE "Objective 11.2";
RUN; QUIT;

DATA Obj11_3;
INPUT DMI @@;
DATALINES;
18 20 22
;
DATA beef2;
SET Obj11_3 beef;
RUN; QUIT;
PROC REG DATA = beef2 PLOTS (ONLY) = FIT;
MODEL cwt = dmi / P CLI CLM ALPHA = 0.01;
ID dmi;
TITLE "Objective 11.3";
RUN; QUIT;

/* The CORR Procedure */
PROC CORR DATA = beef;
VAR dmi adg cwt backfat rea;
TITLE "Objective 11.4";
RUN; QUIT;

PROC CORR DATA = beef;
VAR cwt;
WITH backfat rea;
TITLE "Objective 11.5 (Option 1)";
RUN; QUIT;

PROC CORR DATA = beef;
VAR backfat rea;
WITH cwt;
TITLE "Objective 11.5 (Option 2)";
RUN; QUIT;

PROC CORR DATA = beef SPEARMAN PLOTS =
MATRIX(HISTOGRAM);
VAR dmi adg cwt;
TITLE "Objective 11.6";
RUN; QUIT;
```

Chapter 12: SAS Libraries and Permanent Data Sets

```
/* Chapter 12 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE "Objective 12.1";
DATA one;
INPUT X Y Z @@@;
DATALINES;
25 27 34 28 31 29 41 58 29 37 28 83
;
RUN; QUIT;

LIBNAME SasLib "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\Course Book";
DATA SasLib.one;
INPUT X Y Z @@;
DATALINES;
25 27 34 28 31 29 41 58 29 37 28 83
;
PROC PRINT DATA = SasLib.one;
TITLE "Objective 12.2";
RUN; QUIT;

/* Using Permanent SAS Library */
LIBNAME SasLib "C:\Users\bmishra\Dropbox\OSU\PhD\Fall
2021\STAT5193 SAS R\Course Book";
/* Run the Procedure After this line */
PROC MEANS DATA = SasLib.one;
RUN; QUIT;
```

Chapter 13: Data Step Information 4 - SAS Probability Functions

```
/* Chapter 13 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
/* 13.1: Discrete Probability Distribution */
/* 13.1.1: Binomial Distribution */
* PROBBNML (p, n, m);
* p = Probability of Success;
* n = # of Independent Trials, n > 0.
* m = # of Success in Trials 0 <= m <= n;
DATA one;
a = PROBBNML (0.3, 8, 5);
                        * P(Y <= 5);
b = PROBBNML (0.3, 8, 5) - PROBBNML (0.3, 8, 4);
* P(Y = 5);
c = PROBBNML (0.3, 8, 6) - PROBBNML (0.3, 8, 3);
          * P(4 <= Y = 6);
d = 1 - PROBBNML (0.3, 8, 3);              * P(Y > 3) =
P(Y >= 4) = 1 - P(Y <= 3);
PROC PRINT DATA = one;
TITLE "Objective 13.1: Binomial Probability
Distribution";
RUN; QUIT;

/* 13.1.2: Poisson Distribution */
* POISSON (m, n);
* m = Mean, m > 0;
* n = Value of random variable Y, n = 0 ... INF;
DATA two;
```

```
a = POISSON(5.2, 7);                              * P(Y <=
7);
b = POISSON(5.2, 7) - POISSON(5.2, 6);      * P(Y = 7);
c = 1 - POISSON(5.2, 0);                          * P(Y =
0);
PROC PRINT DATA = two;
TITLE " Objective 13.2: Poisson Probability
Distribution";
RUN; QUIT;

/* 13.2: Continuous Probability Distribution */
/* 13.2.1: The Normal Distribution */
* PROBNORM(x);
* x = Numeric value of standard normal random variable.
* PROBIT(p);
* p = Numeric Probability, 0 < p < 1;

DATA three;
a = PROBNORM(-1) - PROBNORM(-2);    * p(Z < -1) - p(Z <
-2);
b = 1 - PROBNORM(2);                        * p(Z > 2) = 1
= p(z < 2);
c = PROBIT (0.90);                              * 0.90 =
p( Z < c);
d = PROBNORM(1.645);                        * p(Z <
1.645);
;

PROC PRINT DATA = three;
TITLE " Objective 13.3";
RUN; QUIT;

/* 13.2.2: The t Distribution */
* PROBT (x, df, <, nc>);
* x = numeric random variable.
* df = degree of freedom;
* nc = non-centrality parameter, nc >= 0;

/* Inverse of T-Distribution */
* TINV (p , df, <,nc>);
* p = numeric probability, 0 < p < 1;
* df = degree of freedom , df > 0;
* nc = Non-centrality parameter, nc >= 0;

/* 13.2.3: Chi Squared Distribution */
* PROBCHI (x, df, <,nc>);
* x = numeric random variable.
* df = degree of freedom;
* nc = non-centrality parameter, nc >= 0;

/* Inverse of Chi Squared Distribution */
* CINV (p , df, <,nc>);
* p = numeric probability, 0 < p < 1;
* df = degree of freedom , df > 0;
* nc = Non-centrality parameter, nc >= 0;

/* 13.2.2: The F Distribution */
* PROBF (x, ndf, ddf, <, nc>);
* x = numeric random variable.
* ndf = Numerator degree of freedom;
* ddf = Denominator degree of freedom;
* nc = non-centrality parameter, nc >= 0;

/* Inverse of F Distribution */
* FINV (p, ndf, ddf, <,nc>);
* p = numeric probability, 0 < p < 1;
* ndf = Numerator degree of freedom;
* ddf = Denominator degree of freedom;
* nc = Non-centrality parameter, nc >= 0;

DATA four;
a = TINV (0.88, 14);                    * p(t > a) = 0.12 or
0.88 = p(t < a), df = 14;
b = 1 - PROBT (2.104, 20);              * p( t > 2.104, df =
20) = b;
c = CINV (0.96, 14);                    * p(x2 > c) = 0.04
or p(x2 < c) = 0.96;
d = 2 *(1- PROBCHI(17.04, 6));  * 2p(x2 > 17.04) = d,
df = 6;
e = 1 - PROBF (7.83, 4, 16);  * 1 - p (F > 7.78) = e,
ndf = 4, ddf = 16;
f = FINV(0.97, 4, 16);                  * 0.03 = P(F > f),
ndf = 4, ddf = 16 ;

PROC PRINT DATA = four;
TITLE " Objective 13.4";
RUN; QUIT;
```

Chapter 14: Reading and Writing Data Files

```
/* Chapter 14 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
/* The IMPORT Procedure */
FILENAME fileref "path\to\file\file.xlsx";
PROC IMPORT OUT = WORK.DEMO
DATAFILE = "path\tofile\file.xlsx" (or fileref)
DBMS = EXCEL REPLACE;
SHEET = "Sheet1$";
GETNAMES = YES;
MIXED = NO;
SCANTEXT = YES;
```

```
USEDATA = YES;
SCANTIME = YES;
STARTROW = 10; /* In older version of excel */
LABEL X = " Label for variable x"
      y = " Label for variable y";
RUN; QUIT;

/* The EXPORT Procedure */
FILENAME fileref "path\to\file\file.xlsx";
PROC EXPORT OUT = WORK.DEMO
DATAFILE = "path\tofile\file.xlsx" (or fileref)
DBMS = EXCEL REPLACE;
SHEET = "Sheet Name";
RUN;  QUIT;
```

Chapter 15: DATA Step Information 5 – DO Loops, ARRAY, and Random Number Generators

```
/* Chapter 15 Do loops, ARRAY, RANNOR*/
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
/* Do Loops- DO and END Statements */

TITLE "Objective 15.1 With OUTPUT Statement";
DATA loopex1;
DO Trt = 1 to 6;
DO Rep = 1 to 2;
INPUT y @@;
OUTPUT;
END;
END;
DATALINES;
12 14 18 16 12 11 17 19 20 11 13 15
;
PROC PRINT DATA = loopex1;
RUN; QUIT;

TITLE "Objective 15.2 Without OUTPUT Statement";
TITLE2 "Only last loop output was saved";
DATA loopex1_2;
DO Trt = 1 to 6;
DO Rep = 1 to 2;
INPUT y @@;
END;
END;
DATALINES;
12 14 18 16 12 11 17 19 20 11 13 15
;
PROC PRINT DATA = loopex1_2;
RUN; QUIT;

TITLE "Objective 15.3";
DATA loopex2 loopex3;
DO Trt = 1 to 6;
DO Rep = 1 to 2;
INPUT y @@;
y_log = log(y);
IF Trt <= 3 THEN OUTPUT loopex2;
ELSE OUTPUT loopex3;
END;
END;
DATALINES;
12 14 18 16 12 11 17 19 20 11 13 15
;
PROC PRINT DATA = loopex2;
TITLE "Objective 15.3 Trt 1-3 Data";
PROC PRINT DATA = loopex3;
TITLE "Objective 15.3 Trt 4-6 Data";
RUN; QUIT;

* TITLE "Objective 15.4";
DATA four;
DO p = 0.4, 0.45, 0.5;
DO y = 0 TO 6;
cp = PROBBNML (p, 6, y); *(y <= y);
IF y = 0 THEN prob = cp;
ELSE prob = PROBBNML (p, 6, y) - PROBBNML (p, 6, y-1);
* P (y = y);
OUTPUT;
END; END;
lABEL Y = "Y Success in 6 Trials"
cp = "Cumulative Probability, P(Y <= y)"
prob = " Probability, P(Y = y)";
PROC SORT DATA = four; BY p y;
PROC PRINT DATA = four LABEL NOOBS; BY p;
VAR y prob cp;
TITLE "Objective 15.4";
RUN; QUIT;

* TITLE "Objective 15.4";
DATA a;
r = 16;
DO WHILE (r < 20);
r + 1;
OUTPUT;
END;
PROC PRINT DATA = a;
TITLE "Objective 15.4";
RUN; QUIT;

DATA a;
r = 16;
```

```
DO WHILE (r < 20);
r + 1;
END;
PROC PRINT DATA = a;
TITLE "Objective 15.4";
RUN; QUIT;

DATA a;
r = 16;
DO WHILE (r <= 20);
r + 1;
OUTPUT;
END;
PROC PRINT DATA = a;
TITLE "Objective 15.4";
RUN; QUIT;

DATA a;
r = 16;
DO WHILE (r < 20);
r + 1;
END;
PROC PRINT DATA = a;
TITLE "Objective 15.4";
RUN; QUIT;

DATA a;
r = 16;
DO WHILE (r >= 20);
r + 1;
OUTPUT;
END;
PROC PRINT DATA = a;
TITLE "Objective 15.4";
RUN; QUIT;

DATA a;
r = 16;
DO WHILE (r > 20);
r + 1;
END;
PROC PRINT DATA = a;
TITLE "Objective 15.4";
RUN; QUIT;

/* The ARRAY Statement */
DATA a;
INPUT y x1 x2 x3;
ARRAY v{3} x1 x2 x3;
DATALINES;
;
RUN; QUIT;

/* Use of ARRAY */
DATA beef2;
SET beef;
Adg_kg = adg/2.2046;
Dmi_kg = dmi/2.2046;
Cwt_kg = cwt/2.2046;
RUN; QUIT;
/* Equivalently */
DATA beef2;
SET beef;
ARRAY b{3} dmi adg cwt;
ARRAY c{3} Dmi_kg Adg_kg Cwt_kg;
Do i = 1 TO 3;
c{i} = b {i} / 2.2046;
END; RUN; QUIT;

* Notes:
Double dashes (--) between variables be used to denote
old variables.
Double dashes (--) between variables cannot be used to
define new variables.
Generate: number of new variables = number of old
variabels.;

/* Random Number Generator */
TITLE "Objective 15.7";
DATA seven;
seed1 = 2120;
seed2 = 2120;
seed3 = 2120;
DO i = 1 to 10;
CALL RANNOR (seed1, x1);
CALL RANNOR (seed2, x2);
y1 = RANNOR (seed3);
y2 = RANNOR (2120);
IF i = 6 THEN DO;
seed2 = 17;
seed3 = 17;
END; OUTPUT; END;
PROC PRINT DATA =seven;
TITLE "objective 15.7";
RUN; QUIT;

/* Continuous Distributions */
/* Normal Distribution */
DATA cont;
x = RANNOR(seed);
```

```
CALL RANNOR (seed, x);

/* Uniform Distribution */
x = RANNOR(seed);
CALL RANUNI(seed, x);

/* Cauchy Distribution */
x = RANNOR(seed);
CALL RANCAU(seed, x);

/* Exponential Distribution */
x = RANEXP(seed);
CALL RANUNI(seed, x);

/* Gamma Distribution */
x = RANGAM(seed);
CALL RANUNI(seed, x);

DATA eight;
DO i = 1 to 10;
X = RANNOR(28374); * X ~ N(0, 1);
Y = 5 + RANNOR(39587209); * Y ~ N(5, 1);
W = SQRT(6) * RANNOR(659363); * W ~ N(0, 6);
U = 8 + SQRT(10) * RANNOR(494703); * U ~ N(8, 10);
OUTPUT;
END;
RUN;
PROC PRINT DATA = eight NOOBS N;
TITLE "Objective 15.8";
RUN; QUIT;

DATA eight;
SEED = 6474983
DO i = 1 to 10;
CALL RANNOR (seed, x); * X ~ N(0, 1);
X = 5 + x ; * Y ~ N(5, 1);
W = SQRT(6) * X; * W ~ N(0, 6);
U = 8 + SQRT(10); * U ~ N(8, 10);
OUTPUT;
END;
TITLE "Objective 15.8";
RUN; QUIT;

/* Discrete Distributions */
/* Binomial Distribution */
x = RANBIN (seed, n, p);
CALL RANBIN (seed, n, p, x);
/* Poisson Distribution */
x = RANPOI (seed, m);
CALL RANPOI (seed, m, x);
/* Tabled Probability Distribution */
x = RANTBL (seed, p1, p2, ..., pn);
CALL RANTBL (seed, p1, p2, ..., pn, x);

DATA nine;
DO Sample = 1 TO 2;
DO Day = 1 to 5;
C = RANPOI(739284, 14); * RANPOI(seed, mean);
OUTPUT;
END; END;
PROC MEANS DATA = nine;
CLASS sample;
VAR c;
TITLE 'Objective 15.9 - Option 1';
RUN; QUIT;

PROC MEANS DATA = nine;
BY sample;
VAR c;
TITLE 'Objective 15.9 - Option 2';
RUN;
```

Chapter 16: Statistical Graphics Procedure

```
/* Chapter 16 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
/* SGPLOT procedure */
PROC SGPLOT DATA = beef3;
HBAR sex;
TITLE 'Objective 16.1 - Default Horizontal Bar Chart';
RUN;

PROC SORT DATA = beef3; BY Producer;
PROC SGPLOT DATA = beef3; BY Producer;
HBAR sex;
TITLE 'Objective 16.2 - BY Producer';
RUN;

PROC SGPLOT DATA = beef3;
WHERE Producer = 3;
HBAR sex / RESPONSE = adg STAT = MEAN ALPHA = 0.10
LIMITS = BOTH ;
TITLE 'Object-ive 16.3 - Bar Length is ADG mean' ;
TITLE2 '90% CI for ADG Mean - Producer 3';
RUN;

PROC SGPLOT DATA = beef ;
WHERE Producer = 3;
HBAR sex / RESPONSE = adg STAT = MEAN FILLATTRS = (COLOR
= BLACK) ;
```

```
HbAR sex / RESPONSE = dmi STAT = MEAN BARWIDTH = 0.70
TRANSPARENCY = 0.50
FILLATTRS =(COLOR = BLACK);
TITLE 'Objective 16.4 - Overlaying Bar Charts - Producer
3';
RUN;

PROC SGPLOT DATA = beef3;
HISTOGRAM cwt ;
TITLE 'Objective 16.5 - Default Histogram';
RUN;

PROC SGPLOT DATA = beef3;
HISTOGRAM cwt;
DENSITY cwt / TYPE = N0RMAL;
TITLE 'Objective 16.6';
RUN;

PROC SGPLOT DATA = beef3;
VBOX backfat/CATEGORY = producer CONNECT = MEAN;
TITLE 'Objective 16.7';
RUN;

PROC SGPLOT DATA = beef3;
WHERE producer = 3 ;
SCATTER Y = cwt X = dmi / MARKERATTRS =(COLOR = BLACK
SYMBOL = CIRCLEFILLED) ;
TITLE 'Objective 16.8 - Option 1';
RUN;
PROC SGPLOT DATA = beef3;
WHERE producer = 3 ;
SCATTER Y = CWt X = dmi / DATALABEL = sex;
TITLE 'Objective 16.8 - Option 2';
RUN;
PROC SGPLOT DATA = beef3;
WHERE producer = 3;
SCATTER Y = cwt X = dmi / MARKERCHAR = sex;
TITLE 'Objective 16.8 - Option 3';
RUN;

PROC SGPLOT DATA = beef3means;
WHERE _STAT_ = "MEAN";
SERIES Y = cwt X = producer / GROUP = sex LINEATTRS =
{COLOR = BLACK PATTERN = 3);
SCATTER Y = cwt X = producer / MARKERCHAR = sex;
TITLE 'Objective 16.9';
RUN;

/* SGSCATTER procedure */
PROC SGSCATTER DATA = beef3;
COMPARE Y = (cwt rea backfat) X = (dmi adg) / GRID
MARKERATTRS = (SYMBOL = CIRCLE COLOR = BLACK) ;
TITLE 'Objective 16.10';
RUN;

PROC SGSCATTER DATA = beef3;
MATRIX cwt rea backfat dmi adg /MARKERATTRS = (SYMBOL =
CIRCLE COLOR = BLACK) ;
TITLE 'Objective 16.11';
RUN;

/* SGPANEL procedure */
PROC SGPANEL DATA = beef3;
PANELBY producer;
VBAR sex / FILLATTRS = (COLOR = WHITE);
TITLE 'Objective 16.12';
RUN;

PROC SGPANEL DATA = beef3;
PANELBY producer / LAYOUT = COLUMNLATTICE ;
VBAR Sex / RESPONSE adq STAT = MEAN FILLATTRS = (COLOR =
LIGHTGRAY) ;
TITLE 'Objective 16.13' ;
RUN;

PROC SGPANEL DATA = beef3;
PANELBY producer sex / LAYOUT = LATTICE;
HISTOGRAM cwt;
DENSITY cwt / TYPE = NORMAL;
TITLE 'Objective 16.14';
RUN;

PROC SGPANEL DATA = beef3;
PANELBY producer sex / LAYOUT = LATTICE ROWHEADPOS =
BOTH;
VBOX backfat / CATAGORY = producer CONNECT = MEAN;
TITLE 'Objective 16.15',-
RUN;
```

Chapter 17: SAS/GRAPH Procedures

```
/* Chapter 17 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE "chapter 17, Bijesh Mishra";
DATA beef;
Input DMI ADG CWT BackFat REA;
DATALINES;
17.0090 3.05 819.000 0.54571 13.3821
21.4372 3.25 859.000 0.52818 13.9649
;
```

```
PROC GCHART DATA=demo.beef 3 ;
VBAR sex / GROUP=producer;
TITLE 'Objective 17.1';
RUN;

TITLE;
PATTERN;
GOPTIONS FTEXT="Times New Roman";

PROC GCHART DATA = demo.beef3;
VBAR producer / DISCRETE TYPE = MEAN SUMVAR = rea
PATTERNID = MIDPOINT,
TITLE 'Objective 17.2 - Chart 1';
PATTERN1 COLOR = BLACK VALUE = SOLID;
PATTERN2 COLOR = BLACK VALUE = EMPTY;
PATTERN3 COLOR = BLACK VALUE = X3;
RUN;

VBAR producer / DISCRETE TYPE = MEAN SUMVAR = rea
PATTERNID = MIDPOINT
GROUP = sex;
TITLE 'Objective 17.2 - Chart 2';
RUN;

VBAR sex / TYPE = MEAN SUMVAR = rea PATTERNID = GROUP
GROUP = producer;
TITLE 'Objective 17.2 - Chart 3' ;
RUN;

TITLE;
PATTERN;
GOPTIONS PTEXT = "Times New Roman" HTEXT = 1.5;

PROC GCHART DATA = demo.beef3 ;
VBAR producer / DISCRETE SUBGROUP = sex PATTERNID =
SUBGROUP;
TITLE 'Objective 17.3';
PATTERN1 COLOR = GRAY VALUE = SOLID ; *heifers;
PATTERN2 COLOR = GRAY VALUE = EMPTY ; *steers;
RUN;

PROC GCHART DATA = demo.beef3;
VBAR sex / SUBGROUP = producer PATTERNID = SUBGROUP;
PATTERN1 COLOR = GRAY VALUE = SOLID ; * Producer=l ;
PATTERN2 COLOR = GRAY VALUE = EMPTY ; *Producer=2;
PATTERN3 COLOR = GRAY VALUE = X3 ; *Producer=3;
RUN;

GOPTIONS FTEXT = SWISSB HTEXT = 1.5;
PROC GCHART DATA = demo.beef3;
PIE producer / DISCRETE TYPE = PERCENT VALUE = INSIDE
SLICE = OUTSIDE;
TITLE 'Objective 17.4 Pi Chart - Percent';
PATTERN1 COLOR = GRAY VALUE = EMPTY;
PATTERN2 COLOR = GRAY VALUE = SOLID;
PATTERN3 COLOR = SILVER VALUE = EMPTY;
RUN;

PIE producer / DISCRETE TYPE = FREQ VALUE = INSIDE;
TITLE 'Objective l7.4: Pie Chart - FREQ';
RUN;

GOPTIONS RESET=ALL;
PROC GPLOT DATA=demo.beef3 ;
WHERE producer = 3;
PLOT cwt * dmi ;
SYMBOL1 VALUE = DOT CV = BLACK I = NONE;
TITLE 'Objective 17.5';
RUN;

AXIS1 ORDER = (700 TO 1000 BY 50) LABEL = (A=90);
PROC GPLOT DATA = demo.beef3;
PLOT cwt * dmi = sex / VAXIS = AXISl;
SYMBOL1 VALUE = CIRCLE CV = BLACK I = NONE; *heifers;
SYMBOL2 VALUE = TRIANGLE CV = BLACK I = NONE; *steers;
TITLE "Objective 17.6 - Option 1";
RUN;

PLOT cwt * dmi = sex / VAXIS = AXIS1 GRID;
SYMBOL1 VALUE = "H" CV = BLACK I = NONE;
SYMBOL2 VALUE = "S" CV = BLACK I = NONE;
TITLE 'Objective 17.6 - Option 2';
RUN;

PROC GPLOT DATA = demo.beef3means;
WHERE _STAT_ = "MEAN";
PLOT cwt * producer = sex / VAXIS = AXIS1;
SYMBOL1 VALUE = DOT CV = BLACK I = JOIN L = 1;
SYMBOL2 VALUE = NONE C = BLACK I = JOIN L = 3;
TITLE 'Objective 17.7';
RUN;

LEGEND1 POSITION = (TOP INSIDE) LABEL = NONE;
LEGEND2 POSITION = ( TOP RIGHT OUTSIDE );
LEGEND3 POSITION = ( TOP RIGHT INSIDE ) ACROSS = 1;
LEGEND4 POSITION = (LEFT) ;
LEGEND5 POSITION = (LEFT) D0WN = 4 LABEL = ("Gender");

GOPTIONS FTEXT = "Arial" HTEXT = 1.5 CTEXT = BLACK;
AXIS2 ORDER = (700 TO 875 BY 25) ;
```

```
PROC GPLOT DATA = demo.beef3means ;
WHERE _STAT_ = "MEAN";
PLOT cwt * producer = sex / VAXIS = AXIS2 LEGEND =
LEGEND3 ;
SYMBOL1 VALUE = DOT CV = BLACK I = JOIN L = 1 ;
SYMBOL2 VALUE = NONE C = BLACK I = JOIN L = 3 ;
TITLE 'Objective 17.8 - Legend 3';
RUN;

PLOT cwt * producer = sex / VAXIS = AXIS2 LEGEND =
LEGEND5 ;
TITLE 'Objective 17.8 - Legend 5';
RUN;

GOPTIONS RESET = ALL FTEXT = ZAPF;
LEGEND1 POSITION = (TOP INSIDE) LABEL = NONE;
AXIS3 LABEL = NONE;
AXIS4 LABEL =(A = 90) ;
AXIS5 LABEL =(A = 270) ;

PROC GPLOT DATA = demo.beef3 ;
WHERE producer = 3 ;
plot (adg dmi) * rea / OVERLAY LEGEND = LEGEND1 VAXIS =
AXIS3 ;
SYMBOL1 VALUE = CIRCLE CV = BLACK I = NONE;
SYMBOL2 VALUE = DOT CV = BLACK I = NONE;
TITLE 'Objective 17.9 - OVERLAY Option';
RUN;

PLOT adg * rea / VAXIS = AXIS4 LEGEND = LEGEND6;
PLOT2 dmi * rea / LEGEND = LEGEND1 VAXIS = AXIS5;
TITLE 'Objective 17.9 - PL0T2 Statement';
RUN;

GOPTIONS RESET = ALL;
PROC G3D DATA = demo.beef3;
WHERE producer = 3 and sex = "H";
SCATTER adg * dmi = cwt ;
TITLE 'Objective 17.10 - Default G3D Scatter Plot';
RUN;

PROC SORT DATA=demo.beef3;
by sex;
PROC G3D DATA=detno.beef3 ;
WHERE producer = 3;
BY sex;
SCATTER adg * dmi = cwt / ZMIN = 700 ZMAX = 1000
                          ZTICKNUM = 7 GRID
                                                SHAPE = "PILLAR"
COLOR = "GRAY";
TITLE 'Objective 17.11';
RUN;

DATA twelve;
X = -2.5;
y = -1;
DO X = -2.5 TO 2.5 BY 0.1;
DO y = -1 TO 1 BY 0.1;
z =  X*X*X + 3*X*Y*Y + 3*Y*Y - 15*X;
OUTPUT;
END;
END;
RUN;

PROC G3GRID DATA = twelve;
grid y * X = Z;
PROC G3D DATA = twelve;
PLOT y * X = z / GRID;
TITLE 'Objective 17.12 - View 1' ;
RUN;

PROC G3GRID DATA = twelve;
grid X * y = z;
PROC G3D DATA = twelve;
PLOT X * y = z / GRID XYTYPE = 2
                                  XTICKNUM = 6
                                  ZMIN = -25
                                  ZMAX = 25
                                  ZTICKNUM = 11 ;
TITLE 'Objective 17.12 - View 2' ;
RUN; QUIT;

PROC GCONTOUR DATA = twelve;
PLOT X * y = z / AUTOLABEL NOLEGEND
                           LEVELS = -20 -15 -10 -5 0 5 10
15 20;
SYMBOL1 I = JOIN C = BLACK L = 1 REPEAT = 9;
TITLE 'Objective 17.13 - Contour Plot for View 2' ;
RUN;

ODS LISTING;
ODS LISTING CLOSE;
```

Chapter 18: Formatting Responses

```
/* Chapter 18 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE "Chapter 18, Bijesh Mishra";
DATA one;
```

```
INPUT ID exp Salary Region Gender BS MS PhD Group $;
salary2 = salary*1000; *convert salary in Sthousands to
dollars;
LABEL exp = "Work Experience" salary = "Salary
($thousands) "
region = "Region of US";
DATALINES;
000171831 8.5 41.5 1 1 1 1 2 E
077889999 10 53.4 3 2 1 1 1 E
111223333 13 65.0 4 . 1 1 2 B
222334444 20 75.0 2 3 2 1 1 S
/
RUN;
PROC FORMAT;
*** templates for the ID and Salary variables ***;
PICTURE a OTHER = '000-00-0000' (FILL = "0");
PICTURE b OTHER = '000.00';
PICTURE c OTHER = '000000' (PREFIX = '$');

*** create salary intervals for original salary variable
***;
VALUE salfmt low - 49.9 = 'Below 50,000'
             50.0 - 59.9 = '50,000 - 60,000'
             60 - high = '60,000 and over';

             *** Identify numerical regions to
parts of the US ***;
VALUE regfmt 1 = 'Northwest'
             2 = 'Central'
             3-4 = 'Southern';

*** Note that a format can assist with missing or
miscoded values***;
VALUE gen 1 = 'Male'
          2 = 'Female'
          . = 'Missing value'
      OTHER = 'Miscoded';

*** Format can translate numerica values into antoher
language if needed ***;
VALUE degree 1 = "yes" 2 = "No";
VALUE degree 1 = "Da" 2 = "Nyet"; * Yes/No Russian;
VALUE degree 1 = "Si" 2 = "No"; * Yes/No Spanish;

*** $ is needed to create a format when the responeses
are character strings ***;
VALUE $grp "E" =  "Engineering"
           "B" =  "Business"
           "S" =  "Science";
```

```
PROC PRINT DATA:= one ;
FORMAT id a. salary b.;
TITLE 'Objective 18.1' ;
RUN;

PROC PRINT DATA=one LABEL;
VAR id exp salary salary2 region gender bs ms phd group;
format id SSN. salary salfmt. salary2 c. region regfmt.
gender gen.
bs ms phd degree, group $grp.;
TITLE 'Objective 18.2' ;
RUN;

PROC FORMAT;
VALUE likert 1 = "Strongly Disagree"
             2 = "Disagree"
             3 = "Neutral"
             4 = "Agree"
             5 = "Strongly Agree";
PROC PRINT DATA = one;
VAR id salary2 salary;
FORMAT is SSN.salary2 DOLLAR9.2 salary 7.3;
TITLE 'Objective 18.3';
RUN;

LIBNAME demo ' G:\CLGoad\' ;
PROC FORMAT;
VALUE $sexfmt "H" = "Heifers" "S" = "Steers";
VALUE pfmt 1= "Rocking K Ranch" 2="Superior Beef Co." 3
= "Royal Beef
Producers";
GOPTION RESET = ALL; * restore the default graph
settings;

PROC GCHART DATA = demo.beef 3;
VBAR sex / GROUP = producer ;
TITLE 'Objective 18.4' ;
PATTERN1 VALUE = SOLlD COLOR = GRAY;
FORMAT sex $sexfmt. producer pfmt. ;
RUN;

PROC MEANS DATA=demo.beef3 MEAN RANGE;
CLASS sex producer;
VAR dmi rea;
TITLE 'Object:ive 18.5';
FORMAT sex $sexfnit. producer pfmt. ;
*CLASS producer sex; * Change class to see difference;
RUN;
```

Chapter 19: Output Delivery System (ODS)

```
/* Chapter 19 */
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE "Chapter 19, Bijesh Mishra";
data instruction ;
INPUT program $ score @@;
DATALINES;
A 71 A 82 A 88 A 64 A 59 A 78 A 72
A 81 A 83 A 66 A 83 A 91 A 79 A 70
B 65 B 88 B 92 B 76 B 87 B 89 B 85
B 90 B 81 B 91 B 78 B 81 B 86 B 82
B 73 B 79
i
ODS TRACE ON;
PROC UNIVARIATE DATA = instruction CIBASIC ALPHA = 0.01
NORMAL;
CLASS program;
VAR score;
HISTOGRAM score / NORMAL;
TITLE 'Objective 19.1 - Tracing Objective 3 2';
RUN;
ODS TRACE OFF;
QUIT;

ODS LISTING;
ODS TRACE ON / LISTING;
PROC UNIVARIATE DATA= inst:ruction CIBASIC ALPHA=0.01
NORMAL;
CLASS program;
VAR score;
HISTOGRAM score / NORMAL;
TITLE 'Objective 19.1 - Tracing Objective 3.2';
RUN;
ODS TRACE OFF;

PROC UNIVARIATE DATA = instruction CIBASIC ALPHA = 0.01
NORMAL;
CLASS program;
VAR score;
HISTOGRAM score / NORMAL;
TITLE 'Objective 19.2 - SELECT Output';
ODS HTML SELECT TESTSFORNORMALITY HISTOGRAM;
RUN;

PROC UNIVARIATE DATA = instruction CIBASIC ALPHA = 0.01
NORMAL;
CLASS program;
VAR score;
HISTOGRAM score / NORMAL;
TITLE 'Objective 19.3 - EXCLUDE Output';
ODS HTML EXCLUDE EXTREMEOBS;
RUN;

ODS LISTING CLOSE;
PROC UNIVARIATE DATA = instruction CIBASIC ALPHA = 0.01
NORMAL;
CLASS program;
VAR score;
HISTOGRAM score / NORMAL;
ODS OUTPUT MOMENTS = basicstat TESTFORLOCATION =
meantests;
TITLE "Objective 19.4";

PROC PRINT DATA = Basicstat;
TITLE2 "The Recovered Moments Table";
PROC PRINT DATA = meantests LABEL;
TITLE2 "The Recovered Tests for Moment Table";
RUN; QUIT;

PROC UNIVARIATE DATA = instruction CIBASIC ALPHA = 0.01
NORMAL;
CLASS program;
VAR score;
HISTOGRAM score / NORMAL;
ODS HTML SELECT NONE; *or ODS HTML EXCLUDE ALL;
ODS OUTPUT BASICINTERVALS = ci99;
TITLE 'Objective 19.5';
RUN;
PROC PRINT DATA = ci99;
TITLE2 '99% CIs for Mean, Std Dev, and Variance';
ODS HTML SELECT ALL;
RUN;

DATA Ci99_mean ci99_sd ci99_var;
SET Ci99;
IF PARAMETER = "Mean" THEN OUTPUT ci99_mean;
IF PARAMETER = "Std Deviation" THEN OUTPUT ci99_sd;
IF PARAMETER = "Variance" THEN OUTPUT ci99_var;
RUN; QUIT;

LIBNAME ch19 "G:\CGoad" ;
ODS PDF FILE = 'G:\CGoad\Obj19_6.pdf';
PROC UNIVARIATE DATA = instructiion CIBASIC ALPHA = 0.01
NORMAL;
CLASS program;
VAR score;
HISTOGRAM score / NORMAL;
ODS HTML SELECT ALL;
ODS PDF SELECT TESTSFORNORMALITY HISTOGRAM;
ODS OUTPUT MOMENTS = ch19.basicstats TESTSF0RL0CATI0N =
ch19.meantests;
```

```
TITLE 'Objective 19.6' ;
RUN;
ODS PDF CLOSE;
PROC PRINT DATA = ch19.basicstats;
TITLE2 'The Recovered Momerits Table' ;
PROC PRINT DATA=ch19.meantests LABEL;
TITLE2 'The Recovered Tests for Location Table' ;
RUN;
QUIT;
```

Chapter 20: Miscellaneous Topics (Not Summarized)

Book Summary:

Chapter 1: Introduction

```
DM 'LOG; CLEAR; ODSRESUTLS; CLEAR;';
TITLE; FOOTNOTE; ODS GRAPHICS OFF/ON;
RUN; QUIT;
```

Chapter 2: Data Step

```
DATA dataset;
INPUT var1 $ 1-5 var2 6-8 var3 9-13;
DATALINES;

PROC PRINT <options>;
VAR var1 var2;
ID var1;
BY var2;
SUM var3;
TITLE 'TITLE'; Up to 10 titles.
FOOTNOTE 'FOOTNOTE'; Up to 10 footnotes.

PROC PRINT options are:
DATA =: dataset (Print Dataset)
DOUBLE: Double spaced printing.
NOOBS: No observation number.
UNIFORM: Uniform column width.
LABEL: Print variable labels.
N: Print number of observations. Also use with BY
statement but Sort data first.

PROC SORT DATA = dataset;
BY var1 DESCENDING var2 var3;
```

Chapter 3: Summarizing Data Basics:

3.1: The Univariate Procedure

```
PROC UNIVARIATE DATA = dataset <options>;
BY var1 DESCENDING var2 var3;
WHERE condition;
CLASS variable(s);
VAR var1 var2 </options>
OUTPUT OUT = output_dataset <options and new variable
names>;

PROC UNIVARIATE statement Options:
ALPHA = 0.05: Type I error probability.
CIBASIC: Confidence interval for means
FREQ: Frequency table
MU0 =: Ho to test mean and location
NORMAL: Tests Ho: Population normally distributed.
NOPRINT: Suppresses output table (used with OUTPUT
statement)

HISTOGRAM Statement options:
NORMAL: Overlay normal curve on histogram
MIDPOINTS: midpoints for histogram bars.

OUTPUT Statement options:
PROC UNIVARIATE DATA = dataset;
VAR var1 var2;
OUTPUT OUT = output_dataset stat1 = stat1var1 stat1var2
stat2 = stat2var1 stat2var2;

Stat can be: N, NMISS, NOOBS, MIN, MAX, RANGE, SUM,
MEAN, VAR(iance), STD, STDMEAN (Std. Er.), CV, MEDIAN,
Q1, Q3, P1, P5, P10, P90, P95, P99, MODE, T (t-Stat),
PROB (p-value).
```

3.2: The Means Procedure

```
PROC MEANS DATA = dataset <options>;
BY var1 DESCENDING var2 var3;
WHERE condition;
CLASS variable(s);
VAR var1 var2 </options>
OUTPUT OUT = output_dataset <options and new variable
names>;

PROC MEANS Statement (only few) Options:
N, NMISS, NOOBS, MIN, MAX, RANGE, SUM, MEAN, VAR(iance),
STD, STDERR, CV, T (t-Stat), PRT (p-value), ALPHA = P,
CLM, UCLM, LCLM, NOPRINT
```

Chapter 4: Data Step

Commonly used SAS DATA Step Functions:

Algebraic functions:

```
MOD: y = MOD(dividend, divisor) remainder.
SIGN: y = SIGN(argument) sign or 0.
SQRT: y = SQRT(argument)Square root
ABS: y = ABS(argument) Absolute value
INT: y = INT(argument) greatest integer.
ROUND: y = ROUND(argument) nearest round.
EXP: y = EXP(argument)Exponential.
LOG: y = LOG(argument) natural log
LOG2: y = LOG2(argument) Log base 2
LOG10: y = LOG10(argument) log base 10.
```

Statistical functions:

```
MEAN: y = MEAN(argument) Arithmetic mean
STD: y = STD(argument) Standard Deviation
MAX: y = MAX(argument)Maximum value
MIN: y = MIN(argument)Minimum value
N: y = N(argument)Counts non-missing values
NMISS: y = NMISS(argument)Counts Missing
```

Trigonometric functions:

SIN: y = SIN(argument) Sine of angle
COS: y = COS(argument) Cosine of angle
TAN: y = TAN(argument)Tangent of angle
ARCOS: y = ARCOS(argument)Inverse cosine
ARSIN: y = ARSIN(argument) Inverse Sine
ATAN: y = ATAN(argument) Inverse tangent

Text and Miscellaneous functions:

UPCASE: y = UPCASE(argument)UPPER CASE
LOWCASE: y = LOWCASE(argument) lower case
PROPCASE: y = PROPCASE(argument)Proper Case
_n_: y = _n_ Value of observation
LAGn y = LAGn(argument) N^{th} Lag Value

Syntax for Specifying Conditions:

Condition	Symbol/Text	Simple Syntax
Equal	EQ, =	X = 5; X EQ 5
Not Equal	^= NE	Color NE 'Red'
Less than	< LT	Age LT 20
Greater Than	> GT	Age GT 40
LT or EQ	<= LE	Age LE 20
GT or EQ	>= GE	Age GE 40
Inclusion	IN	Time IN (3, 6)
OR	OR	X < 3 OR X GE 7
AND	AND	X < 3 AND X GE 7

IF THEN, ELSE, AND WHERE Statements:

```
IF length < 60 THEN size = 'small';
ELSE size = 'Large';
IF Condition THEN DELETE;
IF Condition THEN KEEP;
PROC PRINT DATA = dataset;
WHERE age LE 15;
```

SET and MERGE Statements:

```
PROC SORT DATA = dataset1; BY var1;
PROC SORT DATA = dataset2; BY var1;
DATA newdataset;
SET olddataset;
MERGE dataset1 dataset2; by var1;
```

Chapter 5: Beginning Charts

The GCHART Procedure

```
PROC GCHART DATA = dataset;
BY var(s);
WHERE conditions;
VBAR vars/<Options>;
HBAR vars/<Options>;
BLOCK vars/<Options>;
PIE vars/<Options>;
```

Statement specific Options for GCHART:

Options	HBAR	VBAR	BLOCK	PIE
STANDARD OPTIONS:				
AXIS =	X	X	X	X
FREQ =	X	X	X	X
DISCRETE	X	X	X	X
LEVELS =	X	X	X	X
MIDPOINTS =	X	X	X	X
TYPE =	X	X	X	X
SUMVAR =	X	X	X	X
SEPARATE INTO GROUPS:				
GROUP =	X	X	X	
SUBGROUP =	X	X	X	
REQUEST STATISTICAL ANALYSIS				
FREQ	X	X		
CFREQ	X	X		
CPERCENT	X	X		
SUM	X	X		
MEAN	X	X		
NOSTATS	X			
CONTROL CHART APPEARANCE				
AUTOREF	X	X		
CLIPREF	X	X		
REF =	X	X		
SPACE =	X	X		
NOHEADER			X	X

DESCRIPTIONS OF OPTION STATEMENTS:

AXIS = 2 4 6 or AXIS = 1 TO 10 BY 2
MIDPOINTS = 2 4 6 or
MIDPOINTS = 1 TO 10 BY 2
DISCRETE used for discrete numeric variable.
TYPE = FREQ, CFREQ, CPERCENT/CPCT MEAN, PERCENT/PCT or SUM.
SUMVAR: summarize the numeric variable.
GROUP: produce side-by-side chart.
Subgroup: divides each bar into sections
AUTOREF: Produce reference line
CLIPREF: Reference line behind the bars
REF: Draw single reference line
SPACE: Amount of space between bars.
NOHEADER: Suppresses default header.
NOSTATS: Suppresses statistics in chart.
LEVEL = 4: Four bars; use with MIDPOINTS.
Use TYPE and SUMVAR together.

Chapter 6: One and Two Population Hypothesis Tests

6.1: The TTEST Procedure:

```
PROC TTEST DATA = dataset <options>;
```

```
CLASS variable;
PAIRED var1*var2;
VAR variables;
RUN; QUIT;
```

PROC TTEST Options:

```
ALPHA = p where 0 < p < 1 Type I error.
CI = EQUAL or NONE
H0 = 0 (Default) or specified value.
PLOTS (ONLY) = NONE, ALL, HISTOGRAM, BOXPLOT, INTERVAL,
QQ, PROFILES
SIDES OR SIDED OR SIDE = 2 | L | U
```

6.2: One Population Test and CI for Mean:

CLASS Statement: Identify grouping variable.
VAR Statement: Variable to test.
PAIRED Statement:

PAIRED STATEMENT	WHAT STATEMENT DOES?
PAIRED A*B;	A-B
PAIRED A*BC*D;	A-B and C-D
PAIRED (A B) *(C D)	A-C, A-D, B-C, B-D
PAIRED (A B) *(B C)	A-B, A-C, B-C

Summary of Hypothesis Testing and Confidence Limits for a Single Population Mean

Hypotheses	Test Statistic	Reject H_0 if	
H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$ (μ_0 is specified)	$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$	$\lvert t \rvert \geq t_{\alpha/2,df}$	(1 – α)100% Confidence Interval for μ $\bar{y} \pm t_{\alpha/2,df} \frac{s}{\sqrt{n}}$
H_0: $\mu = \mu_0$ H_1: $\mu > \mu_0$	$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$	$t \geq t_{\alpha,df}$	(1 – α)100% Confidence Lower Bound for μ $\bar{y} - t_{\alpha,df} \frac{s}{\sqrt{n}}$
H_0: $\mu = \mu_0$ H_1: $\mu < \mu_0$	$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$	$t \leq -t_{\alpha,df}$	(1 – α)100% Confidence Upper Bound for μ $\bar{y} + t_{\alpha,df} \frac{s}{\sqrt{n}}$

where $t_{p,df}$ is the critical t-value that determines a right tail area of p and df = n – 1 in this application.

6.3: Overview: T-tests by TTEST, UNIVARIATE, & MEANS Procedure:

The TTEST Procedure:

```
PROC TTEST DATA = dataset;
VAR var1; RUN; QUIT;
```

The UNIVARIATE Procedure:

```
PROC UNIVARIATE DATA = dataset ALPHA = 0.05 CIBASIC MU0
= 75;
VAR var1; RUN; QUIT;
```

The MEANS Procedure:

```
PROC MEANS DATA = dataset MEAN STD STDERR N ALPHA = 0.05
CLM T PRT;
VAR var1; RUN; QUIT;
```

6.4: Two Population Tests and Confidence Intervals for the Difference between Means:

Summary of Hypothesis Testing and Confidence Intervals for Two Normal Population Means

Hypotheses	Test Statistic	Reject H_0 if	(1–α)100% CI
Dependent Samples H_0: $\mu_1 - \mu_2 = 0$ H_1: $\mu_1 - \mu_2 \neq 0$	$t = \frac{\bar{d}}{s_d/\sqrt{n}}$	$\lvert t \rvert \geq t_{\alpha/2,df}$	$\bar{d} \pm t_{\alpha/2,df}\sqrt{s_d^2/n}$ where df = n – 1
Independent Samples H_0: $\mu_1 - \mu_2 = 0$ H_1: $\mu_1 - \mu_2 \neq 0$ if $\sigma_1^2 = \sigma_2^2$	$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	$\lvert t \rvert \geq t_{\alpha/2,df}$	$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2,df}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$
where $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$ is the pooled variance estimate and df = $n_1 + n_2 - 2$			
H_0: $\mu_1 - \mu_2 = 0$ H_1: $\mu_1 - \mu_2 \neq 0$ if $\sigma_1^2 \neq \sigma_2^2$	$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$	$\lvert t \rvert \geq t_{\alpha/2,df^*}$	$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2,df^*}\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$
			where $df^* = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left[\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1-1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2-1)}\right]}$

Summary of Hypothesis Testing for Two Normal Population Variances

Hypotheses	Test Statistic	Reject H_0 if
H_0: $\sigma_1^2 = \sigma_2^2$ H_1: $\sigma_1^2 \neq \sigma_2^2$	$F = \frac{s_1^2}{s_2^2}$	$F \geq F_{\alpha/2,n_1-1,n_2-1}$ α/2 is the right-side area under the F-distribution, n_1 – 1 is the numerator df, and n_2 – 1 is the denominator df.

Chapter 7: One-way ANOVA METHODS, Non-parametric Methods and Ranking

7.1: ANOVA: Hypothesis testing:

Summary of Hypothesis Testing in ANOVA

Hypothesis	Test Statistic	Reject H_0 if
H_0: $\mu_1 = \mu_2 = \ldots = \mu_t$ H_1: at least one μ, is different	$F_{t-1,N-t} = \frac{\frac{1}{t-1}\sum_{i=1}^{t} n_i(\bar{y}_i - \bar{y})^2}{\frac{1}{N-t}\sum_{i=1}^{t}(n_i - 1)s_i^2} = \frac{MSTrt}{MSE}$	$F \geq F_{\alpha,ndf,ddf}$

where N is the grand sample size; $N = \sum_{i=1}^{t} n_i$ and $F_{\alpha,ndf,ddf}$ is the critical F-value determining a right-side probability of α, ndf is the numerator df, and ddf is the denominator df for the application.

7.2: The GML Procedure:

```
PROC GLM DATA = dataset;
CLASS group;
MODEL response1 response2 = group var1 var2;
MEANS group / <options>;
LSMEANS group/ STDERR;
BY and WHERE Statements as necessary.
```

PROC GLM Statement Options:

```
PLOTS = NONE | DIAGNOSTICS RESIDUALS BOXPLOT
PLOTS(UNPACK) = (DIAGNOSTICS RESIDUALS)
```

MEANS Options: CLM, ALPHA = 0.05, LSD, TUKEY, BON, or SCHEFFE.

- CLDIFF or LINES for illustrating the pairwise comparison conclusion.

7.3: Non-Parametric Tests:

Rank Sum Test Methods for Comparing Two or More Groups

Hypotheses	Test Statistic	Reject H_0 if	Comment
H_0: $\mu_1 = \mu_2 = \ldots = \mu_t$ H_1: at least one μ_i is different	$\chi^2 = \frac{12}{N(N+1)}\sum_{i=1}^{t}\frac{R_i^2}{n_i} - 3(N+1)$ where $N = \sum_{i=1}^{t} n_i$	$\chi^2 \geq \chi^2_{\alpha,t-1}$	This is referred to as the **Kruskal-Wallis** test for $t \geq 2$ groups.
H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	$\mu = \frac{n_1(n_1+n_2+1)}{2}$ and $\sigma = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}$ $z = \frac{R_1 - \mu}{\sigma}$	$\lvert z \rvert \geq z_{\alpha/2}$	This is referred to as the **Wilcoxon Rank Sum Test** or the Mann-Whitney test for $t = 2$. This is algebraically equivalent to the Kruskal-Wallis Test when $t = 2$.

where $\chi^2_{\alpha,df}$ is the critical χ^2 determined by right side probability α, and $df = t - 1$ in this application. Also, z_p is the critical standard normal value determined by right side probability p.

7.4: The NPAR1WAY Procedure:

```
PROC NPAR1WAY DATA = dataset <options>;
CLASS variable; RUN; QUIT;
BY and WHERE Statements are optional.
```

PROC NPAR1WAY Options:

```
ANOVA PLOTS = ANOVABOXPLOT
WILCOXON PLOTS = WILCOXONBOXPLOT
PLOTS = ALL, ANOVA, WILCOXON

Also available but not included here:
EDF(EDFPLOT): Empirical distribution plot
MEDIAN (MEDIANPLOT)
SAVAGE(SAVAGEPLOT)
VW(VWBOXPLOT): Van der Waerden Analysis
```

7.4: The RANK Procedure:

```
PROC RANK DATA = dataset <options>;
VAR var1 var2 var3;
RANKS newvar1 newvar2 newvar3;
RUN; QUIT;
```

PROC RANK Options:

```
OUT = outsetname: New Dataset.
DESCENDING (Ordering of data)
TIES = MEAN | HIGH | LOW
```

Chapter 8: Data Step: Reading Data Files and Labeling Variables:

8.1: The INFILE Procedure:

```
FILENAME name "pathway\to\file\datafile";
DATA dataset;
INFILE name <Options>;
INPUT var1 var2 var3;
Data transformations as per necessary
```

INFILE Options:

```
DELIMITER or DLM = " " or "," or '09`X or '8888' or '##
or "%20";
DSD: Two Adjacent DLM indicates missing.
FIRSTOBS = 10; First observation in 10th row.
LRECL = 256; 256 bytes record length.
MISSOVER: Missing until the end of row.
OBS = 30; 30th observation is last.
STOPOVER: Stop data processing at end of current record.
```

8.2: The LABEL Statement:

```
DATA Dataset;
INPUT var1 $ var2 var3;
LABEL var1 = "Name of variable 1"
      var2 = "Name of variable 2"
      var3 = "Name of variable 3"
DATALINES;
```

8.3: View Table and Table Editor

8.4: DROP, KEEP AND RENAME Statements:

```
DATA Dataset;
<List of DATA Step commands>
DROP var1;
KEEP var2;
RENAME oldnamevar3 = newnamevar3
       oldnamevar4 = newnamevar4;
```

Chapter 9: Frequency Analysis:

9.1.1 & 9.1.2: One-Way Frequency Analysis: Goodness of Fit Test for Equal Proportion and for a Nominal Distribution:

Ho: All class levels have same proportion, $P_1 = \ldots = P_K$
Ha: At least one is different.
Test Statistics: Chi=squared Test.

9.1.3: Two-Way Frequency Analysis: X^2- Contingency Analysis:

Ho: Two variables are independent.
Ha: Two variables are related or dependent.
Test Statistics: Chi-squared test.

9.1: The FREQ Procedure:

```
PROC FREQ DATA = dataset <Options>;
TABLES requests / <Options>;
WEIGHT variable;
<Optional Conditional WHERE IF THEN Statements>
```

PROC FREQ Statements Options:

```
COMPRESS: Multiple One way table in single new page.
PAGE: Print one table per page. (Don't use: compress)
NLEVELS: # of levels of variables in table statement.
```

ORDER = DATA | FREQ (Mutually Exclusive, Order table as in data | descending order of frequency | Alphabetical)

TABLES Statements Options:

TASK	Options	One-Way Table	Two Way Table
Control Printed Results	CROSSLIST		X
	LIST		X
	NOCOL		X
	NOCUM	X	X*
	NOFREQ		X
	NOPERCENT	X	X
	NOROW		X
	PLOTS = Requests	X	X
Further information	CELLCHI2		X
	EXPECTED		X
Statistical Analysis	CHISQ	X	X
	TESTF = (#)	X	
	TESTP = (#)	x	
Works with the LIST Options			

CROSSLIST: Prints two to n-way table in list format.
LIST: Two to n-way table in a list format (no cross tab)
NOCOL: No column percentage.
NOCUM: No Cumulative frequency and percentage.
NOFREQ: No observed frequency.
NOPERCENT: No percentage and cumulative percentage.
NOROW: No row percentage.
PLOTS (ONLY) = ALL | NONE | FREQPLOT | DEVIATIONPLOT (Pick One)
PLOTS = FREQPLOT (TYPE = BARCHART | DOTPLOT)
CELLCHI2: Cell contribution for Chi-squared Statistics
EXPECTED: Cell Expected frequencies.
CHISQ: Chi-squared statistics.
TESTF: GOF test for specified frequency of each class.
TESTP: GOF test for specified % of each class levels.
WEIGHT: Frequency weight. Increase N by the weight.

Chapter 10: Summarizing a Data Table in A Formal Report:

10.1: Revisiting the PRINT, FREQ and MEANS Procedure:

(See previous chapters)

10.2: The REPORT Procedure:

```
PROC REPORT DATA = dataset <Options>;
COLUMN var1 var2 var3; * Order is important.
DEFINE report-intem / <Options>;
RBREAK BEFORE | AFTER / <OPTIONS>;
<BY and WHERE Statements as per necessary>
```

PROC REPORT Statements Options:

WINDOWS | NOWINDOWS or (WD | NOWD)
COLUMN: Selects variables in table.
DEFINE: Modify name of variables in table.

DEFINE Statements Options:

ORDER: Ascending order of rows.
DESCENDING ORDER: Descending order of rows.
WIDTH = # Specify width of column.
GROUP var1 var2: Grouping variables.
SUM, MEAN, N MAX, MIN (Choose one Statistics).
RBREAK: Includes grand total in report.

RBREAK Statements Options:

BEFORE | AFTER: At the top | bottom.
SUMMARIZE: Print total.
OL: Print single line over total.
DOL: Print double line over total.
UL: Print single line under total.
DUL: Print double line under total.
OL, DOL, UL, DUL are effective in LISTING Options.

Chapter 11: Regression and Correlation Analysis:

11.1: Simple Linear Regression:

TABLE 11.1

Testing Regression Model Parameters

Hypotheses	Test Statistic	Reject H_0 if	$(1-\alpha)100\%$ CI for β_i
H_0: $\beta_i = 0$ H_1: $\beta_i \neq 0$	$t = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}$	$\lvert t\rvert \geq t_{\alpha/2,df}$ or $t^2 > F_{\alpha,1,df}$	$\hat{\beta}_i \pm t_{\alpha/2,df}\, se(\hat{\beta}_i)$
H_0: $\beta_i = b_{i0}$ H_1: $\beta_i \neq b_{i0}$ where $b_{i0} \neq 0$	$t = \frac{\hat{\beta}_i - b_{i0}}{se(\hat{\beta}_i)}$	$\lvert t\rvert \geq t_{\alpha/2,df}$ or $t^2 > F_{\alpha,1,df}$	

where i = 0 or 1, $se(\hat{\beta}_i)$ is the standard error of the $\hat{\beta}_i$ parameter estimate, and df is the error degrees of freedom. In simple linear regression df = n − 2.

$$(1-\alpha)100\%\ \text{CI for } \mu_{Y|x} \qquad (\hat{\beta}_0+\hat{\beta}_1 x_0) \pm t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{\sum(x_i-\bar{x})^2}\right)}$$

and

$$(1-\alpha)100\%\ \text{PI for } Y|X \qquad (\hat{\beta}_0+\hat{\beta}_1 x_0) \pm t_{\alpha/2,n-2}\sqrt{MSE\left(1+\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{\sum(x_i-\bar{x})^2}\right)}$$

11.2: The PROC REG Procedure:

```
PROC REG DATA = dataset < Options>;
Label: MODEL dependents = regressors </Options>;
ID var(s);
Label: TEST equation;
RUN; QUIT;
```

PROC REG Statements Options:

PLOTS = NONE| DIAGNOSTICS | FIT (Pick one)
PLOTS(ONLY) = FIT | DIAGNOSTICS
PLOTS(UNPACK) = DIAGNOSTICS /PLOTS = DIAGNOSTICS(UNPACK)
SIMPLE: Mean, Summation, Uncorrected SS, etc.

MODEL Statements Options:

ALPHA = p;
CLI: Upper & Lower CI for **individual predicted** value.
CLM: Upper & Lower CI for **Mean** response /Expected value.
CLB: Upper & Lower CI for Regression Parameters (β_s).
P: Predicted value for input data and estimated model.
COLLIN: Value inflation factor (VIF).
INFLUENCE: Influence Diagnostics.
NOINT: No intercept of regression model.
R: Listing of Residuals.
SELECTION = FORWARD (or F)| BACKWARD (or B)| STEPWISE | MAXR | MINR | RSQUARE | ADJRSQ | CP (Choose one method)
SS1: Type I error
SS2: Type II error. Default of PROC REG: Type III.

11.3: Correlation Coefficient:

TABLE 11.3

Test for the Pearson or Spearman Correlation Coefficient

Hypotheses	Test Statistic	Reject H_0 if
H_0: $\rho = 0$ (No linear association.) H_1: $\rho \neq 0$ (Linear association)	$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	$\lvert t\rvert \geq t_{\alpha/2,n-2}$

11.4: The PROC CORR Procedure:

```
PROC CORR DATA = dataset <Options>;
VAR var1 var2 var3;
WITH var4 var5 var6;
RUN; QUIT;
```

Note: "VAR var1 var2 var3 var4 var5 var6;" can be used without "WITH" statement.

PROC CORR Statements Options:
PEARSON: Pearson Correlation Coefficient (Default).
SPEARMAN: Spearman Correlation Coefficient.
PLOTS = NONE | MATRIX <Options> | SCATTER <Options>;
MATRIX Options: HISTOGRAM (or HIST) NVAR = ALL (or n)
SCATTER Options: ELLIPSE = NONE

Chapter 12: SAS Library and Permanent Data Sets:

11.4: The LIBNAME Statement:

```
LIBNAME libref "drive:\folder";
DATA LIBRARY.DATA; ➔ Data.sas7bdat dataset in library.
DATA LIBRARY.DATA2; ➔ Data2 in library
SET LIBRARY.DATA; ➔ Data from library
```

Note: Permanent SAS dataset file extension: ".sas7bdat".

```
LIBNAME libref "drive:\folder"; ➔ V9 Engine.
LIBNAME libref V6 "drive:\folder"; ➔ V6 Engine.
```

Chapter 13: SAS Probability Functions:

13.1: Discrete Probability Distribution:

13.1.1: Binomial Distribution:

$$P(Y = m|n) = \binom{n}{m} P^m (1-P)^{n-m} = \frac{n!}{m!(N-m)!} P^m (1-P)^{n-m}$$

Where, 0 ≤ Probability (P) ≤ 1, n = # of trial, m = # of success events, n > m.

PROBBNML (p, n, m) Computes binomial distribution.

13.1.2: Poisson Distribution:

$$P(Y = n) = \frac{e^{-\lambda}\lambda^n}{n!}$$

Where, 0 ≤ Probability (P) ≤ 1, $\lambda = Mean; \lambda > 0$. n = value of random variable; n = 0, 1, 2, … , ∞.

POISSON(λ, n) Computes Poisson distribution.

13.2: Continuous Probability Distribution:

13.2.1: The Normal Distribution:

PROBNORM(x) Computes Normal distribution probabilities.
Probit function is the inverse of normal distribution
PROBIT(P) Computes quantiles. Where, 0 < P < 1.

13.2.2: The t-Distribution:

PROBT(x, df, <nc>): t-distribution probabilities.
x is numeric random variable, df = degree of freedom, nc is non-centrality parameter (currently, nc = 0).
Inverse of t-distribution:
INVT(p, df, <nc>

13.2.3: The Chi-Squared Distribution:

PROBCHI(x, df, <nc>): X^2 distribution probabilities.
x is numeric random variable, df = degree of freedom, nc is non-centrality parameter (currently, nc = 0).
Inverse of X^2 distribution: CINV(p, df, <nc>

13.2.4: The F-Distribution:

PROBF(x, ndf, ddf, <nc>): X^2 distribution probabilities.
x is numeric random variable, ndf = numerator degree of freedom, ddf = denominator degree of freedom, nc is non-centrality parameter (currently, nc = 0).
Inverse of F- distribution: FINV(p, ndf, ddf, <nc>

Chapter 14: Reading and Writing Data Files:

14.1: The Import Wizard:

14.2: The IMPORT Procedure:

```
FILENAME fileref "path\to\file\file.xlsx";
PROC IMPORT OUT = WORK.DEMO
DATAFILE = "path\tofile\file.xlsx" (or fileref)
DBMS = EXCEL REPLACE ;
SHEET = "Sheet1$";
```

```
GETNAMES = YES;
MIXED = NO;
SCANTEXT = YES;
USEDATA = YES;
SCANTIME = YES;
STARTROW = 10; /* In older version of excel */
LABEL X = " Label for variable x"
      y = " Label for variable y";
RUN;
```

14.3: The Export Wizard:

14.4: The EXPORT Procedure:

```
FILENAME fileref "path\to\file\file.xlsx";
PROC EXPORT OUT = WORK.DEMO
DATAFILE = "path\tofile\file.xlsx" (or fileref)
DBMS = EXCEL REPLACE;
SHEET = "Sheet Name";
RUN;
```

Chapter 15: DO Loops, ARRAY Random Number Generator:

15.1: DO Loops: DO and END Statements:

```
DO <Options>;
      SAS Statements
END;
```

Each DO statement must have an END statement to close loop. Loop can be nested on each other.

DO Examples:

```
DO I = 8;
DO day = "Mon", "Wed", "Fri";
DO Month = "3", "6", "9", "12";
DO I = 1 to 100 BY 5;
DO day = 75 TO 50 BY -1;
DO WHILE (N GE 5);
DO UNTIL (N LE 5);
END; END; END; END; END; END; END;
```

15.2: The OUTPUT Statement:

15.3: The ARRAY Statement:

```
ARRAY arrayname {subscript} <$> <array elements or
variables>;
```

15.4: Random Number Generator:

```
RANNOR (seed);
CALL RANNOR (seed, x);
DATA example;
SEED1 = 2021;
SEED2 = 2021;
DO I = 1 TO 10 BY 1;
CALL RANNOR (SEED1, X1);
CALL RANNOR (SEED2, X2);
CALL RANNOR (S1250, X3);
IF I = 6 THEN DO;
SEED2 = 17;
SEED3 = 17
END; OUTPUT; END;
```

Chapter 16: Statistical Graphic Procedures:

16.1: The SGPLOT Procedure:

```
PROC SGPLOT DATA = sasdataset;
<Chart statements>; RUN; QUIT;
```

Chart Statement for SGPLOT Procedure

Chart/Plot Type	SGPLOT Statement
Bar Chart	HBAR or VBAR
Histogram	HISTOGRAM
Density	DENSITY
Box Plot	HBOX or VBOX
(X, Y) Plots	SCATTER or SERIES

Options for HBAR and VBAR

Options	Explanation
BARWIDTH = B	0<b<1. 1 is default.
FILL | NOFILL	Fill color inside bar.
GROUP = categorical variable	Grouping variable.
LIMIT = BOTH | UPPER | LOWER ALPHA = p	Specify interval. Pick one; 0<p<1
RESPONSE = summary variable STAT = FREQ | MEAN | SUM | MEDIAN	Both statements needed. Pick one STAT
TRANSPARENCY = t FILLATTRS (COLORS = color)	0<t<1; Default = 1; BLACK, PURPLE, BLUE, RED, GREEN, ORGANGE, GRAY

Options for HISTOGRAM and DENSITY

Options	Explanation
HISTOGRAM var </option>	
BARWIDTH = b	0<b<1. 1 is default.
NBINS = #	# of Bins
SCALE = COUNT|PERCENT|PROPORTION	Pick one. Specify bar heights
DENSITY var </option>	
TYPE = NORMAL	Normal density curve

TYPE = NORMAL(MU = # SIGMA = #)	Normal density with specified parameters.

Options for HBOX and VBOX

Options	Explanation
HBOX|VBOX var </option>	
BOXWIDTH = n	0<n<1. 0.4 is default.
CATEGORY = cat. Var.	Box plot by categories.
CONNECT = MEAN|MEDIAN|Q1|Q3|MIN|MAX	Pick one, draw connecting lines.
FILL|NOFILL	Pick One; Fill fillattrs
FILLATTERS = (COLOR = color)	BLACK, BLUE, RED, GREEN, ORANGE, PURPLE, GRAY
MEANATTRS = style element	Shape outlines: CIRCLE, DIAMOND, HOMEDOWN, SQUARE, STAR & TRIANGLE.
NOMEAN NOMEDIAN NOOUTLIERS	Suppress stats.

Options for SCATTER and SERIES

Options	Explanation

TABLE 16.5

Introductory Options for the SCATTER and SERIES Statements

Options	Explanation
DATALABEL= *<variable>*	The value of the specified variable is to be used as a label in the SCATTER. If no variable is specified, the value of Y is the label.
GROUP = *category-variable*	For the SCATTER statement – this produces different color plotting symbols for each level of the category variable. For the SERIES statement – the plotted points are connected by a line for each level of the category variable.
LINEATTRS=(COLOR=*color* PATTERN=*pattern*)	For the SERIES statement – this produces different color lines using different line patterns. For the colors, keep it simple when starting out: BLACK, BLUE, RED, GREEN, ORANGE, PURPLE, GRAY, and so on. More information on color selection is given in Section 17.1. The choices for *pattern* are: 1, 2, . . . , 42. 1 is a solid line, and the remaining are patterns of dots and dashes.
MARKERATTRS=(COLOR=*color* SYMBOL=*symbol*)	For the SCATTER statement – plotting symbols are called markers. The marker color is controlled by *color*, and color choices are as in LINEATTRS. Symbols are outlines or filled shapes. Shape outlines are: CIRCLE, DIAMOND, HOMEDOWN, SQUARE, STAR, and TRIANGLE. Filled shapes are the previous shapes followed by "FILLED", such as CIRCLEFILLED (no spaces).
MARKERCHAR= *variable*	The value of the specified variable is the plotting symbol.

16.2: The SGSCATTERPLOT Procedure:

```
PROC SGPLOT DATA = sasdataset<Options>;
COMPARE X = varlist Y = varlist </Options>;
MATRIX varlist </Options>; RUN; QUIT;
```

16.2: The SGPANEL Procedure:

```
PROC SGPANEL DATA = sasdataset;
PANALBY var </options>;
<Chart Statements>
```

TABLE 16.6

Options for the PANELBY Statement in the SGPANEL Procedure

Options	Explanation
BORDER | NOBORDER	Pick one. Adds or suppresses cell borders within the panel. BORDER is the default.
COLHEADERPOS = TOP | BOTTOM | BOTH	Pick one. Columns in the panel are captioned with the variable name or label. Specify the location. TOP is the default. If PANEL=LAYOUT, then this option is not effective.
COLUMNS = c	Chooses the number of columns in the panel.
ROWHEADERPOS = LEFT | RIGHT | BOTH	Pick one. Rows in the panel are captioned with the variable name or label. Specify the location. RIGHT is the default. If PANEL=LAYOUT, then this option is not effective.
ROWS = r	Chooses the number of rows in the panel.
LAYOUT = LATTICE | PANEL | COLUMNLATTICE | ROWLATTICE	Pick one. LATTICE creates an r × c panel when two classification variables are specified in this statement. PANEL forms an r × c panel with row and column headings available. COLUMNLATTICE arranges the graph cells into a single row, and ROWLATTICE arranges the graph cells into a single column. The default setting is PANEL.

Chapter 17: SAS/GRAPH Procedures

Chapter 18: SAS/GRAPH Procedures

Chapter 19: SAS/GRAPH Procedures

Chapter 20: SAS/GRAPH Procedures