30. *Hint*: Be careful how the totals are calculated on records with missing data.

*Answer:*

DATA hotel;

\*\* Part a);

INFILE 'c:\MyRawData\Hotel.dat';

INPUT Room Guests InM InD InY OutM OutD OutY

Internet $ InternetDays 49-51 RoomType $ 53-67

RoomRate;

\*\* Part b);

CheckIn = MDY(InM,InD,InY);

CheckOut = MDY(OutM,OutD,OutY);

FORMAT CheckIn CheckOut MMDDYY8.;

\*\* Part c);

\*\* Using the SUM function prevents the rooms with

missing values for Internet service from counting

as a missing value in the total;

SubTotal = SUM( ((CheckOut - CheckIn) \* RoomRate),

(10 \* (Guests - 1) \* (CheckOut - CheckIn)),

(9.95 + (4.95 \* (InternetDays))) );

\*\* Part d);

GrandTotal = ROUND(SubTotal \* 1.0775,.01);

RUN;

\*\* Part e);

\*\* The grand total for room 211 is 1357.65;

(sections 3.1, 3.2, 3.4, 3.8, 3.9)

31. *Hint:* Consider using functions to facilitate the creation of new variables. In addition, DO statements will help with grouping observations with similar conditions.

*Answer:*

DATA emp;

\*\* Part a);

INFILE 'c:\MyRawData\Employees.dat' TRUNCOVER;

INPUT @1 SSN $11. @16 Name $30. @47 DOB DATE9.

@57 Grade $ @62 Salary DOLLAR10.2

@73 Title $27.;

\*\* Part b);

FORMAT DOB MMDDYY8.;

\*\* Part c);

\*\* Using a FORMAT or the ROUND function for the age

will not report it correctly as an integer;

Age = INT(YRDIF(DOB,MDY(12,31,YEAR(TODAY())),'AGE'));

\*\* Part d);

IF Grade = 'GR20' THEN DO;

MinSalary = 50000;

MaxSalary = 70000;

END;

ELSE IF Grade = 'GR21' THEN DO;

MinSalary = 55000;

MaxSalary = 75000;

END;

ELSE IF Grade = 'GR22' THEN DO;

MinSalary = 60000;

MaxSalary = 85000;

END;

ELSE IF Grade = 'GR23' THEN DO;

MinSalary = 70000;

MaxSalary = 100000;

END;

ELSE IF Grade = 'GR24' THEN DO;

MinSalary = 80000;

MaxSalary = 120000;

END;

ELSE IF Grade = 'GR25' THEN DO;

MinSalary = 100000;

MaxSalary = 150000;

END;

ELSE IF Grade = 'GR26' THEN DO;

MinSalary = 120000;

MaxSalary = 200000;

END;

\*\* Part e);

AnnSal = ROUND(MIN(

(Salary \* 12) \* 1.025,MaxSalary),.01);

\*\* Part f);

IF INDEX(Title,'Lead') >= 1 OR

INDEX(Title,'Manager') >= 1 OR

INDEX(Title,'Director') >= 1 THEN Bonus = 1000;

ELSE bonus = 0;

RUN;

\*\* Part g);

\*\* The expected annual salary of William Stone is

88648.68 with a bonus of 1000;

\*\* The expected annual salary of Mark Harrison is

85000.00 with a bonus of 0;

(sections 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.8, 3.9)

32. *Hint:* IF-THEN/ELSE statements will result in less code for the phone number listing.

*Answer:*

DATA confinfo;

\*\* Part a);

INFILE 'c:\MyRawData\Conference.dat' MISSOVER

TRUNCOVER;

INPUT FirstName :$20. LastName :$20. AttendeeID

@47 BusinessPhone $13. @61 HomePhone $13.

@75 MobilePhone $13. ContactBusiness $

ContactHome $ ContactMobile $ Rate WedMixer $

ThursLunch $ Volunteer $ Restrictions $ 117-140;

\*\* Part b);

IF Rate IN (350,200,450) THEN Group = 'Regular';

ELSE IF Rate IN (295,150,395) THEN Group = 'Early';

ELSE IF Rate = 550 THEN Group = 'On-Site';

\*\* Part c);

IF LENGTH(BusinessPhone) = 13 THEN

AreaCode = SUBSTR(BusinessPhone,2,3);

ELSE IF LENGTH(MobilePhone) = 13 THEN

AreaCode = SUBSTR(MobilePhone,2,3);

ELSE IF LENGTH(HomePhone) = 13 THEN

AreaCode = SUBSTR(HomePhone,2,3);

\*\* Part d);

IF INDEX(PROPCASE(Restrictions),'Vegetarian') > 0

THEN VegMeal = 1;

ELSE IF INDEX(PROPCASE(Restrictions),'Vegan') > 0

THEN VegMeal = 1;

ELSE VegMeal = 0;

RUN;

\*\* Part e);

\*\* Tina Gonzales is registered On-Site, has an area code

of 650, and has a special meal flag = 1;

\*\* Patrick Anderson is registered Early, has an area

code of 408, and has a special meal flag = 0;

(sections 3.1, 3.2, 3.3, 3.5, 3.6)

33. *Hint:* You may need to create a new variable to assist in tracking when the winning team name changes as you step through the data set.

*Answer:*

DATA rosebowl;

\*\* Part a);

INFILE 'c:\MyRawData\RoseBowl.dat';

INPUT GameDate MMDDYY10. WinTeam $ 12-37 WinPts

LoseTeam $ 41-66 LosePts;

\*\* Part b);

DiffPts = WinPts - LosePts;

\*\* Part c);

FORMAT GameDate WEEKDATE29.;

\*\* Part d);

TotalCount + 1;

\*\* Part e);

\*\* Create a new variable to help track when there is a

new team so the subset counter can be reset to 1;

\*\* There are techniques later in chapter 6 that can

deal with this type of scenario more easily;

RETAIN WinTemp;

\*\* Condition when there is a new occurrence of a team;

IF WinTeam ~= WinTemp THEN DO;

\*\* Reset subset counter;

WinCount = 1;

\*\* Reset the team tracker variable to the next team

name;

WinTemp = WinTeam;

END;

\*\* Otherwise it is the same team so increase counter;

ELSE WinCount + 1;

RUN;

\*\* Part f);

\*\* The first game won by Southern California was on

Monday, January 1, 1923 with a score difference of 11

points;

(sections 3.1, 3.5, 3.6, 3.8, 3.9, 3.10)

34. *Hint:* Consider the use of arrays and range lists to shorten your program.

*Answer:*

DATA gym;

\*\* Part a);

INFILE 'c:\MyRawData\NewYears.dat' DLM = ','

LRECL = 2500 DSD;

INPUT ID (InDay1 - InDay119) (:TIME8.)

(OutDay1 - OutDay119) (:TIME8.);

\*\* Part b);

ARRAY inD(119) Inday1 - Inday119;

ARRAY outD(119) Outday1 - Outday119;

ARRAY timeD(119) Ttime1 - Ttime119;

\*\* Part c);

Promo = 'Eligible';

\*\* Parts b) and c);

DO i = 1 TO 119;

\*\* Calculating times/day and divide by 60 to get

minutes;

timeD(i) = (outD(i) - inD(i)) / 60;

\*\* Flagging those who fall out of promo;

IF timeD(i) < 30 THEN Promo = 'Out';

END;

\*\* Part d);

AvgTime = MEAN(OF Ttime1 - Ttime119);

RUN;

\*\* Part e);

\*\* If you choose to use PROC PRINT, then the values of

average time may be displayed with fewer decimal

places;

\*\* Member 330 has a value of Eligible for the refund and

and an average time of 124.11680672;

\*\* Member 331 has a value of Out for the refund and

and an average time of 16.355042017;

(sections 3.1, 3.2, 3.4, 3.5, 3.11, 3.12)

35. *Hint:* You will need to tell SAS to preserve information across iterations of the DATA step.

*Answer:*

DATA baj2;

\*\* Part a);

INFILE 'c:\MyRawData\BenAndJerrys.dat' LRECL = 350

MISSOVER DLM = ',' DSD;

INPUT Flavor :$75. Portion Cal CalFat Fat SatFat

TransFat Chol Sod Carb Fib $ Sugar Protein

YearIntro YearRetire Desc :$150. Notes :$200.;

\*\* Part b);

IF YearRetire = . AND

Notes ~= 'Scoop Shop Exclusive';

\*\* Part c);

TB = (Cal / Portion) \* 15;

IF TB >= 0;

\*\* Part d);

TBtotal + TB;

\*\* Part e);

RETAIN HighCal 0;

IF Cal > HighCal THEN HighCal = Cal;

RUN;

\*\* Part f);

\*\* If you choose to use PROC PRINT, then the values of

total calories may be displayed with fewer decimal

places;

\*\* The final value for total calories is 2135.9939433

and highest calories are 340;

(sections 3.1, 3.5, 3.7, 3.10)

36. *Hint:* The raw data file uses a delimiter other than spaces. Think about the order of events when recoding the data. Arrays and DO loops of various sizes can help with the many groupings and calculations.

*Answer:*

DATA wls;

\*\* Part a);

INFILE 'c:\MyRawData\WLSurveys.dat'

dlm = '09'x;

INPUT ID Height Weight1 - Weight5 Q1 - Q30;

\*\* Part b);

ARRAY switch(15) Q2 Q3 Q5 Q8 Q9 Q11 Q14 Q15 Q17 Q20

Q21 Q23 Q26 Q27 Q29;

\*\* Convert questions 2,3,5 to reverse scale;

\*\* Loop runs 15 times for 3\*5 questions that need to

be converted;

DO i = 1 TO 15;

\*\* Use IF-THEN/ELSE to prevent overwriting of

previously reclassified variables;

IF switch(i) = 0 THEN switch(i) = 3;

ELSE IF switch(i) = 1 THEN switch(i) = 2;

ELSE IF switch(i) = 2 THEN switch(i) = 1;

ELSE IF switch(i) = 3 THEN switch(i) = 0;

END;

\*\* Part c);

ARRAY reset(30) Q1 - Q30;

\*\* Reset missing to .;

DO j = 1 TO 30;

IF reset(j) = -99 THEN reset(j) = .;

END;

\*\* Part d);

ARRAY wt(5) Weight1 - Weight5;

ARRAY bodymass(5) BMI1 - BMI5;

DO k = 1 TO 5;

bodymass(k) = (wt(k) / Height \*\* 2) \* 703;

END;

\*\* Part e);

IF BMI5 >= 25;

RUN;

\*\* Part f);

\*\* If you choose to use PROC PRINT, then the values of

final BMI may be displayed with fewer decimal

places;

\*\* The tenth observation has an ID of 13, one missing

value for the question variables, and a final BMI of

25.983241506;

(sections 3.1, 3.5, 3.6, 3.7, 3.11, 3.12)