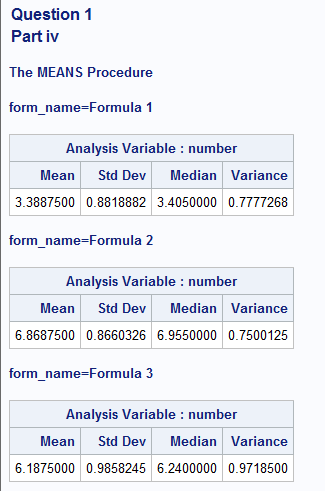
STATISTICS 147 Exam I, Part 1: SAS

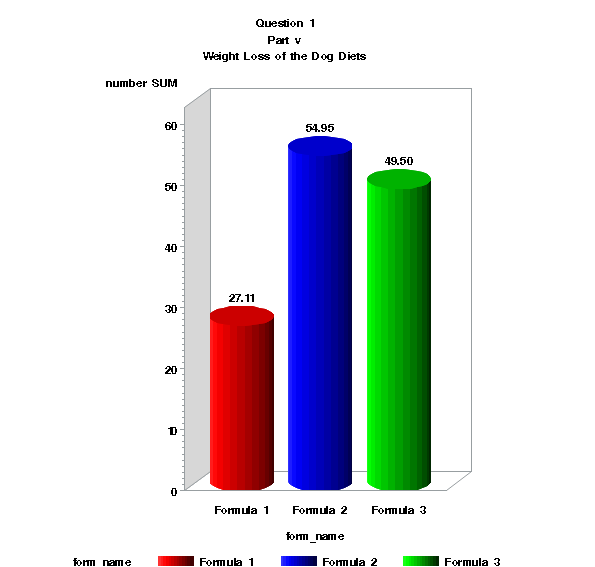
Summer 2020

Wesley Chang

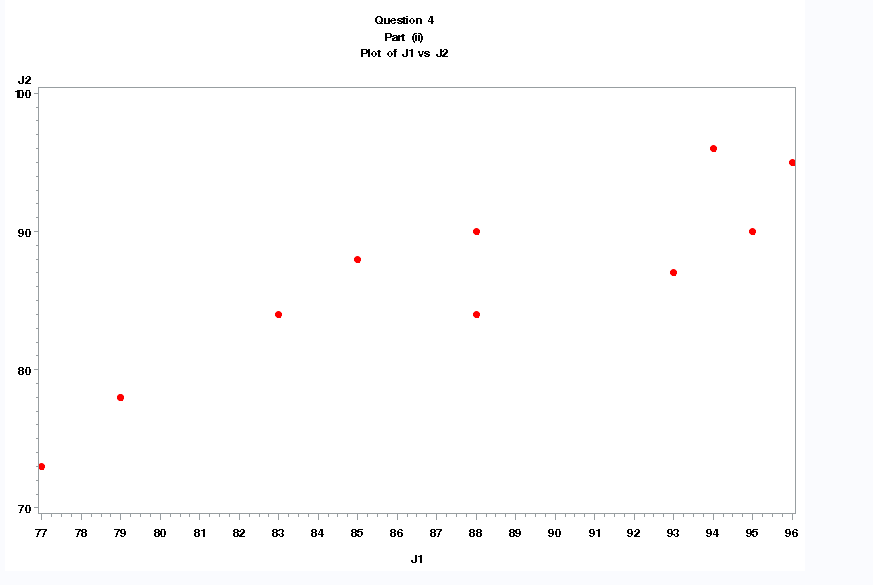
1. Question 1 Part iv



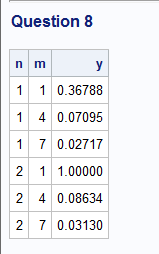
1. Question 1 Part v



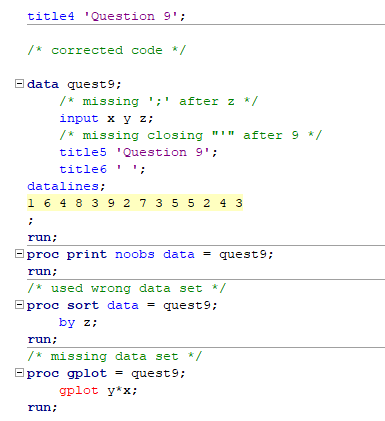
1. Question 4 Part 2:



1. Question 8



1. Question 9



SAS Code:

/\* Statistics 147 Exam I, Part 1: SAS

Summer 2020

Wesley Chang \*/

title1 'Statistics Exam I, Part 1: SAS';

title2 'Summer 2020';

title3 'Wesley Chang';

/\* Set up format of the output \*/

options nocenter ps = **55** nocenter ls = **78** nodate nonumber formdlim='\*';

/\* add line of code to clear all windows except the editor window \*/

DM log "odsresults; clear; out; clear; log; clear;";

ods graphics off;

/\* goptions settings \*/

goptions reset = all colors = (blue,red,green,purple)

ftitle = swissb ftext=swissb htitle=**1.5** htext=**1.0**;

title4 'Question 1';

/\* Carly and Shannon work in R&D are considering 3 recipes for a new diet food, and have decided to test them using severely overweight dogs

24 severely overweight dogs were selected at random and then each randomly assigned to one of the three recipes of dog food

after an appropropriate amount of time, the weight loss, measured in pounds, was recorded \*/

title5 'Part i and Part ii';

**data** diet\_all;

/\* read file dogdiet.dat \*/

infile 'c:\Users\wesle\iCloudDrive\Summer 2020 (UCR)\STAT 147 (Session A)\Midterm\dogdiet.dat' firstobs = **3**;

/\* do loops for 8 rows of data \*/

do rows = **1** to **8**;

/\* do loop for 3 columns of data \*/

do column = **1** to **3**;

if column = **1** then form\_name = 'Formula 1';

else if column = **2** then form\_name = 'Formula 2';

else form\_name = 'Formula 3';

/\* take inputs and output them \*/

input number @@;

output;

end;

end;

**run**;

/\* print output with no observation numbers \*/

**proc** **print** noobs data = diet\_all;

var form\_name number;

**run**;

title5 'Part iii';

/\* sort data by form\_name \*/

**proc** **sort** data = diet\_all;

by form\_name;

**run**;

/\* print sorted data \*/

**proc** **print**;

**run**;

title5 'Part iv';

/\* generate descriptive statistics for each formula, for mean, sd, and median for each formula \*/

**proc** **means** mean stddev median var;

by form\_name;

var number;

**run**;

title5 'Part v';

/\*

generate 3d vertical bar chart (high res) for the total weight loss of each of the dog formulas

a. include following title: Weight Loss of the Dog Diets

Rotate the title by 25%

Use swissb as the font

Use a height of 2

b. make the colors of the bars: red, blue, and green

c. make the shape of the bars: cylinder

d. print the sum of the weights above each bar

\*/

/\* use proc gchart to generate bar chart \*/

**proc** **gchart** data = diet\_all;

/\* add title with 25% rotation, swissb as font, and a height of 2 \*/

title6 'Weight Loss of the Dog Diets'

height = **2** font = swissb rotate=**25**;

/\* vbar3d for vertical chart \*/

vbar3d form\_name /

/\* print sum of weights above each bar \*/

sumvar = number

outside = sum

/\* shape of bars is cylinder \*/

shape = cylinder

subgroup = form\_name;

/\* make colors of the bars red, blue, and green \*/

pattern1 color = red;

pattern2 color = blue;

pattern3 color = green;

**run**;

title4 'Question 2';

title5;

title6;

/\* create a data set with only Formula 1 \*/

**data** onlyFormula1;

set diet\_all;

title5 'Only Formula 1 data';

if form\_name = 'Formula 1';

**proc** **print** noobs data = onlyFormula1;

var number;

**run**;

title4 'Question 3';

title5;

/\* create a data set with only formula 1 and formula 3 \*/

**data** bothF1F3;

set diet\_all;

title5 'Both Formula 1 and Formula 3 data';

if form\_name = 'Formula 1' or form\_name = 'Formula 3';

**proc** **print** noobs data = bothF1F3;

var form\_name number;

**run**;

title4 'Question 4';

title5;

/\* use dogjudge.dat

two judges Khaldoun and Thomas (J1, J2) \*/

title5 'Part (i)';

/\* write the appropriate code to read in and print out data \*/

**data** dogjudge;

/\* read file dogjudge.dat \*/

infile 'c:\Users\wesle\iCloudDrive\Summer 2020 (UCR)\STAT 147 (Session A)\Midterm\dogjudge.dat' firstobs=**3**;

/\* take input from variables dog, J1, J2 and output \*/

input dog $ **1**-**11** J1 J2;

output;

**run**;

/\* print data \*/

**proc** **print** noobs data = dogjudge;

var dog J1 J2;

**run**;

title5 'Part (ii)';

/\* create a high res scatterplot of J1 vs J2, using J1 as the horiz axis, include a descriptive title \*/

/\* add some color and style to the plot \*/

symbol1 color = red

value = dot

height = **1**;

/\* use proc gplot to create scatter plot \*/

**proc** **gplot** data = dogjudge;

title6 'Plot of J1 vs J2';

plot J2\*J1;

**run**;

title4 'Question 5';

title5;

title6;

/\* suppose Abby and MaxS were disqualified. Create a temp SAS data set that reads in and prints out the obs for dogs that have NOT been disqualified

add a descriptive title \*/

/\* create data set q5 that excludes Abby and MaxS \*/

**data** q5;

title5 'Data for Dogs that have not been Disqualified';

set dogjudge;

if dog = 'Cody' or dog = 'Dexter' or dog = 'Dusty' or dog = 'Kali' or dog = 'Korra' or dog = 'Lakota' or dog = 'MaxG' or dog = 'Mercedes' or dog = 'Shadow';

**run**;

/\* print data with no observation numbers \*/

**proc** **print** noobs data = q5;

**run**;

title4 'Question 6';

title5;

/\* specification limits are 40.0 +- 2.50

appropriate level: 37.5 to 42.5 ounces

suppose process is normally distributed with mean of 40 and sd of 1.75

they select a bottle at random

\*/

/\* create data set that contains parts i to iii \*/

**data** q6;

/\* use the normal distribution \*/

/\* cdf('Distribution',X,mean,sd) \*/

/\* Part i \*/

/\* find probability that the bottle is overfilled

P(X>42.5)= 1 - P(X<=42.5) \*/

q6p1 = **1** - cdf('Normal',**42.5**,**40**,**1.75**);

/\* Part ii \*/

/\* find probability X is between 37 and 41

P(37<X<41) = P(X<=41) - P(X<=37) \*/

q6p2 = cdf('Normal',**41**,**40**,**1.75**) - cdf('Normal',**37**,**40**,**1.75**);

/\* Part iii \*/

/\* find the 96th percentile

find x s.t. P(X<=x) = 0.96 \*/

q6p3 = quantile('Normal',**.96**,**40**,**1.75**);

**run**;

/\* print the results for Parts i to iii \*/

title5 'Part (i)';

title6 'P(X>42.5)';

**proc** **print** noobs data = q6;

var q6p1;

**run**;

title5 'Part (ii)';

title6 'P(37<X<41)';

**proc** **print** noobs data = q6;

var q6p2;

**run**;

title5 'Part (iii)';

title6 '96th Percentile';

**proc** **print** noobs data = q6;

var q6p3;

**run**;

title4 'Question 7';

title5;

title6;

/\* p = 35%, random sample of 24 students \*/

/\* Use Binomial distribution \*/

/\*

P(X=x) = pdf('Binom',x,p,n)

P(X <= x) = cdf('Binom',x,p,n)

P(X>x) = sdf('Binom',x,p,n)

\*/

/\* create data set that contains part i and part ii \*/

**data** q7;

/\* Part i \*/

/\* find prob that exactly 10 of the students prefer the change in class times

P(X=10) \*/

q7p1 = pdf('Binom',**10**,**.35**,**24**);

/\* Part ii \*/

/\* find prob that between 7 to 10 students prefer the change in class time

P(7<X<10) = P(X<=10) - P(X>7) \*/

q7p2 = cdf('Binom',**10**,**.35**,**24**) - sdf('Binom',**7**,**.35**,**24**);

**run**;

/\* print the results from part i and part ii \*/

title5 'Part (i)';

title6 'P(X=10)';

**proc** **print** noobs data = q7;

var q7p1;

**run**;

title5 'Part (ii)';

title6 'P(7<X<10)';

**proc** **print** noobs data = q7;

var q7p2;

**run**;

title4 'Question 8';

title5;

title6;

/\* calculate y = e^-sqrt(2m-n) for m = 1,4,7 and n=1,2, using nested DO loops \*/

/\* create data set for loop \*/

**data** q8;

/\* declare n variable with do loop \*/

do n = **1** to **2**;

/\* declare m variable with do loop \*/

do m = **1** to **7** by **3**;

/\* evaluate expression for each n and m \*/

y = exp(-sqrt((**2**\*m)-n));

/\* output \*/

output;

end;

end;

**run**;

/\* print the data set for the loop \*/

**proc** **print** noobs data = q8;

**run**;

title4 'Question 9';

/\* corrected code \*/

**data** quest9;

/\* missing ';' after z \*/

/\* also missing ‘@@’ in input statement \*/

input @@ x y z;

/\* missing closing "'" after 9 \*/

title5 'Question 9';

title6 ' ';

datalines;

1 6 4 8 3 9 2 7 3 5 5 2 4 3

;

**run**;

**proc** **print** noobs data = quest9;

**run**;

/\* used wrong data set \*/

**proc** **sort** data = quest9;

by z;

**run**;

/\* missing data set \*/

**proc** **gplot** = quest9;

gplot y\*x;

**run**;

/\* original data with errors\*/

/\* errors are described above the line in comment \*/

/\*

data quest9;

/\* missing ';' after z \*/

input x y z

/\* missing closing "'" after 9 \*/

title5 'Question 9;

title6 ' ';

datalines;

1 6 4 8 3 9 2 7 3 5 5 2 4 3

;

run;

proc print noobs data = quest9;

run;

proc sort data = diet\_all;

by z;

run;

proc gplot;

gplot y\*x;

run;

\*/