Statistics 147 Assignment #2

Wesley Chang

0996

**R**

1. (Use **R** for this problem!) **(13 pts total)** A production process that creates specialized dog collars operates with 7% nonconforming (defective) output. A sample of 12 collars is selected at random and the number of nonconforming collars counted.

Let Y = number of nonconforming collars. Then *Y* ∼ *Bin*(*n* = 12*,p* = 0*.*07). Use this information to complete the following.

* 1. **(1 pt)** Generate the sequence 0 through 12 and store the sequence in a variable named **y**. (Be sure to print the values of y.)

A picture containing meter

Description automatically generated

* 1. **(1 pt)** Generate the *probability distribution function* of Y and store the values the variable **pdf1**. (Be sure to print the values of pdf1.)

A picture containing drawing

Description automatically generated

* 1. **(1 pt)** Generate the *cumulative distribution function* of Y and store the values in the variable **cdf1**. (Be sure to print the values of cdf1.)

A picture containing drawing

Description automatically generated

* 1. **(1 pt)** Generate the values of *P*(*Y > y*) and store the values in the variable **sdf1**. (Be sure to print the values of sdf1.)

A screenshot of a cell phone

Description automatically generated

* 1. **(1 pt)** Use the **cbind** function to generate a data frame with y1, pdf1, cdf1 and sdf1 as the columns. (Be sure to print the data set.)

A picture containing computer, table

Description automatically generated

* 1. **(1 pt)** Find the probability that *exactly 3* of the 12 collars are nonconforming (defective)

A close up of a logo

Description automatically generated

* 1. **(1 pt)** Find the probability that *at most 3* of the 12 collars are nonconforming (defective).

A close up of a logo

Description automatically generated

* 1. **2 pts)** Find the probability that *between, and including*, 2 and 4 of the 12 collars are nonconforming

(defective).

A picture containing drawing, food

Description automatically generated

* 1. **(2 pts)** Generate a plot of y vs pdf1. Be sure to give your plot an appropriate title.

A close up of text on a black background

Description automatically generated A screenshot of a cell phone

Description automatically generated

* 1. **(2 pts)** Generate a plot of y vs cdf1. Be sure to give your plot an appropriate title.

A screenshot of a social media post

Description automatically generated

A close up of text on a black background

Description automatically generated

1. **(6 pts total)** (Use **R** for this problem!) Suppose the filling process of a particular type of dry dog food bag is normally distributed with a mean of 40 pounds and a standard deviation of 0.50 pounds. The filling process is considered to be functioning at an appropriate level (functioning “in control”) if the amount of fill in the dog food bags is between 38.75 pounds and 41.25 pounds. Ruihan selects a dog food bag at random from the assembly line.
   1. **(1 pt)** Find the probability that the bag is under-filled.

A screenshot of a cell phone

Description automatically generated

* 1. **(1 pt)** Find the probability that the the bag is not under-filled.

A screenshot of a cell phone

Description automatically generated

* 1. **(1 pt)** Find the probability that the bag is over-filled or under-filled. (1 pt) (iv) **(2 pts)** Find the probability that the bag has a fill amount that meets specifications.

A picture containing drawing

Description automatically generated

* 1. **(1 pt)** Find the 97*th* percentile.

A picture containing drawing

Description automatically generated

1. **(3 pts total)** (Use **R** for this problem!) Linda has three dogs (Cody, Dusty and Shadow) that she is training for a national agility championship. Linda records their times to finish the course (in seconds) for 30 runs. Luke does not believe there is a significant difference in mean finishing times between the three dogs. To test this claim, Luke takes three independent random samples of 8 times for each of the three dogs, yielding the following data:

File name: agility.dat

Cody Dusty Shadow

75 69 83

85 79 93

70 77 87

79 51 72

73 53 82

81 69 72

84 59 62

70 64 77

**NOTE:**

♠ The data is located in a datafile named **agility.dat**.

♠ The headings are included in the data file. The actual data begins on line 3.

♠ **Assume** the *Cody* is dog 1, *Dusty* is dog 2, and *Shadow* is dog 3.

♠ *µi* = true average finishing time for dog *i* and *σi*2 = variance of the finishing time for dog *i, i* = 1*,*2*,*3

1. **(2 pts)** Write the R code to read the data into a data frame and then print it out.

A screenshot of a cell phone

Description automatically generated

1. **(1 pt)** Add the appropriate lines of code to make the columns of the data frame accessible individually and obtain the column headers/names.

A screenshot of a cell phone

Description automatically generated

**SAS**

1. **(5 pts total)** (Use **SAS** for this problem!) According to a recent survey, 65% of dog owners prefer a male dog. A random sample of 25 dog owners is selected.

Let X = # of dog owners that prefer a male dog. Use this information to complete the following.

A screenshot of a cell phone

Description automatically generated

* 1. **(1 pt**) Find the probability that exactly 15 of the dog owners prefer a male dog.

**Probability = 0.14085**

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

* 1. **(1 pt**) Find the probability that no more than 15 of the dog owners prefer a male dog.

**Probability = 0.36969**

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

* 1. **(2 pts)** Find the probability that between 16 and 19, inclusively, of the dog owners prefer a male dog.

**Probability = 0.83475**

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

* 1. **(1 pt)** Find the average number of dog owners that prefer a male dog.

**Average = 16.25**

A picture containing bird

Description automatically generated A screenshot of a cell phone

Description automatically generated

1. **(6 pts total)** (Use **SAS** for this problem!) Suppose the filling process of a particular type of canned dog food is normally distributed with a mean of 13.25 ounces and a standard deviation of 0.10 ounces. The filling process is considered to be functioning at an appropriate level (functioning “in control”) if the amount of fill in the dog food cans is between 13.15 ounces and 13.35 ounces. Ruihan selects a can of dog food at random from the assembly line.

A screenshot of a social media post

Description automatically generated

* 1. **(1 pt)** Find the probability that the can is underfilled.

**Probability = 0.15866**

A screenshot of a cell phone

Description automatically generated A picture containing knife

Description automatically generated

* 1. **(1 pt)** Find the probability that the can is *not* underfilled.

**Probability = 0.84134**

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

* 1. **(2 pts)** Find the probability that the can has a fill amount that meets specifications.

**Probability = 0.68269**

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

* 1. **(1 pt)** Find the probability that the can is overfilled or underfilled (does not meet specifications).

**Probability = 0.31731**

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

* 1. **(1 pt)** Find the 97*th* percentile.

**97th Percentile = 13.4381**

A screenshot of a cell phone

Description automatically generated A picture containing knife

Description automatically generated

1. **(5 pts)** (Use **SAS** for this problem!) Let Let for n = 1,3,5 and m = 2,5,8. Use nested

DO loops to calculate the values of y. (Be careful of [P.E.M.D.A.S.](https://en.wikipedia.org/wiki/Order_of_operations#Mnemonics) [/](https://en.wikipedia.org/wiki/Order_of_operations#Mnemonics) [order-of-operations](https://en.wikipedia.org/wiki/Order_of_operations#Mnemonics) when you code!)

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. **(13 pts total)** (Use **SAS** for this problem!) Linda has three dogs (Cody, Dusty and Shadow) that she is training for a national agility championship. Linda records their times to finish the course (in seconds) for 30 runs. Ruihan does not believe there is a significant difference in mean finishing times between the three dogs. To test this claim, Ruihan takes three independent random samples of 8 times for each of the three dogs, yielding the following data:

File name: agility.dat

Cody Dusty Shadow

75 69 83

85 79 93

70 77 87

79 51 72

73 53 82

81 69 72

84 59 62

70 64 77

**NOTE:**

♠ The data is located in a datafile named **agility.dat**.

♠ The headings are included in the data file. The actual data begins on line 3.

♠ **Assume** the *Cody* is dog 1, *Dusty* is dog 2, and *Shadow* is dog 3.

♠ *µi* = true average finishing time for dog *i* and *σi*2 = variance of the finishing time for dog *i, i* = 1*,*2*,*3

1. **(4 pts)** Write the SAS code to read in and print out the data. Use **agility1** as your temporary SAS dataset name. Use nested do loops! **(NOTE: DO NOT COPY AND PASTE THE DATA INTO YOUR SAS PROGRAM. READ THE DATA IN FROM THE EXTERNAL DATA FILE!)** (5 pts)

A screenshot of a cell phone

Description automatically generated A close up of a device

Description automatically generated A screenshot of a social media post

Description automatically generated

1. **(2 pts)** Using if-then-else structures, name the dogs as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **dog** | 1 | 2 | 3 |
| **name** | **Cody** | **Dusty** | **Shadow** |

A picture containing holding

Description automatically generated

1. **(1 pt)**Add the appropriate lines of code to sort the data by the **name** of the dog.

A close up of a telephone

Description automatically generated A close up of a telephone

Description automatically generated \A screenshot of a cell phone

Description automatically generated

1. **(2 pts)** Add the appropriate lines of code to your program to generate the mean, standard deviation and median for each of the dogs and complete the following table. You may need to create a table of your own in your writeup.

|  |  |  |  |
| --- | --- | --- | --- |
| **name** | **Mean** | **Standard Deviation** | **Median** |
| *Cody* | 77.125 | 5.987 | 77.000 |
| *Dusty* | 65.125 | 10.343 | 66.500 |
| *Shadow* | 78.500 | 9.813 | 79.500 |

A screenshot of a cell phone

Description automatically generated A picture containing bird

Description automatically generated

1. **(3 pts)** Create a new temporary SAS data set, named **onlyC**, in which the data is restricted to the times of *Cody* only. (Be sure to print the data as a check.)

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. **(3 pts)** Create a new temporary SAS data set, named **bothDS**, in which the data is restricted to the times of **Dusty** and **Shadow** only. (Be sure to print the data as a check.)

A close up of a telephone

Description automatically generated A screenshot of a cell phone

Description automatically generated

**R Code:**

# Statistics 147 Assignment #2

# Summer 2020

# Wesley Chang

# change working directory to assignment folder for easy access

setwd("/Users/wes\_chang/Library/Mobile Documents/com~apple~CloudDocs/Summer 2020 (UCR)/STAT 147 (Session A)/Assignments/2/")

# R Question 1

## Subpart i

# generate sequence 0 through 12 and store sequence in a variable named y

y <- seq(0,12)

y

## Subpart ii

# generate the pdf of y and store in var pdf1

# Y~Bin(n=12,p=0.07)

pdf1 <- dbinom(y,size=12,prob=0.07)

pdf1

## Subpart iii

# generate cdf of y and store in var cdf1

# y~bin(n=12,p=0.07)

cdf1 <- pbinom(y,size=12,prob=0.07)

cdf1

## Subpart iv

# Generate values of P(Y > y) and store in var sdf1

# Y~Bin(n=12,p=0.07)

sdf1 <- pbinom(y, size=12,prob=0.07, lower=FALSE)

sdf1

## Subpart v

# use cbind function to generate a data frame with y1, pdf1, cdf1, and sdf1

all\_together <- cbind(y,pdf1,cdf1,sdf1)

all\_together

## Subpart vi

# P(X = 3)

p6 <- dbinom(3,size=12,prob=0.07)

p6

## Subpart vii

# P(X <= 3)

p7 <- pbinom(3,size=12,prob=0.07, lower=TRUE)

p7

## Subpart viii

# P( 2 <= X <= 4) = P( X <= 4) - P(X <= 2)

p8 <- pbinom(4,size=12,prob=0.07, lower=TRUE) - pbinom(2,size=12,prob=0.07, lower=TRUE)

p8

## Subpart ix

plot(y, pdf1,

xlab = "y",

ylab = "pdf1",

main = "Plot of y vs pdf1"

)

## Subpart x

plot(y, cdf1,

xlab = "y",

ylab = "cdf1",

main = "Plot of y vs cdf1"

)

# R Question 2

# normally distributed, mean = 40 pounds, std dev = 0.50

# functioning level = 38.75 to 41.25

# randomly selected from assembly line

## Subpart i

# Prob bag is underfilled

# P(X<38.75)

# pnorm(x,mean,sd)

p1 <- pnorm(38.75,mean=40,sd=0.50)

p1

## Subpart ii

# Prob bag is not underfilled

# P(X > 38.75), lower tail is FALSE

p2 <- pnorm(38.75,mean=40,sd=0.50, lower=FALSE)

p2

## Subpart iii

# Prob bag is overfilled or underfilled

# P(X<38.75) + P(X>41.25)

p3 <- pnorm(38.75,mean=40,sd=0.50) + pnorm(38.75,mean=40,sd=0.50, lower=FALSE)

p3

## Subpart iv

# Prob bag has a fill amount that meets specifications

# P(38.75<X<41.25)

# = P(X < 41.25) - P(X < 38.75)

p4 <- pnorm(41.25,mean=40,sd=0.50) + pnorm(38.75,mean=40,sd=0.50)

p4

## Subpart v

# Find the 97th percentile

# use the quantile function qnorm(quantile, mean, sd)

p5 <- qnorm(0.97,mean=40,sd=0.50)

p5

# R Question 3

## Subpart i

# Read from file into p1 with header, skip one line

p1 = read.table(file="agility.dat",header=TRUE,skip=1,sep="")

p1

## Subpart ii

# add code that allows the columns to be accessible individually

attach(p1)

# display columns to verify that the code above worked

names(p1)

**SAS Code:**

title1 'Statistics 147 Assignment #2';

title2 'Summer 2020';

title3 'Wesley Chang';

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

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

ods graphics off;

title4 'Question 1';

/\* pdf of Y~Bin(n=25,p=0.65) \*/

/\* P(X=15) \*/

data question1;

/\* Create variables using pdf, cdf and sdf functions

Format:

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

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

P(X > x) = sdf(íBinomí,x,p,n) \*/

/\* Part (i)

p1 = P(X = 15) \*/

p1 = pdf('Binom',15,.65,25);

/\* Part (ii

p2 = P(X <= 15) \*/

p2 = cdf('Binom',15,.65,25);

/\* Part (iii)

p3 = P(16 < X < 19) =

= P( X <= 19) - P( X > 16) \*/

p3 = cdf('Binom',19,.65,25) - sdf('Binom',19,.65,25);

/\*Part (iv)

Find average number of dog owners

mu = n \* p \*/

p4 = 25 \* 0.65;

run;

/\* Print the results \*/

proc print noobs data = question1;

/\* Revise title4 and title5 \*/

title5 'Subpart (i) P(X = 15)';

var p1;

run;

proc print noobs data = question1;

title5 'Subpart (ii) P(X <= 15)';

var p2;

run;

proc print noobs data = question1;

title5 'Subpart (iii) P(X 16 < X <= 19)';

var p3;

run;

proc print noobs data = question1;

/\* Modify title5 \*/

title5 'Subpart (iv) Average number of dog owners that prefer a male dog';

var p4;

run;

title4 'Question 2';

/\* cdf of normal distribution, mean = 13.25, std dev = 0.10

appropriate level is between 13.15 and 13.35

\*/

data question2;

/\* Subpart (i) find probability that the can is underfilled

X~N, P( X <= 13.15)\*/

p1 = cdf('Normal',13.15,13.25,.1);

/\* Subpart (ii) find probability that can is not underfilled

X~N, P(X > 13.15) \*/

p2 = 1 - cdf('Normal',13.15,13.25,.1);

/\*Subpart (iii) find probability that can meets specification

X~N, P( 13.15 < X < 13.35)

= P(x <= 13.35) - P(X <= 13.15) \*/

p3 = cdf('Normal',13.35,13.25,.1) - cdf('Normal',13.15,13.25,.1);

/\* Subpart (iv) find probability that can is overfilled or underfilled

X~N, P(X < 13.15) + P(X > 13.35)

= P(X <= 13.15) + 1 - P(X <= 13.35) \*/

p4 = cdf('Normal',13.15,13.25,.1) + 1 - cdf('Normal',13.15,13.25,.1);

/\* Subpart (v) find the 97th percentile \*/

p5 = quantile('Normal',.97,13.25,.1);

run;

proc print noobs data = question2;

title5 'Subpart (i)';

var p1;

run;

proc print noobs data = question2;

title5 'Subpart (ii)';

var p2;

run;

proc print noobs data = question2;

title5 'Subpart (iii)';

var p3;

run;

proc print noobs data = question2;

title5 'Subpart (iv)';

var p4;

run;

proc print noobs data = question2;

title5 'Subpart (v)';

var p5;

run;

quit;

title5;

title4 'Question 3';

/\* y = 3n + sqrt(4.5m^-0.5n), for n=1,3,5 and m = 2,5,8 \*/

data question3;

/\* do loop for n \*/

do n = 1 to 5 by 2;

/\* nest m loop \*/

do m = 2 to 8 by 3;

/\* calculate equation given n and m and store in y \*/

y = 3 \* n + sqrt(4.5\*m\*\*(-0.5\*n));

output;

end;

end;

run;

/\* print results \*/

proc print noobs data = question3;

run;

quit;

title4 'Question 4';

title5 'Subpart (i)';

data question4;

/\* read file \*/

infile '\\apporto.com\dfs\UCR\Users\wchan061\_ucr\Desktop\agility.dat' firstobs = 3;

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

do rows = 1 to 8;

/\* do loop for 3 columns of data, assigned by owner \*/

do column = 1 to 3;

if column = 1 then owner = 'Cody';

else if column = 2 then owner = 'Dusty';

else owner = 'Shadow';

/\* input statement to read observations \*/

input time @@;

/\* output \*/

output;

end;

end;

run;

/\* print data \*/

proc print noobs data = question4;

var owner time;

run;

/\* Subpart (iii) \*/

title5 'Subpart (iii)';

proc sort;

by owner;

run;

proc print;

var owner time;

/\* Subpart (iv) \*/

title5 'Subpart (iv)';

proc means mean stddev median;

by owner;

var time;

quit;

/\* Subpart (v) \*/

title5 'Subpart (v)';

data onlyC;

set question4;

if owner = 'Cody';

run;

proc print data = onlyC;

run;

/\* Subpart (vi) \*/

title5 'Subpart (vi)';

data bothDS;

set question4;

if owner = 'Dust' or owner = 'Shad';

run;

proc print data = bothDS;

run;

quit;