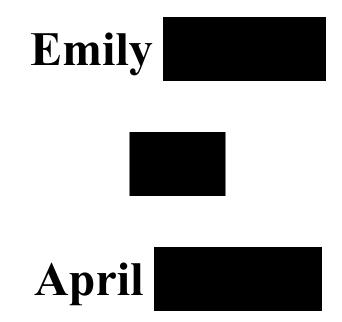
Statistics 157 Assignment #1

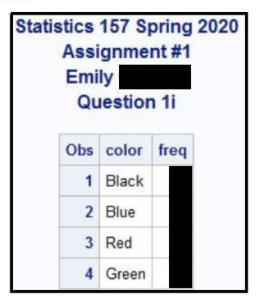


WRITE-UP

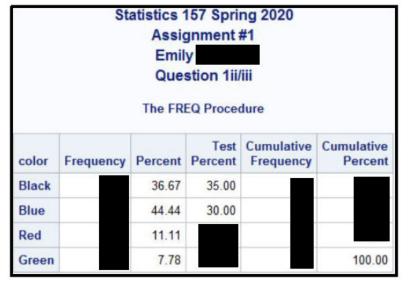
A national survey stated that 35% of the population prefer to use a black whiteboard marker, 30% prefer a
blue whiteboard marker, 25% prefer a red whiteboard marker and 10% prefer a green whiteboard marker.
Jill, Dan and James took a random sample of 90 instructors and asked them to state their whiteboard color
preference. The following data was recorded: Color Black Blue Red Green

Frequency 33 40 10 7

i. Write the SAS code necessary to read in and print out the data.



ii. Modify the code to generate the appropriate goodness of fit test information (output).



Chi-Square Test for Specified Proportions								
Chi-Square	13.7196							
DF	3							
Pr > ChiSq	0.0033							
Sample Si	ze = 90							

- iii. Perform the appropriate test of hypothesis, using $\alpha = 0.05$, to see if it's reasonable to assume the data supports the national claim. Use $\alpha = 0.05$.
 - H₀:
 - H_a:
 - p-value =
 - Rejection Region: Let $\alpha = 0.05$. Reject H if p-value $< \alpha = 0.05$
 - Conclusion:
- 2. Jun, Shuheng and Barry were interested in determining the distribution of birth months for students at UCR. They obtained a random sample of students and recorded the data in data file named **birthmonth_s20.dat**
 - i. Write a SAS program to read in and print out the data.

```
title4 'Question 2i';
/*read in the file*/
   infile 'C:\Users\person\Downloads\birthmonth_s20.dat'

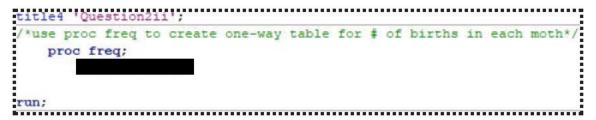
   /*inputthe data*/
   input

/*use proc print to print out data*/
   proc print;
run;
```

Statistics 157 Spring 2020
Assignment #1
Emily Question 2i

Obs	months	21	1	42	10	63	8	84	4
1		22	6	43	9	64	8	85	12
2		23	3	44	1	65	12	86	8
3		24	10	45	0	66	5	87	
4		25	12	46		67	6	88	
5		26	12	47		68	1	89	
6		27	10	48		69	10	90	
7		28		49		70	9	91	
8		29		50		71	4	92	
9		30		51		72	11	93	
10		31		52		73	10	94	
11		32		53		74	12	95	
12	1	33		54		75	1		
13	2	34		55		76	2		
14	9	35		56		77	12		
15	10	36		57	5	78	10		
16	2	37		58	6	79	12		
17	1	38		59	11	80	8		
18	9	39		60	7	81	5		
19	8	40		61	10	82	2		
20	2	41		62	12	83	10		

ii. Create a one-way table for the number of births in each of the 12 months



Statistics 157 Spring 2020
Assignment #1
Emily Question2ii
The FREQ Procedure

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1		7.37	7	7.37
2		8.42	15	15.79
3		3.16		18.95
4		7.37		26.32
5		9.47		35.79
6		6.32		42.11
7				47.37
8				58.95
9	7			66.32
10	14		77	81.05
11	7		84	88.42
12	11		95	100.00

- iii. What month has the highest number of births? (Be sure to give the month and the frequency.)

 Based on the results in part 2ii, the highest number is
- iv. What month has the smallest number of births? (Be sure to give the month and the frequency.)

 Based on the results in part 2ii, the highest number is
- 3. Luke, Shujie and Karen are the operations managers of a company that manufactures protective face masks. They want to determine whether there are significant differences in the quality of workmanship among the four daily shifts. They randomly select face masks from the three shifts and carefully inspect them. They classify each face mask as either perfect, satisfactory, or defective. They record this information along with the shift that produced the chains.

	Face Mask Classification							
Shift	Perfect	Satisfactory	Defective					
Shift 1	124	106	15					
Shift 2	95	35	10					
Shift 3	75	45	12					
Shift 4	35	25	5					

The data is located in a file named (facemask_s20.dat). Complete the following to assist Luke, Shujie and Karen with the analysis.

i. Use nested do-loops to read in and print out the data. Use if-then-else structures to give the appropriate names to face mask classification and shift.

```
/*read in the file*/
   infile 'C:\Users\person\Downloads\facemask s20.dat' firstobs = 2;
/*use nested do loops to read in the data*/
/*do loop for rows*/
   do row = 1 to 4;
   /*Use if-then-else statements to put names to each row*/
           if row = 1 then shift = 'Shift 1';
       else if row = 2 then shift = 'Shift 2';
       else if row = 3 then shift = 'Shift 3';
       else
                        shift = 'Shift 4';
   /*do loop for the columns*/
       do col = 1 to 3;
       /*Use if-then-else statements to put names to each col*/
              if col = 1 then cond = 'Perfect ';
           else if col = 2 then cond = 'Satisfactory';
                             cond = 'Defective ';
       /*input the data*/
          input mask @@;
       /*output the results*/
/*end the do loops*/
    end;
/*print out the data*/
   proc print;
run;
```

Statistics 157 Spring 2020 Assignment #1 Emily Question3i									
Obs	row	shift	col	cond	mask				
1	1	Shift 1	1	Perfect	124				
2	1	Shift 1	2	Satisfactory	106				
3	1	Shift 1	3	Defective	15				
4	2	Shift 2	1	Perfect	95				
5	2	Shift 2	2	Satisfactory	35				
6	2	Shift 2	3	Defective	10				
7	3	Shift 3	1	Perfect	75				
8	3	Shift 3	2	Satisfactory	45				
9	3	Shift 3	3	Defective	12				
10	4	Shift 4	1	Perfect	35				
11	4	Shift 4	2	Satisfactory	25				
12	4	Shift 4	3	Defective	5				

6 14.3212 0.0262

6 14.5670 0.0239

0.1255 0.7231

Prob

ii. Modify your code to generate the appropriate χ 2 test of independence information (output).

title4 'Question 3ii'; /*create two-way table using proc freq and the chi-quare info -used order to override SAS alphabetizing categories*/ proc freq weight col;

		1	he FREQ	Procedure			Statistics for Table of	shift	by cond	
	equency		Tab	le of shift by	cond		Statistic	DF	Value	
Ex	pected			cond	I				I American Conference	ŀ
		shift	Perfect	Satisfactory	Defective	Total	Chi-Square	6	14.3212	
		Shift 1	124	106	15	245	Likelihood Ratio Chi-Square		14.5670	
tistics 157 Sp Assignmen	CONTRACTOR OF THE PARTY OF THE		138.5	88.823	17.68	10.500	Mantel-Haenszel Chi-Square	1	0.1255	
Emily Question		Shift 2	95 79.141	35 50.756	10 10.103	140	Phi Coefficient		0.1569	
		Shift 3	75	45	12	132	Contingency Coefficient		0.1550	
			74.619 47.856 9.52		9.5258		Cramer's V		0.1109	ĺ
		Shift 4	35 36.744	25 23.565	5 4.6907	65	and the top gas reasons			
		Total	329	211	42	582	Sample Size	= 582	2	

111.	Perfo	orm the	approp	riate tes	t of hy	pothesis	to	determine	whethe	er tl	iere is	a r	elationship	between	face
	mask	classif	ication	and shift											
	•	H_0 :													

• H_a: Fa ndent (i.e. two)

• p-value =

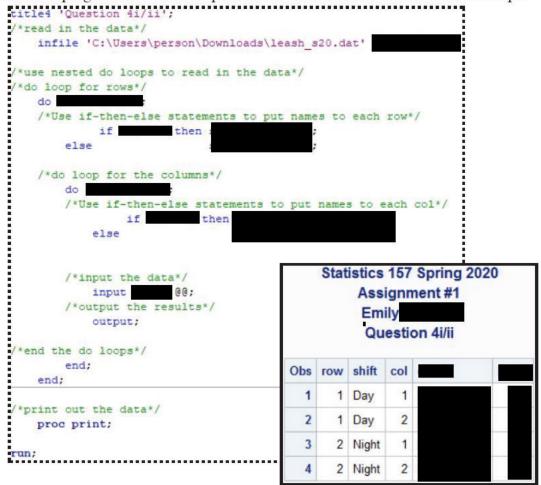
• Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$

Conclusion:

4. Analisa, Wenxiu and Subir work as quality control engineers for a company that manufactures lighted dog leashes. They are concerned that the proportion of defective items differs between the day shift and the night shift. They obtain a random sample of data. Perform the appropriate test. They obtain a random sample of data and cross-classify it (shown below). This data is located in a file named **leash s20.dat**.

	Dog Leash Classification						
Shift	Nondefective	Defective					
Day	71	15					
Night	84	28					

i. Write a SAS program to read in and print out the data. Be sure to use nested do loops.



ii. Use if-then-else structures to name your rows and columns appropriately.

iii. Modify your code to generate the appropriate χ 2 test of independence information (output).

title4 'Question 4iii';
/*create two-way table using proc freq and the chi-quare info -used order to override SAS alphabetizing categories*/ proc freq run;

8		cs 157 Spring						
		ssignment #1 mily						
		Question 4iii						
	The	FREQ Procedu	re					
Frequency Expected	Table of by							
		Non-defective	Defective	Total				
	Day							
	Night							
	Total							

Statis	tic	DF	Value	Prob
Chi-S	quare	1	1.6346	
Likeli	hood Ratio Chi-Square	1	1.6600	
Conti	nuity Adj. Chi-Square	1	1.2203	
Mante	el-Haenszel Chi-Square	1	1.6263	
Phi C	oefficient		0.0909	
Conti	ngency Coefficient		0.0905	
Cram	er's V		0.0909	
	Fisher's Exact Cell (1,1) Frequency			
	Left-sided Pr <= F			
	Right-sided Pr >= F			
	Table Probability (P)			
	Two-sided Pr <= P			

- iv. Perform the appropriate test of hypothesis.
 - H_0 :
 - H_a:
 - p-value =

 - Rejection Region: Reject H₀ if p-value < α = 0.05
 Conclusion: Since the p-value = reject $H_0 \rightarrow it$ is reasonable to assume

5. Xinping, Yehua and Weixin believe that the time between hits, measured in seconds, on a popular Web site follows an exponential distribution with β = 65. They obtain a random sample of time between hits and record the information in a data file named time s20.dat.

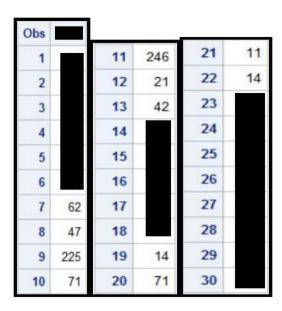
```
time
23 261 87 7 120 14 62 47 225 71 246 21 42 20 12
120 11 5 14 71 11 14 11 16 90 16 52 95 6 30
```

i. Write the SAS code necessary to read in and print out the data.

```
title4 'Question 5i';
/*read in the data*/
   infile 'C:\Users\person\Downloads\time_s20.dat'
   /*inputthe data*/

   /*use proc print to print out data*/
   proc print;
run;
```

Statistics 157 Spring 2020
Assignment #1
Emily
Question 5i



ii. Modify your code to generate the appropriate information (output) to test the hypothesis that the time between hits follows an exponential distribution with $\beta = 65$.

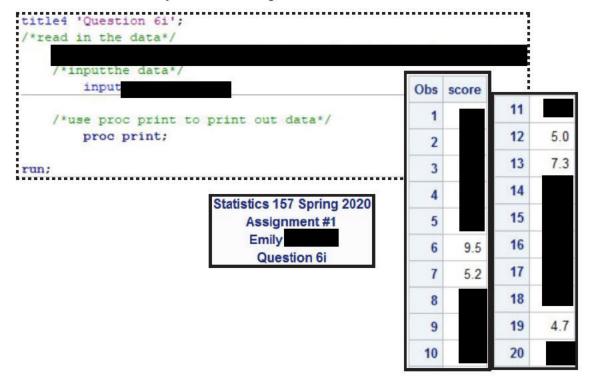


- iii. Perform the appropriate test, using $\alpha = 0.05$
 - H₀:
 - H_a:
 - p-value =
 - Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$
 - Conclusion:
- 6. Linda and Esra were interested in determining whether the quiz scores in a large introductory statistics course are normally distributed. They obtained the following random sample of quiz scores:

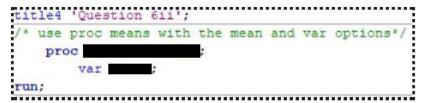
```
scores
  4.0
        8.5
              6.4
                    6.1
                          5.8
                                9.5
                                     5.2
                                           6.7
                                                 8.3
                                                       9.2
  9.1
        5.0
              7.3
                    7.4
                          5.5
                               8.6
                                     7.0
                                           4.3
                                                 4.7
                                                       8.0
```

The data has been saved in a data file named quiz s20.dat.

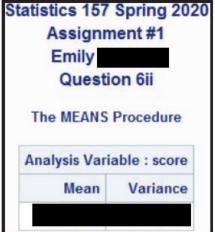
i. Write the SAS code necessary to read in and print out the data.



ii. Write the SAS code necessary to generate the mean and variance for the quiz scores. Be sure to state the values

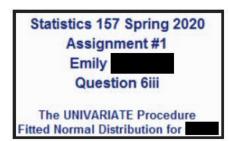


Mean: Variance:



iii. Modify your code to generate the appropriate information (output) to test the hypothesis that the quiz scores in a large introductory statistics course are normally distributed.

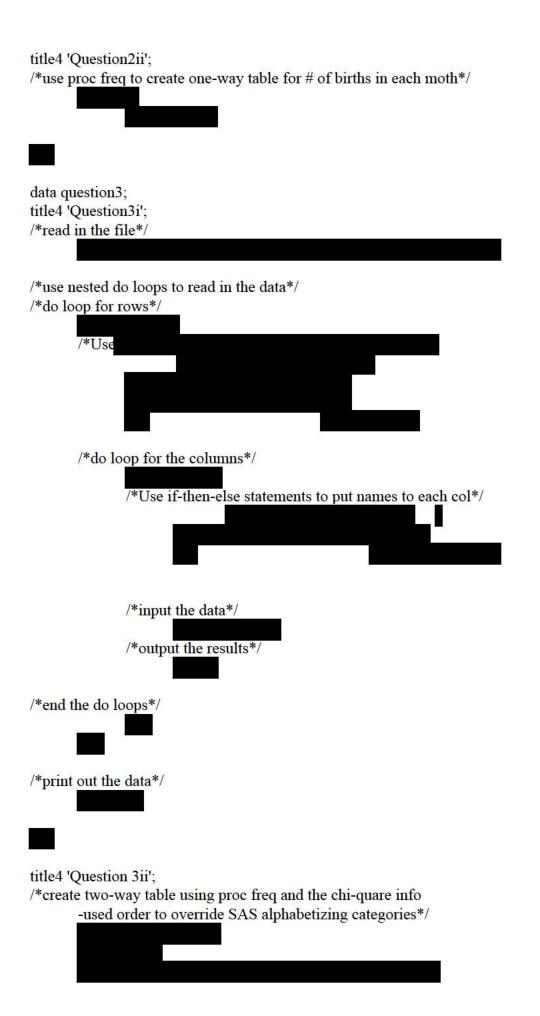
```
title4 'Question 6iii';
/*check to see if distribution follows normal using proc univariate*/
    proc univariate;
    ods selece GoodnessOfFit;
    var score;
    histogram score/normal;
run;
```



Goodness-of-F	it Tests for N	Iormal Distribution
Test	Statist	ic p Value
Kolmogorov-Smirnov	D	Pr > D
Cramer-von Mises	W-Sq	Pr > W-Sq
Anderson-Darling	A-Sq	Pr > A-Sq

- iv. Perform the appropriate test, using $\alpha = 0.05$.
 - H₀:
 - H_a:
 - p-value =
 - Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$
 - Conclusion:

```
The Code
DM log "odsresults; clear;out;clear;log; clear;";
ods graphics off;
title1 'Statistics 157 Spring 2020';
title2 'Assignment #1';
title3 'Emily
title4 'Question 1i';
/*input variable list*/
       /*read in data up to 5 letters*/
/*input the data into the list*/
/* print out the data*/
title4 'Question 1ii/iii';
/* use proc freq w/ order = data --> generates goodness of fit info
       tables classification variable
       test[ = (percentages_to_be_tested)
       weight wt = data frequencies */
data question2;
title4 'Question 2i';
/*read in the file*/
       /*inputthe data*/
       /*use proc print to print out data*/
```



```
/*read in the data*/
/*use nested do loops to read in the data*/
/*do loop for rows*/
       /*Use if-then-else statements to put names to each row*/
       /*do loop for the columns*/
               /*Use if-then-else statements to put names to each col*/
               /*input the data*/
               /*output the results*/
/*end the do loops*/
/*print out the data*/
                       using proc freq and the chi-quare info
       -used order to override SAS alphabetizing categories*/
data question5;
title4 'Question 5i';
/*read in the data*/
       /*inputthe data*/
```

