Statistics 157 Assignment #1

Emily [redacted]

[redacted]

April [redacted]

WRITE-UP

1. 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.

```
title4 'Question 1i';

/*input variable list*/
    input co[redacted]eq;
    /*read in data up to 5 letters*/
        length color $ 5;

/*input the data into the list*/
    datalines;
    Black 33
    [redacted]
    ;

/* print out the data*/
    proc print;
run;
```

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question 1i

Obs	color	freq
1	Black	[redacted]
2	Blue	
3	Red	
4	Green	

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

```
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 */
    proc freq [redacted];
        tabl[redacted]/chisq [redacted];
        wei[redacted];

run;
```

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question 1ii/iii

The FREQ Procedure

color	Frequency	Percent	Test Percent	Cumulative Frequency	Cumulative Percent
Black	[redacted]	36.67	35.00	[redacted]	[redacted]
Blue		44.44	30.00		
Red		11.11	[redacted]		
Green		7.78			100.00

Chi-Square Test for Specified Proportions	
Chi-Square	13.7196
DF	3
Pr > ChiSq	0.0033

Sample Size = 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 $\alpha = 0.05$.

- H_0: [redacted]
- H_a: [redacted]
- **p-value** = [redacted]
- Rejection Region: Let $\alpha = 0.05$. Reject H if p-value $< \alpha = 0.05$
- **Conclusion**: [redacted].

---

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' [redacted]

    /*inputthe data*/
        input [redacted]

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

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question 2i

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

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

```
title4 'Question2ii';
/*use proc freq to create one-way table for # of births in each moth*/
    proc freq;
        [redacted]

run;
```

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question2ii

The FREQ Procedure

[redacted]	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	[redacted]	7.37	7	7.37
2		8.42	15	15.79
3		3.16	[redacted]	18.95
4		7.37		26.32
5		9.47		35.79
6		6.32		42.11
7		[redacted]		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 [redacted]

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 [redacted]

---

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.

```
title4 'Question3i';
/*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';
            else                 cond = 'Defective   ';


        /*input the data*/
            input mask @@;
        /*output the results*/
            output;

/*end the do loops*/
        end;
    end;

/*print out the data*/
    proc print;

run;
```

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
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

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 [redacted]
    weight [redacted]
    t[redacted]ol;

run;
```

tistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question 3ii

The FREQ Procedure

Frequency Expected

Table of shift by cond

shift	cond: Perfect	Satisfactory	Defective	Total
Shift 1	124 138.5	106 88.823	15 17.68	245
Shift 2	95 79.141	35 50.756	10 10.103	140
Shift 3	75 74.619	45 47.856	12 9.5258	132
Shift 4	35 36.744	25 23.565	5 4.6907	65
Total	329	211	42	582

Statistics for Table of shift by cond

Statistic	DF	Value	Prob
Chi-Square	6	14.3212	0.0262
Likelihood Ratio Chi-Square	6	14.5670	0.0239
Mantel-Haenszel Chi-Square	1	0.1255	0.7231
Phi Coefficient		0.1569	
Contingency Coefficient		0.1550	
Cramer's V		0.1109	

Sample Size = 582

iii. Perform the appropriate test of hypothesis to determine whether there is a relationship between face mask classification and shift.

- H_0: [redacted]
- H_a: Fa[redacted]ndent (i.e. [redacted] two)
- **p-value** = [redacted]
- Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$
- **Conclusion**: [redacted]

---

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**.

Shift	Dog Leash Classification	
	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.

```
title4 'Question 4i/ii';
/*read in the data*/
   infile 'C:\Users\person\Downloads\leash_s20.dat' [redacted]

/*use nested do loops to read in the data*/
/*do loop for rows*/
   do [redacted];
   /*Use if-then-else statements to put names to each row*/
            if [redacted] then [redacted];
       else [redacted];

   /*do loop for the columns*/
       do [redacted];
       /*Use if-then-else statements to put names to each col*/
                if [redacted] then [redacted]
           else


       /*input the data*/
           input [redacted] @@;
       /*output the results*/
           output;

/*end the do loops*/
       end;
   end;

/*print out the data*/
   proc print;

run;
```

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question 4i/ii

Obs	row	shift	col	[redacted]	[redacted]
1	1	Day	1		
2	1	Day	2		
3	2	Night	1		
4	2	Night	2		

ii. Use if-then-else structures to name your rows and columns appropriately.
See Question 4.ii

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 [redacted]
    we[redacted]
    ta[redacted];

run;
```

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question 4iii

The FREQ Procedure

Frequency Expected	Table of [redacted] by [redacted]		
	[redacted]		
[redacted]	Non-defective	Defective	Total
Day	[redacted]		
Night			
Total			

Statistics for Table of shift by cond

Statistic	DF	Value	Prob
Chi-Square	1	1.6346	[redacted]
Likelihood Ratio Chi-Square	1	1.6600	
Continuity Adj. Chi-Square	1	1.2203	
Mantel-Haenszel Chi-Square	1	1.6263	
Phi Coefficient		0.0909	
Contingency Coefficient		0.0905	
Cramer's V		0.0909	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	[redacted]
Left-sided Pr <= F	
Right-sided Pr >= F	
Table Probability (P)	
Two-sided Pr <= P	

Sample Size = 198

iv. Perform the appropriate test of hypothesis.

- H_0: [redacted]
- H_a: [redacted]
- **p-value** = [redacted]
- Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$
- **Conclusion:** Since the p-value = [redacted] reject H_0 → it is reasonable to assume [redacted]

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' [redacted]
    /*inputthe data*/
        [redacted]

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

run;
```

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question 5i

Obs					
1	[redacted]	11	246	21	11
2		12	21	22	14
3		13	42	23	[redacted]
4		14	[redacted]	24	
5		15		25	
6		16		26	
7	62	17		27	
8	47	18		28	
9	225	19	14	29	
10	71	20	71	30	

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

```
title4 'Question 5ii';
/*check to see if distribution follows exponential using proc univariate*/
    proc [redacted]
        ods [redacted];
        var [redacted]
            [redacted]

run;
```

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question 5ii

The UNIVARIATE Procedure
Fitted Exponential Distribution for time

Goodness-of-Fit Tests for Exponential Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.20198296	Pr > D	[redacted]
Cramer-von Mises	W-Sq	[redacted]	Pr > W-Sq	
Anderson-Darling	A-Sq		Pr > A-Sq	>0.250

iii. Perform the appropriate test, using $\alpha = 0.05$

- H_0: [redacted]
- H_a: [redacted]
- **p-value** = [redacted]
- Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$
- **Conclusion:** [redacted]

---

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.

```
title4 'Question 6i';
/*read in the data*/
    [redacted]
    /*inputthe data*/
        input [redacted]

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

run;
```

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question 6i

Obs	score
1	[redacted]
2	
3	
4	
5	
6	9.5
7	5.2
8	
9	
10	
11	[redacted]
12	5.0
13	7.3
14	
15	
16	
17	
18	
19	4.7
20	[redacted]

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

```
title4 'Question 6ii';
/* use proc means with the mean and var options*/
    proc [redacted];
        var [redacted];
run;
```

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question 6ii

The MEANS Procedure

Analysis Variable : score	
Mean	Variance
[redacted]	

Mean: [redacted]
Variance: [redacted]

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;
```

Statistics 157 Spring 2020
Assignment #1
Emily [redacted]
Question 6iii

The UNIVARIATE Procedure
Fitted Normal Distribution for [redacted]

Goodness-of-Fit Tests for Normal Distribution

Test	Statistic		p Value	
Kolmogorov-Smirnov	D	[redacted]	Pr > D	[redacted]
Cramer-von Mises	W-Sq		Pr > W-Sq	
Anderson-Darling	A-Sq		Pr > A-Sq	[redacted]

iv. Perform the appropriate test, using $\alpha = 0.05$.

- H_0: [redacted]
- H_a: [redacted]
- **p-value** = [redacted]
- Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$
- **Conclusion:** [redacted]

The Code

```
DM log "odsresults; clear;out;clear;log; clear;";
ods graphics off;

title1 'Statistics 157 Spring 2020';
title2 'Assignment #1';
title3 'Emily [redacted]';

[redacted];
title4 'Question 1i';

/*input variable list*/
        [redacted]
        /*read in data up to 5 letters*/
                [redacted]

/*input the data into the list*/
        [redacted]

/* print out the data*/
        [redacted]
[redacted]

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 */
        [redacted]

[redacted]

data question2;
title4 'Question 2i';
/*read in the file*/
        [redacted]

        /*inputthe data*/
                [redacted]

        /*use proc print to print out data*/
                [redacted]
[redacted]
```

```
title4 'Question2ii';
/*use proc freq to create one-way table for # of births in each moth*/
        [redacted]

[redacted]

data question3;
title4 'Question3i';
/*read in the file*/
        [redacted]

/*use nested do loops to read in the data*/
/*do loop for rows*/
        [redacted]
        /*Use [redacted]

        /*do loop for the columns*/
                [redacted]
                /*Use if-then-else statements to put names to each col*/
                        [redacted]


                /*input the data*/
                        [redacted]
                /*output the results*/
                        [redacted]

/*end the do loops*/
                [redacted]
        [redacted]

/*print out the data*/
        [redacted]

[redacted]

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

[redacted]

[redacted]
/*read in the data*/
[redacted]

/*use nested do loops to read in the data*/
/*do loop for rows*/
[redacted]
/*Use if-then-else statements to put names to each row*/
[redacted]

/*do loop for the columns*/
[redacted]
/*Use if-then-else statements to put names to each col*/
[redacted]

/*input the data*/
[redacted]
/*output the results*/
[redacted]

/*end the do loops*/
[redacted]

/*print out the data*/
[redacted]

[redacted]

[redacted] using proc freq and the chi-quare info
-used order to override SAS alphabetizing categories*/
[redacted]

[redacted]

data question5;
title4 'Question 5i';
/*read in the data*/
[redacted]
/*inputthe data*/

[redacted]

/*use proc print to print out data*/
[redacted]

[redacted]

title4 'Question 5ii';
/*check to see if distribution follows exponential using proc univariate*/
[redacted]

[redacted]

data question6;
title4 'Question 6i';
/*read in the data*/
[redacted]
/*inputthe data*/
[redacted]

[redacted]

[redacted]

title4 'Question 6ii';
/* use proc means [redacted] */
[redacted]

[redacted]

title4 'Question 6iii';
/*check to see if distribution follows normal using proc univariate*/
[redacted]

[redacted]
quit;