Nikhil Jagatia
18055146

Assignment 5

1.

a)

Individual confidence limits for fatalities:

```
DATA EBOLA;
INPUT COUNTRY$ COUNT;
DATALINES;
GUINEA 837
GUINEADEATHS 1327
;
RUN;


PROC FREQ DATA = EBOLA;
            WEIGHT COUNT;
            TABLES COUNTRY / BINOMIAL (LEVEL=2);
RUN;
```

Exact Conf Limits	
95% Lower Conf Limit	0.5923
95% Upper Conf Limit	0.6338

```
DATA EBOLA;
INPUT COUNTRY$ COUNT;
DATALINES;
LIBERIA 4490
LIBERIADEATHS 3145
;
RUN;

PROC FREQ DATA = EBOLA;
            WEIGHT COUNT;
            TABLES COUNTRY / BINOMIAL (LEVEL = 2) ;
RUN;
```

Exact Conf Limits	
95% Lower Conf Limit	0.4008
95% Upper Conf Limit	0.4231

Differences in fatality rates between Guinea and Liberia :

```
DATA EBOLA;
INPUT COUNTRY $ DEATHS $ COUNT;
DATALINES;
GUINEA DEATHS 1327
GUINEA NODEATHS 837
LIBERIA DEATHS 3145
LIBERIA NODEATHS 4490
;
RUN;

PROC FREQ DATA=EBOLA
ORDER=DATA;
WEIGHT COUNT;
TABLES COUNTRY*DEATHS / CHISQ
RISKDIFF;
```

Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Difference is (Row 1 - Row 2)						
Row 1	0.6132	0.0105	0.5927	0.6337	0.5923	0.6338
Row 2	0.4119	0.0056	0.4009	0.4230	0.4008	0.4231
Total	0.4564	0.0050	0.4465	0.4662	0.4465	0.4663
Difference	0.2013	0.0119	0.1780	0.2246		

This shows that the 95% confidence interval for difference is 0.1780, 0.2246

Nikhil Jagatia
18055146

b)
The confidence intervals show that Guinea has greater 95% intervals compared to that of Liberia with no overlapping, however the range of the intervals is higher for Guinea. This suggests that the rate in Guinea is higher than Liberia with over half of the diagnosed patients had fatalities.

c)
Sierra Leone 95% confidence interval

Exact Conf Limits	
95% Lower Conf Limit	0.2071
95% Upper Conf Limit	0.2261

```
DATA EBOLA;
INPUT COUNTRY$ COUNT;
DATALINES;
SL 5729
SLDEATHS 1583
;
RUN;
PROC FREQ DATA = EBOLA;
                WEIGHT COUNT;
                TABLES COUNTRY / BINOMIAL (LEVEL=2);
RUN;
```

The data and confidence intervals show that the rate was more than double in Guinea (61.32% fatalities) than Sierra Leone (21.65%) for deaths from Ebola. In addition the P value of the Chi-Squared is <0.05 therefore we can state that if the null hypothesis was that there was a significant difference in the rates it would be accepted.
A reason for this could be because the virus originally began in Guinea therefore there had been more chance of spreading the virus. In addition these values found show the reported diagnosed cases, worryingly it could indicate the actual number of deaths in Guinea could be greatly higher.

Frequency
Percent
Row Pct
Col Pct

Table of COUNTRY by DEATHS

COUNTRY	DEATHS: DEATHS	DEATHS: NODEATHS	Total
GUINEA	1327	837	2164
	14.00	8.83	22.84
	61.32	38.68	
	45.60	12.75	
SIERRALE	1583	5729	7312
	16.71	60.46	77.16
	21.65	78.35	
	54.40	87.25	
Total	2910	6566	9476
	30.71	69.29	100.00

Statistics for Table of COUNTRY by DEATHS

Statistic	DF	Value	Prob
Chi-Square	1	1235.0912	<.0001

Nikhil Jagatia
18055146

2.

a) The appropriate test to use is paired t-test this is because firstly of the equal number of observations and the mutual plot number associated with each type of fertiliser. Also that the two sets of data have a mutual relationship being that they represent sequential plots, where A and B shared the same plot in each instance.

b)

$H_0 : \mu_{indifferent}$ = There is not a significant difference in yields of fertilisers = 0

$H_1 : \mu_{different}$ = There is a significant difference in yields of fertilisers ≠ 0

c)

```
DATA FERTILISERS;
INPUT A B;
LABEL A = FERTILISER A;
LABEL B = FERTILISER B;
DATALINES;
56 67
62 72
74 79
94 86
52 71
94 90
97 86
80 65
78 85
44 56
52 61
51 66
;
RUN;

PROC TTEST DATA=FERTILISERS ALPHA=0.05 H0=0;
PAIRED A*B;
RUN;
```

DF	t Value	Pr > |t|
11	-1.32	0.2150

The P value found is 0.2150 which is not less than 0.05 so we cannot reject the H_0 at the 5% significance level.

This shows there is not a significant difference between fertiliser A and B

Nikhil Jagatia
18055146

d)

The mean yield for each fertiliser found by the proc univariate is A= 69.5 B=73.67 showing a difference of 4.17

Knowing these values are not significant either Fertiliser A or B should be chosen as although there is a slight difference it is not significant.

Basic Confidence Limits Assuming Normality FERTILISER A

Parameter	Estimate	95% Confidence Limits	
Mean	69.50000	57.37943	81.62057
Std Deviation	19.07640	13.51363	32.38940
Variance	363.90909		

Basic Confidence Limits Assuming Normality FERTILISER B

Parameter	Estimate	95% Confidence Limits	
Mean	73.66667	66.52780	80.80553
Std Deviation	11.23577	7.95936	19.07696
Variance	126.24242		

```
PROC UNIVARIATE DATA=FERTILISERS
CIBASIC;
RUN
```

3.

If we want to analyse the data to identify if there is a change in the weight of rats that have undergone Ozone treatment compared to those rats in the control group that were not exposed to ozone:

H_0 : μ = There is not a significant change in the weights of rats being treated with Ozone = 0
H_1 : μ = There is a significant change in the weights of rats being treated with Ozone $\neq$ 0

Since we are dealing and comparing with two population proportions, in SAS we can use the TTest where we can see the appropriate value and then the 95% confidence interval.

As we are not sure of the variances we can use the proc ttest to evaluate the P value and the 95% confidence interval. Following the code, we can clearly see from the diagrams the variances differ and are unequal so the P value is 0.0192 which is less than 0.05 which is significant at the 5% level therefore we can reject the null hypothesis as accept the alternate hypothesis. In essence we can see that there is a significant change in the weights of rats being treated with ozone.
Furthermore we can see the difference in mean weight change is a little over 11. In addition we can see the 95% confidence interval as 1.9850, 20.84 for the mean differences and that the sample population means

From the visual representation of the distribution plots (shown below of weight change) and the inter-quartile ranges the variances are not the same, hence the broader distribution in

Nikhil Jagatia
18055146

the treatment groups indicating greater levels of fluctuation. Furthermore there seems to more outliers in the control group.
The data given does indicate that there were more rats in the control group, with only similar ages therefore not only the number could have had an effect on results but also the individual differences of the rats.

```
FILENAME REFFILE '/folders/myfolders/sasuser.v94/ozr.csv';
        PROC IMPORT DATAFILE=REFFILE
                DBMS=CSV
                OUT=WORK.IMPORT;
                GETNAMES=YES;
        RUN;
        PROC CONTENTS DATA=WORK.IMPORT; RUN;
        RUN;
        PROC PRINT DATA= WORK.IMPORT; RUN;

        PROC TTEST DATA=WORK.IMPORT;
        CLASS GROUP;
        VAR WEIGHT;
        RUN;
```

The TTEST Procedure

Variable: Weight

Group	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Control		23	22.4261	10.7768	2.2471	-16.9000	41.0000
Treatment		22	11.0091	19.0171	4.0545	-15.9000	54.6000
Diff (1-2)	Pooled		11.4170	15.3636	4.5817		
Diff (1-2)	Satterthwaite		11.4170		4.6355		

Group	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Control		22.4261	17.7659	27.0863	10.7768	8.3347	15.2529
Treatment		11.0091	2.5774	19.4408	19.0171	14.6308	27.1767
Diff (1-2)	Pooled	11.4170	2.1772	20.6568	15.3636	12.6937	19.4660
Diff (1-2)	Satterthwaite	11.4170	1.9850	20.8489			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	43	2.49	0.0166
Satterthwaite	Unequal	32.918	2.46	0.0192

Equality of Variances				
Method	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Folded F	21	22	3.11	0.0107

Nikhil Jagatia
18055146