

Homework #1

**Question 1**

Note that the sample size on the homework sheet is 21, while that provided in lab is 25. The results are similar.

1) Answers and tables shown below.

a) The percent of men in the sample is 52%.

**sex**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid male	13	52.0	52.0	52.0
female	12	48.0	48.0	100.0
Total	25	100.0	100.0	

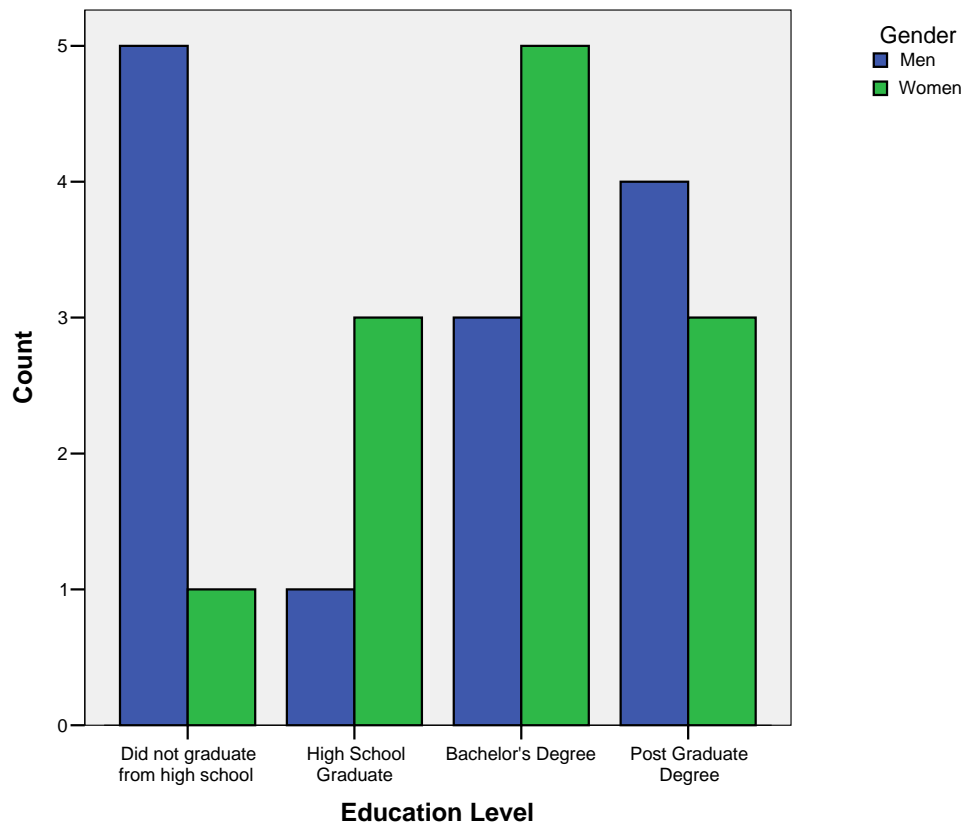
b) The mode for marital status is divorced.

**Marital status**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid married	9	36.0	36.0	36.0
divorced	11	44.0	44.0	80.0
never married	5	20.0	20.0	100.0
Total	25	100.0	100.0	

c) The frequency of divorced people in the sample is 11.

2) There are more men with below high school and postgraduate education, and in contrast there are more women at the high school graduates and college graduate levels.

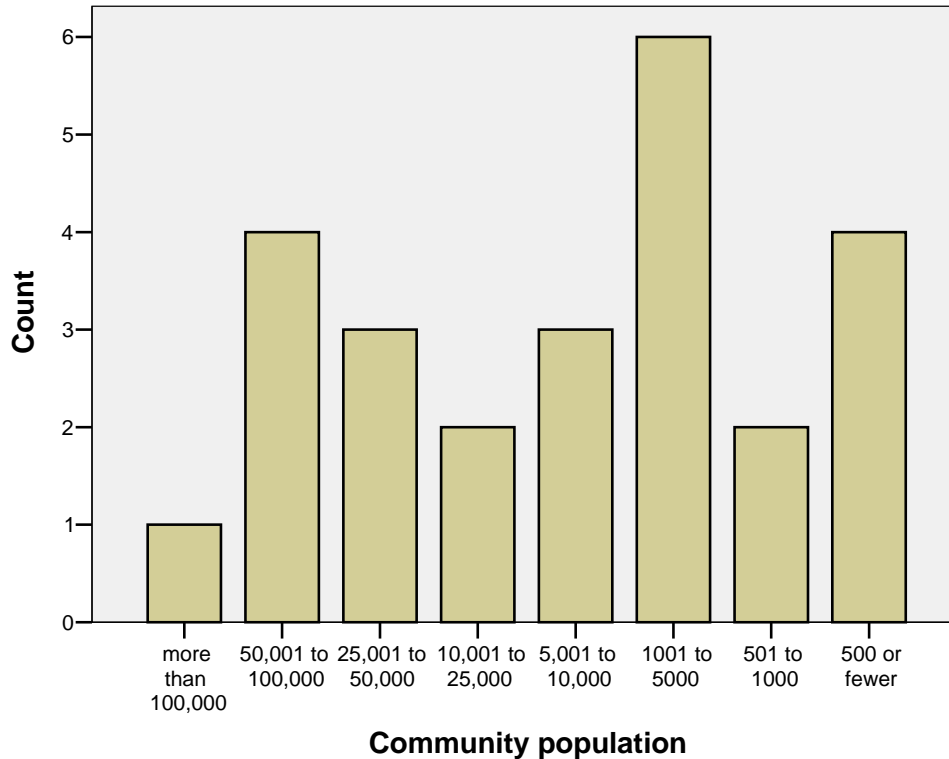


3)

**Education**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	below high school	6	24.0	24.0	24.0
	high school graduate	4	16.0	16.0	40.0
	college graduate	8	32.0	32.0	72.0
	post graduate degree	7	28.0	28.0	100.0
	Total	25	100.0	100.0	

4)



5) There are about the same number of men as women. In the Ann sample, divorced people are more prevalent and never-married people are less prevalent. A large number of men did not graduate from high school while a large number of women are college graduates. Most people live in communities with a population between 1,001 to 5,000 people.

**Question 2.**

2) Here I attach only histograms (Figures 1 and 2) and boxplot (Figure 3) for REACT at 2 dose levels and provide solutions for REACT and LIVER\_WT by dose levels. The rest is very similar. Please refer to the solutions for ST4950 if you want more details.

The normality tests for the variables react and liver\_wt are shown below. React is not normal at either dose level (small p-values) and liver\_wt remains normal at both dose levels (not small p-values).

**Tests of Normality**

	dose	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
React	1.00	.375	5	.021	.708	5	.012
	2.00	.394	5	.011	.655	5	.003
Liver_wt	1.00	.189	5	.200(*)	.942	5	.679
	2.00	.201	5	.200(*)	.950	5	.736

\* This is a lower bound of the true significance.

a Lilliefors Significance Correction

The histograms for react are both skewed and the ones for liver\_wt are not too skewed. The side-by-side boxplots also indicate asymmetry for react. However, the side-by-side boxplots for liver weights are not both symmetric. The boxplot for dose two is asymmetric while that for dose one is symmetric.

3) From the SPSS output:

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
React	Equal variances assumed	.177	.685	-.092	8	.929	-.3000	3.2691	-7.8386	7.2386
	Equal variances not assumed			-.092	7.625	.929	-.3000	3.2691	-7.9036	7.3036

The equal variance is a reasonable assumption since the p-value 0.685 is not small. However, the normal assumption for REACT is failed. Therefore we should conduct distribution-free test. Since we have not taught it yet, here I still conduct t-test. According to the t-test, we can not reject the null hypothesis (p-value =.929). That is, we have insufficient evidence to claim there is a significant difference between REACT for DOSE 1 and REACT for DOSE 2.

4) From the SPSS output:

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Liver_WT	Equal variances assumed	.573	.471	-.783	8	.456	-.6400	.8172	-2.5244	1.2444
	Equal variances not assumed			-.783	7.010	.459	-.6400	.8172	-2.5718	1.2918

The equal variance is a reasonable assumption since the p-value 0.471 is not small. In addition, based on the previous parts, the normal assumption for LIVER\_WT is reasonable. Therefore, we can use (pooled) t-test. The p-value is 0.456 and so we cannot reject the null hypothesis. In conclusion, we have insufficient evidence to claim there is a significant difference between LIVER\_WT for DOSE 1 and LIVER\_WT for DOSE 2.

### Optional question

1) SAS draw pie chart. Suppose we use the original 25 pieces of data and want to show a pie chart of educational level. The required SAS code would be:

```
data education;  
input education;  
datalines;
```

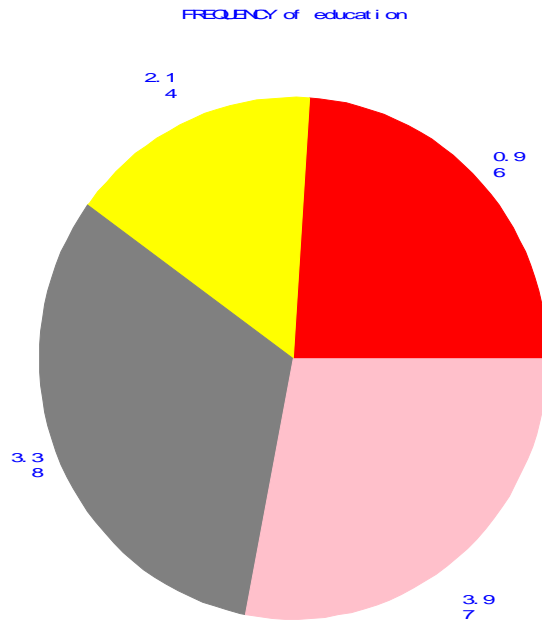
```
4  
4  
3  
1  
2  
1  
3  
4  
1  
2  
3  
4  
1  
3  
1  
4  
3  
2  
3  
3  
3  
4  
1  
2  
3  
4
```

```
;
```

```
proc freq data=education;  
tables education;
```

```
proc gchart data=education;  
pie education; run;
```

The outputted graph is:



2) SPSS different base intervals. Have SPSS draw the histogram automatically. For example draw the histogram for the variable react in the biomedical data. Double click on the histogram until it is in edit mode. Click on the scale that you wish to change, for example the x axis. The interval values appear in blue boxes. In the properties window, click on the histogram options tab. Click custom value for anchor and change value, for example to zero. Click bin size to custom and enter the number of bins desired or the width of the bins, for example width of 5. Close the property window and close the chart editor. The new histogram appears such as the one for react below.

Appendix:

Figure 1

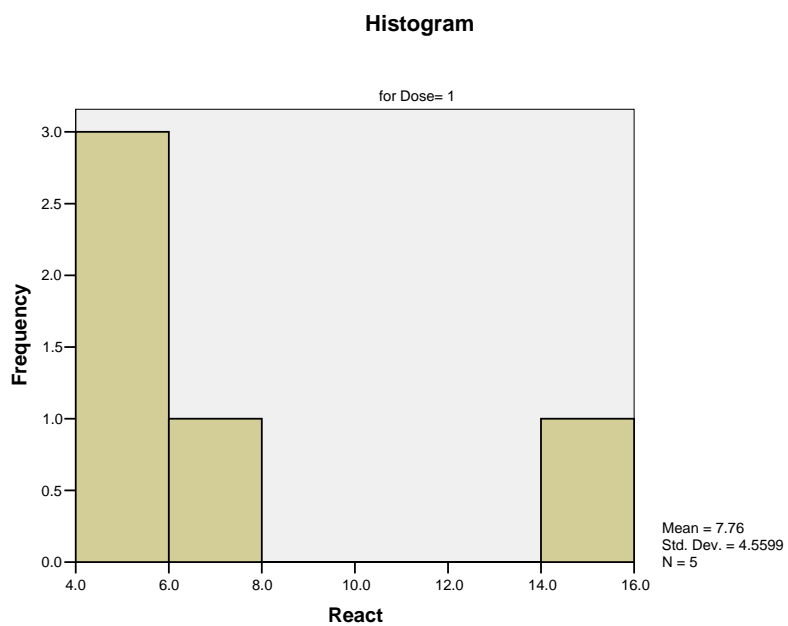


Figure 2

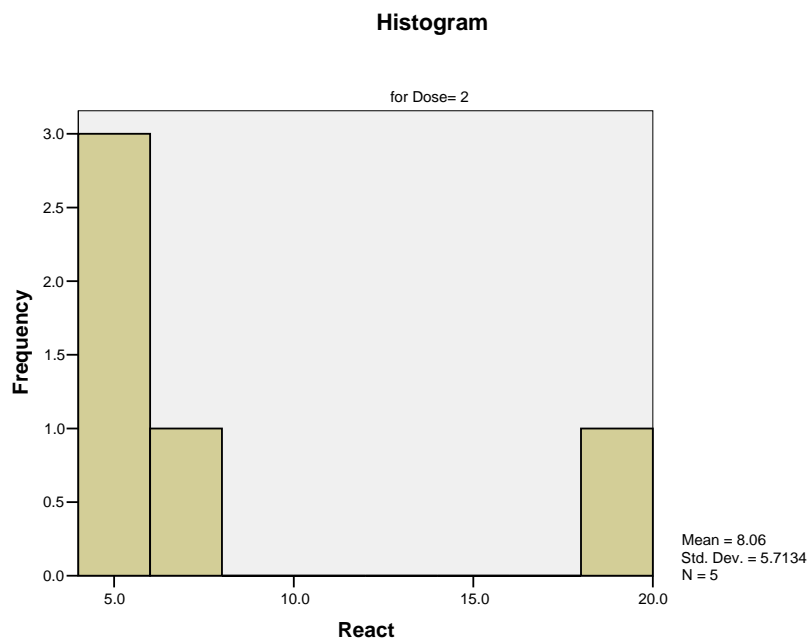


Figure 3

