MSDS 6372: Homework 1

1. Chapter 23 Problem 13:
   The standard treatment of ovarian cancer by the chemicals cytoxan and cisplatin substantially shrinks or eliminates tumors in about 60% of patients. Taxol, derived from the bark of yew trees, is expected to improve upon those response rates. If a clinical trial is conducted to randomly assign patients to the standard chemicals or to taxol, with equal numbers in each group, how large a sample size would be needed to distinguish a 75% response rate for taxol, at 95% confidence? (Data from "Yew Drugs Show Their Mettle," Science News 143 (May 29, 1993): 344.)

```
proc power;
twosamplefreq test = pchi
groupproportions = (0.6 0.75)
nullproportiondiff = 0
power = 0.8
npergroup = .;
run;
```

The POWER Procedure
Pearson Chi-square Test for Proportion Difference

Fixed Scenario Elements	
Distribution	Asymptotic normal
Method	Normal approximation
Null Proportion Difference	0
Group 1 Proportion	0.6
Group 2 Proportion	0.75
Nominal Power	0.8
Number of Sides	2
Alpha	0.05

Computed N per Group	
Actual Power	N per Group
0.800	152

2. Chapter 23 Problem 18:
   Nibbling and LDL Cholesterol. Some studies by J. Mann at the University of Otago, New Zealand, suggest that nibbling—eating small amounts of food frequently rather than eating three larger meals each day—may be a way to lower cholesterol levels. These were small studies involving hospital patients. (Data from J. Mann, "Cholesterol Worries? Nibble More on Less," Science News, 132, March 13 1993, p. 165.) Design a study that would answer the question, "Does consuming nine small snack-sized meals each day (nibbling), rather than consuming three meals, reduce LDL cholesterol levels?" Assume that LDL cholesterol will need no transformation. Assume that the mean of the reduction in LDL cholesterol under the three-meal diet will be small and that its standard deviation will be approximately equal to its mean value. Assume also that the

standard deviation of the reduction under the nibbling diet will be approximately the same as under the three-meal diet. And assume that you wish to be able to detect a mean reduction under the nibbling diet that is 20% greater than the mean reduction under the three-meal diet.

Make sure and describe your experimental design including any treatments, factors, levels etc. Also remember to include a sample size estimate with respect to the given assumptions. Use SAS proc glm power to get your estimated sample size and use proc power to verify. Provide your code and screenshots of the output for each proc. Finally, calculate an estimate of the sample size using the formula on page 709. Note that the textbook says that the 'n' on page 709 represents the sample size per group when it actually represents the total sample size of all the groups.

Treatments: LDL levels with 3 and 9 meals
Factors: 3 meal, 9 meal

```
data result;
do diet = 1 to 2;
input LDL @@;
output;
end;
datalines;
100
80
;
run;
proc print data=result;

proc glmpower data=result;
class diet;
model LDL = diet;
power
stddev=100
ntotal = .
power =.8;
run;
```

The GLMPOWER Procedure

Fixed Scenario Elements	
Dependent Variable	LDL
Source	diet
Error Standard Deviation	100
Nominal Power	0.8
Alpha	0.05
Test Degrees of Freedom	1

Computed N Total		
Error DF	Actual Power	N Total
786	0.801	788

```
proc power;
twosamplemeans
power = .8
ntotal = .
stddev = 100
meandiff = 20
alpha = .05;
run;
```

The POWER Procedure
Two-Sample t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Alpha	0.05
Mean Difference	20
Standard Deviation	100
Nominal Power	0.8
Number of Sides	2
Null Difference	0
Group 1 Weight	1
Group 2 Weight	1

Computed N Total	
Actual Power	N Total
0.801	788

3. Chapter 23Problem 19:
   Seasonal Dyslexia. Researchers at the University of Arkansas theorize that exposure to influenza and other viral diseases in the second trimester of pregnancy may be a factor increasing the odds that the child will be dyslexic. (Dyslexia is a reading disability, defined as a reading score on standard tests falling at least two years behind the expected level, despite a normal IQ.) One piece of evidence that could support this would be a high rate of dyslexia in children born in the summer months, because the second trimester would have fallen in the influenza season. Design an observational study to examine this conclusion. Assume that this will be a case–control study with a random sample of dyslexic boys compared to an equally sized random sample of non-dyslexic boys. Compare the frequency of births in June, July, and August with the frequency of births in December, January, and February. Ignore births in the other months. In the control group, you expect about 50% to be born in the summer months. How large should each sample be if you wish to detect a twofold difference in the odds?

   Factors: Frequency of births and months of births
   Treatments: Frequency of births in June July August December January and February
   Blocks: the months you want to ignore

```
proc power;
twosamplemeans
power = .8
ntotal = .
stddev = 100
meandiff = 50
alpha = .05;
run;
```

The POWER Procedure
Two-Sample t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Alpha	0.05
Mean Difference	50
Standard Deviation	100
Nominal Power	0.8
Number of Sides	2
Null Difference	0
Group 1 Weight	1
Group 2 Weight	1

Computed N Total	
Actual Power	N Total
0.801	128