This week’s prelive assignment is very simple and will just piggy back off of Unit 10.

Below is some sas code utilizing a data set involving two fictional colleges. For our purposes, 4000 students from CRANE and EAGLE college had a friendly competition to see who had the better science knowledge from their respective student bodies.  The 4000 students were randomly selected and then administered a science exam of which they were given a pass/fail score.   Does a student’s overall science knowledge differ between the two schools?

Pre Live Questions to address

* What is the response and what is the explanatory variable?

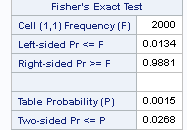
**Response = Exam Pass/Fail**

**Explanatory = College**

* What type of sampling scheme/ study design is this?

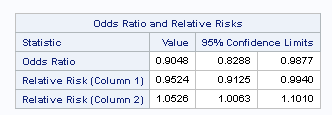
**Observational/Random sample**

* From the SAS output provided by the code below. Find the Fisher’s exact table and make note of the test result? What is the decision and conclusion of the test?



**Fisher’s test calculates the exact hypergeometric probability of observing this particular outcome give the assumed priors (i.e 50/50 in this case). By calculating the row and column totals, the exact probability can be calculated for the outcome. In this case, the test is significant. Given the expected probability of 2000 students each passing/failing at each school, the outcome with eagle’s students passing at a rate of 2100/1900 and Crane passing at 2000/2000 is not statistically likely unless there is a difference between the two groups.**

* Output is provided for CI’s for proportions as well as odds ratios and relative risk. Make note of the odds ratio value and CI. Can you interpret it?



**The odds ratio gives the ratio of the likelihood that a student in column 1 (Crane) will pass the exam to the likelihood that a student in column 2 (Eagle) will pass the exam. In this case – the ratio is less than 1 – hence the odds of a student from Crane passing are less than that of a student from Eagle passing.**

* Take a look at the second set of SAS code. It turns out that the students who participated were sitting in either a math or physics class. The second analysis takes into account this additional factor as a covariate. Compare the ODDS ratio result of the first model with that of the second. What is this saying? Provide some thoughts and explore based on what you got from the videos on the Mantel-Haenszel test. Look at the summary statistics.

proc format;

value ExpFmt 1='Crane C.'

0='Eagle C.’;

value RspFmt 1='Pass'

0='Fail’;

run;

data CH19\_9;

   input Exposure Response Count;

   label Response='Exam Status';

   datalines;

0 0  1900

0 1  2100

1 0  2000

1 1  2000

;

proc sort data=CH19\_9;

   by descending Exposure descending Response;

run;

proc freq data=CH19\_9 order=data;

   format Exposure ExpFmt. Response RspFmt.;

   tables Exposure\*Response / chisq  riskdiff(equal var=null) relrisk;

   exact pchi or fisher;

   weight Count;

   title ' Prospective Study ';

run;

data CH19\_92;

   input Exposure Response Count Subject $;

   label Response='Exam Status';

   datalines;

0 0 1600  Math

0 1 2000  Math

1 0 600  Math

1 1 1200  Math

0 0 300  Physics

0 1 100  Physics

1 0 1400  Physics

1 1 800  Physics

;

proc sort data=CH19\_92;

   by descending Exposure descending Response;

run;

proc freq data=CH19\_92 order=data;

   format Exposure ExpFmt. Response RspFmt.;

   tables Subject\*Exposure\*Response / CMH chisq  riskdiff(equal var=null) relrisk;

   exact pchi or fisher;

   weight Count;

   title ' Prospective Study ';

run;