Name Jerry

Student Number

STA 312 f2022 Quiz 4

1. At-risk youth who wanted help with homework were randomly assigned to either Tutoring Program A, Tutoring Program B, or a wait-list control condition. Here is a table of observed frequencies.

	Graduated High School	Did not graduate
Program A	n_{11}	n_{12}
Program B	n_{21}	n_{22}
Control	n_{31}	n_{32}

(a) (1 point) What kind of design is this? Circle one.

i. Retrospective
ii. Prospective
iii. Introspective
iv. Cross-sectional
v. Multi-sectional

(b) (1 point) Make another 3×2 table. Put model parameters or functions of the parmeters in each cell. These are the parameters *unrestricted by any null hypothesis.*

	Graduated	Did not
A	π_{11}	$1-\pi_{11}$
B	π_{21}	$1-\pi_{21}$
control	π_{31}	$1-\pi_{31}$

(It would be okay to call the three probabilities something else, like π_1, π_2, π_3)

(c) (2 points) Write the unrestricted likelihood. That is, the likelihood is *unrestricted by any null hypothesis.* The answer is a function of the parameters and the n_{ij}. You don't need to show any work. Just think about it and write it down.

$$L = \pi_{11}^{n_{11}}(1-\pi_{11})^{n_{12}}\,\pi_{21}^{n_{21}}(1-\pi_{21})^{n_{22}}\,\pi_{31}(1-\pi_{31})^{n_{32}}$$

(d) (1 point) You want to know whether either tutoring program (or maybe both) had any effect. In terms of the symbols in your table, what is the null hypothesis?

$H_0: \pi_{11} = \pi_{21} = \pi_{31}$

(e) (1 point) How many unknown parameters are there in the restricted model — that is, the model for which the null hypothesis is true? The answer is a number.

1

(f) (2 points) Write the likelihood for the restricted model. Simplify. **Circle your final answer.**

$$\ell_0 = \pi_{11}^{n_{11}}(1-\pi_{11})^{n_{12}}\,\pi_{11}^{n_{21}}(1-\pi_{11})^{n_{22}}\,\pi_{11}^{n_{31}}(1-\pi_{11})^{n_{32}}$$

$$= \pi_{11}^{n_{11}+n_{21}+n_{31}}(1-\pi_{11})^{n_{12}+n_{22}+n_{32}}$$

(No marks off for stopping here.)

$$= \pi_{11}^{n_{+1}}(1-\pi_{11})^{n_{+2}}$$

2. (2 points) In homework question nine (problem 2.16 in the text), you tested for an association between smoking and lung cancer.

(a) *In the space below*, write the likelihood ratio test statistic. The answer is a number on your printout. *On your printout*, circle the number and write "Question 2" beside it.

$G^2 = 19.878$

(b) In plain, non-statistical language, what do you conclude? Give a directional conclusion if possible.

Cancer patients were more likely to be smokers.

Attach your printout to the quiz paper. Make sure your name and student number are on the printout.

Direction is important. It was a retrospective study

```
>
> # Problem 2.16 in the text, my problem 9
>
> cancer = rbind(c(688,650),
+                c( 21, 59))
> rownames(cancer) = c('Smoked','Never')
> colnames(cancer) = c('Cancer','No Cancer')
> addmargins(cancer)
       Cancer No Cancer  Sum
Smoked    688       650 1338
Never      21        59   80
Sum       709       709 1418
> round(prop.table(cancer,margin=2),4)
       Cancer No Cancer
Smoked 0.9704    0.9168
Never  0.0296    0.0832
>
> ctest = chisq.test(cancer,correct=FALSE); ctest

    Pearson's Chi-squared test

data:  cancer
X-squared = 19.129, df = 1, p-value = 1.222e-05

> muhat = ctest$expected
> G2 = 2*sum(cancer*log(cancer/muhat)); G2
[1] 19.87802
> pval = 1-pchisq(G2,df=1); pval
[1] 8.25441e-06
>
> # Estimated P(Cancer given Smoke) if P(Cancer) = 0.01
> 0.9704*0.01/(0.9704*0.01 + 0.9168*.099)
[1] 0.09658874
>
```

Question 2