

## Let Master of Public Health Students Experience Statistical Reasoning

By Qi Zheng\*

*Biostatistics is an integral part of any public health student's curriculum. Most master-level public health students, lacking undergraduate training in calculus-centered mathematics, are often regarded as unprepared for conceptual knowledge, and hence their biostatistics education is dominated by procedural skills. Quoting classroom-tested examples, this article argues for the feasibility of teaching conceptual knowledge using a first principle-guided, computation-driven hands-on approach. Examples are drawn from an online course in categorical data analysis, and it is shown that making the likelihood function comprehensible is key to imparting successfully a majority of the important abstract concepts encountered in an introductory course in categorical data analysis. Data from end-of-semester course surveys were found to support the proposition that master-level public health students are receptive to conceptual knowledge of biostatistics and are able to appreciate the beauty of statistical reasoning.*

**Keywords:** *Biostatistics, Recipe method, Video lecture, Categorical data analysis, Likelihood function*

### Introduction

Biostatistics is an indispensable tool in public health practice and research, and hence is an integral part of any public health student's curriculum. Master-level public health students (commonly known as MPH students) constitute a sizable portion of consumers of today's biostatistics education. However, only recently did this distinct population of students become targets of focused pedagogical research by statistics educators (e.g., Loux et al. 2016). This reflects an increasing awareness among statistics education researchers that students in different disciplines require different statistical curricula and accordingly different statistical pedagogical methods. Among recent examples are statistical education studies aiming at students in medicine (Bland 2004), in pharmacy (McLaughlin and Kang 2017), in nursing (Schwartz 2014, Shillam et al. 2014), in social work (Marson 2007, Mirick and Davis 2017), and in ecology (McGraw and Chandler 2015).

The present research echoes an earlier, relatively richer literature on statistical education for general non-statistics majors (Snee 1993, Garfield 1995, Stuart 1995, Bradstreet 1996, Moore 1997, Sahai 1990). These early research efforts share an urgent call to shift attention from merely teaching statistical methods to earnestly enabling students to experience substantive statistical reasoning. Bradstreet (1996) went so far as to advocate preceding, if not superseding, statistical methods with statistical reasoning in any introductory statistical courses. While most statistics educators embrace this pedagogical philosophy enthusiastically, scant research exists into the nuts and bolts of how to design an introductory biostatistics course

---

\*Professor, School of Public Health, Texas A&M University, USA.

that effectively promotes the teaching of conceptual knowledge, a prerequisite for meaningful statistical reasoning. Even more rarely researched is the teaching of an introductory course in categorical data analysis (CDA), an important sequel to a first introductory biostatistics course (often dubbed Biostats-1) for MPH students. In a meticulously designed survey of the public health research literature, Hayat et al. (2017) found that logistic regression, a leading modeling technique for categorical data, was the most commonly employed data analysis tool in public health research, employed in 38.4% of the studies that were randomly sampled by the survey's authors. When another important categorical data tool, Poisson regression, was also included in this statistic, categorical data analysis techniques figured in 45.8% of the sampled public health studies.

Current statistics education research, particularly research aiming at non-statistics majors, focuses disproportionately on pedagogy assessment. Most existing assessment research took a binary comparison approach, and the investigators seemed to be content with detecting a significant difference between a conventional and a novel pedagogy. This tendency has left a critical gap in current understanding of biostatistics education for non-statistics majors. The following fictitious scenario helps throw this critical gap into sharp relief. In a study of a problem-based approach (Neville 1999) to teaching biostatistics, promising conclusions may be drawn concerning a particular implementation of the problem-based paradigm. But the implementation details are unknown to the reader. The extent to which problem-based learning has been integrated appears to be of no direct concern to the investigators. It is therefore up to the instructors to devise a concrete problem-based approach that can be employed to achieve their teaching goals in a particular setting. Clearly some implementations of the problem-based paradigm will be more effective than others. In other words, the instructors still need to invent the nuts and bolts to allow the problem-based pedagogy to work to their satisfaction, or find the nuts and bolts somewhere else and deftly assemble them according to the instructor's tastes and the students' backgrounds. The following real-world example further underscores the need to fill the foregoing critical gap. The flipped classroom approach is increasingly popular among biostatistics educators (Loux et al. 2016, McGraw and Chandler, 2015, McLaughlin and Kang 2017, Styers et al. 2018). However, a study by Jensen et al. (2015) suggests that the known benefits of the flipped classroom may not be attributable to the flipped classroom per se. Instead, the putative benefits of the flipped classroom may be effected by active learning (Prince 2004, Styers et al. 2018). Such a disagreement about a major teaching strategy highlights the importance of practical implementation details. The effectiveness of the flipped classroom may well depend on a myriad of implementation details, which may in turn depend on the subject matter and the targeted students. The teaching of conceptual knowledge of CDA to MPH students is almost an untrodden realm for pedagogical research, and novel practical details are important in their own right.

The increasing popularity of distance education at public health schools has raised new barriers to teaching conceptual knowledge of biostatistics to MPH students. While the online setting is helpful in imparting application-related procedural knowledge, it does not seem conducive to stimulating students' interest

in mathematical or statistical reasoning (Brimacombe 2014: p. 4). I first taught an introductory CDA course to an on-campus class of 20 MPH students in the fall semester of 2010. The teaching methods were nearly traditional, and the class used the popular textbook by Agresti (2007). I began teaching the same course online in the fall semester of 2015. Faced with the unfamiliar challenges unique to online teaching, I devised a new teaching approach, but relied on the same textbook. I considerably shifted emphasis to conceptual knowledge, and found the outcomes promising. Using roughly the same materials, e.g., videos, homework problems, I repeated the online course in the fall semesters of 2016, 2017, 2018 and 2019. In total, I taught about 75 MPH students using the new approach.

This paper offers detailed methods that proved effective in catalyzing statistical reasoning among MPH students in my online CDA courses. This novel approach requires the instructor to carefully distinguish between abstract mathematical deduction and first principle-guided statistical thinking. This distinction is important because MPH students are normally weaker in the former due to lack of calculus-centered mathematical training, but they are no less strong in the latter than most other students. In the next section, I describe a computing-based method for teaching Fisher's exact test to MPH students, which epitomizes this important distinction. Afterwards, I delineate an elementary way of teaching the concept of the likelihood function, which lies at the heart of my teaching approach. I then demonstrate how I solidify students' understanding of the likelihood function by showing them how to calculate the deviance using the definition. It is followed by a demonstration of my approach to teaching the concept of the likelihood ratio. I subsequently outline students' responses to the new teaching approach and I conclude my paper by remarks that may help the interested reader to adapt or extend the new teaching approach.

### **An Elementary Approach to Fisher's Exact Test**

Fisher's exact test should be in every MPH student's repertoire of investigative tools. But due to seeming mathematical and computational complexity, most introductory biostatistics textbooks eschew detailed discussion of Fisher's exact test. For instance, the popular introductory textbook of Pagano and Gauvreau (2000, p. 347) gives students this succinct advice: "Because the computations involved can be arduous, it is not presented in this text. However, a number of statistical software packages perform Fisher's exact test for  $2 \times 2$  tables in addition to the chi-square test." On further reflection an instructor may realize that the chi-square test requires far more advanced mathematics than does Fisher's exact test. A rudimentary knowledge of the binomial coefficient, a standard topic in any high school algebra curriculum, is the only prerequisite for imparting the ideas underpinning Fisher's exact test to an MPH student. Furthermore, the widespread availability of powerful statistical packages should unshackle students from tedious computations. An MPH student can acquire a solid understanding of Fisher's exact test by studying the example of a gastroenteritis outbreak following a lunch served in the cafeteria of a US high school. The background context is

summarized in the same textbook (Pagano and Gauvreau 2000: p. 360) as follows: "Among a sample of 263 students who bought lunch in the school cafeteria on the day in question, 225 ate prepared sandwiches and 38 did not." Table 1 shows the numbers of cases of gastroenteritis in each group.

**Table 1.** *An Outbreak of Gastroenteritis at a US High School*

Got ill	Ate sandwich		Total
	Yes	No	
Yes	109	4	<b>113</b>
No	116	34	<b>150</b>
<b>Total</b>	<b>225</b>	<b>38</b>	<b>263</b>

The elegant underlying idea is illustrated by a simple urn model. To facilitate students' learning, I first presented a smaller urn model in which the urn contained just 4 black balls and 2 white balls. And 3 balls were to be sampled. The desired probability was the probability of finding 2 black balls among the 3 sampled, which could be explained via the multiplication rule for independent events. Once students got a handle on this simplified problem, they were asked to consider a larger urn containing 225 black balls and 38 white balls. An imaginary goddess of health decides to randomly choose 150 balls to be protected from harm. What is the probability that she ended up with 34 white balls? Students could write down an expression for that probability, but they used the SAS (2012) statement `pdf("hyper",34,263,38,150)` to compute the probability (SAS Institute Inc. 2012):

$$P(34) = \frac{\binom{38}{34} \binom{225}{116}}{\binom{263}{150}} = 3.5 \times 10^{-6} \quad (1)$$

This sets the stage for students to deepen their understanding of p-values. A list of the 39 probabilities  $P(0), P(1), \dots, P(38)$  was produced via SAS. It was straightforward for students to see that the right-side p-value is  $P(34) + \dots + P(38) = 0.00000397..$  But the definition of the p-value had to be reviewed before calculating the left-side p-value. It was emphasized that having 8 white balls is more likely to happen than having 34 white balls. Consequently, the left-side p-value is  $P(0) + \dots + P(8) = 0.00000125$ . Finally, students saw why the (usual) two-sided p-value should be  $5.22 \times 10^{-6}$ .

This example shows the pedagogical power of first principle-driven statistical reasoning when aided by today's computing technology. To further harness the power of this method of teaching CDA to MPH students, I found it fruitful to focus attention on the likelihood function.

### **The Likelihood Function is Key to Comprehension**

Because the concept of the likelihood function spawns most other difficult concepts in CDA, I made students in the first three weeks concentrate on acquiring an ability to write on paper and code in SAS the log-likelihood function for a

simplified logistic regression model, which was adapted from a well-known research problem of beetle deaths caused by a pesticide. In this simplified example six groups of beetles were exposed to a pesticide at different concentrations. Table 2 displays the data.

**Table 2.** *Simplified Data from a Pesticide Experiment*

Concentration	10.8	11.6	12.1	12.6	13.1	13.5
Death	15	24	26	24	29	29
Group size	50	49	50	50	50	49

Students were asked to demonstrate their understanding of the likelihood function by solving several homework problems derived from this example. Before tackling the homework problems, students learned two crucial facts from a lecture about a similar problem. First, from a graph of the logistic function they learned why the problem required the following logistic model:

$$\text{logit} [P(\text{death at concentration } C)] = \beta_0 + \beta_1 \times C \quad (2)$$

Here  $C$  denotes the concentration of the pesticide. To appreciate the logit model (2), students also need to understand the binomial distribution. I discuss the binomial distribution in an introductory biostatistics course, which is a prerequisite for the CDA course. Because the binomial coefficient is a standard topic in any high school algebra curriculum, the binomial distribution may also be smoothly introduced in a CDA course.

Once the role of the likelihood function is explained, students quickly understand why multiplying factors free from the two parameters  $\beta_0$  and  $\beta_1$ , e.g., the binomial coefficients, can be dropped from the likelihood function. For example, students can easily write down the term in the likelihood function contributed by the first group of beetles as

$$\left( \frac{1}{1 + e^{-(\beta_0 + 10.8\beta_1)}} \right)^{15} \left( \frac{1}{1 + e^{(\beta_0 + 10.8\beta_1)}} \right)^{35} \quad (3)$$

Figure 1 shows this expression written by a student. After recalling the monotonic increasing property of the logarithmic function, students were receptive to the fact that maximizing the log-likelihood function is the same as maximizing the likelihood function itself. Students were further reminded of the following properties of the logarithmic function they learned in high school.

$$\begin{aligned} \log(a \times b) &= \log(a) + \log(b) \\ \log\left(\frac{a}{b}\right) &= \log(a) - \log(b) \\ \log(e^x) &= x \\ \log(a^b) &= b \times \log(a) \\ \log\left(\frac{1}{a}\right) &= -\log(a) \text{ with } \log(1) = 0 \\ \frac{e^a}{e^b} &= e^{a-b} \text{ with } e^0 = 1 \end{aligned} \quad (4)$$

**Figure 1.** A Student's Handwritten Work Deriving the Expression in Equation (3)

→ a) likelihood function from the first group (death = 15, size = 50, c = 10.8)

$$L(p) = p^{15} \times (1-p)^{35}$$

given  $\begin{cases} \eta = \text{logit}[P(\text{death at conc. } c)] = \beta_0 + \beta_1 c \\ c = 10.8 \end{cases}$

$$\Rightarrow L(p) = \left( \frac{1}{1+e^{-\eta}} \right)^{15} \times \left( 1 - \frac{1}{1+e^{-\eta}} \right)^{35}$$

$$= \left( \frac{1}{1+e^{-(\beta_0+10.8\beta_1)}} \right)^{15} \times \left( \frac{1}{1+e^{(\beta_0+10.8\beta_1)}} \right)^{35}$$

Note that the student first wrote the linear predictor  $\eta$  before recasting it as a component probability of the likelihood function.

Applying these properties, students were able to transform a likelihood function to the corresponding log-likelihood function and then code it using the SAS fcmp procedure. Figure 2 shows a student's SAS rendition of the log-likelihood function for the beetle example. Acquiring an ability to code the log-likelihood function for the beetle example was a milestone for MPH students learning CDA, because they began to relish concretizing abstract concepts by an accessible hands-on approach, thereby bolstering confidence in their ability to master theoretical basics of CDA. Finding values of  $\beta_0$  and  $\beta_1$  that maximize the likelihood function (they are called maximum likelihood estimates, or MLEs) is beyond the scope of an introductory CDA course, and students were forthrightly informed that a statistical software package like SAS uses sophisticated iterative algorithms to achieve that feat. This only further piqued students' curiosity. Students obtained MLEs of  $\beta_0$  and  $\beta_1$  with SAS proc logistic, and then used them to verify the maximum value of the log-likelihood function by comparing their result with SAS output. (See Figure 3 for a student's actual work.) Through this process students developed an intuitive sense about the meaning of an MLE, the existence and uniqueness of MLEs. This exercise paved the way for some more challenging concepts.

### A Firsthand View of the Deviance

As the concept of the likelihood is a key pillar of conceptual knowledge of CDA methods, students should be given ample opportunities to view it from different angles. When all predictors are categorical in a model, the likelihood function can appear utterly different from the one student saw in the previous example. The following computational exercises, based on the well-known adolescent sexual behavior study (Agresti 2007: 113), aims to solidify students' understanding of the likelihood function, which also opens doors for students to

grasp another elusive concept, the deviance. Table 3 reproduces the data for ease of reference.

**Table 3.** Data on Teenagers and Sex

Race	Gender	Intercourse	
		Yes	No
White	Male	43	134
	Female	26	149
Black	Male	29	23
	Female	22	36

**Figure 2.** A Student's SAS Rendition of the Log-Likelihood Function for the Beetle Example. A Pair of Testing Values were used to Test whether the SAS Code was Executable

```

/* PHEB 609 HW2 */
proc fcmp outlib=work.hw2.a;
function likel(a,b);
l1=-15*log(1+exp(-(a+10.8*b)))-35*log(1+exp(a+10.8*b));
l2=-24*log(1+exp(-(a+11.6*b)))-25*log(1+exp(a+11.6*b));
l3=-26*log(1+exp(-(a+12.1*b)))-24*log(1+exp(a+12.1*b));
l4=-24*log(1+exp(-(a+12.6*b)))-26*log(1+exp(a+12.6*b));
l5=-29*log(1+exp(-(a+13.1*b)))-21*log(1+exp(a+13.1*b));
l6=-29*log(1+exp(-(a+13.5*b)))-20*log(1+exp(a+13.5*b));
return (l1+l2+l3+l4+l5+l6);
endsub;
run;
options cmplib=work.hw2;

data beetle;
a=2; b=3; lik=likel(a,b); output;
run;

proc print data=beetle;
run;

```

The SAS System

Obs	a	b	lik
1	2	3	-5795.3

After studying a similar example from a video lecture, students understand the rationale for adopting the logit model

$$\text{logit}(P(\text{intercourse})) = \beta_0 + \beta_1 \times \text{Race} + \beta_2 \times \text{Gender} \quad (5)$$

Here both Race and Gender are dummy variables. Race takes the value zero when the subject is black, and unity when the subject is white. Similarly, Gender equals zero when the subject is female, and unity when the subject is male. Although this model bears close resemblance to the previous logit model (2), students took a fresh look at the likelihood out of curiosity about the two dummy variables. For example, in the while-female scenario the linear component is  $\beta_0 + \beta_1$ . Therefore, the likelihood function should include the expression:

$$\left( \frac{1}{1 + \exp(-(\beta_0 + \beta_1))} \right)^{26} \times \left( \frac{1}{1 + \exp(\beta_0 + \beta_1)} \right)^{149} \quad (6)$$

which should be in the following form when it appears in the corresponding log-likelihood function:

$$-26 \log(1 + \exp(-(\beta_0 + \beta_1))) - 149 \log(1 + \exp(\beta_0 + \beta_1)) \quad (7)$$

**Figure 3.** A Student Verifies the Maximized Value of a Log-Likelihood Function that SAS Produces

Model Fit Statistics			
Criterion	Intercept Only	Intercept and Covariates	
		Log Likelihood	Full Log Likelihood
AIC	415.062	408.038	32.245
SC	418.759	415.432	39.640
-2 Log L	413.062	404.038	28.245

$-2\log(L)=404.038$ , hence  $\log(L)= -202.019$

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.8097	1.6210	8.8037	0.0030
conc	1	0.3893	0.1315	8.7618	0.0031

**Confirmed:**  
`data beetle;`  
`a=-4.81; b=0.39; lik=likel(a,b); output;`  
`run;`  
`proc print data=beetle;`  
`run;`

**(-4.81, 0.39) = -202.021**

Obs	a	b	lik
1	-4.81	0.39	-202.021

Note that the student uses a pair of parentheses to signify the operation of division by 2, which is non-standard.

This algebraic exercise further familiarized students with the likelihood principle. Figure 4 shows how one student coded the log-likelihood function and how she used it to confirm the value that the log-likelihood function should assume when the three arguments of the function coincide with the maximum likelihood estimates of the three model parameters.

At this stage the concept of the saturated model can be introduced. In the context of a simple model like (5), students would find the concept quite approachable. For example, in the white-female scenario, the probability of intercourse should be  $26/(26+149)$  under the saturated model, for the model to be "closest" to data. Thus, the log-likelihood function should contain the quantity  $26 \log(26/175) + 149 \log(149/175)$ . The maximized log-likelihood value for the saturated model, denoted by  $l_{sat}$ , is easier for students to calculate than  $l_{model}$ , the maximized log-likelihood value for the actual model – the logit model (5) in the present context. Figure 5 shows a student’s calculation of  $l_{sat}$ . The definition of the deviance:

$$D = -2(l_{model} - l_{sat}) \quad (8)$$

can now be introduced using a simple example like this one. A student’s work calculating the deviance of the model (5) is shown in Figure 6.

**Figure 4.** A Student's Work to Evaluate the Log-Likelihood Function at the Maximum Likelihood Estimates of the Three Model Parameters

```
proc fcmp outlib=work.pheb609.hw4;
function mylike(b0,b1,b2);
lik1=-43*log(1+exp(-(b0+b1+b2)));
      -134*log(1+exp(b0+b1+b2));
lik2=-29*log(1+exp(-(b0+b2)))-23*log(1+exp(b0+b2));
lik3=-26*log(1+exp(-(b0+b1)))-149*log(1+exp(b0+b1));
lik4=-22*log(1+exp(-b0))-36*log(1+exp(b0));
return(lik1+lik2+lik3+lik4);
endsub;
run;

options cmplib=work.pheb609;

data Behavior;
b0=-0.4555; b1=-1.3135; b2=0.6478; like=mylike(b0,b1,b2); output;
run;

proc print data=Behavior;
run;
```

The SAS System				
Obs	b0	b1	b2	like
1	-0.4555	-1.3135	0.6478	-245.897

**Figure 5.** A Student's Work Showing the Calculation of the Maximized Log-Likelihood Value for the Saturated Model,  $l_{sat}$ , in the Adolescent Sexual Behavior Study

```
data saturated;
l=43*log(43/177)+134*log(134/177)
  +26*log(26/175)+149*log(149/175)
  +29*log(29/52)+23*log(23/52)
  +22*log(22/58)+36*log(36/58); output;
run;

proc print;
title 'log-likelihood for the saturated model.';
run;
```

log-likelihood for the saturated model.	
Obs	l
1	-245.868

**Figure 6.** A Student's Work Showing the Calculation of the Deviance of the Logit Model (5)

$$j) \text{ Deviance} = -2(l_{\text{model}} - l_{\text{sat}}) = -2(-245.897 - (-245.868)) = 0.058$$

### Anatomizing the Likelihood Ratio Test

The likelihood ratio (LR) test is another bewildering concept for MPH students. However, once they digest the concept of the likelihood function, they

can explore and demystify the LR test. In my CDA classes, students also used the beetle example to see how the LR principle works in testing the null hypothesis  $H_0: \beta_0 = 0$  against the alternative hypothesis  $H_1: \beta_0 \neq 0$ . Due to their newly acquired appreciation of the likelihood function, students now readily saw the precise meaning of the null hypothesis. By setting  $\beta_1 = 0$  in the expression (3), they obtained the expression:

$$\left(\frac{1}{1 + e^{-\beta_0}}\right)^{15} \left(\frac{1}{1 + e^{\beta_0}}\right)^{35} \tag{9}$$

as the contribution of the first group of beetles to the likelihood function under  $H_0$ . Using the elementary properties of the logarithmic function in equation (4), students were able to show that the log-likelihood function under  $H_0$  is of the form

$$-147 \log(1 + e^{-\beta_0}) - 151 \log(1 + e^{\beta_0}). \tag{10}$$

Figure 7 displays a student’s derivation of the above likelihood function. At this stage, a homework problem asked students to plot the expression (10) as a function of  $\beta_0$ , and to use SAS proc logistic to find  $\widetilde{\beta_0}$  (by specifying an intercept-only model) that maximizes the expression (10). This exercise gave students a chance to visualize a log-likelihood function, enhancing their understanding of the concept of maximum likelihood estimates. Students were now in a position to grasp the rationale for forming the ratio

$$\frac{L_0}{L_1} = \frac{\text{maximized value of likelihood without } \beta_1}{\text{maximized value of likelihood with } \beta_1} \tag{11}$$

**Figure 7.** A Student’s Derivation of the Log-Likelihood Function Given in Equation (10)

$$\begin{array}{r} \left(\frac{1}{1 + e^{-\beta_0}}\right)^{15} * \left(\frac{1}{1 + e^{\beta_0}}\right)^{35} \\ \left(\frac{1}{1 + e^{-\beta_0}}\right)^{24} * \left(\frac{1}{1 + e^{\beta_0}}\right)^{25} \\ \left(\frac{1}{1 + e^{-\beta_0}}\right)^{26} * \left(\frac{1}{1 + e^{\beta_0}}\right)^{24} \\ \left(\frac{1}{1 + e^{-\beta_0}}\right)^{24} * \left(\frac{1}{1 + e^{\beta_0}}\right)^{26} \\ \left(\frac{1}{1 + e^{-\beta_0}}\right)^{29} * \left(\frac{1}{1 + e^{\beta_0}}\right)^{21} \\ \left(\frac{1}{1 + e^{-\beta_0}}\right)^{29} * \left(\frac{1}{1 + e^{\beta_0}}\right)^{20} \\ \hline \left(\frac{1}{1 + e^{-\beta_0}}\right)^{147} * \left(\frac{1}{1 + e^{\beta_0}}\right)^{151} \\ = -147 * \log(1 + e^{-\beta_0}) - 151 * \log(1 + e^{\beta_0}) \end{array}$$

Note that the student first derives the likelihood function and then translates it to the log-likelihood function. An equality sign is loosely employed to signify the log-transform.

Because now students saw this ratio as a means of comparing two probabilities, they appreciated the idea that a close to unity LR statistic suggests non-significance of the parameter  $\beta_1$ . Students only needed to accept one theoretical result to see the whole picture: the quantity  $-2 \times \log(L_0/L_1)$  approximately obeys a chi-squared distribution with one degree of freedom. Using the machinery they had already built, e.g., SAS code for the restricted and unrestricted log-likelihood functions, students computed  $L_0$  and  $L_1$  easily. They then obtained the LR test statistics and compared it with SAS output. At this point the equation (11) served as a springboard for introducing students to the concept of the deviance. The rationale for the deviance is similar to the LR principle with which the students were already familiar.

The above learning process paved the way for students to understand the concept of a likelihood ratio confidence interval (CI). In a homework assignment based on the data given in Table 3, students were asked to determine whether a particular parameter value  $\beta_2^* = 2.345$  should be inside or outside an LR-based 95% CI for  $\beta_2$ . This otherwise difficult goal was made reachable to students by a brief discussion about the following  $\Lambda$  quantity:

$$\Lambda(\beta_2^*) = \frac{\max \text{ over } \beta_0, \beta_1 L(\beta_0, \beta_1, \beta_2^*)}{\max \text{ over } \beta_0, \beta_1, \beta_2 L(\beta_0, \beta_1, \beta_2)} \quad (12)$$

Due to their hands-on experience with likelihood functions, students at this point immediately understood the meaning of the quantity in the denominator, despite the strange mathematical notation that were unfamiliar to most students prior to their enrollment in the course. The quantity in the numerator was also easily explainable conceptually, but how to make it tangible? My solution was to let students use the offset option available in SAS proc logistic (Figure 8). This may seem an excessive burden on MPH students at first glance, but it is within their grasp. Students already grasped the concept of an offset in a preceding lecture about Poisson regression. The unconventional use of the offset only enhances students' understanding of the concept. Because the rationale for equation (12) is similar to that for equation (11), students readily saw that if  $-2\log(\Lambda(\beta_2^*)) > \chi_{1,0.05}^2 = 3.8415$ , then  $\beta_2^*$  should be outside the 95% CI for  $\beta_2$ . In this exercise students found that:

$$-2 \log \Lambda(2.345) = 541.996 - 491.795 = 50.201. \quad (13)$$

So, students concluded that the point  $\beta_2^* = 2.345$  should be outside a 95% likelihood ratio CI for  $\beta_2$ , which they diligently checked against SAS output. A student succinctly summarized this reasoning process in Figure 9.

**Figure 8.** A Student's SAS Work to Extract from SAS Output the Constrained Likelihood Value that Corresponds to the Numerator in the Likelihood Ratio Defined by Equation (12)

```

data ASB;
input race gender Intercourse NoIntercourse ;
tot= Intercourse + NoIntercourse;

myoff=2.345*gender;

datalines;
1 1 43 134
1 0 26 149
0 1 29 23
0 0 22 36
;
run;

proc logistic data=ASB;
model Intercourse/tot= race gender;
run;

proc logistic data=ASB;
model Intercourse/tot= race/offset=myoff;
run;

```

Model Fit Statistics			
Criterion	Intercept Only	Intercept and Covariates	
		Log Likelihood	Full Log Likelihood
AIC	584.272	545.996	73.387
SC	588.408	554.267	81.658
-2 Log L	582.272	541.996	69.387

**Figure 9.** A Student's Demonstration of Why a Particular Point Should Be Outside the 95% Confidence Interval for the Parameter  $\beta_1$

$$95\% \text{ CI} \rightarrow \chi_{1,0.05}^2 = 3.842$$

The offset is

```
myoff=2.345*male;
```

$$-2\log\Lambda = 541.996 - 491.795 = 50.201 > 3.842$$

$\rightarrow -2\log\Lambda$  is larger than the critical value, therefore the point is outside of the CI

### Evaluation by Students

The novel approach has been well received by a great majority of students in all five CDA classes that I taught in the past five years. Students particularly savored the triumphant moment when they understood how a once-unfathomable statistical package like SAS actually solved a particular problem. In contrast to the 2010 class in which I adopted a more or less traditional approach, the new

approach generated more enthusiasm and better outcomes as measured by exam results and student evaluations. Because I myself graded all homework assignments and examinations, students' test performance may not be an objective yardstick to evaluate the effectiveness of the new approach. However, the school's routine end-of-semester course surveys provide somewhat objective clues to the new approach's effectiveness. In the fall of 2017, over 400 responses were collected on each item in the end-of-semester course surveys by the Department of Epidemiology and Biostatistics. The average scores based on such a large sample provide a benchmark against which the new approach can be assessed.

A course survey comprises about 20 assertions that are not targeted to a specific course. The students respond to each assertion by choosing an integer score between 1 and 5. A score of "1" indicates that the student strongly disagrees with the statement, while a score of "5" signifies the student's strong agreement with the assertion. Responses to five of the assertions recorded for the five CDA classes are summarized in Table 4.

Response rates varied from 45% to 82%. These five assertions are selected because they may shed light on students' attitude toward the new approach by which I taught the five CDA classes. For example, responses to the second assertion in Table 4, "On the whole, the information learned in this course was valuable to me," could register students' opinions about learning conceptual knowledge and statistical reasoning. The third assertion, "The instructor stimulated my interest in the subject" measures students' enthusiasm about the course content, which depended on the ways the course content was delivered. The fourth assertion, "The instructor gave clear presentations," measures how easily students digest the course content, which in the last five surveys would also indicate students' ability to understand abstract concepts and experience meaningful statistical reasoning.

**Table 4.** Summary of Student Evaluation

Year	PHEB17f	2015	2016	2017	2018	2019
<i>Class size</i>	~400	13	7	11	20	24
<i>A1</i>	4.28	3.88 (0.62)	5.00 (0.07)	4.56 (0.22)	4.22 (0.72)	4.24 (0.53)
<i>A2</i>	4.41	3.71 (0.55)	5.00 (0.07)	4.56 (0.14)	4.44 (0.22)	4.41 (0.28)
<i>A3</i>	4.11	3.88 (1.00)	5.00 (0.07)	4.56 (0.22)	4.33 (0.62)	4.12 (0.74)
<i>A4</i>	4.14	3.50 (0.44)	4.75 (0.19)	4.50 (0.71)	4.22 (0.81)	4.12 (0.74)
<i>A5</i>	4.34	4.25 (0.83)	5.00 (0.07)	4.63 (0.10)	4.67 (0.05)	4.41 (0.23)

PHEB17f: Average scores based on over 400 responses collected by the Department of Epidemiology and Biostatistics in the fall of 2017.

A1: Assertion 1, "On the whole, this was a good course."

A2: Assertion 2, "On the whole, the information learned in this course was valuable to me."

A3: Assertion 3, "The instructor stimulated my interest in the subject."

A4: Assertion 4, "The instructor gave clear presentations."

A5: Assertion 5, "On the whole this is a good instructor."

The parenthesized numbers are p-values based on the one-sample Wilcoxon test comparing scores of a given year with the corresponding departmental average scores based on over 400 responses collected in the fall of 2017.

As the average scores in Table 4 suggest, the introduction of the new teaching approach was in the main a success. Most MPH students in the five CDA classes showed a favorable attitude toward conceptual knowledge of biostatistics, and they evinced a remarkable ability to digest abstract concepts and appreciate the beauty of statistical reasoning. To help the reader assess the evidence, the one-sample Wilcoxon test was applied to the student survey responses with the resulting p-values exhibited in Table 4. This case study involves about 75 students, and some caveats should be noted. First, the five classes were taught sequentially by a single instructor, and the instructor's accumulating experience may interfere with possible differential effects between the two approaches. But the four online classes used the same video files and similar homework and examination problems, the effect of such interference may be minimal. Second, as in any survey, there was a considerable number of non-respondents in each class. A student's reason for not participating in a course survey has not been accounted for. Lastly, a PhD candidate registered for one of my CDA classes. Because the surveys were anonymous, this possible respondent cannot be excluded.

## Discussion and Conclusions

Biostatistics is inherently laden with abstract concepts that are hurdles to most MPH students. While today's biostatistics instructors for MPH students are increasingly aware of the importance of conceptual knowledge, their efforts to weave conceptual knowledge and statistical reasoning into an MPH biostatistics curriculum are often hampered by a lack of easily accessible case-studies. This paper offers detailed, classroom-tested examples that may act as a springboard for improving MPH biostatistics education across public health schools. In particular, these examples may help inexperienced instructors resist the temptation to resort to what Lee (1999: p. 194) dubbed a recipe approach, when MPH students' lack of an adequate mathematical background looms like an insurmountable obstacle.

The need for conceptual statistical knowledge in MPH students' education has not been adequately debated at public health schools, partly because MPH students were rarely able to articulate such a need. For example, in more than 20 end-of-semester course surveys asking for ways to improve biostatistics teaching, no student expressed an interest in learning conceptual knowledge. But when I taught conceptual statistical knowledge by this new approach, most students were quick to recognize the importance of learning it. Procedural skills taught without

imparting adequate underlying conceptual knowledge can be harmful to the unwary student. Moreover, teaching procedural skills as if they were recipes stifles easily a student's interest in learning.

The novel hands-on approach outlined here can effectively pique MPH students' interest in learning statistical reasoning in a CDA course. This approach is anchored in two observations. First, from a would-be public health researcher's or practitioner's point of view, most of the abstract concepts are best understood by appealing to first principles. To them, rigorous mathematical formulations of these concepts are nonessential, if not superfluous. However, MPH students find learning the essence of these concepts intellectually exhilarating and conducive to fruitful statistical applications. Luckily, these concepts are often explicable using high school mathematics. Second, these concepts can be rendered tangible by simple computer experiments, and teaching of the necessary computer coding skill can be seamlessly integrated into a CDA course. The likelihood function is central to successfully imparting conceptual statistical knowledge to MPH students in a CDA class. Despite lacking calculus-centric mathematical training as an undergraduate, most MPH students are receptive to fundamental statistical concepts derived from the likelihood function, once these concepts are rendered tangible by concrete numerical examples. Verifying results of a major statistical package through writing simple computer code is especially conducive to stimulating an MPH student's interest in learning conceptual statistical knowledge. The fine distinction between general cognitive ability and high-order mathematical sophistication is often underappreciated by biostatistics educators teaching MPH students.

The computational exercises quoted and others not presented here lie at the heart of the proposed approach to teaching CDA to MPH students. These exercises help students concretize abstract statistical concepts and basic principles, compelling them to participate in a hands-on process of constructing their own statistical knowledge (Garfield 1995). Computing is woven into these exercises, but it is meant for learning statistical reasoning, not for implementing statistical recipes. Such a hands-on approach bolsters students' confidence by making conceptual knowledge accessible to them, which shape not only how enthusiastically MPH students learn biostatistics, but also how deftly they apply biostatistics in their public health practice and research after graduation. Finally, note that the choice of the SAS software was largely due to its popularity among epidemiologists, and the instructor may choose other software platforms like the R software to serve the same purpose (R Core Team 2017). It is hoped that this unique approach can be adapted to teach other specialized branches of biostatistics, e.g., the analysis of longitudinal data.

## References

- Agresti A (2007) *An introduction to categorical data analysis*. 2<sup>nd</sup> ed. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Bland JM (2004) Teaching statistics to medical students using problem-based learning: the Australian experience. *BMC Medical Education* 4(31).

- Bradstreet TE (1996) Teaching introductory statistics courses so that non-statisticians experience statistical thinking. *The American Statistician* 50(1): 69-78.
- Brimacombe MB (2014) Biostatistical and medical statistics graduate education. *BMC Medical Education* 14(18).
- Garfield J (1995) How students learn statistics. *International Statistical Review* 63(1): 25-34.
- Hayat MJ, Powell A, Johnson T, Cadwell BL (2017) Statistical methods used in the public health literature and implications for training of public health professionals. *PLoS ONE* 12(6): e0179032.
- Jensen JL, Kummer TA, Godoy PD (2015) Improvement from a flipped classroom may simply be the fruits of active learning. *CBE Life Sciences Education* 14(1): 5.
- Lee C (1999) Computer-assisted approach for teaching statistical concepts. *Computers in the Schools* 16(1): 193-208.
- Loux TM, Varner SE, VanNatta M (2016) Flipping an introductory biostatistics course: a case study of student attitudes and confidence. *Journal of Statistics Education* 24(1): 1-7.
- Marson SM (2007) Three empirical strategies for teaching statistics. *Journal of Teaching in Social Work* 27(3/4): 199-212.
- McGraw JB, Chandler JL (2015) Flipping the biostatistics classroom, with a twist. *Bulletin of the Ecological Society of America* 96(2): 375-383.
- McLaughlin JE, Kang I (2017) A flipped classroom model for a biostatistics short course. *Statistics Education Research Journal* 16(2): 441-453.
- Mirick RG, Davis A (2017) Making meaning of MSW students' statistical ability: the role of self-efficacy and knowledge-based assessment. *Journal of Social Work Education* 53(2): 212-221.
- Moore DS (1997) New pedagogy and new content: the case of statistics. *International Statistical Review* 65(2): 123-165.
- Neville AJ (1999) The problem-based learning tutor: teacher? Facilitator, evaluator? *Medical Teacher* 21(4): 393-401.
- Pagano M, Gauvreau K (2000) *Principles of biostatistics*. 2<sup>nd</sup> ed. Duxbury: Chapman and Hall/CRC.
- Prince M (2004) Does active learning work? A review of the research. *Journal of Engineering Education* 93(3): 223-231.
- R Core Team (2017) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Sahai H (1990) Some problems of teaching an introductory biostatistics course to graduate students in health sciences: coping with the diversity of student aptitudes, interests and objectives. *The Statistician* 39(4): 341-347.
- SAS Institute Inc. (2012) *SAS/STAT Software*. Version 9.4. Cary, NC.
- Schwartz TA (2014) Flipping the statistics classroom in nursing education. *Journal of Nursing Education* 53(4): 199-206.
- Shillam CR, Ho G, Commoodore-Mensah Y (2014) Online biostatistics: evidence-based curriculum for master's nursing education. *Journal of Nursing Education* 53(4): 229-232.
- Snee RD (1993) What's missing in statistical education? *The American Statistician* 47(2): 149-154.
- Stuart M (1995) Changing the teaching of statistics. *The Statistician* 44(1): 45-54.
- Styers ML, Van Zandt PA, Hayden KL (2018) Active learning in flipped life science courses promotes development of critical thinking skills. *CBE Life Sciences Education* 17(3): 39.