Characterizing Statistics Understanding And Attitudes In Graduate Students And Postdocs In The Life Sciences

Abha Ahuja[1,2], Kyle Dillon[3], Melanie Stefan[1], Yan Liu[1], Johanna Gutlerner[1] & David Van Vactor[1,2]

[1]Curriculum Fellows Program,[2]Department of Cell Biology, Harvard Medical school; [3]Harvard Psychology

Introduction

A basic understanding of statistics is essential in biological research, to analyze data and to understand and interpret results reported in the literature. However, many research papers are published with clear misuse or misinterpretation of statistical analyses[1]. Therefore, it is imperative to train life scientists in the proper use and interpretation of statistical methodology. Statistical misconceptions have been described in detail in the education research literature and the nature of these misconceptions is well documented among varied populations[2]. However, a systematic study of statistics understanding among life science post-graduates has not been performed. Furthermore, little research has been conducted on the relationship between statistics self-efficacy, statistics anxiety, and performance for these students. Gaining an understanding of the specific types of misconceptions in and attitudes towards statistics in our students will enable us to develop curriculum that forces students to confront their misconceptions directly and supports a corrected understanding of statistical concepts. Here we present the results of a pilot survey administered to graduate students and postdocs enrolled in a short course in introductory biostatistics at Harvard Medical School.

Methods

Survey Instrument

The CAOS test (Comprehensive Assessment of Outcomes In Statistics)[3] is a confirmed valid and reliable assessment consisting of multiple choice items that cover concepts taught in an introductory statistics course. In addition we designed 18 questions to assess student background, motivation and attitudes towards statistics.

Survey Administration

Students were required to complete the survey in order to enroll in a short course on data analysis for biologists. Students completed a web-based survey in-class.

Results

- Mean (SD) completion time = 25.7 (4.33) minutes
- n = 39 ; Male =19; Female = 20
- Median (IQR) Age = 26(3)
- Internal consistency (Cronbach's α) = 0.68

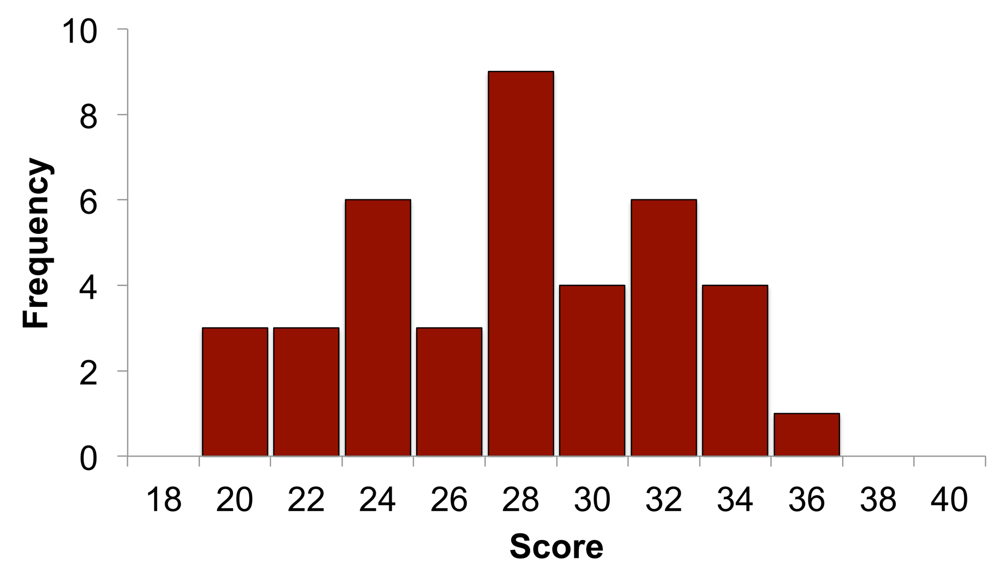

Figure 1. Frequency distribution of test scores (Maximum score = 40) (n=39).
Mean Male Score (SD) = 27.5 (3.76)
Mean Females Score (SD) = 27.0 (4.95)

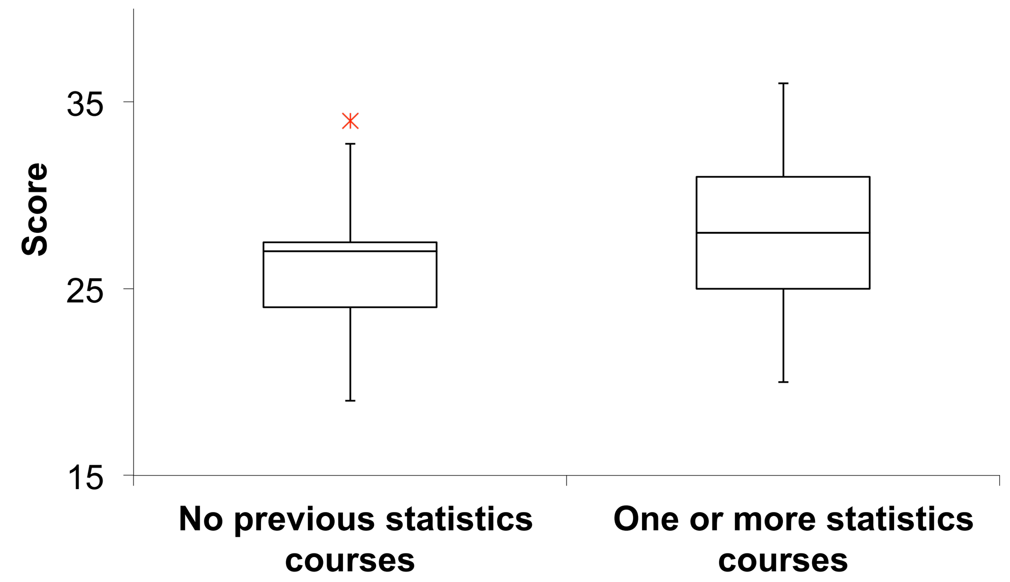

Figure 2. Median (IQR) of score of students with no previous statistics courses (n=16), and one or more courses (n =23). Red cross indicates outlier.

A

#	1	2	3	4	CAOS Measured Outcome
16	0.59	0.47	44	74	Understanding that statistics from small samples vary more than statistics from large samples
17	0.76	0.01	69	77	Understanding of expected patterns in sampling variability.
18	0.76	0.25	88	83	Understanding of the meaning of variability in the context of repeated measurements and in a context where small variability is desired.
32	0.06	0.12	8	10	Understanding of how sampling error is used to make an informal inference about a sample mean.
34	0.94	0.25	88	82	Understanding of the law of large numbers for a large sample by selecting an appropriate sample from a population given the sample size.
35	0.59	0.70	33	71	Understanding of how to select an appropriate sampling distribution for a particular population and sample size.

B

#	1	2	3	4	CAOS Measured Outcome
28	0.59	0.50	75	65	Ability to detect a misinterpretation of a confidence level (the percentage of sample data between confidence limits)
29	0.76	-0.16	50	86	Ability to detect a misinterpretation of a confidence level (percentage of population data values between confidence limits)
30	0.41	-0.24	55	50	Ability to detect a misinterpretation of a confidence level (percentage of all possible sample means between confidence limits)
31	0.82	-0.16	73	81	Ability to correctly interpret a confidence interval

1	Item Difficulty
2	Item Discrimination
3	Percent correct in students with no statistics courses
4	Percent correct in students with one or more statistics courses

Table 1. Breakdown of scores in (A) Sampling Variability and (B) Confidence Intervals categories

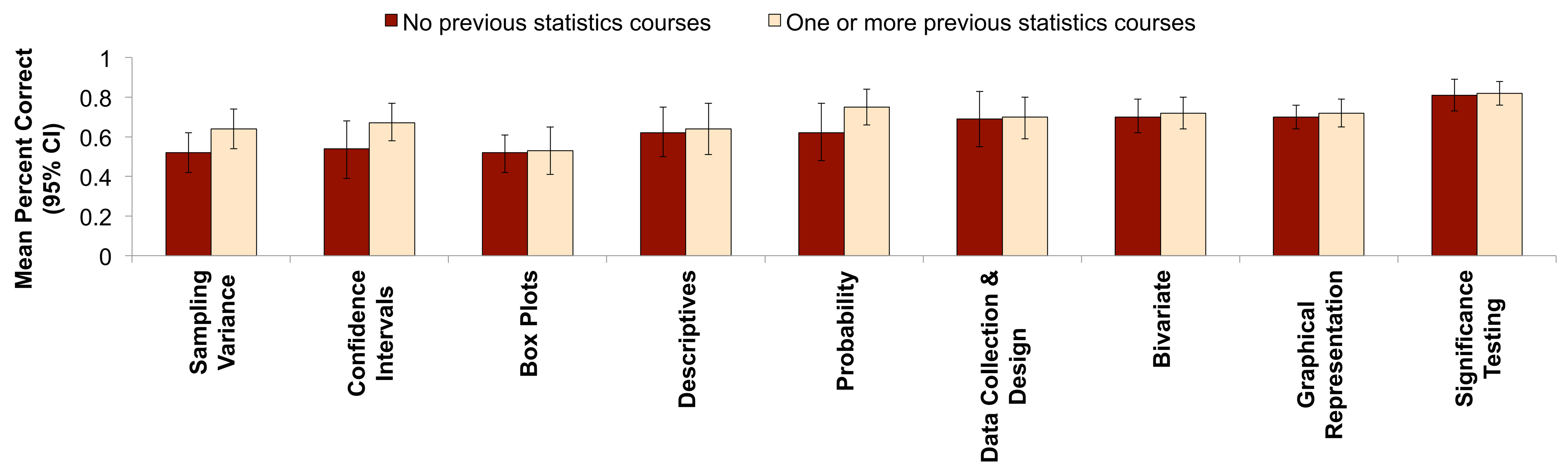

Figure 3. Correct responses by conceptual category & previous statistics course work

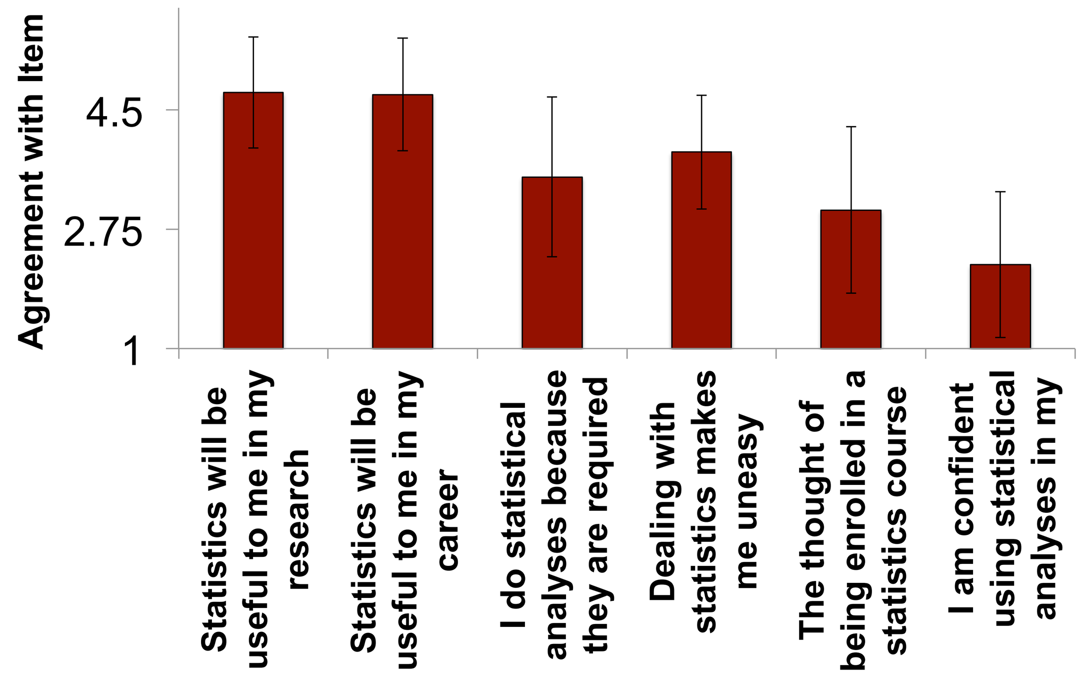

Figure 4. Mean (SD) agreement with items to assess statistics attitudes (1= Strongly disagree, 5 = Strongly agree)

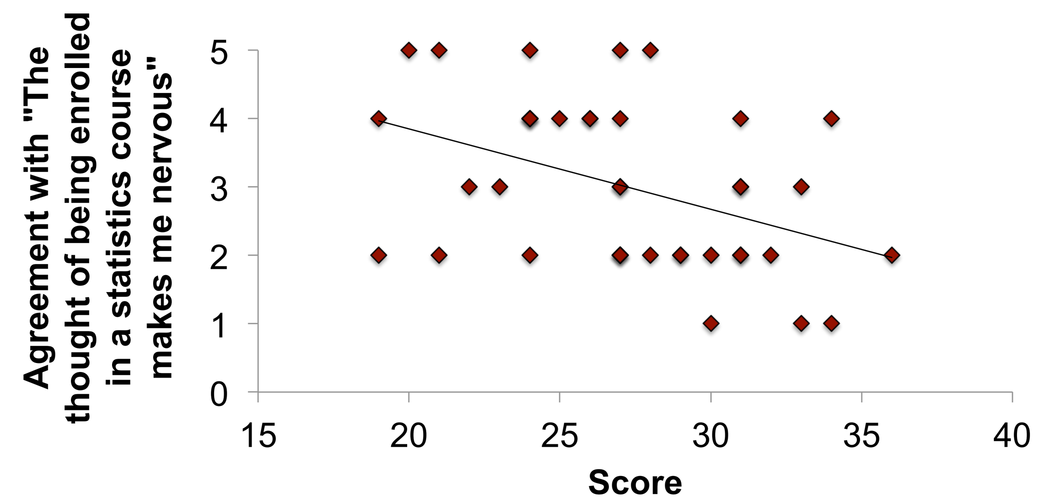

Figure 5. Negative correlation between actual performance and self-reported nervousness

Conclusions

Students display a good understanding of the concepts around significance testing and graphical representation. However, misconceptions regarding sampling distributions and confidence intervals were seen. Student are uneasy dealing with statistics and not confident in their ability to use statistical techniques in research. Future directions include refining the survey, and administering to a larger student population.

References & Acknowledgement

1) Vaux DL. Nature 2012; 492: 180-181.
2) Sotos AEC, Vanhoof S, Van den Noortgate W, Onghena P. Educ Res Rev 2007; 2: 98-113
3) Delmas R, Garfield J, Ooms S, Chance, B 2007. Statistics Educ Res J 6: 28-58.

4) We thank the Harvard Initiative for Learning and Teaching (HILT) initiative for funding.