

SPECIAL TOPIC

# Assessing the statistical training in animal science graduate programs in the United States: survey on statistical training

Nick V. L. Serão,<sup>†,1</sup> Amy L. Petry,<sup>†,‡</sup> Leticia P. Sanglard,<sup>†</sup>  
Mariana C. Rossoni-Serão,<sup>†</sup> and Jennifer M. Bundy<sup>†</sup>

<sup>†</sup>Department of Animal Science, Iowa State University, Ames, IA 50011, USA, <sup>‡</sup>Department of Animal and Food Sciences, Texas Tech University, Lubbock, TX 79415, USA

<sup>1</sup>Corresponding author: [serao@iastate.edu](mailto:serao@iastate.edu)

ORCID number: [0000-0002-6758-208X](https://orcid.org/0000-0002-6758-208X) (N. V. L. Serão).

## Abstract

Statistical analysis of data and understanding of experimental design are critical skills needed by animal science graduate students (ASGS). These skills are even more valuable with the increased development of *high-throughput* technologies. The objective of this study was to evaluate the perceived statistical training of U.S. ASGS. A survey with 38 questions was shared across U.S. universities, and 416 eligible ASGS from 43 universities participated in this study. The survey included questions on the demographics and overall training, graduate education on statistics, and self-assessment on statistics and career path of ASGS. Several analyses were performed: relationship between perceived received education (PRE; i.e., how ASGS evaluated their graduate education in statistics) and perceived knowledge (PK; i.e., how ASGS evaluated their knowledge in statistics from their education); ranking of statistical topics based on PRE, PK, and confidence in performing statistical analyses (CPSA); cluster analysis of statistical topics for PRE, PK, and CPSA; and factors (demographic, overall training, interest in statistics, and field of study) associated with the overall scores (OS) for PRE, PK, and CPSA. Students had greater ( $P < 0.05$ ) PRE than PK for most of the statistical topics included in this study. The moderate to high repeatability of answers within statistical topics indicates substantial correlations in ASGS answers between PRE and PK. The cluster analysis resulted in distinct groups of “Traditional” and “Nontraditional” statistical topics. ASGS showed lower ( $P < 0.05$ ) scores of PRE, PK, and CPSA in “Nontraditional” compared with “Traditional” statistical methods. Several factors were associated ( $P < 0.05$ ) with the OS of PRE, PK, and CPSA. In general, factors related to greater training and interest in statistics of ASGS were associated with greater OS, such as taking more credits in statistics courses, having additional training in statistics outside the classroom, knowing more than one statistics software, and more. This study provided comprehensive information on the perceived level of education, knowledge, and confidence in statistics in ASGS in the United States. Although objective measurements of their training in statistics are needed, the current study suggests that ASGS have limited statistical training on topics of major importance for the current and future trends of data-driven research in animal sciences.

**Key words:** graduate education, learning, statistical methods, statistics, survey, teaching

## Abbreviations

95% CI	95% credible interval
ASGS	animal science graduate students
CPSA	confidence in performing statistical analyses
ICC	intra-class correlation
OS	overall scores
OS <sub>CPSA</sub>	overall scores based on questions related to confidence in performing statistical analyses
OS <sub>PK</sub>	overall scores based on questions related to perceived knowledge
OS <sub>PRE</sub>	overall scores based on questions related to perceived received education
PK	perceived knowledge
PRE	perceived received education

## Introduction

Statistical analysis of data and designing of experiments are critical to the education of animal science graduate students (ASGS). Within the United States, ASGS are generally expected to take at least one course in statistics and experimental designs during their graduate training, regardless of their specialization area. ASGS are immersed in experiential learning opportunities where they are expected to conduct experiment(s), apply the basic concepts of experimental design, develop scientific hypotheses, and perform unbiased sampling of experimental units, randomization, and data collection. Afterward, the collected data are subjected to rigorous statistical analyses, where the assumptions of the statistical methods used must be met to validate the results obtained.

Most of the animal performance data obtained in experiments performed by ASGS include well-established traits with normal distributions, such as milk production in cows, growth in calves, feed intake in pigs, and more. However, with the development of *high-throughput technologies* for phenotyping and the increasing emphasis on *Big Data*, *Phenomics*, *-omics technologies*, and *Precision Livestock Farming*, the novel data generated from these present many statistical challenges. To begin with, large amounts of data are generated from these technologies. Although the use of spreadsheets, such as in Microsoft Excel, is essential for ASGS to gain familiarity with data management, its use for managing large datasets is limited (Auker and Barthelmess, 2020). The use of spreadsheets to manage large datasets increases human error compared with hard-coding and limits the ability of users to track errors (Panko, 2016). Hence, dealing with such large datasets requires ASGS to be proficient in computer coding. Another statistical challenge associated with *Big Data* is using proper methods and models to analyze the data. For example, sensor data obtained on the same animal results in correlated responses that must be appropriately accounted for in statistical models (Chitakasepornkul et al., 2018) and may result in datasets that have a large number of parameters ( $p$ ) to be estimated compared with the number of observations ( $n$ ). This results in the “ $p \gggg n$ ” issue, where there are a large number of traits to be analyzed (i.e.,  $p$ ), whereas the number of “independent” observations (i.e.,  $n$ ) is relatively small. This issue is not addressed using classical approaches and requires the use of appropriate and complex methods to deal with the high dimensionality of the data (Hastie et al., 2017; Morota et al., 2018).

Also, many types of data generated by these technologies are not suitable to be used in a classical linear model, even if it is not uncommon to find such examples in the literature for animal science research (Goulart et al., 2020; Jang et al., 2020; Ran et al., 2020). For example, data generated from RNA-seq (gene expression; *transcriptomics*), isotope labeling-based mass spectrometry (*proteomics*), and 16S rRNA (*microbiome*) studies require normalization (not related to the data having normal distribution) of the data to allow proper comparison between treatments using generalized statistical methods, such as negative binomial, Poisson, and others (Li et al., 2019; Sanglard et al., 2020).

The push for research projects to be promptly finalized in the form of published peer-reviewed publications relies on the statistical software proficiency of at least one of the authors. R, SAS, and SPSS are the most used statistical software in scientific papers (Muenchen, 2012). Overall, the use of statistical software enhances the learning of statistical concepts (Chance et al., 2007). However, the effective learning of statistical concepts using statistical software is questionable when the focus is on the software and not on the subject (Caple, 1996; Moore, 1997; Ozgur et al., 2015). Many ASGS rely on built-in procedures or external packages available on these software without knowing or understanding the statistical methodology behind them. Post graduation, many ASGS will take positions where they are required to make statistical decisions as researchers, peer reviewers of manuscripts and grants, instructors, major professors, etc. Hence, our animal science community must be training ASGS to have not only statistical software skills but also proficiencies in the basic and advanced concepts of data analysis.

Therefore, the aim of this study was to evaluate the perceived statistical training of ASGS at U.S. institutions by performing a nationwide survey, including questions about their general graduate education, statistical training and skills, and their professional goals. The objectives of this study were to 1) quantify the quality of their self-assessed education and knowledge, 2) quantify their self-assessed confidence in performing statistical analyses (CPSA), 3) identify groups of statistical topics for which differences exist in the self-assessed education, knowledge, and CPSA of ASGS, and 4) identify factors associated with their self-assessed education, knowledge, and CPSA.

## Materials and Methods

This study and data collection protocols were approved by the Institutional Review Board from Iowa State University (Protocol# 20-131-00). The voluntary survey was carried by Qualtrics (Provo, UT).

### Recruitment of participants

Information about the survey (<https://faculty.sites.iastate.edu/serao/page/assessing-statistical-training-animal-science-graduate-programs-us>) was shared with potential participants via multiple communications. An e-mail message was sent out to 50 Department Heads and Chairs of Animal Science-related Departments via the North Central Regional Association and National Information Management and Support System administrators. The Poultry Science Association advertised the survey on its April 2020 *eNewsletter*, and the American Dairy Science Association advertised the survey through its listserv of graduate students. Finally, *snowball* sampling was used through

graduate students and other professionals in Animal Science, in which information about the survey was shared among peers (Johnson, 2014). The survey was open for 12 d, between April 9 and 24, 2020.

### Eligibility

Based on their status as of May 15, 2020, participants must have answered “yes” to the following prescreening questions to participate in this study: 1) “Are you currently enrolled in a MS and/or PhD program in the US?”; 2) “Are you conducting, or have you conducted research as part of the graduate requirements for your MS and/or PhD degrees?”; 3) “Will your research (not necessarily all of it) be included in a thesis/dissertation as part of the graduate requirements for your MS and/or PhD degrees?”; 4) “Are you currently pursuing a MS and/or PhD degree in an Animal Science-related Department (Dept.), such as, but not limited to, Dept. of Poultry Science, Dept. of Dairy Science, Dept. of Animal Bioscience, etc?”; and 5) “Have you completed at least one semester and one graduate-level course in statistics during your graduate education?”

Participants who answered “Yes” to all 5 screening questions had access to the remaining 33 questions within the survey. Participants who answered “No” to any of the screening questions were ineligible to participate in the survey.

### Survey, compensation, and data security

The complete survey questionnaire is available in [Supplementary Material 1](#). Information regarding compensation and data security is available in [Supplementary Material 2A](#). The 33 remaining questions were divided into three groups:

1. Demographics and overall training (15 questions). This included questions about their degree, academic institution, field of study, etc;
2. Graduate Education on Statistics (8 questions). This included questions about the topics in statistics that they have learned, their assessment of quality of education in statistics, number of credits taken in statistics, etc;
3. Self-Assessment on Statistics and Career Path (11 questions). This included questions about their self-assessment of knowledge in statistics and statistical software, their level of comfort/CPSA, their career goals in statistics, etc.

Out of these 33 questions, one from group I and two from group II above were answered for multiple topics in statistics based on scores. These scores ranged from 0 (not available/covered) to 5 (high quality/knowledge/confidence). The summary statistics are shown in [Supplementary Table S1](#). These questions were:

- Q20) Evaluate the overall quality of the Graduate Education on Statistics that you have received on each of the items below. This question had 30 sub-questions related to perceived received education (PRE). This can be interpreted as how ASGS evaluated their graduate education received in statistics;
- Q28) Based on the same items included in Q20, evaluate your knowledge on those items. This question had the same 30 sub-questions as above. However, ASGS answered them based on their perceived knowledge (PK). This can be interpreted as how ASGS evaluated their knowledge in statistics from their graduate education;

- Q32) Given your knowledge on statistics, how confident do you feel by performing each of the tasks described below? This question had 31 sub-questions related to their CPSA;

### Data from survey

The survey was accessed by 485 participants, with 435 completing it and 50 being non-eligible based on the five prescreening questions. Data screening indicated that an additional 19 participants were ineligible (e.g., participants enrolled in a Doctorate of Veterinary Medicine degree only from Department of Plant Science) and were removed from the data, resulting in 416 eligible participants. Information on the academic institution was used to create three new variables: *Land-Grant* (with values “Yes” and “No”), *Research Type* (with values “R1” and “Not R1”), and *Type* (with values “Public” and “Private”).

### Quality control

Prior to statistical analyses, explanatory variables and responses for the PRE, PK, and CPSA sub-questions from the 416 eligible students were subject to quality control to remove groups with low representation and answers showing bias, respectively. Explanatory variables with two levels and with less than 5% of participants in one of the levels were not used. In this step, *Land-Grant* (with levels “Yes” and “No”), *Type* (with levels “Public” and “Private”), and *Have you learned any stats software in these courses?* (with levels “Yes” and “No”) were removed. For the explanatory variables with more than two levels, the levels with less than 5% of participants in them were combined and classified as “Other.” This was only observed for variables *Species* and *Field of study*. The *Species* information was not used in this study.

Within the PRE, PK, and CPSA groups of sub-questions, students with more than 10% of missing answers were removed from the dataset. In this step, 14 students had answers set to missing for PRE, PK, and CPSA groups of sub-questions. In addition, the remaining students showing the same scores for all sub-questions (e.g., score of 4 in all questions) within each group (i.e., PRE, PK, or CPSA) were removed to avoid bias due to lack of interest of the participant, which may be captured by a sequence of repeated answers (Akay and Karabulut, 2020). In this step, 2, 6, and 8 students were removed for PRE, PK, and CPSA groups of sub-questions, respectively. After quality control, data from 400, 396, and 394 students were used for the PRE, PK, and CPSA groups, respectively, with a completion rate of 99.8%, 99.8%, and 99.9%, respectively. The summary statistics of the data after quality control is presented in [Supplementary Table S2](#). The overall response rate of the eligible participants ranged from 5% to 100%, with an average (SD) of 96.93% (14.94%) and a median of 100%. The name of the academic institutions and departments and their respective number of eligible participants are presented in [Supplementary Table S3](#).

### Overall score variables

To obtain variables representing the overall assessment of ASGS on the statistical topics included in this study, overall scores (OS) were created for the PRE, PK, and CPSA groups of sub-questions. For each participant with complete data for these questions, the OS were obtained by simply summing their scores across the 30 (for PRE and PK) and 31 (for CPSA) sub-questions. These OS were then named as PRE scores ( $OS_{PRE}$ ), PK scores ( $OS_{PK}$ ), and CPSA scores ( $OS_{CPSA}$ ) based on the PRE, PK, and CPSA groups of sub-questions, respectively. Prior

to analyses, within each OS, the data were scaled to represent scores from 0 to 100, where 100 represented the maximum OS. This strategy allowed for a more direct comparison between the OS. The distribution of the scaled OS is shown in Figure 1 along with their summary statistics. All three OS were normally distributed (Shapiro–Wilk’s  $P \geq 0.078$ ), with similar means, SDs, and minimums. The OS were used in part of the statistical analyses described below.

### Statistical analysis

The survey data from eligible participants were subject to multiple statistical analyses. All the management of the data and results were carried out using R version 3.6.3 (R Core Team, 2017) in RStudio Version 1.1.383 (RStudio Team, 2020) through personal codes and the *tidyverse* package (Wickham et al., 2019). All plots presented in this study were generated in R using the *tidyverse* and *gridExtra* (Auguie and Antonov, 2017) packages.

### Relationship between PRE and PK scores

This analysis aimed to evaluate the differences between PRE and PK within each statistical topic. An ordinal probit regression was used to test the differences in responses for each statistical

topic between PRE and PK. The analysis was carried for each of the 30 statistical topics separately. The statistical model was :

$$\text{probit} [\text{Pr}(Y \leq i)] = \alpha_0 + \beta'X + u'Z, \quad 0 \leq i \leq 5 \quad (1)$$

where  $Y$  is the response of the participant (i.e., scores),  $i$  is the score level (0 to 5);  $\alpha_0$  is the overall intercept;  $\beta$  is the vector of solutions of the fixed effects;  $X$  is the incidence matrix of fixed effects (described below);  $u$  is the vector of solution of the random effects; and  $Z$  is the incidence matrix of random effects (described below). The model above included the fixed effects of graduate *Degree* being pursued (MS or PhD), *Years of graduate education* (covariate), type of question (PRE and PK), and each *Field of study* as a separate categorical effect (16 different effects; Supplementary Table S2). The participant was used as the random effect for all analyses to account for repeated records between type of question, assuming  $u \sim N(0, \sigma_u^2)$ , where  $\sigma_u^2$  represents the participant variance. Analyses were performed using a Bayesian framework. In addition, the intra-class correlation (ICC) for each statistical topic was calculated to measure the impact of participants on the results as the posterior mean of the ratio:

$$\text{ICC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2} \quad (2)$$

Where  $\sigma^2$  represents the residual variance, assumed to be 1 for probit models, and  $\sigma_u^2$  represents the participant variance. Details about number of iterations, model convergence and assumptions, estimation of effects, and more are shown in Supplementary Material 2B.

### Difference in scores between statistical topics

Three analyses were performed, one for each group of sub-questions (i.e., PRE, PK, and CPSA) to rank statistical topics based on their scores. The ordinal mixed probit model used was similar to equation 1 and included the fixed effects of graduate *Degree* being pursued (MS or PhD), *Years of graduate education* (covariate), each *Field of study* as a separate categorical effect (16 different effects; Supplementary Table S2), and the statistical topics, with participant as a random effect. There were 30 statistical topics in the analysis for PRE and PK and 31 for CPSA. Additional information about the analyses is included in Supplementary Material 2C.

### Hierarchical cluster analyses

Hierarchical cluster analyses were performed to identify statistical methods with similar score patterns based on the answers of ASGS. This was performed separately for PRE, PK, and CPSA. Prior to these analyses, the missing scores were imputed since cluster analysis requires complete data. Detailed information regarding the imputation of data is shown in Supplementary Material 2D. After imputation of the data, the score data were pre-adjusted for the fixed effects of graduate *Degree* being pursued (MS or PhD), *Years of graduate education*, and *Field of study* based on the analyses described in Relationship between PRE and PK scores. In other words, the data used for these analyses corresponded to the sum of estimates for the residual and participant effects obtained for each statistical topic.

The pre-adjusted data were used to calculate one Manhattan dissimilarity matrix (Carmichael and Sneath, 1969) for each analysis (PRE, PK, and CPSA) to be used in the DIANA (DIvulsive ANalysis Clustering) algorithm (Kaufman and Rousseeuw, 1990) implemented in the cluster package (Maechler et al., 2019) in R.

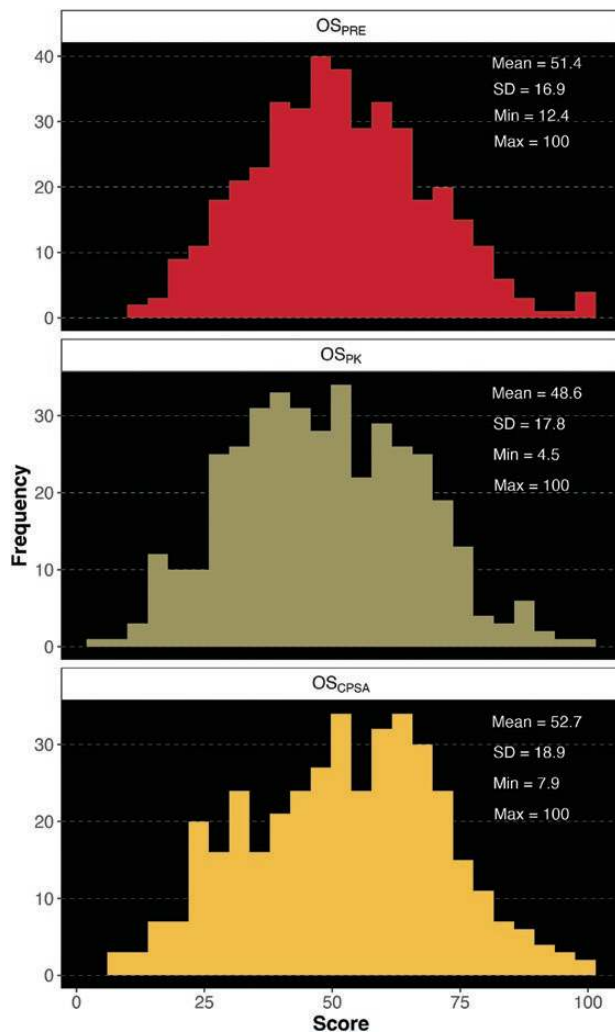


Figure 1. Distribution and summary statistics of the overall scores (OS) after scaling them from 0 to 100. OS<sub>PRE</sub>, OS for perceived received education (PRE); OS<sub>PK</sub>, OS for perceived knowledge (PK); and OS<sub>CPSA</sub>, OS for confidence in performing statistical analyses (CPSA).

The clustering structure of the data, that is, how well the data clustered together, was obtained by computing the divisive coefficient for each analysis (Kaufman and Rousseeuw, 1990). For the PRE and PK clusters, the similarity between their dendrograms was compared by computing the entanglement score (Galili, 2015). Information on the computations of the divisive coefficient and entanglement score is detailed in [Supplementary Material 2D](#). Finally, the optimal number of clusters was defined through the silhouette method (Rousseeuw, 1987) available in the *factoextra* package (Kassambara and Mundt, 2020) in R.

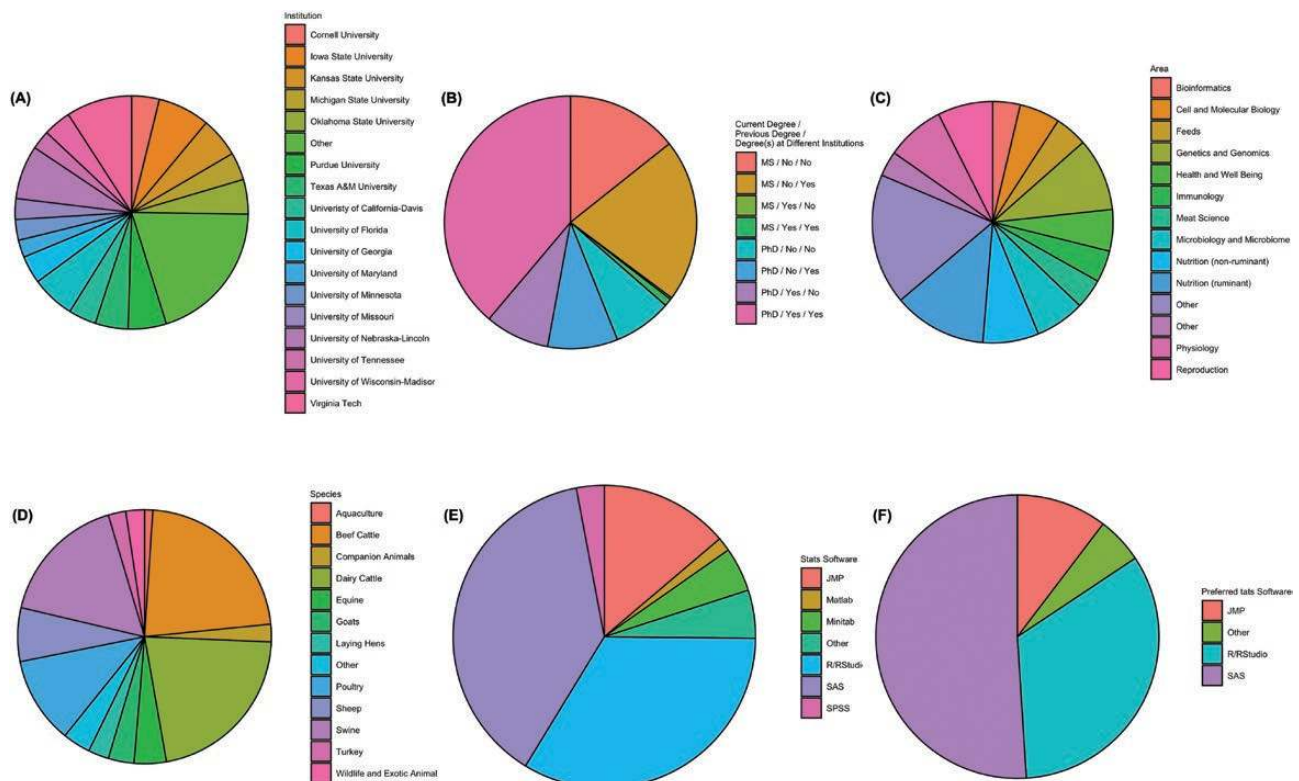
### Factors associated with the OS

The aim of this analysis was to identify factors that better explained the variation associated with the OS (i.e.,  $OS_{PRE}$ ,  $OS_{PK}$ , and  $OS_{CPSA}$ ) of ASGS. For this, information on 38 factors ([Supplementary Table S4](#)), including information about their demographics, overall training, career path, and *Field of study*, was used. The model included all 38 effects as main fixed effects (i.e., no interactions) and underwent backward selection based on Akaike Information Criterion (Akaike, 1974) to identify the most important information explaining variation on the OS. In these analyses, information of whether ASGS were users of SAS and/or R (referred to as *Stats software user*), due to their much greater frequency compared with the users of other software ([Figure 2](#)), was used instead of information on each software. This was performed separately for each OS. Additional information regarding the computational methods

used and model assumptions is available in [Supplementary Material 2E](#).

## Results and Discussion

The data obtained from this survey included participants (ASGS) from major animal science institutions in the United States, working in research from a range of areas and species, and utilizing major statistical software, and, thus, met the expectation of obtaining a representative sample of ASGS in the United States. The distribution of participants across academic institutions, degrees, statistical software, and fields of study and species of graduate research are shown in [Figure 2](#). Eligible participants in this study came from 43 U.S. academic institutions ([Figure 2A](#)). Of these, Virginia Polytechnic Institute and State University (Virginia Tech), *University of Nebraska-Lincoln*, and *Iowa State University* had 38, 31, and 30 ASGS participants, respectively. These represented about 24% of the all participants. The average ( $\pm$ SD) number of participants per institution was  $9.7 \pm 9.6$ , with a median of 6. Most of the participants in this study were PhD students (64% vs. 36% MS students) having previously obtained graduate degree and degree(s) from other institutions (39% of the total; [Figure 2B](#)). A total of 42 fields of study (i.e., areas of research) were selected by ASGS, with most representation from *Ruminant Nutrition* (12.5%), *Genetics and Genomics* (9.8%), and *Physiology* (7.8%). There were 22 species selected by the participants, with greater frequency of *Beef Cattle* (22.3%), *Dairy Cattle* (21.6%), and *Swine* (16.7%). From the 22 statistical software used by ASGS, SAS, R/RStudio, and JMP were the most



**Figure 2.** Distribution of responses for different categories. (A) U.S. academic institutions from participants; (B) participants based on their current degree (MS or PhD), on having graduate degree completed (No or Yes), and on having any degrees obtained at an institution other than the current institution (No and Yes); (C) research areas (i.e., fields of study) that participants are currently working on in their graduate education; (D) species that participants worked with during their whole graduate education; (E) statistical software that participants have experience with; (F) favorite statistical software from participants. The responses categorized as “other” in panel (A) are available in [Supplementary Table S3](#) and in C through F are available in [Supplementary Table S2](#).

representative, with 38.3%, 33.6%, and 13.8%, respectively, of all answers. These were also the preferred software of 51% (SAS), 33.3% (R/RStudio), and 10.3% (JMP) of the participants.

### Relationship between PRE and PK scores

The ICCs for each of the statistical questions included in PRE and PK groups, along with their score differences and 95% credible interval [CI], are shown in Figure 3. The complete results showing ICC and averages are shown in Supplementary Table S5. All ICCs [95% CI] were moderate to high and different than zero ( $P < 0.05$ ), ranging from 0.51 [0.44;0.58] for *Any Stats Software* to 0.792 [0.75;0.84] for *Matrix Algebra*. These results suggest a repeatable perception within ASGS about their PRE and PK across statistical topics. Hence, these results indicate a strong correlation of scores between the PRE and PK. This might be in part driven by a self-esteem component, with some ASGS showing general over- or under-confidence, whereas others not much. There was no relationship (Spearman's  $r = 0.16$ ;  $P = 0.388$ ) between ICCs and differences between statistical topics. Although ICCs were moderate to high, this does not mean that there were no differences between the average scores between PRE and PK. Therefore, whether ASGS indicated having greater PRE or PK for each of these statistical topics was also tested.

There was a consistency in the average scores between PRE and PK across statistical topics. Most of the differences were positive, indicating that students perceive to have greater education than knowledge in these statistical topics. The only negative difference was for *Data Management*, although not significant (Bonferroni-corrected  $P$ -value [ $P_{\text{Bonferroni}} = 1$ ]). In addition to this statistical topic, there were no differences between PK and PRE for *Computer Coding* ( $P_{\text{Bonferroni}} = 1$ ), *Correlation* ( $P_{\text{Bonferroni}} = 0.316$ ), *Heterogeneous Variances* ( $P_{\text{Bonferroni}} = 0.126$ ), *Machine Learning* ( $P_{\text{Bonferroni}} = 0.062$ ), *Principal Component Analysis* ( $P_{\text{Bonferroni}} = 1$ ), and *Response Surface Methodology* ( $P_{\text{Bonferroni}} = 0.663$ ).

All other statistical topics were significant at  $P_{\text{Bonferroni}} \leq 0.032$ . Of these, there was not a clear division of statistical topics based on the differences between PRE and PK. For example, the highest difference [95% CI] was for *Multiple Testing*, with 0.66 [0.50; 0.82], which can be considered a typical statistical topic used in many of the studies. However, most of the large differences between PRE and PK were found for statistical topics that are not usually used in most of the studies. For example, *Poisson Regression*, with 0.60 [0.43; 0.76], which can be used in studies of bacterial counts (Capps et al., 2020); *Theoretical Statistics*, with 0.55 [0.39; 0.72], which can be used in methodological studies (See et al., 2020); *Observational Studies*, with 0.53 [0.38; 0.70], which are vastly used in quantitative genetic studies (Scanlan et al., 2019); and *Bayesian Statistics*, with 0.49 [0.32; 0.66], which can be used for a variety of analysis in place of the traditional frequentist approach (Hong et al., 2020; Marina et al., 2020; Sanglard et al., 2020), are among the statistical topics showing the largest differences between PRE and PK. However, other methods that are vastly used in animal science research, such as *Non-Parametric Methods*, with 0.58 [0.41; 0.74], which can be used as an alternative for nonnormal data (Sterndale et al., 2020), and *ANOVA*, with 0.58 [0.42; 0.74], which is used in virtually every study, showed a larger difference between PRE and PK.

Differences between perceived education and knowledge are well documented in the literature. Although the learning experience is highly dependent on the instructor (Jaques, 2003), it is well accepted that students perceive knowing well their self-assessment of actions, but they are often wrong about it (Bowman and Seifert, 2011). The learning environment of students also plays a role in how students and instructors perceive their ability to learn and to teach, respectively, impacting, therefore, the relationship between PRE and PK (Frenzel et al., 2007). Students perceived greater relationship between PRE and PK when they are more academically

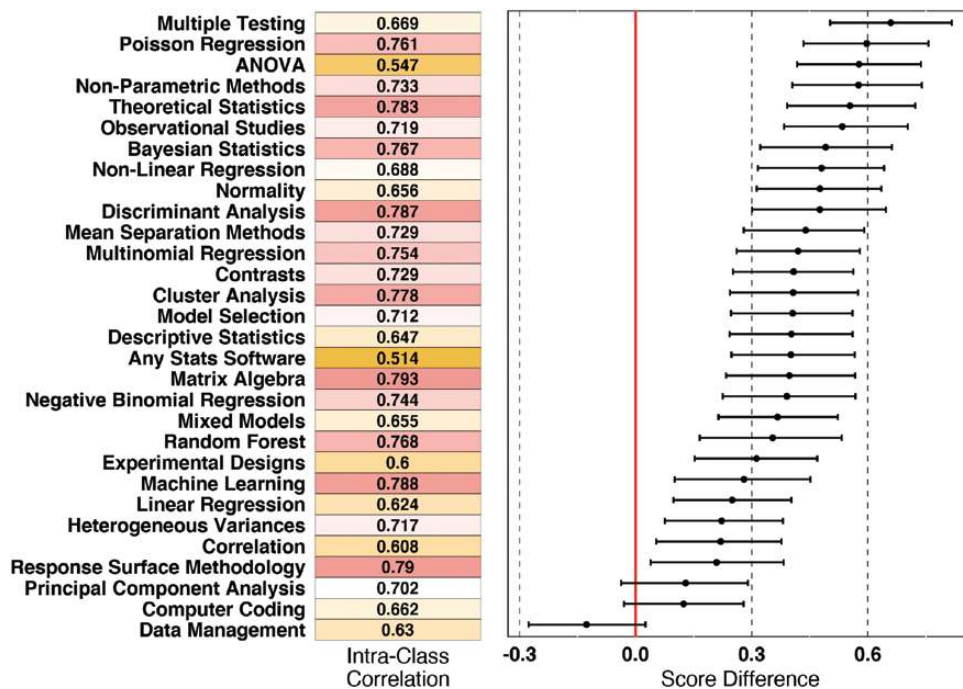


Figure 3. Relationships between Perceived Received Education (PRE) and Perceived Knowledge (PK) scores. Results are presented for each statistical topic included in PRE and PK. The left panel shows the intra-class correlation (ICC) of the scores of participants between their PRE and PK. Tiles are color-coded from gold to cardinal, as ICC values increase. The right panel shows the score difference in scores between each statistical topic. Positive and negative values represent greater scores for PRE and PK, respectively. Error bars represent the 95% CI (analogous to confidence intervals) of the score differences.

successful (Gabel and National Science Teachers Association, 1994), which is in accordance with other results described in the following sections, where ASGS showing greater self-reported CPSA also had reported having greater PRE and PK. In addition, statistics might not be among the favorite topic for ASGS, which could impact how they respond with regard to PRE and PK. Another factor impacting these outcomes is the interaction between instructors and students (Sebastianelli et al., 2015), but this information was not included in the survey questionnaire.

In addition, the order of the questionnaire may have also potentially caused some of these differences (Israel and Taylor, 1990). All participants received questions in the same order, with PRE sub-questions being answered prior to PK sub-questions, although they had the option to freely forward and return to questions as desired. In addition, the sub-questions related to PRE and PK were the same, with the only difference being their respective focus; in the former, ASGS were asked about their learning experience on these statistical topics, whereas in the latter, they were asked about their knowledge on the topics. Hence, in addition to this similar interpretation, which is supported by the moderate to high ICCs between PRE and PK (Figure 3) and the similar dendrograms (Figure 5A), having PRE questions answered prior to PK questions could have resulted in some bias in their answers. Unfortunately, this potential bias cannot be accounted for in our analyses. Nonetheless, the

moderate to high ICCs and the consistently greater values of PK compared with PRE might provide the support that any potential bias, if present, was minimal.

### Difference in scores between statistical topics

The results showing the expected probabilities for each score and statistical topics are shown in Figure 4. The complete results are available in Supplementary Table S6. For the statistical topics in the PRE sub-questions (Figure 4A), the lowest scores ( $P < 0.05$ ) were obtained for *Machine Learning*, *Random Forest*, and *Response Surface Methodology*, which had average responses [95% CI] of 1.20 [1.05; 1.34], 1.22 [1.08; 1.36], and 1.31 [1.17; 1.44], respectively ( $P > 0.05$ ), and the highest score ( $P < 0.05$ ) was obtained for ANOVA, with 4.06 [3.92; 4.20]. Likewise, for statistical topics in the PK sub-questions (Figure 4B), the same statistical topics with lowest scores in the PRE questions were found in this analysis ( $P < 0.05$ ), with 1.00 [0.85; 1.14], 1.03 [0.88; 1.17], and 1.18 [1.04; 1.33], for *Random Forest*, *Machine Learning*, and *Response Surface Methodology*, respectively ( $P > 0.05$ ), whereas ANOVA again had the highest score 3.77 [3.62; 3.91] ( $P < 0.05$ ). Overall, both analyses showed that basic statistical methods, such as ANOVA, *Linear Regression*, *Correlation*, and others, had higher average scores than more complex statistical methods, such as those used in *Big Data* technologies (e.g., *Random Forest*, *Machine Learning*, and *Discriminant Analysis*).

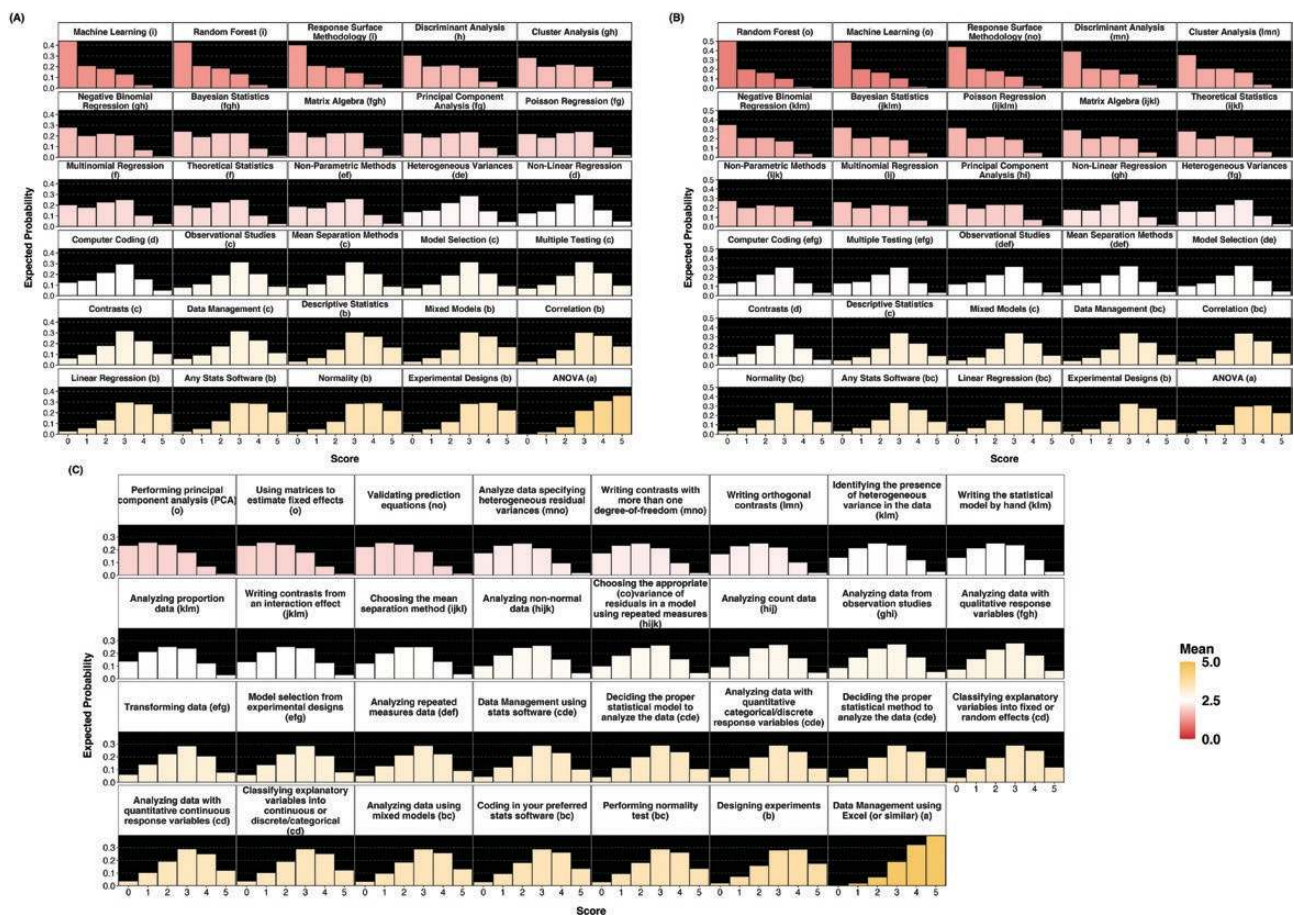


Figure 4. Distribution of predicted probabilities of scores for each statistical topic. Panels (A), (B), and (C) show the results from the analyses for Perceived Received Education (PRE), Perceived Knowledge (PK), and Confidence in Performing Statistical Analyses (CPSA), respectively. The y-axis shows the predicted probabilities. The x-axis shows the scores, representing answers from participants. These ranged from 0 (statistical topic not available/covered) to 5 (high quality/knowledge/confidence). Statistical topics with different letters within parenthesis indicate statistical difference ( $P < 0.05$ ) between them.

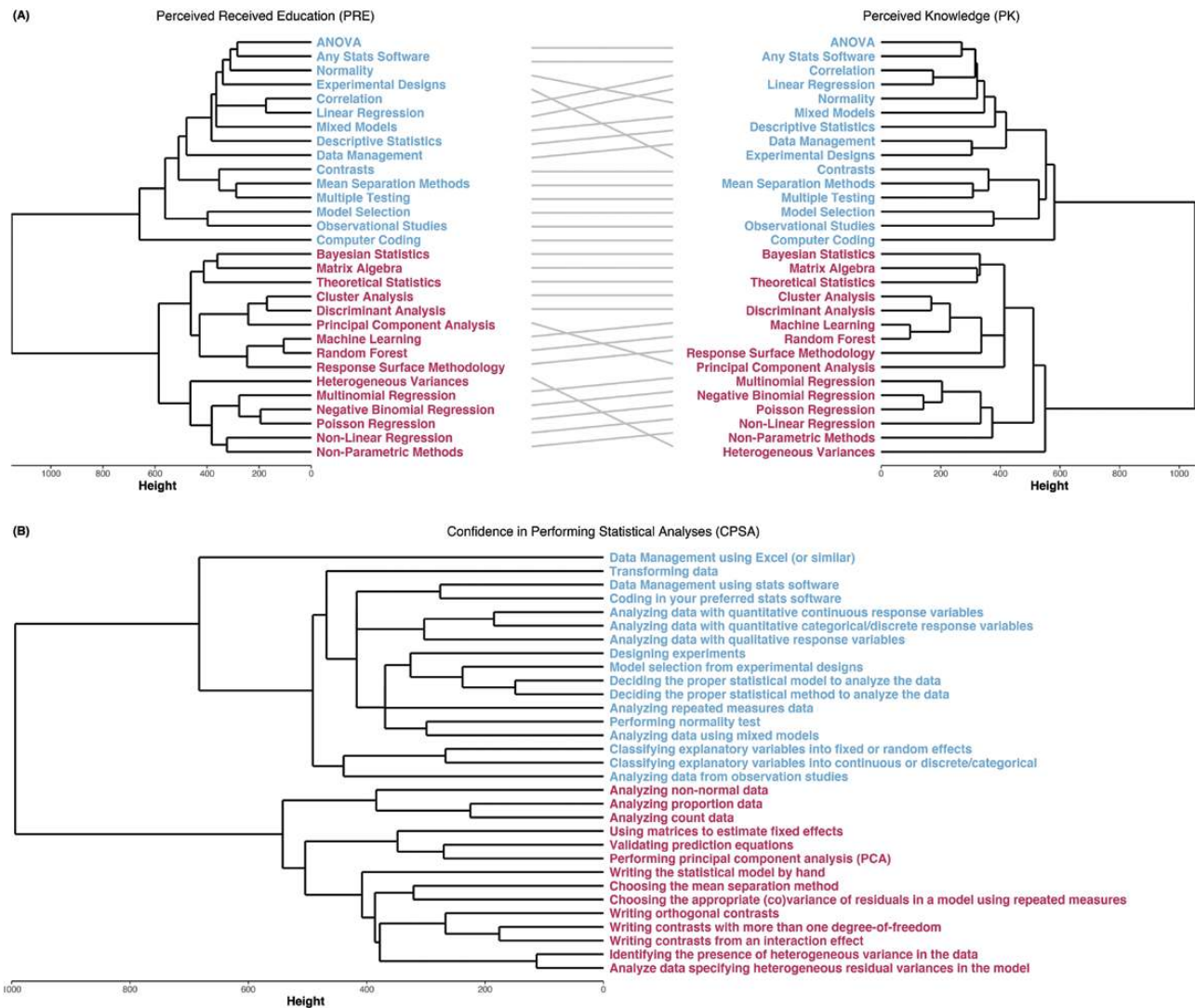


Figure 5. Dendrograms from hierarchical cluster analyses. The results in panel (A) represent the cluster of statistical topics based on the answers for Perceived Received Education (PRE) and Perceived Knowledge (PK). The gray lines show the connection of the same statistical topic between the two dendrograms. The results in panel (B) represent the cluster of statistical topics based on the answers for Confidence in Performing Statistical Analyses (CPSA). In all analyses, two clusters were formed: one representing “Traditional Statistical Topics” (blue cluster) and the other “Nontraditional Statistical Topics” (pink cluster).

For the statistical topics in CPSA, ASGS were most confident ( $P < 0.05$ ) in *Data Management using Excel (or similar)*, with 4.62 [4.47; 4.77] and least confident ( $P < 0.05$ ) in *Performing principal component analysis*, *Using matrices to estimate fixed effects*, *Validating prediction equations*, *Analyze data specifying heterogeneous residual variances*, and *Writing contrasts with more than one degree-of-freedom*, with averages of 1.98 [1.83; 2.11], 1.98 [1.84; 2.12], 2.03 [1.88; 2.17], 2.25 [2.11; 2.41], and 2.26 [2.11; 2.41], respectively ( $P > 0.05$ ). Interestingly, some of the statistical topics showing low scores were somewhat unexpected. For example, *Writing orthogonal contrasts*, with 2.30 [2.14; 2.45], and *Writing contrasts from an interaction effect*, with 2.50 [2.36; 2.65], were not different from each other ( $P > 0.05$ ), had overall low scores, and are somewhat simple procedures to be performed. Although not needed, knowledge in matrix algebra is useful to write any type of contrasts, and *Using matrices to estimate fixed effects* had statistically the lowest score, with 1.98 [1.83; 2.11], which could help explain how statistical topics including the use of contrasts had low scores.

In addition, *Theoretical Statistics* and *Matrix Algebra* had low scores for PRE and PK questions. These statistical topics ranked 19<sup>th</sup>, with 2.07 [1.93; 2.21], and 23<sup>rd</sup>, with 1.92 [1.78; 2.06], respectively, for PRE, and 21<sup>st</sup>, with 1.75 [1.62; 1.89], and 22<sup>nd</sup>, with 1.69 [1.55; 1.84], respectively, for PK, out of the 30 statistical topics used in this study. With these lower scores, it is expected that ASGS do not feel comfortable in using methods that require some knowledge on *Theoretical Statistics* and *Matrix Algebra*. Another statistical topic that had unexpectedly low score was *Validating prediction equations*, with 2.03 [1.88; 2.17]. The use of prediction equations is a standard practice in areas such as animal nutrition (Gutierrez et al., 2014; Adeola and Kong, 2020). However, as in these studies, many other animal science studies that develop prediction equations do not validate their equations using independent datasets or cross-validation. Hence, although validation of prediction equations is an important aspect in research, ASGS might have given low scores for this statistical topic since this is usually overlooked in animal science research.



The statistical topics showing greater average scores in CPSA are generally simple. However, from the theoretical statistics point of view, *Analyzing data using mixed models*, with 3.48 [3.32; 3.63] is quite complex. This statistical topic requires proper specification of the (co)variance structure in the model and convergence of the model to be achieved due to the iterative estimation of (co)variance components (usually through restricted maximum likelihood) given the fixed effects in the model. With the popularization of statistical software, advanced theoretical knowledge in mixed models is not needed by ASGS to analyze their data using mixed models. Also, over 20 yr ago, the American Society of Animal Science, including partnerships with other societies, started offering the “Mixed Models Workshop” during its annual meeting. Hence, the increased use of statistical software and additional training offered to ASGS could help explaining why *Analyzing data using mixed models* had one of the largest average scores. The use of statistical software for data analysis could also explain the high average score of *Coding in preferred stats software*, with 3.50 [3.32; 3.63]. The ASGS in this study could have interpreted *Coding* very broadly, such as the ability to write commands to perform statistical analyses in their statistical software of choice. With 97.5% of participants indicating to having learned at least one statistical software during their graduate classes, it was expected that *Coding in preferred stats software* had a high average score.

Although not statistically different than *Coding in preferred stats software* ( $P > 0.05$ ), *Data management using stats software* had a numerically lower average score, with 3.27 [3.12; 3.42]. This statistical topic could be a better measurement of the capabilities of ASGS for hard-coding than *Coding in preferred stats software*. As shown later in this manuscript, these two statistical topics were the only two clustered (i.e., leaves) in the same clade, further indicating how correlated the scores given by ASGS were for them. Hence, the numerical differences between these two statistical topics could suggest that ASGS are more comfortable in using statistical software to perform standard statistical analyses than managing data. Overall, these results for PRE, PK, and CPSA questions provide novel information with regard to statistical topics taught, learnt, and performed, respectively, by ASGS in the United States. The differences among statistical topics in each analysis well depict the main statistical topics being covered in animal science graduate programs across the country. This information should be used to improve the statistical training of topics showing low average scores and/or those needed for the statistical challenges faced in this research community moving toward using *high-throughput technologies*.

Although the overall objective of these analyses was to identify differences in scores among statistical topics based on the PRE, PK, and CPSA, the impact of the other effects in the model on the OS was also investigated. In general, only a few associations were found for these other effects. For instance, *Degree* being pursued was significant ( $P = 0.01$ ) only for the CPSA analysis, in which the scores of PhD ASGS were 0.32 [0.07; 0.57] points greater than those from MS ASGS. *Years of graduate education* was also associated ( $P < 0.001$ ) with CPSA scores, and it had a tendency on the scores for PRE ( $P = 0.073$ ) and PK ( $P = 0.051$ ). In all analyses, as the *Years of graduate education* increased, the scores increased by 0.15 [0.07; 0.22] for PRE, 0.07 [0; 0.14] for PK, and 0.15 [0.07; 0.22] for CPSA. The results found for *Degree* being pursued and *Years of graduate education* are reasonable, as it should be expected that more experienced students have greater scores for these parameters. As for the fields of study, for PRE, ASGS who selected *Housing and Management* tended ( $P = 0.085$ ) to have scores that were 0.33 [−0.04; 0.68] points greater than those who did not select this field of study. ASGS who selected *Immunology* had significantly lower

scores than those who did not select this field of study for PK (−0.35 [−0.66; −0.04];  $P = 0.023$ ) and CPSA (−0.52 [−0.86; −0.2];  $P = 0.003$ ), whereas a tendency was found for PRE (−0.27 [−0.56; 0.01];  $P = 0.056$ ). The reasons why these associations were found are unclear and a deeper investigation of the training of these ASGS would be required to better understand why these differences were found for *Immunology*. Nonetheless, it is interesting that, in all three analyses, students who work in *Immunology* perceive to have the overall lower statistical capabilities than those not working in this field of study.

The ICCs for within the PRE, PK, and CPSA questions were 0.382 [0.349; 0.422], 0.404 [0.367; 0.439], and 0.453 [0.414; 0.493], respectively, indicating a moderate repeatability, and hence, moderate correlation between answers from the same ASGS across statistical topics. It is important to note that the results presented in this study are based on the perception of ASGS on these statistical topics. Hence, in this study, we did not objectively test whether they are knowledgeable on these. As part of this project, a follow-up study was conducted in which part of the same ASGS were subjected to objective testing on several statistical topics, with the hope of shedding some light on how much of what they perceive is true.

### Hierarchical cluster analyses

The dendrograms from the cluster analyses are shown in Figure 5. For all analyses, two clusters were identified, and these could be broadly divided into “Traditional” (blue cluster) and “Nontraditional” (maroon cluster) statistical topics. There were 15 “Traditional” and 15 “Nontraditional” statistical topics for both PRE and PK clusters (Figure 5A). For CPSA, 17 and 14 statistical topics were included in these two clusters, respectively (Figure 5B). The divisive coefficients for these analyses were moderate high, with 0.74, 0.72, and 0.71, for the analysis using PRE, PK, and CPSA questions, respectively, indicating good clustering structures. This dimensionless coefficient ranges from 0 to 1, with larger values indicating stronger distinction between the clusters obtained in each analysis.

The entanglement between the PRE and PK dendrograms was measured to evaluate how similar the clustering of statistical topics was between the two analyses. The entanglement score measures how identical two dendrograms are, ranging from 0 to 1. When two clusters are perfectly identical (i.e., the order of the leaves are the same in both), the entanglement score is 0. The entanglement score was 0.02, indicating great similarities on how these statistical topics clustered together between the two analyses (Figure 5A). In fact, both clusters were included the same statistical topics between PRE and PK. The differences between the two dendrograms were only regarding the classification of each statistical topic (i.e., each leaf) in the clades, well aligning with the very low entanglement score obtained between the two dendrograms.

These results are in accordance with those previously reported in this study. The “Traditional” methods also showed overall greater PK and PRE scores in Figure 4. In fact, the cluster results (PRE and PK in Figure 5A) were used along with those described in *Difference in scores between statistical topics* to write a contrast to test the difference in scores between these two groups. For PRE, the “Traditional” and “Nontraditional” clusters had scores ( $P < 0.001$ ) of 3.17 [2.43; 4.08] and 1.84 [1.15; 2.48], respectively. For PK, the “Traditional” and “Nontraditional” clusters had scores ( $P < 0.001$ ) of 3.00 [2.35; 3.79] and 1.61 [0.94; 2.30], respectively. In addition to further supporting the greater PRE compared with PK, there is a clear difference in PRE and PK for the two groups of statistical methods.

**Table 1.** Effects associated<sup>1</sup> with the overall scores (OS)<sup>2</sup>

Effect	OS <sub>PRE</sub>		OS <sub>PK</sub>		OS <sub>CPSA</sub>	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
Age	-0.003 (0.24)	0.989	-0.004 (0.24)	0.968	-0.36 (0.28)	0.197
Years of previous professional experience			-0.95 (0.62)	0.126		
Graduate-level courses on statistics					1.76 (0.69)	0.001
Credits in graduate-level courses on statistics	0.29 (0.12)	0.016	0.38 (0.16)	0.020		
Degree						
MS	64.1 <sup>a</sup> (3.2)	0.007	73.7 (3.4)	0.104		
PhD	57.7 <sup>b</sup> (3.0)		68.5 (3.2)			
Research type						
Not R1	64.2 <sup>A</sup> (4.4)	0.099	74.8 (5.7)	0.166	89.9 <sup>a</sup> (5.0)	0.027
R1	57.6 <sup>B</sup> (2.5)		67.4 (2.7)		78.8 <sup>b</sup> (1.7)	
Graduate degrees completed						
No	60.2 (3.0)	0.535	71.2 (3.8)	0.934	82.6 (3.0)	0.164
Yes	61.6 (3.2)		71.0 (4.0)		86.1 (3.1)	
Degrees completed at other institutions						
No					86.6 <sup>a</sup> (3.2)	0.046
Yes					82.1 <sup>b</sup> (2.8)	
Previous professional experience						
No	62.3 <sup>A</sup> (2.9)	0.092				
Yes	59.5 <sup>B</sup> (3.1)					
Additional training on statistics						
No	58.7 <sup>b</sup> (2.9)	0.012	68.8 <sup>b</sup> (3.6)	0.047	82.8 (2.8)	0.151
Yes	63.0 <sup>a</sup> (3.1)		73.4 <sup>a</sup> (3.9)		85.9 (3.1)	
Courses taken at the Department of Statistics						
No	63.0 <sup>a</sup> (3.2)	0.014	72.7 (3.9)	0.168	86.5 <sup>A</sup> (3.2)	0.052
Yes	58.8 <sup>b</sup> (2.9)		69.5 (3.6)		82.2 <sup>B</sup> (2.7)	
Gives statistical advices to lab peers						
No			68.7 <sup>B</sup> (3.8)	0.060	79.7 <sup>b</sup> (3.1)	<0.001
Yes			73.5 <sup>A</sup> (3.7)		89.0 <sup>a</sup> (3.0)	
Uses codes from others in the lab for statistical analyses						
No	62.0 (3.0)	0.161				
Yes	58.9 (3.0)					
Comfortable in giving statistical advice in desired career path						
No	55.0 <sup>b</sup> (3.3)	<0.001	62.6 <sup>c</sup> (4.1)	<0.001	74.6 <sup>b</sup> (3.5)	<0.001
Somewhat Comfortable	61.7 <sup>a</sup> (3.0)		71.1 <sup>b</sup> (3.7)		85.8 <sup>a</sup> (3.1)	
Yes	65.9 <sup>a</sup> (3.4)		79.5 <sup>a</sup> (4.4)		93.0 <sup>a</sup> (3.6)	
Comfortable in reviewing the stats section of a manuscript						
No	72.8 <sup>b</sup> (3.6)	0.004	63.5 <sup>b</sup> (4.1)	<0.001	71.3 <sup>c</sup> (3.2)	<0.001
Somewhat Comfortable	82.7 <sup>a</sup> (3.5)		74.1 <sup>a</sup> (3.8)		84.9 <sup>b</sup> (3.1)	
Yes	83.2 <sup>a</sup> (4.3)		75.8 <sup>a</sup> (4.4)		97.0 <sup>a</sup> (3.8)	
Stats software						
Other	71.1 <sup>b</sup> (4.7)	<0.001	64.0 <sup>b</sup> (4.7)	0.027	73.3 <sup>c</sup> (4.1)	<0.001
R	79.7 <sup>ab</sup> (3.8)		71.7 <sup>ab</sup> (4.1)		83.2 <sup>bc</sup> (3.4)	
SAS	79.8 <sup>b</sup> (3.5)		72.8 <sup>ab</sup> (3.8)		86.8 <sup>b</sup> (3.1)	
R and SAS	87.8 <sup>a</sup> (3.5)		75.9 <sup>a</sup> (4.0)		94.2 <sup>a</sup> (3.1)	
Field of study (Cell and Molecular Biology)						
No			73.6 (3.2)	0.149		
Yes			68.6 (4.6)			
Field of study (Housing and Management)						
No	58.7 (2.5)	0.163				
Yes	63.1 (3.9)					
Field of study (Meat Science)						
No	57.9 <sup>b</sup> (2.7)	0.029	67.6 <sup>b</sup> (3.1)	0.063		
Yes	63.9 <sup>a</sup> (3.6)		74.6 <sup>a</sup> (4.6)			
Field of study (Microbiology and Microbiome)						
No					86.4 (2.8)	0.131
Yes					82.3 (3.4)	
Field of study (Other)						
No	58.7 <sup>b</sup> (2.8)	0.017				
Yes	63.1 <sup>a</sup> (3.2)					

<sup>1</sup>Effects were included in the model based on AIC.<sup>2</sup>OS<sub>PRE</sub>, overall score for perceived received education (PRE); OS<sub>PK</sub>, overall score for perceived knowledge (PK); OS<sub>CPSA</sub>, overall score for confidence in performing statistical analyses (CPSA).<sup>a-c</sup>Means lacking a common lowercase superscript were statistically different at  $P < 0.05$ .<sup>A,B</sup>Means lacking a common uppercase superscript were statistically different at  $P < 0.10$ .

## Factors associated with OS

The factors associated with the OS (i.e.,  $OS_{PRE}$ ,  $OS_{PK}$ , and  $OS_{CPSA}$ ) are presented in Table 1. These 3 OS were computed with the objective of summarizing the overall PRE, PK, and CPSA of participants in this study. Previous analyses showed moderate to high repeatability of the PRE and PK answers of statistical topics, indicating a clear consistency in how ASGS answered their questions. In fact, the marginal correlation estimates between these 3 OS were significant ( $P < 0.001$ ) and moderate, with 0.774 between  $OS_{PRE}$  and  $OS_{PK}$ , 0.645 between  $OS_{PRE}$  and  $OS_{CPSA}$ , and 0.795 between  $OS_{PK}$  and  $OS_{CPSA}$ , further indicating some consistency in their answers.

There were 15 effects selected for  $OS_{PRE}$ , 14 for  $OS_{PK}$ , and 12 for  $OS_{CPSA}$ , for a total of 16 different factors selected across analyses. The  $R^2$  of these models were low to moderate, with 0.297, 0.334, and 0.587, respectively. Of the selected factors, eight were included in all analyses, with *Age* ( $P \geq 0.197$ ) and *Graduate degrees completed* ( $P \geq 0.164$ ) not being significant in any of them. In contrast, three of the eight were significant for all three analyses. For *Comfortable in giving statistical advice in desired career path* ( $P \leq 0.001$ ) and *Comfortable in reviewing the stats section of a manuscript* ( $P \leq 0.004$ ), ASGS answering “Yes” and “Somewhat comfortable” had greater ( $P < 0.05$ )  $OS_{PRE}$ ,  $OS_{PK}$ , and  $OS_{CPSA}$  than those answering “No.” Albeit expected, this consistency in results across the OS is interesting and reaffirming, as both results are relevant from the standpoint of how ASGS have different interests and perceived competencies for statistics. It will be interesting to evaluate how these ASGS answering “Yes” for these questions perform in the follow-up study in which they were objectively tested on a range of statistical procedures.

Finally, the last significant factor across the 3 OS was *Stats software user* ( $P \leq 0.027$ ), in which “R and SAS” users had greater ( $P < 0.05$ )  $OS_{PRE}$ ,  $OS_{PK}$ , and  $OS_{CPSA}$  than those who do not use these software. There was no clear pattern across the OS when ASGS indicated being users of only one of the two software. Those who indicated being “SAS” users had lower ( $P < 0.05$ )  $OS_{PRE}$  and  $OS_{CPSA}$  than “R and SAS” users but no difference ( $P > 0.05$ )  $OS_{PK}$  than “R and SAS” users. Interestingly, there was no difference across all OS ( $P > 0.05$ ) between ASGS who use only “SAS” or “R”, and the only difference ( $P < 0.05$ ) between “Other” software users and “R” and “SAS” users was for  $OS_{CPSA}$ . Although no differences ( $P > 0.05$ ) between “R and SAS” users and users of one of these two software were identified for  $OS_{PRE}$  and  $OS_{PK}$ , it seems that ASGS who use both “R and SAS” might have an advantage, or at least perceive having an advantage, in statistics competency. In contrast, knowing just one software, regardless of which one, might not result in greater statistics competencies. Nonetheless, based on these data, it is clear how SAS and R are the most traditional software used by ASGS in the United States. Furthermore, although knowing statistical software is an important skill for data analysis, knowledge on the statistical concepts should be the focus in statistics courses instead of software (Caple, 1996; Moore, 1997; Ozgur et al., 2015).

There was no significant pattern for the other three effects included in all three analyses. Although not significant ( $P = 0.151$ ) for  $OS_{CPSA}$ , and trending for, there was a significant effect of *Additional training on statistics* on  $OS_{PRE}$  ( $P = 0.012$ ) and  $OS_{PK}$  ( $P = 0.047$ ), where ASGS answering “Yes” had greater OS than those answering “No.” These results align well with those from *Courses taken at the Department of Statistics*, in which ASGS who answered “Yes” had greater  $OS_{PRE}$  ( $P = 0.014$ ) and tended to have greater  $OS_{CPSA}$  ( $P = 0.052$ ) than those answering it with “No.” Although not significant ( $P = 0.168$ ), the same pattern was observed in

$OS_{PK}$ , as expected, given that these OS were moderately to highly correlated among each other. Additional training and taking courses taught by professors in the Department of Statistics could not only increase their exposure to additional statistical skills but also expose them to statistics methods in greater depth compared with courses taught by professors from other departments.

A similar consistent pattern was observed for the last factor included in all analyses. *Research Type*, which is based on the Carnegie classification of institutions of higher education (Indiana University Center for Postsecondary Research, 2018) in the United States, was significant for  $OS_{CPSA}$  ( $P = 0.027$ ) and had a tendency for  $OS_{PRE}$  ( $P = 0.099$ ). Nonetheless, in all three analyses, ASGS from “R1” institutions had lower OS than those from “Not R1” institutions. These results are unexpected; one would think that ASGS from “R1” institutions should have greater exposure to statistical training than those from “Not R1,” and, thus, the former should have greater OS. Although this result might be true, we must reiterate that this study is based on the perception of participants. Hence, it could be ASGS from “R1” institutions underestimate their training due to potentially having a broader idea of the breadth and depth of statistical methods. Nonetheless, the true reason behind this association is beyond the scope of this study. In this study, only 5.04% of participants were from “Not R1” institutions. This low number resulted in the large SE for “Not R1” means in the  $OS_{PRE}$  and  $OS_{PK}$  analyses, causing the lack of significant association ( $P \geq 0.099$ ) of *Research Type* with these OS.

A similar unexpected association was found for  $OS_{PRE}$ : MS ASGS indicated having greater ( $P = 0.007$ )  $OS_{PRE}$  than PhD ASGS. This could be an example of the *Dunning-Kruger Effect* (Kruger and Dunning, 1999), in which MS ASGS might have an inflated perception of their capabilities in statistics compared with PhD ASGS due to their limited exposure to the topic. In addition, the greater ( $P < 0.001$ ) number of *Credits in graduate-level courses on statistics* taken by PhD ASGS ( $9.60 \pm 0.46$ ) compared with MS ASGS ( $6.09 \pm 0.64$ ) could also support this hypothesis. This factor was associated ( $P \leq 0.02$ ) with both  $OS_{PRE}$  and  $OS_{PK}$ , and, as expected, the greater was the number of *Credits in graduate-level courses on statistics*, the greater were both OS. Although this factor was not selected in the analysis of  $OS_{CPSA}$ , a similar factor with a comparable pattern was selected. The greater was the number of *Graduate-level courses on statistics* taken by ASGS was  $OS_{CPSA}$  ( $P = 0.001$ ).

For the other factors selected in the analyses, a few associations were interesting too. ASGS who had *Previous professional experience* tended ( $P = 0.092$ ) to have lower  $OS_{PRE}$  than those without it. Two potential explanations exist for this association. First, similarly to what could be explaining other associations in this study, the limited experience of ASGS without *Previous professional experience* could have led them to overestimate how much they have learned in statistics courses. However, it could be that ASGS with *Previous professional experience* had learned some of the statistical methods from statistics courses during their professional experience, and, thus, they do not think their PRE was as high. Another associated factor related to having broader experience was *Degrees completed at other institutions*. ASGS with *Degrees completed at other institutions* had lower ( $P = 0.046$ )  $OS_{CPSA}$  than those without it. Although this association may not be a case of a true effect due to previous institutions offering worse training in statistics, ASGS who changed institutions could have a broader view of existing topics in statistics. Thus, ASGS who changed institutions might more accurately acknowledging their CPSA compared with those who have only been in the same institutions. Gives

statistical advices to lab peers tended ( $P = 0.06$ ) to have an effect on  $OS_{PK}$  and had a significant ( $P < 0.001$ ) effect on  $OS_{CPSA}$  ( $P < 0.001$ ). For both analyses, as expected, ASGS answering “Yes” had a greater OS than those answering “No.” Finally, five fields of study were selected in these analyses. However, only two had significant associations. ASGS who selected *Meat Science* had significantly greater ( $P = 0.029$ )  $OS_{PRE}$  and tended to have greater ( $P = 0.063$ )  $OS_{PK}$  than those from other areas. Interestingly (and ironically), ASGS who answered doing research in the field of study not listed or with low frequency in the survey (i.e., *Other*) had greater ( $P = 0.017$ )  $OS_{PRE}$  than those not from *Other* fields of study. The reasons why these areas of research were associated with these OS are hard to discuss. Overall, these results show how the overall PRE, PK, and CPSA of ASGS share a common and consistent set of factors explaining them.

### What to do about the current status of statistical training?

The USDA has recently indicated that there is a need to improve personnel education and training in data science to move the U.S. agricultural industry toward data-driven decisions (United States Department of Agriculture, 2018). The question then becomes: should ASGS be more heavily trained on animal data science and experimentation or should they continue taking the current number of courses on statistics? Certainly, this should be evaluated on a case-by-case basis.

The results from this study indicate that major statistical topics used in data generated by *high-throughput* technologies in *Big Data* are not covered in required statistical courses that ASGS take at U.S. institutions. For example, *Random Forest*, *Discriminant Analysis*, and *Cluster Analysis* are all traditional methods used in *Machine Learning*. These showed the second-, fourth-, and fifth-lowest average scores, respectively, whereas *Machine Learning* had the lowest average score in PRE (with similar results in PK). With the move of animal science research toward the generation of hundreds to thousands of data points, this is concerning. Another related concerns are the average PRE and PK scores of *Computer Coding*, and how much more confident ASGS were with *Data Management using Excel (or similar)* compared with *Data Management using stats software*. It is imperative to manage large using computer coding, due to its computational efficiency and more easily traceable human errors. But how to improve their training, knowledge, and confidence across these topics?

One way is by having ASGS taking more credits on statistics courses. This could mean that they would be taking fewer credits in classes in their area of specialization. The average  $\pm$  SD and median of *Graduate-level courses on statistics* were  $2.6 \pm 1.8$  and 2, respectively, and of *Credits in graduate-level courses on statistics* were  $8.3 \pm 7.1$  and 7, respectively. These values show that, on average, ASGS do not take many graduate courses on statistics, but there is a considerable variation in the number of courses and credits taken by ASGS. Another way is to engage ASGS to increase their independent study time on statistics outside the classroom. By spending more time working on their research data, and reading books and online resources, ASGS could further improve their knowledge in statistics. Nonetheless, it is important to note that, in many cases, statistics is simply a tool for these future researchers. In general, these ASGS are being trained in specific fields of study within Animal Sciences, which might include a major biological component in this training. Hence, the ability to connect statistics and the biological aspects of their training is much needed for their professional success.

But how much should ASGS know about statistics, experimental designs, and more? Close to graduation, especially for PhD degrees, ASGS should have at least a minimum level of statistical knowledge to properly read, write, and communicate research as well as to develop hypotheses and design simple experiments. However, are they being trained to achieve this minimum level? Also, how much should they be trained on statistics? To begin with, this study was based on the self-assessment of ASGS. Hence, their perception on all the questions included in this study must be evaluated objectively. This study is part of a larger project that included a follow-up study where 80 of the 416 participants in this study were subjected to two comprehensive exams. With this, we will not only obtain a better idea of the level of statistical knowledge of a range of statistical topics, but also how their perceptions align with their answers in the exams. This information will be useful to more accurately measure their levels of statistical training, critical statistical topics that need to be improved in their education, and to propose strategies to tackle these.

Second, ASGS are not, and should not, have the same level of training in statistics as graduate students in statistics. Although a minimum level of statistical knowledge is expected at graduation, it is important that they acknowledge any limitations in statistics and properly interact with data scientists and statisticians in their research projects. In other words, the authors do not think that all ASGS should be highly trained in statistics. It is more important that they know about their limitations in statistics and work together in a team of scientists that could provide the needed expertise in statistics. Hence, the amount of courses and credits in graduate-level statistics courses that ASGS should take depends on the needs, interest, and career goals of each one of them. With this, it would also be interesting to identify groups of ASGS showing contrasting perceptions (i.e., PRE, PK, and/or CPSA) and contrasting objective performance (i.e., based on the follow-up study). With this, we could identify the overall training and personal characteristics associated with ASGS with greater interest and performance in statistics. Along with the large variability in formal training in statistics from ASGS, information of these factors could be capitalized to propose training strategies for those ASGS to meet the current and future trend of data-driven animal science research.

### Conclusions

This was the first study to assess the current statistical training of ASGS in the United States to the best of our knowledge. Our results showed that U.S. ASGS perceive to have a greater education than knowledge for most of the statistical topics included in this study. However, the repeatability of answers within the same student was moderate to high across statistical topics, indicating a substantial correlation between their PRE and PK. These students mostly learn traditional statistical topics, whereas more complex methods commonly used in *Big Data*, such as *Machine Learning* and *Generalized Models*, are not covered or are limitedly discussed in statistics courses taken by these students. This resulted in students having greater CPSA using traditional statistical topics than more complex ones. The results from this study should be used to change the statistics curriculum of ASGS to better prepare them in statistical methods needed for them to be more proficient in the statistical techniques necessary to meet the current and future trend of data-driven animal science research.

## Supplementary Data

Supplementary data are available at *Journal of Animal Science* online.

## Acknowledgments

We would like to acknowledge Lizzy Trulson for her support in gathering all the data from participants, Christina Hamilton (North Central Regional Association Assistant Director and National Information Management & Support System Administrator) for sharing information on the survey with Heads and Chairs of Animal Science-related Departments in the United States, the Poultry Science Association (PSA) and American Dairy Science Association (ADSA) for sharing information of the survey with their members, and Dr. Jarrod Hadfield (University of Edinburgh) for his statistical advices using MCMCglmm. A.L.P. would like to express her appreciation to the United States Department of Agriculture National Institute of Food and Agriculture for the Food Research Initiative Competitive Grant (2016-38420-25496) in support of graduate studies.

## Conflict of interest statement

The authors declare no real or perceived conflicts of interest.

## Literature Cited

- Adeola, O., and C. Kong. 2020. Energy values of triticale or sorghum distillers' dried grains with solubles and rye fed to broiler chickens. *J. Anim. Sci.* **98**:skaa018. doi: [10.1093/jas/skaa018](https://doi.org/10.1093/jas/skaa018)
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **19**:716–723. doi: [10.1109/TAC.1974.1100705](https://doi.org/10.1109/TAC.1974.1100705)
- Akay, A., and G. Karabulut. 2020. Personality and positionality-evidence from survey experiments with alternative goods. *Eurasian Bus. Rev.* **10**:123–156. doi: [10.1007/s40821-020-00149-7](https://doi.org/10.1007/s40821-020-00149-7)
- Auguie, B., and A. Antonov. 2017. gridExtra: miscellaneous functions for "Grid" graphics. Available from <https://CRAN.R-project.org/package=gridExtra>. Accessed April 6, 2021.
- Auker, L. A., and E. L. Barthelmess. 2020. Teaching R in the undergraduate ecology classroom: approaches, lessons learned, and recommendations. *Ecosphere.* **11**:e03060. doi: [10.1002/ecs2.3060](https://doi.org/10.1002/ecs2.3060)
- Bowman, N. A., and T. A. Seifert. 2011. Can college students accurately assess what affects their learning and development? *J. Coll. Stud. Dev.* **52**:270–290. doi: [10.1353/csd.2011.0042](https://doi.org/10.1353/csd.2011.0042)
- Caple, C. 1996. The effects of spaced practice and spaced review on recall and retention using computer assisted instruction. Available from <https://eric.ed.gov/?id=ED427772>. Accessed April 6, 2021.
- Capps, K. M., R. G. Amachawadi, M. B. Menegat, J. C. Woodworth, K. Perryman, M. D. Tokach, S. S. Dritz, J. M. DeRouchey, R. D. Goodband, J. Bai, et al. 2020. Impact of added copper, alone or in combination with chlortetracycline, on growth performance and antimicrobial resistance of fecal enterococci of weaned piglets. *J. Anim. Sci.* **98**:skaa003. doi: [10.1093/jas/skaa003](https://doi.org/10.1093/jas/skaa003)
- Carmichael, J. W., and P. H. A. Sneath. 1969. Taxometric maps. *Syst. Biol.* **18**:402–415. doi: [10.2307/2412184](https://doi.org/10.2307/2412184)
- Chance, B., D. Ben-Zvi, J. Garfield, and E. Medina. 2007. The role of technology in improving student learning of statistics. *Technol. Innov. Stat. Educ.* **1**. Available from <https://escholarship.org/uc/item/8sd2t4rr>. Accessed April 6, 2021.
- Chitakasempornkul, K., M. W. Sanderson, E. Cha, D. G. Renter, A. Jager, and N. M. Bello. 2018. Accounting for data architecture on structural equation modeling of feedlot cattle performance. *J. Agric. Biol. Environ. Stat.* **23**:529–549. doi: [10.1007/s13253-018-0336-7](https://doi.org/10.1007/s13253-018-0336-7)
- Frenzel, A. C., R. Pekrun, and T. Goetz. 2007. Perceived learning environment and students' emotional experiences: a multilevel analysis of mathematics classrooms. *Learn. Instr.* **17**:478–493. doi: [10.1016/j.learninstruc.2007.09.001](https://doi.org/10.1016/j.learninstruc.2007.09.001)
- Gabel, D., and National Science Teachers Association, eds. 1994. *Handbook of research on science teaching and learning*. New York (NY): Macmillan; Toronto: Maxwell Macmillan Canada; New York (NY): Maxwell Macmillan International.
- Galili, T. 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**:3718–3720. doi: [10.1093/bioinformatics/btv428](https://doi.org/10.1093/bioinformatics/btv428)
- Goulart, R. S., R. A. M. Vieira, J. L. P. Daniel, R. C. Amaral, V. P. Santos, S. G. Toledo Filho, E. H. Cabezas-Garcia, L. O. Tedeschi, and L. G. Nussio. 2020. Effects of source and concentration of neutral detergent fiber from roughage in beef cattle diets on feed intake, ingestive behavior, and ruminal kinetics. *J. Anim. Sci.* **98**:skaa107. doi: [10.1093/jas/skaa107](https://doi.org/10.1093/jas/skaa107)
- Gutierrez, N. A., N. V. Serão, B. J. Kerr, R. T. Zijlstra, and J. F. Patience. 2014. Relationships among dietary fiber components and the digestibility of energy, dietary fiber, and amino acids and energy content of nine corn coproducts fed to growing pigs. *J. Anim. Sci.* **92**:4505–4517. doi: [10.2527/jas.2013-7265](https://doi.org/10.2527/jas.2013-7265)
- Hastie, T., R. Tibshirani, and J. H. Friedman. 2017. *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. (corrected at 12th printing 2017). New York (NY): Springer.
- Hong, J., S. P. Ndou, S. Adams, J. Scaria, and T. A. Woyengo. 2020. Canola meal in nursery pig diets: growth performance and gut health. *J. Anim. Sci.* **98**:skaa338. doi: [10.1093/jas/skaa338](https://doi.org/10.1093/jas/skaa338)
- Indiana University Center for Postsecondary Research. 2018. Bloomington (IN): The Carnegie Classification of Institutions of Higher Education. Available from <http://carnegieclassifications.iu.edu/downloads/CCIHE2018-FactsFigures.pdf>. Accessed April 6, 2021.
- Israel, G. D., and C. L. Taylor. 1990. Can response order bias evaluations? *Eval. Program Plann.* **13**:365–371. doi: [10.1016/0149-7189\(90\)90021-N](https://doi.org/10.1016/0149-7189(90)90021-N)
- Jang, K. B., J. H. Kim, J. M. Purvis, J. Chen, P. Ren, M. Vazquez-Anon, and S. W. Kim. 2020. Effects of mineral methionine hydroxy analog chelate in sow diets on epigenetic modification and growth of progeny. *J. Anim. Sci.* **98**:skaa271. doi: [10.1093/jas/skaa271](https://doi.org/10.1093/jas/skaa271)
- Jaques, D. 2003. Teaching small groups. *Bmj* **326**:492–494. doi: [10.1136/bmj.326.7387.492](https://doi.org/10.1136/bmj.326.7387.492)
- Johnson, T. P. 2014. Snowball sampling: introduction. In: Wiley statsref: statistics reference online. American Cancer Society. Available from <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat05720>. Accessed April 6, 2021.
- Kassambara, A., and F. Mundt. 2020. factoextra: extract and visualize the results of multivariate data analyses. Available from <https://CRAN.R-project.org/package=factoextra>. Accessed April 6, 2021.
- Kaufman, L., and P. J. Rousseeuw, eds. 1990. *Finding groups in data*. Hoboken (NJ): John Wiley & Sons, Inc. Available from <http://doi.wiley.com/10.1002/9780470316801>. Accessed April 6, 2021.
- Kruger, J., and D. Dunning. 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* **77**:1121–1134. doi: [10.1037//0022-3514.77.6.1121](https://doi.org/10.1037//0022-3514.77.6.1121)
- Li, M., S. Tu, Z. Li, F. Tan, J. Liu, Q. Wang, Y. Zhang, J. Xu, Y. Zhang, F. Zhou, et al. 2019. MAP: model-based analysis of proteomic data to detect proteins with significant abundance changes. *Cell Discov.* **5**:40. doi: [10.1038/s41421-019-0107-9](https://doi.org/10.1038/s41421-019-0107-9)
- Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, M. Studer, P. Roudier, J. Gonzalez, K. Kozłowski, E. Schubert, et al. 2019. Finding groups in data: cluster analysis extended rousseeuw et al. Available from <https://CRAN.Rproject.org/package=cluster>. Accessed April 6, 2021.

- Marina, H., A. Reverter, B. Gutiérrez-Gil, P. A. Alexandre, R. Pelayo, A. Suárez-Vega, C. Esteban-Blanco, and J. J. Arranz. 2020. A multiple-phenotype imputation procedure as a method for prediction of cheese-making efficiency in Spanish Assaf sheep. *J. Anim. Sci.* **98**:skaa370. doi: [10.1093/jas/skaa370](https://doi.org/10.1093/jas/skaa370)
- Moore, D. S. 1997. New pedagogy and new content: the case of statistics. *Int. Stat. Rev.* **65**:123–165. doi:[10.1111/j.1751-5823.1997.tb00390.x](https://doi.org/10.1111/j.1751-5823.1997.tb00390.x)
- Morota, G., R. V. Ventura, F. F. Silva, M. Koyama, and S. C. Fernando. 2018. Big data analytics and precision animal agriculture symposium: machine learning and data mining advance predictive big data analysis in precision animal agriculture. *J. Anim. Sci.* **96**:1540–1550. doi: [10.1093/jas/sky014](https://doi.org/10.1093/jas/sky014)
- Muenchen, A. 2012. The popularity of data science software. r4stats.com. Available from <http://r4stats.com/articles/popularity/>. Accessed April 6, 2021.
- Ozgur, C., M. Kleckner, and Y. Li. 2015. Selection of statistical software for solving big data problems: a guide for businesses, students, and universities. *SAGE Open*. **5**:215824401558437. doi: [10.1177/2158244015584379](https://doi.org/10.1177/2158244015584379)
- Panko, R. 2016. What we don't know about spreadsheet errors today: the facts, why we don't believe them, and what we need to do. ArXiv160202601 Cs. Available from <http://arxiv.org/abs/1602.02601>. Accessed April 6, 2021.
- R Core Team. 2017. *R: a language and environment for statistical computing*. Vienna (Austria): R Foundation for Statistical Computing. Available from <https://www.R-project.org/>. Accessed April 6, 2021.
- Ran, T., P. Jiao, O. AlZahal, X. Xie, K. A. Beauchemin, D. Niu, and W. Yang. 2020. Fecal bacterial community of finishing beef steers fed ruminally protected and non-protected active dried yeast. *J. Anim. Sci.* **98**:skaa058. doi: [10.1093/jas/skaa058](https://doi.org/10.1093/jas/skaa058)
- Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**:53–65. doi: [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- RStudio Team. 2020. *RStudio: integrated development for R*. Boston (MA): RStudio, PBC. Available from <http://www.rstudio.com/>. Accessed April 6, 2021.
- Sanglard, L. P., P. Canada, B. E. Mote, P. Willson, J. C. S. Harding, G. S. Plastow, J. C. M. Dekkers, and N. V. L. Serão. 2020. Genomic analysis of igg antibody response to common pathogens in commercial sows in health-challenged herds. *Front. Genet.* **11**:593804. doi: [10.3389/fgene.2020.593804](https://doi.org/10.3389/fgene.2020.593804)
- Sanglard, L. P., S. Schmitz-Esser, K. A. Gray, D. C. L. Linhares, C. J. Yeoman, J. C. M. Dekkers, M. C. Niederwerder, and N. V. L. Serão. 2020. Vaginal microbiota diverges in sows with low and high reproductive performance after porcine reproductive and respiratory syndrome vaccination. *Sci. Rep.* **10**:3046. doi: [10.1038/s41598-020-59955-8](https://doi.org/10.1038/s41598-020-59955-8)
- Scanlan, C. L., A. M. Putz, K. A. Gray, and N. V. L. Serão. 2019. Genetic analysis of reproductive performance in sows during porcine reproductive and respiratory syndrome (PRRS) and porcine epidemic diarrhea (PED) outbreaks. *J. Anim. Sci. Biotechnol.* **10**:22. doi: [10.1186/s40104-019-0330-0](https://doi.org/10.1186/s40104-019-0330-0)
- Sebastianelli, R., C. Swift, and N. Tamimi. 2015. Factors affecting perceived learning, satisfaction, and quality in the online MBA: a structural equation modeling approach. *J. Educ. Bus.* **90**:296–305. doi: [10.1080/08832323.2015.1038979](https://doi.org/10.1080/08832323.2015.1038979)
- See, G. M., B. E. Mote, and M. L. Spangler. 2020. Impact of inclusion rates of crossbred phenotypes and genotypes in nucleus selection programs. *J. Anim. Sci.* **98**:skaa360. doi: [10.1093/jas/skaa360](https://doi.org/10.1093/jas/skaa360)
- Sterndale, S. O., D. W. Miller, J. P. Mansfield, J. C. Kim, and J. R. Pluske. 2020. Increasing dietary tryptophan in conjunction with decreasing other large neutral amino acids increases weight gain and feed intake in weaner pigs regardless of experimental infection with enterotoxigenic *Escherichia coli*. *J. Anim. Sci.* **98**:skaa190. doi: [10.1093/jas/skaa190](https://doi.org/10.1093/jas/skaa190)
- United States Department of Agriculture. 2018. USDA Strategic Plan FY 2018–2022. 64. Available from <https://www.usda.gov/sites/default/files/documents/usda-strategic-plan-2018-2022.pdf>. Accessed April 6, 2021.
- Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemond, A. Hayes, L. Henry, J. Hester, et al. 2019. Welcome to the Tidyverse. *J. Open Source Softw.* **4**:1686. doi: [10.21105/joss.01686](https://doi.org/10.21105/joss.01686)