

Integrating Writing into the Introductory Statistics Course

KB BOOMER

BUCKNELL UNIVERSITY

Writing Across the Curriculum (WAC) - As old as the modern American university

- o Land grant universities (Morrill Act 1862):
 - Shift from classical to vocational education
 - Discipline-specific language develops
- o From writing composition to writing curricula: 1900s – 1970s
 - Writing as a unifier to ensure all students had basic writing skills
 - Instruction came from the English departments
 - Early development of WAC
- o WAC as a pedagogical approach – 1980's and beyond
 - Students discover a deeper understanding of the discipline through writing
 - Writing isn't only taught in the English department
 - Different types of writing: Writing-to-learn, writing-to-communicate

Student writing and the GAISE recommendations

0 2016 Recommendations

1. Teach statistical thinking “...the development of skills that will allow students to think critically about statistics”
2. Focus on conceptual understanding
3. Integrate real data with a context and a purpose
4. Foster active learning The act of writing helps to cement ideas explored in active learning
5. Use technology to explore concepts and analyze data
6. Use assessments to improve and evaluate student learning
“Expand the traditional exam-based view of assessment”

Potential roadblocks

- o I'm not an English professor, how can I teach writing?
 - We do know how to use data to answer real-world questions
 - o Discipline-specific writing
 - Start with short, low-stakes assignments
 - Use available resources, e.g. campus writing centers
- o Oh, the grading!!
 - Rubrics, rubrics, rubrics!
- o There is no time in the course to add in extra pieces
 - Use writing to support what is currently done, not as a new task

Introductory Statistics at Bucknell

- o Four meetings per week
 - One meeting in a computer lab
 - 24 – 28 students per class
- o Topics
 - Exploratory data analysis
 - Inference for one & two means, one & two proportions
 - ANOVA, Regression, Chi-square
- o Grading
 - 25% formative assessment (homework, labs, quizzes, low stakes writing, clicker questions)
 - 60% exams: two midterms and a final
 - 15% semester group project

Semester group project

o Typically 4-5 students per group

o Potential Topics (Week 5)

- Incorporate an external source
- First draft of the project introduction

Variable Type	Number of Variables	
	Quantitative	Categorical
Response	1	1
Independent	1 (2)	1 (2)

o Exploratory Data Analysis Report (Week 10)

- Revised Potential Topics assignment
- Univariate EDA analysis on 4 variables
- Bivariate EDA analysis on three relationships (ANOVA, Regression, χ^2)

o Inference report & class presentation (Week 15)

- Revised Exploratory Data Analysis Report
- Revised lab reports
- Conclusions

Writing requirement at Bucknell

- o One first-year seminar course and two discipline-specific writing courses
- o The discipline-specific writing course
 - Requirements:
 1. Provide writing instruction
 2. Support the writing process
 3. Teach the techniques of writing needed in the discipline
 4. Require frequent writing
 5. Use writing to teach the subject matter

2. Support the writing process

- o In other words, have students revise drafts
 - o In lab, 2-3 students complete specific analyses; I provide written feedback; project-groups (4-5 students) compare results, revise, and re-submit in EDA / inference report
 - Lab 4: EDA on a common PEW dataset
 - Lab 7: EDA on project dataset
-
- Lab 8: Two-sample t tests
 - Lab 9: Chi-square test
 - Lab 10: Two-factor ANOVA
 - Lab 11: Multiple regression

Groups now have 2 explanatory factors and 2 quantitative predictors

1. Provide writing instruction &

3. Teach the techniques of writing needed in the discipline

- o We model writing in the discipline in our teaching (e.g. interpret the results of a inferential procedure in context)
- o Focus: Interpretation – minus jargon
- o Approach #1: Provide support materials
 - Project instructions give examples of clunky writing
 - o The response variable, number of cigarettes smoked per week, differed depending on whether kids answered yes or no to the question of ‘Is alcohol available in your house?’
 - Examples of better writing
 - o The first objective is to assess whether adolescents with access to alcohol at home smoke, on average, more than those without such access.

Approach #2: In-class activity: Compare writing samples

- o Worksheet #1: Introductory paragraphs
- o Worksheet #2: Describe the distribution of a variable
- o Each project group divides into subgroups
- o One worksheet per subgroup
- o Read sample student response, evaluate within a subgroup
- o Come back together as a group and discuss writing samples

Approach #2: In-class activity:

Compare writing samples

- 0 **Writing Sample #1:** The first variable is a categorical explanatory variable of the number of days a person ate breakfast in the last week. The second variable is a quantitative response variable BMI. A potential problem that could arise is that there is skewness with the BMI values. The mean of the quantitative variable, BMI, is 22.03 and the standard deviation is 4.672. The graph is right skewed so the median is a better representation of the center of the graph and the IQR better represents the spread. The median of the quantitative variable is 21.819 and the IQR is 5.102.

Rating (1: strongly disagree, 5: strongly agree)					
	1	2	3	4	5
Fully discusses the distribution of the variable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Concise yet informative	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Free from technical errors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Clear writing style, minimizes jargon	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Overall, this is a well-written paragraph	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. Require frequent writing

5. Use writing to teach subject matter

- ◊ Writing-to-learn activities - ask students to put course content into their own words
- ◊ Informal, low-stakes work
- ◊ Based on textbook readings:

Entrance slip....

Name: _____

Why is $\mu_1 = \mu_2$ a reasonable null hypothesis for the two sample t-test?

4. Require frequent writing
 5. Use writing to teach subject matter
-

o Integrated into homework assignments

Statistics in the News: Read either article below and summarize the sampling bias in one well-written paragraph.

From the notes: In calculating the sample size needed to achieve a specific margin of error for inference on a true proportion, why do we use $p^*=0.5$ if no value of p^* is known? Hint: plot the function $f(p) = p(1-p)$

Using writing to teach subject matter: The Dolphin Activity (Rossman & Chance)

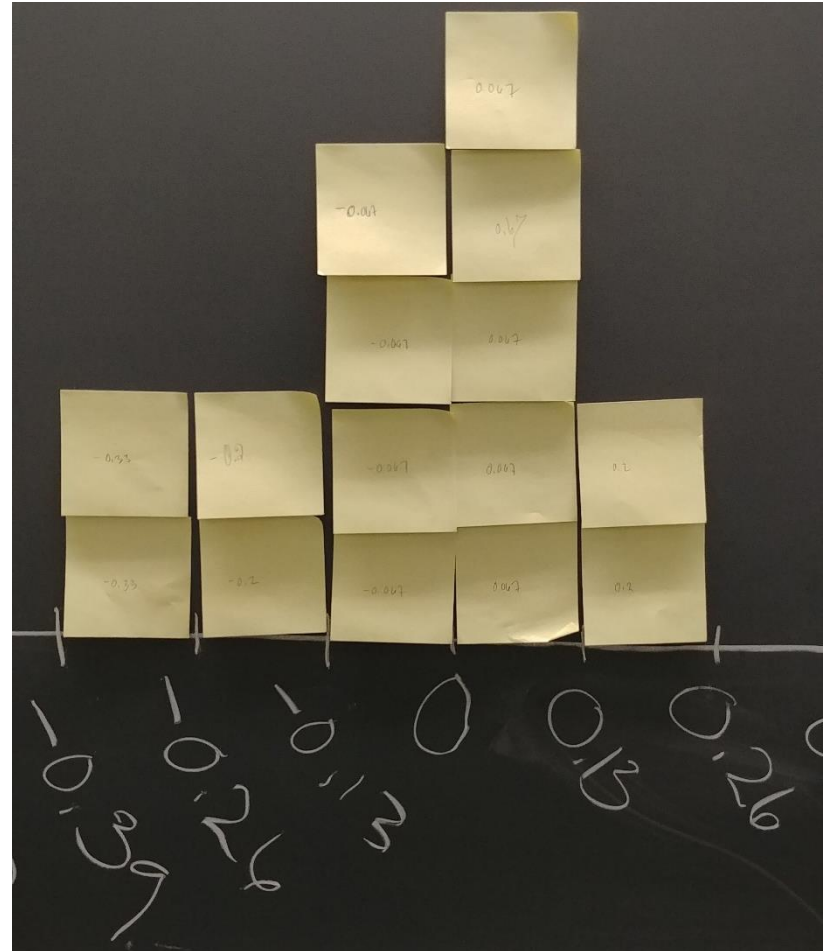
- Background: Thirty individuals with mild to moderate depression flown to Honduras. Random assignment to two treatment groups: (1) dolphin care program and (2) outdoor nature program. Response: Change in depression score.
- Study results:

Improvement?	Dolphin Therapy	Control Group	Total
Subject improved	10	3	13
Subject did not improve	5	12	17
Total	15	15	30

Observed difference in the proportion of improvers: $0.667 - 0.2 = 0.467$

The Dolphin Activity

- Classroom activity: Given a deck of 13 red cards (improvers) and 17 black cards (non-improvers), randomly assign to treatment groups. What is the difference in the proportion who improve?
- Rossman – Chance applet
- <http://www.rossmanchance.com/applets/ChiSqShuffle.html?dolphins=1>



Rossman/Chance Applet Collection

Show Shuffle Options ☒

Number of Shuffles

☒ Cards ☐ Data ☐ Plot

Most Recent Shuffle

Group A

Success

6



Failure

9



Group B

Success

7

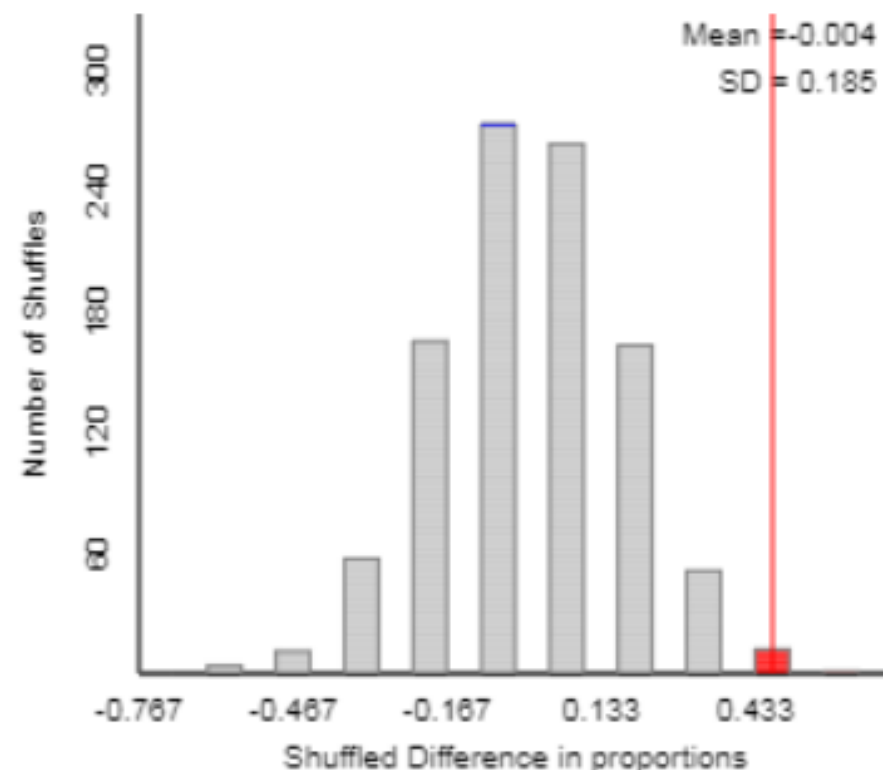


Failure

8



Total Shuffles = 1000



Count Samples

Ideas when thinking about prompts

◌ Bloom's taxonomy

- Remember
- Understand
- Apply
- Analyze
- Evaluate
- Create

Ideas when thinking about prompts

◉ Bloom's taxonomy

- Remember
- Understand Explain ideas or concepts:
Describe, discuss, explain, identify, predict
- Apply Use information in new situations:
Solve a different problem, demonstrate, sketch
- Analyze Draw connections among ideas:
Compare & contrast, organize, examine
- Evaluate
- Create

Dates on Pennies – Sampling Distributions

- Students obtain two random (?) samples, of 5 pennies, and of 25 pennies.
- Tasks:
 - Calculate sample mean and sample standard deviation
 - Plot a sample histogram; describe shape, center, and spread
 - Write the mean on a post-it and plot on histogram on the blackboard
 - “Make a prediction” questions



Dates on Pennies – Sampling Distributions

Make a prediction:

Do you think the shape of the *distribution of the sample means* will be:

- | | | | | | |
|----|-------------|-----|-----------|------|--------------|
| i. | Left-skewed | ii. | Symmetric | iii. | Right-skewed |
|----|-------------|-----|-----------|------|--------------|

How do you think the spread of the *distribution of the sample means* will compare to your *sample histogram*?

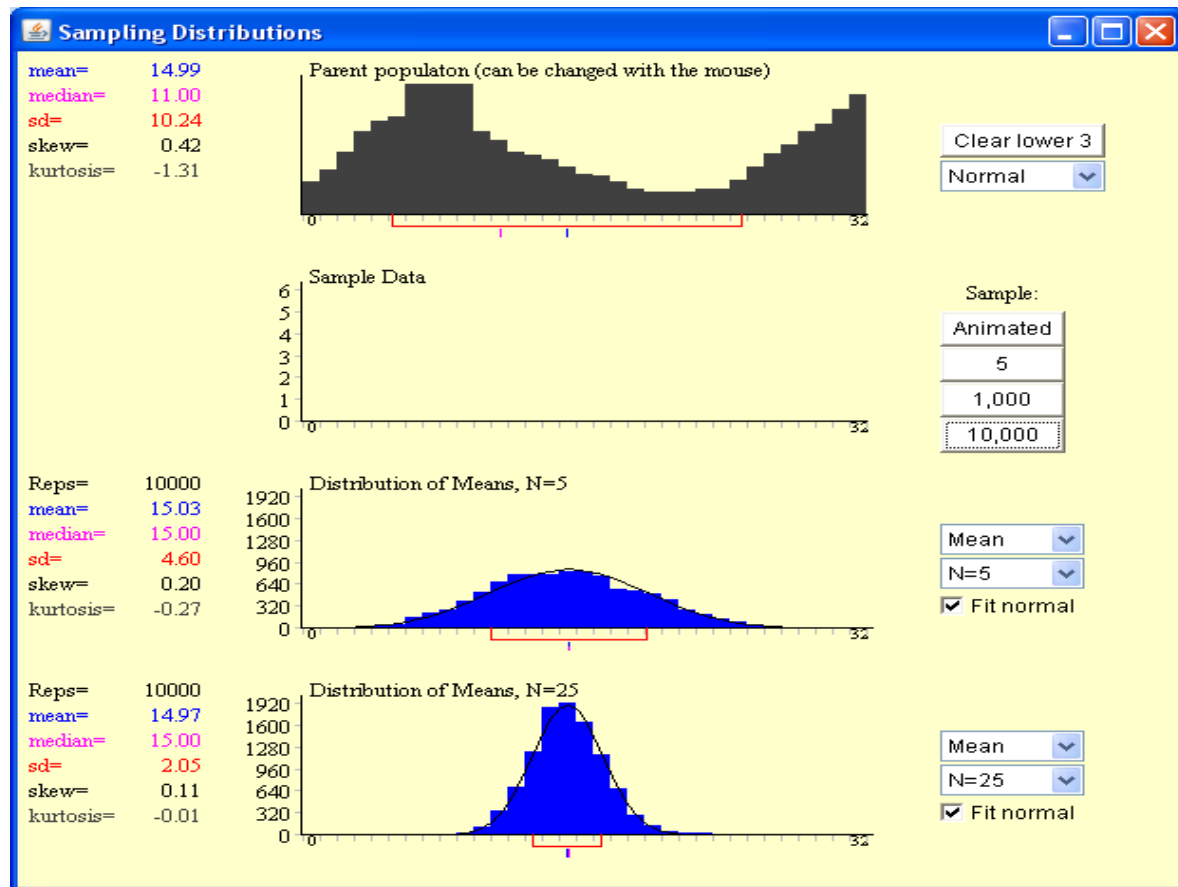
- | | | | | | |
|----|-------------|-----|-----------------------------|------|-------------|
| i. | Less varied | ii. | About the same
variation | iii. | More varied |
|----|-------------|-----|-----------------------------|------|-------------|

How do you think the center of the *distribution of the sample means* will compare to your *sample histogram*?

- | | | | | | |
|----|-------------------|-----|-------------------------------|------|--------------------|
| i. | It will be lower. | ii. | It will be about
the same. | iii. | It will be higher. |
|----|-------------------|-----|-------------------------------|------|--------------------|

Dates on Pennies – Sampling Distributions

0 http://onlinestatbook.com/stat_sim/sampling_distribution/index.html



Ideas when thinking about prompts

◉ Bloom's taxonomy

- Remember
- Understand Explain ideas or concepts:
Describe, discuss, explain, identify, predict
- Apply Use information in new situations:
Solve a different problem, demonstrate, sketch
- Analyze Draw connections among ideas:
Compare & contrast, organize, examine
- Evaluate
- Create

Rubrics – from simple ...

Two Sample T Test Lab Rubric - 8 pts; each item is worth 1 point

- Format (no steps, fluid variable descriptions)
- Define groups
- Check conditions (need to demonstrate that they know why they are looking at the two histograms)
- Include histograms
- Write the form of the alternative hypothesis in words (not notation)
- Include MTB output
- Report mean, sample size, and standard error for each group
- Report test statistic, df, and p-value

.. To more complex

	Max Points	Points Earned	Comments
Introduction (21 pts = 30%)			
Describe research question, citation	3		
Motivation: Why of interest? Who benefits?	2		
Describe the study & observational units	2		
Describe population to which results apply	1		
Introduce four variables	4		
Introduce X-quantitative vs Y-categorical	2		
Introduce X-categorical vs Y-quantitative	2		
Introduce X-categorical vs Y-categorical	2		
Clearly written, incorporates previous feedback	3		

Strategies to help students improve their writing

- o Include a variety of forms
- o Base feedback on communication not writing mechanics
- o Encourage peer review within groups
- o Encourage the use of campus writing centers
- o Have them read it out loud to a roommate

Response: Drug risk; Explanatory: physical activity, parental affection

Adolescents received a one for every question they responded “yes” to and a 0 for every question they responded “no” to. The point totals for all six questions were summed for each individual to create a cumulative drug risk score. This score will be used to examine the relationship between physical activity and perceived parental affection.

Adolescent's risk of drug use is measured using the drug risk score which counts the six risk factors for drug abuse. This includes whether the teen has smoked 5 or more cigarettes in the last 30 days, whether they got drunk more than twice last month, whether they used marijuana more than twice in the last year, whether they used cocaine, inhalants and other drugs more than twice in the last year, and if they used drugs with a needle in the last year. Adolescents received a one for every question they responded “yes” to and a 0 for every question they responded “no” to. The point totals for all six questions were summed for each individual to create a cumulative drug risk score. This score will be used to examine the relationship between physical activity and perceived parental affection. Both perceived parental affection and physical activity have been identified as protective factors for drug use by *The National Institute on Drug Abuse*. Comparing these factors with adolescents drug risk scores will gauge the effectiveness of these factors as protective factors.

awk. Sounds like
relationship is
between physical
activity + affection

awk. Sounds like
relationship is
between physical
activity + affection

Reporting descriptive statistics

Descriptive statistics for adolescents feelings towards mothers support:

The level of mothers support for the 4832 adolescents sampled has a mean of 4.2 with a standard deviation of 0.67. It is important to note that the standard deviation is relatively small meaning that the levels of mother's support are closely spread in this study. The level of mother's support has a minimum of 1.0 and a maximum of 5.0 meaning that the sample population in this investigation had a range of 5.0. With a Q1 of 3.9, a Q3 of 4.7 and a median of 4.1 it is evident that 50 percent of the adolescents surveyed feel somewhat supported by their mothers.

incorrect
at
write at - writing
See - writing
guidelines
that demonstrates
how to put these in
context

(after statement of a significant chi-square result). Within private-for-profit schools, less schools award financial aid to less than 25% of student and between 25% and 50% of students than is expected. However, more private-for-profit schools award financial aid to between 50% and 75% of students and above 75% of students than is expected.

Approach #2: In-class Activity: Compare Writing Samples

- o Activity: Give students examples of student writing. Ask them to evaluate and compare writing. Encourage discussion of how these examples can help them improve their writing.
- o Questions:
 - How would a student approach this task?
 - What concerns are likely easiest for students to identify?
Most difficulty?

Student challenges:

Interpretation in context

The Bonferroni method shows no significant difference between the true mean retention rate of institutions with greater than 60% women and 40-60% women. However, it shows that there is significant difference between institutions with less than 40% women and both institutions greater than 60% women and 40-60% women. The true mean retention rate of institutions with less than 40% women were almost 4 percentage points less than those with more than 40% women. Therefore, schools that have a lower percentage of female students appear to have lower retention rates than schools that have larger percentages of women enrolled.

Percentage of Female Undergraduates	N	Mean	Grouping
>60% women	948	0.727519	A
40%-60% women	828	0.713216	A
<40% women	334	0.673559	B

Table 2. Bonferroni Method Results for the Percentage of Females Enrolled test