Motivating Learning and Measuring Outcomes

Bob Stine
Department of Statistics, Wharton

Wharton
Department of Statistics

Issues

What can we do to encourage student learning?

How do we measure what they have learned?

Issues

What can we do to encourage student learning?

- Hands-on examples, experiential
- Exciting, real world cases
- Engaging classroom discussion
- Student initiated projects

How do we measure what they have learned?

- Weekly, in-class quizzes
- Multiple choice exams
- Projects, assignments, class participation

Issues

What can we do to encourage student learning?

Hands-on examples, experiential

Exciting, real world cases

Engaging classroom discussion

Student initiated projects

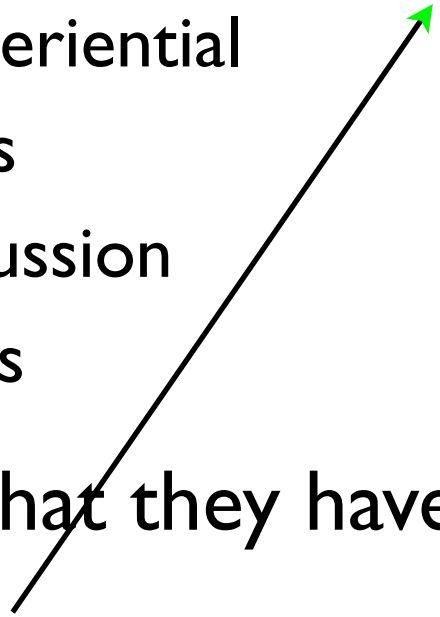
How do we measure what they have learned?

- Weekly, in-class quizzes
- Multiple choice exams

Projects, assignments, class participation

Context

Big classes (800 in 12 sections)



Weekly In Class Quizzes

Weekly Quizzes

In class

- Simple way to enforce attendance
- Forces you to keep them short
- Ten minutes at start of class

Format

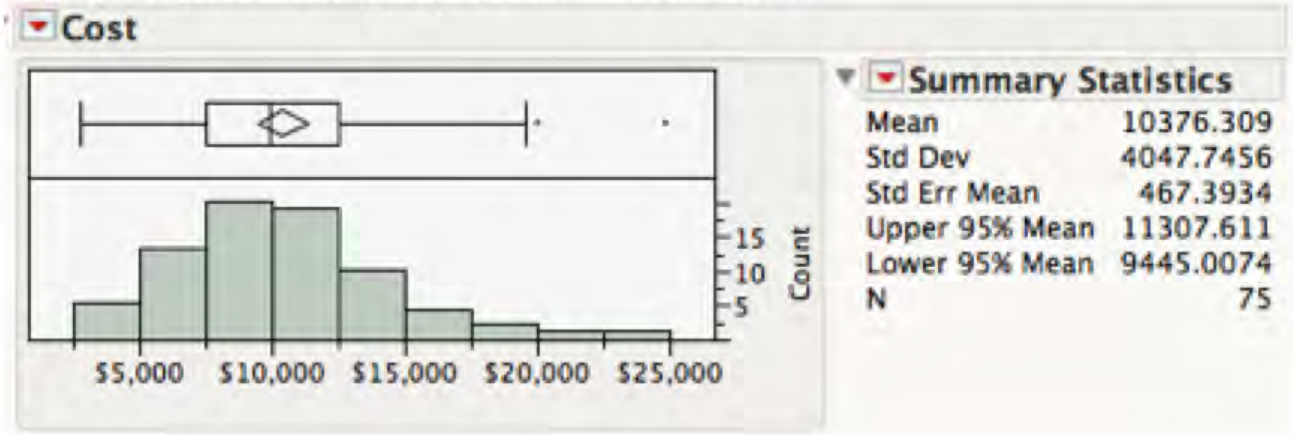
- 'Two sentence' introduction
- Five multiple choice questions in style of exam

Outcomes

- Keep up: squeeky wheel, immediate feedback
- "I was ready for the exam without cramming"
- "Thanks for making me keep up"

Partial Example

A large company computed the costs of providing health-care benefits to a sample of 75 of its employees. The following output summarizes these data.



Little calculation

Conceptual

Classes average
3-4 right per quiz

(1) The procedure used to build the shown 95% confidence interval for the mean μ
- (a) Should not be used because the sample of costs is right-skewed.
- (b) Is designed to produce intervals that hold 95% of sample averages.
- (c) Is designed to produce intervals that hold the population mean for 95% of samples.
- (d) Produces intervals that contain the sample mean with probability 0.95.

(2) If the sample size were doubled from $n = 75$ to $n = 150$, then the 95% confidence interval based on the sample with $n = 150$ would
- (a) Be more likely to contain μ than the shown interval for $n = 75$.
- (b) Be about half the width of the shown interval for $n = 75$.
- (c) Be about 71% of the width of the shown interval for $n = 75$.
- (d) Be less likely to contain μ than the shown interval for $n = 75$.

(3) Industry standards suggest that the cost for this benefit should average $10,000 per employee. These results show that
- (a) This sample of data is biased and should be discarded.
- (b) Costs at this company are statistically significantly higher on average.
- (c) Costs at this company are comparable on average to those in the industry.
- (d) Costs at this company are statistically significantly lower on average.

+ two more...

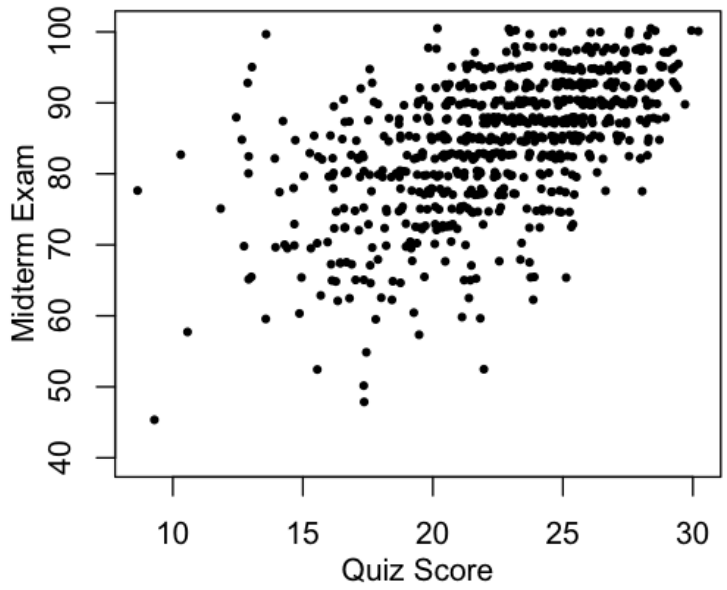
Results

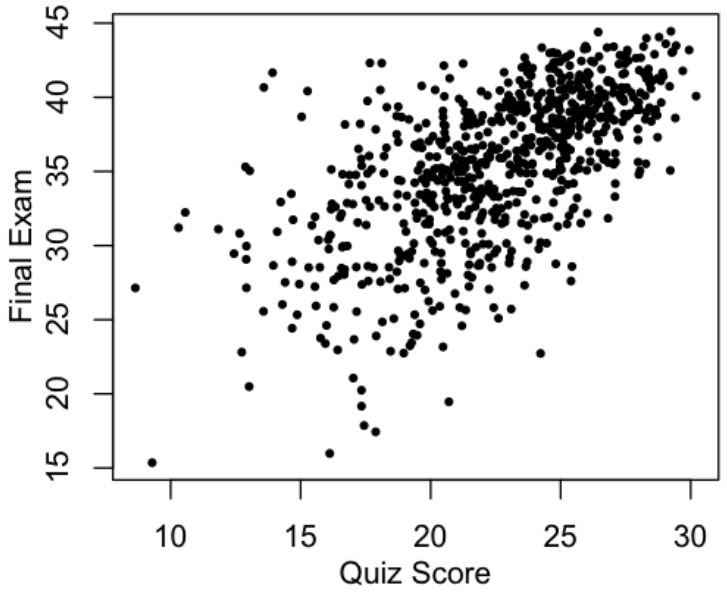
Count 20-25% of final grade

- Enough to be worthwhile, without panic
- Cumulative counts as if another exam

Predictive of exams

- Correlation ≈ 0.6 with exams, same as correlation between midterm and final exams





Multiple Choice Exams

Multiple Choice?

Where would you rather spend time?

- Writing questions or grading

Lessons from prior exams

- Consistency/reliability hard to maintain
- Great answers to wrong questions
- Partial credit: standards and conversations
- Importance of writing standards: Teach that too?

DeVeaux paradox

Switch from open-ended questions

- Inadvertent experiments on regrading tests
- Mix of half essay and half multiple choice

Multiple Choice?

Provides data...

Examples of Questions

6. If the variable X_2 is added to a simple regression that includes X_1, then which of the following must happen in the multiple regression if X_1 and X_2 are uncorrelated?
   a. RMSE must get smaller.
   b. The partial regression coefficient for X_2 will be smaller than that for X_1.
   c. The overall ANOVA F-statistic will have a larger p-value.
   d. The partial regression coefficient for X_1 will be the same as the marginal coefficient.
   e. The overall ANOVA F-statistic will have a smaller p-value.

General knowledge questions

Relate to data analysis skills

Performance

Corr(Ques Right, Course Total) = 0.34

69% correct

Examples of Questions

9. If the t-ratio for a slope in a multiple regression model is -2.6 then
   a. The p-value for the slope is greater than 0.05.
   b. The 95% confidence interval for the slope will contain zero.
   c. The standard error for this coefficient is negative.
   d. Removing this variable will produce a significant decrease in R^2.
   e. This estimated slope is less than zero, but not significantly so.

General knowledge questions

Relate to data analysis skills

Performance

Corr(Ques Right, Course Total) = 0.39

87% correct

Examples of Questions

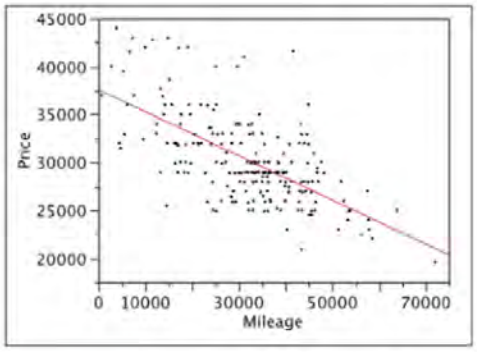
Mini-case business context

Avoid ambiguous language: 6 readers

Detail varies from sparse to paragraphs

Project "reward" question

The data shown in the following scatterplot and simple regression report the price and the number of miles driven for 218 used cars (all in the BMW 325 series) offered for sale in the San Francisco area.



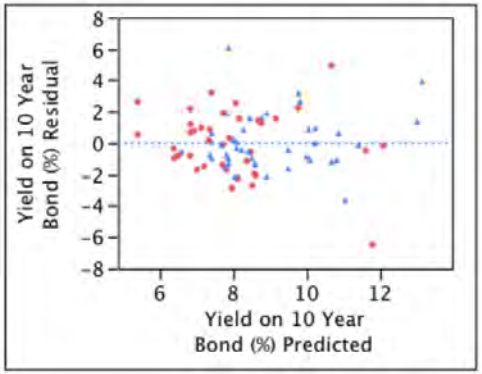
Price ($) = 37650 - 0.24 Mileage

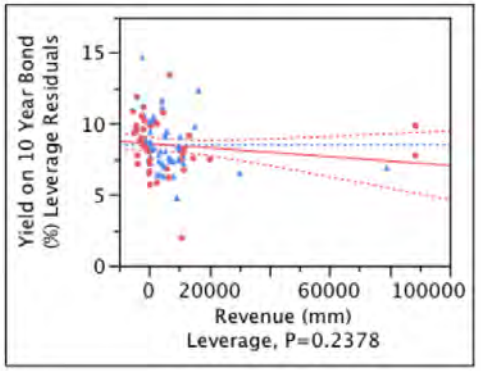
RSquare	0.42
Root Mean Square Error	3500
Mean of Response	30269
Observations	218

A banking analyst was interested in predicting the yield on a ten-year bond for a new issue from a company in one of two South American countries (Argentina or Brazil). The analyst collected data on 85 previous bond issues from different companies in the region. Output on the next page shows a model that includes a country effect, the annual revenue of the bond-issuing company in millions of $US in the previous year and a measure of the financial leverage of the company (calculated as the ratio of total liabilities to net worth). Note that the variable "Financial Leverage" has nothing to do whatsoever with the concept of statistical leverage.

Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	7.4679013	0.3939733	18.96	<.0001*
Revenue (mm)	-0.0000153	0.0000129	-1.19	0.2378
Financial Leverage	0.7118301	0.1119102	6.36	<.0001*
Country[Argentina]	-1.0083861	0.4233810	-2.38	0.0196*





Wharton
Department of Statistics

Examples of Questions

12. According to the fitted equation, cars like these with 40,000 miles would be expected to cost on average

    a. $37,650.
    b. $30,450.
    c. $28,050.
    d. $36,690.
    e. $25, 050.

Don't want many questions that are too easy

But you need some!

Be careful that exam is not too hard!

Performance

Corr(Ques Right, Course Total) = 0.07

98% correct

Examples of Questions

24. With reference to the leverage plot for Revenue, which of the following is a reasonable conclusion?
    a. There is severe collinearity in this dataset.
    b. Autocorrelation is likely.
    c. There are no leveraged outliers in the dataset.
    d. Leveraged observations produce a significant effect for Revenue in the model.
    e. Leveraged observations reduce the standard error of Revenue in the model.

Don't want many questions that are too hard

No partial credit for these…

Not good to have average exam score 40/100

Performance

Corr(Ques Right, Course Total) = 0.25

38% correct

Information in Answers

Regression on correct/incorrect answers for 10 questions has

$R^2 \approx 70\%$ for course total!

- Final with 44 questions is 25% of course total
- Don't need a lot of questions (we use about 40)

Which questions?

- Categorical variables in regression (corr = 0.43)
- The two free-form questions we saw
- Project related mult regr questions
- End of semester questions
- Correct answers, not distractors

Other Benefits

Find mistakes in exam

- Questions with low correlation to overall score often have an error or ambiguity.

Learn what you did not teach well

- Either that or you like to keep secrets

Compare instructors

- "Inversions" in question difficulty vary over sections implies some teach differently
- Touchy issue

Closing Remarks

Exam needs to carry on spirit of course

- We emphasize interpretation over computation.
- Exam must share this emphasis.

Takes a lot of time to write exam

- Good questions carry a lot of signal

Track teaching outcomes over time

Cheating

Thanks!