Assessment Strategies in Statistics

Beth L. Chance
Cal Poly – San Luis Obispo
bchance@calpoly.edu
Overview

- Dimensions of assessment
- Goals of assessment
- Assessment strategies
- Alternative assessments
- Examples
“Assessment”

• Evaluation of program
  – Capstone experiences

• Evaluation of instructional technique
  – Classroom based research

• Evaluation of students
  – Testing and grading
  – Judge student ability, understanding
  – Promote student progress
Evaluating Statistics Students
Some of My Goals

- Understanding of statistical content, process
- Statistical literacy, reasoning, thinking
  - consumers vs. producers
- Computer literacy
- Communication and collaboration skills
- Appreciation, interest level in statistics
- Promoting student progress...
Promoting Student Progress

- Document and *enhance* student learning
- Element of instruction
- Interactive feedback loop
  - Diagnostic with indicators for change
  - Throughout the course
  - To student and instructor
  - Encourage self-evaluation
- Multiple indicators
Design Decisions

• Student activity
  – Calculation
  – Writing
  – Quantitative arguments

• Student environment
  – In class/Out of class - With technology?
  – Timed/Open ended - With notes?
  – Individual/Group work - Student choice?
Example

Given the numbers 5, 9, 11, 14, 17, 29
(a) Find the mean
(b) Find the median
(c) Find the mode
(d) Calculate a 95% confidence interval for μ
Example

A bag contains 100 balls numbered 1 to 100. One ball is removed. What is the probability that the number of the ball is even or less than 30?

(a) .1   (d) .7
(b) .5   (e) .8
(c) .6
Plusses

• Consistent and fast scoring
• “Objective”
• Straight-forward for student
  – Not too time consuming
  – Expected?
• Close match with course coverage?
Minuses

- Focus too much on calculation
- Do not reveal intuition or reasoning
- Lack of meaningful context
- Mislead students as to what we feel is important
Designing Assessments

• Interpretation and Explanation
  – Capture student reasoning and thinking
• More variety, less predictability
  – End of chapter context
• “Statistical” in nature
  – Meaningful context
  – Purpose to calculation
  – Match desired emphasis of course
“Authentic Assessment”

• Assess the student’s performance on worthwhile, significant, meaningful tasks similar to what they will need to do in “real-life” situations beyond school
  – Want we want them to be able to do
  – Production not reproduction
  – Beyond the textbook
“Authentic Assessment”

- Can they apply their knowledge?
- Can they explain their knowledge?
- What are the limitations of their knowledge?
- Can they make good decisions?
- Can they evaluate?
- Can they deal with ill-defined problems?
Writing Assignments

• In own words
  – vocabulary of statistics (‘variable’)
  – internalize their learning
  – enhance ability to judge level of understanding

• Explanations of what learned
  – follow-up
Writing Assignments

• Minute papers
• Projects
• Case studies
• Journals
• Write-up of lab or class activity
• Concept maps
• Critiques
Example

• Matching variables to histograms (*Activity-Based Statistics*)
  – Write a paragraph explaining how you matched the graphs. For example, what features helped you decide?
Example

• Part I: Design data collection
  – Does store A tend to have lower prices than B?
  – Do stats students tend to prefer soda A to B?

• Part II: Full lab report
  – Introduction
  – Data Collection Methods
  – Summarize results
  – Conclusions for the manager
Example

• *Extension*: Discuss one experience with statistics outside of class
  – Bingo with grandma
  – RJ Reynolds talk
Exam Questions

• Multiple choice
  – without options
  – with options but also require explanation
  – with options but require explanation of one of the false statements
  – with options and choices for explanations
Example

Which graph best represents a distribution of sample means for 500 samples of size 4?
A  B  C  D  E
Example (cont.)

• What do you expect to see for the shape of the sampling distribution
  – shaped more like a normal distribution
  – shaped more like the population

• I expect the sampling distribution to have less than/ as the population
  variability than/as the population

  the same

  more
Example

(a) For each of the following statements, identify them as Valid or Invalid interpretations of this p-value.

– The probability that the freshmen score higher than the upper classmen is .04.

– ....

(b) Explain what is wrong with the reasoning for one of the incorrect statements above.
Exam Questions

• Objective-format questions
  – matching problems
    • match verbal description to plots
    • match boxplots to normal probability plots
  – comparative judgments
    • e.g. strength of relationship in two-way table
  – match question with explanation
  – missing pieces of output
Example

- Which data set has a negative correlation?

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>
Example

Order the p-values from largest to smallest
“Rubrics”

- 5 points: discussion of effects of sample size, differences in means, and variability
- 4 points: ignores sample size
- 3 points: only focuses on differences in means
- 2 points: ordering with no explanation
Exam Questions

• Short answer questions
  – Explain in own words
    • confidence level
  – What if…
  – Critique other responses
  – Working backwards
Example

After calculating a 95% confidence interval, how would the sample size need to be changed (larger, smaller, same size) if you wanted:

– (i) to have a larger margin of error?
– (ii) to increase the confidence level?
– (iii) to estimate the mean yield per tree of this fruit variety for just Californian trees (instead of all trees in the country)?
Examples

• True or False? A sample histogram will have a normal distribution if the sample size is large enough.
  – If false, explain why.

• What is wrong with the hypothesis
  \[ H_0: \bar{X}_1 - \bar{X}_2 = 0 \]
Examples

• Construct a data set with 10 values between 0 and 100 so mean > median
• Make up a problem that could be solved with the following output.

```
MTB> ZTEST 25 3.4 C1;
SUBC> ALTERNATIVE  -1.

<table>
<thead>
<tr>
<th>N</th>
<th>MEAN</th>
<th>STDEV</th>
<th>SE MEAN</th>
<th>Z</th>
<th>PVALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>23.63</td>
<td>3.95</td>
<td>0.538</td>
<td>-2.55</td>
<td>0.0054</td>
</tr>
</tbody>
</table>
```
Exam Questions

• Open-Ended Questions
  – Take-home final exam
  – “Demonstrate that…”
  – Student projects…
Example

- A university is interested in studying reasons many of their students were failing to graduate. They found that most attrition was occurring during the first three semesters so they recorded various data on the students when they entered the school and their GPA after three semesters. [Students given data set with numerous variables.]
Example (cont.)

(a) Describe the distribution of GPA for these students.

(b) Is SAT-Math score a statistically significant predictor of GPA for students at this school?

(c) Is there a statistically significant difference between the average GPA values among the majors at this school?
Example

• The FBI reports that nationally 55% of all homicides were the result of gunshot wounds. In a recent sample taken in one community, 66% of all homicides were the result of gunshot wounds. Is this convincing evidence that this community has a higher rate of homicides due to gunshot wounds? Explain.
Example

- The underlying principle of all statistical inference is that one uses sample statistics to learn something (i.e., to infer something) about the population parameters. Convince me that you understand this statement by writing a short paragraph describing a situation in which you might use a sample statistic to infer something about a population parameter. Clearly identify the sample, population, statistic, and parameter in your example. Be as specific as possible, and do not use any example which we have discussed in class.
Projects

• Performance assessment
• Best way to demonstrate and assess the practice of statistics
  – experience finer points, “messiness” of data
• Integration of ideas throughout the course
  – Constant reference
• Communication and collaboration
Making Projects Work

- In-class activities before group selection
- Student choice/ownership
- Full guidelines, examples early
- Periodic progress reports
- Peer review
- Individual accountability
- Oral presentations to each other
Example – Peer Review

• Bottled water vs. tap water
  – Question: Can people really taste the differences between bottled waters and does the packaging that the water comes in have any effect on what the person perceives to be the best water?
Other Project Considerations

• Coverage of statistical tools
• Workload
• Expect/encourage revision
• Expect/encourage “insignificant” results
• Reward the effort
Features of Good “Formative” Assessment

• *Constructive*, timely feedback
• Develop good habits
  – use of context
  – limitations of procedures
  – big and small picture ideas
• Progressive in expectations
• Challenge students to extend knowledge
Features of Good “Summative” Assessment

• Closely aligned with course goals
• Synthesize ideas from different parts of course
• Require students to think in different ways
• Big picture ideas
• Grading focuses on purpose of question and observable student behavior
More Information?

- Do they “appreciate” statistics?
- Have their conceptions changed in this course?
- Do they apply their knowledge outside of the course?
- Do they retain their knowledge?
Cautions

- Inter-related, complementary
- Well-integrated into course
- Well-defined, goals understood by students
- Consistency
- Prepare students
  - detailed guidelines, model papers, rework
Feasibility

• Not all at once, not too much
• Collaboration
  – Instructor level
  – Student level
    • peer review, paired assignments, self-evaluation
• My favorite gold mines:
  – ABS; Rossman; Utts; Iversen & Gergen;
    Freedman, Pisani & Purves; Ramsey & Schafer
Summary

• Assessment must be consistent with learning goals

• Assess what you value
  – Assessment drives instruction
  – Put students in (unprompted) situations that require the skills you are trying to assess

• Remember the goal – increase student learning
Thank You

- bchance@calpoly.edu
Handouts

- projects with peer review form from lsu (3 pages)
- assessment references (1 page)
- slides? (8 pages)
- sample final exam? (5 pages)
to add

- more example exam questions?
- final exam critique?
- other example assignments?
  - good vs. bad
- more discussion of “measureable” and brainstorming?
- how to assess instruction?
“Authentic Assessment”

• Assess the student’s performance on worthwhile, significant, meaningful tasks similar to what they will need to do in “real-life” situations beyond school
  – Want we want them to be able to do
  – Production not reproduction
  – Beyond the textbook
  – Communication and collaboration
Dimensions of Assessment

- Modify Program
- Evaluating Programs
- Making instructional decisions
- Improve Instruction
- Evaluating Students Achievements
- Recognize Accomplishment
References for Assessment in Statistics


Chance course teacher’s guide: [~/chance/teaching_aids/Guide/teaching_chance.pdf](~/chance/teaching_aids/Guide/teaching_chance.pdf)

Activities with follow-up questions:

*STAR Library* [http://starlibrary.net](http://starlibrary.net)
Student Projects in Statistics (sample assignment)

Goal: To collect, describe, and analyze data to answer a question of your choice.

Teams: For the class projects, you will work in groups of 3-5. It is up to the members of the group to make sure everyone contributes equally. Form teams by Sept. 12.

Topics: You are free to choose your own question. The question may be related to your major or some other topic of interest. You should choose a topic so that it will be straightforward to gather the data. You also want to make sure the topic is interesting to you! Be creative! We will discuss some previous topics in class, and some ideas can be found in the EESEE computer package (see lab manual). You may look at previous project topics on the course web page.

Project Reports: The goal of the project reports is to keep you thinking about the projects as the term progresses. Keep in mind that your project may change and evolve as the course progresses. Still, with each project report I would like to hear about your progress and ideas. Turn in one project report for each team, including team members names and previous project reports, preferably typed. Below are some guidelines on what I would like to see in each report.

The first project report is due Sept. 17. For this report you should identify your topic/question of interest, the population you plan to use, and how you plan to collect the data (e.g. survey or experiment and source of your sample).

The second project report is due Sept. 29. Your data collection techniques should be more clearly defined. If an experiment, give a tentative design. If a survey, give the preliminary questionnaire. You should indicate why this study is appropriate to answer your question and what precautions you will take (e.g. nonresponse, sampling bias, wording). Each team is to bring in a typed report. These reports will be peer reviewed by other students in the course.

The third project report is due Nov. 3. You should have completed collecting your data. Include a description of your variables, their units, possible ranges/responses, as well as preliminary descriptive statistics and graphs and what information you would like to gain from your data analysis, that is, what questions do you want to answer.

The fourth project report is due Nov. 21. Identify the statistical tests you will be using and why you chose them. I will work with groups who want to use statistical techniques we have not covered in class yet.

Rough Draft (optional) If you turn in a rough draft by Dec. 1, I will review the paper, providing comments and suggestions.

Final Reports: Final reports are due Dec. 11, but can be turned in earlier. Reports must be typed. Turn in one report per group. Incorporate computer output into the body of the paper or as an appendix. Raw data does not itself need to be included except perhaps as an appendix. You may assume your audience will understand all statistical terminology.

Grading Criteria for Final Report:
10%: Quality of written report
20%: Design of survey/experiment - was data collection adequately explained, were the appropriate data collected to answer the questions and/or test the hypotheses and estimate parameters.
25%: Correctness of statistical analysis and assumptions checks
20%: Correctness of interpretations of the results of the statistical calculations and conclusions

Presentations: (other 25%) Each group will give a 10-15 minute presentation (at the most) to the class on Dec. 8 or 10 (order will be determined randomly). Any number of group members may present part of the report. The presentation should not include extensive details, but provide the audience with an overview of what was done, what conclusions can be drawn, any drawbacks of the techniques and future recommendations. Feel free to be creative. The presentation should be done on PowerPoint, transparencies, or a poster board.

Evaluation: Each person’s grade will be 85% group grade and 15% individual grade. Individual grades will be determined by the instructor and team member evaluations.
First Project Report

1) Candy Challenge: Skittles vs. M&M
Topic question: Which candy has more red candies, skittles or M&Ms? Find the average amount of each color per package of each candy? Find the cost of each individual candy?

Population: all skittles and M&Ms that exist

Sample: skittles and M&M packages from randomly selected grocery stores, convenient stores, gas stations in San Luis Obispo and vending machines on campus.

Type of study: Experiment

2) Bottled water vs. tap water
Question: Can people really taste the differences between bottled waters and does the packaging that the water comes in have any effect on what the person perceives to be the best water?

Population – Cal Poly students

Sampling – Random students on campus a mixture of probably 10 guys and 10 girls who will volunteer to take taste test

Experiment: Have an assortment of bottled waters, tap water, and filtered tap water (Britta) set out and let people taste and rate which ones they like the best. Blindfold the person and give them the same bottled waters and ask person to rate water again. Will the water gets the same rating as it did the first time?

3) Time of day and units taken
Topic of interest: If time of day influences the number of units taken by students

Population planned – Cal Poly students

Sample – 100 people

Sampling frame – People enrolled at Cal Poly State University SLO

Type of Student - survey

4) Candy faithfulness
Possible topic: box office reports for movies in the United States vs. box office reports for movies in SLO county
Finding out which candy tends to sell more, which type brings in the most amount of money in concern with the total amount brought in by concessions
- survey – movie goers favorite candy
- sample – theaters (actual concession income) (compare both sets of data)
5) **Recreational Activities**
We are going to take a survey on students’ recreation patterns. Since we are recreation administration majors, our population will be Cal Poly Rec students. Our sample will be 40 Rec majors.

6) **Sodas**
Topic: What types of sodas do people drink?
Sample: Age, gender, occupation

7) **Air Attack vs. Omaha**
Topic: Which bat do Cal Poly baseball players prefer
Population: Cal Poly baseball players
Sample: Two players from each position
Sampling frame: Cal Poly baseball roster
Type of study: survey

8) **Internet usage**
We are going to be taking a measurement of how many students use the internet, and if it is an abundant amount of time do they feel this affects their grades? We plan to ask students on the Cal Poly campus, and our method is going to be a survey.

9) **Lifestyle habits**
Population: Cal Poly students and Cuesta students
Sample: 100 Poly students and 100 Cuesta students
Conducting a survey of the lifestyle habits, i.e. Nutrition, exercise, sleep, study, drinking, smoking, and then compare the 2 populations.

10) **Alcohol habits**
We are thinking about doing a survey on alcohol related topics for students at Cal Poly. We could ask how much a student drinks in a week, what they prefer to drink, where do they drink, etc.

11) **Rec Center**
Project: Amount the Rec Center is used by males and Females and times used.
Type: survey
Sample: Rec Center Visitors
Sampling Frame: People around campus and people entering the Rec Center.
Sample Second Project Report
We will be conducting an observational study to see who uses their turn signal more often while driving: males or females. By standing on the corner of Alpine and Pershing Ave., we will observe cars as they pass by us. Our group will split up, and while half of us are on one intersection, the other half will be observing the cars from the other side of the street. That way, we will be able to record more cars. We will pick a day out of the week at random, possibly Tuesday and/or Thursday, and conduct our study for about an hour. We will use 5:00pm as our time, since traffic will be fairly heavy then, and we will be able to get a large sample. We will also utilize Sunday afternoon, since many people also drive to church. We will not be imposing any kind of treatment on the driver, since we will only be taking observations of those cars that signal when changing lanes, and the sex of the driver. By doing this, we will see who is more courteous of the other drivers on the road.

Sample Peer Review Questions (based on sample by Anne Sevin)
1. Did the group do a good job of describing the project goal/question to be answered? Why/why not?

2. Did the group clearly describe the population they are trying to describe? Why/why not?

3. Did the group do a good job of selecting a representative sample? Why or why not? Are there any possible sources of selection bias which could make the sample atypical or nonrepresentative of the population? Did the group many any provisions for how they intend to deal with possible biases?

4. Did the group do a good job of describing the design of their study? Why or why not? Are there possible sources of bias?

5. Do you see any potential biases in how the measurements will be taken? What precautions does the group need to make? Do you believe that the data they plan to collect will be relevant to the objectives of the project? Why or why not?
Sample Final Exam
(closed book, ~120 minutes)

1) The following data are the point totals for the Men’s Basketball team in their first 8 victories this season:

80 72 68 55 80 78 90 85

(a) (5 pts) Make a stemplot of these winning point totals and describe the shape of the distribution.
(b) (3 pts) Would the Five Number Summary or the mean and standard deviation be a better summary for this distribution? Explain your choice.

2) Two investigators wanted to study the heights of 18-24 year old men in Stockton. One investigator, Happy Harry, took a random sample of 100 men. The other investigator, Tired Tony, took a random sample of 1000 men.

(a) (2 pts) If each investigator finds the average height of the men in his sample, which investigator, Harry or Tony, should expect a larger average, or should they be about the same? Explain.
(b) (3 pts) Which sample, Harry or Tony’s, do you expect will have less bias or will they be about the same? Explain.
(c) (3 pts) Which estimate of the population mean, Harry or Tony’s, should have higher precision, or will they be about the same? Explain.

3) In 1988, men averaged about 500 on the math SAT, the standard deviation was about 100, and their scores followed a Normal distribution. One of the men who took the math SAT in 1988 will be picked at random, and you have to guess his test score. You will be given 50 dollars if you guess it right to within 50 points.

(a) (2 pts) What one number should you guess?
(b) (5 pts) With this guess, what is your probability of winning the 50 dollars? Explain. Extra Credit: What is your expected winnings?
4) The distribution for a population of test scores is displayed below on the left. Each of the other five graphs, labeled A to E represent possible sampling distributions of sample means for 500 random samples drawn from the population. (Justify choices)

(a) (2 pts) Which graph represents a sampling distribution of sample means for samples of size 1?
   A  B  C  D  E

(b) (2 pts) Which graph represents a sampling distribution of sample means for samples of size 9?
   A  B  C  D  E
5) A social research scientist wants to test whether the percentage of Republicans who favor the death penalty is greater than the percentage of Democrats who are in favor of the death penalty. Suppose the sample data showed that the percentage of Republicans who are in favor of the death penalty is 42% and the percentage of Democrats who are in favor of the death penalty is 40%.

(a) (2 pts) Write down the null and alternative hypotheses for this test.
(b) (3 pts) The p-value for this test is .0021. The 95% confidence interval for \( p_1 - p_2 \) is (.00637, .03363). Which of the following conclusions do you think is more appropriate to draw?

1. There is evidence of a large difference in the two proportions.
2. There is strong evidence of a difference in the two proportions.

Explain.

(c) (2 pts) Which conclusion does a p-value better address? Explain.
(d) (2 pts) Which conclusion does a confidence interval better address? Explain.

6) In a clinical trial, data collection usually starts at “baseline,” when the subjects are recruited into the trial but before they are randomized to treatment and control groups. Data collection continues until the end of follow-up. Two clinical trials on prevention of heart attacks report baseline data on weight, shown below.

<table>
<thead>
<tr>
<th></th>
<th>Number of persons</th>
<th>Average weight</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 1</td>
<td>Treatment</td>
<td>1,012</td>
<td>185 lb</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>997</td>
<td>143 lb</td>
</tr>
<tr>
<td>Trial 2</td>
<td>Treatment</td>
<td>995</td>
<td>166 lb</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>1,017</td>
<td>163 lb</td>
</tr>
</tbody>
</table>

(a) (4 pts) In one of these trials, the randomization did not achieve the desired result. Which trial and why do you say so? How will this affect our results and conclusions for this study? (Hint: make sure you focus on the most serious difficulty)

(b) (4 pts) Below are ten people and their weights. Randomly assign them to one treatment group and a control group (start with line 139 of the random number table). Clearly show your work.

Bob 148   Tom 174   Joe 148   Fred 133   Sam 157
Curt 177   Al 162   Harry 188  Gami 160  Dan 188

7) Can pleasant aromas help a student learn better? Two researchers believed that the presence of a floral scent could improve a person’s learning ability in certain situations. They had ten people work through a pencil and paper maze 2 times, first wearing an unscented mask and then wearing a scented mask. Tests measured the length of time it took subjects to complete each of the two trials. They reported that, on average, subjects wearing the floral-scented mask completed the maze more quickly than those wearing the unscented mask.

(a) (3 pts) Is this an observational study, survey, or experiment? Explain.
(b) (2 pts) Identify the response and explanatory variables.
(c) (4 pts) Explain how confounding makes the results of this study worthless.
(d) (4 pts) Sketch an outline of a more appropriate design for the study.

8) NCAA collected data on graduation rates of athletes in Division I in the mid-1980s. Among 2,332 men, 1,343 had not graduated from college, and among 959 women, 441 had not graduated.
(a) (3 pts) Set up a two-way table to examine the relationship between gender and graduation.
(b) (3 pts) Calculate a couple of conditional percentages to describe the relationship between gender and graduation.
(c) (3 pts) Identify a test procedure would be appropriate for analyzing this relationship? State the null and alternative hypotheses.
(d) (3 pts) What type of distribution does the test statistic you describe in (c) follow? For what values of this test statistic will you reject the null hypothesis at the 5% level?
(e) (2 pts) If the above result is significant, would this mean that if some people have a sex change they will increase their chance of graduating? Explain briefly.

9) A panel of trained testers judged the flavor quality of different vanilla frozen desserts (frozen yogurts, ice milks, other frozen desserts) measured on a scale from 0 to 100. The data are from a *Consumer Reports* article “Low-fat frozen desserts: Better for you than ice cream?” (August, 1992). Below is a graphical summary of the data.

Here is most of the ANOVA output from the computer:

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYPE</td>
<td>6</td>
<td>6364</td>
<td>1061</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERROR</td>
<td>24</td>
<td>3031</td>
<td>126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>9395</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) (2 pts) Explain briefly why ANOVA was the appropriate analysis for these data.
(b) (2 pts) State the null and alternative hypotheses.
(c) (4 pts) Finish the ANOVA table giving the F-statistic, degrees of freedom, and approximating the p-value. Show your work. What is your conclusion about the flavor quality of the different desserts?
(d) (2 pts) Based on the graph, do you feel the technical conditions needed for the validity of this test procedure are valid? Explain.

10) A random sample of 7 households was obtained, and information on their income and food expenditures for the past month was collected. The data (in hundreds of dollars) are given below.

<table>
<thead>
<tr>
<th>Income ($100’s)</th>
<th>22</th>
<th>32</th>
<th>16</th>
<th>37</th>
<th>12</th>
<th>27</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food Expend ($100’s)</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

Here’s a scatterplot of these data with the regression line superimposed.
Here’s the Minitab output:

The regression equation is  
\[
  \text{expend} = 1.87 + 0.202 \text{ income}
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>Stdev</th>
<th>t-ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.8690</td>
<td>0.9068</td>
<td>2.06</td>
<td>0.094</td>
</tr>
<tr>
<td>income</td>
<td>0.20195</td>
<td>0.03661</td>
<td>5.52</td>
<td>0.003</td>
</tr>
</tbody>
</table>

s = 0.8181  R-sq = 85.9%  R-sq(adj) = 83.1%

(a) (2 pts) Describe the direction and strength of the association.
(b) (2 pts) On the graph, identify the point which you think has the largest residual. Explain.
(c) (2 pts) On the graph, identify the point which you think has the most influence on the position of the regression line, and how the line would change if it was removed. Explain.
(d) (3 pts) Provide an interpretation of the number .202 in the regression equation in the context of these data. Exactly what does this value tell us?
(e) (4 pts) Is there evidence of a statistically significant relationship between income and food expenditure? Make sure you clearly explain the basis for your answer.
(f) (2 pts) Explain why you would not recommend using this relationship to predict the food expenditure for a household with an income of $5,200.

11) National data show that, on the average, college freshmen spend 7.5 hours a week going to parties. President DeRosa doesn’t believe that these figures apply at UOP. He takes a simple random sample of 50 freshmen, and interviews them. He finds that the 95% confidence interval for the mean number of hours spent a week going to parties is (5.72, 7.42).

(a) (4 pts) Explain to the President what he means by the phrase “95% confidence.”

Now he wants to test the hypothesis that the mean for UOP is different from the national mean at a 5% significance level.

(b) (2 pts) Specify the null and alternative hypotheses for this test.
(c) (2 pts) Indicate a test procedure he could use to conduct this test.
(d) (3 pts) Eager to gain favor with the president, you tell him that you can save him lots of time because, based on the data already presented, you know what he will conclude and he doesn’t have to perform any additional calculations. Does he reject or fail to reject the null hypothesis at the 5% level? Explain.

**Extra Credit**

Suppose you take 50 measurements on the speed of cars on Interstate 5, and that these measurements follow roughly a Normal distribution. Do you expect the standard deviation of these 50 measurements to be about 1 mph, 5 mph, 10 mph, or 20 mph? Explain.
The 2002 AP Statistics Exam

Beth L. Chance
Cal Poly – San Luis Obispo
bchance@calpoly.edu
Outline

• Brief overview of exam structure
• Discussion of grading process
• 2002 rubrics, sample papers
• Preparing students for the exam
Resources

• APCentral.collegeboard.com
  – copies of questions, rubrics
  – comments from CR
• New teacher’s guide
• Multiple choice questions released
AP Statistics Exam

- 50% multiple choice (90 minutes)
  - 40 questions
- 50% free response
  - five short answer (65 minutes, 75%)
  - one investigative tasks (25 minutes, 25%)
  - holistic scoring (0-4 scale)
- Calculator policy
2002 Free Response Questions

- Question 1: estimation, margin of error
- Question 2: experimental design
- Question 3: normal probability, combination of random variables
- Question 4: correlation, regression
- Question 5: hypotheses, two-sample t test
- Question 6: confidence intervals
Holistic Scoring

• p. 65
Question 1

• Understanding of “precision”

• Use displayed intervals to assess degree of support for each of the two competing theories
  – Conclusion and support
  – Intervals as range of plausible values

• Emphasis on good communication
  – Consistency in statements
Question 1 (cont.)

- Margin of error = number vs. interval
  - interquartile range
- Sloppy terminology
  - observations, estimates, data, experiments
  - impossible, certain, proved
- Emphasis on point estimate
  - ignored uncertainty in estimates
  - considered interval of observations?
Question 2

• Describe experimental design
  – details of implementation (randomization, blocking, response variable)
• Double blind
Question 2 (cont.)

• Missed “obvious” paired nature
  – Try to block on several variables
  – Incorrect usage of “confounding” variables

• Did not use information provided
  – 100 volunteers
  – Failed to verbally explain design
    • Did not describe appropriate randomization, SRS

• Incomplete understanding of who is blind
Question 3

• Normal probability calculation
• Distribution of sum of random variables
• Interpret calculated probability
Question 3 (cont.)

• Judging by standard score vs. when judge by normal probability vs. hypothesis test
  – “unusual” vs. “likely” vs. “evidence”

• Difficulty in applying rules for expected values
  – Recognizing what was being asked for
  – Total time vs. average

• Did not apply answer to (b) in (c)
Question 3 (cont.)

• Calculation errors
  – two-sided
  – wrong direction
  – continuity correction
Question 4

• Interpreting computer output
  – Minitab
  – In context
  – Equation in terms of y-hat
• Relationship between $r$ and $R^2$
  – interpretation of $r$
• Reasonableness of model for subset of data
Question 4 (cont.)

• Many not familiar with computer output
  – slope
  – $R^2$ vs. $R^2$(adj)
  – equation for y-hat

• Not answering question asked
  – defining variables
  – not addressing *effect* of removing points
  – restricting range vs. extrapolation
Question 4 (cont.)

• Unreasonable values for $r$
  – Transcription errors
  – No check of $-1 \leq r \leq 1$
  – Confusion between $r$ and $R^2$

• Failure to consider $r = .755$ moderately strong

• Confusion of $r$ and $R^2$
  – Strength, direction, form, context vs. $R^2$
Question 5

• Setting up hypotheses for other types of parameters
  – Two distinct pairs

• Carrying out two-sample t test
  – Check necessary conditions
  – State conclusions in context
Question 5 (cont.)

1) Identify procedure (name but at least formula)
   Check conditions
   – Independent random samples – stated
   – Normal population or large samples – state the two sample sizes and decide if are large
   – If pooled, compare sample standard devs

2) Mechanics

3) Conclusion
Question 5 (cont.)

- New hypotheses vs. rearrangement
  - Invalid use of two-sided alternative
  - Which hypotheses to use in (b)
  - Did not clarify symbols used in (a)
- Hypotheses about population values
- Link between calculation and conclusion
- State conclusion in context
  - Confuse significance level and confidence
Question 6

• Calculation of one-sample confidence interval
  – including check of conditions
• Interpretation of confidence interval
• Interpretation of confidence level
• Comparison of two population proportions
• Investigation of combining results
Question 6 (cont.)

- Missing or incorrect interpretation
  - Separate interpretations for interval and level
  - What’s in interval, 95% of what
- Checking conditions (both interval and test)
- Stating hypotheses about parameters
- Did better in part (d)!
- Correct calculations alone did not receive much credit
Preparing Students

• Practice communicating in context throughout the course
  – Statistical terminology, notation, it
  – Verbalize each step (in symbols and in words)
  – Coherent organization (linkage)
  – Discourage “TI-speak”

• Emphasize concepts and communication over mechanics
Preparing Students (cont.)

• Require students to show work
  – Start with a picture
  – Reasonable computation errors not detrimental
    • Do no “over round”
  – Be able to justify choice of procedure
  – Support conclusions

• Have students mimic environment

• Grade solutions with these rubrics(students)
Preparing Students (cont.)

• Emphasize problem areas
  – *Checking* conditions of procedure
  – Paired vs. independent samples
  – Reading computer output
  – Explaining overall statistical *process*
    • e.g., why do we randomize
  – Explaining a procedure
  – Working backwards
Preparing Students (cont.)

• Teach good problems solving strategies
  – Read entire question before starting
  – Setting up the problem
    • Provide “context-free” practice
    • Familiarity with question phrasings
    • Multiple representations, variety, integration
  – Answer the question asked, know when to quit!
  – Check answer for credibility