AP® Statistics
2002 Scoring Guidelines
Form B

These materials were produced by Educational Testing Service® (ETS®), which develops and administers the examinations of the Advanced Placement Program for the College Board. The College Board and Educational Testing Service (ETS) are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

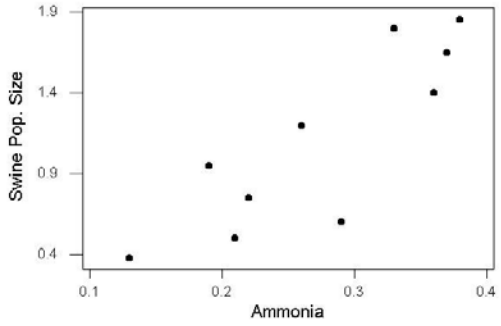
The College Board is a national nonprofit membership association dedicated to preparing, inspiring, and connecting students to college and opportunity. Founded in 1900, the association is composed of more than 4,200 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three million students and their parents, 22,000 high schools, and 3,500 colleges, through major programs and services in college admission, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of equity and excellence, and that commitment is embodied in all of its programs, services, activities, and concerns.

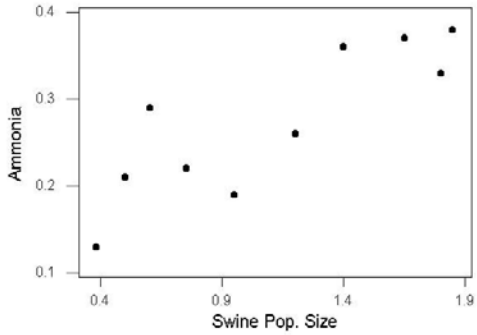
AP® STATISTICS
2002 SCORING GUIDELINES (Form B)

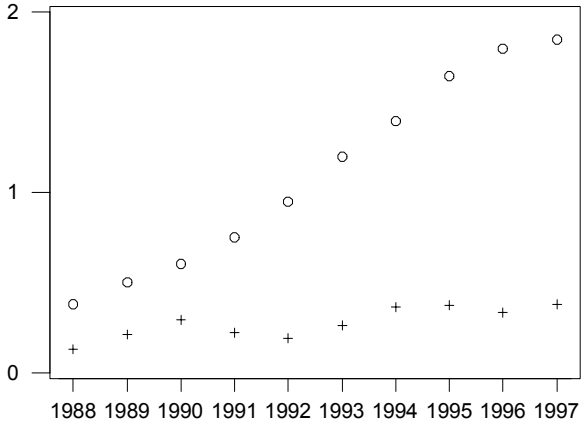
Question 1

Solution

Part (a):







Part (b):

There is a strong, positive, linear relationship between swine population size and atmospheric ammonia.

Part (c):

Both the value of the correlation coefficient and the pattern in the scatterplot indicate that there is a positive linear relationship between the size of the swine population and atmospheric ammonia.

Part (d):

r^2 = .72 or 72% or 72

Note: Just writing r^2 is not sufficient.

Question 1 (cont'd.)

Scoring

Part (a) is considered

Essentially correct if a scatterplot is given that includes axis labels and scales.

Partially correct if a scatterplot is given but only one of (i) the axis labels or (ii) the scales is included.

Incorrect if neither labels or scales are included.

Parts (b) and (c) should be read "collectively."

Part (b) is considered

Essentially correct if the interpretation of the correlation coefficient includes both strong and positive, and is interpreted in context.

Partially correct if the interpretation mentions two of the components: strong, positive, and context.

Parts (c) is considered

Essentially correct if comments are correct and are based on both the value of the correlation coefficient and the scatterplot.

Partially correct if it includes a correct comment based on only one of the components: correlation coefficient and the scatterplot.

Part (d) is considered

Essentially correct or **incorrect.**

NOTE: In order to recoup strength or direction for part (b) in part (c), the comments must be directly tied to the correlation coefficient. For example, saying the variables increase together does not count as saying the correlation tells us the association is positive. Interpretations in context in (c) can recoup missing context in (b).

NOTE: A construction such as "as x increases, y increases" counts as indicating a positive association, but does not by itself count as indicating linearity.

NOTE: A statement such as "judging by the scatterplot and correlation" without appealing to some characteristic of the plot or that the correlation is .85, is scored as partially correct.

Essentially correct responses count as 1 part and partially correct responses count as ½ part. If a paper is between two scores (for example 2 ½ parts), use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.

Question 1 (cont'd.)

4 **Complete Response**
Four parts correct.

3 **Substantial Response**
Three parts correct.

2 **Developing Response**
Two parts correct.

1 **Minimal Response**
One part correct.

AP® STATISTICS
2002 SCORING GUIDELINES (Form B)

Question 2

Solution

Part (a):

P (everyone gets a seat) = P ($X \leq 38$) = .46 + .30 + .16 = .92

OR = 1 – (.05 + .02 + .01) = .92

Part (b):

Y = number of no shows

y	0	1	2	3	4	5
p(y)	.01	.02	.05	.16	.30	.46

E(Y) = 0(.01) + 1(.02) + 2(.05) + 3(.16) + 4(.30) + 5(.46) = 4.1

OR

E(X) = 36(.46) + 37(.30) + 38(.16) + 39(.05) + 40(.02) + 41(.01) = 36.9

E(Y) = 41 – E(X) = 4.1

Part (c):

P(X=36|not all seats are filled) = $P(X = 36 \mid X < 38) = \frac{P(X = 36)}{P(X < 38)} = \frac{.46}{.76} = .605$

Question 2 (cont'd.)

Scoring

Part (a) is

Essentially correct if computes P($X \le 38$) (except for minor arithmetic errors)

Partially correct if computes P($X = 38$) = 0.16 or P($X > 38$) = .08 or P($X < 38$) = 0.76 or $P(X \ge 38) = 0.24$ or gives 0.92 but does not show any work.

Incorrect if no calculation or nonsensical calculation (e.g., 0.921, $\frac{38}{41} = 0.927$, $\frac{36.9}{38} = 0.971$) or only pure expected value calculations.

Part (b) is

Essentially correct if (except for minor arithmetic errors) correctly computes expected value for number of no shows and indicates fully where the 36.9 and 4.1 come from.

Rounding this value to 4 is considered a minus, though can be forgiven with "about" or "approximately."

Partially correct if correctly computes expected value of X = number of passengers who show up for the flight (instead of no shows) and shows work,

OR if incorrectly computes the first expected value but "subtracts,"

OR if correctly computes the first expected value but "subtracts" from the wrong number,

OR if does not show work for 36.9 but subtracts to get 4.1.

Incorrect if is not an expected value or does not use all six outcomes in the expected value.

Part (c) is

Essentially correct if correctly computes the conditional probability. Complete notation could be considered a "plus."

Partially correct if correctly computes $P(X = 36 \mid X < 41) = \frac{0.46}{0.99} = 0.465$ or $P(X = 36 \mid X \le 38) = \frac{0.46}{0.92} = 0.5$ or incorrectly tries to solve the conditional probability (e.g., multiplies probabilities in numerator).

Incorrect if an unconditional probability is computed.

Question 2 (cont'd.)

4 Complete Response

Essentially correct on all three parts.

3 Substantial Response

Essentially correct on two parts and partially correct on the other part or incorrect on (a) and essentially correct on (b) and (c).

2 Developing Response

Essentially correct on two parts and incorrect on the other (except IEE – see above).
OR
Essentially correct on one part and partially correct on the other two parts.
OR
Partially correct on all three parts.
OR
Essentially correct on one part, partial on another, and incorrect on a third part.

1 Minimal Response

Essentially correct on one part and in correct on the other two parts.
OR
Partially correct on one or two parts and incorrect on the other.
PII should be graded holistically (needs something elsewhere for a 1).

AP® STATISTICS
2002 SCORING GUIDELINES (Form B)

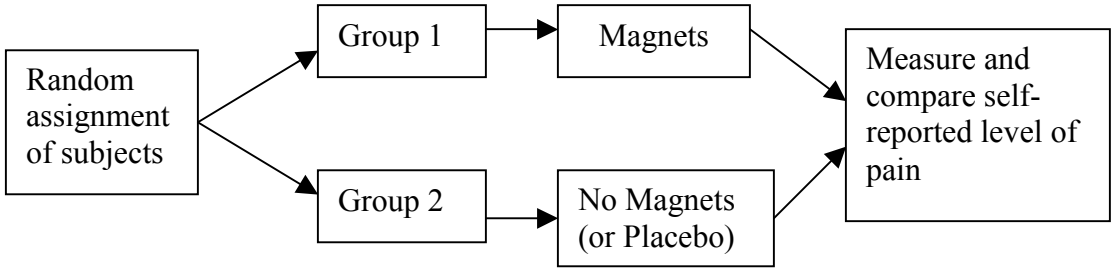
Question 3

Solution

Part (a):

1. Two treatments: magnets and no magnets (or magnets and placebo). Subjects in the no magnet group would be handled in the same way as the magnet group, but there would be no magnets embedded in the pads used.
2. There must be random assignment of subjects to treatments (or treatments to subjects). How the randomization would be carried out does not need to be specified, but it must be clear *what* is being randomized.
3. Variable measured: Self-reported level of pain or reduction in pain.

The design may be described by a diagram, but the treatments and the variable measured must be included and the randomization must be very clear.



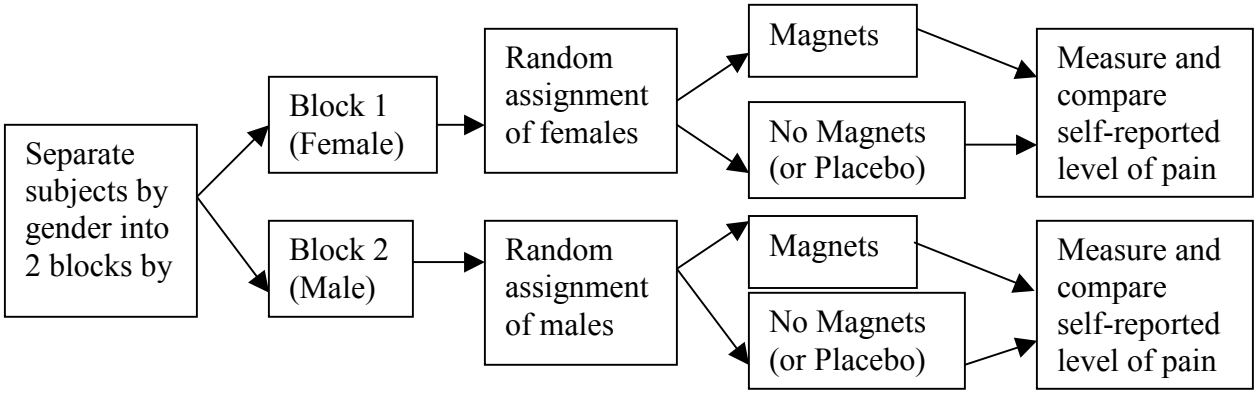
Part (b): Either one of the following approaches is acceptable.

1. Saying yes and indicating how they would alter the design: Separating the subjects into the two gender groups and then randomizing subjects to treatments within each group. This may also be described using a diagram, as shown below, but the blocking factor and randomization must be clearly indicated.

   OR

2. Saying no and describing why. For example, indicating that the randomization in (a) should equalize the effects of gender in the two groups or assuming gender does not have a strong effect and since the sample size is large
   OR
   providing a good explanation for *why* gender does not have a strong differential effect on the outcome.

Question 3 (cont'd.)



Scoring

Part (a) is

Essentially correct if all three of the following are adequately addressed: treatments, random assignment, and response variable.

It is acceptable for them to select the first group with an SRS, but only if it's clear that the remaining subjects are automatically assigned to the second treatment group as opposed to taking a separate SRS.

Note, if they describe a randomization scheme that results in an unequal number of subjects in each treatment group or does not involve randomization for each individual (such as flipping a coin to determine which group gets the placebo), this can be considered a minus.

The student may also describe the experimental protocol in more detail – double blind, placebo effect, etc. These are not necessary but can be considered a plus.

Partially correct if only one or two of the items listed above are adequately addressed. But, to receive the partial with only *one* item, the design must be described exceptionally well. For example, discussion of the placebo effect as reason for the control group, or a "before and after" measurement for the response, or a detailed discussion of randomization. (A well done "before and after" design in the new context can be considered partially correct.)

NOTE: Students do not get credit for discussing treatments if their design does not ever specify a control group or fails to define the treatment.

Question 3 (cont'd.)

Part (b) is

Essentially correct if either of the two acceptable approaches is taken (no and why OR yes and how) and the explanation/justification is clear and complete. For a yes, the response must clearly indicate and maintain the group separation and specify the separate randomization within each group.

Partially correct if either of the two acceptable approaches is taken, but the explanation/justification is weak. For example, the response states yes but it is not clear a separate randomization within each block is carried out, or the response states no but does not fully appeal to the randomization in part (a) or only states that gender does not matter and does not clearly state gender does not matter and why the groups will still be sufficiently balanced.

If done well, the response can appeal to the randomization in part (a) ("divide and proceed for each subgroup" vs. "divide and proceed"). A two factor design that includes randomization rather than a block is considered partially correct.

Note, using the term "stratified random sample" is considered extraneous and is not sufficient for an explanation of how randomization occurs in the new design.

Incorrect if all the response mentions is now having four groups of data to analyze. (No randomization, not even a reference to the randomization in (a), and no clear separation of the groups in the design.)

NOTE: If no randomization is specified anywhere, the highest possible score is a 2.

4 Complete Response

Both parts essentially correct.

3 Substantial Response

Part (a) is essentially correct and part (b) is partially correct.

OR

Part (a) is partially correct and part (b) is essentially correct.

Question 3 (cont'd.)

2 **Developing Response**

Both parts are partially correct.

OR

Part (a) is essentially correct and part (b) is correct.

1 **Minimal Response**

Part (a) is partially correct and part (b) is incorrect.

OR

Part (a) is incorrect and part (b) is essentially or partially correct.

NOTE: PP is between a 1 and a 2. If only the before and after design is present, it will be rounded down to a 1.

AP® STATISTICS
2002 SCORING GUIDELINES (Form B)

Question 4

Solution

Part (a):

Large sample confidence interval for a population proportion.

Requires large population ($N \geq 20n$) and large sample. Can consider sample large since
$n\hat{p} = 1026(.39) = 400.14 \geq 10$, $n(1-\hat{p}) = 1026(.61) = 625.86 \geq 10$

OR $n\hat{p} \geq 5, n(1-\hat{p}) \geq 5$

OR $\hat{p} \pm 3\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .39 \pm .0298 = (.3602, .4198)$ is within the interval (0, 1).

Computation of 95 percent confidence interval for one population proportion:

$$.39 \pm 1.96\sqrt{\frac{(.39)(.61)}{1026}} = .39 \pm .0298 = (.3602, .4198)$$

If use the calculator with x = 400, get (.36002, .41971)

Interpretation of confidence interval:

Based on the sample, we estimate that the proportion of U.S. adults who would respond "make some changes" is between .36 and .42.
OR
We are 95 percent confident that the proportion of all U.S. adults who would respond "make some changes" is between .36 and .42.

Interpretation of the confidence level:

In the long run, 95 percent of the (different) intervals generated using this method will contain the true population proportion.
OR
The *method* used to produce this estimate has a 5 percent error rate.

Question 4 (cont'd.)

Part (b):

The statement is misleading because in addition to the 39 percent who wanted major change, 19 percent wanted a complete overhaul of the system. These two groups combined represent 58 percent (more than half) of the sample.

OR

Since the confidence interval from part (a) for the percentage of all U.S. adults who would respond "make some major changes" includes 41 percent, we can't conclude that the two population percentages are different.

OR

The advocate is claiming that we should not change the system on the basis of adding the 30 percent who want to make some minor adjustments to the 11 percent who want to leave the system the way it is now. It is misleading to claim that these 30 percent don't want to change the system.

Question 4 (cont'd.)

Scoring

There are four parts to this question. You should read (2) and (3) together.

1. Determining the 95 percent confidence interval is
   Essentially correct if it
   - identifies the large sample confidence interval for a population proportion either by name or by formula (numbers plugged in is OK);
   - checks to make sure that the sample size is large enough; and
   - and has correct computations (.3602, .4198)

   Partially correct if just one of the above is incorrect or missing.

2. Interpretation of the confidence interval is
   Essentially correct if it is clear, correct, and in context (include the numbers).
   Partially correct if unclear or no context is given.
   Incorrect if it says "The probability or chance that the true proportion is between .36 and .42 is .95," or if it looks like the sample proportion is in the confidence interval ["…the proportion of U.S. adults who said…" (past tense)].
   OK to say "95 percent sure."

3. Interpretation of the confidence level is
   Essentially correct if it is clear and correct (no context needed).
   Partially correct if it is somewhat unclear but nothing is clearly incorrect.
   Incorrect if it says "In 95 of every 100 samples, the population proportion will be between .36 and .52," or "will be in this interval."

4. Answer to (b) may be scored essentially correct, partially correct, or incorrect.
   For example, if the answer only says the 19 percent was ignored and does not make any new comparisons or suggests a better calculation, then score this as a partial.

Essentially correct responses count as one part and partially correct responses count as ½ part. If a paper is between two scores, use a holistic approach, considering overall strength and communication, to determine whether the student gets the higher and lower score.

4 **Complete Response**
Four parts correct.

3 **Substantial Response**
Three parts correct.

2 **Developing Response**
Two parts correct.

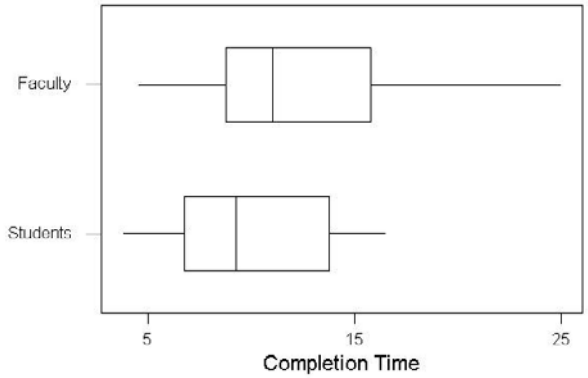
1 **Minimal Response**
One part correct.

AP® STATISTICS
2002 SCORING GUIDELINES (Form B)

Question 5

Solution

Part (a):



Part (b):

The range for faculty completion times is larger than for the students, but the IQR is the same for both groups. The spread is similar in the middle 50 percent of the data, but the smallest 25 percent and the largest 25 percent are more spread out for the faculty members than for students.

Part (c):

Students should comment on at least two of center (mean, median, or general location), variation, or shape. The statements should be correct and clear and suitable for the school newspaper.

Example: Although some faculty negotiated the obstacle course quickly, in general students tended to have shorter completion times. (location) The student completion times, ranging from 3.75 minutes to 16.5 minutes, were more consistent than the faculty times, which ranged from 4.5 minutes to 25 minutes. (variation) Many students and faculty finished relatively quickly, but the slower half of each group tended to spread out. (shape)

Question 5 (cont'd.)

Scoring

Part (a) is considered

Essentially correct if the boxplots are drawn correctly with labels and scale given.

Partially correct if boxplots have no more than one error such as these: missing labels, missing scales, not drawn to scale, or drawn showing outliers.

NOTES:
- It is a considered a minus if "time" label is omitted.
- Any graphic that is not clearly a boxplot is incorrect.

Part (b) is considered

Essentially correct if response notes both that spread in the center is similar for the two groups and that spread in the tails is greater for faculty than for students.

Partially correct if response only comments that variability is greater for the faculty than for the students.

NOTE: It is considered a minus if the 25 is called an outlier.

Part (c) is considered

Essentially correct if there is a clear and coherent statement that comments on at least two of center (mean, median, or general location), variation, or shape. Both faculty and students must be mentioned in the response.

Partially correct if only comments on one aspect of the distribution or if communication is weak.

Question 5 (cont'd.)

4 Complete Response

Essentially correct on all three parts.

3 Substantial Response

Essentially correct on two parts and partially correct on the other part.

2 Developing Response

Essentially correct on two parts and incorrect on the other.

OR

Essentially correct on one part and partially correct on the other two parts.

OR

Partially correct on all three parts.

OR

Essentially correct on one part, partially correct on one part, and incorrect on one part.

1 Minimal Response

Essentially correct on one part and incorrect on the other two parts.

OR

Partially correct on one or two parts and incorrect on the other.

AP® STATISTICS
2002 SCORING GUIDELINES (Form B)

Question 6

Solution

Part (a):

Part 1: States a correct pair of hypotheses, with symbols defined.

μ_{1990} = mean number of lab classes for Biology majors in 1990
μ_{2000} = mean number of lab classes for Biology majors in 2000

$H_0 : \mu_{1990} = \mu_{2000}$
$H_0 : \mu_{1990} \neq \mu_{2000}$

OR

$H_0 : \mu_{1990} - \mu_{2000} = 0$
$H_0 : \mu_{1990} - \mu_{2000} \neq 0$

Part 2: Identifies a correct test (by name or by formula) and checks (not just lists) appropriate assumptions.

Two-sample t test $\quad t = \dfrac{\bar{x}_{1990} - \bar{x}_{2000} - 0}{\sqrt{\dfrac{s^2_{1990}}{n_1} + \dfrac{s^2_{2000}}{n_2}}} = \dfrac{1.83 - 1.93}{\sqrt{\dfrac{1.29^2}{200} + \dfrac{1.37^2}{200}}}$

Conditions:

1. Random samples – given in problem statement
2. Normal population distributions or large samples. Populations aren't normal, but since $n_1 = 200$ and $n_2 = 200$ are both large, it is OK to perform the t test.

OR

Student may select the two-sample z test, using $z = \dfrac{\bar{x}_{1990} - \bar{x}_{2000} - 0}{\sqrt{\dfrac{s^2_{1990}}{n_1} + \dfrac{s^2_{2000}}{n_2}}}$

as long as they comment on the large sample size.

Question 6 (cont'd.)

Part 3: Correct mechanics, including the value of the test statistic, df, and P-value (or rejection region).

```
          N      Mean     StDev   SE Mean
1990    200      1.83      1.29     0.091
2000    200      1.93      1.37     0.097
```

t = -.751 df = 396 P-value = .453

(or based on tables P-value > .20) Also OK to use conservative df of 199

OR Calculator using sample statistics:
t = .7515, P-value = .4527, df = 396.57
Calculator using raw data
t = .7512, P-value = .4529, df = 396.67

OR use a confidence interval approach:

```
Difference = mu 1990 - mu 2000
Estimate for difference:  -0.100
90 percent CI for difference: (-0.319, 0.119)
```

OR use a "pooled" procedure
t = -.751 p=.452 df = 398

OR if use rejection region, $t_{.05} = \pm 1.660$

Part 4: States a correct conclusion in the context of the problem, linking to the result of the statistical test.

Because the P-value > α, fail to reject H_0. There is not convincing evidence that the mean number of lab classes in 2000 is different than it was in 1990.

Part (b):

Answers question about (a): The test in (a) does not answer this question because the distributions could be different even though the means are the same.

Part 1: States correct hypotheses.
H_0: The distribution of number of lab classes is the same in 2000 as it was in 1990.
H_a: The distribution of number of lab classes is not the same in 2000 as in 1990.

Question 6 (cont'd.)

Part 2: Identifies the correct test (by name or by formula) and checks (not just lists) appropriate assumptions.

Chi-square test (of homogeneity): $\chi^2 = \sum \frac{(observed - \exp ected)^2}{\exp ected}$

Conditions: Sample size is large. To check, must compute the expected counts:

1990 Expected	2000 Expected
24	24
67	67
59	59
19	19
21	21
10	10

All expected counts are greater than or equal to 5 (or 10), so it is OK to proceed with the chi-square test.

Part 3: Correct mechanics.

```
Chi-Sq =  0.667 +  0.667 + 0.373 +  0.373 +
          0.017 +  0.017 + 4.263 +  4.263 +
          1.190 +  1.190 + 0.400 +  0.400 = 13.821
DF = 5, P-Value = 0.017
```

Part 4: States a correct conclusion in context using the result of the statistical test.

Because the P-value < α (problem states $\alpha = .10$), reject H_0. There is convincing evidence that the distribution of number of lab classes is not the same in 2000 as it was in 1990.

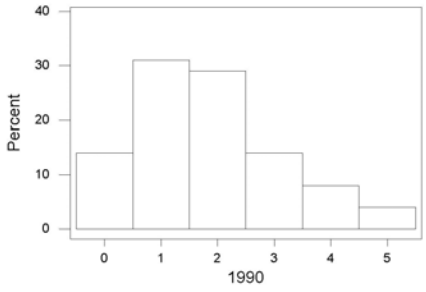
Question 6 (cont'd.)

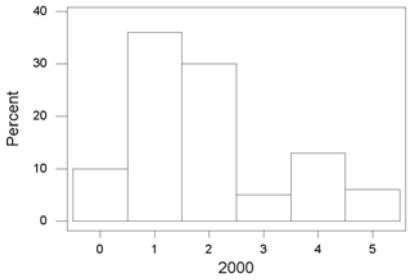
Part (c):

Although the two distributions are centered in about the same place, the shape of the distributions is different. In 2000, the proportion of students taking three labs was smaller than in 1990.

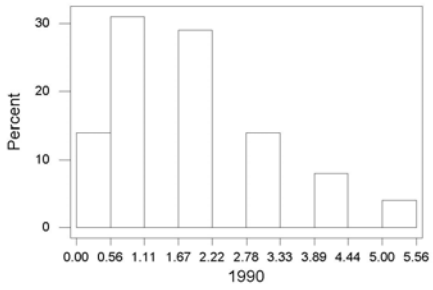
The statement must be backed up by appropriate graphs, such as two histograms (with appropriate scales) or two boxplots.

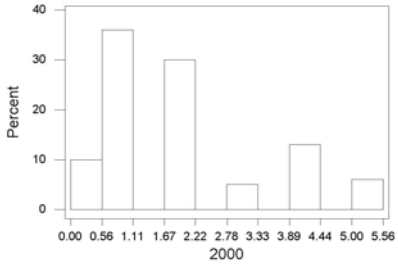
Histograms





NOTE: If the student just uses the default intervals on the calculator, they will get odd looking histograms. (This is a minus.)

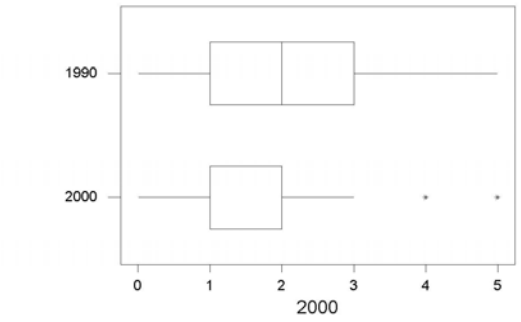




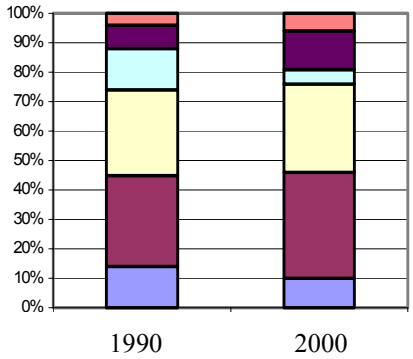
Question 6 (cont'd.)

Other acceptable graphs include:

Boxplots:



Segmented Bar Graph



Scoring

Part (a) is

Essentially correct if
All four parts of the test are essentially correct (except minor arithmetic errors). Since there are so many parts, do not count off if degrees of freedom are omitted. Not defining μ_1 and μ_2 is a minus.

Almost essentially correct if
Three parts are essentially correct and communication is very good.

Partially correct if
Two or three of the parts of the test are essentially correct.

NOTES:
1. Do not need to state that random samples are used since it is in problem statement.
2. If student checks on normality and comments that the histogram does not look normal or that a boxplot has outliers and then says the t test is not appropriate (i.e., fails to recognize that it is OK to proceed with a t test based on the large sample size), part two is not correct.
3. Mentioning samples are independent is a plus.

Part (b) is

Essentially correct if
The student recognizes that part (a) does not answer the question posed and has all four parts of the chi-square test completely correct.

Question 6 (cont'd.)

Almost essentially correct if

The student recognizes that part (a) does not answer the question posed and has three parts essentially correct, including mechanics, and communication is very good.

Partially correct if

The student recognizes that part (a) does not answer the question posed and has two or three parts of the chi-square test completely correct.

NOTES:

1. Part (b) is incorrect if the student says the test in (a) does not answer the question posed. However, if the student rejects the null hypothesis in (a), a "yes" answer to (b) could be correct.
2. If the student says, "No. We should do a chi-square test," but doesn't do one, this is a plus.
3. If student uses the 1990 data as the expected counts and the 2000 data as the observed counts and finds χ^2=23.789, the student does not receive credit for parts two or three (but can still get a P).

Part (c) is

Essentially correct if

response has a coherent description of the difference between the 1990 and 2000 distributions and gives appropriate graphs.

NOTE: If part (b) is incorrect because the student thinks that the test in (a) does answer the question posed, but does good graphs and bases the argument in part (c) on the graphs, part (c) can be scored as essentially correct.

Partially correct if

explanation is weak (such as describing the difference only in terms of comparing the means) or missing but reasonable graphs are included
OR
incorrect graphs are included but the statements made about the distributions are consistent with the given graphs.
OR
graphs are missing but analysis extends the results of (a) and (b).

NOTE: Part (c) should be considered incorrect if graphs that use the frequencies as data are constructed.

If student falls between two scores, read the paper holistically, keeping in mind the number of major errors and quality of communication (4 = complete response, 3 = substantial response, 2 = developing response, 1 = minimal response).

Question 6 (cont'd.)

4 Complete Response

Essentially correct on all 3 parts.

3 Substantial Response

Essentially correct on any two parts and partially correct or incorrect on the other part.

2 Developing Response

Essentially correct on one part and partially correct on at least one of the other two parts

OR

Partially correct on all three parts.

1 Minimal Response

Essentially correct on one part and incorrect on the other two

OR

Partially correct on one or two parts and incorrect on the other(s).