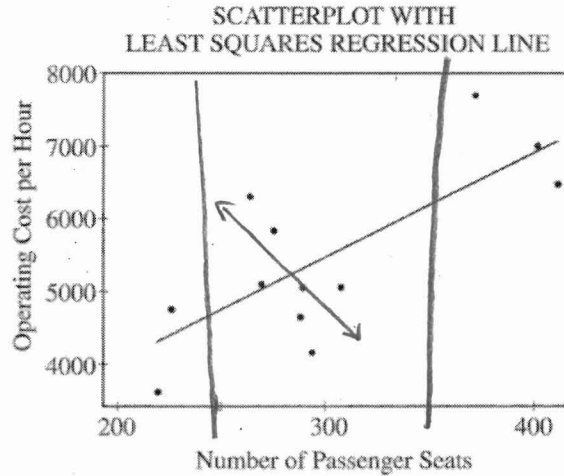




Chapter 3 "FRAPPY" {Free Response AP Problem...Yay!}

The following problem is taken from an actual Advanced Placement Statistics Examination. Your task is to generate a complete, concise statistical response in 15 minutes. You will be graded based on the AP rubric and will earn a score of 0-4. After grading, keep this problem in your binder for your AP Exam preparation.

Commercial airlines need to know the operating cost per hour of flight for each plane in their fleet. In a study of the relationship between operating cost per hour and number of passenger seats, investigators computed the regression of operating cost per hour on the number of passenger seats. The 12 sample aircraft used in the study included planes with as few as 216 passenger seats and planes with as many as 410 passenger seats. Operating cost per hour ranges between \$3,600 and \$7,800. Some computer output from a regression analysis of these data is shown below.



Predictor	Coef	StDev	T	P
Constant	1136	1226	0.93	0.376
Seats	14.673	4.027	3.64	0.005

S=845.3 R-Sq=57% R-Sq(adj)=52.7%

Scoring:

E P I

- (a) What is the equation of the least squares regression line that describes the relationship between operating cost per hour and number of passenger seats in the plane? Define any variables used in this equation.

$$\widehat{\text{op-cost}} = 1136 + 14.673(\text{seats})$$

E P I

- (b) What is the value of the correlation coefficient for operating cost per hour and number of passenger seats in the plane? Interpret this correlation.

$$= \pm \sqrt{R\text{-Sq}} = \pm \sqrt{.57} = \pm .755$$

uses slope for sign

The relationship between operating costs and number of seats is moderately strong and positive.

E P I

- (c) Suppose that you want to describe the relationship between operating cost per hour and number of passenger seats in the plane for planes only in the range of 250 to 350 seats. Does the line shown in the scatterplot still provide the best description of the relationship for data in this range? Why or why not?

No, in the interval of 250 to 350, the relationship between the seats and cost would be negative.

Total: ___/4

AP[®] STATISTICS
2002 SCORING GUIDELINES

Question 4 (cont'd.)

Part (a) addresses the first element.

Element one is:

essentially correct if the solution has the correct equation and variables are defined correctly.

partially correct if only the equation is correct.

incorrect if the equation is not stated correctly.

Part (b) addresses the second and third elements.

Element two is:

essentially correct if the student's solution states that $r = 0.755$.

partially correct if the student's solution only states that $r = \pm 0.755$.

incorrect if the student states any other value of r including $r = 0.726$ (square root of R-Sq (adj)).

Element three is:

essentially correct if the student's solution

addresses, based on a correct understanding of the correlation coefficient, three or four of the following:

- type of relationship
- strength
- direction
- context

OR

states, based on a correct understanding of r^2 :

- that 57 percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats

AND

- that the relationship is positive.

Note: If the student gives a correct interpretation of r but then incorrectly explains r^2 , this is considered a parallel solution and is incorrect.

AP[®] STATISTICS
2002 SCORING GUIDELINES

Question 4 (cont'd.)

partially correct if the student's solution

addresses exactly two of the following – type of relationship (linear), strength, direction, and context (based on a correct understanding of the correlation coefficient).

OR

only states that 57 percent of the variability in operating cost per hour can be explained by a linear relationship between cost and number of passenger seats (based on a correct understanding of r^2) – BUT – does not state that the relationship is positive.

NOTE: Element three may be scored essentially or partially correct if the student uses a reasonable r (between 0 and 1) or R-Sq (adj) value.

Part (c) addresses the fourth element.

Element four is **essentially correct** if the student's solution

states that the existing line is not a good fit for the remaining seven points and correctly explains that the restricted data has a negative correlation or the recalculated least-squares regression line has a negative slope.

Element four is **partially correct** if the student's solution

explains why the existing line is not a good fit for the remaining seven points but does not communicate that the restricted data has a negative correlation or the recalculated least-squares regression line has a negative slope.

OR

removes fewer than the specified five points, but gives a correct interpretation of the effect on the correlation or slope of the least-squares regression line.

4 = 4 elements correct
3 = 3 elements correct
⋮
ect

This one will NOT be on the final



Chapter 4 "FRAPPY" {Free Response AP Problem...Yay!}

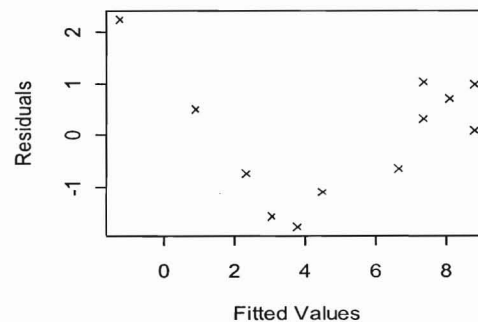
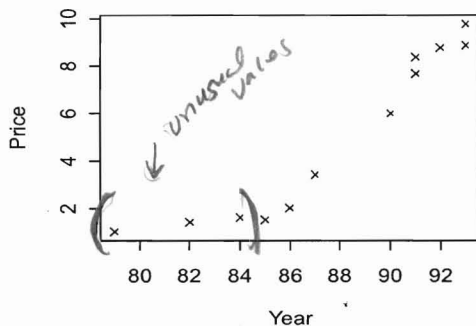
The following problem is taken from an actual Advanced Placement Statistics Examination. Your task is to generate a complete, concise statistical response in 30 minutes - You should expect to spend 30-35 minutes on the "Investigative Task". You will be graded based on the AP rubric and will earn a score of 0-4. After grading, keep this problem in your binder for your AP Exam preparation.

You are planning to sell a used 1988 automobile and want to establish an asking price that is competitive with that of other cars of the same make and model that are on the market. A review of newspaper advertisements for used cars yields the following data for 12 different cars of this make and model. You want to fit a least squares regression model to these data for use as a model in establishing an asking price for your car.

Production Year	1990	1991	1992	1987	1993	1991	1993	1985	1984	1982	1986	1979
Asking Price (thousands of dollars)	6.0	7.7	8.8	3.4	9.8	8.4	8.9	1.5	1.6	1.4	2.0	1.0

The computer printouts for three different linear regression models are shown below. Model 1 fits the asking prices as a function of the production year, Model 2 fits the natural logarithm of the asking price as a function of the production year, and Model 3 fits the square root of the asking price as a function of the production year. Each printout also includes a plot of the residuals from the linear model versus the fitted values, as well as additional descriptive data produced from the least squares procedure.

Model 1



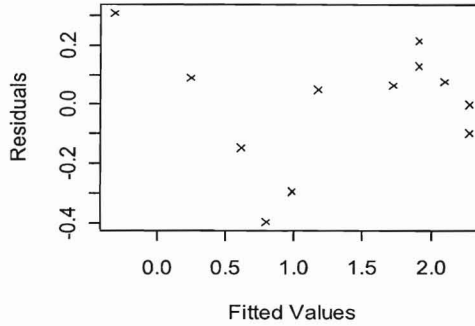
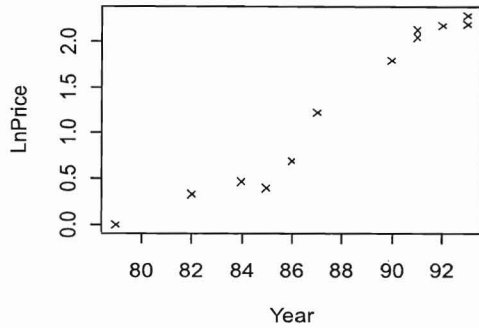
The regression equation is $\text{Price} = -58.1 + 0.719 \text{ Year}$.

$s = 1.255$ $R\text{-sq} = 88.5\%$

Variable	Coefficient	Stdev	t-ratio	prob
Constant	-58.0503	7.205	-8.06	0.000
Year	0.718997	0.082	8.77	0.000

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	121.097	1	121.097	76.9
Residual	15.7521	10	1.57521	

Model 2



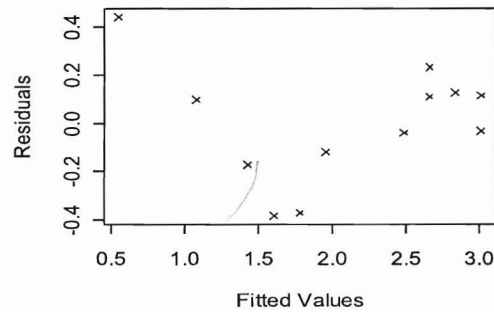
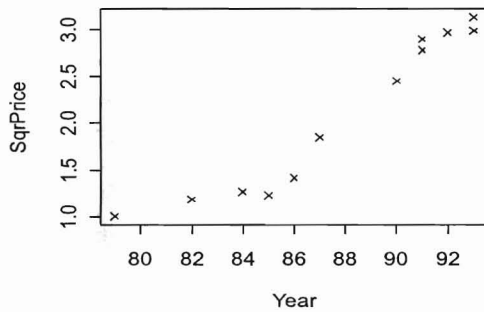
The regression equation is $\text{LnPrice} = -14.9 + 0.185 \text{ Year}$.

$s = 0.213$ $R\text{-sq} = 94.6\%$

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	8.01898	1	8.01898	177
Residual	0.453645	10	0.0453645	

Variable	Coefficient	Stdev	t-ratio	prob
Constant	-14.9244	1.223	-12.2	0.000
Year	0.185021	0.01392	13.3	0.000

Model 3



The regression equation is $\text{SqrPrice} = -13.3 + 0.176 \text{ Year}$.

$s = 0.252$ $R\text{-sq} = 91.9\%$

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	7.22212	1	7.22212	114
Residual	0.635106	10	0.0635106	

Variable	Coefficient	Stdev	t-ratio	prob
Constant	-13.3133	1.447	-9.2	0.000
Year	0.175587	0.01647	10.7	0.000

E P I

(a) Use Model 1 to establish an asking price for your 1988 automobile.

$$\hat{\text{price}} = -58.1 + 0.719 \text{ year} \quad (8888)$$

$$\hat{\text{price}} = -58.1 + 0.719(88) = 5.172 \text{ thousands}$$

I should ask for \$5,172.00

E P I

(b) Use Model 2 to establish an asking price for your 1988 automobile.

$$\ln \hat{\text{price}} = -14.9 + 0.185(88)$$

$$\ln \hat{\text{price}} = 1.38$$

$$\hat{\text{price}} = e^{1.38} = 3.97 \text{ thousands}$$

I should ask for \$3,974.90

E P I

(c) Use Model 3 to establish an asking price for your 1988 automobile.

$$\sqrt{\hat{\text{price}}} = -13.3 + 0.176 \text{ year}$$

$$\sqrt{\hat{\text{price}}} = -13.3 + 0.176(88)$$

$$\sqrt{\hat{\text{price}}} = 2.188^2 \quad \hat{\text{price}} = 4.787344 \text{ thousands}$$

I should ask for \$4,787.34

E P I

(d) Describe any shortcomings you see in these three models.

The residuals plot for all 3 models show a pattern. None of the models are "appropriate."

E P I

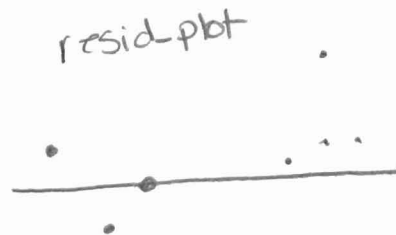
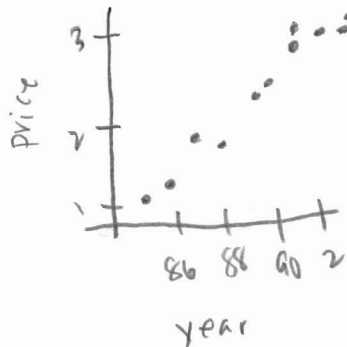
(e) Use some or all of the given data to find a better method for establishing an asking price for your 1988 automobile. Explain why your method is better.

Dropping the years 86-84 the line follows a more linear pattern.

$$\hat{\text{price}} = -87.2837 + 1.04215(\text{year})$$

$$r^2 = .978332$$

$$r = 0.989106$$



Total: ___/4

This residual plot doesn't appear to have a pattern and has a higher r^2 value.

$$\hat{\text{price}} = -87.2837 + 1.04215(88) = 4.4255$$

I should ask for \$4,425.50

Question 6 (Investigative Task) Scoring Guidelines

4 Complete Response

- I. Correctly estimates the asking price in dollars for at least two of the three models, including back transforming the predicted value for at least one of models b or c.

a. $\text{Price} = -58.1 + 0.719(88) = 5.172$
asking price is \$5,172

b. $\ln(\text{price}) = -14.9 + 0.185(88) = 1.380$
 $\exp(1.38) = 3.9749$
asking price = \$3,975

c. $\text{Sqrt}(\text{price}) = -13.3 + 0.176(88) = 2.188$
 $(2.188)^2 = 4.7873$
asking price = \$4,787

- II. Describes the major shortcoming to be the non-linear pattern in the scatterplot or residual plot for all three models.

- III. Suggests a new model that successfully deals with the non-linearity of the data. The prime contenders for this model are:

Fit a simple linear model after the first two or three years are dropped.

($R^2 = .978$ with little pattern in the residuals after dropping 79, 82 and 84.)

Fit separate linear models to the earlier and later years.

($R^2 = .985$ for the years 86 to 93, with little pattern in the residuals.)

Fit a model that attempts to model the curvature in the data. For example, fit a quadratic model to all of the data.

($R^2 = .974$ with little pattern in the residuals for a quadratic model.)

- IV. Justifies why this model is better. For example, comments that there is no pattern in the residuals as seen from looking at the fitted model on the scatterplot or from looking at the residual plot.

3 Substantial Response

Fails to do one of the following satisfactorily:

- Part I (for example, by failing to back transform the prediction)
- describe the major shortcomings of the three models
- justify why the new model is better (for example, by failing to comment on the residual plot)

2 Developing Response

Fails to do one of the following satisfactorily:

- Part I and describe the major shortcomings of the three models
- Part I and justify why the new model is better
- suggest a new model and justify it
- describe the major shortcomings of the three models and justify why the new model is better.

1 Minimal Response

Gives correct responses to one of the items I, II, or III above. **Note:** Students cannot get IV correct without first specifying a new model.



Chapter 5.1 "FRAPPY" {Free Response AP Problem...Yay!}

The following problem is taken from an actual Advanced Placement Statistics Examination. Your task is to generate a complete, concise statistical response in 15 minutes. You will be graded based on the AP rubric and will earn a score of 0-4. After grading, keep this problem in your binder for your AP Exam preparation.

At a certain university, students who live in the dormitories eat at a common dining hall. Recently, some students have been complaining about the quality of the food served there. The dining hall manager decided to do a survey to estimate the proportion of students living in the dormitories who think that the quality of the food should be improved. One evening, the manager asked the first 100 students entering the dining hall to answer the following question.

Many students believe that the food served in the dining hall needs improvement. Do you think that the quality of food served here needs improvement, even though that would increase the cost of the meal plan?

Yes No No opinion

Scoring:

(a) In this setting, explain how bias may have been introduced based on the way this convenience sample was selected and suggest how the sample could have been selected differently to avoid that bias.

E P I

The manager asked only the 1st 100 students (a convenience sample), there may be a difference in the way these students feel about the food quality as compared to the student population in general. (For example, people who show up earlier may be more enthusiastic about the food or people who don't like the food may not show up at all.)

The manager should have taken a SRS of all dormitory residents.

(b) In this setting, explain how bias may have been introduced based on the way the question was worded and suggest how it could have been worded differently to avoid that bias.

E P I

The question mentions that improving the quality of the food would increase the cost of the meal plan, biasing students to choose no.

This bias could be avoided by wording it, "Do you think the quality of food served here needs improvement."

Total: /4

**AP[®] STATISTICS
2004 SCORING GUIDELINES (Form B)**

Question 2 (cont'd.)

Part (b) is essentially correct if the response

1. points out at least one of the two possible problems with the question wording
2. proposes reasonable alternate wording that addresses the concern(s) raised.

Note: The student only needs to identify one problem and take care of it. If only one of the two wording concerns is raised, the alternate wording need only address the one wording problem. However, if the student identifies both problems, appropriate alternate wording must be provided for both problems.

Part (b) is partially correct if the response

identifies one or both of the potential wording problems, but does not propose a new wording that adequately addresses an identified problem.

Part (b) is incorrect if the response

points out both wording problems, but then argues that the question wording is OK as is because the biases are in opposing directions and so will balance each other.

4 Complete Response

Both parts essentially correct

3 Substantial Response

One part essentially correct and the other part partially correct

2 Developing Response

One part essentially correct and the other part incorrect
OR
Both parts partially correct

1 Minimal Response

One part partially correct



Chapter 5.2 "FRAPPY" {Free Response AP Problem...Yay!}

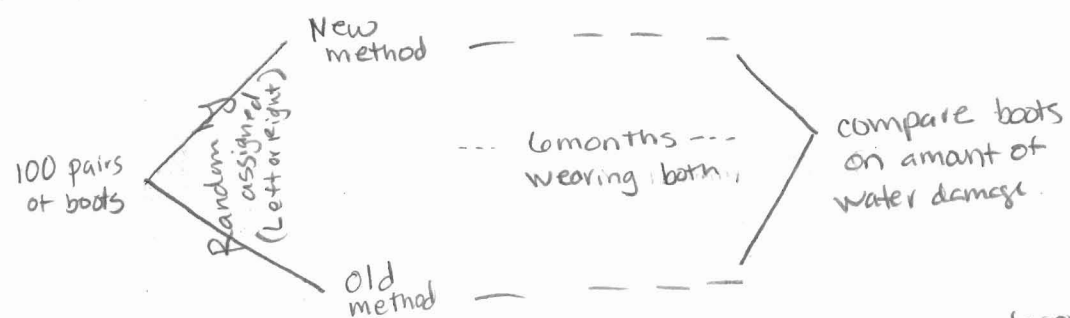
The following problem is taken from an actual Advanced Placement Statistics Examination. Your task is to generate a complete, concise statistical response in 15 minutes. You will be graded based on the AP rubric and will earn a score of 0-4. After grading, keep this problem in your binder for your AP Exam preparation.

A manufacturer of boots plans to conduct an experiment to compare a new method of waterproofing to the current method. The appearance of the boots is not changed by either method. The company recruits 100 volunteers in Seattle, where it rains frequently, to wear the boots as they normally would for 6 months. At the end of the 6 months, the boots will be returned to the company to be evaluated for water damage.

Scoring:

(a) Describe a design for this experiment that uses the 100 volunteers. Include a few sentences on how it would be implemented.

E P I



- Each subject will receive a pair of boots - one with each waterproofing method. To decide which shoe has the new method a coin will be flipped: heads = Left foot new ; tails = right foot new, left foot old. (This is called a matched-pair design).

After 6 months of wear - the left and right will be compared for degree of wear.

(b) Could your design be double blind? Explain.

E P I

Yes. Neither the person wearing the shoes nor the people evaluating the amount of wear need to know which treatment method was assigned to which boot!

Total: ___/4

AP[®] STATISTICS
2002 SCORING GUIDELINES

Question 2

Solution

Part (a):

A **paired design** is used in which each subject receives a pair of boots where one boot is treated with the new method and the other with the current method.

Subjects should be randomly assigned to one of two groups. Group 1 would have the new method applied to the right boot; group 2 would have the new method applied to the left boot.

OR

For each subject, whether the new method is applied to the right or left boot is determined at random.

OR

A **crossover design** is used in which each subject receives a pair of boots, both of which were treated with one treatment. The boots are used for three months and then exchanged for a second pair of boots, both of which were treated with the other treatment. These boots are then used for the next three months.

Subjects should be randomly assigned to one of two groups. One group receives boots with the new treatment first and the other group receives boots with the current method first.

NOTE: Additional appropriate blocking schemes are considered extraneous.

Part (b):

The design could be double blind, as long as both the *subjects* and the person *evaluating* the boots for water damage do not know which boots were treated with the new method and which were treated with the current method.

NOTE: If the student does something unexpected in part (a) and gives a design that actually cannot be double blind, then part (b) could be considered correct provided the response explains why the design could not be double blind.

Scoring

A student response is scored as **E** (essentially correct), **P** (partially correct), or **I** (incorrect) for each of the following key elements:

1. **Design**

- **E** - paired design (may be described as blocking on individual) or crossover design
- **P** - 2 or more groups (e.g., Completely Randomized Design)
- **I** - no grouping or grouping with no treatments specified

2. **Implementation:** Randomization appropriate to the design

- **E** - Written description of appropriate randomization
- **P** - Incomplete or incorrect description of randomization
- **I** - No description of randomization

NOTE: (1) Diagram alone can be scored at most a **P**.

(2) The randomization must apply to the allocation or assignment of subjects to the treatment groups or the allocation of treatments to the subjects.

(3) Randomization to select the 100 volunteers without assignment to the treatment groups is scored an **I**.

**AP[®] STATISTICS
2002 SCORING GUIDELINES**

Question 2 (cont'd.)

3. **Double blind:** Explanation in parts (a) and/or (b) that shows understanding of what it means for an experiment to be double blind.
- **E** - response indicates that blinding applies to both the evaluator and subjects.
 - **P** - response recognizes that blinding applies to the subjects and at least one other party, whether or not they think that this can be accomplished; the other party may not be correctly identified.
 - **I** - response fails to recognize that both the subject and another party must be blinded or is missing or irrelevant.

Score as Design - Randomization - Double Blind

4 Complete Response

E E E

3 Substantial Response

Any one of the following combinations:

E E P P E E P E P *
E E I
E P E

2 Developing Response

Any one of the following combinations:

E P P P E I I E E P E P *
E P I P P E I P E
E I E P P P
E I P P I E

1 Minimal Response

Any one of the following combinations:

E I I P P I I E P
I E I P I P
I I E I P P

AP[®] STATISTICS
2002 SCORING GUIDELINES

Question 2 (cont'd.)

0 No Credit

P II I I I
I P I
I I P

* **P E P** may be scored as either a **2** or a **3**:

- (1) If the description of the randomization *only* says, “Randomly allocate”, then score P E P a 2.
- (2) If the description of the randomization says, “Randomly allocate”, but also contains greater detail about the randomization or the inclusion of blocking in the design or other statistical thinking, then score P E P a 3.



Chapter 6a "FRAPPY" {Free Response AP Problem...Yay!}

The following problem is taken from an actual Advanced Placement Statistics Examination. Your task is to generate a complete, concise statistical response in 15 minutes. You will be graded based on the AP rubric and will earn a score of 0-4. After grading, keep this problem in your binder for your AP Exam preparation.

A simple random sample of adults living in a suburb of a large city was selected. The age and annual income of each adult in the sample were recorded. The resulting data are summarized in the table below.

Age Category	Annual Income			Total
	\$25,000-\$35,000	\$35,001-\$50,000	Over \$50,000	
21-30	8	15	27	50
31-45	22	32	35	89
46-60	12	14	27	53
Over 60	5	3	7	15
Total	47	64	96	207

Scoring:

(a) What is the probability that a person chosen at random from this sample will be in the 31-45 age category?

E P I

$$P(31-45) = 89/207 = 0.42995$$

(b) What is the probability that a person chosen at random from those in this sample whose incomes are over \$50,000 will be in the 31-45 age category? Show your work.

E P I

$$P(31-45 | > 50,000) = 35/96 = 0.36458$$

(c) Based on your answers to parts (a) and (b), is annual income independent of age category for those in this sample? Explain.

E P I

$$P(31-45) \neq P(31-45 | > 50,000)$$

$$0.42995 \neq 0.36458$$

Annual income is not independent of age category, individuals are less likely to be 31-45 if their income is > \$50,000

Total: ___/4

AP[®] STATISTICS
2003 SCORING GUIDELINES (Form B)

Question 2

Solution

Part (a): $P(\text{age } 31 - 45) = \frac{89}{207} = 0.42995$

Part (b): $P(\text{age } 31 - 45 | \text{income over } 50,000) = \frac{35}{96} = 0.36458$

Part (c):

If annual income and age were independent, the probabilities in (a) and (b) would be equal. Since these probabilities are not equal, annual income and age category are not independent for adults in this sample.

Scoring

Part (a) is scored as either essentially correct (E) (may be minor arithmetic errors) or incorrect (I).

Part (b) is

Essentially correct (E) if the conditional probability is correctly calculated.

Partially correct (P) if the student reverses the conditioning, calculating

$$P(\text{income over } 50,000 | \text{age } 31 - 45) = \frac{35}{89} = 0.3933$$

OR

calculates the correct probability for the wrong column, e.g., $\frac{32}{64}$

Incorrect (I) if the student calculates the joint probability: $\frac{35}{207} = 0.169$

Part (c) is

Essentially correct (E) if the student

1. indicates that the two variables are not independent
AND
2. the explanation is tied to the fact that the probabilities in parts (a) and (b) are not equal (the answer must be based on parts (a) and (b))

AP[®] STATISTICS
2003 SCORING GUIDELINES (Form B)

Question 2 (cont'd)

Partially correct (P) if the student indicates that the two variables are not independent, but the explanation is incorrect, or is not based on the answers to parts (a) and (b); i.e., performing new correct calculations instead of referring to those in parts (a) and (b). For example: determining the probability of the intersection and comparing to the two individual probabilities

$\left(\frac{35}{207} = 0.169, \text{ which does not equal } \frac{96}{207} \cdot \frac{89}{207} = (0.43)(0.46) \right)$, reversing conditions

$\left(\text{e.g., } \frac{96}{207} = 0.464, \text{ which does not equal } \frac{35}{89} = 0.393 \right)$, or other conditional probability comparisons.

Incorrect (I) if the student fails to give a numerical justification to support the argument.

OR

Incorrect if the student does one of the following

- performs an incorrect additional calculation
- says the variables are independent based entirely on the context.
- performs a chi-square test ($\chi^2 = 5.38$, $p\text{-value} = 0.496$) since this addresses independence in the population instead of the sample
- only states “yes, independent” with no justification

NOTE: If either of the probabilities calculated in (a) or (b) are incorrect, part (c) should be scored as if those probabilities were correct. For example, if the student incorrectly calculated the same answer for parts (a) and (b), part (c) would be scored as correct if the student states that you can't tell if the two variables are independent because you would need to check all age-gender combinations.

4 Complete Response (EEE)

All three parts essentially correct

3 Substantial Response (EEP, EPE, EPP, IEE)

Part (a) essentially correct and parts (b) and (c) at least partially correct

OR

Part (a) incorrect and parts (b) and (c) essentially correct

2 Developing Response (EEI, EIE, EPI, EIP, IEP, IPE, IPP)

Part (a) essentially correct and one (but not both) of parts (b) and (c) correct

OR

Part (a) incorrect and both parts (b) and (c) at least partially correct

1 Minimal Response (EII, IPI, IIP, IEI, IIE)

Part (a) essentially correct and parts (b) and (c) incorrect

OR

Part (a) incorrect and one of parts (b) and (c) partially correct



Chapter 6b "FRAPPY" {Free Response AP Problem...Yay!}

The following problem is taken from an actual Advanced Placement Statistics Examination. Your task is to generate a complete, concise statistical response in 15 minutes. You will be graded based on the AP rubric and will earn a score of 0-4. After grading, keep this problem in your binder for your AP Exam preparation.

Airlines routinely overbook flights because they expect a certain number of no-shows. An airline runs a 5 P.M. commuter flight from Washington, D.C., to New York City on a plane that holds 38 passengers. Past experience has shown that if 41 tickets are sold for the flight, then the probability distribution for the number who actually show up for the flight is as shown in the table below.

Number who actually show up	36	37	38	39	40	41
Probability	0.46	0.30	0.16	0.05	0.02	0.01

Scoring:

Assume that 41 tickets are sold for each flight.

(a) There are 38 passenger seats on the flight. What is the probability that all passengers who show up for this flight will get a seat?

E P I

$$P(\text{all get a seat}) = P(\# \text{ show up} \leq 38) = 0.46 + 0.30 + 0.16 = .92$$

(b) What is the expected number of no-shows for this flight?

E P I

no shows = 41 - # show up

$X = \text{no shows}$	5	4	3	2	1	0
$P(x)$	0.46	0.3	0.16	0.05	0.02	0.01

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

(c) Given that not all passenger seats are filled on a flight, what is the probability that only 36 passengers showed up for the flight?

E P I

$$P(36 \mid \text{not all filled, ie. } < 38) = \frac{P(36 \cap \text{not filled})}{P(\text{not filled})} = \frac{.46}{.46 + .3} = .6653$$

4 = each answer correct & work shown!

Total: ___/4

AP[®] STATISTICS
2002 SCORING GUIDELINES (Form B)

Question 2 (cont'd.)

Scoring

Part (a) is

Essentially correct if computes $P(X \leq 38)$ (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 $P(X \geq 38) = 0.24$ 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 $P(X = 36 | X < 41) = \frac{0.46}{0.99} = 0.465$ or

$P(X = 36 | X \leq 38) = \frac{0.46}{0.92} = 0.5$ or incorrectly tries to solve the conditional probability (e.g., multiplies probabilities in numerator).

Incorrect if an unconditional probability is computed.

**AP[®] STATISTICS
2002 SCORING GUIDELINES (Form B)**

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).



Chapter 7.1 "FRAPPY" {Free Response AP Problem...Yay!}

The following problem is taken from an actual Advanced Placement Statistics Examination. Your task is to generate a complete, concise statistical response in 15 minutes. You will be graded based on the AP rubric and will earn a score of 0-4. After grading, keep this problem in your binder for your AP Exam preparation.

A department supervisor is considering purchasing one of two comparable photocopy machines, A or B. Machine A costs \$10,000 and Machine B costs \$10,500. This department replaces photocopy machines every three years. The repair contract for Machine A costs \$50 per month and covers an unlimited number of repairs. The repair contract for Machine B costs \$200 per repair. Based on past performance, the distribution of the number of repairs needed over any one-year period for Machine B is shown below.

Number of Repairs	0	1	2	3
Probability	0.50	0.25	0.15	0.10

You are asked to give a recommendation based on overall cost as to which machine, A or B, along with its repair contract, should be purchased. What would your recommendation be? Give statistical justification to support your recommendation.

Scoring:

E P I

$$E(\overset{\#}{\text{repairs}}) = 0(.5) + 1(.25) + 2(.15) + 3(.1) = .85$$

E P I

$$E(\# \text{ repairs over 3 years}) = 2.55$$

Machine A total cost = (including repairs)
(3 years, 36 repairs)
\$10,000 + \$50(36) = \$11,800

E P I

machine B total cost = (including repairs)
(cost of repairs)
10,500 + 200(2.55) = \$11,010

E P I

They should choose machine B, since the expected total cost is lower!

Total: ___/4

AP[®] STATISTICS
2001 SOLUTIONS AND SCORING GUIDELINES

Question 2 - Solution

For A:

Total 3-year cost: $\$10,000 + 36(\$50) = \$10,000 + \$1,800 = \$11,800$
This cost is fixed.

For B:

Expected number of repairs in 3 years = $3[0(.5) + 1(.25) + 2(.15) + 3(.1)] = 3(.85) = 2.55$
Expected cost of repairs in 3 years = $3(\$200)(0.85) = \510
Expected 3-year cost = $\$10,500 + \$510 = \$11,010$

Choice:

Choose B because it has a lower expected (or average) cost. (A has a fixed cost that is \$790 (\$11,800 - \$11,010) higher than the expected cost of B.)

Scoring

The solution should include the following four elements:

1. Correct calculation of 3-year cost for A.
2. Correct calculation of a relevant expected **value** for B (expected number of repairs per year or per 3 years or expected cost of repairs per year or per 3 years). Calculation of expected value must be shown.
3. Correct calculation of expected **total cost** for B.
4. Choice of B with a *complete and coherent* explanation that is based on student's prior calculations for A & B.

"Complete and coherent" means that:

- costs for A & B are compared
AND
- B's cost has been indicated as "expected" or "average" or "mean" or "estimated" or "approximate" or "predicted," etc.

- 4 Complete Response**
Solution includes all four of the required elements.
- 3 Substantial Response**
Solution includes three of the required elements.
- 2 Developing Response**
Solution includes two of the required elements.
- 1 Minimal Response**
Solution includes one of the required elements.

AP® STATISTICS
2001 SOLUTIONS AND SCORING GUIDELINES

Question 2 (cont'd.)

Notes:

1. If calculations are based on 1-year costs rather than 3-year costs, and then the student chooses A with explanation, the student can earn a score of up to 3.

Total 1-year cost for A: $\$10,000 + 12(\$50) = \$10,000 + \$600 = \$10,600$

And for B:

Expected number of repairs in 1 year = $0(.5) + 1(.25) + 2(.15) + 3(.1) = .85$

Expected cost of repairs in 1 year = $\$0(.5) + \$200(.25) + \$400(.15) + \$600(.1) = \$170$

Expected 1-year cost for B: $\$10,500 + \$170 = \$10,670$

2. If initial purchase prices are omitted from the calculations, the student can earn a score of up to 3.
3. Rounded calculations of the expected number of repairs for B: If a student rounds the expected number of repairs per year (.85) to 1, or rounds the expected number of repairs in 3 years (2.55) to 3, the maximum score is a 3. If the student identifies the rounded value as an upper bound on the expected cost, the paper may earn a maximum score of 4.
4. If choice of A or B is not based on expected cost for B, the student can still present a complete response. To earn 4 points with this solution, relevant and complete statistical reasoning must be demonstrated. This solution must include:

- a. Decision based on break-even analysis:

1 point -- Correct calculation of 3-year fixed cost for A (\$11,800)

1 point -- Correct calculation that 7 or more repairs in the 3-year period would be necessary for B's cost to exceed A's 3-year cost

1 point -- Says or calculates that the probability of 7 or more repairs for B is small and therefore chooses B,

OR

Chooses A because the probability of 7 or more repairs for B is not 0 and they want to guard against the possibility of paying more for B than A's fixed cost

1 point -- Correctly calculates that the probability of 7 or more repairs for B is 0.01975 (or about 0.02, or about 2 percent of the time B's cost will exceed A's cost)

AND

States that this analysis depends on the assumption that repairs from year to year are independent

**AP[®] STATISTICS
2001 SOLUTIONS AND SCORING GUIDELINES**

Question 2 (cont'd.)

b. Decision based on minimax analysis may earn a maximum of 3 points:

1 point -- Correct calculation of 3-year fixed cost for A (\$11,800)

1 point -- Correct calculation of range of possible 3-year costs for B
 $\$10,500 \leq \text{cost of B} \leq \$12,300$

AND

Probability calculation showing that chance of observing maximum cost is small
(e.g., 0.001 or 0.1 percent that B costs \$12,500, assuming independence)

1 point -- Relevant statistical justification for choice of A or B:

Gives convincing reasoning for minimizing maximum cost (minimax) and therefore chooses A. Student might argue, for example, that a company may prefer a known fixed cost to a variable cost that could be smaller but also has the chance of being larger.



Chapter 7.2 "FRAPPY" {Free Response AP Problem...Yay!}

The following problem is taken from an actual Advanced Placement Statistics Examination. Your task is to generate a complete, concise statistical response in 15 minutes. You will be graded based on the AP rubric and will earn a score of 0-4. After grading, keep this problem in your binder for your AP Exam preparation.

There are 4 runners on the New High School team. The team is planning to participate in a race in which each runner runs a mile. The team time is the sum of the individual times for the 4 runners. Assume that the individual times of the 4 runners are all independent of each other. The individual times, in minutes, of the runners in similar races are approximately normally distributed with the following means and standard deviations.

	Mean	Standard Deviation
Runner 1	4.9	0.15
Runner 2	4.7	0.16
Runner 3	4.5	0.14
Runner 4	4.8	0.15

Scoring:

E P I

- (a) Runner 3 thinks that he can run a mile in less than 4.2 minutes in the next race. Is this likely to happen?



$$P(X < 4.2) = P\left(Z < \frac{4.2 - 4.5}{.14}\right) \\ = P(Z < -2.14) \\ \approx 0.016$$

There is only a 1.6% chance that Runner 3 will run a mile in less than 4.2 minutes. This is not very likely.

- (b) The distribution of possible team times is approximately normal. What are the mean and standard deviation of this distribution?

E P I

$$\mu_{\text{team time}} = 4.9 + 4.7 + 4.5 + 4.8 = 18.9$$

$$\sigma_{\text{team time}} = \sqrt{.15^2 + .16^2 + .14^2 + .15^2} = \sqrt{0.0902} = 0.3003$$

- (c) Suppose the team's best time to date is 18.4 minutes. What is the probability that the team will beat its own best time in the next race?

E P I



$$P(X < 18.4) \\ = P\left(Z < \frac{18.4 - 18.9}{.3003}\right) \approx 0.048$$

beating a time = getting a lower time

Total: ___/4

**AP[®] STATISTICS
2002 SCORING GUIDELINES**

Question 3 (cont'd.)

Scoring

Each part is scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is *essentially correct* if:

the probability is calculated correctly, it is correctly assessed as unlikely, and the justification is acceptable

OR

the student does not compute the probability, but appeals to the fact that a time of 4.2 minutes is more than 2 standard deviations below the mean of a normal distribution and then uses this information to reach a conclusion with appropriate communication.

Part (a) is *partially correct* if:

the probability computed is not correct (for example, $P(z > -2.14)$ or $P(z < +2.14)$ might be computed), but the given probability is correctly assessed

OR

an argument is based on the number of standard deviations from the mean without invoking normality.

Part (b) is *essentially correct* if both the mean and the standard deviation of the team time distribution are correctly computed (except for purely arithmetic mistakes).

Part (b) is *partially correct* if only one of these is correctly computed (except for purely arithmetic mistakes).

CAUTION: A standard deviation of .3 (numerically correct) can arise from this incorrect

calculation: $\frac{(.15+.16+.14+.15)}{\sqrt{4}} = 0.3$

← don't do this it's wrong!

Part (c) is *essentially correct* if the probability is correctly calculated using a mean which is either correct or carried from (b) as well as a standard deviation which is either correct or carried from (b).

Part (c) is *partially correct* if:

both the mean and standard deviation are correct or carried from (b), but the computed probability is incorrect

OR

the mean or standard deviation is incorrectly derived from (b) but the subsequent probability calculation is correct.

**AP[®] STATISTICS
2002 SCORING GUIDELINES**

Question 3 (cont'd.)

4 Complete Response

All three parts essentially correct.

3 Substantial Response

Two parts essentially correct and one part partially correct.

2 Developing Response

Two parts essentially correct and no parts partially correct.

OR

One part essentially correct and two parts partially correct.

OR

One part essentially correct and one part partially correct.

OR

Three parts partially correct.

1 Minimal Response

One part essentially correct and zero parts partially correct.

OR

No parts essentially correct and two parts partially correct.



Chapter 8 "FRAPPY" {Free Response AP Problem...Yay!}

The following problem is taken from an actual Advanced Placement Statistics Examination. Your task is to generate a complete, concise statistical response in 15 minutes. You will be graded based on the AP rubric and will earn a score of 0-4. After grading, keep this problem in your binder for your AP Exam preparation.

Every Monday a local radio station gives coupons away to 50 people who correctly answer a question about a news fact from the previous day's newspaper. The coupons given away are numbered from 1 to 50, with the first person receiving coupon 1, the second person receiving coupon 2, and so on, until all 50 coupons are given away. On the following Saturday, the radio station randomly draws numbers from 1 to 50 and awards cash prizes to the holders of the coupons with these numbers. Numbers continue to be drawn without replacement until the total amount awarded first equals or exceeds \$300. If selected, coupons 1 through 5 have a cash value of \$200, coupons 6 through 20 have a cash value of \$100, and coupons 21 through 50 have a cash value of \$50.

Scoring:

E P I

- a) Explain how you would conduct a simulation using the random number table provided below to estimate the distribution of the number of prize winners each week.

I will model picking a ticket by picking 2-digit #s 00-99. 01-50 represent their corresponding ticket #'s. (00, 51-99 will be ignored)

For each trial: Numbers will be chosen (ignoring repeats and 00, 51-99) notating the total amount won - as determined by game rules. We will continue picking #'s until the amount awarded equals or exceeds 300.

* The response variable is the total # of 2-digit #'s chosen. This will be repeated many times and graphed on a dotplot.

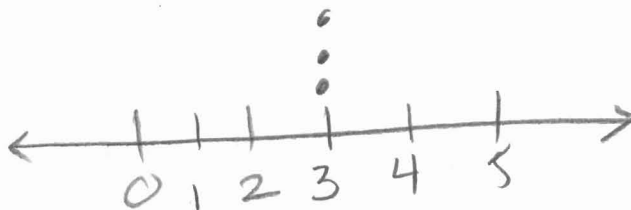
E P I

- b) Perform your simulation 3 times. (That is, run 3 trials of your simulation.) Start at the leftmost digit in the first row of the table and move across. Make your procedure clear so that someone can follow what you did. You must do this by marking directly on or above the table. Report the number of winners in each of your 3 trials.

Total = 100 300 250 300
 trial = 50 50 200 200 50 50 100 150 350
 72749 18347 65030 26128 49067 02904 49953 74674 94617 13317
 End trial # 1 = 3 winners
 End trial # 2 = 3 winners
 End trial # 2 = 3 winners

81638 36566 42709 33717 59943 12027 46547 61303 46699 76423

38449 46438 91579 01907 72146 05764 22400 94490 49833 09258



Total: ___/4

AP[®] STATISTICS
2001 SOLUTIONS AND SCORING GUIDELINES

Question 3 - Solution

Part (a):

1. **Scheme:** Obtain a two-digit random number from the random number table. If it is between 01 and 50, use it to represent the selected ticket. Ignore numbers 00 and 51 – 99.
2. **Stopping Rule:** Determine the amount of the prize associated with the chosen ticket, and add this amount to the total amount awarded so far. If the total amount awarded so far is less than \$300, repeat this process.
3. **Count:** Note the total number of winners.
4. **Non-Replacement:** Ignore any ticket number that has already been awarded a prize in this trial.

Repeats steps 1 – 4 above a large number of times.

Note: It is OK to also devise a scheme that uses 2 two-digit numbers to represent each ticket (for example, 01 and 51 both representing ticket 1; 02 and 52 both representing ticket 2; etc.) that also addresses the issue of assigning 2 two-digit numbers to each coupon correctly.

Part (b):

Solution will depend on answer to part (a).

For example, using scheme above:

Trial 1	Trial 2	Total so far	Trial 3	Total so far
72 ignore 0	02 200	200	06 100	100
74 ignore 0	61 ignore 200	200	70 ignore 100	100
91 ignore 0	28 50	250	29 50	150
33 50 50	48 50	300	04 200	350
47 50 100	Total number of winners: 3		Total number of winners: 3	
65 ignore 100				
03 200 300				
Total number of winners: 3				

Students should perform 3 trials. You will have to look at each student response carefully. Some will continue on 1st row, some will use 2nd row for second trial, etc.

Scoring

There are five components to the solution of this problem:

1. **Scheme:** must include a clear, correct statement of the assignment of two-digit random numbers to the coupon numbers (and/or values) AND clear directions as to how the table is to be used in the simulation.
2. **Stopping Rule:** must state that the trial ends when a total value of \$300 is attained or exceeded.
3. **Count:** must state or demonstrate that the number of winners is the outcome of the simulation.

AP[®] STATISTICS
2001 SOLUTIONS AND SCORING GUIDELINES

Question 3 (cont'd.)

4. **Non-Replacement:** must state that coupon numbers chosen cannot be used more than once in the same trial.
5. **Execution of #1 and #2:** must demonstrate a correct execution of a scheme with a stopping rule.
 - **Credit for components #1 and #2, is given for statements in part (a).**
 - **Credit for component #3 and/or #4 may be given for statements or demonstrations in parts (a) or (b).**
 - **Credit for component #5 is given for a clear demonstration in part (b).**

Scoring Guide:

- 4 **Complete Response**
Essentially correct on all five components.
- 3 **Substantial Response**
Essentially correct on four of the five of components.
- 2 **Developing Response**
Essentially correct on three of five components.
- 1 **Minimal Response**
Essentially correct on component #5 only
or
Essentially correct on any two of the other components.

One **INCORRECT** solution is to use the random digit 1 to represent a \$200 prize, random digits 2, 3, and 4 to represent a \$100 prize, and random digits 6, 7, 8, 9, and 0 to represent a \$50 prize.

Then a trial might look like	number	amount	total so far
	7	50	50
	2	100	150
	7	50	200
	4	100	300

A student using this scheme can merit **at most a score of 2** for the entire problem.