2011 AP Exam Solutions

2011 #1

1. A professional sports team evaluates potential players for a certain position based on two main characteristics, speed and strength.

(a) Speed is measured by the time required to run a distance of 40 yards, with smaller times indicating more desirable (faster) speeds. From previous speed data for all players in this position, the times to run 40 yards have a mean of 4.60 seconds and a standard deviation of 0.15 seconds, with a minimum time of 4.40 seconds, as shown in the table below.

2011 #1

- Based on the relationship between the mean, standard deviation, and minimum time, is it reasonable to believe that the distribution of 40-yard running times is approximately normal? Explain.

	Mean	Standard Deviation	Minimum
Time to run 40 yards	4.60 seconds	0.15 seconds	4.40 seconds

2011 #1

	Mean	Standard Deviation	Minimum
Time to run 40 yards	4.60 seconds	0.15 seconds	4.40 seconds

- Z-score for the minimum is (4.4-4.6)/0.15=-1.3
- We would expect about 10% of the data to be at this point or less if the data were normal. An assumption of normality does not seem appropriate here.

2011 #1

(b) Strength is measured by the amount of weight lifted, with more weight indicating more desirable (greater) strength. From previous strength data for all players in this position, the amount of weight lifted has a mean of 310 pounds and a standard deviation of 25 pounds, as shown in the table below.

	Mean	Standard Deviation
Amount of weight lifted	310 pounds	25 pounds

- Calculate and interpret the *z*-score for a player in this position who can lift a weight of 370 pounds.

2011 #1

- Calculate and interpret the *z*-score for a player in this position who can lift a weight of 370 pounds.

	Mean	Standard Deviation
Amount of weight lifted	310 pounds	25 pounds

- Z=(370-310)/25 = 2.4
- A player with a 370 pound bench press is 2.4 standard deviations better than the mean score.

2011 #1

(c) The characteristics of speed and strength are considered to be of equal importance to the team in selecting a player for the position. Based on the information about the means and standard deviations of the speed and strength data for all players and the measurements listed in the table below for Players A and B, which player should the team select if the team can only select one of the two players? Justify your answer.

	Player A	Player B
Time to run 40 yards	4.42 seconds	4.57 seconds
Amount of weight lifted	370 pounds	375 pounds

2011 #1

	Player A	Player B
Time to run 40 yards	4.42 seconds	4.57 seconds
Amount of weight lifted	370 pounds	375 pounds

Player A

40 yd dash time is 1.2 sd above average.

Weight lift is 2.4 sd above average.

Player B

40 yd dash time is 0.2 sd above average.

Weight lift is 2.6 sd above average.

Player A is the best choice, with a small deficit in the weight lift more than made up for by a far superior 40 yd dash time.

2011 #2

2. The table below shows the political party registration by gender of all 500 registered voters in Franklin Township.

PARTY REGISTRATION – FRANKLIN TOWNSHIP

	Party W	Party X	Party Y	Total
Male	60	120	120	300
Female	28	124	48	200
Total	88	244	168	500

2011 #2

	Party W	Party X	Party Y	Total
Male	60	120	120	300
Female	28	124	48	200
Total	88	244	168	500

(a) Given that a randomly selected registered voter is a male, what is the probability that he is registered for Party Y?

2011 #2

	Party W	Party X	Party Y	Total
Female	60	120	120	300
Male	28	124	48	200
Total	88	244	168	500

(a) Given that a randomly selected registered voter is a male, what is the probability that he is registered for Party Y?

48/200 or 0.24

2011 #2

	Party W	Party X	Party Y	Total
Female	60	120	120	300
Male	28	124	48	200
Total	88	244	168	500

(b) Among the registered voters of Franklin Township, are the events “is a male” and “is registered for Party Y” independent? Justify your answer based on probabilities calculated from the table above.

2011 #2

	Party W	Party X	Party Y	Total
Female	60	120	120	300
Male	28	124	48	200
Total	88	244	168	500

(b) Among the registered voters of Franklin Township, are the events “is a male” and “is registered for Party Y” independent? Justify your answer based on probabilities calculated from the table above.

Independence means P(male) and P(male | registered for party Y) are equal and vice versa.

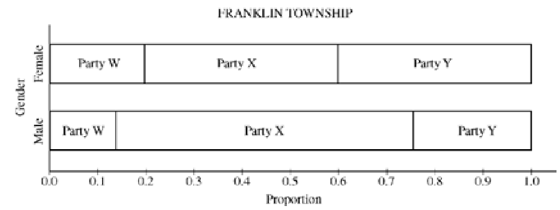
From part A, P(party Y | male) = 0.24.

P(is registered for party Y)=168/500=0.336

Therefore the events are not independent.

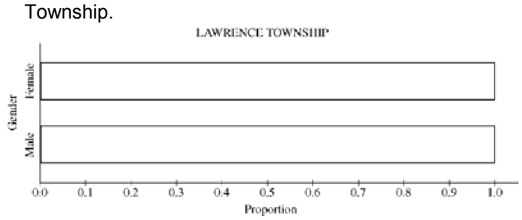
2011 #2

(c) One way to display the data in the table is to use a segmented bar graph. The following segmented bar graph, constructed from the data in the party registration–Franklin Township table, shows party-registration distributions for males and females in Franklin Township.



2011 #2

In Lawrence Township, the proportions of all registered voters for Parties W, X, and Y are the same as for Franklin Township, and party registration is independent of gender. Complete the graph below to show the distributions of party registration by gender in Lawrence Township.



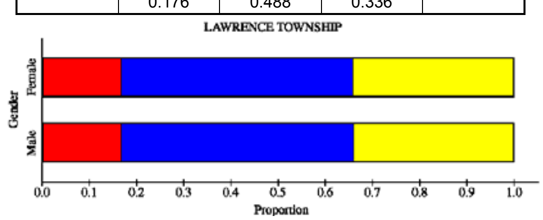
2011 #2

- Independence means that for both genders the proportions of people registered in each party are the same.

	Party W	Party X	Party Y	Total
Male	60	120	120	300
Female	28	124	48	200
Total	88	244	168	500
Proportion	88/500= 0.176	244/500= 0.488	168/500= 0.336	

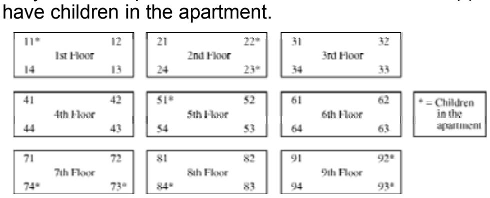
2011 #2

	Party W	Party X	Party Y	Total
Male	60	120	120	300
Female	28	124	48	200
Total	88	244	168	500
Proportion	88/500= 0.176	244/500= 0.488	168/500= 0.336	

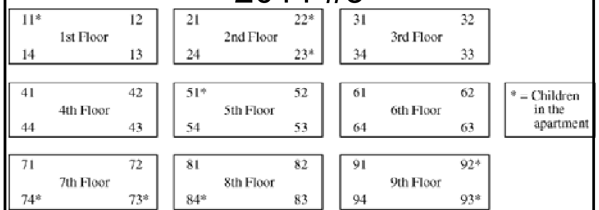


2011 #3

3. An apartment building has nine floors and each floor has four apartments. The building owner wants to install new carpeting in eight apartments to see how well it wears before she decides whether to replace the carpet in the entire building. The figure below shows the floors of apartments in the building with their apartment numbers. Only the nine apartments indicated with an asterisk (*) have children in the apartment.

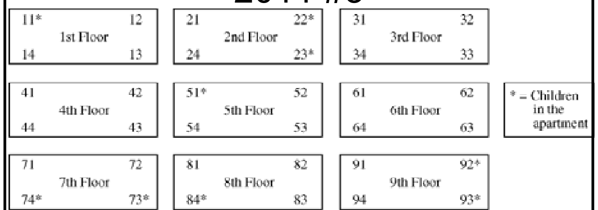


2011 #3



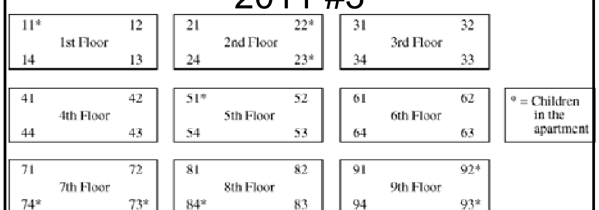
(a) For convenience, the apartment building owner wants to use a cluster sampling method, in which the floors are clusters, to select the eight apartments. Describe a process for randomly selecting eight different apartments using this method.

2011 #3



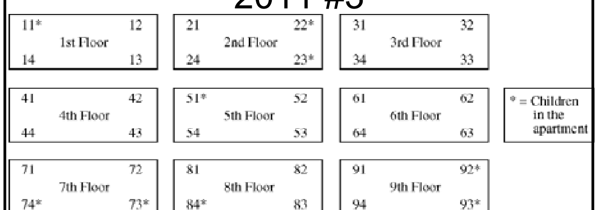
(a) *Use a random number generator to issue a random number from 1-9, ignoring 0. Use the first two numbers and select all the rooms on that corresponding floor.*

2011 #3



(b) An alternative sampling method would be to select a stratified random sample of eight apartments, where the strata are apartments with children and apartments with no children. A stratified random sample of size eight might include two randomly selected apartments with children and six randomly selected apartments with no children. In the context of this situation, give one statistical advantage of selecting such a stratified sample as opposed to a cluster sample of eight apartments using the floors as clusters.

2011 #3



(b) *Stratifying the sample makes sense because it allows a comparison between the apartments with children and those without. By using the cluster it is possible that no apartments with children are selected.*

2011 #4

4. High cholesterol levels in people can be reduced by exercise, diet, and medication. Twenty middle-aged males with cholesterol readings between 220 and 240 milligrams per deciliter (mg/dL) of blood were randomly selected from the population of such male patients at a large local hospital. Ten of the 20 males were randomly assigned to group A, advised on appropriate exercise and diet, and also received a placebo. The other 10 males were assigned to group B, received the same advice on appropriate exercise and diet, but received a drug intended to reduce cholesterol instead of a placebo. After three months, posttreatment cholesterol readings were taken for all 20 males and compared to pretreatment cholesterol readings. The tables below give the reduction in cholesterol level (pretreatment reading minus posttreatment reading) for each male in the study.

2011 #4

Group A (placebo)
Mean Reduction: 10.20
Standard Deviation of Reductions: 7.66

Reduction (in mg/dL)	2	19	8	4	12	8	17	7	24	1

Group B (cholesterol drug)
Mean Reduction: 16.40
Standard Deviation of Reductions: 9.40

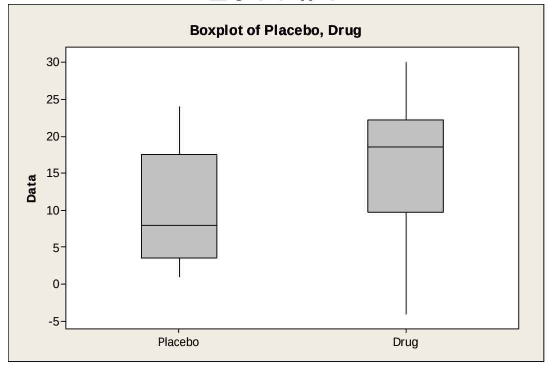
Reduction (in mg/dL)	30	19	18	17	20	-4	23	10	9	22

Do the data provide convincing evidence, at the $\alpha = 0.01$ level, that the cholesterol drug is effective in producing a reduction in mean cholesterol level beyond that produced by exercise and diet?

2011 #4

- The two groups are separate and randomly selected. A 2-sample t test is appropriate.
- Conditions:
  - The data come from an SRS.
  - With n_1=n_2=10, we'll check each data set for deviations from normality.

2011 #4



2011 #4

- The two groups are separate and randomly selected. A 2-sample t test is appropriate.
- Conditions:
  - The data come from an SRS and are selected independently.
  - With n=10, we'll check each data set for deviations from normality. The data has some skew but no signs of outliers. Proceed with caution.

2011 #4

- H_0: Mu_p=Mu_d The true mean reduction in cholesterol for the placebo group is the same as for the drug group.
- H_a: Mu_p < Mu_d The true mean reduction in cholesterol for the placebo group is less than the drug group.

2011 #4

- Two-sample T for Placebo vs Drug

	N	Mean	StDev	SE Mean
Placebo	10	10.20	7.66	2.4
Drug	10	16.40	9.40	3.0

Difference = mu (Placebo) - mu (Drug)
Estimate for difference: -6.20000
95% CI for difference: (-14.28610, 1.88610), although not appropriate for a 1 sided test.
T-Test of difference = 0 (vs not =): T-Value = -1.62, 1 sided P-Value = 0.0619
DF = 17

Since my calculated (0.0619) p is greater than my alpha of 0.01, We fail to reject the null hypothesis and cannot conclude the drug is better than the placebo.

2011 #5

Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237 R-Sq = 0.873 R-Sq (adj) = 0.868

(a) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.
(b) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at your answer.
(c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity?
(d) Is there statistically convincing evidence that electricity production by the windmill is related to wind velocity? Explain.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237 R-Sq = 0.873 R-Sq (adj) = 0.868

(a) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.
Y-hat = 0.137 + 0.240x, where *x* is the wind velocity in mph and *y*-hat is the estimate of the electricity production in amperes.

(b) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at your answer.
y-hat = 0.137 + 0.240(15) = 3.887 & y-hat = 0.137 + 0.240(25) = 6.387
There are about 2.5 more amperes of electricity generated when the wind velocity is 25 mph versus 15 mph.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237 R-Sq = 0.873 R-Sq (adj) = 0.868

(c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity?
There is about 87.3% of variation in electricity production explained by the linear relationship with wind velocity.

(d) Is there statistically convincing evidence that electricity production by the windmill is related to wind velocity? Explain.
Assumptions have already been checked, proceed with statistical test.
H_0: $\beta = 0$ & H_a: $\beta \neq 0$
$t = b/SE_b = 0.240/0.019 = 12.63$ p-value = 0.000
We can conclude that there is a linear relationship between wind velocity and electricity production with a p-value equal to zero.

2011 #6

Every year, each student in a nationally representative sample is given tests in various subjects. Recently, a random sample of 9,600 twelfth-grade students from the United States were administered a multiple-choice United States history exam. One of the multiple-choice questions is below. (The correct answer is C.)

In 1935 and 1936 the Supreme Court declared that important parts of the New Deal were unconstitutional.

President Roosevelt responded by threatening to
(A) impeach several Supreme Court justices
(B) eliminate the Supreme Court
(C) appoint additional Supreme Court justices who shared his views
(D) override the Supreme Court's decisions by gaining three-fourths majorities in both houses of Congress

Of the 9,600 students, 28 percent answered the multiple-choice question correctly.

(a) Let *p* be the proportion of all United States twelfth-grade students who would answer the question correctly. Construct and interpret a 99 percent confidence interval for *p*.

2011 #6

(a) Let *p* be the proportion of all United States twelfth-grade students who would answer the question correctly. Construct and interpret a 99 percent confidence interval for *p*.

We want to estimate the true proportion of U.S. twelfth-grade students who would answer the question correctly using a one-proportion z-interval.

Conditions: The population of all U.S. twelfth-grade students is at least 96,000.
np-hat > 10 & n(1-p-hat) > 10: 2688 > 10 & 6912 > 10

$$.28 \pm 2.576\sqrt{\frac{.28(1-.28)}{9600}}$$

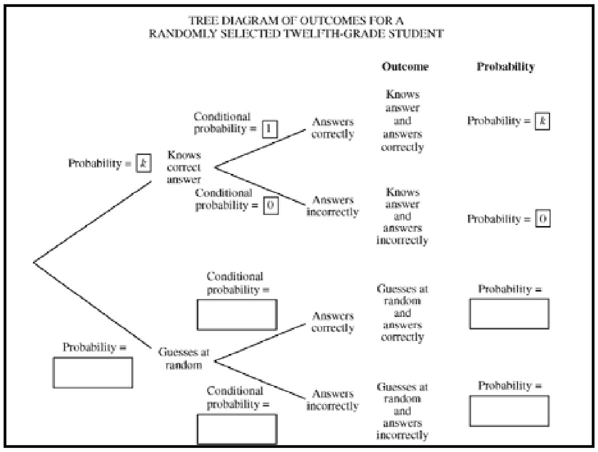
Interval: (.2682, .2918)

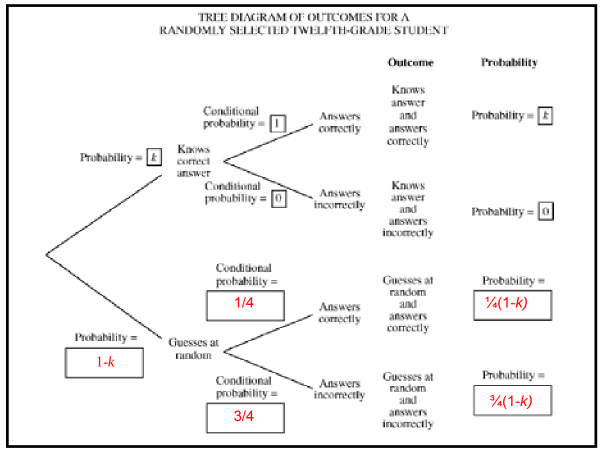
Conclusion: We are 99% confident that the true proportion of U.S. twelfth-grade students who would answer the question correctly is between 26.8% and 29.2%.

Assume that students who actually know the correct answer have a 100 percent chance of answering the question correctly, and students who do not know the correct answer to the question guess completely at random from among the four options.

Let *k* represent the proportion of all United States twelfth-grade students who actually know the correct answer to the question.

(b) A tree diagram of the possible outcomes for a randomly selected twelfth-grade student is provided below. Write the correct probability in each of the five empty boxes. Some of the probabilities may be expressions in terms of *k*.





(c) Based on the completed tree diagram, express the probability, in terms of *k*, that a randomly selected twelfth-grade student would correctly answer the history question.

(c) Based on the completed tree diagram, express the probability, in terms of *k*, that a randomly selected twelfth-grade student would correctly answer the history question.

P(answers correctly) = $k + \frac{1}{4}(1 - k)$

$= 3/4k + 1/4$

(d) Using your interval from part (a) and your answer to part (c), calculate and interpret a 99 percent confidence interval for *k*, the proportion of all United States twelfth-grade students who actually know the answer to the history question. You may assume that the conditions for inference for the confidence interval have been checked and verified.

(d) Using your interval from part (a) and your answer to part (c), calculate and interpret a 99 percent confidence interval for *k*, the proportion of all United States twelfth-grade students who actually know the answer to the history question. You may assume that the conditions for inference for the confidence interval have been checked and verified.

Answer from part (a): (.2682, .2918)
Answer from part (c): P(correct)= ¾ k + ¼

Use CI from (a) & set = to (c), solve for k.

¾ k + ¼ = .2682 → k = .0243
¾ k + ¼ = .2918 → k = .0557
CI for k: (.0243, .0557)