2017 AP® STATISTICS FREE-RESPONSE QUESTIONS

STATISTICS
SECTION II
Part A
Questions 1-5
Spend about 65 minutes on this part of the exam.
Percent of Section II score—75

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. Researchers studying a pack of gray wolves in North America collected data on the length x, in meters, from nose to tip of tail, and the weight y, in kilograms, of the wolves. A scatterplot of weight versus length revealed a relationship between the two variables described as positive, linear, and strong.

   (a) For the situation described above, explain what is meant by each of the following words.

   (i) Positive: AS THE LENGTHS OF WOLVES INCREASE, ON AVERAGE THEIR WEIGHTS INCREASE.

   (ii) Linear: FOR EACH ADD'L METER IN THE WOLVES' LENGTHS, THE AVG. WEIGHTS OF WOLVES INCREASES AT A CONSTANT RATE.

   (iii) Strong: THE WEIGHTS OF WOLVES ARE CLOSE TO WHAT YOU WOULD PREDICT USING THEIR LENGTHS + THE LEAST SQS REGR. LINE.

   The data collected from the wolves were used to create the least-squares equation $\hat{y} = -16.46 + 35.02x$.

   (b) Interpret the meaning of the slope of the least-squares regression line in context.

   (c) One wolf in the pack with a length of 1.4 meters had a residual of -9.67 kilograms. What was the weight of the wolf?

b) FOR EACH ADD'L METER IN WOLF LENGTH, MY PREDICTION FOR WOLF WEIGHT INCREASES BY 35.02 kg

c) $\widehat{\text{WEIGHT}} = -16.46 + 35.02(1.4) \approx 32.568$ kg

$\text{ACTUAL WEIGHT} = 32.568 - 9.67 = 22.898$ kg

2. The manager of a local fast-food restaurant is concerned about customers who ask for a water cup when placing an order but fill the cup with a soft drink from the beverage fountain instead of filling the cup with water. The manager selected a random sample of 80 customers who asked for a water cup when placing an order and found that 23 of those customers filled the cup with a soft drink from the beverage fountain.

(a) Construct and interpret a 95 percent confidence interval for the proportion of all customers who, having asked for a water cup when placing an order, will fill the cup with a soft drink from the beverage fountain.

(b) The manager estimates that each customer who asks for a water cup but fills it with a soft drink costs the restaurant $0.25. Suppose that in the month of June 3,000 customers ask for a water cup when placing an order. Use the confidence interval constructed in part (a) to give an interval estimate for the cost to the restaurant for the month of June from the customers who ask for a water cup but fill the cup with a soft drink.

a) 95% CI FOR A PROPORTION — 1 PROP. Z ~~INTERVAL~~

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- RANDOM SAMPLE OF 80 CUSTOMERS GIVEN
- $n\hat{p} = 80 \cdot \frac{23}{80} = 23 \geq 10$
  $n(1-\hat{p}) = 57 \geq 10$
- WHETHER ONE ~~A~~ SELECTED CUSTOMER GETS A SOFT DRINK IS INDEPENDENT FROM THE NEXT.

$$\frac{23}{80} \pm (1.96)\sqrt{\frac{23/80 \cdot 57/80}{80}}$$

$$(0.188, 0.387)$$

I'M 95% CONFIDENT THE TRUE PROPORTION OF CUSTOMERS ASKING FOR A WATER CUP THAT ACTUALLY GET SODA IS BETWEEN 0.188 AND 0.387.

b) $3000 \times 0.188 \times 0.25 = 141$, $3000 \times 0.388 \times 0.25 = 290.25$

I'M 95% CONFIDENT THAT CUSTOMERS TAKING SOFT DRINKS USING WATER CUPS DURING THE MONTH OF JUNE COST THE RESTAURANT BETWEEN $141 AND $290.25.

3. A grocery store purchases melons from two distributors, J and K. Distributor J provides melons from organic farms. The distribution of the diameters of the melons from Distributor J is approximately normal with mean 133 millimeters (mm) and standard deviation 5 mm.

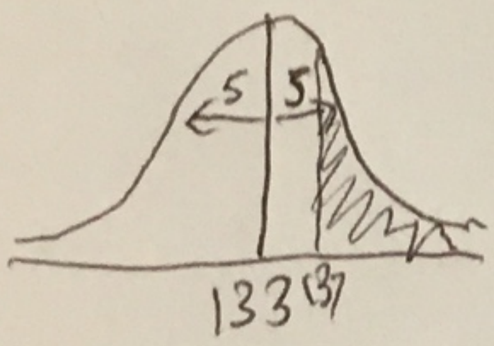
(a) For a melon selected at random from Distributor J, what is the probability that the melon will have a diameter greater than 137 mm?

Distributor K provides melons from nonorganic farms. The probability is 0.8413 that a melon selected at random from Distributor K will have a diameter greater than 137 mm. For all the melons at the grocery store, 70 percent of the melons are provided by Distributor J and 30 percent are provided by Distributor K.

(b) For a melon selected at random from the grocery store, what is the probability that the melon will have a diameter greater than 137 mm?

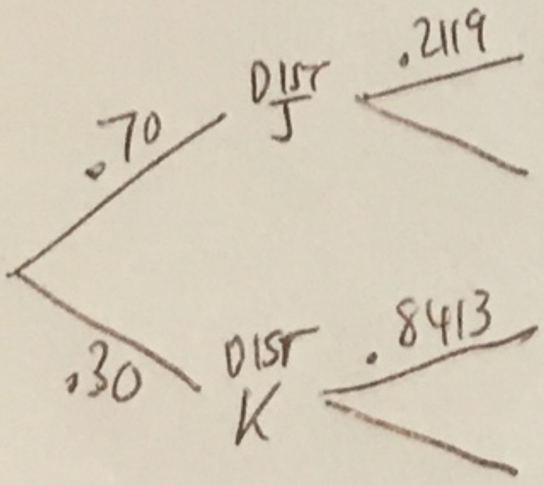
(c) Given that a melon selected at random from the grocery store has a diameter greater than 137 mm, what is the probability that the melon will be from Distributor J?

a)



$P(x > 137) = 0.2119$

b)

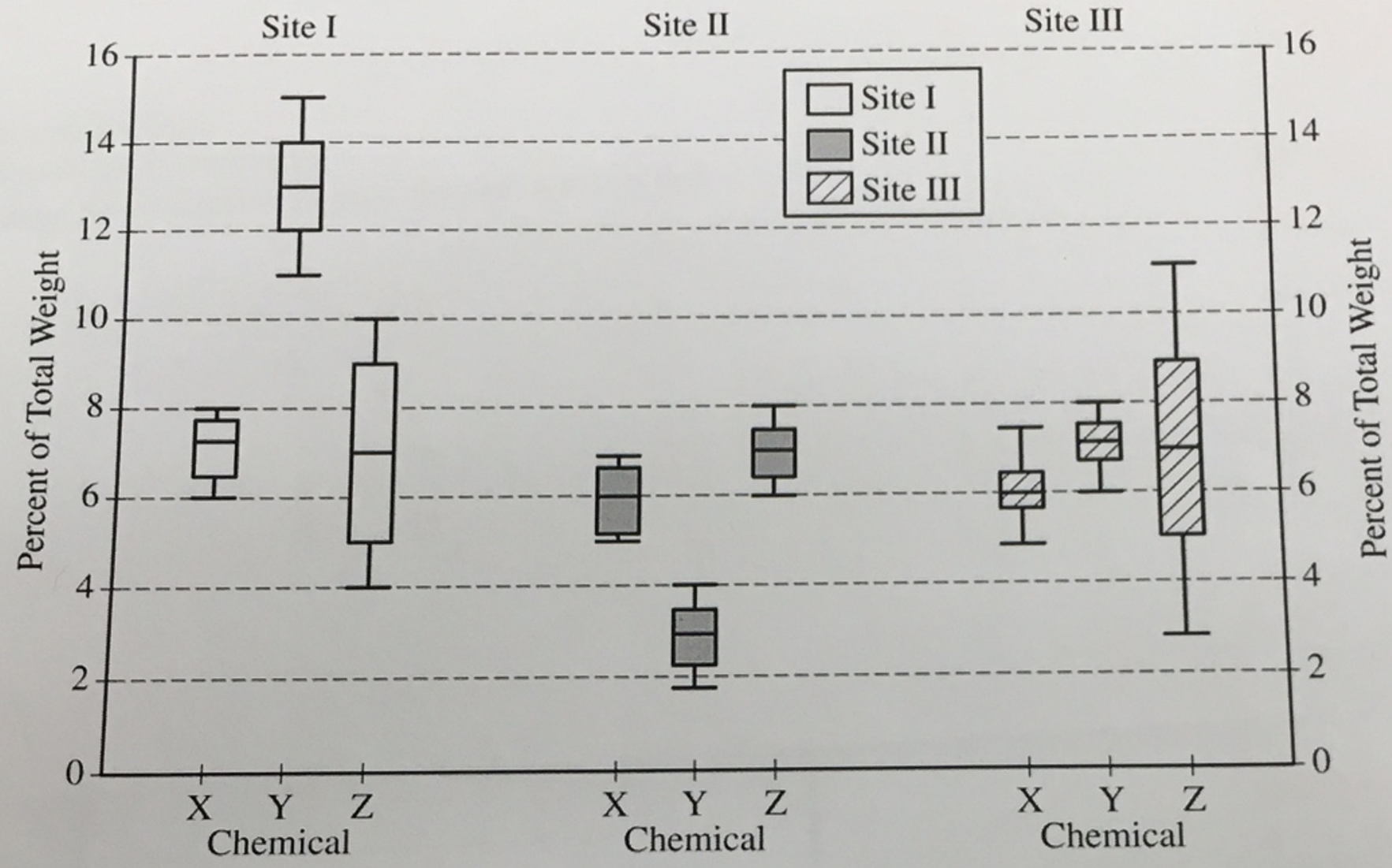


$0.1483 + 0.2523 = 0.4006$

c)

$$P(\text{DIST J} \mid > 137) = \frac{.1483}{0.1483 + 0.2523} = 0.3701$$

4. The chemicals in clay used to make pottery can differ depending on the geographical region where the clay originated. Sometimes, archaeologists use a chemical analysis of clay to help identify where a piece of pottery originated. Such an analysis measures the amount of a chemical in the clay as a percent of the total weight of the piece of pottery. The boxplots below summarize analyses done for three chemicals—X, Y, and Z—on pieces of pottery that originated at one of three sites: I, II, or III.



(a) For chemical Z, describe how the percents found in the pieces of pottery are similar and how they differ among the three sites.

(b) Consider a piece of pottery known to have originated at one of the three sites, but the actual site is not known.

(i) Suppose an analysis of the clay reveals that the sum of the percents of the three chemicals X, Y, and Z is 20.5%. Based on the boxplots, which site—I, II, or III—is the most likely site where the piece of pottery originated? Justify your choice.

(ii) Suppose only one chemical could be analyzed in the piece of pottery. Which chemical—X, Y, or Z—would be the most useful in identifying the site where the piece of pottery originated? Justify your choice.

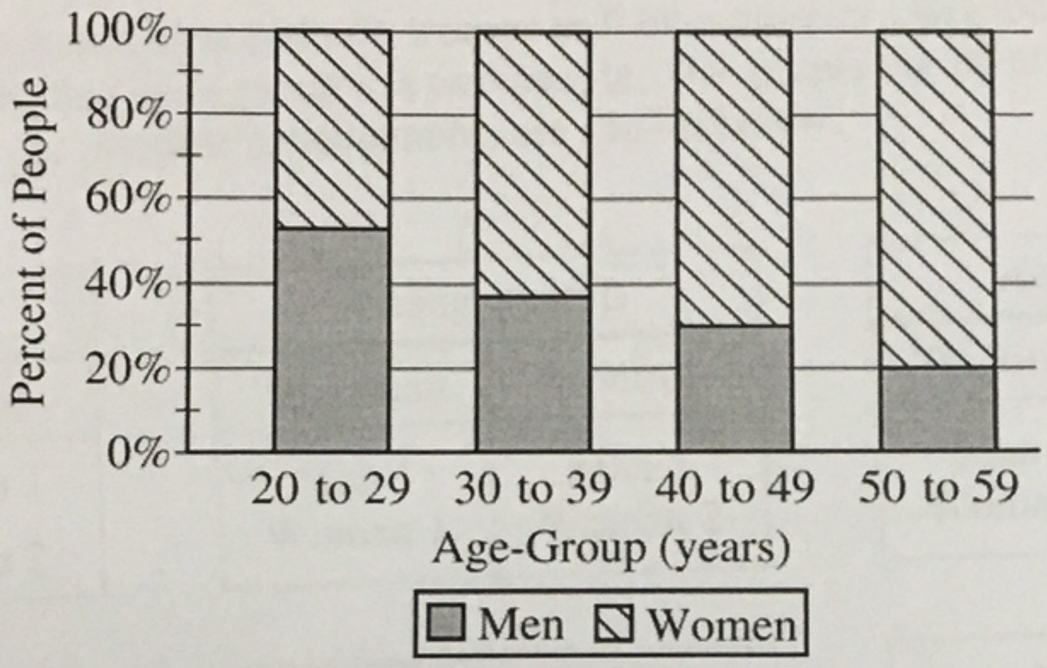
a) THE MEDIAN PERCENT OF THE WT FOR CHEMICAL Z IS ABOUT THE SAME AT THE 3 SITES, BUT THE RANGE AND IQR FOR THE %. OF CHEMICAL Z ARE MUCH LESS AT SITE II THAN AT THE OTHER SITES AT ALL THREE SITES, THE DISTRIBUTION OF %s OF CHEM Z IS ROUGHLY SYMMET

b) i) IF THE SUM OF THE %s OF THE 3 CHEMICALS WAS 20.5%, THIS PROBABLY CAME FROM SITE III, AS THE % SUMS FROM SITE I ARE LIKELY TO BE MUCH LARGER WHILE AT SITE II IT'S LIKELY TO BE MUCH SMALLER.

ii) CHEMICAL Y WOULD BE THE MOST USEFUL TO DISTINGUISH A SITE AS ITS DIST IS SIGNIFICANTLY DIFFERENT AT THE 3 SITES.

5. The table and the bar chart below summarize the age at diagnosis, in years, for a random sample of 207 men and women currently being treated for schizophrenia.

	Age-Group (years)				
	20 to 29	30 to 39	40 to 49	50 to 59	Total
Women	46	40	21	12	119
Men	53	23	9	3	88
Total	99	63	30	15	207



Exp. Count $= \frac{88 \times 15}{207} \approx 6.38$

Do the data provide convincing statistical evidence of an association between age-group and gender in the diagnosis of schizophrenia?

H_0: AGE-GROUP & GENDER ARE NOT ASSOCIATED IN THE DIAG. OF SCHIZO.

H_a: " " " ARE ASSOCIATED " " " "

- GIVEN THE SAMPLE OF 207 MEN & WOMEN IS RANDOM
- THE LOWEST EXP. COUNT IS $\frac{88 \times 15}{207} = 6.38$ WHICH IS > 5.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(46-56.913)^2}{56.913} + \cdots + \frac{(3-6.38)^2}{6.38} \approx 10.88$$

$df = 3$

P-VALUE $= 0.012$

BECAUSE MY P-VALUE 0.012 (IS SMALL / IS LESS THAN) $\alpha = 0.05$, I REJECT H_0.
THERE IS CONVINCING STATISTICAL EVIDENCE OF AN ASSOCIATION BETWEEN AGE-GROUP AND GENDER... IN THE DIAGNOSIS OF SCHIZOPHRENIA.

STATISTICS
SECTION II
Part B
Question 6
Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Consider an experiment in which two men and two women will be randomly assigned to either a treatment group or a control group in such a way that each group has two people. The people are identified as Man 1, Man 2, Woman 1, and Woman 2. The six possible arrangements are shown below.

Arrangement A	
Treatment	Control
Man 1 Man 2	Woman 1 Woman 2

Arrangement B	
Treatment	Control
Man 1 Woman 1	Man 2 Woman 2

Arrangement C	
Treatment	Control
Man 1 Woman 2	Man 2 Woman 1

Arrangement D	
Treatment	Control
Woman 1 Woman 2	Man 1 Man 2

Arrangement E	
Treatment	Control
Man 2 Woman 2	Man 1 Woman 1

Arrangement F	
Treatment	Control
Man 2 Woman 1	Man 1 Woman 2

Two possible methods of assignment are being considered: the sequential coin flip method, as described in part (a), and the chip method, as described in part (b). For each method, the order of the assignment will be Man 1, Man 2, Woman 1, Woman 2.

(a) For the sequential coin flip method, a fair coin is flipped until one group has two people. An outcome of tails assigns the person to the treatment group, and an outcome of heads assigns the person to the control group. As soon as one group has two people, the remaining people are automatically assigned to the other group.

(i) Complete the table below by calculating the probability of each arrangement occurring if the sequential coin flip method is used.

Arrangement	A	B	C	D	E	F
Probability	$1/2 \cdot 1/2 = 1/4$	$1/2 \cdot 1/2 \cdot 1/2 = 1/8$	$1/2 \cdot 1/2 \cdot 1/2 = 1/8$	$1/2 \cdot 1/2 = 1/4$	$1/2 \cdot 1/2 \cdot 1/2 = 1/8$	$1/2 \cdot 1/2 \cdot 1/2 = 1/8$

(ii) For the sequential coin flip method, what is the probability that Man 1 and Man 2 are assigned to the same group?

$P(A \text{ or } D) = 1/4 + 1/4 = 1/2$

The six arrangements are repeated below.

Arrangement A	
Treatment	Control
Man 1	Woman 1
Man 2	Woman 2

Arrangement B	
Treatment	Control
Man 1	Man 2
Woman 1	Woman 2

Arrangement C	
Treatment	Control
Man 1	Man 2
Woman 2	Woman 1

Arrangement D	
Treatment	Control
Woman 1	Man 1
Woman 2	Man 2

Arrangement E	
Treatment	Control
Man 2	Man 1
Woman 2	Woman 1

Arrangement F	
Treatment	Control
Man 2	Man 1
Woman 1	Woman 2

(b) For the chip method, two chips are marked "treatment" and two chips are marked "control." Each person selects one chip at random without replacement.

(i) Complete the table below by calculating the probability of each arrangement occurring if the chip method is used.

Arrangement	A	B	C	D	E	F
Probability	$2/4 \cdot 1/3 = 1/6$	$2/4 \cdot 2/3 \cdot 1/2 = 1/6$	$2/4 \cdot 2/3 \cdot 1/2 = 1/6$	$2/4 \cdot 1/3 = 1/6$	$2/4 \cdot 2/3 \cdot 1/2 = 1/6$	$2/4 \cdot 2/3 \cdot 1/2 = 1/6$

(ii) For the chip method, what is the probability that Man 1 and Man 2 are assigned to the same group?

$P(A \text{ or } D) = 1/6 + 1/6 = 1/3$

(c) Sixteen participants consisting of 10 students and 6 teachers at an elementary school will be used for an experiment to determine lunch preference for the school population of students and teachers. As the participants enter the school cafeteria for lunch, they will be randomly assigned to receive one of two lunches so that 8 will receive a salad, and 8 will receive a grilled cheese sandwich. The students will enter the cafeteria first, and the teachers will enter next. Which method, the sequential coin flip method or the chip method, should be used to assign the treatments? Justify your choice.

PARTS a + b SHOW THAT THE CHIP METHOD IS LESS LIKELY TO ASSIGN ALL TEACHERS TO THE SAME GROUP, OR TO CREATE UNBALANCED TREATMENT GROUPS, SINCE IT'S LIKELY THAT STUDENT/TEACHER STATUS IS ASSOCIATED WITH LUNCH PREFERENCE, THE CHIP METHOD IS THE PREFERRED METHOD, AS IT'S LESS LIKELY TO ALLOW THIS ASSOCIATION TO AFFECT THE RESULTS OF THE EXPERIMENT.

STOP

END OF EXAM