Mr. V Solutions

2017 AP® STATISTICS FREE-RESPONSE QUESTIONS

STATISTICS
SECTION II
Part A
Questions 1-5

3 parts
· i, ii, iii, context, pred.
· (b) int, context
· (c)

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 length of a wolf increases, the weight of the wolf also tends to increase

   (ii) Linear: the relationship increases at a generally constant rate

   (iii) Strong: the actual points on the scatterplot come fairly close to the prediction 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 every meter the wolf grows it's predicted weight increases by 35.02 kg

(c) $\hat{y} = -16.46 + 35.02(1.4) = 32.568$

the prediction is 9.67 kg above the actual

$-9.67 = \text{actual} - 32.568$

The wolf actually weighed 22.898 kg

3 parts
· cond.
· int
· int

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) We must assume 80 is less than 10% of all the customers

This was a random sample.

We must assume each person selected was independent of eachother in choosing whether to steal the soda or not. This may not be safe since they may be observing other people doing it and feel more emboldened to steal.

$np = 23 > 10$ $nq = 57 > 10$ so this sample is large enough.

Conditions are ok to do a confidence interval for proportions

$p = \frac{23}{80} = .2875$ $\quad .2875 \pm 1.96\sqrt{\frac{.2875(.7125)}{80}}$

$.2875 \pm .0992$ or $(.1883, .3867)$

b)

$.1883\,(3000)(.25) = \$141.23$

$.3867\,(3000)(.25) = \$276.53$

$(\$141.23, \$276.53)$

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?

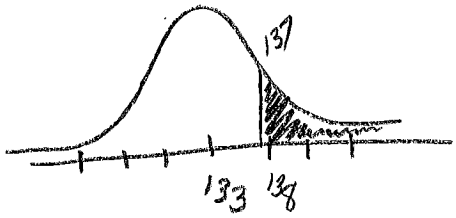
   Distributor K provides melons from nonorganic farms. The probability is 0.8413 that a melon selected at random from Distributor K will have a diameter greater than 137 mm. For all the melons at the grocery store, 70 percent of the melons are provided by Distributor J and 30 percent are provided by Distributor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

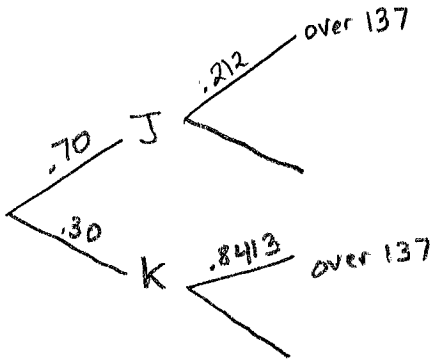
3 parts

(a) J $\mu = 133_{mm}$ $\sigma = 5_{mm}$



$P(x > 137) = .212$ for Distributor J

(b) $P(x > 137) = .8413$ for Distributor K



$P(x > 137) = P(J \text{ and over } 137) + P(K \text{ and over } 137)$

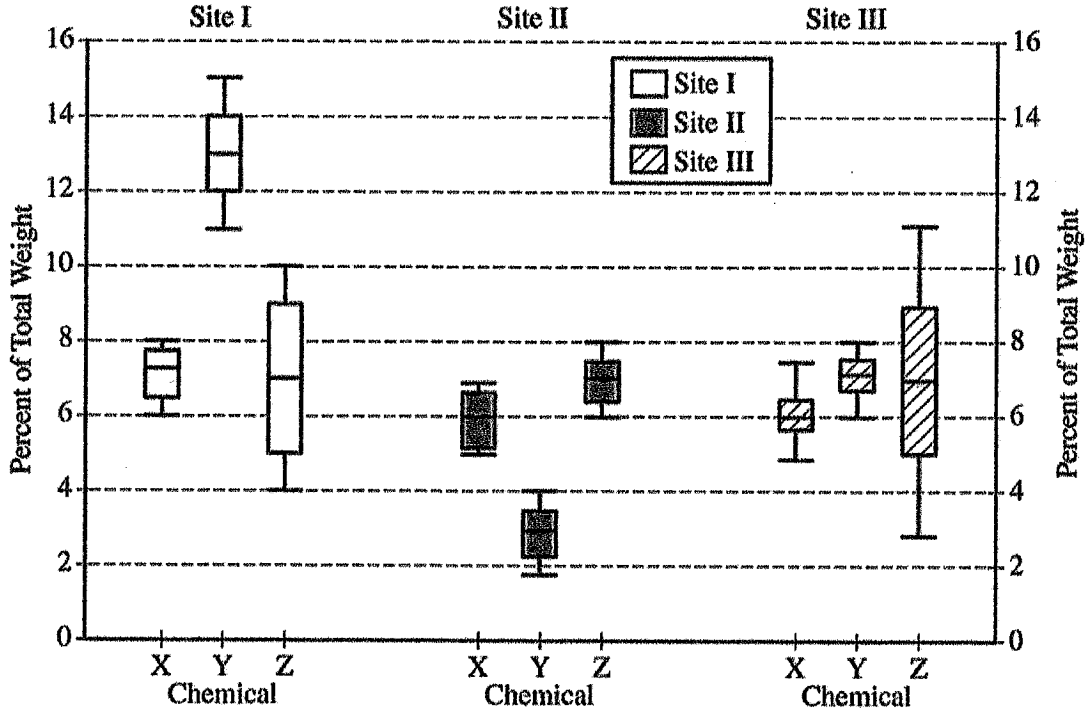
$= .70(.212) + .30(.8413)$

$= .401$

(c) $P(J \mid \text{over } 137) = \dfrac{P(J \text{ and over } 137)}{P(\text{over } 137)} = \dfrac{.70(.212)}{.401} = .370$

3 parts

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) All 3 sites have about the same mean for chemical Z of about 7%
The spread is greatest for chemical Z at site III with an IQR of about 5%, site I had the next greatest spread with an IQR of also 5% but a slightly smaller range (6% for site I, 8% for site III), and site II had the smallest spread with an IQR of about 1%.
All 3 sites had a fairly symmetric shape for chemical Z

(b) i) if the sum of chemical X, Y, and Z was 20.5% then it most likely came from site III since the medians for the 3 chemicals at that site are about 6, 7+, and 7 or a sum of a little over 20%. Site I has a sum of it's median's much farther away at about 27% while site II had the sum of its medians at about 16%

ii) Chemical Y would be the best chemical to test since the median at all three sites is widely different whereas chemical Z comes out about the same at all 3 sites and chemical X is ~~pretty~~ close the same average at site II and III
Also, chemical Y does not have any overlap in it's spread (range) so it would be pretty unlikely to predict the wrong site based on its outcome.

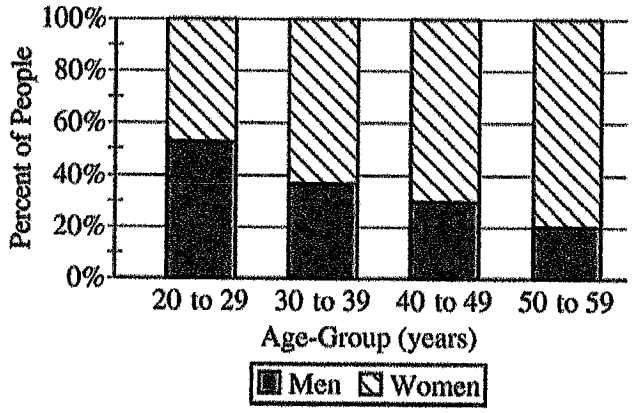
9-15% would be site I
6-8% would be site III
1.8-4% would be site II

4 parts

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



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

H_0: there is no association between age-group and gender in diagnosis of schizophrenia

H_A: there is an association between age-group and gender in diagnosis of schizophrenia

Expected outcomes if no association

	20-29	30-39	40-49	50-59
Women	56.9	36.2	17.2	8.6
Men	42.1	26.8	12.8	6.4

All expected counts are ≥ 5

This is a random sample

Each person sampled was independent in their incidence of schizophrenia

Conditions are ok to do a χ^2 Test of independence

$\chi^2 = 10.88$ $p = .012$ $df = 3$

There is pretty strong evidence to reject H_0 with a pvalue this low, .012 ($<.05$). There seems to be an association between age of diagnosis for schizophrenia and gender. At earlier ages men are diagnosed as often or more than women, as people get older women are diagnosed for schizophrenia more and more often than men.

STATISTICS
SECTION II
Part B
Question 6

3 parts

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	T T $\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$	T H T $\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$	T H H $\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$	H H $\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$	H T H $\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$	H T T $\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$

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

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

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	T T $\frac{2}{4}\cdot\frac{1}{3}=\frac{1}{6}$	T C T $\frac{2}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}=\frac{1}{6}$	T C C $\frac{2}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}=\frac{1}{6}$	C C $\frac{2}{4}\cdot\frac{1}{3}=\frac{1}{6}$	C T C $\frac{2}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}=\frac{1}{6}$	C T T $\frac{2}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}=\frac{1}{6}$

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

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

(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.

The chip method is equally likely to give any of possible arrangements so it should be used. The coin method doesn't necessarily give equal likelihood to all possible arrangements so it should not be used as it is unfair.

STOP

END OF EXAM