AP Statistics

Review #3
Anticipating Patterns

Name ________________

AP Statistics **Anticipating Patterns Review**

III. Anticipating Patterns: Exploring random phenomena using probability and simulation (20%-30%)
Probability is the tool used for anticipating what the distribution of data should look like under a given model.

A. Probability

1. Interpreting probability, including long-run relative frequency interpretation

2. 'Law of Large Numbers' concept

3. Addition rule, multiplication rule, conditional probability, and independence

4. Discrete random variables and their probability distributions, including binomial and geometric

5. Simulation of random behavior and probability distributions

6. Mean (expected value) and standard deviation of a random variable, and linear transformation of a random variable

B. Combining independent random variables

1. Notion of independence versus dependence

2. Mean and standard deviation for sums and differences of independent random variables

C. The normal distribution

1. Properties of the normal distribution

2. Using tables of the normal distribution

3. The normal distribution as a model for measurements

D. Sampling distributions

1. Sampling distribution of a sample proportion

2. Sampling distribution of a sample mean

3. Central Limit Theorem

4. Sampling distribution of a difference between two independent sample proportions

5. Sampling distribution of a difference between two independent sample means

6. Simulation of sampling distributions

7. t-distribution

8. Chi-square distribution

Tips and Hints:

➢ It is likely you will be asked a simulation problem on the AP test. For simulations, you need to be able to describe how you will perform a simulation in addition to actually doing it. Make sure that you:
  - Create a correspondence between random numbers and outcomes.
  - Explain how you will obtain the random numbers (e.g., move across the rows of the random digits table, examining pairs of digits), and how you will know when to stop.
  - Make sure you understand the purpose of the simulation, and when a trial ends -- counting the number of trials until you achieve "success" or counting the number of "successes" or some other criterion.
  - Are you drawing numbers with or without replacement? Be sure to mention this in your description of the simulation and to perform the simulation accordingly.

➢ Speaking of simulations, if you're not sure how to approach a probability problem on the AP Exam, see if you can design a simulation to get an approximate answer.

➢ One of the most common errors students make is independence vs. mutually exclusive events. It is important to note: Independent events are not the same as mutually exclusive (disjoint) events!!
  - Two events, A and B, are *independent* if the occurrence or non-occurrence of one of the events has no effect on the probability that the other event occurs.
  - Events A and B are *mutually exclusive* if they cannot happen simultaneously.
  - For example, if I know you are a senior in HS, then I know you cannot be a sophomore (mutually exclusive events). Therefore, knowing that you are a senior AFFECTS the probability of you being a sophomore, so they are dependent events.

➢ Know the basic rules of probability, such as the probability of each event is between 0 and 1 inclusive and the total probability of the sample space is 1.

➢ Be familiar with the strategies of probability – whenever you are stuck, try referring to the formulas on the formula sheet, try making a Venn diagram, a contingency table, or a tree diagram.

➢ The union (U) symbol stands for OR – the intersection symbol ($\cap$) stands for AND

➢

- ➢ Recognize a discrete random variable setting when it arises. Be prepared to calculate its mean (expected value) and standard deviation.
  - $E(x) = \mu_x = \Sigma(x \bullet p(x))$
  - $Var(x) = \sigma_x^2 = \Sigma((x-\mu)^2 \bullet p(x))$
  - $SD(x) = \sigma_x = \sqrt{\sigma_x^2}$
  - To use your calculator to find the mean and standard deviation, put the x values into L1, the probabilities into L2, then do 1-var stats L1, L2
- ➢ Transforming variables: When adding or subtracting values, only the mean changes. When multiplying or dividing values, both mean and standard deviation change.
  - $\mu_{aX+b} = a(\mu) + b$
  - $\sigma_{aX+b} = a(\sigma)$
- ➢ Combining variables:
  - $\mu_{X \pm Y} = \mu_X \pm \mu_Y$
  - $\sigma_{X \pm Y}^2 = \sigma_X^2 + \sigma_Y^2$ **Note: variances ALWAYS add!!
- ➢ The four requirements for a random phenomenon to be a binomial situation are:
  1. There are a fixed number of trials.
  2. On each trial, there are two possible outcomes: "success" and "failure."
  3. The probability of a "success" on each trial is constant.
  4. The trials are independent.
- ➢ The calculational binomial probability formula ($p(X=x) = {}_nC_r p^r q^{n-r}$) is given on your formula sheet – **Know how to use it!!!**
- ➢ To save yourself some time, you should use the binompdf(and binomcdf(functions from your 2nd:Distr menu. However, in order to receive full credit on the exam, you cannot just use the "TI-speak". If you don't remember how to **find the probability by hand**, the *minimum* you should put is: This is a binomial situation with number of trials = ___, the probability of success = ____, and the trial of interest = ___." You can put the "TI-talk" on your paper and use the calculator for the probability.
- ➢ Binomial mean and standard deviation formulas are also given on your formula sheet!
- ➢ Realize that a binomial distribution can be approximated well by a normal distribution if the number of trials is sufficiently large. If n is the number of trials in a binomial setting, and if p represents the probability of "success" on each trial, then a good rule of thumb states that a normal distribution can be used to approximate the binomial distribution if np is at least 10 and n(1-p) is at least 10.
- ➢ The primary difference between a binomial random variable and a geometric random variable is what you are counting. A binomial random variable counts the **number of successes** in n trials. A geometric random variable counts the **number of trials** up to and including the first success.
- ➢ To find a geometric probability, simply multiply the number of failure probabilities time the success. For example, if the probability of success is 0.4 and you are interested in the first success on the 3rd trial, then you had two failures and one success – (0.6)(0.6)(0.4)
- ➢ Again, you can use the functions geometpdf(and geometcdf(from your calculator, but again should write a statement similar to the binomial statement above.
- ➢ Technically the Normal distribution fits under the "probability umbrella" as well. Make sure you know how to find a z-score for a Normal model and use your tables to find a probability
- ➢ The idea of sampling distributions also falls under this topic and bridges the material into inference. Remember that the Central Limit Theorem assures us of *approximate* Normality when our sample size is sufficiently large (for proportions, that means 10 successes/failures… for means, that means n>30).
- ➢ The mean and standard deviation for sampling distributions is also on the formula sheet. This idea has shown up in the past, so make sure you know how to find these values.

Final note: It is *vital* that you show a probability statement P(?) = how = what and your work for all probability problems. An answer without supporting work is partial credit at best.

2010B #5

5. An advertising agency in a large city is conducting a survey of adults to investigate whether there is an association between highest level of educational achievement and primary source for news. The company takes a random sample of 2,500 adults in the city. The results are shown in the table below.

Primary Source for News	HIGHEST LEVEL OF EDUCATIONAL ACHIEVEMENT: Not High School Graduate	High School Graduate But Not College Graduate	College Graduate	Total
Newspapers	49	205	188	442
Local television	90	170	75	335
Cable television	113	496	147	756
Internet	41	401	245	687
None	77	165	38	280
Total	370	1,437	693	2,500

(a) If an adult is to be selected at random from this sample, what is the probability that the selected adult is a college graduate or obtains news primarily from the internet?

(b) If an adult who is a college graduate is to be selected at random from this sample, what is the probability that the selected adult obtains news primarily from the internet?

(c) When selecting an adult at random from the sample of 2,500 adults, are the events "is a college graduate" and "obtains news primarily from the internet" independent? Justify your answer.

(d) The company wants to conduct a statistical test to investigate whether there is an association between educational achievement and primary source for news for adults in the city. What is the name of the statistical test that should be used?

What are the appropriate degrees of freedom for this test?

2009 B #2

2. The ELISA tests whether a patient has contracted HIV. The ELISA is said to be positive if it indicates that HIV is present in a blood sample, and the ELISA is said to be negative if it does not indicate that HIV is present in a blood sample. Instead of directly measuring the presence of HIV, the ELISA measures levels of antibodies in the blood that should be elevated if HIV is present. Because of variability in antibody levels among human patients, the ELISA does not always indicate the correct result.

As part of a training program, staff at a testing lab applied the ELISA to 500 blood samples known to contain HIV. The ELISA was positive for 489 of those blood samples and negative for the other 11 samples. As part of the same training program, the staff also applied the ELISA to 500 other blood samples known to not contain HIV. The ELISA was positive for 37 of those blood samples and negative for the other 463 samples.

(a) When a new blood sample arrives at the lab, it will be tested to determine whether HIV is present. Using the data from the training program, estimate the probability that the ELISA would be positive when it is applied to a blood sample that does not contain HIV.

(b) Among the blood samples examined in the training program that provided positive ELISA results for HIV, what proportion actually contained HIV?

(c) When a blood sample yields a positive ELISA result, two more ELISAs are performed on the same blood sample. If at least one of the two additional ELISAs is positive, the blood sample is subjected to a more expensive and more accurate test to make a definitive determination of whether HIV is present in the sample. Repeated ELISAs on the same sample are generally assumed to be independent. Under the assumption of independence, what is the probability that a new blood sample that comes into the lab will be subjected to the more expensive test if that sample does not contain HIV?

2008 #3

3. A local arcade is hosting a tournament in which contestants play an arcade game with possible scores ranging from 0 to 20. The arcade has set up multiple game tables so that all contestants can play the game at the same time; thus contestant scores are independent. Each contestant's score will be recorded as he or she finishes, and the contestant with the highest score is the winner.

After practicing the game many times, Josephine, one of the contestants, has established the probability distribution of her scores, shown in the table below.

Josephine's Distribution				
Score	16	17	18	19
Probability	0.10	0.30	0.40	0.20

Crystal, another contestant, has also practiced many times. The probability distribution for her scores is shown in the table below.

Crystal's Distribution			
Score	17	18	19
Probability	0.45	0.40	0.15

(a) Calculate the expected score for each player.

(b) Suppose that Josephine scores 16 and Crystal scores 17. The difference (Josephine minus Crystal) of their scores is -1. List all combinations of possible scores for Josephine and Crystal that will produce a difference (Josephine minus Crystal) of -1, and calculate the probability for each combination.

(c) Find the probability that the difference (Josephine minus Crystal) in their scores is -1.

(d) The table below lists all the possible differences in the scores between Josephine and Crystal and some associated probabilities.

Distribution (Josephine minus Crystal)						
Difference	-3	-2	-1	0	1	2
Probability	0.015			0.325	0.260	0.090

Complete the table and calculate the probability that Crystal's score will be higher than Josephine's score.

2007 #3

3. Big Town Fisheries recently stocked a new lake in a city park with 2,000 fish of various sizes. The distribution of the lengths of these fish is approximately normal.

   (a) Big Town Fisheries claims that the mean length of the fish is 8 inches. If the claim is true, which of the following would be more likely?

   - A random sample of 15 fish having a mean length that is greater than 10 inches

   or

   - A random sample of 50 fish having a mean length that is greater than 10 inches

   Justify your answer.

   (b) Suppose the standard deviation of the sampling distribution of the sample mean for random samples of size 50 is 0.3 inch. If the mean length of the fish is 8 inches, use the normal distribution to compute the probability that a random sample of 50 fish will have a mean length less than 7.5 inches.

   (c) Suppose the distribution of fish lengths in this lake was nonnormal but had the same mean and standard deviation. Would it still be appropriate to use the normal distribution to compute the probability in part (b) ? Justify your answer.

2005B #2

2. For an upcoming concert, each customer may purchase up to 3 child tickets and 3 adult tickets. Let C be the number of child tickets purchased by a single customer. The probability distribution of the number of child tickets purchased by a single customer is given in the table below.

c	0	1	2	3
$p(c)$	0.4	0.3	0.2	0.1

   (a) Compute the mean and the standard deviation of C.

   (b) Suppose the mean and the standard deviation for the number of adult tickets purchased by a single customer are 2 and 1.2, respectively. Assume that the numbers of child tickets and adult tickets purchased are independent random variables. Compute the mean and the standard deviation of the total number of adult and child tickets purchased by a single customer.

   (c) Suppose each child ticket costs \$15 and each adult ticket costs \$25. Compute the mean and the standard deviation of the total amount spent per purchase.

2003 #3

3. Men's shirt sizes are determined by their neck sizes. Suppose that men's neck sizes are approximately normally distributed with mean 15.7 inches and standard deviation 0.7 inch. A retailer sells men's shirts in sizes S, M, L, XL, where the shirt sizes are defined in the table below.

Shirt size	Neck size
S	$14 \leq$ neck size < 15
M	$15 \leq$ neck size < 16
L	$16 \leq$ neck size < 17
XL	$17 \leq$ neck size < 18

(a) Because the retailer only stocks the sizes listed above, what proportion of customers will find that the retailer does not carry any shirts in their sizes? Show your work.

(b) Using a sketch of a normal curve, illustrate the proportion of men whose shirt size is M. Calculate this proportion.

(c) Of 12 randomly selected customers, what is the probability that exactly 4 will request size M ? Show your work.