***Curriculum Topic III – PROBABILITY***

***Chapters 6-9***

**The tool used for anticipating what the distribution of data should look like under a given model.**

Random phenomena are not haphazard: they display an order that emerges only in the long run and is described by a distribution. The mathematical description of variation is central to statistics. The probability required for statistical inference is not primarily axiomatic or combinatorial, but is oriented toward describing data distributions.

**Probability and Random Variables**

The outcome of some events is random, meaning that while the outcome of a specific event is unpredictable, in the long run the different outcomes seem to settle down to a pattern of specific values. The proportion of times that a specific outcome would occur in a very long series of repetitions is the probability of the event.

A numerical variable whose value is the outcome of a random phenomenon is called a random variable. A random variable is discrete if the number of outcomes is a collection of isolated points on the number line, and is continuous if the outcomes include an interval on the number line.

Examples of discrete random variables include the outcome of rolling dice in a game of Monopoly or the number of babies born in your local hospital on Wednesdays throughout the year. Two specific types of discrete random variables that students should be familiar with are binomial and geometric random variables.

Examples of continuous random variables are the weight of a tomato you pick up at the produce market or the time you spend waiting in line at the drive through window at your favorite fast food restaurant.

**Statistics and Sampling Distributions**

A parameter is some numerical characteristic of a population. A statistic is the corresponding quantity computed from a sample of the population. If a sample is randomly selected, then statistics computed from different samples will vary (but in predictable ways for certain conditions). This is called sampling variability. The value of the statistic is therefore a random variable, and the distribution that describes the pattern of values for the statistic is its sampling distribution.

Calculating the probabilities of various outcomes in a sampling distribution allows decisions to be made in sampling settings as to the confidence placed in a specific observation.

**Simulations**

Students should understand the role of simulations in estimating probabilities, the shape of sampling distributions, and the probabilities associated with a sampling distribution.

**Anticipating patterns: probability, simulations, and random variables**

You need to be able to describe how you will perform a simulation in addition to actually doing it.

* 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 -- 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.

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.

*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.

Example: Roll two fair six-sided dice. Let A = the sum of the numbers showing is 7,
B = the second die shows a 6, and C = the sum of the numbers showing is 3.

By making a table of the 36 possible outcomes of rolling two six-sided dice, you will find that P(A) = 1/6, P(B) = 1/6, and P(C) = 2/36.

* Events A and B are independent. Suppose you are told that the sum of the numbers showing is 7. Then the only possible outcomes are {(1,6), (2,5), (3,4), (4,3), (5,2), and (6,1)}. The probability that event B occurs (second die shows a 6) is now 1/6. This new piece of information did not change the likelihood that event B would happen. Let's reverse the situation. Suppose you were told that the second die showed a 6. There are only six possible outcomes: {(1,6), (2,6), (3,6), (4,6), (5,6), and (6,6)}. The probability that the sum is 7 remains 1/6. Knowing that event B occurred did not affect the probability that event A occurs.
* Events A and B are not disjoint. Both can occur at the same time.
* Events B and C are mutually exclusive (disjoint). If the second die shows a 6, then the sum cannot be 3. Can you show that events B and C are not independent?



**Recognize a *discrete random variable* setting when it arises. Be prepared to calculate its mean (expected value) and standard deviation.**

Example:

Let X = the number of heads obtained when five fair coins are tossed.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Value of x | 0 | 1 | 2 | 3 | 4 | 5 |
| Probability | 1/32 = 0.03125 | 5/32 = 0.15625 | 10/32 = 0.3125 | 10/32 = 0.3125 | 5/32 = 0.15625 | 1/32 = 0.03125 |

 |



Recognize a *binomial situation* when it arises.

The four requirements for a chance phenomenon to be a binomial situation are:

1. There are a fixed number of trials.
2. On each trial, there are two possible outcomes that can be labeled "success" and "failure."
3. The probability of a "success" on each trial is constant.
4. The trials are independent.

Example: Consider rolling a fair die 10 times. There are 10 trials. Rolling a 6 constitutes a "success," while rolling any other number represents a "failure." The probability of obtaining a 6 on any roll is 1/6, and the outcomes of successive trials are independent.

Using the TI-83, the probability of getting exactly three sixes is (10C3)(1/6)3(5/6)7 or binompdf(10,1/6,3) = 0.155045, or about 15.5 percent.

The probability of getting less than four sixes is binomcdf(10,1/6,3) = 0.93027, or about 93 percent. Hence, the probability of getting four or more sixes in 10 rolls of a single die is about 7 percent.

If X is the number of sixes obtained when 10 dice are rolled, then

If X is the number of 6's obtained when ten dice are rolled, then E(X) =



x = 10(1/6) = 1.6667, and



Did you notice that the coin-tossing example above is also a binomial situation?

**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."