AP Statistics – Formula Sheet Study Guide Name: ______________________________________

I. Descriptive Statistics	The branch of statistics that summarizes data using numbers, graphs and patterns, and describes the main features.
$\bar{x} = \frac{\sum x_i}{n}$	Pronounced "x bar". This is the mean of a sample of size n. Each x_i is an individual observation.
$s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$	The standard deviation of a sample of size n. Each x_i is an individual observation.
$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}}$	Pooled standard deviation for two samples, one of size n_1 and one of size n_2 . (NEVER USE!)
$\hat{y} = b_0 + b_1 x$	Equation for the least squares regression line. Pronounced "y hat". $\hat{y}$ is the predicted value for the response variable y, for a given value of the input explanatory variable x. b_0 is the y-intercept and b_1 is the slope. (Note that in our calculators we use the regression equation $\hat{y} = a + bx$, where a and b are the y-intercept and slope, respectively.)
$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$	Formula for the slope of the least squares regression line.
$b_0 = \bar{y} - b_1\bar{x}$	Formula for the y-intercept of the least squares regression line. Notice that it uses the slope, and the sample mean of each variable. This formula should remind you that the ordered pair $(\bar{x}, \bar{y})$ is always a point on the least squares regression line.
$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$	Formula for correlation. Note that this involves summing the products of the standardized scores of each ordered pair (x_i, y_i).
$b_1 = r\frac{s_y}{s_x}$	An alternate formula for the slope of the least squares regression line. Note that this formula illustrates the relationship between correlation and slope of the regression line. Since the standard deviations can never be negative, the slope and correlation will always have the same sign – both negative, or both positive.
$s_{b_1} = \frac{\sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}}}{\sqrt{\sum(x_i - \bar{x})^2}}$	This is the standard error of the slope of the regression line. Note that the numerator is the standard deviation of the residuals.

II. Probability	Probability is what allows us to build inference tools.
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	The general addition rule. This is the probability of event A or event B occurring.
$P(A\|B) = \frac{P(A \cap B)}{P(B)}$	Conditional probability. This is the probability of event A occurring, given that event B has occurred.
$E(X) = \mu_x = \sum x_i p_i$	This is the formula for the expected value of a discrete random variable X. The expected value is another way of saying the "mean".
$Var(X) = \sigma_x^2 = \sum (x_i - \mu_x)^2 p_i$	This is the formula for the variance of a discrete random variable X. To get σ, the standard deviation of the random variable, first find the variance and then take the positive square root. (N0te that neither standard deviation nor variance can ever be negative.)
If X has a binomial distribution with parameters n and p, then:	
$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$	This is the formula for a the probability of a discrete random variable X with a binomial distribution. Remember there is a four-point check to see if a variable has the binomial setting. This corresponds to the calculator function "binompdf". If you want a cumulative probability (with $\geq$ or $\leq$), then use "binomcdf". In this formula, k is the number of "successes" and n is the number of trials. p is the probability of success on any given trial.
$\mu_x = np$	This is the formula for the mean of a binomial random variable. n is the number of trials. p is the probability of success on any given trial.
$\sigma_x = \sqrt{np(1-p)}$	This is the formula for the standard deviation of a binomial random variable. n is the number of trials. p is the probability of success on any given trial.
$\mu_{\hat{p}} = p$	This is the formula for the mean of the sampling distribution of sample proportions $\hat{p}$ taken from samples of size n. p is the true proportion of the population. Recall that $\hat{p} = \frac{X}{n}$ where X is the number of successes in the sample.
$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	This is the formula for the standard deviation of the sampling distribution of sample proportions $\hat{p}$ taken from samples of size n. p is the true proportion of the population. (This formula is true when the 10% condition hold, if it applies.)
If $\bar{x}$ is the mean of a random sample of size n from an infinite population with mean μ and standard deviation σ, then:	
$\mu_{\bar{x}} = \mu$	This is the formula for the mean of the sampling distribution of sample means $\bar{x}$ taken from samples of size n. μ is the true mean of the population.
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	This is the formula for the standard deviation of the sampling distribution of sample means $\bar{x}$ taken from samples of size n. σ is the true standard deviation of the population. (This formula is true when the 10% condition hold, if it applies.)

III. Inferential Statistics

The branch of statistics that uses sample data to generalize about a population, or to draw a cause and effect relationship conclusion.

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$

This is the general formula for standardizing any statistic. Recall that a statistic is a number calculated from sample data. The parameter is the true population value. We use the standardized test statistic to conduct significance tests about a population parameter.

Confidence interval: $\text{statistic} \pm (\text{critical value}) \bullet (\text{standard deviation of statistic})$

This is the general formula for a confidence interval used to estimate an interval of plausible values for an unknown parameter.
The statistic is the point estimate of the parameter.
The critical value is determined by the confidence level and the distirbution of the test statistic.
In practice, we often use the standard error of the statistic in place of the standard deviation of the statistic, because the standard deviation requires unknown parameters. The standard error uses sample data to estimate these unknown values.
The product of critical value and standard deviation together form the margin of error, sometimes abbreviated ME.

Single-Sample

Statistic	Standard Deviation of Statistic	
Sample Mean	$\frac{\sigma}{\sqrt{n}}$	The standard error of this statistic is $\frac{s}{\sqrt{n}}$
Sample Proportion	$\sqrt{\frac{p(1-p)}{n}}$	The standard error of this statistic is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Two-Sample

Statistic	Standard Deviation of Statistic	
Difference of sample means	$\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ Special case when $\sigma_1 = \sigma_2$ $\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	
Difference of sample proportions	$\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$ Special case when $p_1 = p_2$ $\sqrt{p(1-p)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	

Chi-square test statistic = $\sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$