# Formulas for AP Statistics

## I. Descriptive Statistics

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{\sum x_i}{n}$$

$$s_x = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2} = \sqrt{\frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}}$$

$$\hat{y} = a + bx$$

$$\bar{y} = a + b\bar{x}$$

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$$b = r\frac{s_y}{s_x}$$

## II. Probability and Distributions

$$P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(A \cap B\right)$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

| Probability Distribution | Mean | Standard Deviation |
|---|---|---|
| Discrete random variable, $X$ | $\mu_X = E\left(X\right) = \sum x_i \cdot P\left(x_i\right)$ | $\sigma_X = \sqrt{\sum\left(x_i - \mu_X\right)^2 \cdot P\left(x_i\right)}$ |
| If $X$ has a **binomial** distribution with parameters $n$ and $p$, then: $P\left(X = x\right) = \binom{n}{x} p^x \left(1-p\right)^{n-x}$ where $x = 0, 1, 2, 3, \ldots, n$ | $\mu_X = np$ | $\sigma_X = \sqrt{np\left(1-p\right)}$ |
| If $X$ has a **geometric** distribution with parameter $p$, then: $P\left(X = x\right) = \left(1-p\right)^{x-1} p$ where $x = 1, 2, 3, \ldots$ | $\mu_X = \frac{1}{p}$ | $\sigma_X = \frac{\sqrt{1-p}}{p}$ |

## III. Sampling Distributions and Inferential Statistics

Standardized test statistic: $\frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$

Confidence interval: $\text{statistic} \pm \left(\text{critical value}\right)\left(\text{standard error of statistic}\right)$

Chi-square statistic: $\chi^2 = \sum\frac{\left(\text{observed} - \text{expected}\right)^2}{\text{expected}}$

Sampling distributions for proportions:

| Random Variable | Parameters of Sampling Distribution | | Standard Error* of Sample Statistic |
|---|---|---|---|
| For one population: $\hat{p}$ | $\mu_{\hat{p}} = p$ | $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$ | $s_{\hat{p}} = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ |
| For two populations: $\hat{p}_1 - \hat{p}_2$ | $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$ | $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$ | $s_{\hat{p}_1-\hat{p}_2} = \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$<br>When $p_1 = p_2$ is assumed:<br>$s_{\hat{p}_1-\hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$<br>where $\hat{p}_c = \dfrac{X_1 + X_2}{n_1 + n_2}$ |

Sampling distributions for means:

| Random Variable | Parameters of Sampling Distribution | | Standard Error* of Sample Statistic |
|---|---|---|---|
| For one population: $\overline{X}$ | $\mu_{\overline{X}} = \mu$ | $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$ | $s_{\overline{X}} = \dfrac{s}{\sqrt{n}}$ |
| For two populations: $\overline{X}_1 - \overline{X}_2$ | $\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2$ | $\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ | $s_{\overline{X}_1-\overline{X}_2} = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$ |

Sampling distributions for simple linear regression:

| Random Variable | Parameters of Sampling Distribution | | Standard Error* of Sample Statistic |
|---|---|---|---|
| For slope: $b$ | $\mu_b = \beta$ | $\sigma_b = \dfrac{\sigma}{\sigma_x\sqrt{n}}$,<br>where $\sigma_x = \sqrt{\dfrac{\sum(x_i - \bar{x})^2}{n}}$ | $s_b = \dfrac{s}{s_x\sqrt{n-1}}$,<br>where $s = \sqrt{\dfrac{\sum(y_i - \hat{y}_i)^2}{n-2}}$<br>and $s_x = \sqrt{\dfrac{\sum(x_i - \bar{x})^2}{n-1}}$ |

*Standard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the standard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.