AP Statistics Formula Sheet

Collect data you can trust

Random sample
supports population generalization
Random assignment
supports causal comparison
Blocking **randomize within similar groups**

A large convenience sample can still be biased.
In an observational study, confounding can explain an association.

Center, spread and outliers

Mean	$\bar{x} = \sum x_i / n$
Sample variance	$s^2 = \sum (x_i - \bar{x})^2/(n-1)$
Interquartile range	$IQR = Q_3 - Q_1$
Potential outlier fences	$Q_1 - 1.5\,IQR,\ Q_3 + 1.5\,IQR$

Divide by N for the variance of a whole population.
Sample SD is sqrt(s^2).
Median and IQR are less sensitive to extreme observations.

Events and their complements

Complement	$P(\text{not } A) = 1 - P(A)$
A or B	$P(A) + P(B) - P(A \text{ and } B)$
Disjoint events	$P(A \text{ and } B) = 0$
Independent events	$P(A \text{ and } B) = P(A)P(B)$

Disjoint events with positive probabilities cannot be independent.

Random variable mean and spread

Discrete mean	$E(X) = \sum x\,P(X = x)$
Variance	$Var(X) = E(X^2) - E(X)^2$
Linear transformation	$Var(aX + b) = a^2\,Var(X)$
Independent sum	$Var(X + Y) = Var(X) + Var(Y)$

SD is the square root of variance. Expectation adds even without independence.

Binomial and geometric trials

Binomial: k successes	$P(X = k) = C(n,k)p^k(1-p)^{n-k}$
Binomial moments	$E(X) = np,\ Var(X) = np(1-p)$
First success at trial k	$P(X = k) = (1-p)^{k-1}p$
Geometric mean	$E(X) = 1/p$

Use independent trials with constant success probability p. Geometric k starts at 1.

Normal model and z-scores

Standardize	$z = (x - \mu)/\sigma$
Convert back	$x = \mu + z\,\sigma$
Within 1, 2, 3 SDs	**about 68%, 95%, 99.7%**

Use positive sigma. The percentages describe a normal distribution.

Sampling distributions

Sample mean	$E(\bar{x}) = \mu,\ SD = \sigma/\sqrt{n}$
Sample proportion	$E(\hat{p}) = p,\ SD = \sqrt{p(1-p)/n}$

Standard error
replace unknowns with estimates

The SD formulas assume independent observations.
Check the 10% rule for samples without replacement.
Check approximate normality before using a normal critical value.

Regression and residuals

Predict y from x	$\hat{y} = a + bx$
Least-squares slope	$b = r\,s_y/s_x$
Intercept	$a = \bar{y} - b\,\bar{x}$
Residual	$e = y - \hat{y}$

With an intercept, r^2 is the fraction of y variation explained by this line.
Use a residual plot to check the model and avoid extrapolating far beyond the data.

AP Statistics Formula Sheet

Confidence intervals for proportions

One proportion	$\text{phat} \pm z^* \sqrt{(\text{phat}(1-\text{phat})/n)}$
Two proportions	$\text{estimate} = \text{phat}_1 - \text{phat}_2$
Two proportions	$\text{interval} = \text{estimate} \pm z^* \text{ SE}$
Two-group SE	$\sqrt{(p_1 q_1/n_1 + p_2 q_2/n_2)}$

In the SE, p_i are sample proportions and q_i = 1 − p_i. Do not pool for an interval.

Check independence and at least ten observed successes and failures in each group.

Test statistics and decisions

Standardize	$(\text{estimate} - \text{null})/\text{SE}$
One-proportion null SE	$\sqrt{(p_0(1 - p_0)/n)}$
One-mean t	$(\text{xbar} - \text{mu}_0)/(s/\sqrt{(n)})$
Reject H_0 if	$\text{p-value} \le \text{significance level}$

Use null SE for z and estimated SE for t.

For a two-proportion test of equality, pool counts.

AP: Compare two proportions

Pool under equality	$p_{pool} = (x_1 + x_2)/(n_1 + n_2)$
Pooled SE	$\sqrt{(p_{pool}(1-p_{pool})(1/n_1+1/n_2))}$
Standardize	$z = (\text{phat}_1 - \text{phat}_2)/\text{SE}$

x_i counts successes. Check independent groups and expected successes and failures.

For a confidence interval, use separate observed proportions in the standard error.

Chi-square for a two-way table

Expected count	$(\text{row total})(\text{col total})/n$
Test statistic	$\sum (\text{observed} - \text{expected})^2/\text{expected}$
Degrees of freedom	$(r - 1)(c - 1)$

Check independence and expected counts at least five.

O and E are cell counts. The p-value is the right tail.

Confidence intervals for means

One mean	$\text{xbar} \pm t^* s/\sqrt{(n)}$
Paired mean difference	$\text{dbar} \pm t^* s_d/\sqrt{(n)}$
Independent groups	$\text{estimate} = \text{xbar}_1 - \text{xbar}_2$
Independent group SE	$\sqrt{(s_1^2/n_1 + s_2^2/n_2)}$

Use t with appropriate degrees of freedom. For paired data analyze the differences.

Check randomization, independence and the shape or sample size of each sample.

AP: Choose a comparison for means

Paired design	$\text{form one list of differences}$
Paired t, null 0	$t = \text{dbar}/(s_d/\sqrt{(n)})$
Two independent groups	$t = (\text{xbar}_1 - \text{xbar}_2)/\text{SE}$
Independent group SE	$\sqrt{(s_1^2/n_1 + s_2^2/n_2)}$

For a paired test, check the differences and use df = n − 1.

For independent groups, check both samples and use Welch df.

AP: State, check and conclude

Random process **identify sample or assignment**

Independence

check design and sampling fraction

Approximation

justify expected counts or shape

Conclusion **answer in context using the result**

A confidence interval estimates a parameter. A significance test evaluates a stated null.

Failing to reject does not establish that the null is true.