AP Statistics

Formula Sheet · 公式表

AP Statistics supplies every formula and every table. Not one point on the exam is available for recalling them. The graded decisions are which procedure the situation calls for, whether its conditions are met, and how to state the conclusion in context – so that is what this booklet lists. 公式与统计表全部提供，考的是选择正确的方法、检验条件并结合情境作结论。

Given on the AP formula sheet

Descriptive statistics: the mean, the sample standard deviation, and the least-squares regression slope and correlation coefficient.

Probability: the addition and multiplication rules, conditional probability, the binomial and geometric distributions, and the mean and standard deviation of each.

Sampling distributions and inference: the standard error for every procedure tested, the test statistic in the form (statistic minus parameter) over standard error, and the chi-square statistic.

Tables for the standard normal, t and χ^2 distributions.

Choosing the procedure – the decision the sheet cannot make

One proportion or two — Counting successes in categories and asking about a percentage means a proportion procedure. One group compared against a claimed value is one-sample; two groups compared with each other is two-sample.

One mean or two, paired or independent — Measurements with units mean a mean procedure. If each observation in one group is naturally matched with one in the other – before and after, twins, left and right – it is PAIRED, and you analyse the differences with a one-sample t. Treating paired data as two independent samples is the most costly single error on the exam.

Chi-square, and which one — Goodness of fit compares one sample against a claimed distribution. Homogeneity compares the same variable across several populations. Independence tests two variables measured on one sample. Naming the wrong one loses the point even when the arithmetic is identical.

Regression inference — A question about whether a linear relationship exists in the population is a t-test for the slope, with the null hypothesis that the slope is zero.

Conditions – state them with the study's own numbers

Random — Say how the data were collected and quote it: "the problem states the sample was randomly selected". If it was not random, say so and note that inference to the population is not justified.

Independence and the 10 per cent rule — Sampling without replacement requires the sample to be under 10 per cent of the population. Write the actual comparison: $50 < 0.10 \times 20{,}000$.

Normality — For proportions, both np and $n(1-p)$ must be at least 10 – compute both. For means, either the population is normal, or $n \geq 30$, or the given plot shows no strong skew and no outliers, and you must say which of the three you are relying on.

Expected counts — Every expected count in a chi-square procedure must be at least 5, and the expected counts are not the observed ones. Check them explicitly.

The conclusion – a fixed shape, scored strictly

With a p-value — Compare it to alpha, then reject or fail to reject, then state what that means about the ORIGINAL question in context. "Since $p = 0.03 < 0.05$, we reject the null hypothesis; there is convincing evidence that the mean commute time exceeds 25 minutes."

Never accept the null — A large p-value gives "we fail to reject" and "we do not have convincing evidence that". It never gives "the means are equal" or "we accept the null hypothesis" – that phrasing loses the point outright.

Interpreting a confidence interval — "We are 95 per cent confident that the interval from 2.1 to 4.7 grams captures the true mean weight of all the bars." The confidence is in the method, not in this one interval, and it is about a parameter, not about individuals.

Interpreting the confidence level — It means that in repeated sampling, about 95 per cent of the intervals built this way would capture the true parameter. This is a different question from the previous one and the exam asks both.

The distinctions that carry the most points

An observational study can establish association; only a randomised experiment with random assignment can establish causation. Random SELECTION allows generalisation; random ASSIGNMENT allows causal claims. These two are examined together to see whether you can separate them.

Interpret a slope as the predicted change in y per one-unit increase in x, with both units named, and always say "predicted". Interpret r^2 as the percentage of variation in y explained by the linear relationship with x.

Extrapolating beyond the observed range of x is unjustified, and saying so is worth a point when the question invites it.