

AP Statistics is marked on communication as much as computation: name the procedure, state the conditions, and interpret the result IN CONTEXT. A correct number with no context earns partial credit at best. Learn each definition and practise attaching “...in this study” to it. 统计题重在表述: 写出方法、检验条件, 并结合题目情境解释结果。

Exploring data

- Statistic** 统计量 — a number computed from a sample, used to estimate a population parameter
- Parameter** 参数 — a number describing a population, usually unknown
- Outlier (1.5 IQR rule)** 异常值 — an observation more than $1.5 \times \text{IQR}$ below Q1 or above Q3
- Resistant measure** 稳健统计量 — a statistic little affected by extreme values; the median and IQR are resistant, the mean and standard deviation are not
- Standard deviation** 标准差 — a measure of the typical distance of observations from the mean
- z-score** 标准分 — the number of standard deviations an observation lies from the mean

Collecting data

- Simple random sample** 简单随机抽样 — a sample chosen so that every group of the given size has an equal chance of being selected
- Stratified random sample** 分层随机抽样 — a sample formed by dividing the population into homogeneous strata and taking a random sample within each
- Confounding variable** 混杂变量 — a variable whose effect on the response cannot be separated from that of the explanatory variable
- Blocking** 区组化 — grouping experimental units that are similar and randomising treatments within each group, to reduce the effect of a known source of variability
- Random assignment** 随机分配 — assigning treatments to units by chance; it is what permits a conclusion of CAUSATION
- Bias** 偏倚 — a systematic tendency for a sampling or measurement method to favour certain outcomes

Probability and distributions

- Independent events** 独立事件 — events for which knowing one occurred does not change the probability of the other
- Binomial setting** 二项分布条件 — a fixed number of independent trials, two outcomes per trial, and the same probability of success on each
- Geometric setting** 几何分布条件 — independent trials with constant success probability, counting trials until the FIRST success
- Sampling distribution** 抽样分布 — the distribution of a statistic over all possible samples of a given size from the population

Central limit theorem 中心极限定理 — for a sufficiently large sample, the sampling distribution of the sample mean is approximately normal whatever the population’s shape

Inference

- Confidence interval** 置信区间 — an interval computed from sample data that would capture the true parameter in a stated percentage of repeated samples
- Confidence level (interpretation)** 置信水平 — the percentage of all possible samples whose interval would contain the parameter – it is a statement about the METHOD, not about one interval
- p-value** p 值 — the probability of obtaining a result at least as extreme as the observed one, ASSUMING the null hypothesis is true
- Significance level** 显著性水平 — the threshold α below which the p-value leads to rejecting the null hypothesis
- Type I error** 第一类错误 — rejecting a null hypothesis that is in fact true
- Type II error** 第二类错误 — failing to reject a null hypothesis that is in fact false
- Power** 检验效能 — the probability of correctly rejecting a false null hypothesis; it increases with sample size and with effect size