

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.

Exploring data

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

Collecting data

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

Probability and distributions

- Independent events of the other** – events for which knowing one occurred does not change the probability
- Binomial setting** – a fixed number of independent trials, two outcomes per trial, and the same probability of success on each
- Geometric setting** – independent trials with constant success probability, counting trials until the FIRST success
- Sampling distribution from the population** – the distribution of a statistic over all possible samples of a given size

Central limit theorem – for a sufficiently large sample, the sampling distribution of the sample mean is approximately normal whatever the population's shape

Inference

Confidence interval – an interval computed from sample data that would capture the true parameter in a stated percentage of repeated samples

Confidence level (interpretation) – 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 – the probability of obtaining a result at least as extreme as the observed one, ASSUMING the null hypothesis is true

Significance level – the threshold α below which the p-value leads to rejecting the null hypothesis

Type I error – rejecting a null hypothesis that is in fact true

Type II error – failing to reject a null hypothesis that is in fact false

Power – the probability of correctly rejecting a false null hypothesis; it increases with sample size and with effect size