A: Confidence interval: tells you both location and precision of a statistic.
a 95% interval includes all parameters for which p-value > 0.05
Many statistics:

$$\text{estimate} \pm T_{prob} \times \text{se}$$

T_{prob} is a quantile of the appropriate T distribution
T_{prob} approx 2 for 95% interval unless df small

How quantiles are labelled:
T_{prob} is the value (quantile) such that $P[T \leq T_{prob}] = prob$
so *prob* of the T distribution is less than or equal to the quantile
For a 95% confidence interval, we need 2 T quantiles
one has 2.5% probability below it
other has 2.5% probability above it = 97.5% probability below it
We need $T_{0.025}$ and $T_{0.975}$ for the appropriate degrees of freedom
For a 99% interval, we need $T_{0.005}$ and $T_{0.995}$
These have 99% of the T distribution between them
Often written as $T_{\alpha/2}$ and $T_{1-\alpha/2}$ for a $1 - \alpha$% confidence interval
The T distribution is symmetrical, so $T_{0.025} = -T_{0.975}$
So sometimes see: $(-T_{0.975}, T_{0.975})$

CI connected to (some) information about the p-value:
95% interval includes 0, e.g., (-2, 4) ⇒ p-value for test of $0 > 0.05$
95% interval does not include 0, e.g., (-2, -1) or (3, 100)⇒ p-value for test of $0 < 0.05$
99% interval does not include 0 ⇒ p-value for test of $0 < 0.01$

B: Two schools of inference:
Frequentist: parameters have fixed but unknown values
Bayesian: parameters are random variables, so have a distribution

Confidence intervals are frequentist:
The interval is random; it depends on random values (mean, se)
Different samples give different confidence intervals
Most include the true parameter; some do not. Don't know which is which
Coverage is P[interval includes true but unknown parameter]
Example: Calculate a 95% ci using estimate $\pm T_{quantile}$se and get (5,10).
How can you interpret this?
Correct: The interval calculated by estimate $\pm T_{quantile}$se has an 0.95 probability of including the true but unknown parameter
technically Incorrect: claim P[parameter between 5 and 10] = 0.95
Because probabilities are properties of random variables.
Parameter is not a random variable

Bayesian interval called a credible interval:
Allowed to talk about P[parameter between 5 and 10] or P[parameter > 0]

Different calculations,
need additional information (the prior distribution), and lots more math
but in almost all cases, can set up the problem so that
the credible interval is exactly the same as the confidence interval
So, I don't make a distinction
you can interpret a confidence interval as a credible interval.

More methods for two-sample comparisons:
Chapter 3: assumptions of t-test, methods to assess, transformations
Chapter 4: alternatives to a t-test

C: Nonparametric test: third way to compare two groups
First two were randomization test and T-test
also called distribution-free test, we will ignore a subtle distinction
do not assume a specific distribution

Why consider more methods?
sometimes hard to use other methods: example: Case study 4.2 cognitive load study
some values recorded as > 300 seconds - can't (easily) compute mean
sometimes data violate assumptions: example hamburger E coli study
one value of 2.5 cfu/gm; rest are 0.03 - 0.57 cfu/gm

Resistant method:
results don't change when a small number of observations altered a lot
median is resistant; mean is not

Wilcoxon rank-sum test: compares locations of two groups
also called Mann-Whitney test or Mann-Whitney U test
resistant to outliers
only requires ranking observations, so > 300 is not an issue
the test statistic is sum of ranks (or average rank) for one group
extreme when that group tends to have all the largest obs. or all the smallest obs.
get p-value by randomization (small n_1 or n_2) or approximation (large n_1 and n_2)

why resistant? Consider the hamburger data
2.5 in the control group is the largest value, gets rank 12
next largest value is 0.57 so 0.6 would get rank 12 and 1000 would get rank 12.
change that value a lot and still have same sum of ranks for the control group
Null hypothesis being tested: various ways to state this
Practically useful version: H0: two groups have same median (or no diff. in medians)
difficult to obtain CI for difference in medians. See someone in Stat if you need this.

Paired data nonparametric tests: based on difference within each pair (like paired T test)
sign test: # positive differences (so value in trt A $>$ value in trt B)
ignores magnitude of the difference

only requires $>$ or $<$; can use when can't compute the diff.'s, e.g., censored values
Test based on # pairs with positive differences
Wilcoxon signed rank test: includes size of the difference, not just + or −
Details in the book - not relevant
when both sign and signed rank tests are possible,
signed rank test is more powerful ⇔ more likely to find a difference

D: Review and comparisons of assumptions

Assumptions of a randomization test:
independence: observations randomly assigned to treatments,
done individually one observation at at time
or: two independent simple random samples

Assumptions of a t-test:
independence: observations randomly assigned to treatments, one at at time
or: two independent simple random samples
or: errors are independent
errors have equal variances
errors are normally distributed

Assumptions of a Wilcoxon rank sum test: (various ways to describe them)
independence: observations randomly assigned to treatments, one at at time
or: two independent simple random samples
or: errors are independent
errors have the same distribution, perhaps after a transformation

Model that leads to the equal variance T test:
Notation:
Observations, Y_{ji} are indexed by group j: A or B, and observation # within group, i.

$$\text{For group A} \quad Y_{Ai} = \mu_A + \varepsilon_{Ai}, \quad \varepsilon_{Ai} \sim \text{independent } N(0,\, \sigma^2)$$
$$\text{For group B} \quad Y_{Bi} = \mu_B + \varepsilon_{Bi}, \quad \varepsilon_{Bi} \sim \text{independent } N(0,\, \sigma^2)$$

Assumptions:
model represents assumptions in words
normally distributed: $N()$
equal variances: $N(0, \sigma^2)$ (one variance for both groups),
not $N(0, \sigma_A^2)$ and $N(0, \sigma_B^2)$ (each group has own variance)
assumptions concern errors

E: Errors and residuals
Errors: unknown population quantities
error for observation Y_{Ai}: $\varepsilon_{Ai} = Y_{Ai} - \mu_A$

unknown because need μ_A (unknown pop. mean) to compute Residuals:
estimates of the errors, called residuals, can be computed from data
residual is $Y_{Ai} - \overline{Y}_A$
because estimate of μ_A is $\overline{Y}_A$

F: Diagnosing concerns about assumptions:
Goal: when is T-based inference reasonable?
Answer: when assumptions are reasonable
because then T distribution matches randomization distribution
use residuals to assess assumptions
Emphasize informal approaches, not statistical tests, no p-values

G: Equal variance: diagnosis
numeric: easiest with only two groups
s_1 sd in group 1; s_2 sd in group 2. $s_1 > s_2$
$\frac{s_1}{s_2} \leq 2$: fine
$\frac{s_1}{s_2} \geq \sqrt{10} \approx 3.16$: concern about unequal variances
graphical:
plot residuals (Y axis) against predicted values (X axis)
predicted values for a two-group study are $\overline{Y}_A$ and $\overline{Y}_B$
so residual vs predicted value plot has two columns of dots
look for unequal spreads
See plots from lecture:
simulated data sets with sd's equal (1:1), 1:2, and 1:3.16

Treatment of outliers (Display 3.6):
Don't just delete! Hamburger study: 2.5 cfu/gm is the most important value
Book (and my) recommendations:
Is there any error (in measurement, transcription, ...): fix
Is that obs from a different population:
focus on the majority, remove all obs from that minor population
Analyze data with and without the "outliers":
similar results, report with all obs
different results, report both, probably emphasize one set.
CONSORT diagram for flow of individuals through study