**A:** One- and two-sided null hypotheses:
Notation:
$\delta$ = population difference in the means
$d$ = sample estimated difference in averages
two-sided: $H_0$: $\delta = 0$, $H_a$: $\delta \neq 0$
reject $H_0$ when $d$ sufficient large ($d > 0$) or sufficiently small ($d < 0$)
one-sided: two possibilities, depending on which "side"
$H_0$: $\delta \leq 0$, $H_a$: $\delta > 0$
reject $H_0$ when $d$ sufficient large ($d > 0$)
$H_0$: $\delta \geq 0$, $H_a$: $\delta < 0$
reject $H_0$ when $d$ sufficient small ($d < 0$)
p-value for two-sided test "counts" both tails
Twice the smaller one-sided p-value, when randomization distribution symmetric
Two-sided tests much more common, even when only one-side is interesting
Less opportunity to "fudge" the result
Can't pick the side after seeing the data

**B:** Model based inference: T-test.
Could use difference of means, but every data set would have a different distribution
Null says $\delta = 0$, doesn't say anything about spread
Stat theory $\Rightarrow$ a statistic with a known distribution

$$T = \frac{\text{statistic} - \text{parameter}}{\text{se of statistic}}$$

Usually, parameter = 0 (H0 of no difference) so $T$ = statistic/se
When assumptions (chapter 3) appropriate, $T$ has a theoretical T distribution
Many T distributions: which one depends on (error) degrees of freedom
Reject H0 when $T$ unusually large or unusually small (for 2 sided test)
T $> 2$ or $< -2 \Rightarrow$p $< 0.05$ (approximately, unless df small)

One sample: statistic is the average:

$$se = s/\sqrt{n}$$

**C:** Two independent samples: commonly used statistic is the difference in averages

$$\begin{aligned}
\text{se} &= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \text{when variances unequal} \\
&= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\
&= \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \text{when variances equal} \\
&= s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \text{estimated pooled variance}
\end{aligned}$$

Note: used "plug-in" principle: using $s^2$ as the estimate of $\sigma^2$
Strong theoretical justification for doing this

Pooled error variance:
$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

**D:** Degrees of freedom: how much information used to estimate a variance (or sd)
how much data minus number of "things" you need to estimate
"how much data" = # observations used to compute $s$
one group:
"how much data" = $n$, need to estimate mean (1 parameter), so df = $n - 1$
two groups, pooled variance:
"how much data" = $n_1 + n_2$
need to estimate 2 means, so df = $n_1 + n_2 - 2$
two groups, unequal variances:
df is complicated, usual approach is the Satterthwaite approximation
not an integer, covered in Stat 571
Analysis often called the Welch T-test

**E:** Paired data: study design to increase precision and reduce confounding (in obs. study)
compute difference for each subject (pair).
compute mean difference and sd of differences (across subjects)
se = $s/\sqrt{n}$
Note: This is the se **of the difference** because it is computed directly from differences
df = $n - 1$ because there $n$ values (each a pair)
Serious error to ignore pairing when it exists (Design matters!)
Or to use paired analyses when data aren't

**F:** Mis-interpretations of p-values:
Frequently mis-interpreted, leading many to reject any use of p-values
Does not prove null hypothesis (even if p-value large)
Does not show that alternative more likely than null.

Hypothesis test really doesn't tell you much
if $p$ small, test gives you evidence that the parameter is **not zero**.
even though p-values are all over the scientific literature

**G:** 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) $\Rightarrow$ p-value for test of 0 $> 0.05$
95% interval does not include 0, e.g., (-2, -1) or (3, 100)$\Rightarrow$ p-value for test of 0 $< 0.05$
99% interval does not include 0 $\Rightarrow$ p-value for test of 0 $< 0.01$

**H:** 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.