A: Spread: variability of observations:
sample variance: $s^2 = \Sigma(X_i - \overline{X})^2/(n-1)$
why $(n-1)$? Data used twice: to compute $\overline{X}$ and then to compute s^2
account for this by adjusting the sample size
called "degrees of freedom" or df. More later
standard deviation: $s = \sqrt{\text{Variance}}$
coefficient of variation, cv: $s/\overline{X}$, measures relative variation.

B: Inference: two types: Confidence interval, hypothesis test
Spoiler: multiple opinions and lots of controversy.
Lots of tricky epistemological and practical issues.
I give the traditional approach
Some perspectives in the refeences to extra material section of the non-canvas web site

C: Hypothesis test and p-values:
Assume something (null hypothesis) is correct, ask whether observations are unusual
If so, either a fluke happened or the null hypothesis assumption is probably wrong
p-value: probability of observing data or something more extreme given null hypothesis

Scale of evidence: Display 2.12, applied to difference of two group means
Null hypothesis is no difference in means $\Leftrightarrow$ equal means
p $>$ 0.10: no evidence of a difference
not: "null is true", e.g., not: "X has no effect", not "Two groups have the same mean"
0.05 $<$ p $<$ 0.10: weak evidence of a difference
0.01 $<$ p $<$ 0.05: evidence of a difference
0.001 $<$ p $<$ 0.01: strong evidence of a difference
p $<$ 0.001: very strong evidence of a difference

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.
Limited knowledge: small p only tells you parameter not zero (for most frequent H0).

D: Randomization test: design-based inference
randomly reassign labels, compute S = statistic of interest
e.g., difference in means or absolute value of difference in means: $\mid \overline{Y}_A - \overline{Y}_B \mid$
p = P[obs S more extreme than random S], e.g., (two-sided hypothesis)
P[$\mid \overline{Y}_A - \overline{Y}_B \mid \geq$ random $\mid \overline{Y}_A - \overline{Y}_B \mid$]

One- and two-sided null hypotheses:
Notation:
δ = 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
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

Two very similar versions of a randomization test:
permutation test: enumerate all possible treatment arrangements
randomization test: randomly sample from all possible treatment arrangements
want/need to sample when too many possibilities:
creativity: >16,123,800,000,000 arrangements

Calculating the p-value from a permutation or randomization test:
need to know:
R: number of as or more extreme randomizations,
N: number of randomizations
permutation: observed arrangement is one of the possibilities: $p = R/N$
randomization: observed arrangement is one more possibility: $p = (R + 1)/(N + 1)$
which is why commonly see 999 or 9999 randomizations

Randomization/Permutation tests can be very useful, especially for simple studies:
logic is clear
no model, few assumptions.
key assumptions:
random assignment of treatments to each unit
observations are independent $\Leftrightarrow$ treatment assigned separately to each observation
applicable when standard models aren't easy:
e.g., inference on difference of medians or ratio of means
But can be (very) hard to implement for non-simple (most) studies

E: How precise are estimates?
standard deviation: estimated variability in **observations**
does not get systematically smaller with larger sample sizes
standard error: estimated precision of a statistic (e.g., mean)
for simple random sample: $\text{se}_{average} = s/\sqrt{n}$
Generally smaller with larger sample sizes
Can use se to determine number of replicates to use in a study
Spoiler: sd and se can be defined in multiple ways. My use focuses on the "endpoints".
you might see "standard deviation of the mean": se, but with known sd, not estimated

Which should you report?
My suggestion: what do you want to describe?
variability in observations: report sd
precision of the average (or other statistic): report se
Correctly say which statistic you are reporting
editor's note on Med diet retraction:
reviewed 934 manuscripts in NEJM
found 11 with curious results.
5 of those were because se was incorrectly reported as sd or vice-versa

When reporting an average, should you report 106.4 or 106.3867190?
Kelley's rule: Science 60:524 (1924)
report a statistic (e.g., an average) to the leading digit of se/3
Estimate from computer: 106.3867190,

se	se/3	report:
0.0008	0.00026	106.3867
0.02	0.0066	106.387
0.06	0.02	106.39
9	3	106
50	17	110

F: Model based inference: T-test

$$T = \frac{\text{statistic} - \text{H}_0 \text{ parameter}}{\text{se of statistic}}$$

Many T distributions: which depends on (error) degrees of freedom
T > 2 or $< -2 \Rightarrow$ two-sided p < 0.05 (approximately, unless df small)
Need to know se of various statistics
se for mean from a single population: se $= s/\sqrt{n}$
Now, need to know se for a difference of two means

SE of a difference of two means
Two independent samples: statistic is difference in averages

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

Pooled error variance:

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

Degrees of freedom: how much information used to estimate a variance (or sd)
n observations in a group minus number of estimates
one group: need to estimate mean (1 parameter), so $n - 1$ for one group
$n_1 + n_2 - 2$ for two groups (estimate two means)

G: Paired data:
compute difference for each subject (pair).
compute mean difference and sd of differences (across subjects)
se $= s/\sqrt{n}$
df $= n - 1$

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

H: 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: estimate $\pm\, T_{quantile} \times$ se
$T_{quantile}$ approx 2 for 95% interval unless df small
95% interval includes 0 $\Rightarrow$ p-value for test of 0 > 0.05
95% interval does not include 0 $\Rightarrow$ p-value for test of 0 < 0.05