Confidence intervals give a range of plausible values for the true parameter, while hypothesis tests give an idea about how likely our data is, if H_0 was true.

Connection between CIs and HTs

For a two-sided hypothesis test and CIs, the results will always match. I.e.

1) If a two sided HT rejects $H_0: \mu_1 = \mu_2$ for a specific α, the $(1-\alpha)100\%$ CI will not contain $\emptyset$ (or vice versa).

2) If a $(1-\alpha)100\%$ CI contains zero, we will fail to reject H_0 at α.

Ex: The recovery time (in days) for a particular virus is measured for two groups, with the following summary statistics:

	Drug Group (1)	Placebo Group (2)	
$\bar{y}$	11.4	12.3	
s	1.1	1.25	$r = 17$
n	10	10	

a) State H_0, H_A for testing if the drug group has a significantly lower average recovery time.

$H_0: \mu_1 \geq \mu_2$ vs $H_A: \mu_1 < \mu_2$

b) Calculate the test-statistic, and interpret it.

$$t_S = \frac{(11.4 - 12.3) - 0}{\sqrt{1.1^2/10 + 1.25^2/10}} = -1.724$$

Our observed difference is 1.724 estimated std. dev's below the null value of 0.

c) Calculate the range for the p-value.

Our p-value formula is $Pr\{t < t_S\}$ since $H_A: \mu_1 < \mu_2$, or $Pr\{t < -1.724\} = Pr\{t > +1.724\}$ by symmetry.

At d.f = 17 (row: 17), 1.333 is in column 0.10, and 1.740 is in column 0.05. Thus, our p-value is:
$0.05 < \text{p-value} < 0.10$

d) Interpret your p-value from c) in terms of the problem.

If in reality the true average recovery time for the drug group was NOT lower than the placebo group, we would observe our data or more extreme with probability between 0.05 and 0.10.

e) State your decision, and conclusion in terms of the problem, assuming $\alpha = .10$

Decision: p-value $< \alpha$, so we reject H_0.

We conclude that we support the claim that the drug group has a lower average recovery time than the placebo group.

Errors in Hypothesis Testing

Since μ_1, μ_2 are unknown, it is possible that we made an error when we made a decision about H_0. There are two types of error we could make:

Type I error: We reject H_0, when in reality H_0 was true

Type II error: We failed to reject H_0, when in reality H_0 was false (H_A true).

In a table, we have four possibilities:

	Decision	
Reality ↓	Reject H_0	Fail to reject H_0
H_0 true	Type I error	correct
H_0 false	correct	Type II error

We can control the probability of one type of error, but not the other.

Definitions:

$\alpha = Pr\{\text{Type I error}\} = Pr\{\text{Reject } H_0 \mid H_0 \text{ true}\}$

$\beta = Pr\{\text{Type II error}\} = Pr\{\text{Fail to reject } H_0 \mid H_0 \text{ false}\}$

We can control α ($\alpha = .10, .05, .01$) since it assumes H_0 is true, which is a very specific value. "H_0 false" on the other hand has too many possibilities, so the distribution of t_S is non-specific. We cannot calculate it.

Related to errors is the power of a test:

power $= 1 - \beta = 1 - Pr\{\text{Fail to reject } H_0 \mid H_0 \text{ false}\}$

$= Pr\{\text{Reject } H_0 \mid H_0 \text{ false}\}$ = The prob of a correct decision (one of them).

Facts:

i) As α increases, β decreases

ii) As β increases, α decreases

iii) Higher power is ideal, and lower errors.

iv) Depending on which error you believe is worse, you can choose different α's.

Ex: A treatment for a terminal disease is thought to prolong life (increase ave. time to death) compared to an old treatment. Let time to death be measured in months.

a) Write down H_0, H_A.

Let group 1 = new, group 2 = old. $H_0: \mu_1 \leq \mu_2$, $H_A: \mu_1 > \mu_2$

b) Describe what a Type I error is in terms of the problem.

Type I = reject H_0 | H_0 true

= We concluded that the new drug prolonged life, when in reality it did not (did not change, or decreased).

c) Describe what a Type II error is in terms of the problem.

Type II = fail to reject H_0 | H_0 false

= We concluded that the new drug did not prolong life, when in reality it did.

* Assumptions of a HT for $\mu_1 - \mu_2$

The same assumptions used for a CI:

1) A random sample was taken from both populations
2) Populations 1 and 2 are independent
3) $\bar{Y}_1$ and $\bar{Y}_2$ are normally distributed, either because
   i) $n_1 \geq 30$ and $n_2 \geq 30$ or
   ii) Populations 1 and 2 are normally distributed.

Important reminders:

1) We assume H_0 is true so that mathematically we can assume the "true" center of $\bar{Y}_1 - \bar{Y}_2$.
2) The p-value is not the probability H_0 is true. It is the probability of our data or more extreme, if H_0 is true.
3) The "claim" of interest does not have to be in H_0

* Comparing means of more than two groups

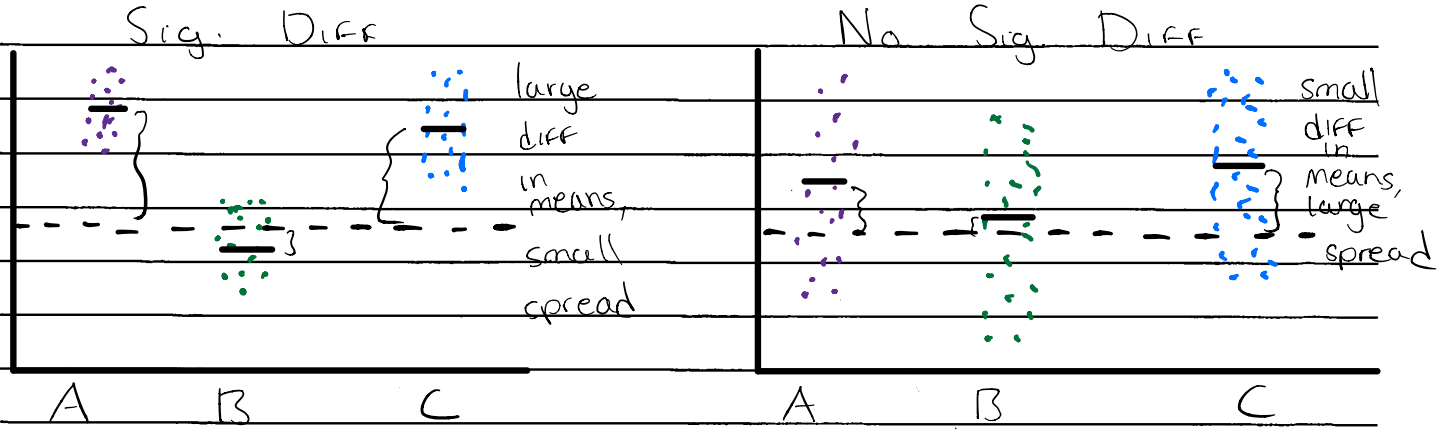
What if we had three groups, say 1 = drug, 2 = control, 3 = placebo. What could we do?

One option (not the ideal) is to do 3 hypothesis tests:
Test 1: $H_0: \mu_1 = \mu_2$
Test 2: $H_0: \mu_1 = \mu_3$
Test 3: $H_0: \mu_2 = \mu_3$
Then based on these three tests, we could determine which of the three groups differ.
But, our type I error increases with every test we do, so our overall type I error is huge.
Instead, we use a different approach.

Consider comparing the averages of each group, and taking into account the variance of each group as well.

But, instead of comparing group means to eachother, we compare them to what the mean is regardless of group (the overall mean).

If there is a large difference in the group means to the overall mean, and the group variances are low, we can say there is a "significant" difference in means:



Now we need to quantify the difference in means to overall mean, and the overall spread of the data (in all groups combined).

Notation:

Let I = # of groups

y_{ij} = jth observation from group i

n_i = sample size of i^{th} group

$\bar{y}_i$ = sample mean of i^{th} group

s_i = sample std. dev of i^{th} group

$n_\cdot$ = total sample size regardless of groups
$= \sum_{i=1}^{I} n_i$

$\bar{\bar{y}}$ = overall sample average, regardless of groups
$= (\sum_{i=1}^{I} n_i \bar{y}_i)/n_\cdot$

Ex: Three groups of men tried different diet programs, and their weight loss was measured (6 months later):

(1) A (by themselves): 3, 4, 5, 3, 7, 6

(2) B (with a friend): 7, 8, 5, 9, 2, 8, 5

(3) C (with a program): 1, 3, 7, 3, 4, 8, 2, 8, 4

a) Find I, all n_i's, and $n_\cdot$.

$I = 3$, $n_1 = 6$, $n_2 = 7$, $n_3 = 9$, $n_\cdot = 6+7+9 = 22$

b) Find all $\bar{y}_i$'s

$\bar{y}_1 = 28/6 = 4.66$, $\bar{y}_2 = 44/7 = 6.28$, $\bar{y}_3 = 40/9 = 4.44$

c) Find $\bar{\bar{y}}$

$\bar{\bar{y}} = (\sum n_i \bar{y}_i)/n_\cdot = (4.66(6) + 6.28(7) + 4.44(9))/22$
$= (28 + 44 + 40)/22 = 5.09$

Further, $s_1 = 1.63$, $s_2 = 2.42$, $s_3 = 2.60$

d) Which group has the highest mean, and which the highest variance?

Mean = B, variance = C.

The groups differ from the overall mean by:

A	B	C
Diff: $(4.66-5.09)=-0.43$	$(6.28-5.09)=1.19$	$(4.44-5.09)=-0.65$

* Typical Deviation from a Group

To have an "overall deviation from group" that is NOT specific to a particular group, we calculate a weighted average of s_i's:

$$s_p^2 = \left(\sum_{i=1}^{n}(n_i-1)s_i^2\right)/(n.-I)$$

$$= \frac{\text{(total squared deviation by group, summed)}}{\text{(adjusted sample size)}}$$

"p" stands for "pooled"

Then, $s_p = \sqrt{s_p^2}$ = typical deviation from any group to its group mean.

For our example

$$s_p^2 = \left((5)(1.63^2)+(6)(2.412)^2+8(2.60)^2\right)/(22-3)$$

$$= (102.5029)/19 = 5.39$$

$$s_p = \sqrt{5.39} = 2.32.$$

ANalysis Of VAriance (ANOVA)

Analysis of variance splits the TOTAL (overall) variance in a group INTO TWO parts
i) Variance between groups
ii) Variance within groups

I) Total Variance, denoted SSTO or SS(Total)
$SSTO = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2$ = total squared deviation from overall mean
SSTO = Sum of Squares TOtal

II) Variance between groups, denoted SSB or SS(Between)
$SSB = \sum_{i=1}^{I} n_i(\bar{y}_i-\bar{y})^2$ = total squard deviation of group means to overall mean.
SSB = Sum of Squares Between

III) Variance within groups, denoted SSW or SS(within)
$SSW = \sum_{i=1}^{n}(n_i-1)(s_i)^2$ = total squared deviation by group to its group mean
SSW = Sum of Squares Within

Fact: SSTO = SSB + SSW

A problem with all of these is that they are going to get larger and larger. To "stabilize" them, we divide by each ones d.f.:
d.f.{TO} = n-1, d.f{W} = n.-I, d.f.{B} = I-1
Fact: d.f.{TO} = d.f.{B} + d.f{W}

Then, the "stabilized" versions of all the SS (Sum of Squares) are called MS (Mean Squares)

I) MSTO = SSTO / d.f.{TO} = SSTO/(n-1) = sample variance
II) MSB = SSB / d.f{B} = between group variance
III) MSW = SSW / d.f.{W} = within group variance

If MSB is large relative to MSW, then we may say there is a sig. diff in group means.

If MSB and MSW are approx. equal, then we may say there is no sig. diff in group means

Ex: Diet continued...

$SSW = \sum_{i=1}^{I}(n_i-1)s_i^2 = 102.5029$, $d.f.\{W\} = 22-3 = 19$,

$MSW = s_p^2 = 5.39$

$SSB = \sum_{i=1}^{I} n_i(\bar{y}_i-\bar{\bar{y}})^2 = 6(4.66-5.09)^2 + 7(6.28-5.09)^2 + 9(4.44-5.09)^2 = 14.82$

$d.f.\{B\} = I-1 = 3-1 = 2$, $MSB = 14.82/2 = 7.41$

$SSTO = SSW + SSB = 102.5029 + 14.82 = 117.3229$,

$d.f\{TO\} = 22-1 = 21$ $MSTO = 117.3229/21 = 5.5868$

Here, MSW and MSB are pretty close. But how do we know if they are too close?

* HT for ANOVA

The question of interest for ANOVA is, "are all the means equal" vs "is at least one mean different".

Step 1) State H_0, H_A

H_0: $\mu_1 = \mu_2 = \mu_3 = \cdots = \mu_I$ (all means equal)

H_A: At least one μ_i is not equal.

Step 2) Calculate the test-statistic

We directly compare MSB and MSW:

F_S = MSB/MSW

= ("stabilized" difference in group means to overall mean) / ("stabilized" combined estimated std. dev's)

Notice if all group means were equal, $F_S = 0$

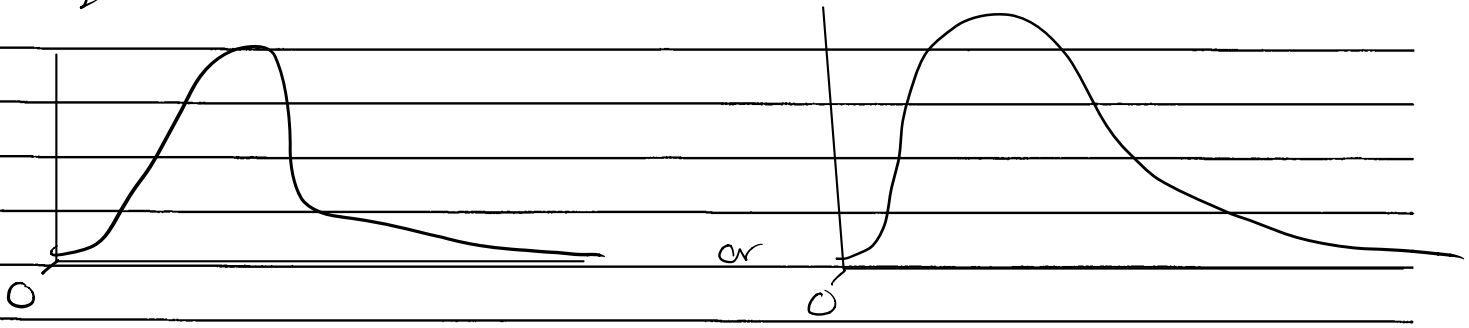
(since $\sum n_i(\bar{y}_i - \bar{\bar{y}})^2 = 0$), just like t_s. But, this is clearly not t-distributed, since $F_s \geq 0$ since MSW, MSB are sum of squared values.

Step 3) Calculate the p-value

If H_0 is true, then statisticians know that F_s is distributed F, with two d.f.:

d.f.{numerator} = d.f.{B}, d.f.{denominator} = d.f.{W}

An F distribution is approximately a t distribution, squared:



(depending on the d.f.)

And, "extreme" means F_s is further from 0, so

p-value = $P_r\{F > F_s\}$

The F-tables are exactly like the t-tables, but you have to match up both d.f.'s.

Step 4) Make decision and state conclusion

If p-value $< \alpha$ => reject H_0

If p-value $\geq \alpha$ => Fail to reject H_0

Ex: Diet continued....

a) State H_0, H_0

H_0: $\mu_1 = \mu_2 = \mu_3$ (all average weight loss is equal)

H_A: At least one μ_i is not equal (at least one diet has different weight loss)

b) Calculate the test-statistic

Recall MSW = 5.39, MSB = 7.41

$F_s = 7.41/5.39 = 1.37$, d.f{num} = 2, d.f{denom} = 19

c) Calculate the p-value

p-value = $Pr\{F > F_s\} = Pr\{F > 1.37\}$

In table with d.f{num} = 2, row 19,

1.75 is in column 0.20 $\Rightarrow Pr\{F > 1.75\} = 0.20$

Our $F_s < 1.75$, so our probability is > 0.20

p-value > 0.20

d) Interpret the p-value in terms of the problem

If in reality the true average weight loss was the same for all groups, we would observe our data or more extreme more than 20% of the time.

e) State your decision

Since p-value $> \alpha$ {.10, .05, .01}, fail to reject H_0

There is not enough evidence to suggest weight loss differs between groups.