Ex: The efficacy of the delivery system for an anti-viral medication was being assessed using the time (in days) to full recovery. Summary statistics follow:

Method	A	B	C	Overall
mean	5.4	5.3	6.3	5.67
std. dev.	2.7	2.8	2.9	2.83
n	85	88	89	262

a) Find SSTO, SSW, SSB

$SSTO = S_y^2 (n.-1) = (2.83)^2(261) = 2088.9$

$SSW = \sum_{i=1}^{I} S_i^2(n_i-1) = (2.7)^2(84) + (2.8)^2(87) + (2.9)^2(89)$

$= 2042.93$

$SSB = \sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2 = 85(5.4-5.67)^2 + 88(5.3-5.67)^2 + 89(6.3-5.67)^2$

$= 53.57$

b) Find MSW and MSB

$$MSW = \frac{SSW}{n.-I} = \frac{2042.93}{262-3} = 7.888$$

$$MSB = \frac{SSB}{I-1} = \frac{53.57}{3-1} = 26.835$$

c) Conduct a hypothesis test to see if the average recovery time is equal for all groups. State H_0, H_A, F_s, the p-value, and your decision in terms of the problem. Assume $\alpha = 0.05$

$H_0: \mu_1 = \mu_2 = \mu_3$ vs H_A: At least one μ_i is not equal.

$F_s = MSB/MSW = 26.835/7.888 = 3.402$

At d.f.{num} = 2, d.f.{denom} = 140 (rounded down)

3.06 is in column 0.05 ($Pr\{F > 3.06\} = 0.05$) and

4.02 is in column 0.02 ($Pr\{F > 4.02\} = 0.02$)

Thus, $0.02 < \text{p-val} < 0.05$

Since p-value $< \alpha$, reject H_0.

We conclude at least one of the groups has a different ave. recovery time.

* Assumptions for ANOVA

There are 4 main assumptions for ANOVA:

1) The I groups are independent
2) Random samples are taken from each of the I groups
3) The I populations are normally distributed.
4) $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_I^2$ (population variances are equal)

Note: 1, 2 we have to assume to be true.

3) we can assess through a QQ plot, or using the Shapiro-Wilks test. This tests if data is normal (H_0) or not (H_A). Can only be done in R.

For 4) we usually assume if $s_1 \approx s_2 \approx \cdots \approx s_I$, then this condition roughly holds.

* CIs for ANOVA

Once we have rejected $H_0: \mu_1 = \cdots = \mu_I$, the next logical question is "Which means are different?"

We could use many independent two sample CIs, for example lets say $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$. The possible pairs are:

$$\left.\begin{array}{lll} (\mu_1 - \mu_2) & (\mu_2 - \mu_3) & (\mu_3 - \mu_4) \\ (\mu_1 - \mu_3) & (\mu_2 - \mu_4) & \\ (\mu_1 - \mu_4) & & \end{array}\right\} \text{Six total.}$$

In general, if you have I groups, there are $I(I-1)/2$ possible pairs of $(\mu_a - \mu_b)$

But! Recall each one has an α associated with it, where $\alpha = \Pr\{\text{Type I error}\}$.

The overall error rate for the 6 CIs is defined to be:

$\Pr\{\text{At least one Type I error}\}$

We can actually calculate this.

Let k = # of CIs. For each CI, let a "success" be a Type I error occured.
Let Y = # of Type I errors out of k CIs.
$\Pr\{Y \geq 1\}$ = overall error = $1 - \Pr\{Y=0\} = 1-(1-\alpha)^k$
(using binomial distribution)

Lets say $\alpha = 0.05$. The overall error rate for k CIs is:

k	1	3	6	10	15	30
overall error	0.05	0.1426	0.2649	0.401	0.53	0.78

So, the "overall error rate" increases, even though each individual CI has an error rate of 0.05 (or α).

* Bonferroni Correction

A way to correct for T is error accumulation is to adjust α for every CI, so that the overall error rate is $\leq \alpha$.

Bonferroni suggested that if we make k CIs, make them all have individual confidence level $(1-\alpha/k)100\%$, to have an "overall" confidence of $(1-\alpha)100\%$.

In short, use α/k instead of α. Lets see what happens to "overall error rate" = $1-(1-\alpha)^k$:

$\alpha = 0.05$

k	1	3	6	10	15	30
$(1-(1-\alpha/k)^k)$	0.05	0.0492	0.0490	0.0489	0.0489	0.04881

Thus, by making k CIs at $(1-\alpha/k)100\%$ confidence level, we control the "overall error" as $\leq \alpha$.

Together, we say we are overall/simultaneous/family-wise $(1-\alpha)100\%$ confident that each true difference in means is in its respective CI, but individually they are $(1-\alpha/k)100\%$ CIs.

* Corrected CIs for k pairwise means

Because we assume $\sigma_1^2 = \cdots = \sigma_I^2$, we use a "common" variance for a CI for some $\mu_a - \mu_b$. That common variance is MSW.

An overall / simultaneous / family-wise CI for $\mu_a - \mu_b$ (out of k pairwise) is:

$$(\bar{y}_a - \bar{y}_b) \pm t_{\alpha/(2k)} \sqrt{MSW\left(\frac{1}{n_a} + \frac{1}{n_b}\right)} \qquad d.f. = n_\cdot - I$$

Notice $t_{\alpha/(2k)}$ uses the Bonferroni correction, and

$$\sqrt{MSW\left(\frac{1}{n_a} + \frac{1}{n_b}\right)} = \sqrt{\frac{MSW}{n_a} + \frac{MSW}{n_b}}$$ assumes a common variance.

Note: $t_{\alpha/(2k)}$ will be given on exams, and we use R to find it for homework (qt(1-alpha/(2*k), df = n.- I)) where you input alpha, k, n., and I

Ex: A lab manager wants to know if paying for different levels of training results in a difference for how long it takes to complete a procedure (in minutes). Summary statistics are:

	High	Medium	Low
$\bar{y}_i$	24.2	27.1	30.2
n_i	10	10	10

With MSE = 19.313, $Pr\{F > F_s\} = 0.0188$

a) Calculate CIs for all possible pairwise differences, with overall / simultaneous / family-wise level 95%.

Individually all CIs will be $(1-0.05/3)100\% = 98.333\%$.
Simultaneously they will be 95%.

$t_{0.05/(2*3)}$ = qt(1-0.05/(2*3), 30-3) = 2.552 (using R)

Low vs High: $(30.2 - 24.2) \pm (2.552)\sqrt{19.313\left(\frac{1}{10}+\frac{1}{10}\right)} \Rightarrow (0.983, 11.016)$
Low vs M: $(30.2 - 27.1) \pm (2.552)\sqrt{19.313\left(\frac{1}{10}+\frac{1}{10}\right)} \Rightarrow (-2.116, 7.916)$
M vs H: $(27.1 - 24.2) \pm (2.552)\sqrt{19.313\left(\frac{1}{10}+\frac{1}{10}\right)} \Rightarrow (-1.916, 8.116)$

b) Which CIs suggest a true difference in means?
The CI for the high vs low does not contain a zero, so that this interval suggests a difference in means. However, the other two contain 0, which means they suggest no sig. difference in means.

Note: Notice that the intervals suggest $\mu_H \neq \mu_L$, but $\mu_M \approx \mu_L$, and $\mu_M \approx \mu_H$.
This is weird! But, what happened was this:
i) The p-value was 0.0188. I.e, we would fail to reject for any α where $0.0188 > \alpha$
ii) The α we used for the CIs was $0.05/3 = 0.0166$

iii) Notice that our p-value is larger than 0.0166!

This is what is called a "border line" rejection of H_0, when p-value < 0.05 but > 0.01, and it may result in seemingly "contradictory" conclusions based on corrected CIs.

c) Interpret the interval for high vs low in terms of the problem.
We are overall/simultaneously/family-wise 95% confident that the "high" training group is faster than the "low" training group by between .983 and 11.06 minutes on average.

Note: You can also say we are $(1-\alpha/k)100\%$ confident.... which in this case is 98.33%.

Ex: Say we had two CIs at an overall/simultaneous/family-wise level of 90%: $(\mu_1-\mu_2)$ CI => (8.5, 10)
$(\mu_1-\mu_3)$ CI => (2.8, 3.2)
We are overall 90% confident that:
μ_1 is larger than μ_2 by between 8.5 and 10 on ave
and (simultaneously)
μ_1 is larger than μ_2 by between 2.8 and 3.2 on ave
I.e, both statements are true at the same time
But individually, we have $(1-\frac{0.10}{2})100\%$ confidence:
We are 95% confident that μ_1 is larger than μ_2 by between 8.5 and 10 on ave
OR (I.e NOT simultaneously).
We are 95% confident that μ_1 is larger than μ_3 by between 2.8 and 3.2 on ave

Assumptions for the CIs are the same as those for ANOVA.

Categorical Data - Proportions

Now we are switching from numeric y to categorical y. Consider a two-category Y. The goal is to estimate the prob. of a trait (success), p.

Estimating p (traditional)

Let $\hat{p} = y/n$, where y = # of subjects in the sample with the trait

You can show $\mu_{\hat{p}} = p$, and $\sigma^2_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

So that $SE(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n}$

However, a corrected value of $\hat{p}$ is used in practice, because it results in narrower confidence intervals (with the same or better confidence levels).

Wilsons-Adjusted CI for p

The Wilson-Adjusted estimate for p is:

$\tilde{p} = (y+2)/(n+4)$, with $SE(\tilde{p}) = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$

$\tilde{p}$ is slightly skewed towards 0.50 compared to $\hat{p}$, but results in better CIs for p (the true proportion).

We assume $\tilde{p}$ has a <u>normal</u> distribution, and then a a $(1-\alpha)100\%$ CI for p is:

$$\tilde{p} \pm Z_{\alpha/2}\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$$

Assumptions for this CI:

1) A random sample was taken
2) $y \geq 5$, and $(n-y) \geq 5$

Notes:

1) We can use a Z distribution for two reasons:
   - i) p is bounded between 0 and 1 (unlike μ)
   - ii) We are not estimating another parameter, so there is no "extra error"
   - iii) $Z_{\alpha/2} = t_{\alpha/2}$ at $d.f. = \infty$

Ex: Out of a random sample of 90 dogs, 74 of them were spayed/neutered ("fixed").

a) Find a 99% CI for the true proportion of dogs who are "fixed"

$\tilde{p} = (y+2)/(n+4) = (74+2)/(90+4) = 0.809$

CI: $t_{.01/2}$ at $d.f. = \infty$ $= 2.576$,

$.809 \pm (2.576)\sqrt{.809(1-.809)/(90+4)}$

$\Rightarrow (.7046, 0.9134)$

b) Interpret your CI from (a)

We are 99% confident that the true proportion of dogs who are "fixed" is between .7046 and 0.9134.

c) Does this interval suggest that the majority of dogs are "fixed"?

Yes, since both bounds are above 0.5.