Chapter 13

One-way Analysis of Variance

We know how to test fit the difference in 2 means

$$\begin{aligned} H_0 : \mu_1 = \mu_2 \qquad &\text{i.e} \qquad \mu_1 - \mu_2 = 0. \\ H_a : \mu_1 \neq \mu_2 \qquad &\text{i.e} \qquad \mu_1 - \mu_2 \neq 0 \end{aligned}$$

Now we endeavor to construct a test to detect difference in means over multiple groups.

Example: 24 expert typists test three new keyboard designs. Randomly assign 8 to each type of keyboard and assigned the same document to type up. The time of the task is recorded with the idea that a group with significantly less time on task would imply a better keyboard design

design	times
KB1	364 366 394 386 379 398 371 370
KB2	355 359 374 342 378 355 376 358
KB3	360 345 374 390 386 373 393 366

An obvious stat to look at is means

$$\bar{A} = 378.5 \qquad \bar{B} = 362.125 \qquad \bar{C} = 373.375$$

B "looks" better, but can we design a test to determine if it is "significantly" so? This is our goal.

The standard null hypothesis is that the true population means are the same for each group, $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu_0$. The alternative H_a is that at least one of the μ_i's are different.

Notation: Let i indicate the "group". E.g. k_i with $i = 1, 2, 3$. (k for keyboard). Let j denote the data point in that group. i.e. $Y_{1\,3} = 394$, $Y_{2\,5} = 378$.

Let n_i be the number of data points in the i^{th} group. Here $n_1 = n_2 = n_3 = 8$. Let n be the total of all data points. $n = \sum_{i=1}^{k} n_i$. Here $n = 24$.

Formally, we assume $Y_{i\,j} \sim N(\mu_i, \sigma^2)$ where each group has mean μ_i, but all groups have the same variance σ^2.
We need a stat.

To find "a" stat, we construct a likelihood ratio test. For H_0 we have our usual MLEs. If there is only μ_0

$$\widehat{\mu}_0 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}$$

and

$$S_0{}^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \widehat{\mu}_0)^2.$$

Under the alternative hypothesis, for the individual means, we again use the standard MLE

$$\widehat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} = \bar{Y}_i.$$

Aside: In our motivational example, $\bar{A} = \bar{Y}_1 = 378.5$, $\bar{B} = \bar{Y}_2 = 362.125$, etc.

But we are still assuming a unique σ^2 for all groups. So the MLE for

$$S_a^2 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \widehat{\mu}_i)^2.$$

Remark: Proving that this is an MLE is a good review and will be coming to a homework soon.

Construct a likelihood function:

$$\begin{aligned}
L(\text{All } Y_{ij} \mid \widehat{\mu}_0, S_0^2) &= \prod_{i=1}^{k}\prod_{j=1}^{n_i}\left(\frac{1}{2\pi S_0^2}\right)^{1/2} \exp\left(\frac{-(Y_{ij-\widehat{\mu}_0})^2}{2S_0^2}\right) \\
&= \left(\frac{1}{2\pi S_0^2}\right)^{n/2} \exp\left(-\frac{1}{2S_0^2}\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \widehat{\mu}_0)^2}_{nS_0^2}\right) \\
&= \left(\frac{1}{2\pi S_0^2}\right)^{n/2} \exp\left(-\frac{nS_0^2}{2S_0^2}\right) \\
&= \left(\frac{1}{2\pi S_0^2}\right)^{n/2} e^{-n/2}
\end{aligned}$$

Under H_a, same "algebra" occurs.

$$L(\text{All } Y_{ij} \mid \widehat{\mu}_0, S_a^2) = \left(\frac{1}{2\pi S_a^2}\right)^{n/2} e^{-n/2}$$

Then the ratio of likelihood functions yields

$$\frac{L(\text{All } Y_{ij} \mid H_0)}{L(\text{All } Y_{ij} \mid H_a)} = \frac{\left(\frac{1}{2\pi S_0^2}\right)^{n/2} e^{-n/2}}{\left(\frac{1}{2\pi S_a^2}\right)^{n/2} e^{-n/2}} = \left(\frac{S_a^2}{S_0^2}\right)^{n/2} < k$$

Implies the statistic $T = \dfrac{S_a^2}{S_0^2}$ and small values of T support the alternative hypothesis (define our RR). This makes sense because if the true means are different then S_a^2 will be the correct estimator for σ^2 and S_0^2 will be larger (on average).

We have a problem... This is an entirely new stat for us.

One thing we should show is that (a)

$$S_a^2 \text{ is an estimator of } \sigma^2$$

moreover, we need it unbiased.

Note that via multiplying by the correct degrees of freedom ν, and dividing by σ^2, the numerator and denominator are χ^2 distributed. This lead to the advent of the F-distribution.

Topic: The F-distribution. Suppose $X_1 \sim \chi^2(m)$ and $X_2 \sim X^2(n)$ and are independent. Define

$$Y_i := \frac{X_1/m}{X_2/n}$$

to be an F random variable with m numerator degrees of freedom and n denominator degrees of freedom. We usually write $Y_i \sim F(m, n)$.

The derivation is akin to the derivation of the T-stat. We used the Jacobian method of transformations. (§6.6 in the text. This is usually skipped in MTH 325.)

Letting $Y_2 = X_2$, we can find the joint density function of Y_1 and Y_2, then we integrate the joint density to get the marginal density of Y_i...

$$f(y_1) = \frac{\Gamma\left(\frac{m+n}{n}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{n/2} \frac{y_1^{m/2-1}}{\left(1 + \left(\frac{my_1}{n}\right)^{(m+n)/2}\right)} \text{ and } y_i > 0.$$

Last day: Looking at at designing a test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu_0$ versus H_a not all equal.

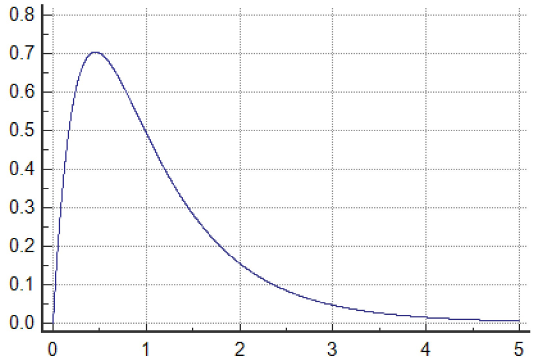
Theorem: Likelihood ration, find a new (for us) statistic that looked like a $\dfrac{\chi^2 \text{ r.v.}}{\chi^2 \text{ r.v.}}$. This lead us to the F-distribution.

$$Y \sim F(m, n) \qquad \text{when} \qquad Y = \frac{X_1/m}{X_2/n}$$

and

$$X_1 \sim \chi^2(m) \qquad \text{and} \qquad X_2 \sim \chi^2(n).$$

Graph $f(y), y > 0$:



For $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$.

$$\begin{aligned}
\mathrm{E}(Y) &= \mathrm{E}\left(\frac{X_1/m}{X_2/n}\right) \\
&= \mathrm{E}\left(\frac{n}{m} \cdot \frac{X_1}{X_2}\right) \\
&= \frac{n}{m}\,\mathrm{E}\left(X_1 \cdot \frac{1}{X_2}\right) \\
&= \frac{n}{m}\,\mathrm{E}(X_1)\,\mathrm{E}\left(\frac{1}{X_2}\right) && \text{(by independence)} \\
&= \frac{n}{m} \cdot m\,\mathrm{E}\left(\frac{1}{X_2}\right) \\
&= n\,\mathrm{E}\left(\frac{1}{X_2}\right) && \text{(numerator df has no impact)} \\
&= n \cdot \underbrace{\frac{1}{n-2}}_{\text{by thm.}}
\end{aligned}$$

$$\mathrm{E}(Y) = \frac{n}{n-2}.$$

Also, with proof omitted,

$$\mathrm{Var}(Y) = \frac{2n^2(m+n-2)}{n(n-2)^2(n-4)}.$$

End F-distribution background. Then,

$$T = \frac{{S_0}^2}{{S_a}^2} = \frac{\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \widehat{\mu}_0)^2}{\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \widehat{\mu}_i)^2}$$

and the $\frac{1}{n}$ can cancel. If S_0^2 and S_a^2 were independent, then dividing by their degrees of freedom would yield an F-distribution. Sadly, they aren't. But we can algebra to independent ratios. Start with the top,

$$
\begin{aligned}
{S_0}^2 &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \widehat{\mu}_0)^2 \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left((Y_{ij} - \widehat{\mu}_i) + (\widehat{\mu}_i - \widehat{\mu}_0)\right)^2 \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[(Y_{ij} - \widehat{\mu}_i)^2 + 2(Y_{ij-\widehat{\mu}_i})(\widehat{\mu}_i - \widehat{\mu}_0) + (\widehat{\mu}_i - \widehat{\mu}_0)^2\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \widehat{\mu}_i)^2 + 2\sum_{i=1}^{k}\Big[(\widehat{\mu}_i - \widehat{\mu}_0) \underbrace{(Y_{ij} - \widehat{\mu}_i)}_{\substack{=0 \text{ as it's the} \\ \text{sum of all} \\ \text{deviations for} \\ \text{mean within a} \\ \text{group from group mean}}} \Big] + \sum_{i=1}^{k} n_i(\widehat{\mu}_i - \widehat{\mu}_0)^2 \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \widehat{\mu}_i)^2 + \sum_{i=1}^{k} n_i(\widehat{\mu}_i - \widehat{\mu}_0)^2 \\
&= {S_a}^2 + \sum_{i=1}^{k} n_i(\widehat{\mu}_i - \widehat{\mu}_0)^2
\end{aligned}
$$

Then

$$\frac{{S_0}^2}{{S_a}^2} = 1 + \frac{\sum_{i=1}^{k} n_i(\widehat{\mu}_i - \widehat{\mu}_0)^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \widehat{\mu}_i)^2}$$

We are pretty close to independence. Recall we have Fisher's Theorem: $\chi^2 \sim N(\mu, \sigma^2)$ then $\bar{X}$ and S^2 are independent random variables and $\bar{X} \sim N(\mu, \sigma^2/n)$.

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi(1)$$

Thus the $\widehat{\mu}_i$'s up top are independent of all the variances ${S_i}^2$ in sum of the denominator. Also, $\widehat{\mu}_0$ <u>is</u> a function of $\widehat{\mu}_i$

$$\widehat{\mu}_0 = \frac{n_1\widehat{\mu}_1 + n_2\widehat{\mu}_2 + \cdots + n_k\widehat{\mu}_k}{n}$$

And the entire numerator is independent of the variance terms in the numerator when H_0 is true. Now, when H_0 is true, all the resultant variable (top and bottom) are χ^2 distributed. The last thing to do is figure out degrees of freedom. Given μ_0,

$$\frac{\widehat{\mu}_1 - \mu_0}{\sigma/\sqrt{n_i}} \sim N(0,1) \qquad \text{and} \qquad \frac{n_i(\widehat{\mu}_i - \mu_0)^2}{\sigma^2} \sim \chi^2(1)$$

So, the sum of χ^2 random variables is also a χ^2 rv and the degrees of freedom add. Thus

$$\frac{1}{\sigma^2}\sum_{i=1}^{k} n_i(\widehat{\mu}_i - \mu_0)^2 \sim \chi^2(k).$$

On the other hand, to use these facts, use the say "trick" again

$$\widehat{\mu}_1 - \mu_0 = (\widehat{\mu}_1 - \widehat{\mu}_0) + (\widehat{\mu}_0 - \mu_0)$$

And

$$\sum_{i=1}^{k} n_i(\widehat{\mu}_i - \mu_0)^2 = \underbrace{\sum_{i=1}^{k} n_i(\widehat{\mu}_i - \widehat{\mu}_0)^2}_{\text{the numerator}} + n\underbrace{(\widehat{\mu}_0 - \mu_0)^2}_{\substack{\text{divide by } \sigma^2, \\ \text{this is } \chi^2(1)}}$$

In the end

$$\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{k} n_i(\widehat{\mu}_i - \mu_0)^2}_{\chi^2(k)} = \underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{k} n_i(\widehat{\mu}_i - \widehat{\mu}_0)^2}_{\substack{\text{the numerator is} \\ \chi^2(k-1)}} + \underbrace{n\frac{(\widehat{\mu}_0 - \mu_0)^2}{\sigma^2}}_{\chi^2(1)}$$

Denominator is much easier,

$$\sum_{i=1}^{k} \underbrace{\sum_{j=1}^{n_i} \frac{(Y_{ij} - \mu_i)^2}{\sigma^2}}_{S_i^2 \sim \chi^2(n_i - 1)}$$

and by independences, summing all the $\chi^2(n_i - 1)$ variables yield a

$$\underbrace{\chi^2\left(\sum_{i=1}^{k}(n_i - 1)\right)}_{\text{add all df}} = \chi^2(n - k)$$

This is a ton of work, but in the end we have

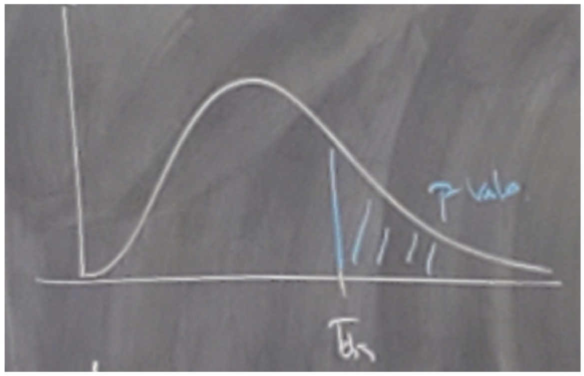
$$F = \frac{\sum_{i=1}^{k} n_i(\widehat{\mu}_i - \mu_0)^2 \Big/ k - 1}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \widehat{\mu}_i)^2 \Big/ n - k} \sim F(k-1,\, n-k)$$

when H_0 is true.

Topic: The ANOVA test.

1. Compute the null and alternative estimate of all the sample means.
2. Use these to compute the sum of squares in our F-statistic
3. Construct a table (R or Excel?) to compute the p-value.
   F_{obs} is F observed. We computed this.

$$\Pr\left(F > F_{\text{obs}}\right), \; F - F(k-1, n-k)$$



④ Conclusion. There is standard language for all this. In the denominator,

$$\mathrm{SSW} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{i\,j} - \widehat{\mu}_i)^2$$

the sum of squares <u>WITHIN</u> a group. And the numerator

$$\mathrm{SSA} = \sum_{i=1}^{k} n_i(\widehat{\mu}_i - \mu_0)^2$$

the sum of the squares <u>AMONG</u> groups. A measure of variance among the sample means. Recall originally,

$$\frac{{S_0}^2}{{S_a}^2} = \frac{{S_a}^2 + \mathrm{SSA}}{\mathrm{SSW}}$$

Thus we showed the factorization yield <u>total variation</u>.

$$S_0^2 = \mathrm{SSW} + \mathrm{SSA} \qquad \text{or} \qquad \mathrm{SST} = \mathrm{SSW} + \mathrm{SSA}$$

ANOVA Test

$$\mathrm{SSA} = \sum_{i=1}^{k} n_i(\widehat{\mu}_1 - \widehat{\mu}_0)^2 \qquad \text{(SS among groups)}$$

$$\mathrm{SSW} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{i\,j} - \widehat{\mu}_i)^2 \qquad \text{SS within groups}$$

$${S_0}^2 = \mathrm{SST} = \mathrm{SSA} + \mathrm{SSW}$$

$$R^2 = \frac{\mathrm{SSA}}{\mathrm{SST}} \quad \text{coef. of determination}$$

Discussion: The ANOVA table

Source	SS	df	Mean Squares	F_{obs}	p-value
among	SSA	$k-1$	SSA $/(k-1)$	MSA / MSW	$\Pr\left(F(k-1, n-k) > F_{\text{obs}}\right)$
within	SSW	$n-k$	SSW $/(n-k)$		
total	SST	$n-1$			

Example:

Treatment	Data		
1	2	4	3
2	6	4	
3	3	5	4

Mean squares $k = 3, n_1 = 3, n_2 = 2, n_3 = 3, n = 8, \widehat{\mu}_1 = 3, \widehat{\mu}_2 = 5, \widehat{\mu}_3 = 4$.

$$\widehat{\mu}_0 = \frac{\sum\sum Y_{ij}}{n} = \frac{n_1\widehat{\mu}_1 + n_2\widehat{\mu}_2 + n_3\widehat{\mu}_3}{n} = \frac{9+10+12}{8} = 3.875.$$

$$\begin{aligned}
\text{SSA} &= \sum_{i=1}^{k} n_i(\widehat{\mu}_1 - \widehat{\mu}_0)^2 \\
&= 3(3-3.875)^2 + 2(5-3.875)^2 + 3(4-3.875)^2 \\
&= 4.875. \\
\text{SSW} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \widehat{\mu}_i)^2 \\
&= \underbrace{(2-3)^2 + (4-3)^2 + (3-3)^2}_{i=1} + \underbrace{(6-5)^2 + (4-5)^2}_{i=2} + \underbrace{(3-4)^2 + (5-4)^2 + (4-4)^2}_{i=3} \\
&= 6 \\
\text{SST} &= \text{SSW} + \text{SSA} = 10.875.
\end{aligned}$$

Note that $R^2 = \dfrac{\text{SSA}}{\text{SST}} \approx 0.326$. Make table:

Source	SS	df	Mean Squares	F_{obs}	p-value
among	4.875	2	2.4375	2.03125	0.2261023
within	6	5	1.2		
total	10.875	7			

$$\begin{aligned}
p &= \Pr\left(F(2,5) > F_{\text{obs}} = 2.03125\right) \\
&= \texttt{pf(Fobs, k-1, n-k, lower.tail=FALSE)} \qquad \text{(R code)} \\
&= 0.2261023.
\end{aligned}$$

Hence a large p-value does not support the alternative. No evidence that the means are different.

Example: Recall the typists/keyboards designs:

design	times
KB1	364 366 394 386 379 398 371 370
KB2	355 359 374 342 378 355 376 358
KB3	360 345 374 390 386 373 393 366

Then $\widehat{\mu}_1 = 378.5, \widehat{\mu}_2 = 362.125, \widehat{\mu}_3, 373.375$. $k = 3, n_1 = n_2 = n_3 = 8, n = 24$.

R-code, with little c for column. y = c (364, 355, 360, 366, 359, 345, ..., 370, 358, 366). typists = rep (1:3, 8) anova (lm (y factor(typists))). R spits out:

	df	SS	Mean Squares	F_{obs}	p-value
factor(typists)	2	532.0	266.00	1.5181	0.2422
residuals	21	3679.6	175.22		

Conclusion, again, fairly large p value. Cannot reject H_0. Practical conclusion: No evidence that our keyboard design is different in use that another. Remark: Missing from R's table is $\text{SST} = \text{SSW} + \text{SSA} = 4211.6$ and $R^2 = \text{SSA} / \text{SST} = 0.126$.

Example: Background sounds and impact on memory. 30 students are randomly divided into 3 sets of 10. Silence, classical, and jazz. They are told to "study". The "data ish"

	Quiet	Classical	Jazz
$\widehat{\mu}_i$ sample means	89.5	89.7	79.4
S_i sample standard dev.	8.91	10.25	7.06

Need $k = 3$ treatments. $n_1 = n_2 = n_3 = 10$ and $n = 30$. Then

$$\widehat{\mu}_0 = \frac{n_1\widehat{\mu}_1 + n_2\widehat{\mu}_2 + \cdots + n_k\widehat{\mu}_k}{n} = 86.2$$

$$\begin{aligned} \text{SSA} &= \sum_{i=1}^{k} n_i(\widehat{\mu}_1 - \widehat{\mu}_0)^2 \\ &= 10(89.5 - 86.2)^2 + 10(89.7 - 86.2)^2 + 10(79.4 - 86.2)^2 \\ &= 680.3 \end{aligned}$$

Note we have $S_i^2 = \dfrac{\sum_j (Y_{i\,j} - \widehat{\mu}_i)^2}{n_i - 1}$.

$$\begin{aligned} \text{SSW} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{i\,j} - \widehat{\mu}_i)^2 \\ &= \sum_{j=1}^{10}(Y_{1\,j} - \widehat{\mu}_1)^2 + \sum_{j=1}^{10}(Y_{2\,j} - \widehat{\mu}_2)^2 + \sum_{j=1}^{10}(Y_{3\,j} - \widehat{\mu}_3)^2 \\ &= 9\left(\frac{1}{9}\sum_{j=1}^{10}(Y_{1\,j} - \widehat{\mu}_1)^2 + \frac{1}{9}\sum_{j=1}^{10}(Y_{2\,j} - \widehat{\mu}_2)^2 + \frac{1}{9}\sum_{j=1}^{10}(Y_{3\,j} - \widehat{\mu}_3)^2\right) \end{aligned}$$

So $\text{SSW} = 9(S_1^2 + S_2^2 + S_3^2)$. Fact, $\text{SSW} = \sum_{i=1}^{k}(n_i - 1){S_i}^2$. Here, $\text{SSW} = 2108.65$.

Source	SS	df	Mean Squares	F_{obs}	p-value
among	680.3	2	340.15	4.3598	0.022863
within	2108.65	27	78.09815		
total	2788.95	7			

Example: Background sounds (again)

	Quiet	Classical	Jazz
$\widehat{\mu}_i$ sample means	89.5	89.7	79.4
S_i sample standard dev.	8.91	10.25	7.06

$n_1 = n_2 = n_3 = 10$ and $n = 30$ and $k = 3$. Then

$$\widehat{\mu}_0 = \frac{\sum n_i \widehat{\mu}_i}{n} = 86.2$$

$$\text{SSA} = \sum_{i=1}^{3} n_i \left(\widehat{\mu}_1 - \widehat{\mu}_0\right)^2 = 680.3$$

$$\text{SSW} = \sum_{i=1}^{3}\sum_{j=1}^{10} (Y_{i\,j - \widehat{\mu}_i})^2 = \sum_{i=1}^{3} 0.95_i^2 = 2108.65$$

FACT: $\text{SSW} = \sum_{i=1}^{k} (n_i - 1) {S_i}^2$. Recall:

Source	SS	df	Mean Squares	F_{obs}	p-value
among	680.3	2	340.15	4.3598	0.022863
within	2108.65	27	78.09815		
total	2788.95				

Where p-value is $\Pr\left(F(2,27) > 4.3598\right) \approx 2.28\%$.

At $\alpha = 0.05$ level, we cannot reject H_0. Conclusion: "At least one of the group means is different than the others."

Topic: Confidence interval (§ 13.7)
We may be interested in a CI for the group mean. We clearly have an estimate for μ_i and $\widehat{\mu}_i$.

$$\text{C.I.} \equiv \widehat{\theta} \pm t^* \sqrt{\text{Var}\left(\widehat{\theta}\right)}.$$

Big question: What do we use for variance?

We have estimators ${S_0}^2, {S_a}^2$ which are MLE.
Issue: As with the original $S^1 = \frac{1}{n}\sum (X_i - \bar{X})^2$, our MLEs are biased.

H is true if H_0 is true. Both alone are "good" estimators. If H_a is true, it stands to reason that ${S_a}^2$ is "better". Using a bit of algebra, we can make an unbiased estimator of ${S_a}^2$.

Last day: we showed

$$\text{SSW} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (Y_{i\,j} - \widehat{\mu}_i)^2 = \sum_{i=1}^{k} (n_i - 1) {S_i}^2.$$

This is a weighted sum of unbiased estimators ${S_i}^2$ for σ^2. As individually, each ${S_i}^2 \sim \sigma^2$ (unbiased). Then the collective average should be an even better approximation.

$$\sigma^2 = \frac{\sum_{i=1}^{k}(n_i - 1){S_i}^2}{(n_1 - 1) + \cdots + (n_k - 1)} = \frac{\sum_{i=1}^{k}(n_i - 1){S_i}^2}{n - k} = \text{MSW}$$

FACT: We can use the MSW as an unbiased estimator for σ^2.
Aside: Isn't this just the pooled estimator written largely?

$${S_p}^2 = \frac{(n_1 - 1){S_1}^2 + (n_2 - 1){S_2}^2}{n_1 + n_2 + k}$$

Here

$$S_{\cdot}^2 = \frac{(n_1 - 1){S_1}^2 + \cdots + (n_k - 1){S_k}^2}{n_1 + n_2 + \cdots + n_k - k}$$

We are constructing hypothesis of different of CI. The "t" inherits the df from S^2. For our CI we need df $= n - k$.

Example: Background sounds (again)

a.) Find a 95% CI for the mean score for someone listening to Jazz.

Here, $n_3 = 10$ and $t_{\alpha/2}(\text{df}) = t_{0.025}(27) \approx 2.052$.
Then, $\widehat{\mu}_3 \pm t_{0.025}(27)\sqrt{S^2/n_3}$ with S^2/n being the total variance affiliated with samling dist.

We have $79.4 \pm 2.052\sqrt{\dfrac{78.09815}{10}}$ which is

$$79.4 \pm 5.734 \qquad \text{or} \qquad (73.665,\ 85.134)$$

Note $\widehat{\mu}_1 = 89.5$ and $\widehat{\mu}_2 = 89.7$ this is still a couple standard errors away.

b.) Is the proof of the other 2 μ_i's are different? Note: We could ANOVA again without Jazz, or do chapter 9 stuff:

Doing Ch. 9 stuff, consider $\widehat{\mu}_1 - \widehat{\mu}_2 = -0.2$. Then to make a CI we use the exact same formula from before (Ch9) but we use the "ANOVA" estimator for $S^2 = \text{MSE}$ with df $= n - k$.

$$\begin{aligned}
\widehat{\mu}_1 - \widehat{\mu}_2 \pm t_{0.025}(27)\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}} &= \widehat{\mu}_1 - \widehat{\mu}_2 \pm t_{0.025}(27)S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\
&= -0.2 \pm 20.052\sqrt{\frac{75.09815}{5}} \\
&= -0.2 \pm 7.95 \\
&\equiv (-8.15,\ 7.75)
\end{aligned}$$

And notice that $0 \in$ CI.

Summary: 2-sided $1-\alpha$ level CI.

$$\widehat{\mu}_i \pm t_{\alpha/2}(\mathrm{df})\frac{S}{\sqrt{n_i}} = \widehat{\mu}_i - \widehat{\mu}_j \pm t_{\alpha/2}(\mathrm{df})S\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$$

Where $S^2 = \mathrm{MSW} = \dfrac{\mathrm{SSW}}{n-k}$ with $\mathrm{df} = n-k$.

Example: Arbitrary numbers (no story)

A	80	85	71	64
B	70	72	75	
C	83	70		

Then $n_1 = 4$, $n_2 = 3$, $n_3 = 2$, $n = 9$, $\bar{A} = 75$, $\bar{B} = 217/3$, $\bar{C} = 153/2$ and

$$\widehat{\mu}_0 = \frac{4\bar{A}+3\bar{B}+2\bar{C}}{9} = \frac{690}{9}.$$

	SS	df	MS	F_{obs}
among	SSA $= 23.0556$	2	11.5278	0.192575
within	SSW $= 359.167$	6	39.8611	

The above was computed on mathematica. Using R for the p-value, $p \approx 82.97\%$. We cannot reject H_0.

The R code:

```
scores = c(80, 85, 71, 64, 70, 72, 75, 83, 70)
treat = c(rep("A", 4), rep("B", 3), rep("C", 2))
table = data.frame(saved, treat)
results = aov(scores ~ treat, data = table)
summary(results)
```

The output is

	df	Sum Square	Mean Square	F	Pr $(> F)$
treat	2	23.1	11.53	0.193	0.83
residual	6	359.2	59.86		