Multiple (more than 2) groups: Analysis of Variance (ANOVA):
Chapter 5: two basic concepts, Chapter 6: "after the F test" analyses

Treatments often (usually) chosen to answer specific questions
Diet and longevity study (case study 5.1): 6 treatments,
5 questions, each a difference between two means (shown in Display 5.3)

Two new statistical concepts in Chapter 5:
Pooling variability from more than 2 groups
Comparing models $\rightarrow$ the F test and Analysis of Variance (ANOVA)

Pooling variability:
Assume population variance same for all groups
treatments may change means; do not change variances
sample variances of each group, s_i^2, all estimate the common population variance, σ^2
Pool into a single estimate
weighted average of variances, using df as the weights
just like pooling two groups, but now have k groups

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

pooled df = sum of sample-specific df
or: $N-k$ where N is total # obs, k is # of groups

Comparison of specific pairs of means, eg $\overline{Y}_a$ and $\overline{Y}_b$:
Only change (and a small one) is to the df

$$\text{se } = s_p\sqrt{\frac{1}{n_a}+\frac{1}{n_b}}$$

Note: s_p based on all N obs, has df based on all N obs
"other part" of se based only on sample sizes for the two groups being compared
Estimates, T statistic for test, confidence intervals same as with 2 groups

F test: answers a more general question
Do all groups have the same population mean? (the null)
if not, at least one group has a different mean
Note: **not** all means are different from each other
most groups may have identical means, with just one group different This is a model comparison!

Models for k groups:
Same idea as two-group model comparison, just with more groups
model I: all groups have population mean μ
model II: each group has its own population mean μ_i
Not saying $\mu_1 \neq \mu_2$. Just saying they aren't forced to be the same

How well does a model fit the data?: Sum of squared errors = SSE
also known as sum of squared residuals
Residuals depend on predicted values, which depend on the model.
$\hat{Y}_{ij}$ is the predicted value for observation Y_{ij}
SSE = $\Sigma_{ij}(Y_{ij} - \hat{Y}_{ij})^2$
When errors normally distributed with equal variance,
SSE is the best measure of fit of a model to data
compute SSE for the two models, compare them in a specific way
Why not use chi^2?
Because we are back to normally distributed observations

Mean Square Error: MSE and root Mean Square Error: rMSE
Connection between the fit of model II and pooled variance
$s_p^2 = \frac{\text{SSE}}{\text{df}}$, when SSE computed using model II predicted values
s_p^2 also called MSE (see why a bit later)
rMSE = $\sqrt{\text{MSE}}$, so rMSE for model II is the pooled sd, s_p
s_p^2 and s_p more commonly used names when 2 groups
MSE and rMSE more commonly used names when more than 2 groups

Full and reduced models
Model I usually called "reduced model"
reduced model always describes the null hypothesis
Model II usually called "full model"

F test:
logic: (same as logic for a Chi-square test, just different measures of fit)
If Ho is true, model II will fit almost as well as Model I
If Ho is false, model II will fit substantially better than Model I
What is "almost as well" and what is "substantially worse"?
Need to look at change in SSE, change in error df,
and compare to the MSE from model II

$$
\begin{aligned}
F &= \frac{\text{Difference in SSE/Difference in error df}}{s_p^2} \\
&= \frac{(SSE_{reduced} - SSE_{full})/(df_{reduced} - df_{full})}{SSE_{full}/df_{full}}
\end{aligned}
$$

When model assumptions appropriate, F has a known theoretical distribution
called the F distribution
Many F distributions: depend on:
numerator df = difference in error df
denominator df = df for MSE
Values larger than 4 are usually associated with $p < 0.05$, unless small df

ANOVA table:

A way to organize the computations.
Start with numbers from the two models

Source	df	SS
Between groups	$\text{df}_{reduced}$ - df_{full}	$\text{SSE}_{reduced}$ - SSE_{full}
Within groups (error, residual)	df_{full}	SSE_{full}
Corrected total	$\text{df}_{reduced}$	$\text{SSE}_{reduced}$

Then compute the last two columns, using the relationships:

$$\text{MS} = \frac{\text{SS}}{\text{df}}, \text{ for any row}$$
$$F = \frac{\text{MS}_{between}}{\text{MS}_{within}}$$

Source	df	SS	MS	F
Between groups	$\text{df}_{between}$	$\text{SSE}_{between}$	$\text{SS}_{between}$ / $\text{df}_{between}$	$\text{MS}_{between}/\text{MS}_{within}$
Within groups (error, residual)	df_{full}	SSE_{full}	SS_{within} / df_{within}	
Corrected total	$\text{df}_{reduced}$	$\text{SSE}_{reduced}$		

What do you learn from the "usual" F test:
If accept Ho, no evidence of any difference in means
If reject Ho, at least one group has a different mean
follow up, "after the ANOVA" methods is Chapter 6

What is the difference between targeted questions and the "usual" F test?
Targeted questions are comparisons of specific groups
more likely to find a difference when that effect is present
usual F test looks for any difference
i.e., you aren't sure which effects to look at
less likely to find a specific difference