Stat 414 – ANOVA Handout

Notation: Assume a *balanced, one-way ANOVA* study design

- y_{ij} is the i^{th} observation of our response variable in group *j*
- n_j is the number of observations in the j^{th} group
- J is the number of groups

$i = 1 .. n_j = n$
$N = nJ$

Complete the SS and DF columns in the table below using this notation

Source	SS	Df	Mean Square	Interpretation
Groups	$n\Sigma(\bar{y}_{\cdot j} - \bar{y})^2$	$J - 1$	$SSG/(J-1)$	variability among groups
Error	$\Sigma\Sigma(y_{ij} - \bar{y}_{\cdot j})^2$	$N - J$	$SSE/(N-J)$	within group variability aka random noise
Total	$\Sigma\Sigma(y_{ij} - \bar{y})^2$	$N - 1$		variance of Y

SS total = SS groups + SS error

(a) How would you explain the " – 1" parts of the degrees of freedom values?

because of the reference group, only J-1 or N-1 "independent" pieces of information in the sum

(b) How is R^2 calculated and what does it tell us?

$R^2 = \frac{SS\ groups}{SS\ total}$ in this dataset $\frac{\sigma_g^2}{\sigma_e^2}$

aka "eta-squared"

(c) How is the *F*-ratio calculated and what does it tell us?

$F = \frac{MS\ groups}{MS\ error}$ MS groups is F times higher than variation w/in groups

Reminders: Let $S = n\Sigma(\mu_j - \mu)^2$ represents the "true variability" among the groups on the SS scale (like aggregated across all the observations)

$E(SSgroups) = (J-1)\sigma^2 + S;$

$E(MSgroups) = \sigma^2 + S/(J-1) = \sigma^2 + n\sigma_g^2$ where $\sigma_g^2 = 1/\{J-1)\Sigma(\mu_j - \mu)^2$

$E(MSerror) = \sigma^2;$

$E(SStotal) = (N-1)\sigma^2 + S$

So (N-1)sigma^2 + S^2 + sigma^2 will estimate N(sigma^2) + S on the SS scale

So the SSgroups includes both variation between groups and "random noise" from the sampling process (e.g., the group means varying from sample to sample)

Turns out, R^2 is a biased estimator of the between group variation / total variation in the population. Another popular statistic corrects for that bias. In particular, SSgroups is expected to be larger than zero even when there are no genuine group differences, due to the random noise in the data. So we will adjust the numerator by subtracting off the expected noise across the *J* groups: $(J-1)MSError$. So the denominator better estimates the expected total variance (think $S + N\sigma^2$), we will add one more $MSError$.

between group vs. within group (times the number of observations in the data set)

Definition: "Omega-squared" (ω^2) is an (approximately) unbiased estimator of $\frac{S}{S+N\sigma^2}$ for the specific groups in the study:

only between group

$$\widehat{\omega}^2 = \frac{SSgroups - (J-1)MSError}{SStotal \ + MSerror}$$

total variability (inclues random noise)

$0 \le \hat{\omega}^2 \le 1$

This statistic is usually smaller than R^2, software can even create confidence intervals.

Alternatively (see Section 3.3)

Instead of calculating $MSgroups$, let's literally calculate the variance of the group means.

Variance of group means = $S^2_{groups} = \frac{1}{J-1}\Sigma(\bar{y}_i - \bar{y})^2$ = *MSgroups/n* (*n* is common group size)

But this again mixes together the true variation in the group means and the sample to sample variance in the sample means: σ^2/n.

So to adjust, we subtract the sampling variability: MSgroups/n - sigma^2/n = (MSG-MSE)/n

Then the denominator will add the between group variation to the within group variation

denominator = (MSG-MSE)/n + MSE = (MSG/n + (n-1)/n MSE) = 1/n(MSG + (n-1)MSE)

"only between group"

"within group" so sum is a measure of "total variation"

Definition: Another estimator of the between/total variance ratio:

$$ANOVA\ ICC = \frac{(MSgroups - MSerror)}{(MSgroups) + (n-1)MSerror} = \frac{F-1}{F+n-1}$$

which doesn't depend on the number of groups. Consider it a rescaled effect size. Later we will see how this can be used in some cases to estimate the amount of "agreement" between observations in the same group. (Is also some discussion in Ch. 3 of your text. This formula actually matches the formula it gives, e.g., equation 3.12.) You can also show that $ICC = J\omega^2/((J-1) + \omega^2)$, so that $ICC \ge \omega^2 = (F-1)/(F+n-1+\frac{n}{J-1})$.

Bottom line, these are two different "effect size" measures but this second one will become very important to us later...