Comparing Several Means (One-way ANOVA) ①

Recap: Last week we discussed comparing 2 means using t-tests

what if we want to compare multiple means?

$\Rightarrow$ Queue in the <u>An</u>alysis <u>o</u>f <u>Va</u>riance (ANOVA)

Suppose we have multiple groups and we want to compare the means (μ) between groups.

↳ this is very similar to the analysis using t-tests, but we <u>must account for variability from different sources</u>

In this case our null and alternative hypothesis are:
(see example I)

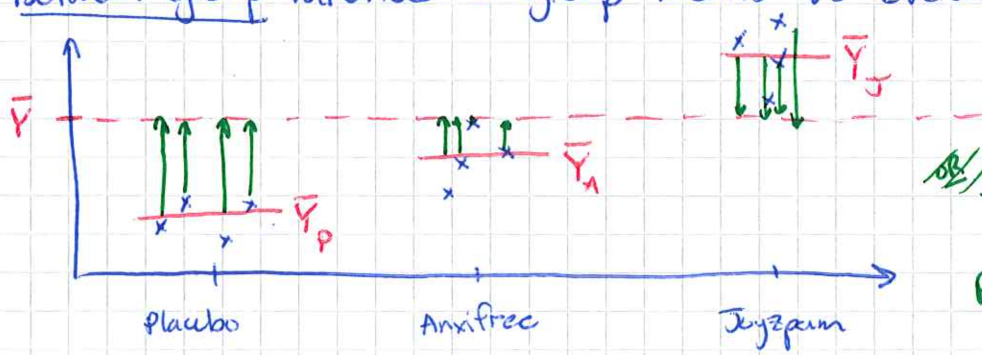
H_0: it is true that $\mu_P = \mu_A = \mu_J$

H_A: it is not true that $\mu_P = \mu_A = \mu_J$

where
μ_P = population mean for placebo group
μ_A = pop. mean for Anxifree group
μ_J = pop. mean for Joyzepam group

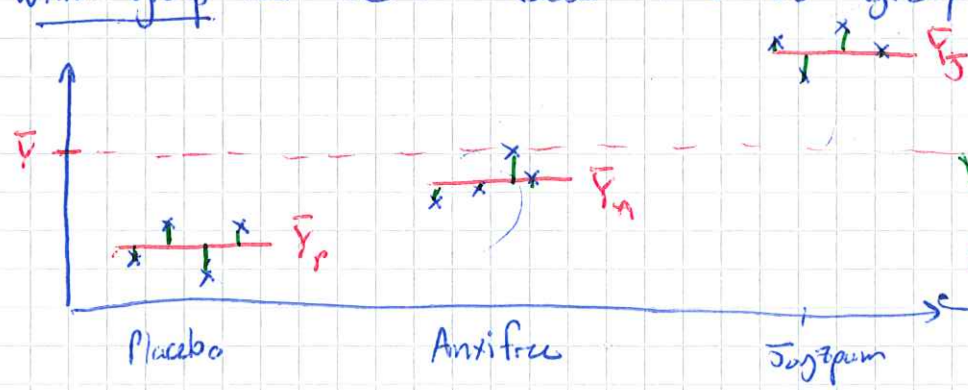
Where does variability come from?

<u>Between group</u> variance → group means vs overall mean



Residual: $\bar{Y}_k - \bar{Y}$

<u>Within group</u> variance → observations vs group mean



Residuals: $Y_{ik} - \bar{Y}_k$

Total Variability → observations vs overall mean

Residuals: $Y_{ik} - \bar{Y}$

Variability is

$$Var(Y) = \frac{1}{N} \sum_{k=1}^{G} \sum_{i=1}^{N_k} (Y_{ik} - \bar{Y})^2$$

i.e. weighted sum of residuals squares.

We want to compare within group variability to between group variability.

⇒ Measure using Sums of Squares (SS)
↳ this is very similar to variance.

$$SS(\text{Total}) = SS(\text{Between}) + SS(\text{Within})$$

$$\underset{SS_B}{SS(\text{Between})} = \sum_{k=1}^{G} \sum_{i=1}^{N_k} (\bar{Y}_k - \bar{Y})^2$$

$$\underset{SS_W}{SS(\text{Within})} = \sum_{k=1}^{G} \sum_{i=1}^{N_k} (Y_{ik} - \bar{Y}_k)^2$$

$$\underset{SS_T}{SS(\text{Total})} = \sum_{k=1}^{G} \sum_{i=1}^{N_k} (Y_{ik} - \bar{Y})^2$$

Notation

G = number of Groups

N_k = # observations in group k

k = group number

i = observation number

$\bar{Y}$ = total average

$\bar{Y}_k$ = ~~group~~ average of group k

Y_{ik} = variable value of observation i in group k

Our overall variation is the sum of variation due to "differences in sample means for the different groups" plus "all the rest of variation"

→ If H_0 is true, then SS_b would be small (relative to SS_w)

→ If H_A is true, then SS_b would be very high

How do we determine what is small vs ~~large~~ large/high?

↳ The F-test

The F-test

→ we need to know the degrees of freedom associated with SS_b and SS_w

⇒ d.f. corresponds to # of unique "data points" minus the number of "constraints" that they need to satisfy.

Within Group: N is the # of unique data points around the group means (G constraints)

Between Groups: we are looking at our group means (G data points) around the grand mean (1 constraint)

Thus

$$d.f._b = G - 1$$

$$df_w = N - G$$

This brings us to our "Mean Squares (MS)" values

$$MS_b = \frac{SS_b}{df_b} \quad ; \quad MS_w = \frac{SS_w}{df_w}$$

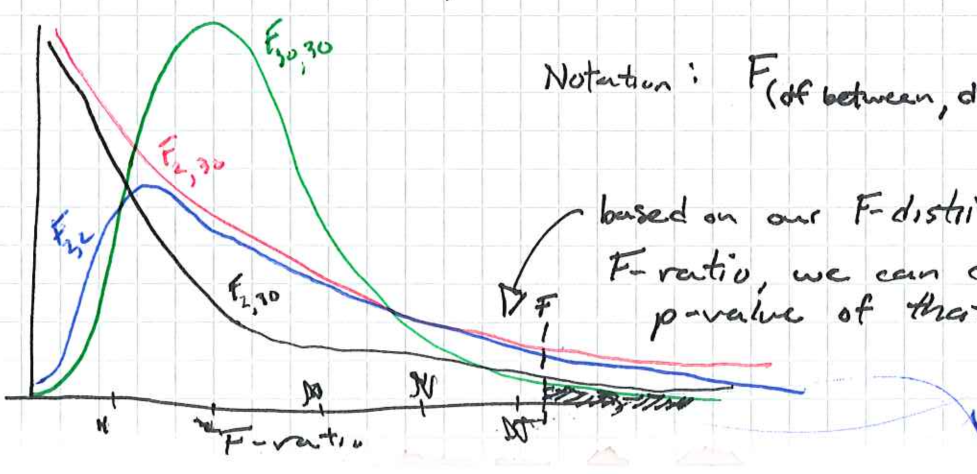
These are used to calculate the F-ratio

$$F = \frac{MS_b}{MS_w}$$

→ this is also the pooled estimate of variance s_p^2 (weighted avg. of group variances weighted by the group's deg of freedom)

↳ intuitively a bigger F-ratio means between-group variation is large relative to within-group variations (i.e. common variance)

↳ similar to t-tests' t-distributions, there is an F-distribution of F-ratios for if the H_0 were true. It is dependent on the d.f. between ; within groups



Notation: $F_{(\text{df between},\ \text{df within})}$

based on our F-distribution and F-ratio, we can determine the p-value of that result

The ANOVA Table

→ The key quantities involved in an ANOVA are usually organized into a standard table

	d.f.	Sum of Squares	mean squares	F-Stat	p-value
between groups	$df_b = G-1$	$SS_b = \sum_{k=1}^{G} N_k (\bar{Y}_k - \bar{Y})^2$	$MS_b = \frac{SS_b}{df_b}$	$F = \frac{MS_b}{MS_w}$	Naw too hard...
within groups	$df_w = N-G$	$SS_w = \sum_{k=1}^{G} \sum_{i=1}^{N_k} (Y_{ik} - \bar{Y}_k)^2$	$MS_w = \frac{SS_w}{df_w}$		
Total	$N-1$	$SS_{tot} = \sum_{k=1}^{G} \sum_{i=1}^{N_k} (Y_{ik} - \bar{Y})^2$			

→ this model tells us if one or more groups are different from the others, **BUT** it does not tell us which one is different

The null hypothesis claims that all 3 of the following are true:

$\mu_A = \mu_P$ & $\mu_J = \mu_P$ & $\mu_J = \mu_A$

→ we now need to do post-hoc testing to prove our H_A of ~~A~~ Joyzepam performing better than the others

we could run ~~4~~ pairwise t-tests to compare the following

μ_A vs μ_P $\quad$ μ_J vs μ_P $\quad$ μ_J vs μ_A

But say we have 10 groups ... that's 45 t-tests!!

what problem will we encounter?

Each individual t-test is designed to have 5% Type I error rate with 45 t-tests you'd expect 2-3 to return a "significant" ~~find~~ difference by chance alone.

=> usually we use adjustments to the p-value i.e. a correction of multiple comparisons

Methods For Correction for Multiple Comparisons

Bonferroni Corrections $\Rightarrow$ $p' = m \times p$ where m is the number of tests

Holm corrections

↳ we imagine we're doing the tests sequentially, starting from the smallest p-value to the largest
↳ we then adjust the p-value by either:

$$p'_j = j \times p_j \quad \text{OR} \quad p'_j = p'_{j+1}$$ – and take the largest.

Example:

raw p	rank j	$p \times j$	Holm p
0.001	5	0.005	0.005
0.005	4	0.020	0.020
0.019	3	0.057	0.057
0.022	2	0.044	0.057
0.103	1	0.103	0.103

Assumptions of One-way ANOVA

1) The populations have normal distributions – histogram or qq plot
   ↳ we look at residuals to check this (Sharpiro-Wilk test)
2) the populations standard deviation (or variance) are all the same (rarely true – we should check for homogeneity)
   ↳ Levene test or Brown-Forsythe test
3) Observations are independent (within groups and between groups)
   ↳ i.e. repeated measures violates this and needs different model)

Finally: An ANOVA with only 2 groups is identical to the Student t-test