

Analysis of Variance

Psychology 3256

Introduction

- We have t and z tests to deal with differences with one or two groups
- what if we have more than two groups?

an example

	A1	A2	A3
	85	67	52
	90	80	60
	77	75	65
\bar{x}	84	74	59

Why do the scores vary?

- Or, what are the sources of variation
- Well individual difference
- and of course group differences
-

I never said there'd be no math..

- any score = being human + group differences + individual difference
- $x = \mu + \tau + \epsilon$

The structural model of ANOVA

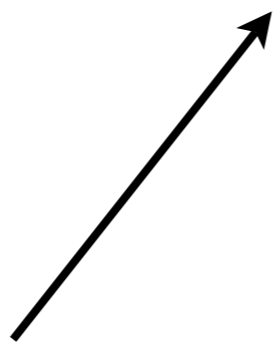
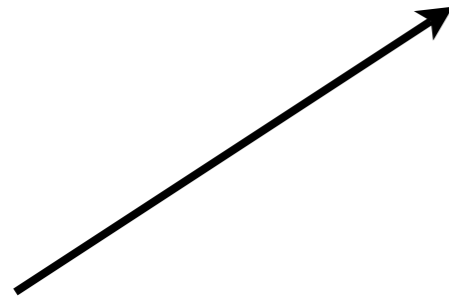
$$x = \mu + \tau + \varepsilon$$

any score

grand mean

treatment effect

error



Let's make an assumption

$$\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

$$\therefore \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_k^2$$

H₀ true

- This is the null hypothesis assumption

More assumptions

- Scores are randomly and normally distributed around the grand mean
- independent observations
- all sources of variation are in the model

Let's look at variance

$$\sigma_1^2 \approx s_1^2$$

$$\sigma_2^2 \approx s_2^2$$

etc

$$\sigma_\varepsilon^2 \approx s^2 = \frac{1}{k} \sum s_j^2$$

Remember, by the CLT

$$\text{var}(\bar{x}) = \frac{\sigma^2}{n}$$

$$\therefore s_x^2 = \frac{\sigma_\varepsilon^2}{n}$$

$$s_x^2 n = \sigma_\varepsilon^2$$

We now have two
estimates of σ_{ε}^2

$$s_x^2 n = \sum \frac{s_j^2}{k} = \sigma_{\varepsilon}^2$$

MS_{Treat}

MS_{Error}

So, if H_0 is true..

- $E(\text{MSE}) = \sigma_{\varepsilon}^2$
- $E(\text{MST}) = \sigma_{\varepsilon}^2$

If H_0 is not true

- $E(\text{MSE}) = \sigma_{\varepsilon}^2$
- $E(\text{MST}) = \sigma_{\varepsilon}^2 + n\sigma_{\tau}^2$
- $E(\text{MSE}) < E(\text{MST})$

SO...

- If we were to divide MST by MSE (MSE/MST) we would have some estimate of how much extra variation MST is measuring
- i.e. F
- This is precisely what is done in ANOVA

The F word

- $F = MST / MSE$
- $E(F|H_0 \text{ true}) ?$
- $E(F|H_a \text{ true}) ?$
- If H_0 is true, then MST/MSE will be distributed as $F(df_t, df_e)$
- if not it will be distributed some other way

Partitioning SS and df

- $SS_{\text{Total}} = SS_{\text{Treatment}} + SS_{\text{Error}}$
- $df_{\text{Total}} = df_{\text{Treatment}} + df_{\text{Error}}$

More Precisely..

$$\frac{\sum (x - \bar{x}_g)^2}{N - 1} = n \frac{\sum (\bar{x}_j - \bar{x}_g)^2}{k - 1} + \frac{\sum \sum (x - \bar{x}_j)^2}{N - k}$$

ANOVA Summary Table

Source of Variation	df	MS	F
Between Groups	$k-1$	$SSBG/k-1$	$MSBG/MSWG$
Within Groups	$N-k$	$SSWG/N-k$	
TOTAL	$N-1$		