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

The structural model of ANOVA

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

any score

grand mean

treatment effect

error

Let's make an assumption

$$\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$$
$$\therefore \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots = \sigma_k^2$$
$$H_0 true$$

- 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

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

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

$$etc$$

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

Remember, by the CLT

$$\mathrm{var}(\bar{x}) = \sigma^2 / n$$

$$\therefore s_{\bar{x}}^2 = \sigma_\varepsilon^2 / n$$

$$s_{\bar{x}}^2 n = \sigma_\varepsilon^2$$

We now have two estimates of σ_ε^2

$$s_{\bar{x}}^2 n = \sum \frac{s_j^2}{k} = \sigma_\varepsilon^2$$

MS_{Treat} $\qquad\qquad$ MS_{Error}

So, if Ho is true..

- E(MSE) = σ_ε^2
- E(MST)= σ_ε^2

If Ho is not true

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

so...

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

The F word

- $F = MST / MSE$
- $E(F|H_o \text{ true})$?
- $E(F|H_a \text{ true})$?

The F word

- If Ho is true, then MST/MSE will be distributed as $F(df_t, df_e)$
- if not it will be distributed some other way
  - i.e. it will be unlikely to be distributed that way, or $p<.05$ (or whatever your alpha is)

Partitioning SS and df

- SSTotal = SSTreatment + SSError
- dfTotal = dfTreatment +dfError

More Precisely..

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

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		

Conclusions

- Easy to do
- Only tells you that two means are different, not which two, or three or whatever
- It’s about the pattern overall
- You can figure that out, we’ll do that next week.