Analysis of Variance

- ANOVA and its terminology
- Within and between subject designs
- Case study

Analysis of Variance (Anova)

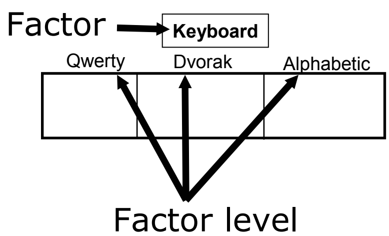
Statistical Workhorse

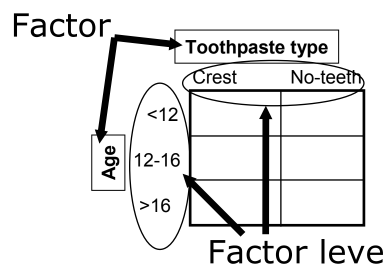
- supports moderately complex experimental designs and statistical analysis
- Lets you examine multiple independent variables at the same time
- Examples:
  - There is no difference between people's mouse typing ability on the Dvorak, Alphabetic and Qwerty keyboard
  - There is no difference in the number of cavities of people aged under 12, between 12-16, and older than 16 when using Crest vs No-teeth toothpaste

Analysis of Variance (Anova)

Terminology

- Factor = independent variable
- Factor level = specific value of independent variable





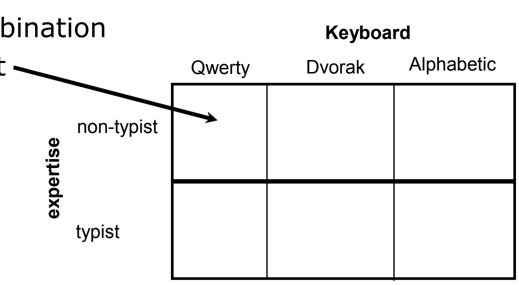
Anova terminology

Factorial design

- cross combination of levels of one factor with levels of another
- eg keyboard type (3) x expertise (2)

Cell

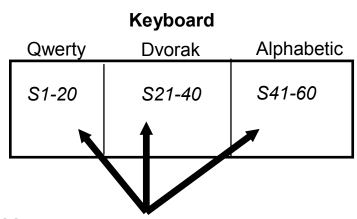
- unique treatment combination
- eg qwerty x non-typist



Anova terminology

Between subjects (*aka* nested factors)

- subject assigned to only one factor level of treatment
- control is general population
- advantage:
  - –guarantees independence i.e., no learning effects
- problem:
  - –greater variability, requires more subjects



different subjects in each cell

Anova terminology

Within subjects (*aka* crossed factors)

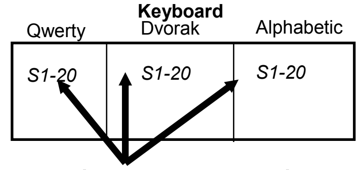
–subjects assigned to all factor levels of a treatment

–advantages

- requires fewer subjects
- subjects act as their own control
- less variability as subject measures are paired

–problems:

- order effects



same subjects in each cell

Anova terminology

Order effects

- within subjects only
- doing one factor level affects performance in doing the next factor level, usually through learning

- example:
  - learning to mouse type on *any* keyboard improves performance on the next keyboard
  - Alphabetic > Dvorak > Qwerty performance
    even if there was really no difference between keyboards!

S1: Q then D then A
S2: Q then D then A
S3: Q then D then A
S4: Q then D then A …

Anova terminology

Counter-balanced ordering

- mitigates order problem
- subjects do factor levels in different orders
- distributes the order effect across all conditions, but does not remove them
- Fails if order effects are not the equal between conditions
  - people's performance improves when starting on Qwerty
    but worsens when starting on Dvorak

S1: Q then D then A
S2: D then A then Q
S3: A then Q then D
S4: Q then A then D…

Anova terminology

Mixed factor

- contains both between and within subject combinations

- within subjects: keyboard type
- between subjects: expertise

expertise	Keyboard: Qwerty	Dvorak	Alphabetic
non-typist	*S1-20*	*S1-20*	*S1-20*
typist	*S21-40*	*S21-40*	*S21-40*

Single Factor Analysis of Variance

Compare means between two or more factor levels within a single factor

example:

- dependent variable: mouse-typing speed
- independent variable (factor): keyboard
- between subject design

Keyboard

Qwerty	Alphabetic	Dvorak
S1: 25 secs S2: 29 … S20: 33	S21: 40 secs S22: 55 … S40: 33	S51: 17 secs S52: 45 … S60: 23

Anova

Compares relationships between many factors

In reality, we must look at multiple variables to understand what is going on

Provides more informed results
- considers the *interactions* between factors

Anova Interactions

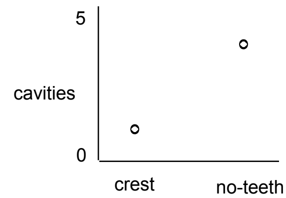
Example interaction
- typists are faster on Qwerty than the other keyboards
- non-typists perform the same across all keyboards
- cannot simply say that one keyboard is best

	Qwerty	Alphabetic	Dvorak
non-typist	S1-S10	S11-S20	S21-S30
typist	S31-S40	S41-S50	S51-S60

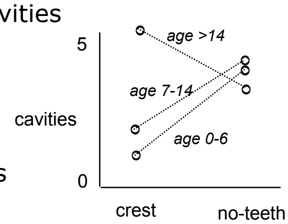
Anova - Interactions

Example:

- t-test: crest *vs* no-teeth
  - subjects who use crest have fewer cavities



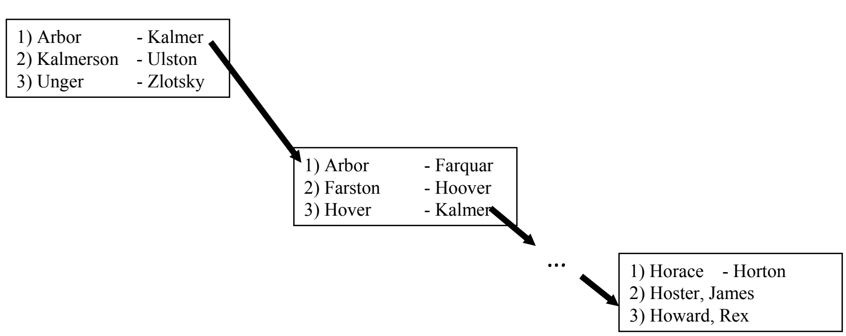
- anova: toothpaste x age
  - subjects 14 or less have fewer cavities with crest.
  - subjects older than 14 have fewer cavities with no-teeth.
- interpretation?
  - the sweet taste of crest makes kids use it more, while it repels older folks



Anova case study

The situation

- text-based menu display for large telephone directory
- names listed as a range within a selectable menu item
- users navigate menu until unique names are reached



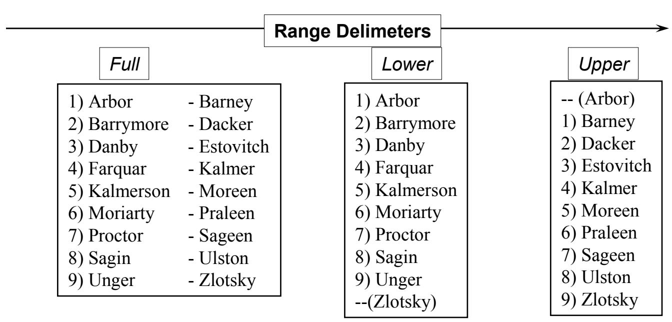
Anova case study

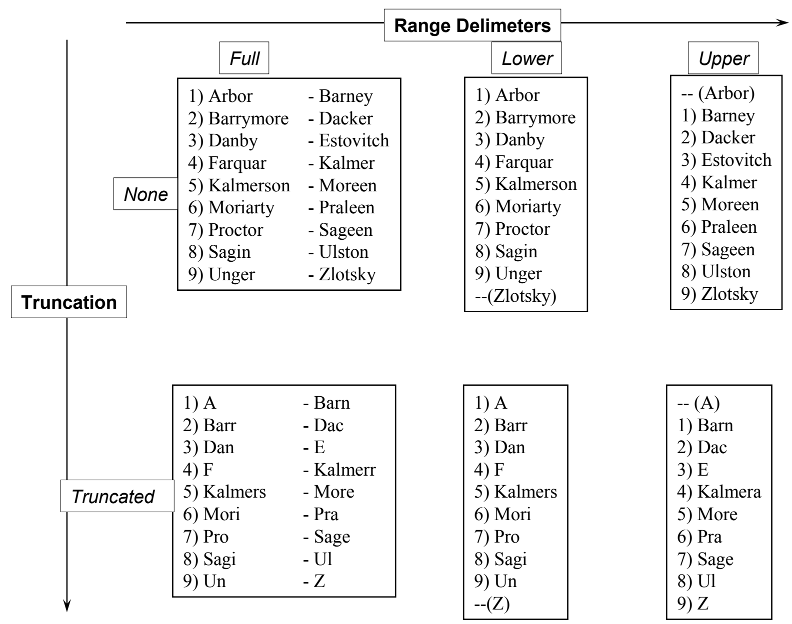
The problem

- we can display these ranges in several possible ways
- expected users have varied computer experiences

General question

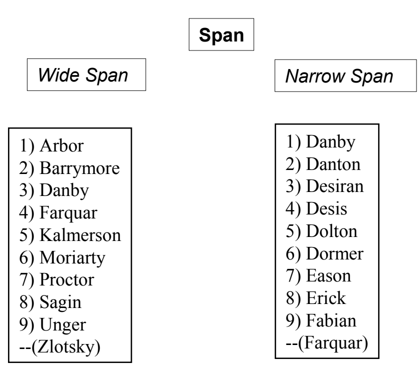
- which display method is best for particular classes of user expertise?





Span

as one descends the menu hierarchy, name suffixes become similar



Null Hypothesis

- six menu display systems based on combinations of ***truncation*** and ***range delimiter*** methods do not differ significantly from each other as measured by people's **scanning speed** and ***error rate***
- ***menu span*** and ***user experience*** has no significant effect on these results

- 2 level (truncation) x
  2 level (menu span) x
  2 level (experience) x
  3 level (delimiter)

		Truncated		Not Truncated	
		narrow	wide	narrow	wide
Full	Novice	S1-8	S1-8	S1-8	S1-8
	Expert	S9-16	S9-16	S9-16	S9-16
Upper	Novice	S17-24	S17-24	S17-24	S17-24
	Expert	S25-32	S25-32	S25-32	S25-32
Lower	Novice	S33-40	S33-40	S33-40	S33-40
	Expert	S40-48	S40-48	S40-48	S40-48

Statistical results

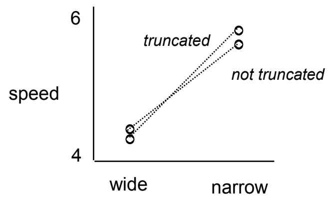
Scanning speed

	F-ratio.	*p*
Range delimeter (R)	**2.2***	**<0.5**
Truncation (T)	0.4	
Experience (E)	**5.5***	**<0.5**
Menu Span (S)	**216.0****	**<0.01**
RxT	0.0	
RxE	1.0	
RxS	3.0	
TxE	1.1	
TxS	**14.8***	**<0.5**
ExS	1.0	
RxTxE	0.0	
RxTxS	1.0	
RxExS	1.7	
TxExS	0.3	
RxTxExS	0.5	

Statistical results

Scanning speed:

- Truncation x Span



Main effects (means)

		Full	Lower	Upper
Full		----	1.15*	1.31*
Lower			----	0.16
Upper				----
Span:	Wide	4.35		
	Narrow	5.54		
Experience	Novice	5.44		
	Expert	4.36		

Results on Selection time

- Full range delimiters slowest
- Truncation has very minor effect on time: ignore
- Narrow span menus are slowest
- Novices are slower

Statistical results

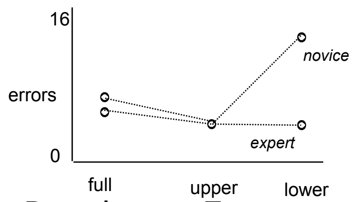
Error rate

	F-ratio.	*p*
Range delimeter (R)	**3.7***	**<0.5**
Truncation (T)	2.7	
Experience (E)	**5.6***	**<0.5**
Menu Span (S)	**77.9****	**<0.01**
RxT	1.1	
RxE	**4.7***	**<0.5**
RxS	**5.4***	**<0.5**
TxE	1.2	
TxS	1.5	
ExS	2.0	
RxTxE	0.5	
RxTxS	1.6	
RxExS	1.4	
TxExS	0.1	
RxTxExS	0.1	

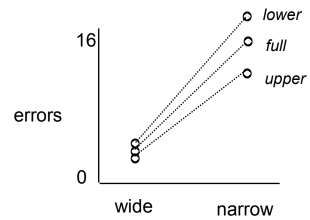
Statistical results

Error rates

Range x Experience



Range x Span



Results on Errors

- more errors with lower range delimiters at narrow span
- truncation has no effect on errors
- novices have more errors at lower range delimiter

Conclusions

Upper range delimiter is best

Truncation up to the implementers

Keep users from descending the menu hierarchy

Experience is critical in menu displays

You now know

Anova terminology

- factors, levels, cells
- factorial design
  - between, within, mixed designs

You should be able to:

Find a paper in CHI proceedings that uses Anova

Draw the Anova table, and state

dependant variables

independant variables / factors

factor levels

between/within subject design