12.2-3: Factorial ANOVA and Blocking

15 Nov 2011
BUSI275
Dr. Sean Ho

- HW8 due next Tue
- Please download:
  19-Eukanuba.xls
  19-Applebees.xls

Outline for today

- Factorial ANOVA (multiple nominal IVs)
  - Assumptions
  - Graphing
  - Model and calculations
  - Main effects
  - Interaction / moderation
- Randomized Complete Block ANOVA
  - Fixed vs. random effects
  - Model and calculations
  - Post-hoc analysis: Fisher's LSD

Factorial ANOVA

DV: Purch Amt

IV1: Gender

IV2: Src		
	$40 $20 $25 $32	$17 $21 $19 $22
	$30 $26 $25	$22 $19
	$40 $45 $52 $43 $39 $48	$50 $60 $55

- Equivalent to multiple regression
  - Except with nominal predictors
  - N-way ANOVA for N predictors
- IVs are “between-groups” factors:
  - Divide up sample into cells
  - Each participant in only 1 cell
- If your IVs are mixed continuous / nominal, try regression using dummy variables
  - Although this may result in many IVs!
- You can also try ANCOVA:
  - Continuous “covariates” are first factored out then regular ANOVA is done on residuals

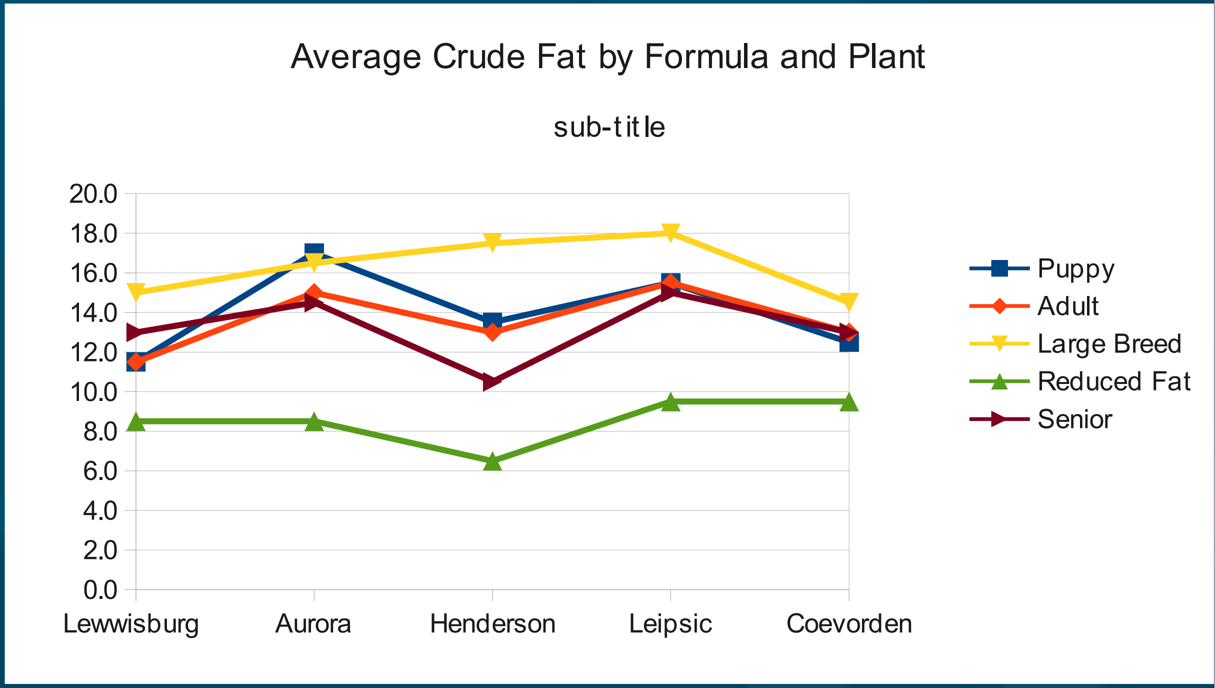
Assumptions

		DV: Purch Amt	
		IV1: Gender	
IV2: Src		$40 $20 $25 $32	$17 $21 $19 $22
		$30 $26 $25	$22 $19
		$40 $45 $52 $43 $39 $48	$50 $60 $55

- Same as for regular ANOVA, per cell:
  - DV continuous
  - Independent observations, independent cells (groups)
  - DV normal within each cell
  - Variance of DV similar across all cells:
    - (largest SD) / (smallest SD) < 2
- The last two are less important as long as:
  - Total sample size is reasonably large (>50)
  - Balanced design: all cells similar sample size
  - No rows/cols are completely empty

Graphing 2-way ANOVA data

- Dataset: 19-Eukanuba.xls
- The DV has a different distribution in each cell
- One way to visualize: condense it down to the average of DV within each cell
- Pivot Table:
  - Formula (row)
  - Plant (col)
  - Average of Fat (data)
- Try a line chart:



2-way ANOVA: model

- Main effects on each IV, plus interactions:
  - Purchase = b_0 + (Gender effect) + (Src effct) + (Gender*Source effect) + ε
- Decomposition of variance:
  - $SS_{tot} = SS_{Gen} + SS_{Src} + SS_{Gen*Src} + SS_{resid}$
- Global F-test looks for any effect of IVs on DV
  - If not significant, check for violations of assumptions
- Effect size η^2 is akin to R^2: $1 - (SS_{resid} / SS_{tot})$

2-way ANOVA: calculating

-	IV_1 (a levels)	IV_2 (b levels)	IV_1*IV_2 (Interaction)
SS	$bn\sum_{i=1}^{a}(\bar{x}_i-\bar{x})^2$	$an\sum_{j=1}^{b}(\bar{x}_j-\bar{x})^2$	$n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij}-\bar{x}_i-\bar{x}_j+\bar{x})^2$
df	a - 1	b - 1	(a - 1) * (b - 1)

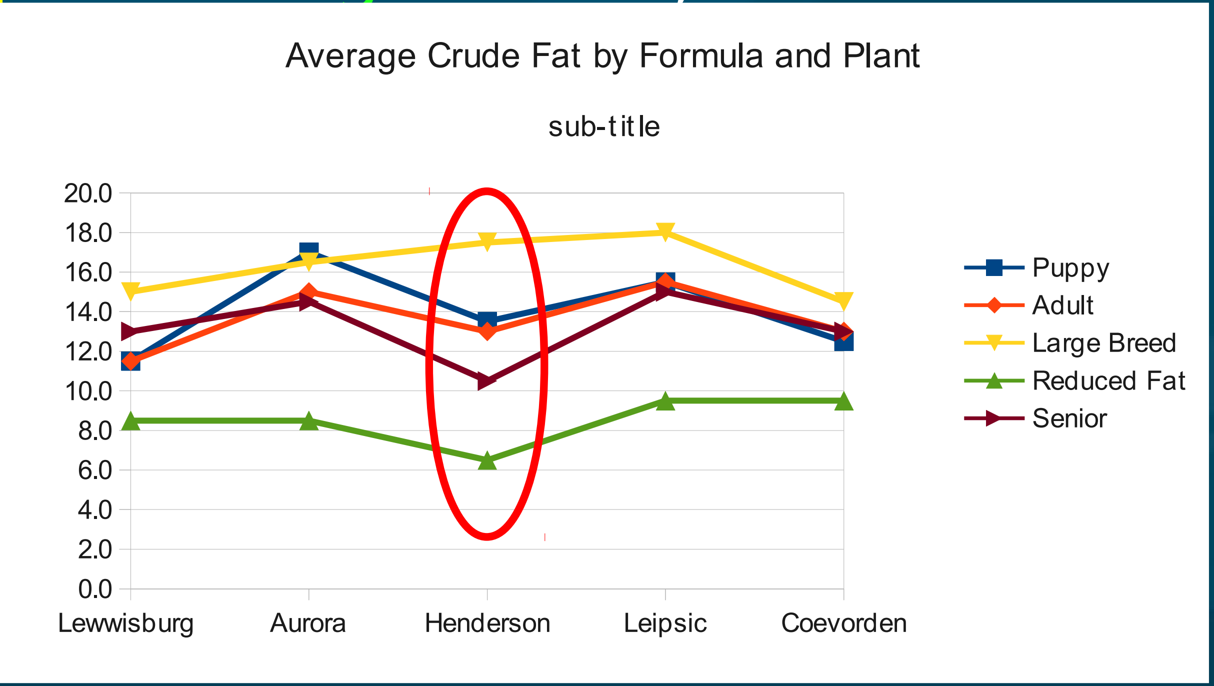
- Also find SS_{tot} as before, and SS_{res}
  - $df_{tot} = n - 1$, and $df_{res} = n - ab$
  - The SS and df always add up:
    - Tot = IV_1 + IV_2 + (IV_1*IV_2) + Resid
- 3 F-tests: IV_1, IV_2, and interaction
  - e.g., main effect on IV_1: $F = MS_1 / MS_{res}$

Main effects

- A main effect is a one-way ANOVA on one IV, after controlling for the other predictors
  - Analogous to t-tests on slope for each IV in multiple regression
  - Here, the main effects are themselves F-tests
- E.g., do females spend more at your site, after accounting for source?
  - 2-way ANOVA on both Gender and Source, then look at main effect of Gender
- E.g., do different formulas have different fat content, across all plants?

Interactions

- When the effect of one IV on the DV changes, depending on the level of the moderator
- e.g., females spend more in response to print ads, but males spend more in response to web
- e.g., Henderson generally has lower fat than the other plants, except for Large Breed, where it has the second-highest fat:
- Plot means, note change in shape of the curves



Randomized Complete Block

- Special case of 2-way ANOVA, where each cell has only 1 observation
- IV1 is the factor: typically a fixed effect
  - Fixed: levels are set in the hypothesis: e.g., gender, province, store branch, plant
- IV2 is the block: typically a random effect
  - Random: levels are sampled from a population: e.g., customer, truck, day
- e.g, 19-Applebees.xls:
  - Factor: Restaurant
  - Block: Week
  - DV: Revenue

	Restaurant		
Week	8.34	6.79	9.18
	10.7	10.0	12.8
	…	…	…
	…	…	…
	…	…	…

Randomized Block model

- A complete 2-way ANOVA on this data would have zero residual in each cell
  - So the interaction term serves as “residual”
  - Tot = Factor + Blocking + Residual
    - $df_{res} = (a - 1)(b - 1)$
- Factor effect (IV_1): $F = MS_1 / MS_{res}$
  - This is usually what we're most interested in
- Blocking effect (IV_2): $F = MS_2 / MS_{res}$
  - If non-sig, then blocking was not necessary and we could've just done a 1-way ANOVA

Post-hoc: Fisher's LSD test

- If the factor effect is significant, one post-hoc test we can use is Fisher's least sig. diff. test
  - Like Tukey-Kramer, but for equal-size cells
- Critical range:
  - t: 2-tails, use df_{res}
  - b: # blocks (IV_2)

$$LSD = t\sqrt{\frac{2\,MS_{res}}{b}}$$

- For all pairs of levels of the main factor, if the difference of means $|\bar{x}_i - \bar{x}_j|$ exceeds LSD, then those two groups differ significantly
  - Use the results to cluster the factor levels

TODO

- HW8 (ch15,12): due next Tues
- Projects:
  - Presentations in two weeks!
  - If you don't know what analysis to perform, or how to perform it, ask me for help