Categorical Data Analysis: Logistic Regression and Log-Linear Regression

For discussion:
- ***Myers & Hayes***
- ***Horowitz***

For the lecture:
- ***GenderDepr.sav***
- ***Fitzpatrick et al.***

26 Nov 2010
CPSY501
Dr. Sean Ho
Trinity Western University

TRINITY WESTERN UNIVERSITY

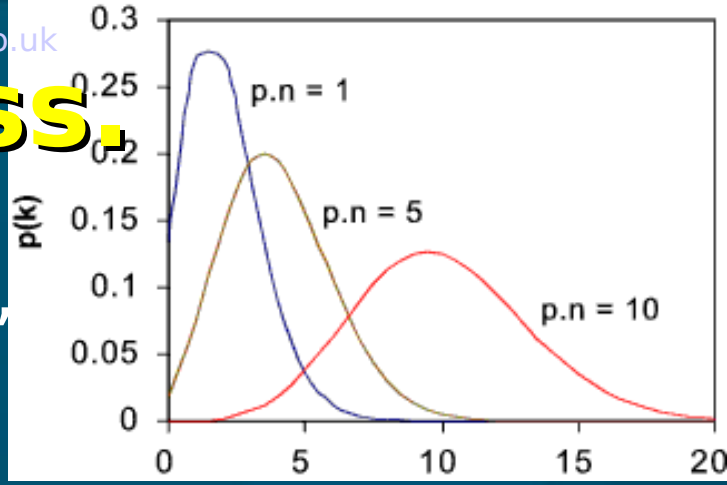
Outline for today

- Linear models:
  - Logistic regression
  - Log-linear regression
- Categorical Data Analysis
  - 2 vars: chi-squared test, effect sizes
  - Multiple vars: log-linear analysis
  - Example: Fitzpatrick '01

Generalized Linear Model

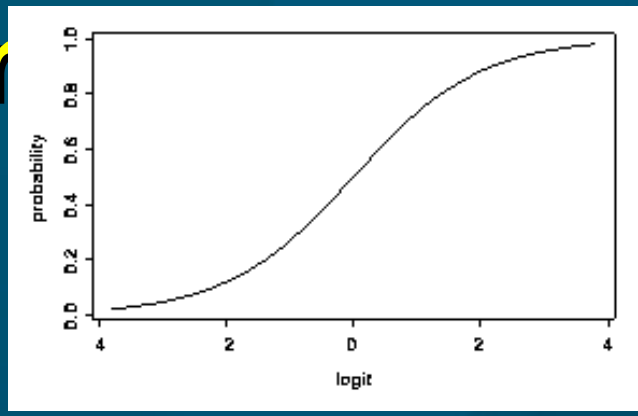
- To deal with a categorical DV, we need the Generalized Linear Model:
  - f(Y) ~ X_1 + X_2 + ...
  - The linear model predicts not Y directly, but the link function f() applied to Y
- Examples of link functions:
  - f(Y) = log(Y): log-linear regression
    Used when Y represents counts/frequencies
  - f(Y) = logit(Y): logistic regression
    Used when Y represents a probability (0..1)

GLM: log-linear regress.

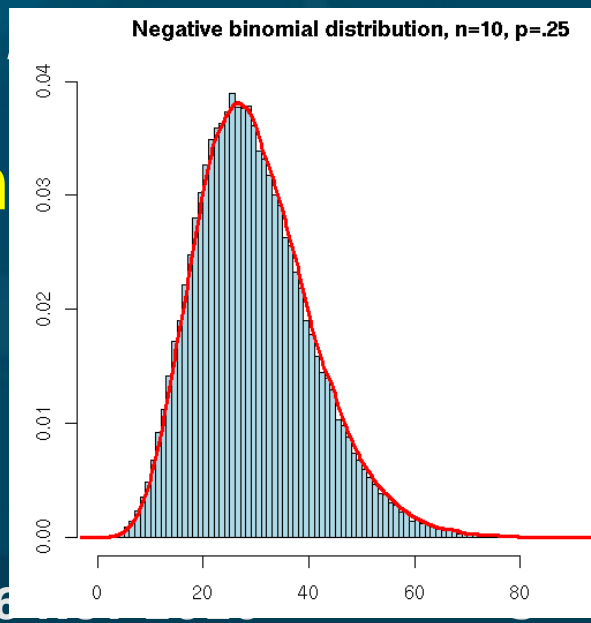


- When DV is counts/frequencies, its distribution is often not normal, but Poisson
  - e.g., DV = # violent altercations
  - If mean is large, Poisson → normal
- e.g., "log(violent_alts) ~ depression"
  - residuals (ε) are also Poisson distributed
- Log-linear is also used to look at many cat. vars
  - IVs are all categorical (factorial cells)
  - DV = # people in each cell
  - Fitzpatrick, et al. example paper later

GLM: logistic regressior



- When DV is a probability (0 to 1), the distribution is binomial
  - e.g., DV = "likelihood to develop depress."
  - Probability of Y: P(Y). Odds of Y: $\frac{P(Y)}{1-P(Y)}$
  - Logit link function: logit(Y) = log(odds(Y))
- Also works for DV = # out of total
  - e.g., DV = "# correct out of 100
  - As #tot → ∞, binomial → Poisson
- Also works for binary (dichot.) DV
  - e.g., DV = "is pregnant"



Outline for today

- Linear models:
  - Logistic regression
  - Log-linear regression
- Categorical Data Analysis
  - 2 vars: chi-squared test, effect sizes
  - Multiple vars: log-linear analysis
  - Example: Fitzpatrick '01

Contingency tables

- When comparing two categorical variables, all observations can be partitioned into cells of the contingency table
  - e.g., two dichotomous variables: 2x2 table
  - Gender vs. clinically depressed:

	Depressed	Not Depressed
Female	**126**	**154**
Male	**98**	**122**

  - RQ: is there a significant relationship between gender and depression?

SPSS: frequency data

- Usually, each row in the Data View represents one participant
  - In this case, we'd have 500 rows
- For our example, each row will represent one cell of the contingency table, and we will specify the frequency for each cell
- Open: GenderDepr.sav
- Data → Weight Cases: Weight Cases by
  - Select "Frequency" as Frequency Variable

2 categorical vars: χ^2 and φ

- Chi-squared (χ^2) test: Two categorical variables
  - Asks: is there a significant relationship?
- Requirements on expected cell counts:
  - No cells have expected count ≤ 1, and
  - <20% of cells have expected count < 5
  - Else (for few counts) use Fisher's exact test
- Effect size:
  - φ is akin to correlation: definition: $\varphi^2 = \chi^2 / n$
  - Cramer's V extends φ for more than 2 levels
  - Odds ratio: #yes / #no

SPSS: χ^2 and φ

- Analyze → Descriptives → Crosstabs:
  - One var goes in Row(s), one in Column(s)
  - Cells: Counts: Observed, Expected, and Residuals: Standardized, may also want Percentages: Row, Column, and Total
  - Statistics: Chi-square, Phi and Cramer's V
  - Exact: Fisher's exact test: best for small counts, computationally intensive
- If χ^2 is significant, use standardized residuals (z-scores) to follow-up which categories differ

Reporting χ^2 results

- As in ANOVA, IVs with several categories require follow-up analysis to determine which categories show the effect
  - The equivalent of a single pairwise comparison is a 2x2 contingency table!
- Report:

  "There was a significant association between gender and depression, $\chi^2(1) =$ ___, $p < .001$. Females were twice as likely to have depression as males."
  - Odds ratio: (#F w/depr) / (#M w/depr)

Outline for today

- Linear models:
  - Logistic regression
  - Log-linear regression
- Categorical Data Analysis
  - 2 vars: chi-squared test, effect sizes
  - Multiple vars: log-linear analysis
  - Example: Fitzpatrick '01

Many categorical variables

- Need not have IV/DV distinction
- Use log-linear: Generalized Linear Model
  - Include all the categorical vars as IVs
  - DV = # people in each cell
  - e.g., "count ~ employment * gender * depr"
- Look for moderation / interactions:
  - e.g., employment * gender * depression
- Then lower-level interactions and main effects
  - e.g., employment * depression

Goodness of Fit

- Two χ^2 metrics measure how well our model (expected counts) fits the data (observed):
  - Pearson χ^2 and likelihood ratio (G) (likelihood ratio is preferred for small n)
- Significance test looks for deviation of observed counts from expected (model)
  - So if our model fits the data well, then the Pearson and likelihood ratio should be small, and the test should be non-significant
- SPSS tries removing various effects to find the simplest model that still fits the data well

Hierarchical Backward Select'n

- By default, SPSS log-linear regression uses automatic hierarchical “backward” selection:
- Starts with all main effects and all interactions
  - For a “saturated” categorical model, all cells in contingency table are modelled, so the “full-factorial” model fits the data perfectly: likelihood ratio is 0 and p-value = 1.0.
- Then removes effects one at a time, starting with higher-order interactions first:
  - Does it have a significant effect on fit?
  - How much does fit worsen? (ΔG)

Example: Fitzpatrick et al.

- Fitzpatrick, M., Stalikas, A., Iwakabe, S. (2001). Examining Counselor Interventions and Client Progress in the Context of the Therapeutic Alliance. *Psychotherapy, 38(2), 160-170.*

- Exploratory design with 3 categorical variables, coded from session recordings / transcripts:
  - Counsellor interventions (VRM)
  - Client good moments (GM)
  - Strength of working alliance (WAI)
- Therapy: 21 sessions, male & female clients & therapists, expert therapists, diverse models.

Fitzpatrick: Research Question

- RQ: For expert therapists, what associations exist amongst VRM, GM, and WAI?
- Therapist Verbal Response Modes:
  - 8 categories: encouragement, reflection, self-disclosure, guidance, etc.
- Client Good Moments:
  - Significant (I)nformation, (E)xploratory, or (A)ffective-Expressive
- Working Alliance Inventory
  - Observer rates: low, moderate, high

Fitzpatrick: Abstract

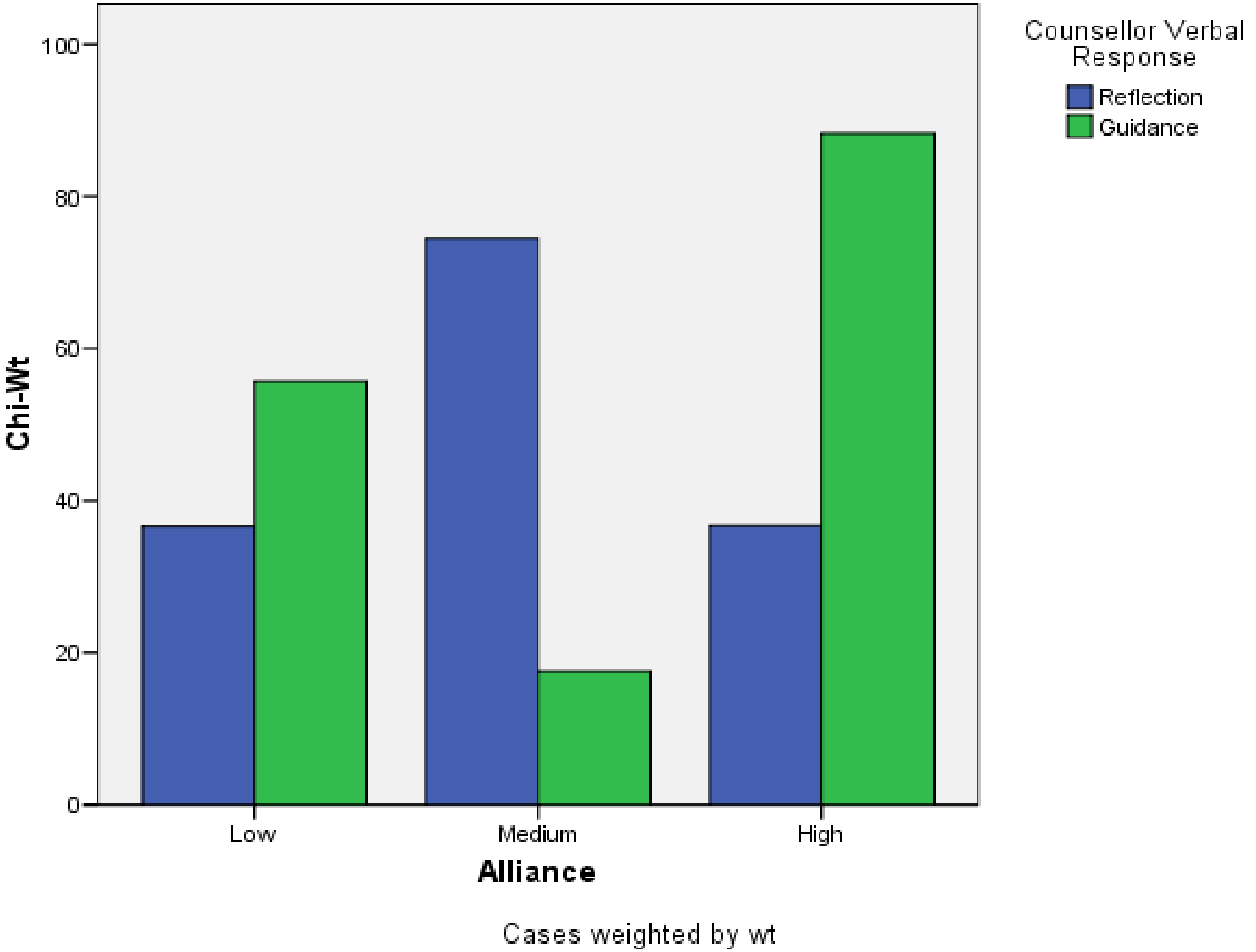
- Client "good moments" did not necessarily increase with Alliance
- Different interventions fit with good moments of client information (GM-I) at different Alliance levels.
- "Qualitatively different therapeutic processes are in operation at different Alliance levels."
- Explain each statement and how it summarizes the results.

Top-down Analysis: Interaction

- As in ANOVA and Regression, Loglinear analysis starts with the most complex interaction ("highest order") and tests if it adds incrementally to the overall model fit
  - Compare with ΔR^2 in regression analysis
- Interpretation focuses on:
  - 3-way interaction: VRM * GM * WAI
  - Then the 2-way interactions: GM * WAI, etc.
- Fitzpatrick did separate analyses for each of the three kinds of good moments: GM-I, GM-E, GM-A

Results: Interactions

- 2-way CGM-E x WAI interaction:
  - Exploratory Good Moments tended to occur more frequently in High Alliance sessions
- 2-way WAI x VRM interaction:
  - Structured interventions (guidance) take place in Hi or Lo Alliance sessions, while
  - Unstructured interventions (reflection) are higher in Moderate Alliance sessions
  - Describes shared features of “working through” and “working with” clients, different functions of safety & guidance.


Counsellor Verbal Response
Reflection
Guidance
100
80
60
40
20
0
Chi-Wt
Low
Medium
High
Alliance
Cases weighted by wt

Formatting Tables in MS-Word

- Use the “insert table” and “table properties” functions of Word to build your tables; don’t do it manually.
- General guidelines for table formatting can be found on pages 147-176 of the APA manual.
- Additional tips and examples: see NCFR site: http://oregonstate.edu/~acock/tables/
- In particular, pay attention to the column alignment article, for how to get your numbers to align according to the decimal point.