Meta-Analysis, Generalized Linear Models, and Categorical Data Analysis

Download:
- ***Hill & Lent***
- ***GenderDepr.sav***
- ***Fitzpatrick et al.***

19 Nov 2010
CPSY501

Dr. Sean Ho
Trinity Western University

Outline for today

- Meta-Analysis
  - Example: Hill & Lent (2006)
- Linear models:
  - Covers every test we know so far
  - Logistic regression
  - Log-linear regression
- Categorical Data Analysis
  - 2 vars: chi-squared test, effect sizes
  - Multiple vars: log-linear analysis
  - Example: Fitzpatrick '01

Meta-Analysis

- The APA journal has basic standards for literature review in many areas
- Meta-Analysis (MA) is a tool for combining results of quantitative studies in a systematic, quantitative way.
- Example Meta-Analysis journal article:
  - Hill, C. E., & Lent, R. W. (2006). A narrative and meta-analytic review of helping skills training: Time to revive a dormant area of inquiry. *Psychotherapy: Theory, Research, Practice, Training*, 43(2), 154–172.

MA focuses on effect sizes

- Choose groups of studies and subgroups of studies to combine and compare
- Each individual study might not have significance, due to low sample size
  - Combine results from multiple studies
  - Must be careful that studies are comparable
- g : difference between the means divided by the pooled standard deviation
- d : unbiased estimates of the population effect size as reported by each study

Combining effect sizes: ex.

- Example: two correlation studies, with
  - $r_1 = .22$ and $r_2 = .34$
  - $N_1 = 125$ and $N_2 = 43$
- Combine studies to estimate r
- Unweighted average: $(.22 + .34) / 2 = .28$
- Weighted average by sample size:
  $[\ .22(125) + .34(43)\ ] / (125 + 43) = .25$
- The larger sample has a smaller effect size!

Persuasiveness of MA

- Quality of studies (design, etc.)
- Comparability of studies:
  - Variables, measures, participants, etc.
  - Pay attention to moderating factors!
- RQ: Differences among types of training? (instruction, modeling, feedback)
- Do we know the "amount" of training time examined in each study?
- What impact might these factors have on the interpretation of the meta-analysis?

Hill & Lent (2006)

- p.159: Summary of strategy and symbols used
- p.160: List of studies being summarized ($k = 14$), including outcome measures, etc.
- Within each study, aggregate multiple measures by calculating mean effect size and standard error
- Use Cohen's (1988) criteria for effect size: d=0.20 (small), d=0.50 (med), d=0.80 (large)

Global analysis: outlier

- Hill & Lent chose to exclude one entire study as an outlier: p.161:
- "Given its potential to disproportionately influence effect sizes, especially in a relatively small set of studies, the outlier study was omitted in our subsequent analyses."
- Now only 13 studies left …
- Pros & cons of this omission?

Questions… pre-assignment

- Note: The same group of studies is used in all sections of Hill & Lent…
- How do the different research questions shape the MA calculations?
- How do confidence intervals help us interpret effect sizes (ES)?
- How do we integrate the results of different research questions?

Outline for today

- Meta-Analysis
  - Example: Hill & Lent (2006)
- Linear models:
  - Covers every test we know so far
  - Logistic regression
  - Log-linear regression
- Categorical Data Analysis
  - 2 vars: chi-squared test, effect sizes
  - Multiple vars: log-linear analysis
  - Example: Fitzpatrick '01

Introduction to linear models

- All these techniques use the same framework:
  - Regression and Curvilinear Regression
  - ANOVA
  - Multiple Regression, Factorial ANOVA
  - ANCOVA
- Even techniques with categorical DV:
  - Log-linear Regression
  - Logistic Regression
- RM-ANOVA is related, using multi-level models

Linear model: notation

- Recall the linear model of 1-way regression:
  - $Y = b_0 + b_1 X + \varepsilon$
  - Y and X are random vars (DV and IV)
  - b_0 and b_1 are parameters of the model
  - ε are the residuals (assumed to be "IID": independent, identically distributed, normal)
- May also be specified in shorthand:
  - $Y \sim X$, e.g., "depression ~ age"
  - Constant term (intercept) is implied

Linear model: multiple regress.

- Multiple regression, with interaction terms:
  - $Y \sim X_1 + X_2 + X_1{:}X_2$
  - Or simply $Y \sim X_1 * X_2$
  - This expands to: $Y = b_0 + b_1X_1 + b_2X_2 + b_{1,2}X_1{*}X_2 + \varepsilon$
- e.g., "depression ~ age * self_esteem"
  - Includes constant term, main effects, and interaction term (plus residuals)

Linear model: ANOVA

- ANOVA, using “dummy coding”: $Y \sim X$
  - if X is cat. w/ 3 levels: e.g., {Ctrl, CBT, CSG}
  - then it expands to: $Y = b_0 + b_1X_1 + b_2X_2 + \varepsilon$
  - where X_1 and X_2 are dummy variables: e.g., Ctrl-vs-CBT, Ctrl-vs-CSG (both 0/1 vars)
- Factorial ANOVA: $Y \sim X_1 * X_2$
  - e.g., X_1: {Ctrl, CBT, CSG}, X_2: {M, F} → 2 dummy vars for X_1, 1 for X_2.
- Planned comparisons: only include selected dummy vars

Linear model: ANCOVA

- ANCOVA = Regr! No problem to mix-and-match
  scale-level and categorical predictors!
- e.g., "depression ~ age * treatment"
  - age is scale; trmt. is cat. {Ctrl, CBT, CSG}
  - Dummy-code treatment as 2 binary vars
- If age is viewed as a covariate (not predictor), then the interaction should be non-significant
  - ANCOVA: partial out effect of age, so just focus on main effect of treatment
  - Model: "depression ~ age + treatment"

Linear model: curvilinear regr.

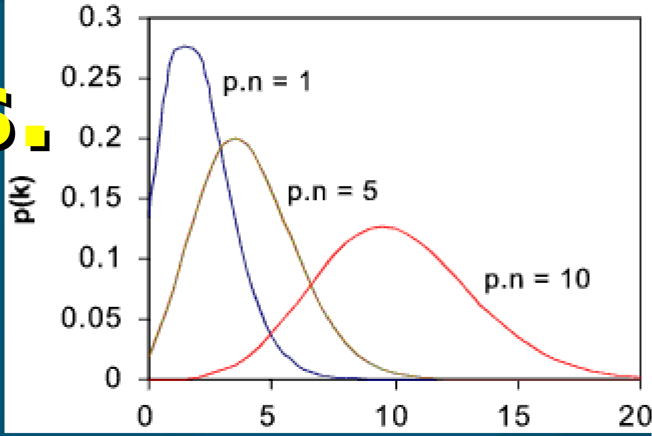
- Model data with a polynomial curve:
  - 1 IV, quadratic: $Y \sim X_1 + X_1^2$
  - 2 IVs, quadratic: $Y \sim (X_1 + X_2)^2$
    (full model has 6 terms, including constant)
- More complex curves exist, e.g., splines
  - Generalized Additive Model
- Visualize data (e.g., scatterplots)
  to guide choice of curve
- Try several models: best fit, fewest terms
  - → Tradeoff specificity vs. generalizability

Categorical data: GLM

- To deal with a categorical DV, we need the Generalized Linear Model:
  - $f(Y) \sim X_1 + X_2 + \ldots$
  - 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)

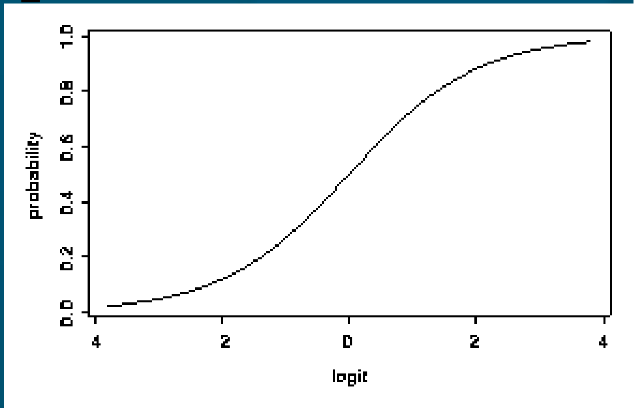
GLM: log-linear regress.

roymech.co.uk

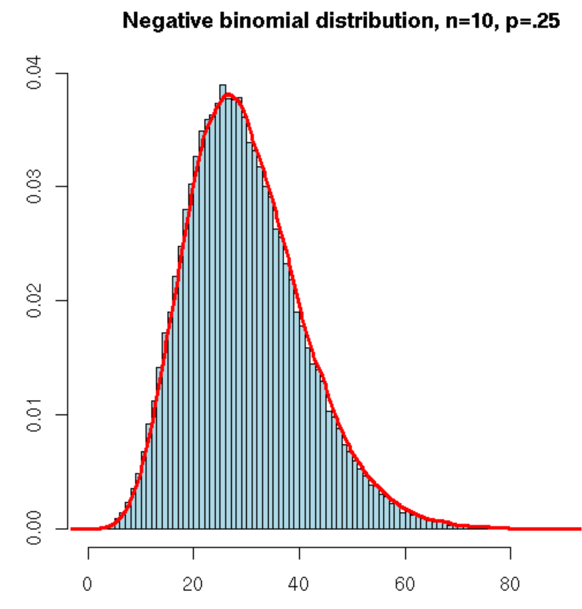


- When DV is counts/frequencies, its distribution is often not normal, but Poisson
  - e.g., DV = # violent altercations
  - If mean is large, Poisson → normal
- e.g., "log(violent_alts) ~ depression"
  - residuals (ε) are also Poisson distributed
- Log-linear is also used to look at many cat. vars
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

GLM: logistic regression



- 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: $\dfrac{P(Y)}{1-P(Y)}$
  - 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"



zoonek2.free.fr

Outline for today

- Meta-Analysis
  - Example: Hill & Lent (2006)
- Linear models:
  - Covers every test we know so far
  - 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

- Meta-Analysis
  - Example: Hill & Lent (2006)
- Linear models:
  - Covers every test we know so far
  - 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

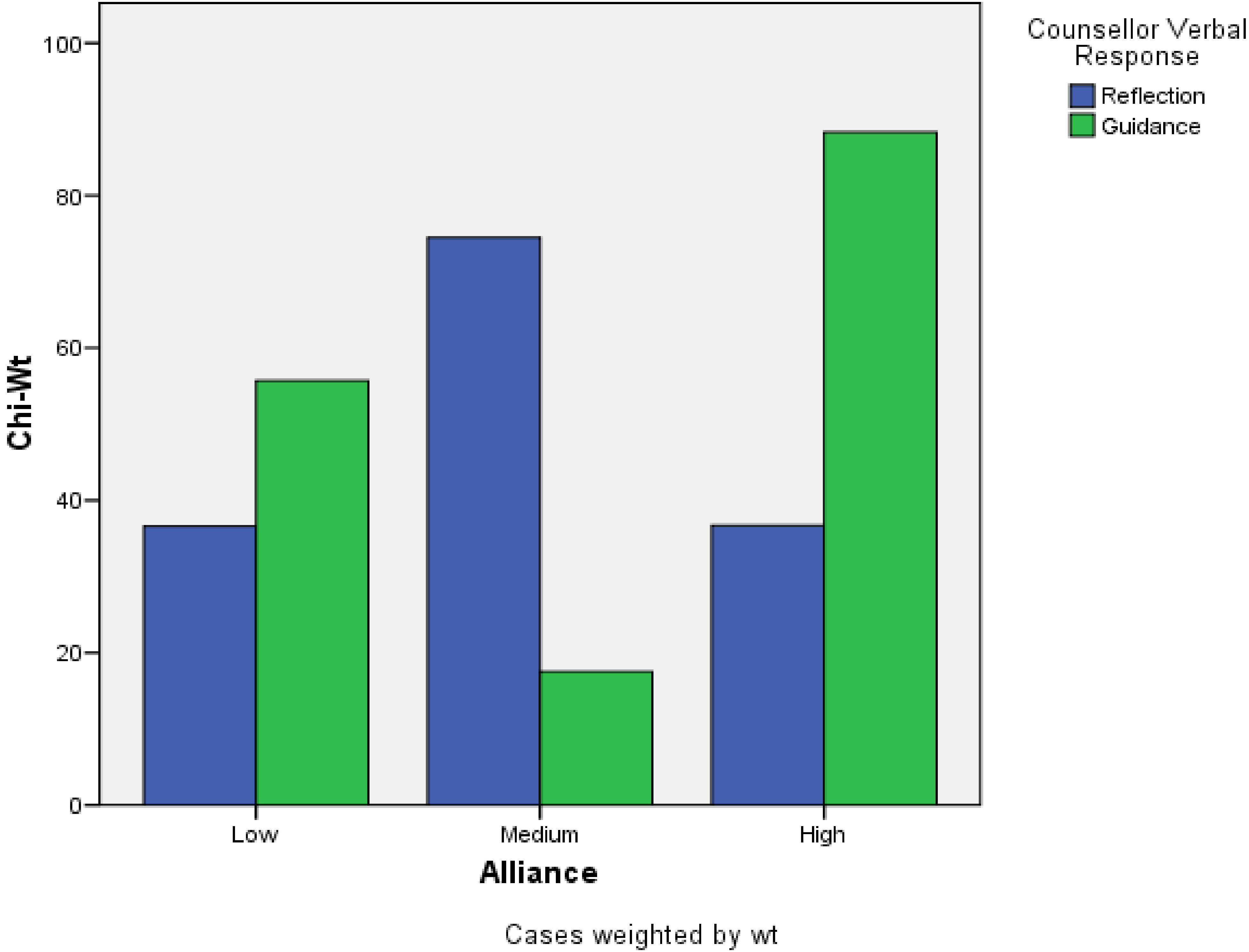
- Client "good moments" did not necessarily increase with Alliance
- Different interventions fit with good moments of client information (GM-I) at different Alliance levels.
- "Qualitatively different therapeutic processes are in operation at different Alliance levels."
- Explain each statement and how 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 whether 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
Chi-Wt
100
80
60
40
20
0
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.