

# Linkage Notes

Fritz Scheuren

March 2012

# Organization of Ideas

- Some Guesses about your Interests
- Context Brief
- Paradigm Brief
- Illustration Brief
- Play it Again?

# Context Briefly Covered

- Purpose of linkages
- Administrative Data?
- Statistical Information?
- E.g., Content v. Coverage
- Mixed Purpose (EITC and Decennial)

# Linkage Externalities

- Cost of editing
- Confidentiality
- Complexity of Inference
- Explainability and Comparability
- No Free Lunch, Of Course

# Paradigm Briefing

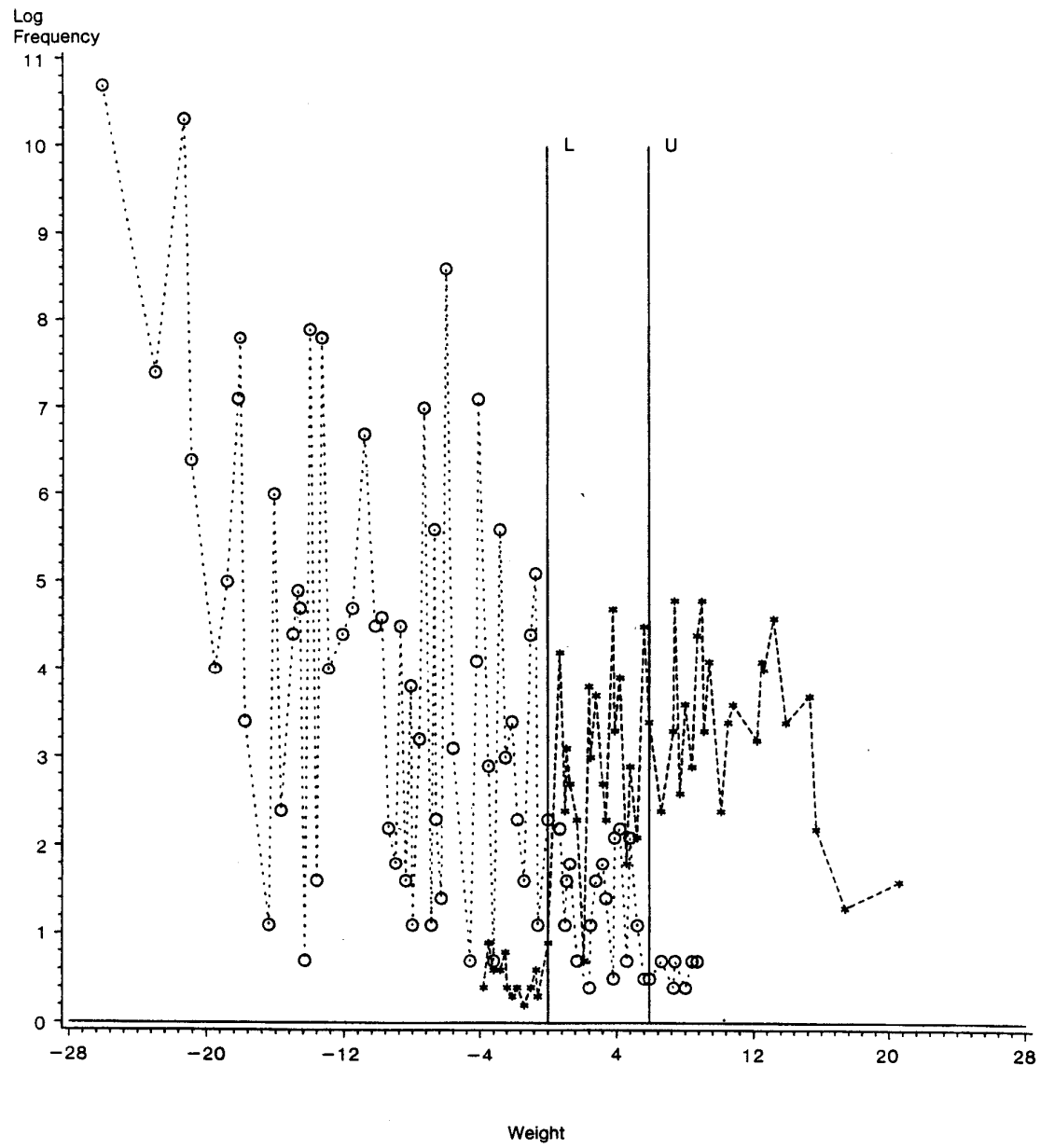
- Links of Matches two or more sources uniquely to put records together that are for the same entity
- Contrast Statistical Matching or Data Fusion

# Some Possible Themes

- Fellegi-Sunter (Mantra)
- Technology/Computer Advances
- Role of Paradata
- Role of Imputation
- How to Validate Results
- How to Complete Inference

# Fellegi-Sunter (Mantra)

- Essential Idea is that there be two bounds:
  - True Links (upper),
  - True Nonlinks (lower)
  - Middle Uncertainty
- Emphasis Error Bounds



0 = nonlink, \* = link cutoff "L" = 0 and cutoff "U" = 6



# More on Bounds

- Bound depends on Match variables available
- Linkage Probabilities are Model based, weakness
- But all variables in files to be used in inference

# Technology/Computer Advances

- Use Computer much more
- Clerical Review now Unaffordable, except for Sample Checking
- With Multiple System Estimation, multiple (Say Lower) bounds may make sense

# Role of Paradata

- Always important but Survey Uses push us to do much more
- Modeling True Link in Middle Uncertainly Area
- Moving away from Counts to Estimates

# Role of Imputation

- Multiple Imputation is recommended to Handle Uncertainty in Middle
- Allows us to use multiple models based on Paradata and capture variability

# How to Validate Results

- Small Samples needed to go to Ground to check
- Changes must be KAISEN
- “Best” to Benchmark Internationally

# How to Complete the Inference

- Rubin's Multiple Methods Help
- Bishop, Fienberg and Holland Book Helps too
- Test your explanation - Hard? Illustration!

# CPS DER Earnings Link

- Linkage Goal to Focus on 2005 earnings Results  
Comparing CPS to DER
- Robustness of CPS Poverty estimates to Changes in Imputation Rates over time
- Relation of CPS to DER by imputation and Proxy Status

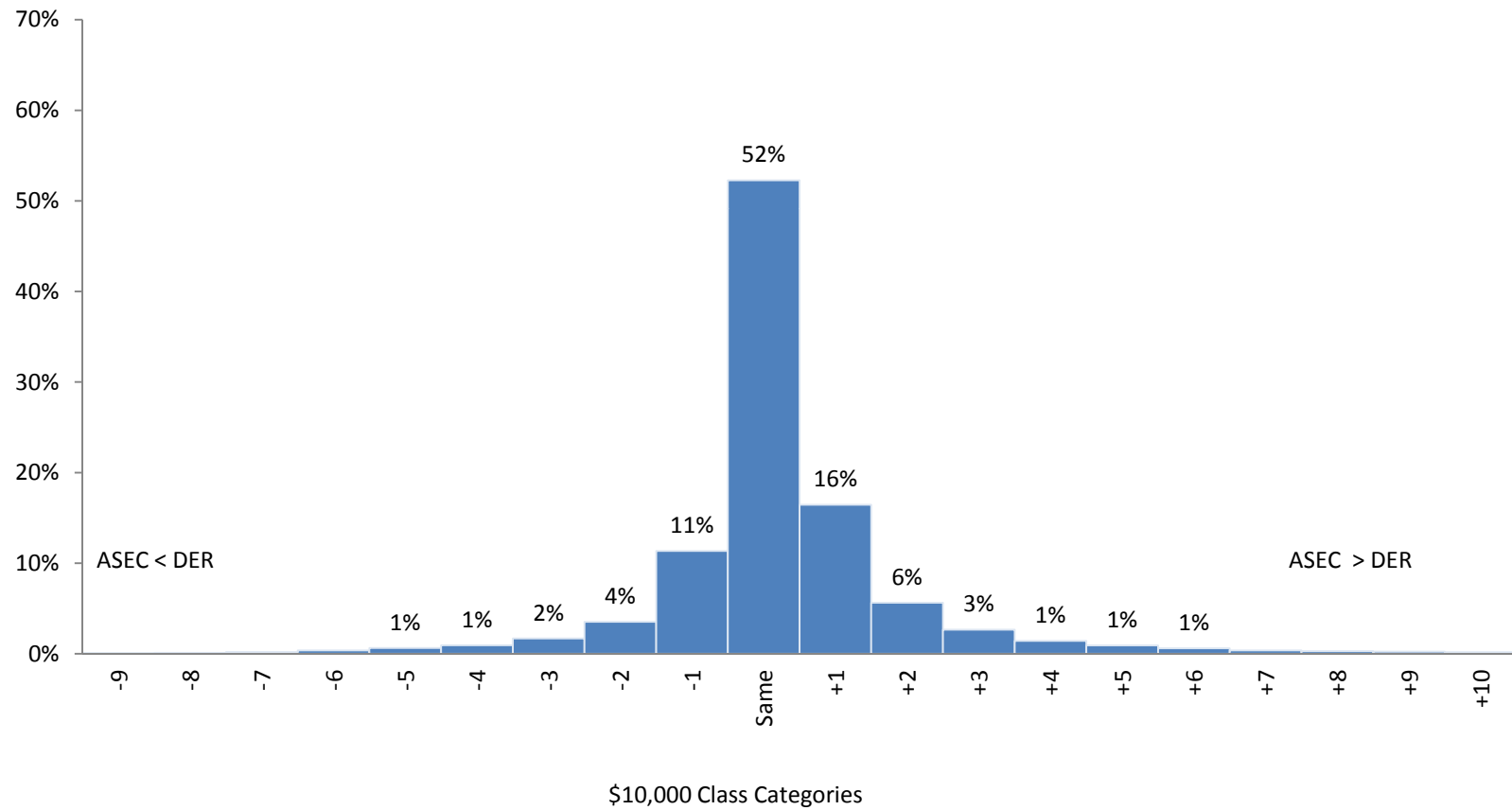
# CPS/DER Matched Results

- The two measures are cross-classified for matched individuals
- Equal Dollar Size Intervals (\$10,000 wide) are Used
- Classes tallied in both measures as agreeing on class or higher/lower and by how many

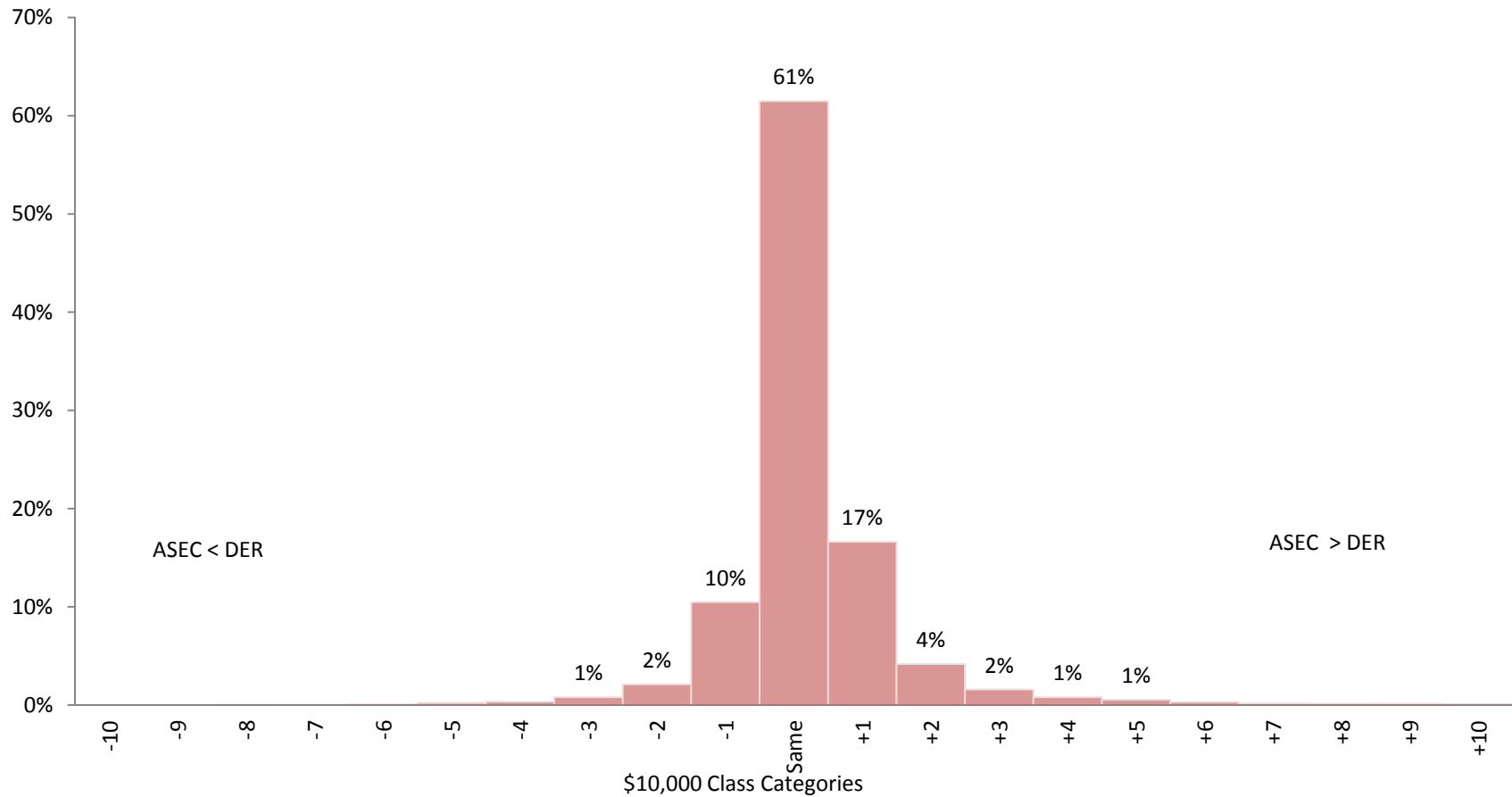


# All Earners: Distribution of ASEC and DER Earning Class Differences

## All: 2005

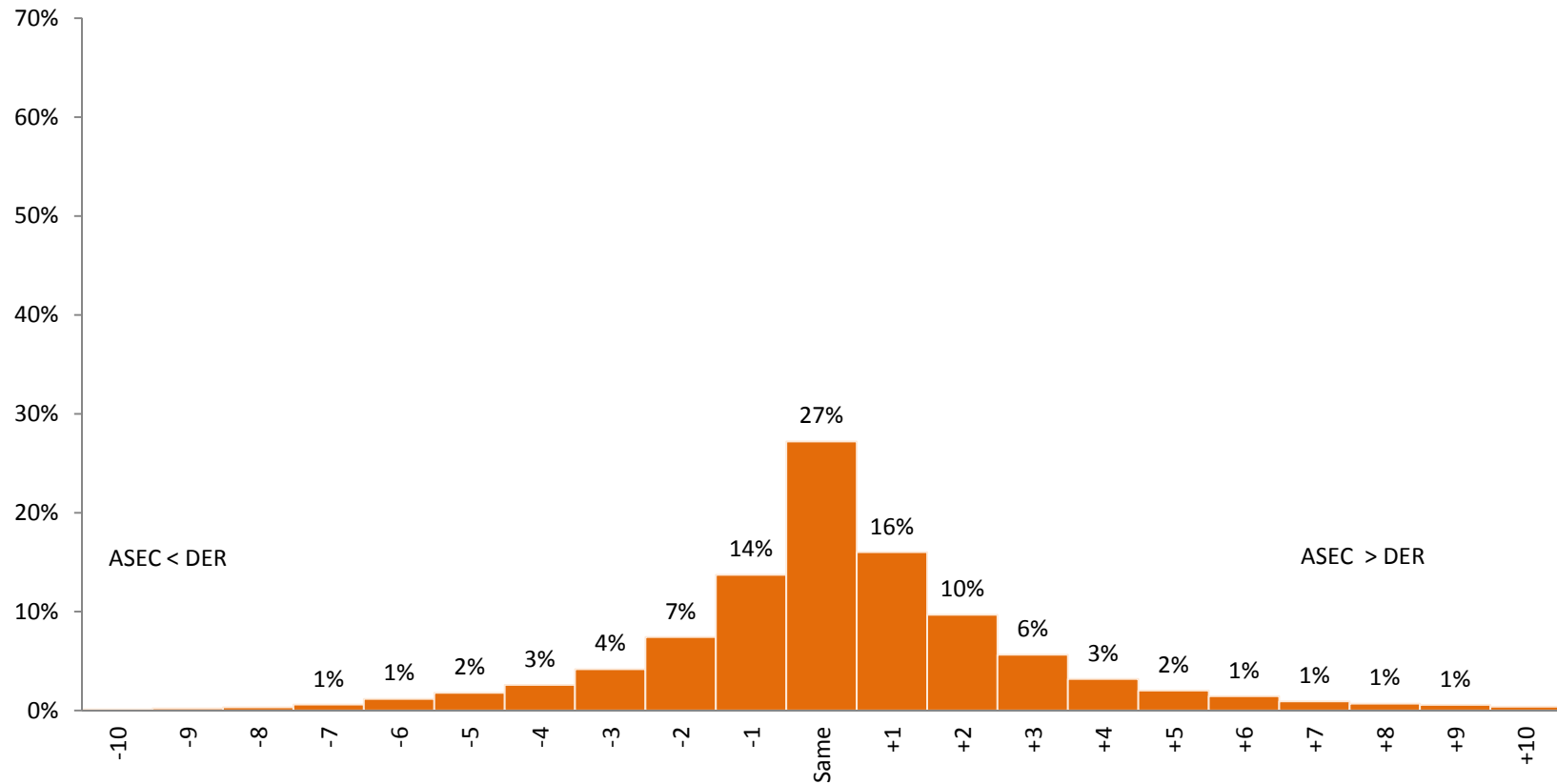


# Earners without Imputations: Distribution of ASEC and DER Earning Class Differences 2005: No Imputes



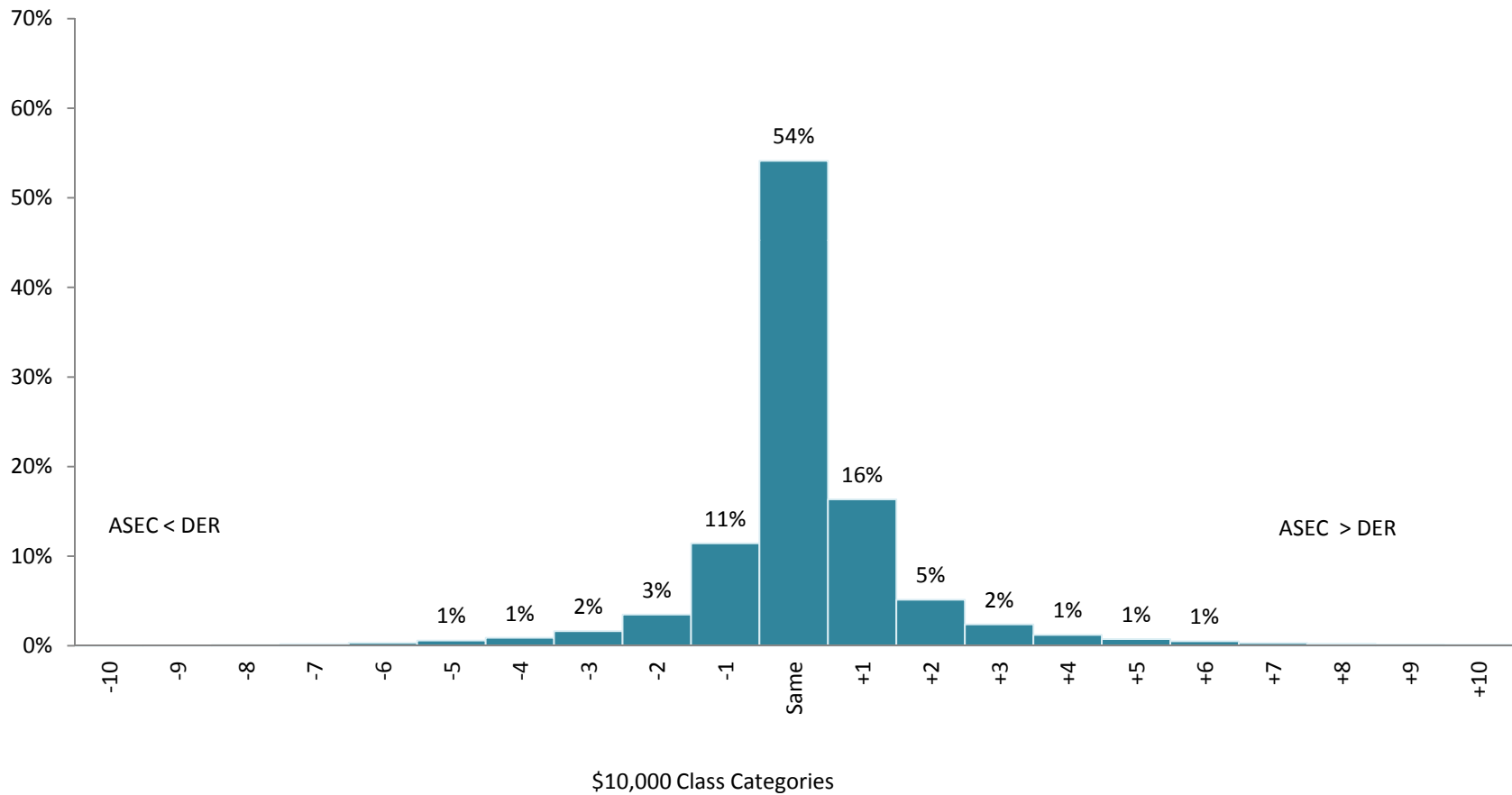
# Imputed Earners: Distribution of ASEC Earning Class Differences

## 2005: Whole and Item Imputes



# Wage Earners: Distribution of ASEC and DER Wage Class Differences

## 2005: All



# Schematic for Diagonal Charts

CPS-ASEC Reported Earnings	Matched DERS Earnings											
	\$1 to \$9999	\$10,000 to \$19,999	\$20,000 to \$29,999	\$30,000 to \$39,999	\$40,000 to \$49,999	\$50,000 to \$59,999	\$60,000 to \$69,999	\$70,000 to \$79,999	\$80,000 to \$89,999	\$90,000 to \$99,999	\$100,000 or more	Sum of Diagonals
\$1 to \$999	Same	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	
\$10,000 to \$19,999	+1	Same	-1	-2	-3	-4	-5	-6	-7	-8	-9	Sum -10
\$20,000 to \$29,999	+2	+1	Same	-1	-2	-3	-4	-5	-6	-7	-8	Sum -9
\$30,000 to \$39,999	+3	+2	+1	Same	-1	-2	-3	-4	-5	-6	-7	Sum -8
\$40,000 to \$49,999	+4	+3	+2	+1	Same	-1	-2	-3	-4	-5	-6	Sum -7
\$50,000 to \$59,999	+5	+4	+3	+2	+1	Same	-1	-2	-3	-4	-5	Sum -6
\$60,000 to \$69,999	+6	+5	+4	+3	+2	+1	Same	-1	-2	-3	-4	Sum -5
\$70,000 to \$79,999	+7	+6	+5	+4	+3	+2	+1	Same	-1	-2	-3	Sum -4
\$80,000 to \$89,999	+8	+7	+6	+5	+4	+3	+2	+1	Same	-1	-2	Sum -3
\$90,000 to \$99,999	+9	+8	+7	+6	+5	+4	+3	+2	+1	Same	-1	Sum -2
\$100,000 or more	+10	+9	+8	+7	+6	+5	+4	+3	+2	+1	Same	Sum -1
Sum of Diagonals		Sum +10	Sum +9	Sum +8	Sum +7	Sum +6	Sum +5	Sum +4	Sum +3	Sum +2	Sum +1	Sum Same

# **Many, Many Thanks**

**Play it Again Sam?  
Scheuren@aol.com  
+1-202-320-3446**

# Size of Sampling

- Unavoidable Linkage Error can be minor if samples are not too small provided MSE is calculated properly and includes response and nonresponse errors