

# 2009 German Stata Users Group Meeting



## Performing within and between analysis (WABA) in Stata

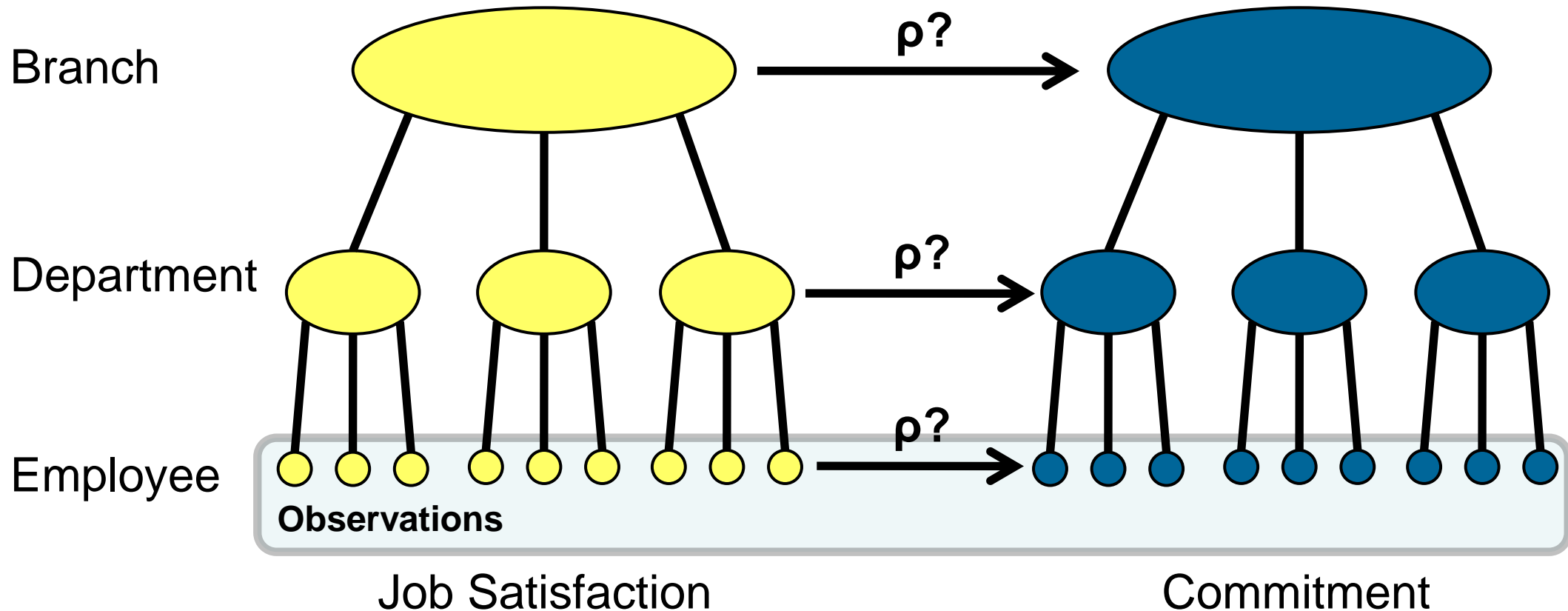
Sven-Oliver Spieß



- **Introduction**  
**Where are constructs in hierarchical data really associated?**
- **Within and between analysis**
  - The basic idea
  - Simple ANOVA
  - Partitioning of correlations
  - Graphical demonstration
- **wabacorr.ado in Stata**

- **Many names for one common problem:**  
Fallacy of composition, ecological fallacy, atomistic fallacy, individualistic fallacy, Simpson's paradox, ...  
→ **Fallacies of the wrong level**
- **Therefore global correlations potentially misleading**
  - E.g.  $\text{Corr}(\text{Job Satisfaction, Commitment}) = .72$
  - But at which level is the association?  
Individuals? Work groups? Departments?
- **Particularly problematic in applied settings**
  - No simple random samples
  - **Interventions**

- $r_{\text{Job Satisfaction, Commitment}} = .72$



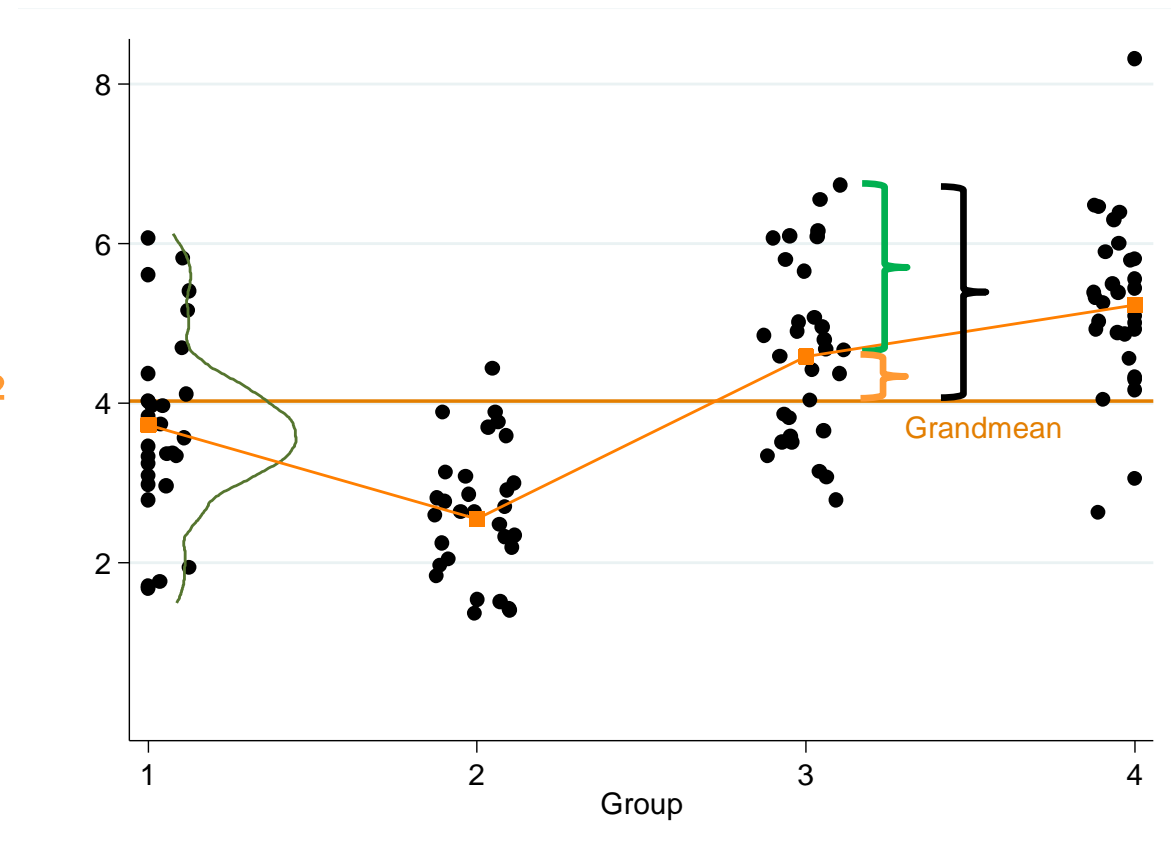
- At which level is the association?

- **Introduction**  
**Where are constructs in hierarchical data really associated?**
- **Within and between analysis**
  - The basic idea
  - Simple ANOVA
  - Partitioning of correlations
  - Graphical illustration
- **wabacorr.ado in Stata**

- **The idea:**  
Let's split up the total correlation into a component **within** the groups and another component **between** the groups  
  
→ similar to idea behind analysis of variance (ANOVA)
- **Simply needs to be adjusted to correlations**
- **Data prerequisites:**  
variables in question must be metric and levels must be nested

- $X_{ij} = (X_{ij} - \mu_{\cdot j}) + \mu_{\cdot j}$
- $SS_{\text{Total}} = SS_{\text{Error}} + SS_{\text{Group}}$   
 $(X_{ij} - \mu_{\cdot\cdot})^2 = (X_{ij} - \mu_{\cdot j})^2 + (\mu_{\cdot j} - \mu_{\cdot\cdot})^2$
- $\eta^2 = SS_{\text{Group}} / SS_{\text{Total}}$

==>  $\eta^2$  **between** measure



- Adjusted to correlations:

$$r_{xy} = \eta_{Bx} * \eta_{By} * r_{Bxy} + \eta_{Wx} * \eta_{Wy} * r_{Wxy} \quad \Leftrightarrow$$
$$r_{xy} = C_B + C_W$$

- $\eta_B = \text{corr}(\mu_{\cdot j}, X_{ij})$
  - $\eta_W = \text{corr}[(X_{ij} - \mu_{\cdot j}), X_{ij}]$
- Central question: is **between** or **within** component (i.e. **higher** or **lower** level, or both) of total correlation more important?

- **Typical procedure in 3 steps:**

1. Univariate comparison of the within and between variances

$$r_{xy} = \eta_{Bx} * \eta_{By} * r_{Bxy} + \eta_{Wx} * \eta_{Wy} * r_{Wxy}$$

2. Bivariate comparison of the within and between correlations

$$r_{xy} = \eta_{Bx} * \eta_{By} * r_{Bxy} + \eta_{Wx} * \eta_{Wy} * r_{Wxy}$$

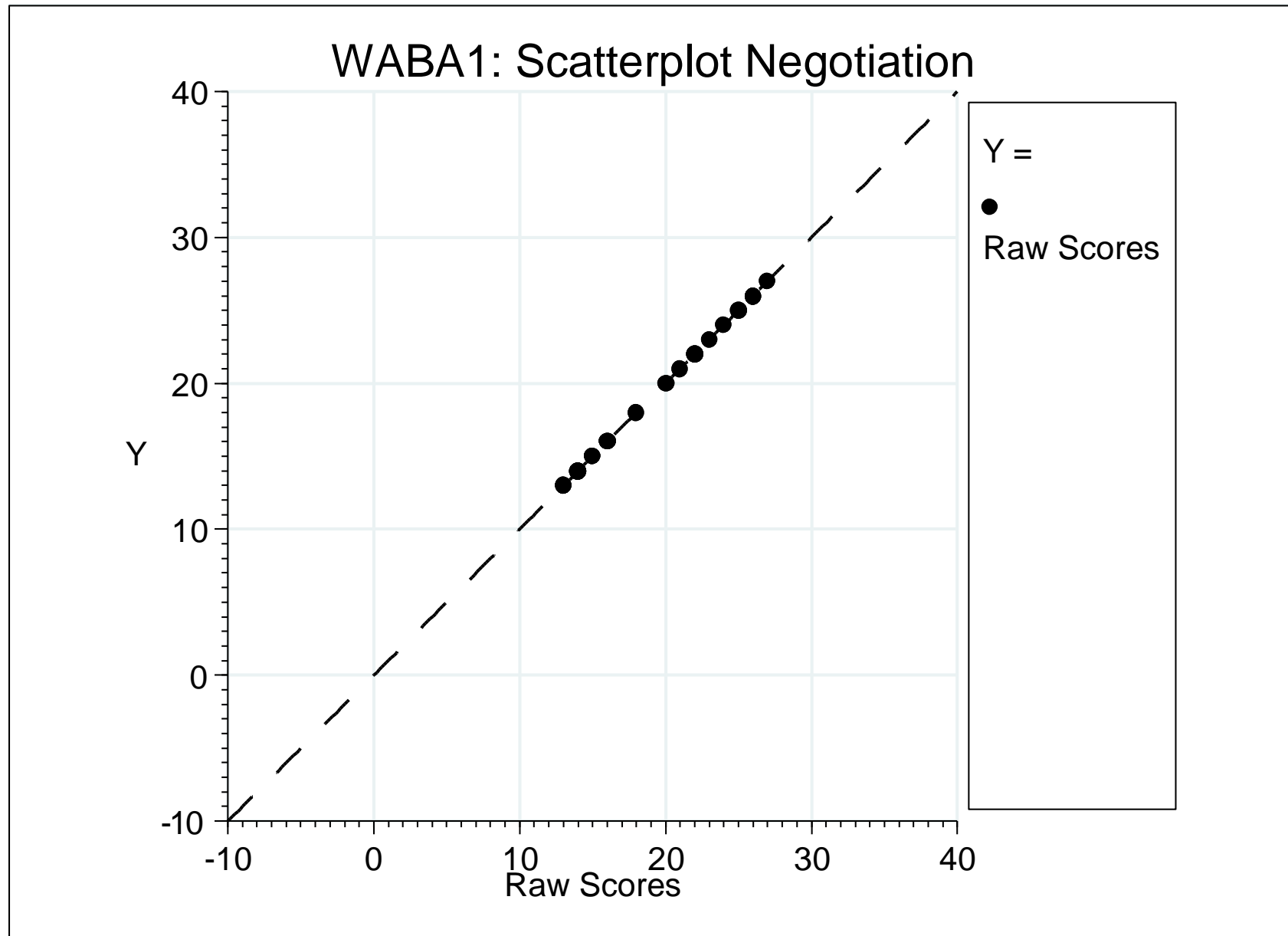
3. Summary judgment on the importance of the within and between components for the total correlation

$$r_{xy} = C_B + C_W$$

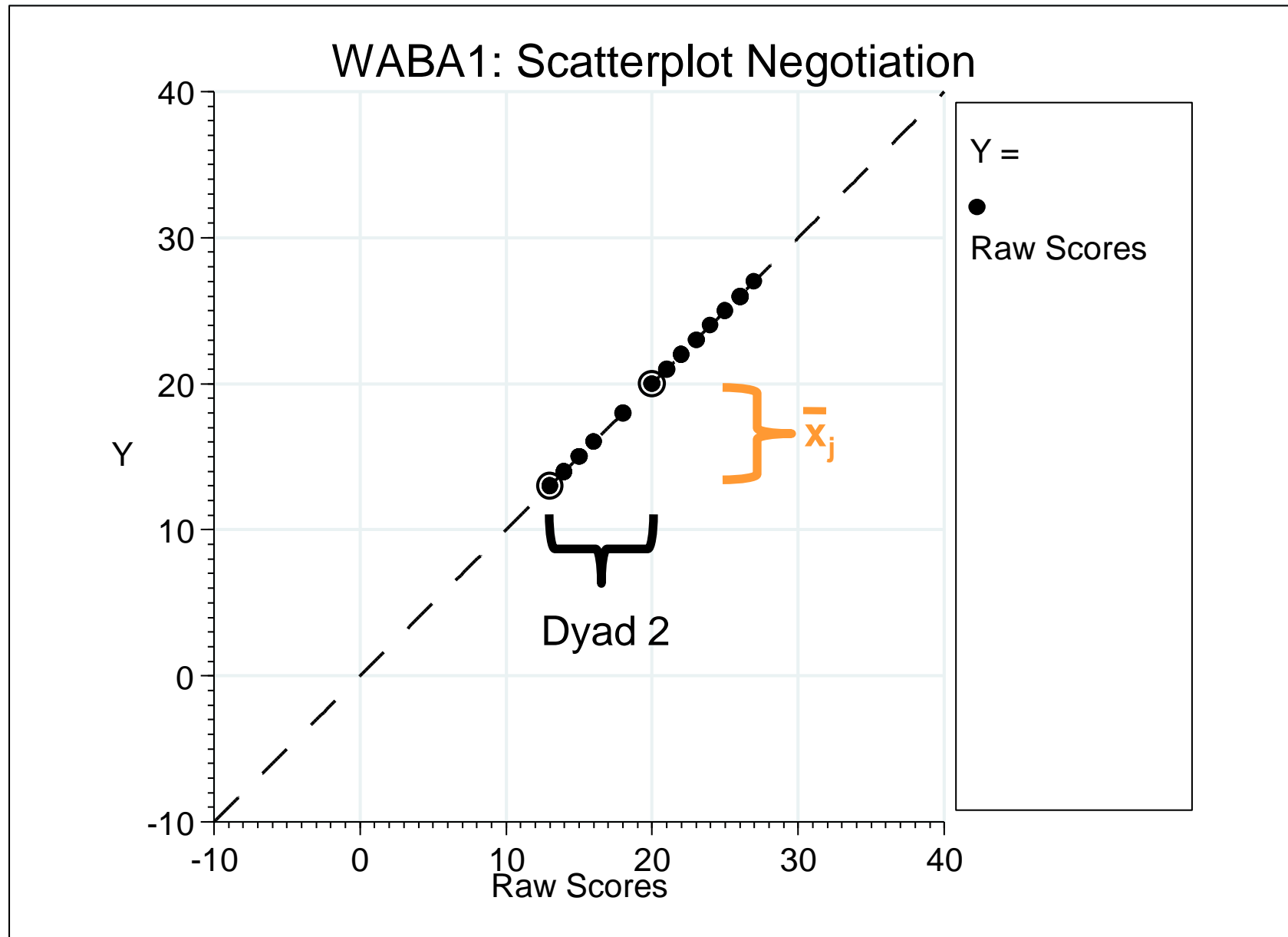
## 4 possible outcomes/inductions:

1. **Parts** → lower level/within
2. **Wholes** → higher level/between
3. **Equivocal** → meaningful association at both levels
4. **Inexplicable** → noise

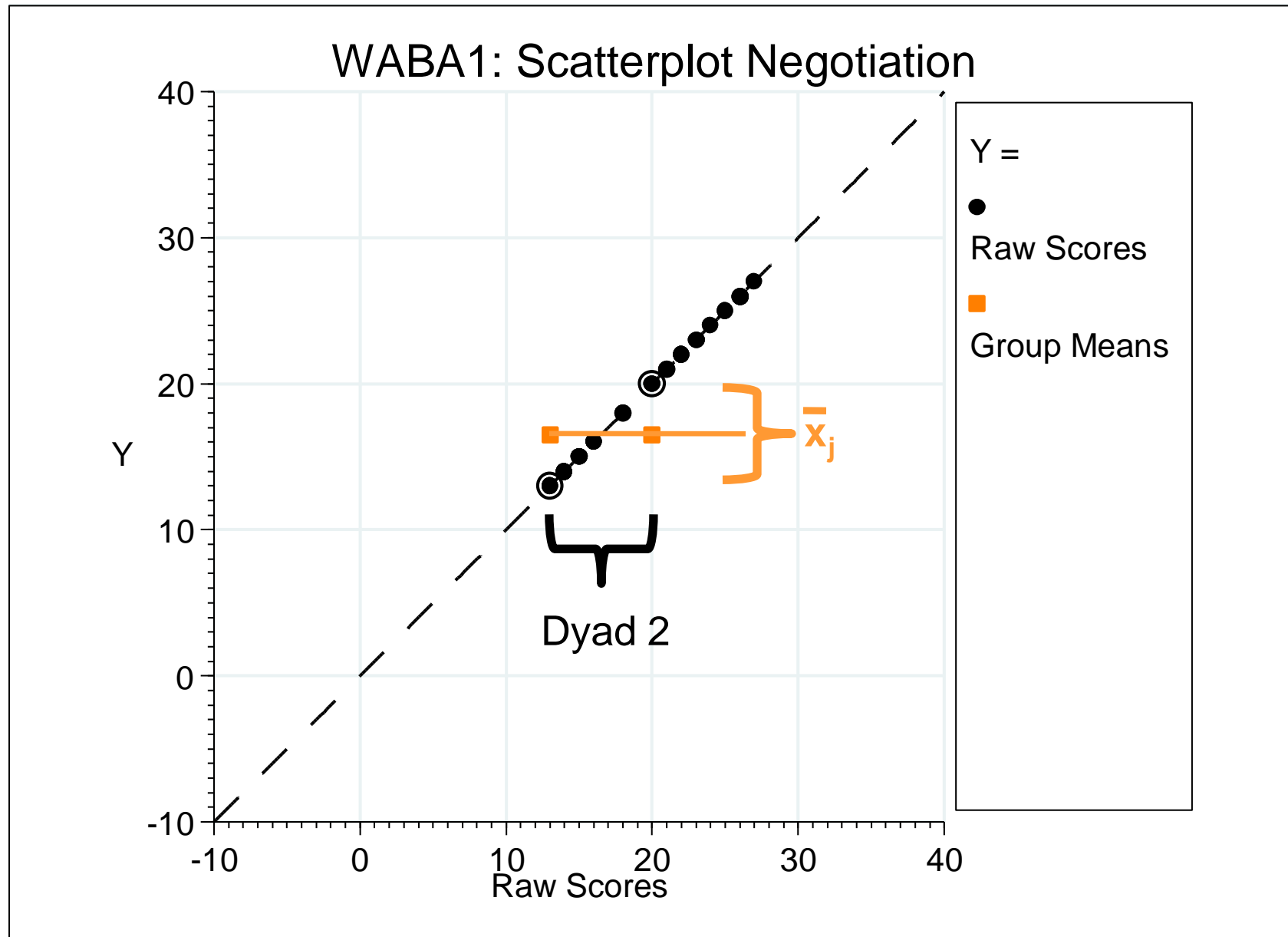
# Graphical illustration: Step 1



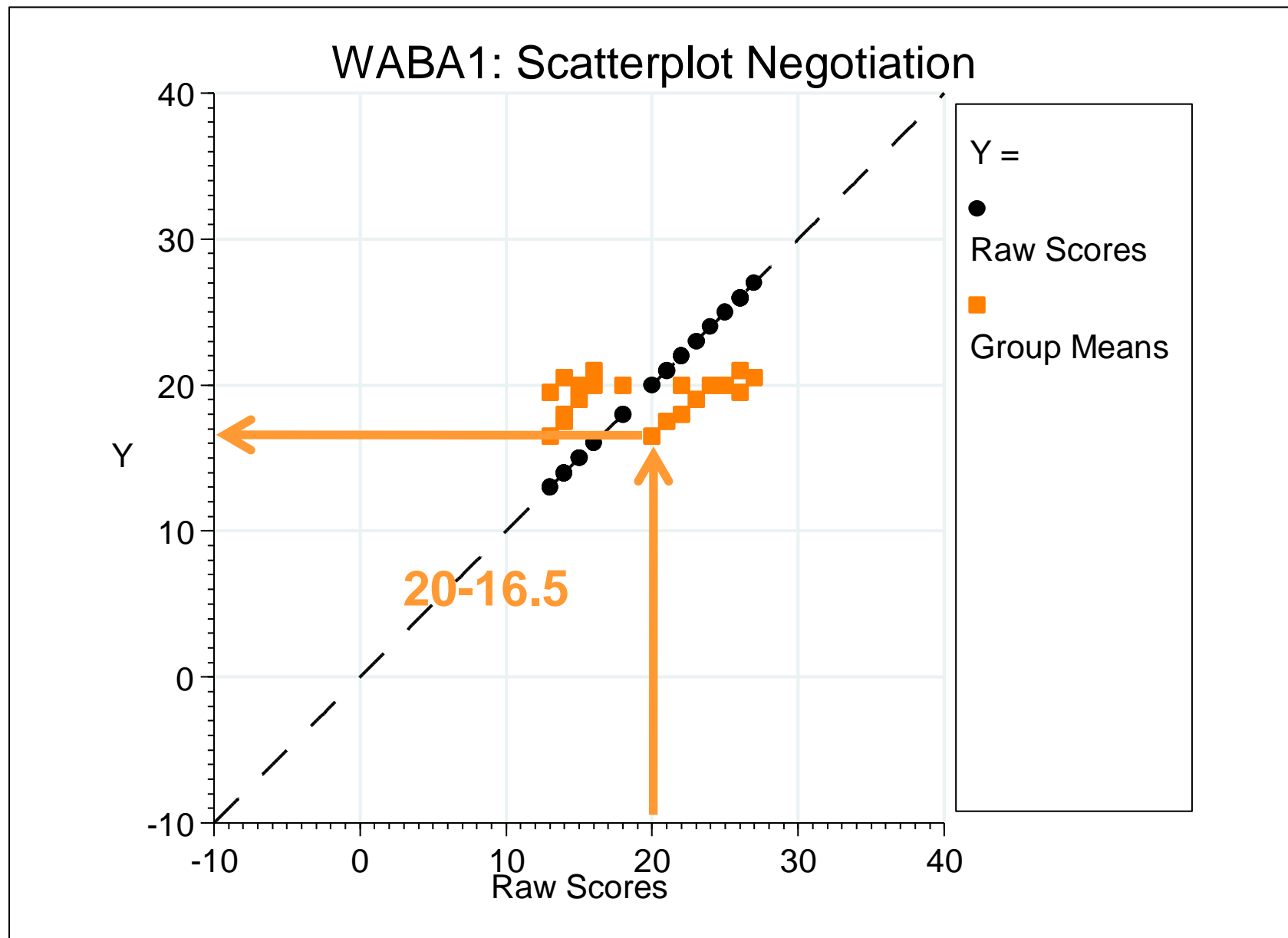
# Graphical illustration: Step 1 ( $\eta_B$ )



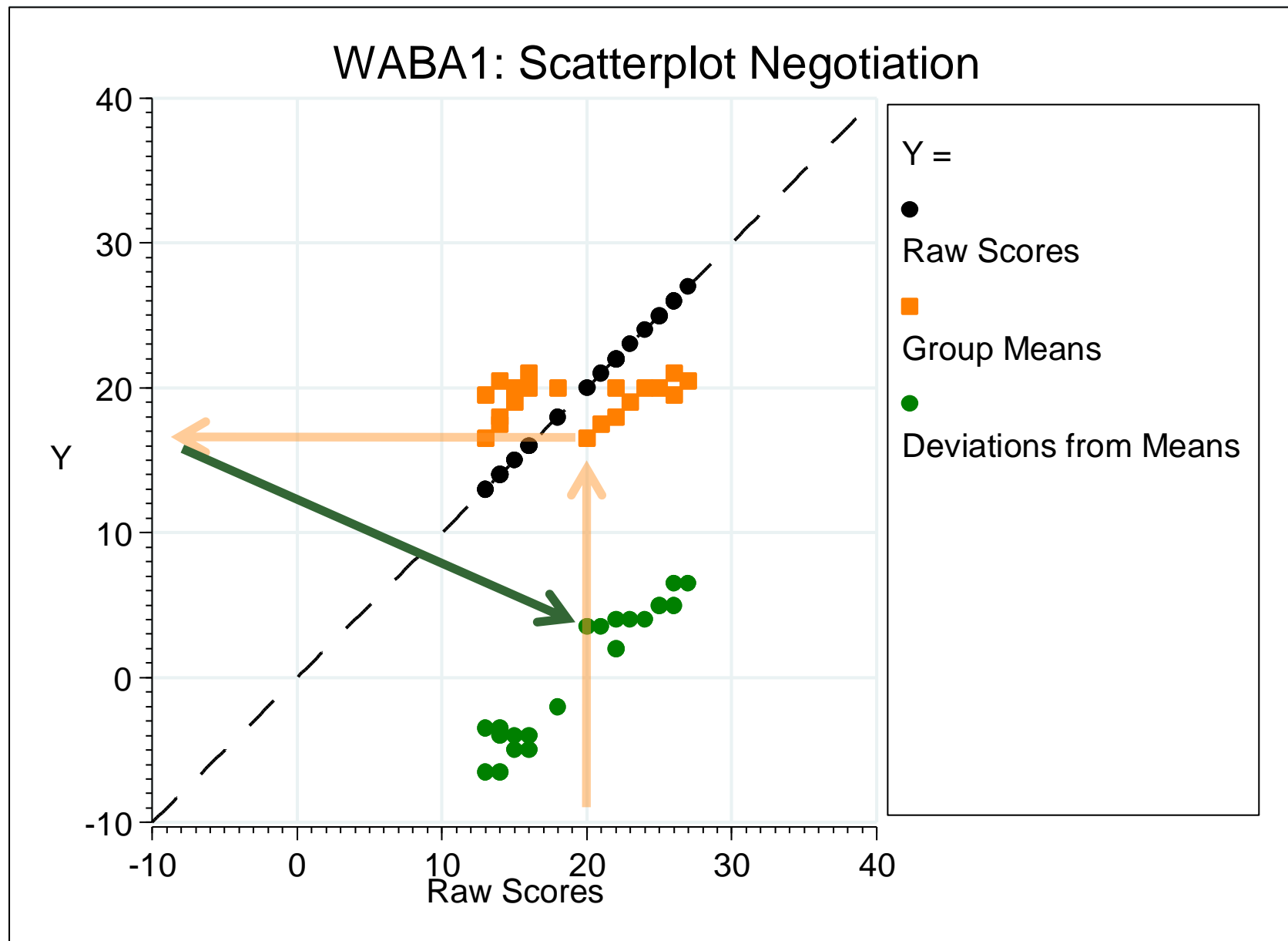
# Graphical illustration: Step 1 ( $\eta_B$ )



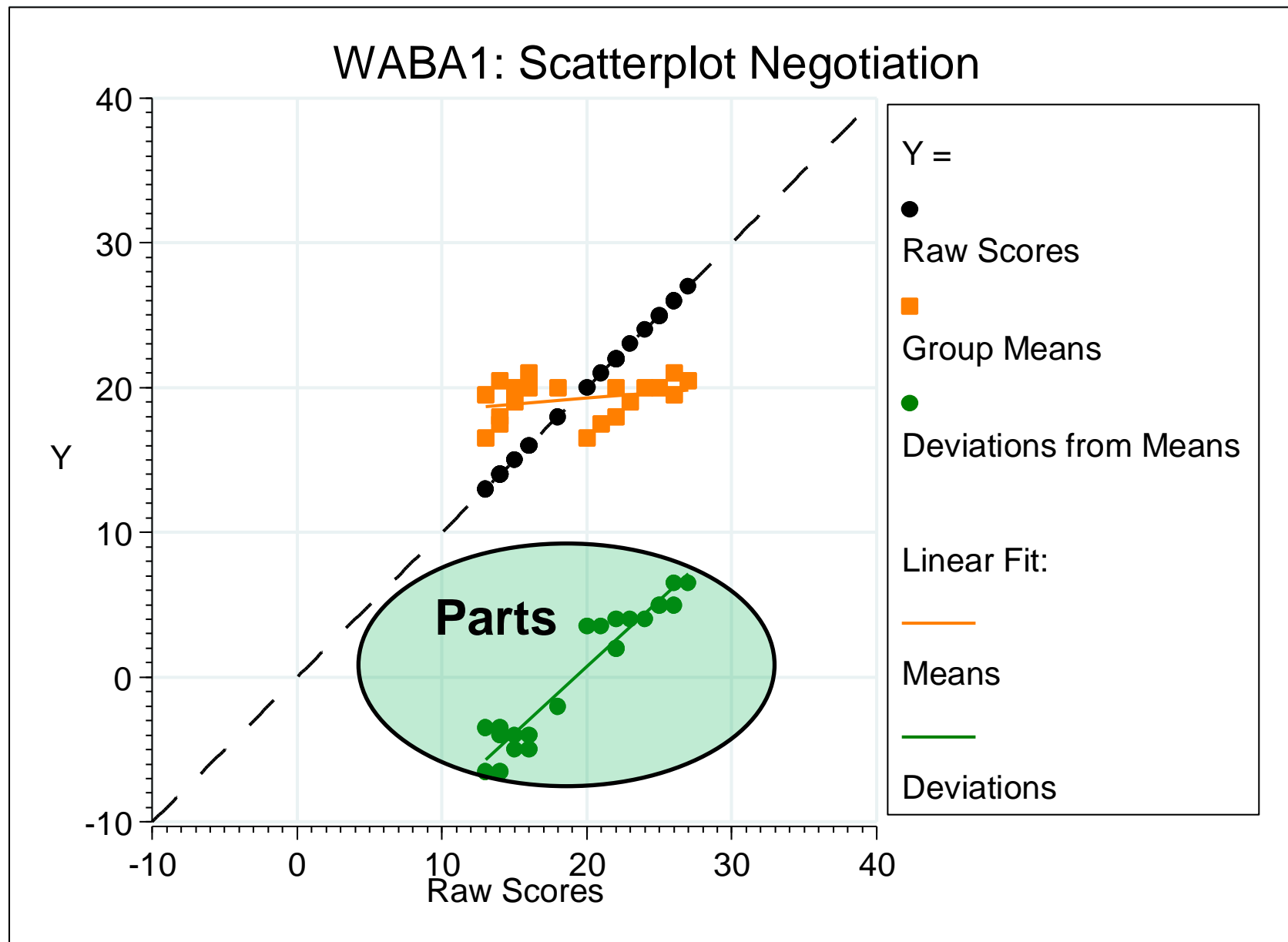
# Graphical illustration: Step 1 ( $\eta_B$ )



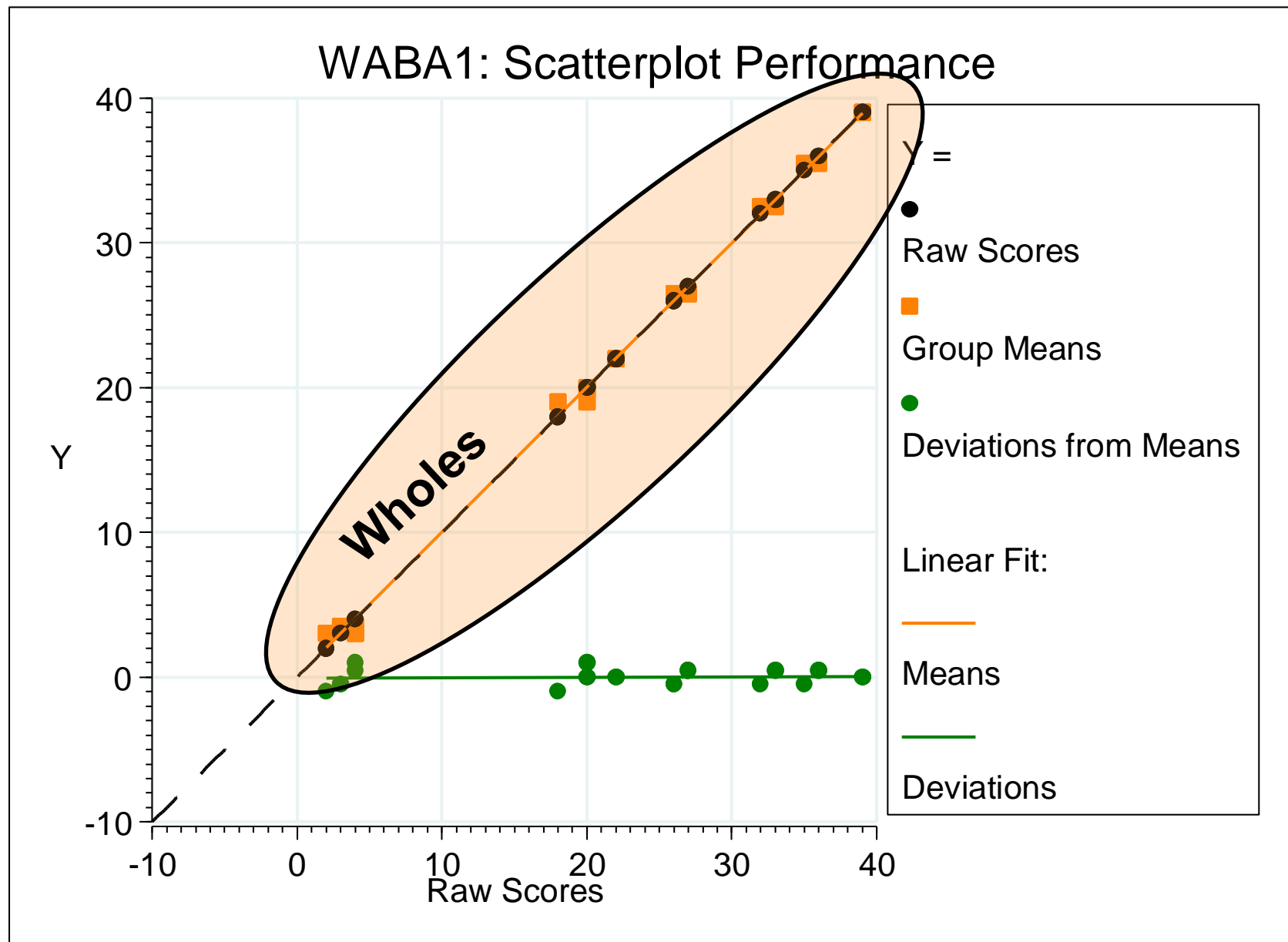
# Graphical illustration: Step 1 ( $\eta_w$ )

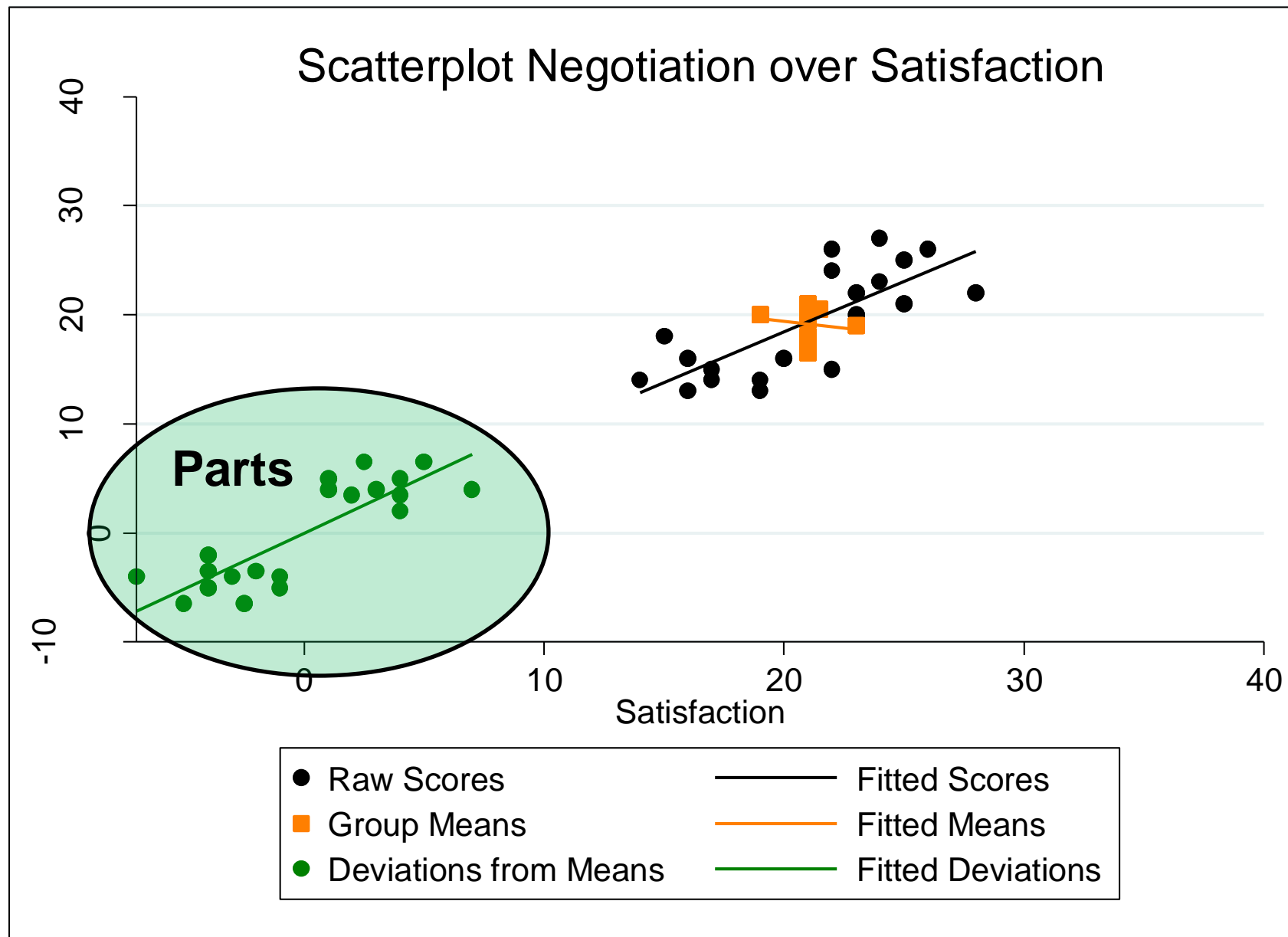


# Graphical illustration: Step 1

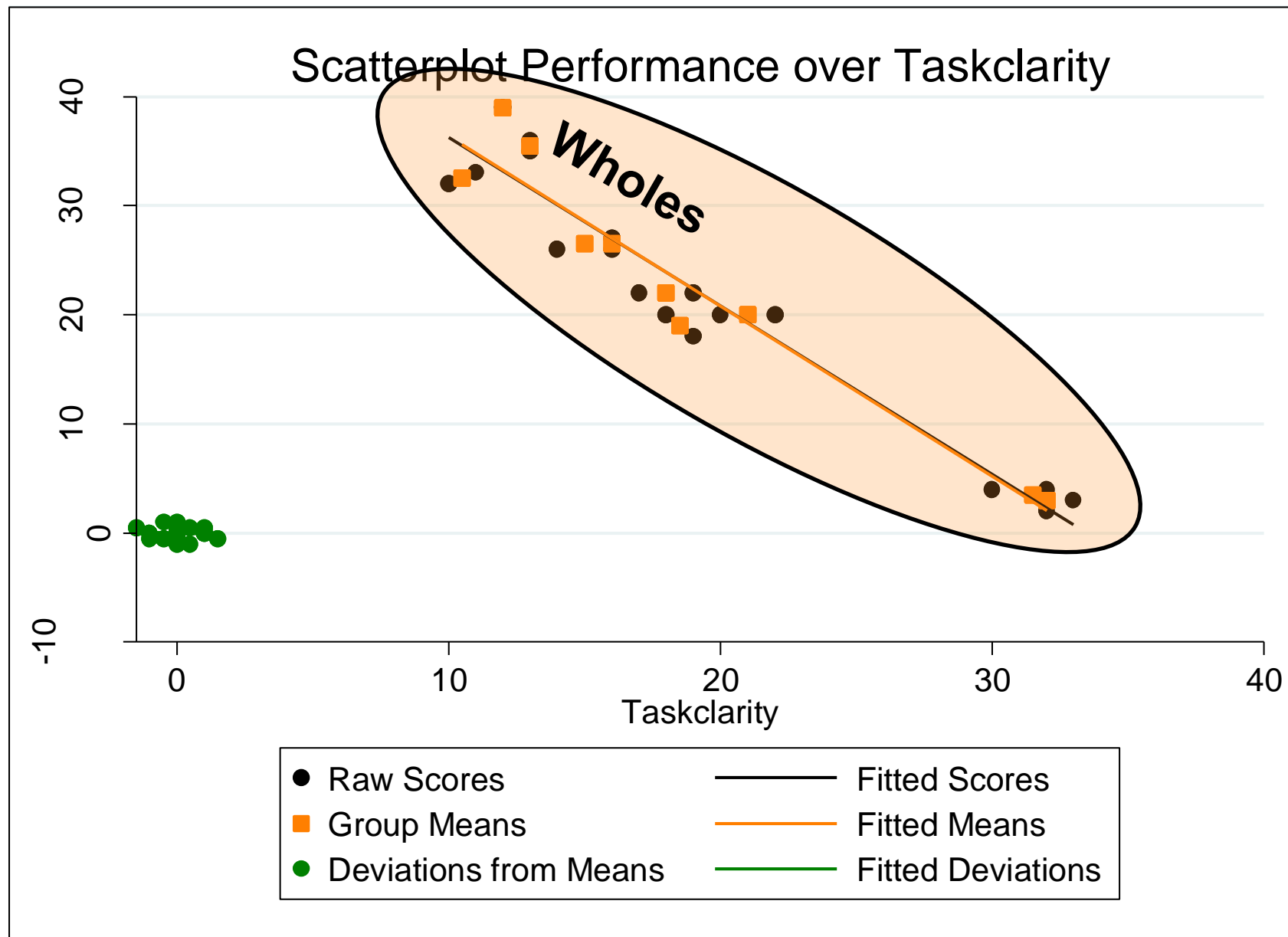


# Graphical illustration: Step 1





# Graphical illustration: Step 2



- **Introduction**  
**Where are constructs in hierarchical data really associated?**
- **Within and between analysis**
  - The basic idea
  - Simple ANOVA
  - Partitioning of correlations
  - Graphical illustration
- **wabacorr.ado in Stata**

- **General syntax:**

```
wabacorr varlist [if] [in] [fweight], by(grpvar) [detail]
```

- **Examples based on Detect Data set A**

- 40 persons in 20 dyads in 10 groups in 4 collectivities
- 4 metric variables: negotiation, satisfaction, performance, taskclarity

```
. wabacorr negotiation satisfaction performance taskclarity, by(dyad)
```

Within and between analysis

Group variable: dyad

Number of obs = 40

Number of groups = 20

Obs per group: min = 2

avg = 2.0

max = 2

Within- and between-groups Etas and Eta-squared values:

Variable	Eta-betw	Eta-with	Eta-b <sup>2</sup>	Eta-w <sup>2</sup>	F	p>F
negotiation	0.2846	0.9586	0.0810	0.9190	10.7769	0.0000
satisfaction	0.2783	0.9605	0.0774	0.9226	11.3170	0.0000
performance	0.9988	0.0493	0.9976	0.0024	431.6194	0.0000
taskclarity	0.9944	0.1054	0.9889	0.0111	93.7529	0.0000

⇒ Parts

Within- and between-groups correlations:

Variables	r-betw	r-with	z'	p>z'	
negotiation-satisfaction	-0.1973	0.8441	-3.0614	0.0011	⇒ Parts
negotiation-performance	0.1413	-0.0477	0.2794	0.3900	⇒ Noise
negotiation-taskclarity	-0.0589	0.0502	0.0257	0.4897	
satisfaction-performance	-0.0346	0.0695	-0.1037	0.4587	
satisfaction-taskclarity	0.0526	0.1568	-0.3119	0.3776	⇒ Wholes
performance-taskclarity	-0.9679	-0.1157	5.7429	0.0000	

Total correlation and components:

Variables	r-total	betw-comp	with-comp	z'	p> z'
negotiation-satisfaction	0.7616	-0.0156	0.7772	-3.0239	0.0025
negotiation-performance	0.0379	0.0402	-0.0023	0.1122	0.9107
negotiation-taskclarity	-0.0116	-0.0167	0.0051	0.0343	0.9726
satisfaction-performance	-0.0063	-0.0096	0.0033	0.0187	0.9851
satisfaction-taskclarity	0.0304	0.0145	0.0159	-0.0039	0.9969
performance-taskclarity	-0.9620	-0.9614	-0.0006	5.8045	0.0000

⇒ Parts

- Induction for the correlation between negotiation and satisfaction is **parts**
- Thus variables should not be aggregated, but higher level information could be disregarded without a big loss

- What if induction is **wholes** (as with performance and taskclarity) or **equivocal**?
- If possible repeat WABA at the next higher level until induction is **parts**
  - New number of cases N equals the number of groups M during the previous analysis
  - Input/initial values are correspondingly the means  $\mu_j$  of the previous analysis
  - This is called *multiple WABA*
  - In *unbalanced data* the means must be *weighted* to avoid distortions (`wabacorr` supports frequency weights)
  - Aggregate data no higher than level of first **parts** induction, but do not disregard levels where inductions were **equivocal**
- **Stata again:**

```
. collapse (mean) performance taskclarity group collectivity  
  (count) obs=performance, by(dyad)
```

```
. wabacorr performance taskclarity [fweight=obs], by(group)
```

Within and between analysis

Group variable: group

Number of obs = 20

Number of groups = 10

Obs per group: min = 2

avg = 2.0

max = 2

Number of weighted obs = 40

Weighted obs per group: min = 4

avg = 4.0

max = 4

:::

Output omitted

:::

```
. collapse (mean) performance taskclarity collectivity  
  (rawsum) obs [fweight=obs], by(group)
```

```
. wabacorr performance taskclarity [fweight=obs], by(collectivity)
```

```
. wabacorr performance taskclarity [fweight=obs], by(collectivity)
```

```
Within and between analysis          Number of obs      =          10
Group variable: collectivity         Number of groups   =           4

                                     Obs per group: min =           2
                                     avg =           2.5
                                     max =           3

Number of weighted obs =           40      Weighted obs per group: min =           8
                                     avg =          10.0
                                     max =           12
```

```
:::
Output omitted
:::
```

- Induction remains **wholes** even at the highest level
- Data could thus be aggregated by collectivities

# Example for an Analysis: Dansereau et al. (2006)

**TABLE 3**  
**Hypothesis 2: Illustration of Group Parts Result for**  
**Trustworthiness (x) and Delegated Activities (y)**

Number of persons	40	
Number of groups	20	
Total correlation (individual level; from Table 1)	.76**	→ Initial value
<b>Between-group model</b>		
Between-group correlation ( $r_{BY}$ )	-.197	→ Step 2
<b>Between-group variation</b>		
Trustworthiness ( $\eta_{BX}$ )	.285	→ Step 1
Delegated activities ( $\eta_{BY}$ )	.278	→ Step 1
<b>Between-group component</b>		
$(r_{BY})(\eta_{BX})(\eta_{BY})$	$(-.197)(.285)(.278) = -.02$	→ Step 3
<b>Within-group model</b>		
Within-group correlation ( $r_{WXY}$ )	.844**	→ Step 2
<b>Within-group variation</b>		
Trustworthiness ( $\eta_{WX}$ )	.959	→ Step 1
Delegated activities ( $\eta_{WY}$ )	.961	→ Step 1
<b>Within-group component</b>		
$(r_{WXY})(\eta_{WX})(\eta_{WY})$	$(.844)(.959)(.961) = .78$	→ Step 3
<b>Differences</b>		
<b>Between correlation versus within correlation</b>		
A test	-.81**	→ Step 2
Z test	-3.06**	→ Step 2
<b>Trustworthiness</b>		
<b>Between variation versus within variation</b>		
E ratio = $\eta_B/\eta_W$	.30**	→ Step 1
F ratio = $(1/E^2)(J-1)/(N-J)$	10.78**	→ Step 1
<b>Delegated activities</b>		
<b>Between variation versus within variation</b>		
E ratio = $\eta_B/\eta_W$	.29**	→ Step 1
F ratio = $(1/E^2)(J-1)/N-J$	11.32**	→ Step 1
Induction	Group parts	

<sup>†</sup> $\theta > 15^\circ$ . <sup>\*\*</sup> $\theta > 30^\circ$ . \* $p < .05$ . \*\* $p < .01$ .

- **Within and between analysis**
  - provides a detailed picture of patterns of associations between variables at different levels in nested hierarchical data instead of an all-or-nothing decision as with ANOVA or intra-class correlations (ICC)
  - has its greatest added value in equivocal cases
  - can reveal important results even if total correlation is nil
  - can be employed at two levels (single WABA) or successively at more levels (multiple WABA)
  - can also be employed in multivariate contexts like regression analysis (cf. Dansereau et al. (2006))
  - can inform further analyses, like the choice of levels in multi level modeling (MLM), and selection of starting points for interventions

- **wabacorr . ado**
  - performs WABA of correlations in Stata 9.2 or higher
  - also provides tests of practical significances with 'detail' option
  - supports frequency weights to allow multiple WABA with unbalanced data
  - stores results for further use by the user

- **Method:**

- Dansereau, F., Cho, J. and Francis J. Yammarino. (2006). Avoiding the "Fallacy of the Wrong Level": A Within and Between Analysis (WABA) Approach. *Group & Organization Management*, 31, 536 - 577.
- O'Connor, B. P. (2004). SPSS and SAS programs for addressing interdependence and basic levels-of-analysis issues in psychological data. *Behavior Research Methods, Instrumentation, and Computers*, 36 (1), 17-28.
- Detect software: <http://www.levelsofanalysis.com>

- **wabacorr . ado:**

- <http://www.wip-mannheim.de/>
- <http://www.svenoliverspiess.net/stata>
- Soon: Statistical Software Components

Thank you!

