

Rolf Steyer
Ulf Kröhne

Probability and Causality

Version: June 13, 2016

University of Jena

Contents

Part I Introduction

1	Introductory Examples	3
1.1	Example 1 — Simpson's Paradox	3
1.1.1	Prima Facie Effect	4
1.1.2	Prima Facie Effects Controlling for Sex	5
1.1.3	Prima Facie Effect vs. Average of the Prima Facie Effects ...	6
1.1.4	How to Evaluate the Treatment?	8
1.2	Example 2 — Nonorthogonal Two-Factorial Experiment	8
1.2.1	Prima Facie Effects	9
1.2.2	Prima Facie Effects Controlling for Neediness	10
1.2.3	Prima Facie Effects vs. Average of the Prima Facie Effects .	11
1.2.4	How to Evaluate the Treatment?	12
1.3	Example 3 — Direct Effect in a Randomized Experiment	12
1.3.1	Conditional Expectation of Y Given Treatment and Intermediate Variables	12
1.3.2	Conditional Expectation of Y Given Treatment, Intermediate, and Pre-Test Variables	14
1.3.3	Conditional Expectation of Y Given Treatment Variable ...	15
1.3.4	How to Analyze Direct Effects?	15
1.4	Summary and Conclusions	16
1.5	Exercises	19
2	Some Typical Random Experiments	25
2.1	Simple Experiments	26
2.1.1	Sampling a Unit	27
2.1.2	Treatment Variable	28
2.1.3	Covariates	28
2.1.4	Outcome Variable	30
2.1.5	Causal Effects and Causal Dependencies	30
2.2	Experiments With Fallible Covariates	31
2.3	Two-Factorial Experiments	34
2.4	Multilevel Experiments	36
2.5	Experiments With Intermediate Variables	37
2.6	Experiments With Latent Outcome Variables	39
2.7	Summary and Conclusions	40

2.8	Exercises	42
Part II Basic Concepts		
3	Causality Space	47
3.1	Filtration	48
3.2	Priority Relation	53
3.3	Simultaneity Relation	57
3.4	Causality Space	60
3.5	Summary and Conclusions	61
3.6	Proofs	64
3.7	Exercises	65
4	True-Outcome Variables and Atomic Effect Variables	69
4.1	Potential-Confounder σ -Algebra and Covariates	69
4.1.1	Examples	71
4.1.2	Intermediate Variable	74
4.2	True-Outcome Variable and Atomic Effect Variables	75
4.2.1	Conditional Expectation With Respect to $P^{X=x}$	76
4.2.2	Total-Effect True-Outcome Variables	77
4.2.3	Atomic Total-Effect Variable	77
4.2.4	Direct-Effect True-Outcome Variables	78
4.3	Numerical Examples	80
4.3.1	Joe and Ann With Random Assignment	80
4.3.2	No Treatment for Joe	84
4.3.3	Jim and Jane	86
4.4	Summary and Conclusions	92
4.5	Proofs	94
4.6	Exercises	96
5	Causal Effects	101
5.1	Average Total Effect	102
5.1.1	Numerical Example	103
5.2	Conditional Total Effect	103
5.2.1	Conditional Total Effects given a Covariate	105
5.2.2	Individual Total Effects	106
5.2.3	Conditional Total Effects Given a Value of X	107
5.2.4	Conditional Total Effects Given Values of X and Z	110
5.3	Average and Conditional Direct and Indirect Effects	111
5.4	Summary and Conclusions	115
5.5	Exercises	118

6	Unbiasedness	121
6.1	Unbiasedness With Respect to Total Effects	121
6.1.1	τ_x -Unbiasedness of Conditional Expectations	122
6.1.2	$\delta_{xx'}$ -Unbiasedness of Prima Facie Effects	123
6.2	Numerical Examples	125
6.2.1	Assumptions in all Examples	125
6.2.2	Description of the Examples	127
6.2.3	Average Total Effect	128
6.2.4	Conditional Total Effect	131
6.2.5	Computing Average Total Effect From Conditional Total Effects	131
6.2.6	First Conclusions	132
6.3	Bias With Respect to Total Effects	132
6.3.1	Theory	132
6.3.2	Numerical Examples	134
6.3.3	Baseline Bias and Effect Bias	139
6.3.4	Numerical Examples	141
6.3.5	Another Example	142
6.4	Unbiasedness With Respect to Direct Effects	144
6.4.1	$\tau_{x,t}$ -Unbiasedness of Conditional Expectations	144
6.4.2	$\delta_{xx',t}$ -Unbiasedness of Prima Facie Effects	145
6.5	Summary and Conclusions	148
6.6	Proofs	151
6.7	Exercises	152
7	Independent Cause and Mean-Independent Outcome	159
7.1	Independent Cause Conditions	160
7.1.1	Independence and Conditional Independence	160
7.1.2	Independent Cause Conditions for Total Effects	160
7.1.3	Independent Cause Conditions for Direct Effects	161
7.1.4	Falsifiability of the Independent Cause Conditions	161
7.2	Mean-Independent Outcome Conditions	162
7.2.1	Conditional Mean Independence	162
7.2.2	Mean-Independent Outcome Conditions for Total Effects	162
7.2.3	Mean-Independent Outcome Conditions for Direct Effects	163
7.2.4	Falsifiability of the Mean-Independent Outcome Conditions	164
7.3	Implications on Unbiasedness	165
7.3.1	Unbiasedness With Respect to Total Effects: No Covariates	165
7.3.2	Conditioning on a Covariate	167
7.3.3	Unbiasedness With Respect to Direct Effects	169
7.4	Examples	171
7.4.1	Examples for the Independent Cause Conditions	171
7.4.2	Examples for the Mean-Independent Outcome Conditions	175
7.5	Methodological Implications	177

7.6	Proofs	181
7.7	Exercises	185
8	Unconfoundedness	189
8.1	Unconfoundedness of Regressions	190
8.1.1	Unconfoundedness With Respect to Total Effects	190
8.1.2	Unconfoundedness With Respect to Direct Effects	192
8.2	Conditions Implying Unconfoundedness	193
8.3	Numerical Example	194
8.4	Implications of Unconfoundedness on Unbiasedness	197
8.4.1	Unbiasedness With Respect to Total Effects	197
8.4.2	Unbiasedness With Respect to Direct Effects	201
8.5	Implications Between Causality Conditions	203
8.6	Summary and Conclusions	204
8.7	Proofs	207
8.8	Exercises	209
9	Other Causality Conditions	215
9.0.1	Strong Ignorability With Respect to Total Effects	215
9.0.2	Strong Ignorability With Respect to Direct Effects	216
9.0.3	Weak Ignorability With Respect to Total Effects	216
9.0.4	Weak Ignorability With Respect to Direct Effects	217
9.1	Independence of X and True Outcomes	218
9.1.1	Theory	218
9.1.2	Substantive Meaning	220
9.1.3	Numerical Example	221
9.2	Regressive Independence	224
9.2.1	Theory	224
9.2.2	Substantive Meaning	227
9.2.3	Numerical Example	227
9.3	Implications Between Causality Conditions	232
9.4	Summary and Conclusions	233
9.5	Proofs	233
9.6	Exercises	236
10	Identification of Causal Effects and Effect Functions	241
10.1	Identification of Total Effects	241
10.1.1	Theory	242
10.1.2	Methodological Implications	244
10.1.3	Numerical Example	247
10.2	Identification of Direct Effects	250
10.2.1	Theory	251
10.2.2	Methodological Implications	253
10.3	Summary and Conclusions	256
10.4	Proofs	258
10.5	Exercises	261

11 Propensities	263
11.1 True Propensities for Total Effects	263
11.1.1 Theory	264
11.1.2 Methodological Implications	266
11.1.3 Numerical Example	267
11.2 Conditional Propensities for Total Effects	269
11.2.1 Theory	270
11.2.2 Methodological Implications	272
11.2.3 Numerical Examples	275
11.3 Conditional Propensities for Direct Effects	278
11.3.1 Theory	278
11.3.2 Methodological Implications	281
11.4 Weighting the Outcome Variable	283
11.4.1 Adjusting for Total Effects by Weighting the Outcome Variable	283
11.4.2 Theory	283
11.4.3 Substantive Meaning	285
11.4.4 Numerical Example	286
11.4.5 Adjusting for Direct Effects by Weighting the Outcome Variable	286
11.4.6 Theory	287
11.4.7 Substantive Meaning	288
11.5 Summary and Conclusions	289
11.6 Proofs	290
11.7 Exercises	293
12 Analysis of Change Scores	297
12.1 Theory	297
12.2 Numerical Examples	301
12.3 Summary and Conclusions	304
12.4 Proofs	305
12.5 Exercises	307
13 Analysis of Covariance and its Generalizations	307
13.1 Analysis of Covariance (ANCOVA)	307
13.1.1 Theory	307
13.1.2 Numerical Examples	310
13.1.3 Conclusions	312
13.1.4 Statistical Programs	312
13.2 Generalized ANCOVA	313
13.2.1 Theory	313
13.2.2 Numerical Examples	316
13.2.3 Conclusions	317
13.2.4 Statistical Programs	318
13.3 Generalized ANCOVA With Latent Variables	318

13.3.1 Theory	318
13.3.2 Conclusions	325
13.3.3 Statistical Programs	326
13.3.4 Conclusions	326
13.3.5 Statistical Programs	326
13.4 Summary and Conclusions	327
13.5 Proofs	327
13.6 Exercises	328
References	331

List of Figures

1.1	Probability of success given treatment	4
1.2	Probabilities of success given treatment (and sex)	6
1.3	Probabilities of success given treatment and sex	8
1.4	Conditional expectation values of Y given treatment and neediness	11
1.5	Path diagram of $E(M X)$ and $E(Y X, M)$	13
1.6	Path diagram of $E(M X, Z, W)$ and $E(Y X, M, Z, W)$	15
2.1	A simple experiment or quasi-experiment.	27
2.2	Experiment or quasi-experiment with a fallible covariate	32
2.3	Experiment or quasi-experiment with an intermediate variable	38
3.1	Venn-diagram of a filtration with $T = \{1, 2, 3\}$	50
3.2	Venn-diagram of a filtration with $T = \{1, 2, 3, 4\}$	53
4.1	Conditional probabilities of success given treatment	84
13.1	Generalized ANCOVA model with manifest variables	313
13.2	Path diagram of a generalized ANCOVA model with latent variables	321
13.3	Path diagrams of a generalized ANCOVA model with latent variables	325

List of Tables

1.1	Joint Probabilities of Treatment and Success	4
1.2	Joint Probabilities of Treatment, Sex and Success	6
1.3	Conditional Expectations Given Treatment	9
1.4	Conditional Expectations Given Treatment and Neediness	10
1.5	Covariances, Correlations, and Expectations (Omitting Pre-Tests) ...	13
1.6	Covariances, Correlations, and Expectations (Including Pre-Tests) ..	14
3.1	Joe and Ann With Self-Selection to Treatment Conditions	49
3.2	Joe and Ann With Perfect Dependence of Y on X	56
4.1	Joe With Two Independent Treatments	72
4.2	Joe and Ann With Random Assignment to Treatment	82
4.3	No Treatment for Joe	85
4.4	Jim and Jane: An Example With Bias at the Individual Level	88
4.5	An Example With an Intermediate Variable	91
6.1	Biased Treatment and Covariate-Treatment Regressions	128
6.2	Unbiased Treatment and Covariate-Treatment Regressions	129
6.3	Unbiased Covariate-Treatment Regression	130
6.4	Accidental Unbiasedness	143
7.1	Mean-Independent Outcome Condition	175
8.1	Covariate-Treatment Regression That is C_X -Unconfounded	195
8.2	Implications Between Causality Conditions	205
9.1	Conditional independence of X and true outcomes	222
9.2	Conditional regressive independence	228
9.3	Conditional probabilities $P(U=u X=x, Z=z)$ for Table 9.2	229
9.4	Implication structure between causality conditions	232
10.1	Strong Ignorability	247
10.2	Conditional Expectations of Y Given Treatment and Covariates	248
11.1	Expectations of the Outcome Variable Given Treatment and True Propensity in the Example of Table 10.1	269

11.2 Conditional expectations of Y given treatment and Z -conditional propensity scores in the example of Table 10.1	276
12.1 Covariances, Correlations, and Expectations in Example 1	301
12.2 Covariances, Correlations, and Expectations in Example 2	303
12.3 Covariances, Correlations, and Expectations in Example 3	304
13.1 Expectations Within Treatment and Neediness Conditions	311

Chapter 12

Analysis of Change Scores

In chapter 10, we treated the theoretical foundations of a number of data analysis procedures that may be used for estimating and testing conditional and average causal effects, especially in quasi-experimental studies. In this chapter, we will study the assumptions under which the analysis of change scores can be used to estimate conditional and average total effects.

12.1 Theory

Suppose there is a pre-test Z , a treatment variable X with at least two values and an outcome variable (or ‘post-test’) Y assessed with the same measurement instrument as the pre-test. In such a design it seems natural to analyze the differences $E(Y - Z | X = x) - E(Y - Z | X = x')$ between the conditional expectation values of the difference scores, or, in other words, the conditional expectation $E(Y - Z | X)$ of the difference between post-test and pre-test given the treatment variable and to test the hypothesis that this conditional expectation is a constant. In fact, this is exactly what we do in a two-factorial ANOVA with the within-factor *pre-post* and the between-factor *treatment*. If X is dichotomous and we choose $Y - Z$ as the dependent variable, the conditional expectation $E(Y - Z | X)$ is also estimated and tested in the t -test for independent observations, where X is the grouping variable. The intuitive idea is that if the expectation of the change variable $Y - Z$ in the treatment condition is greater than its expectation in the control condition, then this difference in the expected changes must be due to the treatment.

Can we rely on this intuition? Although this idea has some intuitive appeal, it can be shown that causal inference based on such an analysis of change scores rests on two strong assumptions.

Theorem 12.1 (Change Scores Considering two Values of X)

Let $\langle (\Omega, \mathcal{A}, P), (\mathcal{F}_t, t \in T), X, Y \rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, and let Z be a covariate of X . If $P(X = x), P(X = x') > 0$ and $E(Y | X)$ is unbiased w.r.t. total effects, then

$$\begin{aligned} & E(Y - Z | X = x) - E(Y - Z | X = x') \\ &= E(\delta_{xx'}) - [E(Z | X = x) - E(Z | X = x')]. \end{aligned} \tag{12.1}$$

(Proof p. 305)

Remark 12.2 (Implications for Data Analysis) Hence, if the conditional expectation $E(Y|X)$ is unbiased w.r.t. total effects, then we can conclude that

$$E(\delta_{xx'}) = E(Y-Z|X=x) - E(Y-Z|X=x')$$

only if $E(Z|X=x) = E(Z|X=x')$. In other words, if $E(Y|X)$ is unbiased w.r.t. total effects, then the difference $E(Y-Z|X=x) - E(Y-Z|X=x')$ between the expectations of the difference scores is $\delta_{xx'}$ -biased *unless the conditional expectation values $E(Z|X=x)$ and $E(Z|X=x')$ of the pre-test Z are identical*. If the expectations of the pre-test Z do depend on X and $E(Y|X)$ is unbiased w.r.t. total effects, then the average total effects can be identified by

$$E(\delta_{xx'}) = E(Y-Z|X=x) - E(Y-Z|X=x') + [E(Z|X=x) - E(Z|X=x')], \quad (12.2)$$

which is identical to $E(\delta_{xx'}) = E(Y|X=x) - E(Y|X=x')$. Hence, if the conditional expectation $E(Y|X)$ is unbiased w.r.t. total effects and $E(Z|X) \stackrel{p}{=} E(Z)$, then we may analyze change scores or test the *time* \times *group* interaction in ANOVA if we aim at testing the hypothesis that the total treatment effect is zero. However, in this case the ordinary *t*-test for independent groups tests the same hypothesis, and for this test we only have to assume unbiasedness of $E(Y|X)$ and not additionally $E(Z|X) \stackrel{p}{=} E(Z)$. \triangleleft

In Theorem 12.1 we considered only two values x and x' of a treatment variable X . Let us generalize this result to the case in which X is discrete with values $0, 1, \dots, J$. In this case, we use the random variables $1_{X=x}$ indicating with 1 if the event $\{X=x\} = \{\omega \in \Omega: X(\omega) = x\}$ occurs. Otherwise $1_{X=x}$ takes on the value 0.

Corollary 12.3 (Change Scores)

Let $\langle (\Omega, \mathcal{A}, P), (\mathcal{F}_t, t \in T), X, Y \rangle$ be a causality space, let Y be numerical and nonnegative or with finite expectation, and let Z be a covariate of X . Furthermore, let X be discrete with a finite number of values $x = 0, 1, \dots, J$, and let $P(X=x) > 0$ for all $x = 0, 1, \dots, J$. Then there are $\beta_0, \beta_x \in \mathbb{R}$ such that

$$E(Y-Z|X) = \beta_0 + \sum_{x=1}^J \beta_x 1_{X=x}. \quad (12.3)$$

and unbiasedness of $E(Y|X)$ w.r.t. total effects implies

$$\beta_x = E(\delta_{x0}) - [E(Z|X=x) - E(Z|X=0)], \quad x = 1, \dots, J. \quad (12.4)$$

(Proof p. 305)

Remark 12.4 (Time \times Group Interaction in ANOVA) With the *time* \times *group* interaction in ANOVA we test the hypothesis

$$H_0: \beta_x = 0, \quad x = 1, \dots, J. \quad (12.5)$$

Furthermore, under the assumptions about X specified in Corollary 12.3, the condition

$$E(Z|X=x) = E(Z|X=0) \quad \text{for all } x = 1, \dots, J,$$

is equivalent to $E(Z|X) = E(Z)$. Therefore, with the *time* \times *group* interaction in ANOVA we also test the hypothesis

$$H_0: E(\delta_{x0}) = 0, \quad x = 1, \dots, J, \quad (12.6)$$

provided that $E(Y|X)$ is unbiased w.r.t. total effects and $E(Z|X) = E(Z)$. \triangleleft

Corollary 12.3 immediately implies the following proposition.

Corollary 12.5 (A Necessary and Sufficient Condition for Unbiasedness)

Let the assumptions of Corollary 12.3 be true and let $E(Z|X) = E(Z)$. Then

$$\beta_x = E(\delta_{x0}), \quad x = 1, \dots, J, \quad (12.7)$$

if and only if

$$x - z = E(\delta_{x0}), \quad x = 1, \dots, J. \quad (12.8)$$

(Proof p. 306)

Remark 12.6 (One-Factorial ANOVA) Note that, under the assumptions about X specified in Corollary 12.3, the conditional expectation $E(Y|X)$ can be written

$$= \alpha_0 + \sum_{x=1}^J \alpha_x 1_{X=x}, \quad (12.9)$$

where

$$\alpha_x = E(Y|X=x) - E(Y|X=0), \quad x = 1, \dots, J. \quad (12.10)$$

A one-factorial ANOVA with dependent variable Y and between factor X tests the hypothesis

$$H_0: \alpha_x = 0, \quad x = 1, \dots, J. \quad (12.11)$$

Hence, if $E(Z|X) = E(Z)$ and $E(Y|X)$ is unbiased, then a one-factorial ANOVA with dependent variable Y and between-factor X also tests the hypothesis of no average total treatment effects. Oftentimes, there is a strong correlation between pre- and post-test. Therefore, statistical power will be higher for the interaction test mentioned in Remark 12.4. \triangleleft

Remark 12.7 (Randomized Experiment) In a randomized experiment, i. e., in an experiment with randomized assignment of the unit to one of the treatment conditions, X and the global covariate C_X are independent, implying that X and Z are independent as well. Then $E(Y|X)$ is unbiased w.r.t. total effects and $E(Z|X) = E(Z)$ will hold, i. e., the expectations of the pre-test Z will not differ between treatment conditions. In this case, analyzing change scores yields in fact estimates of average total treatment effects and tests the hypothesis that the average total treatment effects are zero. \triangleleft

Remark 12.8 (Mean-Independent Outcome Condition) Suppose that the mean-independent outcome condition $Y \perp C_X | X$ holds, i. e., suppose

$$E(Y|X, C_X) \stackrel{p}{=} E(Y|X).$$

Then the conditional expectation $E(Y|X)$ is unbiased w.r.t. total effects but $E(Y-Z|X)$ will be biased unless $E(Z|X) = E(Z)$. Hence, unless the expectations of the pre-test do not differ between treatment conditions, the analysis of change scores *does not test the hypothesis that the average total effects of the treatments are zero*, even if $E(Y|X)$ is unbiased w.r.t. total effects. \triangleleft

Remark 12.9 (Unbiasedness of Z) In Theorem 12.1 and Corollary 12.3 we presuppose that the conditional expectation $E(Y|X)$ is unbiased w.r.t. total effects. Of course, we can also analyze change scores under the assumption that Z is unbiased w.r.t. total effects. However, in this case we need an additional assumption that is specified in the following theorem. \triangleleft

Theorem 12.10 (Change Scores II)

Let the assumptions of Corollary 12.3 be true. If $E(Y|X, Z)$ is unbiased w.r.t. total effects and

$$E(Y|X, Z) \stackrel{p}{=} \beta_0 + 1 \cdot Z + \sum_{x=1}^J \beta_x 1_{X=x}, \quad (12.12)$$

then

$$E(Y-Z|X) = \beta_0 + \sum_{x=1}^J \beta_x 1_{X=x}, \quad (12.13)$$

where

$$\beta_x = E(\delta_{x0} | Z) = E(\delta_{x0}), \quad x = 1, \dots, J. \quad (12.14)$$

(Proof p. 306)

Remark 12.11 (A Caveat) Of course, the assumption that Equation (12.12) holds will rarely be true in empirical applications. Neither randomization techniques, nor techniques of covariate selection can be used to secure that this equation holds.

Therefore, it is often a better idea to analyze the conditional expectation $E(Y-Z|X, Z)$, which yields estimates of the same effects of X on Y as the analysis of $E(Y|X, Z)$, because

$$E(Y-Z|X, Z) = E(Y|X, Z) - Z.$$

The analysis of the conditional expectations $E(Y|X, Z)$ and $E(Y-Z|X, Z)$ will be treated in more detail in sections 13.1 and 13.2. \triangleleft

Table 12.1. Covariances, Correlations, and Expectations in Example 1

		Z	X	Y
<i>Covariate</i>	Z	100.00	.000	.582
<i>Treatment (yes=1, no=0)</i>	X	0.00	0.25	.364
<i>Outcome</i>	Y	80.00	2.50	189
Expectations		100.00	0.50	185.00

Note. Correlations (in italics) are rounded. A concrete sample generated from these parameters can be downloaded at www.causal-effects.de. The file name is *PC1.Table12.1.dat*.

12.2 Numerical Examples

In this section we will present three examples. In the first one, the conditions that are sufficient for unbiased w.r.t. total effectsness of the conditional expectation are $E(Y-Z|X)$ are satisfied. In this case, the analysis of change scores as well as the analysis of the outcome variable yield the average total treatment effect. In the second and third examples, the sufficient conditions are not met and in these cases, the analysis of change scores yields severely biased results.

Example 12.12 (Example 1) In our first example, we consider a treatment with two treatment conditions, $X=0$ and $X=1$, a pre-test Z , and a post-test (outcome variable) Y . In this example, the conditional expectation $E(Y|X)$ is unbiased w.r.t. total effects and the expectations of the covariate Z in the two treatment conditions are identical with each other. This situation occurs in a randomized experiment. Table 12.1 displays the variances, covariances, correlations, and expectations of the variables involved. These parameters are implied by

$$E(Y|X, Z) = 100 + 10X + .80Z, \quad (12.15)$$

with $E(X) = .50$ and $\text{Var}(Y|X) = 100$. Furthermore,

$$E(Z|X) = E(Z) = 100 \quad (12.16)$$

and

$$\text{Var}(Z|X) = 100.$$

The filtration $(\mathcal{F}_t, t \in T)$ is specified by

- (a) $\mathcal{F}_1 = \sigma(Z)$,
- (b) $\mathcal{F}_2 = \sigma(Z, X)$, and
- (c) $\mathcal{F}_3 = \sigma(Z, X, Y)$.

Furthermore, in this example, the covariate Z is identical with the global covariate C_X .

In this example, the conditional expectation

$$\begin{aligned}
 &= E[E(Y|X, Z)|X] && \text{[PR-Box 10.2 (v)]} \\
 &= E(100 + 10X + .80Z|X) && \text{[(12.15)]} \\
 &= 100 + 10X + .80E(Z|X) && \text{[PR-Box 10.2 (xvi)]} && (12.17) \\
 &= 100 + 10X + .80E(Z) && \text{[(12.16)]} \\
 &= 180 + 10X && \text{[}E(Z) = 100\text{]}
 \end{aligned}$$

is unbiased w.r.t. total effects and the average total treatment effect is 10. Therefore, according to Theorem 12.1, the conditional expectation $E(Y-Z|X)$ of the difference score on X is unbiased w.r.t. total effects as well. In fact,

$$\begin{aligned}
 E(Y-Z|X) &= E(Y|X) - E(Z|X) && \text{[PR-Box 10.2 (xvi)]} \\
 &= E(Y|X) - E(Z) && \text{[Eqs. (12.17), (12.16)]} \\
 &= 180 + 10X - 100 \\
 &= 80 + 10X.
 \end{aligned}$$

Hence, the average total treatment effect is 10, and it can be analyzed using $E(Y|X)$, but also using the conditional expectation $E(Y-Z|X)$ of the change score variable $Y-Z$ on X . Furthermore, in this first example, the t -test for independent groups defined by X with Y or $Y-Z$ as the dependent variable and the ANOVA *time* \times *group* interaction test also test the hypothesis of no average total treatment effect.

<

Example 12.13 (Example 2) In this second example, we again consider a treatment with two conditions, $X=0$ and $X=1$, a pre-treatment variable Z , and an outcome variable Y . In this second example, the conditional expectation $E(Y|X)$ is again unbiased w.r.t. total effects, but this time the expectations of the covariate Z in the two treatment conditions *are not identical* with each other, a situation frequently occurring in non-randomized experiments. Table 12.2 displays the variances, covariances, correlations, and expectations of the variables involved. These parameters are implied by

$$= 100 + 10X \quad (12.18)$$

with $E(X) = .50$ and $Var(Y|X) = 100$, as well as

$$E(Z|X) = 90 + 20X \quad (12.19)$$

and $Var(Z|X) = 25$. The filtration $(\mathcal{F}_t, t \in T)$ and the global covariate C_X are defined as in Example 12.12.

In this second example, the conditional expectation $E(Y|X)$ is unbiased w.r.t. total effects and the average total treatment effect is 10. However, $E(Z|X) \neq E(Z)$.

Table 12.2. Covariances, Correlations, and Expectations in Example 2

		Z	X	Y
<i>Covariate</i>	Z	125.00	.894	.400
<i>Treatment (yes=1, no=0)</i>	X	5.00	0.25	.447
<i>Outcome</i>	Y	50.00	2.50	125
Expectations		100.00	0.50	105.00

Note. Correlations (in italics) are rounded. A concrete sample generated from these parameters can be downloaded at www.causal-effects.de. The file name is *PC1.Table12.2.dat*.

Therefore, according to Corollary 12.3, the conditional expectation $E(Y - Z | X)$ of the difference score on X is biased. In fact,

$$\begin{aligned}
 E(Y - Z | X) &= E(Y | X) - E(Z | X) && \text{[PR-Box 10.2, (xvi)]} \\
 &= 100 + 10X - (90 + 20X) && \text{[Eqs. (12.18), (12.19)]} \\
 &= 10 - 10X.
 \end{aligned}$$

Hence, instead of the average total treatment effect of 10, analyzing the difference scores yields the alleged treatment effect of -10 . Obviously, this effect is seriously biased.

◁

Example 12.14 (Example 3) In Example 12.13 the outcome variable has been X -conditionally regressively independent of Z (i. e., $Y \perp\!\!\!\perp Z | X$), and because we assumed $Z = C_X$, the causality condition $Y \perp\!\!\!\perp C_X | X$ has been satisfied. However, if Z represents a pre-test and Y the post-test, this assumption is not realistic. Therefore, we study a third example, which is typical in a quasi-experiment in which pre- and post-tests are assessed: Now Y depends on X and Z , and the expectations of Z are different in the two treatment conditions.

Hence, we again consider a treatment with two conditions, $X=0$ and $X=1$, a pre-test Z , and a post-test Y . In this third example, the conditional expectation $E(Y | X)$ is not unbiased w.r.t. total effects, and again the expectations of the covariate Z in the two treatment conditions are not identical with each other. Table 12.3 displays the variances, covariances, correlations, and expectations of the variables involved. Now these parameters are implied by

$$E(Y | X, Z) = 100 + 10X + .80Z, \tag{12.20}$$

with $E(X) = .50$ and $Var(Y | X) = 100$. Furthermore,

$$E(Z | X) = 90 + 20X \tag{12.21}$$

Table 12.3. Covariances, Correlations, and Expectations in Example 3

		Z	X	Y
<i>Covariate</i>	Z	125.00	.894	.795
<i>Treatment (yes=1, no=0)</i>	X	5.00	0.25	.770
<i>Outcome</i>	Y	150.00	6.50	285
Expectations		100.00	0.50	105.00

Note. Correlations (in italics) are rounded. A concrete sample generated from these parameters can be downloaded at www.causal-effects.de. The file name is *PC1.Table12.3.dat*.

and $\text{Var}(Y|X) = 25$. The filtration $(\mathcal{F}_t, t \in T)$ and the global covariate C_X are defined as in Example 12.12.

In this third example, the average total treatment effect is still 10. However, the conditional expectations

$$\begin{aligned}
 &= E[E(Y|X, Z)|X] && \text{[PR-Box 10.2 (v)]} \\
 &= E(100 + 10X + .80Z|X) && \text{[(12.20)]} \\
 &= 100 + 10X + .80E(Z|X) && \text{[PR-Box 10.2 (xvi)]} && (12.22) \\
 &= 100 + 10X + .80(90 + 20X) && \text{[(12.21)]} \\
 &= 172 + 26X
 \end{aligned}$$

and

$$\begin{aligned}
 E(Y-Z|X) &= E(Y|X) - E(Z|X) && \text{[PR-Box 10.2 (xvi)]} \\
 &= 172 + 26X - (90 + 20X) && \text{[(12.22), (12.21)]} \\
 &= 82 + 6X
 \end{aligned}$$

are both seriously biased. Hence, instead of the average total treatment effect of 10, analyzing the conditional expectation of the post-test Y or the conditional expectation of the change score variable $Y-Z$ on treatment variable X yields invalid results about the average total effect of X . \triangleleft

12.3 Summary and Conclusions

In this section, we have shown that the analysis of difference scores yields correct estimates of the average total effects, provided that

- the conditional expectation $E(Y|X)$ is unbiased w.r.t. total effects, and
- there are no differences in the expectations of the pre-test between treatment conditions.

These prerequisites are met in the perfect randomized experiment, but usually not in other research situations. And of course, if $E(Y|X)$ is unbiased w.r.t. total effects, in order to estimate the total treatment effects, we can also estimate and test the conditional expectation $E(Y|X)$ instead of $E(Y-Z|X)$. However, statistical tests based on the analysis of change score will often have a higher power than the analysis of the outcome variable.

A second set of assumptions under which the analysis of difference scores yields correct estimates of the average total effects is

- (c) the conditional expectation $E(Y|X, Z)$ is unbiased w.r.t. total effects, and
- (d) $E(Y|X, Z) = \bar{y} + \beta_0 + 1 \cdot Z + \sum_{x=1}^J \beta_x 1_{X=x}$.

However, assumption (b) will usually not hold in empirical applications.

It has also been suggested to use analysis of covariance with difference scores as a dependent variable and the pre-test as the covariate (see, e. g., Jamieson, 2004). This suggestion will be dealt with in section 13.1.

Statistical Programs

The analysis of difference scores can be accomplished with every major statistical program package and with every program for multiple linear regression. Repeated measures ANOVA and, if X is dichotomous, t -tests for independent samples with the dependent variable $Y-Z$ will analyze and test hypothesis about the average total effect of the treatment conditions *if the assumptions (a) and (b) mentioned above are met*.

12.4 Proofs

Proof of Theorem 12.1

If $E(Y|X)$ is unbiased w.r.t. total effects, then by definition,

$$E(Y|X=x) = E(\tau_x) \quad \text{for each value } x \text{ of } X \text{ for which } P(X=x) > 0. \quad (12.23)$$

Because

$$\begin{aligned} E(Y-Z|X=x) &= E(Y|X=x) - E(Z|X=x) \\ &= E(\tau_x) - E(Z|X=x), \end{aligned} \quad [(12.23)]$$

$$\begin{aligned} E(Y-Z|X=x') &= E(Y|X=x') - E(Z|X=x') \\ &= E(\tau_{x'}) - E(Z|X=x'), \end{aligned} \quad [(12.23)]$$

and $E(\delta_{xx'}) = E(\tau_x) - E(\tau_{x'})$, taking the difference $E(Y-Z|X=x) - E(Y-Z|X=x')$ proves the theorem.

Proof of Corollary 12.3

First we show that Equation (12.3) holds under the assumptions about X . If $X=0$, then Equation (12.3) yields

$$E(Y-Z|X=0) = \beta_0, \quad (12.24)$$

and if $X=x$, then it yields

$$E(Y-Z|X=x) = \beta_0 + \beta_x, \quad x = 1, \dots, J. \quad (12.25)$$

Hence, the conditional expectation $E(Y-Z|X)$ can always be parameterized by Equation (12.3) and

$$\begin{aligned} \beta_x &= E(Y-Z|X=x) - E(Y-Z|X=0) \\ &= E(Y|X=x) - E(Z|X=x) - [E(Y|X=0) - E(Z|X=0)] \\ &= E(\delta_{x0}) - [E(Z|X=x) - E(Z|X=0)], \quad x = 1, \dots, J, \end{aligned} \quad (12.26)$$

because unbiasedness of $E(Y|X)$ w.r.t. total effects implies $E(Y|X=x) - E(Y|X=0) = E(\delta_{x0})$, for all $x = 1, \dots, J$ [see Defs. 6.1 (i) and 6.6 (i)].

Proof of Corollary 12.5

We presume $E(Z|X) = E(Z)$, which implies $E(Z|X=x) = E(Z)$ for all $x = 0, 1, \dots, J$, and

$$\begin{aligned} \beta_x &= E(Y-Z|X=x) - E(Y-Z|X=0) && [(12.26)] \\ &= E(Y|X=x) - E(Z|X=x) - [E(Y|X=0) - E(Z|X=0)] && [\text{PR-Box 9.2 (xvi)}] \\ &= E(Y|X=x) - E(Z) - [E(Y|X=0) - E(Z)] && [E(Z|X=x) = E(Z)] \\ &= E(Y|X=x) - E(Y|X=0). \end{aligned}$$

Hence, $\beta_x = E(\delta_{x0})$ if and only if $E(Y|X=x) - E(Y|X=0) = E(\delta_{x0})$.

Proof of Theorem 12.10

First we show that Equation (12.13) follows from Equation (12.12).

$$\begin{aligned} E(Y-Z|X) &= E[E(Y-Z|X, Z)|X] && [\text{PR-Box 9.2 (v)}] \\ &= E[E(Y|X, Z) - Z|X] && [\text{PR-Box 9.2 (xvi), (vii)}] \\ &= E\left(\beta_0 + 1 \cdot Z + \sum_{x=1}^J \beta_x 1_{X=x} - Z|X\right) && [(12.12)] \\ &= \beta_0 + \sum_{x=1}^J \beta_x 1_{X=x} && [\text{PR-Box 9.2 (v)}] \end{aligned}$$

Now we show that Equation (12.14) follows from Equation (12.12) and the additional assumption that $E(Y|X, Z)$ is unbiased w.r.t. total effects. If Equation (12.12) holds, then, according to Theorem PR-15.3 [with $g_i(Z) = \beta_x$],

$$E^{X=0}(Y|Z) \stackrel{p^{X=0}}{=} \beta_0 + 1 \cdot Z \quad (12.27)$$

and

$$E^{X=x}(Y|Z) \stackrel{p^{X=x}}{=} \beta_0 + 1 \cdot Z + \beta_x, \quad x = 1, \dots, J. \quad (12.28)$$

Therefore,

$$\beta_x \stackrel{p}{=} E^{X=x}(Y|Z) - E^{X=0}(Y|Z) \stackrel{p}{=} E(\delta_{x0}|Z), \quad (12.29)$$

the latter equation following from unbiasedness of $E(Y|X, Z)$ w.r.t. total effects (see Def. 6.6). The second equation follows from

$$\begin{aligned} E(\delta_{x0}) &= E[E(\delta_{x0}|Z)] && \text{[PR-Box 9.2 (iv)]} \\ &= E(\beta_x) && \text{[PR-Box 9.2 (i)]} \\ &= \beta_x. \end{aligned}$$

12.5 Exercises

Solutions

References

- Abraham, W. T., & Russell, D. W. (2004). Missing data: A review of current methods and applications in epidemiological research. *Current Opinion in Psychiatry, 17*, 315–321.
- Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.
- Aiken, L. S., & West, S. G. (1996). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.
- Aitkin, M. (1978). The analysis of unbalanced cross-classifications. *Journal of the Royal Statistical Society, Series A: Statistics in Society, 141*, 195–223.
- Allen, M. P. (1997). *Understanding regression analysis*. New York, NY: Plenum Press.
- Appelbaum, M. I., & Cramer, E. M. (1974). Some problems in the nonorthogonal analysis of variance. *Psychological Bulletin, 81*, 335–343.
- Arbuckle, J. L. (2006). *AMOS 7.0 user's guide* [Computer software manual]. Chicago, IL: SPSS.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.
- Bauer, H. (1996). *Probability theory*. Berlin, Germany and New York, NY: de Gruyter.
- Bentler, P. M. (1995). *EQS Structural Equations program manual* [Computer software manual]. Encino, CA: Multivariate Software.
- Bentler, P. M., & Wu, E. J. C. (2002). *EQS 6 for Windows guide*. Encino, CA: Multivariate Software.
- Berger, M. P. E., & Wong, W. K. (Eds.). (2005). *Applied optimal designs*. Chichester, England: Wiley.
- Biesanz, J. C., Deeb-Sossa, N., Aubrecht, A. M., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9*, 30–52.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*, 605–634.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach*. Hoboken, NJ: Wiley.
- Bonney, G. E. (1987). Logistic regression for dependent binary observations. *Biometrics, 43*, 951–973.

- Borooh, V. K. (2001). *Logit and probit: Ordered and multinomial models*. Thousand Oaks, CA: Sage.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219. doi: 10.1037/0033-295X.110.2.203
- Browne, M. W., & Mels, G. (1998). Path analysis: RAMONA. In *SYSTAT for Windows: Advanced Applications (Version 8) [Computer software manual]*. Evanston, IL: SYSTAT.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago, IL: Rand McNally.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.
- Carlson, J. E., & Timm, N. H. (1974). Analysis of nonorthogonal fixed-effects designs. *Psychological Bulletin*, *81*, 563–570.
- Cartwright, N. (1979). Causal laws and effective strategies. *Noûs*, *13*, 419–437.
- Cheng, J., & Small, D. S. (2006). Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *68*, 815–836.
- Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics*, *13*, 261–281.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression / correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.
- Cox, D. R., & Wermuth, N. (2004). Causality: A statistical view. *International Statistical Review*, *72*, 285–305.
- Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *41*(1), 1–31.
- Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. New York, NY: Wiley.
- Dunn, G., Maracy, M., Dowrick, C., Ayuso-Mateos, J. L., Dalgard, O. S., Page, H., ... Wilkinson, G. (2003). Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. *British Journal of Psychiatry*, *183*, 323–331.
- Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on Generalized Linear Models* (2nd ed.). New York, NY: Springer.
- Feller, W. (1968). *An introduction to probability theory and its applications* (2nd ed., Vol. 1). New York, NY: Wiley.
- Feller, W. (1971). *An introduction to probability theory and its applications* (2nd ed., Vol. 2). New York, NY: Wiley.
- Fichman, M., & Cummings, J. N. (2003). Multiple imputation for missing data: Making the most of what you know. *Organizational Research Methods*, *6*, 282–308.

- Fisher, R. A. (1925/1946). *Statistical methods for research workers* (10th ed.). Edinburgh, England: Oliver and Boyd.
- Gelman, A., & Hill, J. L. (2007). *Data analysis using regression and multilevel / hierarchical models*. New York, NY: Cambridge University.
- Georgii, H.-O. (2008). *Stochastics – Introduction to probability and statistics*. Berlin, Germany: de Gruyter.
- Gosslee, D. G., & Lucas, H. L. (1965). Analysis of variance of disproportionate data when interaction is present. *Biometrics*, *21*, 115–133.
- Graham, J. W., & Donaldson, S. I. (1993). Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. *Journal of Applied Psychology*, *78*, 119–128.
- Green, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Greenland, S. (2000). Causal analysis in the health sciences. *Journal of the American Statistical Association*, *95*, 286–289.
- Greenland, S. (2004). An overview of methods for causal inference from observational studies. In A. Gelman & X.-L. Meng (Eds.), *Applied bayesian modeling and causal inference from incomplete-data perspectives* (pp. 3–14). Chichester, England: Wiley.
- Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, *10*, 37–48.
- Greenland, S., & Robins, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*, *15*, 413–419.
- Györfi, L., Kohler, M., Krzyzak, A., & Walk, H. (2002). *A distribution-free theory of nonparametric regression*. New York, NY: Springer.
- Hancock, G. R. (2004). Experimental, quasi-experimental, and nonexperimental design and analysis with latent variables. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences*. Thousand Oaks, CA: Sage.
- Hernán, M. A., Clayton, D., & Keiding, N. (2011). The Simpson's paradox unraveled. *International Journal of Epidemiology*, 1–6. (Advance Access published March 30, 2011) doi: 10.1093/ije/dyr041
- Höfler, M. (2005). Causal inference based on counterfactuals. *BMC Medical Research Methodology*, *5*, 1–12.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960.
- Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equation models. *Sociological Methodology*, *18*, 449–484.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley.
- Hoyer, J., & Klein, A. (2000). Self-reflection and well-being: Is there a healthy amount of introspection? *Psychological Reports*, *86*, 135–141.
- Huet, S., Bouvier, A., Poursat, M.-A., & Jolivet, E. (2004). *Statistical tools for non-linear regression: A practical guide with S-PLUS and R examples* (2nd ed.).

- New York, NY: Springer.
- Jamieson, J. (2004). Analysis of covariance ANCOVA with difference scores. *International Journal of Psychophysiology*, *52*, 277–283.
- Jennings, E., & Green, J. L. (1984). Resolving nonorthogonal ANOVA disputes using cell means. *Journal of Experimental Education*, *52*, 159–162.
- Jo, B. (2002a). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics*, *27*, 385–409.
- Jo, B. (2002b). Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance. *Statistics in Medicine*, *21*, 3161–3181.
- Jo, B. (2002c). Statistical power in randomized intervention studies with non-compliance. *Psychological Methods*, *7*, 178–193.
- Jo, B., Asparouhov, T., Muthén, B. O., Jalongo, N. S., & Brown, C. H. (2008). Cluster randomized trials with treatment noncompliance. *Psychological Methods*, *13*, 1–18.
- Jöreskog, K. G., & Sörbom, D. (1996/2001). LISREL 8: User's Reference Guide. [Computer software manual].
- Kaplan, D. (2000). *Structural Equation Modeling: Foundations and Extensions*. Thousand Oaks, CA: Sage.
- Keele, L. (2008). *Semiparametric regression for the social sciences*. Chichester, England: Wiley.
- Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychology Bulletin*, *82*, 887–903.
- Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210.
- Kenny, D. A., & Judd, C. M. (1996). A general procedure for the estimation of interdependence. *Psychological Bulletin*, *119*(1), 138–148.
- Keren, G., & Lewis, C. (1976). Nonorthogonal designs: Sample versus population. *Psychological Bulletin*, *83*, 817–826.
- King, G., & Zeng, L. (2001). Improving forecasts of state failure. *World Politics*, *53*, 623–658.
- Klauer, K. J., Willmes, K., & Phye, G. D. (2002). Inducing inductive reasoning: Does it transfer to fluid intelligence? *Contemporary Educational Psychology*, *27*, 1–25.
- Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474.
- Klein, A. G., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, *42*, 647–673.
- Klenke, A. (2013). *Probability theory – A comprehensive course* (2nd ed.). London, England: Springer. doi: 10.1007/978-1-4471-5361-0
- Kolmogorov, A. N. (1956). *Foundations of the theory of probability* (2nd ed.; N. Morrison, Trans.). New York, NY: Chelsea.

- Kramer, C. Y. (1955). On the analysis of variance of a two-way classification with unequal sub-class numbers. *Biometrics*, *11*, 441–452.
- Langsrud, Ø. (2003). ANOVA for unbalanced data: Use Type II instead of Type III sums of squares. *Statistics and Computing*, *13*, 163–167.
- Lechner, M. (2000, November). A note on the "common support problem" in applied evaluation studies. Retrieved from <http://ssrn.com/abstract=259239>
- Lee, B. K., Lessler, J., & Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine*, *29*, 337–346. doi: 10.1002/sim.3782
- Lee, S.-Y. (2007). *Structural equation modeling: A bayesian approach*. Chichester, England: Wiley.
- Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models*. London, England: Sage.
- Little, T. D., Card, N. A., Bovaird, J. A., Preacher, K. J., & Crandall, C. S. (2007). Structural equation modeling of mediation and moderation with contextual factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 207–230). Mahwah, NJ: Lawrence Erlbaum.
- Loève, M. (1977). *Probability theory I* (4th ed.). New York, NY: Springer.
- Loève, M. (1978). *Probability theory II* (4th ed.). New York, NY: Springer.
- Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, *68*, 304–305.
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201–226.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Lawrence Erlbaum.
- Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, *9*, 275–300.
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Mayer, A., Dietzfelbinger, L., Rosseel, Y., & Steyer, R. (in press). The effectlitter approach for analyzing average and conditional effects. *Multivariate Behavioral Research*.
- Mayer, A., Thoemmes, F., Rose, N., Steyer, R., & West, S. G. (2014). Theory and analysis of total, direct, and indirect causal effects. *Multivariate Behavioral Research*, *49*(5), 425–442.
- McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 341–380). Lincolnwood, IL: Scientific Software International.
- McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational

- studies. *Psychological Methods*, 9(4), 403–425.
- McCullagh, P., & Nelder, J. A. (1989). *Monographs on statistics and applied probability: Vol. 37. Generalized linear models* (2nd ed.; D. R. Cox, D. V. Hinkley, N. Reid, D. B. Rubin, & D. V. Silverman, Eds.). Chapman & Hall.
- Meredith, M., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.
- Mill, J. S. (1843/1865). Of the four methods of experimental inquiry. In *A system of logic, ratiocinative and inductive: Volume 1. Being a connected view of the principles of evidence, and the methods of scientific investigation*. London, England: Longmans, Green, and Co.
- Morgan, S. L., & Harding, D. J. (2006). Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods & Research*, 35, 3–60. Retrieved from <http://smr.sagepub.com> doi: 10.1177/0049124106289164
- Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference. methods and principles for social research*. New York, NY: Cambridge University.
- Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus User's Guide* (5th ed.) [Computer software manual]. Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599–620.
- Nagengast, B. (2009). *Causal inference in multilevel models* (Unpublished doctoral dissertation). Friedrich-Schiller-Universität Jena, Thüringen, Germany.
- Nagengast, B., Kröhne, U., Bauer, M., & Steyer, R. (2007). *Causal Effects Explorer: A didactic tool for teaching the theory of individual and average causal effects* [Computer software manual]. University of Jena, Thüringen, Germany.
- Nelder, J. A., & Lane, P. W. (1995). The computer analysis of factorial experiments: In memoriam – Frank Yates. *The American Statistician*, 49, 382–385.
- OpenMx. (2009). *OpenMx - Advanced Structural Equation Modeling [Computer Software]*. Retrieved from <http://openmx.psyc.virginia.edu/>.
- Overall, J. E., & Spiegel, D. K. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin*, 72(5), 311–322.
- Overall, J. E., & Spiegel, D. K. (1973a). Comment on "Regression analysis of proportional cell data". *Psychological Bulletin*, 80, 28–30.
- Overall, J. E., & Spiegel, D. K. (1973b). Comments on Rawlings' nonorthogonal analysis of variance. *Psychological Bulletin*, 79, 164–167.
- Overall, J. E., Spiegel, D. K., & Cohen, J. (1975). Equivalence of orthogonal and nonorthogonal analysis of variance. *Psychological Bulletin*, 82, 182–186.
- Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4), 669–688.
- Pearl, J. (1998). Graphs, causality, and structural equation models. *Sociological Methods & Research*, 27, 226–284.

- Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge, UK.
- Pearson, K., Lee, A., & Bramley-Moore, L. (1899). Genetic (reproductive) selection: Inheritance of fertility in man, and of fecundity in thoroughbred racehorses. *Series A, Containing Papers of a Mathematical or Physical Character: Philosophical Transactions of the Royal Society of London*, *192*, 257–330.
- Pukelsheim, F. (2006). *Optimal design of experiments*. Philadelphia, PA: SIAM (Society for Industrial and Applied Mathematics).
- Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York, NY: Wiley.
- Reichardt, C. S. (1979). The statistical analysis of data from nonequivalent group designs. In T. D. Cook & D. C. Campbell (Eds.), *Quasi-experimentation: Design and analysis issues for field settings*. Orlando, FL: Houghton Mifflin.
- Reichardt, C. S. (2005). Nonequivalent group design. In B. Everitt & D. Howell (Eds.), *Encyclopedia of behavioral statistics*. New York, NY: Wiley.
- Robins, J., & Rotnitzky, A. (2004). Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika*, *91*, 763–783.
- Robins, J. M. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine*, *17*, 269–302.
- Rogosa, D. (1980a). Comparing nonparallel regression lines. *Psychological Bulletin*, *88*, 307–321.
- Rogosa, D. (1980b). A critique of cross-lagged correlation. *Psychological Bulletin*, *88*, 245–258.
- Rosenbaum, P. R. (2002). Attributing effects to treatment in matched observational studies. *Journal of the American Statistical Association*, *97*(457), 183–192. doi: 10.1198/016214502753479329
- Rosenbaum, P. R., & Rubin, D. B. (1983a). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *45*, 212–218.
- Rosenbaum, P. R., & Rubin, D. B. (1983b). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55. doi: 10.1093/biomet/70.1.41
- Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*, 516–524.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.
- Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701.
- Rubin, D. B. (1984). Assessing the fit of logistic regressions using the implied discriminant-analysis: Comment. *Journal of the American Statistical Association*, *79*, 79–80.

- Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, *31*, 161–170.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*, 322–331. doi: 10.1198/016214504000001880
- Rubin, D. B. (2006). *Matched sampling for causal effects*. New York, NY: Cambridge University.
- Senn, S. (2006). Change from baseline and analysis of covariance revisited. *Statistics in Medicine*, *25*, 4334–4344.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *13*, 238–241.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University.
- Sobel, M. E. (2008). Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics*, *33*, 230–251. doi: 10.3102/1076998607307239
- Song, X. Y., & Lee, S. Y. (2006). Bayesian analysis of structural equation models with nonlinear covariates and latent variables. *Multivariate Behavioral Research*, *41*, 337–365.
- Sörbom, D. (1978). An alternative to the methodology for analysis of covariance. *Psychometrika*, *43*, 381–396.
- Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: MIT.
- Splawa-Neyman, J. (1923/1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9 (Reprinted from *roczniki nauk rolniczych* tom 10, pp. 1–51, 1923). *Statistical Science*, *5*, 465–480.
- Spohn, W. (1980). Stochastic independence, causal independence, and shieldability. *Journal of Philosophical Logic*, *9*, 73–99.
- Stegmüller, W. (1983). *Erklärung, Begründung, Kausalität: Probleme und Resultate der Wissenschaftstheorie und analytischen Philosophie. [Explanation, justification, causality: Problems and findings of philosophy of science and analytic philosophy]*. Berlin, Germany: Springer.
- Steyer, R. (1984). Causal linear stochastic dependencies: An introduction. In J. R. Nesselroade & A. von Eye (Eds.), *Individual development and social change: Explanatory analysis*. New York, NY: Academic Press.
- Steyer, R. (1992). *Theorie kausaler Regressionsmodelle [Theory of causal regression models]*. Stuttgart, Germany: Fischer.
- Steyer, R. (2001). Classical test theory. In C. Ragin & T. D. Cook (Eds.), *International encyclopedia of the social and behavioural sciences: Logic of inquiry and research design* (pp. 481–520). Oxford, England: Pergamon.

- Steyer, R. (2005). Analyzing individual and average causal effects via structural equation models. *Methodology*, *1*, 39–54.
- Steyer, R., & Eid, M. (2001). *Messen und Testen: Ein Lehrbuch [Measurement and testing]*. Berlin, Germany: Springer.
- Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research Online*, *2*, 21–33.
- Steyer, R., Gabler, S., von Davier, A. A., & Nachtigall, C. (2000). Causal regression models II: Unconfoundedness and causal unbiasedness. *Methods of Psychological Research Online*, *5*, 55–87.
- Steyer, R., Mayer, A., Geiser, C., & Cole, D. (2015). A theory of states and traits - revised. *Annual Review of Clinical Psychology*, *11*, 71–98.
- Steyer, R., & Nagel, W. (in press). *Probability and conditional expectation: Fundamentals for the empirical sciences*. Chichester, England: Wiley.
- Steyer, R., & Partchev, I. (2007). EffectLite: User's manual – A program for the uni- and multivariate analysis of unconditional, conditional and average mean differences between groups [Computer software manual]. University of Jena, Thüringen, Germany.
- Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam, Netherlands: North-Holland.
- Tisak, J., & Tisak, M. S. (2000). Permanency and ephemerality of psychological measures with application to organizational commitment. *Psychological Methods*, *5*, 175–198.
- van Breukelen, G. J. P. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology*, *59*, 920–925.
- von Eye, A., & Schuster, C. (1998). *Regression analysis for social sciences*. San Diego, CA: Academic Press.
- Wainer, H. (1991). Adjusting for differential base rates: Lord's paradox again. *Psychological Bulletin*, *109*, 147–151.
- Wall, M. M., & Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology*, *56*, 47–63.
- Watkins, M. W., Lei, P., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence*, *35*, 59–68.
- West, S. G., & Aiken, L. S. (2005). Multiple linear regression. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science*. Chichester, England: Wiley.
- West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization in field settings experimental and quasi-experimental designs. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 40–84). New York, NY: Cambridge University.
- Williams, J. D. (1972). Two way fixed effects analysis of variance with disproportionate cell frequencies. *Multivariate Behavioral Research*, *7*, 57–83.

- Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, 25, 659–707.
- Wolf, F. M., Chandler, T. A., & Spies, C. J. (1981). A crossed-lagged panel analysis of quality of school life and achievement responsibility. *Journal of Educational Research*, 74, 363–368.
- Woo, M.-J., Reiter, J. P., & Karr, A. F. (2008). Estimation of propensity scores using generalized additive models. *Statistics in Medicine*, 27, 3805–3816.
- Woodward, J. A., & Bonett, G. B. (1991). Simple main effects in factorial designs. *Journal of Applied Statistics*, 18(2), 255–264. doi: 10.1080/02664769100000019
- Wright, S. (1918). On the nature of size factors. *Genetics*, 3, 367–374.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20(7), 557–585.
- Wright, S. (1923). The theory of path coefficients: A reply to Niles's criticism. *Genetics*, 8, 239–255.
- Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161–215.
- Wright, S. (1960a). Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics*, 16(2), 189–202.
- Wright, S. (1960b). The treatment of reciprocal interaction, with or without lag, in path analysis. *Biometrics*, 16(3), 423–445.
- Yule, G. U. (1903). Notes on the theory of association of attributes of statistics. *Biometrika*, 2, 121–134.