Vorlesung Statistik II

Florian Wickelmaier

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Sommersemester 2026

Kovarianzanalyse

Grundlagen

Die Kovarianzanalyse (ANCOVA) kombiniert den Ansatz der (multiplen) linearen Regression mit einer varianzanalytischen Auswertung (ANOVA/MANOVA)

Im entsprechenden Versuchsplan sind neben den abhängigen Variablen auch sogenannte *Kovariablen* zu erheben (d. h. Variablen, die mit den abhängigen Variablen kovariieren)

Beispiele

- Pretest-Werte in einem Pretest-Posttest-Design
- IQ-Werte bei der Beurteilung der Effektivität von Treatments

Achtung

- Die Kovariablen dürfen nicht vom Treatment abhängen (post-treatment bias)

Modell

Univariate, einfaktorielle Kovarianzanalyse mit einer Kovariable

Wir wollen also die Wirkung einer unabhängigen Variable auf die abhängige Variable erfassen und dabei den Einfluss einer Kovariable berücksichtigen

Für die j-te Beobachtung in der i-ten Faktorstufe ($i = 1, \ldots, m$ und $j = 1, \ldots, n$) werden die Werte y_{ij} der abhängigen Variable und x_{ij} der Kovariable als Realisierungen der Zufallsvariablen Y_{ij} und X_{ij} betrachtet

Modellgleichung

$$Y_{ij} = \mu + \alpha_i + \beta \cdot X_{ij} + \varepsilon_{ij}$$

Annahmen

- Die Fehler ε_{ij} sind dabei stochastisch unabhängig und identisch $N(0, \sigma^2)$-verteilt

Modell

Univariate, einfaktorielle Kovarianzanalyse mit einer Kovariable

Dabei bezeichnen

β den Effekt der Kovariable

α_i die Effekte der Faktorstufen, für die gilt

- entweder $\alpha_1 := 0$, damit ist μ der bedingte Erwartungswert der Referenzkategorie (Stufe 1), die Effekte sind die Differenzen zur Referenzkategorie (Dummykodierung)
- oder $\sum_i \alpha_i = 0$, damit ist μ der bedingte Gesamterwartungswert (grand mean), die Effekte sind die Differenzen zum Gesamterwartungswert (Effektkodierung)
- auch andere, weniger gebräuchliche Kontraste sind denkbar

Beispiel

Lowry (2000, Chap. 17) liefert folgendes Anwendungsbeispiel

- Anhand der Werte Y eines Abschlusstests sollen $m = 3$ Lehrmethoden zum Erwerb grundlegender Programmierkenntnisse verglichen werden (d. h. $i = 1, 2, 3$)
- In jeder von drei ausgewählten Schulen wurden hierzu für einen sechswöchigen Kurs je $n = 12$ Schüler der 5. Klasse zufällig ausgewählt ($j = 1, \ldots, 12$)
- Vor Beginn des Kurses wurde in einem Eingangstest die Vertrautheit im Umgang mit dem Computer erhoben, dessen Werte die Kovariable X definieren

Beispiel

Ergebnisse

Methode													
1	x_{1j}	14	10	7	18	14	16	13	15	5	18	16	10
	y_{1j}	29	24	14	27	27	28	27	32	13	35	32	17
2	x_{2j}	6	16	9	19	13	14	15	18	17	8	15	16
	y_{2j}	15	28	13	36	29	27	31	33	32	15	30	26
3	x_{3j}	15	9	7	12	12	9	12	3	13	10	11	8
	y_{3j}	32	27	15	23	26	17	25	14	29	22	30	25

X Eingangstest
Y Abschlusstest

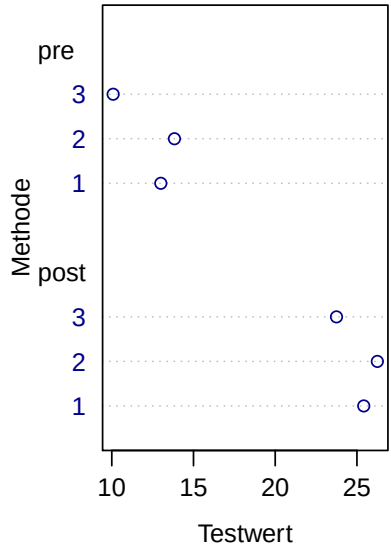
Beispiel

Mittlere Testwerte vor und nach den Kursen

```
dat <- read.table("ancova-data.txt",
                  header = TRUE)
dat$method <- factor(dat$method)

aggregate(cbind(pre, post) ~ method,
          dat, mean)
```

```
  method   pre  post
1      1 13.00 25.42
2      2 13.83 26.25
3      3 10.08 23.75
```



Adjustierung der Mittelwerte

Berechnung derjenigen Gruppenmittelwerte der abhängigen Variable, die resultieren würden, wenn die Gruppen bezüglich der Kovariable identische Mittelwerte aufweisen würden

Die adjustierten Mittelwerte sind die Vorhersagen des Modells für die Faktorstufen und $X_{ij} = \bar{x}$

$$\hat{y}_{i\cdot} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta} \cdot \bar{x}$$

Vorgehensweise

Es werden drei hierarchische lineare Modelle betrachtet

$$Y_{ij} = \mu + \beta \cdot X_{ij} + \varepsilon_{ij} \tag{1}$$

$$Y_{ij} = \mu + \alpha_i + \beta \cdot X_{ij} + \varepsilon_{ij} \tag{2}$$

$$\begin{aligned} Y_{ij} &= \mu + \alpha_i + \beta \cdot X_{ij} + (\alpha\beta)_i \cdot X_{ij} + \varepsilon_{ij} \\ &= \mu + \alpha_i + (\beta + (\alpha\beta)_i) X_{ij} + \varepsilon_{ij} \end{aligned} \tag{3}$$

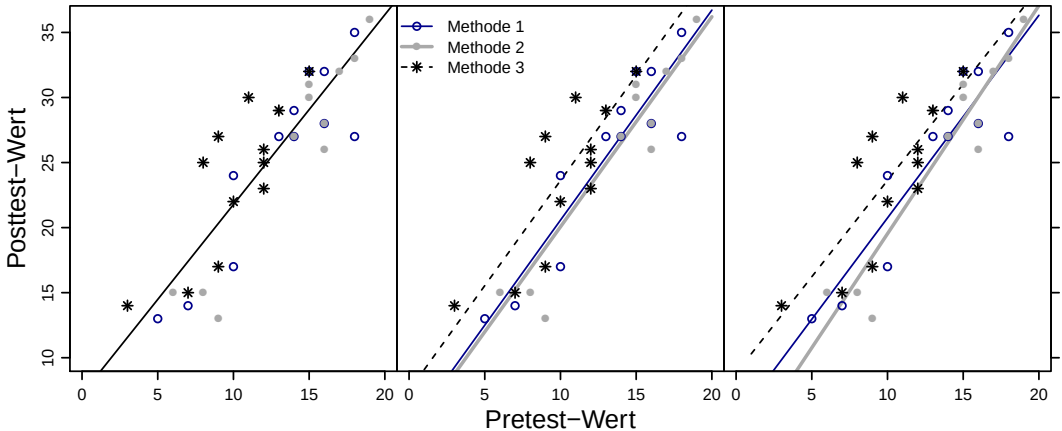
Modell 1 ist eine einfache Regression auf die Kovariable

Modell 2 besitzt drei y-Abschnitte, aber jeweils die gleiche Steigung

Modell 3 besitzt zusätzlich drei unterschiedliche Steigungen

Hypothesentests resultieren aus dem Vergleich der Modelle

Beispiel: Vergleich der Modelle



Hypothesentests

Der Test der Effekte des Faktors, $H_0\colon\ \alpha_i = 0\,\forall i$, resultiert aus dem Vergleich von Modell 1 mit Modell 2

Der Test auf gleiche Steigung, $H_0\colon\ (\alpha\beta)_i = 0\,\forall i$, resultiert aus dem Vergleich von Modell 2 mit Modell 3

Logik der Modellvergleiche

- Das bessere Modell klärt signifikant mehr Varianz auf (R^2)
- Das schlechtere Modell hat eine signifikant höhere Fehlerstreuung ($SS_{\hat{\varepsilon}}$)
- Die Komplexität (Anzahl der Parameter) der Modelle wird bei der Beurteilung der Güte berücksichtigt

Inkrementeller F-Test

Die Hypothesentests können auch als Spezialfälle einer allgemeinen Testprozedur verstanden werden: dem Vergleich von Modellen anhand ihrer Likelihood

Mit dem inkrementellen F-Test werden zwei verschachtelte lineare Modelle miteinander verglichen

Bezeichne M_1 das komplexere Modell und M_0 das einfachere, das durch die Parameterrestriktion $\vartheta_p = 0$ für ein oder mehrere p aus M_1 entstanden ist

Teststatistik

$$F = \frac{(SS_{\hat{Y}1} - SS_{\hat{Y}0})/(q_1 - q_0)}{SS_{\hat{\varepsilon}1}/(N - q_1)} = \frac{(SS_{\hat{\varepsilon}0} - SS_{\hat{\varepsilon}1})/(q_1 - q_0)}{SS_{\hat{\varepsilon}1}/(N - q_1)} = \frac{N - q_1}{q_1 - q_0} \cdot \frac{R_1^2 - R_0^2}{1 - R_1^2}$$

Inkrementeller F-Test

Anhand der Teststatistik wird überprüft, ob das einfachere zugunsten des komplexeren Modells verworfen werden kann, wobei unter der Nullhypothese $F \sim F(q_1 - q_0, N - q_1)$

Wenn das einfachere Modell signifikant weniger Varianz aufklärt als das komplexere, ist es zur Beschreibung der Daten ungeeignet

Es gelten folgende Definitionen

$SS_{\hat{Y}0}, SS_{\hat{\varepsilon}0}$	erklärte und Residualstreuung von M_0
$SS_{\hat{Y}1}, SS_{\hat{\varepsilon}1}$	erklärte und Residualstreuung von M_1
q_0, q_1	Anzahl der Parameter in M_0 und M_1
μ_{0i}, μ_{1i}	Erwartungswert von Y_i unter M_0 und M_1

Man kann zeigen, dass sich die Teststatistik aus dem allgemeinen Likelihoodverhältnistest herleiten lässt

Inkrementeller F-Test: Herleitung

Seien L_1 bzw. L_0 die Likelihoodfunktionen des komplexeren bzw. des einfacheren Modells

Dann gilt für normale lineare Modelle

$$
\begin{aligned}
G^2 &= 2 \log \frac{L_1}{L_0} \\
&= 2 \log \frac{\prod_i \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - \mu_{1i})^2}{2\sigma^2}\right)}{\prod_i \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - \mu_{0i})^2}{2\sigma^2}\right)} = 2\left(\sum_i -\frac{(y_i - \mu_{1i})^2}{2\sigma^2} - \sum_i -\frac{(y_i - \mu_{0i})^2}{2\sigma^2}\right) \\
&= 2\left(\frac{1}{2\sigma^2} \underbrace{\sum_i (y_i - \mu_{0i})^2}_{SS_{\hat{\varepsilon}0}} - \frac{1}{2\sigma^2} \underbrace{\sum_i (y_i - \mu_{1i})^2}_{SS_{\hat{\varepsilon}1}}\right) = \frac{SS_{\hat{\varepsilon}0} - SS_{\hat{\varepsilon}1}}{\sigma^2} \sim \chi^2(q_1 - q_0)
\end{aligned}
$$

Inkrementeller F-Test: Herleitung

Mit

$$\hat{\sigma}^2 = \frac{q_1 - q_0}{N - q_1} SS_{\hat{\varepsilon}1}$$

ergibt sich dann die unter der Nullhypothese F-verteilte Teststatistik

$$F = \frac{(SS_{\hat{\varepsilon}0} - SS_{\hat{\varepsilon}1})/(q_1 - q_0)}{SS_{\hat{\varepsilon}1}/(N - q_1)} \sim F(q_1 - q_0, N - q_1)$$

Beispiel: Regression von Y auf X

Summen der Abweichungsquadrate

	SS	df	MS	F	p
Regression	1237.718	1	1237.718	108.296	<0.001
Residuen	388.587	34	11.429		
Total	1626.306	35			

Modellgleichung

$$\hat{y}_{ij} = 7.188 + 1.459 \cdot x_{ij}$$

Determinationskoeffizient

$$R^2 = 0.761$$

Beispiel: Regression von Y auf X

```
m1 <- lm(post ~ pre, dat)
anova(m1)

Analysis of Variance Table

Response: post
          Df  Sum Sq Mean Sq F value    Pr(>F)
pre        1 1237.72 1237.72   108.3 4.163e-12
Residuals 34  388.59   11.43
```

Beispiel: Regression von Y auf X

```
summary(m1)

Call:
lm(formula = post ~ pre, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-7.3169 -2.4316  0.2656  2.0040  6.7656

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.1880     1.8147   3.961 0.000362
pre           1.4588     0.1402  10.407 4.16e-12

Residual standard error: 3.381 on 34 degrees of freedom
Multiple R-squared:  0.7611,  Adjusted R-squared:  0.754
F-statistic: 108.3 on 1 and 34 DF,  p-value: 4.163e-12
```

Parameterschätzung im ANCOVA-Modell

Zur Schätzung des für alle Gruppen gemeinsamen Regressionskoeffizienten β betrachten wir

- Die Summe der Kreuzprodukte von Y und X innerhalb der Gruppen

$$SC_W = \sum_{i=1}^{m} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.}) \cdot (x_{ij} - \bar{x}_{i.})$$

($\bar{x}_{i.}$ bezeichnet den Mittelwert von X in der i-ten Gruppe)

- Die Summe der Abweichungsquadrate von X innerhalb der Gruppen

$$SS_W(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{i.})^2$$

Parameterschätzung im ANCOVA-Modell

Den für alle Gruppen gemeinsamen Regressionskoeffizienten β schätzt man durch

$$\hat{\beta} = \frac{SC_W}{SS_W(X)}$$

Die gruppenspezifischen Ordinatenabschnitte α_i errechnet man für $i = 1, \ldots, m$ aus

$$\bar{y}_{i.} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta} \cdot \bar{x}_{i.}$$

Dabei gilt in Dummykodierung mit $\alpha_1 := 0$

$$\hat{\mu} = \bar{y}_{1.} - \hat{\beta} \cdot \bar{x}_{1.}$$

und für $i = 2, \ldots, m$

$$\hat{\alpha}_i = \bar{y}_{i.} - \hat{\mu} - \hat{\beta} \cdot \bar{x}_{i.}$$

Beispiel: Parameterschätzung

Summen der Kreuzprodukte der Abweichungen vom jeweiligen Mittelwert für Y und X innerhalb der Gruppen

Gruppe	SC_W
1	299.00
2	325.50
3	164.25
Summe	788.750

Quadratsumme innerhalb der Gruppen bzgl. X

$$SS_W(X) = 488.583$$

Beispiel: Parameterschätzung

Den Regressionskoeffizienten β schätzt man durch

$$\hat{\beta} = \frac{SC_W}{SS_W(X)} = \frac{788.750}{488.583} = 1.614$$

Gruppenspezifische Regressionsgeraden mit gleichen Steigungen

$$\begin{aligned}
\hat{y}_{1j} &= 4.43 + 1.614 \cdot x_{1j} \\
\hat{y}_{2j} &= 4.43 + (-0.51) + 1.614 \cdot x_{2j} \\
\hat{y}_{3j} &= 4.43 + 3.04 + 1.614 \cdot x_{3j}
\end{aligned}$$

Beispiel: ANCOVA-Modell

Summen der Abweichungsquadrate

	SS	df	*MS*	*F*	*p*
X	1237.718	1	1237.718	126.101	<0.001
A	74.498	2	37.249	3.795	0.033
Residuen	314.089	32	9.815		
Total	1626.306	35			

Determinationskoeffizient

$$R^2 = 0.807$$

Beispiel: ANCOVA-Modell

```
m2 <- lm(post ~ pre + method, dat)
anova(m2)

Analysis of Variance Table

Response: post
          Df  Sum Sq Mean Sq F value    Pr(>F)
pre        1 1237.72 1237.72 126.101 1.234e-12
method     2   74.50   37.25   3.795   0.03319
Residuals 32  314.09    9.82
```

Beispiel: ANCOVA-Modell

```
summary(m2)

Call:
lm(formula = post ~ pre + method, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-6.4885 -1.8357  0.4896  1.7718  4.9989

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.4300     2.0526   2.158   0.0385
pre           1.6144     0.1417  11.390 8.56e-13
method2      -0.5120     1.2845  -0.399   0.6928
method3       3.0419     1.3442   2.263   0.0306

Residual standard error: 3.133 on 32 degrees of freedom
Multiple R-squared:  0.8069,  Adjusted R-squared:  0.7888
F-statistic: 44.56 on 3 and 32 DF,  p-value: 1.565e-11
```

Beispiel: gruppenspezifische Steigungen

Summen der Abweichungsquadrate

	SS	df	*MS*	*F*	*p*
X	1237.718	1	1237.718	120.592	<0.001
A	74.498	2	37.249	3.629	0.039
$A \times X$	6.179	2	3.089	0.301	0.742
Residuen	307.911	30	10.264		
Total	1626.306	35			

Determinationskoeffizient

$$R^2 = 0.811$$

Beispiel: gruppenspezifische Steigungen

```
m3 <- lm(post ~ pre*method, dat)
anova(m3)

Analysis of Variance Table

Response: post
           Df  Sum Sq Mean Sq  F value    Pr(>F)
pre         1 1237.72 1237.72 120.5920 4.949e-12
method      2   74.50   37.25   3.6292   0.03877
pre:method  2    6.18    3.09   0.3010   0.74229
Residuals  30  307.91   10.26
```

Beispiel: gruppenspezifische Steigungen

```
summary(m3)

Call:
lm(formula = post ~ pre * method, data = dat)

Residuals: ...

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.17188    3.14476   1.645    0.110
pre          1.55729    0.23121   6.735 1.82e-07
method2     -3.17367    4.61771  -0.687    0.497
method3      3.64631    4.48923   0.812    0.423
pre:method2  0.19585    0.32975   0.594    0.557
pre:method3 -0.07645    0.38209  -0.200    0.843

Residual standard error: 3.204 on 30 degrees of freedom
Multiple R-squared:  0.8107,  Adjusted R-squared:  0.7791
F-statistic: 25.69 on 5 and 30 DF,  p-value: 5.292e-10
```

Beispiel: inkrementeller F-Test

ANOVA-Tabelle

	$SS_{\hat{\varepsilon}}$	df	ΔSS	Δdf	F	p
Modell 1	388.59	34				
Modell 2	314.09	32	74.498	2	3.629	0.039
Modell 3	307.91	30	6.179	2	0.301	0.742

Determinationskoeffizienten

	R^2
Modell 1	0.761
Modell 2	0.807
Modell 3	0.811

Beispiel: inkrementeller F-Test

```
anova(m1, m2, m3)

Analysis of Variance Table

Model 1: post ~ pre
Model 2: post ~ pre + method
Model 3: post ~ pre * method
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     34 388.59
2     32 314.09  2    74.498 3.6292 0.03877
3     30 307.91  2     6.179 0.3010 0.74229
```

Adjustierung der Gruppenmittelwerte

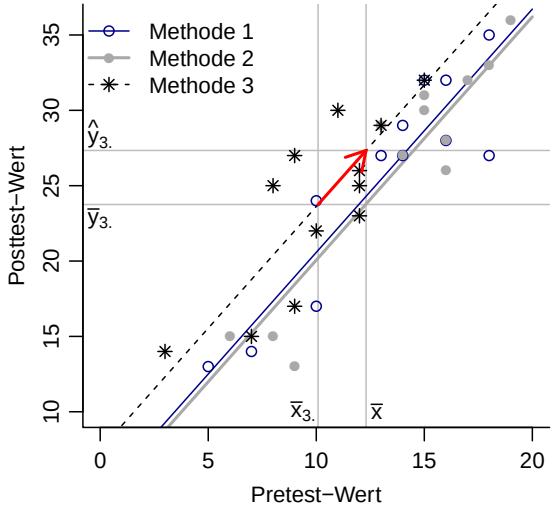
Die Berücksichtigung des Einflusses der Kovariable X erfolgt mittels der geschätzten Regressionskoeffizienten über die Gleichung

$$\hat{y}_{i.} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta} \cdot \bar{x}$$

für alle $i = 1, \ldots, m$

Die adjustierten Mittelwerte werden als diejenigen Gruppenmittel interpretiert, die resultieren würden, wenn die Gruppen bezüglich der Kovariable denselben Mittelwert hätten

Beispiel: Adjustierung der Gruppenmittelwerte



Beispiel: Adjustierung der Gruppenmittelwerte

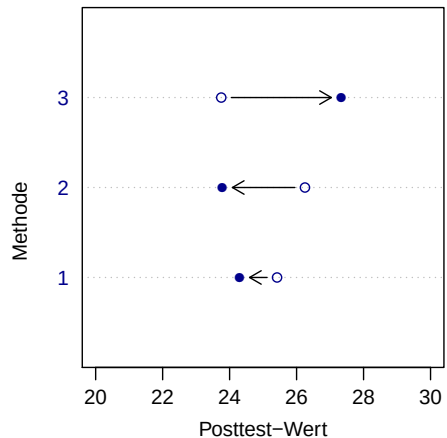
Man errechnet

$$\begin{aligned}\hat{y}_{1.} &= \hat{\mu} + \hat{\alpha}_1 + \hat{\beta} \cdot \bar{x} \\ &= 4.430 + 0 + 1.614 \cdot 12.306 = 24.3\end{aligned}$$

bzw.

$$\hat{y}_{2.} = 23.9 \quad \text{und} \quad \hat{y}_{3.} = 27.3$$

```
predict(
  m2, data.frame(method = factor(1:3),
                 pre = mean(dat$pre))
)
    1     2     3
24.30 23.78 27.34
```



Literaturhinweise

Stevens, J. (2001). *Applied multivariate statistics for the social sciences* (Kap. 9). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lowry, R. (2000). *Concepts and applications of inferential statistics* (Kap. 17) [On-line Lehrbuch]. Verfügbar unter: `http://vassarstats.net/textbook/`

Vickers, A. J., & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *British Medical Journal*, 323, 1123–1124. `https://doi.org/10.1136/bmj.323.7321.1123`