Zürcher Hochschule für Angewandte Wissenschaften ZHAW

Master of Science in Engineering MSE

Sina Rüeger
sina.rueeger@gmail.com

Auswertung Konsumententest Lindt & Sprüngli

Projektarbeit I: September bis November 2009

Betreuerin

Prof. Dr. phil. Marianne Müller
Zürcher Hochschule für Angewandte Wissenschaften, Winterthur

Um bei neuen Produkten das Marktpotential abschätzen zu können, lässt die Sensorikabteilung von Lindt & Sprüngli ausgewählte Produkte von Konsumenten testen. Die Testpersonen geben pro Produkt eine Wertung ab, die Produktereihenfolge pro Person ist randomisiert. Ein marktreifes Produkt sollte mindestens die Bewertung 8 erhalten. Um die Punktschätzungen besser beurteilen zu können, werden Vertrauensintervalle der Punktschätzungen verwendet. Um einen guten Überblick über die Schätzungen und die Vertrauensintervalle zu haben, wird eine Grafik dazu erstellt.

Das Design dieser Konsumententests entspricht einem Repeated Measures, resp. einem Split-Plot Design. Bei der Analyse dieser Daten muss auf die Abhängigkeit der Bewertungen innerhalb einer Personen geachtet werden, weil sonst die Effekte durch falsch F-Tests nicht richtig interpretiert werden können. Diese Projektarbeit zeigt die Methoden zur Analyse von Repeated Measures Daten auf. Dabei gibt es einige Herausforderungen wie die Analyse von einem unbalancierten Design, Datensätze mit fehlenden Werten, die Forderung einer homogenen Varianz-Kovarianzstruktur und - bei einem unbalancierten Design - die Erstellung von Vertrauensintevallen durch die Boostrap-Methode, weil die t-Werte nur approximativ t-verteilt sind.

Für eine benutzerfreundliche Anwendung wird eine Routine zur Auswertung dieser Konsumententestdaten in der Statistik-Software SPSS geschrieben, mit der Konsumententest-Daten über die bisherige Auswertung hinaus standardisiert analysiert werden können. Die Auswertungen in diesem Bericht wurden mit der Software `R` erstellt.

Inhaltsverzeichnis

1. Einleitung

Um bei neuen Produkten das Marktpotential abschätzen zu können oder um herauszufinden, ob ältere Produkte noch Anklang finden, lässt die Sensorikabteilung von Lindt & Sprüngli ausgewählte Produkte von Konsumenten testen. Die Testpersonen geben pro Produkt eine Wertung ab, die Produktereihenfolge pro Person ist randomisiert. Es handelt sich um wiederholte Wertungen einer Person die als abhängig voneinander betrachtet werden. Die Struktur eines solchen Datensatzes wird auch *Repeated Measures*, *Longitudinal* oder *Cross-over* genannt.

Der Datensatz für die Bewertung von J Produkten kann in drei Variablengruppen aufgeteilt werden. Weil die erklärenden Variablen alle kategoriell sind, werden sie auch *Faktoren* genannt. Es gibt *between-subject* Faktoren, wie beispielsweise das Alter oder das Geschlecht, die sich pro Person ändern und in einen *within-subject* Faktor, das Produkt, welcher sich bei jeder Person wiederholt.

subject	*within-subject*			*between-subject*	*between-subject*
Subject	Produkt 1	...	Produkt J	Geschlecht	Alter
Person 1	*10*	...	*7*	weiblich	-34
⋮	⋮	⋱	⋮	⋮	⋮
Person N	*3*	...	*8*	männlich	55-

Tabelle 1.1.: Datensatzstruktur von Lindt & Sprüngli mit der Unterteilung in *within-subject* und *between-subject* Faktoren.

Repeated Measures Datensätze können schematisch als Split-Plot wie in der Tabelle 1.2 dargestellt werden. Das Split-Plot Design entstand durch Feldversuche[1]. Ein konkretes Beispiel: Für ein Experiment steht ein grosses Feld zur Verfügung. Es sollen 4 Maissorten und 3 Bewässerungstechniken getestet werden. Das Feld wird in 8 Rechtecke unterteilt. Jedes Rechteck wird widerum in 3 Teile unterteilt, sogenannte *Main-Plots.* Pro Main-Plot werden die Maissorten randomisiert zugeteilt. Dann wird jedem Main-Plot wird in 4 Teile unterteilt, sogenannte *Sub-Plots.* Innerhalb jedes Main-Plots wird jedem Sub-Plot randomisiert eine Bewässerungstecknik zugeteilt. Die Beobachtungen innerhalb eines Main-Plots werden als abhängig voneinander betrachtet. Diese zweistufige Randomisierung muss von der vollständigen Randomisierung unterschieden werden. Bei einer vollständigen Randomisiertung wird das Rechteck zwar auch in 12 Teile unterteilt, aber jedem Rechteck

[1]Plot engl. Landstück

wird zufällig eine Maissorte und eine Bewässerungstechnik zugeteilt. Die Ordnung in Main- und Sub-Plots geht also verloren.

Bei Split-Plot Experimenten ist oft vom Agrar-Vokabular die Rede (Feldern, Pflanzen, Sorten, Techniken, Bewässerung, ...). Bei Repeated Measures geht es um Soziologie, Psychologie, Pharmazie, Medizin (Menschen, Behandlungen, Produkte). Obwohl diese zwei Designs in den Anwendungsgebieten weit auseinander liegen, gibt es Paralellen. Tabelle 1.2 zeigt die Übersetzung eines Split-Plot Designs auf die Repeated Measures Daten von Lindt & Sprüngli. Wobei Rechteck = Personen mit Altersgruppe k und Geschlecht h, Main-Plots = Personen und Sub-Plots = Produkte. Pro Kombination von Geschlecht und Alter gibt es ein Anzahl Personen die zur Verfügung stehen. Jeder Personen wird die Produktereihenfolge randomisiert zugeteilt. Der Faktor Produkt wird auch *within-subject* Faktor genannt und das Alter und das Geschlecht *between-subject* Faktor. Die Personen sind die *subjects*.

	Main-Plot 1	**Main-Plot** 2	**Main-Plot** 3	....	**Main-Plot** N
	Person 1	Person 2	Person 3	....	Person N
Sub-Plot 1	Produkt 1	Produkt 4	Produkt 4	....	Produkt 1
Sub-Plot 2	Produkt 4	Produkt 3	Produkt 3	....	Produkt 5
Sub-Plot 3	Produkt 2	Produkt 2	Produkt 5	....	Produkt 2
Sub-Plot 4	Produkt 5	Produkt 5	Produkt 1	....	Produkt 3
Sub-Plot 5	Produkt 1	Produkt 2	Produkt 2	....	Produkt 4

Tabelle 1.2.: Split-Plot Design für Alter k und Geschlecht h, N_{kh} Personen testen 5 Produkte.

Es gibt verschiedene Ansätze, einen solchen Datensatz zu analysieren. Weil es sich bei den erklärenden Variablen ausschliesslich um kategorielle Variablen handelt, liegt eine Varianzanalyse nahe (ANOVA, Analysis of Variance engl. Varianzanalyse).

1. **ANOVA** ohne Personeneffektx
   Die Varianzanalyse kann mit den Variablen Geschlecht und Alter und Produkt gemacht werden. Die Unterschiede zwischen Personen werden dabei nicht berücksichtigt.

2. **ANOVA mit vollständig randomisierten Blockdesign**
   Die Varianzanalyse wird mit den Variablen Geschlecht, Alter, Produkt und Person gemacht. Jede Person stellt einen Block dar. Die Wertungen einer Person werden als *unabhängig* voneinander betrachtet.

3. **Repeated Measures ANOVA**
   Die Varianzanalyse wird mit den Variablen Geschlecht, Alter, Produkt und Person gemacht. Die Bewertungen einer Person werden als *abhängig* voneinander betrachtet. Es gibt zwei Fehlerstreuungen. Die F-Tests für die *between-subject* Faktoren werden

mit der Variabilität zwischen den Personen gemacht. Die F-Tests für die *within-subject* Faktoren werden mit der Variabilität innerhalb einer Person gebildet.

Diese Projektarbeit soll darüber Aufschluss geben, welcher Ansatz nun der beste ist. Vorweg die Antwort: Repeated Measures ANOVA liefert die richtigen Resultate. Repeated Measures ANOVA behandeln die Faktoren getrennt als *between-subject* und *within-subject*, was schlussendlich entscheidend ist für die F-Tests.

Vergleichen wir Ansatz 1 und 2 mit dem korrekten Modell Ansatz 3. Die P-Werte der *between-subject* Faktoren werden in Ansatz 1 und 2 unterschätzt. Hingegen wird der P-Wert des *within-subject* Faktors in Ansatz 1 überschätzt. Die Gefahr ist also gross, dass *within-subject* Effekte übersehen und *between-subject* Effekte fälschlicherweise als statistisch signifikant angesehen werden. In Kapitel 7.4 werden die Resultate dieser Modelle aufgeführt und verglichen.

	F-Wert <u>Ansatz 1</u> statt Ansatz 3	F-Wert <u>Ansatz 2</u> statt Ansatz 3
Within-subject Faktor (Produkt)	zu klein	gleich
Between-subject Faktoren (Alter, Geschlecht)	viel zu gross	viel zu gross
Subject (Person)	-	gleich

Bisher wurde als Auswertungsmethode Ansatz 1 verwendet. Der Einbezug des Personeneffekts ist einerseits wichtig wegen der Korrektheit der F-Tests. Ein anderer Aspekt ist die Messung der Variabilität zwischen den Personen. Häufig ist die Bewertung der Produkte innerhalb der Personen sehr homogen, zwischen den Personen aber heterogen. Und diese Heterogenität der Bewertungen kann nicht durch die bestehenden Faktoren erklärt werden. σ_ϵ^2 ist die Variabilität zwischen den Bewertungen derselben Person. σ_S^2 ist die Variabilität der Bewertungen zwischen den Personen. Ist σ_ϵ^2 kleiner als σ_S^2, ist die Streuung innerhalb der Personen tatsächlich kleiner wie zwischen den Personen. Diese zwei Fehlerterme implizieren ein *Mixed Model*. Ein Repeated Measures ANOVA entspricht also sowohl einem Split-Plot Design als auch einem Mixed Model mit einem Intercept-Fehlerterm (*traditionelles* Mixed Model). Aber - nicht alle Mixed Models entsprechen zwingend ein Repeated Measures ANOVA. Mixed Models können beliebig erweitert werden, beispielsweise mit einem Slope-Term.

Zentral für die Auswertung der Lindt & Sprüngli Daten ist vor allem eine Grafik siehe Abbildung 1.1, die die Schätzungen für Subgruppen der Konsumententestenden anzeigt, sowie Vertrauensintervalle, die den Bereich anzeigen, wo mit 95% Wahrscheinlichkeit der wahre Wert liegt. Oder anders gesagt, wo die Wahrscheinlichkeit 95% ist, dass man einen Balken konstruiert, in dem der wahre Wert liegt. Als Nebeneffekt soll die Analyse Faktoren indentifizieren, die einen statistisch signifikanten Effekt haben und den Personeneffekt beziffern.

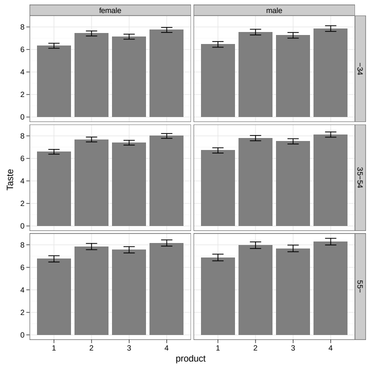


Abbildung 1.1.: Grafik mit den Schätzungen der Produkte als graue Flächen und schwarze Balken als Vertrauensintervalle für die Subgruppen Alter und Geschlecht

Kapitel 2 beschreibt zwei Datensätze von Lindt & Sprüngli, ein vollständigen und einen mit fehlenden Werten. Kapitel 3 geht auf verschiedene Modellansätze ein und zeigt, wann welche angewendet werden. In Kapitel 4 wird kurz die bisherige Auswertungsmethode ausgeführt. Kapitel 5 und 6 widmen sich Repeated Measures ANOVA Modellen und Mixed Models. Kapitel 4, 5 und 6 gliedern sich jeweils in einen Theorieteil, in die Umsetzung am Beispiel des Lindt & Sprüngli Datensatzes und deren wichtigste Ergebnisse. Kapitel 7 geht auf die praktische Umsetzung mit `R` und SPSS ein und zeigt eine Analyse eines Teils der Lindt & Sprüngli Daten mit dem Modell, das schlussendlich gewählt wurde. Kapitel 8 diskutiert weitere Möglichkeiten und Änderungen, die vorgenommen werden könnten.

Vorab noch etwas Notation. Wird in diesem Bericht von einem Regressionsmodell geschrieben, wird es stets als *linear* angenommen. Grundsätztlich wird versucht, deutschsprachige Begriffe zu verwenden. Falls aber der englische Ausdruck üblich ist, wird dieser englische Begriff verwendet (z.B. Mixed Model). $y_{...}$ ist der Mittelwert über alle drei Indizes, $y_{hi.}$ ist der Mittelwert des Gruppe h und der Gruppe i über den dritten Index. Die eckigen Klammern $\langle \ldots \rangle$ deuten auf eine Funktion hin (z.B. $\exp\langle \ldots \rangle$).

2. Daten

Der Datensatz von Lindt & Sprüngli setzt sich aus der Wertung des Produktes (metrisch) punkto Geschmack und Beschaffenheit und erklärenden Faktoren zusammen. Faktoren sind in der Regel das Alter und das Geschlecht. Manchmal auch die Sprache, wenn der Test auch mit Einwohnern aus der französischsprachigen Schweiz durchgeführt wurde oder der Monat des Tests, wenn an verschiedenen Zeitpunkten im Jahr derselbe Test durchgeführt wurde.

Beim Datensatz muss auf 3 Dinge geachtet werden, die einen Einfluss auf die Wahl des Modells haben.

1. Welche Faktoren sind geschachtelt, welche gekreuzt
2. Fehlende Werte
3. Balanciertes Design

Fehlende Werte können auf zwei Arten auftauchen. Einerseits in der Zielvariable, wenn eine Person ein Produkt nicht bewertet. Andererseits in den *between-subject* Faktoren, wenn eine Person beispielsweise das Alter nicht angibt.

Bei der Balancierung der Daten gilt ähnliches. Für gewisse Modelle gilt die Bedingung, dass alle Personen alle Produkte testen müssen. Die Kombinationen der Faktoren wie Alter und Geschlecht können ebenfalls balanciert sein.

Zusätzlich kann eine vom Produkt abhängige Variable dazukommen. Im Fall von Lindt & Sprüngli wäre das die Kaufabsicht vor und nach dem Testen eines Produktes. Diese kann von Produkt zu Produkt variieren. Diese Analysevariante wird im Ausblick in Kapitel 8 besprochen.

2.1. Datensatz A

519 Personen haben jeweils 4 Produkte getestet.

Überblick

Variable	Bedeutung
`Taste`	Geschmack
`Texture`	Beschaffenheit
`product`	Produkt
`subject`	Nummerierung der Testperson
`age`	Altersgruppe der Testperson
`gender`	Geschlecht der Testperson
`run`	Zeitpunkt des Tests

Die Variable `product` ist der *within-subject* Faktor. Die Variablen `age`, `gender` und `run` sind die *between-subject* Faktoren.

Levels

- `Taste`: metrische Variable von `1` bis `10`, (ursprünglich eine kategorielle Varibable von *sehr schlecht* bis *sehr gut*).
- `Texture`: metrische Variable von `1` bis `10`, (ursprünglich eine kategorielle Varibable von *sehr schlecht* bis *sehr gut*).
- `NJ`: Nummerierung der Personen von `1` bis `519`.
- `product`: 4 Pralinés, `1` bis `4`.
- `age: -34, 35-54, 55-`
- `gender: female, male`
- `run: VT0803, VT0809`

Die Faktoren `age`, `gender` und `product` und `run` sind gekreuzt. Der Faktor `subject` ist mit `product` gekreuzt und in `age` und `gender` und `run` geschachtelt.

Der Datensatz ist vollständig, es gibt keine fehlenden Werte.

Die Daten insofern balaciert, dass jede Person den `Taste` und `Texture` für alle 4 Produkte bestimmt hat. Pro Kombination der Faktoren `age`, `gender` und `run` sind aber unterschiedlich viele Beobachtungen vorhanden.

Bei der Auswertung dieses Datensatzes werden keine Wechselwirkungen berücksichtigt, weil keine vorhanden sind.

2.2. Datensatz B: Fehlende Werte

Die 199 Testpersonen haben jeweils 8 verschiedene Produkte getestet. Nicht alle Personen haben alle 8 Produkte bewertet, es gibt fehlende Werte.

Überblick

Variable	Bedeutung
`Taste`	Geschmack
`Texture`	Beschaffenheit
`subject`	Nummerierung der Testperson
`product`	Produkt
`age`	Alter der Testperson
`gender`	Geschlecht der Testperson
`lng`	Sprache der Testperson

Die Variable `product` ist der *within-subject* Faktor. Die Variablen `age`, `gender` und `lng` sind die *between-subject* Faktoren.

Levels

- `Taste`: metrische Variable von `1` bis `10`, (ursprünglich eine kategorielle Varibable von *sehr schlecht* bis *sehr gut*).
- `Texture`: metrische Variable von `1` bis `10`, (ursprünglich eine kategorielle Varibable von *sehr schlecht* bis *sehr gut*).
- `subject`: Nummerierung der Personen von `1` bis `199`.
- `product: Coop, Frey, Cailler, Lindt CH, Lindt F, Lindt D, Aldi, Ritter`
- `age: -34, 35-54, 55-`
- `gender: female, male`
- `lng: german (1), french (2)`

Die Faktoren `age`, `gender` und `product` und `lng` sind gekreuzt. Der Faktor `subject` ist mit `prd` gekreuzt und in `age` und `gender` und `lng` geschachtelt.

Im Datensatz gibt es bei der Zielvariable fehlende Werte, als auch bei den erklärenden Variablen. Ausser einer Person, die drei Produkte nicht bewertet hat, haben alle Personen jeweils zwei Produkte nicht bewertet (ein sogenanntes *Balanced Incomplete Design*). Drei Personen haben ihr Geschlecht nicht angegeben.

Wenn die Zeilen mit den fehlenden Werten in der Zielvariable herausgelöscht werden, gibt es eine Unbalanciertheit im Sinn der Produkte. Pro Kombination der Faktoren `age`, `gender` und `lng` sind unterschiedlich viele Beobachtungen vorhanden.

2.3. Graphische Darstellung

Um sich die Daten besser vorstellen zu können oder Modellannahmen zu überprüfen, können die Daten graphisch dargestellt werden. Abbildung 2.1 zeigt einen Spaghetti-Plot, wo die Wertungen einer Person mit einer Linie verbunden wird. Die Darstellung zeigt ein heterogenes Bild der Bewertungen der Personen. Die Bewertungen variieren stark, es gibt keinen klaren Trend. Produkt 2 und Produkt 8 scheinen besonders schlecht bewertet worden sein. Ansonsten bewegen sich die Wertungen eher auf höherem Niveau.

Abbildung 2.2 bildet die Kovarianzmatrizen der Produkte ab. Die Grösse der Punkte gibt die Kovarianz wider. Die Diagonale bildet die Varianz ab. Verschiedene Modelle fordern eine Homogenität der Varianz. Die Varianz soll dabei konstant sein und die Kovarianz zwischen allen Produkten dieselbe. Bei beiden Abbildungen wird die Homogenitätsannahme nicht grob verletzt.

Der Mosaicplot in 2.3 zeigt das Verhältnis zwischen kategoriellen Variablen. Es hat haben mehr Frauen wie Männer am Test teilgenommen und die mittlere Altersgruppe ist mit dem grössten Anteil vertreten. Mosaicplots können beliebig erweitert werden, beispielsweise durch den Faktor Sprache. Ab mehr als drei Faktoren werden sie aber unübersichtlich.

In Abbildung 2.4 sind Boxplots für die einzelnen Produkte dargestellt. Das Produkt von Coop und Lindt D hat eine geringere Streuung wie die Produkte. Die Lage der Produkte variiert ebenfalls ein wenig. Die Forderung der konstanten Varianz ist aber nicht verletzt. Die Box umfasst das 1. bis zum 3. Quartil. Der horizontale Strich kennzeichnet den Median. Die senkrechten Linien gehen bis zum 1. Quartil - 1.5 der Interquartilsdifferenz, resp. 3. Quartil + 1.5 der Interquartilsdifferenz. Die Interquartilsdifferenz ist die Differenz zwischen dem 3. und dem 1. Quartil. Werte, die sich ausserhalb dieses Bereichs befinden, werden als schwarze Punkte gekenn- und als Ausreisser bezeichnet. Ein Boxplot beinhaltet viel Information über die Verteilung. Die Lage des Medians gibt an, ob die Verteilung symmetrisch ist oder schief. Die Boxlänge und die senkrechten Linien geben Information über die Lage der Werte und ob die Verteilung eher kurz- oder langschwänzig ist. Werden mehrere Boxplots nebeneinander aufgeführt können die Verteilungen miteinander verglichen werden.

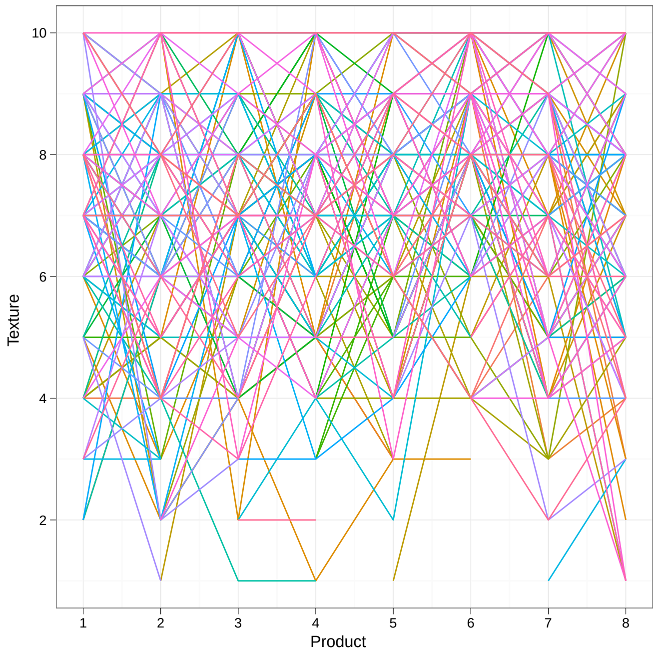
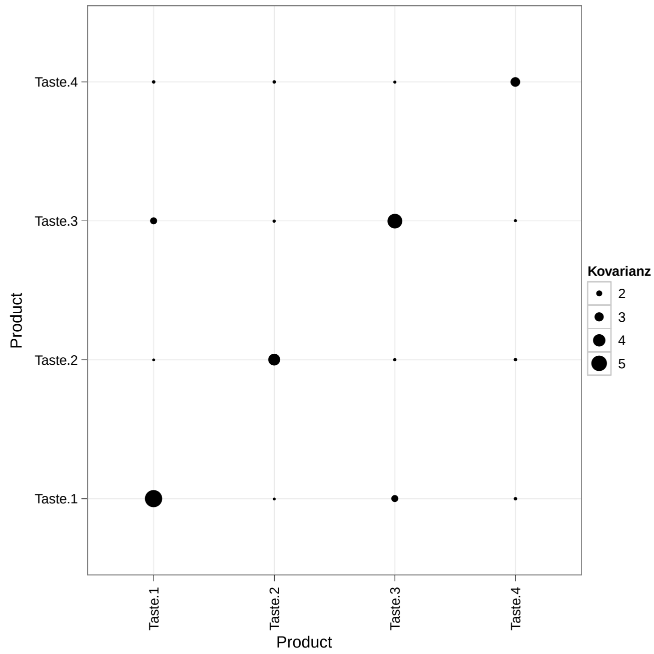


Abbildung 2.1.: Spaghetti-Plot der Produkte. Die Linien stellen die einzelnen Personen dar. Die Nummerierung ist wie folgt: Coop, Frey, Cailler, Lindt CH, Lindt F, Lindt D, Aldi, Ritter



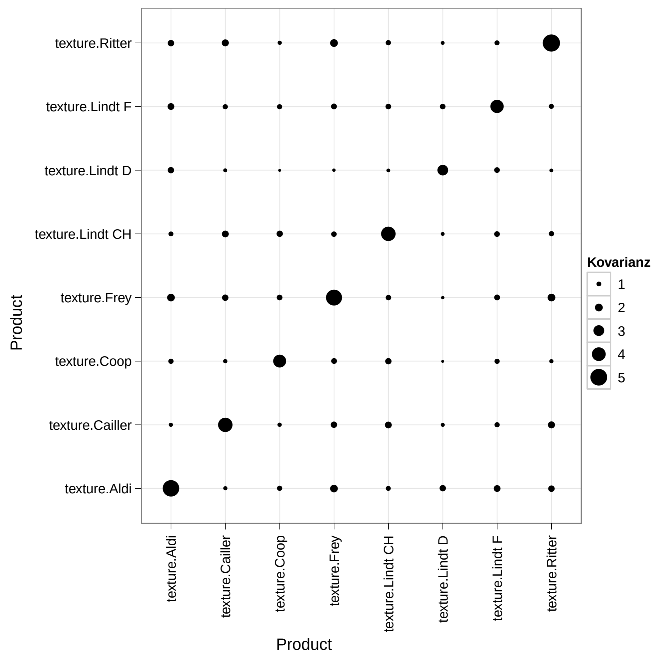


Abbildung 2.2.: Kovarianz der Produkte des Datensatz A für die Wertung des Geschmacks (oben) und Kovarianz der Produkte des Datensatz B für die Wertung der Beschaffenheit (unten)

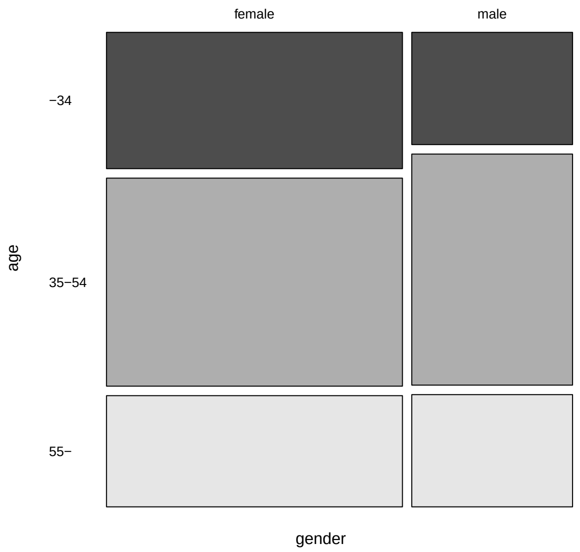


Abbildung 2.3.: Mosaicplot der Alters- und Geschlechtergruppen des Datensatz B

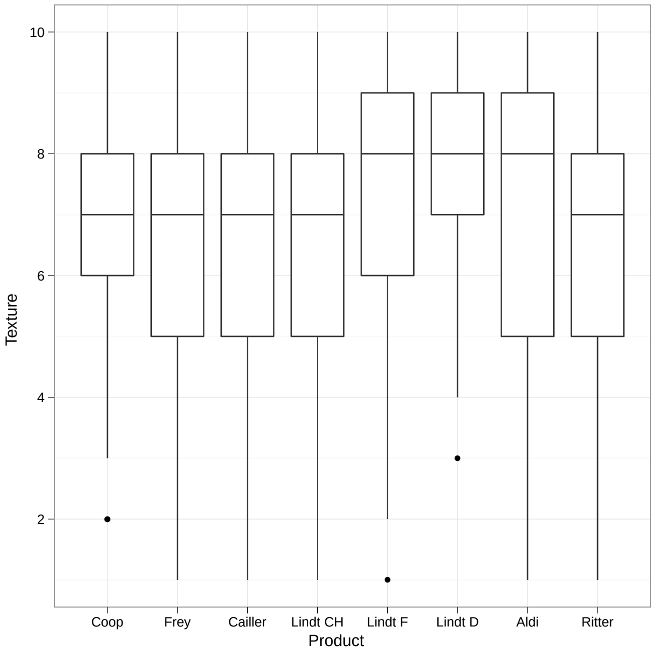


Abbildung 2.4.: Boxplots der Wertungen der einzelnen Produkte des Datensatz B

3. Modellübersicht

Wie in der Einleitung des Berichts bereits angesprochen entstprechen die Datensätze von Lindt & Sprüngli der Auswertung von Repeated Measures, resp. Longitudinal-Daten.

Für Longitudinal-Daten gibt es mehrere mögliche Auswertungsmöglichkeiten. *Longitudinal Data Analysis* von Hedeker und Gibbons (2006) [3] liefert einen guten Überblick über die Möglichkeiten, ebenfalls Kapitel 13, 14 und 15.4 in *Design and Analysis of Experiments* von Montgomery (2005) [4].

Die verschiedenen Modelle unterscheiden sich durch verschiedene Modellannahmen und die Macht des Modells. Es gibt 5 zentrale Annahmen, die variieren können.

- Varianz-Kovarianz-Struktur
- Umgang mit fehlenden Werten
- Restriktion an die Balancierung des Designs
- Verteilung der Zielvariablen

a. ANOVA ohne Personeneffekt

Es wird eine einfache ANOVA-Tabelle berechnet, ohne Berücksichtigung des Personeneffekts. D.h. das Cross-Over-Design wird zum Cross-Sectional-Design reduziert. Produkteeffekte werden in der Regel unterschätzt und die restlichen Effekte überschätzt. Diese Variante wurde bisher angewendet und wird in Kapitel 4 nochmals aufgegriffen.

Beispiel Lindt & Sprüngli: Die Zielvariable ist der Geschmack oder die Beschaffenheit. Die erklärenden Faktoren sind die Produkte, das Alter und das Geschlecht.

b. Repeated Measures ANOVA

Bei dem Repeated Measures ANOVA wird in der ANOVA-Tabelle der Personeeffekt berücksichtigt. Wichtig ist die Abgrenzung zum vollstängig randomisierten Design, weil die F-Tests unterschiedlich gebildet werden. Ein Repeated Measures ANOVA ist ein einfaches Mixed Model mit einem zufälligen Term. Es gilt die Homogenitätsannahme (konstante Varianz der Produkte und konstante Kovarianz zwischen den Produkten). Die Kovarianzen für ähnliche Objekte sind oft höher als für unähnliche Objekte. Details dazu in Kapitel 5 oder in d. Separate Mixed Models.

Beispiel Lindt & Sprüngli: Die Zielvariable ist der Geschmack oder die Beschaffenheit. Die erklärenden Variablen bestehen aus festen Faktoren und einem zufälligen Faktor. Die festen Faktoren sind die Produkte, das Alter und das Geschlecht. Der zufällige Term ist der Personeneffekt. Der Personeneffekt wird als zufälliger Faktor modelliert. Die F-Tests der within-subject *(Produkte) und* between-subject *(Alter, Geschlecht) Faktoren werden unterschiedlich konstruiert.*

c. Mixed Model (Linear Mixed Effects Regressionmodel)

Mixed Models sind Regressionsmodelle mit festen Effekten und einem oder mehreren zufälligen Effekten. Sie eignen sich sowohl für normalverteilte, als auch für nicht-normalverteilte Zielvariablen. Mixed Models können gut mit fehlenden Werten umgehen und können produkteab- und produkteunabhängige Faktoren beinhalten. Details dazu in Kapitel 6.

Beispiel Lindt & Sprüngli: Die Zielvariable ist der Geschmack oder die Beschaffenheit. Die erklärenden Variablen bestehen aus festen und einem oder mehreren zufälligen Faktoren. Die festen Faktoren sind die Produkte, das Alter und das Geschlecht. Der zufällige Term ist der Personeneffekt (Intercept) und der Personeneffekt innerhalb der Altersgruppen (Slope). Der Personeneffekt innerhalb einer Altersgruppe beziffert die Variabilität der Bewertungen zwischen den Personen in dieser Altersgruppe variieren.

d. Separate Mixed Models

Falls die Homogenitätsannahme verletzt wird, können die Analysen für die Produkte separat durchgeführt werden.

Beispiel Lindt & Sprüngli: Es werden 4 Produkte getestet. Zwei davon sind mit Vanille versetzt und haben eine weisse Schicht, sie unterscheiden sich nur geringfügig voneinander. Die anderen zwei Produkte sind dunkel und sind mit Haselnusssplitter versetzt und schmecken nach dunkler Schokolade, sie unterscheiden sich nur geringfügig voneinander. Falls die Testpersonen in ihrer Bewertung ebenfalls Untergruppen bilden und diese ähnlich bewerten, wird die Kovarianz innerhalb der Untergruppen Vanille und Haselnuss ähnlich sein, zwischen einander aber nicht. Hier ist die Homogenitätsannahme verletzt. Eine Möglichkeit für eine Analyse bietet eine getrennte Auswertung nach Vanille und Haselnuss.

Diese Analyse wird nicht weiter ausgeführt.

e. weitere Möglichkeiten

Die Varianzanalyse für multivariate Zielvariablen MANOVA lässt keine Missings zu in der Zielvariable zu. Es besteht keine Homogenitätsannahme. Als weitere mögliche Modelle sind die Stichworte GEE Models oder Covariance Pattern Models gegeben.

4. ANOVA ohne Personeneffekt

Die einfachste Analyse beruht auf der Idee von Mittelwertvergleichen. Dabei werden Faktormittel mit dem Gesamtmittel verglichen. Sind diese quadrierten Abweichungen im Verhältnis zu den Abweichungen innerhalb eines Faktors genug gross, so ist ein Effekt vorhanden.

Seite 42 bis 45 aus *Experimental Design* von Müller (2006) [1] zeigt den Aufbau einer Varianzanalyse mit mehreren Faktoren im Detail. An dieser Stelle ein Auszug davon.

Formel (4.1) enthält eine Zielgrösse y, das Gesamtmittel, dazu eine oder mehrere abhängige erklärende Faktorvariablen mit Wechselwirkungen und einen Fehlerterm ϵ. Die erklärenden Faktoren sind gekreuzt.

$$y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \epsilon_{ijk} \tag{4.1}$$

y_{ijk} ist die k-te Beobachtung mit Faktor A auf Level i und Faktor B auf Level j, $k = 1, \dots, K$
μ ist das Gesamtmittel
A_i ist der Effekt des i-ten Levels des Faktors A, $i = 1, \dots, I$
B_j ist der Effekt des j-ten Levels des Faktors B, $j = 1, \dots, J$
$(AB)_{ij}$ ist die Interaktion zwischen des i-ten Levels von A und des j-ten Levels von B
ϵ_{ijk} ist der zufällige Fehlerterm mit $\epsilon_{ijk} \backsim$ iid. $N(0, \sigma_\epsilon^2)$

Die Nebenbedingungen sind $\sum_i A_i = \sum_j B_j = \sum_i (AB)_{ij} = \sum_j (AB)_{ij} = 0$.

a. Fehlende Werte

Beobachtungen mit fehlenden Werte werden aus der Analyse ausgeschlossen.

b. Balanciertheit

Die F-Tests in der ANOVA-Tabelle im nächsten Abschnitt c. ANOVA gelten nur, falls jede Faktorkombination gleich viele Replikate hat, sprich, die Daten balanciert sind. Ist das nicht der Fall ändern sich die Parameterschätzer und die ANOVA-Tabelle und damit die Schätzungen und dazugehörigen Vertrauensintervalle.

Für ANOVA treten im unbalancierten Fall drei Arten auf, wie die F-Tests zu bestimmen sind. Auf den Seiten 34 und 35 im Skript *Angewandte Varianzanalyse und Versuchsplanung* von Roth (2009) [2] ist erklärt, wie sich die Zusammensetzung der Sum of Squares ändern.

- Typ I
  Als Referenz wird immer das vorhergehende Modell verwendet, die Reihenfolge wie das Modell geschrieben wird, ist also wesentlich.
- Typ II
  Als Referenz dient das Modell mit allen Haupteffekten.
- Typ III
  Als Referenz dient das volle Modell.

Typ I wird nur verwendet, wenn ein Modell mit einer klar vorgegebenen Reihenfolge vorhanden ist.

c. ANOVA

Das Modell aus 4.1 lässt sich auch mit *Sum of Squares* ausdrücken. Damit erhält man die klassische ANOVA-Tabelle, deren Nullhypothese der F-Tests ist, dass alle Levels eines Faktors gleich Null sind. Die Alternativhypothese ist, dass mindestens ein Level eines Faktors ungleich Null ist.

Faktor	df	SS	MS	E(MS)	F-Werte
A	$(I-1)$	$\mathrm{SS}_A = \sum_i \sum_j \sum_k (y_{i..} - y_{...})^2$	$\frac{SS_A}{I-1}$		$\frac{MS_A}{MS_R}$
B	$(J-1)$	$SS_B = \sum_i \sum_j \sum_k (y_{.j.} - y_{...})^2$	$\frac{SS_B}{J-1}$		$\frac{MS_B}{MS_R}$
AB	$(I-1)(J-1)$	$SS_{AB} = \sum_i \sum_j \sum_k (y_{ij.} - y_{i..} - y_{.j.} - y_{...})^2$	$\frac{SS_{(AB)}}{(I-1)(J-1)}$		$\frac{MS_{(AB)}}{MS_R}$
R	$IJK - IJ$	$SS_R = SS_{tot} - SS_A - SS_B - SS_{AB}$	$\frac{SS_R}{IJK-IJ}$	σ_ϵ^2	
Total	$IJK - 1$	$\mathrm{SS}_{Total} = \sum_i \sum_j \sum_k (y_{ijk} - y_{...})^2$			

d. Schätzungen und Vertrauensintervalle

Die Schätzungen der Zielvariable entsprechen der Formel (4.2), die Varianz der Schätzung wird durch die Kovarianzmatrix der Koeffizienten wie in Formel[1] (4.3) und (4.4) berechnet[2]. Gemeinsam bilden die Schätzung und die Varianz das Vertrauensintervall aus Formel (4.5), wobei n die Anzahl Beobachtungen sind und p die Anzahl geschätzter Parameter.

$$\widehat{y}_{ijk} = \hat{\mu} + \hat{A}_i + \hat{B}_j + (\widehat{AB})_{ij} \tag{4.2}$$

$$\beta_w = [\hat{\mu}, \hat{A}_i, \hat{B}_j, (\widehat{AB})_{ij}] \tag{4.3}$$

$$\widehat{Var\langle y_{ijk}\rangle} = \sum_w \sum_{w'} Cov\langle \beta_w, \beta_{w'}\rangle \qquad \forall \quad w \tag{4.4}$$

$$\widehat{y}_{ijk} \pm q^t_{1-\alpha/2,n-p} \sqrt{\widehat{Var\langle y_{ijk}\rangle}} \tag{4.5}$$

e. Residuenanalyse

Mit der Residuenanalyse wird überprüft, ob die Residuen normalverteilt sind (Q-Q-Plot) und ob der Erwartungswert der Residuen Null ist und die Varianz konstant (Tukey-Anscombe-Plot).

4.1. Anwendung Lindt & Sprüngli

Auf den Datensatz A und B angewendet, lässt sich Formel (4.1) wie in Formel (4.6) schreiben. Es werden nur die Haupteffekte eingesetzt, die Faktoren `run` und `lng` werden nicht verwendet. In diesem Modell wird der Personeneffekt nicht betrachtet. Diese Auswertung ist für die Lindt & Sprüngli Daten zwar nicht adäquat, aber für einen Vergleich der Resultate mit den Repeated Measures ANOVA hilfreich.

$$T_{hijk} = \mu + P_j + A_k + G_h + \epsilon_{hijk} \tag{4.6}$$

wobei:
T_{hijk} ist Bewertung des 'Taste', bzw. 'Texture' der i-ten Beobachtung mit der Altersgruppe k, dem Geschlecht h, dem Produkt j, $i = 0 \ldots 519$
μ ist das Mittel über die Beobachtungen mit Alter $k = 1$, Geschlecht $h = 1$, Produkt

[1] β_w enthält nur die Parameterschätzer aus $\widehat{y}_{hij}$
[2] Die Klammern '[' und ']' fassen die Koeffizienten zu einem Vektor zusammen.

$j = 1$
A_k ist der Effekt der k-ten Altersgruppe, $k = 1, 2, 3$
G_h ist der Effekt des h-ten Geschlechts, $h = 1, 2$
P_j ist der Effekt des j-ten Produkt, $j = 1, 2, 3, 4$

$\epsilon_{hijk} \backsim N(0, \sigma^2_\epsilon)$

Die Nebenbedingungen lauten $A_1 = G_1 = P_1 = 0$.

4.1.1. Ergebnisse

Das Modell für den Datensatz A hat nur Haupteffekte, jedes für den Datensatz B (im Anhang) auch Wechselwirkungen. Die restlichen Resultate sind im Anhang A.1.

taste $\backsim$ product + age + gender, data = datA

```
aov(taste ~ product + age + gender , data = datA)

# Analysis of Variance Table
# ---------------------------
# Response: Taste
#             Df Sum Sq Mean Sq F value    Pr(>F)
# product      3  567.3   189.1 44.0546 < 2.2e-16
# age          2   51.9    26.0  6.0467  0.002408
# gender       1    6.8     6.8  1.5894  0.207549
# Residuals 2069 8880.4     4.3
```

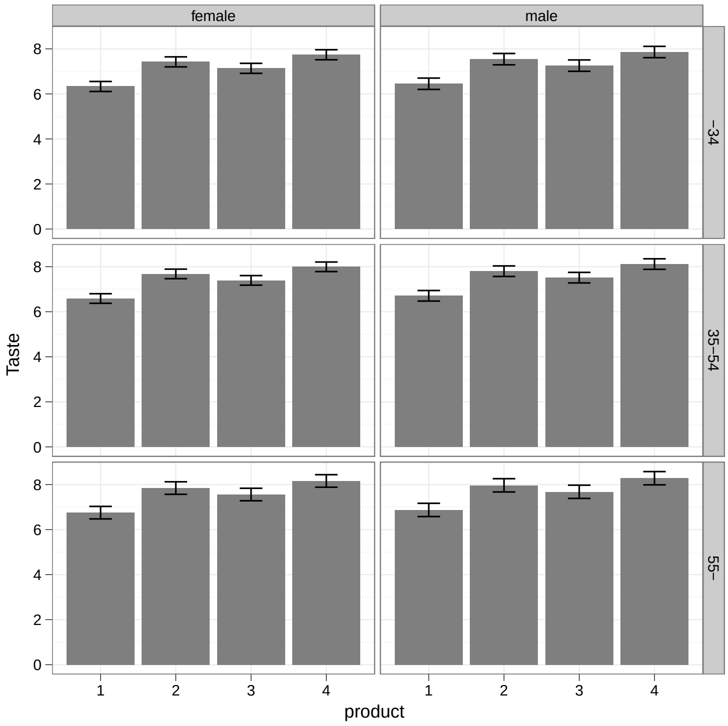


Abbildung 4.1.: Datensatz A: Schätzungen der Bewertungen des Geschmacks für das Alter, Geschlecht und Produkt

4.2. Bemerkung

Die Anwendung der einfachen Varianzanalyse ohne Berücksichtigung des Personeneffekts ist insofern unbefriedigend, als das erstens in diesem Modell nicht alle vorhandene Information genutzt wird. Und zweitens werden die F-Werte nicht korrekt berechnet: gundsätzlich wird bei diesem Design der P-Wert des Faktor A (Altersgruppe und/oder Geschlecht) aus der ANOVA-Tabelle der Formel 4.1 zu klein, jener von Faktor B (Produkt) zu gross. Die Punktschätzungen der Effekte sind zwar richtig, aber das Vertrauensintervall der Produkteeffekte wird zu gross, jenes der Altersgruppen und des Geschlechts zu klein.

5. Repeated Measures ANOVA

Bei Repeated Measures ANOVA[1] wird zusätzlich noch der Personeneffekt als zufälliger Effekt eingeführt. Bei der Hinzunahme des Personeneffekts ist die korrekte Erkennung der Datenstruktur wichtig. Repeated Measures Modelle haben einen zweistufigen Randomisierungsprozess und daher zwei Fehlerterme, nämlich die Variabilität zwischen den Personen (*between*) und die Variabilität innerhalb der Beobachtungen einer Person (*within*). Die F-Tests der *between-subject* Faktoren werden anders konstruiert als diejenigen der *within-subject* Faktoren. Hingegen wird bei vollständig randomisierten Designs nicht zwischen *between* und *within* unterschieden. Mit der Hinzunahme des Personeneffekts werden zwei Dinge gewonnen: Erstens stimmen die P-Werte der F-Tests (die im vorangehenden Kapitel unter- bzw. überschätzt wurden), zweitens ist die Streuung zwischen der Personen bekannt. Weil es sich beim Personeneffekt um einen zufälligen Effekt handelt, wird das Design auch *Univariates Mixed Model* genannt. Im Kapitel 6 werden erweiterte Mixed Model behandelt.

In Kapitel 15.4 in *Design and Analysis of Experiments* von Montgomery (2005) [4] ist ein Beispiel eines Repeated Measures ANOVA aufgeführt. Kapitel 14 zeigt Split-Plot Modelle. Kapitel 2 in *Longitudinal Data Analysis* von Hedeker und Gibbons (2006) [3] zeigen ein *einfaches* Repeated Measures Modell und die Modellstruktur eines *multiplen*[2] Repeated Measures Modells. Von diesem Kapitel stammen auch die untenstehenden Ausführungen in Kapitel 5.1.

5.1. Theorie

Bei *multiplen* Repeated Measures ANOVA gibt es im einfachsten Fall einen zufälligen und zwei feste Faktoren. Zum Beispiel: Personen (zufällig), die Produkte (fest) und eine Gruppierung (fest). Beim Personeneffekt gibt es verschiedene Gründe, weshalb man diesen Faktor als zufällig modelliert. Diese Bedingung ist einerseits durch das Repeated Measures Modell so vorgegeben. Andererseits wurden die Personen zufällig ausgewählt. Das Experiment ist also nicht reproduzierbar und genaue Parameterschätzungen pro Person machen daher keinen Sinn. Hinzu kommt, dass genaue Parameterschätzungen oft

[1] Weitere Bezeichnungen für Repeated Measures ANOVA: Split-Plot-Design, Univariates Mixed Model

[2] Die Theorie der Repeated Measures ANOVA teilt sich in Modelle mit *einem* und solche mit *mehreren* festen Faktoren auf. Das Modell von Lindt & Sprünli handelt es sich um ein *multiple* Repeated Measures ANOVA.

gar nicht interessieren, sondern nur, wie fest sie voneinander abweichen und das wird mit der Variabilität zwischen den Personen ausgedrückt.

Beim Repeated Measures ANOVA Modell in der Gleichung 5.1 werden die Personen S die ein Produkt P testen einer Gruppe G zugeordnet. Die Personen sind in den Gruppen geschachtelt[3], was durch durch die Schreibweise $S_{i(h)}$ angedeutet wird.

$$y_{hij} = \mu + G_h + P_j + (GP)_{hj} + S_{i(h)} + \epsilon_{hij} \tag{5.1}$$

y_{hij} ist die Beobachtung der Person i in der Gruppe h und dem Produkt j
μ ist das Gesamtmittel
G_h ist der Effekt der h-ten Gruppe, $h = 1, \ldots, H$
P_j ist der Effekt des j-ten Produkts, $j = 1, \ldots, J$
$(GP)_{hj}$ ist der Effekt der Interaktion zwischen der Gruppe h und der Produkt j
$S_{i(h)}$ ist der Effekt der Person i aus der h-ten Gruppe, $i = 1, \ldots, N_h$, $\sum_h N_h = N$
ϵ_{hij} ist der Fehler für Person i in der Gruppe h beim Produkt j

Es gibt $N = IJ$ Beobachtungen.

$S_{i(h)} \backsim N(0, \sigma_S^2)$
$\epsilon_{hij} \backsim N(0, \sigma_\epsilon^2)$

σ_ϵ^2 ist die Variabilität von Bewertung zu Bewertung innerhalb der gleichen Personen. σ_S^2 ist die Variabilität zwischen den Personen.

Die Nebenbedingungen lauten $\sum_j P = \sum_h G = \sum_h (GP) = \sum_j (GP) = 0$.

a. Varianzstruktur

Für das Modell ist eine Varianzstruktur wie in (5.2) vorgegeben. Die Homogenitätsannahme fordert eine konstante Varianz über alle Beobachtungen die sich aus der Variabilität innerhalb der Personen und zwischen den Personen zusammensetzt und eine konstante Kovarianz zwischen den Produkten die der Variabilität zwischen den Personen entspricht. Die Überprüfung erfolgt auf den Originaldaten, indem die Kovarianzen zwischen den Produkten ausgerechnet werden. Die Darstellung kann wie in Kapitel 2.3 erfolgen.

$$\begin{aligned} V\langle y_{ij}\rangle &= \sigma_\epsilon^2 + \sigma_S^2 \quad \forall \quad j, \\ Cov\langle y_{ij}, y_{ij'}\rangle &= \sigma_S^2 \quad \forall \quad j \text{ und } j' \quad (j \neq j') \end{aligned} \tag{5.2}$$

$$Corr\langle y_{ij}, y_{ij'}\rangle = \frac{\sigma_S^2}{\sigma_\epsilon^2 + \sigma_S^2} \tag{5.3}$$

[3] Faktor A ist in Faktor B geschachtelt bedeutet, dass pro Level von Faktor B nur ein Teil der Levels vom Faktor A vorkommen.

Ein Beispiel für ein Design, bei denen eventuell Vorsicht geboten ist: Es werden 6 Produkte getestet. Sie teilen sich in zwei Gruppen auf, die innerhalb der Gruppe sehr homogen in den Produkteeigenschaften sind, sich untereinander aber stark unterscheiden. Wenn die Personen die Produkte innerhalb der Produktegruppen ähnlich bewerten, dann gibt es beispielsweise eine nicht-konstante Kovarianzstruktur.

Die Homogenitätsannahme kann auch durch statistische Tests überprüft werden. Leider sind die Tests sehr sensitiv und verlangen gewisse Rahmenbedingungen. Beide Tests sind in Hedeker und Gibbons (2006) [3] erklärt.

Wird die Homogenitätsannahme abgelehnt, kann auf multivariate Repeated Measures (MANOVA) zurückgegriffen werden. Als Alternative bietet sich eine separierte Analyse von Repeated Measures ANOVA an.

Formel (5.4) zeigt, dass die Bewertungen zwischen einzelnen Personen unabhängig voneinander sind.

$$Cov\langle y_{ij}, y_{i'j}\rangle = 0 \quad \forall \quad i \text{ and } i' \quad (i \neq i') \tag{5.4}$$

b. Fehlende Werte

Für ein Repeated Measures ANOVA gilt eine Balancierung im Sinn der Produkte - alle Personen müssen alle Produkte bewerten. Ist das nicht der Fall, werden die Personen aus dem Datensatz entfernt, weil bei blosser Streichung der fehlenden Beobachtungen der Datensatz im Sinn der Produkte unbalanciert wäre.

c. Balanciertheit

Im Sinn der Anzahl Produkte muss der Datensatz balanciert sein. Allerdings kann in den Gruppen die Anzahl Personen N_h variieren.

d. ANOVA

Die ANOVA-Tabelle gilt für den Fall, dass alle Personen alle Produkte bewertet haben. Die Gruppengrösse darf variieren.

Faktor	df	SS	MS	E(MS)	Tests
G	$(H-1)$	$SS_G = J\sum_h N_h(y_{h..} - y_{...})^2$	$\frac{SS_G}{H-1}$	$\sigma_\epsilon^2 + n\sigma_P^2 + D_G$	(5.7)
P	$(J-1)$	$SS_P = N\sum_j(y_{..j} - y_{...})^2$	$\frac{SS_P}{J-1}$	$\sigma_\epsilon^2 + D_P$	(5.6)
GP	$(H-1)(J-1)$	$SS_{GP} = \sum_h\sum_j N_h(y_{h.j} - y_{h..} - y_{..j} + y_{...})^2$	$\frac{SS_{(GP)}}{(H-1)(J-1)}$	$\sigma_\epsilon^2 + D_{GP}$	(5.5)
S(G)	$(N-H)$	$SS_{S(G)} = J\sum_h\sum_i(y_{hi.} - y_{h..})^2$	$\frac{SS_{S(G)}}{N-H}$	$\sigma_\epsilon^2 + J\sigma_P^2$	(5.8)
R	$(N-H)(J-1)$	$SS_R = \sum_h\sum_i\sum_j(y_{hij} - y_{h.j} - y_{hi.} + y_{h..})^2$	$\frac{SS_R}{(N-H)(J-1)}$	σ_ϵ^2	
Total	$NJ-1$	$SS_{Total} = \sum_h\sum_i\sum_j(y_{hij} - y_{...})^2$			

D_G, D_P, D_{GP} sind die summierten Differenzen über die Gruppen, die Produkte und die Interaktion Gruppe und Produkt.

Tests für Gruppe G und Produkt P

Die F-Tests der *between-subject* Faktoren werden mit $MS_{S(G)}$ gebildet, jene der *within-subject* Faktoren mit MS_R. Als *within-subject* Faktor gilt auch die Wechselwirkung zwischen dem Produkt und der Gruppe. Falls mehrere Gruppierungen vorhanden sind (z.B. Alter und Geschlecht) werden diese mit $MS_{S(G)}$ getestet.

Die Interaktion zwischen Gruppe und dem Produkt wird wie in Formel (5.5) getestet. Falls H_{GP} verworfen wird, wird angenommen, das der Gruppeneffekt pro Produkte unterschiedlich ist.

$$H_{GP} : D_{GP} = 0, \qquad F_{GP} = \frac{MS_{GP}}{MS_R} \overset{\text{unter } H_{GP}}{\backsim} F_{(H-1)(J-1),(N-H)(J-1)} \tag{5.5}$$

Falls H_{GP} nicht verworfen wird, können H_G und H_P getestet werden.

$$H_P : P_1 = \ldots = P_n = 0, \qquad F_P = \frac{MS_P}{MS_R} \overset{\text{unter } H_P}{\backsim} F_{(J-1),(N-H)(J-1)} \tag{5.6}$$

$$H_G : G_1 = \ldots = G_s = 0, \qquad F_G = \frac{MS_G}{MS_{S(G)}} \overset{\text{unter } H_G}{\backsim} F_{(H-1),(N-H)} \tag{5.7}$$

Tests für Personeneffekt $S(G)$

Der Personeneffekt wird wie in Formel (5.8) getestet.

$$H_{S(G)} : \sigma_\epsilon^2 = 0, \qquad F_{S(G)} = \frac{MS_{S(G)}}{MS_R} \overset{\text{unter } H_{S(G)}}{\backsim} F_{(N-H),(N-H)(J-1)} \tag{5.8}$$

Zusätzlich oder alternativ zum Testen kann die *Intraclass Correlation* wie in Formel (5.9) berechnet werden - die Variabilität zwischen den Personen kann auch als einen Anteil an der gesamten Rest-Variabilität der Beobachtungen angesehen werden (Variabilität zwischen den Personen und Variabilität innerhalb der Personen). Die *Intraclass Correlation* nimmt Werte zwischen 0 und 1 an. Ist die *Intraclass Correlation* nahe bei 0, kann angenommen werden, dass die Bewertungen zwischen den Personen nicht stark variiieren, die Korrelation zwischen den Produkten klein und kein Personeneffekt vorhanden ist. Liegt die *Intraclass Correlation* nahe bei 1 variieren die Bewertungen zwischen den Personen stark und die Wertungen innerhalb einer Person sind hoch korreliert.

$$\widehat{ICC} = \frac{\hat{\sigma}_S^2}{\hat{\sigma}_\epsilon^2 + \hat{\sigma}_S^2} \tag{5.9}$$

e. Schätzungen und Vertrauensintervalle

Für die Schätzung des Modells werden nur die festen Effekte berücksichtigt. Die Schätzungen entsprechen der Formel (5.10), die Varianz der Schätzung wird durch die Kovarianzmatrix der Koeffizienten wie in Formel (5.11) und (5.12) berechnet[4]. Gemeinsam bilden die Schätzung und die Varianz das Vertrauensintervall in Formel (5.13), wobei n die Anzahl Beobachtungen sind und p die Anzahl geschätzter fester Effekte (inklusive Gesamtmittel).

$$\widehat{y}_{hij} = \hat{\mu} + \hat{G}_h + \hat{P}_j + (\widehat{GP})_{hj} \tag{5.10}$$

$$\beta_w = [\hat{\mu}, \hat{G}_h, \hat{P}_j, (\widehat{GP})_{ij}] \tag{5.11}$$

$$\widehat{Var\langle y_{hij}\rangle} = \sum_w \sum_{w'} Cov\langle \beta_w, \beta_{w'} \rangle \qquad \forall \quad w \tag{5.12}$$

$$\widehat{y}_{hij} \pm q^t_{1-\alpha/2,n-p} \sqrt{\widehat{Var\langle y_{hij}\rangle}} \tag{5.13}$$

[4] β_k enthält nur die Parameterschätzer aus y_{hij}

f. Residuenanalyse

Bei Mixed Models gibt es durch die zwei Fehlerterme zwei Sorten von Residuen.

- Residuum I = Beobachtung - feste Effekte - zufälliger Effekt
- Residuum II = Beobachtung - feste Effekte

Mit dem Tukey-Anscombe-Plot kann die konstante Varianz und der Erwartungswert gleich 0 des Fehlerterms anhand der Residuen überprüft werden.

5.2. Anwendung Lindt & Sprüngli

Das Repeated Measures Modell für die Lindt & Sprüngli Daten ist in der Gleichung (5.14) aufgeführt. Der Term $S_{i(kh)}$ bezeichnet den Personeneffekt.

$$y_{hijk} = \mu + G_h + A_k + (AG)_{kh} + P_j + (PG)_{jh} + (PA)_{jk} + S_{i(kh)} + \epsilon_{hijk} \tag{5.14}$$

y_{hijk} ist die Bewertung des 'Taste' bzw. 'Texture' des Produkt j von der Person i mit der Altersgruppe k, dem Geschlecht h
μ ist das Mittel über die Beobachtungen mit Alter $k = 1$, Geschlecht $h = 1$, Produkt $j = 1$
G_h ist der Effekt des h-ten Geschlecht, $h = 1, 2$
A_k ist der Effekt der k-ten Altersgruppe, $k = 1, 2, 3$
P_j ist der Effekt des j-ten Produkts, $j = 1, ..., 8$
$(PA)_{jk}$ ist die Wechselwirkung zwischen dem j-ten Produkt und der k-ten Altersgruppe
$(PG)_{jh}$ ist die Wechselwirkung zwischen dem j-ten Produkt und dem h-ten Geschlecht
$(AG)_{kh}$ ist die Wechselwirkung zwischen der k-ten Altersgruppe und dem h-ten Geschlecht
$S_{i(kh)}$ ist der Effekt der Person i, die Personen sind in den Altersgruppen k und dem Geschlecht h geschachtelt, $i = 1, \ldots, N_{kh}$, $\sum_k \sum_h N_{kh} = N$
ϵ_{hijk} ist der Fehler für Person i mit dem Geschlecht h in der Altersgruppe k beim Produkt j.

$S_{i(h,k)} \backsim N(0, \sigma_S^2)$
$\epsilon_{hij} \backsim N(0, \sigma_\epsilon^2)$

Die Nebenbedingungen der Koeffizienten lauten $G_1 = A_1 = P_1 = (AG)_{11} = (AG)_{21} = (AG)_{31} = (AG)_{12} = (PA)_{11} = (PA)_{12} = (PA)_{13} = (PA)_{21} = (PA)_{31} = (PA)_{41} = (PA)_{51} = (PA)_{61} = (PA)_{71} = (PA)_{81} = (PG)_{11} = (PG)_{12} = (PG)_{21} = (PG)_{31} = (PG)_{41} = (PG)_{51} = (PG)_{61} = (PG)_{71} = (PG)_{81} = 0$

5.2.1. Ergebnisse

Als Beispiel für ein Ergebnis einer Repeated Measures ANOVA wird das Modell mit der Zielvariable Taste vom Datensatz A gezeigt. Die restlichen Resultate sind im Anhang A.2.

taste ∽ product + age + gender + Wechselwirkungen + Error(subject), data=datA

```
mod <- aov(taste ~ (product + age + gender)^2 + Error(subject), data = datA)

#Error: subject
#            Df Sum Sq Mean Sq F value  Pr(>F)
#age          2   51.9    26.0  3.1313 0.04449 *
#gender       1    6.8     6.8  0.8231 0.36470
#age:gender   2    3.8     1.9  0.2295 0.79498
#Residuals  513 4251.9     8.3
#---
#Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#
#Error: Within
#                 Df Sum Sq Mean Sq F value Pr(>F)
#product           3  567.3   189.1 63.4361 <2e-16 ***
#product:age       6   16.5     2.7  0.9207 0.4788
#product:gender    3    3.0     1.0  0.3340 0.8008
#Residuals      1545 4605.3     3.0
```

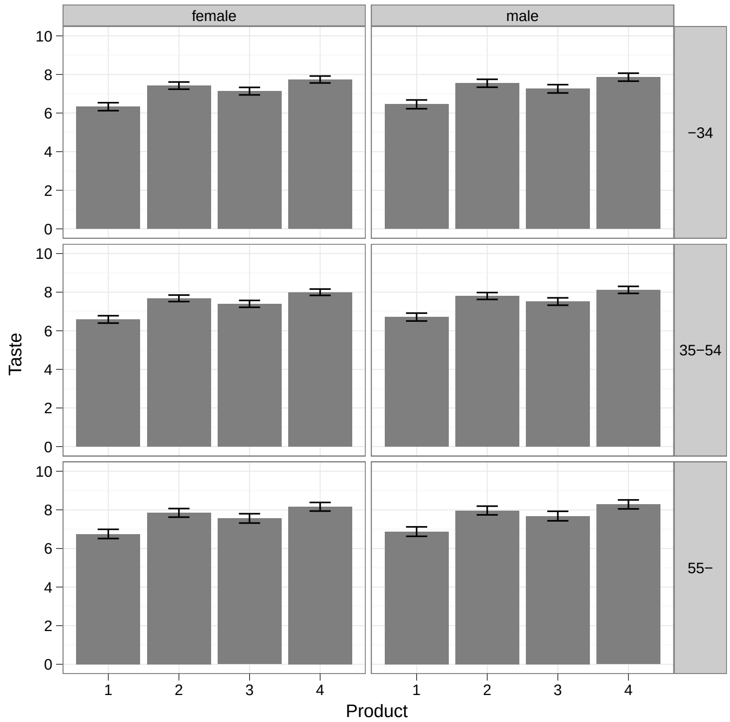


Abbildung 5.1.: Datensatz A: Schätzungen der Bewertungen des Geschmacks durch ein Repeated Measures Modell für das Alter, Geschlecht und Produkt.

5.3. Bemerkung

Repeated Measures ANOVA bieten bei den Daten von Lindt & Sprüngli eine korrekte Analyse. Falls eine Unterteilung des Personeneffekts auf verschiedene Altersgruppen oder das Geschlecht nötig ist oder unbalancierte Daten vorliegen (Datensatz B) können erweiterte Mixed Models angewendet werden.

6. Linear Mixed Effects Regression Model

Die Erweiterung des Repeated Measures ANOVA - sogenannten traditionellen Mixed Models - mit einem zusätzlichen zufälligen Slope-Effekt führt zu einem *Linear Mixed Effects Regression Model*. Im Vergleich zum univariaten Mixed Model sind mehrere zufällige Terme erlaubt.

Im Zusammenhang mit Repeated Measures Daten wird die Implementation von Mixed Models dann benötigt, wenn ein Datensatz fehlende Werte beinhaltet. Ein Mixed Model kann aber auch angebracht sein, wenn die Unterteilung des Personeneffkts auf eine Gruppierung nötig ist, weil sich die Variabilität der Bewertungen der Personen innerhalb der Gruppenlevels stark unterscheiden.

Einen Überblick über Mixed Models bietet sowohl Kapitel 13 aus *Design and Analysis of Experiments* von Montgomery (2005) [4], als auch Kapitel 4 aus *Longitudinal Data Analysis* von Hedeker und Gibbons (2006) [3], aus dem die nachführenden Ausführungen teilweise entnommen sind. *Mixed-effects models in S and S-Plus* von Pinheiro und Bates (2000) [5] ist ein Buch über die Anwendung von Mixed Models in `R`.

6.1. Theorie

Im Repeated Measures Modell der Gleichung (5.1) gibt es den zufälligen Term $S_{i(h)}$, dessen geschätzte Varianz $\hat{\sigma}_S^2$ die Variabilität zwischen den Personen beschreibt. In derselben Formel gibt es auch einen Faktor G, der die Personen in Gruppen einteilt. S ist in G geschachtelt. Wenn die Variabilität der Bewertungen der Personen innerhalb einer Gruppierung G_h interessiert, kann das im Model durch den zufälligen Term $S_{i(h)}^h$ ausgedrückt werden. Das entspricht quasi einer Wechselwirkung des zufälligen Terms $S_{i(h)}$ mit dem Term G_h. $S_{i(h)}^h$ ist ebenfalls normalverteilt mit Erwartungswert 0 und der Varianz $\sigma_{S^h}^2$. Die Variabilität der Bewertungen der Personen im ersten Level der Gruppe G_1 entspricht dem Term σ_S^2. $S_{i(h)}^h$ ist ein sogenannter Slope-, $S_{i(h)}$ ein Intercept-Fehlerterm. Gleichung (6.1) zeigt das ergänzte Modell.

$$y_{hij} = \mu + G_h + P_j + (GP)_{hj} + S_{i(h)} + S_{i(h)}^h + \epsilon_{hij} \tag{6.1}$$

y_{hij} ist die Beobachtung der Person i in der Gruppe h und dem Produkt j
μ ist das Gesamtmittel
G_h ist der Effekt der h-ten Gruppe, $h = 1, \ldots, H$

P_j ist der Effekt des j-ten Produkts, $j = 1, \ldots, J$
$(GP)_{hj}$ ist der Effekt der Interaktion zwischen der Gruppe h und der Produkt j
$S_{i(h)}$ ist der Effekt der Person i, i ist in h geschachtelt, $i = 1, \ldots, N_h, \sum_h N_h = N$
$S^h_{i(h)}$ ist der Effekt der Person i in der Gruppe h, i ist in h geschachtelt
ϵ_{hij} ist der Fehler für Person i in der Gruppe h beim Produkt j.

Die Nebenbedingungen lauten $\sum_j P = \sum_h G = \sum_j (GP) = \sum_h (GP) = 0$.

Der Term $S_{i(h)}$ wird in $S^h_{i(h)}$ integriert, weil $\sigma^2_{S_1} = \sigma^2_S$.

$S^h_{i(h)} \backsim \mathbf{N}(0, \Sigma_S)$
$\epsilon_{hij} \backsim N(0, \sigma^2_\epsilon)$

σ^2_ϵ ist die Variabilität von Bewertung zu Bewertung innerhalb der gleichen Personen. $\sigma^2_{S_h}$ ist die Variabilität der Personen in der Gruppe h. Σ_S ist die Kovarianzmatrix der multivariaten Normalverteilung der Terme $S_{i(h)}$ und $S^h_{i(h)}$ mit Erwartungswert 0.

a. Varianz-Kovarianz-Struktur

Für jedes Level h der Gruppe G_h gibt es ein anderes $\sigma^2_{S^h}$, ausser für $\sigma^2_{S^1}$, welches dem Gesamtmittel σ^2_S entspricht. Die geschätzte Varianz gibt der Terme gibt darüber Aufschluss, ob ein solcher Term überhaupt notwendig ist, also ob überhaupt ein Personeneffekt in diesem Gruppenlevel überhaupt exisitert. Die Korrelation gibt an, ob zwischen den zwei Termen eine positive oder negative Korrelation herrscht.

$$\mathbf{\Sigma_S} = \begin{pmatrix} \sigma^2_S & \sigma_S \sigma_{S^2} & \ldots & \sigma_S \sigma_{S^h} \\ \sigma_{S^2} \sigma_S & \sigma^2_{S^2} & \ldots & \sigma_{S^2} \sigma_{S^h} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{S^h} \sigma_S & \sigma_{S^h} \sigma_{S^2} & \ldots & \sigma^2_{S^h} \end{pmatrix}$$

Die geschätzte Korrelationsmatrix der zufälligen Terme wird beim Modelloutput der Programme ausgegeben und sollte beachtet werden um die Form der Varianz-Kovarianzstruktur zu überprüfen. Die Korrelation zwischen σ^2_S und $\sigma^2_{S_2}$ beträgt $(\sigma_S \sigma_{S_2})/(\sigma^2_S \sigma^2_{S_2})$. Für die Korrelationsmatrix werden die Seitenelemente also durch die entsprechenden Diagonalelemente geteilt.

Die Varianz-Kovarianz-Struktur wird nicht mehr vorab wie bei der Repeated Measures ANOVA, sondern im Nachhinein überprüft.

Hier geht man von korrelierten zufälligen Faktoren aus. Falls die Faktoren nicht miteinander korrelieren, können sie auch als unkorreliert modelliert werden.

a. ANOVA

Die Tests für *between-subject* und *within-subject* Faktoren entsprechen der ANOVA-Tabelle in Kapitel 5.

b. Fehlende Werte

Beobachtungen mit fehlenden Werten werden im Modell nicht berücksichtigt.

c. Schätzungen und Vertrauensintervalle

Die Schätzungen entsprechen der Formel (6.2), die Varianz der Schätzung wird durch die Kovarianzmatrix der Koeffizienten wie in Formel (6.3) und (6.4) berechnet[1]. Gemeinsam bilden die Schätzung und die Varianz das Vertrauensintervall aus Formel (6.5), wobei n die Anzahl Beobachtungen sind und p die Anzahl geschätzter fester Effekte (inklusive Gesamtmittel).

$$\widehat{y}_{hij} = \hat{\mu} + \hat{G}_h + \hat{P}_j + (\widehat{GP})_{hj} \tag{6.2}$$

$$\beta_w = [\hat{\mu}, \hat{G}_h, \hat{P}_j, (\widehat{GP})_{hj}] \tag{6.3}$$

$$\widehat{Var\langle y_{ijh}\rangle} = \sum_w \sum_{w'} Cov\langle \beta_w, \beta_{w'}\rangle \qquad \forall \quad w \tag{6.4}$$

$$\widehat{y}_{ijh} \pm q^t_{1-\alpha/2,n-p} \sqrt{\widehat{Var\langle y_{ijh}\rangle}} \tag{6.5}$$

Die Verteilung der Parameterschätzer und damit das Vertrauensintervall kann auch mit Bootstrap simuliert werden. Diese Methode wird in Kapitel 6.2 detaillierter besprochen.

d. Residuenanalyse

Bei Mixed Models gibt es durch die zwei Fehlerterme zwei Sorten von Residuen.

- Residuum I = Beobachtung - feste Effekte - zufälliger Effekt
- Residuum II = Beobachtung - feste Effekte

[1] β_w enthält nur die Parameterschätzer aus $\widehat{y}_{hij}$

Beide Typen können für die Residuenanalyse verwendet werden. Mit dem Tukey-Anscombe-Plot kann die konstante Varianz und der Erwartungswert gleich 0 des Fehlerterms anhand der Residuen überprüft werden.

e. Schätzmethoden

REML und ML sind verschiedene Schätzmethoden. REML bedeutet *Restricted Estimated Maximum Likelihood*[2] und ML *Maximum Likelihood.* Als Schätzmethode wird REML bevorzugt, weil ML-Schätzungen einen Bias haben. Bei balancieren Daten sind die REML- und ML-Schätzungen identisch.

6.2. P-Werte und Bootstrap

Im Fall eines balancierten Designs (alle Personen bewerten alle Produkte) sind die P-Werte der F-Tests in der ANOVA-Tabelle und die P-Werte der t-Tests der Parameterschätzungen korrekt. Im unbalancierten Design ist das nicht der Fall. Dass die F- und t-Verteilungen nur noch approximativ sind, hängt mit der Veränderung in der Schätzung der festen und zufälligen Parameter zusammen. Der Grad der Approximation der F- und t-Verteilungen hängt mit der Datenlage (Unbalanciertheit, fehlenden Werten, Anzahl Beobachtungen) zusammen. Bei der Funktion `lmer()` in der Software `R` werden keine P-Werte angegeben. Um dennoch ein Vertrauensintervall zu erhalten, kann mit der Bootstrap-Methode nachgeholfen werden, mit der die Verteilung von 'verteilungslosen' Schätzern approximativ bestimmt werden können.

Seite 87 bis 90 und 195 bis 201 in Harrell (2001) [6] erklären die Bootstrap-Methode. Bootstrap ist eine nicht-parametrische Methode um die Verteilung einer Zufallsvariable X oder deren Funktion $g(X)$ zu bestimmen[3].

Von einer Zufallsvariable X gibt es Realisationen, von denen aber die Verteilung unbekannt ist. Bei Bootstrap werden k Stichproben Λ_k mit folgendem Schema generiert: n Werte aus den Realisationen von X mit Zurücklegen ziehen (es gibt Elemente, die nicht in der Stichprobe erscheinen und welche, die mehrfach vorhanden sind). Für jede Stichprobe Λ_k kann die zu interessierende Statistik $g(\Lambda_i)$ berechnet werden (beispielsweise der Mittelwert). So hat man am Schluss die Verteilung von $g(X)$.

In unserem Anwendungsfall der Vertrauensintervalle verhält sich der oben beschriebene Algorithmus etwas anders. Wir haben einen Datensatz mit N Zeilen. Daraus werden 5000 Stichproben der Grösse N mit Zurücklegen gezogen. Mit jeder Stichprobe Λ_i wird ein Mixed Model mit k Parametern gerechnet und daraus jeweils die Parameter in einer Matrix gespeichert. Aus der Matrix der Grösse $k \times i$ können dann die Mittelwerte und die gewünschten Quantile pro Parameter k berechnet werden. Schematisch kann das gut

[2] Siehe auch Seite 75 in *Mixed-effects models in S and S-Plus* von Pinheiro und Bates (2000) [5].
[3] Bootrap engl. Steigbügel

mit einem for-loop dargestellt werden. (Bei der For-Schlaufe muss die Grösse des Modells bedacht werden. Ein Intercept-Term Mixed Model rechnet sehr schnell, mit Slope-Termen kann die Berechnung sehr lange dauern.)

```
# ------------------------
n = 5000

for (i von 1 bis n) {

(1a) Daten_i = Stichprobe aus den Daten ziehen mit Zurücklegen
(1b) bcoef_i= Modell mit den Daten_i schätzen und die Koeffizienten
in einer Matrix speichern

}

(2) Mittelwerte, alpha/2-quantil und 1-alpha/2-quantil
aus bcoef_i für jeden Parameter berechnen.

(3) Aus (2) kann das Vertrauensintervall berechnet werden.
# ------------------------
```

Nun gibt es drei Möglichkeiten:

- Die Mittelwerte und die Standardabweichungen unterscheiden sich nicht stark von den Schätzungen des Koeffizienten des Modells. → Vertrauensintervalle, Tests aus den Schätzungen im Modell.
- Die Mittelwerte des Bootstrapping weichen nicht stark von den geschätzten Koeffizienten des Modells ab, die Standardabweichungen jedoch schon. → Mittelwerte aus dem Modell, die Quantile aus der Bootstrap-Methode. Dieser Punkt gilt bedingt im Fall der Lindt & Sprüngli Daten. Bei den *between-subject* Faktoren (Alter, Geschlecht) (die oft gar nicht in das Modell einbezogen werden) weichen die Boostrap-Vertrauensintervalle von den geschätzten Vertrauensintervallen ab. Bei den *within-subject* Faktoren (Produkte) stimmen die Vertrauensintervalle überein.
- Die Mittelwerte und Standardabweichungen des Bootstrapping weichen von den geschätzten Koeffizienten des Modells ab. → Mittelwerte für Vertrauensintervalle, Tests aus der Bootstrap-Methode

6.3. Anwendung Lindt & Sprüngli

Gleichung (6.6) zeigt das Mixed Model für die Lindt & Sprüngli Daten mit einem Slope-Term.

$$T_{hijk} = \mu + A_k + G_h + P_j + (AG)_{kh} + (PA)_{jk} + (PG)_{jh} + S_{i(hk)} + S^k_{i(hk)} + \epsilon_{hijk} \quad (6.6)$$

T_{hijk} ist die Bewertung des 'Taste' bzw. 'Texture' des Produkt j von der Person i mit der Altersgruppe k und dem Geschlecht h
μ ist das Mittel über die Beobachtungen mit Alter $k = 1$, Geschlecht $h = 1$ und Produkt $j = 1$
G_h ist der Effekt des h-ten Geschlecht, $h = 1, 2$
A_k ist der Effekt der k-ten Altersgruppe, $k = 1, 2, 3$
P_j ist der Effekt des j-ten Produkts, $j = 1, ..., 8$
$(PA)_{jk}$ ist die Wechselwirkung zwischen dem j-ten Produkt und der k-ten Altersgruppe
$(PG)_{jh}$ ist die Wechselwirkung zwischen dem j-ten Produkt und dem h-ten Geschlecht
$(AG)_{kh}$ ist die Wechselwirkung zwischen der k-ten Altersgruppe und dem h-ten Geschlecht
$S_{i(kh)}$ ist der Effekt der Person i, die Personen sind in den Altersgruppen k und dem Geschlecht h geschachtelt, $i = 1, \ldots, N_{kh}$, $\sum_k \sum_h N_{kh} = N$
$S^k_{i(kh)}$ ist der Effekt der Person i in der k-ten Altersgruppe, die Personen sind in den Altersgruppen k und dem Geschlecht h geschachtelt
ϵ_{hijk} ist der Fehler für Person i mit dem Geschlecht h in der Altersgruppe k beim Produkt j

Der Term $S_{i(kh)}$ wird in $S^k_{i(kh)}$ integriert, weil $\sigma^2_{S_1} = \sigma^2_S$.

$S^k_{i(kh)} \backsim \mathbf{N}(0, \Sigma_S)$
$\epsilon_{hijk} \backsim N(0, \sigma^2_\epsilon)$

σ^2_ϵ ist die Variabilität von Bewertung zu Bewertung innerhalb der gleichen Personen. $\sigma^2_{S_k}$ ist die Variabilität der Personen in der Altersgruppe k. Σ_S ist die Kovarianzmatrix der multivariaten Normalverteilung der zufälligen Terme $S_{i(kh)}$ und $S^k_{i(kh)}$ mit Erwartungswert 0.

Die Nebenbedingungen der Koeffizienten lauten $G_1 = A_1 = P_1 = (AG)_{11} = (AG)_{21} = (AG)_{31} = (AG)_{12} = (PA)_{11} = (PA)_{12} = (PA)_{13} = (PA)_{21} = (PA)_{31} = (PA)_{41} = (PA)_{51} = (PA)_{61} = (PA)_{71} = (PA)_{81} = (PG)_{11} = (PG)_{12} = (PG)_{21} = (PG)_{31} = (PG)_{41} = (PG)_{51} = (PG)_{61} = (PG)_{71} = (PG)_{81} = 0$.

6.3.1. Ergebnisse

Die Vertrauensintervalle der Ergebnisse des Datensatz A und B werden jeweils mit der Boostrap-Methode gebildet. Die ANOVA-Tabelle entspricht dem gewöhnlichen Modeloutput und ist nicht durch Boostrapping korrigiert worden. Die restlichen Resultate sind im Anhang A.3.

taste ∽ product + age + gender + (1 + age | subject), data=datB

```
mod <- lmer(taste ~ (product + age + gender)^2 + (age | subject), data = datB)

# Linear mixed model fit by REML
# Formula: taste ~ (product + age + gender)^2 + (age | subject)
#    Data: dat
#   AIC  BIC logLik deviance REMLdev
#  5072 5280  -2495     4974    4990
# Random effects:
#  Groups   Name        Variance Std.Dev. Corr
#  subject  (Intercept) 1.11490  1.05589
#               age35-54 0.98064  0.99027  -0.271
#               age55-   0.34574  0.58799  -0.611  0.165
#  Residual              3.46567  1.86163
# Number of obs: 1174, groups: NJ, 196
#
# Fixed effects:
# ...

# Analysis of Variance Table
#                  Df Sum Sq Mean Sq F value
# product           7 466.39   66.63 19.2249
# age               2  53.00   26.50  7.6461
# gender            1  12.60   12.60  3.6366
# product:age      14  44.26    3.16  0.9122
# product:gender    7  42.67    6.10  1.7590
# age:gender        2   9.22    4.61  1.3301

# zusätzlich mit der Bootstrap-Methode Vertrauensintervalle
der Parameterschätzungen bilden
```

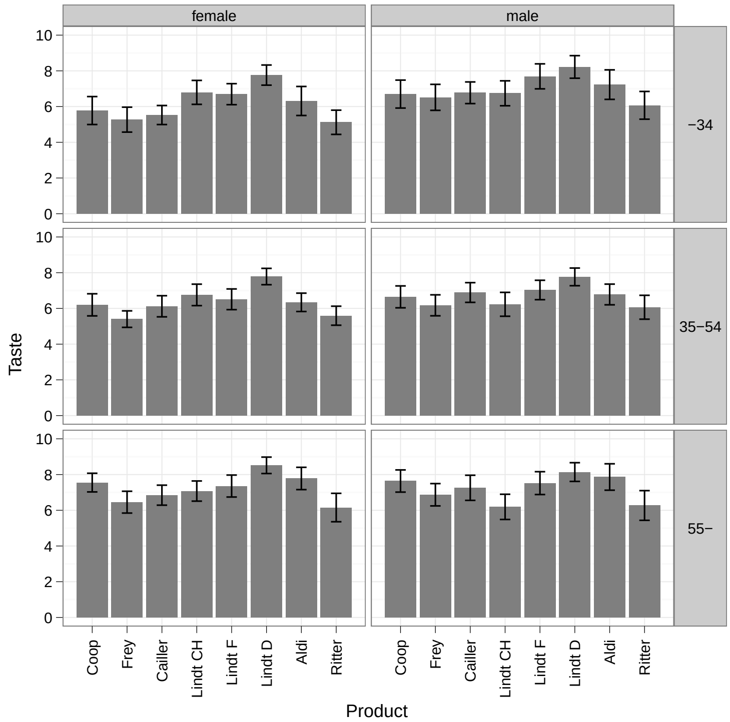


Abbildung 6.1.: Datensatz B: Schätzungen der Bewertungen des Geschmacks durch ein Mixed Model für das Alter, Geschlecht und Produkt.

7. Umsetzung bei Lindt & Sprüngli

Die Analysen der Sensorikabteilung von Lindt & Sprüngli dienen unter anderem als Entscheidungshilfe für das Management. Produkte die eine Wertung über 8 Punkte erreichen, werden als potentielles Verkaufsprodukt gewertet. Für eine solche Entscheidung ist neben der Punktschätzung auch das Vertrauensintervall von Bedeutung.

Von gewissen Effekten sind die Muster bereits bekannt. Jüngere Menschen, bzw. Frauen bewerten die Produkte tendenziell tiefer als ältere Menschen, bzw. Männer. Wenn das Alter oder das Geschlecht mit in das Modell einbezogen und in einer Grafik dargestellt werden, werden Muster dargestellt, die bereits bekannt und daher uninteressant sind (für das Managment). Daher wird für eine Darstellung oft nur das Produkt als erklärende Variable verwendet.

7.1. Praktisches Vorgehen

1. Homogenitätsannahme überprüfen
2. Vollständiges Modell (Haupteffekte/Wechselwirkungen) berechnen
3. Falls bei den Alters- und Geschlechtereffekten die gewöhnlichen Muster auftreten, können diese Faktoren aus dem Modell genommen werden.
4. Reduziertes Modell berechnen (Haupteffekte/Wechselwirkungen)
5. Modellannahmen anhand der Residuenanalyse überprüfen
6. Grafik mit Schätzungen des reduzierten Modells erstellen.

Neben den Faktoren Alter und Geschlecht kann auch die Kaufabsicht der Personen interessieren. Ein Beispiel hierzu wird in Kapitel 8 gezeigt.

Falls die Homogenitätsannahme abgelehnt werden muss, weil zwischen einzelnen Produkten eine hohe Kovarianz besteht, können die Produktegruppen einzeln ausgewertet werden.

7.2. Beispiel: Modell für Taste aus Datensatz B

Als Beispiel einer Analyse wird der Datensatz B verwendet. Es soll eine Grafik mit den Schätzungen der Zielvariable 'Taste' für die Sub-Gruppen Alter, Geschlecht und Sprachgebiet erstellt werden. Formel (7.1) zeigt das Mixed Model mit einem Intercept-Fehlerterm dazu.

$$\begin{aligned} T_{hijkl} = \mu + P_j + A_k + G_h + L_l + (PA)_{jk} + (PG)_{jh} + (PL)_{jl} + (AG)_{kh} \\ +(AL)_{kl} + (GL)_{hl} + S_{i(khl)} + \epsilon_{hijkl} \end{aligned} \tag{7.1}$$

T_{hijkl} ist die Bewertung des 'Taste' des Produkt j von der Person i mit der Altersgruppe k, dem Geschlecht h, dem Sprachgebiet l
μ ist das Mittel über die Beobachtungen mit Alter $k = 1$, Geschlecht $h = 1$, Produkt $j = 1$ und Sprachgebiet $l = 1$
G_h ist der Effekt des h-ten Geschlecht, $h = 1, 2$
A_k ist der Effekt der k-ten Altersgruppe, $k = 1, 2, 3$
P_j ist der Effekt des j-ten Produkts, $j = 1, ..., 8$
L_l ist der Effekt des l-ten Sprachgebiets, $l = 1, 2$
$(PA)_{jk}$ ist die Wechselwirkung zwischen dem j-ten Produkt und der k-ten Altersgruppe
$(PG)_{jh}$ ist die Wechselwirkung zwischen dem j-ten Produkt und dem h-ten Geschlecht
$(PL)_{jl}$ ist die Wechselwirkung zwischen dem j-ten Produkt und dem k-ten Sprachgebiet
$(AG)_{kh}$ ist die Wechselwirkung zwischen der k-ten Altersgruppe und dem h-ten Geschlecht
$(AL)_{kl}$ ist die Wechselwirkung zwischen der k-ten Altersgruppe und dem l-ten Sprachgebiet
$(GL)_{hl}$ ist die Wechselwirkung zwischen dem h-ten Geschlecht und dem l-ten Sprachgebiet
$S_{i(khl)}$ ist der Effekt der Person i, die Personen sind in den Altersgruppen k, dem Sprachgebiet l und dem Geschlecht h geschachtelt, $i = 1, \ldots, N_{khl}$, $\sum_k \sum_h \sum_l N_{khl} = N$
ϵ_{hijkl} ist der Fehler für Person i mit dem Geschlecht h, dem Sprachgebiet l in der Altersgruppe k beim Produkt j

$S_{i(hkl)} \backsim i.i.d.N(0, \sigma_S^2)$
$\epsilon_{hijkl} \backsim i.i.d.N(0, \sigma_\epsilon^2)$

σ_ϵ^2 ist die Variabilität von Bewertung zu Bewertung innerhalb der gleichen Personen. σ_S^2 ist die Variabilität zwischen den Personen.

Die Nebenbedingungen der Koeffizienten lauten $G_1 = A_1 = P_1 = (AG)_{11} = (AG)_{21} = (AG)_{31} = (AG)_{12} = (PA)_{11} = (PA)_{12} = (PA)_{13} = (PA)_{21} = (PA)_{31} = (PA)_{41} = (PA)_{51} = (PA)_{61} = (PA)_{71} = (PA)_{81} = (PG)_{11} = (PG)_{12} = (PG)_{21} = (PG)_{31} = (PG)_{41} = (PG)_{51} = (PG)_{61} = (PG)_{71} = (PG)_{81} = (PL)_{11} = (PL)_{12} = (PL)_{21} = (PL)_{31} = (PL)_{41} = (PL)_{51} = (PL)_{61} = (PL)_{71} = (PL)_{81} = 0$

a. Grafische Darstellung

Zuerst werden die Modellannahmen anhand der Originaldaten überprüft. In Abbildung 7.1 ist die Kovarianzmatrix der Produkte und die Boxplots pro Produkt abgebildet. Die Homogenitätsannahme verlangt eine konstante Varianz und dieselbe Kovarianz zwischen allen Produkten. Die Kovarianz scheint zwar nicht konstant zu sein für alle Kombinationen, aber eine grobe Verletzung scheint nicht vorzuliegen - dasselbe bei der Varianz.

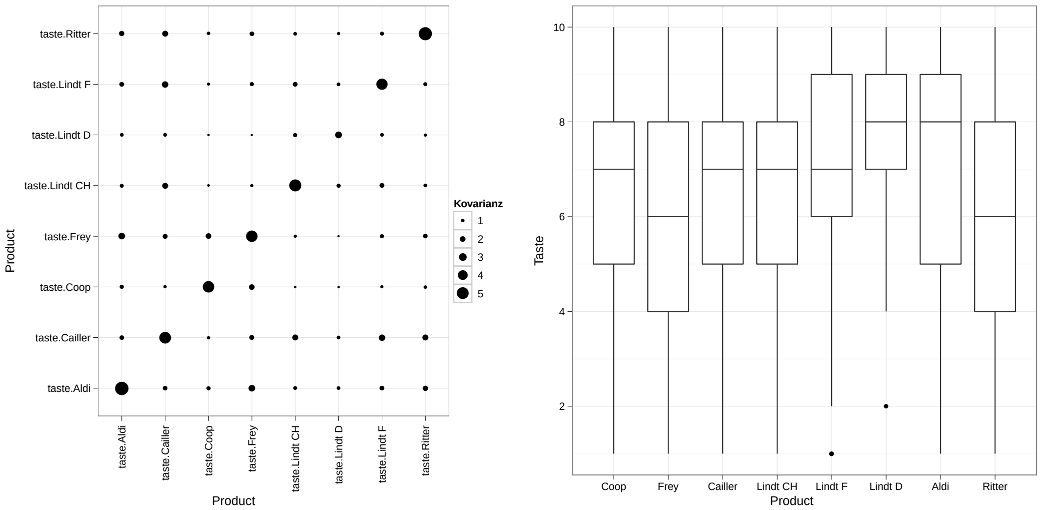


Abbildung 7.1.: Kovarianzmatrix und die Boxplots der 8 Produkte für die Überprüfung der Homogenitätsannahme

Abbildung 7.2 zeigt mit Mosaikplots die Verhältnisse der Gruppengrössen auf. Auf den ersten Blick ist ersichtlich, dass das Design nicht balanciert ist (sonst wären alle Vierecke gleich gross). Von den Testpersonen sind etwas weniger wie 2/3 aus einem deutschsprachigen Gebiet der Schweiz. Die Altersgruppen sind in den Sprachgebiet-Gruppen ungleich verteilt. Es nehmen mehr Frauen wie Männer an den Tests teil, die Altersverteilung ist beim Geschlecht ähnlich. Die dritte Altersgruppe mit Personen über 55 Jahren ist am kleinsten.

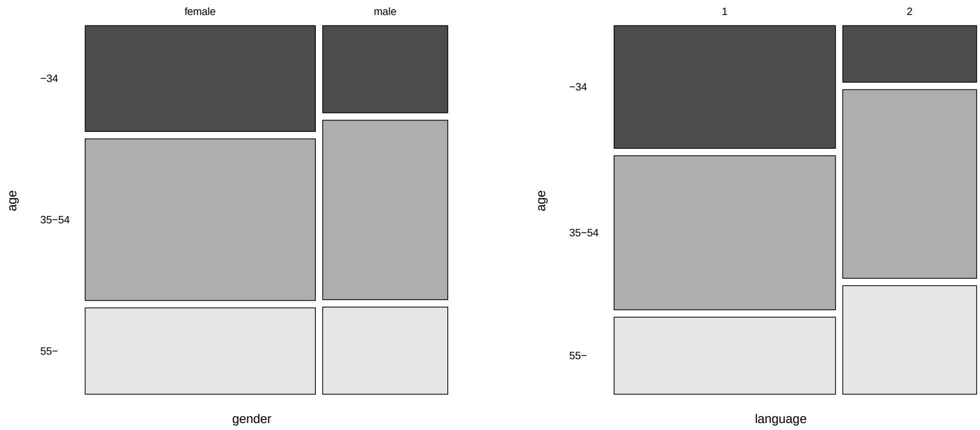


Abbildung 7.2.: Mosaikplots zur Darstellung der Sub-Gruppen-Grösse für Alter, Geschlecht und Sprachgebiet (1=deutschsprachig, 2=französischsprachig)

b. Modellberechnung

Das Modell wird zunächst vollständig berechnet. Nachdem ersichtlich wird, das das Sprachgebiet keinen statistisch signifikanten Einfluss hat, wird dieser Term aus dem Modell ausgeschlossen.

$\hat{\sigma}_S^2$ beträgt 1.21, $\hat{\sigma}_\epsilon^2$ ist gleich 3.47. Die *Intraclass Correlation* beträgt 0.25. Es gibt nur eine mässige Variabilität zwischen den Personen.

Abbildung 7.3 zeigt die Schätzungen für die Level-Kombinationen der Faktoren Alter-, Geschlecht- und Produkt. Das Produkt `Lindt D` wird bei allen Sub-Gruppen am höchsten bewertet, am schlechtesten das Produkt `Ritter`. Die Bewertung 8 erreicht nur das Produkt `Lindt D`, wobei das Vertrauensintervall auch Bewertungen unter 8 umschliesst. In dieser Grafik sieht man, dass Männer ein Produkt nicht zwingend höher bewerten wie Frauen und dass ältere Menschen nicht zwingend höher bewerten als jüngere. Weil Wechselwirkungen zwischen dem Geschlecht und der Altersgruppe, dem Geschlecht und dem Produkt und zwischen der Altersgruppe und dem Produkt eingeführt wurden, heben sich diese Tendenzen etwas auf.

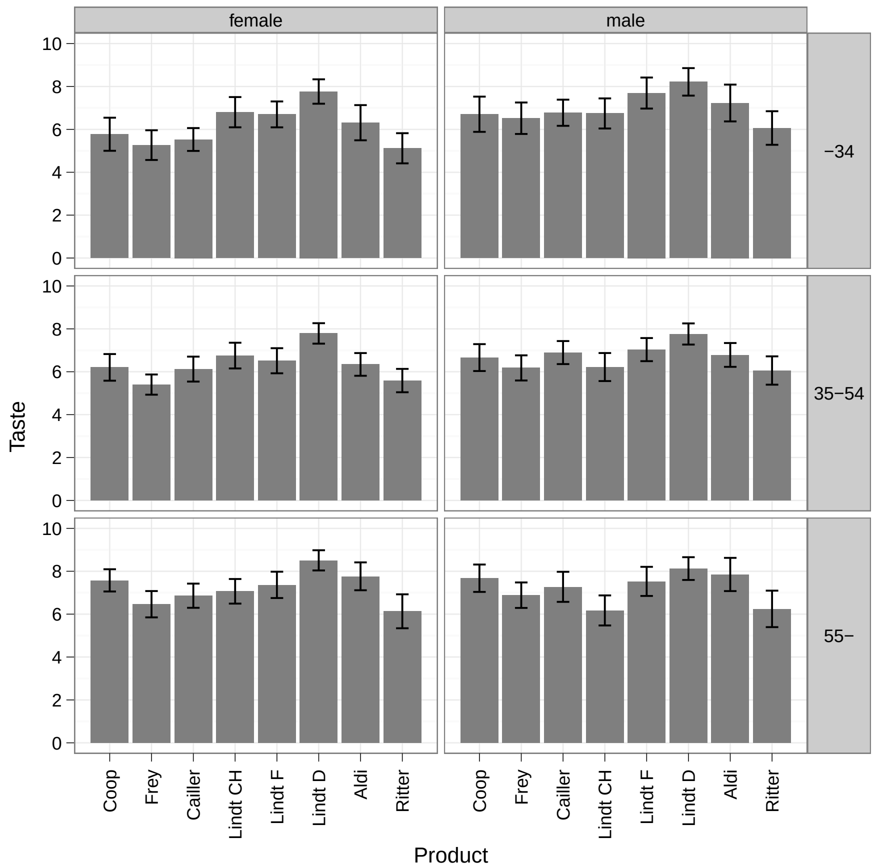


Abbildung 7.3.: Datensatz B: Schätzungen der Bewertungen des Geschmacks durch ein Mixed Model für das Alter, Geschlecht und Produkt.

```
# Modell-Output aus R
# ---
# Linear mixed model fit by REML
# Formula: taste ~ (product + age_new + gen)^2 + (1 | NJ)
#    Data: dat
#   AIC  BIC logLik deviance REMLdev
#  5066 5248  -2497     4978    4994
# Random effects:
#  Groups   Name        Variance Std.Dev.
#  NJ       (Intercept) 1.2137   1.1017
#  Residual             3.4656   1.8616
# Number of obs: 1174, groups: NJ, 196
#
# Fixed effects:
```

```
#                             Estimate Std. Error t value
# (Intercept)                  5.77211    0.36730  15.715
# productFrey                 -0.50662    0.43108  -1.175
# productCailler              -0.24464    0.42346  -0.578
# productLindt CH              1.02883    0.44935   2.290
# productLindt F               0.92609    0.44429   2.084
# productLindt D               1.99259    0.45430   4.386
# productAldi                  0.53887    0.45940   1.173
# productRitter               -0.65358    0.44551  -1.467
# age_new35-54                 0.43143    0.44878   0.961
# age_new55-                   1.80295    0.51164   3.524
# genmale                      0.93259    0.49704   1.876
# productFrey:age_new35-54    -0.29562    0.51552  -0.573
# productCailler:age_new35-54  0.16456    0.51957   0.317
# productLindt CH:age_new35-54 -0.47878    0.53442  -0.896
# productLindt F:age_new35-54  -0.61600    0.53079  -1.161
# productLindt D:age_new35-54  -0.40879    0.53671  -0.762
# productAldi:age_new35-54    -0.40079    0.53916  -0.743
# productRitter:age_new35-54   0.03894    0.52850   0.074
# productFrey:age_new55-      -0.60542    0.59690  -1.014
# productCailler:age_new55-   -0.47129    0.60088  -0.784
# productLindt CH:age_new55-  -1.53961    0.61301  -2.512
# productLindt F:age_new55-   -1.13650    0.59880  -1.898
# productLindt D:age_new55-   -1.05941    0.60777  -1.743
# productAldi:age_new55-      -0.35003    0.63048  -0.555
# productRitter:age_new55-    -0.78827    0.60818  -1.296
# productFrey:genmale          0.32095    0.45955   0.698
# productCailler:genmale       0.31412    0.47278   0.664
# productLindt CH:genmale     -0.99125    0.47332  -2.094
# productLindt F:genmale       0.06264    0.46557   0.135
# productLindt D:genmale      -0.48301    0.46900  -1.030
# productAldi:genmale         -0.01434    0.47458  -0.030
# productRitter:genmale        0.01210    0.47109   0.026
# age_new35-54:genmale        -0.47580    0.48527  -0.980
# age_new55-:genmale          -0.83297    0.56253  -1.481

# Analysis of Variance Table
#                 Df Sum Sq Mean Sq F value
# product          7 464.71   66.39 19.1560
# age_new          2  42.26   21.13  6.0975
# gen              1  13.90   13.90  4.0103
# product:age_new 14  45.50    3.25  0.9379
# product:gen      7  43.39    6.20  1.7885
# age_new:gen      2   7.71    3.85  1.1123
```

c. Residuenanalyse

Mit den Residuen wird die Normalverteilungsannahme überprüft. Als Residuen können bei Mixed Models zwei Typen auftreten. Residuum I ist das allgemein bekannte Residuum, nämlich Beobachtung minus die geschätzten Parameter aus dem Modell. Residuum II ist die Beobachtung minus die festen Effekte. Von den Residuen I ist bekannt, dass sie normalverteilt sein sollten. Die Residuen II sollten wegen der Additivität der Normalverteilung ebenfalls normalverteilt sein. In dieser Residuenanalyse werden die Residuen I verwendet.

Abbildung 7.4 zeigt zwei Grafiken. Die erste hat auf der X-Achse die Beobachtungen und auf der Y-Achse die Residuen I. Die Grafik zeigt eine Verletzung der Normalverteilungsannahme im Hinblick auf den konstanten Erwartungswert 0. Tiefe Bewertungen werden überschätzt, hohe Bewertungen unterschätzt. Dies ist ein bekanntes Muster bei limitierten Antwortemöglichkeiten und kann auch nicht mit einer Transformation der Zielvariablen umgangen werden. Die Varianz ist konstant. Die Grafik rechts zeigt auf der X-Achse die Produktegruppe und auf der Y-Achse die Residuen I. Die Annahme Erwartungswert 0 und konstante Varianz ist nicht verletzt.

In der Abbildung 7.5 sind die theoretischen und die echten Quantile der Residuen I aufgetragen. Auch hier ist keine Verletzung der Normalverteilung erkennbar.

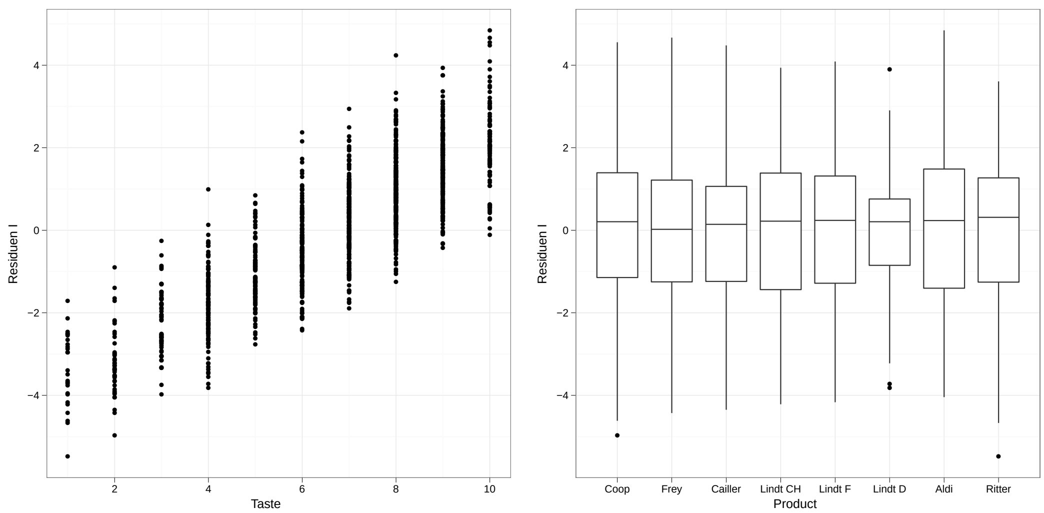


Abbildung 7.4.: Überprüfung der konstanten Varianz und dem Erwartungswert 0.

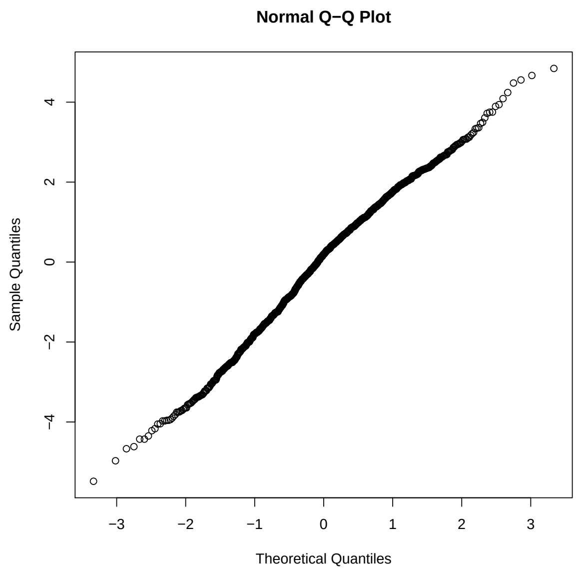


Abbildung 7.5.: Normal Q-Q-Plot.

d. Schlussfolgerung

Falls die Bewertung 8 als Benchmark gilt, hat keine der Schokoladen das Ziel erreicht.

7.3. Umsetzung in SPSS

SPSS ist eine Statistik-Software und Lizenzpflichtig. Das Programm kann via Menüleiste oder auch via Syntax bedient werden. SPSS Menüs können durch eigene Menüpunkte ergänzt werden. Diese werden mit Hilfe von einem Dialogfeld konstruiert. Ein Dialogfeld stellt ein eigenes Objekt dar und kann mit anderen Benutzern ausgetauscht und an einem beliebigen Menüort installiert werden.

7.3.1. Dialogfelder

SPSS berechnet sowohl Repeated Measures ANOVA als auch Mixed Models. Für eine standardisierte Benutzung wurden zwei Dialogfelder generiert.

1. Analyse Konsumententest: Die Zielvariable, die festen Faktoren und der zufällige Faktor werden vorgegeben. Das Dialogfeld berechnet ein Repeated Measures ANOVA (ohne Wechselwirkungen). Output: Parameterschätzungen und ANOVA-Tabelle

2. Grafik Konsumententest: Generiert anhand der Parameterschätzungen des Modell aus Dialogfeld 1 eine Grafik. Kann maximal 4 erklärende Variablen umfassen, mehr kann und soll in einer Grafik nicht dargestellt werden. Output: Grafik.

Für die Installation und Anwendung von Dialogfeldern in SPSS gibt es eine Anleitung. Die Dialogfelder und die Anleitungen liegen auf der beigelegten CD bereit.

7.4. Umsetzung in R

`R` ist eine Open-Source Software (kostenlos) und kann unter http://www.r-project.org/ heruntergeladen genutzt werden. `R` ist eine Implementation von S, die Resultate werden über eine Syntax generiert. Von SPSS weicht `R` unter anderem ab, weil mit `R` objektorientiert programmiert wird. In der Regel wird mit einem einem Editor gearbeitet, z.B. Tinn-R oder NotepadPlus, welche ebenfalls kostenlos sind.

Für das Thema Mixed Models und die Funktion `lmer()` existiert eine eigene R-Help, für die man sich unter http://stat.ethz.ch/CRAN/ > Mailing-List > R-sig-mixed-models anmelden kann und die einen reichen Fundus an bereits gestellten Fragen zu Mixed Models und `lmer()` bietet.

Eine gute Einführung in `R` bietet beispielsweise *An Introduction to R: Software for Statistical Modelling & Computing (Course Materials and Exercisis)* von Kuhnert und Venables (2005) (http://www.csiro.au/resources/Rcoursenotes.html, Zugriff: 2009-11-24) oder *Programmieren mit R* von Ligges (2009) das beim Springer-Verlag in Heidelberg erschienen ist.

7.4.1. Syntaxmöglichkeiten

Die Modelle aus Kapitel 4, 5 und 6 können mit `R` auf verschiedene Arten berechnet werden. In den Beispielen folgt jeweils die Syntax-Schreibweise und der Output dazu.

Zwei Punkte zur Implementation in R:

- In `R` sind die Kontraste anders gesetzt. Statt $\sum_i A_i = 0$ wird jeweils der erste Parameter eines Faktors Null gesetzt: $A_1 = 0$. Bei SPSS wird das letzte Level gleich 0 gesetzt. Mit diesen Änderungen ändert sich die Interpretation des 'Gesamtmittels' μ. Die Default-Kontraste können bei beiden Programmen angepasst werden.
- Um den die Sum of Squares nach Typ II oder III zu generieren, wird die Funktion `Anova()` aus der Library `car` Funktion benötigt.
- Fehlende Werte in einem Repeated Measures Design verändern die Balancierung der Daten. Die Funktion `aov()` verlangt aber nach einer Balancierung im Sinn der Produkte. Deshalb werden im Datensatz diejenigen Personen aus der Analyse ausgeschlossen, die nur einen Teil der zu testenden Produkte bewertet haben. Bei

der Funktion `lmer()` werden lediglich die Beobachtungen mit den fehlenden Werten aus dem Modell ausgeschlossen.

- Für unkorrelierte Fehlerterme wird mehrmals `( | )` verwendet. Zum Beispiel `(1 | subject) + (age -1 | subject)`

Für eine Beispiels-Auswertung in R und die Berechnung der Vertrauensintervalle wurde ein Syntax-File erstellt. Beides ist auf der beigelegten CD.

ANOVA ohne Personeneffekt → `aov()`

Die Faktoren `age, gender, product` sind feste Effekte.
Dieses Modell beschreibt die Situation nur ungenügend. Grundsätzlich sind die P-Werte der Faktoren `gender`, `age` und `gender:age` zu klein und suggerieren damit unter Umständen eine statistische Signifikanz, die gar keine ist. Die P-Werte der Faktoren `product`, `age:product` und `gender:product` sind hingegen zu gross. Vorhandene Effekte werden möglicherweise gar nicht entdeckt, weil nicht signifikant. Die Unbalanciertheit des Modells kann diesen Grundsatz etwas stören.

```
mod <- aov(taste ~ (product + age + gender)^2, data = dat)
mod <- aov(taste ~ product + age + gender + age:gender +
           + age:product + gender:product, data = dat) # ist dasselbe
summary(mod)

# Analysis of Variance Table
# - - - - - - - - - - - - - - - - -
#                   Df Sum Sq Mean Sq F value    Pr(>F)
# product            3  567.3   189.1 43.9355 < 2.2e-16  ## zu grosser P-Wert
# age                2   51.9    26.0  6.0304  0.002447  ## zu kleiner P-Wert
# gender             1    6.8     6.8  1.5851  0.208166  ## zu kleiner P-Wert
# product:age        6   16.5     2.7  0.6377  0.700166  ## zu grosser P-Wert
# product:gender     3    3.0     1.0  0.2313  0.874623  ## zu grosser P-Wert
# age:gender         2    3.8     1.9  0.4420  0.642780  ## zu kleiner P-Wert
# Residuals       2058 8857.1     4.3
```

Man beachte, dass diese ANOVA-Tabelle dem Typ I entspricht, siehe Unterkapitel *Balancierung* in Kapitel 4. Für Typ II oder III kann die Funktion `Anova` aus dem Package `car` verwendet werden.

```
library(car)
help(Anova)   # Informationen über die Bedeutung der verschiedenen Varianten
Anova(mod)
```

```
# Analysis of Variance Table (Type II tests)
# - - - - - - - - - - - - - - - - - - - - - -
# Response: Taste
#                  Sum Sq   Df F value    Pr(>F)
# product           567.3    3 43.9355 < 2.2e-16   ## zu grosser P-Wert
# age                49.6    2  5.7571  0.003211   ## zu kleiner P-Wert
# gender              6.8    1  1.5851  0.208166   ## zu kleiner P-Wert
# product:age        16.6    6  0.6433  0.695640   ## zu grosser P-Wert
# product:gender      3.0    3  0.2313  0.874623   ## zu grosser P-Wert
# age:gender          3.8    2  0.4420  0.642780   ## zu kleiner P-Wert
# Residuals        8857.1 2058
```

Repeated Measures ANOVA → `aov()`

Das korrekte Modell lautet: `age`, `gender`, `product` sind feste Effekte, `subject` ist ein zufälliger Effekt. `subject` ist in `age`, `gender` geschachtelt. `product` ist mit `subject` gekreuzt.

```
mod.aov1 <- aov(taste ~ (product + age + gender)^2 + Error(subject), data = dat)
summary(mod.aov1)
```

```
# Analysis of Variance Table
# - - - - - - - - - - - - - - -
# Error: subject
#              Df Sum Sq Mean Sq F value  Pr(>F)
# age           2   51.9    26.0  3.1313 0.04449
# gender        1    6.8     6.8  0.8231 0.36470
# age:gender    2    3.8     1.9  0.2295 0.79498
# Residuals   513 4251.9     8.3
# ---
# Error: Within
#                   Df Sum Sq Mean Sq F value Pr(>F)
# product            3  567.3   189.1 63.4361 <2e-16
# product:age        6   16.5     2.7  0.9207 0.4788
# product:gender     3    3.0     1.0  0.3340 0.8008
# Residuals       1545 4605.3     3.0
```

Wird die Formel wie in `mod.aov2` geschrieben, entspricht der Modelloutput einem vollständig randomisierten ANOVA Design.

```
mod.aov2 <- aov(taste ~ (age + gender + product)^2 + subject, data = dat)
summary(mod.aov2)
```

```
# Analysis of Variance Table
# - - - - - - - - - - - - - - -
#                   Df  Sum Sq Mean Sq F value    Pr(>F)
# product            3   567.3   189.1 63.4361 < 2.2e-16
# age                2    51.9    26.0  8.7069 0.0001737 ## viel zu kleiner P-Wert
# gender             1     6.8     6.8  2.2887 0.1305246 ## viel zu kleiner P-Wert
# subject          515  4255.7     8.3  2.7723 < 2.2e-16
# product:age        6    16.5     2.7  0.9207 0.4788419
# product:gender     3     3.0     1.0  0.3340 0.8007916
# age:gender         2     3.8     1.9  0.2295 0.79498   ## zu kleiner P-Wert
# Residuals       1545  4605.3     3.0
```

Um die Repeated Measures ANOVA-Tabelle aufzustellen, müssen die korrekten F-Werte berechnet werden.

```
# Analysis of Variance Table
# - - - - - - - - - - - - - - -
#                    Df  Sum Sq Mean Sq   F value    Pr(>F)
# age                 2   51.9     26.0    26/8.3   0.04444  ## 1-pf(26/8.3,2,515)
# gender              1    6.8      6.8   6.8/8.3   0.36581  ## 1-pf(6.8/8.3,1,515)
# age:gender          2    3.8      1.9   1.9/8.3   0.79548  ## 1-pf(1.9/8.3,2,515)
# subject           515 4255.7      8.3    2.7723 < 2.2e-16
# ---
# product             3  567.3    189.1   63.4361 < 2.2e-16
# product:age         6   16.5      2.7    0.9207 0.4788419
# product:gender      3    3.0      1.0    0.3340 0.8007916
# Residuals        1545 4605.3      3.0
```

Der Output wird in R nicht so ausgegeben - die Interaktion zwischen gender und age kann nicht berechnet werden, weil das Modell überbestimmt ist.

Repeated Measures ANOVA → `lmer()`

`age, gender, product` sind feste Effekte, `subject` ist ein zufälliger Effekt. Der Output von `lmer()` hat Informationen zu den Parameterschätzungen, Korrelation zwischen den Parametern, $\hat{\sigma}_S^S$, $\hat{\sigma}_\epsilon^2$. Es werden keine P-Werte angegeben. Für den Grund dafür siehe Unterkapitel 6.2. Die ANOVA-Tabelle entspricht einem Typ I Test, also sequentiell.

```
library(lme4)  # Package für die Funktion lmer()
mod.lmer <- lmer(taste ~ (product + age + gender)^2+ (1 | subject), data = dat)

anova(mod.lmer)

# Analysis of Variance Table
# - - - - - - - - - - - - - - -
# Analysis of Variance Table
#                 Df Sum Sq Mean Sq F value
# product          3 567.26  189.09 63.4360
# age              2  18.67    9.33  3.1310
# gender           1   2.45    2.45  0.8230
# product:age      6  16.47    2.74  0.9207
# product:gender   3   2.99    1.00  0.3340
# age:gender       2   1.37    0.68  0.2295

summary(mod.lmer)

# Linear mixed model fit by REML
# - - - - - - - - - - - - - - - - -
# Formula: taste ~ (product + age + gender)^2 + (1 | subject)
#    Data: datA
#   AIC  BIC logLik deviance REMLdev
#  8741 8854  -4351     8671    8701
# Random effects:
#  Groups   Name        Variance Std.Dev.
#  subject  (Intercept) 1.3269   1.1519    ## sigma^2 (Subject)
#  Residual             2.9808   1.7265    ## sigma^2 (Epsilon)
# Number of obs: 2076, groups: subject, 519

# Fixed effects:  ## Parameterschätzungen
# ...
#
# Correlation of Fixed Effects:  ## Korrelation der Parameterschätzungen
# ...
```

Erweitertes Mixed Model → `lmer()`

```
library(lme4)  # Package für die Funktion lmer()
mod.lmer <- lmer(taste ~ (product + age + gender)^2
     + (age | subject), data = dat)

anova(mod.lmer)

# Analysis of Variance Table
# - - - - - - - - - - - - - - -
#                Df Sum Sq Mean Sq F value
# product         3 567.26  189.09 63.4360
# age             2  18.54    9.27  3.1107
# gender          1   2.27    2.27  0.7625
# product:age     6  16.47    2.74  0.9207
# product:gender  3   2.99    1.00  0.3340
# age:gender      2   1.32    0.66  0.2207

summary(mod.lmer)

# Linear mixed model fit by REML
# - - - - - - - - - - - - - - - - -
# Linear mixed model fit by REML
# Formula: Taste ~ (product + age + gender)^2 + (age | subject)
#    Data: dat
#   AIC  BIC logLik deviance REMLdev
#  8750 8891  -4350     8670    8700
# Random effects:
#  Groups   Name        Variance Std.Dev. Corr
#  subject  (Intercept) 1.49186  1.22142
#           age35-54    0.46828  0.68431  -0.413
#           age55-      1.00943  1.00470  -0.556  0.230
#  Residual             2.98077  1.72649
# Number of obs: 2076, groups: subject, 519

# Fixed effects:  ## Parameterschätzungen
# ...
#
# Correlation of Fixed Effects:  ## Korrelation der Parameterschätzungen
# ...
```

8. Diskussion und Ausblick

Das Modell kann durch Wechselwirkungen und eine weitere Variable ergänzt werden. Wird die Homogenitätsannahme verletzt, muss nach Modellalternativen gesucht werden.

8.1. Wechselwirkungen

In der Implementation des Repeated Measures ANOVA in SPSS wurden keine Wechselwirkungen berücksichtigt. Eine Ergänzung des Modells mit Wechselwirkungen wäre jedoch von Vorteil. Wenn gewisse Produkte nur bei gewissen Altersgruppen oder Geschlechtern ankommen, kann das für das Marketing wichtig sein.

8.2. Kaufabsicht

Die Personen werden manchmal vor dem Testen des Produktes und nach dem Testen des Produktes gefragt, ob sie das Produkt kaufen würden. Damit wird ein neuer Varaiblen-Typ eingeführt. Die Variable kann sich pro Person von Produkt zu Produkt ändern.

- `PIBT`: Purchaseintent before testing
  - sicher nicht kaufen
  - wahrscheinlich nicht kaufen
  - weiss nicht, ob ich kaufen würde
  - wahrscheinlich kaufen
  - sicher kaufen
- `PIAT`: Purchaseintent after testing
  - sicher nicht kaufen
  - wahrscheinlich nicht kaufen
  - weiss nicht, ob ich kaufen würde
  - wahrscheinlich kaufen
  - sicher kaufen

Mit diesen beiden fünfstufigen Faktoren möchte man herausfinden, ob sich das Kaufverhalten mit dem Testen des Produktes ändert. Die Levels können in zwei Kategorien eingeteilt werden: *kaufen* und *nicht kaufen*. Sicher nicht kaufen, wahrscheinlich nicht kaufen, weiss nicht, ob ich kaufen würde gehören zu `nicht kaufen`. Wahrscheinlich kaufen und sicher kaufen gehören zu `kaufen`. Damit sind zwei Faktoren mit je zwei Levels entstanden.

- `PIBT2`: Purchaseintent before testing, `nicht kaufen`, `kaufen`.
- `PIAT2`: Purchaseintent after testing, `nicht kaufen`, `kaufen`.

Die Abbildung 8.1 zeigt einen Mosaik-Plot für die beiden Faktoren. Vor dem Testen würden etwa 80% der Personen das Produkt kaufen, nach dem Testen davon nur noch etwas mehr als 50%. Von den 20% der Personen, die vor dem Testen das Produkt nicht kaufen würden, lässt sich etwa 1/6 umstimmen.

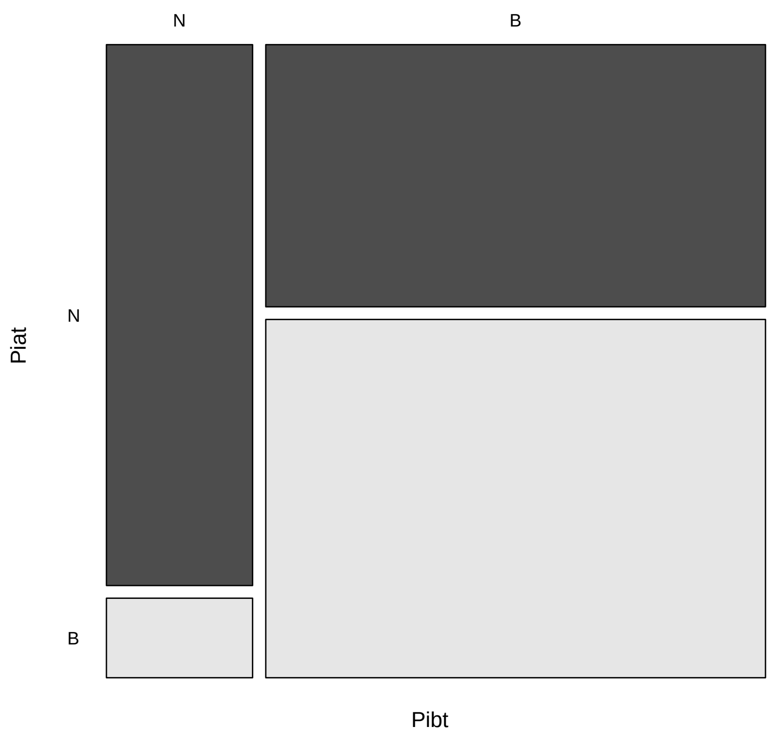


Abbildung 8.1.: Mosaik-Plot der Kaufabsicht vor und nach dem Testen.

Die Einführung dieses Faktors in das Modell ist insofern kritisch, als dass der Faktor sich mit dem *within-subject* Faktor ändern kann - die Kaufabsicht kann sich pro Produkt ändern. Bei diesem Modell waren die Produkte allerdings Pralinés und die Kaufabsicht wurde für alle Produkte auf einmal abgefragt, was den Faktor zu einem *between-subject* Faktor macht.

Der Modelloutput für den Datensatz A, Zielvariable Taste sieht folgendermassen aus:

```
# Linear mixed model fit by REML
# Formula: Taste ~ (product + Pibt2 + Piat2)^2 + (1 | subject)
#    Data: dat
#   AIC  BIC logLik deviance REMLdev
#  8483 8568  -4227     8428    8453
# Random effects:
#  Groups   Name        Variance Std.Dev.
#  subject  (Intercept) 0.67576  0.82204
#  Residual             2.94830  1.71706
# Number of obs: 2068, groups: subject, 517
#
# Fixed effects:
#                  Estimate Std. Error t value
# (Intercept)        4.9451     0.1819  27.190
# product2           1.7344     0.2264   7.660
# product3           1.0205     0.2264   4.507
# product4           2.1510     0.2264   9.499
# Pibt2B             1.1945     0.2251   5.307
# Piat2B             1.9616     0.3578   5.483
# product2:Pibt2B  -0.6807     0.2755  -2.471
# product3:Pibt2B  -0.3523     0.2755  -1.279
# product4:Pibt2B  -0.6918     0.2755  -2.511
# product2:Piat2B  -0.2619     0.2308  -1.135
# product3:Piat2B   0.1071     0.2308   0.464
# product4:Piat2B  -0.4444     0.2308  -1.925
# Pibt2B:Piat2B    -0.5194     0.3500  -1.484

# Analysis of Variance Table
#               Df Sum Sq Mean Sq  F value
# product        3 560.14  186.71  63.3296
# Pibt2          1 308.22  308.22 104.5403
# Piat2          1 423.21  423.21 143.5440
# product:Pibt2  3  48.03   16.01   5.4307
# product:Piat2  3  20.79    6.93   2.3504
# Pibt2:Piat2    1   6.49    6.49   2.2018
```

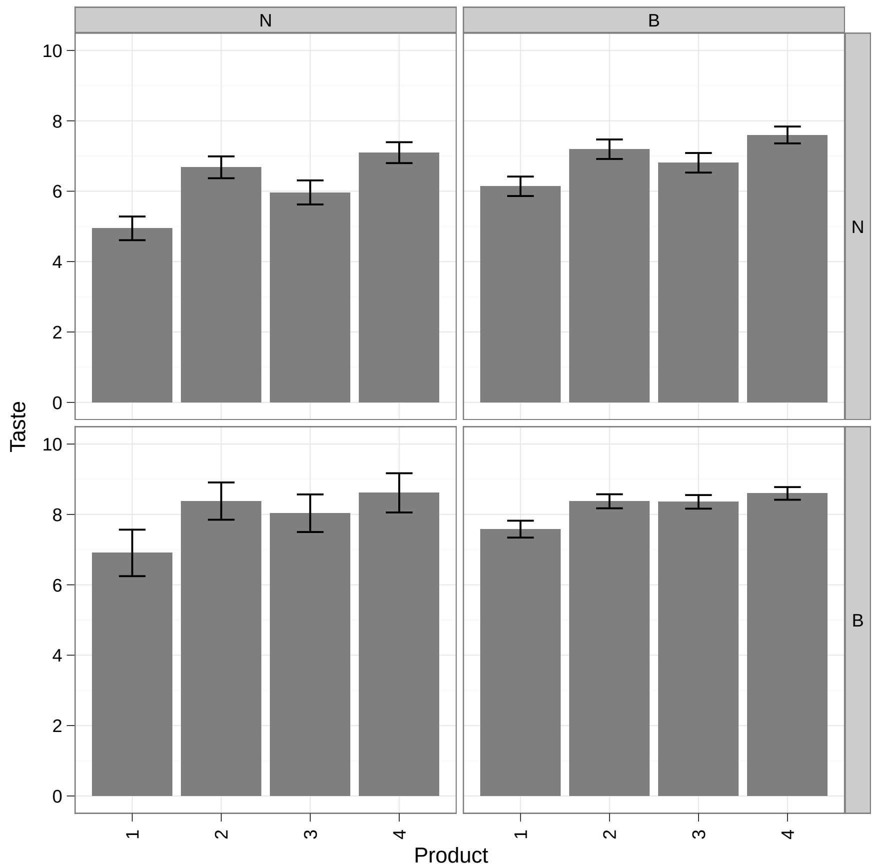


Abbildung 8.2.: Kaufabsicht beim Datensatz A. N bedeutet *Nicht kaufen* und B *Kaufen*. Die Grafik ist in 4 Felder unterteilt. Die Spalten repräsentieren den Faktor `Pibt2` und die Zeilen den Faktor `Piat2`

Die Variabilität der Bewertungen innerhalb der Personen ist gering, die Intraclasscorrelation beträgt bloss 0.18. Das heisst, die Faktoren `Pibt2` und `Piat2` erklären einen grossen Teil der Variabilität zwischen den Personen, sodass der Personeneffekt sehr klein wird. Als Slope-Term könnte noch der Faktor `Pibt2` oder `Piat2` hineingenommen werden um zu sehen, in welcher Faktorkombination am meisten Variabilität unter den Personen herrscht.

Wie im Mosaikplot 8.1 ersichtlich ist, sind die Gruppengrössen ungleich. Die Personengruppe die das Produkt vor dem Testen nicht kaufen würden, danach aber schon, ist am kleinsten. Das wirkt sich auch auf die Vertrauensintervalle dieser Faktorkombination aus, die in der Grafik 8.2 im Feld links unten sehr gross werden. Das Produkt 1 schneidet bei allen Gruppierungen am schlechtesten ab. Das Niveau der Personen, die das Produkt vor und nach dem Testen nicht kaufen würden, ist am tiefsten. Das Niveau der Personen, die die Produkte vor dem Testen nicht kaufen würden und nachher aber schon, ist zum Teil höher wie das Niveau der Personen, die das Produkt vor- und nachher kaufen würden. Das bedeutet: Personen aus der Gruppe die das Produkt vor dem Testen nicht kaufen würden sind schwer zu überzeugen. Ein Produkt muss sie schon sehr im Geschmack überraschen, damit sie sich umentscheiden. Um also die Personengruppe, die sich von *nicht kaufen* zu *kaufen* umentscheidet, zu künftigen Konsumenten zu machen, müssten sie das Produkt testen können.

Literaturverzeichnis

Varianzanalyse

[1] Müller, M. (2006). *Experimental Design*, Unveröffentlichtes Vorlesungsskript, Zürcher Hochschule für Angewandte Wissenschaften.

[2] Roth, H. R. (2009). *Angewandte Varianzanalyse und Versuchsplanung*, Unveröffentlichtes Vorlesungsskript, Eidgenössische Technische Hochschule Zürich.

Longitudinal Data Analysis

[3] Hedeker, D. und Gibbons, D. (2006). *Longitudial Data Analysis*, John Wiley & Sons, Inc. (Kapitel 1 bis 4)

[4] Montgomery, D. (2005). *Design and Analysis of Experiments*, John Wiley & Sons, Inc. (Kapitel 13 und 14 und 15.4)

Mixed Model

[5] Pinheiro, J. C. und Bates, D. (2000). *Mixed-effects models in S and S-Plus*, Springer-Verlag New York, Inc.

Bootstrap

[6] Harrell, F. Jr. (2001), *Regression Modeling Strategies*, Springer-Verlag New York. (Seite 87 bis 90, 195 bis 201)

A. Modelloutput

Die Theorie der Kapitel 4, 5 und 6 wurde auf die Datensätze von Lindt & Sprüngli übertragen. In A.1, A.2 und A.3 sind jeweils die ANOVA-Tabelle und die Grafik für den Geschmack und die Beschaffenheit des Datensatz A und B aufgeführt.
Im Modell des Datensatzes A gibt es keine Wechselwirkungen, im Modell des Datensatzes B schon.

A.1. ANOVA ohne Personeneffekt

A.1.1. taste ∽ product + age + gender, data=datA

```
aov(taste ~ product + age + gender , data = datA)

# Analysis of Variance Table
# --------------------------
# Response: Taste
#             Df Sum Sq Mean Sq F value    Pr(>F)
# product      3  567.3   189.1 44.0546 < 2.2e-16
# age          2   51.9    26.0  6.0467  0.002408
# gender       1    6.8     6.8  1.5894  0.207549
# Residuals 2069 8880.4     4.3
```

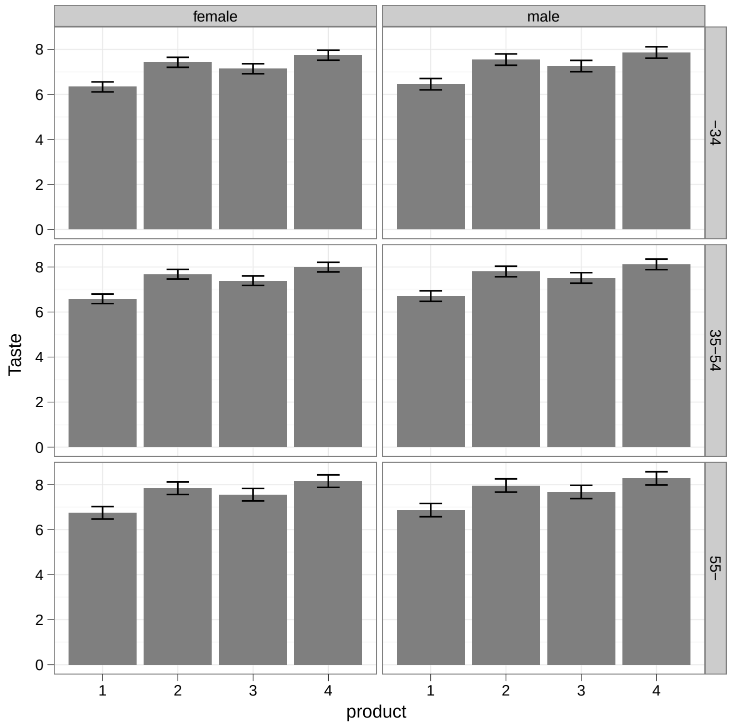


Abbildung A.1.: Datensatz A: Schätzungen der Bewertungen des Geschmacks für das Alter, Geschlecht und Produkt

A.1.2. texture ∽ product + age + gender, data=datA

```
aov(texture ~ product + age + gender, data = datA)

# Analysis of Variance Table
# ---------------------------
# Response: Texture
#             Df Sum Sq Mean Sq F value  Pr(>F)
# product      3   51.2    17.1  4.9000 0.00214
# age          2   20.4    10.2  2.9332 0.05345
# gender       1    4.0     4.0  1.1594 0.28171
# Residuals 2069 7203.9     3.5
```

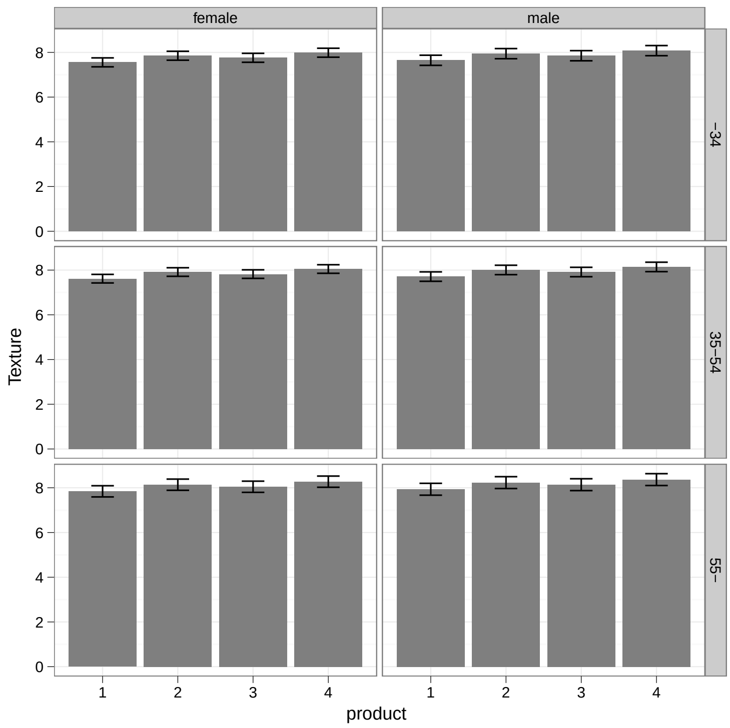


Abbildung A.2.: Datensatz A: Schätzungen der Bewertungen der Beschaffenheit für das Alter, Geschlecht und Produkt.

A.1.3. taste ∽ product + age + gender + Wechselwirkungen, data=datB

```
aov(taste ~ (product + age + gender)^2, data = datB)

# Analysis of Variance Table
# --------------------------
# Response: taste
#                   Df Sum Sq Mean Sq F value    Pr(>F)
# product            7  483.7    69.1 14.8065 < 2.2e-16
# age_new            2  130.9    65.5 14.0302 9.562e-07
# gen                1   42.3    42.3  9.0606  0.002669
# product:age_new   14   33.7     2.4  0.5165  0.924835
# product:gen        7   42.8     6.1  1.3108  0.241455
# age_new:gen        2   23.8    11.9  2.5534  0.078264
# Residuals       1140 5319.8     4.7
```

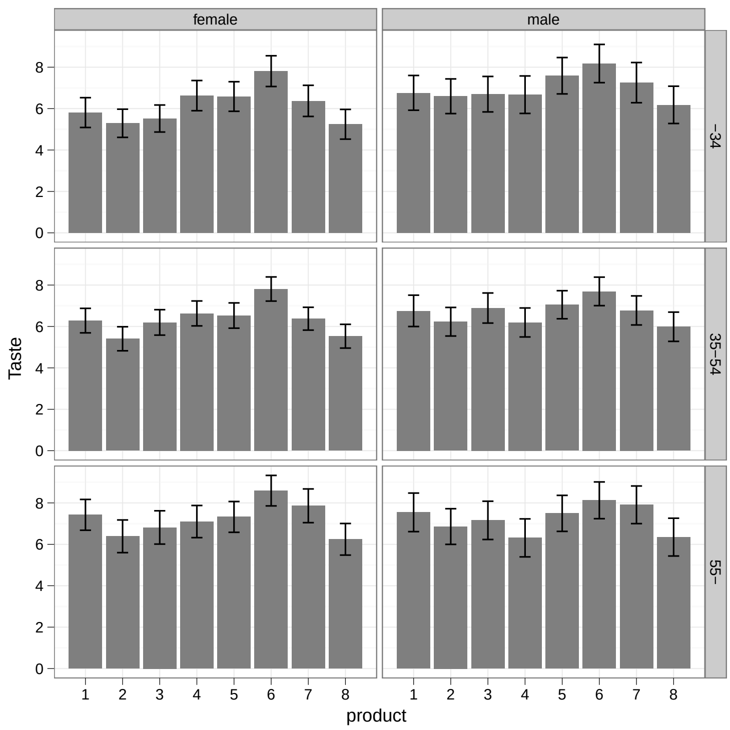


Abbildung A.3.: Datensatz B: Schätzungen der Bewertungen des Geschmacks für das Alter, Geschlecht und Produkt. Die Nummerierung von 1 bis 8 entspricht Coop, Frey, Cailler, Lindt CH, Lindt F, Lindt D, Aldi, Ritter.

A.1.4. texture ∽ product + age + gender + Wechselwirkungen, data=datB

```
aov(texture ~ (product + age + gender)^2 , data = datB)

# Analysis of Variance Table
# --------------------------
# Response: texture
#                      Df Sum Sq Mean Sq F value    Pr(>F)
# product               7  194.8    27.8  6.6242 1.070e-07
# age_new               2   54.4    27.2  6.4792  0.001592
# gen                   1    0.8     0.8  0.2006  0.654321
# product:age_new      14   19.1     1.4  0.3254  0.990944
# product:gen           7   36.5     5.2  1.2418  0.276551
# age_new:gen           2    1.1     0.6  0.1355  0.873336
# Residuals          1139 4785.8     4.2
```

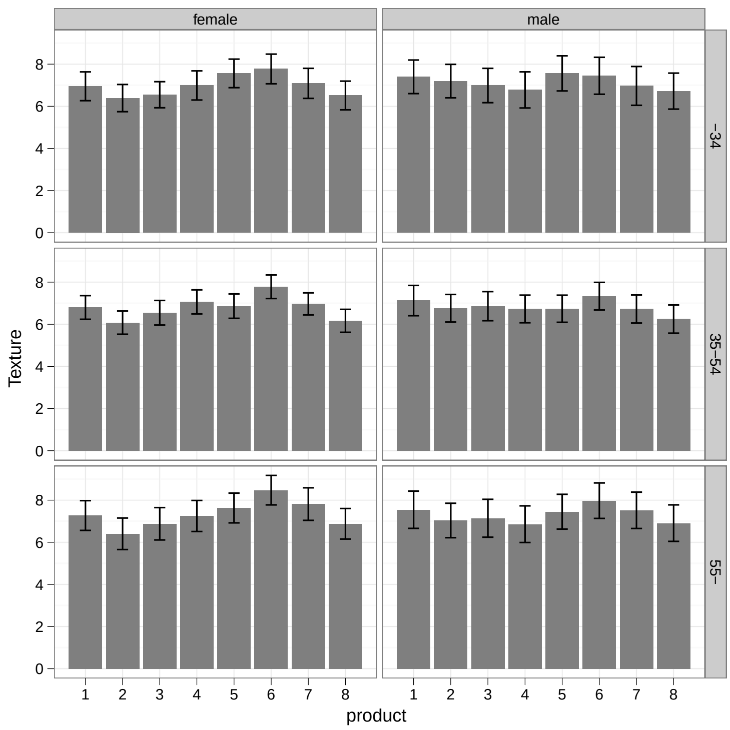


Abbildung A.4.: Datensatz B: Schätzungen der Bewertungen der Beschaffenheit für das Alter, Geschlecht und Produkt. Die Nummerierung von 1 bis 8 entspricht Coop, Frey, Cailler, Lindt CH, Lindt F, Lindt D, Aldi, Ritter.

A.2. Repeated Measures ANOVA

Die Vertrauensintervalle der Ergebnisse des Datensatz A und B sind jeweils mit Boostrap gebildet worden. Die ANOVA-Tabelle entspricht dem gewöhnlichen Modelloutput und ist nicht durch Boostrap korrigiert worden.

A.2.1. taste ∽ product + age + gender + Error(subject), data=datA

```
mod <- aov(taste ~ product + age + gender + Error(subject), data = datA)

# Analysis of Variance Table
# ---------------------------
# Error: subject
#            Df Sum Sq Mean Sq F value  Pr(>F)
# age         2   51.9    26.0  3.1407 0.04408
# gender      1    6.8     6.8  0.8256 0.36398
# Residuals 515 4255.7     8.3
# ---
# Error: Within
#             Df Sum Sq Mean Sq F value    Pr(>F)
# product      3  567.3   189.1  63.537 < 2.2e-16 ***
# Residuals 1554 4624.7     3.0
```

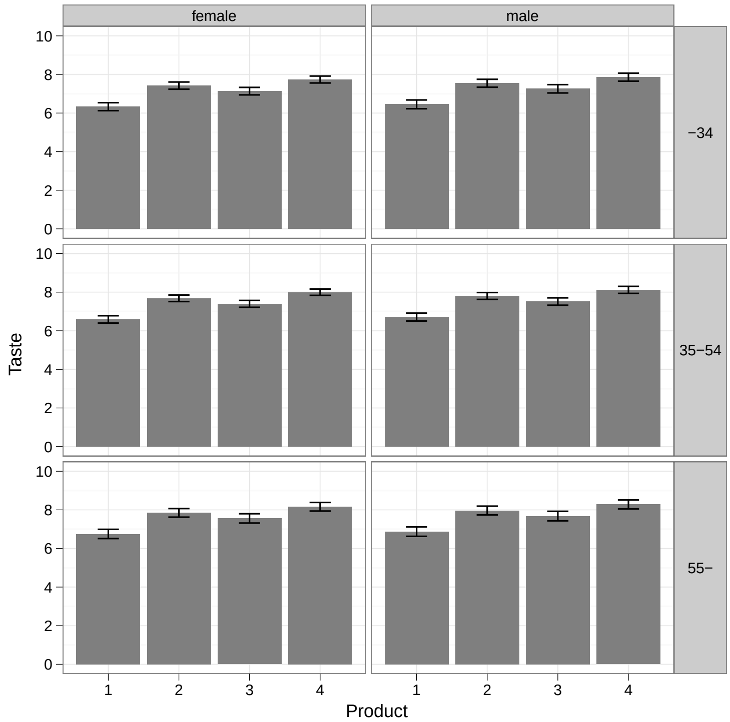


Abbildung A.5.: Datensatz A: Schätzungen der Bewertungen des Geschmacks durch ein Repeated Measures Modell für das Alter, Geschlecht und Produkt.

A.2.2. texture ∽ product + age + gender + Error(subject), data=datA

```
mod <- aov(texture ~ product + age + gender + Error(subject), data = datA)

# Analysis of Variance Table
# ---------------------------
# Error: subject
#            Df Sum Sq Mean Sq F value Pr(>F)
# age         2   20.4    10.2  0.9582 0.3843
# gender      1    4.0     4.0  0.3787 0.5385
# Residuals 515 5489.1    10.7
# ---
# Error: Within
#             Df  Sum Sq Mean Sq F value    Pr(>F)
# product      3   51.18   17.06  15.461 6.508e-10
# Residuals 1554 1714.82    1.10
```

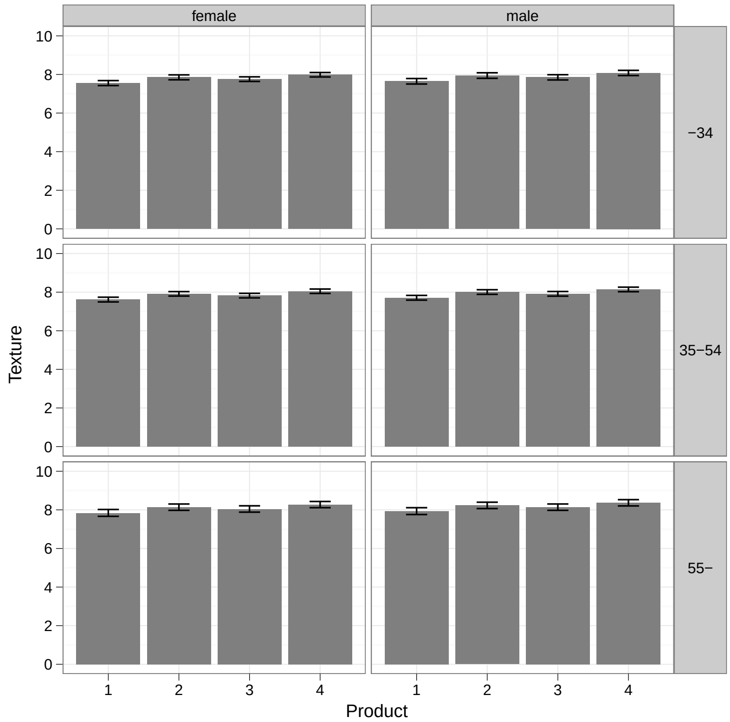


Abbildung A.6.: Datensatz A: Schätzungen der Bewertungen der Beschaffenheit durch ein Repeated Measures Modell für das Alter, Geschlecht und Produkt.

A.2.3. taste ∽ product + age + gender + Wechselwirkungen + Error(subject), data=datB

```
mod <- lmer(taste ~ (product + age + gender)^2 + (1 | subject), data = datB)

# Linear mixed model fit by REML
# Formula: taste ~ (product + age + gender)^2 + (1 | subject)
#    Data: dat
#   AIC  BIC logLik deviance REMLdev
#  5066 5248  -2497     4978    4994
# Random effects:
#  Groups   Name        Variance Std.Dev.
#  subject  (Intercept) 1.2137   1.1017
#  Residual             3.4656   1.8616
# Number of obs: 1174, groups: NJ, 196
#
# Fixed effects:
# ...
```

```
# Analysis of Variance Table
#                Df Sum Sq Mean Sq F value
# product         7 464.71   66.39 19.1560
# age             2  42.26   21.13  6.0975
# gender          1  13.90   13.90  4.0103
# product:age    14  45.50    3.25  0.9379
# product:gender  7  43.39    6.20  1.7885
# age:gender      2   7.71    3.85  1.1123

# zusätzlich mit der Bootstrap-Methode Vertrauensintervalle
der Parameterschätzungen bilden
```

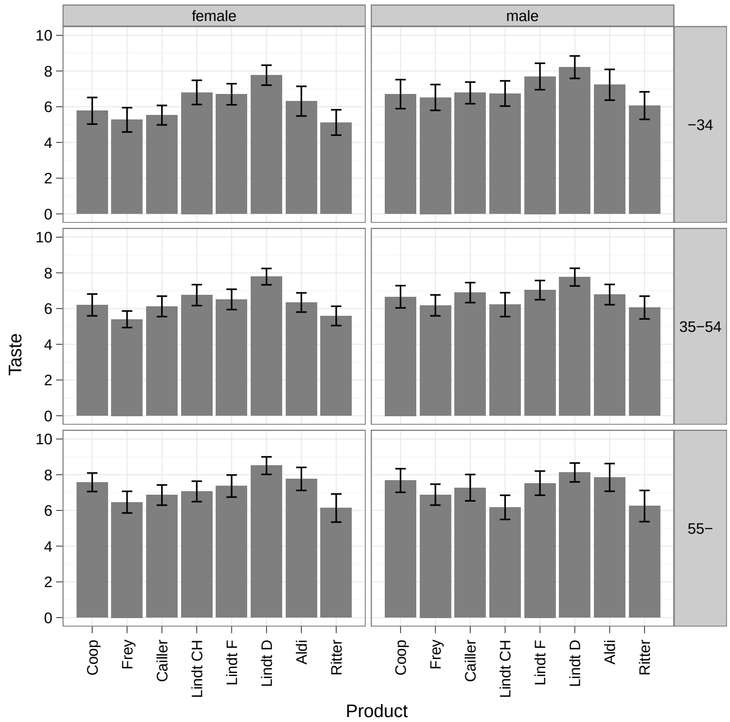


Abbildung A.7.: Datensatz B: Schätzungen der Bewertungen des Geschmacks durch ein Repeated Measures Modell für das Alter, Geschlecht und Produkt.

A.2.4. texture ∽ product + age + gender + Wechselwirkungen +(1 | subject), data=datB

```
mod <- lmer(texture ~ (product + age + gender)^2 + (1 | subject), data = datB)
```

```
# Linear mixed model fit by REML
# Formula: texture ~ (product + age + gender)^2 + (1 | subject)
#    Data: dat
#   AIC  BIC logLik deviance REMLdev
#  4923 5106  -2426     4831    4851
# Random effects:
#  Groups   Name        Variance Std.Dev.
#  subject   (Intercept) 1.1843   1.0883
#  Residual             3.0300   1.7407
# Number of obs: 1173, groups: NJ, 196
#
# Fixed effects:
# ...

# Analysis of Variance Table
#                 Df  Sum Sq Mean Sq F value
# product          7 187.907  26.844  8.8593
# age              2  16.519   8.259  2.7259
# gender           1   0.265   0.265  0.0873
# product:age     14  18.379   1.313  0.4333
# product:gender   7  34.139   4.877  1.6095
# age:gender       2   0.329   0.165  0.0543

# zusätzlich mit der Bootstrap-Methode Vertrauensintervalle
der Parameterschätzungen bilden
```

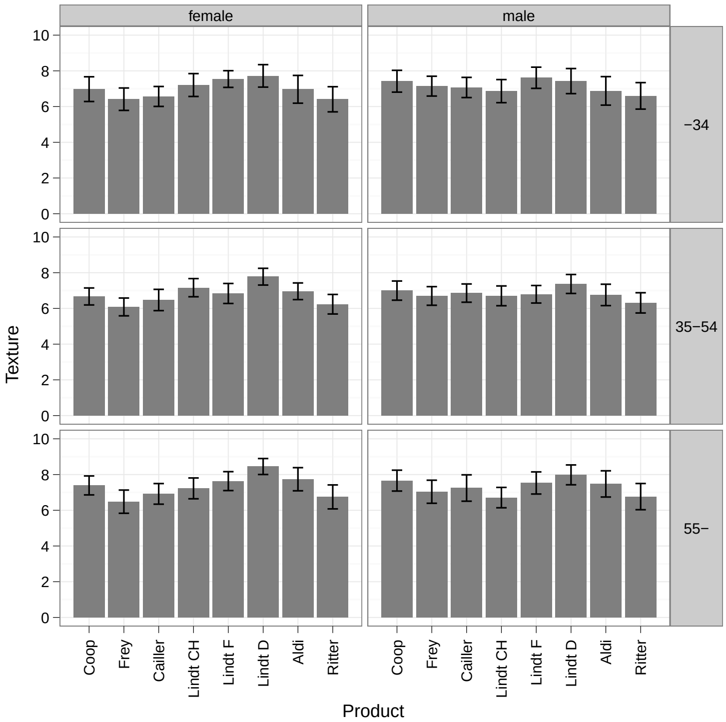


Abbildung A.8.: Datensatz B: Schätzungen der Bewertungen der Beschaffenheit durch ein Repeated Measures Modell für das Alter, Geschlecht und Produkt.

A.3. Mixed Model

Die Vertrauensintervalle der Ergebnisse des Datensatz A und B sind jeweils mit Boostrap gebildet worden. Die ANOVA-Tabelle entspricht dem gewöhnlichen Modelloutput und ist nicht durch Boostrap korrigiert worden.

A.3.1. taste ∽ product + age + gender + (1 + age | subject), data=datA

```
mod <- lmer(Taste ~ (product + age + gender)^2 + (age | subject), data = datA)

# Linear mixed model fit by REML
# Formula: taste ~ product + age + gender + (age | subject)
#    Data: dat
#   AIC  BIC logLik deviance REMLdev
#  8724 8803  -4348     8677    8696
# Random effects:
#  Groups   Name        Variance Std.Dev. Corr
#  subject  (Intercept) 1.48831  1.21996
#           age35-54    0.48215  0.69437  -0.415
#           age55-      0.99164  0.99581  -0.560  0.233
#  Residual             2.97602  1.72511
# Number of obs: 2076, groups: subject, 519
#
# Fixed effects:
#             Estimate Std. Error t value
# (Intercept)   6.3315     0.1341   47.22
# product2      1.0925     0.1071   10.20
# product3      0.8054     0.1071    7.52
# product4      1.4085     0.1071   13.15
# age35-54      0.2583     0.1422    1.82
# age55-        0.4231     0.1845    2.29
# gendermale    0.1160     0.1327    0.87

# Analysis of Variance Table
#         Df Sum Sq Mean Sq F value
# product  3 567.26  189.09 63.5372
# age      2  18.60    9.30  3.1255
# gender   1   2.27    2.27  0.7644

# zusätzlich mit der Bootstrap-Methode Vertrauensintervalle
der Parameterschätzungen bilden
```

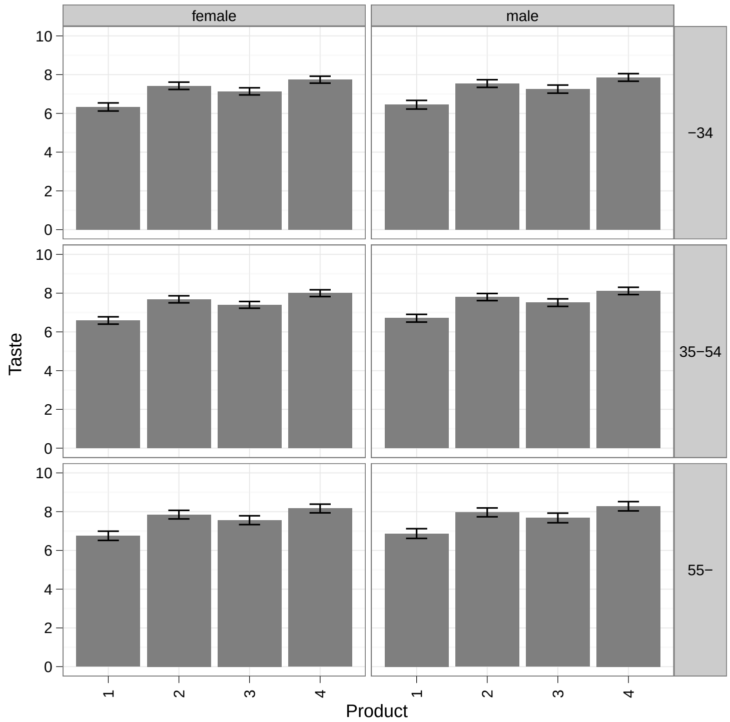


Abbildung A.9.: Datensatz A: Schätzungen der Bewertungen des Geschmacks durch ein Mixed Model für das Alter, Geschlecht und Produkt.

A.3.2. texture ∽ product + age + gender + (1 + age | subject), data=datA

```
mod <- lmer(texture ~ (product + age + gender)^2 + (age | subject), data = datA)

# Linear mixed model fit by REML
# Formula: texture ~ product + age + gender + (age | subject)
#    Data: dat
#   AIC  BIC logLik deviance REMLdev
#  7310 7389  -3641     7262    7282
# Random effects:
#  Groups   Name        Variance Std.Dev. Corr
#  subject  (Intercept) 2.8038   1.6745
#           age35-54    2.1754   1.4749   -0.557
#           age55-      3.8405   1.9597   -0.715  0.398
#  Residual             1.1035   1.0505
# Number of obs: 2076, groups: subject, 519
#
```

```
# Fixed effects:
#             Estimate Std. Error t value
# (Intercept)  7.55274    0.14246   53.02
# product2     0.29672    0.06521    4.55
# product3     0.20424    0.06521    3.13
# product4     0.43160    0.06521    6.62
# age35-54     0.06266    0.16368    0.38
# age55-       0.28802    0.20727    1.39
# gendermale   0.09575    0.14978    0.64

# Analysis of Variance Table
#         Df Sum Sq Mean Sq F value
# product  3 51.183  17.061 15.4611
# age      2  2.386   1.193  1.0811
# gender   1  0.451   0.451  0.4086

# zusätzlich mit der Bootstrap-Methode Vertrauensintervalle
der Parameterschätzungen bilden
```

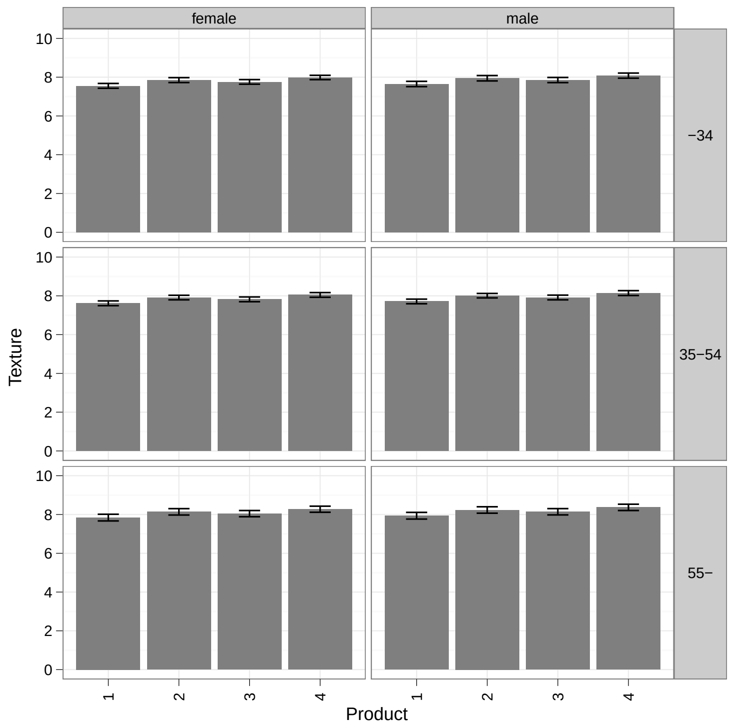


Abbildung A.10.: Datensatz A: Schätzungen der Bewertungen der Beschaffenheit durch ein Mixed Model für das Alter, Geschlecht und Produkt.

A.3.3. taste ∽ product + age + gender + (1 + age | subject), data=datB

```
mod <- lmer(taste ~ (product + age + gender)^2 + (age | subject), data = datB)

# Linear mixed model fit by REML
# Formula: taste ~ (product + age + gender)^2 + (age | subject)
#    Data: dat
#   AIC  BIC logLik deviance REMLdev
#  5072 5280  -2495     4974    4990
# Random effects:
#  Groups   Name         Variance Std.Dev. Corr
#  subject  (Intercept)  1.11490  1.05589
#               age35-54 0.98064  0.99027  -0.271
#               age55-   0.34574  0.58799  -0.611  0.165
#  Residual              3.46567  1.86163
# Number of obs: 1174, groups: NJ, 196
#
# Fixed effects:
# ...

# Analysis of Variance Table
#                 Df Sum Sq Mean Sq F value
# product          7 466.39   66.63 19.2249
# age              2  53.00   26.50  7.6461
# gender           1  12.60   12.60  3.6366
# product:age     14  44.26    3.16  0.9122
# product:gender   7  42.67    6.10  1.7590
# age:gender       2   9.22    4.61  1.3301

# zusätzlich mit der Bootstrap-Methode Vertrauensintervalle
der Parameterschätzungen bilden
```

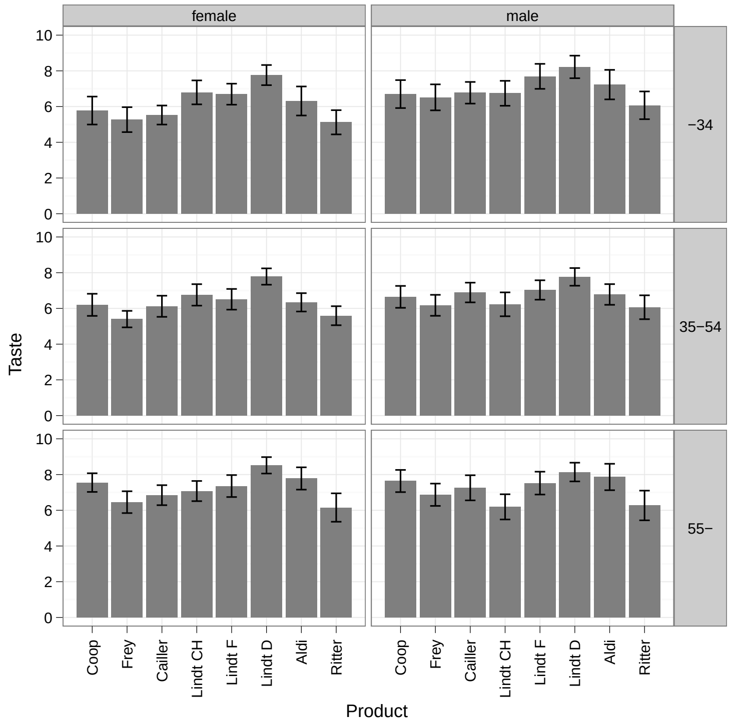


Abbildung A.11.: Datensatz B: Schätzungen der Bewertungen des Geschmacks durch ein Mixed Model für das Alter, Geschlecht und Produkt.

A.3.4. texture ∽ Product + age + gender + (1 + age | subject), data=datB

```
mod <- lmer(texture ~ (product + age + gender)^2 + (age | subject), data = datB)

# Linear mixed model fit by REML
# Formula: texture ~ (product + age + gender)^2 + (age | subject)
#    Data: dat
#   AIC  BIC logLik deviance REMLdev
#  4931 5139  -2424     4828    4849
# Random effects:
#  Groups   Name          Variance Std.Dev. Corr
#  subject  (Intercept)  1.08004  1.03925
#               age35-54 0.89328  0.94514  -0.275
#               age55-   0.48681  0.69772  -0.521  0.143
#  Residual              3.02997  1.74068
# Number of obs: 1173, groups: NJ, 196
#
```

```
# Fixed effects:
# ....

# Analysis of Variance Table
#                 Df  Sum Sq Mean Sq F value
# product          7 187.979  26.854  8.8628
# age              2  17.995   8.997  2.9695
# gen              1   0.264   0.264  0.0870
# product:age     14  18.369   1.312  0.4330
# product:gender   7  33.803   4.829  1.5937
# age:gender       2   0.372   0.186  0.0614

# zusätzlich mit der Bootstrap-Methode Vertrauensintervalle
der Parameterschätzungen bilden
```

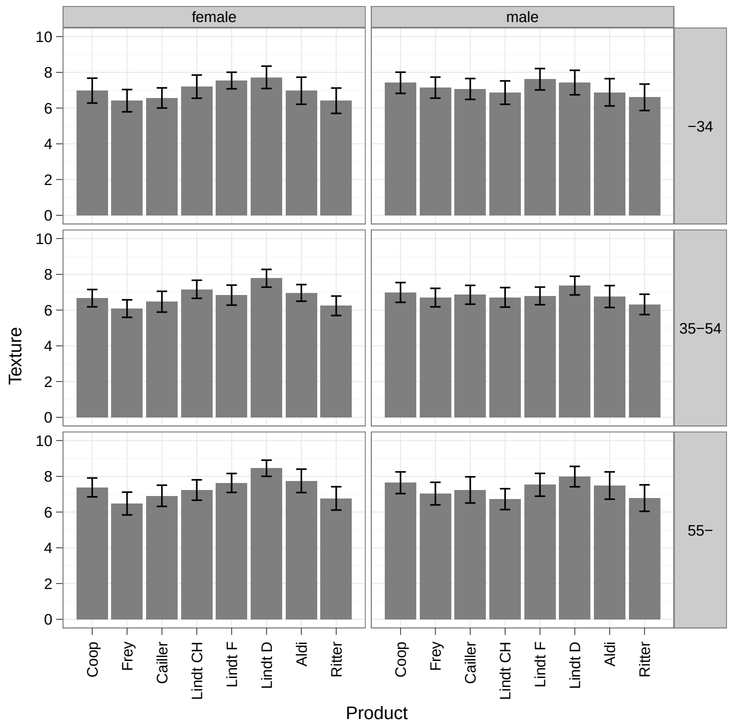


Abbildung A.12.: Datensatz B: Schätzungen der Bewertungen der Beschaffenheit durch ein Mixed Model für das Alter, Geschlecht und Produkt.