



STATISTIK- PROJEKT

VSEVOLOD DORSKIY

AUFGABE ZUM PROJEKT

- Die Homo Ludens GmbH produziert und vertreibt Spielzeug. Auch wenn elektronisches Spielzeug in den letzten Jahren immer mehr Einzug in unseren Kinderzimmern gehalten hat, setzt man weiterhin auch auf klassische Produkte.
- Neben ständiger Beobachtung der Märkte ist auch ein wachsames Auge auf die eigene Produktion zu richten, so dass auf ungünstige Veränderungen schnell und gezielt reagiert werden kann.
- Zusätzlich werden regelmäßig Informationen über die Konkurrenz gesammelt, um sicherzustellen, dass man hier nicht den Anschluss verliert.



AUFGABE ZUM PROJEKT

Untersuchen Sie das vorliegende Material und beantworten Sie dem Unternehmen einige Fragen:

- Wie sieht die aktuelle Fehlerzusammensetzung im Vergleich zur Vergangenheit aus? Gibt es Unterschiede? Falls ja: Wo ist der Brennpunkt?
- Lohnt sich die weitere Untersuchung eines Schweren Fliegers?
- Sind wir bei unserem Flugmodell noch konkurrenzfähig?
- Können wir das Modell soweit verbessern, dass wir konkurrenzfähig sind?

Soweit nicht anders angegeben, beträgt das Signifikanzniveau $1 - \alpha = 95\%$



AUFGABE 1

- **Wie sieht die aktuelle Fehlerzusammensetzung im Vergleich zur Vergangenheit aus? Gibt es Unterschiede? Falls ja: Wo ist der Brennpunkt?**

Zur Verfügung wurde folgendes Datenmaterial gestellt:

1_Produktqualität.xlsx

- Stichprobe der Fehlermeldungen der drei Produktionslinien und historische Daten über die bisherige Zusammensetzung.



A U F G A B E 1

Um Daten zu analysieren, benutzen wir den χ^2 -Test

- Der χ^2 -Test darf eingesetzt werden, um zu überprüfen, ob ein gegebener Datensatz der angenommen Verteilung entspricht (Anpassungstest)

Voraussetzungen zum Test

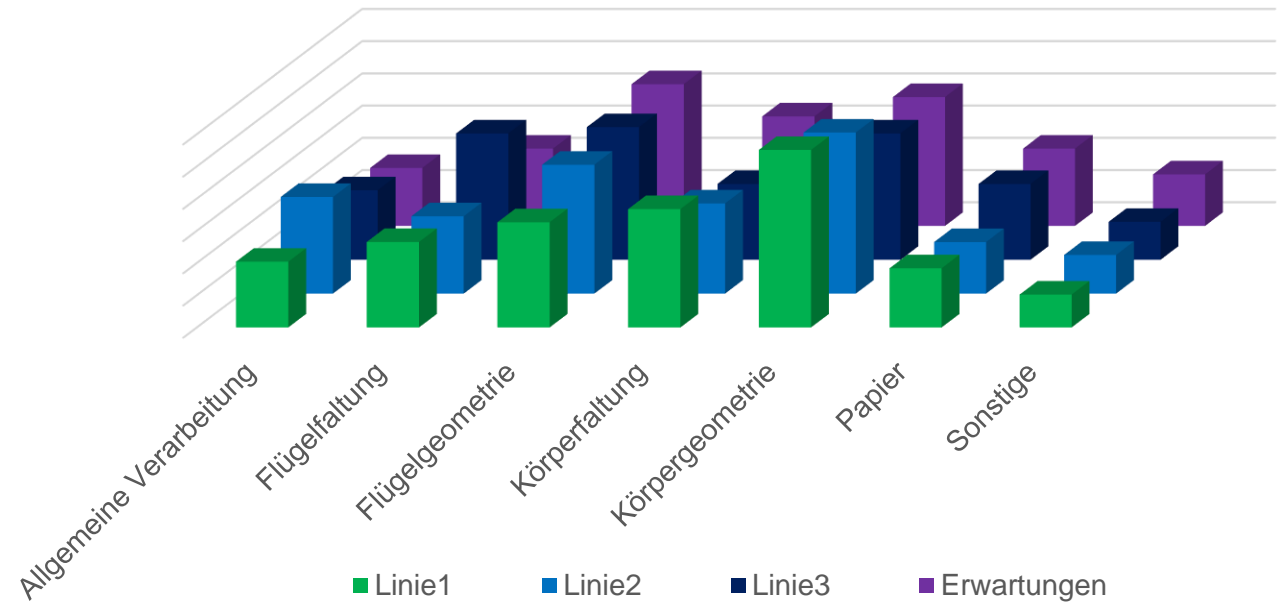
- ✓ Nominale Daten
- ✓ Die Anzahl aller Werte ≥ 5
- ✓ Die Daten sind unabhängig voneinander



AUFGABE 1

Vergleichung der Fehlerzusammensetzungen

Fehler	Linie1	Linie2	Linie3	Erwartungen
Allgemeine Verarbeitung	10%	15%	11%	9%
Flügelfaltung	13%	12%	20%	12%
Flügelgeometrie	16%	20%	21%	22%
Körperfaltung	18%	14%	12%	17%
Körpergeometrie	28%	25%	20%	20%
Papier	9%	8%	12%	12%
Sonstige	5%	6%	6%	8%



Die größten Abweichungen gibt es bei der Körpergeometrie und der Körperfaltung

AUFGABE 1

Wir führen die paarweise Vergleichen der heutigen Fehlermeldungen (Erwartungen) mit den historischen Daten der Linien in Excel und in R-Studio durch

Ursprüngliche Daten

Fehler	Linie1	Linie2	Linie3	Erwartungen
Allgemeine Verarbeitung	10	15	11	9%
Flügelfaltung	13	12	20	12%
Flügelgeometrie	16	20	21	22%
Körperfaltung	18	14	12	17%
Körpergeometrie	27	25	20	20%
Papier	9	8	12	12%
Sonstige	5	6	6	8%

Hypothese₀: Jeder Datensatz entspricht der angenommenen Verteilung (den Erwartungen). Es gibt keinen signifikanten Unterschied zwischen den Datensätzen (den Linien) und den heutigen Fehlermeldungen (den Erwartungen)

Hypothese₁: Nicht jeder Datensatz entspricht der angenommenen Verteilung (den Erwartungen). Es gibt signifikanten Unterschied zwischen einem der Datensätze (einer Linie) und den heutigen Fehlermeldungen (den Erwartungen)



AUFGABE 1

Vergleichung für Linie 1

Excelrechnung

Fehler	Historische Daten (Linie 1)	Erwartungen	Csi_sq (Linie1)
Allgemeine Verarbeitung	10	8,82	0,16
Flügelfaltung	13	11,76	0,13
Flügelgeometrie	16	21,56	1,43
Körperfaltung	18	16,66	0,11
Körpergeometrie	27	19,6	2,79
Papier	9	11,76	0,65
Sonstige	5	7,84	1,03
Summe	98	98	6,30

$$\chi^2 = 6,30; P_{wert} = 0,39; \text{Degrees of Freedom} = 6$$

$$\chi^2_{kritisch} = 12,59; \chi^2 < \chi^2_{kritisch}$$

Wir bleiben bei der *Hypothese₀*. Der Datensatz (Linie 1) entspricht der angenommenen Verteilung (den Erwartungen)

Rechnung in R-Studio

```
with(Dataset, chisq.test(x=Linie1,p=Erwartungen))
```

Chi-squared test for given probabilities

data: Linie1

X-squared = 6.3006, df = 6, p-value = 0.3904

$$P_{wert} = 0,39 > \alpha (0,05)$$



AUFGABE 1

Vergleichung für Linie 2

Excelrechnung

Fehler	Historische Daten (Linie 2)	Erwartungen	Csi_sq (Linie2)
Allgemeine Verarbeitung	15	9	4,00
Flügelfaltung	12	12	-
Flügelgeometrie	20	22	0,18
Körperfaltung	14	17	0,53
Körpergeometrie	25	20	1,25
Papier	8	12	1,33
Sonstige	6	8	0,50
Summe	100	100	7,79

$\chi^2 = 7,79$; $P_{Wert} = 0,25$; Degrees of Freedom = 6

$\chi^2_{kritisch} = 12,59$; $\chi^2 < \chi^2_{kritisch}$

Wir bleiben bei der *Hypothese*₀. Der Datensatz (Linie 2) entspricht der angenommenen Verteilung (den Erwartungen)

Rechnung in R-Studio

```
with(Dataset, chisq.test(x=Linie2,p=Erwartungen))
```

Chi-squared test for given probabilities

data: ObsLinie2

X-squared = 7.7946, df = 6, p-value = 0.2535

$P_{Wert} = 0,25 > \alpha (0,05)$



AUFGABE 1

Vergleichung für Linie 3

Excelrechnung

Fehler	Historische Daten (Linie 3)	Erwartungen	Csi_sq (Linie3)
Allgemeine Verarbeitung	11	9,18	0,36
Flügelfaltung	20	12,24	4,92
Flügelgeometrie	21	22,44	0,09
Körperfaltung	12	17,34	1,64
Körpergeometrie	20	20,4	0,01
Papier	12	12,24	0,00
Sonstige	6	8,16	0,57
Summe	102	102	7,60

$\chi^2 = 7,60$; $P_{wert} = 0,27$; Degrees of Freedom = 6

$\chi^2_{kritisch} = 12,59$; $\chi^2 < \chi^2_{kritisch}$

Wir bleiben bei der *Hypothese*₀. Der Datensatz (Linie 3) entspricht der angenommenen Verteilung (den Erwartungen)

Alle Datensätze (Linien) entsprechen der angenommenen Verteilung (den Erwartungen). Es gibt keinen signifikanten Unterschied zwischen den vorherigen Daten und der heutigen Fehlerzusammensetzung

Rechnung in R-Studio

```
with(Dataset, chisq.test(x=Linie3,p=Erwartungen))
```

Chi-squared test for given probabilities

data: ObsLinie3

X-squared = 7.6018, df = 6, p-value = 0.2688

$P_{wert} = 0,27 > \alpha (0,05)$



AUFGABE 2

Lohnt sich die weitere Untersuchung eines schweren Fliegers?

Zur Verfügung wurde folgendes Datenmaterial gestellt:

2_Gewicht.xlsx

- Zwei Stichproben zum Vergleich des Grundmodells mit einem beschwerten Flieger.



AUFGABE 2

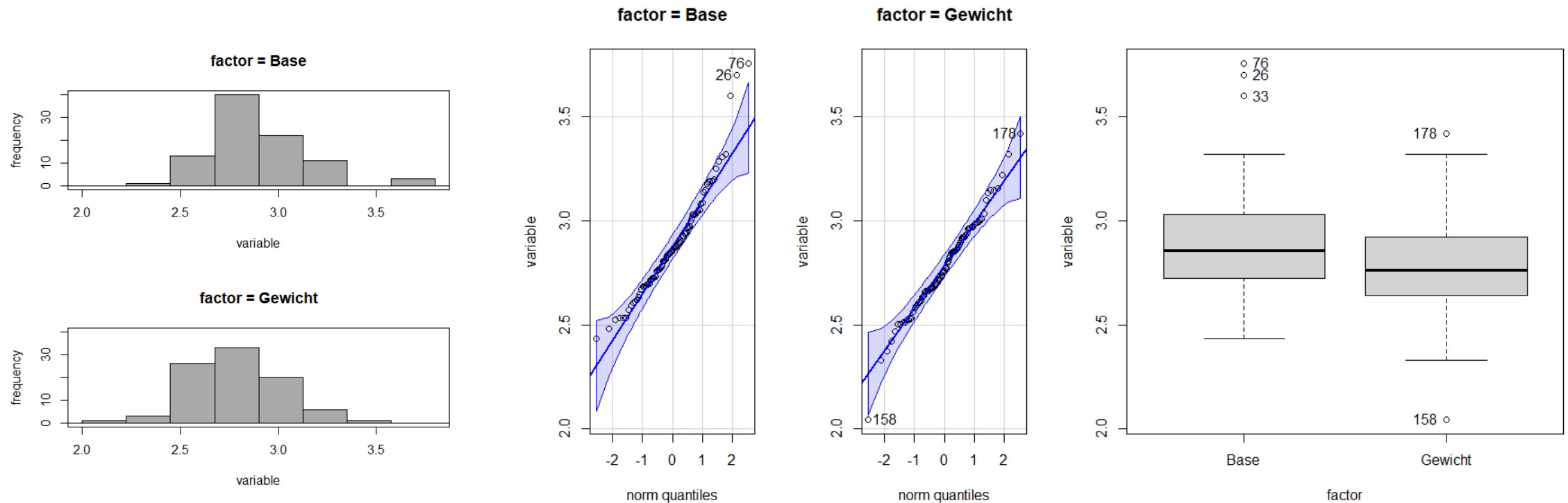
	Base		Gewicht
Min	2,434	Min.	2,046
1st Qu.	2,721	1st Qu.	2,646
Median	2,855	Median	2,76
Mean	2,888	Mean	2,779
3rd Qu.	3,024	3rd Qu.	2,921
Max.	3,753	Max.	3,417

	mean	sd	IQR	0%	25%	50%	75%	100% n	
Base	2,887516	0,247016	0,30315	2,4344	2,72125	2,8554	3,0244	3,7527	90
Gewicht	2,778744	0,222515	0,2757	2,0462	2,6457	2,7597	2,9214	3,4169	90

- **Daten sind intervallskaliert**
- **Die Mittelwerte und die Streuung sind vergleichbar**



AUFGABE 2



- **Kein größer Unterschied zwischen den Mittelwerten, die Varianzgleichheit**
- **Beide Verteilungen deuten auf Ausreißer hin, Datensätze müssen für Normalität überprüft werden**

AUFGABE 2

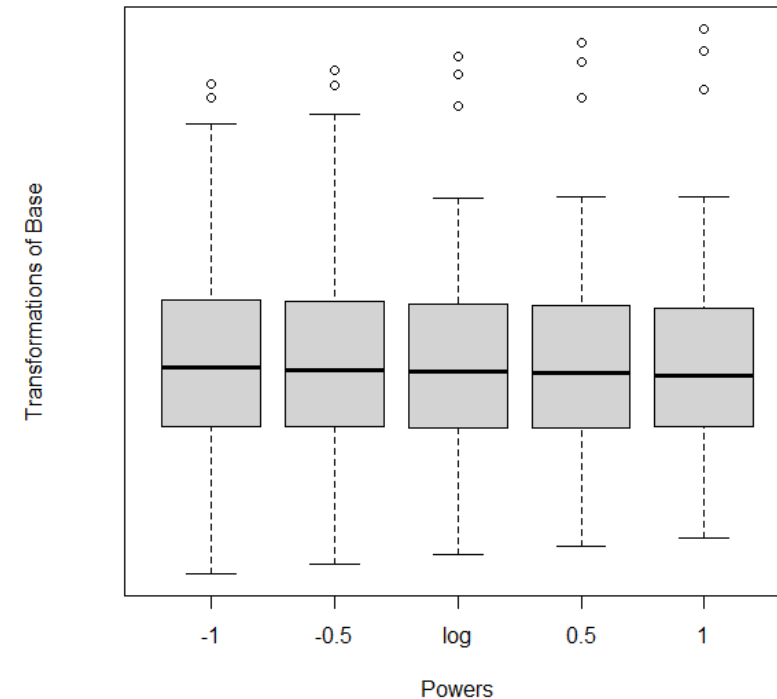
Prüfung der Normalverteilung

Shapiro-Wilk normality test

p-values adjusted by the Holm method:

	unadjusted	adjusted
Base	0,00032931	0,00065861
Gewicht	0,45390486	0,45390486

$P_{Wert}(Base) < 0,05$ – keine Normalverteilung



- Die Stichprobe (Gewicht) ist normalverteilt, aber der Datensatz (Base) hat keine Normalität
- Die Transformationsmöglichkeit zeigt keine Verbesserung bei der Normalisierung der Daten. Es wird ein nichtparametrischer Test durchgeführt

AUFGABE 2

***Hypothese₀*: Median (Base) \geq Median (Gewicht). Die Flugzeit des Basismodells \geq die Flugzeit des beschwerten Fliegers**

***Hypothese₁*: Median (Base) $<$ Median (Gewicht). Die Flugzeit des Basismodells $<$ die Flugzeit des beschwerten Fliegers**

Wilcoxon rank sum test with continuity correction

data: variable by factor

W = 5042, **p-value = 0.9977**

alternative hypothesis: true location shift is less than 0

$P_{Wert} > 0,05$ –wir verbleiben bei der *Hypothese₀*

Es gibt keine genügenden Beweise, dass die Flugzeit des beschwerten Fliegers besser ist als die des Grundmodells. Die weitere Untersuchung eines schweren Fliegers lohnt sich nicht



A U F G A B E 3

- **Sind wir bei unserem Flugmodell noch konkurrenzfähig?**

Zur Verfügung wurde folgendes Datenmaterial gestellt:

3_Technischer Vergleich mit Konkurrenz.xlsx

- Daten zum Vergleich des eigenen Flugmodells mit Konkurrenzmodellen. Der Datensatz bietet Ihnen einen Flugzeitvergleich [s] zwischen unserem Modell und der Konkurrenz.
- Die technischen Parameter der einzelnen Modelle können Sie der anhängenden Tabelle entnehmen.



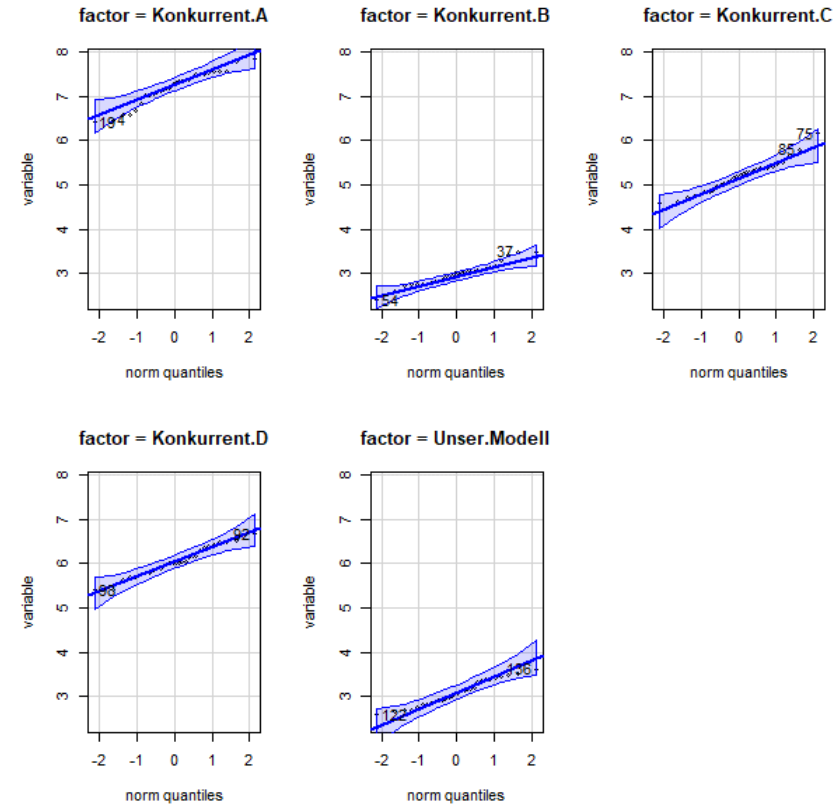
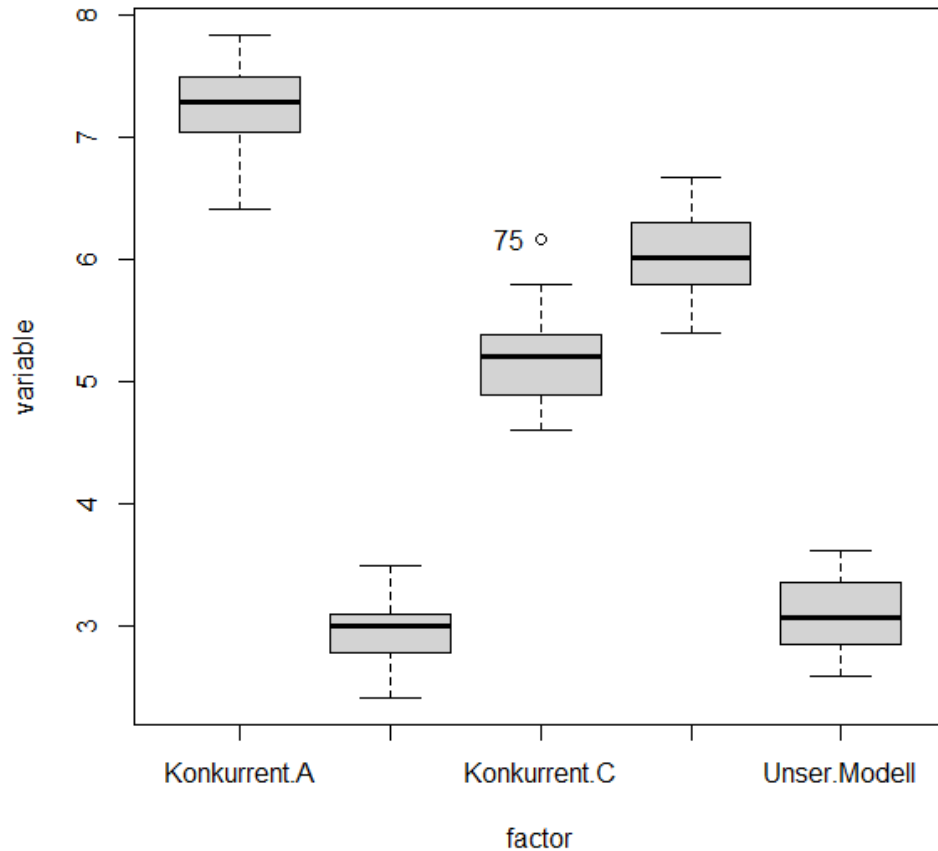
AUFGABE 3

Unser.Modell	Konkurrent.A	Konkurrent.B	Konkurrent.C	Konkurrent.D
Min. :2.590	Min. :6.420	Min. :2.410	Min. :4.600	Min. :5.400
1st Qu. :2.855	1st Qu. :7.045	1st Qu. :2.788	1st Qu. :4.912	1st Qu. :5.817
Median :3.060	Median :7.290	Median :3.000	Median :5.205	Median :6.020
Mean :3.081	Mean :7.210	Mean :2.977	Mean :5.183	Mean :6.037
3rd Qu. :3.345	3rd Qu. :7.500	3rd Qu. :3.083	3rd Qu. :5.378	3rd Qu. :6.265
Max. :3.620	Max. :7.840	Max. :3.490	Max. :6.160	Max. :6.670

	mean	sd	IQR	0%	25%	50%	75%	100%	n
Konkurrent,A	7,210333	0,37577	0,455	6,42	7,045	7,29	7,5	7,84	30
Konkurrent,B	2,977333	0,247678	0,295	2,41	2,7875	3	3,0825	3,49	30
Konkurrent,C	5,182667	0,355061	0,465	4,6	4,9125	5,205	5,3775	6,16	30
Konkurrent,D	6,037333	0,321719	0,4475	5,4	5,8175	6,02	6,265	6,67	30
Unser,Modell	3,081	0,2832	0,49	2,59	2,855	3,06	3,345	3,62	30

- Die Daten sind intervallskaliert
- Die Mittelwerte der Gruppen unterscheiden sich stark, die Streuung in den Gruppen nicht so groß

AUFGABE 3



- Der Unterschied zwischen den Gruppen ist offensichtlich
- Die Datensätze sehen normalverteilt aus



AUFGABE 3

Test auf Normalität

p-values adjusted by the Holm method:

	unadjusted	adjusted
Konkurrent.A	0.11129	0.55644
Konkurrent.B	0.42506	1.00000
Konkurrent.C	0.46068	1.00000
Konkurrent.D	0.69817	1.00000
Unser.Modell	0.43543	1.00000

Alle $P_{\text{Werte}} > \alpha$: Alle Datensätze sind normalverteilt

Bartlett test of homogeneity of variances

data: variable by factor

Bartlett's K-squared = 6.29, df = 4, p-value = 0.1785

$P_{\text{Wert}} > \alpha$: Die Varianzhomogenität ist bestätigt



AUFGABE 3

***Hypothese₀*: Alle Stichproben haben die gleichen Mittelwerte**

***Hypothese₁*: Nicht alle Stichproben weisen die gleichen Mittelwerte auf. Es gibt einen signifikanten Unterschied zwischen mindestens zwei Stichproben**

```
summary(AnovaModel.1)
```

```
          Df Sum Sq Mean Sq F value Pr(>F)
factor      4  411.5   102.9    1004 <2e-16 ***
Residuals  145   14.9     0.1
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
      mean      sd data:n
Konkurrent.A 7.210333 0.3757703    30
Konkurrent.B 2.977333 0.2476780    30
Konkurrent.C 5.182667 0.3550613    30
Konkurrent.D 6.037333 0.3217188    30
Unser.Modell 3.081000 0.2832003    30
```

- **$P_{Wert} < \alpha$: Wir wenden uns der *Hypothese₁* zu. Es gibt einen signifikanten Unterschied zwischen mindestens zwei Stichproben**
- **Wir gehen nun zum paarweisen Vergleich über**

AUFGABE 3

Paarweiser Vergleich

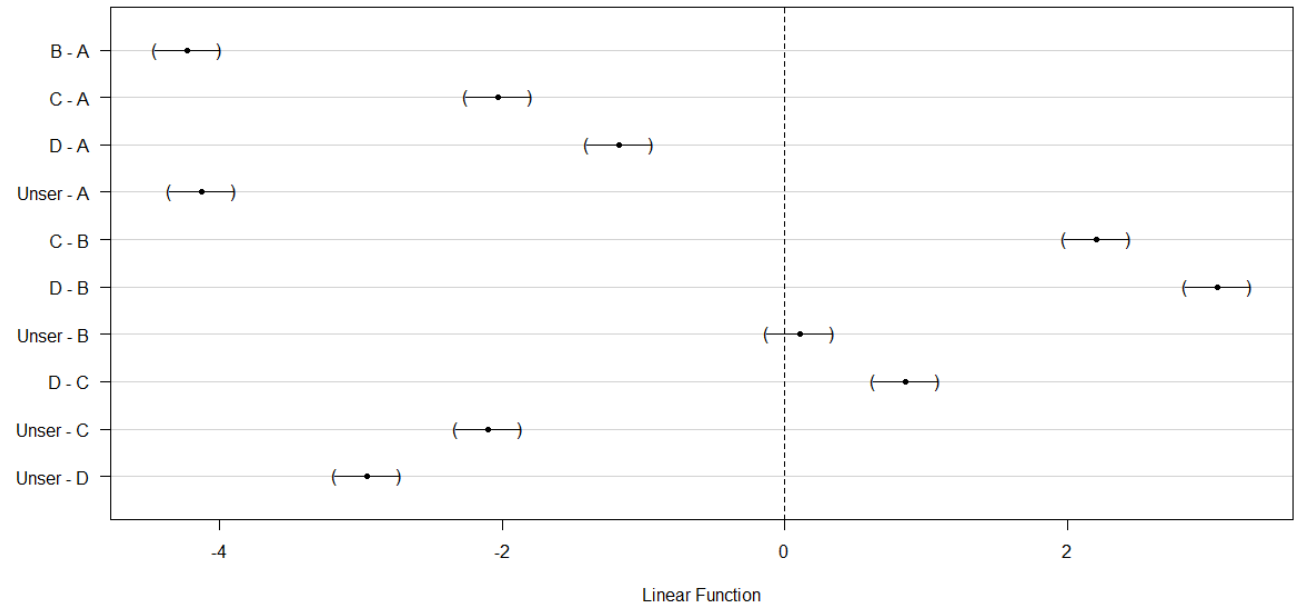
Linear Hypotheses:

	Estimate	lwr	upr
B - A == 0	-4.2330	-4.4613	-4.0047
C - A == 0	-2.0277	-2.2560	-1.7993
D - A == 0	-1.1730	-1.4013	-0.9447
Unser - A == 0	-4.1293	-4.3577	-3.9010
C - B == 0	2.2053	1.9770	2.4337
D - B == 0	3.0600	2.8317	3.2883
Unser - B == 0	0.1037	-0.1247	0.3320
D - C == 0	0.8547	0.6263	1.0830
Unser - C == 0	-2.1017	-2.3300	-1.8733
Unser - D == 0	-2.9563	-3.1847	-2.7280

A	B	C	D	Unser
"a"	"b"	"c"	"d"	"b"

- Unser Modell ist in einer Familie (b) mit dem Konkurrent B

95% family-wise confidence level



Unser Modell ist vergleichbar nur mit dem Konkurrenten B und schlechter als Konkurrenten A,C und D. Es gibt keine Hinweise, dass unser Flugmodell noch konkurrenzfähig ist

AUFGABE 3

	Körper-breite	Flügel-länge	Körper-länge	Papier-gewicht
Unser Modell	20 mm	80 mm	80 mm	80 g
Konkurrent A	35 mm	130 mm	130 mm	80 g
Konkurrent B	20 mm	80 mm	80 mm	80 g
Konkurrent C	35 mm	80 mm	80 mm	80 g
Konkurrent D	35 mm	80 mm	130 mm	90 g

Die Ähnlichkeit der Flugeigenschaften unseres Modells mit dem Konkurrenten B wird durch die gleichen Modellparameter bestätigt



A U F G A B E 4

- **Können wir das Modell soweit verbessern, dass wir konkurrenzfähig sind?**

Zur Verfügung wurde folgendes Datenmaterial gestellt:

4_Produktoptimierung.xlsx

- Es sind bereits Flugdaten eines geänderten Modells erfasst worden, in denen wichtige Parameter für die Flugzeit untersucht wurden



AUFGABE 4

Prüfung der Normalverteilung

Shapiro-Wilk normality test

p-values adjusted by the Holm method:

	unadjusted	adjusted
20 mm,130 mm,130 mm,80 g	0,55958	1
20 mm,130 mm,130 mm,90 g	0,56399	1
20 mm,130 mm,80 mm,80 g	0,68726	1
20 mm,130 mm,80 mm,90 g	0,85775	1
20 mm,80 mm,130 mm,80 g	0,73391	1
20 mm,80 mm,130 mm,90 g	0,73805	1
20 mm,80 mm,80 mm,80 g	0,73938	1
20 mm,80 mm,80 mm,90 g	0,67322	1
35 mm,130 mm,130 mm,80 g	0,67137	1
35 mm,130 mm,130 mm,90 g	0,59614	1
35 mm,130 mm,80 mm,80 g	0,42902	1
35 mm,130 mm,80 mm,90 g	0,87351	1
35 mm,80 mm,130 mm,80 g	0,2986	1
35 mm,80 mm,130 mm,90 g	0,91805	1
35 mm,80 mm,80 mm,80 g	0,18676	1
35 mm,80 mm,80 mm,90 g	0,69987	1

Alle $P_{\text{Werte}} > \alpha$: Alle Datensätze sind normalverteilt



A U F G A B E 4

Prüfung auf Homonität der Varianzen

Bartlett test of homogeneity of variances

data: Zeit by interaction(Flügel.L, Körper.B, Körper.L, Papier)

Bartlett's K-squared = 5.6717, df = 15, p-value = 0.9848

$P_{Wert} > \alpha$: Die Varianzen der einzelnen Gruppen können als gleich angesehen werden



AUFGABE 4

Suche nach dem optimalen Modell

	20 mm	35 mm
Körper-breite	-1	1

	80 mm	130 mm
Flügel-länge	-1	1

	80 mm	130 mm
Körper-länge	-1	1

	80 g	90 g
Papier-gewicht	-1	1

Um ein lineares Regressionsmodell zu erstellen, kodieren wir die Daten mit (-1) und (+1)



AUFGABE 4

Erstellen des linearen Regressionsmodells (Schritt 1)

`lm(formula = Zeit ~ (FL + KB + KL + P)^4, data = Dataset)`

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	5,95202	0,0044487	1337,918	<2e-16	***
FL	1,4344	0,0044487	322,429	<2e-16	***
KB	1,05098	0,0044487	236,243	<2e-16	***
KL	-0,3861	0,0044487	-86,799	<2e-16	***
P	0,83169	0,0044487	186,95	<2e-16	***
FL:KB	0,00852	0,0044487	1,915	0,0561	,
FL:KL	0,0084	0,0044487	1,887	0,0598	,
FL:P	-0,0063	0,0044487	-1,41	0,1593	
KB:KL	0,00056	0,0044487	0,126	0,8994	
KB:P	-0,0089	0,0044487	-1,99	0,0471	*
KL:P	-0,0006	0,0044487	-0,145	0,8846	
FL:KB:KL	-0,0064	0,0044487	-1,438	0,1512	
FL:KB:P	-0,0046	0,0044487	-1,044	0,2969	
FL:KL:P	-0,0083	0,0044487	-1,859	0,0636	,
KB:KL:P	-0,0076	0,0044487	-1,709	0,0881	,
FL:KB:KL:P	0,00694	0,0044487	1,559	0,1196	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09747 on 464 degrees of freedom

Multiple R-squared: **0.9977**, Adjusted R-squared: **0.9976**

F-statistic: **1.349e+04** on 15 and 464 DF, **p-value: < 2.2e-16**

- Die 4 Fach-Wechselwirkungen sind nicht signifikant und können aus dem Modell entfernt werden
- $F = 13490$, $P_{\text{Wert}} < 0,05$ (Das Modell liefert signifikanten Erklärungsbeitrag)
- $R\text{-squared} \sim 1$ (Das zeigt, wie viel Prozent der Varianz der abhängigen Variable (Zeit) erklärt wird. Hier ~ 100%)

- $F = 14400$ (besser als auf dem Schritt 1),
 $P_{Wert} < 0,05$
- $R\text{-squared} \sim 1$

AUFGABE 4

Erstellen des linearen Regressionsmodells (Schritt 3)

lm(formula = Zeit ~ (FL + KB + KL + P)^2, data = Dataset)

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	5,95202	0,0044816	1328,089	<2e-16	***
FL	1,4344	0,0044816	320,06	<2e-16	***
KB	1,05098	0,0044816	234,508	<2e-16	***
KL	-0,3861	0,0044816	-86,162	<2e-16	***
P	0,83169	0,0044816	185,576	<2e-16	***
FL:KB	0,00852	0,0044816	1,901	0,0579	,
FL:KL	0,0084	0,0044816	1,873	0,0616	,
FL:P	-0,0063	0,0044816	-1,399	0,1624	
KB:KL	0,00056	0,0044816	0,126	0,9002	
KB:P	-0,0089	0,0044816	-1,976	0,0488	*
KL:P	-0,0006	0,0044816	-0,144	0,8855	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09819 on 469 degrees of freedom

Multiple R-squared: **0.9977**, Adjusted R-squared: **0.9976**

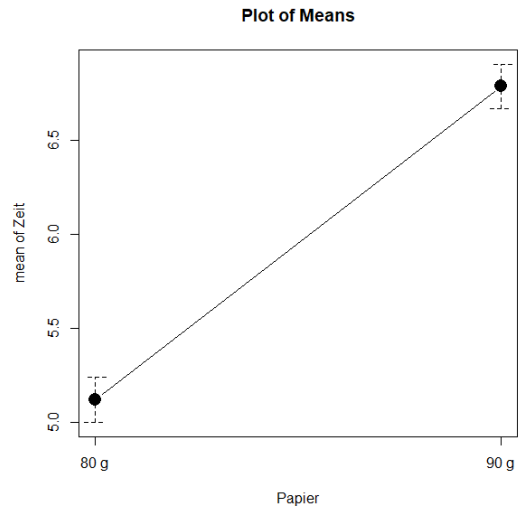
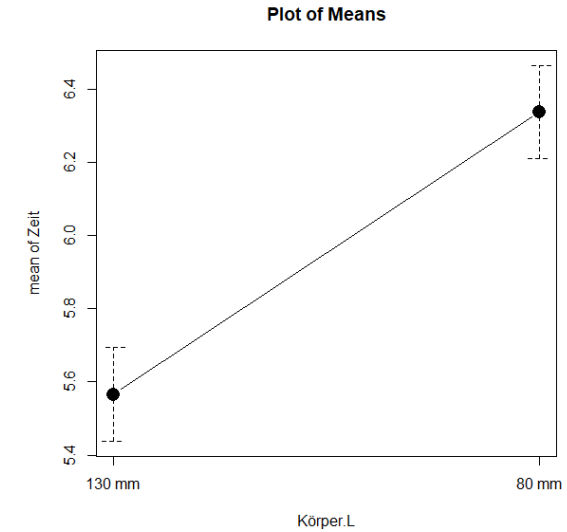
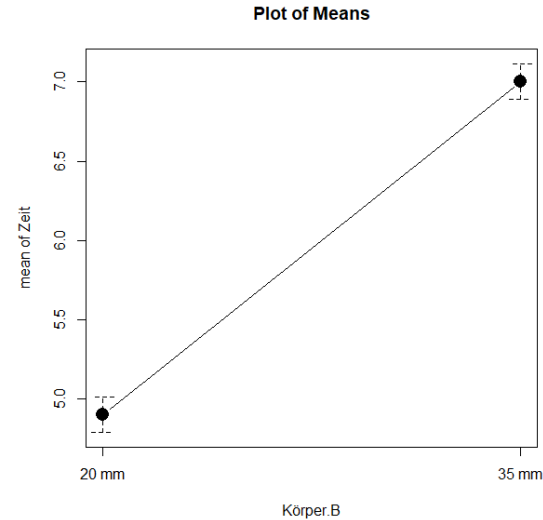
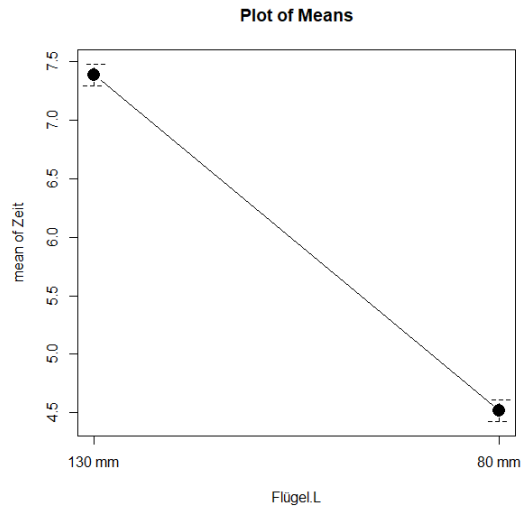
F-statistic: **1.993e+04** on 10 and 469 DF, **p-value: < 2.2e-16**

- Die 2 Fach-Wechselwirkung **KB:P** ist signifikant für das Modell. Die anderen 2 Fach-Wechselwirkungen können aus dem Modell entfernt werden
- **F = 19930** (besser als auf den Schritten 1 und 2), $P_{Wert} < 0,05$
- **R-squared ~ 1**

- Im Modell verbleiben die Konstanten, alle Hauptfaktoren und die Wechselwirkung $KB:P$
- $F = 39520$ (besser als auf den vorherigen Schritten), $P_{Wert} < 0,05$
- $R\text{-squared} \sim 1$
- Es ist das endgültige Modell

AUFGABE 4

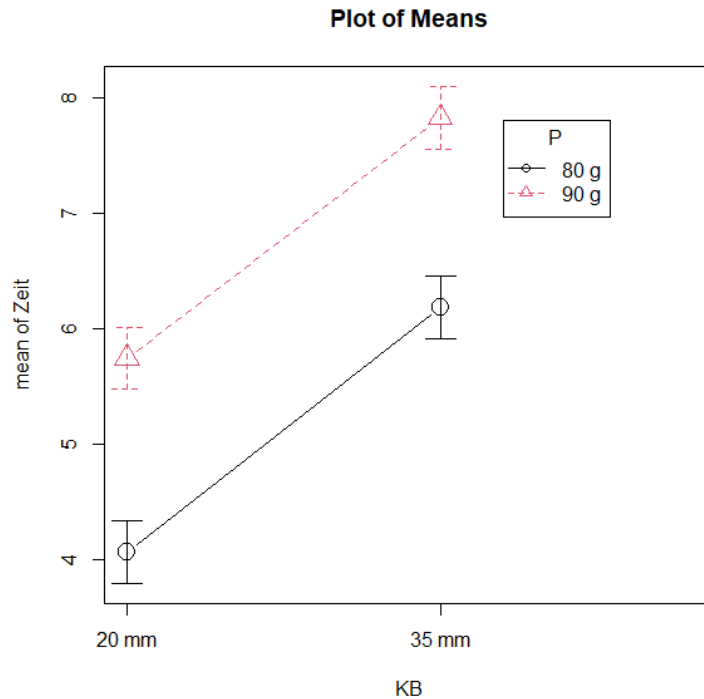
Optimierung des Modells



- Die Änderung der Flügel-länge von 80 mm auf 130 mm ergibt ein besseres Resultat
- Die Änderung der Körper-breite von 20 mm auf 35 mm bringt positive Wirkung
- Die Änderung der Körper-länge von 80 mm auf 130 mm bringt negative Wirkung
- Papier-Gewichtzunahme von 80 g auf 90 g hat einen positiven Effekt

AUFGABE 4

Optimierung des Modells



Das Wechselwirkungsdiagramm der Faktoren (Körper-breite: Papier-gewicht) zeigt, dass diese Interaktion in unserem Fall keine signifikante Rolle spielt (die Graphen konvergieren nur sehr schwach)

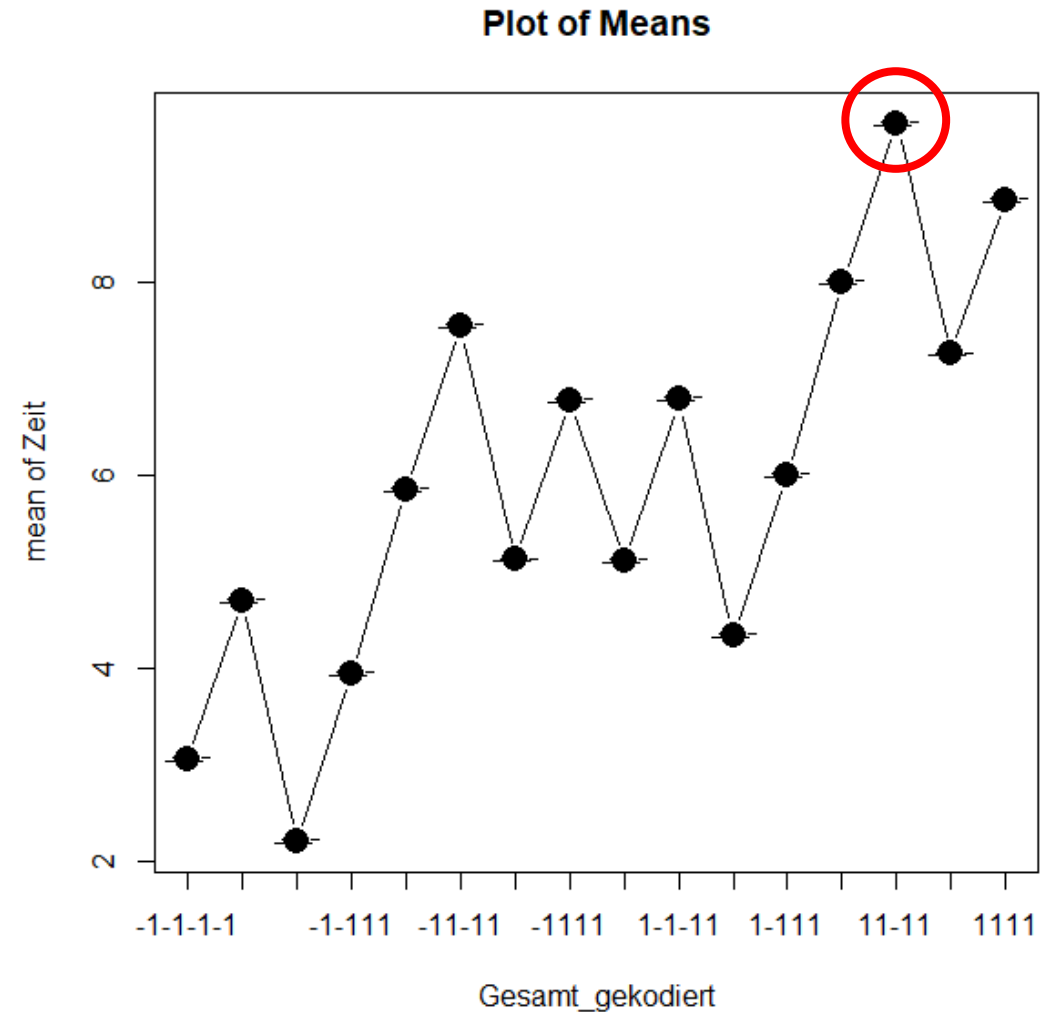


AUFGABE 4

Optimierung des Modells

Optimale Einstellung:

Flügel-länge	+1	130 mm
Körper-breite	+1	35 mm
Körper-länge	-1	80 mm
Papier-gewicht	+1	90 g



$$\text{Zeit}_{\text{maximal}} = 5,95 + 1,43*(1) + 1,05*(1) - 0,39*(-1) + 0,83*(1) - 0,009*(0) = 9,65$$