Projektarbeit Statisik mit R

Von

Nour-Eddine Kzaiber

Vadim Welz

Aufgabe 1:

Wie sieht die aktuelle Fehlerzusammen setzung im Vergleich zur Vergangenheit aus?

Fehler	Linie 1	Linie 2	Line 3	Summe	Aktuelle Fehlerzusammen setzung	Erwartung aus der Vergangenheit
Allgemeine Verarbeitung	10	15	11	36	12%	9%
Flügelfaltung	13	12	20	45	15%	12%
Flügelgeometrie	16	20	21	57	19%	22%
Körperfaltung	18	14	12	44	15%	17%
Körpergeometrie	27	25	20	72	24%	20%
Papier	9	8	12	29	10%	12%
Sonstige	5	6	6	17	6%	8%

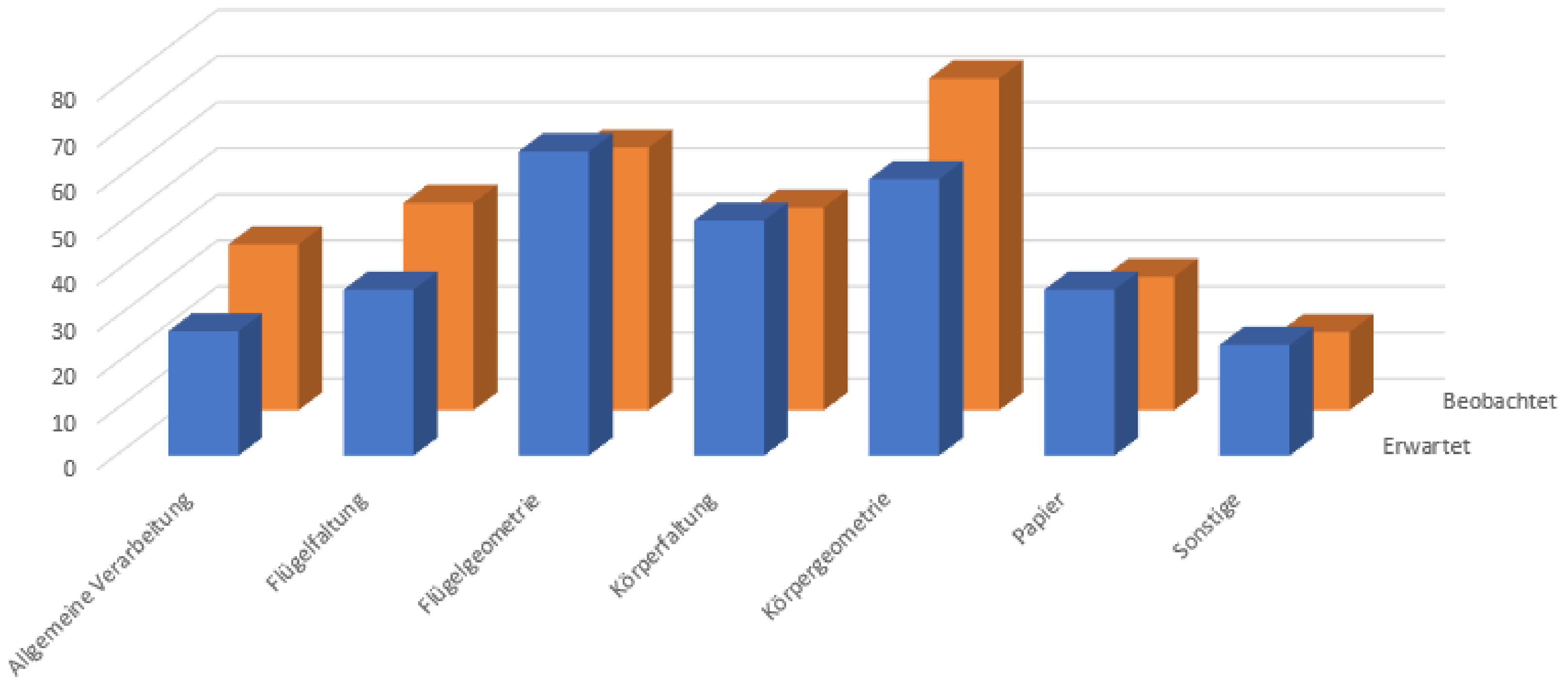
→ **aktuelle Daten weisen große Abweichungen (>1%) von erwarteten Werten aus der Vergangenheit auf**

→ **Bei einigen Fehlerarten liegt eine Verschlechterung vor**

Aufgabe 1:

Wie sieht die aktuelle Fehlerzusammen setzung im Vergleich zur Vergangenheit aus?

Aktuelle Zusammensetzung ggü historischen Werten aus der Vergangenheit



Balkendiagramm zeigt auch, dass bei einigen Fehlern größere Abweichungen ggü Vergangenheit vorliegen

Aufgabe 1:

Wie sieht die aktuelle Fehlerzusammen setzung im Vergleich zur Vergangenheit aus?

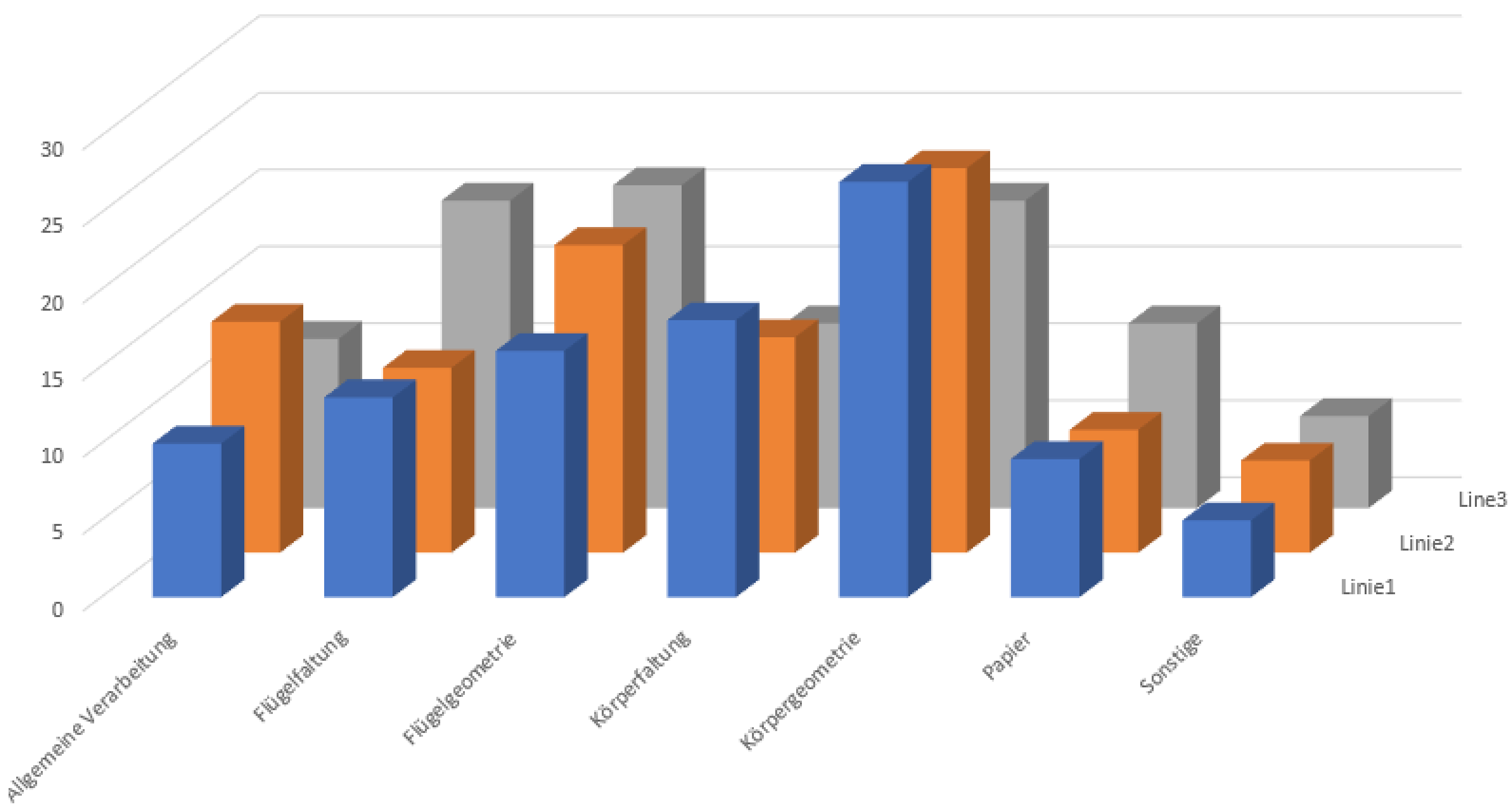
```
> obs<-c(36,45,57,44,72,29,17)
> erw<-c(0.09,0.12,0.22,0.17,0.2,0.12,0.08)
> chisq.test(x=obs,p=erw)

        Chi-squared test for given probabilities

data:  obs
X-squared = 13.241, df = 6, p-value = 0.03937
```

- → **p<5% , H1 gilt**
- → **Die beobachteten Fehlerhäufigkeiten aller Produktionslinien unterscheiden sich signifikant von den Erwartungen. Wir**
- → **müssen die Linien im einzelnen anschauen**

Aufgabe 1:

Wie sieht die aktuelle Fehlerzusammen setzung im Vergleich zur Vergangenheit aus?

Aktuelle Zusammensetzung der einzelnen Linien



Balkendiagramm zeigt unterschiedliche Fehleraufkommen bei einzelnen Linien

Aufgabe 1:

Wie sieht die aktuelle Fehlerzusammen setzung im Vergleich zur Vergangenheit aus?

Ho: Daten einzelner Linien sind von einander unabhängig

H1: Daten einzelner Linien sind von einander abhängig

```
rows                       Linie 1 Linie 2 Linie 3
  Allgemeine Verarbeitung       10      15      11
  Flügelfaltung\n               13      12      20
  Flügelgeometrie\n             16      20      21
  Körperfaltung\n               18      14      12
  Körpergeometrie               27      25      20
  Papier                         9       8      12
  Sonstige                       5       6       6

Rcmdr>  .Test <- chisq.test(.Table, correct=FALSE)

Rcmdr>  .Test

        Pearson's Chi-squared test

data:  .Table
X-squared = 7.7022, df = 12, p-value = 0.8079


Rcmdr>  .Test$expected # Expected Counts
                         columns
rows                         Linie 1   Linie 2 Linie 3
  Allgemeine Verarbeitung 11.760000 12.000000   12.24
  Flügelfaltung\n         14.700000 15.000000   15.30
  Flügelgeometrie\n       18.620000 19.000000   19.38
  Körperfaltung\n         14.373333 14.666667   14.96
  Körpergeometrie         23.520000 24.000000   24.48
  Papier                   9.473333  9.666667    9.86
  Sonstige                 5.553333  5.666667    5.78
```

- **Da p>5% sind die Daten einzelner Linien unabhängig**
- **Wir können wir für die einzelnen Linien die gleichen Erwartungen wie für die Grundgesamtheit aller Fehler nehmen**

Aufgabe 1:

Wie sieht die aktuelle Fehlerzusammen setzung im Vergleich zur Vergangenheit aus?

Fehler	Linie 1	Linie 2	Line 3	Erwartungen aus der Vergangenheit
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%

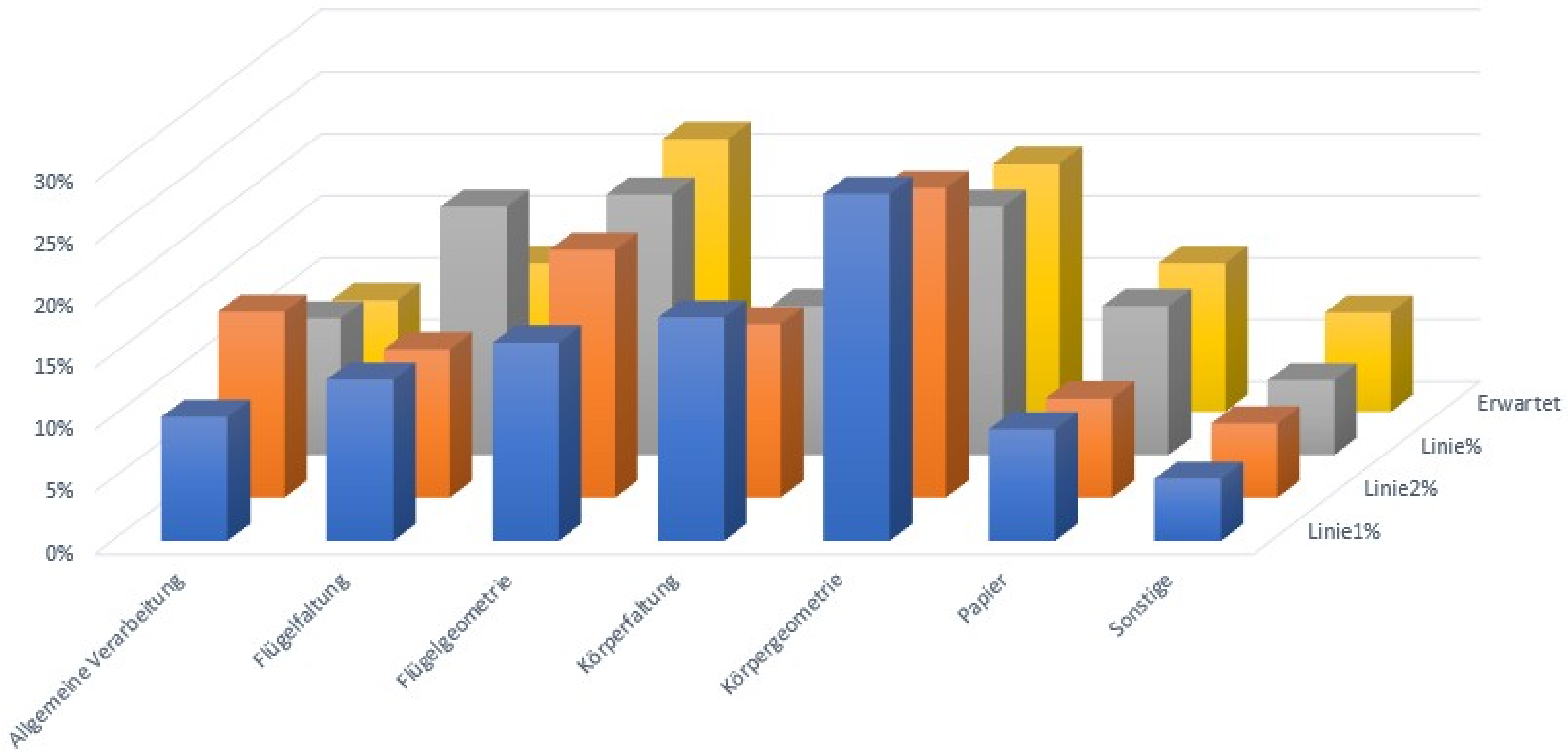
→ **Auch die aktuelle Daten von einzelnen Linien weisen große Abweichungen (>1%) von erwarteten Werten aus der Vergangenheit auf**

→ **Bei einigen Fehlerarten liegt eine Verschlechterung vor**

Aufgabe 1:

Wie sieht die aktuelle Fehlerzusammen setzung im Vergleich zur Vergangenheit aus?

Aktuelle Beobachtungswerte einzelner Linien ggü historischen Werten aus der Vergangenheit



Balkendiagramm zeigt auch, dass bei einigen Fehlern größere Abweichungen ggü Vergangenheit vorliegen

Aufgabe 1:

Wie sieht die aktuelle Fehlerzusammen setzung im Vergleich zur Vergangenheit aus?

H0: Daten folgen dem bekannten Verlauf aus der Vergangenheit
H1: Daten entsprechen nicht der erwarteten Werten

Linie 1

```
> obs<-c(10,13,16,18,27,9,5)
> erw<-c(0.09,0.12,0.22,0.17,0.2,0.12,0.08)
> chisq.test(x=obs,p=erw)

        Chi-squared test for given probabilities

data:  obs
X-squared = 6.3006, df = 6, p-value = 0.3904
```

Linie 2

```
> obs<-c(15,12,20,14,25,8,6)
> erw<-c(0.09,0.12,0.22,0.17,0.2,0.12,0.08)
> chisq.test(x=obs,p=erw)

        Chi-squared test for given probabilities

data:  obs
X-squared = 7.7946, df = 6, p-value = 0.2535
```

Linie 3

```
> obs<-c(11,20,21,12,20,12,6)
> erw<-c(0.09,0.12,0.22,0.17,0.2,0.12,0.08)
> chisq.test(x=obs,p=erw)

        Chi-squared test for given probabilities

data:  obs
X-squared = 7.6018, df = 6, p-value = 0.2688
```

→ **Da bei allen Linien p>5% können wir bei einzelnen Linien keine Unterschiede zu erwarteten Werten feststellen**

Aufgabe 1:

Wo können evtl. die Probleme vorliegen?

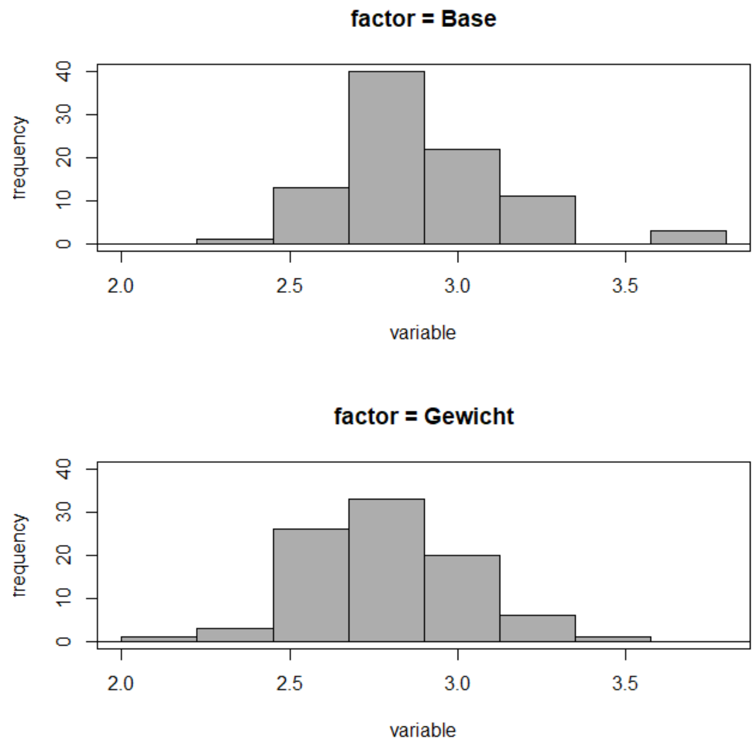
Zusammenfassung

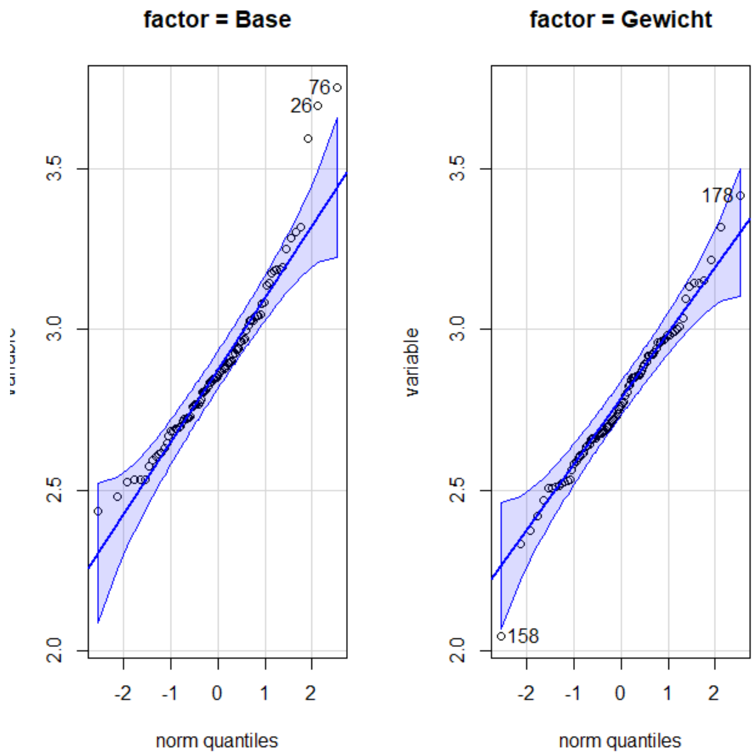
- Es deutet sich an, dass erst durch Summierung das Problem ersichtlich wird
- Daher müssen alle Linien für die folgenden Fehlern angeschaut werden
  - Allgemeine Verarbeitung
  - Flügelfaltung
  - Körpergeometrie

Aufgabe 2:

Lohnt sich die weitere Untersuchung eines Schweren Fliegers?

```
            mean        sd     IQR      0%     25%    50%    75%
Base    2.887516 0.2470161 0.30315 2.4344 2.72125 2.8554 3.0244
Gewicht 2.778744 0.2225149 0.27570 2.0462 2.64570 2.7597 2.9214
          100% variable:n
Base    3.7527         90
Gewicht 3.4169         90
```

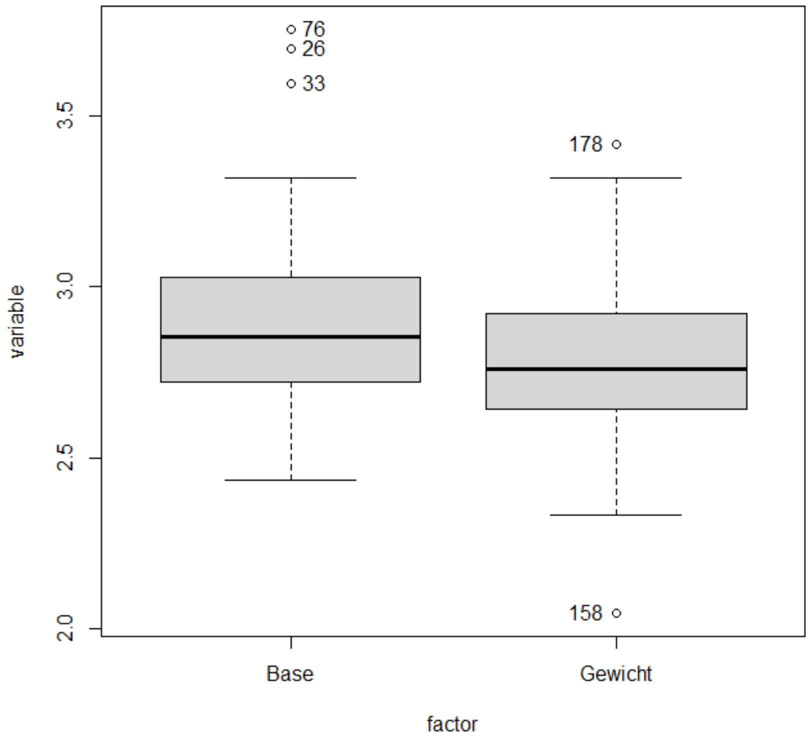




- → Unterschiedliche Mittelwerte und ähnliche Standardabweichungen
- → Bei Histogramm nichts auffälliges bis auf die Lücke bei Base
- → Q-Q Diagramm bei Base liegen einige Werte außerhalb des Bands, Hinweis auf nicht Normalverteilung

Aufgabe 2:

Lohnt sich die weitere Untersuchung eines Schweren Fliegers?



→ **Deutliche Ausreißer bei Base Modell deuten auf keine Normalverteilung, müssen evtl. mit Kunden besprochen werden**

Aufgabe 2:

Lohnt sich die weitere Untersuchung eines Schweren Fliegers?

Shapiro-Wilk normality test

H0: Daten sind normalverteilt
H1: Daten sind nicht normalverteilt

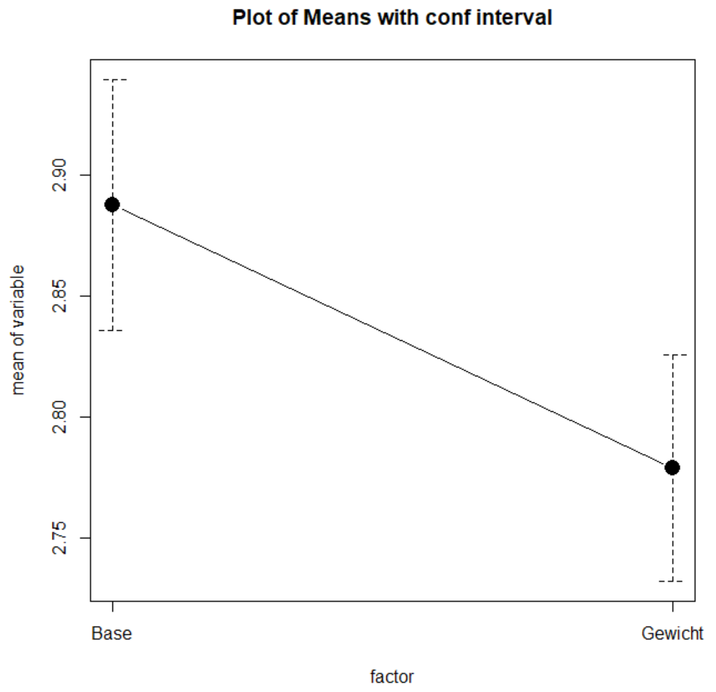
```
 --------
 factor = Base

        Shapiro-Wilk normality test

data:  variable
W = 0.93806, p-value = 0.0003293

 --------
 factor = Gewicht

        Shapiro-Wilk normality test

data:  variable
W = 0.98607, p-value = 0.4539

 --------

 p-values adjusted by the Holm method:
        unadjusted adjusted
Base    0.00032931 0.00065861
Gewicht 0.45390486 0.45390486
```

→ **p< 5%, Nicht parametrisches Verfahren, da keine Normalverteilung**

Aufgabe 2:

Lohnt sich die weitere Untersuchung eines Schweren Fliegers?

```
Levene's Test for Homogeneity of Variance (center = "median")
       Df F value Pr(>F)
group   1  0.0708 0.7904
      178
```



→ **Plot of means deutet an, dass eine evtl. Verschlechterung durch schwerer Gewicht kommt**

Aufgabe 2:

Lohnt sich die weitere Untersuchung eines Schweren Fliegers?

H0: Base $\geq$ Gewicht
H1: Base < Gewicht

```
        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>5%, Wir bleiben in der Nullhypothese, d.h. das zusätzliche Gewicht führt zu einer gleichen oder schlechteren Flugzeit

Aufgabe 2:

Lohnt sich die weitere Untersuchung eines Schweren Fliegers?

Grubbs Test in R zur Identifizierung von Ausreißern

```
> grubbs.test(Dataset2$Base,type = 10, opposite = FALSE, two.sided = FALSE)

        Grubbs test for one outlier

data:  Dataset2$Base
G = 3.50254, U = 0.86061, p-value = 0.01304
alternative hypothesis: highest value 3.7527 is an outlier
```

```
        Grubbs test for one outlier

data:  Dataset2$Base
G = 3.55606, U = 0.85467, p-value = 0.01014
alternative hypothesis: highest value 3.6973 is an outlier
```

→ p<5%, die Werte 3,7527 und 3,6973 als Ausreißer identifiziert

Shapiro-Wilk normality test

Ho: Daten sind normalverteilt
H1: Daten sind nicht normalverteilt

```
 p-values adjusted by the Holm method:
        unadjusted adjusted
Base    0.1771     0.3542
Gewicht 0.4539     0.4539
```

→ p>5%, Normalverteilung liegt vor

Aufgabe 2:

Lohnt sich die weitere Untersuchung eines Schweren Fliegers?

```
        F test to compare two variances

data:  variable by factor
F = 0.92727, num df = 87, denom df = 89, p-value = 0.7247
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6091972 1.4128543
sample estimates:
ratio of variances
         0.9272712
```

→ p>5%, gleiche Varianzen liegen vor

Ho: Base $\geq$ Gewicht
H1: Base < Gewicht

```
        Two Sample t-test

data:  variable by factor
t = 2.7398, df = 176, p-value = 0.9966
alternative hypothesis: true difference in means between group Base and group Gewicht is less than 0
95 percent confidence interval:
     -Inf 0.143897
sample estimates:
   mean in group Base mean in group Gewicht
             2.868482              2.778744
```

p>5%, Wir bleiben in der Nullhypothese, d.h. auch das Rausnehmen der Ausreißern aus Base bestätigt die vorherige Aussage vom nicht parametrischen Verfahren - die weitere Untersuchung des Gewichts lohnt sich nicht

Aufgabe 2:

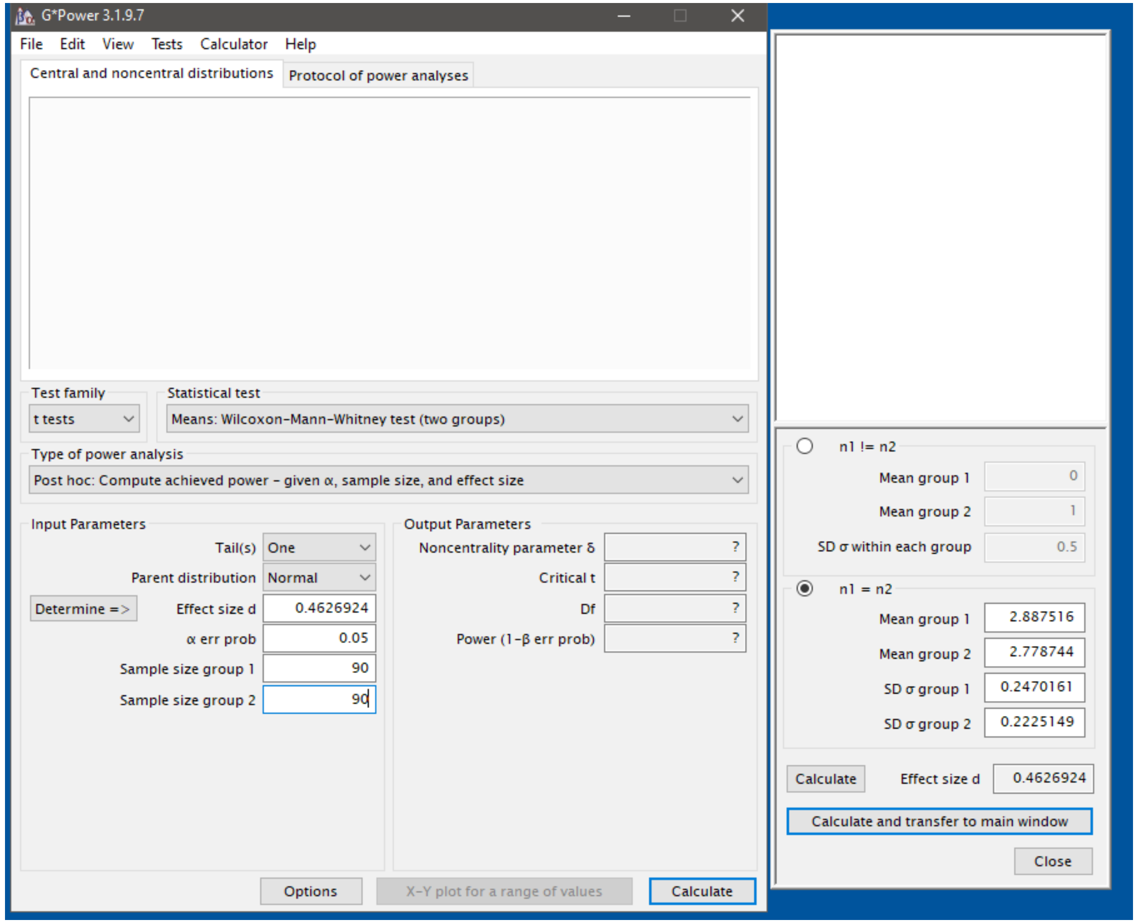
Lohnt sich die weitere Untersuchung eines Schweren Fliegers?

Effektstärke

```
cohensD(x=Dataset2$Base , y=Dataset2$Gewicht)
  0.4626886
```

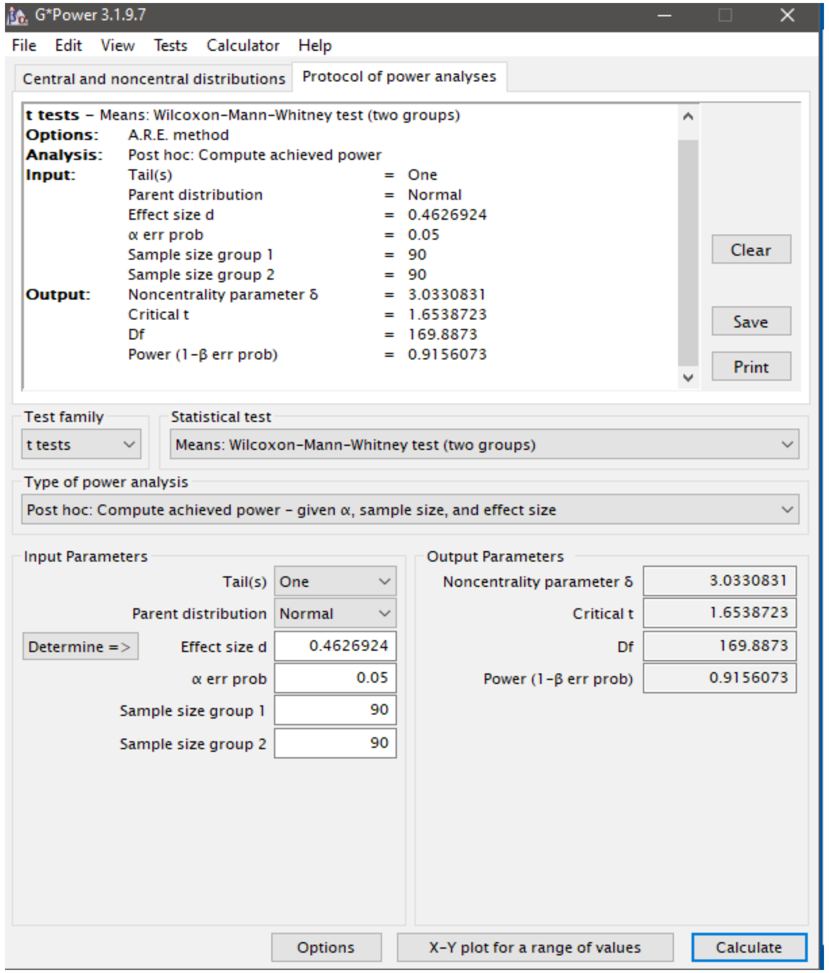
Power Analyse

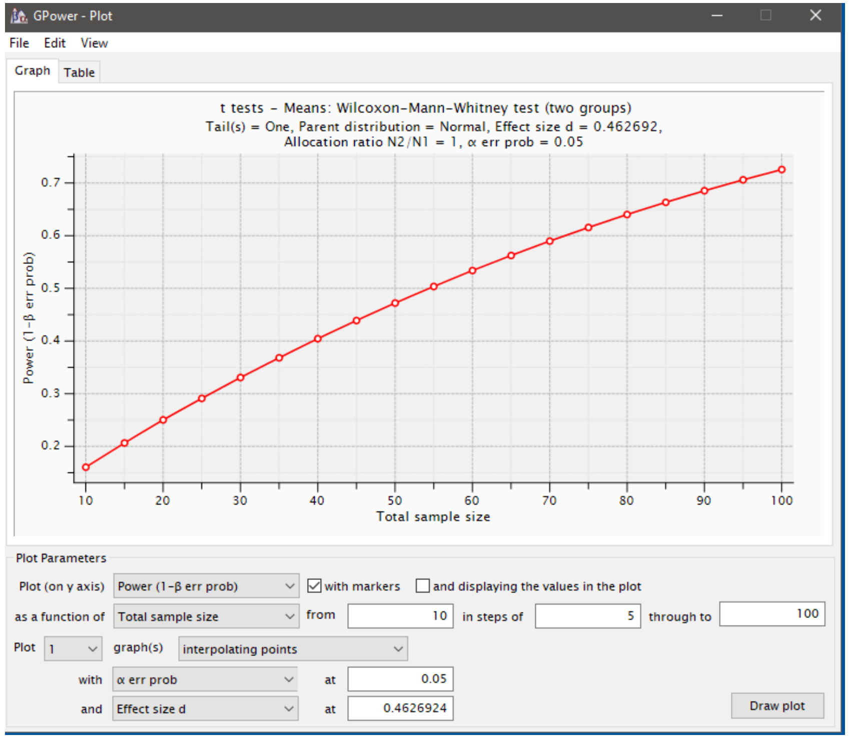
Wilcoxen -Mann-Whitey



Aufgabe 2:

Lohnt sich die weitere Untersuchung eines Schweren Fliegers?



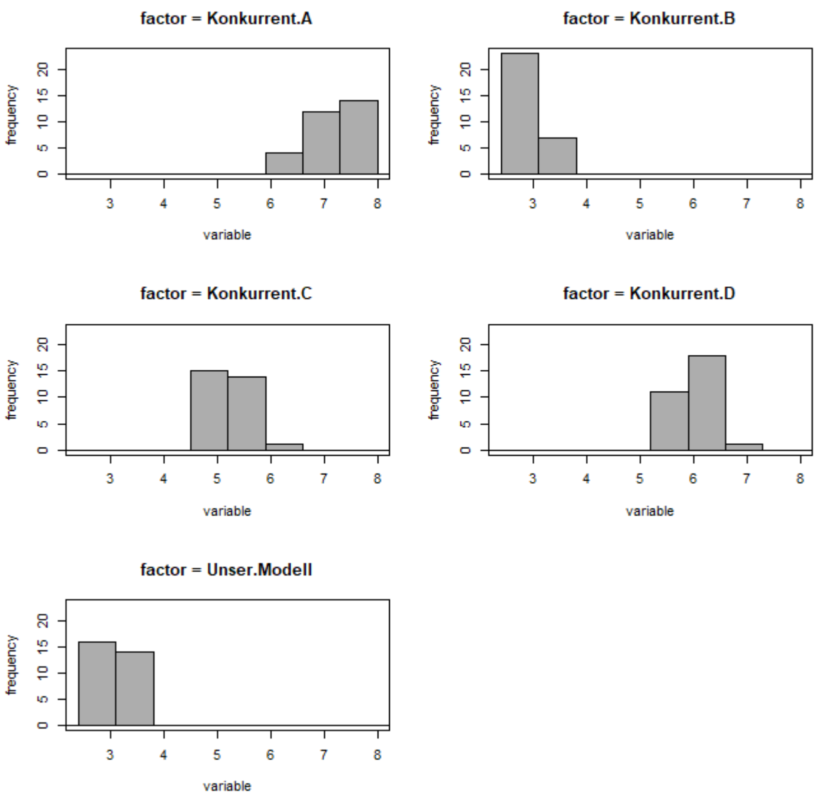


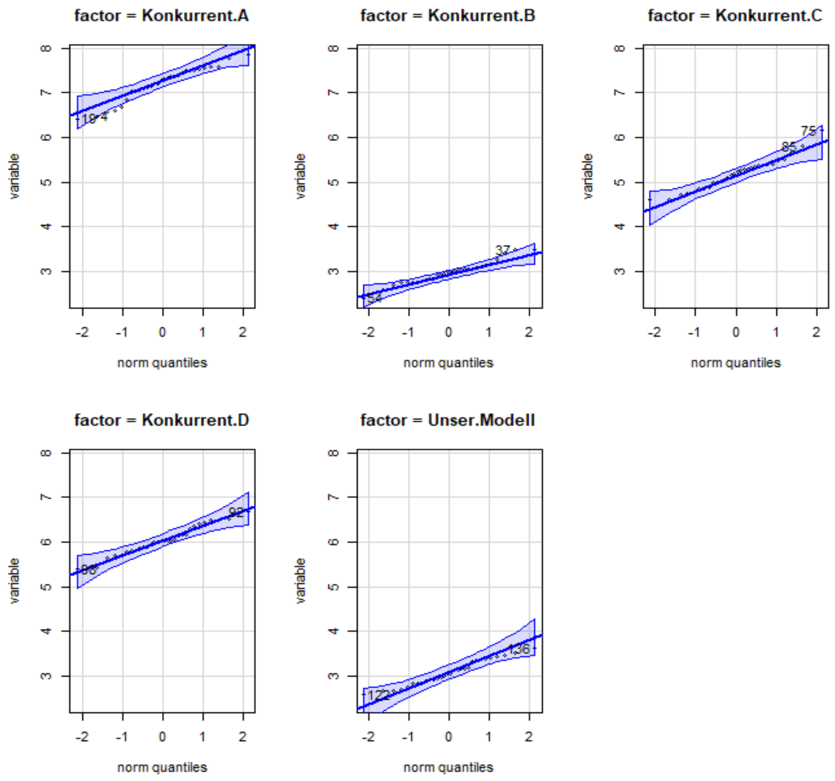
Stichproben von 90 zeigen eine Effektstärke von 0,46 und ein Power von 0,91

Aufgabe 3:

Sind wir bei unserem Flugmodell noch konkurrenzfähig?

	mean	sd	IQR	0%	25%	50%	75%	100%	n
Konkurrent.A	7.210333	0.3757703	0.4550	6.42	7.0450	7.290	7.5000	7.84	30
Konkurrent.B	2.977333	0.2476780	0.2950	2.41	2.7875	3.000	3.0825	3.49	30
Konkurrent.C	5.182667	0.3550613	0.4650	4.60	4.9125	5.205	5.3775	6.16	30
Konkurrent.D	6.037333	0.3217188	0.4475	5.40	5.8175	6.020	6.2650	6.67	30
Unser.Modell	3.081000	0.2832003	0.4900	2.59	2.8550	3.060	3.3450	3.62	30

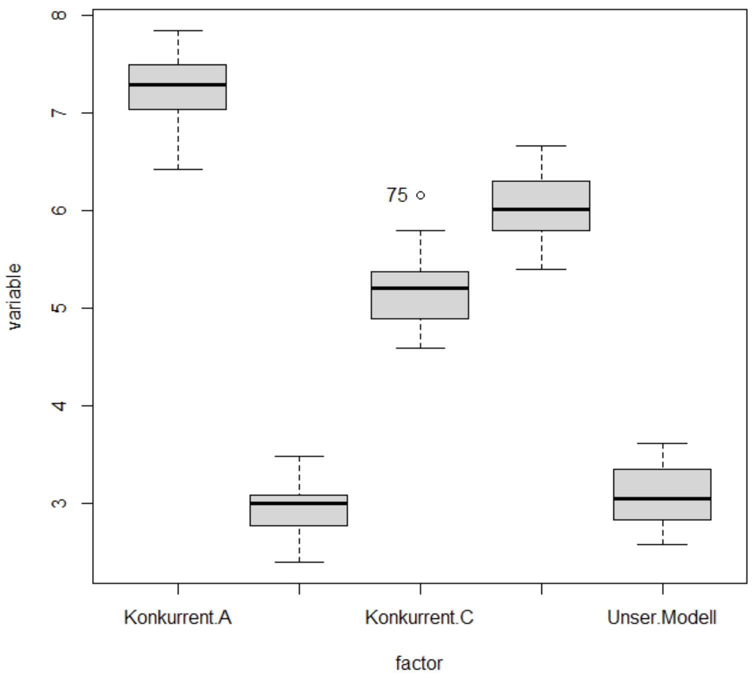


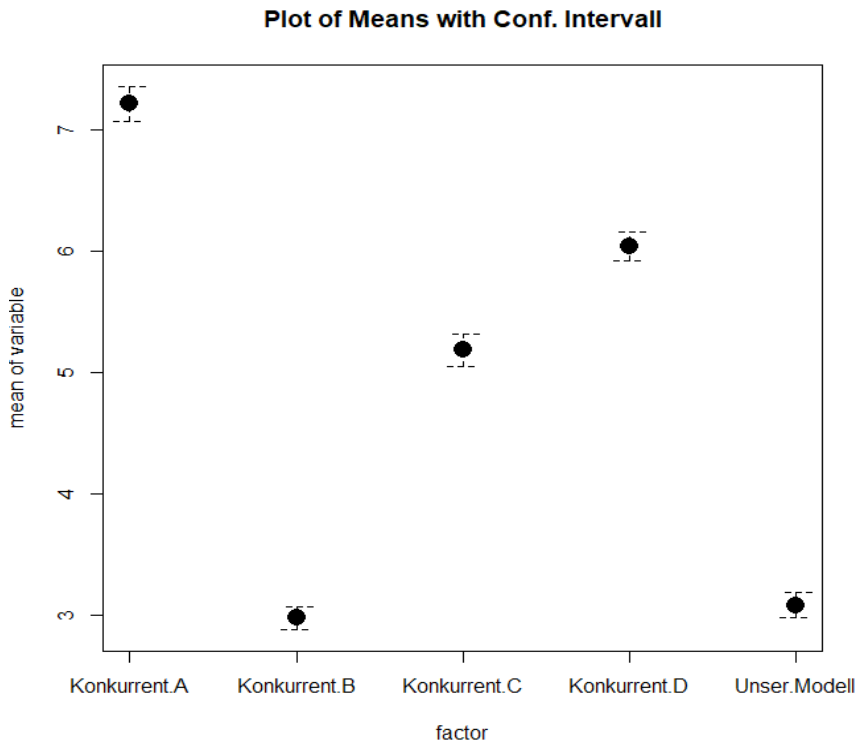


- → Unterschiedliche Mittelwerte und ähnliche Standardabweichungen
- → Histogramme weisen keine Symmetrien auf
- → Q-Q Diagramme nicht auffällig, weisen auf Normalverteilung hin

Aufgabe 3:

Sind wir bei unserem Flugmodell noch konkurrenzfähig?





→ **Keine Symmetrie, Ausreißer bei Konkurrent C**

→ **Erste Betrachtung der Daten weist schon hin, dass es teilweise signifikante Unterschiede zwischen den Konkurrenten zu beobachten sind**

Aufgabe 3:

Sind wir bei unserem Flugmodell noch konkurrenzfähig?

Shapiro-Wilk normality test

Ho: Daten sind normalverteilt
H1: Daten sind nicht normalverteilt

```
 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
```

Bei allen Stichproben ist p>5% → alle Stichproben sind normalverteilt

Ho: gleiche Varianzen
H1: ungleiche Varianzen

```
        Bartlett test of homogeneity of variances

data:  variable by factor
Bartlett's K-squared = 6.29, df = 4, p-value = 0.1785
```

p>5% → gleiche Varianzen

Aufgabe 3:

Sind wir bei unserem Flugmodell noch konkurrenzfähig?

AnovaModel

Ho: alle Mittelwerte sind gleich
H1: mindestens ein Mittelwert unterschiedlich

```
              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

Rcmdr>  with(StackedData3, numSummary(variable, groups=factor, statistics=c("mean", "sd")))
                 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<5%, es gilt eine Alternativhypothese, es gibt mindestens ein Unterschied

→ Paarweise Überprüfung zur Festlegung, wo die Unterschiede genau vorliegen

Aufgabe 3:

Sind wir bei unserem Flugmodell noch konkurrenzfähig?

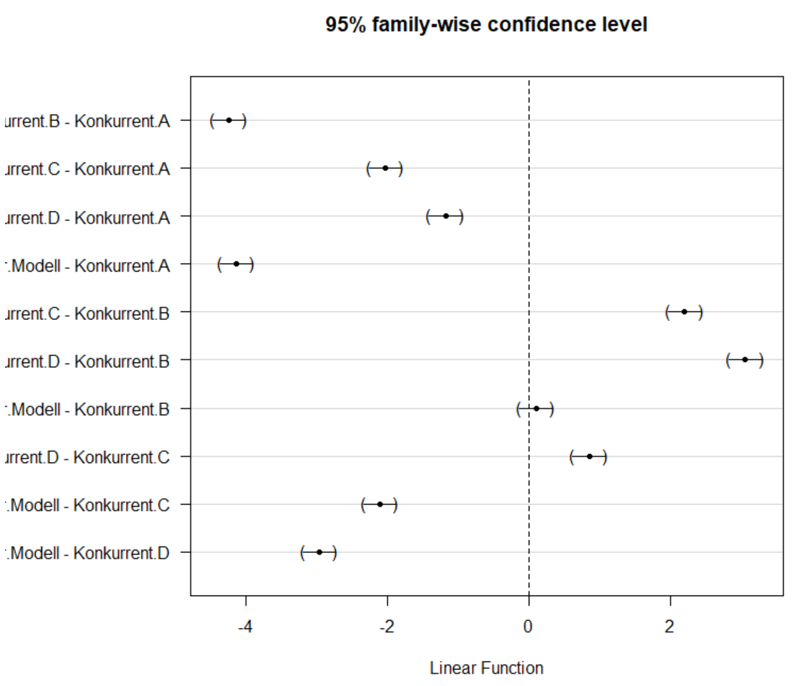
```
         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = variable ~ factor, data = StackedData3)

Linear Hypotheses:
                                  Estimate Std. Error t value Pr(>|t|)    
Konkurrent.B - Konkurrent.A == 0 -4.23300    0.08265 -51.216 <0.00001 ***
Konkurrent.C - Konkurrent.A == 0 -2.02767    0.08265 -24.533 <0.00001 ***
Konkurrent.D - Konkurrent.A == 0 -1.17300    0.08265 -14.192 <0.00001 ***
Unser.Modell - Konkurrent.A == 0 -4.12933    0.08265 -49.962 <0.00001 ***
Konkurrent.C - Konkurrent.B == 0  2.20533    0.08265  26.683 <0.00001 ***
Konkurrent.D - Konkurrent.B == 0  3.06000    0.08265  37.024 <0.00001 ***
Unser.Modell - Konkurrent.B == 0  0.10367    0.08265   1.254    0.719    
Konkurrent.D - Konkurrent.C == 0  0.85467    0.08265  10.341 <0.00001 ***
Unser.Modell - Konkurrent.C == 0 -2.10167    0.08265 -25.429 <0.00001 ***
Unser.Modell - Konkurrent.D == 0 -2.95633    0.08265 -35.769 <0.00001 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

<u>Ergebnis</u>: Ranking: A, D, C, B/Unser Modell.

Aufgabe 3:

Sind wir bei unserem Flugmodell noch konkurrenzfähig?

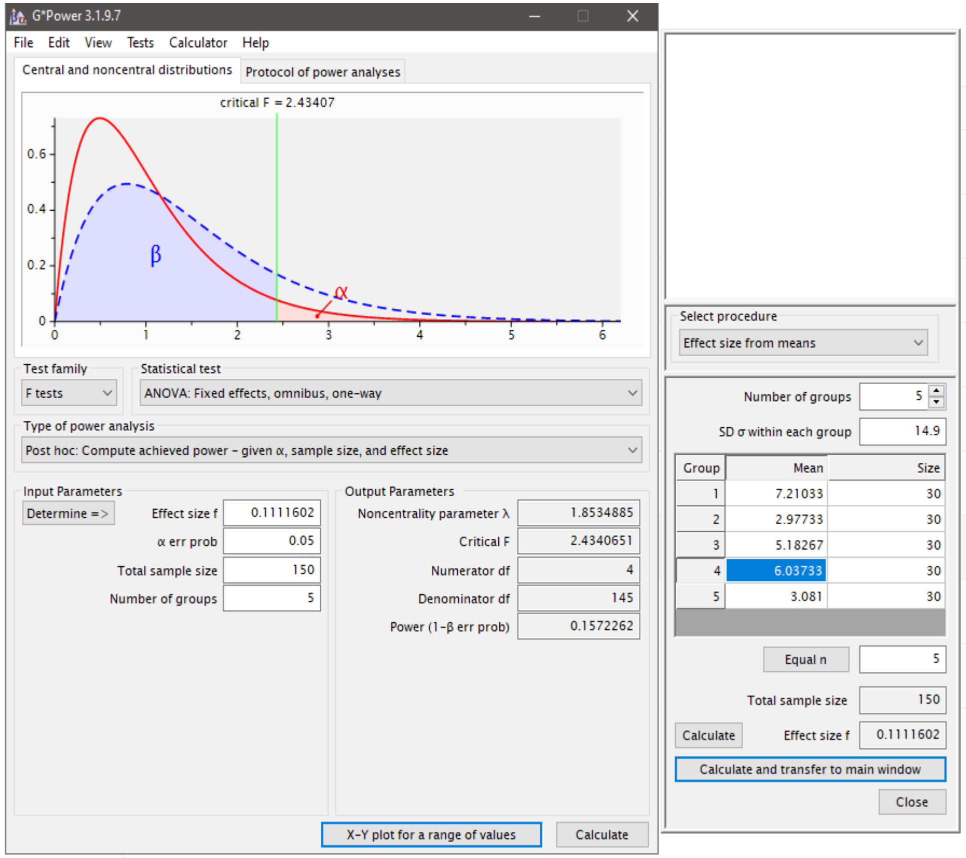
```
         Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = variable ~ factor, data = StackedData3)

Quantile = 2.7623
95% family-wise confidence level


Linear Hypotheses:
                                 Estimate lwr     upr    
Konkurrent.B - Konkurrent.A == 0 -4.2330  -4.4613 -4.0047
Konkurrent.C - Konkurrent.A == 0 -2.0277  -2.2560 -1.7994
Konkurrent.D - Konkurrent.A == 0 -1.1730  -1.4013 -0.9447
Unser.Modell - Konkurrent.A == 0 -4.1293  -4.3576 -3.9010
Konkurrent.C - Konkurrent.B == 0  2.2053   1.9770  2.4336
Konkurrent.D - Konkurrent.B == 0  3.0600   2.8317  3.2883
Unser.Modell - Konkurrent.B == 0  0.1037  -0.1246  0.3320
Konkurrent.D - Konkurrent.C == 0  0.8547   0.6264  1.0830
Unser.Modell - Konkurrent.C == 0 -2.1017  -2.3300 -1.8734
Unser.Modell - Konkurrent.D == 0 -2.9563  -3.1846 -2.7280

Konkurrent.A Konkurrent.B Konkurrent.C Konkurrent.D Unser.Modell 
         "a"          "b"          "c"          "d"          "b" 
```

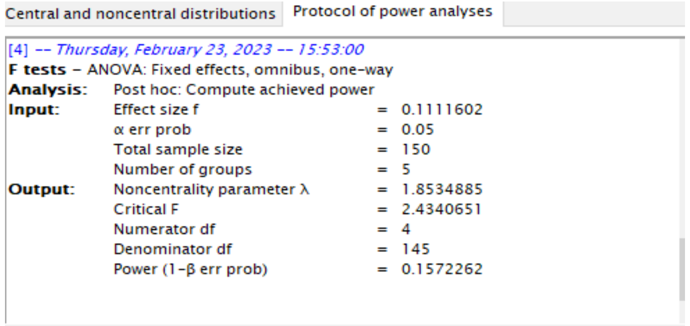


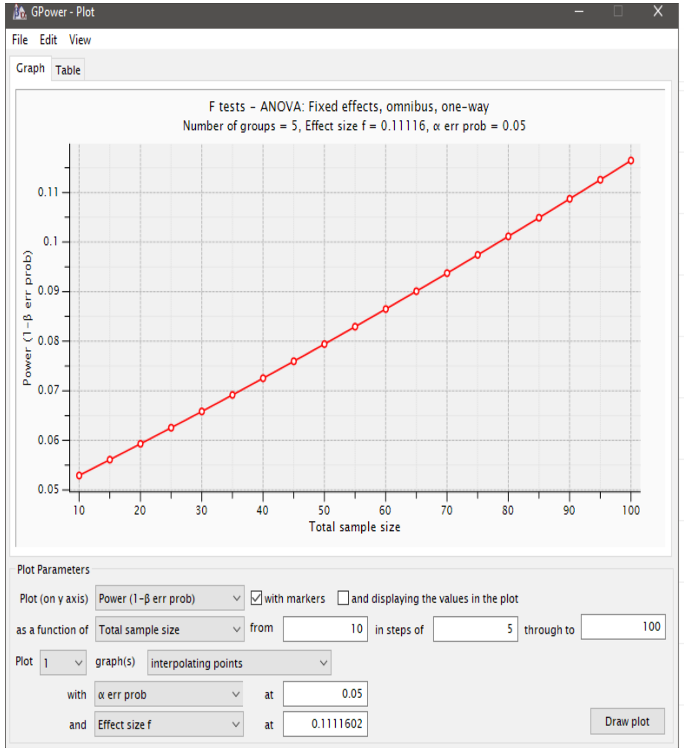
<u>Antwort</u>: Unser Flugmodell ist nicht mehr konkurrenzfähig und muss dringend verbessert werden.

Aufgabe 3:

Sind wir bei unserem Flugmodell noch konkurrenzfähig?



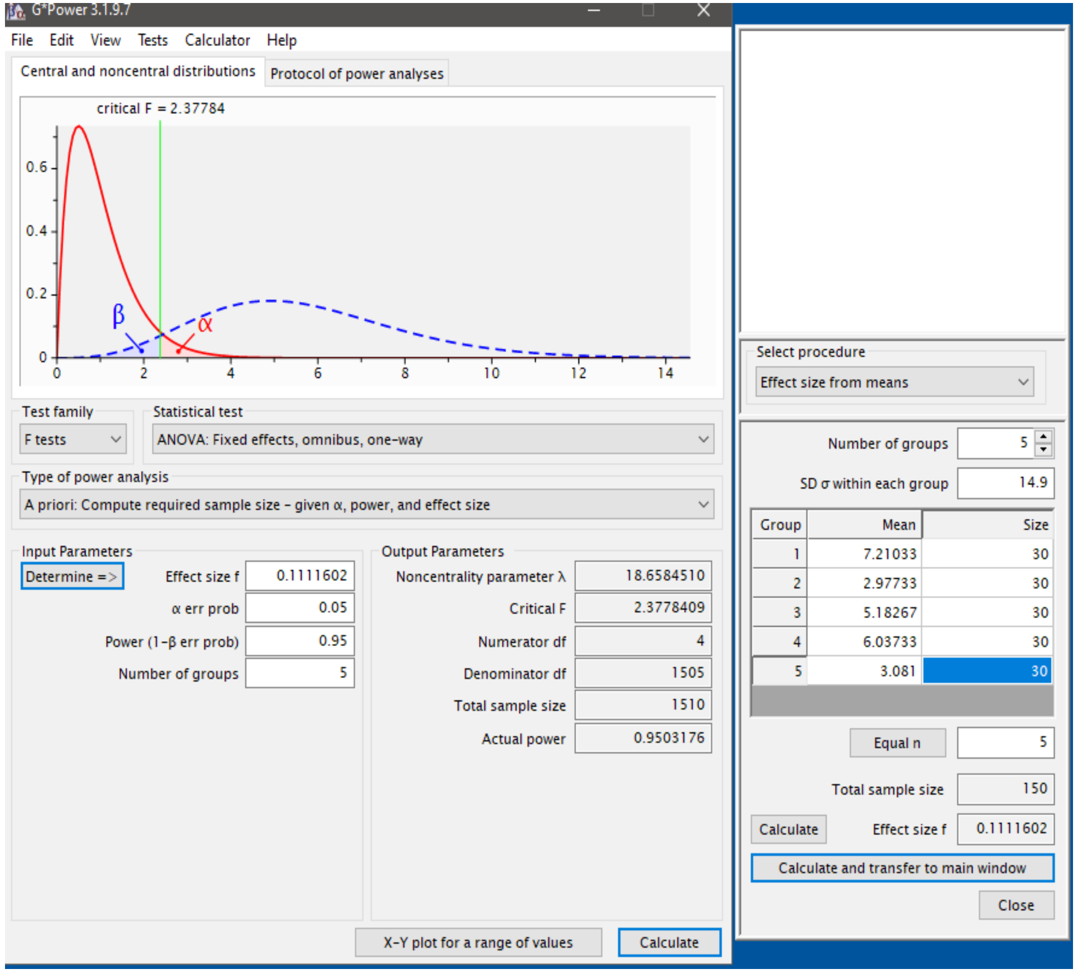


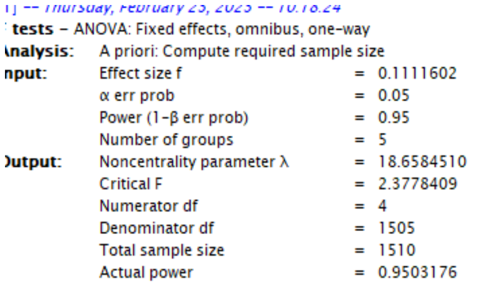


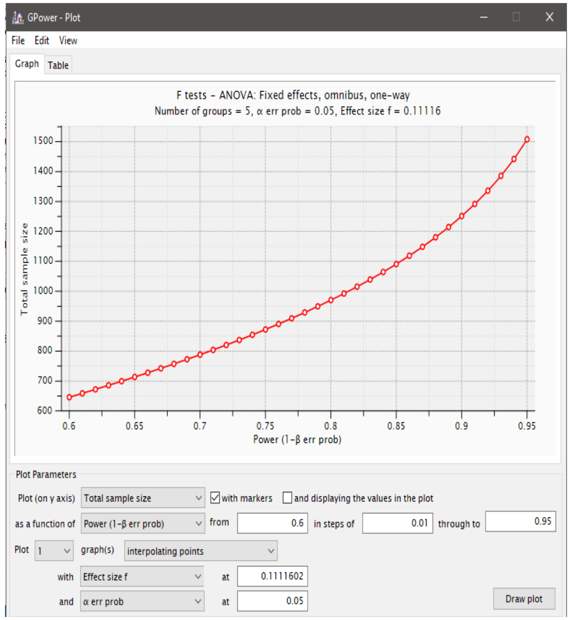
Mit einer Stichprobengröße von jeweils 30 und einer Effektstärke von 0,11 wird Teststärke von 15,7% erreicht

Aufgabe 3:

Sind wir bei unserem Flugmodell noch konkurrenzfähig?



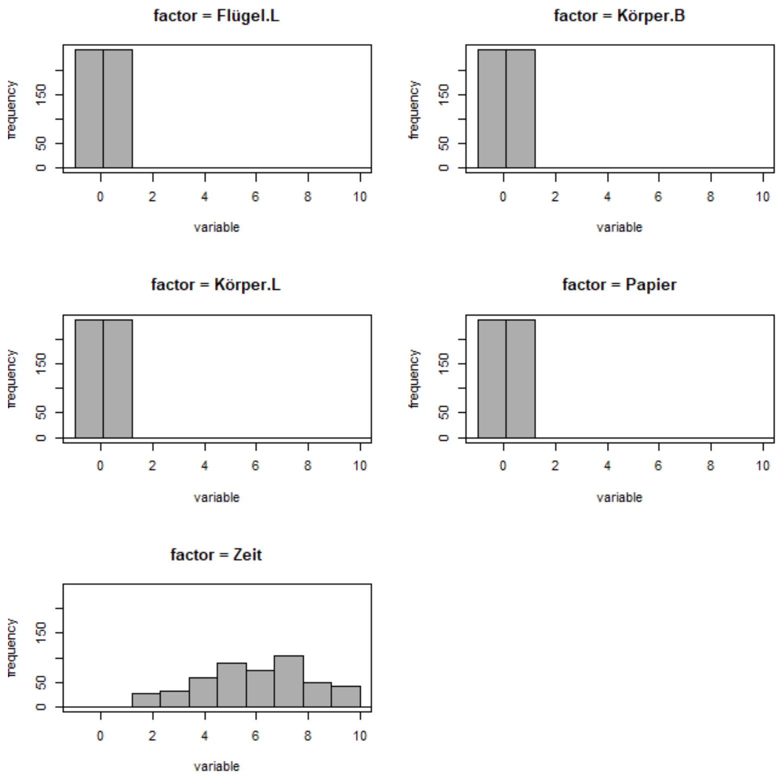


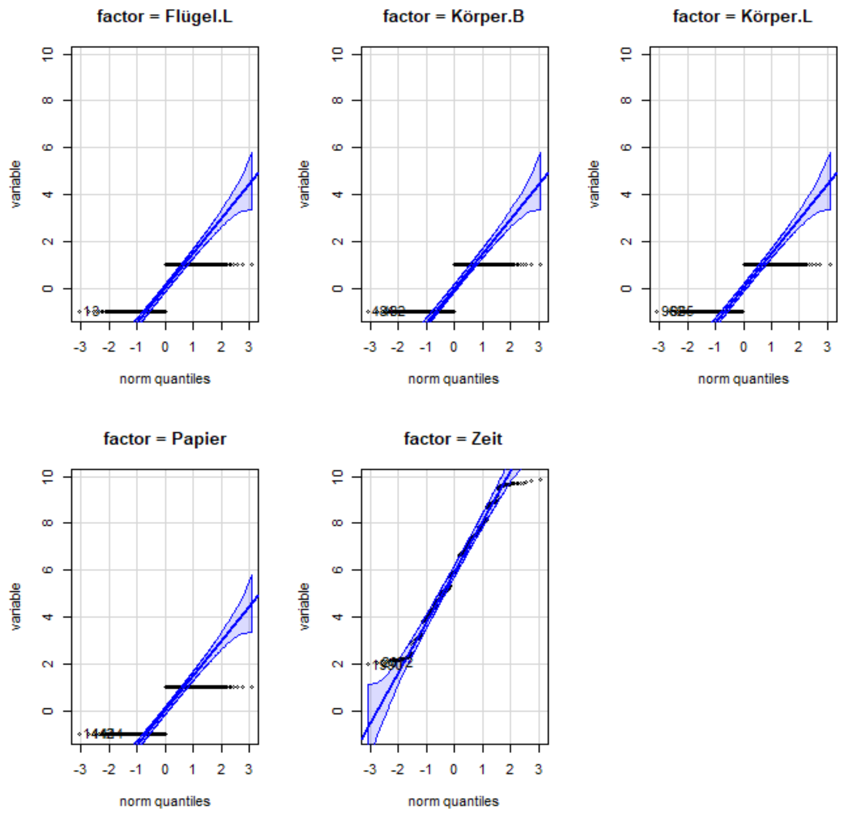


Mit einer Effektstärke von 0,11 benötigen eine Stichprobengröße von jeweils 302, um die Teststärke von 95% zu erreichen

Aufgabe 4:

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

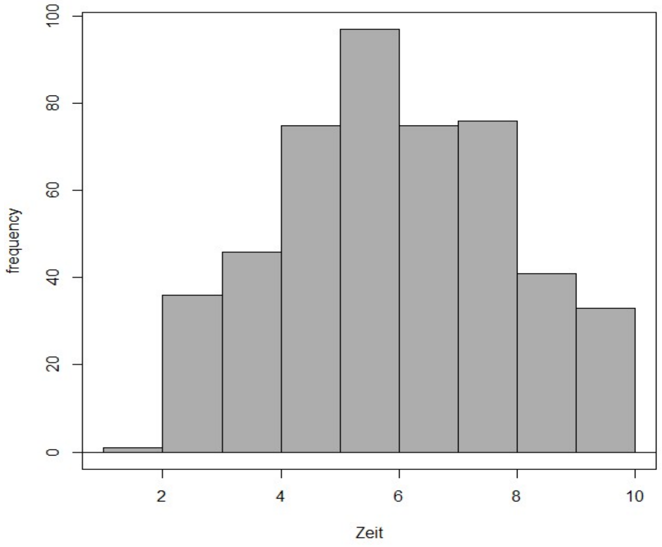


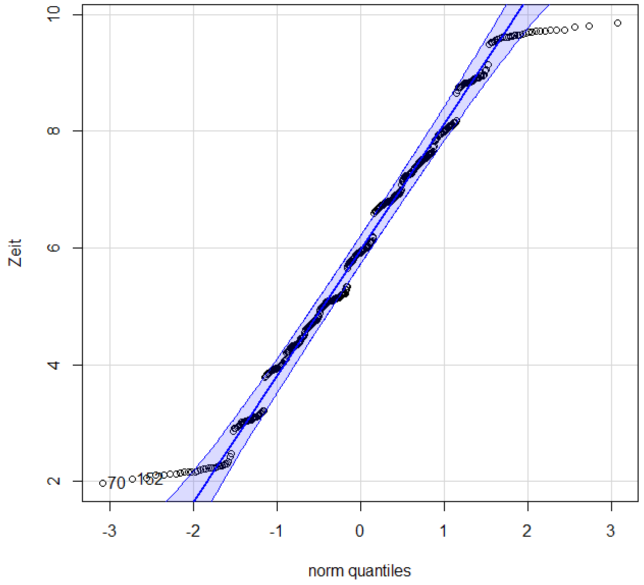


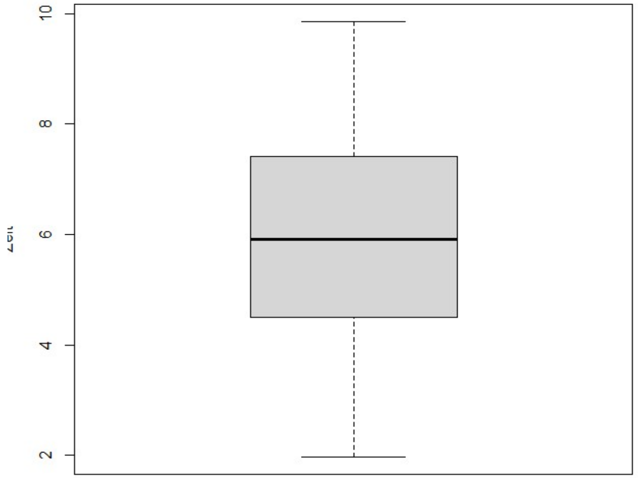
→ **Da die Daten nicht Intervall skaliert sind (außer Zeit) sind die Histogramme und Q-Q Diagramme nicht so aussagekräftig, die weisen nur darauf hin, dass es um Faktoren geht**

Aufgabe 4:

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







→ **Histogramm symmetrisch und normalverteilt**

→ **Q-Q Diagramme weist hier auch auf die Faktoren hin und ein Paar Werte liegen außerhalb des Bands**

→ **Boxplot sieht normal und symmetrisch aus**

Aufgabe 4:

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

Prüfung der Normalverteilung mit Shapiro-Wilk normality test

Ho: Daten sind normalverteilt
H1: Daten sind nicht normalverteilt

```
 p-values adjusted by the Holm method:
                         unadjusted adjusted
130 mm 20 mm 130 mm 80 g 0.55958    1
130 mm 20 mm 130 mm 90 g 0.56399    1
130 mm 20 mm 80 mm 80 g  0.68726    1
130 mm 20 mm 80 mm 90 g  0.85775    1
130 mm 35 mm 130 mm 80 g 0.67137    1
130 mm 35 mm 130 mm 90 g 0.59614    1
130 mm 35 mm 80 mm 80 g  0.42902    1
130 mm 35 mm 80 mm 90 g  0.87351    1
80 mm 20 mm 130 mm 80 g  0.73391    1
80 mm 20 mm 130 mm 90 g  0.73805    1
80 mm 20 mm 80 mm 80 g   0.73938    1
80 mm 20 mm 80 mm 90 g   0.67322    1
80 mm 35 mm 130 mm 80 g  0.29860    1
80 mm 35 mm 130 mm 90 g  0.91805    1
80 mm 35 mm 80 mm 80 g   0.18676    1
80 mm 35 mm 80 mm 90 g   0.69987    1
```

Alle p-Werte > 5% → Alle Stichproben sind normalverteilt

Aufgabe 4:

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

Prüfung auf Homogenität der Varianz

H0: gleiche Varianzen
H1: ungleiche Varianzen

```
        Bartlett test of homogeneity of variances

data:  Zeit by variable
Bartlett's K-squared = 5.6717, df = 15, p-value = 0.9848
```

Alle p-Werte > 5% → Die Varianzen der einzelnen Stichproben können als gleich angesehen werden

Aufgabe 4:

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

```
Call:
lm(formula = Zeit ~ (Flügel.L + Körper.B + Körper.L + Papier)^4,
    data = Dataset4)

Residuals:
      Min        1Q    Median        3Q       Max
-0.277000 -0.059500 -0.005667  0.063500  0.275333

Coefficients:
                                 Estimate Std. Error  t value Pr(>|t|)
(Intercept)                     5.9520208  0.0044487 1337.918   <2e-16 ***
Flügel.L                        1.4343958  0.0044487  322.429   <2e-16 ***
Körper.B                        1.0509792  0.0044487  236.243   <2e-16 ***
Körper.L                       -0.3861458  0.0044487  -86.799   <2e-16 ***
Papier                          0.8316875  0.0044487  186.950   <2e-16 ***
Flügel.L:Körper.B               0.0085208  0.0044487    1.915   0.0561 .
Flügel.L:Körper.L               0.0083958  0.0044487    1.887   0.0598 .
Flügel.L:Papier                -0.0062708  0.0044487   -1.410   0.1593
Körper.B:Körper.L               0.0005625  0.0044487    0.126   0.8994
Körper.B:Papier                -0.0088542  0.0044487   -1.990   0.0471 *
Körper.L:Papier                -0.0006458  0.0044487   -0.145   0.8846
Flügel.L:Körper.B:Körper.L     -0.0063958  0.0044487   -1.438   0.1512
Flügel.L:Körper.B:Papier       -0.0046458  0.0044487   -1.044   0.2969
Flügel.L:Körper.L:Papier       -0.0082708  0.0044487   -1.859   0.0636 .
Körper.B:Körper.L:Papier       -0.0076042  0.0044487   -1.709   0.0881 .
Flügel.L:Körper.B:Körper.L:Papier  0.0069375  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
```

- **Wir fangen von unten an mit 4fach-Wechselwirkungen, und, nehmen schrittweise Wechselwirkungen aus, die nicht signifikant zur Verbesserung des Modells beitragen (p>5%)**
- **Die Werte, die nah an der Grenze liegen behalten wir zuerst**

Aufgabe 4:

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

```
Call:
lm(formula = Zeit ~ (Flügel.L + Körper.B + Körper.L + Papier)^3,
    data = Dataset4)

Residuals:
      Min        1Q    Median        3Q       Max
-0.283937 -0.060062 -0.005104  0.062719  0.282271

Coefficients:
                           Estimate Std. Error  t value Pr(>|t|)
(Intercept)               5.9520208  0.0044556 1335.862   <2e-16 ***
Flügel.L                  1.4343958  0.0044556  321.934   <2e-16 ***
Körper.B                  1.0509792  0.0044556  235.880   <2e-16 ***
Körper.L                 -0.3861458  0.0044556  -86.666   <2e-16 ***
Papier                    0.8316875  0.0044556  186.663   <2e-16 ***
Flügel.L:Körper.B         0.0085208  0.0044556    1.912   0.0564 .
Flügel.L:Körper.L         0.0083958  0.0044556    1.884   0.0601 .
Flügel.L:Papier          -0.0062708  0.0044556   -1.407   0.1600
Körper.B:Körper.L         0.0005625  0.0044556    0.126   0.8996
Körper.B:Papier          -0.0088542  0.0044556   -1.987   0.0475 *
Körper.L:Papier          -0.0006458  0.0044556   -0.145   0.8848
Flügel.L:Körper.B:Körper.L -0.0063958  0.0044556   -1.435   0.1518
Flügel.L:Körper.B:Papier -0.0046458  0.0044556   -1.043   0.2976
Flügel.L:Körper.L:Papier -0.0082708  0.0044556   -1.856   0.0640 .
Körper.B:Körper.L:Papier -0.0076042  0.0044556   -1.707   0.0886 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09762 on 465 degrees of freedom
Multiple R-squared:  0.9977,    Adjusted R-squared:  0.9976
F-statistic: 1.44e+04 on 14 and 465 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Zeit ~ Papier + Körper.L + Körper.B + Flügel.L +
    (Papier + Körper.L + Körper.B + Flügel.L)^2 + Flügel.L:Körper.L:Papier +
    Körper.B:Körper.L:Papier, data = Dataset4)

Residuals:
      Min        1Q    Median        3Q       Max
-0.294979 -0.061073 -0.003146  0.067896  0.271229

Coefficients:
                          Estimate Std. Error  t value Pr(>|t|)
(Intercept)              5.9520208  0.0044610 1334.224   <2e-16 ***
Papier                   0.8316875  0.0044610  186.434   <2e-16 ***
Körper.L                -0.3861458  0.0044610  -86.560   <2e-16 ***
Körper.B                 1.0509792  0.0044610  235.591   <2e-16 ***
Flügel.L                 1.4343958  0.0044610  321.539   <2e-16 ***
Papier:Körper.L         -0.0006458  0.0044610   -0.145   0.8850
Papier:Körper.B         -0.0088542  0.0044610   -1.985   0.0478 *
Papier:Flügel.L         -0.0062708  0.0044610   -1.406   0.1605
Körper.L:Körper.B        0.0005625  0.0044610    0.126   0.8997
Körper.L:Flügel.L        0.0083958  0.0044610    1.882   0.0605 .
Körper.B:Flügel.L        0.0085208  0.0044610    1.910   0.0567 .
Papier:Körper.L:Flügel.L -0.0082708  0.0044610   -1.854   0.0644 .
Papier:Körper.L:Körper.B -0.0076042  0.0044610   -1.705   0.0889 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09774 on 467 degrees of freedom
Multiple R-squared:  0.9977,    Adjusted R-squared:  0.9976
F-statistic: 1.676e+04 on 12 and 467 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Zeit ~ Papier + Körper.L + Körper.B + Flügel.L +
    Papier:Körper.B + Körper.L:Flügel.L + Körper.B:Flügel.L +
    Flügel.L:Körper.L:Papier + Körper.B:Körper.L:Papier,
    data = Dataset4)

Residuals:
      Min        1Q    Median        3Q       Max
-0.302458 -0.061521 -0.004667  0.065625  0.264875

Coefficients:
                          Estimate Std. Error  t value Pr(>|t|)
(Intercept)               5.952021   0.004456 1335.627   <2e-16 ***
Papier                    0.831688   0.004456  186.630   <2e-16 ***
Körper.L                 -0.386146   0.004456  -86.651   <2e-16 ***
Körper.B                  1.050979   0.004456  235.839   <2e-16 ***
Flügel.L                  1.434396   0.004456  321.877   <2e-16 ***
Papier:Körper.B          -0.008854   0.004456   -1.987   0.0475 *
Körper.L:Flügel.L         0.008396   0.004456    1.884   0.0602 .
Körper.B:Flügel.L         0.008521   0.004456    1.912   0.0565 .
Papier:Körper.L:Flügel.L -0.008271   0.004456   -1.856   0.0641 .
Papier:Körper.L:Körper.B -0.007604   0.004456   -1.706   0.0886 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09763 on 470 degrees of freedom
Multiple R-squared:  0.9977,    Adjusted R-squared:  0.9976
F-statistic: 2.24e+04 on 9 and 470 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Zeit ~ Papier + Körper.L + Körper.B + Flügel.L +
    Papier:Körper.B + Körper.L:Flügel.L + Körper.B:Flügel.L,
    data = Dataset4)

Residuals:
      Min        1Q    Median        3Q       Max
-0.301792 -0.065240 -0.002708  0.064792  0.249000

Coefficients:
                   Estimate Std. Error  t value Pr(>|t|)
(Intercept)        5.952021   0.004477 1329.506   <2e-16 ***
Papier             0.831688   0.004477  185.774   <2e-16 ***
Körper.L          -0.386146   0.004477  -86.254   <2e-16 ***
Körper.B           1.050979   0.004477  234.758   <2e-16 ***
Flügel.L           1.434396   0.004477  320.402   <2e-16 ***
Papier:Körper.B   -0.008854   0.004477   -1.978   0.0485 *
Körper.L:Flügel.L  0.008396   0.004477    1.875   0.0614 .
Körper.B:Flügel.L  0.008521   0.004477    1.903   0.0576 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09808 on 472 degrees of freedom
Multiple R-squared:  0.9976,    Adjusted R-squared:  0.9976
F-statistic: 2.853e+04 on 7 and 472 DF,  p-value: < 2.2e-16
```

Aufgabe 4:

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

```
Call:
lm(formula = Zeit ~ Papier + Körper.L + Körper.B + Flügel.L +
    Papier:Körper.B + Körper.B:Flügel.L, data = Dataset4)

Residuals:
      Min        1Q    Median        3Q       Max
-0.293396 -0.068625 -0.003396  0.063698  0.257396

Coefficients:
                   Estimate Std. Error  t value Pr(>|t|)
(Intercept)        5.952021   0.004489 1325.983   <2e-16 ***
Papier             0.831688   0.004489  185.282   <2e-16 ***
Körper.L          -0.386146   0.004489  -86.025   <2e-16 ***
Körper.B           1.050979   0.004489  234.136   <2e-16 ***
Flügel.L           1.434396   0.004489  319.553   <2e-16 ***
Papier:Körper.B   -0.008854   0.004489   -1.973   0.0491 *
Körper.B:Flügel.L  0.008521   0.004489    1.898   0.0583 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09834 on 473 degrees of freedom
Multiple R-squared:  0.9976,	Adjusted R-squared:  0.9976
F-statistic: 3.311e+04 on 6 and 473 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Zeit ~ Papier + Körper.L + Körper.B + Flügel.L,
    data = Dataset4)

Residuals:
      Min        1Q    Median        3Q       Max
-0.293063 -0.060854 -0.001479  0.067062  0.257063

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.952021   0.004515 1318.38   <2e-16 ***
Papier       0.831688   0.004515  184.22   <2e-16 ***
Körper.L    -0.386146   0.004515  -85.53   <2e-16 ***
Körper.B     1.050979   0.004515  232.79   <2e-16 ***
Flügel.L     1.434396   0.004515  317.72   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09891 on 475 degrees of freedom
Multiple R-squared:  0.9976,	Adjusted R-squared:  0.9976
F-statistic: 4.91e+04 on 4 and 475 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Zeit ~ Papier + Körper.L + Körper.B + Flügel.L +
    Papier:Körper.B, data = Dataset4)

Residuals:
      Min        1Q    Median        3Q       Max
-0.301917 -0.063333 -0.002187  0.064385  0.265917

Coefficients:
                 Estimate Std. Error  t value Pr(>|t|)
(Intercept)      5.952021   0.004501 1322.356   <2e-16 ***
Papier           0.831688   0.004501  184.775   <2e-16 ***
Körper.L        -0.386146   0.004501  -85.790   <2e-16 ***
Körper.B         1.050979   0.004501  233.495   <2e-16 ***
Flügel.L         1.434396   0.004501  318.679   <2e-16 ***
Papier:Körper.B -0.008854   0.004501   -1.967   0.0498 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09861 on 474 degrees of freedom
Multiple R-squared:  0.9976,	Adjusted R-squared:  0.9976
F-statistic: 3.952e+04 on 5 and 474 DF,  p-value: < 2.2e-16
```

Das Model mit Hauptfaktoren ohne Wechselwirkungen erweist sich als bestes Modell

Aufgabe 4:

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

```
Call:
lm(formula = Zeit ~ Papier + Körper.L + Körper.B + Flügel.L,
    data = Dataset4)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.293063 -0.060854 -0.001479  0.067062  0.257063 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.952021   0.004515 1318.38   <2e-16 ***
Papier       0.831688   0.004515  184.22   <2e-16 ***
Körper.L    -0.386146   0.004515  -85.53   <2e-16 ***
Körper.B     1.050979   0.004515  232.79   <2e-16 ***
Flügel.L     1.434396   0.004515  317.72   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09891 on 475 degrees of freedom
Multiple R-squared:  0.9976,    Adjusted R-squared:  0.9976 
F-statistic: 4.91e+04 on 4 and 475 DF,  p-value: < 2.2e-16
```

FZ = 5,95 + 0,83 $*$ Papier – 0,39 $*$ Körper.L + 1,05 $*$ Körper.B + 1,43 $*$ Flügel.L
(?? evtl. – 0,01*Papier:KörperB ??)

- **Das Model mit Hauptfaktoren ohne Wechselwirkungen erweist sich als bestes Modell**
- **Die Wechselwirkung Papier:Körper.B ist zwar signifikant, aber schwach (steuert kaum zusätzliche FZ bei)**
- **Durch Herausnahme dieser letzten Wechselwirkung erreichen wir aber Verbesserung des F-Wertes und unseres Model wird einfacher und robuster**

Aufgabe 4:

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

FZ = 5,95 + 0,83 ∗ Papier – 0,39 ∗ Körper.L + 1,05 ∗ Körper.B + 1,43 ∗ Flügel.L

Optimale Einstellung:

- Papier +1
- Körper.L -1
- Körper.B +1
- Flügel.L +1

Maximale Flugzeit:

FZ = 5,95 + 0,83 + 0,39 + 1,05 + 1,43 = **9,646**

Aufgabe 4:

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

Konkurrenzfähigkeit

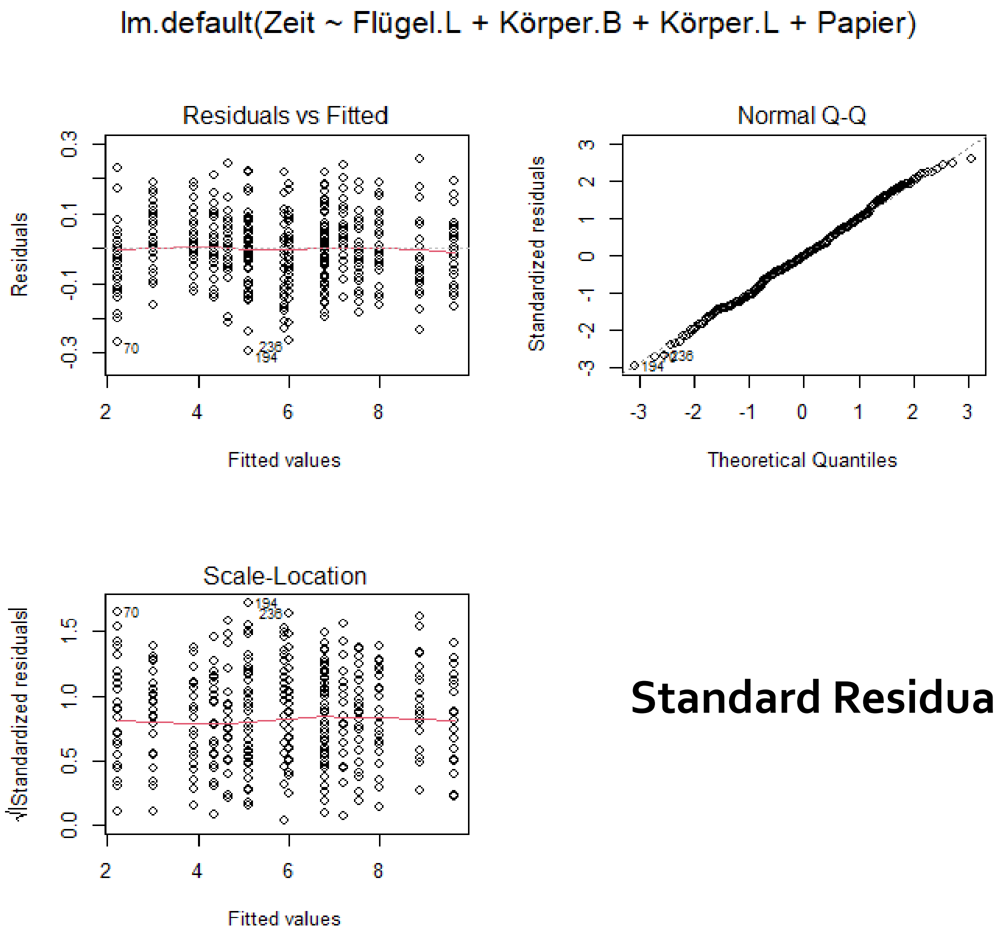
```
One Sample t-test
data:  Konkurrent.A
t = -35.502, df = 29, p-value < 2.2e-16
alternative hypothesis: true mean is less than 9.646
95 percent confidence interval:
     -Inf 7.326904
sample estimates:
mean of x
 7.210333
```

p < a= 0,05: Wir wechseln zur Alternativhypothese, Mittelwert A < 9,646

Unsere Flugzeit liegt über der Zeit unseres schärfsten Konkurrenten.
Wir sind wieder konkurrenzfähig hinsichtlich der Flugzeit!

Aufgabe 4:

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

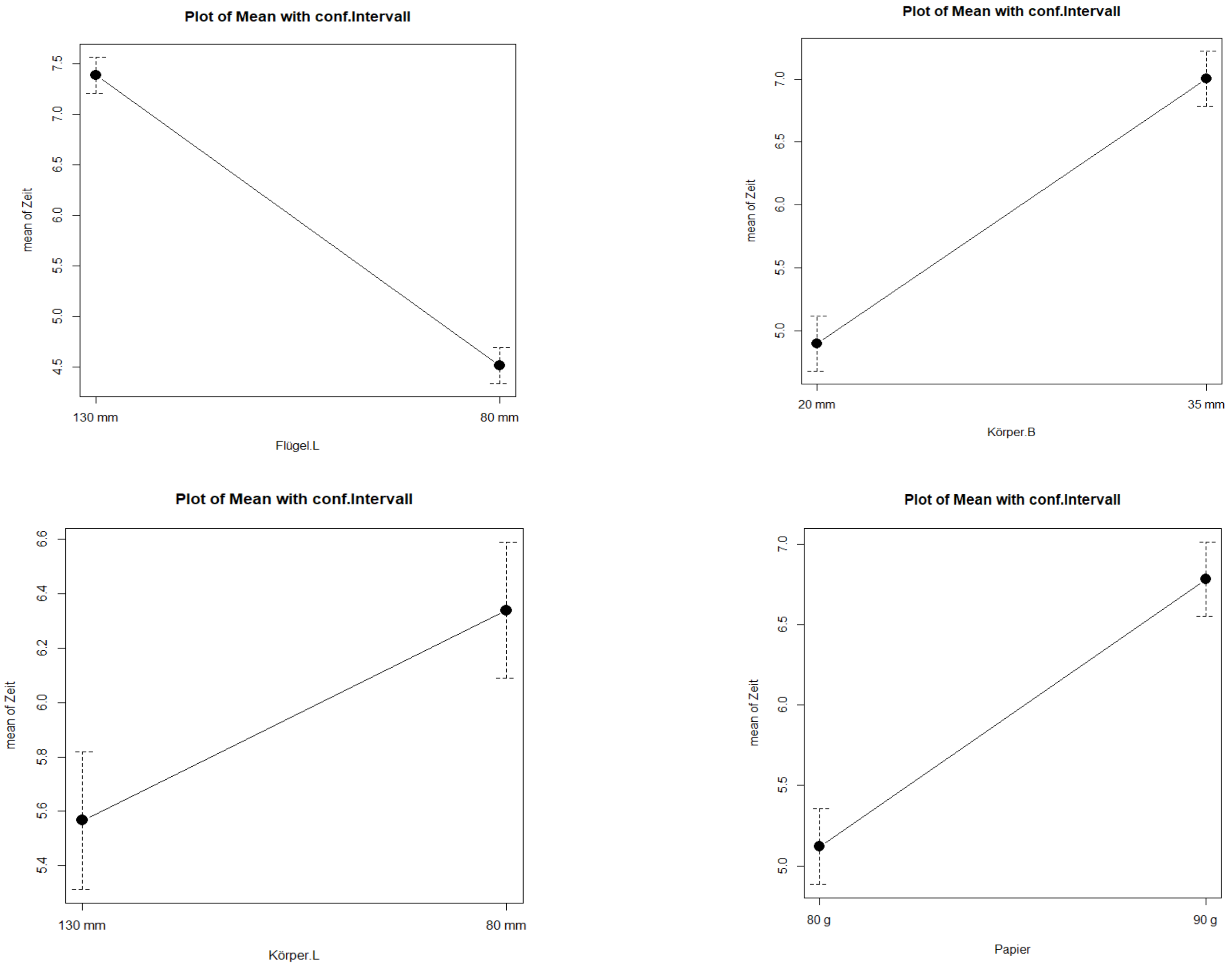


Standard Residuals sind normalverteilt

Aufgabe 4:

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

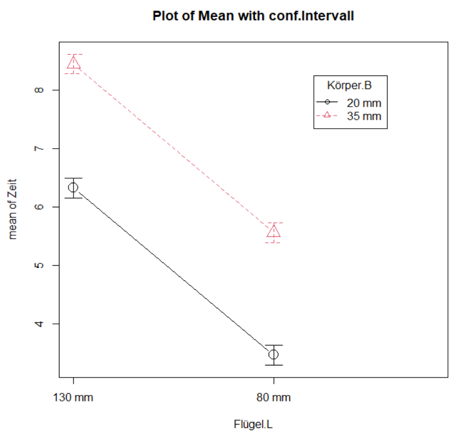
Hauptfaktoren-Diagramme

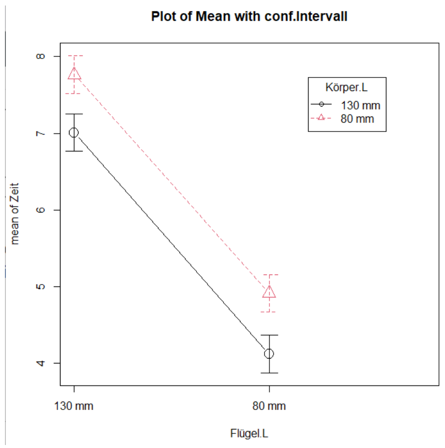


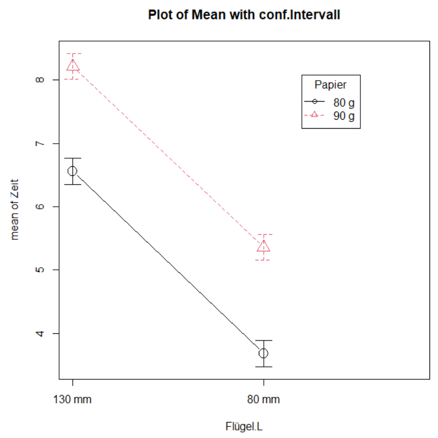
Aufgabe 4:

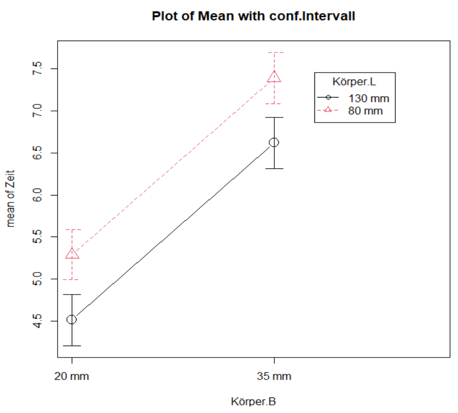
Können wir das Modell soweit verbessern, dass wir konkurrenzfähig sind?

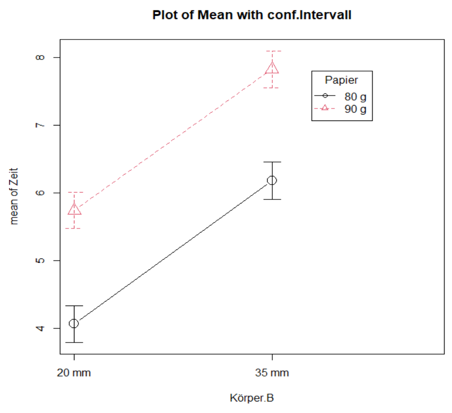
Wechselwirkungsdiagramme

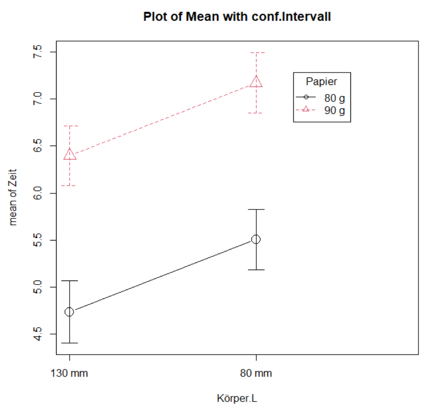












Die grafische Darstellung von Die 2-fachWechselwirkungen deutet auf nicht-signifikante WW hin (die Grafen laufen nur ganz schwach zusammen)

Aufgabe 4:

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

Multikollinearität Prüfung

➔starke Korrelation (|Korr>= 0,7|) von unabhängigen Variablen beim Rechnen einer Multiplen Regression (mehrere Variablen dasselbe messen)

```
Rcmdr>  LinearModel.2 <- lm(Zeit ~ Flügel.L + Körper.B + Körper.L + Papier,
Rcmdr+    data=Data4)

Rcmdr>  summary(LinearModel.2)

Call:
lm(formula = Zeit ~ Flügel.L + Körper.B + Körper.L + Papier,
    data = Data4)

Residuals:
      Min        1Q    Median        3Q       Max
-0.293062 -0.060854 -0.001479  0.067063  0.257063

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.952021   0.004515 1318.38   <2e-16 ***
Flügel.L     1.434396   0.004515  317.72   <2e-16 ***
Körper.B     1.050979   0.004515  232.79   <2e-16 ***
Körper.L    -0.386146   0.004515  -85.53   <2e-16 ***
Papier       0.831688   0.004515  184.22   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09891 on 475 degrees of freedom
Multiple R-squared:  0.9976,    Adjusted R-squared:  0.9976
F-statistic: 4.91e+04 on 4 and 475 DF,  p-value: < 2.2e-16

> vif(LinearModel.2)
Flügel.L Körper.B Körper.L   Papier
       1        1        1        1
```

Werte <10 ➔ es liegt keine Multikollinearität vor

Aufgabe 4:

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

Heteroskedastizität Prüfung

Heteroskedastizität: Zunahme oder Abnahme der Streuung der Residuen des Regressionsmodells

- Verzerrte Standardfehler
- Verzerrte t-Werte
- Verzerrte p-Werte
- Erhöhte Chance für Fehler 1.Art und Fehler 2.Art

Ho: Homoskedastizität liegt vor
H1: Homoskedastizität liegt nicht vor

```
> bptest(LinearModel.2)

        studentized Breusch-Pagan test

data:  LinearModel.2
BP = 0.95031, df = 4, p-value = 0.9172
```

p-Wert>5% Ho: verbleib , wir haben gleichmäßige Streuung → Homoskedastizität