

Statistik – Projektaufgabe 5

Musterlösung

Allgemeine Informationen

- Soweit nicht anders angegeben, beträgt das Signifikanzniveau $1-\alpha = 95\%$
- Geben Sie zu allen Tests die Nullhypothese H_0 und die Alternativhypothese H_1 an, wenn diese wählbar sind
- Sollten während der Bearbeitung Fragen aufkommen, die nicht von der Aufgabenstellung beantwortet werden, wenden Sie sich an den Dozenten

Datenmaterial

5_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.

Ihre Aufgabe

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

- Sind wir bei unserem Flugmodell noch konkurrenzfähig?

Ihre Aufgabe

Bei der Beantwortung der Fragen erwartet das Unternehmen einen aussagekräftigen Bericht.

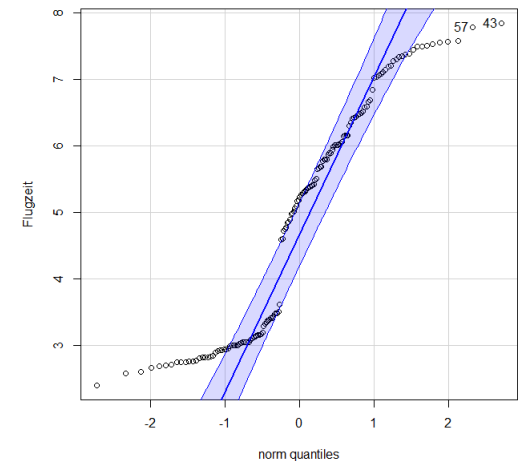
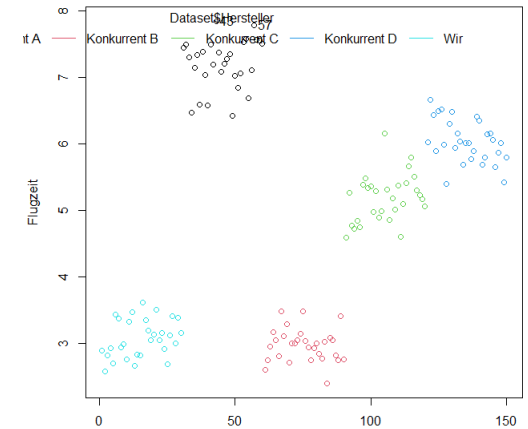
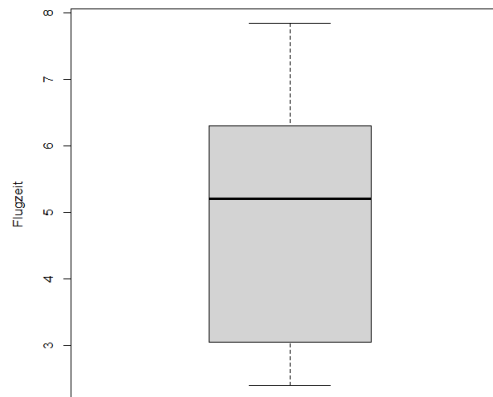
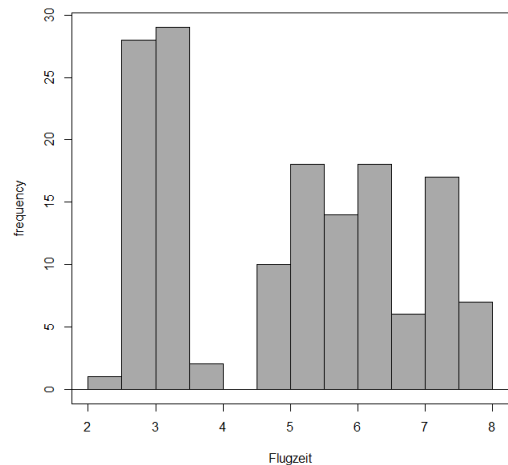
Dieser sollte für jede Fragestellung mit eingehender deskriptiver Statistik starten und anschließend mit geeigneten Werkzeugen statistisch signifikante Aussagen machen.

Sollten bei Tests die Hypothesen wählbar sein, müssen Sie diese im Bericht angeben.

Lösung Aufgabe 5

Deskriptive Statistik – Gesamtdaten

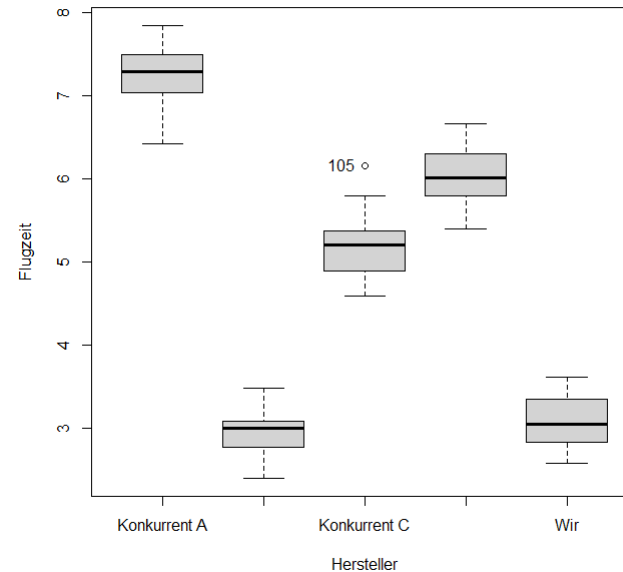
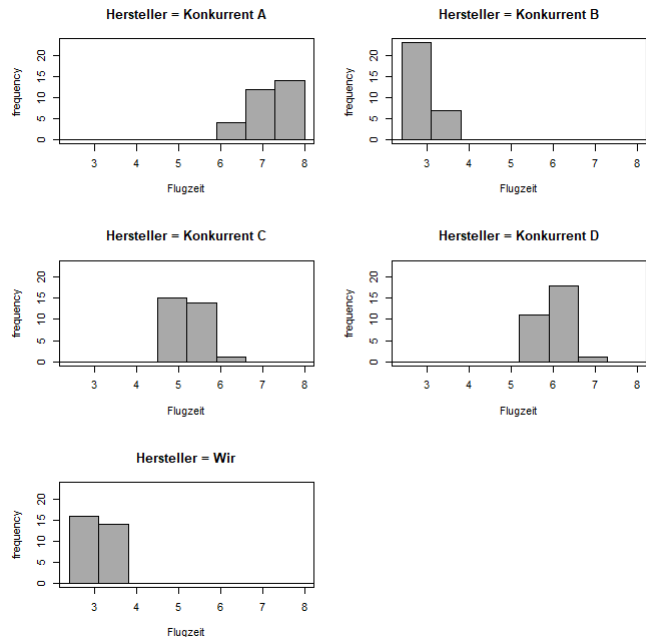
- Boxplot ist unauffällig
- QQ, Histogramm und Index Plot deuten auf unterschiedliche überlagerte Datensätze hin



Lösung Aufgabe 5

Deskriptive Statistik – Einzeldaten

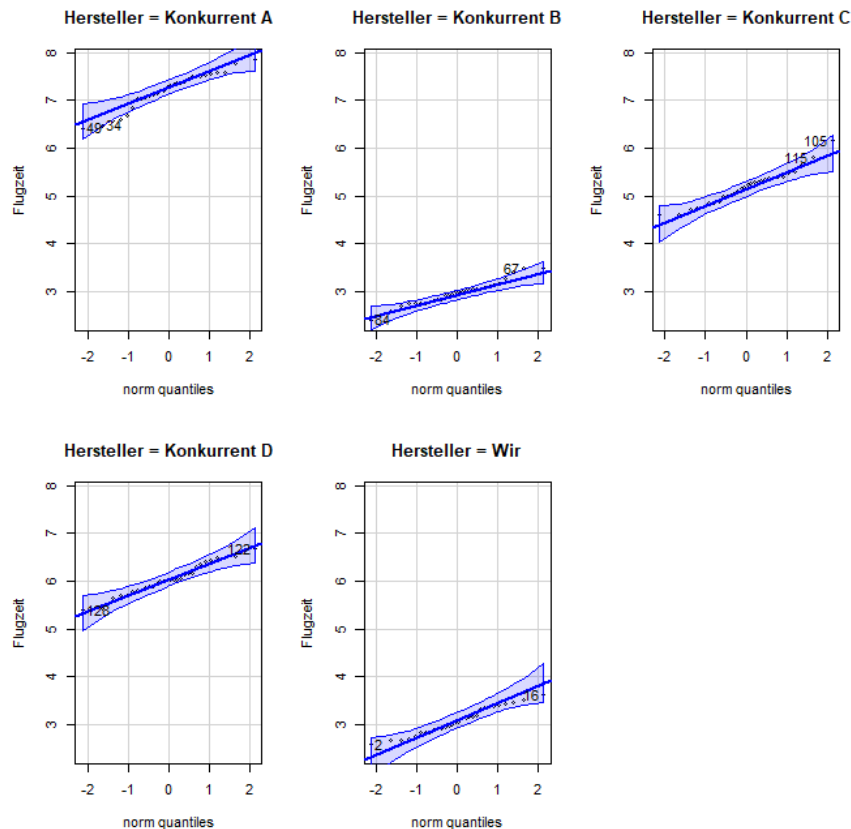
- Histogramme sehen befremdlich aus, das kann aber auf die gemeinsame Skalierung zurückzuführen sein
- Die Boxplots sehen unauffällig aus, unterschiedliche Niveaulagen sind aber deutlich erkennbar



Lösung Aufgabe 5

Deskriptive Statistik – Einzeldaten

- Die QQs deuten alle auf Normalverteilung hin



Lösung Aufgabe 5

Deskriptive Statistik – Einzeldaten

Starke Abweichungen der Mittelwerte, auch Streuung ist unterschiedlich

	mean	sd	Flugzeit: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
Wir	3.081000	0.2832003	30

Lösung Aufgabe 5

Deskriptive Statistik – Einzeldaten

Shapiro-Wilk normality test

data: Flugzeit

p-values

Konkurrent A 0.11129

Konkurrent B 0.42506

Konkurrent C 0.46068

Konkurrent D 0.69817

Wir 0.43543

Für alle Hersteller gilt: $p > \alpha = 0,05$

Wir können alle Datensätze als normalverteilt ansehen

Lösung Aufgabe 5

Deskriptive Statistik – Einzeldaten

Alle Datensätze sind normalverteilt => Bartlett

Bartlett test of homogeneity of variances

data: Flugzeit by Hersteller

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

$p > \alpha = 0,05$

Wir können für alle Datensätze von gleicher Varianz ausgehen

Lösung Aufgabe 5

Einfaktorielle ANOVA

- Alle Voraussetzungen sind erfüllt (metrische Skalierung, Normalverteilung, Varianzgleichheit)

```
Df Sum Sq Mean Sq F value Pr(>F)
Hersteller      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
```

$$p < \alpha = 0,05$$

Es gilt die Alternativhypothese, mindestens zwei Stichproben unterscheiden sich im Mittelwert von einander

Zur Identifizierung wird ein paarweiser Vergleich herangezogen (Tukey: Bestandteil der ANOVA-Rechnung)

Lösung Aufgabe 5

Paarweiser Vergleich – Tukey

Multiple Comparisons of Means: Tukey Contrasts

```
Fit: aov(formula = Flugzeit ~ Hersteller, data = Dataset)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)	
Konkurrent B - Konkurrent A == 0	-4.23300	0.08265	-51.216	<0.0001	***
Konkurrent C - Konkurrent A == 0	-2.02767	0.08265	-24.533	<0.0001	***
Konkurrent D - Konkurrent A == 0	-1.17300	0.08265	-14.192	<0.0001	***
Wir - Konkurrent A == 0	-4.12933	0.08265	-49.962	<0.0001	***
Konkurrent C - Konkurrent B == 0	2.20533	0.08265	26.683	<0.0001	***
Konkurrent D - Konkurrent B == 0	3.06000	0.08265	37.024	<0.0001	***
Wir - Konkurrent B == 0	0.10367	0.08265	1.254	0.719	
Konkurrent D - Konkurrent C == 0	0.85467	0.08265	10.341	<0.0001	***
Wir - Konkurrent C == 0	-2.10167	0.08265	-25.429	<0.0001	***
Wir - Konkurrent D == 0	-2.95633	0.08265	-35.769	<0.0001	***

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

(Adjusted p values reported -- single-step method)

Lösung Aufgabe 5

Paarweiser Vergleich – Tukey

$p > \alpha = 0,05$ für Vergleich Homo Ludens und Konkurrent B

Hier kann von gleichen Mittelwerten ausgegangen werden, die beiden Datensätze gehören zur gleichen Grundgesamtheit

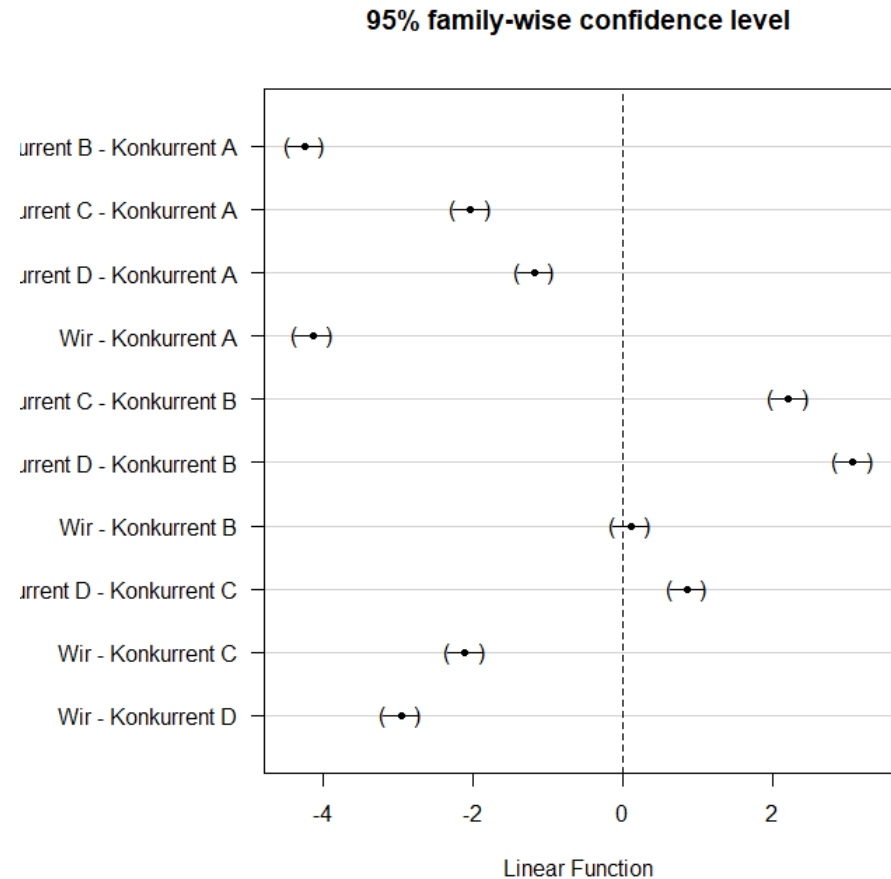
$p < \alpha = 0,05$ für alle anderen Vergleiche

Die Mittelwerte unterscheiden sich signifikant, die folgende Grafik verdeutlicht, dass die „Wir“-Mittelwerte im Vergleich zu den Konkurrenten A, C und D kleiner sind

Lösung Aufgabe 5

Paarweiser Vergleich – Tukey

Die Grafik unterstreicht die Deutung der p-Werte



Lösung Aufgabe 5

Frage der Konkurrenzfähigkeit

Die Ergebnisse der einfaktoriellen ANOVA lassen den Schluss zu, dass wir nur noch mit Hersteller B konkurrieren, alle anderen Hersteller weisen andere und zwar deutlich höhere Mittelwertlagen ihrer Flugzeiten auf.

Die Frage nach dem stärksten Konkurrenten wird nachfolgend mit einer einfaktoriellen ANOVA für die Hersteller A, C und D geklärt.

Lösung Aufgabe 5

Frage der Konkurrenzfähigkeit

Die Deskriptive Statistik der Datensätze A, C und D hat bereits im Vorfeld stattgefunden, auch an der Normalverteilungsannahme hat sich nichts geändert. Die Varianzgleichheit sollte in solchen Fällen aber noch einmal geprüft werden.

```
Bartlett test of homogeneity of variances
```

```
data: Flugzeit by Hersteller
```

```
Bartlett's K-squared = 0.69467, df = 2, p-value = 0.7066
```

$p > \alpha = 0,05$ für Vergleich A, C, D

Wir bleiben bei der Nullhypothese des Varianztests: Die Varianzen sind gleich

Lösung Aufgabe 5

Frage der Konkurrenzfähigkeit

```
          Df Sum Sq Mean Sq F value Pr(>F)
Hersteller  2  62.18  31.089    251.5 <2e-16 ***
Residuals  87  10.75   0.124
```

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

$p < \alpha = 0,05$ für Vergleich A, C, D: Wechsel zur Alternativhypothese, mindestens zwei Mittelwerte unterscheiden sich

Lösung Aufgabe 5

Frage der Konkurrenzfähigkeit

Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = Flugzeit ~ Hersteller, data = Dataset)
Linear Hypotheses:

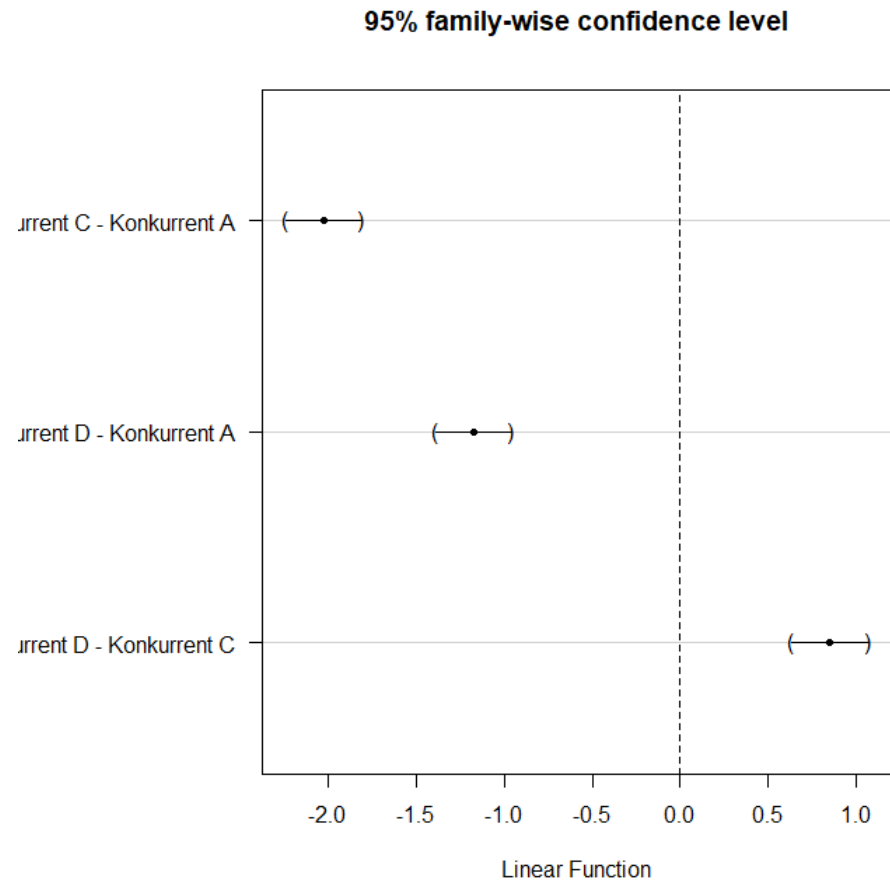
	Estimate	Std. Error	t value	Pr(> t)	
Konkurrent C - Konkurrent A == 0	-2.02767	0.09077	-22.338	<1e-10	***
Konkurrent D - Konkurrent A == 0	-1.17300	0.09077	-12.923	<1e-10	***
Konkurrent D - Konkurrent C == 0	0.85467	0.09077	9.416	<1e-10	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

Alle p-Werte $< \alpha = 0,05$ für Vergleich A, C, D: Wechsel zur Alternativhypothese, alle Mittelwerte unterscheiden sich voneinander

Lösung Aufgabe 5

Frage der Konkurrenzfähigkeit



Lösung Aufgabe 5

Frage der Konkurrenzfähigkeit

Two Sample t-test

data: Flugzeit by Hersteller

t = 12.988, df = 58, p-value < 2.2e-16

alternative hypothesis: true difference in means between group Konkurrent A and group Konkurrent D is not equal to 0

95 percent confidence interval:

0.992214 1.353786

sample estimates:

mean in group Konkurrent A	mean in group Konkurrent D
7.210333	6.037333

Zu guter Letzt ein 2t-Test für die stärksten Modelle A und D

H_0 Die Mittelwerte A,D sind gleich

H_1 Die Mittelwerte A, D sind ungleich

Lösung Aufgabe 5

Frage der Konkurrenzfähigkeit

p-Werte $< \alpha = 0,05$

Wir wechseln zur Alternativhypothese, die Mittelwerte A und D sind nicht gleich.
Konkurrent A hat die höchste Flugzeit