

4.2.5 Box-Whisker-Plot

Will man auf einen Blick ausgewählte Lage- und Streuungsmaße mehrerer Verteilungen miteinander vergleichen, so bietet sich ein Box-Whisker-Plot (BWP) an. Dabei wird jede Stichprobe durch ein Rechteck (Box) dargestellt, dessen Lage und Länge den Interquartilbereich repräsentiert. An beiden Enden der Box werden sogenannte Whiskers (Schnurrhaare) angehängt, die den Abständen von x_{min} bis Q_1 bzw. Q_3 bis x_{max} entsprechen, d.h. die Gesamtlänge der Box mit den beiden Whiskers stellt die Variationsbreite dar. Es existieren viele Modifikationen des BWP, man kann z. B. zur Charakterisierung von Ausreißern eine Art Vertrauensbereich markieren, indem man als Whisker-Längen das 1,5-fache der Abstände ($Z-Q_1$) und (Q_3-Z) nimmt und alle Meßwerte, die außerhalb von Box und Whiskers liegen, gesondert einzeichnet.

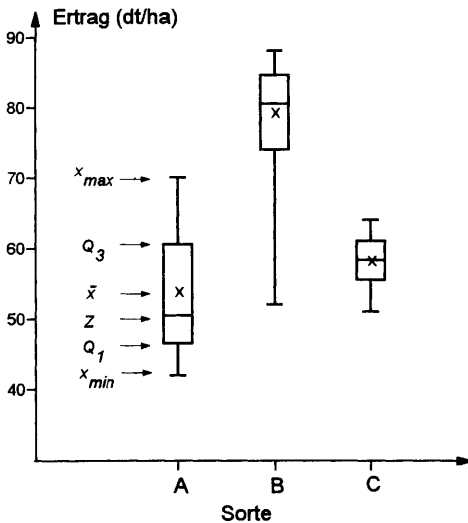


Abb. 4.4: Erträge dreier Sorten A, B, C im Box-Whisker-Plot, Sorte A ist linksgeipflig, Sorte B ist rechtsgeipflig und Sorte C symmetrisch. Alle drei Sorten unterscheiden sich in Lage und Streuung.

4.2.6 Diversität

Für nominalskalierte Merkmale wie z. B. beim kommerziellen Anbau einiger Topf- und Ballenzierpflanzen (Beispiel in Abschn. 3.1.2, Abb. 3.3) charakterisieren die Modalwerte zwar die in Berlin bzw. Hessen am häufigsten angebauten Pflanzen (Pelargonien in Hessen und Tulpen in Berlin), sie stellen aber keine Bezugsgrößen dar, mit deren

Hilft man die Streuung der einzelnen Merkmalsausprägungen um einen Mittelwert quantifizieren kann. In diesen Fällen greift man auf das Konzept der Diversität zurück, indem man die Streuung der Beobachtungshäufigkeiten f_i zwischen den einzelnen Kategorien betrachtet und sogenannte *Diversitätsindizes* berechnet. Sie werden beispielsweise in der Ökologie oft angewandt, und eines der am häufigsten benutzten Diversitätsmaße ist der Shannon-Wiener-Index H_k aus der Informationstheorie. Er wird auch zur Charakterisierung der Artenvielfalt herangezogen. Da er von der Anzahl k der betrachteten Kategorien abhängig ist, sollte er mit der maximal möglichen Diversität von $H_k (= \ln k)$ korrigiert werden. Diese relative Diversität H_{korr} kann für den Vergleich von verschiedenen Untersuchungen herangezogen werden. In der Ökologie wird dieser Index auch mit Evenness E_k bezeichnet. Der Index E_k misst die relative Diversität oder Homogenität in dem Datenmaterial, während $(1-E_k)$ die Heterogenität zwischen den Kategorien bzw. die Dominanz einzelner Kategorien kennzeichnet.

Berechnung des **Diversitätsindex** und der **Evenness** nach Shannon-Wiener

$$H_k = - \sum_{i=1}^k p_i \ln p_i = \frac{1}{n} \left(n \ln n - \sum_{i=1}^k f_i \ln f_i \right) \quad (\text{Formel 4.10})$$

$$E_k = H_{korr} = \frac{H_k}{H_{\max}} = \frac{H_k}{\ln k} \quad 0 \leq E_k \leq 1.0$$

wobei $n = \sum f_i$ der Stichprobenumfang,

$p_i = \frac{f_i}{n}$ die relative Häufigkeit der Kategorie i ,

f_i die Häufigkeit der i -ten Kategorie,

k die Anzahl beobachteter Kategorien,

$H_{\max} = \ln k$ die maximale Diversität bei k Kategorien.

Beispiel 1: Die Diversität im Anbau von fünf verschiedenen Topf- und Ballenzierpflanzen ($k=5$) in Hessen bzw. in Berlin beträgt (Daten aus Abb. 3.3):

$$H_{\text{Hessen}} = -(0.38 \ln 0.38 + 0.28 \ln 0.28 + 0.13 \ln 0.13 + 0.11 \ln 0.11 + 0.10 \ln 0.10) = 1.46$$

$$H_{\text{Berlin}} = -(0.35 \ln 0.35 + 0.01 \ln 0.01 + 0.41 \ln 0.41 + 0.15 \ln 0.15 + 0.08 \ln 0.08) = 0.83$$

und für die relative Diversität bzw. Homogenität (Evenness) folgt mit

$$k=5: E_{\text{Hessen}} = \frac{1.46}{\ln 5} = 0.91 \quad \text{und} \quad E_{\text{Berlin}} = \frac{0.83}{\ln 5} = 0.52. \quad \text{Der kommerzielle}$$

Anbau von Topf- und Ballenzierpflanzen ist in Hessen homogener als in Berlin.

Beispiel 2: In zwei Gebieten in Hessen und Sachsen wurden Musteliden gefangen. Man erhielt folgendes Ergebnis:

Art	Hessen				Sachsen			
	f_i	p_i	$\ln f_i$	$f_i \ln f_i$	f_i	p_i	$\ln f_i$	$f_i \ln f_i$
Marder	5	0.20	1.61	8.05	34	0.85	3.53	119.90
Wiesel	5	0.20	1.61	8.05	2	0.05	0.69	1.39
Iltis	5	0.20	1.61	8.05	2	0.05	0.69	1.39
Mauswiesel	5	0.20	1.61	8.05	2	0.05	0.69	1.39
Summe	20			32.19	40			124.06

Wir berechnen

$$H_{\text{Hessen}} = \frac{1}{20} (20 \ln 20 - 32.19) = 1.39 \text{ und } E_{\text{Hessen}} = \frac{H_{\text{Hessen}}}{\ln 4} = 1.00,$$

$$H_{\text{Sachsen}} = \frac{1}{40} (40 \ln 40 - 124.06) = 0.59 \text{ und } E_{\text{Sachsen}} = \frac{H_{\text{Sachsen}}}{\ln 4} = 0.42.$$

Die Fangergebnisse in Hessen sind offensichtlich völlig homogen (keine Streuung in den Fangzahlen f_i), während im sächsischen Gebiet der Marder dominiert.

4.3 Zur Anwendung der eingeführten Maßzahlen

4.3.1 Standardabweichung und Normalverteilung

Oft kann man davon ausgehen, daß die gegebenen Daten annähernd normal verteilt sind. Die Häufigkeitsverteilung ergibt dann bei stetigem Ausgleich (vgl. Abschn. 3.1.7) eine der Glockenkurve ähnliche Funktion.

Wir wollen uns nun an der graphischen Darstellung der Normalverteilung einige ihrer Eigenschaften veranschaulichen. Die Kurve ist *symmetrisch*, Dichtemittel D , Zentralwert Z und arithmetisches Mittel \bar{x} fallen mit dem Maximum der Funktion zusammen (vgl. § 24.3).

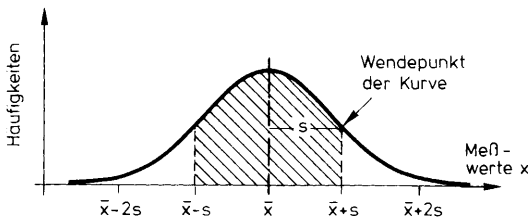


Abb. 4.5: Graphische Darstellung einer Normalverteilung mit Mittelwert \bar{x} und Standardabweichung s . Die schraffierte Fläche macht ca. 68% der Gesamtfläche unter der Kurve aus.

In der letzten Zeit wurden dazu wesentliche Verbesserungen vorgeschlagen, und wir stellen davon das Bonferroni-Holm-Verfahren vor. Darüber hinaus sind viele A-posteriori-Verfahren aufgrund der in Abschn. 15.2 skizzierten Lösungsvorschläge entwickelt worden, von denen wir im Anschluß daran fünf Tests besprechen.

15.2.2 Multiple Testverfahren mit Bonferroni-Korrektur

Wie wir im vorigen Abschnitt am Beispiel des t -Tests aufgezeigt haben, kann im Fall von beliebig vielen, ungeplanten Vergleichen das multiple Risiko, d. h. mindestens eine der geprüften Nullhypothesen irrtümlich abzulehnen, sehr groß werden. Im Gegensatz dazu kontrollieren die A-posteriori-Methoden das multiple Niveau, das tatsächliche Signifikanzniveau im einzelnen Vergleich kann aber erheblich kleiner sein.

Betrachten wir beispielsweise eine Hypothesenfamilie mit $m \geq 2$ paarweisen Vergleichen und überprüfen jede der einzelnen m Hypothesen mit einem Test zum Signifikanzniveau α' , so besteht zwischen α' und dem multiplen Risiko α folgende Ungleichung:

$$\alpha' \leq \alpha \leq m \cdot \alpha'.$$

Diese Beziehung folgt aus der Bonferroni-Ungleichung. Das multiple Risiko ist also nach oben begrenzt. Wählt man beispielsweise $\alpha' = \alpha/m$, so folgt aus der Ungleichung, daß das multiple Signifikanzniveau nicht größer als α sein kann.

Diese Konsequenz aus der Bonferroni-Ungleichheit führt zu Vorgehensweisen, in denen das Signifikanzniveau in jedem einzelnen Test korrigiert wird, um das multiple Signifikanzniveau einzuhalten. Es ist ein allgemein gültiges Prinzip und kann auf alle Testverfahren angewandt werden, die das vergleichsbezogene Niveau einhalten.

Will man beispielsweise nach einer Varianzanalyse mit k Faktorstufen $m \leq k \cdot (k-1)/2$ ungeplante paarweise Vergleiche mit dem multiplen t -Test überprüfen und dabei das multiple Niveau α einhalten, so wählt man für jeden einzelnen Test das Signifikanzniveau $\alpha' = \alpha/m$ (*Bonferroni-Fisher-Verfahren*).

Diese Vorgehensweise ist sehr konservativ und kann dazu führen, daß nicht alle wahren Signifikanzen entdeckt werden. Es wurden deshalb Methoden entwickelt, um diese nachteilige Eigenschaft zu verbessern. Wir stellen hier die Vorgehensweise nach HOLM (*Bonferroni-Holm*) am Beispiel von ungeplanten paarweisen Vergleichen mit dem multiplen t -Test nach Varianzanalyse vor.

Fragestellung: Welche der k Stichprobenmittelwerte $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ unterscheiden sich signifikant?

Voraussetzung: Die Varianzanalyse ergab eine Verwerfung der Nullhypothese. Die Vergleiche sind ungeplant. Es werden jeweils zwei Mittelwerte verglichen, insgesamt $m = k \cdot (k-1)/2$.

Rechenweg:

- (1) Berechne für jeden der m Vergleiche die Teststatistik

$$t_{Vers} = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{MQI}} \cdot \sqrt{\frac{(n_i + n_j)}{n_i \cdot n_j}} \quad (i \neq j)$$

bzw. im balancierten Fall, d. h. für $n_i = n_j = n$:

$$t_{Vers} = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{MQI}} \cdot \sqrt{\frac{n}{2}},$$

wobei n_i (bzw. n_j) der zugehörige Stichprobenumfang zu \bar{x}_i (bzw. \bar{x}_j),
 k die Anzahl der Stichprobenmittelwerte ($i = 1, 2, \dots, k$),
 m die Anzahl der ungeplanten Vergleiche ($l = 1, 2, \dots, m$).

- (2) Berechne $a_l = \frac{\alpha}{m + 1 - l}$

und lies in einer geeigneten t -Tabelle den Wert $t_{Tab}^l = t_{Tab}(FG; a_l)$ ab,

wobei a_l das zugehörige Signifikanzniveau des l -ten
 Vergleichs,
 FG der Freiheitsgrad von MQI .

- (3) Ordne die $t_{Vers} = t_{Vers}^l$ -Werte der Größe nach, beginne mit dem größten $t_{Vers} = t_{Vers}^l$.

- (4) Vergleiche t_{Vers}^l mit t_{Tab}^l und beginne mit $l = 1$:

$$t_{Vers}^l \leq t_{Tab}^l \Rightarrow H_0(\bar{x}_i = \bar{x}_j).$$

$$t_{Vers}^l > t_{Tab}^l \Rightarrow H_1(\bar{x}_i \neq \bar{x}_j).$$

- (5) Wird die Alternativhypothese angenommen, so gehe zurück nach (4) und vergleiche im nächsten Schritt t_{Vers}^l mit t_{Tab}^l für $l = l + 1$. Andernfalls muß die Nullhypothese beibehalten werden, und auch alle noch ausstehenden Vergleiche sind nicht signifikant (Abbruchvorschrift).

Beispiel: In einer einfaktoriellen Varianzanalyse mit vier Faktorstufen (A, B, C, D) und mit jeweils vier Wiederholungen führte der F -Test zur Ablehnung der Nullhypothese. Wir vergleichen mit dem multiplen t -Test alle vier Mittelwerte, ordnen die t_{Vers} -Werte der Größe nach und

erhalten folgende Tabelle ($MQI=12.0$, $FG=15$, $n_i=4$, $k=4$, $m=6$, $\alpha=0.05$):

	Vergleich	t'_{Vers}	$\alpha_l = \alpha / (m+1-l)$	$t'_{Tab}(FG; \alpha_l)$	P_l	Entscheidung
$l=1$	A vs D	5.88	$\alpha_1 = 0.05/6 = 0.0083$	3.04	0.000	H_1
$l=2$	B vs D	4.65	$\alpha_2 = 0.05/5 = 0.0100$	2.95	0.000	H_1
$l=3$	A vs C	4.21	$\alpha_3 = 0.05/4 = 0.0125$	2.84	0.001	H_1
$l=4$	B vs C	2.98	$\alpha_4 = 0.05/3 = 0.0167$	2.69	0.009	H_1
$l=5$	C vs D	1.67	$\alpha_5 = 0.05/2 = 0.0250$	2.49	0.116	H_0 , Abbruch
$l=6$	A vs B	1.60	$\alpha_6 = 0.05/1 = 0.0500$	2.13	0.130	

Bemerkung 1: Sollen Vergleiche gegen die Kontrolle bzw. gegen einen vorgegebenen Standardwert durchgeführt werden, kann man einen geeigneten Zwei- bzw. Einstichprobentest zur Prüfung der Hypothesen wählen und entscheidet entsprechend der in der obigen Tafel dargestellten Vorgehensweise für den multiplen t -Test.

Bemerkung 2: Stehen für jeden Vergleich die entsprechenden Überschreitungswahrscheinlichkeiten P_l zur Verfügung, können die Testentscheidungen durch den direkten Vergleich mit α_l durchgeführt werden und eine Interpolation der Tabellenwerte t_{Tab} entfällt.

15.2.3 Der Newman-Keuls-Test (NK-Test)

Fragestellung: Welche der k Stichprobenmittelwerte $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ unterscheiden sich signifikant?

Voraussetzung: Die Varianzanalyse ergab eine Verwerfung der Nullhypothese. Die Vergleiche sind ungeplant. Es werden jeweils zwei Mittelwerte verglichen.

Rechenweg:

(1) Tafel zum NK-Test

p	$q_\alpha(p; FG)$	R_p
k	$q_\alpha(k; FG)$	R_k
$k-1$	$q_\alpha(k-1; FG)$	R_{k-1}
.	.	.
.	.	.
.	.	.
2	$q_\alpha(2; FG)$	R_2

wobei

k

p

MQI

α

$q_\alpha(p; FG)$

die Anzahl aller Mittelwerte,
die Anzahl beteiligter Mittelwerte,
aus der Varianzanalyse entnommen,
das Signifikanzniveau,
der Tabelle „**studentisierte Variationsbreiten**“ zu entnehmen,

mit $R_p = q_\alpha(p; FG) \cdot \sqrt{\dots}$