

In der Einstichprobensituation in Abschn. 8.2 haben wir einen *asymptotischen* Test der Nullhypothese konstruiert, dass ein behaupteter Wert $\mu^{(0)} = 20m$ der wahre Erwartungswert ist. Für Beobachtungen $\mathbf{x}_n = (x_1, \dots, x_n)^t$ haben wir dort die Statistik

$$T_n(\mathbf{x}_n) = \frac{\bar{x}_n - \mu^{(0)}}{s_n(\mathbf{x}_n)/\sqrt{n}} \quad (9.1)$$

verwendet. Dabei haben wir den Zentralen Grenzwertsatz ausgenutzt, der besagt, dass sich $T_n(\mathbf{x}_n)$ unter schwachen Modellannahmen für $n \rightarrow \infty$ asymptotisch normal verteilt. Die Approximation durch die Normalverteilung kann jedoch besonders für kleine n schlecht sein, vgl. Abb. 9.1. Das ist insbesondere in den Flanken der Verteilung der Teststatistik problematisch, da ja etwa der *P*-Wert über die Flanken beschrieben wird. Es wäre also wünschenswert, auch für kleine n die genaue Verteilung von $T_n(\mathbf{x}_n)$ zu kennen.

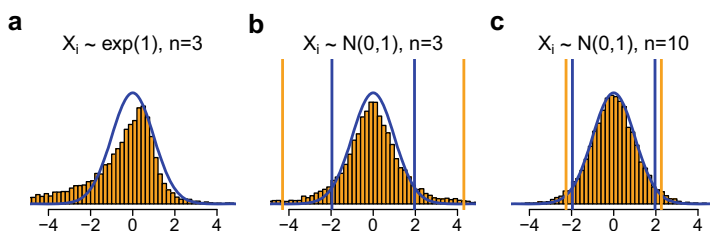


Abb. 9.1 Approximation der empirischen Verteilung von $T_n(\mathbf{X})$ (orange) durch $N(0, 1)$ (blau) für $n = 3$ (a, b) bzw. $n = 10$ (c) und $X_i \sim \exp(1)$ (a) bzw. $X_i \sim \mathcal{N}(0, 1)$ (b,c). Vertikale Linien zeigen die 0.025- und 0.975-Quantile der Standardnormalverteilung (blau) und der Verteilung von $T(\mathbf{X})$ (orange)

Es stellt sich heraus, dass bei einer festen Anzahl an Beobachtungen exakte Aussagen häufig dann möglich sind, wenn im Kontext der statistischen Modellierung geeignete Normalverteilungsannahmen gestellt werden. Im Modell von n unabhängigen und identisch normalverteilten Zufallsvariablen X_1, \dots, X_n leitete William Gosset (Student 1908) die Verteilung von $T_n(\mathfrak{X}_n)$ unter der Nullhypothese her – bei festem $n > 1$. Es ist die sogenannte t -Verteilung mit $n - 1$ Freiheitsgraden, siehe auch Satz 9.9.

Der Schlüssel zum erfolgreichen Umgang mit statistischen Modellen unter Normalverteilungsannahmen ist eine geometrische Betrachtungsweise des Vektors $\mathfrak{X} = (X_1, \dots, X_n)^t$.

Wir leiten im Folgenden zunächst die t -Verteilung her (Abschn. 9.1) und diskutieren die Geometrie des Datenvektors \mathfrak{X} (Abschn. 9.2). Dann werden wir sowohl Students Einstichproben- t -Test (Abschn. 9.3), als auch seinen Zweistichproben- t -Test zum Vergleich von zwei Stichproben (Abschn. 9.4) kennenlernen.

9.1 Die t -Verteilung

Um die t -Verteilung zu definieren, geben wir zunächst eine Reihe verwandter Verteilungen an, deren Beziehungen durch Fisher (1924) dargestellt wurden.

Definition 9.1 (Chiquadratverteilung)

Es seien $n \in \mathbb{N} \setminus \{0\}$ und Z_1, \dots, Z_n unabhängige und identisch verteilte Zufallsvariable mit $Z_1 \sim N(0, 1)$. Eine reellwertige Zufallsvariable X heißt χ^2 -verteilt mit n Freiheitsgraden, kurz $X \sim \chi^2(n)$, falls gilt

$$X \sim Z_1^2 + \dots + Z_n^2.$$

Aufgrund der Beziehung zur Normalverteilung folgen direkt einige elementare Eigenschaften der Chiquadratverteilung.

Lemma 9.2 (Eigenschaften der Chiquadratverteilung)

Es seien X und Y unabhängig und $X \sim \chi^2(n)$ und $Y \sim \chi^2(m)$. Dann gelten folgende Aussagen:

1. $X + Y \sim \chi^2(n + m)$,
2. $\mathbb{E}[X] = n$,
3. $\mathbb{V}\text{ar}(X) = 2n$,
4. X hat die Dichte

$$f(x) = \frac{x^{(n/2)-1}}{\Gamma(n/2)2^{n/2}} \cdot \exp\left(-\frac{x}{2}\right) \mathbb{1}_{(0,\infty)}(x),$$

mit Skalierung via Gammafunktion Γ , vgl. Beispiel 2.1.

Es gilt zudem $X \sim \gamma(n/2, 1/2)$, wie ein Vergleich der Dichten zeigt. Insbesondere gilt für $n = 1$, dass $X \sim \gamma(1/2, 1/2)$, und für $n = 2$, dass $X \sim \exp(1/2)$.

Beweis 1) folgt direkt aus der Definition, 2) gilt wegen $\mathbb{E}[Z_1^2] = \text{Var}(Z_1) + \mathbb{E}[Z_1]^2 = 1$ und der Linearität des Erwartungswerts. Eigenschaft 3) gilt wegen der Unabhängigkeit der Z_i und $\text{Var}(Z_1^2) = \mathbb{E}[Z_1^4] - \mathbb{E}[Z_1^2]^2 = 3 - 1 = 2$. Eigenschaft iv) kann man mithilfe des Dichtetransformationssatzes nachrechnen, siehe zum Beispiel Krengel (2005).

Definition 9.3 (Fisher-Verteilung)

Es seien X und Y unabhängige Zufallsvariablen und $X \sim \chi^2(n)$ und $Y \sim \chi^2(m)$. Eine reellwertige Zufallsvariable F heißt Fisher-verteilt mit n Zähler- und m Nennerfreiheitsgraden, kurz $F \sim \mathcal{F}(n, m)$, falls gilt

$$F \sim \frac{X/n}{Y/m}.$$

Ist $F \sim \mathcal{F}(n, m)$, dann hat F die Dichte

$$f_{n,m}(x) = \frac{\Gamma(n/2 + m/2)}{\Gamma(n/2)\Gamma(m/2)} n^{n/2} m^{m/2} \frac{x^{(n/2)-1}}{(m + nx)^{(n+m)/2}} \cdot \mathbb{1}_{(0,\infty)}(x). \quad (9.2)$$

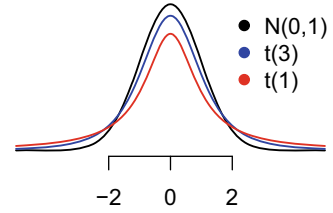
Der Beweis ist zum Beispiel in Georgii (2009) zu finden.

Definition 9.4 (t-Verteilung)

Es seien Z und X unabhängige Zufallsvariablen mit $Z \sim N(0, 1)$ und $X \sim \chi^2(n)$. Eine reellwertige Zufallsvariable T heißt t-verteilt mit n Freiheitsgraden, kurz $T \sim t(n)$, falls gilt

$$T \sim \frac{Z}{\sqrt{X/n}}.$$

Abb. 9.2 Dichte der t -Verteilung



Die t -Verteilung ist eine reellwertige Verteilung, die sich wie die Normalverteilung glockenförmig und symmetrisch um die Null verteilt, siehe Abb. 9.2. Falls $T \sim t(n)$, so hat T die Dichte

$$f_n(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Zudem konvergiert $f_n(x)$ punktweise gegen die Dichte der Standardnormalverteilung.

Zur Herleitung der Dichte der t -Verteilung nutzen wir aus, dass $T^2 \sim \mathcal{F}(1, n)$ mit entsprechender Dichte (vgl. (9.2)). Dann betrachten wir die Abbildung $\Phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ via $\Phi(x) = \sqrt{x}$. Sie ist ein Diffeomorphismus, d.h. bijektiv, stetig differenzierbar und mit stetig differenzierbarer Umkehrabbildung, gegeben durch $\Phi^{-1}(y) = y^2$. Dann hat $|T| = \sqrt{T^2} = \Phi(T^2)$ nach dem Dichtetransformationssatz (vgl. Krengel 2005) die Dichte

$$g(y) = f_{1,n}(\Phi^{-1}(y)) |(\Phi^{-1}(y))'| = f_{1,n}(y^2) 2y \mathbb{1}_{(0,\infty)}(y).$$

Da die Verteilung von T symmetrisch ist, hat sie die Dichte

$$\begin{aligned} h(y) &= f_{1,n}(y^2) |y| \mathbb{1}_{(-\infty,\infty)}(y) \\ &\stackrel{(9.2)}{=} \frac{\Gamma((n+1)/2)}{\Gamma(1/2)\Gamma(n/2)} 1^{1/2} n^{n/2} \frac{(y^2)^{(1/2)-1}}{(n+y^2)^{(n+1)/2}} |y| \mathbb{1}_{(-\infty,\infty)}(y). \end{aligned}$$

Die behauptete Form der Dichte folgt, wenn man ausnutzt, dass $\Gamma(1/2) = \sqrt{\pi}$, dass $n^{n/2} = (1/\sqrt{n})n^{(n+1)/2}$, dass $1/(n+y^2)^{(n+1)/2} = 1/n^{(n+1)/2}(1+y^2/n)^{-(n+1)/2}$ und dass $(y^2)^{-1/2}|y| = 1$.

Mit der gegebenen Form der Dichte der t -Verteilung kann man folgern, dass $f_n(x)$ punktweise gegen die Dichte der Standardnormalverteilung $f(x) = (1/\sqrt{2\pi}) \exp(-x^2/2)$ konvergiert, denn es gilt $(1+x^2/n)^{-(n+1)/2} \rightarrow \exp(-x^2/2)$, und man kann zeigen, dass der erste Faktor gegen $1/\sqrt{2\pi}$ konvergiert für $n \rightarrow \infty$. Diese punktweise Konvergenz von Dichten impliziert übrigens die Konvergenz in Verteilung der entsprechenden Folge von Zufallsvariablen, vgl. Satz von Scheffé zum Beispiel in Shao (2003).

Wir werden im Folgenden sehen, dass die t -Statistik (9.1) unter Normalverteilungsannahme der t -Verteilung mit $(n-1)$ Freiheitsgraden folgt. Die t -Verteilung hat etwas schwerere Flanken, d.h. mehr Gewicht bei extremen Werten als die Normalverteilung. Das ist plausibel, denn in der t -Statistik wird ja zusätzlich die unbekannte Varianz σ^2 durch s^2 geschätzt, sodass zusätzliche Variabilität ins Spiel kommt. Müssten wir σ^2 nicht schätzen

(ersetze in der t -Statistik s durch σ), dann wäre diese Statistik unter Normalverteilungsannahme auch schon bei jedem festen n unter der Nullhypothese $N(0, 1)$ -verteilt.

9.2 Geometrie des Datenvektors

Um zu zeigen, dass die t -Statistik unter Normalverteilungsannahme $t(n - 1)$ -verteilt ist, nutzen wir eine geometrische Betrachtungsweise der n Beobachtungen x_1, \dots, x_n . Wir betrachten dazu im Folgenden einen Vektor $\mathbf{x} = (x_1, \dots, x_n)^t \in \mathbb{R}^n$. Dann leiten wir zum einen eine geometrische Betrachtungsweise des Mittelwerts \bar{x}_n und der empirischen Varianz $s^2(\mathbf{x})$ her. Zum anderen untersuchen wir geometrische Aspekte der mehrdimensionalen Normalverteilung. Damit wird sich die Verteilung der t -Statistik einfach ermitteln lassen.

Wir verstehen den \mathbb{R}^n als euklidischen Vektorraum ausgestattet mit dem Standardskalarprodukt $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n$ via $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$. Dies induziert die euklidische Norm $\|\cdot\| := \sqrt{\langle \cdot, \cdot \rangle}$.

Es sei \mathcal{H} ein linearer Teilraum (Untervektorraum) des \mathbb{R}^n . Dann können wir den \mathbb{R}^n zerlegen in \mathcal{H} und sein orthogonales Komplement \mathcal{H}^\perp . Insbesondere ist \mathcal{H}^\perp selbst wieder ein Untervektorraum. Es gilt $\mathcal{H} \cap \mathcal{H}^\perp = \{0\}$ und $\mathcal{H} + \mathcal{H}^\perp = \mathbb{R}^n$, und wir schreiben abkürzend

$$\mathbb{R}^n = \mathcal{H} \oplus \mathcal{H}^\perp.$$

Folglich setzt sich jedes Element $\mathbf{x} \in \mathbb{R}^n$ zusammen aus einem Anteil $\mathcal{P}_{\mathcal{H}}\mathbf{x}$, der in \mathcal{H} liegt, und einem Anteil $\mathcal{P}_{\mathcal{H}^\perp}\mathbf{x}$ aus \mathcal{H}^\perp ,

$$\mathbf{x} = \mathcal{P}_{\mathcal{H}}\mathbf{x} + \mathcal{P}_{\mathcal{H}^\perp}\mathbf{x}.$$

Die Orthogonalität von \mathcal{H} und \mathcal{H}^\perp impliziert, dass $\mathcal{P}_{\mathcal{H}}\mathbf{x} \perp \mathcal{P}_{\mathcal{H}^\perp}\mathbf{x}$ bzw.

$$\langle \mathcal{P}_{\mathcal{H}}\mathbf{x}, \mathcal{P}_{\mathcal{H}^\perp}\mathbf{x} \rangle = 0,$$

d.h., $\mathcal{P}_{\mathcal{H}}\mathbf{x}$ entspricht der *orthogonalen Projektion* von \mathbf{x} auf \mathcal{H} und $\mathcal{P}_{\mathcal{H}^\perp}\mathbf{x}$ der orthogonalen Projektion von \mathbf{x} auf \mathcal{H}^\perp .

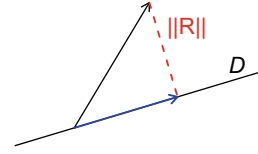
9.2.1 Geometrie von Mittelwert und Standardabweichung

Wir betrachten zunächst einen bestimmten Untervektorraum, die *Diagonale* \mathcal{D} , gegeben durch

$$\mathcal{D} := \{c\mathbb{1} \mid c \in \mathbb{R}\}, \quad (9.3)$$

wobei $\mathbb{1} = (1, \dots, 1)^t \in \mathbb{R}^n$ den Vektor bezeichne, der nur aus Einsen besteht.

Abb. 9.3 Zur Geometrie von \bar{x}_n und $s(\mathbf{x})$



Es sei $\mathbf{x} \in \mathbb{R}^n$. Der Mittelwert \bar{x}_n ist mit der Projektion von \mathbf{x} auf \mathcal{D} assoziiert, und die Standardabweichung $s(\mathbf{x})$ mit der Projektion von \mathbf{x} auf das orthogonale Komplement \mathcal{D}^\perp , vgl. Abb. 9.3.

Lemma 9.5 (Geometrie von Mittelwert und Standardabweichung)

Es sei $n \geq 2$ und $\mathbf{x} \in \mathbb{R}^n$. Dann gilt

1. $\mathcal{P}_{\mathcal{D}}\mathbf{x} = \bar{x}_n \mathbb{1}$,
2. $\|\mathcal{P}_{\mathcal{D}^\perp}\mathbf{x}\| = \sqrt{n-1} s(\mathbf{x})$.

Beweis

1. Wir suchen den Vektor $c\mathbb{1}$ so, dass $R_c := \mathbf{x} - c\mathbb{1}$ senkrecht auf \mathcal{D} steht, d. h.

$$0 \stackrel{!}{=} \langle \mathbb{1}, \mathbf{x} - c\mathbb{1} \rangle = \sum_{i=1}^n (x_i - c) = \sum_{i=1}^n x_i - nc,$$

und damit ist $c = (1/n) \sum_{i=1}^n x_i = \bar{x}_n$.

2. Die Zerlegung $\mathbb{R}^n = \mathcal{D} \oplus \mathcal{D}^\perp$ liefert dann

$$\mathbf{x} = \mathcal{P}_{\mathcal{D}}\mathbf{x} + \mathcal{P}_{\mathcal{D}^\perp}\mathbf{x} = \bar{x}_n \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \begin{pmatrix} x_1 - \bar{x}_n \\ \vdots \\ x_n - \bar{x}_n \end{pmatrix},$$

also ist $\mathcal{P}_{\mathcal{D}^\perp}\mathbf{x} = \mathbf{x} - \bar{x}_n \mathbb{1}$ und somit $\|\mathcal{P}_{\mathcal{D}^\perp}\mathbf{x}\|^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2 = (n-1)s^2(\mathbf{x})$.

Die orthogonale Projektion entspricht der Minimierung des Längenquadrates von $R_c(\mathbf{x}) := \mathbf{x} - c\mathbb{1}$ unter allen $c \in \mathbb{R}$ – dieses Prinzip ist als *Methode der kleinsten Quadrate* bekannt, d. h. der Mittelwert ist der sogenannte *Kleinste-Quadrate-Schätzer* (engl. *least squares estimator*, kurz LSE) von c , denn

$$\hat{c}(\mathbf{x}) := \arg \min_{c \in \mathbb{R}} \|R_c(\mathbf{x})\|^2 = \arg \min_{c \in \mathbb{R}} \sum_{i=1}^n (x_i - c)^2 = \bar{x}_n.$$

Der Minimierer $R(\mathbf{x}) := R_{\hat{c}(\mathbf{x})}(\mathbf{x})$ des Quadrates $\|R_c(\mathbf{x})\|^2$ wird auch häufig als das *Residuum* von \mathbf{x} bzgl. \mathcal{D} bezeichnet. Es gilt $R(\mathbf{x}) = \mathcal{P}_{\mathcal{D}^\perp}(\mathbf{x})$.

Die korrigierte Stichprobenstandardabweichung $s(\mathbf{x})$ lässt sich also durch das Residuum ausdrücken mit $s(\mathbf{x}) = \|R(\mathbf{x})\|/\sqrt{n-1}$.

Die Darstellung von Mittelwert und Standardabweichung durch Projektionen ist intuitiv nachvollziehbar: Dass $\mathcal{P}_{\mathcal{D}}\mathbf{x} = \bar{x}_n \mathbb{1}$ gilt, ergibt Sinn, denn wenn wir nach einem Vektor mit identischen Komponenten $c\mathbb{1}$ suchen, der \mathbf{x} ‚bestmöglich‘ beschreibt (im Sinne der kleinsten Quadrate), dann sollte der Mittelwert \bar{x}_n ein vernünftiger Kandidat für den Faktor c sein. Zudem ergibt es Sinn, dass sich die Variabilität der Daten über das Residuum ausdrücken lässt. Denn je kleiner das Residuum, desto weniger weichen die x_i von \bar{x}_n ab, d. h. desto geringer ist die Variabilität der Daten (Abb. 9.3).

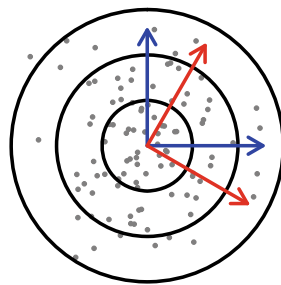
Wir bemerken schließlich, dass die Projektionen $\mathcal{P}_{\mathcal{D}}$ und $\mathcal{P}_{\mathcal{D}^\perp}$ Statistiken sind, d. h. es wird uns nichts davon abhalten, nicht nur deterministisches $\mathbf{x} \in \mathbb{R}^n$, sondern auch einen Zufallsvektor $\mathfrak{X} \in \mathbb{R}^n$ zu projizieren.

9.2.2 Geometrie der mehrdimensionalen Standardnormalverteilung

Wir werden sehen, dass sich die Verteilung der t -Statistik wunderschön über geometrische Argumente im Zusammenhang mit der sogenannten mehrdimensionalen Standardnormalverteilung herleiten lässt. Wir werden Letztere nun definieren und zeigen, dass sie invariant unter orthogonalen Transformationen ist, vgl. Abb. 9.4. Diese wichtige Eigenschaft bildet das Rückgrat sämtlicher Modelle, die wir im Kontext der Normalverteilung diskutieren.

Es sei \mathbb{O}_n die orthogonale Gruppe, d. h. die Menge der orthogonalen $n \times n$ -Matrizen, deren Spalten eine Orthonormalbasis (ONB) des \mathbb{R}^n bilden.

Abb. 9.4 Invarianz der $N(0, I)$ -Verteilung unter orthogonalen Transformationen



Hilfslemma 9.6 (Eigenschaften orthogonaler Matrizen)

Für $M \in \mathbb{O}_n$ gilt

1. $M^{-1} = M^t \in \mathbb{O}_n$,
2. $|\det M| = 1$,
3. $\|M\mathbf{x}\| = \|\mathbf{x}\| \quad \forall \mathbf{x} \in \mathbb{R}^n$.

Beweis

1. Es bezeichne E_n die Einheitsmatrix im \mathbb{R}^n . Dann ist $M^t M = E_n$ also $M^t = M^{-1}$. Damit ist auch $(M^{-1})^t = M$, also $(M^{-1})^t M^{-1} = M M^{-1} = E_n$, d.h. die Spalten von M^{-1} bilden eine ONB, sodass $M^{-1} \in \mathbb{O}_n$.
2. Nach dem Determinantenmultiplikationssatz gilt

$$(\det(M))^2 = \det(M) \cdot \det(M^t) = \det(M M^t) = \det(E_n) = 1.$$

3. $\|M\mathbf{x}\|^2 = (M\mathbf{x})^t (M\mathbf{x}) = \mathbf{x}^t M^t M \mathbf{x} = \|\mathbf{x}\|^2$.
Bedeutung: \mathbb{O}_n beschreibt die Menge der ‚Drehspiegelungen‘ im \mathbb{R}^n . Eine orthogonale Matrix ist eine längen- (wegen 3.)) und winkeltreue (wegen 2.)) lineare Abbildung.

Definition 9.7 (Multivariate Standardnormalverteilung)

Ein Zufallsvektor $\mathfrak{Z} = (Z_1, \dots, Z_n)^t \in \mathbb{R}^n$ heißt *standardnormalverteilt* auf \mathbb{R}^n , falls seine Komponenten Z_1, \dots, Z_n unabhängige und identisch verteilte Zufallsvariable beschreiben, mit $Z_1 \sim N(0, 1)$. Wir schreiben kurz $\mathfrak{Z} \sim N(0, E_n)$, wobei 0 den Nullvektor im \mathbb{R}^n und E_n die $n \times n$ -Einheitsmatrix bezeichnen.

Die Dichte $f_{\mathfrak{Z}}(\mathbf{z})$ von \mathfrak{Z} ist gegeben durch

$$f_{\mathfrak{Z}}(\mathbf{z}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z_i^2\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^n z_i^2\right),$$

mit $\mathbf{z} \in (z_1, \dots, z_n)^t \in \mathbb{R}^n$. Der Wert der Dichte hängt also nur von $\|\mathbf{z}\|$ ab. Das heißt, die Dichte ist rotationssymmetrisch um die Null, die Niveaulinien bilden Kreise.

Aufgrund der Rotationssymmetrie ist die $N_n(0, E_n)$ -Verteilung invariant unter längen- und winkeltreuen linearen Abbildungen, d.h. invariant unter orthogonalen Transformationen. Diese Eigenschaft werden wir zur Herleitung der Verteilung der t -Statistik nutzen. Wir

bemerken zudem, dass viele andere Modelle, mit denen man in der Praxis hantiert – Varianzanalysen (Kap. 10), lineare Regressionsmodelle (Kap. 11) etc. –, auf dieser Eigenschaft basieren, und daher halten wir sie im folgenden wichtigen Satz fest:

Satz 9.8 (Invarianz von $N_n(0, E_n)$ unter orthogonalen Transformationen)

Es sei $\mathfrak{Z} \sim N_n(0, E_n)$, $M \in \mathbb{O}_n$ und \mathcal{H} ein k -dimensionaler Teilraum des \mathbb{R}^n mit $k \in \{1, \dots, n\}$. Dann gelten folgende Aussagen:

1. $\mathfrak{X} := M\mathfrak{Z} \sim N_n(0, E_n)$.
2. In jeder ONB des \mathbb{R}^n sind die Koordinaten von \mathfrak{Z} wieder unabhängig und $N_n(0, E_n)$ -verteilt.
3. $\|P_{\mathcal{H}}\mathfrak{Z}\|^2 \sim \chi^2(k)$.
4. $P_{\mathcal{H}}\mathfrak{Z}$ und $P_{\mathcal{H}^\perp}\mathfrak{Z}$ sind unabhängig.

Beweis

1. Wir betrachten zunächst die Abbildung $M : \mathbb{R}^n \rightarrow \mathbb{R}^n$ via $\mathbf{z} \mapsto M\mathbf{z}$. Sie ist ein Diffeomorphismus, denn sie ist bijektiv und stetig differenzierbar und besitzt eine stetig differenzierbare Umkehrabbildung $M^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ via $\mathbf{y} \mapsto M^{-1}\mathbf{y}$. Da M^{-1} eine lineare Abbildung ist, folgt für die Jacobi-Matrix $(M^{-1}\mathbf{y})' = M^{-1}$. Aus dem Dichtetransformationssatz (siehe Krengel 2005) folgt die Dichte von $\mathfrak{X} = M\mathfrak{Z}$ als

$$f_{\mathfrak{X}}(\mathbf{y}) = f_{\mathfrak{Z}}(M^{-1}\mathbf{y}) |\det((M^{-1}\mathbf{y})')| = f_{\mathfrak{Z}}(\mathbf{y}),$$

da $f_{\mathfrak{Z}}$ rotationssymmetrisch ist, und nach Lemma 9.6 gilt $\|M^{-1}\mathbf{y}\| = \|\mathbf{y}\|$ und $|\det M^{-1}| = 1$.

2. Sei $\mathbf{e}_1, \dots, \mathbf{e}_n$ eine ONB von \mathbb{R}^n und $M := (\mathbf{e}_1, \dots, \mathbf{e}_n) \in \mathbb{O}_n$. Wir verstehen M hier als ‚Basiswechselmatrix‘. Dann sind die Koordinaten von \mathfrak{Z} in der neuen Basis gegeben durch die Komponenten von $\mathfrak{X} = M^t\mathfrak{Z}$. Nach 1. ist $\mathfrak{X} \sim N_n(0, E_n)$, da $M^t \in \mathbb{O}_n$.
3. und 4. Sei eine ONB von \mathcal{H} gegeben durch $\mathbf{e}_1, \dots, \mathbf{e}_k$. Ergänze diese durch $\mathbf{e}_{k+1}, \dots, \mathbf{e}_n$ zu einer ONB des \mathbb{R}^n . Es seien die Koordinaten von \mathfrak{Z} in dieser Basis gegeben durch $X_1, \dots, X_k, X_{k+1}, \dots, X_n$. Diese sind nach 2. unabhängige und identisch verteilte Zufallsvariablen mit $X_1 \sim N(0, 1)$, sodass auch

$$\mathcal{P}_{\mathcal{H}}\mathfrak{Z} = \sum_{i=1}^k X_i \mathbf{e}_i \quad \text{und} \quad \mathcal{P}_{\mathcal{H}^\perp}\mathfrak{Z} = \sum_{i=k+1}^n X_i \mathbf{e}_i$$

unabhängig sind und per definitionem gilt

$$\|\mathcal{P}_{\mathcal{H}} \mathfrak{Z}\|^2 = \sum_{i=1}^k X_i^2 \sim \chi^2(k) \quad \text{und} \quad \|\mathcal{P}_{\mathcal{H}^\perp} \mathfrak{Z}\|^2 = \sum_{i=k+1}^n X_i^2 \sim \chi^2(n-k).$$

9.3 Der Einstichproben-t-Test

Das Lemma 9.5 über die Geometrie des Mittelwerts und der Standardabweichung in Kombination mit Satz 9.8 über die Geometrie der mehrdimensionalen Normalverteilung erlaubt uns nun, die Verteilung der t -Statistik herzuleiten.

Satz 9.9 (Verteilung der Einstichproben- t -Statistik)

Es sei $n \geq 2$ und $\mathfrak{X} = (X_1, \dots, X_n)^t$ ein Zufallsvektor mit unabhängigen und identisch verteilten Komponenten und $X_1 \sim N(\mu, \sigma^2)$, mit $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. Sei $s^2(\mathfrak{X}) = (1/(n-1)) \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Dann gilt

$$T(\mathfrak{X}) := \frac{\bar{X}_n - \mu}{s(\mathfrak{X})/\sqrt{n}} \sim t(n-1).$$

Beweis Wir zeigen, dass T verteilt ist wie der Quotient zweier unabhängiger Zufallsvariablen, wobei der Zähler standardnormalverteilt ist und der Nenner sich als die Wurzel einer mit $n-1$ skalierten $\chi^2(n-1)$ -verteilten Zufallsvariablen schreibt (vgl. Definition 9.4). Wir definieren standardisierte Zufallsvariable $Z_i := (X_i - \mu)/\sigma$ für $i = 1, \dots, n$, sodass $\mathfrak{Z} = (Z_1, \dots, Z_n)^t \sim N_n(0, E_n)$, und finden sofort

$$T(\mathfrak{X}) = \frac{\bar{X}_n - \mu}{s(\mathfrak{X})/\sqrt{n}} = \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\frac{s(\mathfrak{X})}{\sigma}} = \frac{\sqrt{n} \bar{Z}_n}{\left(\frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z}_n)^2\right)^{1/2}}. \quad (9.4)$$

Im Zähler sind wir fertig, denn $\bar{Z}_n \sim N(0, 1/n)$, sodass $\sqrt{n} \bar{Z}_n \sim N(0, 1)$. Wegen Lemma 9.5 und Satz 9.8 finden wir, dass $\bar{Z}_n \mathbb{1} = \mathcal{P}_{\mathcal{D}} \mathfrak{Z}$, sowie

$$\sum_{i=1}^n (Z_i - \bar{Z}_n)^2 = \|\mathcal{P}_{\mathcal{D}^\perp} \mathfrak{Z}\|^2 \sim \chi^2(n-1).$$

Inbesondere sind Zähler und Nenner in (9.4) unabhängig, und der Nenner schreibt sich als Wurzel aus einer mit $n-1$ skalierten $\chi^2(n-1)$ -verteilten Zufallsvariablen, sodass $T(\mathfrak{X}) \sim t(n-1)$.

Interessanterweise folgt aus dem Beweis auch, dass bei unabhängigen und identisch normalverteilten Zufallsvariablen auch der empirische Mittelwert und die empirische Standardabweichung unabhängig sind. Für die konkrete Anwendung formulieren wir Satz 9.9 als folgendes Lemma:

Lemma 9.10 (Einstichproben-t-Test)

Es sei ein statistisches Modell gegeben durch unabhängige und identisch verteilte Zufallsvariable X_1, \dots, X_n mit $X_1 \sim N(\mu, \sigma^2)$ und $(\mu, \sigma^2) \in \Theta := \mathbb{R} \times \mathbb{R}^+$. Zudem sei s wie in (3.3) und $\alpha \in (0, 1)$ sowie q_α das α -Quantil der $t(n-1)$ -Verteilung. Dann gelten folgende Aussagen:

1. Student'sches Konfidenzintervall: Das Intervall I gegeben durch

$$I(\mathbf{x}) := \left[\bar{x}_n - q_{1-\frac{\alpha}{2}} \cdot \frac{s(\mathbf{x})}{\sqrt{n}}, \bar{x}_n + q_{1-\frac{\alpha}{2}} \cdot \frac{s(\mathbf{x})}{\sqrt{n}} \right]$$

ist ein Konfidenzintervall für μ zum Niveau $(1 - \alpha)$.

2. Student's t-Test: Für $\mu^{(0)}$ in \mathbb{R} sei eine Nullhypothese gegeben durch

$$H_0 : (\mu, \sigma^2) \in \Theta_0 = \{(\mu, \sigma^2) \in \Theta \mid \mu = \mu^{(0)}\}.$$

Dann ist die t-Statistik

$$T(\mathbf{x}) := \frac{\bar{x}_n - \mu^{(0)}}{s(\mathbf{x})/\sqrt{n}}$$

eine Teststatistik für einen Test der Nullhypothese H_0 zum Niveau α mit Ablehnungsbereich $\mathcal{R}(\alpha) = (-\infty, q_{\alpha/2}] \cup [q_{1-\alpha/2}, \infty)$.

Wir bemerken, dass wir den α -Fehler $\mathbb{P}_{H_0}(T(\mathfrak{X}) \in \mathcal{R}) = \alpha$ als Konsequenz der Normalverteilungsannahme für jedes n kontrollieren können. Im Vergleich zum nichtparametrischen Beispiel 8.1 kommen wir daher ohne Asymptotiken aus. Schließlich gilt wieder $\{T(\mathbf{x}) \in \mathcal{R}(\alpha)\}$ genau dann, wenn $\{I(\mathbf{x}) \not\ni \mu^{(0)}\}$.

Beispiel: Einstichproben-t-Test bei vorgegebenem $\mu^{(0)}$

Wir kommen zurück zu dem Blauwalbeispiel aus Abschn. 8.2. Dort hatten wir unter allgemeinen Verteilungsannahmen mithilfe eines asymptotischen Verfahrens die Nullhypothese getestet, dass die mittlere Länge von Blauwalen 20 m beträgt. Nun verwenden wir für das gegebene (relativ kleine) $n = 16$ ein exaktes Verfahren unter der Annahme, dass die Längen

der Blauwale einer Normalverteilung folgen. Die Teststatistik nimmt denselben Wert an, $T(\mathbf{x}) = 5.6$, und mit der t -Verteilung mit $n - 1 = 15$ Freiheitsgraden erhalten wir nun einen (unter Normalverteilungsannahme exakten) P -Wert von

$$P(\mathbf{x}) := \mathbb{P}_{H_0}(|T(\mathfrak{X})| \geq 5.6) < 10^{-4}.$$

9.3.1 Gepaarter Zweistichproben- t -Test

Interessanterweise kann der Einstichproben- t -Test auch in einer auf den ersten Blick anders erscheinenden Stichprobensituation auf genau dieselbe Weise eingesetzt werden. Betrachten wir zum Beispiel die Frage, ob die Tauchtiefe eines Blauwals davon abhängt, ob er auf Nahrungssuche ist oder nicht. Um dies zu untersuchen, könnte man Wale mit einem Sender ausstatten und für dieselben Wale die Tiefe je eines Tauchgangs mit und ohne Nahrungssuche vergleichen. Man hat es dann mit sogenannten *gepaarten* Beobachtungen zu tun: Es seien x_1, \dots, x_n die Tauchtiefen der n Wale auf Nahrungssuche, und y_1, \dots, y_n die Tauchtiefen derselben n Wale, wenn sie nicht auf Nahrungssuche waren. Eigentlich haben wir es hier also mit zwei Stichproben zu tun, aber dadurch, dass jeder Messung x_i genau eine Messung y_i zugeordnet werden kann, die zu demselben Wal gehört, spricht man von *gepaarten* Stichproben.

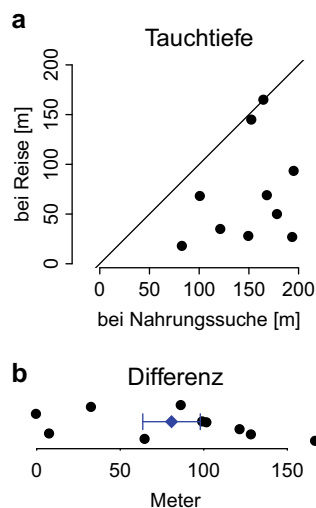
Mathematisch ist die Herangehensweise nun ganz analog zum Einstichprobentest: Untersuchen wir die Nullhypothese, dass die mittlere Tauchtiefe nicht davon abhängt, ob der Wal auf Nahrungssuche ist, so können wir zur Analyse dieser Änderung gerade die Differenz der gepaarten Beobachtungen $d_i := x_i - y_i$ betrachten und so die Beobachtungen auf die Einstichprobensituation zurückführen. Wir modellieren die d_i als Realisierungen von unabhängigen und identisch normalverteilten Zufallsvariablen D_i und testen ganz analog die Nullhypothese, dass der Erwartungswert der D_i in Wahrheit null ist, wobei die D_i in Lemma 9.10 die Rolle der X_i übernehmen.

Ein Datenbeispiel findet sich in Abb. 9.5a. Da bei gepaarten Beobachtungen jedem x_i genau ein y_i zugeordnet ist, werden diese Paare (x_i, y_i) in einem *Streudiagramm* dargestellt. Bei Punkten oberhalb (unterhalb) der Hauptdiagonalen war y_i größer (kleiner) als x_i . Da fast alle Tauchtiefen bei Nahrungssuche größer waren, ist auch die Differenz fast immer positiv (Abb. 9.5b). Mit einer mittleren Differenz von etwa 80 und einer Standardabweichung der Differenz von etwa 55 berechnen wir für $n = 10$ eine t -Statistik von $T(\mathbf{d}) \approx 4.6$, wobei $\mathbf{d} = (d_1, \dots, d_n)^t$. Diese ist unter der Nullhypothese t -verteilt mit $n - 1 = 9$ Freiheitsgraden, was zu einem P -Wert von etwa $P(\mathbf{d}) \approx 0.001$ führt.

Interessant ist an diesem Beispiel auch, dass bei der Analyse lediglich eine Normalverteilungsannahme an die *Differenzen* D_i der gepaarten Beobachtungen gestellt wird. Die Einzelbeobachtungen selbst müssen keiner Normalverteilung folgen.

In Situationen mit zwei *ungepaarten* Stichproben dagegen – zum Beispiel wenn man die Tauchtiefen von n Blauwalen und m Finnwalen vergleichen möchte – kann man den

Abb. 9.5 Gepaarte Beobachtungen



Messungen x_i keine Messung der y_j zuordnen. Die Verallgemeinerung des Einstichproben- t -Tests auf diese Situation zweier ungepaarter Stichproben diskutieren wir in Abschn. 9.4.

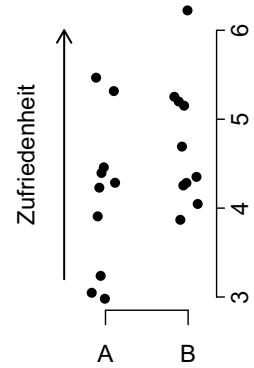
9.4 Der Zweistichproben-t-Test

Die Ergebnisse aus dem Einstichproben- t -Test können wir auf die Situation mit zwei Stichproben übertragen. Im Szenario des Zweistichproben- t -Tests werden zwei Stichproben $\mathbf{x} = (x_1, \dots, x_m)^t$ und $\mathbf{y} = (y_1, \dots, y_n)^t$ miteinander verglichen.

Ein Beispiel: Sie möchten für ein Semester an einer anderen Universität studieren und schwanken zwischen zwei Studienorten A und B. Sie lesen einen Bericht, in dem die Zufriedenheit von Mathematikstudierenden an verschiedenen Universitäten untersucht wurde. Wir nehmen der Einfachheit halber an, dass die Zufriedenheit auf einer metrischen Skala gemessen wurde und die Daten etwa glockenförmig verteilt waren. Daher machen wir im Folgenden die Modellannahme, dass die Daten Realisierungen normalverteilter Zufallsvariablen sind. Die Zufriedenheit von je zehn befragten Studierenden sehen Sie in Abb. 9.6. Die mittlere Zufriedenheit an Ort B war bei den befragten Studierenden etwas höher als an Ort A. Andererseits wäre Ort A für Sie aber leichter zu erreichen, und so fragen Sie sich: Kann dieses Ergebnis vielleicht auch durch Zufall zustande gekommen sein, auch wenn alle Beobachtungen aus derselben (Normal-)Verteilung stammen?

Zur Beantwortung dieser Frage treffen wir wiederum eine Normalverteilungsannahme. Speziell nehmen wir an, dass die Beobachtungen Realisierungen von unabhängigen und normalverteilten Zufallsvariablen $X_1, \dots, X_m, Y_1, \dots, Y_n$ mit gleicher positiver Varianz sind und dass die X_i alle den gleichen Erwartungswert μ_x und die Y_i alle den gleichen

Abb. 9.6 Ungepaarte Beobachtungen



Erwartungswert μ_y besitzen. Wir testen dann die Nullhypothese, dass alle Beobachtungen aus der gleichen Normalverteilung stammen, oder kurz, dass $\mu_x = \mu_y$. Die *Zweistichproben-t-Statistik* wird daher nichts anderes tun, als die Mittelwerte in den beiden Gruppen (geeignet normiert) zu vergleichen.

Da die Varianz als gleich angenommen wird, wird sie global über beide Stichproben geschätzt. Wir betrachten daher für $\mathbf{x} \in \mathbb{R}^m$ und $\mathbf{y} \in \mathbb{R}^n$, zusammengeführt im Vektor $\mathbf{z} \in \mathbb{R}^{m+n}$, die *gepoolte Stichprobenvarianz*

$$s_p^2(\mathbf{z}) := \frac{1}{m+n-2} \left(\sum_{i=1}^m (x_i - \bar{x}_n)^2 + \sum_{j=1}^n (y_j - \bar{y}_n)^2 \right) = \frac{(m-1)s(\mathbf{x}) + (n-1)s(\mathbf{y})}{m+n-2}. \quad (9.5)$$

Die gepoolte Stichprobenstandardabweichung s_p ist eine Statistik, die die Variabilität *aller* Beobachtungen \mathbf{z} um ihre jeweiligen Mittelwerte beschreibt. Diese gepoolte Stichprobenstandardabweichung tritt in der Zweistichproben-t-Statistik auf.

Lemma 9.11 (Verteilung der Zweistichproben-t-Statistik)

Es seien $m, n \geq 2$ und $\mathfrak{Z} = (X_1, \dots, X_m, Y_1, \dots, Y_n)^t$ ein Zufallsvektor mit unabhängigen Komponenten, wobei $X_i \sim N(\mu_x, \sigma^2) \forall i = 1, \dots, m$ und $Y_j \sim N(\mu_y, \sigma^2) \forall j = 1, \dots, n$ mit $\mu_x, \mu_y \in \mathbb{R}$ und $\sigma^2 > 0$. Dann ist

$$T(\mathfrak{Z}) := \frac{(\bar{X}_m - \bar{Y}_n) - (\mu_x - \mu_y)}{\sqrt{(1/m) + (1/n)} s_p(\mathfrak{Z})} \sim t(m+n-2).$$

Die Statistik T heißt die Zweistichproben-t-Statistik unter Annahme der Varianzhomogenität, d. h., wenn wie in Lemma 9.11 alle Komponenten von \mathfrak{Z} die gleiche Varianz haben.

Beweis Es ist zu zeigen, dass

$$T(\mathfrak{Z}) \sim \frac{Z}{\sqrt{W/(m+n-2)}},$$

mit Z und W unabhängig und $Z \sim N(0, 1)$ und $W \sim \chi^2(m+n-2)$. Aufgrund der Unabhängigkeit und der Normalverteilung folgt $\bar{X}_n \sim N(\mu_x, \sigma^2/n)$ und $\bar{Y}_n \sim N(\mu_y, \sigma^2/n)$, sodass

$$\bar{X}_n - \bar{Y}_n \sim N\left(\mu_x - \mu_y, \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right),$$

und damit durch Standardisierung

$$Z := \frac{\bar{X}_m - \bar{Y}_n - (\mu_x - \mu_y)}{\sqrt{(1/m) + (1/n)} \sigma} \sim N(0, 1).$$

Nun bleibt zu zeigen, dass

$$W := \frac{s_p^2(\mathfrak{Z})}{\sigma^2}(n+m-2) \sim \chi^2(n+m-2) \quad \text{und } Z, W \text{ unabhängig.}$$

Wir argumentieren wie im Einstichproben- t -Test mit der Geometrie der mehrdimensionalen Standardnormalverteilung: Wir standardisieren die X_i via $Z_i := (X_i - \mu_x)/\sigma$ für alle $i = 1, \dots, m$. Damit ist $\mathfrak{Z}_x := (Z_1, \dots, Z_m)^t \sim N_m(0, E_m)$, und wir finden nach Satz 9.8, dass

$$\frac{1}{\sigma^2} \sum_{i=1}^m (X_i - \bar{X}_m)^2 = \sum_{i=1}^m (Z_i - \bar{Z}_m)^2 = \|\mathcal{P}_{\mathcal{D}^\perp} \mathfrak{Z}_x\|^2 \sim \chi^2(m-1), \quad (9.6)$$

wobei $\mathcal{D} = \{c\mathbb{1} \mid c \in \mathbb{R}\}$ wieder die Diagonale im \mathbb{R}^m beschreibt. Der Ausdruck ist wegen der Orthogonalität der zugehörigen Untervektorräume zudem unabhängig von \bar{X}_m . Analog folgt

$$\frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 \sim \chi^2(n-1) \quad (9.7)$$

und unabhängig von \bar{Y}_n . Zudem sind (9.6) und (9.7) unabhängig, da alle Komponenten von \mathfrak{Z} unabhängig sind, sodass

$$W = \frac{s_p^2(\mathfrak{Z})}{\sigma^2}(n+m-2) = \frac{1}{\sigma^2} \left[\sum_{i=1}^m (X_i - \bar{X}_m)^2 + \sum_{j=1}^n (Y_j - \bar{Y}_n)^2 \right] \sim \chi^2(m+n-2),$$

und Z und W unabhängig.

Lemma 9.12 (Zweistichproben-t-Test) Für $m, n \geq 2$ sei ein statistisches Modell gegeben durch einen Zufallsvektor $\mathfrak{Z} = (X_1, \dots, X_m, Y_1, \dots, Y_n)^t$ mit unabhängigen Komponenten, wobei $X_i \sim N(\mu_x, \sigma^2) \forall i = 1, \dots, m$ und $Y_j \sim N(\mu_y, \sigma^2) \forall j = 1, \dots, n$ mit $(\mu_x, \mu_y, \sigma^2) \in \Theta := \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$. Es sei s_p wie in (9.5). Zudem sei

$$d := \mu_x - \mu_y$$

und $\alpha \in (0, 1)$ sowie q_α das α -Quantil der $t(m+n-2)$ -Verteilung. Dann gelten folgende Aussagen:

1. Students Zweistichproben-Konfidenzintervall:

$$I(\mathbf{z}) := \left[(\bar{x}_m - \bar{y}_n) - q_{1-\frac{\alpha}{2}} \cdot s_p(\mathbf{z}) \sqrt{\frac{1}{m} + \frac{1}{n}}, (\bar{x}_m - \bar{y}_n) + q_{1-\frac{\alpha}{2}} \cdot s_p(\mathbf{z}) \sqrt{\frac{1}{m} + \frac{1}{n}} \right]$$

ist ein $(1 - \alpha)$ -Konfidenzintervall für d .

2. Students Zweistichproben-t-Test: Für $d^{(0)} \in \mathbb{R}$ sei eine Nullhypothese gegeben durch

$$H_0 : (\mu_x, \mu_y, \sigma^2) \in \{(\mu_x, \mu_y, \sigma^2) \in \Theta \mid \mu_x - \mu_y = d^{(0)}\}.$$

Dann ist die Zweistichproben-t-Statistik T via

$$T(\mathbf{z}) := \frac{(\bar{x}_m - \bar{y}_n) - d^{(0)}}{\sqrt{(1/m) + (1/n)} \cdot s_p(\mathbf{z})}$$

eine Teststatistik für einen exakten Test der Nullhypothese H_0 zum Niveau α mit Ablehnungsbereich $\mathcal{R}(\alpha) = (-\infty, q_{\alpha/2}] \cup [q_{1-\alpha/2}, \infty)$.

Typischerweise ist man in der Praxis an der Nullhypothese $d^{(0)} = 0$, d. h. der Gleichheit der Verteilungen, interessiert. Sofern nicht anders ausgeführt, werden wir dies im Folgenden immer voraussetzen. Zudem gilt wieder $T(\mathbf{z}) \in \mathcal{R}(\alpha)$ genau dann, wenn $I(\mathbf{z}) \not\supset d^{(0)}$.

Datenanalyse Im Beispiel aus Abb. 9.6 finden wir eine mittlere Zufriedenheit von etwa $\bar{x}_m \approx 4.1$ an Ort A und etwa $\bar{y}_n \approx 4.7$ an Ort B, mit einer gepoolten Standardabweichung von etwa $s_p(\mathbf{z}) \approx 0.8$. Mit je zehn Beobachtungen pro Gruppe ergibt das etwa $T(\mathbf{z}) \approx -1.7$ und einen P -Wert von ca. $P(\mathbf{z}) \approx 0.11$. Die Nullhypothese kann also auf dem 5 %-Niveau nicht abgelehnt werden – ein solch extremes oder ein noch extremeres Ergebnis könnte unter der Nullhypothese in etwa einem von zehn Fällen durch Zufall vorkommen.

Bemerkung zur Homoskedastizität

In die Herleitung der Verteilung der t -Statistik in der Zweistichprobensituation (Lemma 9.11) haben wir in der Standardisierung die Gleichheit der Varianzen in beiden Gruppen, auch *Homoskedastizität* genannt, ausgenutzt. Theoretisch bietet eine Anpassung von Welch (1947) in der Situation unterschiedlicher Varianzen ein alternatives Verfahren zu Students t -Test aus Lemma 9.12.

Sind in der Situation von Lemma 9.11 verschiedene Varianzen σ_x^2 und σ_y^2 in den Gruppen möglich, so muss die gepoolte Stichprobenvarianz s_p^2 (9.5) ersetzt werden durch eine separate Schätzung der Varianzen in den Gruppen, die erst danach zusammengefasst werden. Man definiert für die Skalierung der t -Statistik $s_u^2(\mathbf{z}) := s(\mathbf{x})^2/m + s(\mathbf{y})^2/n$, denn

$$\mathbb{V}ar(\bar{X}_m - \bar{Y}_n) = \mathbb{V}ar(\bar{X}_m) + \mathbb{V}ar(\bar{Y}_n) = \frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}.$$

In der Tat lässt sich im Setting ungleicher Varianzen zeigen, dass

$$T(\mathfrak{Z}) := \frac{(\bar{X}_m - \bar{Y}_n) - (\mu_x - \mu_y)}{s_u(\mathfrak{Z})}$$

approximativ $t(v)$ -verteilt ist, mit

$$v = \frac{(s(\mathbf{x})^2/m + s(\mathbf{y})^2/n)^2}{s^4(\mathbf{x})/(m^2(m-1)) + s^4(\mathbf{y})/(n^2(n-1))}$$

Freiheitsgraden (Welch 1947). Vom theoretischen Standpunkt aus erleichtert uns also die Annahme der Homoskedastizität das Leben sehr. Die Gleichheit der Varianzen wird auch im normalen linearen Modell eine wesentliche Modellannahme sein.

9.5 Dialog: Gepaarte und ungepaarte Tests

Ein Mitarbeiter einer Umweltorganisation (U) interessiert sich dafür, ob durch Kläranlagen ein bestimmter Schadstoff in ein Gewässer eingeleitet wird oder nicht. Da er keinen Zugriff auf die Ausleitungsrohre der Kläranlagen hat, misst er im Fließgewässer bei verschiedenen Kläranlagen jeweils vor und nach dem Einleiter des Kläranlagenwassers die Konzentration des Schadstoffs im Wasser. Bei der statistischen Analyse wird er allerdings unsicher und fragt einen Bekannten, der mittlerweile Mathematik studiert (M).

U: Kannst du mir vielleicht helfen? Ich habe hier zwei Stichproben, einmal vor, einmal nach der Kläranlage. Die jeweiligen Verteilungen sehen etwa glockenförmig aus. Damit darf ich doch den t -Test machen, oder? Ich bin nur ein bisschen unsicher, welchen ich da nehmen soll. Mache ich den ungepaarten t -Test, wird das Ergebnis nicht signifikant.

Mache ich den gepaarten, wird es signifikant. Was stimmt denn nun? Leitet die Kläranlage den Stoff ein oder nicht?

An dieser Stelle hakt der Mathematikstudent direkt ein.

M: Moment mal. Als Erstes ist es ganz wichtig, sich darüber im Klaren zu sein, dass wir nie sicher sagen können, was in Wirklichkeit los ist – egal, was beim Test rauskommt! Das Einzige, was wir tun können, ist zu schauen, wie gut die Daten zur Nullhypothese passen.

U: Gut, aber die Nullhypothese ist doch dieselbe bei beiden Tests. Wie kommt es dann, dass die Daten einmal passen und einmal nicht? Das klingt doch stark nach statistischer Willkür!

Aber natürlich gibt es für die Auswahl des Verfahrens gute Gründe, die der Mathematiker sogleich zu erklären beginnt.

M: Lass uns mal eins nach dem anderen durchgehen. Zuerst müssen wir die Voraussetzungen der Tests checken. Die Beobachtungen sehen etwa glockenförmig aus, damit sind wir mit dem t -Test wohl erstmal gut bedient. Jetzt gibt es aber noch mehr Voraussetzungen, die wir beachten müssen. Zum Beispiel wird beim ungepaarten t -Test angenommen, dass die Beobachtungen der beiden Stichproben unabhängig sind. Lass uns das für dein Beispiel mal prüfen...

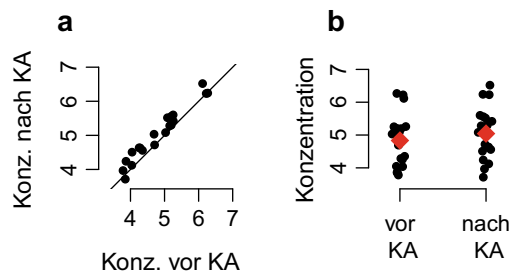
Rasch entwirft der Mathematikstudent etwa Abb. 9.7. Jetzt benötigt er aber noch zusätzliche Informationen.

M: Bei jeder Kläranlage hast du genau zweimal gemessen, richtig?

U: Ja genau, erst eine Probe vor ihrer Einleitung und dann eine Probe nach ihrer Einleitung.

M: Sehr gut! Damit sind deine Stichproben *gepaart*. Schau mal in Bild a, jeder Punkt ist eigentlich ein Paar von Beobachtungen, mit einer x - und einer y -Koordinate. Ist ein Messwert vor der Kläranlage hoch (großes x), so ist er auch oft nach der Kläranlage hoch (großes y), und genauso sieht es bei kleinen Messwerten aus. Statistisch muss man daher wohl eine gewisse Korrelation zwischen den Messungen berücksichtigen. Die Annahme der Unabhängigkeit deiner beiden Stichproben ist hier also eher nicht sinnvoll – der ungepaarte t -Test scheint also nicht angebracht.

Abb. 9.7 Darstellung gepaarter Daten mit (a) und ohne (b) Paarungsstruktur



Der Mitarbeiter der Umweltorganisation ist immer noch ein bisschen verunsichert und hakt nach.

U: Ok, soweit so gut. Aber wieso sagst Du, das sei ‚sehr gut‘? Was wäre denn zum Beispiel passiert, wenn ich bei einigen Kläranlagen nur Proben vor der Einleitung und bei anderen Kläranlagen nur Proben nach der Einleitung genommen hätte?

M: Dann wären deine Stichproben ungepaart, und wir könnten sie gar nicht im Streudiagramm darstellen, weil nicht jeder x -Wert genau einen y -Wert hätte. Wir bräuchten etwa Abbildung b, und das wäre dann ein Fall für den ungepaarten t -Test.

U: Und wäre der dann signifikant geworden?

Da muss der Mathematikstudent ihn enttäuschen.

M: Vermutlich eher nicht. Schau dir mal Abbildung b an. Das sind dieselben Daten, aber die Paarungsstruktur, also dass jede Messung links genau einen Partner rechts hat, wird nicht berücksichtigt. In die Zweistichproben- t -Statistik gehen quasi die Standardabweichungen der Beobachtungen in beiden Gruppen ein, und zwar im Nenner, und da diese im Vergleich zur Differenz der Mittelwerte recht groß sind, wird die t -Statistik keinen besonders extremen Wert annehmen. Eigentlich müsste man im Nenner zusätzlich noch die Korrelation berücksichtigen, aber das tut man beim ungepaarten t -Test nicht, weil man annimmt, dass die Stichproben unabhängig sind.

U: Und was ist anders beim gepaarten t -Test?

M: Beim gepaarten t -Test wird das Problem der Abhängigkeiten zwischen den Stichproben geschickt umgangen. Denn man betrachtet nur die Differenz der Messungen bei jedem einzelnen Datenpaar. Damit bereinigt man automatisch um die großen Schwankungen innerhalb der Gruppen. Manche Flüsse haben vielleicht grundsätzlich eine höhere Schadstoffkonzentration als andere, was dann zu großer Variabilität innerhalb der Gruppen führen kann. Durch den Übergang zu den Differenzen konzentriert man sich wieder auf das Wesentliche, nämlich den Unterschied zwischen vor und nach der Einleitung. Cool, oder?

U: Total! Da kann ich ja froh sein, dass ich gepaart gemessen habe!

M: Ja absolut! Deswegen hilft es auch, wenn du deinen Versuchsaufbau schon vorher mal mit einem Statistiker besprichst. Dann kannst du besser einschätzen, welche Konsequenzen dein Vorgehen später in der Auswertung haben kann.