

Validität und Trennschärfe statistischer Tests in Abhängigkeit der Effektstärke einer Störvariable: Grundlagen

Jonas Cedro Delgado

Georg-August-Universität Göttingen

4. Dezember 2019

- 1 Grundbegriffe
- 2 Gepaarter t -Test
- 3 Zweistichproben- t -Test
- 4 Literatur

Section 1

Grundbegriffe

Validität vs. Trennschärfe eines Tests

Störfaktor:

Ein Störfaktor (z. B. Bodenart) ist ein Faktor, welcher sowohl die abhängige Variable (z. B. Dünger) als auch die unabhängige Variable (z. B. Ertrag) beeinflussen könnte und nicht manipuliert werden kann. Störfaktoren können Merkmale von Untersuchungseinheiten oder äußere Faktoren sein. Störfaktoren sind als alternative oder konkurrierende Erklärungen zur Ausgangshypothese des Forschungsproblems zu sehen. Zur Kontrolle von Störfaktoren kann paarweise Zuordnung (Matching), oder Randomisierung angewandt werden.

- Sind die Versuchseinheiten sehr unterschiedlich, dann wird die Isolierung interessierender Effekte (z. B. Ertrag einer Nutzpflanze) durch die Heterogenität des Materials (z. B. Bodenart) erschwert
- Es empfiehlt sich Untergruppen von Untersuchungseinheiten zu bilden, die in sich gleichförmiger sind als das gesamte Material \Rightarrow sog. *homogene Versuchsblöcke*

Blockbildung:

Die Bildung von Blöcken homogener Untersuchungseinheiten wird *Blockbildung* (Blocking) genannt. Innerhalb eines Blockes sind die Bedingungen homogener als Außerhalb und der Versuchsfehler wird i. A. kleiner. Durch Blocking können andere Effekte innerhalb der Blöcke verglichen werden; der Blockeffekt kann von anderen Effekten isoliert werden. Blöcke können zum Beispiel nebeneinanderliegenden Parzellen eines Feldes in einem landwirtschaftlichen Feldversuch sein.

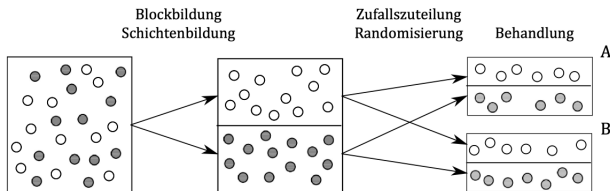
Randomisierung (Zufallszuteilung):

Unter Verwendung eines Zufallsmechanismus werden Untersuchungseinheiten unterschiedlichen Gruppen zugeordnet. Ziel ist das gleichmäßige Verteilen bekannter und unbekannter an die Untersuchungseinheiten gebundene Störfaktoren auf Experimental- und Kontrollgruppe. Die Randomisierung soll die Zufälligkeit der Stichprobe im Sinne der mathematischen Statistik garantieren, entsprechend der Auswahl einer Zufallsstichprobe.

Prinzipien der Versuchsplanung: Blockbildung und Randomisierung

Die Blockbildung und die Randomisierung sind zwei der drei Prinzipien der Versuchsplanung.¹ Beide dienen dazu, für alle anderen Einflussgrößen zu kontrollieren, sie also konstant zu halten.

1/Blockbildung.png



¹Das Dritte Prinzip, die Wiederholung ist hier nicht relevant

Paarweise verbundener Versuchsplan:

Ein *paarweise verbundener Versuchsplan* (paarweise verbundenes Design) ist ein Spezialfall eines *randomisierten Blockversuchsplans* (randomisiertes Blockdesign). Ein paarweise verbundener Versuchsplan kann verwendet werden, wenn das Experiment nur zwei Behandlungsbedingungen hat und die Untersuchungseinheiten basierend auf einer Blockbildungsvariable in Paare zusammengefasst werden können. Die Untersuchungseinheiten werden dann innerhalb jedes Paares unterschiedlichen Behandlungsarten zufallszugeteilt (sog. Randomisierung)

Paarweise verbundener Versuchsplan: Beispiel

Beispiel:

Bei einer medizinischen Studie mit 1000 Versuchsteilnehmern bekommen alle Versuchsteilnehmer eine von zwei möglichen Behandlungsarten (Placebo oder Impfstoff). Die 1000 Teilnehmer sind in 500 verbundene Paare gruppiert. Jedes Paar ist auf Geschlecht und Alter abgestimmt. Zum Beispiel könnte Paar 1 zwei Frauen im Alter von 21 Jahren sein. Paar 2 könnte zwei Männer im Alter von 21 Jahren sein. Paar 3 könnte zwei Frauen im Alter von 22 Jahren sein und so weiter. Der paarweise verbundene Versuchsplan verwendet Randomisierung um für Störfaktoren zu kontrollieren. Hier kontrolliert der paarweise verbundene Versuchsplan für die zwei potenziellen Störfaktoren Alter und Geschlecht. Da der Versuchsplan paarweise verbunden ist kann er mittels eines gepaarten t -Tests ausgewertet werden.

Verbundene Stichproben

Im verbundenen Zweistichprobenproblem werden bei jeder Untersuchungseinheit beide Verfahren betrachtet. Die Daten fallen also paarweise an. Es sei x_i der Wert, der bei der i -ten Untersuchungseinheit beim ersten Verfahren beobachtet wird (z. B. Gabe eines spezifischen Düngers) und y_i der Wert, der bei der i -ten Untersuchungseinheit beim zweiten Verfahren beobachtet wird. Bei der i -ten Untersuchungseinheit wird also das Paar (x_i, y_i) beobachtet (Realisierung der zweidimensionalen Zufallsvariablen (X_i, Y_i)). Es soll überprüft werden, ob sich die beiden Verfahren unterscheiden. Besteht kein Unterschied zwischen den beiden Verfahren, sollte

$$E(X_i) = E(Y_i)$$

bzw.

$$E(D_i) = E(X_i - Y_i) = E(X_i) - E(Y_i) = 0.$$

(Die Verteilung der Differenzen D_i sollte das Zentrum null besitzen) gelten

Section 2

Gepaarter t -Test

Gepaarter t -Test (verbundene Stichproben)

Der gepaarte t -Test testet im Falle, dass die Beobachtungen paarweise verbunden vorliegen, ob die Differenz von zwei Mittelwerten der Grundgesamtheit

$$\mu_D = \mu_X - \mu_Y$$

verschieden von einem spezifischen Wert d_0 ist.

Gepaarter t -Test (verbundene Stichproben)

Seien x_1, x_2, \dots, x_n und y_1, y_2, \dots, y_n zwei verbundene Stichproben, die Realisierungen von $X_1, X_2, \dots, X_n \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(\mu_x, \sigma_x^2)$ und $Y_1, Y_2, \dots, Y_n \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(\mu_y, \sigma_y^2)$ darstellen. Es sei weiterhin angenommen, dass die Daten auf einer Intervall- oder Verhältnisskala gemessen wurden. Definiere $D_i = Y_i - X_i$, $i = 1, \dots, n$. Sind (X_i, Y_i) zweidimensional normalverteilt, dann gilt $D_i \sim \mathcal{N}(\mu_d, \sigma_d^2)$, wobei $\mu_d = \mu_y - \mu_x$ und $\sigma_d^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$, mit $\sigma_{xy} = \text{Cov}(X_i, Y_i)$. Die wahre Varianz σ_d^2 der Differenz wird durch die Stichprobenvarianz der Differenzen S_D^2 geschätzt.

Teststatistik (verbundene Stichproben)

Für die Konstruktion der Teststatistik nutzt man aus, dass sich eine t -verteilte Zufallsvariable aus der Division einer standardnormalverteilten Zufallsvariable durch die Quadratwurzel einer (um die Anzahl ihrer Freiheitsgrade korrigierten) Chi-Quadrat-verteilten Zufallsvariablen ergibt, d. h.

$$t_n \equiv \frac{\mathcal{N}(0, 1)}{\sqrt{\chi_n^2/n}}.$$

Die Teststatistik für den gepaarten t -Test ist gegeben durch:

$$T = \sqrt{n} \frac{\bar{D} - d_0}{S_D} \stackrel{H_0}{\sim} t_{n-1},$$

wobei die Stichprobenstandardabweichung und das Stichprobenmittel gegeben sind durch $S_D = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2}$ und $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$

Hypothesenspezifikation (verbundene Stichproben)

Es ist möglich einen zweiseitigen, rechtsseitigen und linksseitigen Hypothesentest zu spezifizieren:

- $H_0: \mu_d = d_0$ vs. $H_1: \mu_d \neq d_0$
- $H_0: \mu_d \leq d_0$ vs. $H_1: \mu_d > d_0$
- $H_0: \mu_d \geq d_0$ vs. $H_1: \mu_d < d_0$

mit den dazugehörigen Ablehnkriterien und Ablehnbereichen A :

- $|T| > t_{n-1, \alpha/2}$ und $A = (-\infty, -t_{1-\frac{\alpha}{2}; n-1}] \cup [t_{1-\frac{\alpha}{2}; n-1}, \infty)$
- $T > t_{n-1, 1-\alpha}$ und $A = [t_{1-\alpha; n-1}, \infty)$
- $T < t_{n-1, \alpha}$ und $A = (-\infty, -t_{1-\alpha; n-1}]$

Vor- und Nachteile eines gepaarten t -Tests

- bietet den Vorteil, dass das Ergebnis des t -Test valide ist (d. h. sein Ergebnis ist realitätsnah bzw. der Einfluss des Störfaktors wird „herausgerechnet“)
- Nachteil ist der Verlust an Freiheitsgraden. Da die Anzahl der Freiheitsgrade aufgrund der paarweisen Zuordnung geringer ist kann die Trennschärfe (Power) des Tests sinken (echte Unterschiede der Mittelwerte werden mit einer geringeren Wahrscheinlichkeit gefunden)
- Eine Möglichkeit, um die Trennschärfe des Tests zu erhöhen besteht in einem vollen Ausschöpfen der Freiheitsgrade, da mit steigender Anzahl der Freiheitsgrade die Trennschärfe des Tests erhöht wird
- Ein Test der die Freiheitsgrade voll ausschöpft (aber bzgl. der Validität als schlechter zu bewerten ist) ist der Zweistichproben- t -Test für unverbundene Stichproben

Section 3

Zweistichproben- t -Test

Zweistichproben- t -Test (unverbundene Stichproben)

Seien $X_1, X_2, \dots, X_n \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(\mu_x, \sigma_x^2)$ und $Y_1, Y_2, \dots, Y_m \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(\mu_y, \sigma_y^2)$ zwei unabhängige Zufallsstichproben. Es sei zunächst angenommen, dass $\sigma_x = \sigma_y = \sigma^2$. Dann gilt $X_i - Y_i \stackrel{H_0}{\sim} \mathcal{N}(\mu_x - \mu_y, 2\sigma^2)$. Da $\bar{X} \sim \mathcal{N}(\mu_x, \sigma^2/n)$ und $\bar{Y} \sim \mathcal{N}(\mu_y, \sigma^2/m)$ geeignete Schätzer für μ_x und μ_y sind und da $\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_x - \mu_y, \sigma^2/n + \sigma^2/m)$ ist die Teststatistik gegeben durch:

$$T = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \stackrel{H_0}{\sim} t_{n+m-2}$$

Hierbei wird die Größe

$$S_p^2 = \frac{1}{n+m-2} \left(\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2 \right)$$

als *gepoolte Stichprobenvarianz* bezeichnet, da sie die Varianzen aus verschiedenen Stichproben kombiniert, indem sie ihr gewichtetes Mittel verwendet, um die Gesamtvarianz zu erhalten.

Hypothesenspezifikation (unverbundene Stichproben)

Es ist möglich einen zweiseitigen, rechtsseitigen und linksseitigen Hypothesentest zu spezifizieren:

- $H_0: \mu_x = \mu_y$ vs. $H_1: \mu_x \neq \mu_y$
- $H_0: \mu_x \leq \mu_y$ vs. $H_1: \mu_x > \mu_y$
- $H_0: \mu_x \geq \mu_y$ vs. $H_1: \mu_x < \mu_y$

mit den dazugehörigen Ablehnkriterien und Ablehnbereichen A :

- $|T| > t_{n+m-2, \alpha/2}$ und $A = (-\infty, -t_{1-\frac{\alpha}{2}; n+m-2}] \cup [t_{1-\frac{\alpha}{2}; n+m-2}, \infty)$
- $T > t_{n+m-2, 1-\alpha}$ und $A = [t_{1-\alpha; n+m-2}, \infty)$
- $T < t_{n+m-2, \alpha}$ und $A = (-\infty, -t_{1-\alpha; n+m-2}]$

Die Testentscheidung ergibt sich wie gewohnt aus dem Vergleich mit dem kritischen Wert.

Section 4

Literatur

- Taeger, Dirk, und Sonja Kuhnt: *Statistical hypothesis testing with SAS and R*. Chichester, UK, Wiley, 2014. pp. 23
- Lothar Sachs, Jürgen Hedderich: *Angewandte Statistik: Methodensammlung mit R*. 8., überarb. und erg. Auflage. Springer Spektrum, Berlin/ Heidelberg 2018.
- Joseph Adler: *R in a nutshell: A desktop quick reference.*, O'Reilly Media, Inc., 2010.
- Handl, Andreas und Torben Kuhlenkasper: *Einführung in die Statistik: Theorie und Praxis mit R*. Springer-Verlag, 2018.