Institut für Volkswirtschaftslehre (ECON)
Lehrstuhl für Ökonometrie und Statistik

Kapitel XIII - p-Wert und Beziehung zwischen Tests und Konfidenzintervallen

Induktive Statistik

Prof. Dr. W.-D. Heller
Hartwig Senska
Carlo Siebenschuh

Testprobleme

Vorgehensweise bei Testproblemen:

1. Festlegung des Niveaus α

2. Bestimmung eines Tests zum Niveau α entsprechend dem Schema

- stetiger Fall: Niveau α kann exakt eingehalten werden
- diskreter Fall: Niveau (Maximalwert der Fehlerwahrscheinlichkeit 1. Art) ergibt sich aus der Testschranke

3. Ziehung der Stichprobe, Berechnung der Testgröße und Feststellung der Entscheidung

p-Wert (bzw. Signifikanzniveau)

Die Auswertung von Stichprobenergebnissen zur Entscheidung in Testsituationen erfordert entsprechend der obigen Vorgehensweise die Festlegung eines Niveaus.

Soll die Entscheidung durch ein Computerprogramm ermittelt werden, ist vorher die Eingabe des Niveaus erforderlich.

Übliche Vereinfachung bei kommerzieller Software:

Ausgabe des Niveaus, bei dem die Nullhypothese *gerade noch* nicht abgelehnt wird.

Je höher das vorgegebene Niveau, also die zugelassene Fehlerwahscheinlichkeit

1. Art, desto kleiner der Annahmebereich.

gerade noch bedeutet also:

Der p-Wert ist Lösung der Optimierungsaufgabe

Maximiere das Niveau zum vorliegenden Stichprobenergebnis unter der **Nebenbedingung: Entscheidung** d_0

Beispiel:

Monatlicher Umsatz einer Filiale einer Lebensmittelkette sei normalverteilt mit Mittelwert μ und bekannter Standardabweichung von 10000 EURO.

Es soll getestet werden, ob der Umsatz gegenüber dem früheren Mittelwert von 100000 EURO gestiegen ist oder die Abweichung als rein zufällig angesehen werden muss.

Eine Stichprobe vom Umfang 150 ergab einen Mittelwert von 103000 EURO.

Hypothesen:

$$H_0 : \mu \leq \mu_0$$

gegen

$$H_1 : \mu > \mu_0$$

Gleichmäßig bester Test zum Niveau α:

$$\delta(x) = \begin{cases} d_0 & \bar{x} \leq \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha} \\ d_1 & \bar{x} > \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha} \end{cases}$$

mit

$$\mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha} = 100000 + \frac{10000}{\sqrt{150}} u_{1-\alpha}$$

Beispiel: Für $\alpha = 0.05$ ist $u_{1-\alpha} = 1.645$ und die Testschranke ist damit 101343. Zu diesem Niveau wird die Gegenhypothese angenommen.

Die Nullhypothese wird nicht abgelehnt, wenn 103000 nicht größer als

die Testschranke ist:

$$103000 \leq 100000 + \frac{10000}{\sqrt{150}} u_{1-\alpha}$$

$$\text{oder } u_{1-\alpha} \geq \frac{3}{10}\sqrt{150} = 3.674$$

Das größte α mit dieser Eigenschaft ist

$$\alpha = 1 - \Phi(3.674) = 1 - 0.9999 = 0.0001$$

Der p-Wert ist also 0.0001.

Entscheidung auf der Basis des p-Werts:

Vorgabe des Niveaus prinzipiell vor der Entnahme der Stichprobe bzw. der Durchführung der Experimente, spätestens aber vor der Auswertung durch das Programm.

Analyse des berechneten p-Werts:

$\alpha \leq p$-Wert $\Rightarrow$ Nullhypothese wird nicht abgelehnt (d_0)

$\alpha > p$-Wert $\Rightarrow$ Gegenhypothese wird angenommen (d_1)

Im Beispiel: $\alpha = 0,05$, p-Wert$=0.0001$ $\Rightarrow$ Gegenhypothese wird angenommen

Interpretation des p-Werts:

- Ein kleiner p-Wert ist ein Indiz für die Gültigkeit der Gegenhypothese
- Ein hoher p-Wert ist **kein** Argument für die Gültigkeit der Nullhypothese

Begründung:

a) Der p-Wert hängt ab vom zufälligen Stichprobenergebnis, ist also selbst zufällig.

b) Die Verteilung des p-Werts ist zu Testproblemen für stetige Zufallsvariablen bei Gültigkeit der Nullhypothese in der Regel die Gleichverteilung auf dem Intervall [0,1]. Damit hat in diesem Fall jeder Wert zwischen 0 und 1 dieselbe Chance. Große p-Werte kommen also bei Gültigkeit der Nullhypothese nicht häufiger vor als kleine Werte.

$\Rightarrow$ p-Wert basierte Entscheidung beruht auf Konvention.

Beziehung zwischen Tests und Konfidenzintervallen

Ein Vergleich des Tests zum Niveau α beim zweiseitigen Testproblem für den Erwartungswert der Normalvertelung (Varianz bekannt)

$$\delta_c^*(x) = \begin{cases} d_0 & \mu_0 - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \leq \bar{x} \leq \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \\ d_1 & \text{sonst} \end{cases}$$

mit der Realisation des $(1-\alpha)-$Konfidenzintervalls

$$[\bar{x} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}, \bar{x} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}]$$

für den Erwartungswert einer normalverteilten Grundgesamtheit mit bekannter Varianz zeigt, dass die Nullhypothese genau dann angenommen wird, wenn μ_0 im Konfidenzintervall liegt.

Dieses Prinzip lässt sich verallgemeinern:

Sei $I_\alpha(x)$ eine Realisation des $(1-\alpha)-$Konfidenzintervalls zum Stichprobenergebnis x für den Parameter γ einer Zufallsvariable $(\Gamma \subset \mathbb{R})$.

Für die zweiseitige Testsituation

$$H_0 : \gamma = \gamma_0$$

$$\text{gegen}$$

$$H_1 : \gamma \neq \gamma_0$$

ist

$$\delta(x) = \begin{cases} d_0 & \gamma_0 \in I_\alpha(x) \\ \\ d_1 & \gamma_0 \notin I_\alpha(x) \end{cases}$$

eine Entscheidungsfunktion.

Fehlerwahrscheinlichkeit 1.Art von δ:

$$P_I(\delta, \gamma_0) = P_{\gamma_0}(\gamma_0 \notin I_\alpha(X)) = \alpha$$

δ ist also ein Test zu Niveau α.

Umgekehrt:

Sei $\Gamma \subset \mathbb{R}$ der Parameterraum einer Verteilung.
$T(x)$ sei eine suffiziente (und vollständige) Statistik bezüglich $\gamma \in \Gamma$.
Zu jedem γ_0 sei

$$\delta_\alpha^{\gamma_0}$$

ein Test zum Niveau α für das zweiseitige Testproblem

$$H_0 : \gamma = \gamma_0$$

gegen

$$H_1 : \gamma \neq \gamma_0$$

mit der Testgröße $T(x)$.

Der Annahmebereich von $\delta_\alpha^{\gamma_0}$ ist in der Regel ein Intervall $[a(\gamma_0), b(\gamma_0)]$, so dass gilt:

$$P_I(\delta_\alpha^{\gamma_0}, \gamma_0) = 1 - P_{\gamma_0}(a(\gamma_0) \leq T(X) \leq b(\gamma_0)) = \alpha$$

Gelingt es, die Ungleichung

$$a(\gamma_0) \leq T(x) \leq b(\gamma_0)$$

in eine Ungleichung für γ_0 umzuformen, haben wir ein $(1-\alpha)-$Konfidenzintervall für γ_0 gefunden.

Bemerkungen:

1. Jede Eigenschaft von Tests zum Niveau α lässt sich ausdrücken in einer Eigenschaft von $(1-\alpha)-$Konfidenzintervallen.
2. Bei einseitigen Tests ist statt eines Konfidenzintervalls eine Konfidenzschranke (bzw. eine Konfidenzhalbgerade) zu verwenden.