Induktive Statistik

Statistik, Parameter

Eine *Statistik* ist eine Maßzahl einer Stichprobe; *ein Parameter* eine Maßzahl der Grundgesamtheit:

Maßzahl	Statistik (Stichprobe)	Parameter (Grundgesamtheit)
Arithmetisches Mittel	$\bar{x}$	μ
Standardabweichung	s	σ
Anzahl der Elemente	n	N
Anteil	p	P

Stichprobenverteilung

Eine *Stichprobenverteilung* ist die Verteilung der Statistiken einer Stichprobe.

Stichprobenverteilung der Mittelwerte

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}}^2 = \begin{cases} \dfrac{\sigma^2}{n} & \text{mit Zurücklegen} \\ \dfrac{\sigma^2}{n} \cdot \left(\dfrac{N-n}{N-1} \right) & \text{ohne Zurücklegen} \end{cases}$$

Hier sind:

μ: Mittelwert der Grundgesamtheit

σ^2: Varianz der Grundgesamtheit

$\mu_{\bar{x}}$: Mittelwert der Stichprobenverteilung

$\sigma_{\bar{x}}^2$: Varianz der Stichprobenverteilung

Form der Stichprobenverteilung des Mittelwertes 1.

1.
Für eine *normalverteilte Grundgesamtheit* gilt: Die Verteilung des Stichprobenmittelwertes ist *normal* mit

$$\mu_{\bar{x}} = \mu, \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

2.
Für eine „große“ Stichprobe ($n \geq 30$) gewählt aus einer *beliebig verteilten Grundgesamtheit* gilt: Die Verteilung des Stichprobenmittelwertes ist *näherungsweise normal* mit

$$\mu_{\bar{x}} = \mu, \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

Intervallschätzung des Mittelwertes der Grundgesamtheit

Zusammenfassung

Grundgesamtheit	Stichproben-umfang	σ bekannt	σ unbekannt
Normalverteilt	Groß $(n \geq 30)$	$\bar{x} \pm z\sigma_{\bar{x}}$	$\bar{x} \pm ts_{\bar{x}}$ oder $\bar{x} \pm zs_{\bar{x}}$
	Klein $(n < 30)$	$\bar{x} \pm z\sigma_{\bar{x}}$	$\bar{x} \pm ts_{\bar{x}}$
Nicht normalverteilt	Groß $(n \geq 30)$	$\bar{x} \pm z\sigma_{\bar{x}}$	$\bar{x} \pm ts_{\bar{x}}$ oder $\bar{x} \pm zs_{\bar{x}}$
	Klein $(n < 30)$	Nichtparametrische Methoden werden benutzt.	

Hier sind:

n :	Stichprobenumfang
σ :	Standardabweichung der Grundgesamtheit
$\bar{x}$:	Mittelwert der Stichprobe
s :	Standardabweichung der Stichprobe
z :	Kritischer $z-$Wert
t :	Kritischer $t-$Wert

Nullhypothese, Alternativhypothese

Eine *Nullhypothese* ist eine Annahme über einen Parameter der Grundgesamtheit, die als wahr angenommen wird bis sie sich als falsch erweist. Sie wird mit H_0 bezeichnet.

Eine *Alternativhypothese* ist eine Annahme über einen Parameter der Grundgesamtheit, die wahr ist, wenn sich die Nullhypothese als falsch erweist. Sie wird mit H_1 bezeichnet.

Schematische Darstellung der Fehlertyps I und II

		Wirklichkeit	
		H_0 ist wahr	H_0 ist falsch
Schlussfolgerung	H_0 wird *nicht* abgelehnt	richtig	Fehler vom Typ II
	H_0 wird abgelehnt	Fehler vom Typ I	richtig

Schritte der Methode des kritischen Wertes

1. *Formuliere die Hypothesen*

2. *Wähle die*

   i. *Normalverteilung, falls σ bekannt*
   ii. *t-Verteilung, falls σ unbekannt*

3. *Bestimme die kritischen Werte z_{krit} (falls σ bekannt) bzw. t_{krit} (falls σ unbekannt) und damit die Ablehnungs- bzw. Annahmeregionen.*

4. *Berechne die z_{stat} bzw. t_{stat} :*

   i. $z_{stat} = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}}$, $\quad \sigma_{\bar{x}} := \dfrac{\sigma}{\sqrt{n}}$, $\quad$ *falls σ bekannt*

   ii. $t_{stat} = \dfrac{\bar{x} - \mu}{s_{\bar{x}}}$, $\quad s_{\bar{x}} = \dfrac{s}{\sqrt{n}}$ $\quad$ *falls σ unbekannt*

5. *Entscheidung:*
   Vergleiche z_{stat} mit z_{krit} bzw. t_{stat} mit t_{krit} und je nach dem lehne bzw. nicht lehne H_0 ab.

Schritte der p-Wert-Methode

1. *Formuliere die Hypothesen*

2. *Wähle die*

   i. *Normalverteilung, falls σ bekannt*
   ii. *t-Verteilung, falls σ unbekannt*

3. *Bestimme den p-Werte mit*

   i. $z_{stat} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}, \qquad \sigma_{\bar{x}} := \frac{\sigma}{\sqrt{n}},$ *falls σ bekannt*

   ii. $t_{stat} = \frac{\bar{x} - \mu}{s_{\bar{x}}} \qquad s_{\bar{x}} = \frac{s}{\sqrt{n}},$ *falls σ* un*bekannt*

4. *Entscheidung:*

   i. *Lehne H_0 ab, wenn $p - Wert < \alpha$*
   ii. *Lehne H_0 nicht ab, wenn $p - Wert \geq \alpha$*