Statistische Tests von Hypothesen

Fehler 1. Art und Fehler 2. Art:
(für eine Hypothese H_o)

	Wirklichkeit		
Testentscheidung		H_o wahr	H_o falsch
	H_o-Annahme	$1 - \alpha$	β Fehler 2. Art
	H_o-Ablehnung	α Fehler 1. Art	$1 - \beta$ Macht der Tests

(Gelegentlich werden auch β und 1-β „vertauscht".)
α wird vorgegeben
β normalerweise unbekannt, allerdings für einige gängige Tests errechenbar, z.B. mit dem Programm PASS

α=0.05 besagt: in 5 von 100 Tests wird H_0 angelehnt, obwohl es richtig ist.

Regeln: α -> 0 => β -> 1

β -> 0=> α -> 1

n -> oo => β -> 0 (bei konstantem α)

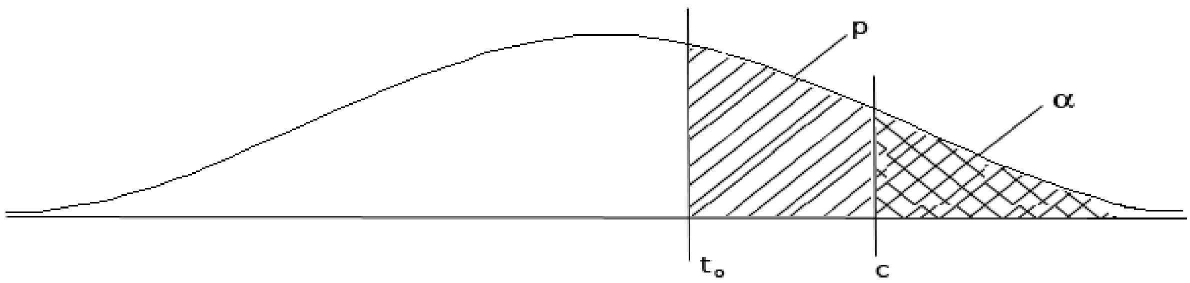
Daher wird empfohlen:

α groß (etwa α=0.10) bei kleinen Stichprobenumfängen (etwas n<30) (bei kleinen Stichproben ist sonst schwer, Unterschiede nachzuweisen)

α normal (etwas α=0.05) bei normalen Stichprobenumfängen (etwa n=30-200)

α klein (etwas α=0.01) bei großen Stichprobenumfängen (etwa n>200) (bei großen Stichproben sind sonst alle Tests signifkant, und es sollten relevante von signifikanten Ergebnissen unterschieden werden.)

P-Wert:
Sei t_o die auf Grund der Hypothese H_o errechnete Prüfgröße, für die folgende Verteilung angenommen wird:



Klassische Vorgehensweise: H_o-Annahme, wenn $t_o \leq c$

H_o-Ablehnung, wenn $t_o > c$

Vorgehensweise mit p-Wert: H_o-Annahme, wenn $p \geq \alpha$

H_o-Ablehnung, wenn $p < \alpha$

Begriffe zu statistischen Tests

- Robuster Test:
  Ein Test heißt *robust*, wenn die Wahrscheinlichkeiten α und β auch dann noch eingehalten werden, wenn die Voraussetzungen (leicht) verletzt sind. Häufig ist allerdings eine leichte Vergrößerung von β in Kauf zu nehmen.

- Konservativer Test:
  Ein Test A heiß *konservativ* (in Bezug auf Test B), wenn β (Test A) > β (Test B), also bei Test A häufiger H_o angenommen wird, obwohl H_o falsch ist, als bei Test B.

- Test A ist besser als Test B,
  wenn β (Test A) < β (Test B), d.h.
  wenn 1-β (Test A) > 1-β (Test B), d.h. die Macht von Test A größer als die von Test B ist, bei konstantem α und n.

- Effizienz eines (nicht-parametrischen) Tests:
  Unter Annahme bestimmter Voraussetzungen (z.B. Normalverteilung, Varianzhomogenität etc) läßt sich zu jeder Hypothese (z.B. $\mu_1 = \mu_2$, d.h. gleiche Mittelwerte zweier Stichproben) genau ein Test A herleiten (z.B. t-Test), der dann der beste ist (in o.a. Sinn). Wird unter schwächeren Voraussetzungen (z.B. beliebige Verteilung) zur gleichen Hypothese ein anderer Test B hergeleitet, so gilt β (Test A) < β (Test B). Wegen β -> 0 (für n -> oo) muß es für jedes β (Test A) zu einem n_A auch ein n_B ($n_B > n_A$) geben, so daß β (Test B (n_B)) = β (Test A (n_A)). D.h. wenn zur Erlangung eines bestimmten β bei dem (optimalen) Test A der Stichprobenumfang n_A erforderlich ist, dann kann durch Vergrößern des Stichprobenumfangs auf n_B auch mit dem (schlechteren nicht-parametrischen) Test B dasselbe β erzielt werden.
  Das Verhältnis n_B/n_A für n -> oo heißt (asymptotische) *Effizienz* des Tests B. 1/Effizienz gibt also an, wie stark die Stichprobe vergrößert werden muß, um mit einem nicht-parametrischen Test genau so gut dieselbe Hypothese testen zu können.