Mögliche Fehler beim Testen

Fehler 1. Art (Irrtumswahrscheinlichkeit α), Zusammenfassung:

- **Die Nullhypothese wird verworfen, obwohl sie zutrifft.**
  „*Wir haben uns blamiert, weil wir etwas Wahres abgelehnt haben.*“
- Dieser Fehler ist kontrollierbar.
- Signifikanzniveau:
  Obere Schranke für die Irrtumswahrscheinlichkeit.
- „***Statistische Signifikanz****:*
  In der Statistik heißen Unterschiede oder Zusammenhänge ***signifikant****, wenn die Wahrscheinlichkeit gering ist, dass sie durch Zufall zustande gekommen sind.*“

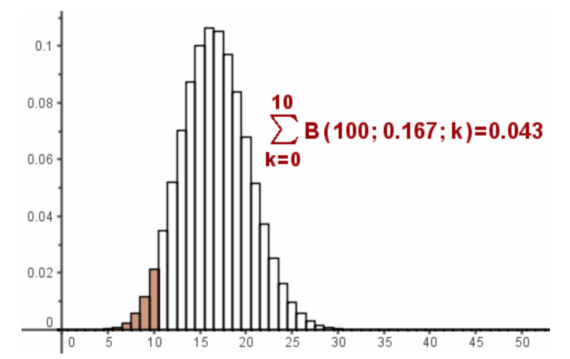
Fehler 2. Art (β):
Der ehemalige Mitarbeiter verrät dem Kommissar, dass die Wahrscheinlichkeit für eine 6 sogar nur 1/8 beträgt.
Der Kommissar kommt ins Grübeln:

Stichprobenumfang: 100;   Stichprobenergebnis: 10 Sechsen
Signifikanzniveau: 5 %;   Ablehnungsbereich: {0; 1; …; 10}
Entscheidung: Ich halte den Würfel für gefälscht, die Irrtumswahrscheinlichkeit beträgt 4,3%.

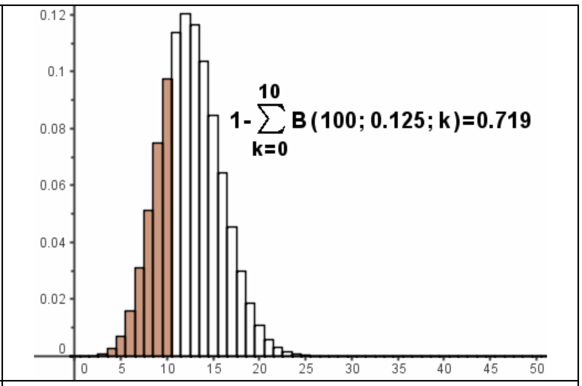
Wie groß war die Gefahr, dass ich solch einen gefälschten Würfel nicht entdecke ?

Die Anzahl der Sechsen ist bei dem gefälschten Würfel ***B100;1/8****-verteilt.*
Bei einem Stichprobenergebnis ***11; 12;…; 100*** *hätte ich den gefälschten Würfel nicht entdeckt.*

Fester Ablehnungsbereich **{0; 1; …; 10} !**



Verteilung gemäß Nullhypothese mit p = 1/6



Tatsächliche Verteilung mit p‘ = 1/8

Mit einer Wahrscheinlichkeit von etwa 72 % wird ein gefälschter Würfel nicht entdeckt !

Mit einer Wahrscheinlichkeit von etwa 72 % wird die Nullhypothese nicht abgelehnt, obwohl sie falsch ist ! (**Fehler 2. Art**)

Rechnerische Bestimmung des Fehlers 2. Art

Die Anzahl der Sechsen ist bei dem gefälschten Würfel $B_{100;1/8}$-verteilt.

Bei einem Stichprobenergebnis 11; 12;…; 100 wird der gefälschte Würfel nicht entdeckt.

Nullhypothese H_0: Der Würfel ist in Ordnung.

Alternativhypothese H_1: Der Würfel ist gefälscht.

Es gelte die Annahme $p' = \frac{1}{8}$.

Gesucht:

Die Wahrscheinlichkeit β dafür, dass H_0 nicht abgelehnt wird, obwohl H_0 falsch ist.

X': Anzahl der Sechsen beim gefälschten Würfel. X' ist $B_{100;1/8}$-verteilt.

$\beta = P(X' > 10)$ mit $p' = \frac{1}{8}$.

$\beta = P(X' > 10) = 1 - P(X' \le 10) \approx 0,719$.

Mit einer Wahrscheinlichkeit von 72 % wird ein gefälschter Würfel nicht entdeckt.

Mit einer Wahrscheinlichkeit von 72 % wird also die Nullhypothese nicht verworfen, obwohl sie falsch ist. Der Fehler 2. Art beträgt 72 %

Fehler 2. Art (β): Die Nullhypothese wird nicht verworfen, obwohl sie falsch ist.

Der Fehler 2. Art hängt ab

a) von der Wahl des Signifikanzniveaus,

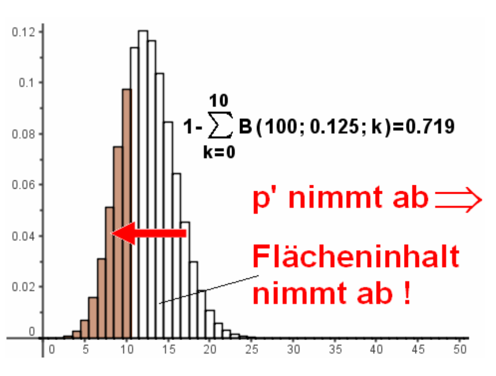
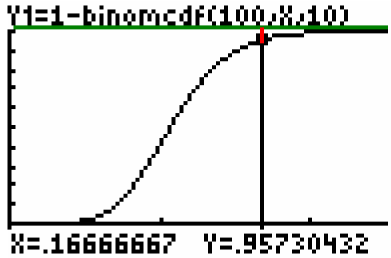
b) von der tatsächlichen Trefferwahrscheinlichkeit p’,

c) vom Stichprobenumfang n.

Die Einflüsse von p’ und n sollen im Folgenden näher untersucht werden.

Einfluss der tatsächlichen Trefferwahrscheinlichkeit p’ auf den Fehler 2. Art

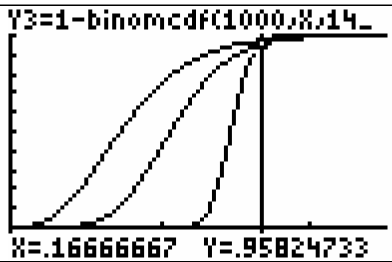
Ein stärker verfälschter Würfel müsste zuverlässiger entdeckt werden !

<table>
<tr>
<td>Mit abnehmendem p’ „wandert die Verteilung nach links“.
Der Ablehnungsbereich bleibt fest.
Nimmt p‘ ab, dann nimmt auch β ab.
Linksseitiger Test:
Für p‘ → 0 gilt: β → 0.
Für p‘→ 1/6 gilt: β → 1 – α.</td>
<td>
</td>
</tr>
<tr>
<td>Bestimmung von β mit Hilfe des GTR:
$\beta(p') = 1 - \sum_{k=0}^{10} \mathrm{B}(100, p', k)$
$= 1 - \mathrm{binomcdf}(100, X, 10)$</td>
<td>
</td>
</tr>
</table>

Einfluss des Stichprobenumfangs auf den Fehler 2. Art

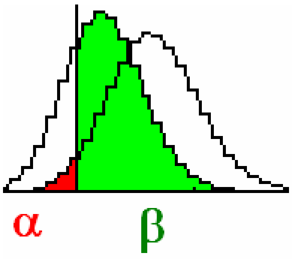
Mit dem Stichprobenumfang wächst auch die Menge der Information über die Grundgesamtheit.

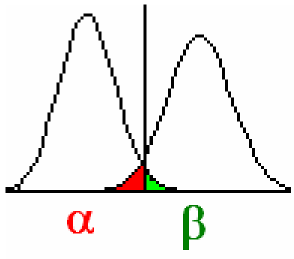
p = 1/6; tatsächliche Wahrscheinlichkeit p‘ = 1/8				
Stichproben-umfang n	Signifikanz-niveau	Ablehnungs-bereich	α	β
50	5 %	{0; 1; 2; 3}	2,4 %	**88,6 %**
100	5 %	{0; 1; …; 10}	4,3 %	**71,9 %**
1000	5 %	{0; 1; …; 146}	4,2 %	**2,2 %**



Fehler 2. Art in Abhängigkeit von der tatsächlichen Trefferwahrscheinlichkeit p‘ für n = 50; 100; 1000 bei einem Signifikanzniveau von 5 %

Die Ursache hierfür liegt darin, dass die Trefferzahlen bei größeren Stichprobenumfängen weniger stark um den Erwartungswert streuen:

<table>
<tr>
<td>n = 100; p = 1/6; p‘ = 1/8
– Ablehnungsbereich: {0; 1; …; 10}
– Verteilungen (GTR):


– α = 4,3 %; β = 72 %
Geringe Teststärke
– Für p = 1/6 gilt: μ = 16,7; σ = 3,7
(starke Streuung um den Erwartungswert).</td>
<td>n = 1000; p = 1/6; p‘ = 1/8
– Ablehnungsbereich: {0; 1; …; 146}
– Verteilungen (GTR):


– α = 4,2 %; β = 2 %
Hohe Teststärke (hohe Trennschärfe)
– Für p = 1/6 gilt: μ = 167 ; σ = 11,8
(schwache Streuung um den Erwartungswert).</td>
</tr>
<tr>
<td colspan="2">(Es gilt: $\mu = n \cdot p$ und $\sigma = \sqrt{n \cdot p \cdot (1-p)}$; d.h. $\mu \sim n$ und $\sigma \sim \sqrt{n}$)</td>
</tr>
</table>

Fehler 2. Art (β), Zusammenfassung:

- **Die Nullhypothese wird nicht abgelehnt, obwohl sie falsch ist.**
  „*Wir sind reingefallen, weil wir etwa Falsches akzeptiert haben.*“
- β hängt ab von der Irrtumswahrscheinlichkeit **α**, von der tatsächlichen Trefferwahrscheinlicheit **p‘** und vom Stichprobenumfang **n**.
  - Dieser Fehler ist i.A. nicht kontrollierbar.
- Eine Beibehaltung der Nullhypothese kann mit einem sehr hohen Fehler behaftet sein.
  - Die Tatsache, dass eine Nullhypothese nicht abgelehnt wird, kann also nicht als Beleg für deren Gültigkeit angesehen werden !
    („*Freispruch aus Mangel an Beweisen*“)
- Aus einer Annahme über p‘ kann β abgeschätzt werden.
  Mit Hilfe des Fehlers 2. Art kann dann entschieden werden, ob ein Test sinnvoll ist (Teststärke).
  (Auch aus diesem Grund ist die Alternativhypothese erforderlich.)

Ein Fehler ist also immer möglich !

- Bei festem Stichprobenumfang bewirkt ein Verkleinerung von α eine Vergrößerung von β und umgekehrt.
    - Ein Verkleinerung der Summe $\alpha + \beta$ kann nur durch Vergrößern des Stichprobenumfangs erreicht werden.
- Nur die Kenntnis **beider Fehler** erlaubt eine Einschätzung der Konsequenzen bei einer Fehlentscheidung !
  Dies ist bei Wahl der Nullhypothese zu berücksichtigen.
- (Für Wirksamkeitsstudien medizinischer Behandlungen schlägt Cohen (1969) für β einen 4-mal so hohen Wert vor wie für das Signifikanzniveau α. (*Wikipedia*))