<u>LB III Beurteilende Statistik</u>
<u>0. Ausgangsniveau</u>

<u>BISHER:</u> Ausgang: Die Wahrscheinlichkeit p von bestimmten Ereignissen (z.B Trefferwahrscheinlichkeit, Anzahl der schwarzen Kugeln in einer Urne…..) ist aus dem Zufallsexperiment bekannt und daraus kann man auf das Ergebnis bei n – facher Durchführung schließen.

Bsp.: Maschine produziert Teile mit einer Fehlerquote von 10%. Aus der laufenden Produktion werden rein zufällig 50 Teile entnommen. Mit welcher Wahrscheinlichkeit erhält man 5 defekte Teile?
B(50, 0,1) verteilt → P(X = 5) = 0,18492

<u>ABER:</u> In vielen Fällen kennt man die Wahrscheinlichkeit nicht.
Ist die Grundgesamtheit klein, ist es kein Problem, da man alles untersuchen kann und damit p feststellt.
Ist die Grundgesamtheit groß, ist es zeitlichen und finanziellen Gründen nicht möglich.
Seit 1920 verwendet man Stichproben, wo man z. B. bei der Qualitätskontrolle relativ wenige Stücke umso genauer untersucht und dann auf die Gesamtheit schließt.
Die Anzahl n der zu untersuchenden Stücke heißt Länge n der Stichprobe.
Logisch, dass dabei Fehlentscheidungen entstehen können.
→ **Aufgabe der Statistik** ist es, diese Fehlentscheidung unter Kontrolle zu bringen und durch Vorgabe von Entscheidungsregeln diese zu begrenzen.

In der mathematischen Statistik geht man davon aus, das die Trefferwahrscheinlichkeit p nicht bekannt ist. Anhand einer Stichprobe (d.h. man führt ein Zufallsexperiment aus) versucht man z. B. aus der auftretenden Trefferzahl auf die Wahrscheinlichkeit p zu schließen.
Dabei sind grundsätzlich zwei Situationen möglich:

Der Ausgang des Zufallsexperimentes (Stichprobe der Länge n) d.h. $X = x_i$ ist bekannt

Vermutung über die Trefferwahrscheinlichkeit p wird angestellt

Was für einen Wert soll für p angenommen werden? **(Schätzproblem)**	Stimmt p mit einem vermuteten Wert p_0 überein? **(Testproblem)**

Das Testproblem wird nun vertieft!

<u>Testproblem:</u>

Man hat von Anfang an eine gewisse Vermutung über die Trefferwahrscheinlichkeit p. Das Stichprobenergebnis (Zufallsexperiment) wird jetzt zu Hilfe genommen, um eine Entscheidung über die Vermutung, also über die aufgestellte **Hypothese** zu treffen.
Aufgrund des vorliegenden Stichprobenergebnisses wird eine Hypothese entweder verworfen oder beibehalten.
Wichtig ist es auch hier sich Gedanken über die Sicherheit eines solchen Urteils zu machen, denn bei einer anderen Stichprobe kann die Entscheidung ganz anders aussehen.
<u>Nichtverwerfen (beibehalten) einer Hypothese heißt nicht, dass wahr ist, sondern nur, dass sie mit dem entsprechenden Stichprobenergebnis verträglich ist.</u>

<u>Bsp1</u>: Ein Würfel wird 100 – mal geworfen. Die Sechs tritt viermal auf.
Wie ist dieser Würfel zu beurteilen?

Bsp.2: Eine Münze soll durch eine Stichprobe der Länge n = 50 überprüft werden, ob sie eine ideale Münze ist. Was ist bei dieser Überprüfung zu beachten?

Was ist ein statistischer Test?

Ein **statistischer Test** ist ein Verfahren, um zu entscheiden, ob die von einer Stichprobe gelieferten Daten einer Hypothese über die unbekannte Grundgesamtheit widersprechen.

Entscheidungen zwischen zwei Hypothesen

(1) Es stehen zwei Hypothesen zur Verfügung
H_0 → Nullhypothese (Arbeitshypothese)
H_1 → Gegenhypothese
nach Festlegung dieser Hypothesen
(2) Bestimmung der Entscheidungsregel
man gibt die Länge n der Stichprobe an sowie den **Annahmebereich A**, indem die **Hypothese H_0 nicht verworfen** (abgelehnt) wird
bzw. den **Ablehnungsbereich $\overline{A}$, in dem H_0 verworfen** (abgelehnt) wird.

Daher ergeben sich folgende **Entscheidungsregeln:**
- Stichprobenergebnis aus A d.h. $k \in A$ → Entscheidung für H_0
- Stichprobenergebnis aus $\overline{A}$ d.h. $k \in \overline{A}$ → Entscheidung gegen H_0
  d.h. Entscheidung für H_1

(Da Sachsen die Testverfahren nur an der **Binomialverteilung** erläutert, setzen wir diese auch im Schema voraus)

H_0: $p = p_0$ H_1: $p = p_1$

Start

H_0 trifft zu H_1 trifft zu

Ergebnis aus A	Ergebnis aus $\overline{A}$	Ergebnis aus A	Ergebnis aus $\overline{A}$
Entscheidung für H_0	Entscheidung für H_1	Entscheidung für H_0	Entscheidung für H_1
Richtige Entscheidung	**Fehlentscheidung** Fehler 1.Art	**Fehlentscheidung** Fehler 2.Art	**Richtige Entscheidung**
$B(n,p_0,X\in A)$	$\alpha = B(n,p_0,X\in \overline{A})$	$\beta = B(n,p_1,X\in A)$	$B(n,p_1,X\in \overline{A})$

Zwei der beiden Entscheidungsmöglichkeiten sind Fehlentscheidungen.
- H_0 trifft zu, aber es stellt sich ein Stichprobenergebnis aus $\overline{A}$ ein: man entscheidet sich irrtümlich gegen H_0. Man begeht einen Fehler 1.Art.
- H_1 trifft zu, aber es stellt sich ein Stichprobenergebnis aus A ein: man entscheidet sich irrtümlich gegen H_1. Man begeht einen Fehler 2.Art.

Zwei Testverfahren: Alternativtest
Signifikanztest

Alternativtest

Bei diesem Test sind zwei Hypothesen (Alternativen) vorgegeben.
Das Ereignis E besitzt entweder p_0 oder die Wahrscheinlichkeit p_1.
Aufgrund einer Stichprobe (mit vorgegebener Länge) und einer ebenfalls vorgegebenen Entscheidungsregel hält man p_0 oder p_1 für wahr. Diese Entscheidung für eine der beiden Hypothesen kann richtig oder falsch sein.

Aufbau eines Alternativtest an einem Beispiel:

Eva hilft bei einem Schulfest an einer Losbude. Sie füllt die Lostöpfe mit jeweils 50 Losen, darunter 10 Gewinnlose und 40 Nieten. Bei der letzten Füllung ist sie sich nicht mehr sicher, ob sie nicht aus Versehen 30 Gewinnlose und 20 Nieten eingefüllt hat. Sie beschließt, den Verkauf der ersten 10 Lose abzuwarten und sich für die Standardfüllung mit 10 Gewinnlose zu entscheiden, falls höchstens 4 Gewinnlose gezogen werden.
Bestimme die Wahrscheinlichkeiten der möglichen Fehlentscheidungen.

1. Festlegung der Zufallsgröße: X. Anzahl der Gewinnlose
2. Festlegung der Hypothesen: H_0: $p = p_0$ H_1: $p = p_1$
   H_0: $p = 0{,}2$ H_1: $p = 0{,}6$
3. Festlegung der Stichprobenlänge n und des Annahmebereiches A bzw. des Ablehnungsbereiches $\overline{A}$ der Hypothese H_0
   $n = 10$ $A = \{0,1,2,3,4\}$ $\overline{A} = \{5,......10\}$

4. Entscheidungsfindung:

(1) H_0 ist wahr und Stichprobenergebnis aus A → **Richtige Entscheidung**

$P_{H_0\ \text{wahr}}(H_0 \text{ wird angenommen}) = B(10,\ 0{,}2,\ X \le 4) = 0{,}96721$

Wertung: Mit einer Wahrscheinlichkeit von 96,72% wird ein richtig gefüllter Topf als solche erkannt.

(2) H_0 ist wahr und die Stichprobe aus $\overline{A}$ → Fehlentscheidung: **Fehler 1.Art**

$\alpha = P_{H_0\ \text{wahr}}(H_0 \text{ wird abgelehnt})$

$= P_{H_0\ \text{wahr}}(X \in \overline{A})) = 1 - B(10,\ 0{,}2,\ X \le 4) = 0{,}03279$

Wertung: Mit einer Wahrscheinlichkeit von 3,28% wird ein richtig gefüllter Topf nicht als solche erkannt.

(3) H_1 ist wahr und Stichprobenergebnis aus $\overline{A}$ → **Richtige Entscheidung**

$P_{H_1\ \text{wahr}}(H_1 \text{ wird angenommen}) = B(10,\ 0{,}6,\ X \ge 5\ ) = 0{,}83376$

Wertung: : Mit einer Wahrscheinlichkeit von 83,38% wird ein falsch gefüllter Topf als solche erkannt.

(4) H_1 ist wahr und die Stichprobe aus A→ Fehlentscheidung: **Fehler 2.Art**

$\beta = P_{H_1\ \text{wahr}}(H_1 \text{ wird abgelehnt})$

$= P_{H_1\ \text{wahr}}(X \in A)) = 1 - B(10,\ 0{,}6,\ X \ge 5\ ) = 0{,}16624$

Wertung: Mit einer Wahrscheinlichkeit von 16,62% wird ein falsch gefüllter Topf für einen richtig gefüllten erkannt.

Anmerkungen:

1. Durch Veränderung der Entscheidungsregel bei gleichem n kann die Wahrscheinlichkeit für einen α – Fehler vergrößert und dadurch die Wahrscheinlichkeit für den β- Fehler verkleinert werden und umgekehrt.
2. Zur Analyse eines Testproblems wird das „gegenseitige Abwägen“ der mit den möglichen Fehlentscheidungen verbundenen Risiken gehören. Bei der Bestimmung des Ablehnungsbereiches muss darauf Rücksicht genommen werden.

3. Ein Test ist umso zuverlässiger, je größer die Länge der Stichprobe n ist. $\alpha = \beta = 0$, wenn alle Lose überprüft werden und dabei kein Beobachtungsfehler gemacht wird.

Signifikanztest

In der Testtheorie wird es selten vorkommen, dass man zwischen zwei einfachen Hypothesen entscheiden muss. Vielmehr ist es häufig so, dass man aufgrund von Überlegungen **eine** Vermutung hat, die man in einem Test überprüfen soll.

Ein Entscheidungsverfahren, bei dem festgestellt wird, ob eine Hypothese H_0 verworfen wird oder nicht heißt **Signifikanztest**. Das Risiko α bei dieser Entscheidung heißt **Signifikanzniveau**.

Anmerkungen:

(1) Da man nur feststellt, ob eine Nullhypothese abgelehnt wird oder nicht, interessiert im Allgemeinen nicht, welche andere Hypothese eventuell wahr ist.

(2) Ein Versuchsergebnis, das zur Ablehnung der Nullhypothese H_0 führt d.h. in bedeutsamer (**signifikanter**) Weise der Nullhypothese widerspricht, heißt **signifikant auf dem Niveau** α.

Beispiel:

Bei einem idealen Würfel tritt die AZ 6 mit einer Wahrscheinlichkeit von $\frac{1}{6}$ auf.

Jemand glaubt nur dann, dass ein idealer Würfel vorliegt,, wenn bei 100 Würfen die die Anzahl der Sechsen

a) mindestens 13
b) mindestens 11 und höchstens 22 beträgt

Wie groß ist in beiden Fällen die Wahrscheinlichkeit dafür, dass ein idealer nicht als solcher erkannt wird?

a) (Ablauf entspricht dem beim Alternativtest)

X: Anzahl der Sechsen → $B\left(100, \frac{1}{6}\right)$ verteilt

H_0: $p_0 = \frac{1}{6}$ $\qquad$ H_1: $p_1 \neq \frac{1}{6}$

$A = \{13, 14, \ldots\ldots 100\}$ $\quad$ $\overline{A} = \{0, 1, \ldots 12\}$

$$\alpha = P_{H_0\text{ wahr}}(H_0 \text{ wird abgelehnt})$$

$$= P_{H_0\text{ wahr}}(X \in \overline{A})) = B(100, \frac{1}{6}, X \leq 12) = 0{,}12967$$

→ Ein idealer Würfel wird mit 12,97% nicht als solcher erkannt.

Anmerkung: Da in diesem Fall der kritische Bereich $\overline{A}$ = Ablehnungsbereich $\overline{A}$ eine „Richtung“ besitzt, d.h. H_0 nur abgelehnt wird, wenn zu wenige Sechsen auftreten, heißt dieser Test einseitiger Test (linksseitiger Test)

b) X: Anzahl der Sechsen → $B\left(100,\frac{1}{6}\right)$ verteilt

H_0: $p_0 = \frac{1}{6}$ H_1: $p_1 \neq \frac{1}{6}$

$\overline{A} = \{0,1,\ldots 10\} \cup \{23,\ldots 100\} \rightarrow A = \{11,\ldots 22\}$

$\alpha = P_{H_0\ \text{wahr}}(H_0 \text{ wird abgelehnt})$

$= P_{H_0\ \text{wahr}}(X \in \overline{A}))$

$= B(100, \frac{1}{6}, X \leq 10) + B(100, \frac{1}{6}, X \geq 23) = B(100, \frac{1}{6}, X \leq 10) + 1 - B(100, \frac{1}{6}, X \leq 22)$

$= 0{,}10575$

Mit einer Wahrscheinlichkeit von 10,58% wird ein idealer Würfel nicht als solcher erkannt.

Anmerkung: Da in diesem Fall der kritische Bereich $\overline{A}$ = Ablehnungsbereich $\overline{A}$ keine „Richtung“ besitzt, d.h. H_0 sowohl abgelehnt wird, wenn sowohl zu wenige als auch zu viele Sechsen auftreten, heißt dieser Test zweiseitiger Test.

Ein Test heißt **einseitig,** wenn der Ablehnungsbereich $\overline{A}$ (Annahmebereich A) ein Intervall ist. Ein Test heißt **zweiseitig,** wenn der Ablehnungsbereich $\overline{A}$ (durch den Annahmebereich) in zwei Intervalle zerfällt.

Beispiel:
Laut Angaben des Herstellers keimen mindestens 90% der Samen seines Saatgutes.
Ein Gärtner sät 100 Samenkörner und beobachtet das Keimverhalten.
a) Wie muss der Ablehnungsbereich $\overline{A}$ für die Behauptung des Herstellers gewählt werden, wenn man sich höchstens mit einer Wahrscheinlichkeit von 5% irren will?
b) Wie groß wäre die Wahrscheinlichkeit für eine Fehlentscheidung, wenn man die Behauptung des Herstellers irrtümlich ablehnt, falls weniger als 80 Samen keimen.

X: Anzahl der keimenden Samen
H_0: $p_0 \geq 0{,}9$ H_1: $p_1 < 0{,}9$ → B(100, 0,9) verteilt
$\overline{A} = \{0,\ldots k\}$ weil H_0 abgelehnt wird, wenn zu wenige Samen keimen
a) Die Ablehnung von H_0 soll mit Wahrscheinlichkeit von höchstens 5% erfolgen.

$\alpha = P_{H_0\ \text{wahr}}(H_0 \text{ wird abgelehnt}) = P_{H_0\ \text{wahr}}(X \in \overline{A})) \leq 0{,}05$

→ $B(100, 0{,}9, X \leq k) \leq 0{,}05$ → GTR (Table oder Solver) → k = 84

d.h. H_0 wird abgelehnt, wenn höchstens 84 Samen keimen d.h. $\overline{A} = \{0,\ldots 84\}$

b) $\overline{A} = \{0,\ldots 79\}$

$\alpha_1 = P_{H_0\ \text{wahr}}(H_0 \text{ wird abgelehnt}) = P_{H_0\ \text{wahr}}(X \in \overline{A})) = B(100, 0{,}9, X \leq 79) = 0{,}00081$

Die Fehlerwahrscheinlichkeit ist auf 0,08% gesunken, weil jetzt viel vorsichtiger entschieden wird.

Signifikanztest läuft (fast) immer nach folgenden Schritten ab:

1. Wie lautet die Nullhypothese H_0(häufig die Verneinung der Vermutung)?
2. Wie groß ist der Stichprobenumfang und welche Irrtumswahrscheinlichkeit (Signifikanzniveau α) ist vorgegeben?
3. Welche Größe X wird verwendet und wie lautet $\overline{A}$ (kritischer Bereich)?
4. Wie wird aufgrund des Stichprobenergebnisses entschieden?

Anmerkungen:

1. Wenn $\overline{A}$ gegeben ist, kann α bestimmt werden und umgekehrt
2. Bei einem zweiseitigem Test d.h. $\overline{A} = \{0,\ldots k_1\} \cup \{k_2,\ldots n\}$ wird die

   Fehlerwahrscheinlichkeit α gleichmäßig auf beide Bereiche von $\overline{A}$ verteilt ($\frac{\alpha}{2}$)

   Ein zweiseitiger Test liegt immer dann vor, wenn die Vermutung keine Richtung besitzt d.h. H_0 die Form H_0: $p = p_0$ ($p \neq p_0$) bzw. H_0: $p_1 < p < p_2$
3. Bei einem einseitigem Test gilt:
   H_0: $p \leq p_0$ ($p < p_0$) wird abgelehnt, wenn die Prüfgröße X sehr große Werte annimmt, d.h. $\overline{A} = \{k,\ldots n\}$ gilt.
   H_0: $p \geq p_0$ ($p > p_0$) wird abgelehnt, wenn die Prüfgröße X sehr klein Werte annimmt, d.h. $\overline{A} = \{0,\ldots k\}$ gilt.
4. Ob einseitig oder zweiseitig getestet wird, muss man aufgrund der Aufgabenstellung genau überlegen d.h. ob irgendwelche Abweichungen nach oben und nach unten oder nur nach einer Seite von Interesse sind.
   Beispielweise wird beim Ankauf einer Ware mit einer bestimmten Fehlerwahrscheinlichkeit kaum interessieren, ob der Anteil der fehlerhaften Ware in der Lieferung kleiner ist als vereinbart.