1.1 ▶ Wahrscheinlichkeit nachweisen (5BE)

Gehen wir die Informationen aus dem Zeitungsartikel durch: in Usse leben 9625 Menschen, von denen 49 % Männer sind. Berechne zunächst, wie viele Männer dies tatsächlich sind:

$0,49 \cdot 9625 = 4716,25 \approx 4716$

In Usse leben also 4716 männliche Einwohner. Statistisch gesehen hätten von diesen 4716 innerhalb der letzten acht Jahre eigentlich nur 5,2 Männer an TIN erkranken dürfen. Berechne die zugehörige relative Häufigkeit:

$\dfrac{5,2}{4716} = 0,001103 = 0,11\,\%$

Auf ganz Deutschland übertragen bedeutet dies: für einen in Deutschland lebenden Mann beträgt die Wahrscheinlichkeit, an TIN zu erkranken, 0,11 %.

1.2 ▶ Rechnung im Zusammenhang erläutern (5BE)

Wir wollen mit X die Anzahl der Männer in der Stichprobe bezeichnen, die innerhalb der letzten acht Jahre an TIN erkrankt sind. Aufgrund der großen Anzahl Männer in Deutschland und der relativ geringen Stichprobengröße kann X näherungsweise als binomialverteilt angenommen werden mit $n = 2.500$ und $p = 0,0011$. (s. Aufgabenteil 1.1)

Betrachten wir unter dieser Annahme nun die Rechnung in der Aufgabenstellung. Hier wird eine Wahrscheinlichkeit berechnet, und zwar die Wahrscheinlichkeit $P(X \leq 3)$:

$$\begin{aligned}
P(X \leq 3) &= \sum_{k=0}^{3} P(X = k) \\
&= \sum_{k=0}^{3} \binom{2.500}{k} \cdot 0,0011^k \cdot (1 - 0,0011)^{2.500-k} \\
&= \sum_{k=0}^{3} \binom{2.500}{k} \cdot 0,0011^k \cdot (0,9989)^{2.500-k}
\end{aligned}$$

Diese Wahrscheinlichkeit nimmt den Wert 70,3 % an. Wir können also sagen: Mit einer Wahrscheinlichkeit von 70,3 % sind von den 2.500 beobachteten Männern innerhalb der letzten acht Jahre **höchstens 3** an der Krankheit TIN erkrankt.

2.1 ▶ Nullhypothese und Gegenhypothese nennen (2BE)

Das deutschlandweite Risiko für einen Mann, an TIN zu erkranken, liegt bei 0,11 %. Dies ist auch der Wert, von dem zunächst in Usse ausgegangen wird und der im Hypothesentest widerlegt werden soll.

Die zu widerlegende Hypothese wählen wir als **Nullhypothese** und sichern sie so auf einem Signifikanzniveau ab. Auf diese Weise können wir die Wahrscheinlichkeit für den Fehler erster Art, also irrtümlicherweise von einem höheren Risiko auszugehen, in den Griff bekommen.

Die Nullhypothese kannst du also so formulieren: $H_0 : p \leq 0,0011$.

Die Gegenhypothese nimmt an, dass ein **erhöhtes** Risiko besteht: $H_1 : p > 0,0011$.

2.2 ▶ Annahme und Ablehnungsbereich bestimmen (8BE)

Sei X die Anzahl der an TIN erkrankten Männer in Usse. Insgesamt leben 4.716 Männer in Usse. Bei wahrer Nullhypothese kann X somit als binomialverteilt angenommen werden mit $n = 4.716$ und $p = 0,0011$.

Die Nullhypothese $H_0 : p \leq 0{,}0011$ wird **abgelehnt**, wenn **besonders viele** Männer an TIN erkranken. Der **Ablehnungsbereich** der Nullhypothese hat also die Form $\overline{A} = \{k; ...; 4.716\}$ und wird durch eine noch unbekannte untere Grenze k begrenzt.

Der Annahmebereich der Nullhypothese ist dann entsprechend $A = \{0; ...; k-1\}$.

Nun soll die Nullhypothese auf dem Signifikanzniveau $\alpha = 0{,}05$ abgesichert werden. Das bedeutet: die Wahrscheinlichkeit, die Nullhypothese **irrtümlich** abzulehnen, soll höchstens 5 % betragen:

$P(X \in \overline{A}) \leq 0{,}05 \;\Leftrightarrow\; P(X \geq k) \leq 0{,}05 \;\Leftrightarrow\; 1 - P(X \leq k-1) \leq 0{,}05 \;\Leftrightarrow\; P(X \leq k-1) \geq 0{,}95$

In der Anlage findest du eine Tabelle zur kumulierten Binomialverteilung für $n = 4.716$ und $p = 0{,}0011$. Gesucht ist nun der **kleinste Wert** für $k-1$, der obige Ungleichung **gerade noch** erfüllt. In der Tabelle findest du die Werte:

$P(X \leq 8) = 0{,}9191 < 0{,}95$; und $P(X \leq 9) = 0{,}9609 > 0{,}95$.

Damit folgt $k - 1 = 9$ und somit $k = 10$. Du erhältst so den **Annahmebereich** $A = \{0; ...; 9\}$ und den **Ablehnungsbereich** $\overline{A} = \{10; ...; 4716\}$.

Die Nullhypothese $H_0 : p \leq 0{,}0011$ wird also abgelehnt, wenn mindestens 10 Männer in Usse an TIN erkrankt sind.

In Usse sind laut Zeitungsbereich **12** Männer an TIN erkrankt. Da 12 im **Ablehnungsbereich** der Nullhypothese liegt, wird diese abgelehnt. Es besteht demnach ein signifikanter Unterschied.

2.3 ▶ Wahrscheinlichkeit für Fehler 2. Art bestimmen (6BE)

Ein Fehler 2. Art liegt vor, wenn die Nullhypothese tatsächlich falsch ist, aber dennoch angenommen wird. In unserem Fall bedeutet das: das Risiko für einen in Usse lebenden Mann, an TIN zu erkranken, liegt in Wirklichkeit bei 0,25 %, dennoch werden in der Stichprobe **weniger als 10** Männer gefunden, die an TIN erkrankt sind.

Wir führen eine neue Zufallsvariable Y ein, die binomialverteilt sei mit $n = 4.716$ und $p = 0{,}0025$. Die Wahrscheinlichkeit für den Fehler 2. Art ist die Wahrscheinlichkeit $P(Y \leq 9)$.

Betrachte wieder die Tabelle im Anhang und sieh dir dieses Mal die Spalte für $p = 0{,}0025$ an: du kannst ihr den Wert $P(Y \leq 9) = 0{,}2609$ entnehmen.

Das bedeutet: die Wahrscheinlichkeit, einen Fehler 2. Art zu begehen und irrtümlicherweise von einem zu niedrigen Krankheitsrisiko auszugehen, ist mit 26,09 % doch recht hoch.

2.4 ▶ Operationscharakteristik beschreiben (4BE)

Im Hypothesentest wird davon ausgegangen, dass das Risiko für einen Mann, an TIN zu erkranken, bei $p_0 = 0{,}0011$ liegt. Abhängig davon, wie hoch das **tatsächliche** Risiko p_1 ist, an TIN zu erkranken, tritt in diesem Hypothesentest nun mit unterschiedlicher Wahrscheinlichkeit ein Fehler 2. Art auf.

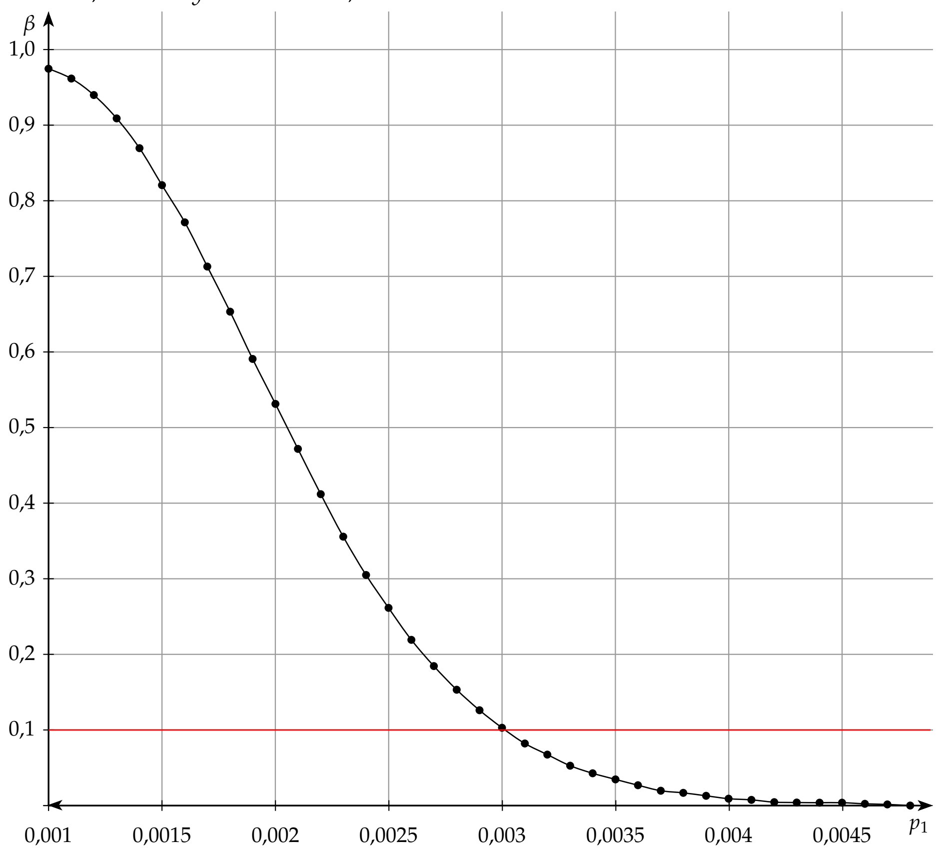
Die Operationscharakteristik zeigt die **Entwicklung der Wahrscheinlichkeit** für den Fehler 2. Art in Abhängigkeit vom tatsächlichen Wert p_1. Diese tatsächlichen Werte p_1 sind auf der x-Achse abgetragen. Auf der y-Achse wird die zugehörige Wahrscheinlichkeit für den Fehler 2. Art abgetragen.

Betrachte z.B. die Kurve an der Stelle $p_1 = 0{,}0025$: sie verläuft hier durch den Punkt $(0{,}0025 \mid 0{,}2609)$. Dies bestätigt das Ergebnis aus Aufgabenteil 2.3.

Die Funktionswerte der Operationscharakteristik werden **kleiner** mit **größer werdendem** Wert p_1. Im Sachzusammenhang interpretiert bedeutet das: je höher das tatsächliche Risiko für einen Mann, an TIN zu erkranken, desto **kleiner** wird die Wahrscheinlichkeit, irrtümlicherweise vom geringeren Risiko 0,0011 auszugehen.

▶ Werte für p_1 bestimmen

Gesucht sind nun die Werte für p_1, für die der Fehler 2. Art **kleiner als** 10 % wird. Diese Aufgabe kannst du graphisch lösen: Zeichne in die Operationscharakteristik eine Parallele zur x-Achse, die die y-Achse bei $0,1$ schneidet:



Du kannst erkennen, dass diese Parallele die OC-Kurve etwa bei 0,003 schneidet. Für alle Werte von p_1, die **kleiner** als 0,003 sind, verläuft die Kurve unterhalb der Parallelen.

Das bedeutet: Für $p_1 > 0,003$ wird die Wahrscheinlichkeit für den Fehler zweiter Art **kleiner als** 10 %.