a) **► Berechnung der Wahrscheinlichkeiten der drei Ereignisse** (11P)

X sei die Anzahl der Lesefans unter den 8 befragten Personen. Da eine Person nur ein Lesefan sein kann oder eben nicht, kann X als binomialverteilt mit $n = 8$ sowie $p = 0,25$ gesehen werden, da laut Zeitungsartikel nun 25% aller Personen zu den Lesefans gehören.

► Wahrscheinlichkeit für genau 2 Lesefans

Hier ist die Wahrscheinlichkeit für $X = 2$ Lesefans gefragt, die nach der Formel von Bernoulli berechnet werden muss:

$$P(X=2) = \binom{8}{2} \cdot 0,25^2 \cdot 0,75^6 \approx 0,3115 = 31,15\%$$

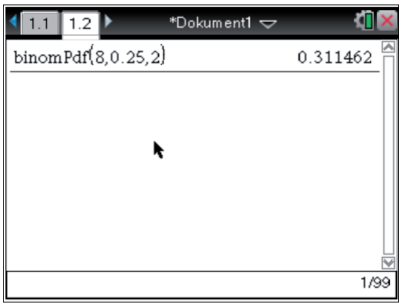
Mit einer Wahrscheinlichkeit von ca. 31% finden sich unter den 8 Personen genau 2 Lesefans.

alternativ

Alternativ kannst du hier die Wahrscheinlichkeit mit dem CAS berechnen. Verwende dazu im Calculator - Modus den `binomPdf(..)` Befehl. Dieser ist so anzuwenden:

`binomPdf(`n`,`p`,`k`)`

Dabei ist n die Anzahl aller Versuche, p das Wahrscheinlichkeitsmaß und k die Anzahl der Ereignisse. In der nebenstehenden Abbildung wurde die Aufgabe, auf die beschriebene Art und Weise gelöst.



► Wahrscheinlichkeit für keinen Lesefan

Die Wahrscheinlichkeit errechnet sich mit der Bernoulli-Formel oder wie oben beschrieben über den CAS:

$$P(X=0) = \binom{8}{0} \cdot 0,25^0 \cdot 0,78^8 = 0,78^8 \approx 0,1001 = 10,01\%$$

► Wahrscheinlichkeit für mindestens 3 Lesefans

Diese Wahrscheinlichkeit wird über das Gegenereignis bestimmt, wobei die Werte für $P(X=2)$ und $P(X=0)$ von den obigen Wahrscheinlichkeiten übernommen werden können:

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) = 1 - [P(X=0) + P(X=1) + P(X=2)] \\ &\approx 1 - \left[0,1001 + \binom{8}{1} \cdot 0,25^1 \cdot 0,75^7 + 0,3115\right] \approx 0,3214 = 32,14\% \end{aligned}$$

Mit einer Wahrscheinlichkeit von ca. 32% finden sich unter den 8 Personen mindestens 3 Lesefans.

Anmerkung

Die Wahrscheinlichkeit $P(X \leq 2)$ kann auch über den Calculator - Modus des CAS bestimmt werden. Verwende dazu im entsprechenden Dokument den Befehl `binomCdf(..)`. Dieser Befehl ist so anzuwenden:

`binomCdf(`n`,`p`,`k`)`
n: Anzahl der Versuche.
p: Wahrscheinlichkeitsmaß.
k: Anzahl der Ereignisse.
(hier: `binomCdf(8,0.25,2)`)

b) **► Bestimmung eines Hypothesentests, den die Firma durchführen wird** (10P)

Die Firma will sicher gehen, dass sie nicht nicht irrtümlich behauptet, dass der Anteil der Lesefans wirklich gesunken ist. Deswegen geht sie in der Nullhypothese davon aus, dass es doch mehr als 22% Lesefans gibt, somit ist H_0: $p_0 \geq 0,22$ und damit H_1: $p_1 < 0,22$. Die Nullhypothese stößt auf Ablehnung, wenn sich unter den 2.500 befragten nur **sehr wenige** Personen ergeben, die wirklich Lesefans sind, daher handelt es sich um einen linksseitigen Signifikanztest.

X sei dabei die Anzahl der Lesefans, die unter den 2.500 befragten Personen. Wenn H_0 wahr ist, ist diese Zufallsgröße binomialverteilt mit $n = 2.500$ sowie $p = 0,22$. Für den Erwartungswert μ sowie die Standardabweichung σ von X ergibt sich:

$$\mu = n \cdot p = 2.500 \cdot 0,22 = 550$$

$$\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{2.500 \cdot 0,22 \cdot 0,78} \approx 20,71$$

Weil hier die Standardabweichung $\sigma > 3$ ist, kann die in der Aufgabenstellung gegebene Näherung angewandt werden. Sie besagt, dass sich X mit einer Wahrscheinlichkeit von ca. 90% im Bereich $[\mu - 1,64\sigma; \mu + 1,64\sigma]$ befindet.
Wegen der symmetrischen Verteilung befinden sich die Werte von X damit mit 5% Wahrscheinlichkeit im Bereich zwischen 0 und dem Wert $\mu - 1,64\sigma$ $[P(X < \mu - 1,64\sigma) \approx 0,05]$ und ebenso mit einer Wahrscheinlichkeit von ca. 5% im Bereich zwischen $\mu + 1,64\sigma$ und 2.500 $[P(X > \mu + 1,64\sigma) \approx 0,05]$.

Da die Nullhypothese ihren Ablehnungsbereich auf der **linken** Seite hat, müssen wir die linke „Randwahrscheinlichkeit“ $P(X < \mu - 1,64\sigma)$ betrachten:

$$\begin{aligned} P(X < \mu - 1,64\sigma) &\approx 0,05 \\ P(X < 550 - 1,64 \cdot 20,71) &\approx 0,05 \\ P(X < 516,032) &\approx 0,05 = \alpha \end{aligned}$$

Die Wahrscheinlichkeit, dass X (unter der wahren Nullhypothese) fälschlicherweise einen Wert kleiner 516 annimmt, ist somit genauso groß wie unser Signifikanzniveau. Damit wird H_0 genau dann abgelehnt, wenn maximal 516 Lesefans unter den 2.500 Personen auftreten.

Da bei dem vorliegenden Test nur 502 Lesefans aufgetreten sind, erweist sich hier die Nullhypothese als falsch. Die Firma hat mit ihrer Behauptung, dass Anteil der Lesefans kleiner als 22% ist, Recht.

► Beschreibung des Fehlers, den die Firma begehen kann

Da sie die Nullhypothese nun ablehnt, kann sie den **Fehler 1. Art** begehen. Das bedeutet, dass sie nun ihre Behauptung als bestätigt sieht, **obwohl** in Wirklichkeit doch mehr als 22% aller Deutschen Lesefans sind.

Wegen des Signifikanzniveaus beträgt die Wahrscheinlichkeit für diesen Fehler $\alpha = 5\%$.

c) (1) ▶ **Nullhypothese und Berechnung der Fehlerwahrscheinlichkeit** (12P)

Die Chefin lehnt ihre Vermutung genau dann ab, wenn sich weniger als 80 Lesefans in der Stichprobe befinden, daher ist der Ablehnungsbereich der Nullhypothese
$\overline{A} = \{0; 1; ...; 79\}$.
Ihre Nullhypothese lautete somit H_0: $p_0 \geq 0,35$.

Um die maximale Wahrscheinlichkeit dafür, dass eine negative Zeitungsmeldung erscheint, zu bestimmen, werden die Lesefans als Zufallsgröße X bezeichnet. Sie sind in dem Test hier binomialverteilt mit $n = 250$ und – bei wahrer Nullhypothese – wenigstens mit $p = 0,35$.
Die Fehlerwahrscheinlichkeit ergibt sich daraus, dass die Nullhypothese zwar stimmt, genau in dem Test aber trotzdem nur maximal 79 Lesefans aufgetreten sind $(X \leq 79)$. Die Wahrscheinlichkeit ist somit:
Gesucht: $P(X \leq 79)$
Diese Wahrscheinlichkeit kannst du wie im zweiten Aufgabenteil vorgestellt, über den Calculator - Modus des CAS berechnen. Dazu musst du den `binomCdf(..)` - Befehl wie folgt anwenden:

$P(X \leq 79) = \texttt{binomCdf}(n, p, k)$

$P(X \leq 79) = \texttt{binomCdf}(250, 0.35, 79)$

$P(X \leq 79) = 0{,}1441$

$P(X \leq 79) = 14{,}41\%$

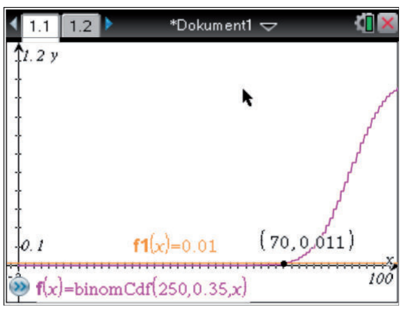
(2) ▶ **Bestimmung einer Entscheidungsregel**

Da nun ein neues Signifikanzniveau vorgegeben ist, ist der Ablehnungsbereich nun
$\overline{A} = \{0; 1; 2; ...; k\}$. Es ist nun das letzte k gesucht,so dass die Ungleichung $P(X \leq k) \leq 0,01$ erfüllt ist, wobei X die Anzahl der Lesefans bei wahrer Nullhypothese angibt und wie oben verteilt ist.

Um herauszufinden ab welchem k dies der Fall ist, empfiehlt es sich hier mit dem Graphs - Mode des CAS zu arbeiten. Hier definierst du die $B_{250;0,35}$-verteilte Zufallsvariable X als Funktion in Abhängigkeit von x. x wird dabei als k verwendet und beschreibt die Anzahl der Lesefans bei insgesamt 250 Befragungen. Um k zu berechnen, musst du hier, wie schon im ersten Aufgabenteil mit dem `binomCdf(..)` - Befehl arbeiten. Die Funktion f, welche die Wahrscheinlichkeiten von X beschreibt, sollte so aussehen:

$f(x) = \texttt{binomCdf(250,0.35,}x\texttt{)}$

Da du bestimmen willst, ab welchem k $P(X \leq k) \leq 0,01$ gilt, lässt du f mit der Geraden $y = 0,01$ schneiden und bestimmst den Schnittpunkt der beiden Graphen. Hast du beide Funktionen in das entsprechende Graphs - Dokument eingetragen, bestimmst du über diese Befehlsfolge den Schnittpunkt:



$\boxed{\text{menu} \rightarrow \text{6: Graph analysieren} \rightarrow \text{4: Schnittpunkt}}$

Somit hast du herausgefunden, dass $k = 70$, dass erste k ist, für welches die oben genannte Bedingung ($P(X \leq k) \leq 0,01$) nicht mehr gilt. Daraus folgt, dass auf dem Signifikanzniveau von 1% ist H_0 genau dann abzulehnen ist, wenn sich unter den 250 Befragten maximal nur 69 Lesefans finden lassen.

(3) ► **Erklärung der Wahl der Nullhypothese bei der Stadtverwaltung**

Um sich gegen die womöglich falsche Behauptung der Chefin abzusichern, muss die Stadtverwaltung genau vom Gegenteil, also von H_0: $p_0 < 0,35$, ausgehen.

Auf einem niedrigen Signifikanzniveau wird man hier den Fehler 1. Art nämlich nur sehr selten begehen. So werden hier jegliche Fehlinvestitionen in den Ausbau der Bibliothek vermieden!

d) (1) ► **Wahrscheinlichkeit für eine falsche Information der Öffentlichkeit** (9P)

Da es in der Stadt nun wirklich nur 20% Lesefans gibt, ist die obige Anzahl X der Lesefans unter den 250 Befragten tatsächlich binomialverteilt mit $n = 250$ und $p = 0,2$. Offensichtlich ist die Vermutung der Chefin falsch.

Die Öffentlichkeit wird falsch informiert, wenn nun trotzdem ein positiver Zeitungsartikel erscheint. Dies passiert, wenn sich unter den 250 Befragten mindestens 60 Lesefans befinden:

$P(X \geq 60) = 1 - P(X \leq 59)$

$P(X \geq 60) = 1 - \texttt{binomCdf(250,0.2,59)}$ Erklärung: siehe Aufgabenteil a)

$P(X \geq 60) = 0{,}0689$

Mit einer Wahrscheinlichkeit von ca. 6,89% wird die Öffentlichkeit somit falsch informiert.

(2) ► **Erklärung einer möglichen Fehlerverringerung**

Die obige Wahrscheinlichkeit für den **Fehler 2. Art** wird genau dann verringert, wenn man den Annahmebereich der Nullhypothese verkleinert. Denn genau dann passiert es auch seltener, dass man die Nullhypothese irrtümlich annimmt, wenn sie eigentlich falsch ist, wie es hier der Fall war.

Wird der Annahmebereich verringert, vergrößert sich logischerweise der Ablehnungsbereich von H_0. Dann steigt wiederum die Gefahr, dass H_0 irrtümlich abgelehnt wird, also der Fehler 1. Art.

In diesem Beispiel bestand dieser darin, eine negative Zeitungsmeldung zu veröffentlichen, obwohl der Anteil der Lesefans in Wirklichkeit bei mindestens 30% lag.

e) (1) **Erklärung, warum Anna das Modell mit der Urne wählt** (8P)

Es gibt mehrere Gründe für das Modellieren der Situation mit einer Urne:

Da Anna die Kugeln ohne Zurückziehen entnimmt, ahmt sie so die Befragung der einzelnen Personen nach – denn: keine Person wird mehrmals, sondern nur **einmal** befragt!

Da es nur zwei Sorten von Kugeln in der Urne gibt, kann Anna so die beiden möglichen Ergebnisse bei der Befragung nachahmen: Lesefan (eine Farbe) oder eben kein Lesefan (die andere Farbe).

Wenn Anna im ersten Zug eine rote Kugel entnimmt, geschieht dies mit einer Wahrscheinlichkeit von $p = \frac{10}{40} = 0,25$. Dies entspricht genau der Wahrscheinlichkeit dafür, einen Lesefan anzutreffen, womit dem Antreffen eines Lesefans in diesem Modell das Ziehen einer roten Kugel gleichgestellt werden könnte.

(2) ► **Begründung, dass das Modell die Situation nicht angemessen simuliert**

Das Problem bei Annas Modell ist, dass die Trefferwahrscheinlichkeit für eine rote Kugel, welche einen Lesefan beschreibt, **nicht konstant** ist und sich von Stufe zu Stufe verringert. Im zweiten Zug zieht sie eine rote Kugel beispielsweise nur noch mit einer Wahrscheinlichkeit von $p = \frac{9}{39} \approx 0,2308$.

In Teilaufgabe wurde eine Binomialverteilung der Lesefans angenommen, die mit diesem Modell allerdings nicht erfüllt ist, da die Trefferwahrscheinlichkeit nicht konstant ist. Daher kann Annas Modell die vorliegende Situation nicht richtig simulieren.