Übungsserie 3

1. Die Datei **distrib.syd** enthält 400 ganzzahlige Zahlen (**X**), die durch einen Zufallsprozess aus einer vorgegebenen Verteilung generiert worden sind.

a) Welchen Verteilungen könnten diese Zahlen entsprechen? Betrachten Sie zu diesem Zweck das Stabdiagramm von **X**. Durch wieviele Parameter werden diese Verteilungen bestimmt?

Mit dem Quickbutton **Bar** das Stabdiagrammenü aktivieren und dort die X-Variable angeben. Für eine schönere Darstellung in **Axes / All Axes ...** im Feld **Tick Mark Style** den Knopf **Outside Axes** anwählen (dadurch werden in der Graphik die Strichlein bei den Werten auf den Achsen nach aussen gerichtet).

b) Schätzen Sie aus den Daten alle Parameter der Verteilung. (Falls Sie unter a) mehrere Verteilungen erwähnen, dann wählen Sie diejenige, die durch die kleinste Anzahl Parameter bestimmt wird.) Überprüfen Sie auf einfache Weise, ob die Daten mit der Verteilung übereinstimmen.

Nützliches befindet sich im Menü **Statistics / Descriptive Statistics / Basic Statistics...**.

2. Auf einer gefährlichen Kreuzung ereignen sich pro Monat durchschnittlich 0.9 schwere Unfälle. Die Anzahl Unfälle pro Monat kann als Poissonverteilt mit Parameter $\lambda = 0.9$ angenommen werden.

a) Berechnen Sie die Wahrscheinlichkeit, dass sich während eines Monats kein Unfall, ein Unfall, mehr als ein Unfall ereignet.

b) Der Gemeinderat beschliesst, die Kreuzung zu sanieren. Im folgenden Jahr wurden auf der Kreuzung noch 4 Unfälle registriert. Hat sich die Situation gebessert?
(Berechnen Sie die Wahrscheinlichkeit, dass auf der **nicht** sanierten Kreuzung 4 oder weniger Unfälle pro Jahr geschehen.)

Man kann diese Wahrscheinlichkeit mit Systat schnell rechnen: öffne ein neues Datenblatt mit **File / New / Data**, definiere eine neue Variable **X** mit Werte von 0 bis mindestens 4, definiere mit **Data / Transform / Let...** eine neue Variable **Kumulativ** durch PCF(X,λ) wobei für λ den richtigen Wert eingesetzt muss.
Bemerkung: PCF(.,.) ist eine Abkürzung für "Poisson cumulative (distribution) function".

3. Bestimmen Sie für jede der folgenden Zufallsvariablen, ob die Binomial- oder die Poissonverteilung oder keine von beiden als Modell sinnvoll wäre.

- $X^{(a)}$ sei die Anzahl Lokomotiven der SBB, die in der nächsten Woche einen Defekt haben werden.
- $X^{(b)}$ sei die Anzahl fauler Äpfel in einer Packung zu 6 Stück.
- $X^{(c)}$ sei die Anzahl Autos, die an einem Tag bei einer bestimmten Ampel über rot fahren.
- Jemand kauft solange Lose, bis er einen Treffer landet. $X^{(d)}$ sei die Anzahl gekaufter Lose.

- $X^{(e)}$ sei die Anzahl VolleyballspielerInnen in der Schweiz, die im nächsten Jahr einen Bänderriss erleiden werden.
- $X^{(f)}$ sei die Anzahl VolleyballspielerInnen aus einer Menge von 100 zufällig ausgewählten, die im nächsten Jahr einen Bänderriss erleiden werden.
- $X^{(g)}$ sei die Länge eines Wurms in einem faulen Apfel.

4. Zwei Lotterien bieten beide je n verschiedene Lose an. In jeder Lotterie gewinnt genau ein Los und ergibt einen Gewinn von je c Franken. Die Zufallsvariable X bezeichnet den Gewinn beim Kauf von zwei Losen der ersten Lotterie und die Zufallsvariable Y den Gewinn beim Kauf von je einem Los der beiden Lotterien.

a) Welche Werte können die Zufallsvariablen X und Y annehmen? Geben Sie die Verteilung von X und die Verteilung von Y an.

b) Berechnen Sie den Erwartungswert von X und von Y.

c) (Freiwillig!) Berechnen Sie die Varianz von X und von Y.

5. Vier Joghurtkostproben mit folgenden ungewöhnlichen Farb-Geschmackkombinationen wurden hergestellt:

Farbe	Geschmack
rot	Apfelsaft
braun	Himbeere
grün	Kaffee
gelb	Zitrone

25 Testpersonen mussten den Geschmack jeder Kostprobe angeben. Es ergab sich folgendes Ergebnis:

erkannt	Anzahl TesterInnen
0	0
1	2
2	15
3	8
4	0

(*Quelle: Genser, Moskowitz et.al (1977). Sensory response to food, sensory workshop*).
Erklärung: keine TesterIn hat alle Proben falsch beurteilt, zwei TesterInnen haben bei einer Probe den Geschmack richtig erkannt, fünfzehn TesterInnen bei zwei, etc.

a) Wir gehen davon aus, dass alle Testpersonen im Prinzip gleich gut sind im Erkennen von Joghurten. Wir wollen nun annehmen, dass die Anzahl richtig erkannter Geschmacksnoten bei jeder Testperson binomialverteilt ist. Welche implizite Annahme machen wir dabei?

b) Wie gross ist die mittlere richtig erkannte Anzahl Joghurte pro Person? Wenn die Testpersonen einfach raten würden, wäre das p der Binomialverteilung 1/4. Wie gross ist das p hier ungefähr?

c) Angenommen die Daten sind exakt binomialverteilt mit dem in b) bestimmten p. Wieviele von den 25 Versuchspersonen müssten 0,1,2,3,4 richtige Antworten geben? Vergleichen Sie diese Werte mit den beobachteten Daten (Ergänzen Sie die obige Tabelle mit einer Kolonne für die erwarteten Daten). Was muss man daraus schliessen? Was sind mögliche Begründungen?

Punkteverteilung:

Aufgabe 1: a) (1 Punkt) Verteilung.
b) (1 Punkt) Parameter und Überprüfung.

Aufgabe 2: a) (1 Punkt) Wahrscheinlichkeiten.
b) (1 Punkt) Wahrscheinlichkeit.

Aufgabe 3: (1 Punkt) 4 von 7 richtig.

Aufgabe 4: (2 Punkt) Verteilungen und Erwartungswerte.

Aufgabe 5: a) (1 Punkt) Kurze Beschreibung der impliziten Annahme.
b) (1 Punkt) Wahrscheinlichkeit p.
c) (1 Punkte) Tabelle. Schlussfolgerung aus Vergleich mit beobachteten Anzahlen.

Abgabe: 1 Woche nach der Übungsstunde in der Vorlesung.

Präsenz: Montag, 12:00-13:00, LEO C12.1.