Aufgabe 1 (10 Punkte)

Markieren Sie die jeweils einzig korrekte oder einzig nicht korrekte Aussage.

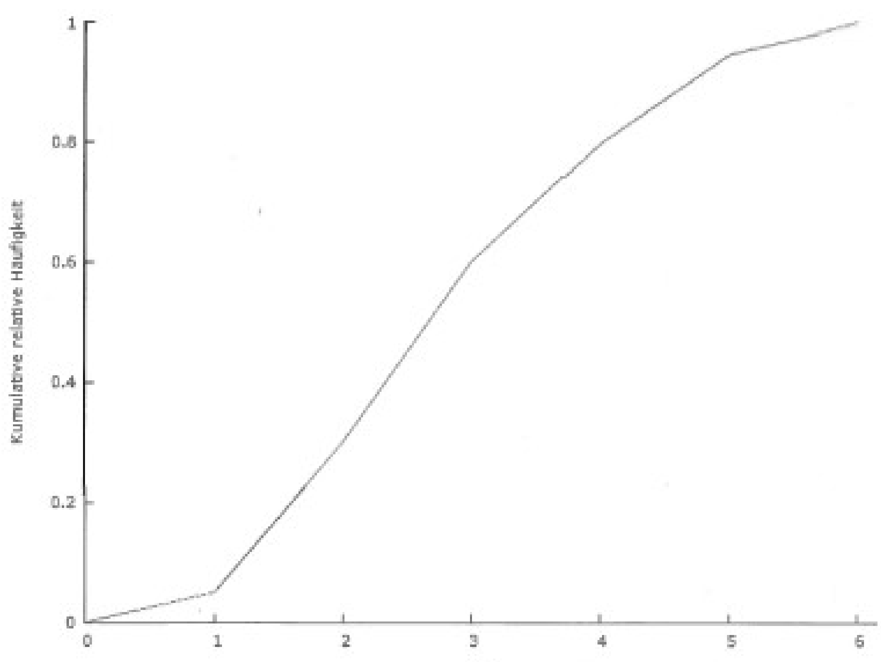
a) Welche Masse der zentralen Tendenz können für kategoriale, ordinalskalierte Daten verwendet werden? (2P)

- ☐ arithmetisches Mittel und Median.
- ☐ arithmetisches Mittel und geometrisches Mittel.
- ☐ arithmetisches Mittel und Modus.
- ☐ geometrisches Mittel und Median.
- ☐ Median und Modus.

b) Gegeben seien folgende Daten: 1, 2, 2, 1, 8, 3, 8. Welche Aussage ist korrekt? (3P)

- ☐ Interquartilsabstand = Spannweite.
- ☐ erstes Quartil = drittes Quartil.
- ☐ Standardabweichung > Varianz.
- ☐ Variationskoeffizient > Standardabweichung.
- ☐ Median > arithmetisches Mittel.

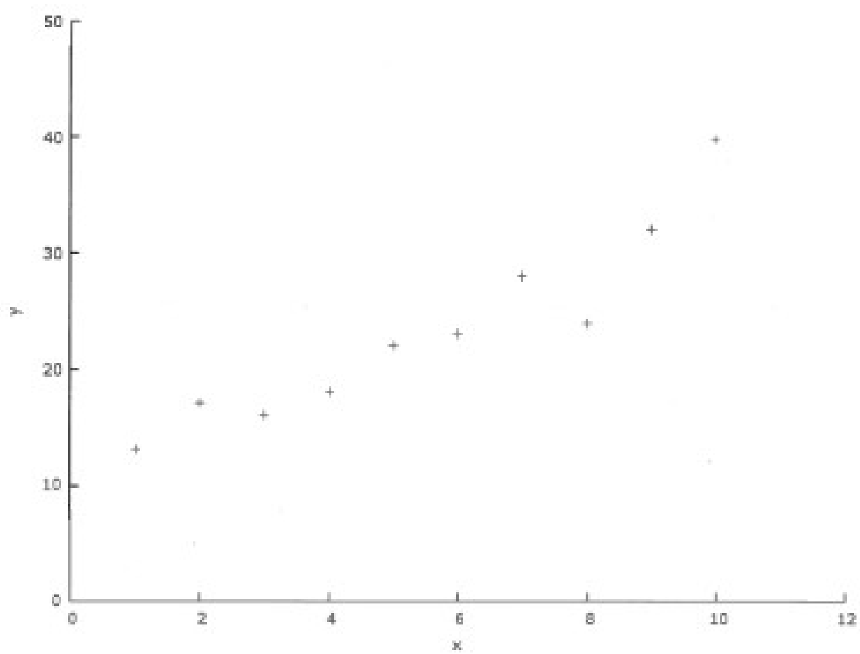
c) Die folgende Abbildung zeigt die Ogive einer metrisch stetigen Variablen X. Die Werte von X sind in sechs Klassen eingeteilt: 0 bis unter 1, 1 bis unter 2, 2 bis unter 3, 3 bis unter 4, 4 bis unter 5, 5 bis unter 6.



Welche Aussage ist korrekt? (2P)

- ☐ 60% der Werte von X sind zwischen 2 und 3.
- ☐ 40% der Werte von X sind mindestens 3.
- ☐ 20% der Werte von X sind zwischen 5 und 6.
- ☐ 60% der Werte von X sind kleiner als 4.
- ☐ 100% der Werte von X sind grösser als 1.

d) Die folgende Abbildung zeigt die Beziehung zwischen zwei metrisch stetigen Variablen X und Y: (3P)



Welche Aussage ist nicht korrekt? (Bezeichnungen: s_x^2 Varianz von X, s_y^2 Varianz von Y, s_{xy} Kovarianz von X und Y, r_{xy} Korrelationskoeffizient von X und Y)

- ☐ $s_x^2 > 0$.
- ☐ $s_x^2 < s_y^2$.
- ☐ $s_{xy} < 0$.
- ☐ $s_x^2 - s_y^2 < 0$.
- ☐ $r_{xy} > 0$.

Aufgabe 2 (10 Punkte)

Markieren Sie die jeweils einzig korrekte oder einzig nicht korrekte Aussage.

a) Angenommen, A und B sind unabhängige Ereignisse mit $P(A) = 0.4$ und $P(B) = 0.5$. Welche Aussage ist <u>nicht</u> korrekt? (2P)

- ☐ $P(A \cap B) = 0.2$
- ☐ $P(A|B) = 0.4$
- ☐ $P(B|A) = 0.5$
- ☐ $P(B) - P(A) = 0.1$
- ☐ $P(A \cup B) = 0.9$

b) Gegeben seien zwei Ereignisse A und B mit $P(A) = 0.4$, $P(\bar{B}) = 0.4$ und $P(\bar{A} \cap \bar{B}) = 0.3$. Welche Aussage ist korrekt? (3P)

- ☐ $P(A|B) = 0.3$
- ☐ $P(A|B) = 0.4$
- ☐ $P(A|B) = 0.5$
- ☐ $P(A|B) = 0.6$
- ☐ $P(A|B) = 0.7$

c) Die Tabelle zeigt die Verteilung einer diskreten Zufallsvariablen X:

x	$P(x)$
1	0.1
2	0.2
3	0.4
4	0.2
5	0.1

Es wird eine lineare Transformation durchgeführt: $Y = 6 - 2X$. Welche Aussage bezüglich Erwartungswert μ_Y und Varianz σ_Y^2 von Y ist korrekt? (3P)

- ☐ $\mu_Y = 0, \quad \sigma_Y^2 = 4.8$
- ☐ $\mu_Y = -4, \quad \sigma_Y^2 = 2.4$
- ☐ $\mu_Y = 0, \quad \sigma_Y^2 = 2.4$
- ☐ $\mu_Y = -4, \quad \sigma_Y^2 = 4.8$
- ☐ $\mu_Y = 0, \quad \sigma_Y^2 = -2.4$

d) Die Tabelle zeigt die Lohnklassen (monatliche Nettolöhne) von allen vollzeitangestellten Frauen und Männern in der Schweiz im Jahr 2010:

Lohnklasse (Schweizer Franken)	Wahrscheinlichkeit (in %) Frauen	Wahrscheinlichkeit (in %) Männer
0 - 4000	29.8	9.0
4001 - 6000	43.3	45.0
6001 - 8000	16.2	23.0
8001 und mehr	10.7	23.0
Total	100.0	100.0

Ein Mann und eine Frau werden zufällig aus der Grundgesamtheit ausgewählt. Die Wahrscheinlichkeit, dass mindestens eine Person mindestens CHF 8001 verdient, ist: (2P)

- ☐ 26%
- ☐ 31%
- ☐ 36%
- ☐ 41%
- ☐ 46%

Aufgabe 3 (10 Punkte)

Markieren Sie die jeweils einzig korrekte oder einzig nicht korrekte Aussage.

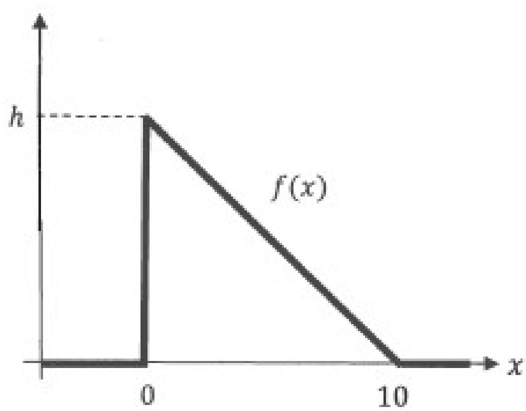
a) Gegeben sei eine stetige uniformverteilte Zufallsvariable X (Gleichverteilung) im Intervall $[-2, 2]$. Welche Aussage ist nicht korrekt? (2P)

- ☐ $E(X) = 0$
- ☐ $Var(X) > 0$
- ☐ $P(X = 0) = 0.1$
- ☐ $P(X < -1) = 0.25$
- ☐ $f(x) = 0.25$ für $-2 \leq x \leq 2$ (0 sonst)

b) Was ist das 3. Quartil einer normalverteilten Zufallsvariable $X \sim N(4,4)$? (2P)

- ☐ 5.35
- ☐ 5.25
- ☐ 5.15
- ☐ 4.95
- ☐ 4.75

c) Die stetige Zufallsvariable habe die folgende Dreiecksverteilung:



i. Die Höhe h der Dreiecksverteilung in der Abbildung ist: (3P)

- ☐ $1/20$.
- ☐ $1/10$.
- ☐ $1/5$.
- ☐ $2/5$.
- ☐ Mit den gegebenen Informationen nicht bestimmbar.

ii. Die kumulative Verteilungsfunktion $F(x)$ der Dreiecksverteilung in der Abbildung ist gegeben durch:

$$F(x) = 1 - \left(\frac{10 - x}{10}\right)^2$$

Der Median beträgt dann: (3P)

- ☐ 1.72
- ☐ 2.93
- ☐ 3.74
- ☐ 4.49
- ☐ 5.00

Aufgabe 4 (10 Punkte)

Markieren Sie die jeweils einzig korrekte oder einzig nicht korrekte Aussage.

a) Um die Intervallbreite eines 90%-Konfidenzintervals für den unbekannten Mittelwert μ einer (grossen) Grundgesamtheit zu halbieren, muss die Stichprobengrösse n: (2P)

- ☐ halbiert werden.
- ☐ verdoppelt werden
- ☐ verdreifacht werden.
- ☐ vervierfacht werden.
- ☐ verfünffacht werden.

b) Angenommen, das 90% Konfidenzintervall für den unbekannten Mittelwert μ der (grossen) Grundgesamtheit beträgt (1, 1.2). Wir testen die folgende Nullhypothese vs. Alternativhypothese: (2P)

$$H_0: \mu \leq 1$$
$$H_1: \mu > 1$$

Welche Aussage ist korrekt?

- ☐ Der p-Wert beträgt 20%.
- ☐ Der p-Wert beträgt 10%.
- ☐ Der p-Wert beträgt 5%.
- ☐ Der p-Wert beträgt 1%.
- ☐ Der p-Wert beträgt 0.1%.

c) Gegeben seien die folgenden vier Aussagen: (2P)

I. Die durchschnittliche Anzahl Jahre, die Vollzeitbeschäftigte vor der Pensionierung arbeiten, beträgt 40.

II. Die durchschnittliche Jahresgebühr von Privatschulen ist nicht 20'000 CHF.

III. Die durchschnittliche Anzahl Autos, die eine Person in ihrem Leben besitzt, ist nicht höher als 10.

IV. Das durchschnittliche Anfangssalär einer ZHAW Absolventin/eines ZHAW Absolventen beträgt mindestens 60'000 CHF.

Welche der obigen Aussagen bezeichnen eine Nullhypothese?

- ☐ (I) and (II)
- ☐ (II) and (III)
- ☐ (I), (II) and (III)
- ☐ (I), (II) and (IV)
- ☐ (I), (III) and (IV)

d) Wir testen die folgende Nullhypothese vs. Alternativhypothese: (2P)

$$H_0: \mu \geq 100$$
$$H_1: \mu < 100$$

i. Der Typ-I-Fehler (α-Fehler) ist gegeben, wenn:

- ☐ Wir schliessen, dass der Mittelwert μ der Grundgesamtheit kleiner ist als 100, wenn dieser tatsächlich mindestens 100 ist.
- ☐ Wir schliessen, dass der Mittelwert μ der Grundgesamtheit mindestens 100 ist, wenn dieser tatsächlich weniger als 100 ist.
- ☐ Wir schliessen, dass der Mittelwert μ der Grundgesamtheit grösser ist als 100, wenn dieser tatsächlich höchstens 100 ist.
- ☐ Wir schliessen, dass der Mittelwert μ der Grundgesamtheit grösser ist als 100, wenn dieser tatsächlich grösser ist als 100.
- ☐ Wir schliessen, dass die Nullhypothese nicht verworfen werden kann.

ii. Für eine Stichprobe der Grösse $n = 100$ erhalten wir den Stichprobenmittelwert $\bar{x} = 96$ und die Stichprobenvarianz $s^2 = 16$. (2P)

Unsere Entscheidung lautet:

- ☐ Die Nullhypothese kann nicht verworfen werden ($\alpha = 0.1$).
- ☐ Die Nullhypothese kann auf dem $\alpha = 0.1$ Signifikanzniveau verworfen werden, nicht aber auf dem $\alpha = 0.05$ Signifikanzniveau.
- ☐ Die Nullhypothese kann auf dem $\alpha = 0.05$ Signifikanzniveau verworfen werden, nicht aber auf dem $\alpha = 0.01$ Signifikanzniveau.
- ☐ Die Nullhypothese kann auf dem $\alpha = 0.01$ Signifikanzniveau verworfen werden.
- ☐ Kein Entscheidung ist möglich.

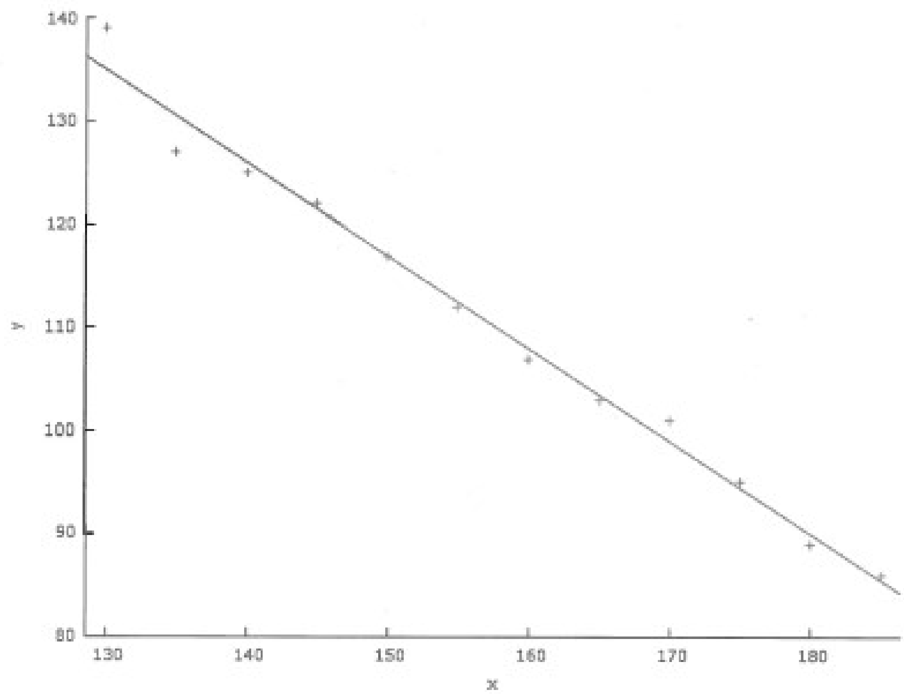
Aufgabe 5 (10 Punkte)

Markieren Sie die jeweils einzig korrekte oder einzig nicht korrekte Aussage.

a) Die OLS Regressionsgleichung $\hat{y} = b_0 + b_1 x$ erfüllt die folgende Eigenschaft: (2P)

- ☐ Die Summe der Residuen ist minimiert.
- ☐ Die Summe der Residuen ist null.
- ☐ Die Summe der quadrierten Residuen ist null.
- ☐ Die Summe der Werte der abhängigen Variable ist minimiert.
- ☐ Der Determinationskoeffizient R^2 ist minimiert.

b) Das lineare Regressionsmodell $y = \beta_0 + \beta_1 x + \varepsilon$ wird verwendet, um die Beziehung zwischen der abhängigen Variable y und der unabhängigen Variable x zu schätzen.



i. Bestimmen Sie anhand der Abbildung, welche der folgenden Parameterschätzungen korrekt sind: (3P)

- ☐ $b_0 = 137, \quad b_1 = -0.9, \quad s_e^2 = 36$
- ☐ $b_0 = 137, \quad b_1 = 0.9, \quad s_e^2 = 3.6$
- ☐ $b_0 = 252, \quad b_1 = -0.9, \quad s_e^2 = 3.6$
- ☐ $b_0 = 252, \quad b_1 = -0.9, \quad s_e^2 = 36$
- ☐ $b_0 = 252, \quad b_1 = 0.9, \quad s_e^2 = 3.6$

ii. Bestimmen Sie anhand der Abbildung, welche der folgenden Aussage korrekt ist: (2P)

- ☐ Das Modell erklärt 99% der Variation der unabhängigen Variablen.
- ☐ 1% der Variation der abhängigen Variablen wird durch das Modell nicht erklärt.
- ☐ Es liegt starke positive Korrelation zwischen der abhängigen und unabhängigen Variablen vor.
- ☐ Es liegt schwach negative Korrelation zwischen der abhängigen und unabhängigen Variablen vor.
- ☐ Es liegt keine Korrelation zwischen der abhängigen und unabhängigen Variablen vor.

c) Ein anderes lineares Regressionsmodell $y = \beta_0 + \beta_1 x + \varepsilon$ zeigt die Beziehung zwischen dem Medikament x (in Milligramm) und der Anzahl Erholungstage y von einer Allergie. Die folgenden zwei Summen sind für eine Zufallsstichprobe von $n = 10$ Patienten gegeben:

$\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y}) = 1008.9$, $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 368.1$

Der OLS Schätzung b_1 des Parameters β_1 ist: (3P)

- ☐ $b_1 = 3.44$
- ☐ $b_1 = 2.74$
- ☐ $b_1 = 2.14$
- ☐ $b_1 = 1.34$
- ☐ $b_1 = 0.34$

ENDE DER PRÜFUNG