Probeklausur: Statistik I

Kurzlösung, Rechenschritte sind vollständig anzugeben

Matrikelnummer: _____
Kontrollieren Sie die Vollständigkeit der Klausur. Sie besteht aus 5 Seiten.
Runden Sie alle Ergebnisse auf drei Nachkommastellen.
Geben Sie alle Berechnungen nachvollziehbar an!
Erlaubte Hilfsmittel: Nicht programmierbarer Taschenrechner

Aufgabe 1: Begriffe (12)

Erklären Sie die folgenden Fachbegriffe in wenigen Stichworten:

qualitativ: **nicht in Größen messbar**

Skalenniveau: **auf welcher Skala können die Werte gemessen werden?**

nominal: **nur Wert ermittelbar/interpretierbar**

ordinal: **Wert und Reihenfolge interpretierbar**

Aufgabe 2: Lage- und Streuungsmaße (15)

Bei einer Umfrage wird ein Zusammenhang zwischen wöchentlichen Überstunden und jährlichen Krankheitstagen gesucht.

Person	1	2	3	4	5
X: Überstunden	0	5	2	8	10
Y: Krankheitstage	1	5	2	3	7

Bestimmen Sie nachvollziehbar, falls möglich

2.1 das arithmetische Mittel der Überstunden. (4)

$\bar{x} = 5$

2.2 das arithmetische Mittel der Krankheitstage. (4)

$\bar{y} = 3{,}6$

2.3 den Modus der Krankheitstage. (2)

Jeder Wert, da alle Werte gleich häufig.

2.4 die Kovarianz zwischen Überstunden und Krankheitstagen. (5)

$\overline{xy} = 24{,}6$	(3)
$Cov(X,Y) = 6{,}6$	(2)

Aufgabe 3: Zusammenhänge (15)

Wie viele Krankheitstage sind auf Basis der vorliegenden Daten für eine Person mit 10 wöchentlichen Überstunden zu erwarten?

$\overline{x^2} = 38{,}6$	(2)
$Var(x) = 13{,}6$	(2)
$b \approx 0{,}485$	(3)
$a \approx 1{,}174$	(3)
$y = 0{,}485x + 1{,}174$	(2)
$y = 0{,}485 \cdot 10 + 1{,}174 = 6{,}024$	(3)

Aufgabe 4: Wahr oder falsch (9)

Geben Sie jeweils an, ob die Aussage wahr (w) oder falsch (f) ist.

4.1 Ordinale Daten sind beispielsweise Notenpunkte, weil 10 Punkte doppelt so viel wie 5 Punkte sind. **F**

4.2 Wenn es keine Korrelation gibt, hat die Regressionsgerade auch keine Steigung. **W**

4.3 Wenn etwas mit einer Wahrscheinlichkeit von 50% passiert, dann passiert es jedes zweite Mal. **F**

Aufgabe 5: Bedingte Wahrscheinlichkeit (15)

An einer international ausgerichteten Universität sind 40% der Stunden Ausländer.
50% aller Studenten sind weiblich. Von den Inländern sind 40% weiblich.

5.1 Erstellen Sie eine 4-Felder-Tafel (4)

	A	$\overline{A}$	*Summe*
W	26%	24%	50%
$\overline{W}$	14%	36%	50%
Summe	40%	60%	100%

Mit welcher Wahrscheinlichkeit

5.2 ist ein beliebig herausgegriffener Student weiblich? (2)

$P(W) = 50\%$ $(1)+(1)$

5.3 ist ein zufällig herausgegriffener männlicher Student Ausländer? (3)

$P(A|\overline{W}) = \frac{14\%}{50\%} = 28\%$ $(1)+(1)+(1)$

5.4 ist ein zufällig herausgegriffener ausländischer Student männlich? (3)

$P(\overline{W}|A) = \frac{14\%}{40\%} = 35\%$ $(1)+(1)+(1)$

5.5 ist ein zufällig herausgegriffener weiblicher Student Inländer? (3)

$P(\overline{A}|W) = \frac{24\%}{50\%} = 48\%$ $(1)+(1)+(1)$

Aufgabe 6: Verteilungen (20)

Im Rahmen einer Umfrage wurde in einem ersten Schritt in einer Szene Kneipe 30 Studenten befragt, anschließend wurde die Befragung in Hessen durchgeführt und abschließend in Indien. Thema war die Zahl der Sportinteressierten während eines dualen Studiums.

6.1 Definieren Sie die zu verwendende Zufallsvariable. (2)

X: Zahl der Sportinteressierten

6.2 In der Kneipe sind von den 300 anwesenden Studenten 200 Sportinteressiert. (6)
Mit welcher Wahrscheinlichkeit befragt man

-0 Sportinteressierte? $P(X=0)=\frac{\binom{200}{0}\cdot\binom{100}{30}}{\binom{300}{30}}\approx 0$ (0,5)+(1)+(0,5)

-20 Sportinteressierte? $P(X=20)=\frac{\binom{200}{20}\cdot\binom{100}{10}}{\binom{300}{30}}\approx 16{,}128\%$ (0,5)+(1)+(0,5)

-30 Sportinteressierte? $P(X=30)=\frac{\binom{200}{30}\cdot\binom{100}{0}}{\binom{300}{30}}\approx 0$ (0,5)+(1)+(0,5)

Hypergeometrische Verteilung

6.3 In Hessen sind 60% der Studenten sportinteressiert. Es werden 50 befragt. (6)
Mit welcher Wahrscheinlichkeit befragt man

-0 Sportinteressierte? $P(X=0)=\binom{50}{0}\cdot 60\%^{0}\cdot 40\%^{50}\approx 0$ (0,5)+(1)+(0,5)

-20 Sportinteressierte? $P(X=20)=\binom{50}{20}\cdot 60\%^{20}\cdot 40\%^{30}\approx 0{,}199\%$ (0,5)+(1)+(0,5)

-30 Sportinteressierte? $P(X=30)=\binom{50}{30}\cdot 60\%^{30}\cdot 40\%^{20}\approx 11{,}456\%$ (0,5)+(1)+(0,5)

Binomialverteilung

6.4 In Indien sind nur 0,1% der Studenten sportinteressiert. Es werden 1.000 befragt. (6)
Mit welcher Wahrscheinlichkeit befragt man

-0 Sportinteressierte? $\lambda=1.000\cdot 0{,}1\%=1\Rightarrow P(X=0)=\frac{e^{-1}}{0!}\approx 36{,}788\%$ (0,5)+(1)+(0,5)

-20 Sportinteressierte? $P(X=20)=\frac{e^{-1}}{20!}\approx 0\%$ (0,5)+(1)+(0,5)

-30 Sportinteressierte? $P(X=30)=\frac{e^{-1}}{30!}\approx 0\%$ (0,5)+(1)+(0,5)

Poissonverteilung

Aufgabe 7: Tests (14)

In einer Befragung unter Studenten soll die Arbeitszeit ihrer Kollegen im Job ermittelt werden. Die Unternehmensführung gibt an, dass ihre Mitarbeiter im Mittel mindestens 40 Stunden arbeiten

Die Auswertung der Angaben der Studenten liefert folgenden Output:

n=28
MITTELWERT: 38
STANDARDABWEICHUNG: 2

Zudem sind auch noch die folgenden beiden Quantilstabellen gegeben:

t-Verteilung

n	$\alpha = 0.60$	0.70	0.80	0.90	0.95	0.975	0.990
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457

Standardnormalverteilung

α	z_α
0.975	1.9600
0.970	1.8808
0.965	1.8119
0.960	1.7507
0.955	1.6954
0.950	1.6449

Zur Info: Formeln für zweiseitige Intervalle:

$$KI_{1-\alpha} = \left[\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right], KI_{1-\alpha} = \left[\bar{x} \pm t(n-1)_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right], KI_{1-\alpha} = \left[p \pm z_{1-\frac{\alpha}{2}} \frac{\sqrt{p\cdot(1-p)}}{\sqrt{n}}\right]$$

7.1 Stellen Sie Null- und Alternativhypothese auf. (2)

$H_0 : \mu \geq 40$
$H_1 : \mu < 40$

7.2 Ermitteln Sie jeweils den Annahmebereich zum gegebenen Signifikanzniveau α (8)

a. $\alpha = 5\%$.

$t(28-1)_{0,95} = 1{,}703$

$\Rightarrow \mu \in \left]-\infty; 38 + 1{,}703 \cdot \frac{2}{\sqrt{28}}\right] \Rightarrow \mu \in]-\infty; 38{,}644]$

(1)
(2)+(1)

b. $\alpha = 2{,}5\%$.

$t(28-1)_{0,975} = 2{,}052$

$\Rightarrow \mu \in \left]-\infty; 38 + 2{,}052 \cdot \frac{2}{\sqrt{28}}\right] \Rightarrow \mu \in]-\infty; 38{,}776]$

(1)
(2)+(1)

7.3 Treffen Sie die Testentscheidung bei einem Signifikanzniveau von (4)

a. $\alpha = 5\%$.
Ablehnen
b. $\alpha = 2{,}5\%$.
Ablehnen

Erläuterung
Annehmen
$\bar{X} \geq \mu$ $\underbrace{-\sigma_{\bar{X}} z_{1-\alpha}}$
wegen Schwankungen kann H_0 stimmen, obwohl $\bar{X}$ kleiner als μ
$\mu \leq \bar{X} + \sigma_{\bar{X}} z_{1-\alpha}$