

Probe-Test SA.01

Donnerstag, 20. 12. 2018

Dauer: 90 Minuten

Name, Vornam	e:					
IDS Nr.:	(Auf Rückseite HS	(Auf Rückseite HSLU-Karte) L				
Unterschrift:						
	Aufgabe	1	2	3	4	Total
	max. Punkte	5	10	10	10	35
	erreichte Punktzahl					

Wichtige Hinweise

- Zugelassene Hilfsmittel:
 - Taschenrechner
 - Zusammenfassung: 10 einseitige A4-Seiten
 - Beliebige käufliche Formelsammlung
 - Papier und Schreibzeug
- Der Lösungsweg muss vollständig schriftlich festgehalten werden.
- Alle Antworten müssen begründet werden. Ohne Lösungsweg gibt es 0 Punkte!

Viel Erfolg! Peter Büchel

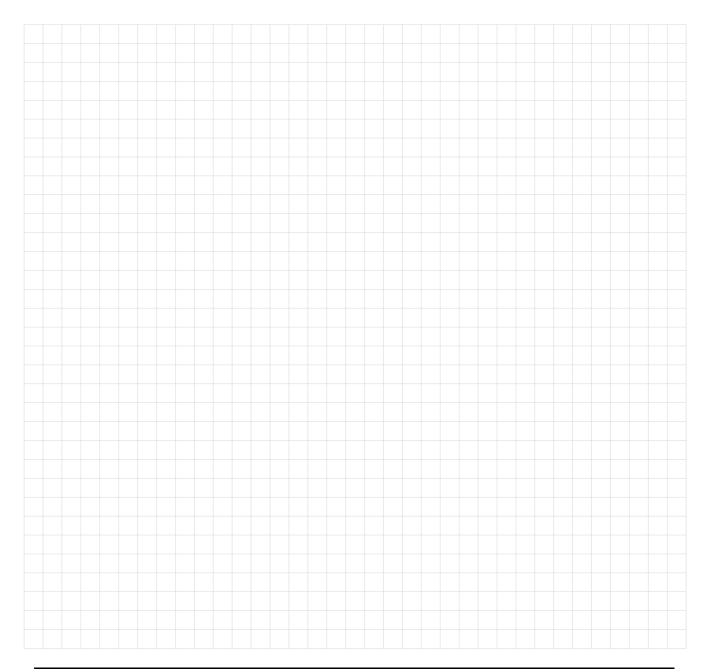


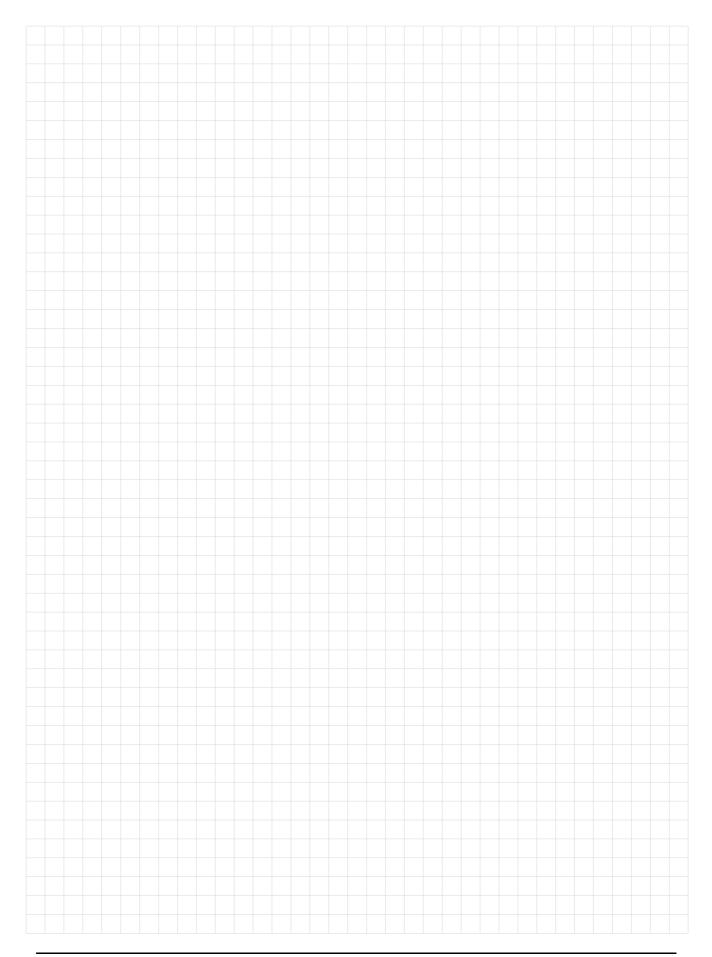
Aufgabe 1:(5 Punkte)

Nach einem internationalen Sportwettkampf werden Dopingtests durchgeführt. Dabei wird ein neues Schnelltestverfahren eingesetzt. Falls ein Sportler gedopt ist, dann ist der Schnelltest mit einer Wahrscheinlichkeit von 90 % positiv. Falls ein Sportler in Tat und Wahrheit nicht gedopt ist, zeigt der Schnelltest dennoch in 5 % der Fälle ein positives Ergebnis. Aus Erfahrung wissen die Dopingkontrolleure, dass 10 % der Sportler gedopt sind. Wir betrachten die folgenden Ereignisse:

D: Der Sportler ist gedopt $\overline{D}:$ Der Sportler ist nicht gedopt $T^+:$ Der Schnelltest ist positiv $T^-:$ Der Schnelltest ist negativ

Wie gross ist die Wahrscheinlichkeit, dass ein Sportler gedopt ist, wenn der Test negativ ist?





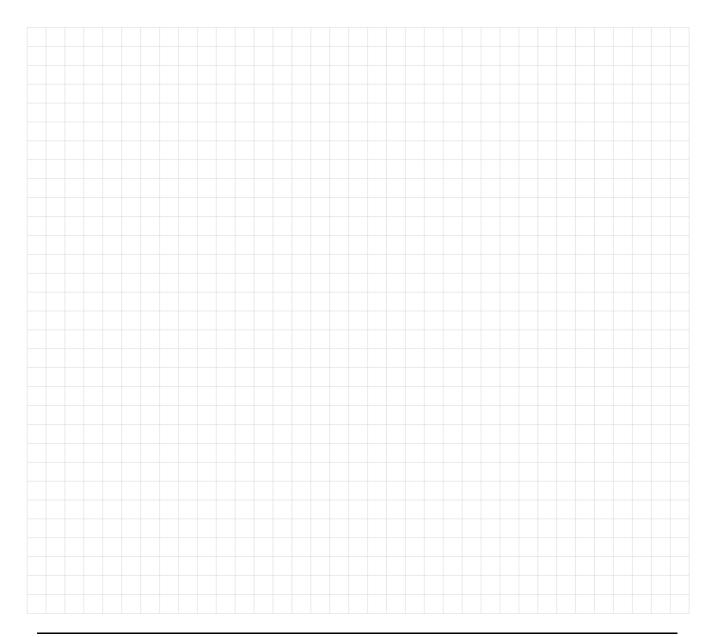


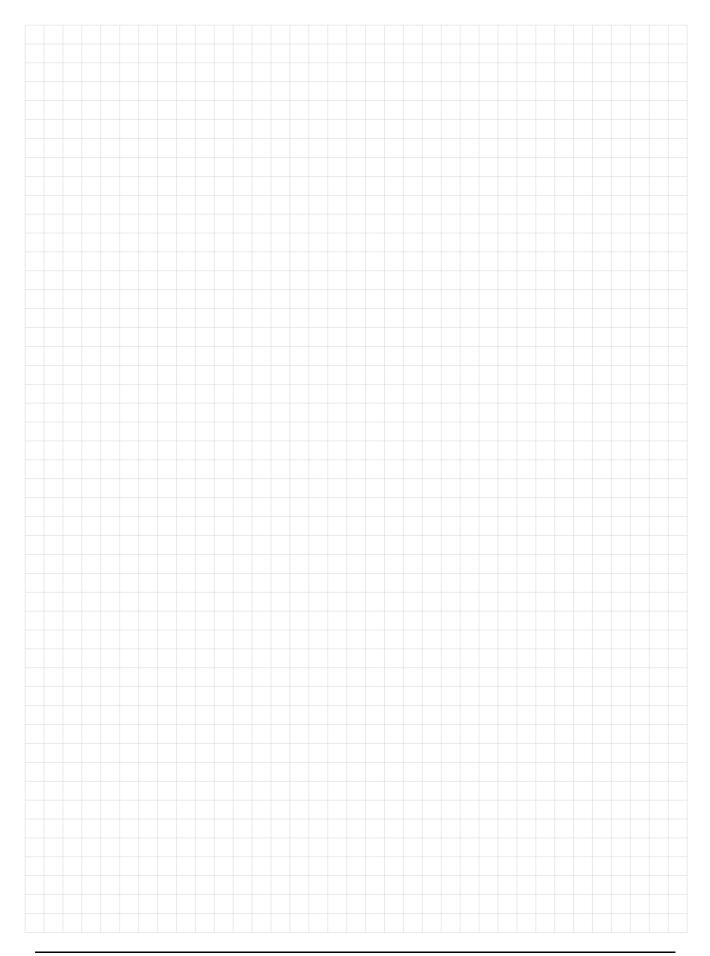
Aufgabe 2:	.(10 Punkte)
------------	--------------

Man vermutet, dass Cholesterinwerte bei Männern mit zunehmenden Alter höher werden. Die Ergebnisse einer Untersuchung der Cholesterinwerte von je 11 Männern in den Altersgruppen 20-30 und 40-50 liegen im Datensatz **cholesterin** vor. Gibt es einen statistisch signifikanten Unterschied zwischen den beiden Altersgruppen?

```
cholesterin <- read.csv("cholesterin.csv")</pre>
```

- a) Handelt es sich um einen gepaarten oder ungepaarten Test? Begründen Sie Ihre Antwort
- b) Welche Art von Test würden Sie hier verwenden? Begründen Sie Ihre Antwort.
- c) Führen Sie diesen Test durch. Formulieren Sie Null- und Alternativhypothese und fällen Sie den Testentscheid.





Aufgabe 3:(10 Punkte)

Im Datensatz **creativity** ist das Alter der höchsten Kreativität von 36 Künstlern aufgeführt. Diese werden in 3 *Kategorien* aufgeteilt: Novelisten, Poeten und Filmregisseure (directors). Zusätzlich wird zwischen zwei *Typen* unterschieden: Conceptualists (finders) / experimentalists (seekers).

Das Kriterium für das Alter der grössten Kreativität:

- Poeten: Medianalter für ihre 5 der am meisten gedruckten Gedichten
- Novelisten: Medianalter der Publikation der 3 wichtigsten Romane
- Regisseure: Medianalter für die Filme mit den höchsten Filmbewertungen

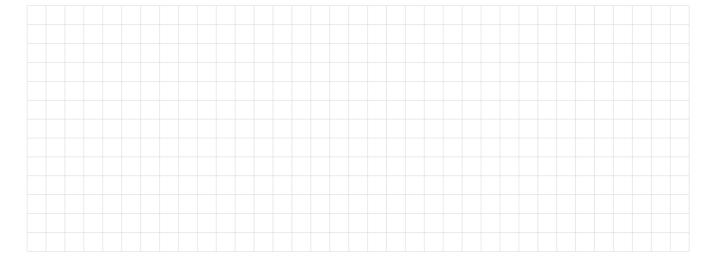
```
df <- read.csv("creativity.csv")
head(df)</pre>
```

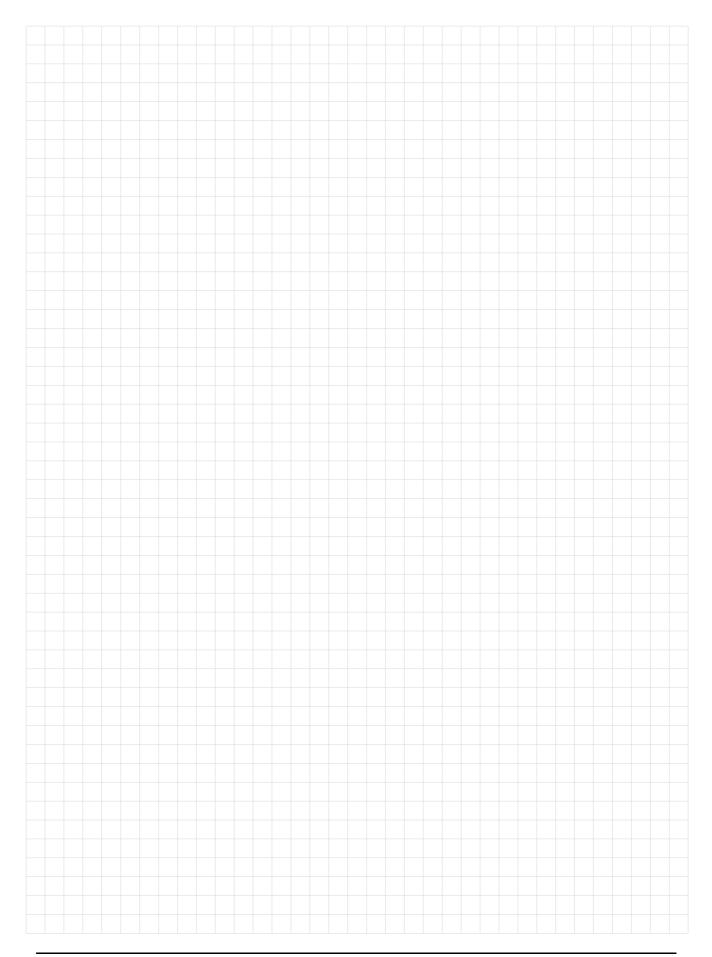
a) Erstellen Sie je ein Boxplot, der das Alter der höchsten Kreativität mit der Kategorie bzw. Typ zeigt. Interpretieren Sie diese beiden Boxplots.

```
boxplot(..., data=df,col=c("lightblue", "forestgreen", "orange"))
```

b) Erstellen Sie einen Interaktionsplot mit Typ, Kategorie und Alter der höchsten Kreativität. Interpretieren Sie diesen Plot.

- c) Gibt es statistisch signifikante Unterschiede (ohne Interaktion) zwischen dem Alter der höchsten Kreativität und Kategorie und Typ?
 Stellen Sie Null- und Alternativhypothese auf und fällen Sie den Testentscheid.
- d) Gibt es statistisch signifikante Interaktion zwischen Kategorie und Typ? Stellen Sie Null- und Alternativhypothese auf und fällen Sie den Testentscheid.







Aufgabe 4:(10 Punkte)

Forscher waren an den Faktoren interessiert, die den Wasserverbrauch (consumption) in 48 US-Staaten beeinflussen. Für jeden Staat wurde der Wasserverbrauch pro Person (in Gallonen pro Tag), das Einkommen pro Kopf (in \$1000), der jährliche durchschnittliche Niederschlag (in inches) und die mittleren Kosten des Wassers von 1000 Gallonen (in Dollar) gemessen. Die Daten sind im Datensatz water aufgeführt

```
df <- read.csv("water.csv", header = T)</pre>
```

- a) Stellen Sie formelmässig ein multiples Regressionsmodell mit dem Wasserverbrauch als Zielvariablen und den restlichen Variablen als Prädiktoren auf.
- b) Bestimmen Sie die Parameter dieses Modelles und interpretieren Sie diese.
- c) Welcher Anteil der Variation wird durch das Regressionsmodell erklärt?
- d) Interpretieren Sie den *p*-Wert des zugehörigen *F*-Wertes.
- e) Wir betrachten nun die einzelnen Regressionskoeffizienten. Deuten die Resultate an, dass wir irgendwelche Variablen vom Modell entfernen können? Begründen Sie Ihre Antwort.

