Lade bitte bis Donnerstag 19.10.2017 20:00 den `R-Code` zu den jeweiligen Aufgaben in `moodle` hoch.

Bitte Name und die jeweilige Beispielnummer im Code notieren. Der Code soll so programmiert werden, dass das (richtige) Ergebnis in der Console ausgegeben wird. Alle Beispiele können und sollen mit den Funktionen der Kapitel 1-10 gelöst werden.

Die Internetplattform `neuwal.com` sammelt Umfrageergebnisse zu Wahlen. Die Umfrageergebnisse für die Nationalratswahlen am 15.Oktober 2017 inklusive Umfragezeitpunkt und Stichprobenumfang stehen in der Datei `neuwal.RDATA` zur Verfügung. Für alle, die die österreichische Innenpolitik nicht im Detail verfolgen: es stehen 6 größere Parteien zur Wahl, wobei die Liste Pilz erst im Juli gegründet wurde und somit nicht in allen Umfragen vorkommt.

1. a) Lade die Datei `neuwal.RDATA` mit Spezifikation des gesamten Pfades. Zeige die vorhandenen Objekte an, gib die Anzahl der Objekte aus und lösche anschließend alle vorhandenen Objekte (die Datei `neuwal.RDATA` enthält mehrere R-Objekte, mit Namen wie `oevp`, `gruene`, etc.).
   b) Lade das Objekt `neuwal.RDATA` mit Hilfe der Funktion `setwd()`
   c) Überprüfe ob alle Objekte dieselbe Länge ausweisen. Die Überprüfung soll mittels R-Befehlen durchgeführt werden und TRUE oder FALSE zurückgeben.
   d) Die Umfrageergebnisse wurden mit Stand 8.10.2017 abgerufen. Füge alle aktuelleren Ergebnisse hinzu.
   e) Berechne sinnvolle deskriptive Statistiken für die Stichprobengröße und speichere sie in einem Vektor.
   f) Berechne den Anteil der Stimmen, die auf sonstige Parteien entfallen und speichere das Ergebnis in einem Vektor.
2. a) Berechne die Korrelation zwischen dem Ergebnis der SPÖ und der Stichprobengröße. Vergleiche die Korrelationskoeffizienten nach Pearson und Spearman.
   b) Berechne den Mittelwert für die Umfragewerte der SPÖ für kleine und große Stichproben. Kleine Stichproben sind jene mit n $<$ 1.Quartil der Verteilung der Stichprobengröße, große Stichproben jene mit n $>$ 3.Quartil. Programmiere möglichst allgemein!
   c) Berechne den Mittelwert und kumulativen Mittelwert (ausgehenden von der ältesten Umfrage) der Umfrageergebnisse für die ÖVP. Wie oft ist der kumulative Mittelwert größer als der Mittelwert?
   d) In wie viel Prozent der Umfragen liegt die FPÖ vor der SPÖ? Schränke anschließend auf Umfragen seit September ein!

e) Wie stark veränderte sich das Ergebnis für die SPÖ von Umfrage zu Umfrage? Was war die stärkste Veränderung, ist sie positiv oder negativ?

f) Wie hoch war die zweitstärkste Veränderung? Gib das Ergebnis inklusive Vorzeichen aus.

g) Gib Tag und Monat der Umfragen aus, bei denen die NEOS vor den Grünen lagen.

h) Gib im vorigen Beispiel das Ergebnis im Format "tag.monat" zurück.

i) Gib im vorvorigen Beispiel das Ergebnis im Format TT.MM zurück, wobei bei einstelligen Tagen und Monaten eine 0 vorangestellt werden soll.

3. Zu Konfidenzintervallen etc. stehen auf der nächsten Seite Hintergrundinformationen.

   a) Berechne für jede Umfrage ein $100 \cdot (1\text{-}\alpha)$ % - Konfidenzintervall für das Ergebnis der ÖVP. Gib das Ergebnis in Prozent folgendermaßen aus: `[unter Grenze;obere Grenze]`, auf ganze Prozent gerundet. Allgemein (vor allem in Bezug auf α) programmieren!

   b) Berechne für jede Umfrage den Standardfehler des Anteilwerts der ÖVP. Berücksichtige dafür, dass ein Teil der Befragten sich noch nicht entschieden hat. Nimm an, dass dieser Anteil 20 % beträgt und sich somit die Stichprobengröße um 20 % reduziert.

   c) Berechne den mittleren Standardfehler aus dem letzten Beispiel und die Standardabweichung der Ergebnisse der ÖVP. Schränke dabei auf Umfragen seit August ein. Interpretiere das Ergebnis aus statistischer Sicht![1]

   d) Modifiziere den Code des vorigen Beispiels so, dass er für jede andere Partei möglichst einfach anwendbar ist.

   e) Angenommen der Anteil der Unentschiedenen verringert sich pro Monat um 2 Prozentpunkte und beginnt im April mit 25 Prozent. Berechne die effektive Stichprobengröße (= Anzahl der deklarierten Personen) pro Umfrage!

4. Verwende für die folgenden Beispiele eine geeignetes Signifikanzniveau α.

   a) Teste mit der Funktion `prop.test()`, ob sich der Anteil der FPÖ-Wähler zwischen den letzten beiden Umfragen signifikant geändert hat. Verwende n als Stichprobengröße.

   b) Wie oben, nur mit 2 zufällig gewählten Umfragen.

   c) Bilde den Mittelwert der FPÖ-Ergebnisse über alle Umfragen im Oktober bzw. im September und teste, ob sich die Anteile zwischen September und Oktober signifikant verändert haben.

[1] Links zum Thema:
`https://derstandard.at/2000064226139/Die-Wahlumfragen-aehneln-einander-verdaechtig?_blogGroup=1`
`https://fivethirtyeight.com/features/heres-proof-some-pollsters-are-putting-a-thumb-on-the-scale/`

Hintergrundinformation: Standardfehler und Konfidenzintervall des Anteilswerts

X misst die Anzahl der Erfolge eines Zufallsexperiment mit:

- n *unabhängigen* Versuchen
- 2 möglichen Ausgängen: Erfolg (1), Misserfolg (0)
- einer Wahrscheinlichkeit für Erfolg für jeden Versuch von p und der Wahrscheinlichkeit für Misserfolg $q = 1 - p$
- Anzahl an Erfolgen X

Dann ist X binomialverteilt: $X \sim B(n, p)$

Der **Schätzer** für den wahren Anteilswert p ist:

$$\hat{p} = \frac{X}{n}$$

Der **Standardfehler** des Schätzer $\hat{p}$ beträgt:

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Das $100 \cdot$ (1-α) % **Konfidenzintervall** für den Anteilswert p ist:

$$[\hat{p} - Q^N(1-\alpha/2)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + Q^N(1-\alpha/2)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$$

wobei $Q^N(1-\alpha/2)$ das 1-α/2-Quantil der Standardnormalverteilung bezeichnet.

Dieses Intervall überdeckt den tatsächlichen Anteil p mit einer Wahrscheinlichkeit von $100 \cdot$ (1-α) %, d. h. *im Mittel* enthalten $100 \cdot$ (1-α) % der Konfidenzintervalle den wahren Parameter p.

Anmerkungen: das Konfidenzintervall für den Anteilswert verwendet eine Normalverteilungsapproximation. Die Approximation ist umso besser, je größer die Stichprobengröße ist und je näher p beim Wert 0.5 liegt. Eine zufriedenstellende Approximation wird für $np(1-p) > 9$ erreicht.