**Übungsaufgaben Biostatistik**

**Verwendung des Datensatzes USsample:**

Der Datensatz USsample basiert auf einer Fall Kontroll Studie mit Fällen, die definiert sind durch einen Body Mass Index von >= 30 (=severe obesity) oder durch den Erhalt eines Magenbypass. Dazu wurden zufällig Kontrollen aus derselben geographischen Region rekrutiert. Der Datensatz beinhaltet die folgenden Variablen: 

Kopieren sie alle Ergebnisse (Grafiken, evtl Tabellen etc.) in ihre Ergebnisdatei und beantworten sie in ganzen Sätzen die gestellten Fragen!

**Aufgabe 1:**

Es soll nun analysiert werden, ob sich der Gesamtcholesterinspiegel (Variable: Total\_Cholesterol) zwischen den Raucherkategorien unterscheidet. Stellen sie den Unterschied zuerst graphisch an (Boxplots des Gesamtcholesterinspiegels geteilt nach Raucherkategorien). Testen sie diesen Unterschied anschließend mit einer ANOVA.

Falls die durchgeführte ANOVA ein signifikantes Ergebnis liefert, führen sie paarweise T-Tests durch (mithilfe der Methode für multiples Testen „Tukey“). Bei welchem Gruppenvergleich kann man nun einen Unterschied sehen?

**Aufgabe 2:**

Im Datensatz Germansample.sav kann man sehen, dass sich das Rauchverhalten zwischen Männern und Frauen unterscheidet, aber dass dies nicht mehr der Fall ist, wenn man die Gruppe der unter 40-Jährigen betrachtet.

Testen Sie, ob dieser Zusammenhang in dem US-sample zwischen Geschlecht und Rauchverhalten (bei allen und der Gruppe der unter 40-Jährigen) vergleichbar mit dem Ergebnis aus dem Germansample Datensatz ist.

**Aufgabe 3:**

Wie sind die Lipide (HDL\_Cholesterol und Total\_Cholesterol) und die Variable ALT (Liver enzyme (Alanine-Aminotransferase)) verteilt? Prüfen Sie dies mithilfe von Histogrammen und QQ-Plots.

**Aufgabe 4:**

Sie haben nun gesehen, dass die Variable ALT schief verteilt ist. Um ALT zwischen den Raucherkategorien vergleichen zu können, brauchen Sie also einen nichtparametrischen Test. Führen Sie den geeigneten Test durch.

**Aufgabe 5:**

In Publikationen zu Fall Kontroll Studien ist es üblich, verschiedene Variablen zwischen Fällen und Kontrollen zu unterscheiden. Man verwendet dazu Lokalisationsmaße (Mittelwert +/- Standardabweichung, zusätzlich Quantile für nicht-normalverteilte Variablen), absolute und relative Häufigkeiten für kategorielle Variablen und wendet die dazu passende Teststatistik an, um zu testen ob es einen Unterschied zwischen Fällen und Kontrollen in Bezug auf diese Variablen gibt.

**Hier eine Beispieltabelle:**

Sie sehen ein Beispiel für verschiedene Testsituationen: T-Test, Chi-Quadrat Test und Wilcoxon Test:



**Füllen Sie nun die entsprechende Tabelle aus** (mit zwei Nachkommastellen):

Verwenden sie die bereits bestehende Variable „Case\_Control“, um zwischen Fällen und Kontrollen zu unterschieden.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Utah obesity case-control study** | | **Statistical test used** | **Comparison between groups p-value** |
|  | **Cases**  **(n=)** | **Controls**  **(n=)** |
| Age in years | ± | ± |  |  |
| Total cholesterol, mg/dL | ± | ± |  |  |
| HDL cholesterol, mg/dL | ± | ± |  |  |
| Alanine-Aminotransferase ALT, U/L  (25th, 50th, 75th percentile) | ±  ( ; ; ) | ±  ( ; ; ) |  |  |
| Diabetes mellitus, yes/no, n (%) | /  ( %/ %) | /  ( %/ %) |  |  |