Aufgabe 1.

a) Stetige Gleichverteilung

Histogramm oder Q-Q-Norm-Plot eignen sich gut dazu, das Vorliegen oder nicht Vorliegen einer Normalverteilung zu erkennen.

Histogramm: absolute Häufigkeit ist in allen Klassen etwa gleich – die Daten sind nicht normalverteilt.

Q-Q-Norm-Plot: starke Abweichungen von der Geraden an den Rändern – die Daten sind nicht normalverteilt.

Box-Plot: nahezu symmetrisch, aber eher nicht normalverteilt, ansonsten wären in dem mittleren Bereich mehr Daten vorhanden, dadurch sollte Quartilsabstand kleiner sein; kein Ausreißer.

Gestapeltes Punktdiagramm: keine Bindungen.

Shapiro-Wilk Test:
H_0 :„Die Daten sind normalverteilt“ $|H_A$:„Die Daten sind nicht normalverteilt“

$p = 0.001288 < \alpha = 0.05 \rightarrow H_0$ wird abgelehnt. Die Verteilung der Daten unterscheidet sich signifikant von einer Normalverteilung.

b) Exponentialverteilung

Alle Grafiken deuten auf eine schiefe Verteilung (rechtschief). Die Daten sind nicht normalverteilt.

Histogramm: rechtsschiefe Verteilung.
Q-Q-Norm-Plot: leichte Bogenform $\rightarrow$ schiefe Verteilung.
Box-Plot: es gibt Ausreißer, einer davon ist ein starker.
Gestapeltes Punktdiagramm: keine Bindungen vorhanden.

Shapiro-Wilk Test:
H_0 :„Die Daten sind normalverteilt“ $|H_A$:„Die Daten sind nicht normalverteilt“
$p = 3.557 \cdot 10^{-8} < \alpha = 0.05 \rightarrow H_0$ wird abgelehnt. Die Verteilung der Daten unterscheidet sich signifikant von einer Normalverteilung.

c) Normalverteilung

Alle Grafiken deuten auf eine nahezu symmetrische Verteilung. Normalverteilung möglich.

Histogramm: symmetrische Verteilung.
Box-Plot: symmetrische Verteilung; es gibt Ausreißer.
Q-Q-Norm-Plot: entspricht gut der vorgegebenen Geraden; einzelne Ausreißer an den Rändern.

Gestapeltes Punktdiagramm: Häufung der Punkte im zentralen Bereich; keine Bindungen vorhanden.

Shapiro-Wilk Test:
H_0 :„Die Daten sind normalverteilt“ $|H_A$:„Die Daten sind nicht normalverteilt“
$p = 0.5687 > \alpha = 0.05 \rightarrow H_0$ wird angenommen. Die Verteilung der Daten unterscheidet sich nicht signifikant von einer Normalverteilung.

d) Weibullverteilung
Histogramm: leicht rechtsschiefe Verteilung. Normalverteilung möglich.
Q-Q-Norm-Plot: leichte Abweichungen von einer Geraden; Normalverteilung möglich.
Box-Plot: Verteilung ist nahezu symmetrisch; es gibt Ausreißer nach oben.
Gestapeltes Punktdiagramm: in der Mitte und links sind mehr Werte als rechts – leicht rechtsschiefe Verteilung; keine Bindungen vorhanden.

Shapiro-Wilk Test:
H_0 :„Die Daten sind normalverteilt“ $|H_A$:„Die Daten sind nicht normalverteilt“
$p = 0.04339 < \alpha = 0.05 \rightarrow H_0$ wird abgelehnt. Die Verteilung der Daten unterscheidet sich signifikant von einer Normalverteilung.

Aufgabe 2.

a)
i) Datenmatrix.
ii) X_1 - Mehrschlaf (Differenz zwischen nach und vor Einnahme);
X_2 - Gruppe (Kontrolle (Placebo), Behandlung).
iii) X_1 - stetig, Intervallskala;
X_2 - diskret, Nominalskala (dichotom).
iv) Die Daten sind repräsentativ für die Grundgesamtheit, falls jedes Individuum
- 1) unabhängig von einander und
- 2) mit der gleichen Wahrscheinlichkeit
ausgewählt wird.
v)
• „Patienten in der Klinik mit Schlafstörungen“ – hier könnte das als Grundgesamtheit betrachtet werden, da die Stichprobe repräsentativ ist (angenommen, dass die Unabhängigkeit auch gegeben ist);
• „Alle Patienten mit Schlafstörungen“ – könnte nicht als Grundgesamtheit betrachtet werden, weil es keine Information über Schlafstörung Grade der Patienten dieser Klinik gibt und ob das für diese Grundgesamtheit repräsentativ ist.
iii) **Lagevergleich.**

• Mehrschlaf ist in der Gruppe „Behandlung“ im Durchschnitt größer als in der Gruppe „Kontrolle“:

• Median bei Gruppe „Behandlung“ ist größer als bei Gruppe „Kontrolle“ und liegt ungefähr bei oberen Viertelwert der Gruppe „Kontrolle“;

• Maximum bei der Gruppe „Behandlung“ ist größer als Maximum bei Gruppe „Kontrolle“;

• Minimum der Gruppe „Behandlung“ ist ungefähr unterer Viertelwert bei der Gruppe „Kontrolle“.

Streuungsvergleich.

Etwas größere Streuung in der Gruppe „Behandlung“:

• Quartilsabstand (Größe der Box) ist in der Gruppe „Behandlung“ größer als bei Gruppe „Kontrolle“;

• Spannweite ist nahzu gleich.

Symmetrie. Verteilung in beiden Gruppen ist leicht rechtsschief (linkssteil).

Aufgabe 3.

a) X_1 - Teufe (Tiefe) / stetig, Verhältnisskala
X_2 - Type: Poren, Kluft / diskret, Nominal (dichotom)
X_3 - Transmissivität (Durchlässigkeit) / stetig, Verhältnisskala

b) **Q-Q-Norm-Plot:**
$X_1, log X_1, log X_3$ - könnten normalverteilt sein.
bei X_3 - keine normalverteilten Daten.

Histogramm:
$X_1, log X_3$ - können normalverteilt sein, bei X_3 - keine Normalverteilung.

Box-Plot:

Teufe (für Kluft und Poren gemeinsam): kein Ausreißer vorhanden, leicht rechtschiefe Verteilung.

Teufe (parallele Box-Plots für Kluft und Poren): bei Typ „Poren“ gibt es einen Ausreißer; Median bei Poren ist größer als bei Kluft; Teufe streut in Kluft mehr als in Poren; bei Kluft Verteilung ist rechtsschief; bei Poren – linksschief.

Transmissivität (für Kluft und Poren gemeinsam): zwei Ausreißer; rechtsschiefe Verteilung.

Transmissivität (parallele Box-Plots für Kluft und Poren): die entsprechenden logarithmierten Daten werden analysiert.

log(Transmissivität) (für Kluft und Poren gemeinsam): kein Ausreißer vorhanden, nahezu symmetrische Verteilung.

log(Transmissivität) (parallele Box-Plots für Kluft und Poren): bei Typ „Poren“ ein Ausreißer; log(Transmissivität) bei Poren ist kleiner als bei Kluft; nahezu symmetrische Verteilung bei Kluft und Poren.

Streudiadramm:

Die Werte log(Transmissivität) ist von der Teufe abhängig. Für jeden Typ gilt: je größer die Teufe ist, umso kleiner die log(Transmissivität). Dabei ist der Zusammenhang für jeden Typ annähernd linear (schwarz in der Kluft, rot in den Poren).