

6. Lösungen weitere Übungsaufgaben Statistik für Ingenieure WiSe 19/20

1. Aufgabe:

- a) Lage: Die mittlere Druckfestigkeit ist in Aachen und Dresden höher als bei den anderen Herstellungsorten
- Mediane von Kassel, Karlsruhe und Leipzig liegen in einem ähnlichen Bereich
 - Mediane von Aachen und Dresden liegen in einem ähnlichen Bereich
 - Die mittlere Druckfestigkeit ist in Aachen und Dresden höher als bei den anderen Herstellungsorten
 - Median von Darmstadt ca. $5 - 10 N/mm^2$ höher als bei Kassel, Karlsruhe und Leipzig, aber unter Aachen und Dresden

Streuung: Die Druckfestigkeiten streuen in Aachen und in Kassel stärker als bei den anderen Herstellungsorten.

- IQR von Aachen, Karlsruhe und Kassel etwa gleich groß
- IQR bei Darmstadt, Dresden und Leipzig etwa gleich und deutlich kleiner als bei Aachen, Karlsruhe und Kassel
- Whisker bei Aachen und Kassel sehr groß, bei Darmstadt sehr klein

Sonstiges:

- Es gibt deutlich mehr Daten für Kassel. Das ist aus den Grafiken nicht ersichtlich.
- Bis auf Aachen und Karlsruhe liegen annähernd symmetrische Verteilungen vor.
- Die Verteilung der Druckfestigkeiten in Aachen ist rechtsschief.
- Kassel: Drei Ausreißer nach unten

- b) Lage: Die mittlere Festbetondichte ist in Dresden höher und in Darmstadt niedriger als an den anderen Standorten.
- Mediane von Aachen, Karlsruhe und Kassel sowie Leipzig in ähnlichem Bereich
 - Median von Dresden liegt deutlich höher (etwa $2,53 g/cm^3$).
 - Median von Darmstadt liegt deutlich niedriger (etwa $2,47 g/cm^3$)

Streuung: Es gibt 3 Herstellungsorte (Darmstadt, Dresden, Kassel) mit vergleichsweise hoher Streuung und 3 Orte (Aachen, Karlsruhe, Leipzig) mit geringerer Streuung.

- IQR und Whisker von Darmstadt, Dresden, Kassel sehr groß (starke Streuung)
- IQR und Whisker von Leipzig, Aachen und Karlsruhe klein (geringere Streuung)
- mit Ausnahme von Dresden ist der untere Whisker länger als der obere

Sonstiges:

- Aachen und Karlsruhe weisen stark unsymmetrische Verteilungen auf
- Die Verteilung in Aachen ist linksschief.
- Die Verteilung in Karlsruhe ist eher rechtsschief.
- Darmstadt, Leipzig und Kassel weisen eine leicht unsymmetrische Verteilung auf
- zugrunde liegende Datenanzahl nicht ersichtlich
- Die Box für Leipzig ist zu klein, deswegen ist der Plot unübersichtlich.
- Karlsruhe: zwei Ausreißer nach unten

2. Aufgabe:

- a) Der Korrelationskoeffizient gibt eine Aussage über lineare Zusammenhänge, weshalb keine Korrelation ermittelt wurde, obwohl eine quadratische Abhängigkeit vorliegt. Der Korrelationskoeffizient ist nahe bei Null, da kein ausgeprägter linearer Zusammenhang vorliegt. Für negative x -Werte liegt es einen fallenden Kurvenverlauf und für positive x -Werte einen wachsenden Zusammenhang. Diese Effekte heben sich gegenseitig auf, so dass der Korrelationskoeffizient nahe bei Null liegt.
- b) Ein Korrelationskoeffizient von -0.969 (negativ, betragsmäßig nahe bei 1) beschreibt einen deutlich ausgeprägten gegenläufigen Zusammenhang, beschreibt also den Zusammenhang zwischen x_2 und y_2 .

Ein Korrelationskoeffizient von 0.058 (nahe bei Null) beschreibt einen nicht oder wenig ausgeprägten linearen Zusammenhang. Den gibt es zwischen x_3 und y_3 .

Ein Korrelationskoeffizient von -0.984 (positiv, betragsmäßig nahe bei 1) beschreibt einen deutlich ausgeprägten gleichläufigen Zusammenhang, beschreibt also den Zusammenhang zwischen x_1 und y_1 .

3. Aufgabe:

- a) x_1 : Diagramm oben links (blau). Die Werte von x_1 scheinen relativ gleichmäßig über das Intervall $[3, 4]$ verteilt zu sein (eventuell links etwas dichter als rechts). Es handelt sich um exakte Werte. Deshalb sind keine Bindungen erkennbar.
 x_2 : Diagramm unten links (rot). Die Werte von x_2 erstrecken sich über einen Bereich von ca. 0.5 bis 6.5, wobei die Werte in der Mitte dichter liegen. Es handelt sich um exakte Werte. Es sind keine Bindungen erkennbar.
 x_3 : Diagramm oben rechts (blau). Gerundete Werte von x_1 . Es sind starke Bindungen erkennbar, da der Wertebereich sehr klein ist. Durch das Runden werden viele unterschiedliche Werte auf den gleichen Wert gerundet.
 x_4 : Diagramm unten rechts (rot). Gerundete Werte von x_2 . Es sind Bindungen erkennbar (Stapel). Diese sind schwächer ausgeprägt als bei x_3 , da der Wertebereich von x_4 größer ist.
- b) Bindungen können bei rangbasierten Tests zu Problemen führen. Beispiele dafür sind Rangsummen- und Rangkorrelationstests.

4. **Aufgabe:** Der erste Plot zeigt einen konvexen Verlauf im Vergleich mit der symmetrischen Normalverteilung. D.h. die Daten entstammen keiner Normalverteilung sondern eher einer rechtsschiefen Verteilung.

Im zweiten Plot liegen die Punkte annähernd auf einer Gerade. D.h. die Daten könnten aus einer normalverteilten Grundgesamtheit entstammen.

Beim dritten Plot ist der Anstieg an beiden Enden flacher als in der Mitte. D.h., die Verteilungsfunktion der zugrunde liegenden Verteilung geht schneller gegen 0 und 1 als die Normalverteilung.

5. **Aufgabe:**

- a) Ja, es gibt Bindungen, erkennbar an Überlappungen im Bereich zwischen 1,8 und 2,0 im gestapelten Punktdiagramm.
- b) Nein, die Verteilung ist nicht symmetrisch. Sie ist linksschief. Der Median im Boxplot liegt sehr weit rechts und auch im Histogramm ist die Häufung der Werte im rechten Bereich erkennbar.
- c) Aus den gegebenen Daten können keine Abhängigkeiten ermittelt werden, da nur eine Stichprobe vorliegt.
- d) Ja, es gibt Ausreißer. Diese sind im Boxplot erkennbar.

6. **Aufgabe:**

- a) X_1 : diskret, dichotom
 X_2 : diskret, Nominalskala
 X_3 : diskret, Nominalskala
 X_4 : diskret, Absolutskala
 X_5 : stetig, Verhältnisskala
- b) (i) kann nicht genutzt werden, weil nur Studenten die in die Mensa gehen, zur Grundgesamtheit gehören und nicht alle Studenten. Außerdem werden die zeitigen Esser bevorzugt.
(ii) kann nicht genutzt werden, da nur Studenten, die die Unirundmail lesen, zur Grundgesamtheit gehören. Auch hier werden die ersten 50 Studenten bevorzugt.
(iii) kann genutzt werden, da aus allen Studenten 50 ausgewählt werden und ein jeder mit gleicher Chance und unabhängig von allen anderen Studenten ausgewählt werden kann.
- c)
 - Die Werte liegen zwischen 300 und 700 €
 - Häufung zwischen 400 und 600 €
 - keine Bindungen
 - 5 Befragte haben weniger als 400 €
 - 2 Befragte haben ca. 700 €