**Einheit 0 – Überblick**

* Hier: Abb\_0\_1\_Titel\_Übersicht.jpg -

In diesem Modul werden Sie in das Data Wrangling eingeführt. Data Wrangling kann grob mit dem deutschen Wort "Datenaufbereitung" übersetzt werden und beschreibt den Prozess der Vorbereitung erhobener Daten zur anschließenden eigenen Analyse sowie der Nachnutzung durch Dritte. Wenn wir z.B. in einer Umfrage Daten erhoben haben, dann liegt uns nach der Eingabe der Daten zunächst ein Rohdatensatz vor, mit dem wir wenig anfangen können.

Teilnehmende an einer Studie haben beispielsweise mehrere Fragen zu ihrer Studienmotivation jeweils durch Ankreuzen auf einer Skala von 1-5 beantwortet. Die zunächst auf Papier beantworteten Fragen müssen wir anschließend in ein Programm zur Datenverarbeitung (z.B. Excel, SPSS) eingeben.

Uns interessiert, wie hoch die Studienmotivation insgesamt ist (um diese vielleicht mit der Motivation in einem anderen Fach zu vergleichen). Dafür müssen wir die verschiedenen Fragen zusammenfassen, wir müssen schauen, ob es fehlende Werte im Datensatz gibt (hat eine Teilnehmerin vielleicht eine oder mehrere Fragen ausgelassen?), wir müssen überlegen, wie wir mit fehlenden Daten umgehen und überprüfen, ob wir die verschiedenen Fragen überhaupt zusammenfassen können. Der Prozess des Data Wrangling besteht aus mehreren Schritten, die alle gleichsam wichtig sind, damit wir am Ende auch etwas mit unseren Daten anfangen können.

* Hier: Abb\_0\_2\_Garbage\_in\_Garbage\_out.jpg –

Unten finden Sie eine Übersicht über die Themen, die in diesem Modul behandelt werden. Sie müssen die Themen nicht in dieser Reihenfolge abarbeiten, es ist aber sicher sinnvoll. Sie werden in den verschiedenen Inhaltsbereichen immer wieder auf andere Bereiche verwiesen, zu denen Sie dann vor- oder zurückspringen können. Nach Bearbeitung dieses Kurses wissen Sie, was mit "garbage in, garbage out" gemeint ist und kennen wichtige Schritte, mit denen Sie dafür sorgen können, dass Ihre Daten und somit auch Ihre Analysen eben kein garbage sind.

* H5P-Element „Timeline“ (10 Zeitpunkte) –

1. **Prozessmodell des Data Wrangling**

Hier wird Ihnen zunächst ein Modell des Prozesses des Data Wrangling präsentiert, an dem wir uns im Laufe des Moduls immer wieder orientieren und die einzelnen Schritte verorten können.

1. **Die Datenmaske**

**Datensatz: ein erster Blick auf die Daten**  
Wir haben Daten erhoben, um die Motivation Studierender in einer Vorlesung genauer zu untersuchen. Diesen Datensatz stellen wir Ihnen zur Verfügung, um die Vorbereitung von Daten und alle Schritte des Data Wrangling einzuüben, bzw. selbst einmal auszuprobieren.  
In diesem Kursabschnitt **stellen wir Ihnen den Datensatz vor,** mit dem Sie arbeiten werden. Wenn Sie mit Daten umgehen, dann müssen Sie wissen, was hinter diesen Daten steckt und zu welchem Zweck Sie erhoben wurden. Sie erhalten also eine kleine **Einführung in die der Datenerhebung zugrundeliegenden Annahmen.   
Zusätzlich werfen wir einen ersten Blick auf den Datensatz.** Wenn wir Daten erhoben haben, dann liegt uns (nach der Eingabe der Daten in eine Datenmaske) zunächst ein Datensatz vor, mit dem wir noch nicht viel anfangen können. Auch sind die Daten wahrscheinlich noch nicht sortiert, nicht verständlich bezeichnet und die Eigenschaften der verschiedenen Datentypen wurden noch nicht festgelegt. In diesem Abschnitt werfen wir einen ersten Blick auf unsere Daten und schaffen ein wenig Ordnung.  
Hier lernen Sie, wie Sie einen **Datensatz in SPSS** erstellen können und was ein **Codebuch** ist.

1. **Skalenniveaus**

Hier schauen wir uns das Konzept der Skalenniveaus noch genauer an. Was sind Skalenniveaus? Welche Rechnungen kann ich mit welchen Daten machen? Was bedeutet das für meine Analyseverfahren? Und was haben sie mit Data Wrangling zu tun?

1. **Daten bereinigen und anreichern**

Was machen wir wenn einzelne Daten fehlen?  
Es kommt häufig vor, dass Daten im Datensatz fehlen. Vielleicht hat eine Teilnehmerin eine Frage übersehen und daher kein Kreuz gesetzt oder vielleicht hat ein Teilnehmer eine Frage nicht beantwortet, weil er sie nicht verstanden hat.  
Was machen wir wenn einzelne Daten stark von den anderen abweichen?  
Es kann auch vorkommen, dass einzelne Daten stark von den anderen abweichen. Eventuell ist beispielsweise eine Person in der Stichprobe, die deutlich älter ist als die anderen. Je nach Fragestellung kann es sein, dass diese Ausreißer die Analyse beeinflussen.   
In diesem Abschnitt lernen Sie, wie Sie fehlende Daten und Ausreißer identifizieren können und wie Sie mit ihnen umgehen.

1. **Maße zentraler Tendenz**

In diesem Abschnitt lernen Sie verschiedene Maße zentraler Tendenz kennen. Sie sagen uns etwas darüber, wie unsere Daten verteilt sind. Sie lernen außerdem, wie Sie diese in SPSS berechnen können.

1. **Verteilungen**

In dieser Einheit werden Sie ausgewählte, wichtige Häufigkeitsverteilungen kennenlernen oder auffrischen, mit deren Hilfe Sie sich einen ersten Eindruck über Ihre Daten verschaffen können und außerdem die Plausibilität Ihrer bisherigen Bearbeitungen überprüfen können.

1. **Skalenbildung**

Es kommt häufig vor, dass wir mehrere Datenpunkte (z.B. die Antworten auf Fragen) zusammenfassen müssen, damit wir unsere Analysen sinnvoll umsetzen können. In diesem Abschnitt lernen Sie, wie Skalen in Fragebögen zusammengefasst werden und wie Sie dies in SPSS umsetzen können.

1. **Reliabilität**

Reliabilität ist ein Maß für die Genauigkeit und Verlässlichkeit von Messungen. Im Prozess des Data Wrangling sollten Sie die Reliabilität Ihrer Instrumente überprüfen. Welche Arten von Reliabilität es gibt, wie sie berechnet werden und wie Sie Ihre Rechnungen in SPSS umsetzen können, lernen Sie in diesem Kursabschnitt.

1. **Standardisierung**

In diesem Abschnitt werden Sie lernen, wie Sie ursprünglich vielleicht weniger vergleichbare Daten so transformieren können, dass Sie sie vergleichen können. Sie lernen Verfahren kennen, mit denen Sie Daten standardisieren können und wozu dies gut ist.

1. **Dokumentation und Zusammenfassung**

Der gesamte Prozess des Data Wrangling muss gewissenhaft dokumentiert werden, damit Andere und nicht zuletzt Sie selbst auch nach der Aufbereitung der Daten noch nachvollziehen können, was Sie an den Daten verändert haben. Wenn Andere beispielsweise Ihre Daten nutzen wollen, um eigene Analysen durchzuführen, dann ist es dringend notwendig, dass Sie wissen, was mit den Daten seit der Erhebung passiert ist.

* Hier neuer Abschnitt –

**Abgaben und Anforderungen in diesem Kurs**

Um diesen Kurs zu bestehen, sollten Sie verschiedene Anforderungen erfüllen. Folgen Sie den Inhalten aufmerksam, werden diese aber keine Herausforderungen für Sie darstellen.

1. Abschlussquizze. Bearbeiten Sie die Quizze am Ende der jeweiligen Abschnitte, um die nachfolgenden Abschnitte freizuschalten.

2. Abgabe: Codebuch. Über den gesamten Kurs hinweg werden Sie gebeten, Ihr sogenanntes Codebuch immer weiter zu aktualisieren. Nach Bearbeitung dieses Kursabschnittes werden Sie gebeten, Ihr Codebuch abzugeben.

- Ende -