

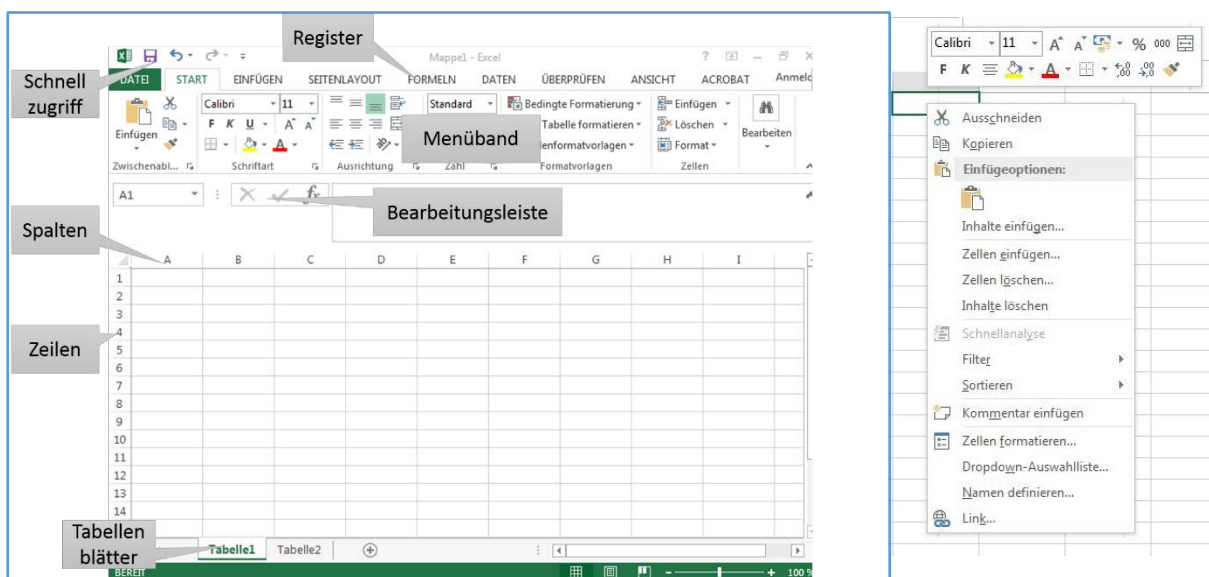
Studiendaten mit Hilfe von MS Excel sinnvoll verwalten und nutzen

Jan-Lukas Plenio, Institut für Veterinär-Epidemiologie und Biometrie, FU Berlin
Dezember 2016

Einführung

Das Handout bezieht sich auf MS Excel 2016. Die Kursinhalte sind problemlos auf die älteren Versionen Excel 2007 bis Excel 2013 übertragbar, zu Excel 2003 und älter (ohne Menüband) sind sie jedoch nicht mehr kompatibel.

- Excel ist hervorragend dafür geeignet, Daten darin abzulegen und zu verwalten. Die so vorbereiteten Daten können dann in ein Statistikprogramm eingelesen und statistisch ausgewertet werden.
- Auch und gerade die Tabellenkalkulation ist eine der Stärken von Excel, hier gibt es viele Möglichkeiten, von denen heute einige vorgestellt werden.
- Auch wenn es möglich ist, statistische Auswertungen mit Excel durchzuführen, ist Excel dafür nicht das geeignete Programm, und wir empfehlen, dies nicht zu tun!
- Auch Datenbanken können in Excel nicht erzeugt oder verwaltet werden, sondern wir legen Daten in Datentabellen ab. Hintergrund: Eine Datenbank besteht aus mehreren Tabellen, die logisch miteinander verknüpft sind und Beziehungen abbilden, die aus mehr als zwei Dimensionen bestehen.
- Mit Hilfe von Excel können schnell und einfach sehr schöne Übersichten erzeugt werden, mit denen Ergebnisse zusammengefasst und präsentiert werden können. Dies darf aber nicht mit den Datentabellen verwechselt werden, um die es heute geht.



Datentabelle

Allgemeines

- Eine Datentabelle besteht aus Zeilen und Spalten.
- Jede Zeile steht für eine Beobachtungseinheit / ein Individuum, z.B. ein Tier, eine Probe, etc.
- Jede Spalte steht für eine Variable, z.B. Geschlecht oder Gewicht des Tieres.
- In der ersten Zeile steht in der Regel der Variablenname.

Beobachtungen = Zeilen

- Jede Zeile repräsentiert eine Beobachtungseinheit, z.B. ein Tier, eine Probe. Dies bedeutet, dass in einer Zeile nur Informationen zu einem Individuum enthalten sein dürfen und dass alle Informationen zu einem Individuum in einer einzigen Zeile sein müssen.
- Daher ist es notwendig, die Beobachtungseinheit im Vorfeld genau zu definieren. Was ist z.B. mit Eutervierteln einer Kuh?
- Wird ein Tier mehrfach untersucht (Wiederholungsmessungen), empfiehlt es sich in der Regel, das Tier als Beobachtungseinheit zu betrachten und die Ergebnisse der einzelnen Untersuchungen in mehreren Spalten abzulegen. Es kann sein, dass diese Daten für bestimmte Auswertungen transformiert werden müssen (jeder Untersuchungszeitpunkt in einer Zeile, das Tier kommt in mehreren Zeilen vor).

Variablennamen

- So kurz wie möglich
- „sprechend“
- Ohne Leerzeichen und Umlaute

Variablen = Spalten

- Laufende Nummer („ID“ oder „lfdNr“)
wird immer in der ersten Spalte vergeben, damit jede Zeile auf jeden Fall eine eindeutige Bezeichnung erhält → so kann man prüfen, ob alle Daten vollständig sind.
- Identifizierungszeichen,
wie z.B. Tiernummer, Probennummer etc. werden in den folgenden Spalten eingetragen. Welche dieser Zeichen sind einmalig? Welche kommen mehrfach vor (wenn z.B. ein Tier mehrmals untersucht wurde)? Der Variablenname sollte darauf hinweisen, dass es sich um ein Identifizierungszeichen handelt, z.B. „ID_Tier“, „Probennr“.
- Weitere Identifizierungszeichen,
falls die Beobachtungseinheiten hierarchisch gruppiert / strukturiert sind, werden diese in den folgenden Spalten eingefügt. Beispiele: Betrieb, aus dem das Tier stammt („Betriebs_ID“); Labor, in dem die Proben untersucht wurden („Labor_ID“).
- Einflussfaktoren,
beschreibenden Merkmale, z.B. Rasse, Alter, Geschlecht, Probenahmedatum, folgen in den weiteren Spalten. Diese Merkmale können das Untersuchungsziel unter Umständen beeinflussen

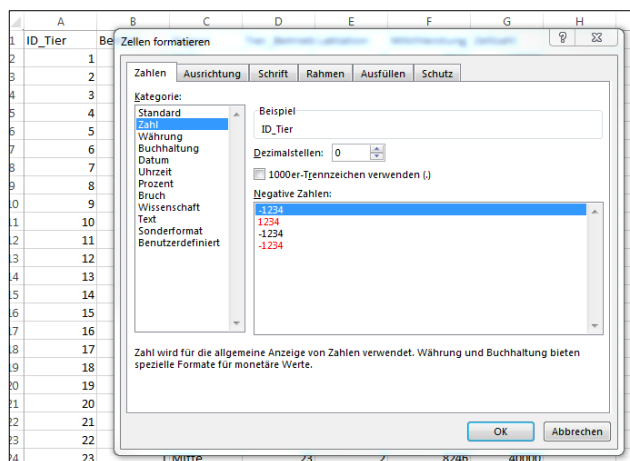
und werden deshalb als Einflussfaktoren bezeichnet.

- Zielvariablen, die Ergebnisse der eigentlichen Untersuchungen werden in den letzten Spalten der Tabelle eingetragen. Hier kommen also die Ergebnisse der Laboruntersuchungen oder die Angaben aus den Fragebögen hin. Diese Variablen stellen den eigentlichen Gegenstand der Studie dar und werden deshalb als Zielvariablen bezeichnet.

Die strikte Einteilung der Variablen in Einfluss- und Zielvariablen hilft bei der Auswertung den Überblick zu bewahren und unterstützt dadurch den Prozess der Hypothesenformulierung.

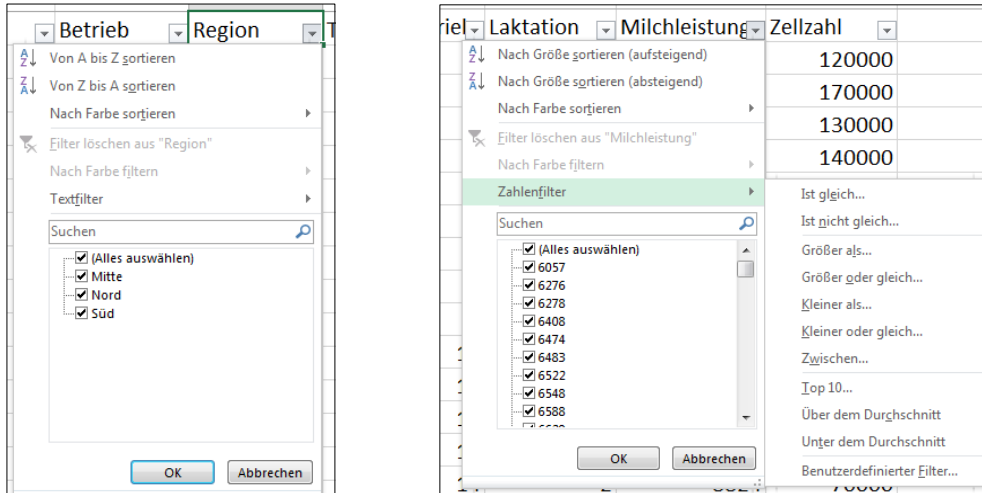
Zellen formatieren

- Für jede Variable wird festgelegt, welche Informationen sie enthalten soll, und danach richtet sich die Formatierung der Zellen.
- Handelt es sich um eine Zahl? Oder um Text? Ein Datum?
- Vor Beginn der Dateneingabe wird die ganze Spalte formatiert.
- Start → Zellen → Format → Zellen formatieren



Filtern und Sortieren

- Filter einschalten: Start → Bearbeiten → Sortieren und Filtern → Filtern
- Alternativ: Daten -> Sortieren und Filtern
- Mit dem Filter können Sie sich Datensätze anzeigen lassen, die bestimmte Bedingungen erfüllen.

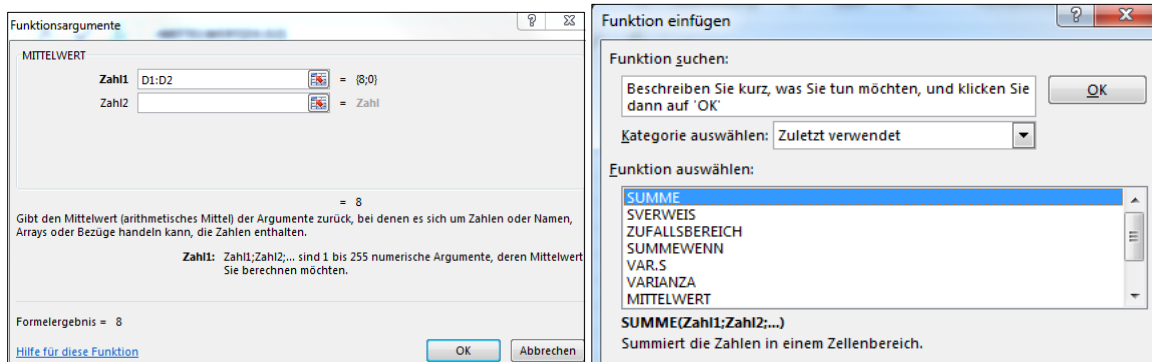


- Fenster fixieren
 - Ansicht → Fenster → Fenster fixieren → Fenster fixieren
 - Die Zeilen und Spalten, die über oder links von der markierten Zelle liegen, werden fixiert und bleiben beim Scrollen durch das Datenblatt sichtbar.
- So kann man sich in großen Datensätzen bewegen: Mit Strg + Pfeiltasten (←↑→↓ oder Bild↑/Bild↓) kann man zum ersten oder letzten Eintrag springen. Pos1 springt zur ersten Spalte in der aktuellen Zeile, Strg + Pos1 springt zur Zelle A1, also der ersten Spalte in der ersten Zeile. Entsprechend Ende und Strg + Ende.

Formeln verwenden

Über die Formelfunktion können Berechnungen angestellt werden, deren Ergebnis in einer Zelle ausgegeben wird. Durch die Eingabe von „=“ in eine Zelle erkennt Excel, dass nun eine Formel folgt. Nach Betätigung der Eingabetaste führt Excel die Berechnung durch und gibt das Ergebnis in der Zelle aus. In der Bearbeitungsleiste wird die verwendete Formel angezeigt.

Es gibt unendlich viele Möglichkeiten, diese Funktion zu nutzen. Hier werden nur wenige Beispiele vorgestellt. Durch Klick auf das Symbol öffnet sich das entsprechende Menü, aus dem man die gewünschte Funktion aufrufen kann.



Nach Auswahl der Funktion öffnet sich ein Dialogfenster, in dem beschrieben ist, welche Angaben zur Ausführung der Funktion benötigt werden.

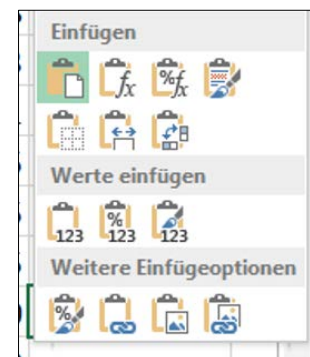
Alternative kann durch Eingabe von „=“ und dem Befehl, z.B. „Summe“ in die Bearbeitungsleiste eine Formel direkt eingegeben werden. Zur Hilfe wird von Excel ein Fenster angezeigt, welches die von Excel erwartenden Werte Ihnen anzeigt.



Der Umgang mit Formeln wird auch in dem Heft „Formeln und Funktionen clever nutzen“ beschrieben. Dieses und weitere Bücher zu Excel, Word etc. gibt es für Mitarbeiter der FU mit ZEDAT-Account kostenlos unter: <https://portal.zedat.fu-berlin.de/services/ebooks.php>.

Kopieren und Einfügen:

Wenn Sie eine Zelle markieren und kopieren (Strg + C) und in einer anderen Zelle wieder ausgeben lassen (Strg + V), dann haben Sie über das Kontextmenü mehrere Auswahlmöglichkeiten. Standardmäßig werden Formeln und auch Formate mit eingefügt. Sie können aber auch nur Inhalte (Formeln werden dann nicht mitkopiert) oder nur die Formel an sich einfügen.

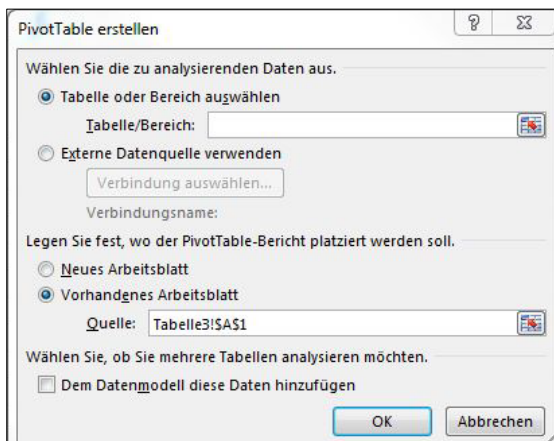


Pivot-Tabellen

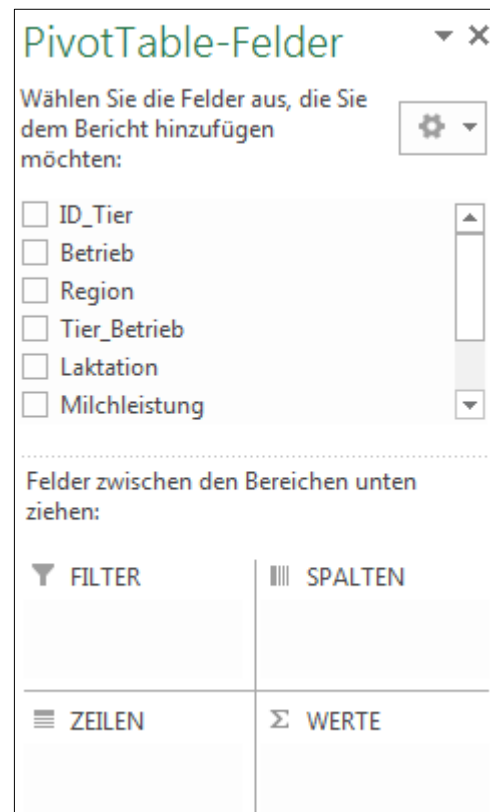
Wo auch immer Sie derartige Berechnungen platzieren – in einer reinen Datentabelle gibt es keinen guten Platz dafür. Besser ist es, diese zusammenfassenden oder beschreibenden Daten in einem separaten Datenblatt abzulegen, damit in dem Tabellenblatt „Daten“ auch wirklich nur die Rohdaten enthalten sind.

Noch schneller geht es mit Hilfe von Pivot-Tabellen. Mit diesen können Sie sich Ihre Daten übersichtlich darstellen lassen. Einfache Berechnungen wie Anzahl, Mittelwert, Summe o.ä. können Sie sich getrennt für verschiedene Kategorien anzeigen lassen. Wir bauen Schritt für Schritt auf.

- Öffnen Sie ein neues Tabellenblatt und platzieren Sie den Cursor oben links.
- Klicken Sie auf Einfügen → Tabellen → Pivot-Tables.
- Es öffnet sich ein Dialogfeld.
- Platzieren Sie den Cursor in das Feld „Tabelle/Bereich“ und wechseln Sie dann in das Tabellenblatt „Daten“.
- Markieren Sie alle Spalten mit Daten vollständig und klicken Sie auf OK.



Rechts im Bild befindet sich nun das Arbeitsfenster „PivotTable-Felder“, mit dem wir weiter arbeiten. Alle Spalten der Tabelle sind abgebildet. Wir können sie auswählen und mit dem Cursor an die gewünschte Position ziehen.



- Ziehen Sie „Betrieb“ in das Feld „Zeilen“. Links im Bild erscheinen die Zahlen 1 bis 9. Wenn das Fenster „PivotTable-Felder“ einmal verschwindet, dann klicken Sie mit dem Cursor in die Pivottable hinein.

- Ziehen Sie nun „Milchleistung“ in das Feld „Werte“. Es erscheint eine Spalte, die mit „Anzahl von Milchleistung“ überschrieben ist. Es zeigt an, wie viele Einträge zu Milchleistung – also wie viele Tiere – es zu jedem Betrieb gibt.
- Benennen Sie die Spalte um, z.B. „Anzahl Tiere“ oder „Betriebsgröße“.
- Ziehen Sie „Milchleistung“ noch einmal in das Feld „Werte“. Öffnen Sie im Kontextmenü die „Wertfeldeinstellungen“ und wählen Sie „Mittelwert“ aus. Benennen Sie die Spalte wieder nach Belieben um und stellen Sie die Anzahl der Nachkommastellen und die Spaltenbreite ein.
- Fügen Sie anschließend noch „Region“ in das Feld „Zeilen“ ein und verschieben Sie „Region“ vor „Betrieb“.
- Auch bei Pivottabellen haben Sie die Möglichkeit, nach bestimmten Daten oder Kategorien zu filtern. Dazu können Sie das kleine Häkchen in der ersten Spalte der Tabelle anklicken und dann eine Region auswählen.
- Alternativ können Sie in dem Fenster „PivotTable-Felder“ die jeweilige Variable anklicken und dort die gewünschte Auswahl treffen.

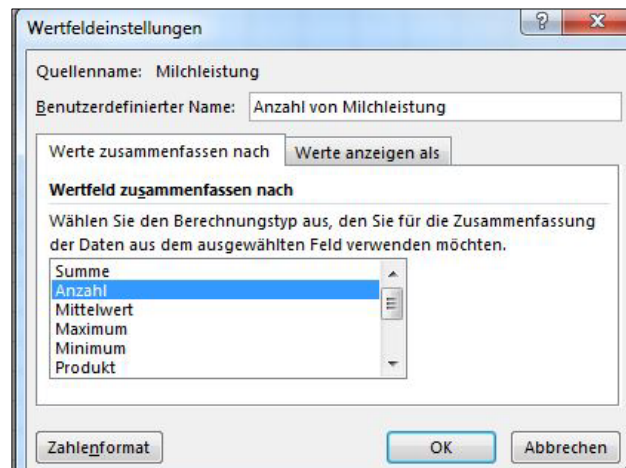
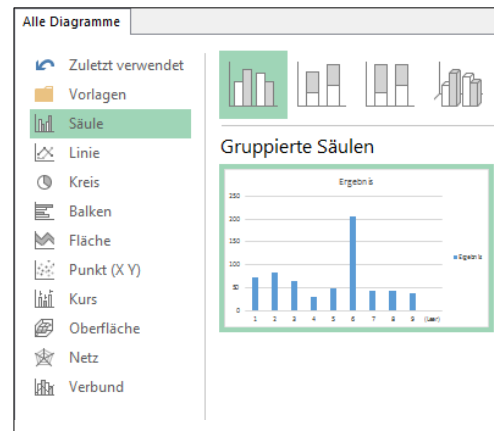


Diagramme erstellen

- Mit Excel lassen sich schnell und einfach Diagramme erstellen und in ein anschauliches Layout bringen.
- Die Auswahl des Diagrammtypen richtet sich u.a. nach dem vorliegenden Datentyp – für kategorielle Daten sind andere Typen geeignet als für stetige Daten.
- Durch die Art der Darstellung können Sie Einfluss auf die Botschaft nehmen, die das Diagramm sendet. Überlegen Sie also im Vorfeld, was Sie mit dem Diagramm zum Ausdruck bringen oder in den Vordergrund stellen wollen.
- Über Einfügen → Diagramme können Sie das Auswahlfenster der Diagramme öffnen.



Beispiel: Anzahl der Tiere je Betrieb

- Wie viele Tiere je Betrieb werden gehalten?
- Möglichkeiten: Säulendiagramm, Balkendiagramm, Kreisdiagramm...

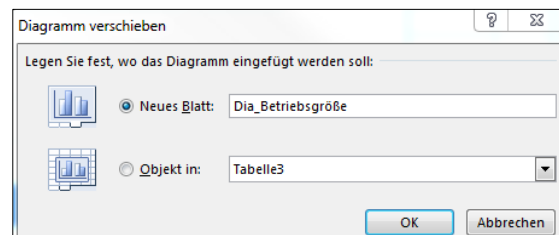
Region	Betrieb	Milchleistung	log. Zellzahl
Mitte	1	7974,5	5,019
	2	7895,7	5,196
	3	8061,5	5,012
Nord	4	7875,1	4,968
	5	7940,6	5,213
	6	7951,0	4,994
Süd	7	8097,1	5,017
	8	8076,8	5,127
	9	7982,1	5,155

■ Diagramm erstellen:

Alle Daten markieren und auf Einfügen →
Diagramme klicken.

Gewünschten Diagrammtyp auswählen →
gruppierte Säulen

Rechtsklick → Diagramm verschieben → als Neues Blatt „Dia_Betriebsgröße“ ablegen.



■ Daten anpassen

Über Diagrammtools → Entwurf → Daten
auswählen oder in Zeichnungsfläche klicken →
Rechtsklick → Daten auswählen

Links stehen die Daten, die im Diagrammfeld
dargestellt werden, rechts die Daten, die als
Kategorien betrachtet werden.

Zum Anpassen (in diesem Fall muss aus „Region“ und „Betrieb“ je eine Kategorie werden) können links die Häkchen entfernt werden. Rechts können über „Bearbeiten“ die Spalten mit den Kategorien ausgewählt werden.

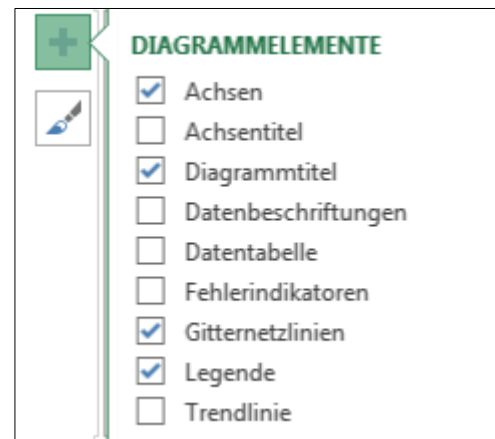


■ Sekundärachse einfügen:

Datenreihe markieren → Rechtsklick → Datenreihen
formatieren → Reihenoptionen → Sekundärachse

Ausrichtung Datenreihen über Abstandsbreite und
Reihenachsenüberlappung

Alternative: Datentyp für eine der Datenreihen z.B. von
Säule auf Linie ändern

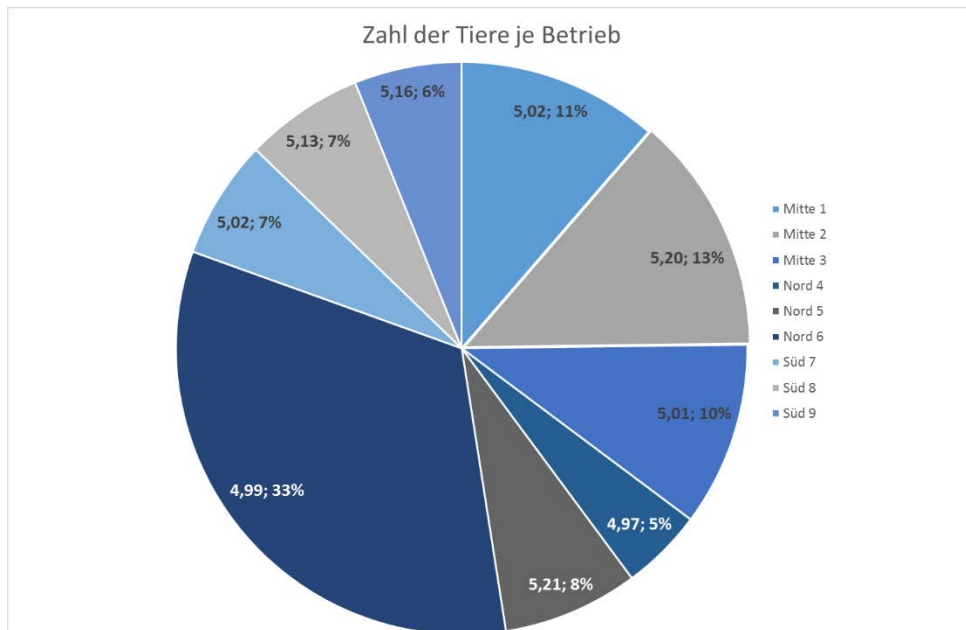


■ Layoutoptionen:

■ Formatvorlage:

- Es gibt viele Formatvorlagen in Excel, viele davon sind im wissenschaftlichen Kontext eher ungeeignet.
- Beschriftungen hinzufügen:
 - über das oder
 - über Diagramm-Tools → Entwurf → Diagrammelement hinzufügen
- Anpassen von Schriftgröße und Position, Einfügen / Entfernen von Rahmen etc.
 - über Diagramm-Tools → Format oder

- Element markieren → Rechtsklick → xxx formatieren
- Diagrammtyp ändern
 - über Diagramm-Tools → Entwurf → Diagrammtyp ändern oder
 - in Zeichnungsfläche klicken → Rechtsklick → Diagrammtyp ändern



Nützliche Tastenkombinationen

ENTER		Auswahl springt zum Anfang der nächsten Zeilen
TAB		Auswahl springt in die nächste Spalte
F2		Bearbeitungsleiste für markierte Zelle aktivieren
STRG + X/C/A		Ausschneiden/Kopieren/Einfügen
STRG + Z/Y		Veränderung rückgängig machen/wiederherstellen

	Cursor zeichenweise nach links bzw. rechts bewegen
Strg	Cursor wortweise nach links bzw. rechts bewegen
Pos 1 / Ende	Cursor an den Anfang bzw. das Ende des Zellinhalts bewegen
	Zeichen links des Cursors löschen
Entf	Zeichen rechts des Cursors löschen

	Nächste Zelle gemäß Pfeilrichtung
Strg (1) Strg (2) Strg (3) Strg (4)	Zum Anfang/Ende des angrenzenden ausgefüllten Tabellenbereichs gemäß der Pfeilrichtung
Bild ↓	Eine Bildschirmseite nach unten
Bild ↑	Eine Bildschirmseite nach oben
Pos 1	Erste Zelle der aktuellen Zeile
Strg Pos 1	Erste Zelle des Tabellenblattes (Zelle A1)
Strg Ende	Letzte Zelle des Tabellenblattes mit Inhalt

Markierung	Tastenkombination
Aktuelle Zeile	
Aktuelle Spalte	Strg
Ausgefüllter Zellbereich um die aktive Zelle	Strg A (einmal betätigen)
Gesamtes Tabellenblatt	Strg A (zweimal betätigen)
Ab aktiver Zelle in die gewünschte Richtung	+ , , oder
Ab aktiver Zelle zur untersten Zelle mit Inhalt	Strg Ende
Ab aktiver Zelle zum Anfang des Tabellenblattes (Zelle A1)	Strg Pos 1

Sie möchten ...	
Zellen mit Tasten aktivieren	←, →, ↑, ↓, Bild ↓, Bild ↑, Pos 1, Strg Pos 1, Strg Ende
Daten eingeben/überschreiben	Zelle aktivieren, Daten in die Zelle oder Bearbeitungsleiste eingeben, ←
Zellinhalte verändern	Zelle doppelt anklicken, Änderungen vornehmen, ←
Zellinhalte löschen	Zellen markieren, Entf
die Spaltenbreite/Zeilenhöhe ändern	Begrenzungslinie im Spalten- bzw. Zeilenkopf mit der Maus ziehen oder doppelt anklicken
einen Zellbereich markieren	Maus bei gedrückter Maustaste über Zellbereich ziehen oder ↑ in Verbindung mit einer Pfeiltaste drücken
Mehrfachbereiche markieren	Strg drücken und Bereiche mit der Maus markieren