

NWW · KSK

Statistik verstehen und anwenden

Begleitheft für die naturwissenschaftliche Forschungswoche

Das Ziel Du beschreibst Daten passend, erkennst Streuung und Verteilungen und wählst einen sinnvollen Signifikanztest. Das Heft funktioniert mit den Kursbeispielen – und später als Leitfaden für deine eigenen Forschungsdaten.

Der Weg von der Frage zur Aussage

1. Forschungsfrage und messbare Variable festlegen.
2. Datenart, Stichprobe und Versuchsplan klären.
3. Rohdaten prüfen und sichtbar machen.
4. Lage und Streuung beschreiben.
5. Voraussetzungen prüfen und Test wählen.
6. Ergebnis mit Zahlen, Unsicherheit und Grenzen berichten.

Name: _____ Projekt: _____

1 • Von der Forschungsfrage zur Datenart

Die Datenart gehört nicht zum Gegenstand, sondern zur Frage und zur Art der Erfassung. Dieselben vier Seifen können deshalb nominale, ordinale oder metrische Daten liefern.

Frage am selben Gegenstand	Datenart	Warum?
Welche Seife gefällt am besten? Jede Person nennt eine Sorte; die Stimmen werden gezählt.	nominal	Kategorien ohne natürliche Reihenfolge
Bringe die Seifen in eine Rangfolge: beste, zweitbeste, ...	ordinal	Reihenfolge ja, gleiche Abstände nein
Wie hoch ist der Schaum nach 30 Sekunden in mm?	metrisch	Messbare, sinnvoll vergleichbare Abstände

Warum sind Schulnoten nicht metrisch?

Bei Schulnoten ist die Reihenfolge klar: Eine 5 ist besser als eine 4. Aber der Leistungsunterschied zwischen 3 und 4 muss nicht gleich gross sein wie zwischen 5 und 6. Auch „Note 6 ist doppelt so gut wie Note 3“ ist keine sinnvolle Aussage. Schulnoten sind daher ordinal.

Zahl $\neq$ Messwert Auch eine Bewertung von 1 bis 9 bleibt meist ordinal. Eine Zahl wird erst metrisch, wenn gleiche Differenzen inhaltlich gleich viel bedeuten.

Welche Beschreibung passt?

Datenart	Gute Darstellung	Typischer Lagewert	Möglicher Vergleich
nominal	Häufigkeit / Balkendiagramm	Modalwert	Chi-Quadrat oder Fisher
ordinal	Häufigkeit / geordnete Balken	Modalwert und Median	Mann-Whitney / Wilcoxon
metrisch	Punktdiagramm, Histogramm, Boxplot	Mittelwert oder Median	t-Test oder Rangtest

2 • Stichprobe und Versuchsplan

Die Stichprobe ist der untersuchte Teil einer grösseren Grundgesamtheit. Die Stichprobengrösse n zählt unabhängige Untersuchungseinheiten – etwa Personen, Keimlinge oder Proben. Zehn Messungen desselben Keimlings sind nicht automatisch $n = 10$ unabhängige Keimlinge.

Unabhängig oder gepaart?

Versuchsplan	Konkretes Creme-Beispiel	Folge für den Test
unabhängige Gruppen	20 Personen bewerten nur Creme 1, andere 20 nur Creme 4: $n_1 = 20$, $n_2 = 20$.	Die Werte der Gruppen haben keine festen Partner; bei ordinalen Daten z. B. Mann-Whitney-U.
gepaarte Messungen	Dieselben 20 Personen bewerten beide Cremes: $n = 20$ Wertepaaren.	Jeder Wert gehört zu derselben Person; bei ordinalen Daten z. B. Wilcoxon.
mehrfach gemessen	Ein Keimling wird täglich gemessen.	Die Tageswerte desselben Keimlings sind abhängig; nicht wie unabhängige Wiederholungen behandeln.

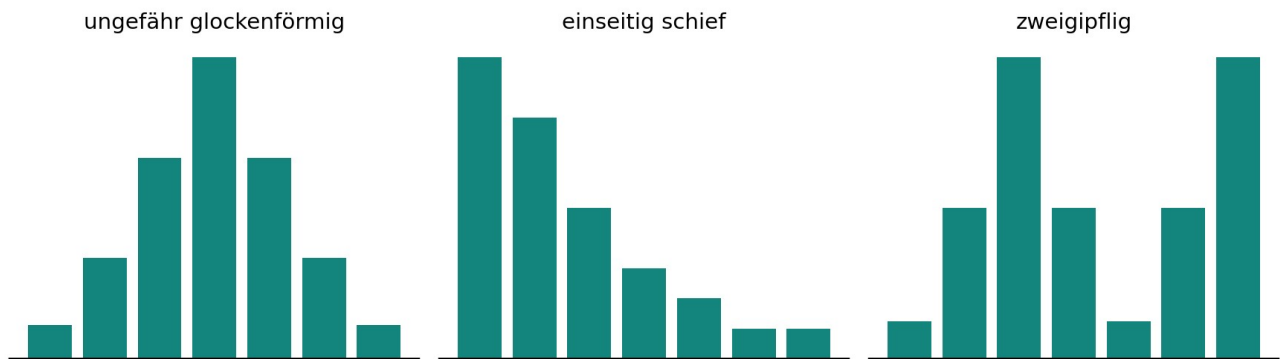
Eine tragfähige Datenerhebung

- Vorher definieren, was eine Untersuchungseinheit ist und wie n gezählt wird.
- Bedingungen möglichst gleich halten; Reihenfolge wenn möglich zufällig wählen.
- Biologische Wiederholungen von technischen Wiederholungsmessungen unterscheiden.
- Einheiten, Messgenauigkeit, fehlende Werte und Abweichungen dokumentieren.
- Keine Daten nachträglich weglassen, nur weil sie nicht zur Erwartung passen.

Merksatz Die Testwahl beginnt nicht beim p-Wert, sondern beim Versuchsplan.

3 • Verteilung und typische Werte

Eine Verteilung zeigt, welche Werte vorkommen und wie häufig sie vorkommen. Ein Häufigkeitsdiagramm macht sichtbar, ob Werte gebündelt, einseitig schief, zweigipflig oder ungefähr glockenförmig liegen.



Modalwert, Median und Mittelwert

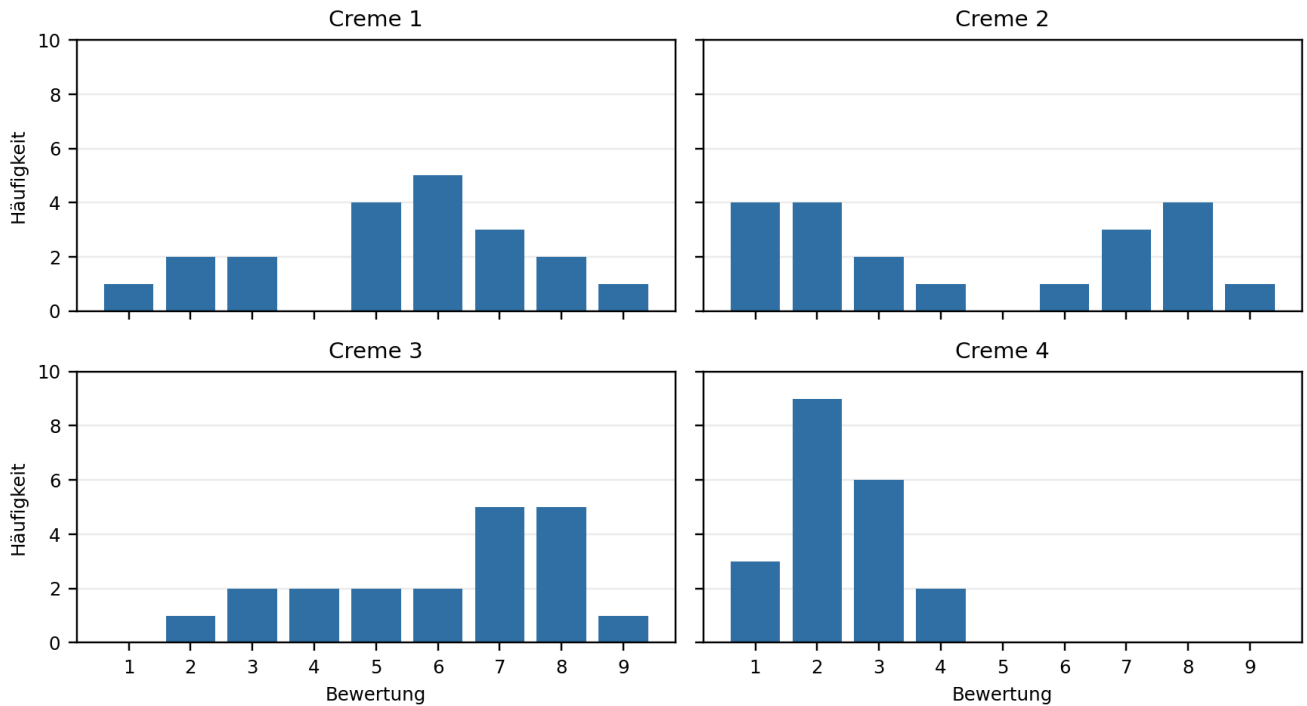
Kennwert	Bedeutung	Wann besonders sinnvoll?	Grenze
Modalwert	häufigster Wert / häufigste Kategorie	auch bei nominalen Daten; direkt aus Häufigkeiten	kann mehrere Werte haben
Median	mittlerer Wert nach dem Sortieren	ordinale Daten; Ausreisser; schiefe Verteilungen	nutzt Abstände nicht
Mittelwert	Summe aller Werte ÷ Anzahl	metrische Daten; ungefähr symmetrische Verteilung	reagiert stark auf Ausreisser

Ausreisser-Beispiel

Wachstum von sechs Keimlingen: 9, 10, 10, 10, 11 und 12 mm. Median = 10 mm, Mittelwert = 10,3 mm. Wird der letzte Wert auf 50 mm gesetzt, bleibt der Median 10 mm, der Mittelwert steigt auf 16,7 mm. Der ungewöhnliche Wert muss geprüft und begründet behandelt werden – nicht automatisch gelöscht.

4 • Das Creme-Beispiel: Häufigkeiten lesen

Je 20 Personen bewerten vier Cremes von 1 bis 9. Die Zahlen sind ordinale Bewertungen. Die Diagramme zeigen die Häufigkeitsverteilungen und damit auch die Modalwerte.



	Creme 1	Creme 2	Creme 3	Creme 4
Modalwert(e)	6	1, 2 und 8*	7 und 8*	2
Median	6	3,5	7	2
Mittelwert**	5,35	4,50	6,10	2,35
Varianz s ^{2**}	4,77	8,89	4,09	0,77
Standardabweichung s**	2,18	2,98	2,02	0,88

* Excel MODE.SNGL zeigt nur einen der gleich häufigen Modalwerte. ** Mittelwert, Varianz und Standardabweichung werden hier zur Demonstration gezeigt; die Bewertungen bleiben ordinal.

Lesebeispiel Creme 2 besitzt keine einzige zentrale Häufung: tiefe und hohe Bewertungen kommen oft vor. Der Mittelwert 4,50 allein würde diese Form verdecken.

5 • Streuung, Varianz und Standardabweichung

Zwei Messreihen können denselben Mittelwert besitzen und trotzdem sehr verschieden sein: 4,8 / 4,9 / 5,0 / 5,1 / 5,2 liegen eng beieinander; 1 / 3 / 5 / 7 / 9 streuen stark. Ein Lagewert braucht deshalb fast immer ein Streumass.

Streumass	Berechnungsidee	Bedeutung
Spannweite	Maximum – Minimum	schnell, aber nur von zwei Extremwerten bestimmt
Varianz s^2	quadrierte Abweichungen vom Mittelwert mitteln	grosse Abweichungen zählen stark; Einheit ist quadriert
Standardabweichung s	Wurzel aus der Varianz	typische Entfernung vom Mittelwert; ursprüngliche Einheit

Warum beginnt die Varianz beim Mittelwert?

7. Mittelwert berechnen.
8. Für jeden Messwert: Abweichung vom Mittelwert bilden.
9. Abweichungen quadrieren – damit sich Plus und Minus nicht aufheben.
10. Quadrierte Abweichungen addieren und bei einer Stichprobe durch $n - 1$ teilen.
11. Wurzel ziehen: Das Ergebnis ist die Standardabweichung.

Excel Für eine Stichprobe: VAR.S und STABW.S (englisch STDEV.S). Für die Grundgesamtheit wären VAR.P und STABW.N/ STDEV.P vorgesehen.

Standardabweichung ist nicht Standardfehler

Die Standardabweichung beschreibt, wie verschieden die Einzelwerte sind. Der Standardfehler beschreibt, wie unsicher der geschätzte Mittelwert ist. Für die Forschungswoche genügt meist: Einzelwerte zeigen und die Standardabweichung korrekt benennen.

6 · Richtigkeit, Präzision und Messfehler

Begriff	Frage	Typische Ursache
Richtigkeit	Liegt der Mittelwert nahe am richtigen Wert?	Kalibrierung, systematische Verschiebung
Präzision	Liegen Wiederholungen eng beieinander?	zufällige Schwankungen, Ablesunsicherheit
zufälliger Fehler	Schwanken Werte unvorhersehbar?	Sensorrauschen, kleine Umweltänderungen
systematischer Fehler	Sind alle Werte ähnlich verschoben?	falsch geeichtes Gerät, verunreinigte Probe
grober Fehler / Ausreisser	Ist ein Einzelwert offensichtlich auffällig?	Verwechslung, Tippfehler, Geräteausfall

Was hilft?

- Zufällige Fehler: mehrere unabhängige Wiederholungen, standardisierte Durchführung.
- Systematische Fehler: Kalibrierung, Kontrollen, Blindproben, Vergleich mit Referenzwerten.
- Grobe Fehler: Ursache prüfen und transparent dokumentieren. Einen Wert nur mit sachlicher Begründung ausschliessen.

Wichtig Viele Wiederholungen können die Präzision verbessern, aber eine systematisch falsche Messung nicht automatisch richtig machen.

Diagramme ehrlich zeichnen

- Achsen mit Messgrösse und Einheit beschriften.
- Skala nicht so abschneiden, dass kleine Unterschiede riesig wirken.
- Einzelwerte zeigen, wenn n klein ist.
- Bei Mittelwertbalken genau angeben, ob Fehlerbalken SD, Standardfehler oder Konfidenzintervall darstellen.

7 • Normalverteilung: warum und wie prüfen?

Manche Tests – besonders t-Tests – beruhen auf Annahmen über metrische Daten. Bei einem t-Test für unabhängige Gruppen sollten die Werte innerhalb der Gruppen ungefähr passend verteilt sein; beim gepaarten t-Test geht es um die Verteilung der Differenzen. Nominale und ordinale Daten werden nicht durch einen Normalverteilungstest „metrisch“.

Vorgehen ohne unnötige Mathematik

12. Nur bei metrischen Daten prüfen.
13. Zuerst Punktdiagramm oder Histogramm ansehen: Symmetrie, Ausreisser, Schiefe, zwei Gipfel?
14. Fachlich überlegen: Können zwei Teilgruppen oder Messgrenzen die Form erklären?
15. Optional Shapiro-Wilk-Test als zusätzlichen Hinweis verwenden.

Shapiro-Wilk-Ergebnis bei $\alpha = 0,05$	Aussage
$p < 0,05$	deutlicher Hinweis gegen Normalverteilung; H_0 wird verworfen
$p \geq 0,05$	keine deutliche Abweichung gefunden; kein Beweis für Normalverteilung

Mechanisches Lernbeispiel mit den Creme-Zahlen

Creme 1	Creme 2	Creme 3	Creme 4
$p = 0,275$	$p = 0,0049$	$p = 0,071$	$p = 0,018$

Die p-Werte zeigen nur die Testlogik. Inhaltlich bleiben die Cremebewertungen ordinal; deshalb wird für Gruppenvergleiche ein Rangtest gewählt. Bei kleinen Stichproben hat der Test zudem wenig Trennschärfe – das Diagramm bleibt wichtig.

8 · Signifikanz, α , p-Wert und Konfidenzintervall

Zwei Werte unterscheiden sich statistisch signifikant, wenn der beobachtete Unterschied unter der Annahme „kein wirklicher Unterschied“ so ungewöhnlich ist, dass eine rein zufällige Erklärung wenig plausibel wird. Das ist keine absolute Gewissheit.

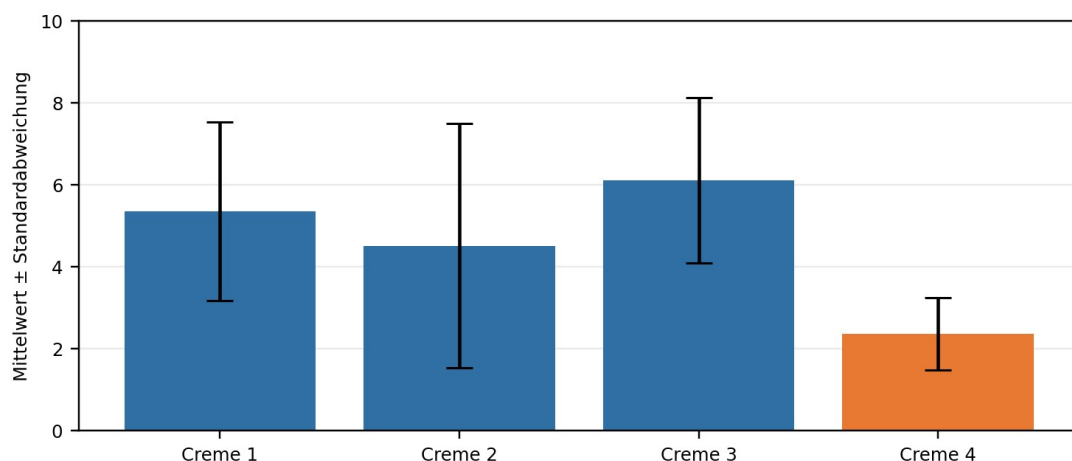
Begriff	Einfache Bedeutung
H_0 – Nullhypothese	kein Unterschied / kein Effekt
H_1 – Alternativhypothese	ein Unterschied / ein Effekt
α – Signifikanzniveau	vorher erlaubtes Fehlalarmrisiko; häufig 0,05
p-Wert	wie ungewöhnlich die Daten oder extremere Daten unter H_0 wären
Konfidenzniveau	$1 - \alpha$; bei $\alpha = 0,05$ also 95 %
95%-Konfidenzintervall	plausibler Wertebereich für den unbekannten Effekt

Nicht verwechseln α ist kein Konfidenzintervall. $\alpha = 5\%$ gehört zu einem Konfidenzniveau von 95 %. Das Konfidenzintervall selbst ist ein Zahlenbereich.

Entscheidungsregel

Wenn α vorab auf 0,05 festgelegt wurde: $p < 0,05 \rightarrow H_0$ verwerfen, statistisch signifikant. $p \geq 0,05 \rightarrow H_0$ nicht verwerfen. Das beweist weder H_0 noch Gleichheit.

Optische Abschätzung mit Standardabweichungsbalken



Stark überlappende SD-Balken sprechen häufig gegen einen klaren Unterschied; nicht überlappende Balken machen einen Unterschied plausibler. Das ist nur eine grobe Vorabschätzung: SD zeigt die Streuung der Einzelwerte, nicht direkt die Unsicherheit des Mittelwerts. n, Versuchsplan und Test fehlen im Bild.

9 • Welcher Signifikanztest passt?

Die folgende Auswahl ist ein schulgeeigneter Wegweiser für zwei Gruppen. Bei komplexeren Designs sollte eine Fachperson oder geeignete Software hinzugezogen werden.

Daten	Gruppen	Bedingung	Passender Test
nominal	unabhängig	ausreichende erwartete Häufigkeiten	Chi-Quadrat-Test
nominal	unabhängig	kleine erwartete Häufigkeiten	Fisher-Exakt-Test
ordinal / metrisch nicht passend verteilt	unabhängig	zwei Gruppen	Mann-Whitney-U-Test
ordinal / metrisch nicht passend verteilt	gepaart	zwei Messungen	Wilcoxon-Vorzeichen-Rang- Test
metrisch	unabhängig	ungefähr normal, zwei Gruppen	Welch-t-Test
metrisch	gepaart	Differenzen ungefähr normal	gepaarter t-Test

Konkrete Creme-Vergleiche

Vergleich	Mann-Whitney p	Entscheidung bei $\alpha = 0,05$
Creme 1 vs. 2	0,461	nicht signifikant
Creme 1 vs. 4	< 0,001	signifikant
Creme 2 vs. 4	0,074	nicht signifikant
Creme 3 vs. 4	< 0,001	signifikant

Diese Zahlen gelten unter der Annahme unabhängiger Gruppen. Haben dieselben Personen alle Cremes bewertet, sind die Daten gepaart und der Wilcoxon-Test ist passender. Dieselben Zahlen können deshalb je nach Versuchsplan anders ausgewertet werden.

Signifikant $\neq$ wichtig Berichte zusätzlich die Grösse des Unterschieds, Lagewerte, Streuung, n und fachliche Bedeutung.

10 • Eigenes Forschungsprojekt auswerten

A • Vor dem Messen

16. Forschungsfrage: Was genau soll beantwortet werden?
17. Hypothese: Welche überprüfbare Erwartung folgt daraus?
18. Abhängige Variable: Was wird beobachtet oder gemessen?
19. Datenart: nominal, ordinal oder metrisch?
20. Untersuchungseinheit und Stichprobengröße n festlegen.
21. Gruppen unabhängig oder Messungen gepaart?
22. Kontrollvariablen, Einheit, Messmethode und Wiederholungen planen.

B • Nach dem Messen

23. Rohdaten vollständig tabellarisch dokumentieren.
24. Fehlende Werte und mögliche Ausreisser prüfen; nichts heimlich entfernen.
25. Passendes Diagramm erstellen und Achsen beschriften.
26. Geeignete Lage- und Streumasse berechnen.
27. Nur bei metrischen Daten und passender Testfrage die Verteilung prüfen.
28. Test anhand von Datenart und Versuchsplan wählen.
29. p -Wert mit vorab festgelegtem α vergleichen.
30. Ergebnis fachlich interpretieren und Grenzen nennen.

Berichtsvorlage

Satzgerüst „Wir untersuchten ... mit $n = \dots$ unabhängigen Einheiten / Paaren. Die Variable ... war nominal / ordinal / metrisch. Gruppe A zeigte ... (Lagewert und Streuung), Gruppe B Der ...-Test ergab $p = \dots$ bei $\alpha = 0,05$. Der Unterschied war damit statistisch signifikant / nicht signifikant. Inhaltlich bedeutet dies Grenzen der Untersuchung waren“

Kurzcheck & Formelsammlung

Frage	Wenn ja ...	Wenn nein ...
Ist die Variable metrisch?	Mittelwert/SD möglich; Verteilung für Test prüfen	Modalwert/Median und passende Häufigkeits- oder Rangtests
Sind die Messungen gepaart?	Partnerstruktur erhalten; gepaarten Test wählen	unabhängigen Test wählen
Gibt es Ausreisser oder Schiefe?	Median und Rangtest erwägen; Ursache prüfen	parametrischer Test kann passen
Ist $p < \alpha$?	H_0 verwerfen; signifikant	H_0 nicht verwerfen; Gleichheit nicht bewiesen

Formeln für metrische Stichprobendaten

Mittelwert: $\bar{x} = (\text{Summe aller Messwerte}) / n$

Varianz der Stichprobe: $s^2 = \text{Summe der quadrierten Abweichungen von } \bar{x} / (n - 1)$

Standardabweichung: $s = \sqrt{s^2}$

Spannweite: Maximum – Minimum

Excel-Funktionen

Aufgabe	Deutsch / international
Anzahl	ANZAHL / COUNT
Modalwert	MODALWERT.EINF / MODE.SNGL
Median	MEDIAN
Mittelwert	MITTELWERT / AVERAGE
Varianz Stichprobe	VAR.S
Standardabweichung Stichprobe	STABW.S / STDEV.S
Häufigkeit einer Bewertung	ZÄHLENWENN / COUNTIF

Letzter Blick Eine gute statistische Aussage verbindet Forschungsfrage, Versuchsplan, Rohdaten, Diagramm, Kennwerte, Test und fachliche Bedeutung. Kein einzelner p-Wert ersetzt diese Kette.

Grundlage der Zusammenfassung

Erarbeitet aus den bereitgestellten Kursmaterialien: „Statistik – Messen“, „Mathematische Betrachtungen der Statistik“, „Richtigkeit und Präzision“, den Cartoon-Crashkursen zu Messwerten und t-Test sowie den Excel-Beispielen zu Cremebewertungen, Boxplots und Normalverteilung. Fachbegriffe und Aussagen wurden für den Einsatz im Gymnasium vereinheitlicht und korrigiert.