



Grundbegriffe der Statistik

Skript Kapitel 6.2, Seite 130

Grundbegriffe der Statistik

Ich glaube nur *der*
Statistik, die ich
selbst gefälscht habe





„Wer misst, misst Mist...“

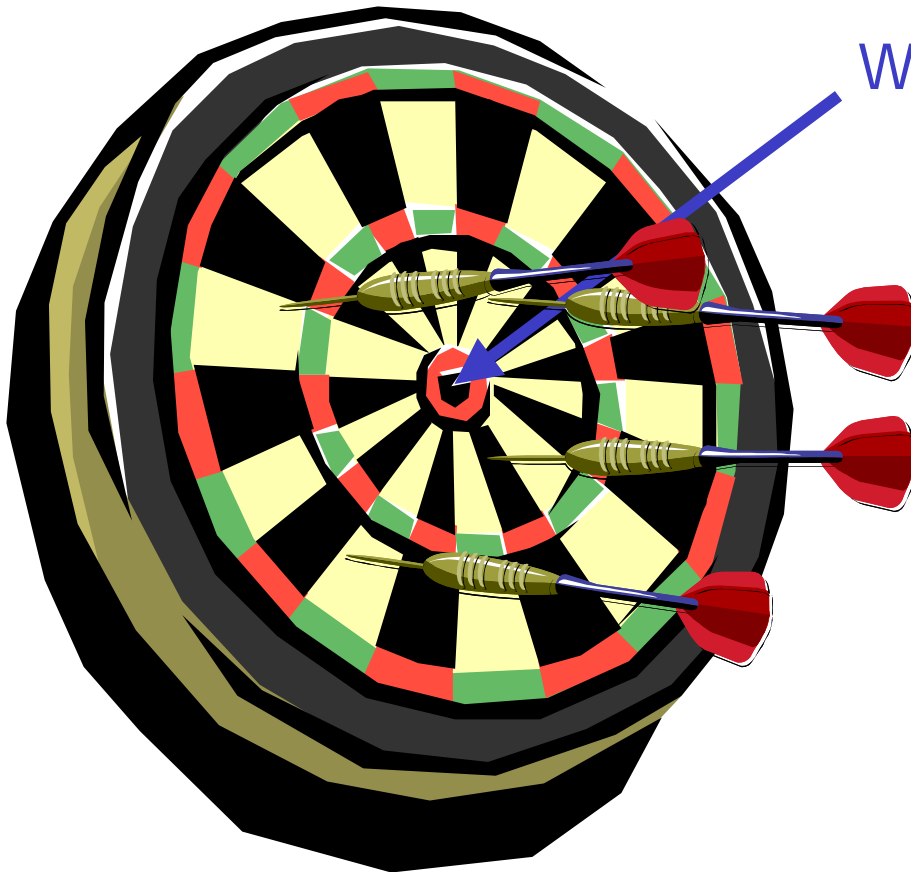
- Analysenergebnisse sind stets mit einem **Fehler** behaftet
- Mit Hilfe von **statistischen** Verfahren lassen sich diese Fehler objektiv beurteilen



Fehler

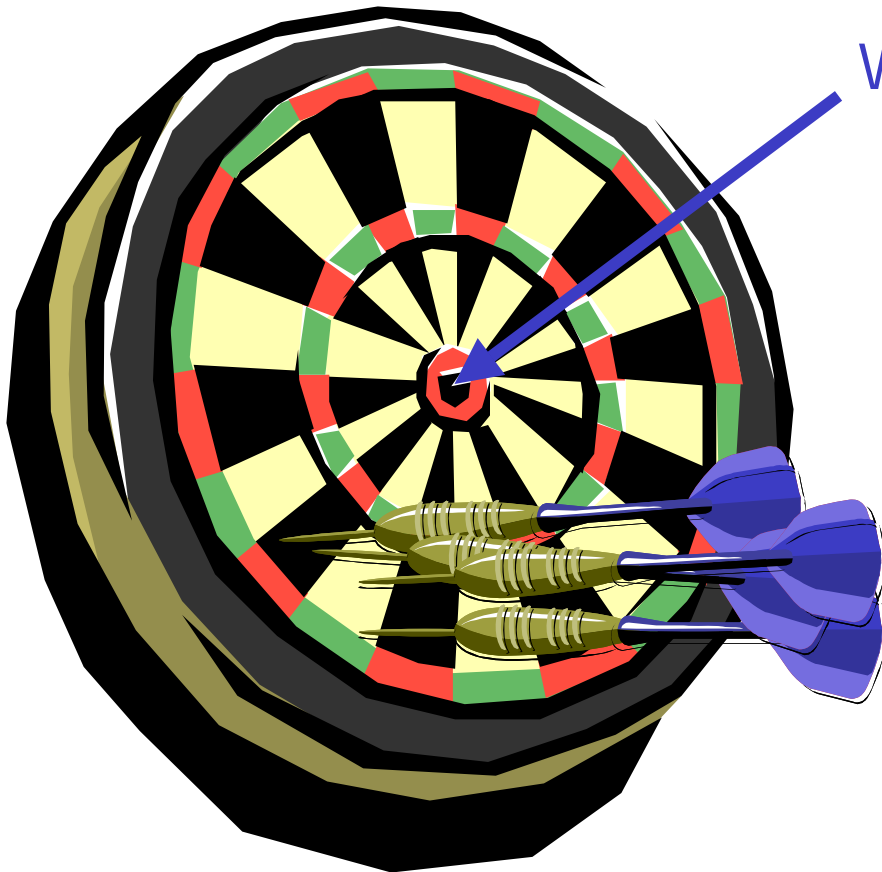
- **Fehler = Differenz** zwischen dem **wahren** und dem **gemessenen Wert**
- Bei „normalen“ Proben kennen wir nur den Messwert, nicht aber den wahren Wert
- Wir unterscheiden zwischen **zufälligen** und **systematischen** Fehlern

Arten von Fehlern: der zufällige Fehler



- Die Werte streuen zufällig um den **Wahren Wert** und häufen sich um einen mittleren Wert.
- Entsteht während der Messung durch nicht beherrschte Änderungen der Probe und der Messgeräte.
- Unregelmässig bezüglich Betrag und Vorzeichen.
- Beispiel: *Zittern* der letzten Ziffer beim Einwägen.

Arten von Fehlern: der **systematische Fehler**



- Die Werte weichen deutlich vom **Wahren Wert** ab.
- Entsteht durch bleibende Unvollkommenheiten im Analysenverfahren und in der Probe.
- Beispiel: Falsche Einwaage bei der Herstellung einer 0.1 molaren Natriumoxalatlösung.



Messunsicherheit: Präzision und Genauigkeit

- Die **Präzision** (auch **Reproduzierbarkeit**) sagt aus, wie gut die Ergebnisse mehrerer Analysen untereinander übereinstimmen.
- Die **Genauigkeit** (auch **Richtigkeit**) beschreibt, wie gut die Ergebnisse mit dem wahren Wert übereinstimmen.

Messunsicherheit: Präzision und Genauigkeit

Zufällige Fehler

Systematische Fehler

Präzision
(Reproduzierbarkeit)

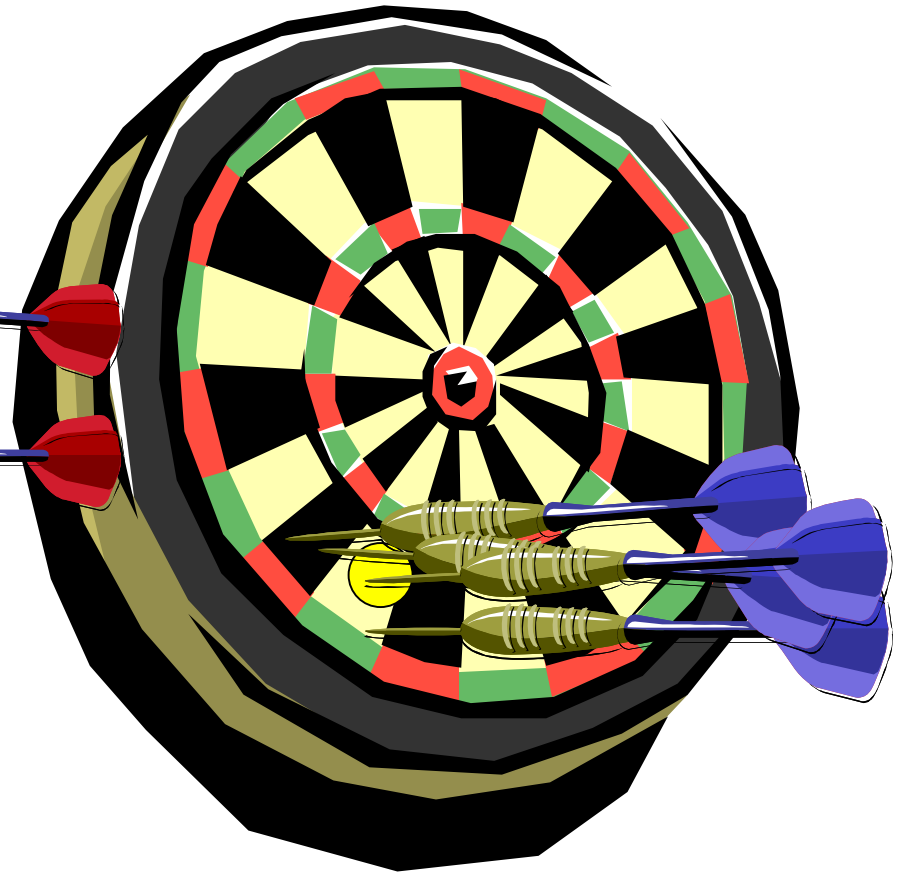
Genauigkeit
(Richtigkeit)

Messunsicherheit

Zufällige und systematische Fehler



Genauigkeit: gut
Präzision: schlecht



Genauigkeit: schlecht
Präzision: gut

Statistische Kenngrößen:

Mittelwert

- Auch „Durchschnitt“ genannt
- Formelzeichen \bar{x} („x quer“)
- Berechnet sich aus der Summe aller Messwerte x_i geteilt durch die Anzahl der Werte n :

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots}{n} = \frac{\sum x_i}{n}$$

Statistische Kenngrößen: Mittelwert

Gesamt-Phosphor
im Belebtschlamm



Messwerte:

5.9 mg P/l

5.4 mg P/l

5.8 mg P/l

5.7 mg P/l

22.8 mg P/l : 4 =

5.7 mg P/l

Statistische Kenngrößen:

Mittlerer Fehler

- Vergleich eines Messwertes mit dem Mittelwert \Rightarrow einzelner absoluter Fehler d
- Mittelwert aus allen absoluten Fehlern \Rightarrow mittlerer Fehler \bar{d}
- Berechnung:

$$\bar{d} = \frac{d_1 + d_2 + d_3 + \dots}{n}$$

$$\bar{d} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + \dots}{n} = \frac{\sum (x_i - \bar{x})}{n}$$

Statistische Kenngrößen: Mittlerer Fehler

Messwert	Fehler (Messwert - Mittelwert)	absoluter Fehler d
5.9	$5.9 - 5.7 = 0.2$	0.2
5.4	$5.4 - 5.7 = -0.3$	0.3
5.8	$5.8 - 5.7 = 0.1$	0.1
5.7	$5.7 - 5.7 = 0.0$	<u>0.0</u>
		0.6 : 4 =

Mittlerer Fehler = 0.15 mg P/l



Statistische Kenngrößen: Standardabweichung

- Wichtiges **Mass für die Streuung**
- Die absoluten Fehler werden vor dem Zusammenzählen quadriert, nach der Mittelwertbildung wird die Wurzel gezogen
- Bei Stichproben wird durch **n-1** dividiert
- Berechnung:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots}{n-1}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Statistische Kenngrößen: Standardabweichung

Messwert	Fehler (Messwert - Mittelwert)	Fehler- Quadrat
5.9	$5.9 - 5.7 = 0.2$	0.04
5.4	$5.4 - 5.7 = -0.3$	0.09
5.8	$5.8 - 5.7 = 0.1$	0.01
5.7	$5.7 - 5.7 = 0.0$	<u>0.00</u>
		0.14 : 3 = 0.0467

n-1

davon die Quadratwurzel \Rightarrow 0.216

Standardabweichung = 0.22 mg P/l



Statistische Kenngrößen: Relative Standardabweichung

- Die relative Standardabweichung (Variationskoeffizient V) erlaubt den Vergleich der Streuung unterschiedlicher Konzentrationen oder Parameter
- Die Standardabweichung wird relativ zum Mittelwert ausgedrückt (in Prozent)
- Berechnung:

$$V = \frac{s}{\bar{x}} \cdot 100\%$$



Statistische Kenngrößen: Relative Standardabweichung

Probe	Mittelwert	Standardabweichung
1	5.7 mg P/l	0.22 mg P/l
2	1.2 mg P/l	0.09 mg P/l

Bei welcher Probe war die Messung reproduzierbarer?
Antwort gibt die **relative Standardabweichung**:

$$\text{Probe 1: } V = \frac{0.22}{5.7} \cdot 100\% = \underline{3.86\%}$$

$$\text{Probe 2: } V = \frac{0.09}{1.2} \cdot 100\% = \underline{7.5\%}$$

Normalverteilung: Tabelle

Kaliumpermanganatverbrauch [mg/l]					
316	341	303	340	260	354
337	320	356	327	300	380
327	316	316	379	319	322
348	344	327	305	274	340
284	403	310	335	350	369
322	360	316	340	319	319
360	272	360	338	310	379
341	250	316	332		



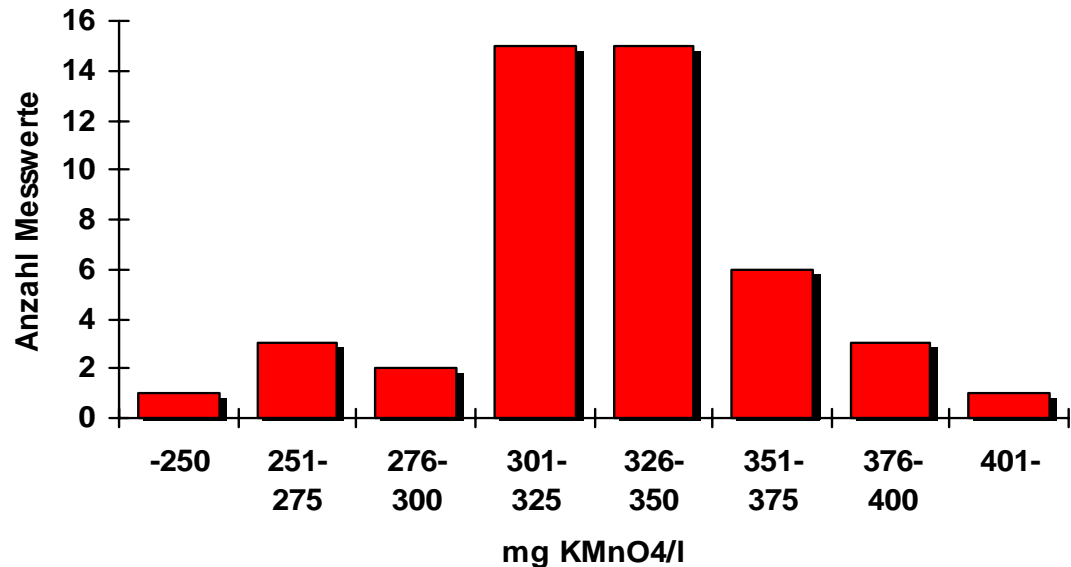
Gruppe	Anzahl
-250	1
251-275	3
276-300	2
301-325	15
326-350	15
351-375	6
376-400	3
401-	1

Um die Verteilung der Daten darstellen zu können, teilen wir die Resultate in acht Gruppen ein und zählen die Messwerte in jeder Gruppe.

Normalverteilung: Häufigkeitsdiagramm

Gruppe	Anzahl
-250	1
251-275	3
276-300	2
301-325	15
326-350	15
351-375	6
376-400	3
401-	1

Häufigkeitsverteilung der KMnO₄-Resultate



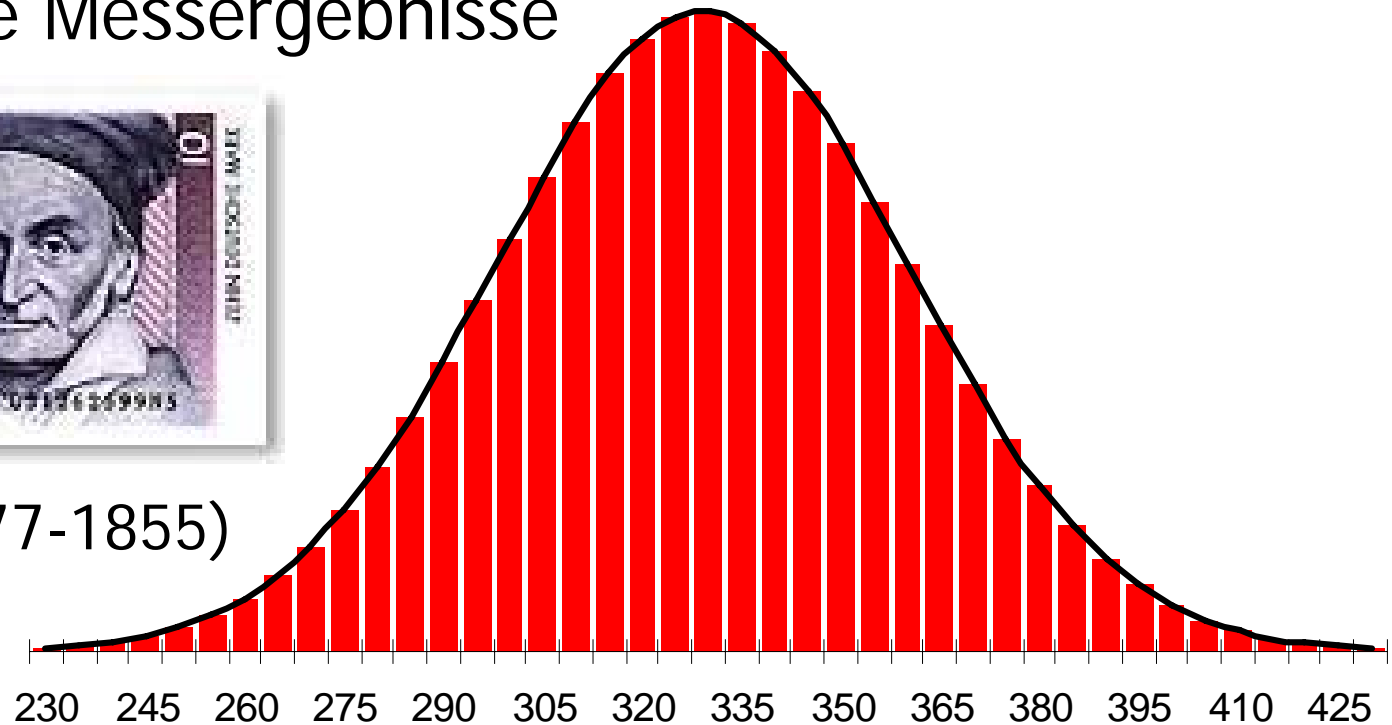
Die grafische Darstellung der gruppierten Daten nennt man Häufigkeitsdiagramm oder **Histogramm**.

Normalverteilung: Gaußsche Glockenkurve

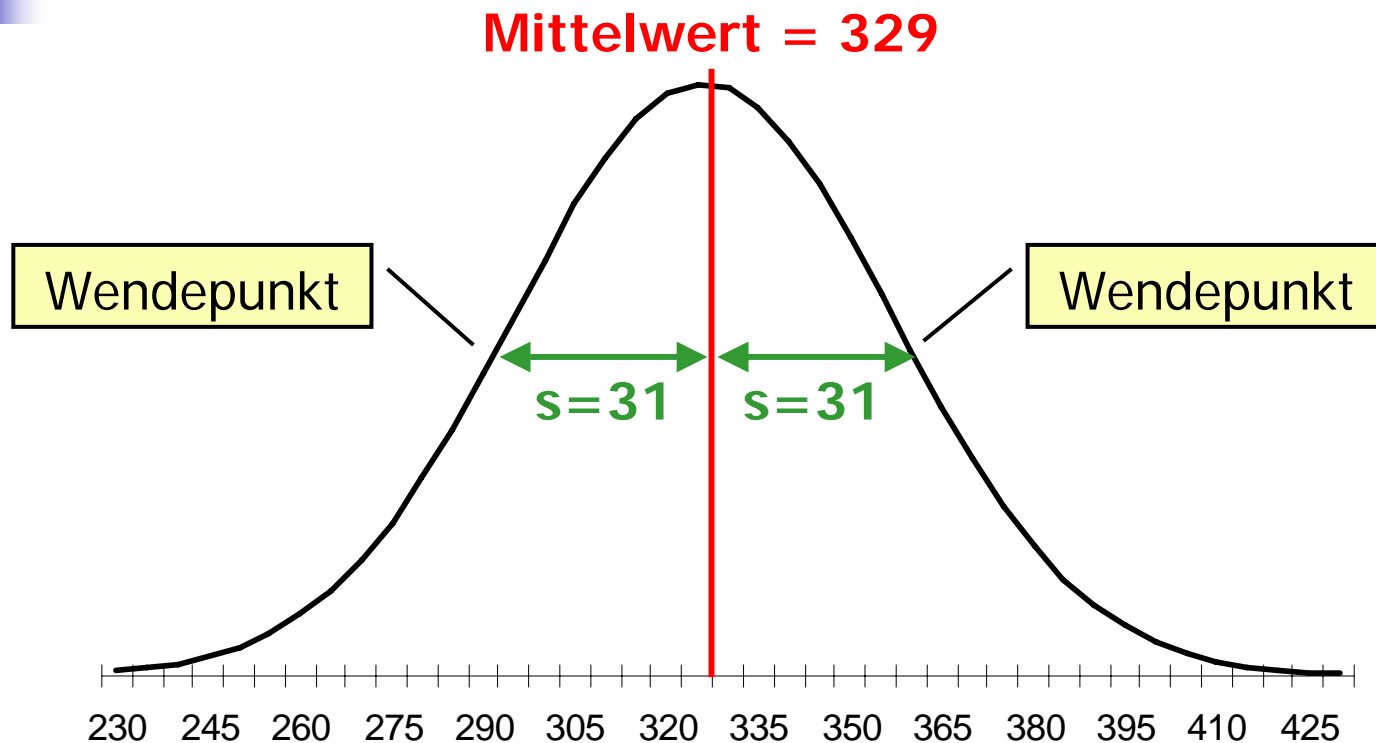
- keine systematischen Fehler
- Gruppenbreite 5 mg/l
- sehr viele Messergebnisse



C.F. Gauß (1777-1855)

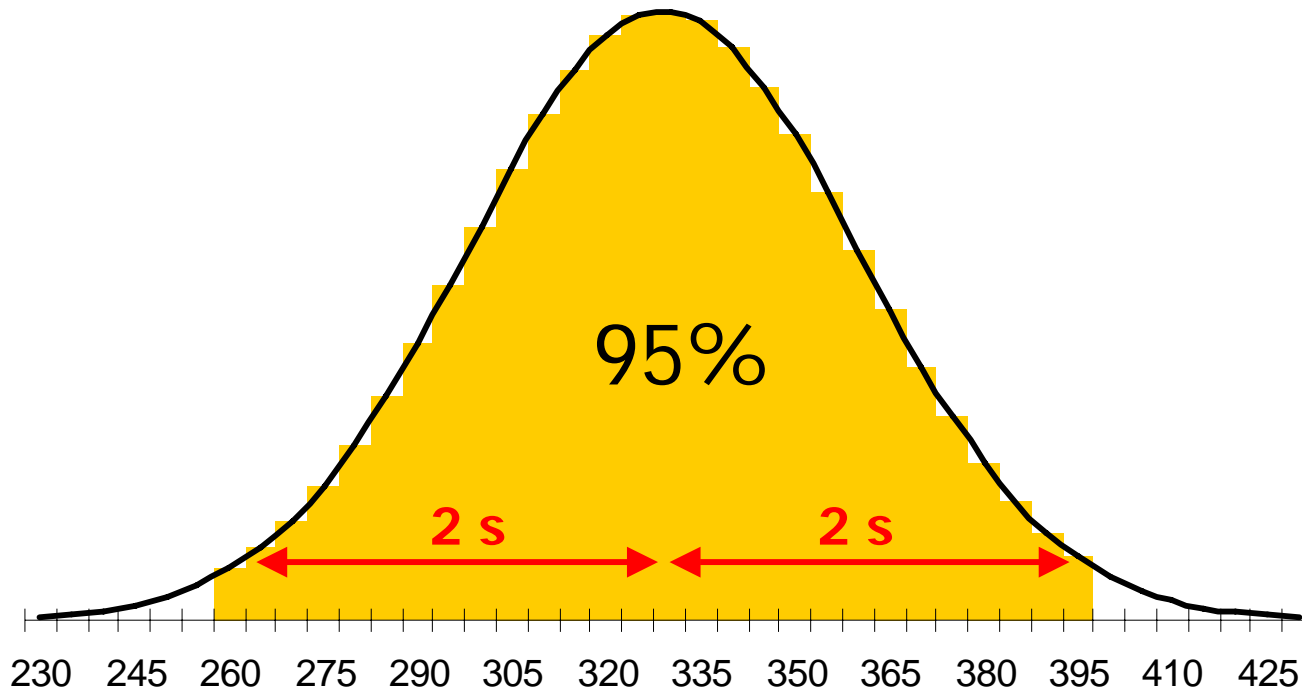


Eigenschaften der Normalverteilung



- Kurve symmetrisch um den **Mittelwert**
- **Standardabweichung** bestimmt die Wendepunkte

Eigenschaften der Normalverteilung



- ca. 2/3 aller Messwerte liegen im Bereich $\bar{x} \pm s$
- 19 von 20 Messwerten liegen im Bereich $\bar{x} \pm 2s$

Statistische Kenngrößen:

Median

- Entspricht bei den nach der Grösse aufgelisteten Messwerten dem in der Mitte liegenden Wert
- Auch „Zentralwert“ genannt
- Unempfindlich gegenüber Ausreissern
- Entspricht bei exakt symmetrischen Verteilungen dem Mittelwert



Statistische Kenngrößen: Spannweite

- Differenz zwischen dem grössten und dem kleinsten Wert einer Reihe von Messwerten
- Formelzeichen R
- Empfindlich gegenüber Ausreissern

Statistische Kenngrößen: Median und Spannweite

Messwerte

2.5

2.6

2.7

2.8

2.7

3.4

Mittelwert

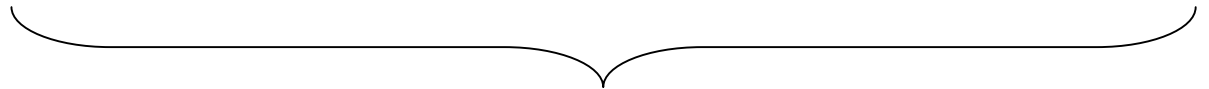
2.8

Median

2.7

Spannweite

0.9



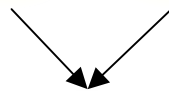
Statistische Kenngrößen: Median

Fall 1: **Ungerade** Anzahl Messwerte:

12.1 12.3 12.6 12.7 13.1

Fall 2: **Gerade** Anzahl Messwerte:

5.4 5.7 5.8 5.9



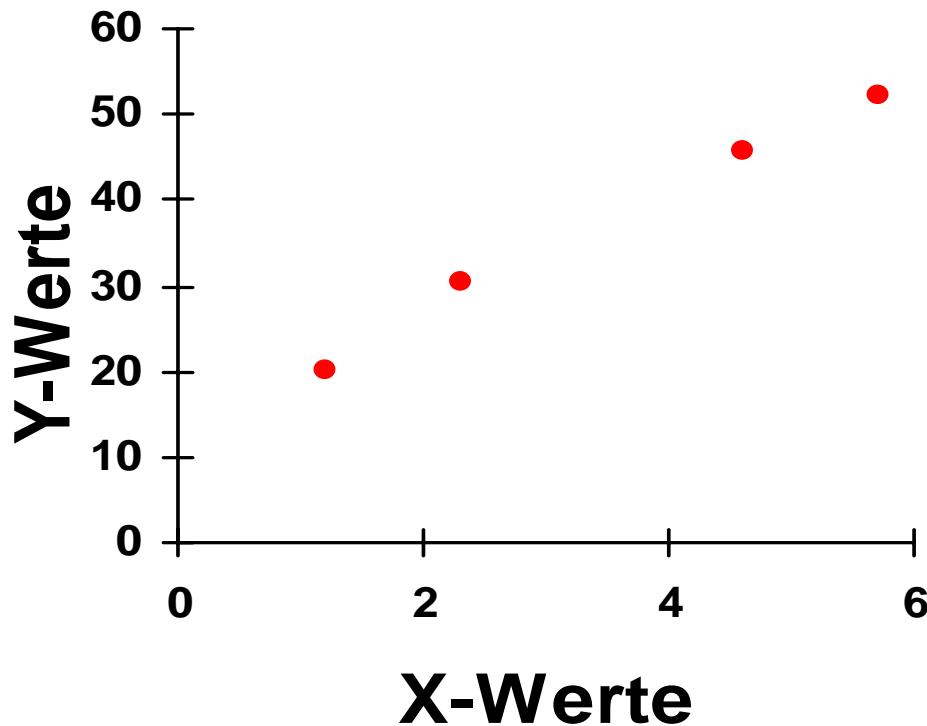
5.75

Grafische Darstellungen: Diagramme

Ein Bild sagt mehr
als tausend Worte!

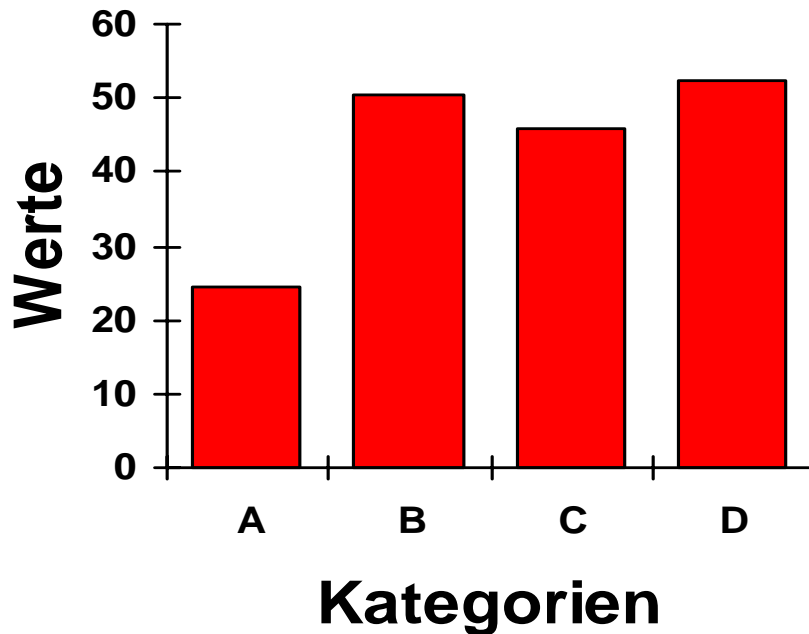


Grafische Darstellungen: XY-Diagramm



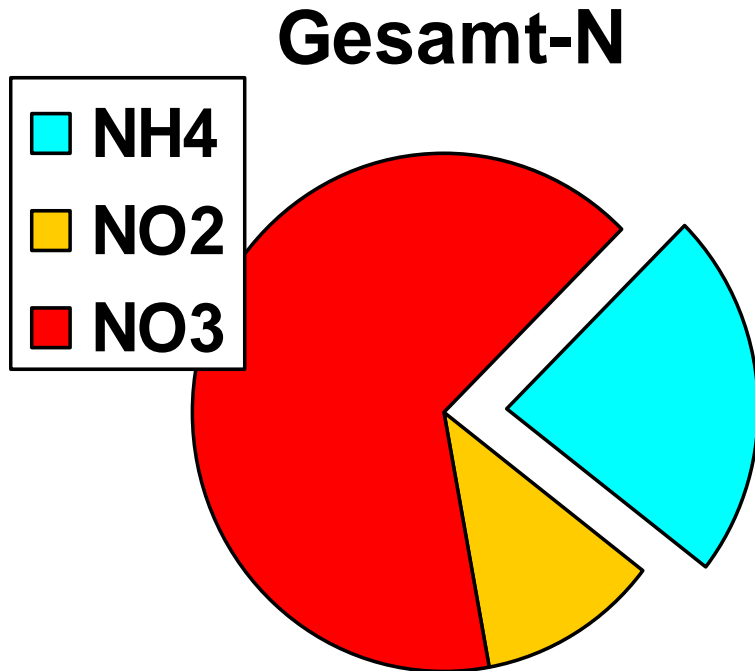
Das XY-Diagramm eignet sich dazu, Zusammenhänge zwischen zwei Zahlenreihen darzustellen

Grafische Darstellungen: Balken/Säulen-Diagramm



Dieser Diagrammtyp eignet sich z.B. für den Vergleich unterschiedlicher Analysemethoden, Proben typen etc.

Grafische Darstellungen: Kuchen-Diagramm



Das Kuchendiagramm zeigt die prozentuelle Aufteilung eines zusammengesetzten Werts in seine Komponenten.



Grafische Darstellungen: Regeln

- Auf optische Spielereien und unübersichtliche 3D-Darstellungen verzichten
- Nicht zu viele Daten ins gleiche Diagramm hineinzwängen
- Das Diagramm soll selbsterklärend sein
- Auf sinnvolle Skalierung der Achsen achten



Qualitätssicherung im Labor

Skript Kapitel 6.3, Seite 134



Analysenkenndaten

Was leistet meine Analysenmethode?
Eignet sie sich überhaupt für die Fragestellung?



Methoden-
validierung

Analysenkenndaten



Analysenkenndaten: Präzision

- Mass für die Reproduzierbarkeit der gesamten analytischen Methode
- Berechnung der relativen Standardabweichung aus mehreren Wiederholmessungen der gleichen Probe

Beispiel: Präzision der PO_4 -P-Messung

$$V = 3.86\% \quad (n=4, c=5.7 \text{ mg/l})$$



Analysenkenndaten: Genauigkeit

- Abweichung eines Resultats vom wahren Wert
- Test mit einer Lösung bekannter Konzentration (z.B. Ringversuch)

Beispiel: Genauigkeit KMnO_4 beim Ringversuch

Sollwert: 377 mg/l Messwert: 416 mg/l

Abweichung: +39 mg/l bzw. +10.3%



Analysenkennndaten: Linearität

- Linearität einer Analysenmethode bedeutet, dass die Messwerte entweder direkt oder nach einer Umrechnung proportional zur Konzentration der interessierenden Substanz sind

Beispiele, wo die Linearität aufhört:

- Extinktion bei hohen Konzentrationen
- KMnO_4 ab 5 ml Titrationsvolumen



Analysenkenndaten: Messbereich

- Derjenige Konzentrationsbereich der interessierenden Substanz in der Probe, in dem eine Messung mit akzeptabler Präzision, Genauigkeit und Linearität möglich ist

Beispiel: BSB₅

Messbereich Sauerstoff in der Verdünnung: 2 bis 6 mg O₂/l



Analysenkenndaten: **Robustheit**

- Kleine Veränderungen im Verfahren sollen keinen Einfluss auf das Resultat haben
- Klärwärter A sollte das gleiche Ergebnis erhalten wie Klärwärter B

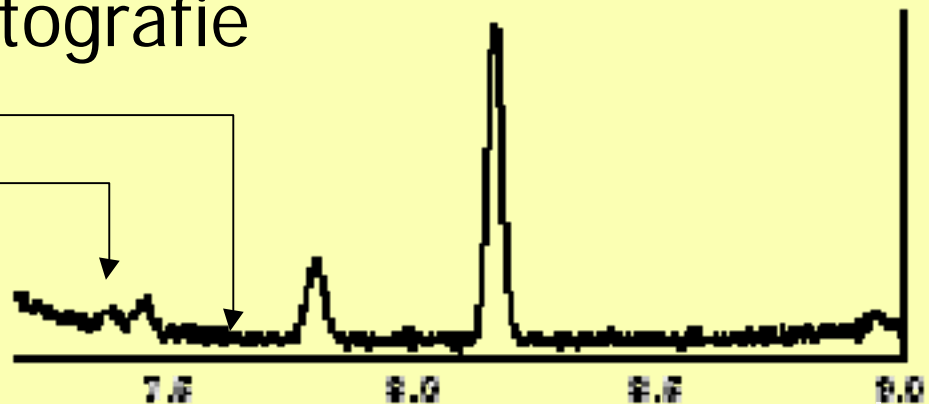
Analysenkenndaten: Nachweisgrenze

- Kleinste Konzentration in einer Probe, die noch nachgewiesen werden kann
- Der gemessene Wert muss sich vom Geräte-*Rauschen* unterscheiden

Beispiel: Chromatografie

Rauschen

Nachweisgrenze



Analysenkenndaten: Bestimmungsgrenze

- Kleinste Konzentration in einer Probe, die sich noch mit einer akzeptablen Präzision und Genauigkeit bestimmen lässt

Beispiel: Unterschiedliche NH_4 -Konzentrationen

- Präzision 5% bei 1 mg/l ✓
- Präzision 10% bei 0.1 mg/l ✓ → Best.grenze
- Präzision 30% bei 0.01 mg/l ☹️

Analysenkennndaten:

Selektivität / Spezifität

- **Selektivität:** Stören andere Komponenten in der Probe die Messung der interessierenden Substanz?
- **Spezifität:** Stammt das gemessene Signal wirklich von der interessierenden Substanz und nicht von anderen Komponenten aus der Probe?



Qualitätssicherung im Labor: **Kontrollkarten**

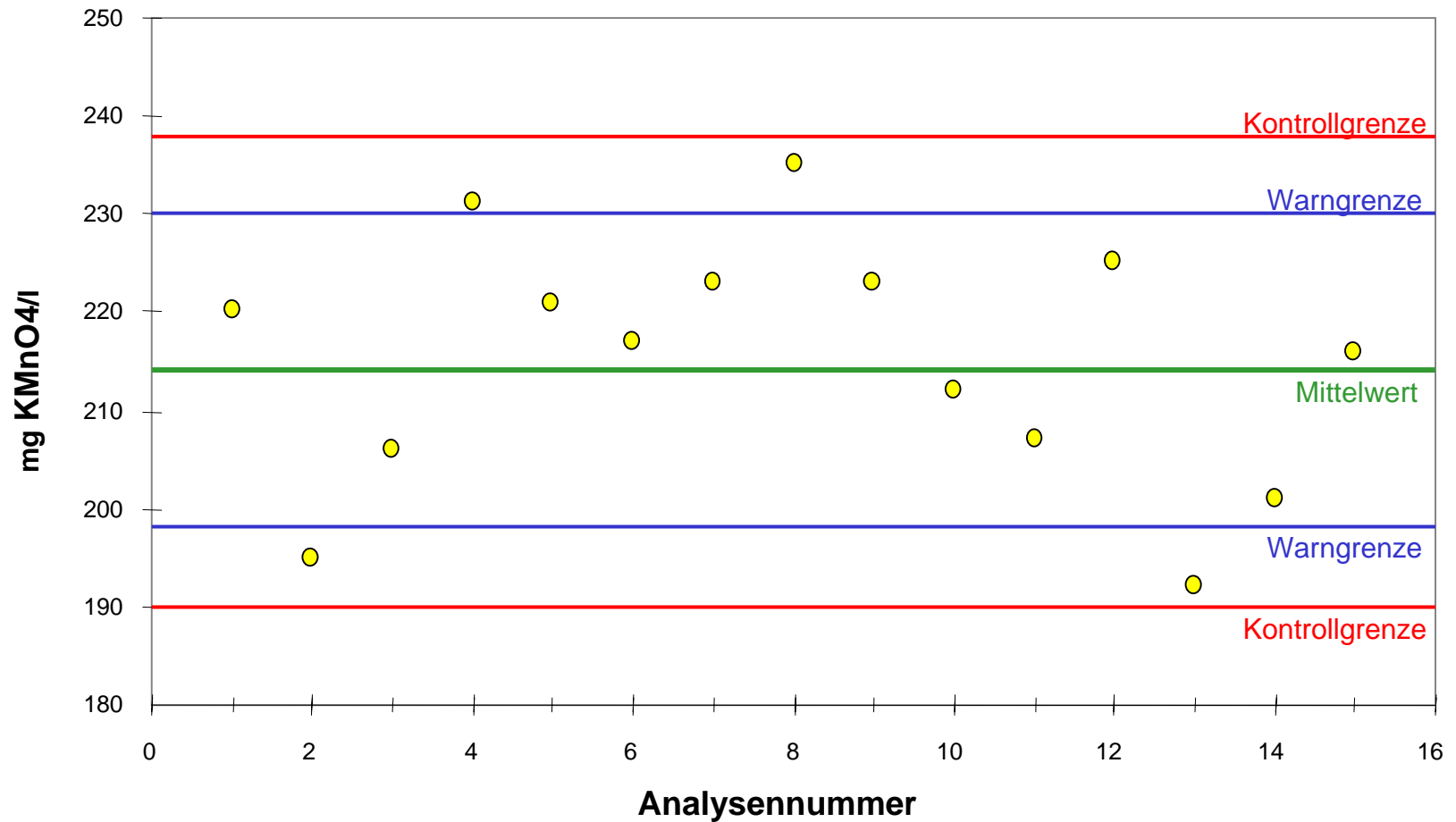
- **Analysenmethode** nicht nur bei der Einführung, sondern auch in der Routine **regelmässig überprüfen**
- Testlösung (Standard oder stabile Probe) mehrmals messen und die Präzision der Analyse bestimmen
- Im Routinebetrieb die Testlösung mitlaufen lassen
- Messergebnisse grafisch auswerten (auf Kontrollkarte eintragen)



Kontrollkarten: Beispiel

- KMnO_4 -Analysenmethode mit Zuckerlösung überprüfen
- Testlösung 7 mal messen:
Mittelwert $\bar{x} = 214$ mg/l
Standardabw. $s = 8$ mg/l
- **Warngrenzen** bei $\bar{x} \pm 2s$ (NV 95%)
→ oben **230** mg/l, unten **198** mg/l
- **Kontrollgrenzen** bei $\bar{x} \pm 3s$ (NV 99.7%)
→ oben **238** mg/l, unten **190** mg/l

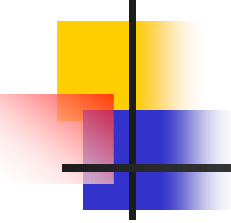
Kontrollkarten: Beispiel



Kontrollkarten:

Methode überprüfen bei...

- 1 Wert ausserhalb der Kontrollgrenze
- 7 aufeinanderfolgende Werte auf einer Seite der Mittelswertlinie
- 7 aufeinanderfolgende Werte mit ansteigender Tendenz
- 7 aufeinanderfolgende Werte mit absteigender Tendenz
- 2 von 3 aufeinanderfolgenden Werten ausserhalb der Warngrenze
- 10 von 11 Werten auf einer Seite der Mittelswertlinie



Aufdecken systematischer Fehler

- **Problem:**

Systematische Fehler werden durch Mehrfachbestimmungen mit der gleichen Methoden nicht entdeckt

- **Lösung:**

- Überprüfung mit einer zweiten, unabhängigen Methode

und/oder

- Vergleichsmessung mit anderen Labors