# Algorithm Engineering 6
# Messen und Vergleichen

Karsten Weicker

F IMN, HTWK Leipzig

# Überblick

# Beispiel

## Quicksort

- Divide&Conquer-Lösung
- nach dem Lehrbuch von Weicker&Weicker
- Pivot-Element als Median von $A[links]$, $A[mitte]$ und $A[rechts]$

## Shellsort

- Iteratives Insertionsort über Teilfolgen mit abnehmender Schrittweite
- Schrittweiten nach Ciura: 44842, 19930, 8858, 3937, 1750, 701, 301, 132, 57, 23, 10, 4, 1

# Beispiel

## Daten

- bis zu 130000 Elemente
- kein Wissen über eine Vorsortierung
- zwei unterschiedliche Datentypen für die Schlüsselwert:
  - `Double`
  - `String`: Länge 2000, nur Zeichen 'b', aber mit jeweils 2% Wahrscheinlichkeit 'a' und 'c', Stringvergleich ist selbst implementiert

# Beispiel

## Experimente

- $n = 2000, 10000, 18000, 26000, 34000, 42000, 50000, 58000, 66000, 74000, 82000, 90000, 98000, 106000, 114000, 122000, 130000$
- pro $n$: 40 Probleminstanzen, für die jeweils beide Algorithmen gemessen werden.
- Ermittlung der Mittelwerte und Standardabweichung (wenigstens!)

# Überblick

# Laufzeiten messen

## Möglichkeiten in Java

- Messung mit

```
long start = System.currentTimeMillis();
aufruf();
long time = System.currentTimeMillis() - start;
```

- für kurze Perioden: `System.nanoTime()`

```
long start1 = System.nanoTime();
long start2 = System.nanoTime();
aufruf();
long stop = System.nanoTime();
long difference = stop - 2*start2 + start1;
```

# Laufzeiten messen

## Weitere Möglichkeiten in Java

- Klasse `StopWatch` aus `org.apache.commons.lang.time`
- Schnittstelle `MethodInterceptor` von Spring

# Laufzeiten messen

## Hinweise für Java

- immer aus dem Programm heraus messen – nicht inklusive JRE-Start
- Just-in-time (JIT) Compiler verändert die gemessenen Zeiten durch Optimierung von Hot-Spots
  - unoptimiert: für jede Messung neue JRE starten
  - optimiert: erst 5–10 Durchläufe ohne Messung – danach messen
- bei mehreren sequentiellen Messungen in einer JRE: dazwischen `Systems.gc()` aufrufen

# Laufzeiten messen

## Speicherbedarf in Java messen

- über das Objekt der Klasse `Runtime` aus `java.lang`

```
Runtime runtime = Runtime.getRuntime();
runtime.gc();
long memory = runtime.totalMemory() -
                          runtime.freeMemory();
```

# Überblick

# Präsentation experimenteller Daten

## Grundsätzliches

- für Publikationen: immer schwarz/weiß
  $\Rightarrow$ ist auch ohne Farbdrucker noch lesbar
- 2D
  abgebildete räumliche Effekte können unterschiedlich interpretiert werden
- keine Animationen
- die Schlussfolgerung soll deutlich werden – es gibt keinen Schönheitspreis

# Präsentation experimenteller Daten

## Tabellen

- i.d.R. nicht für einzelne Messergebnisse
- Übersichtliche Darstellung der exakten Messwerte (Erwartungswert, Standardabweichung)
- ermöglicht spätere statistische Hypothesentests bzw. macht selbige nachvollziehbar
- Darstellung ist eher langweilig, nicht schnell intuitiv lesbar
- sollte immer durch Funktionsplots ergänzt werden

# Präsentation experimenteller Daten

## Beispiel: Double-Werte

- nur Mittelwerte

| $[10^{-6}s]$ | Quicksort | Shellsort |
|---|---|---|
| $n = 2000$ | 143.375 | 125.600 |
| $n = 10000$ | 775.225 | 764.050 |
| $n = 18000$ | 1447.125 | 1494.550 |
| $n = 26000$ | 2156.800 | 2252.175 |
| $n = 34000$ | 2852.625 | 3055.075 |
| $n = 42000$ | 3580.800 | 3866.225 |
| $n = 50000$ | 4313.475 | 4699.300 |
| $n = 58000$ | 5100.225 | 5555.600 |
| $n = 66000$ | 5817.550 | 6427.425 |
| $n = 74000$ | 6674.475 | 7355.950 |

# Präsentation experimenteller Daten

## Beispiel: String-Werte

- mit Standardabweichung für Hypothesentests

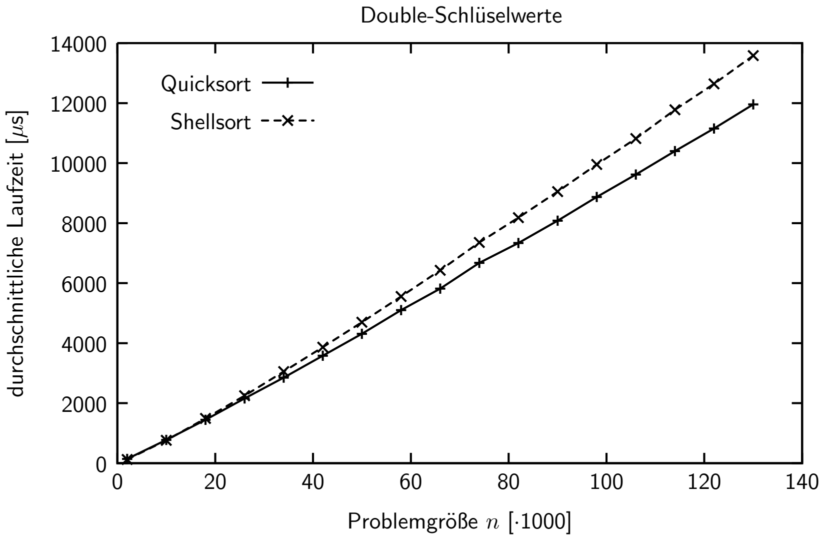
| $[10^{-3}s]$ | Quicksort | Shellsort |
|---|---|---|
| $n = 2000$ | 6.700±0.458 | 1.600±0.490 |
| $n = 10000$ | 37.525±0.632 | 12.825±2.587 |
| $n = 18000$ | 73.950±2.549 | 35.625±4.705 |
| $n = 26000$ | 111.600±5.485 | 71.000±8.680 |
| $n = 34000$ | 148.750±6.390 | 102.150±8.341 |
| $n = 42000$ | 188.900±9.997 | 136.500±10.663 |
| $n = 50000$ | 227.900±11.766 | 170.075±13.180 |
| $n = 58000$ | 268.500±13.719 | 204.850±15.570 |
| $n = 66000$ | 316.350±13.780 | 250.050±17.187 |
| $n = 74000$ | 360.275±17.856 | 285.775±19.605 |

# Präsentation experimenteller Daten

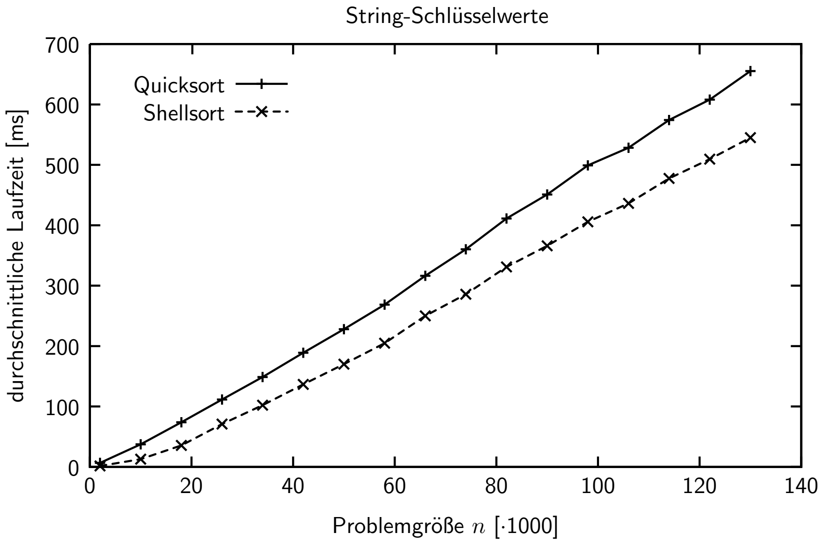
## 2D-Graphen

- X-Achse: i.d.R. Größe der Eingabe
- Y-Achse: benötigte Zeit (Speicher etc.)

# Präsentation experimenteller Daten



# Präsentation experimenteller Daten

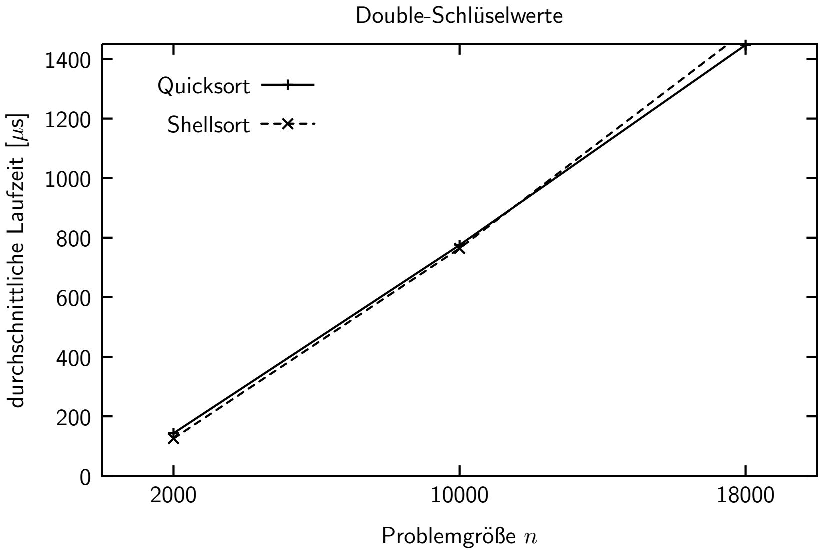


# Präsentation experimenteller Daten

## Unterschiede bei kleinem $n$

- Möglichkeit 1:
  - nur Ausschnitt der x-Werte plotten
  - ggf. große y-Werte abschneiden

# Präsentation experimenteller Daten

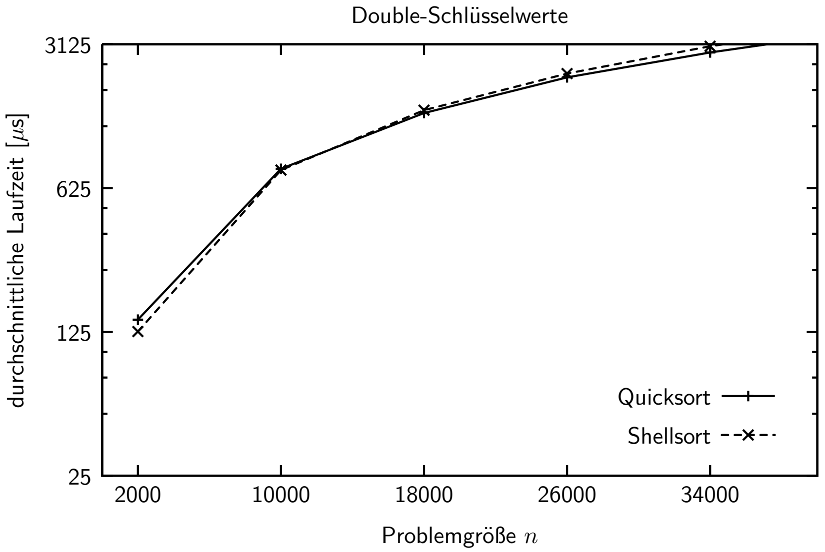


# Präsentation experimenteller Daten

## Unterschiede bei kleinem $n$

- Möglichkeit 2:
    - y-Achse logarithmisch skalieren
    - Vorsicht: Kurven sind nicht mehr intuitiv interpretierbar

# Präsentation experimenteller Daten



# Präsentation experimenteller Daten

## Wichtig

- immer eine Legende (*key*) in einem weißen Bereich angeben
- Achsen beschriften – mit Einheiten
- keine kryptischen Abkürzungen
- zu kleine Schrift vermeiden
- Werte an den Achsen angeben
- nie mehr als 7 Kurven in ein Bild
- gerade Verbindungen implizieren keine Interpolation
  aber immer: Messpunkte kennzeichnen!

# Überblick

# Umgang mit Ausreißern

## Eigenschaften von Messungen

- Ausreißer = Einzelmessungen, die andere Charakteristik aufweisen
- Vermutung: andere Faktoren sind wirksam
- bei Laufzeiten: Garbage-Collector, Systemprozesse etc.
- diese Messungen können ein Ergebnis verfälschen

# Umgang mit Ausreißern

## Lösung 1: Konzentration auf mittlere Performance

- Technik: bei allen Messreihen werden die besten und schlechtesten Messungen nicht berücksichtigt
- Beispiel: Dezil = Zerlegung in 10 sortierte Bereiche – unberücksichtigt bleiben die Messungen im
  - oberen Dezil = beste 10% der Messwerte und im
  - unteren Dezil = schlechteste 10% der Messwerte

| $n$ | Original Erw. | Original Anz. | jeweils $-10\%$ Erw. | jeweils $-10\%$ Anz. |
|---|---|---|---|---|
| 2000 | 6.700 | 40 | 6.750 | 32 |
| 10000 | 37.525 | 40 | 37.500 | 32 |
| 18000 | 73.950 | 40 | 73.625 | 32 |
| 26000 | 111.600 | 40 | 111.094 | 32 |
| 34000 | 148.750 | 40 | 147.688 | 32 |
| 42000 | 188.900 | 40 | 187.531 | 32 |
| 50000 | 227.900 | 40 | 226.031 | 32 |
| 58000 | 268.500 | 40 | 266.938 | 32 |
| 66000 | 316.350 | 40 | 316.469 | 32 |
| 74000 | 360.275 | 40 | 359.000 | 32 |
| 82000 | 411.175 | 40 | 409.813 | 32 |
| 90000 | 450.800 | 40 | 450.313 | 32 |
| 98000 | 499.050 | 40 | 496.781 | 32 |
| 106000 | 528.325 | 40 | 529.438 | 32 |
| 114000 | 574.275 | 40 | 573.594 | 32 |
| 122000 | 608.025 | 40 | 607.063 | 32 |
| 130000 | 655.300 | 40 | 654.406 | 32 |

# Umgang mit Ausreißern

## Lösung 2: statistische Identifikation

- Technik I:
  - Annahme: die Messwerte sind annähernd normalverteilt mit Erwartungswert $\mu$ und Standardabweichung $\sigma$
  - alle Messwerte die außerhalb des Bereichs $[\mu - \sigma, \mu + \sigma]$ nicht berücksichtigen

| | Original | | jenseits $\pm$sdv | | | | jeweils $-10\%$ | |
|---|---|---|---|---|---|---|---|---|
| $n$ | Erw. | Anz. | Erw. | Anz. | $<$ | $>$ | Erw. | Anz. |
| 2000 | 6.700 | 40 | 7.000 | 28 | 12 | 0 | 6.750 | 32 |
| 10000 | 37.525 | 40 | 37.486 | 37 | 1 | 2 | 37.500 | 32 |
| 18000 | 73.950 | 40 | 73.400 | 30 | 4 | 6 | 73.625 | 32 |
| 26000 | 111.600 | 40 | 110.667 | 24 | 8 | 8 | 111.094 | 32 |
| 34000 | 148.750 | 40 | 146.419 | 31 | 2 | 7 | 147.688 | 32 |
| 42000 | 188.900 | 40 | 186.241 | 29 | 4 | 7 | 187.531 | 32 |
| 50000 | 227.900 | 40 | 224.710 | 31 | 3 | 6 | 226.031 | 32 |
| 58000 | 268.500 | 40 | 264.071 | 28 | 4 | 8 | 266.938 | 32 |
| 66000 | 316.350 | 40 | 317.607 | 28 | 7 | 5 | 316.469 | 32 |
| 74000 | 360.275 | 40 | 361.607 | 28 | 8 | 4 | 359.000 | 32 |
| 82000 | 411.175 | 40 | 411.250 | 28 | 7 | 5 | 409.813 | 32 |
| 90000 | 450.800 | 40 | 449.538 | 26 | 7 | 7 | 450.313 | 32 |
| 98000 | 499.050 | 40 | 496.667 | 30 | 5 | 5 | 496.781 | 32 |
| 106000 | 528.325 | 40 | 529.296 | 27 | 6 | 7 | 529.438 | 32 |
| 114000 | 574.275 | 40 | 573.433 | 30 | 5 | 5 | 573.594 | 32 |
| 122000 | 608.025 | 40 | 607.867 | 30 | 6 | 4 | 607.063 | 32 |
| 130000 | 655.300 | 40 | 655.097 | 31 | 5 | 4 | 654.406 | 32 |

# Umgang mit Ausreißern

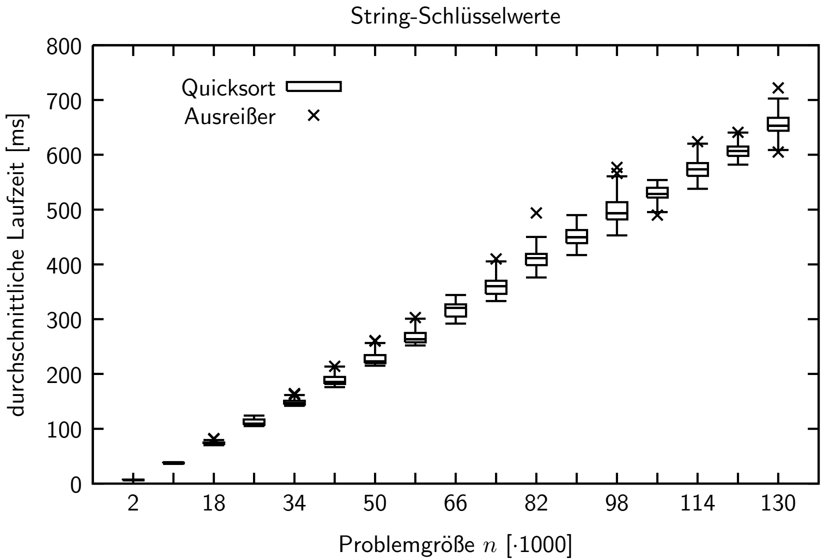
## Lösung 2: statistische Identifikation

- Technik II:
  - betrachten Median $Q_{0.5}$ und Quartilen $Q_{0.25}$ und $Q_{0.75}$
  - Werte außerhalb des Bereichs

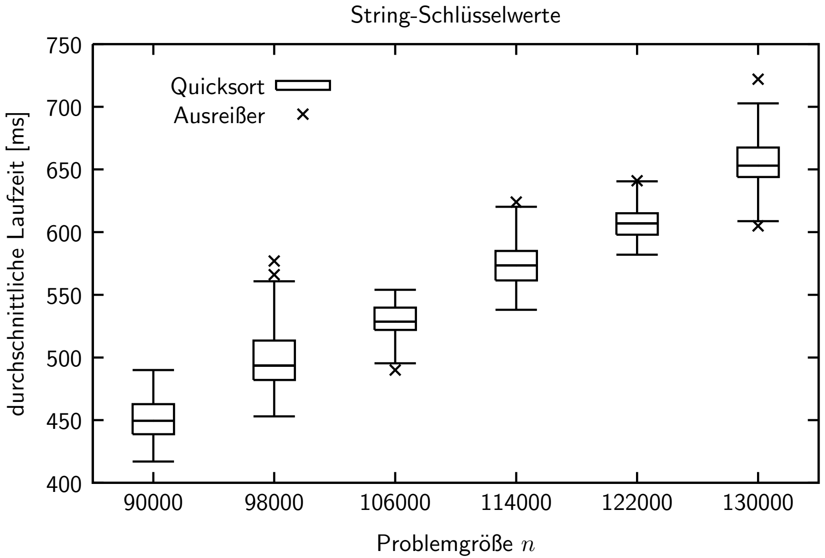
$$[Q_{0.25} - 1.5 \cdot (Q_{0.75} - Q_{0.25}), Q_{0.75} + 1.5 \cdot (Q_{0.75} - Q_{0.25})] = [a, b]$$

    nicht berücksichtigen
  - Grafische Darstellung: Boxplot
    - Rechtecke mit Linie markieren Median und Quartile
    - Whiskers markieren $\min\{b, t_{\max}\}$ und $\max\{a, t_{\min}\}$
    - Ausreißer werden eingezeichnet

# Umgang mit Ausreißern



# Umgang mit Ausreißern



| $n$ | Orig. Erw. | jens. 3·Quart.Diff. Erw. | Anz. | $<$ | $>$ | jens. $\pm$sdv Erw. | Anz. | jeweils $-10\%$ Erw. | Anz. |
|---|---|---|---|---|---|---|---|---|---|
| 2000 | 6.700 | 6.700 | 40 | 0 | 0 | 7.000 | 28 | 6.750 | 32 |
| 10000 | 37.525 | 37.525 | 40 | 0 | 0 | 37.486 | 37 | 37.500 | 32 |
| 18000 | 73.950 | 73.553 | 38 | 0 | 2 | 73.400 | 30 | 73.625 | 32 |
| 26000 | 111.600 | 111.600 | 40 | 0 | 0 | 110.667 | 24 | 111.094 | 32 |
| 34000 | 148.750 | 147.111 | 36 | 0 | 4 | 146.419 | 31 | 147.688 | 32 |
| 42000 | 188.900 | 188.256 | 39 | 0 | 1 | 186.241 | 29 | 187.531 | 32 |
| 50000 | 227.900 | 226.184 | 38 | 0 | 2 | 224.710 | 31 | 226.031 | 32 |
| 58000 | 268.500 | 267.615 | 39 | 0 | 1 | 264.071 | 28 | 266.938 | 32 |
| 66000 | 316.350 | 316.350 | 40 | 0 | 0 | 317.607 | 28 | 316.469 | 32 |
| 74000 | 360.275 | 359.000 | 39 | 0 | 1 | 361.607 | 28 | 359.000 | 32 |
| 82000 | 411.175 | 409.051 | 39 | 0 | 1 | 411.250 | 28 | 409.813 | 32 |
| 90000 | 450.800 | 450.800 | 40 | 0 | 0 | 449.538 | 26 | 450.313 | 32 |
| 98000 | 499.050 | 495.237 | 38 | 0 | 2 | 496.667 | 30 | 496.781 | 32 |
| 106000 | 528.325 | 529.308 | 39 | 1 | 0 | 529.296 | 27 | 529.438 | 32 |
| 114000 | 574.275 | 573.000 | 39 | 0 | 1 | 573.433 | 30 | 573.594 | 32 |
| 122000 | 608.025 | 607.179 | 39 | 0 | 1 | 607.867 | 30 | 607.063 | 32 |
| 130000 | 655.300 | 654.868 | 38 | 1 | 1 | 655.097 | 31 | 654.406 | 32 |

# Überblick

# Statistische Aussagen

## Empirische Forschung

- Erkenntnisse beruhen auf Erfahrungen
- z.B. Beobachtung, Experiment & Induktion
- Problem: Einzelbeobachtungen (deskriptive Aussage) vs. allgemeine Gesetzmäßigkeiten
- Explikative Aussage: Generalisierung aus einzelnen Beobachtungen
- Anwendung von Hypothesentests
- Beispiel: *„Der Hypothesentest zeigt, dass Algorithmus A schneller als Algorithmus B ist (mit einem Fehler von 2%).“*

# Statistische Aussagen

## Algorithmen vergleichen mit Hypothesentests

- Messungen zu zwei Algorithmen $A_1$ $(x_1, \ldots, x_n)$ und $A_2$ $(y_1, \ldots, y_m)$ — *wichtig: alle Messungen mit gleicher Problemgröße!*
- Mittelwerte der Messungen: $\mu_1$ und $\mu_2$
- falls $\mu_1 < \mu_2$: Zufall? Wiederholt beobachtbar?
- Null-Hypothese $H_o$: $\tilde{\mu}_1 \geq \tilde{\mu}_2$
- Alternativ-Hypothese $H_1$: $\tilde{\mu}_1 < \tilde{\mu}_2$
- statistischer Hypothesentest: versucht Null-Hypothese zu widerlegen und gibt einen Fehler $\alpha$ (Signifikanzniveau) für den Irrtum an

# Statistische Aussagen

## Zweistichproben-t-Test für abhängige Stichproben

- $x_i$ und $y_i$ wurden auf derselben Probleminstanz $i$ gemessen ($n = m$)
- berechne Differenz: $d_i = x_i - y_i$
- $H_0$: $\tilde{\mu}_d \leq 0$; $H_1$: $\tilde{\mu}_d > 0$
- Annahme: Werte $d_i$ seien normalverteilt (Erwartungswert $\mu_d$ und Standardabw. $\sigma_d$)
- t-Wert: $t = \sqrt{n}\frac{\mu_d - 0}{\sigma_d}$
- Nullhypothese ablehnen, wenn $t > t(1-\alpha, n-1)$

# Statistische Aussagen

## Beispiel: Double-Schlüsselwerte, $n = 2000$

**Anderer Rechner! andere Messreihe!**

- Quicksort: $\mu_1 = 891.625$, $\sigma_1 = 40.994$
- Shellsort: $\mu_2 = 727.9$, $\sigma_2 = 37.603$
- Differenz: $\mu_d = 163.725$, $\sigma_d = 59.050$
- t-Wert: $t = \sqrt{40}\frac{163.725}{59.050} = 17.536$
- Signifikanzniveau $\alpha = 0.999$, d.h. maximale Irrtumswahrscheinlichkeit von 0.01
- Nullhypothese abgelehnt wegen: $t(0.999, 39) \approx 3.313 < t$

$\Rightarrow$ Shellsort ist hier schneller als Quicksort

# Statistische Aussagen

## Beispiel: Double-Schlüsselwerte, $n = 18000$

**Anderer Rechner! andere Messreihe!**

- Shellsort: $\mu_1 = 9332.775$, $\sigma_1 = 65.496$
- Quicksort: $\mu_2 = 9164.4$, $\sigma_2 = 122.965$
- Differenz: $\mu_d = 168.375$, $\sigma_d = 139.202$
- t-Wert: $t = \sqrt{40}\frac{168.375}{139.202} = 7.65$
- Signifikanzniveau $\alpha = 0.999$, d.h. maximale Irrtumswahrscheinlichkeit von 0.01
- Nullhypothese abgelehnt wegen: $t(0.999, 39) \approx 3.313 < t$

$\Rightarrow$ Quicksort ist hier schneller als Shellsort

| Anz. der Freih. grade | $(1-\alpha)$ für zweiseitigen Test | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0,5 | 0,75 | 0,8 | 0,9 | 0,95 | 0,98 | 0,99 | 0,998 |
| | $(1-\alpha)$ für einseitigen Test | | | | | | | |
| | 0,75 | 0,875 | 0,90 | 0,95 | 0,975 | 0,99 | 0,995 | 0,999 |
| 1 | 1,000 | 2,414 | 3,078 | 6,314 | 12,706 | 31,821 | 63,657 | 318,309 |
| 2 | 0,816 | 1,604 | 1,886 | 2,920 | 4,303 | 6,965 | 9,925 | 22,327 |
| 3 | 0,765 | 1,423 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 | 10,215 |
| 4 | 0,741 | 1,344 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 | 7,173 |
| 5 | 0,727 | 1,301 | 1,476 | 2,015 | 2,571 | 3,365 | 4,032 | 5,893 |
| 6 | 0,718 | 1,273 | 1,440 | 1,943 | 2,447 | 3,143 | 3,707 | 5,208 |
| 7 | 0,711 | 1,254 | 1,415 | 1,895 | 2,365 | 2,998 | 3,499 | 4,785 |
| 8 | 0,706 | 1,240 | 1,397 | 1,860 | 2,306 | 2,896 | 3,355 | 4,501 |
| 9 | 0,703 | 1,230 | 1,383 | 1,833 | 2,262 | 2,821 | 3,250 | 4,297 |
| 10 | 0,700 | 1,221 | 1,372 | 1,812 | 2,228 | 2,764 | 3,169 | 4,144 |
| 11 | 0,697 | 1,214 | 1,363 | 1,796 | 2,201 | 2,718 | 3,106 | 4,025 |
| 12 | 0,695 | 1,209 | 1,356 | 1,782 | 2,179 | 2,681 | 3,055 | 3,930 |
| 13 | 0,694 | 1,204 | 1,350 | 1,771 | 2,160 | 2,650 | 3,012 | 3,852 |
| 14 | 0,692 | 1,200 | 1,345 | 1,761 | 2,145 | 2,624 | 2,977 | 3,787 |
| 15 | 0,691 | 1,197 | 1,341 | 1,753 | 2,131 | 2,602 | 2,947 | 3,733 |

| Anz. der Freih. grade | $(1-\alpha)$ für zweiseitigen Test | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0,5 | 0,75 | 0,8 | 0,9 | 0,95 | 0,98 | 0,99 | 0,998 |
| | $(1-\alpha)$ für einseitigen Test | | | | | | | |
| | 0,75 | 0,875 | 0,90 | 0,95 | 0,975 | 0,99 | 0,995 | 0,999 |
| 16 | 0,690 | 1,194 | 1,337 | 1,746 | 2,120 | 2,583 | 2,921 | 3,686 |
| 17 | 0,689 | 1,191 | 1,333 | 1,740 | 2,110 | 2,567 | 2,898 | 3,646 |
| 18 | 0,688 | 1,189 | 1,330 | 1,734 | 2,101 | 2,552 | 2,878 | 3,610 |
| 19 | 0,688 | 1,187 | 1,328 | 1,729 | 2,093 | 2,539 | 2,861 | 3,579 |
| 20 | 0,687 | 1,185 | 1,325 | 1,725 | 2,086 | 2,528 | 2,845 | 3,552 |
| 21 | 0,686 | 1,183 | 1,323 | 1,721 | 2,080 | 2,518 | 2,831 | 3,527 |
| 22 | 0,686 | 1,182 | 1,321 | 1,717 | 2,074 | 2,508 | 2,819 | 3,505 |
| 23 | 0,685 | 1,180 | 1,319 | 1,714 | 2,069 | 2,500 | 2,807 | 3,485 |
| 24 | 0,685 | 1,179 | 1,318 | 1,711 | 2,064 | 2,492 | 2,797 | 3,467 |
| 25 | 0,684 | 1,178 | 1,316 | 1,708 | 2,060 | 2,485 | 2,787 | 3,450 |
| 26 | 0,684 | 1,177 | 1,315 | 1,706 | 2,056 | 2,479 | 2,779 | 3,435 |
| 27 | 0,684 | 1,176 | 1,314 | 1,703 | 2,052 | 2,473 | 2,771 | 3,421 |
| 28 | 0,683 | 1,175 | 1,313 | 1,701 | 2,048 | 2,467 | 2,763 | 3,408 |
| 29 | 0,683 | 1,174 | 1,311 | 1,699 | 2,045 | 2,462 | 2,756 | 3,396 |
| 30 | 0,683 | 1,173 | 1,310 | 1,697 | 2,042 | 2,457 | 2,750 | 3,385 |

| Anz. der Freih. grade | $(1-\alpha)$ für zweiseitigen Test | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0,5 | 0,75 | 0,8 | 0,9 | 0,95 | 0,98 | 0,99 | 0,998 |
| | $(1-\alpha)$ für einseitigen Test | | | | | | | |
| | 0,75 | 0,875 | 0,90 | 0,95 | 0,975 | 0,99 | 0,995 | 0,999 |
| 40 | 0,681 | 1,167 | 1,303 | 1,684 | 2,021 | 2,423 | 2,704 | 3,307 |
| 50 | 0,679 | 1,164 | 1,299 | 1,676 | 2,009 | 2,403 | 2,678 | 3,261 |
| 60 | 0,679 | 1,162 | 1,296 | 1,671 | 2,000 | 2,390 | 2,660 | 3,232 |
| 70 | 0,678 | 1,160 | 1,294 | 1,667 | 1,994 | 2,381 | 2,648 | 3,211 |
| 80 | 0,678 | 1,159 | 1,292 | 1,664 | 1,990 | 2,374 | 2,639 | 3,195 |
| 90 | 0,677 | 1,158 | 1,291 | 1,662 | 1,987 | 2,368 | 2,632 | 3,183 |
| 100 | 0,677 | 1,157 | 1,290 | 1,660 | 1,984 | 2,364 | 2,626 | 3,174 |
| 200 | 0,676 | 1,154 | 1,286 | 1,653 | 1,972 | 2,345 | 2,601 | 3,131 |
| 300 | 0,675 | 1,153 | 1,284 | 1,650 | 1,968 | 2,339 | 2,592 | 3,118 |
| 400 | 0,675 | 1,152 | 1,284 | 1,649 | 1,966 | 2,336 | 2,588 | 3,111 |
| 500 | 0,675 | 1,152 | 1,283 | 1,648 | 1,965 | 2,334 | 2,586 | 3,107 |
| $\infty$ | 0,674 | 1,150 | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 | 3,090 |

# Statistische Aussagen

## Zweistichproben-t-Test für unabhängige Stichproben

- unterschiedliche Probleminstanzen und/oder $n \neq m$
- Annahme: $x_i$ und $y_i$ sind normalverteilt
- $H_0$: $\tilde{\mu}_x - \tilde{\mu}_y \leq 0$; $H_1$: $\tilde{\mu}_x - \tilde{\mu}_y \leq 0$
- t-Wert: $t = \sqrt{\frac{nm}{n+m}} \cdot \frac{\mu_x - \mu_y - 0}{\sigma}$ mit $\sigma = \sqrt{\frac{(n-1)\sigma_x^2 + (m-1)\sigma_y^2}{n+m-2}}$
- Nullhypothese ablehnen, wenn $t > t(1-\alpha, n+m-2)$

# Statistische Aussagen

## Beispiel: Double-Schlüsselwerte, $n = 18000$

- je 10% Ausreißer entfernt, d.h. $n = m = 32$
- Shellsort: $\mu_1 = 9319.813$, $\sigma_1 = 25.971$
- Quicksort: $\mu_2 = 9155.156$, $\sigma_2 = 82.983$
- gemeinsame Sdv:
  $\sigma = \sqrt{\frac{(31)25.971^2+(31)82.983^2}{32+32-2}} = 61.484$
- t-Wert:
  $t = \sqrt{\frac{32 \cdot 32}{32+32}} \cdot \frac{9319.813-9155.156-0}{61.484} = 10.71211$
- Nullhypothese abgelehnt wegen:
  $t(0.999, 78) \approx 3.198 < t$

$\Rightarrow$ Quicksort ist hier schneller als Shellsort

# Statistische Aussagen

## Veränderte Rahmenbedingungen

- Beispiel: ein Algorithmus wird auf speziellen Daten angewandt
- Frage: wirkt sich die Spezifik der Daten (im Gegensatz zu zufälligen Daten) auf die Laufzeit aus?
- zweiseitiger Test – $H_0$: $\tilde{\mu}_1 = \tilde{\mu}_2$; $H_1$: $\tilde{\mu}_1 \neq \tilde{\mu}_2$
- Test ist analog – nur mit dem zweiseitigen Vertrauensbereich in den t-Quantil-Tabellen

# Statistische Aussagen

## Irrtumswahrscheinlichkeit

- der erlaubte Fehler für einen Hypothesentest wird vorgegeben
- sog. Signifikanzniveau $\alpha$
- typische Werte:
  - $\alpha \leq 0.05$: signifikanter Unterschied
  - $\alpha \leq 0.01$: hoch signifikanter Unterschied
  - $\alpha \leq 0.001$: höchst signifikanter Unterschied

# Statistische Aussagen

## Bessere Testverfahren

- Wilcoxon-Vorzeichen-Rang-Test (Wilcoxon signed-rank test) – eine Population, weniger strikte Voraussetzungen
- Mann-Whitney-Wilcoxon Test – X und Y beziehen sich auf unterschiedliche Populationen
- Software: SPSS, Matlab, Mathematica, R
- Beispiel in R: `wilcox.test(x, y, paired = TRUE, alternative = "greater")`

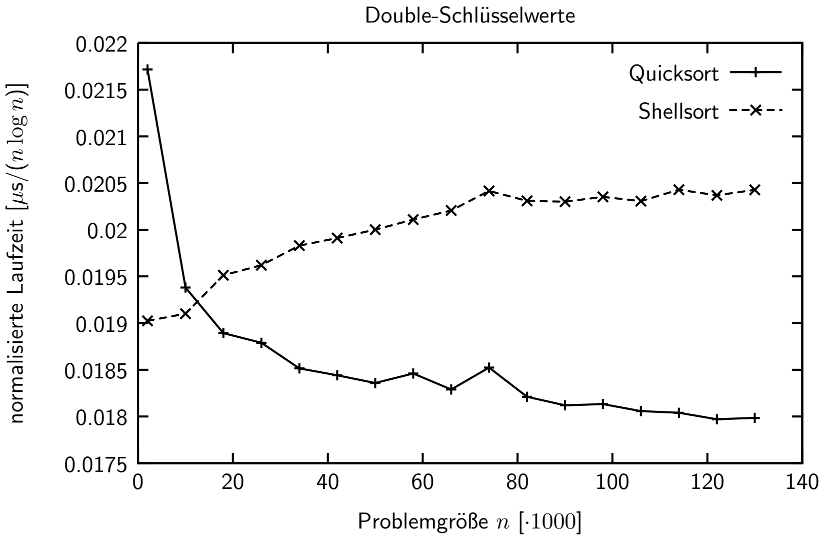
# Überblick

# Vergleich mit asymptotischer Laufzeit

## Präsentation experimenteller Daten

- Y-Achse zur theoretischen Laufzeit skalieren
- Technik: alle Datenpunkte durch die asymptotische Laufzeit teilen
- dann: alle Kurven müssten waagerecht sein
- im Beispiel:
  - bei Quicksort: $n \log n$ (im Average-Case)
  - bei Shellsort mit Ciura-Schrittfolge: unbekannt

# Vergleich mit asymptotischer Laufzeit

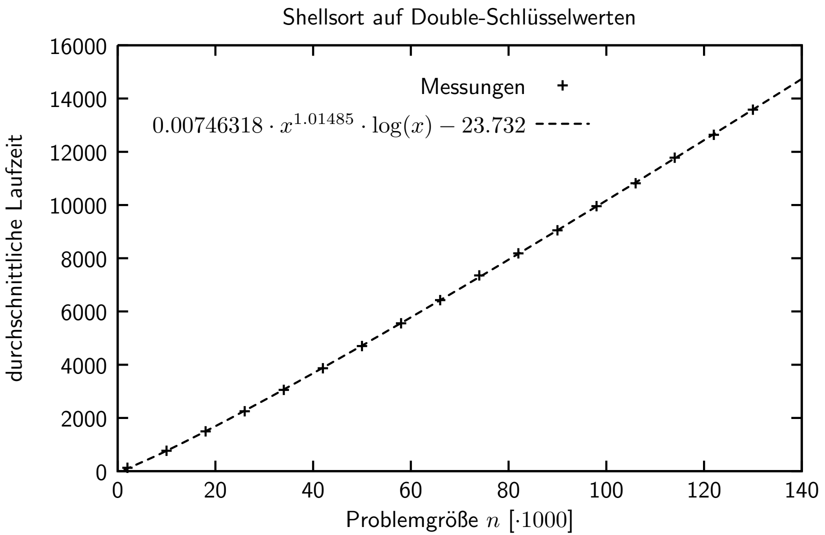


# Vergleich mit asymptotischer Laufzeit

## Empirisch gestütztes Laufzeitmodell: Ping-Pong-Regression

1. Laufzeiten messen
2. Ein Modell mit offenen Konstanten formulieren z.B.: $f(x) = a \cdot x^b$
3. per Regression (z.B. in Matlab, R, Gnuplot) das Modell an die Messdaten angleichen lassen
4. vorherige Schritte solange (mit unterschiedlichen Startwerten für die Konstanten oder anderen Modellen) iterieren, bis Datenpunkte und Kurven tendentiell passen

# Vergleich mit asymptotischer Laufzeit



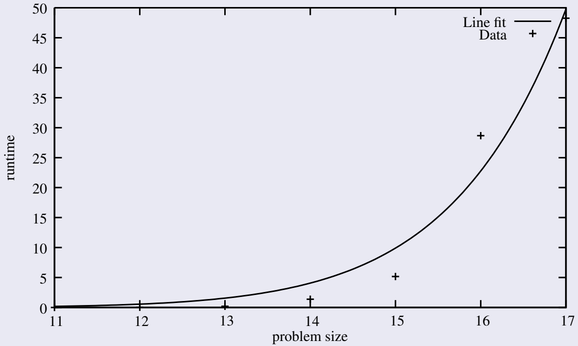
# Vergleich mit asymptotischer Laufzeit

## Zweites Beispiel: Branch&Bound

- Backtracking-Algorithmus für TSP, der Branch&Bound benutzt
- Backtracking selbst hat Laufzeit $\Theta(n!)$
- Welche Laufzeit erreicht der Branch&Bound-Algorithmus?

# Vergleich mit asymptotischer Laufzeit

## Bestes Ergebnis in Gnuplot




$f(x) = a \cdot x^b,\ a = 6.6 \cdot 10^{-15},\ b = 12,9$

$\Rightarrow$ unzureichende Qualität