

Übungen zur Vorlesung  
**Einführung in die angewandte  
Bioinformatik**

Sommersemester 2009

**Übungsblatt 3**

Bearbeitungszeit:  
30.04.2009

Sie finden die Links zu den für diese Aufgaben nötigen Datenbanken auf der **Übungs-Webseite**  
<http://bioinfo.wikidot.com/> .

**Aufgabe 3.1 – Hirsch-Index**

Geben Sie in eigenen Worten eine kurze Definition des Hirsch-Index (*h*-Index) eines Wissenschaftlers an.

---

---

---

**Aufgabe 3.2 – H1N1**

Auf der NCBI-Homepage steht ein ganz aktueller Link zu den "Newest swine influenza A (H1N1) sequences". Wenn Sie dem Link folgen, sehen Sie eine Liste mit verschiedenen Influenza-Datensätzen. Betrachten Sie das PB2-Protein des "Influenza A virus A/California/07/2009(H1N1)". Wie alt ist der Mensch, dem dieses

Virus entnommen wurde? \_\_\_\_\_

Laden Sie sich die FASTA-Datei herunter (dazu rechts oben auf "Download" klicken). Starten Sie eine Konsole und bestimmen Sie an der Kommandozeile mit geeigneten Befehlen die Länge der kodierenden

DNA-Sequenz. \_\_\_\_\_

**Aufgabe 3.3 – J Abbrev**

In vielen Literaturlisten wird aus Platzgründen statt des vollen Journaltitels nur eine abgekürzte Version angegeben. Die möglichen Abkürzungen sind standardisiert und lassen sich zum Beispiel auf der JournalSeek-Seite nachsehen. Wie lautet die Abkürzung für das Journal *Human Molecular Genetics*? \_\_\_\_\_

Welches Journal steckt hinter der Abkürzung *Int J Biol Sci*? \_\_\_\_\_

**Aufgabe 3.4 – R: Zehn zufällige Zahlen ziehen**

Starten Sie R. Schauen Sie sich die Hilfe zu `rnorm` an und benutzen Sie dann diese Funktion, um zehn normalverteilte Zufallszahlen zu generieren. Die Normalverteilung soll einen Mittelwert von 100 haben und eine Varianz von 3. Speichern Sie die erzeugten Zahlen in einer Variable ab.

---

Bestimmen Sie den Mittelwert und die Varianz der generierten Zahlen. Was fällt Ihnen auf? Wie erklären Sie den Unterschied?

---

Berechnen Sie die Standardabweichung. Wie könnten Sie diese einfach berechnen, wenn Sie die R-Funktion `sd` nicht kennen würden? \_\_\_\_\_

### Aufgabe 3.5 – Korrelationskoeffizient

Geben Sie in R ein:

```
x = seq(1, 10, 0.125)
y = x^3/128
plot(x, y)
```

Was für einen Korrelationskoeffizienten zwischen  $x$  und  $y$  erwarten Sie nach Betrachten des Plots?

---

Berechnen Sie jetzt mit `cor` den Korrelationskoeffizienten zwischen  $x$  und  $y$ . Lagen Sie mit Ihrer Vermutung richtig? Wenn nein, versuchen Sie zu erklären, warum Sie sich geirrt haben.

Setzen Sie jetzt

```
u = log(x)
v = log(y)
```

Was erwarten Sie zu sehen, wenn Sie `plot(u,v)` eingeben? Warum?

---

Welchen Korrelationskoeffizienten erwarten Sie zwischen  $u$  und  $v$ ?

---

Überprüfen Sie Ihre Vermutungen und erklären Sie Ihre Beobachtungen.

### Aufgabe 3.6 – PubMed

Suchen Sie in der PubMed-Datenbank des NCBI nach Veröffentlichungen von Philippe Bastiaens. Wie viele Artikel werden gefunden?

---

Sehen Sie sich den Artikel an, den er 2005 zusammen mit E. Karsenti veröffentlicht hat. In welcher Zeitschrift war dies? \_\_\_\_\_

Finden Sie auf der Ergebnisseite den passenden Link, um sich den Artikel anzusehen. Laden Sie sich dort den Artikel als PDF herunter. Wie lautet das erste Wort der Beschreibung von Figure 2? \_\_\_\_\_ Auf

welcher Seite ist Figure 2 abgedruckt? \_\_\_\_\_ Gehen Sie zurück zu PubMed und nutzen Sie den *Limits*-Reiter, um die Suche nach Artikeln von Herrn Bastiaens auf das Erscheinungsjahr 2008 einzugrenzen.

Wie viele Artikel werden gefunden? \_\_\_\_\_

Deaktivieren Sie die Limits wieder und führen Sie die Suche erneut durch. Nehmen Sie die Einschränkung auf 2008 diesmal durch das *Search Field Tag* [dp] durch. Für Hilfe zu diesen Tags siehe Link auf der Übungs-Webseite.

Suchen Sie nach dem 1953 erschienenen Nature-Artikel von J. D. Watson und F. H. Crick über die Molekularstruktur von DNA. Versuchen Sie dabei, die Ergebnisliste durch Nutzen von Limits oder Search Field Tags auf höchstens zwei Artikel einzuschränken. Wie lauten die Seitenzahlen des Artikels? \_\_\_\_\_