
runs

syntax: `runs(data)` or `[lens, vals] = runs(data)`

purpose: Counts the lengths of all “runs” in a vector. A run is a set of consecutive repeats of a single value. (Runs of length 1, which aren’t really runs, are also included.)

```
examples: >> x = [1 1 1 2 3 3 3 3 2 1 1 1 1 1];
>> runs(x) => ans: 3 1 4 1 5
>> [lens, vals] = runs(x)
    lens: 3 1 4 1 5
    vals: 1 2 3 2 1
```

Note that `lens` is the same in both cases. `vals` tells the value involved in each run.

A baseball team with a .500 record (that is, the team wins half of its games) has just been observed to win 10 games in a row. The press declares that this team is having an “incredible streak” and is hot. A reporter who, once, somewhere, took a probability and statistics course computes that there is only a one in 1024 chance (“astronomically low odds”) of this happening. (His computation is based on the fact that $\frac{1}{2}^{10} = \frac{1}{1024}$.) However, his reasoning is only correct for a fixed series of 10 games, whereas in a baseball season there are many possible series of 10 consecutive games any one of which would qualify as a streak if the team happened to win all 10 games. If a .500 team plays 150 games per season, in what fraction of seasons will the team have a streak of 10 games or more?

```
games = [0 1]; % a .500 record,
          % or we could write [.5 0; .5 1]
z = starttally;
Ntrials = 1000;
for trials = 1:Ntrials
    season = sample(150,games); 150 games
    [runlen,winorlose] = runs(season);
    we consider only the winning runs (==1)
    bestrun = max( runlen( find( winorlose == 1) ));
    tally bestrun z;
end
count(z>=10)/Ntrials
```

When we carried out this simulation, we find that in 5% of seasons, the .500 team had a streak of 10 games or more, and in a quarter of all seasons there was a streak of 8 games or more.

This document is an excerpt from
Resampling Stats in MATLAB
Daniel T. Kaplan
Copyright (c) 1999 by Daniel T. Kaplan, All Rights Reserved
This document differs from the published book in pagination and in the omission (unintentional, but unavoidable for technical reasons) of figures and cross-references from the book. It is provided as a courtesy to those who wish to examine the book, but not intended as a replacement for the published book, which is available from
Resampling Stats, Inc.
www.resample.com
703-522-2713

