## Exercise Solutions

Solution 18-2
URN 36\#1 36\#0 pit
URN 77\#1 52\#0 chi
REPEAT 1000
SAMPLE 72 pit pit\$
SAMPLE 129 chi chi\$
MEAN pit\$ $p$
MEAN chi\$ c
SUBTRACT pcd
SCORE d scrboard

## END

HISTOGRAM scrboard

## PERCE NTI LE scrboard (2.5 97.5) interval

PRI NT interval


Results:
INTERVAL $=-0.259210 .039083$ [estimated 95 percent confidence interval]

Solution 21-1
REPEAT 1000
GENERATE 200 1,100 a
COUNT $\mathrm{a}<=\mathbf{7}$ b
DIVIDE b 200 c
SCORE c scrboard
END
HISTOGRAM scrboard

## PERCENTILE z (2.5 97.5) interval

PRINT interval


Result:
INTERVAL $=0.0350 .105 \quad$ [estimated 95 percent confidence interval]

## Solution 21-2

We use the "bootstrap" technique of drawing many bootstrap re-samples with replacement from the original sample, and observing how the re-sample means are distributed.

NUMBERS (30 32312831292924303128283231242331 2727 31) a

## REPEAT 1000

Do 1000 trials or simulations

## SAMPLE 20 ab

Draw 20 lifetimes from a, randomly and with replacement

## MEAN bc

Find the average lifetime of the 20

## SCORE c scrboard

Keep score

## END

## HISTOGRAM scrboard

Graph the experiment results

## PERCENTILE scrboard ( 2.5 97.5) interval

Identify the 2.5 th and 97.5 th percentiles. These percentiles will enclose 95 percent of the resample means.

## PRI NT interval



Result:
INTERVAL $=27.730 .05$ [estimated 95 percent confidence interval]

Solution 21-3
NUMBERS (.02 .026 .023 .017 .022 .019 .018 . 018 . 017 .022) a REPEAT 1000

SAMPLE 10 ab
MEAN bc
SCORE c scrboard

## END

HISTOGRAM scrboard
PERCE NTILE scrboard (2.5 97.5) interval
PRINT interval


Result:
INTERVAL $=0.01870 .0219$ [estimated 95 percent confidence interval]

## Solution 23-1

1. Create two groups of paper cards: 25 with participation rates, and 25 with the spread values. Arrange the cards in pairs in accordance with the table, and compute the correlation coefficient between the shuffled participation and spread variables.
2. Shuffle one of the sets, say that with participation, and compute correlation between shuffled participation and spread.
3. Repeat step 2 many, say 1000, times. Compute the proportion of the trials in which correlation was at least as negative as that for the original data.

DATA (67.5 65.665 .759 .339 .876 .173 .681 .675 .585 .080 .3
54.579 .194 .080 .389 .644 .782 .789 .783 .684 .976 .374 .7
68.8 79.3) partic1

DATA (13 19181220511235654813181322121726 6) spread1

CORR partic1 spread1 corr
compute correlation - it's -. 37

## REPEAT 1000

shuffle the participation rates

# CORR partic2 spread1 corrtria 

compute re-sampled correlation

## SCORE corrtria z

keep the value in the scoreboard

## END

## HISTOGRAM z

## COUNT $z<=-.37$ n

count the trials when result $<=-.37$

## DIVIDE n 1000 prob

compute the proportion of such trials

## PRINT prob

Conclusion: The results of 5 Monte Carlo experiments each of a thousand such simulations are as follows:
prob $=0.028,0.045,0.036,0.04,0.025$.
From this we may conclude that the voter participation rates probably are negatively related to the vote spread in the election. The actual value of the correlation (-.37398) cannot be explained by chance alone. In our Monte Carlo simulation of the null-hypothesis a correlation that negative is found only 3 per-cent-4 percent of the time.

Distribution of the test statistic's value in 1000 independent trials corresponding to the null-hypothesis:


Solution 23-2
NUMBERS (14 200389382231331140515323295 32) homeruns

NUMBERS (135 153120161138175126200205147165124 169156369882 131) strikeout

MULTIPLY homerun strikeout r
SUM rs
REPEAT 1000
SHUFFLE strikeout strikout2
M ULTI PLY strikout2 homeruns c
SUM c cc
SUBTRACT scc d
SCORE d scrboard

## END

HISTOGRAM scrboard
COUNT scrboard $>=s k$
DIVIDE k 1000 kk

## PRINT kk



Result: kk = 0
Interpretation: In 1000 simulations, random shuffling never produced a value as high as observed. Therefore, we conclude that random chance could not be responsible for the observed degree of correlation.

Solution 23-3
NUMBERS (14 200389382231331140515323295 32) homeruns

NUMBERS (135 153120161138175126200205147165124 169156369882 131) strikeou

## CORR homeruns strikeou r

REPEAT 1000
SHUFFLE strikeou strikou2
CORR strikou2 homeruns $\mathbf{r}$ \$
SCORE r\$ scrboard

## END

HISTOGRAM scrboard
COUNT scrboard >=62 k
DIVIDE k 1000 kk
PRINT kk r


Result: $\mathrm{kk}=.001$
Interpretation: A correlation coefficient as high as the observed value (.62) occurred only 1 out of 1000 times by chance. Hence, we rule out chance as an explanation for such a high value of the correlation coefficient.

## Solution 23-4

READ FILE "noreen2.dat" exrate msuppl
read data from file

## CORR exrate msuppl stat

compute correlation stat (it's .419)

## REPEAT 1000

## SHUFFLE msuppl msuppl\$

shuffle money supply values
CORR exrate msuppl\$ stat\$
compute correlation

## SCORE stat\$ scrboard

keep the value in a scoreboard

## END

## PRINT stat

## HISTOGRAM scrboard

COUNT scrboard $>=419 \mathrm{k}$

## DIVIDE k 1000 prob

## PRINT prob

Distribution of the correlation after permutation of the data:


Result: prob = . 001
Interpretation: The observed correlation (.419) between the exchange rate and the money supply is seldom exceeded by random experiments with these data. Thus, the observed result 0.419 cannot be explained by chance alone and we conclude that it is statistically significant.

