

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2006, The Johns Hopkins University and Karl W. Broman. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.

Statistics for laboratory scientists II

Solutions for the homework problems for lecture 2

1. Here is the R code for creating the table and calculating the expected counts under the null hypothesis of $p_i=1/10$ for all i .

```
mydata <- c(98, 99, 100, 89, 107, 114, 100, 112, 85, 96)
n <- sum(mydata)
ex <- n * rep(1/10, 10)
```

- a. Here is the code for calculating the chi-square test statistic and corresponding P-value.

```
chi <- sum( (mydata-ex)^2 / ex)      # value = 7.56
1 - pchisq(chi, 9)                  # P-value = 0.58
```

- b. Here is the code for calculating the LRT statistic and corresponding P-value.

```
lrt <- 2 * sum( mydata * log(mydata/ex) )  # value = 7.58
1 - pchisq(lrt, 9)                        # P-value = 0.58
```

- c. Here is some code for using computer simulations to estimate P-values.

```
n.sim <- 1000
results <- matrix(ncol=2, nrow=n.sim)
for(i in 1:n.sim) {
  # simulate data
  simdat <- rmultinom(1, n, rep(0.1,10))

  # calculate statistics
  chi.sim <- sum( (simdat - ex)^2 / ex)
  lrt.sim <- 2 * sum( simdat * log(simdat/ex) )

  results[i,] <- c(chi.sim, lrt.sim)
}

# p-value for chi-square test
mean(results[,1] >= chi)  # p-value = 0.57 (for my simulations)

# p-value for likelihood ratio test (LRT)
mean(results[,2] >= lrt)  # p-value = 0.58 (for my simulations)
```

- d. We *fail to reject* the null hypothesis of equal frequencies. The observed differences in the counts could reasonably be ascribed to chance variation.

2. With the second version of the table, in applying either the chi-square test or likelihood ratio test, we would get identical statistics and P-values. However, these tests look for general differences from equal frequencies, and not the sort of trend in this observed table.

With such data, we would want to use a different statistical test, such as of whether the mean (or median) is equal to 4.5 or not. Of course, one would ideally choose to perform such a test before obtaining the data rather than after having observed such a trend.