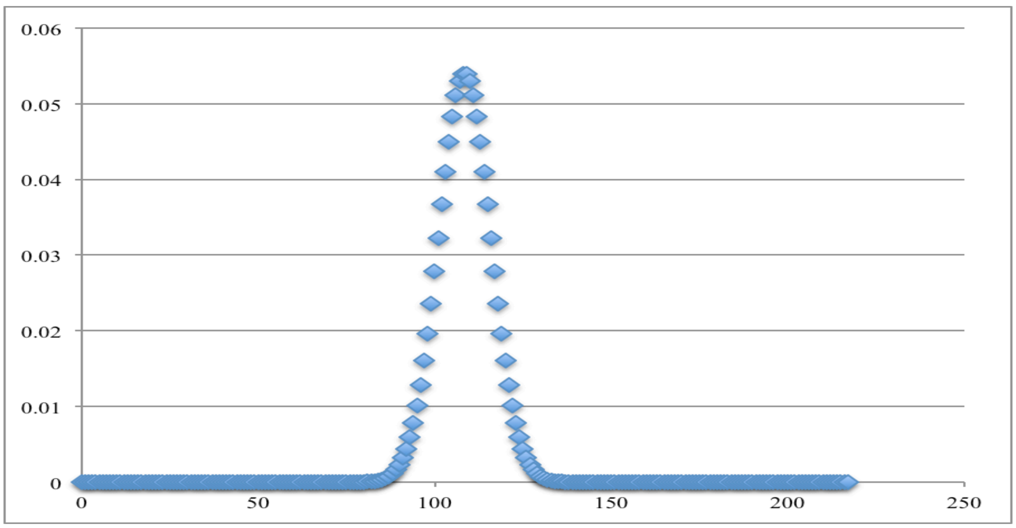
Homework IV
41000 - Business Statistics
Spring 2013
SOLUTION

Problem I: Let us check the claim that the daily closing returns of Apple Inc. (AAPL) and Microsoft Corporation (MSFT) in 2009 are just as likely to be positive or negative. Assume that, day to day, positive returns are all independent Bernoulli(p), where p is the probability of positive return. Of the 217 business days in 2009, AAPL was positive for 124 days, while MSFT was positive for 120 days.

a) Draw the Binomial(217,0.5) distribution. This is the distribution of the number of days out of 217 with positive returns should the probability of positive return for a given day be 0.5. Use the excel function BINOMDIST.



b) Locate the observed data on the graph, i.e. 124 days in AAPL and 120 days in MSFT;

c) Obtain 95% confidence intervals for pA;
Phat=124/217=0.571 and e=2*sqrt(0.571*0.429/217)=0.0672, so the limits of the 95% C.I. for pA are 0.5038 and 0.6382.

d) Obtain 95% confidence intervals for pM;
Phat=120/217=0.553 and e=2*sqrt(0.553*0.447/217)=0.0675, so the limits of the 95% C.I. for pM are 0.4855 and 0.6205.

e) Test the hypothesis H0: pA = 0.5 for AAPL at the 5% level;
The test statistic is (0.571-0.5)/sqrt(0.5*0.5/217)=2.09. Since it is greater than 2 we would reject the hypothesis.

f) Test the hypothesis H0: pM = 0.5 for MSFT at the 5% level;
The test statistic is (0.553-0.5)/sqrt(0.5*0.5/217)=1.56. Since it is not greater than 2 we would fail to reject the hypothesis.

g) Compute the P-value for e);
The p-value equals 2*Pr(Z>2.09) = 0.037

h) Compute the P-value for f).
The p-value equals 2*Pr(Z>1.56) = 0.119

Problem II: Let x1,…,x1000 be sample A and y1,…,y1000 be sample B (see Figure 2 below).

a) Two statisticians, Shapiro and Wilk, devised in 1965 a statistical procedure to test normality of a data set. I implemented the Shapiro-Wilk's normality test for both samples, i.e. I tested H0: "data is normal". The P-values are equal to 0.0000000000426 and 0.2917, respectively. What can we say about the assumption of normality for the data sets?

Regarding X, since the pvalue is extremely small, we would be highly confident when rejecting the null hypothesis that the data come from a normal distribution. Regarding Y, since the pvalue is quite large, we would not be willing (or confident) to reject the null, so we fail to reject it.

b) Two other statisticians, Durbin and Watson, devised in 1950 a statistical procedure to test serial correlation of a data set. If there is serial correlation then today's data is similar to yesterday's data and so forth. I implemented the Durbin-Watson's serial correlation test for both samples, i.e. we tested H0: "data is not serially correlated". The P-values are to 0.08906 and 0.4622, respectively. What can we say about the assumption of no serial correlation for the data sets?

Based on the p-values, we reject the null for dataset X and fail to reject the null for dataset Y.

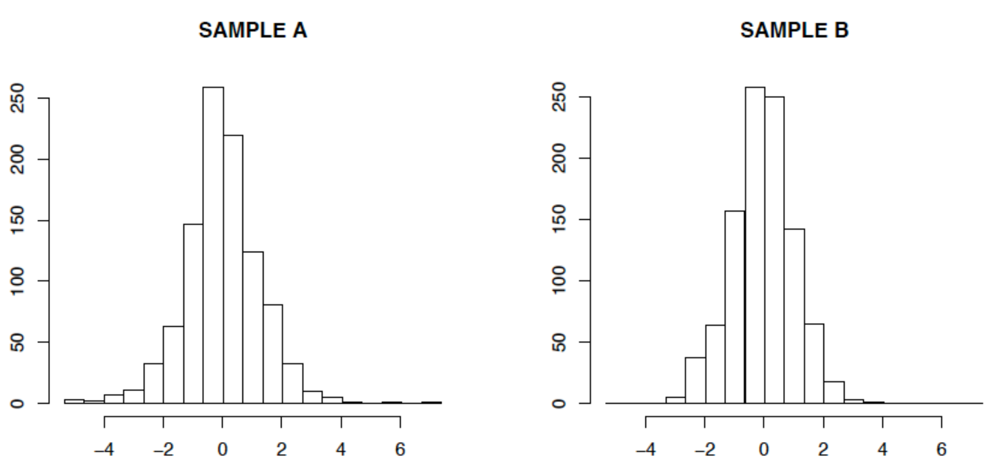


Figure 2: Sample A and sample B.