Lesson 07: Sampling and Sampling Distributions

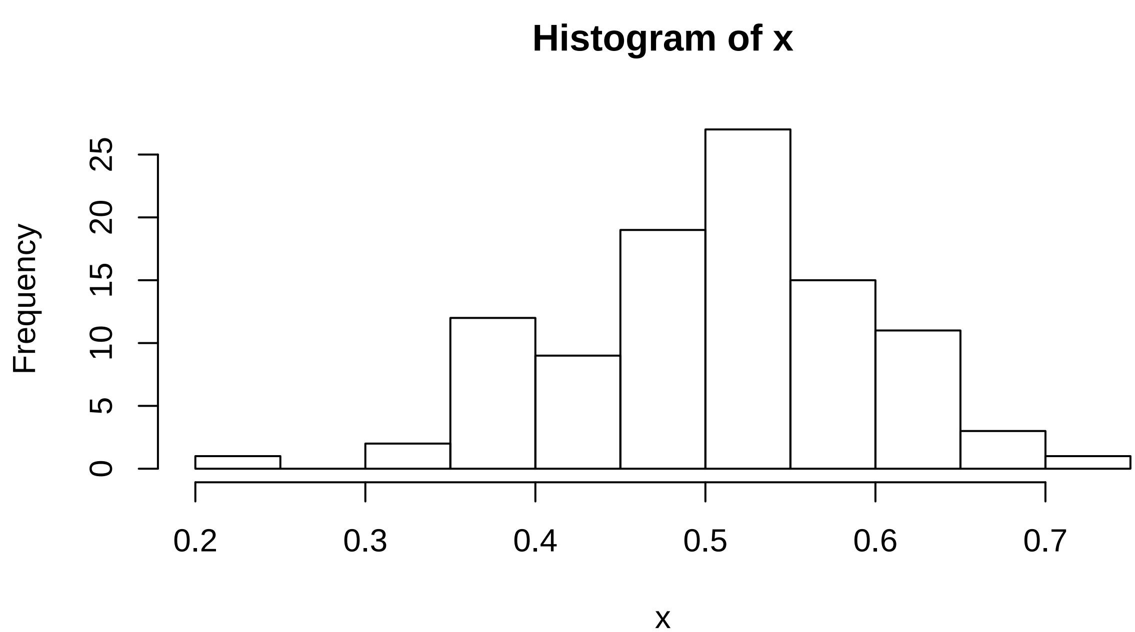
References

- Black, Chapter 7 Sampling and Sampling Distributions (pp. 224-254)
- Davies, Chapter 17 Sampling Distributions and Confidence (pp. 367-377)

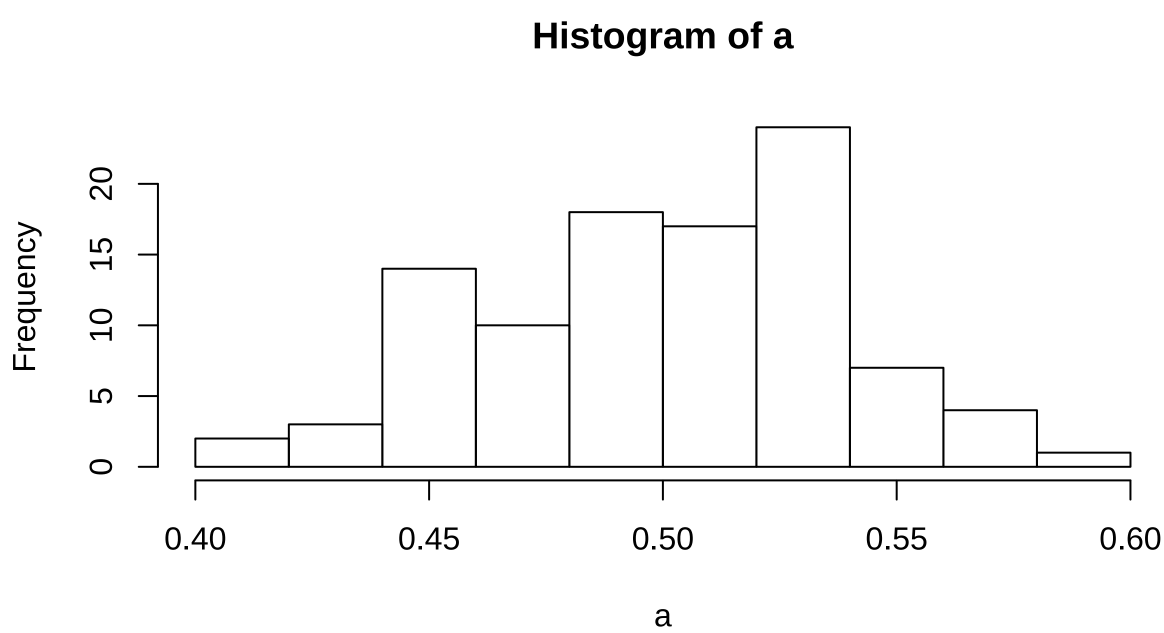
Exercises:

1) Use the uniform distribution over the interval 0 to 1. Draw 100 random samples of size 10. Calculate the means for each sample. Using the 100 mean values plot a histogram. Repeat with 100 random samples of size 50. Repeat with 100 samples of size 500. Present the three histograms using par(). Calculate the variance of each histogram and compare to the original uniform distribution. What do you conclude?

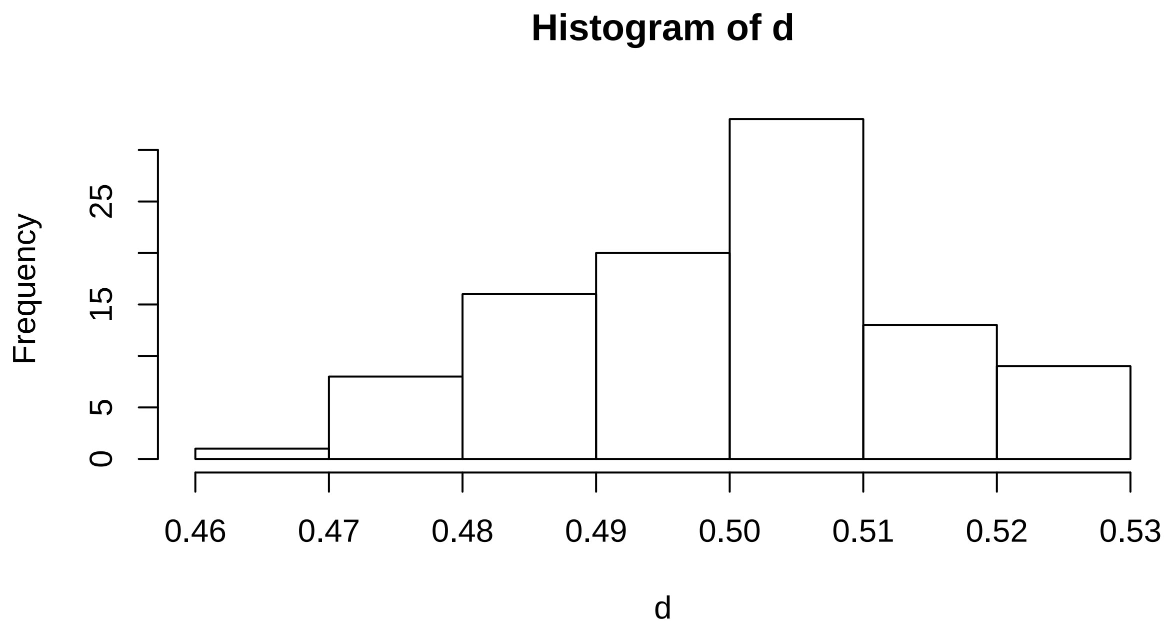
```
set.seed(1234)

x <- c() # creates empty vector
y <- c() # creates empty vector

for (i in 1:100) {
    z <- runif(10)
    x <- append(x, mean(z)) # vector "x" will contain our 100 means
    y <- append(y, var(z))  # vector "y" will contain our 100 variances
}

#plot histogram of means
hist(x)
```



```r
# repeat with 100 samples of size 50
a <- c() # creates empty vector
b <- c() # creates empty vector

for (i in 1:100) {
    c <- runif(50)
    a <- append(a, mean(c))
    b <- append(b, var(c))
}

#plot histogram of means
hist(a)
```

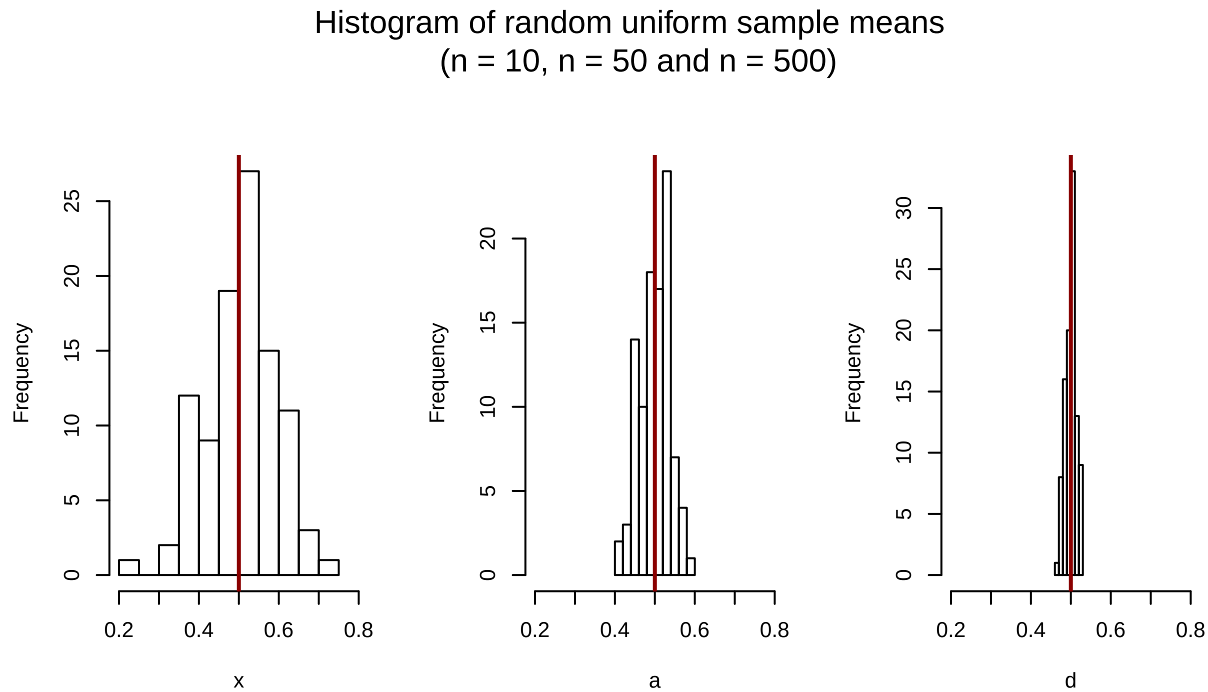



```r
# repeat with 100 samples of size 50
d <- c() # creates empty vector
e <- c() # creates empty vector

for (i in 1:100) {
    f <- runif(500)
    d <- append(d, mean(f))
    e <- append(e, var(f))
}

#plot histogram of means
hist(d)
```



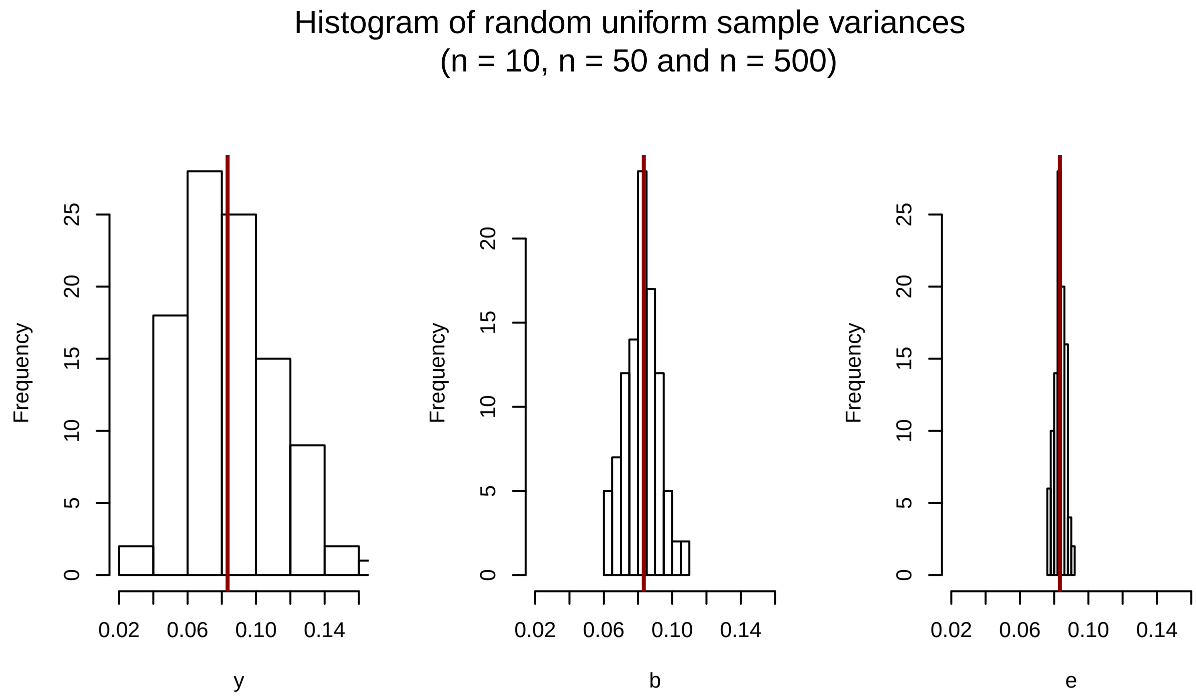

```r
# present the three histograms using par() and mfrow()
par(mfrow=c(1,3), oma=c(0,0,2,0))
hist(x, main = "", xlim = c(0.2,0.8))
abline(v = 0.5, col = "darkred", lwd = 2)
hist(a, main = "", xlim = c(0.2,0.8))
abline(v = 0.5, col = "darkred", lwd = 2)
hist(d, main = "", xlim = c(0.2,0.8))
abline(v = 0.5, col = "darkred", lwd = 2)
mtext("Histogram of random uniform sample means
    (n = 10, n = 50 and n = 500)",side = 3, outer = T, line = -1)
```

```
## Warning in mtext("Histogram of random uniform sample means\n\t(n = 10, n =
## 50 and n = 500)", : font width unknown for character 0x9
```




```
par(mfrow=c(1,1))

# create and present histograms of the sample variances
par(mfrow=c(1,3), oma=c(0,0,2,0))
hist(y, main = "", xlim = c(0.02,0.16))
abline(v = 0.0833333, col = "darkred", lwd = 2)
hist(b, main = "", xlim = c(0.02,0.16))
abline(v = 0.0833333, col = "darkred", lwd = 2)
hist(e, main = "", xlim = c(0.02,0.16))
abline(v = 0.0833333, col = "darkred", lwd = 2)
mtext("Histogram of random uniform sample variances
    (n = 10, n = 50 and n = 500)",side = 3, outer = T, line = -1)
```

```
## Warning in mtext("Histogram of random uniform sample variances\n\t(n = 10,
## n = 50 and n = 500)", : font width unknown for character 0x9
```




```r
par(mfrow=c(1,1))

# NOTE:  abline() for mean and variance histograms equal to "true" values
# for a uniform distribution (0,1).

# mean = a + b / 2 = 0.5
# variance = (b - a)^2 / 12 = 0.08333333
```

2) Using the histogram determined above for samples of size 50, find the quartiles. Using the normal distribution with the true mean and variance for a uniform distribution over the interval 0 to 1, determine the theoretical quartiles for a sample mean from 50 observations. Compare the two sets of quartiles. What do you conclude?

```r
j <- c()

for (i in 1:100) {
    k <- rnorm(50, 0.5, sqrt(0.08333333))
    j <- append(j, mean(k))
}

quantile(a)
```

```
##        0%       25%       50%       75%      100%
## 0.4021239 0.4674217 0.5029306 0.5290173 0.5946016
```

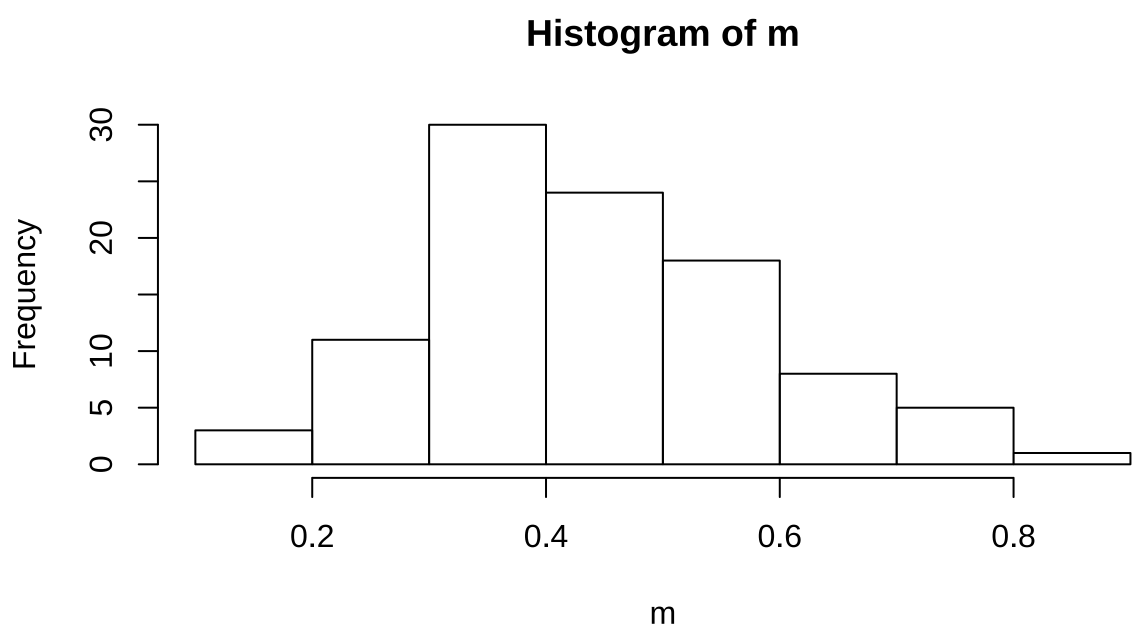
```r
quantile(j)
```

```
##        0%       25%       50%       75%      100%
## 0.4158362 0.4810670 0.5022179 0.5252440 0.6535995
```

```
# The two sets of quartiles are very similar, likely to converge as sample sizes
# are increased.
```

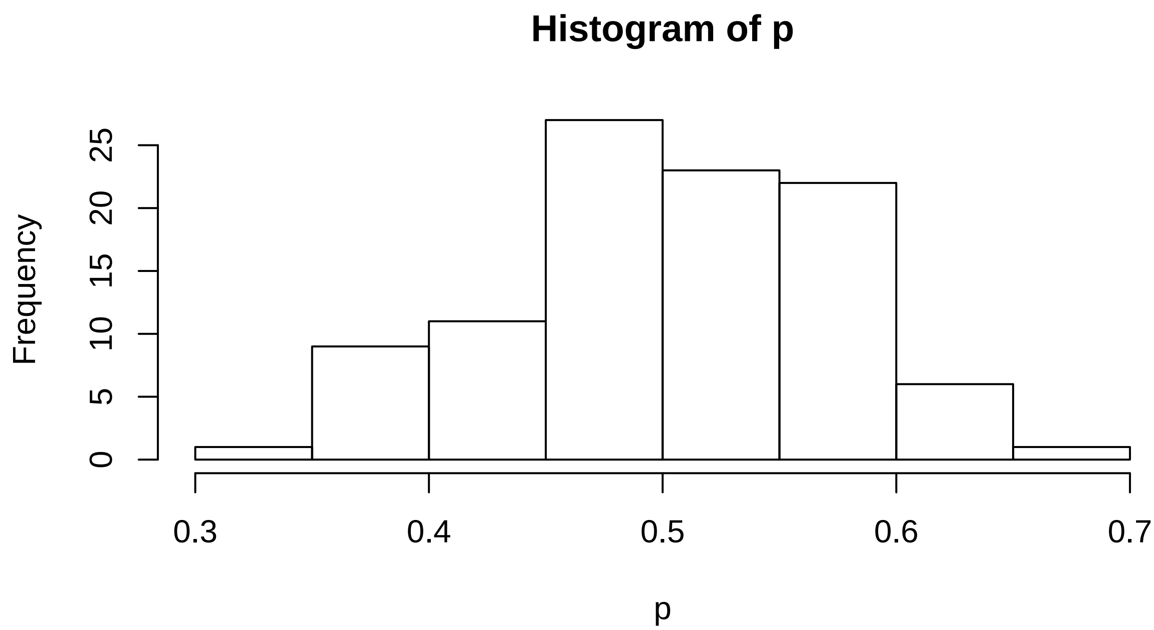
3) Use the binomial distribution with p = 0.5. Draw 100 random samples of size 10. Calculate the means for each sample. Using the 100 mean values plot a histogram. Repeat with 100 random samples of size 50. Repeat with 100 samples of size 500. Present the three histograms using par(). Calculate the variance of each histogram and compare to the original mean and variance for the binomial. What do you conclude?

```
m <- c()
n <- c()

for (i in 1:100) {
    o <- rbinom(10, 1, p = 0.5)
    m <- append(m, mean(o))
    n <- append(n, var(o))
}

# plot histogram of means
hist(m)
```

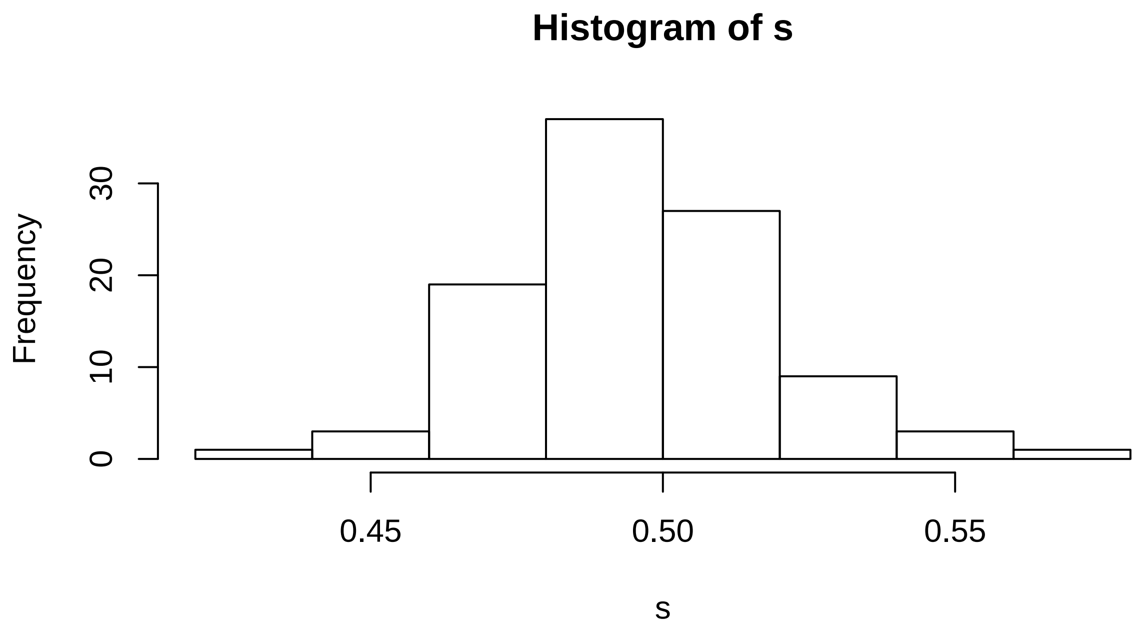



```
# repeat with 100 samples of size 50
p <- c()
q <- c()

for (i in 1:100) {
    r <- rbinom(50, 1, p = 0.5)
    p <- append(p, mean(r))
    q <- append(q, var(r))
```

```
}

hist(p)
```

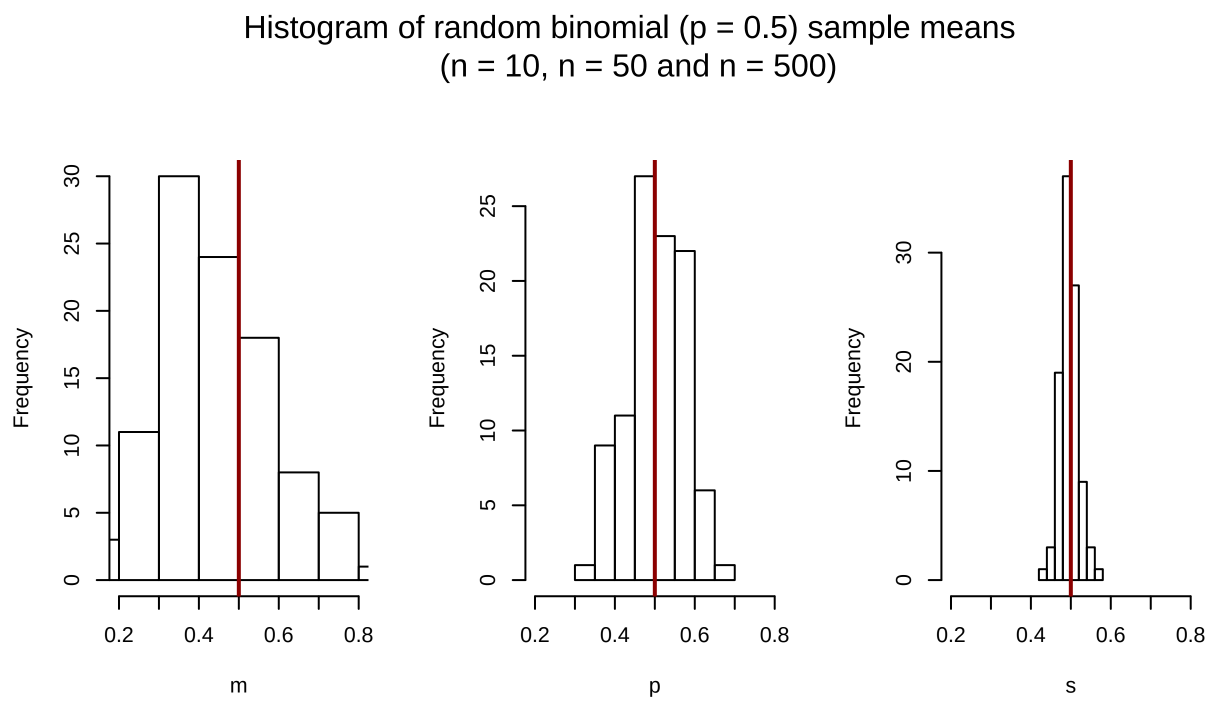



```
# repeat with 100 samples of size 500
s <- c()
t <- c()

for (i in 1:100) {
    u <- rbinom(500, 1, p = 0.5)
    s <- append(s, mean(u))
    t <- append(t, var(u))
}

hist(s)
```



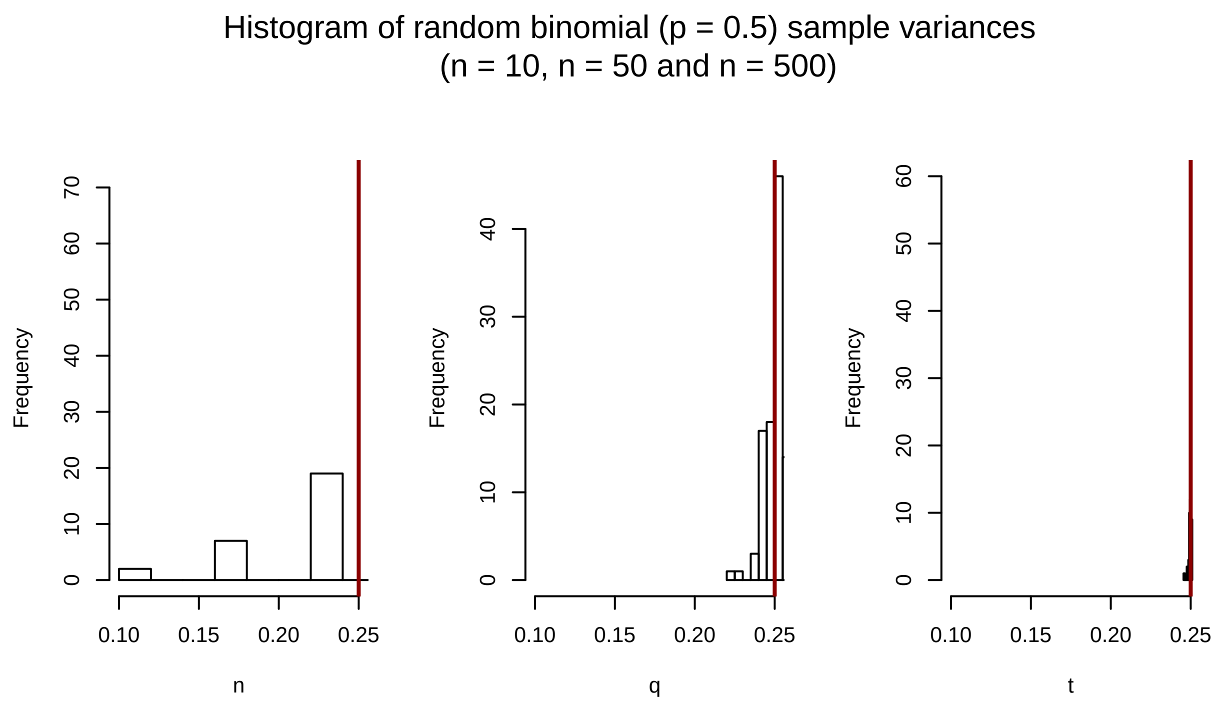

```r
# present the three histograms using par() and mfrow()
par(mfrow=c(1,3), oma=c(0,0,2,0))
hist(m, main = "", xlim = c(0.2,0.8))
abline(v = 0.5, col = "darkred", lwd = 2)
hist(p, main = "", xlim = c(0.2,0.8))
abline(v = 0.5, col = "darkred", lwd = 2)
hist(s, main = "", xlim = c(0.2,0.8))
abline(v = 0.5, col = "darkred", lwd = 2)
mtext("Histogram of random binomial (p = 0.5) sample means
    (n = 10, n = 50 and n = 500)",side = 3, outer = T, line = -1)
```

```
## Warning in mtext("Histogram of random binomial (p = 0.5) sample means\n\t(n
## = 10, n = 50 and n = 500)", : font width unknown for character 0x9
```




```
par(mfrow=c(1,1))

# create and present histograms of the sample variances
par(mfrow=c(1,3), oma=c(0,0,2,0))
hist(n, main = "", xlim = c(0.10,0.25))
abline(v = 0.25, col = "darkred", lwd = 2)
hist(q, main = "", xlim = c(0.10,0.25))
abline(v = 0.25, col = "darkred", lwd = 2)
hist(t, main = "", xlim = c(0.10,0.25))
abline(v = 0.25, col = "darkred", lwd = 2)
mtext("Histogram of random binomial (p = 0.5) sample variances
    (n = 10, n = 50 and n = 500)",side = 3, outer = T, line = -1)
```

```
## Warning in mtext("Histogram of random binomial (p = 0.5) sample variances\n
## \t(n = 10, n = 50 and n = 500)", : font width unknown for character 0x9
```




```r
par(mfrow=c(1,1))
```

4) Using the histogram determined above for samples of size 50, find the quartiles. Using the normal distribution with the true mean and variance for a binomial distribution with p = 0.5, determine the theoretical quartiles for a sample mean from 50 observations. Compare the two sets of quartiles. What do you conclude?

```r
v <- c()

for (i in 1:100) {
    w <- rnorm(50, 0.5, sqrt(0.25))
    v <- append(v, mean(w))
}

quantile(p)
```

```
##   0%  25%  50%  75% 100%
## 0.32 0.46 0.52 0.56 0.66
```

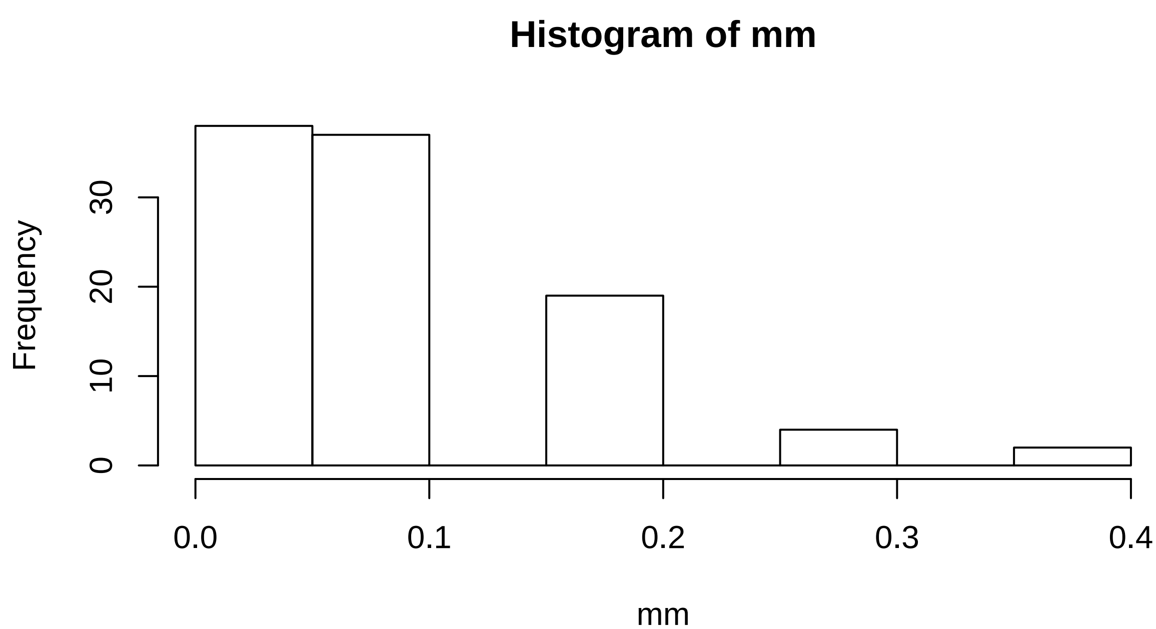
```r
quantile(v)
```

```
##        0%       25%       50%       75%      100%
## 0.3328496 0.4511897 0.4955020 0.5505300 0.6580152
```

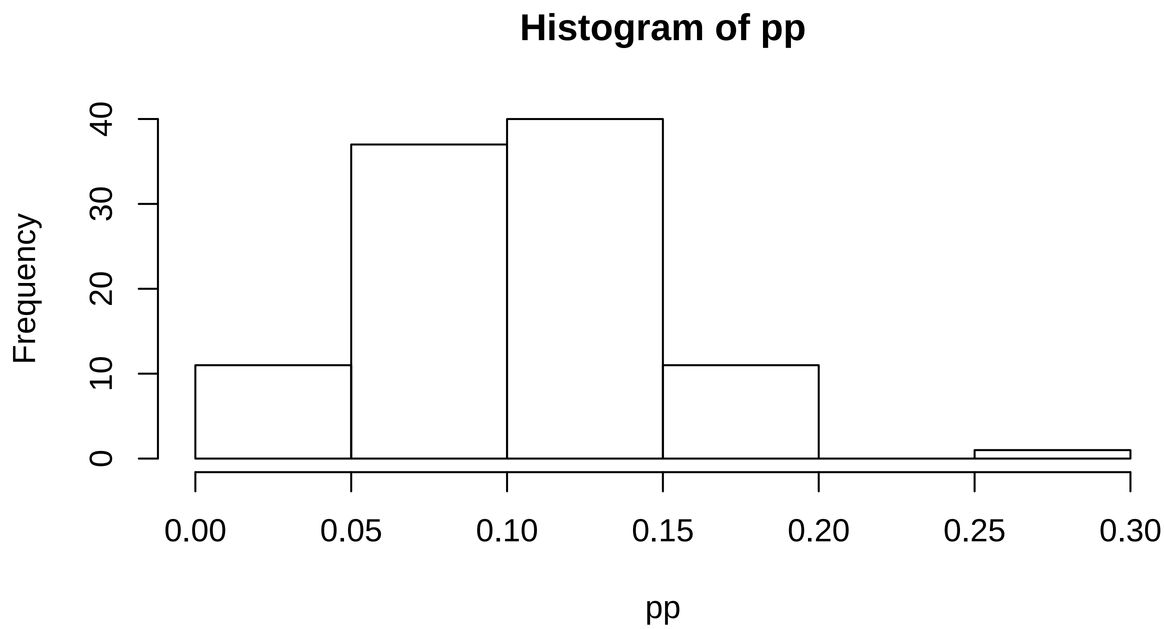
```r
# The two sets of quartiles are very similar, likely to converge as sample sizes
# are increased.
```

5) Use the binomial distribution with p = 0.1. Draw 100 random samples of size 10. Calculate the means for each sample. Using the 100 mean values plot a histogram. Repeat with 100 random samples of size 50. Repeat with 100 samples of size 500. Present the three histograms using par(). Calculate the variance of each histogram and compare to the original mean and variance for the binomial. What do you conclude?

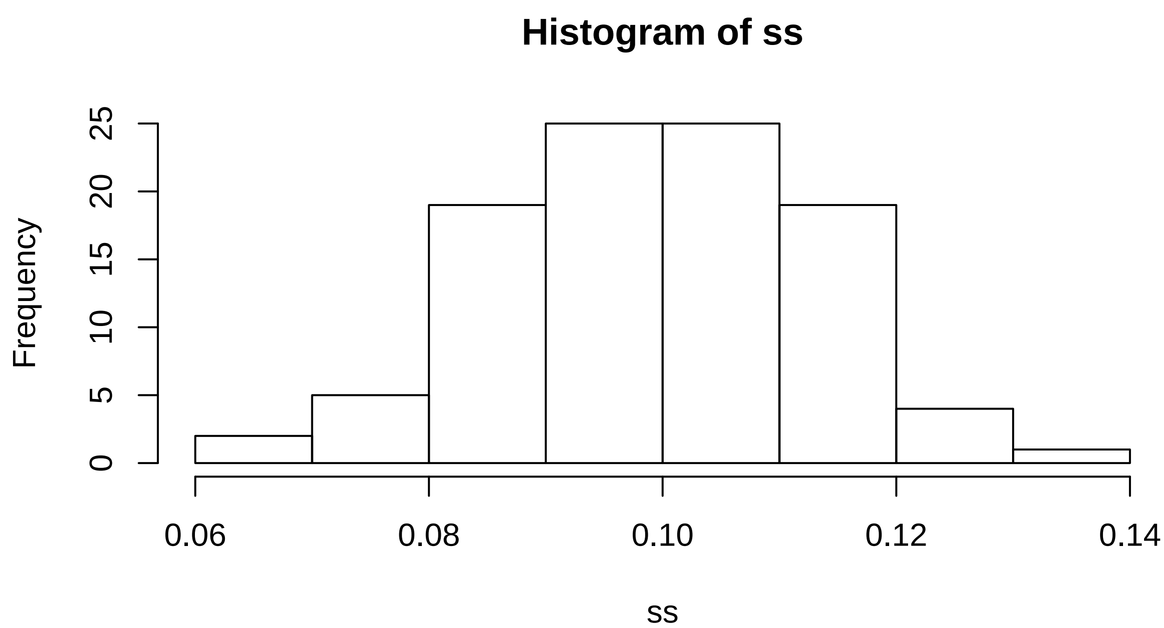
```r
mm <- c()
nn <- c()

for (i in 1:100) {
    oo <- rbinom(10, 1, p = 0.1)
    mm <- append(mm, mean(oo))
    nn <- append(nn, var(oo))
}

# plot histogram of means
hist(mm)
```




```r
# repeat with 100 samples of size 50
pp <- c()
qq <- c()

for (i in 1:100) {
    rr <- rbinom(50, 1, p = 0.1)
    pp <- append(pp, mean(rr))
    qq <- append(qq, var(rr))
}

hist(pp)
```

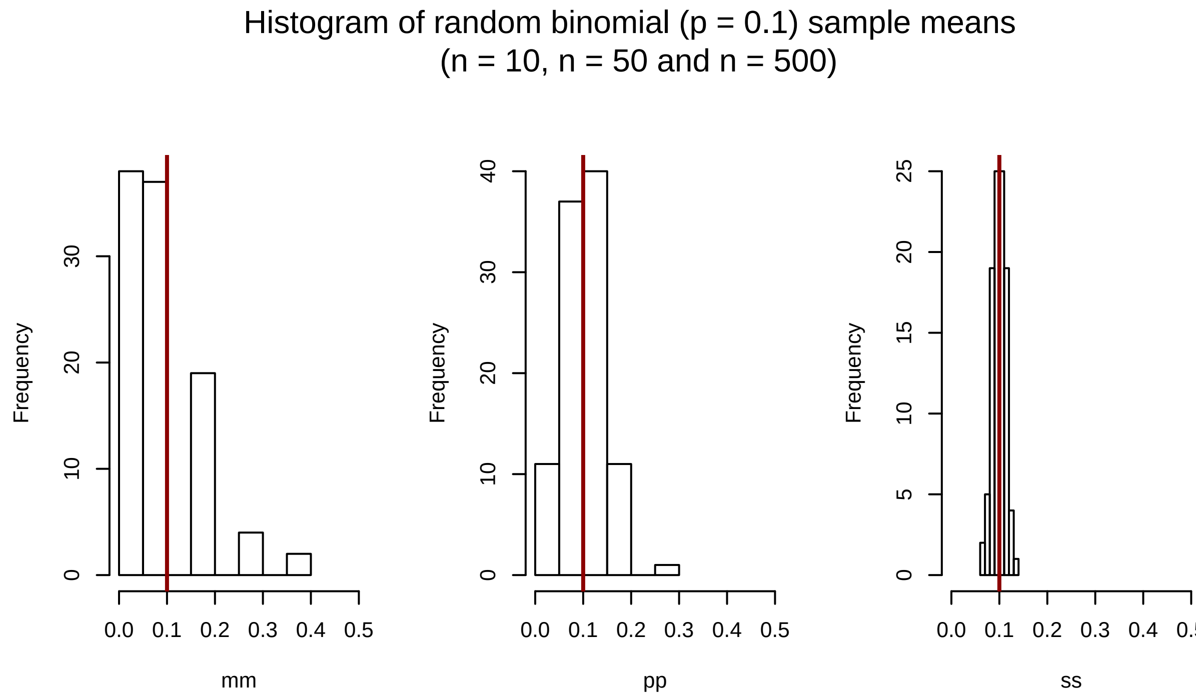



```r
# repeat with 100 samples of size 500
ss <- c()
tt <- c()

for (i in 1:100) {
    uu <- rbinom(500, 1, p = 0.1)
    ss <- append(ss, mean(uu))
    tt <- append(tt, var(uu))
}

hist(ss)
```



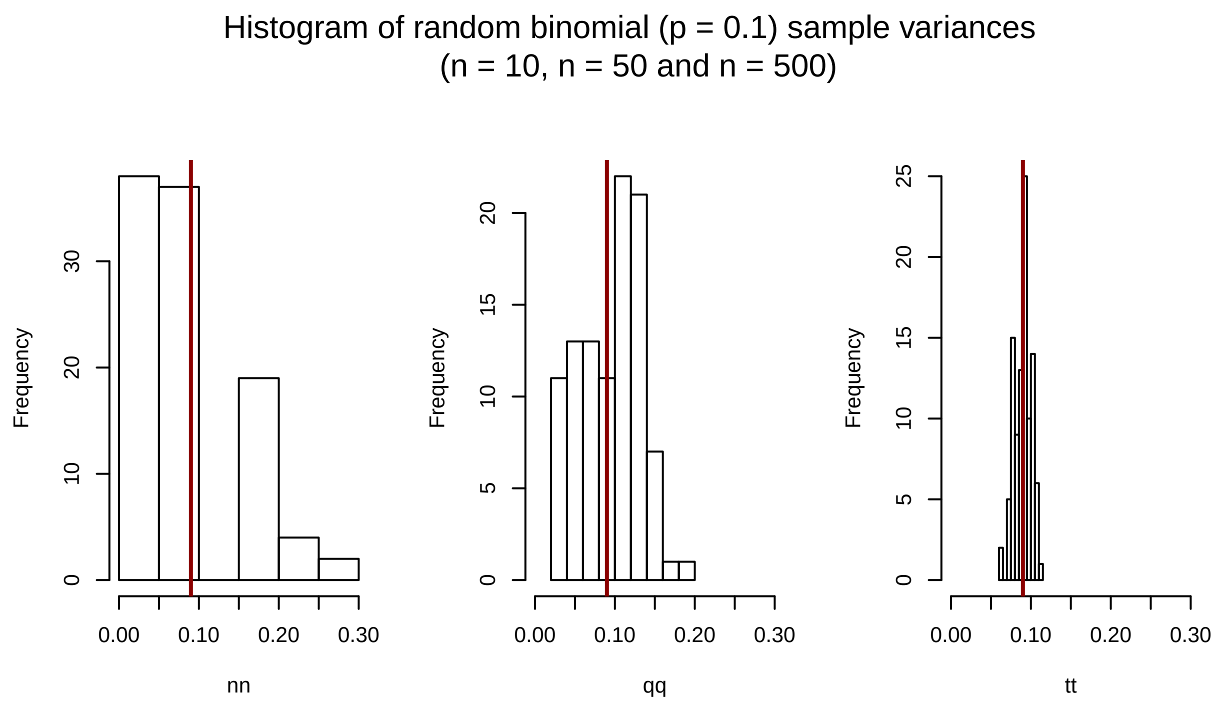

```
# present the three histograms using par() and mfrow()
par(mfrow=c(1,3), oma=c(0,0,2,0))
hist(mm, main = "", xlim = c(0,0.5))
abline(v = 0.1, col = "darkred", lwd = 2)
hist(pp, main = "", xlim = c(0,0.5))
abline(v = 0.1, col = "darkred", lwd = 2)
hist(ss, main = "", xlim = c(0,0.5))
abline(v = 0.1, col = "darkred", lwd = 2)
mtext("Histogram of random binomial (p = 0.1) sample means
    (n = 10, n = 50 and n = 500)",side = 3, outer = T, line = -1)
```

```
## Warning in mtext("Histogram of random binomial (p = 0.1) sample means\n\t(n
## = 10, n = 50 and n = 500)", : font width unknown for character 0x9
```




```
par(mfrow=c(1,1))

# create and present histograms of the sample variances
par(mfrow=c(1,3), oma=c(0,0,2,0))
hist(nn, main = "", xlim = c(0,0.30))
abline(v = 0.09, col = "darkred", lwd = 2)
hist(qq, main = "", xlim = c(0,0.30))
abline(v = 0.09, col = "darkred", lwd = 2)
hist(tt, main = "", xlim = c(0,0.30))
abline(v = 0.09, col = "darkred", lwd = 2)
mtext("Histogram of random binomial (p = 0.1) sample variances
    (n = 10, n = 50 and n = 500)",side = 3, outer = T, line = -1)
```

```
## Warning in mtext("Histogram of random binomial (p = 0.1) sample variances\n
## \t(n = 10, n = 50 and n = 500)", : font width unknown for character 0x9
```




```r
par(mfrow=c(1,1))
```

6) Using the histogram determined above for samples of size 50, find the quartiles. Using the normal distribution with the true mean and variance for a binomial distribution with p = 0.1, determine the theoretical quartiles for a sample mean from 50 observations. Compare the two sets of quartiles. What do you conclude?

```r
vv <- c()

for (i in 1:100) {
    ww <- rnorm(50, 0.1, sqrt(0.09))
    vv <- append(vv, mean(ww))
}

quantile(pp)
```

```
##   0%  25%  50%  75% 100%
## 0.02 0.08 0.12 0.14 0.26
```

```r
quantile(vv)
```

```
##          0%         25%         50%         75%        100%
## -0.02883012  0.08130462  0.10593952  0.13540200  0.19509431
```

```r
# The two sets of quartiles are very similar, likely to converge as sample sizes
# are increased.
```