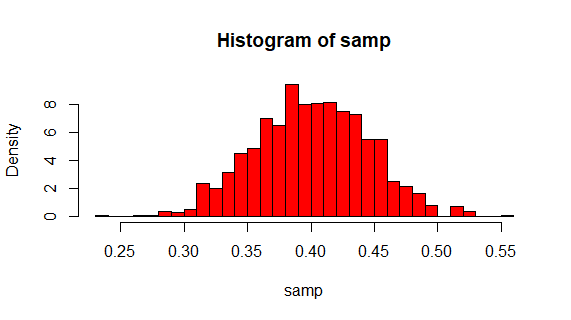
Cameron Chandra

9/20/17

Stats 449

Homework 1

* 1. We know that the expected value of x (or mu) for a Bernoulli random variable is E[X]=p and the variance of X is Var(X)= p(1-p). So, given p = 0.4, E[X] = **0.4** and Var(X) = .4(1-.4) = **0.24.**
  2. The expected value of the sample mean (E[X-bar]) is equal to mu or the expected value of X. So, E[X-bar] = **0.4**. The variance of X-bar is the variance of x divided by n. Because we are not given n, Var[X-bar] = **0.24/n**.
  3. To generate X-bar for each column, we use the R code >colMeans(mat). The mean of the sample is 0.4015833. The standard deviation of the sample is 0.04666546. The central limit theorem states that, given a large sample size, the mean of the sample will approximately equal the mean of the population. Shown by the histogram and the previously given mean (0.40123), this graph supports the CLT as the mean is near 0.4.



Code Used:

> x <- rbinom(96000,1,0.4)

> mat <- matrix(x,nrow=96,ncol=1000)

> samp <- colMeans(mat)

> mean(samp)

> sd(samp)

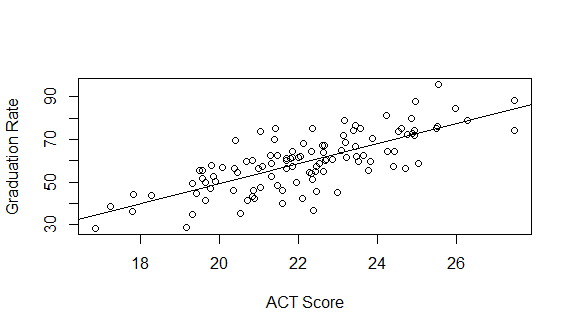
> hist(samp,freq=F,col= “red”, main= “”, breaks= 30

* 1. The probability that a certain machine will produce less than or equal to a defective item in a sample of 100 items is as follows:

P(X<1) = P(X=0) + P(X=1) 🡺 P(X=0) = .99; P(X=1) = 0.01 🡺 P(X<1) = .99 + .01 = **1**

* 1. The Poisson approximation for the same situation is as follows:

The Poisson approximation is close to the probability computed in part a.

* 1.  Fitted Regression Line: Y = -45.6486 + 4.7428 (scatterplot below)

Code used:

> college\_data <- read.csv(“C:/Users/camer/OneDrive/Documents/Stats 449/Homework 1/college.csv”)

> model1 <- lm(gradrat~act, data= college\_data)

> summary(model1)

> plot(college\_data$act,college\_data$gradrat,xlab=”ACT Score”,ylab=”Graduation Rate”)

>abline(model1)

* 1. The 95% confidence interval is as follows:

Code used:

> qt(.95,108)

> confint(model1,level= .95)

* 1. Comparing computed t statistic (qt(.95,108)= 1.659085)) and the t statistic computed from the t-test (11.708), we decide to reject the null hypothesis that Beta = 0.

Code used:

> summary(model1)$coef

> t.test(college\_data$act)

Analysis of Variance Table

Model 1: gradrat ~ act

Model 2: gradrat ~ 1

Res.Df RSS Df Sum of Sq F Pr(>F)

1 108 8443

2 109 19160 -1 -10717 137.09 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

T-statistic = 11.708 🡺 T^2 = 137.077

F-statistic = 137.09

T^2 and F are equal.

Code Used:

> m1 <- lm(gradrat~act, college\_data)

> m2 <- lm(gradrat~1, college\_data)

> anova(m1,m2)

* 1. The correlation coefficient of this model is: 0.7478901

Code used:

> cor(college\_data$gradrat, college\_data$act)

f. The 95% confidence and prediction interval for a university with mean ACT score equal to 23 are as follows:

Confidence interval: (61.64099,65.23251)

Prediction interval: (45.81917,81.05433)