

Statistical Methods: Mid Term

Cameron McIntyre

October 2018

1 Question 1

A researcher finds that, of 982 men who died in 2002, 221 died from some heart disease. Also, of the 982 men, 334 had at least one parent who had some heart disease. Of the latter 334 men, 111 died from some heart disease. A man is selected from the group of 982. Given that neither of his parents had some heart disease, find the conditional probability that this man died of some heart disease.

Answer:

Define Events

H - Died of Heart Disease.

P - Parents Have Heart Disease

$$P(H|\bar{P}) = \frac{P(H \cap \bar{P})}{P(\bar{P})} = \frac{\frac{221-111}{982}}{\frac{982-334}{982}} = .1697$$

2 Question 2

At the beginning of a certain study of a group of persons, 15% were classified as heavy smokers, 30% as light smokers, and 55% as nonsmokers. In the five-year study, it was determined that the death rates of the heavy and light smokers were five and three times that of the nonsmokers, respectively. A randomly selected participant died over the five-year period; calculate the probability that the participant was a nonsmoker.

Answer:

Events:

H - Heavy smoker. $P(H) = .15$

L - Light Smokers. $P(L) = .30$

N - Non Smoker $P(N) = .55$

We also know

$P(D|N) = X$

$P(D|L) = 3X$

$P(D|H) = 5X$

We need to find $P(N|D)$.

$$P(N|D) = \frac{P(D|N) \cdot P(N)}{P(D)} = \frac{X \cdot .55}{\sum_{i \in (N,L,H)} P(D|i)P(i)} = \frac{X \cdot .55}{X \cdot .55 + 3X \cdot .3 + 5X \cdot .15}$$

$$P(N|D) = \frac{X \cdot .55}{X(.55 + 3 \cdot .3 + 5 \cdot .15)} = \frac{.55}{2.2} = .25$$

3 Question 3

An insurance company sells an automobile policy with a deductible of one unit. Let X be the amount of the loss for a policy, with X having a probability density function

$$P_X(x) = .9, \text{ if } x = 0$$

$$P_X(x) = \frac{c}{x}, \text{ if } x = 1, 2, 3, 4, 5, 6$$

(a) Determine c to make the probability density valid:

$$1 = .9 + \sum_1^6 \frac{c}{x} = .9 + c\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}\right) \leftrightarrow .1 = c \cdot 2.45$$

$$c = \frac{.1}{2.45}$$

Therefore our distribution becomes:

$$P_X(x) = .9, \text{ if } x = 0$$

$$P_X(x) = \frac{1}{2.45} \cdot \frac{1}{x}, \text{ if } x = 1, 2, 3, 4, 5, 6$$

(b) Expected value of the amount of insurance a company must pay:

The payout function is 0, $X \in (0, 1)$ and $X - 1$, $X \in (2, 3, 4, 5, 6)$.

$$E[\text{Payout}] = 0 \cdot .9 + 0 \cdot \frac{1}{2.45} \cdot \frac{1}{1} + 1 \cdot \frac{1}{2.45} \cdot \frac{1}{2} + 2 \cdot \frac{1}{2.45} \cdot \frac{1}{3} + 3 \cdot \frac{1}{2.45} \cdot \frac{1}{4} + 4 \cdot \frac{1}{2.45} \cdot \frac{1}{5} + 5 \cdot \frac{1}{2.45} \cdot \frac{1}{6}$$

$$E[\text{Payout}] = 1.44898$$

4 Question 4

Suppose that 2000 points are selected independently and at random from the unit square $\{0 \leq x \leq 1, 0 \leq y \leq 1\}$. Let W equal the number of points that fall into $A = X^2 + Y^2 < 1$

- (a) How must W be distributed?

The points are distributed uniformly over the unit square. The area that we care about is in the unit circle between $0 \leq \theta \leq \frac{\pi}{2}, 0 \leq r \leq 1$. The probability of a single dot being within that area is going to be the ratio of the areas because the dots are uniformly distributed.

$$\text{Area of Unit Square} = 1 * 1 = 1$$

$$\text{Area of Circle in unit square} = \int_0^{\frac{\pi}{2}} \int_0^1 r dr d\theta = \int_0^{\frac{\pi}{2}} \frac{1}{2} d\theta = \frac{\pi}{4}$$

Therefore the probability of being within the area A is $\frac{\pi}{4}$

The probability of a single dot being in the intersection of the unit circle and the unit square is a bernoulli trial with probability $p = \frac{\pi}{4}$. The number of dots within that area can be thought of as a binomial distribution with a successful probability $p = \frac{\pi}{4}$. Because there are 2000 bernoulli trials, each of which are independent and identically distributed. We can use the central Limit Theorem, because the margin of error in convergence is very small for such a large value of N . Also large factorials can become computationally inefficient.

We have estimates for μ and σ .

$$\mu = NP = 2000 * \frac{\pi}{4} = 1570.797$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{2000 * \frac{\pi}{4} * (1 - \frac{\pi}{4})} = 18.36$$

There for $W \sim N(1570.797, 18.36)$

- (b) What is the mean and standard deviation for W ?

As we calculated for the distribution above.

$$\mu = n * p = 2000 * \frac{\pi}{4} = 1570.797$$

$$\sigma = \sqrt{n * p(1-p)} = \sqrt{2000 * p * (1-p)} = 18.36$$

5 Question 5

Let the random variable X be equal to the number of days that it takes a high-risk driver to have an accident. Assume that X has an exponential distribution.

If $P(X < 50) = 0.25$, compute $P(X > 100|X > 50)$.

Answer:

We rely on the Markovian Property (memoryless-ness) of the exponential distribution. In essence for an exponential random variable $P(s + t < X|t < X) = P(s < X)$.

So,

$$P(100 < X|50 < X) = P(50+50 < X|50 < X) = P(50 < X) = 1 - P(X < 50) = 1 - .25 = .75$$

Alternative Solution:

We solve for λ .

$$.25 = 1 - e^{-\lambda 50} \leftrightarrow \lambda = -1 * \frac{\ln(.75)}{50} \leftrightarrow \lambda = 0.005753641$$

$$P(X > 100|X > 50) = \frac{P(X > 100) \cap P(X > 50)}{P(X > 50)} = \frac{e^{-\lambda * 100}}{e^{-\lambda * 50}} = e^{-\lambda * 50} = .75$$

6 Question 6

A painting contractor was contracted to paint 25 housing units. However, to cut corners the contractor watered down the paint and failed to give a second coat in 10 of the units. A housing unit inspector sampled 5 units at random and would be able to tell if the painting job was not according to specifications. Assume the business model is to approve the work if there is no more than 20% found on inspection to be subpar. What is the probability that the contractor's work is approved?

Answer: This Inspectors selection of a sub par house is distributed hypergeometrically. If we define the random variable H to be the number of houses selected that fail the inspection, we are interested in $P(H \leq 1)$.

We are interested in this quantity because 20% of the five houses selected is 1. Therefore at maximum he can select 1 faulty house out of the maximum of 5 houses he selects.

$$P(H \leq 1) = P(H = 0) + P(H = 1) = \frac{\binom{5}{0}\binom{20}{5}}{\binom{25}{5}} + \frac{\binom{5}{1}\binom{20}{4}}{\binom{25}{5}} = .313439$$

7 Question 7

Two admirers of a new worker in the office will ask the new worker out according to a pdf of $\frac{4}{x^5}$, $1 < x < \infty$ where X is measured in weeks. If the admirers act independently, determine

- (a) The pdf for the maximum wait to be asked out by both and
- (b) The expected time to be asked out by both.

Answer:

- (a) We know from the text $f_{Ymax}(y) = n(F(y))^{n-1}f_Y y$.
So,

$$F_X(x) = \int_1^x \frac{4}{x^5} dx = -\frac{1}{x^4} \Big|_1^x = 1 - \frac{1}{x^4}$$

Therefore,

$$f_{Xmax}(x) = 2(1 - \frac{1}{x^4})^1 \frac{4}{x^5} = \frac{8}{x^5} - \frac{8}{x^9} = 8(\frac{1}{x^5} - \frac{1}{x^9})$$

- (b) The expected value is calculated below:

$$E[X] = \int_1^\infty x f_{Xmax}(x) = 8 \int_1^\infty \frac{1}{x^4} - \frac{1}{x^8} = -8 \frac{7x^4 - 3}{21x^7} \Big|_1^\infty = -8(0 - \frac{4}{21}) = \frac{32}{21}$$

8 Question 8

Consider a fair four-sided die numbered 1 – 4 and a fair six-sided die numbered 1 – 6, where X is the number appearing on the four-sided die and Y is the number appearing on the six-sided die. Define $W = X + Y$ when they are rolled together. Assuming X and Y are independent, Use the moment generating function technique to find the expectation and variance.

- (a) Find the moment generating function for W . First we find

$$M_X(t) = \sum_{i=1}^4 \frac{e^{ti}}{4} = \frac{1}{4}(e^{1t} + e^{2t} + e^{3t} + e^{4t})$$

Similarly,

$$M_Y(t) = \sum_{i=1}^6 \frac{e^{ti}}{6} = \frac{1}{6}(e^{1t} + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t})$$

We use a result shown in the textbook that when X and Y are independent and $W = X + Y$ then, $M_W(t) = M_X(t) \cdot M_Y(t)$.

So,

$$M_W(t) = \frac{1}{4}(e^{1t} + e^{2t} + e^{3t} + e^{4t}) \cdot \frac{1}{6}(e^{1t} + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t})$$

$$M_W(t) = \frac{1}{24}(e^{1t} + e^{2t} + e^{3t} + e^{4t}) \cdot (e^{1t} + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t})$$

(b) Find the expectation $E(W)$.

$$E(W) = M'_w(t) \Big|_{t=0}$$

$$E(W) = \frac{1}{24}(1e^{1t} + 2e^{2t} + 3e^{3t} + 4e^{4t}) \cdot (e^{1t} + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t}) \\ + \frac{1}{24}(e^{1t} + e^{2t} + e^{3t} + e^{4t}) \cdot (1e^{1t} + 2e^{2t} + 3e^{3t} + 4e^{4t} + 5e^{5t} + 6e^{6t}) \Big|_{t=0}$$

$$E(W) = \frac{1}{24}(1 + 2 + 3 + 4) \cdot (6) + \frac{1}{24}(4 \cdot (1 + 2 + 3 + 4 + 5 + 6))$$

$$E(W) = \frac{10}{4} \frac{6}{6} + \frac{21}{6} \frac{4}{4} = \frac{72}{12} = 6$$

(c) Find the variance $\text{Var}(W)$.

Firstly, we find the second derivative of the moment generating function.

$$\frac{d}{dt} M'_w(t) = \frac{1}{24} \left[\frac{d}{dt} (1e^{1t} + 2e^{2t} + 3e^{3t} + 4e^{4t}) \cdot (e^{1t} + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t}) \right. \\ \left. + (e^{1t} + e^{2t} + e^{3t} + e^{4t}) \cdot (1e^{1t} + 2e^{2t} + 3e^{3t} + 4e^{4t} + 5e^{5t} + 6e^{6t}) \right] \\ = \frac{1}{24} \left[(1e^{1t} + 4e^{2t} + 9e^{3t} + 16e^{4t}) \cdot (e^{1t} + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t}) \right. \\ + (1e^{1t} + 2e^{2t} + 3e^{3t} + 4e^{4t}) \cdot (1e^{1t} + 2e^{2t} + 3e^{3t} + 4e^{4t} + 5e^{5t} + 6e^{6t}) \\ + (1e^{1t} + 2e^{2t} + 3e^{3t} + 4e^{4t}) \cdot (1e^{1t} + 2e^{2t} + 3e^{3t} + 4e^{4t} + 5e^{5t} + 6e^{6t}) \\ \left. + (e^{1t} + e^{2t} + e^{3t} + e^{4t}) \cdot (1e^{1t} + 4e^{2t} + 9e^{3t} + 16e^{4t} + 25e^{5t} + 36e^{6t}) \right] \\ E[W^2] = M''_w(t) \Big|_{t=0} = \frac{1}{24} [30 * 6 + 10 * 21 + 10 * 21 + 4 * 91] = \frac{964}{24}$$

To find the variance, we use the variance formula $E[X^2] - E[X]^2 = \text{Var}[X]$

$$\text{Var}[W] = E[W^2] - (E[W])^2 = \frac{964}{24} - \frac{864}{24} = \frac{100}{24} = 4.16$$

9 Question 9

Let $f_{X,Y}(x,y) = \frac{4}{3}$, $0 < x < 1, x^3 < y < 1$.

- (a) Determine the marginal density $f_X(x)$.
- (b) Determine the marginal density $f_Y(y)$.
- (c) Determine the conditional density density $f_{X|Y}(x|y)$.

Answer:

(a)

$$f_X(x) = \int_{x^3}^1 \frac{4}{3} dy = \frac{4}{3} y \Big|_{x^3}^1 = \frac{4}{3}(1 - x^3), 0 < x < 1$$

(b)

$$f_Y(y) = \int_0^{y^{\frac{1}{3}}} \frac{4}{3} dx = \frac{4}{3} x \Big|_0^{y^{\frac{1}{3}}} = \frac{4}{3} y^{\frac{1}{3}}, 0 < y < 1$$

(c) Note, by definition:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{4}{3}}{\frac{4}{3}y^{\frac{1}{3}}} = \frac{1}{y^{\frac{1}{3}}} \text{ We need to add boundary conditions}$$

So since $x > 0$,

$$x^3 < y < 1 \leftrightarrow x < y^{\frac{1}{3}} < 1$$

$$f_{X|Y}(x|y) = \begin{cases} \frac{\frac{4}{3}}{\frac{4}{3}y^{\frac{1}{3}}} = \frac{1}{y^{\frac{1}{3}}} & x < y^{\frac{1}{3}} \\ 0 & x \geq y^{\frac{1}{3}} \end{cases}$$

10 Question 10

An urn contains b black balls and r red balls. One of the balls is drawn at random, but when it is put back in the urn c additional balls of the same color are put in with it. Now suppose that we draw another ball. Show that the probability that the first ball drawn was black given the second ball drawn was red is $\frac{b}{b+r+c}$.

Answer:

Define X_1 to be the first ball drawn. Define X_2 to be the second ball drawn.

$$P(X_1 = b | X_2 = r) = \frac{P(X_1 = b \cap X_2 = r)}{P(X_2 = r)}$$

$$= \frac{P(X_1 = b) * P(X_2 = r | X_1 = b)}{P(X_2 = r | x_1 = b)P(X_1 = b) + P(X_2 = r | x_1 = r)P(X_1 = r)}$$

$$\begin{aligned}
&= \frac{\frac{b}{b+r} \cdot \frac{r}{b+r+c}}{\frac{r}{b+r+c} \cdot \frac{b}{b+r} + \frac{r+c}{b+r+c} \cdot \frac{r}{b+r}} \\
&= \frac{\frac{b}{b+r} \cdot \frac{r}{b+r+c}}{\frac{r \cdot (b+r+c)}{(b+r+c) \cdot (b+r)}} \\
&= \frac{b}{b+r} \cdot \frac{r}{b+r+c} \cdot \frac{(b+r+c)(b+r)}{r(b+r+c)} \\
&= \frac{r}{b+r+c}
\end{aligned}$$

The result is as desired.

11 Question 11

During a visit to a primary care physician's office, the probability of having neither lab work nor referral to a specialist is 0.21. Of those coming to that office, the probability of having lab work is 0.41 and the probability of having a referral is 0.53. What is the probability of having both lab work and a referral?

Answer:

Define:

L is the event of having Lab work ordered.

R is the event of having a referral to a specialist.

Find $P(L \cap R)$

$$P(L \cup R)^c = .21 \leftrightarrow P(L \cup R) = .79$$

$$P(L \cup R) = P(L) + P(R) - P(L \cap R)$$

$$P(L \cap R) = P(A) + P(R) - P(L \cup R)$$

$$P(L \cap R) = .41 + .53 - .79$$

$$P(L \cap R) = .15$$

12 Question 12

An elliptical jungle clearing has been identified as a military target for a bombing run. Let the equation for the ellipse (in meters) be $\frac{X^2}{400^2} + \frac{Y^2}{200^2} = 1$. Assume the approach of the bomber will be along the major axis, and the target would be the point (0,0). Assume also that variation of the drop in the X and Y directions are independent, with $X \sim N(0, 175^2)$ and $Y \sim N(50, 150^2)$, with the bias in Y due to unexpected cross breezes. Any strike within the ellipse is considered a hit. Simulate the probability of hit by using 1000 simulations for the X and Y positions of the bomb impact and determining whether that point

is inside the ellipse.

Note: You will likely want to express the equation in terms of X and Y to compare an observed X to the boundary for X and the observed Y to the boundary for Y . I recommend that you first screen the observed X and Y to see if they individually have the proper domain for the ellipse: then see if the point is inside. Turn in your code. Run your simulation of 1000, ten times, and list in increasing order your sample proportions to the hundredths place.

Answer:-Starts on next page.

Trial Number	Accuracy	Hits	Misses	Total
TRIAL 9	0.703	703	297	1000
TRIAL 2	0.704	704	296	1000
TRIAL 7	0.71	710	290	1000
TRIAL 3	0.714	714	286	1000
TRIAL 10	0.718	718	282	1000
TRIAL 1	0.722	722	278	1000
TRIAL 5	0.726	726	274	1000
TRIAL 8	0.728	728	272	1000
TRIAL 4	0.735	735	265	1000
TRIAL 6	0.745	745	255	1000

Visualizations Below Code Sample

R SCRIPT:

```

inside_target <- function(x,y) {
  if (x^2/400^2 + y^2/200^2<= 1){
    return(TRUE)
  }
  else{
    return(FALSE)
  }
}
draw_target <- function(x, s) {
  if ( s == -1){
    return( -200*(1-x^2/(400^2))^(1/2) )
  }
  else{
    return(200*(1-x^2/(400^2))^(1/2))
  }
}

trial_results = c()

for (j in (1:10)) {
  x_coords=rnorm(1000, 0, 175)
  y_coords=rnorm(1000, 50, 150)

  hits_x = c()
  hits_y = c()
  misses_x = c()
  misses_y = c()

  for (i in (1:1000)){
    if (inside_target(x_coords[i],y_coords[i])){
      hits_x = c(hits_x, x_coords[i])

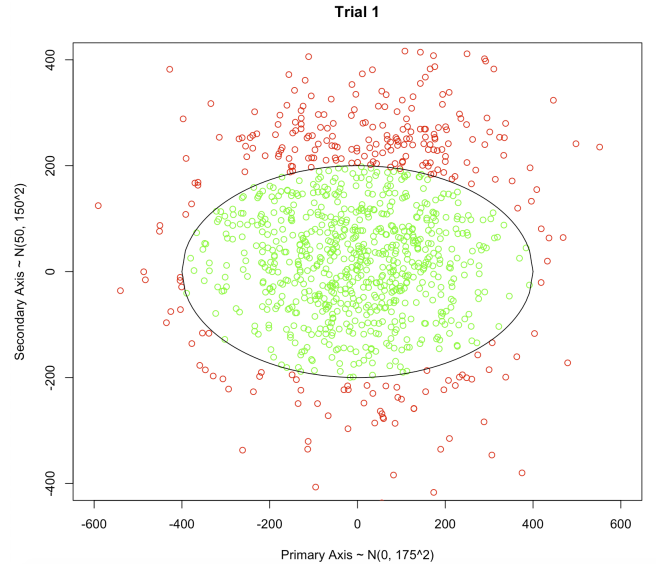
```

```

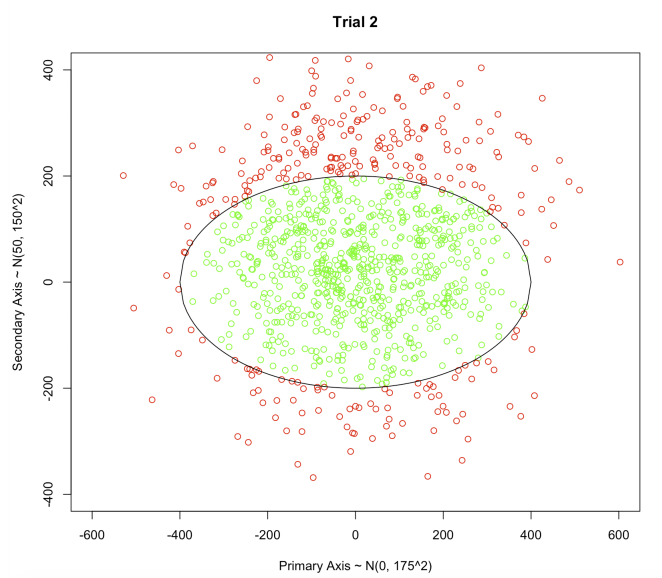
    hits_y = c(hits_y, y_coords[i])
  }
  else{
    misses_x = c(misses_x, x_coords[i])
    misses_y = c(misses_y, y_coords[i])
  }
}
trial_results = c(trial_results,length(hits_x)/1000.000)
plot(hits_x, hits_y , type = 'p', from=-400,to = 400, add = TRUE, col
     = 'green',xlim=c(-600,600), ylim=c(-400,400), xlab="Primary Axis
     ~ N(0, 175^2)",ylab="Secondary Axis ~ N(50, 150^2)",main =
     paste("Trial", toString(j)[1]))
points(misses_x, misses_y , type = 'p', from=-400,to = 400, add =
      TRUE, col = 'red',xlim=c(-600,600), ylim=c(-600,600))
curve(draw_target(x,-1),from=-400,to = 400, add = TRUE,
      xlim=c(-600,600), ylim=c(-400,400), xlab="Primary Axis ~ N(0,
      175^2)",ylab="Secondary Axis ~ N(50, 150^2)")
curve(draw_target(x,1),from=-400,to = 400, add = TRUE,
      xlim=c(-600,600), ylim=c(-400,400), xlab="Primary Axis ~ N(0,
      175^2)",ylab="Secondary Axis ~ N(50, 150^2)")
}

```

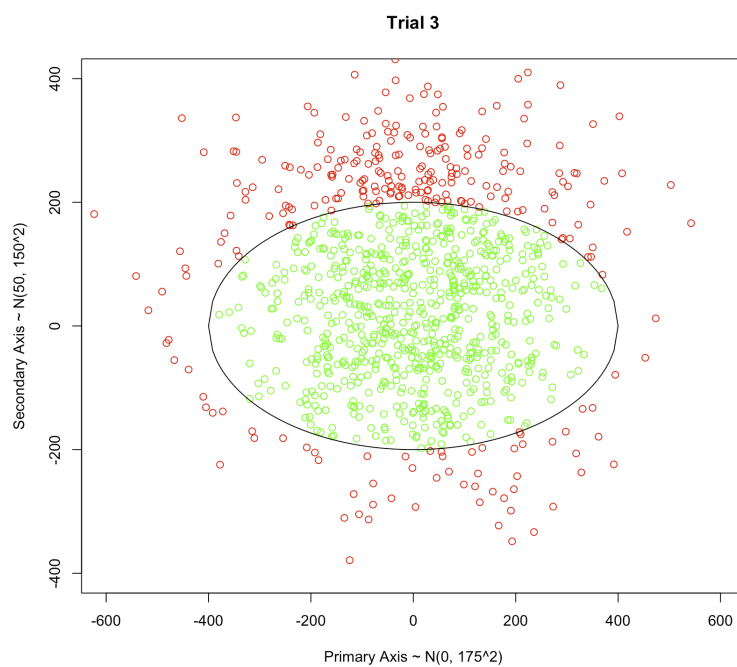
TRIAL 1



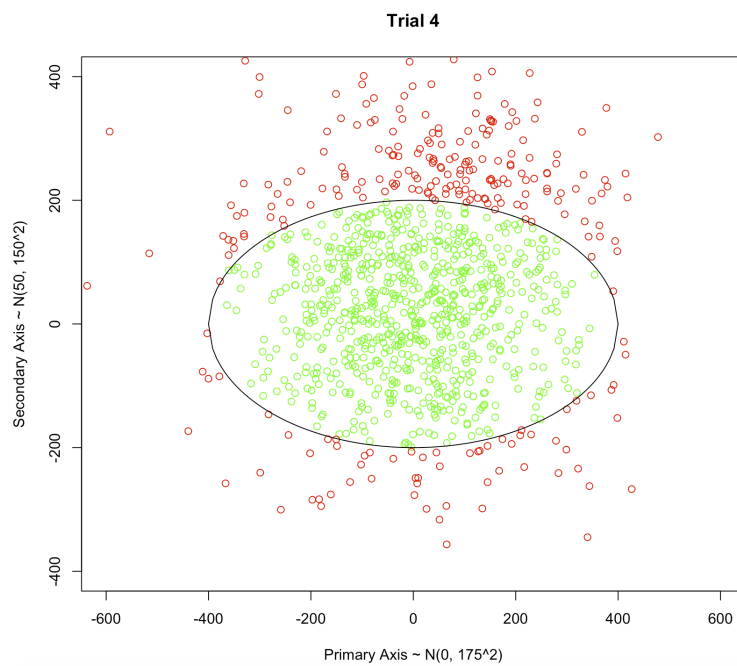
TRIAL 2



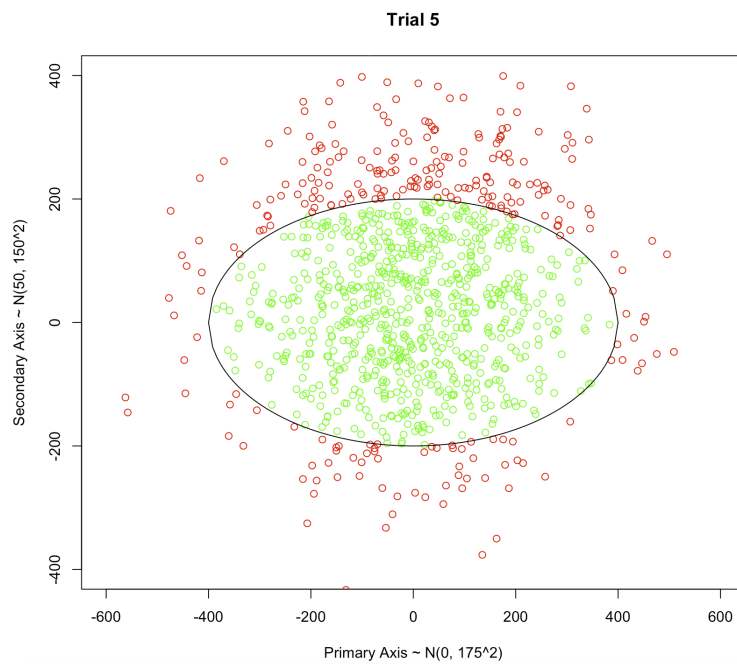
TRIAL 3



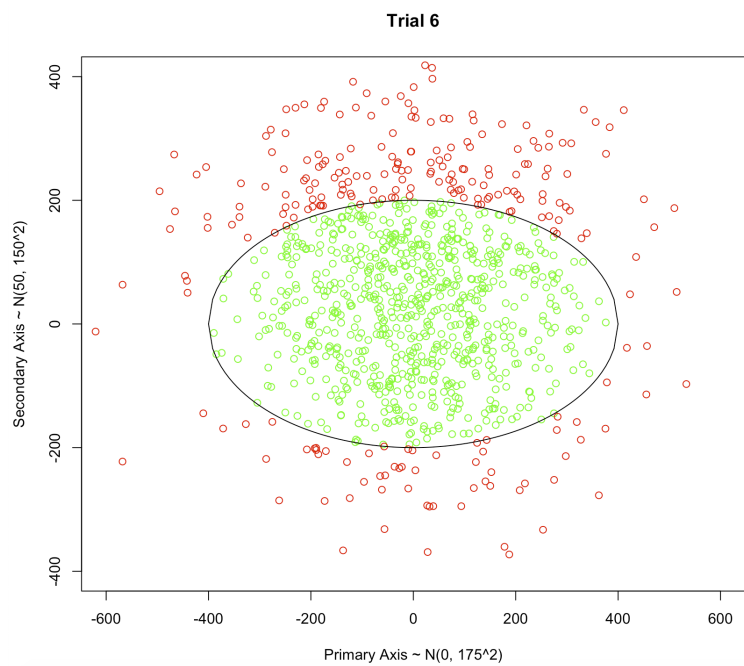
TRIAL 4



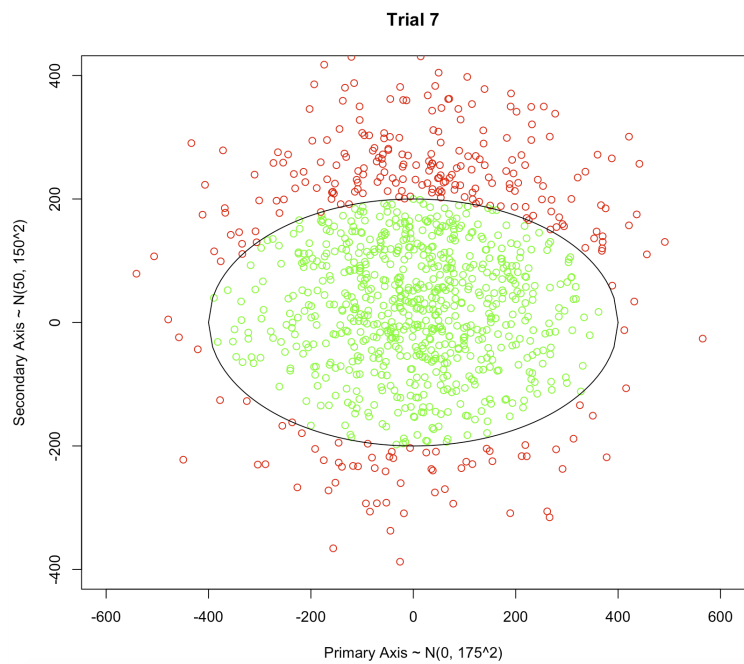
TRIAL 5



TRIAL 6

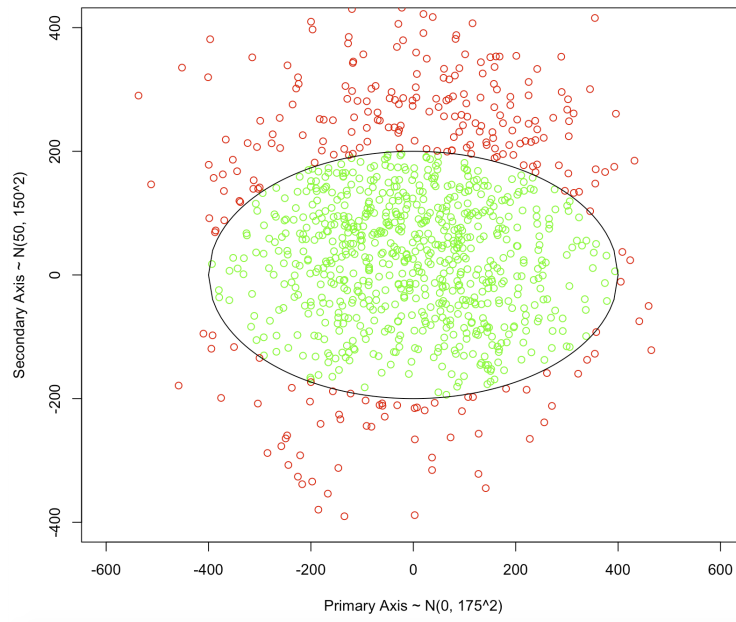


TRIAL 7



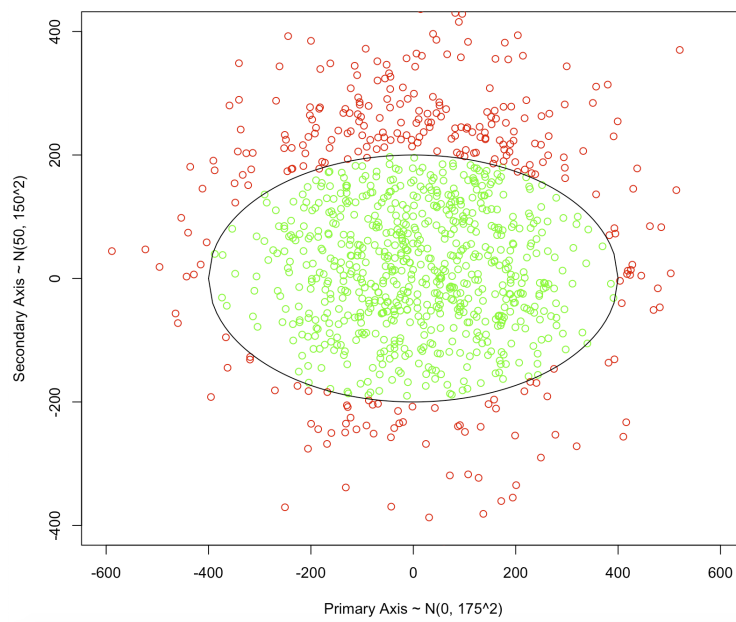
TRIAL 8

Trial 8



TRIAL 9

Trial 9



TRIAL 10

