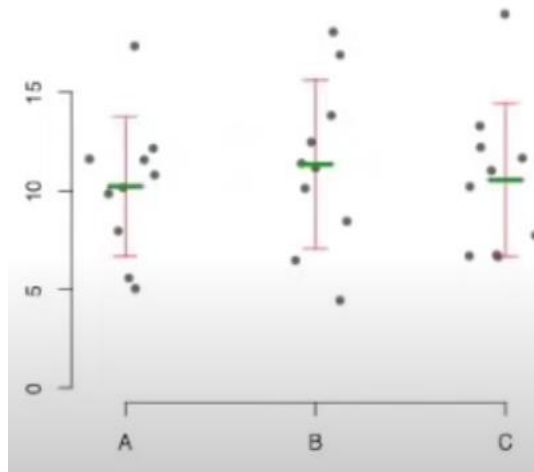
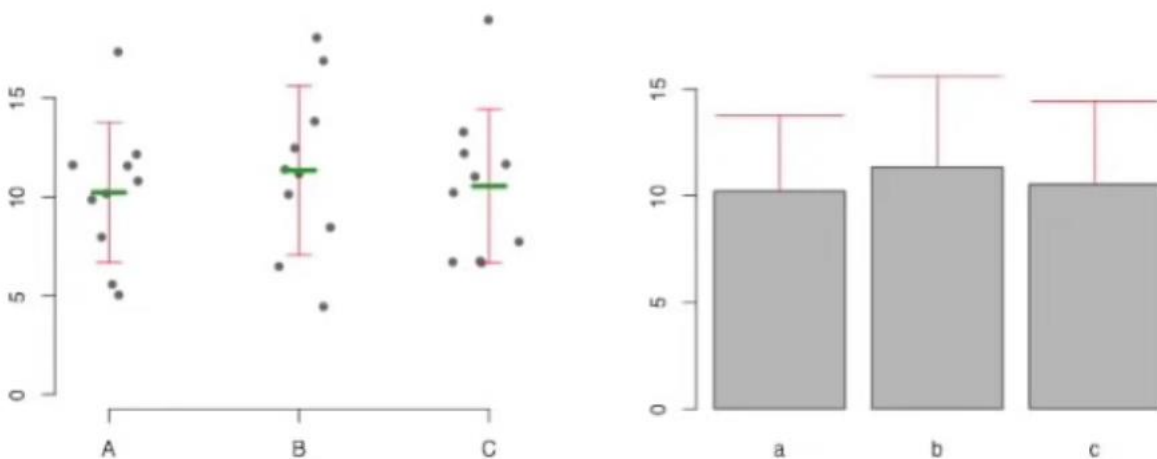


The Standard Error

Standart hatalarla çok yakından ilişkili olan error barlardan bahsederek başlayacağız. Örneğin, A, B ve C olarak etiketlenmiş üç numuneden ölçümler toplayabilir ve bunları tıpkı burada gördüğümüz gibi bir scatter plot üzerinde çizebiliriz. Üç datasetinin ortalamalarını hesaplayıp bunları veri noktalarının yaklaşık yarısında üç yeşil yatay çubukla gösterdik. Daha sonra standart sapmaları hesaplayıp grafiğe ekledik, bunları kırmızı error barlarla gösterdik.



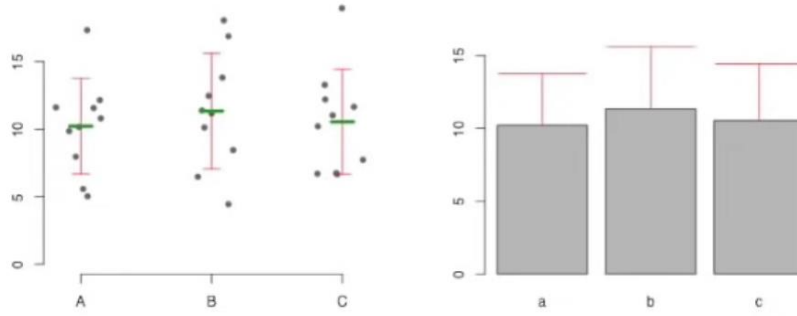
Sunumlarda, insanlar genellikle orijinal verileri göstermezler, bunun yerine sadece standart sapmadaki ortalamayı ve dinamit grafiği denilen şeyi gösterirler çünkü çizimdeki her sütun bir dinamit çubuğunun ateşleyicisi gibi görünür.



Üç yaygın hata çubuğu türü vardır:

Standard Deviations

Az önce gördüğümüz ilk standart sapma türü. Bunlar, verilerin ortalama etrafında nasıl dağıldığını bize söyler. Büyük standart sapmalar, bazı veri noktalarının ortalamadan oldukça uzak olduğunu söyler. Çoğu durumda, bize verileriniz hakkında bilgi verdiği için grafiklerimizde kullanmanız gereken budur.



Standard Errors

İkinci tip hata çubuğu, standart hatalardan gelir, bunlar bize sadece verilerin değil, ortalamasının nasıl dağıldığını söyler.

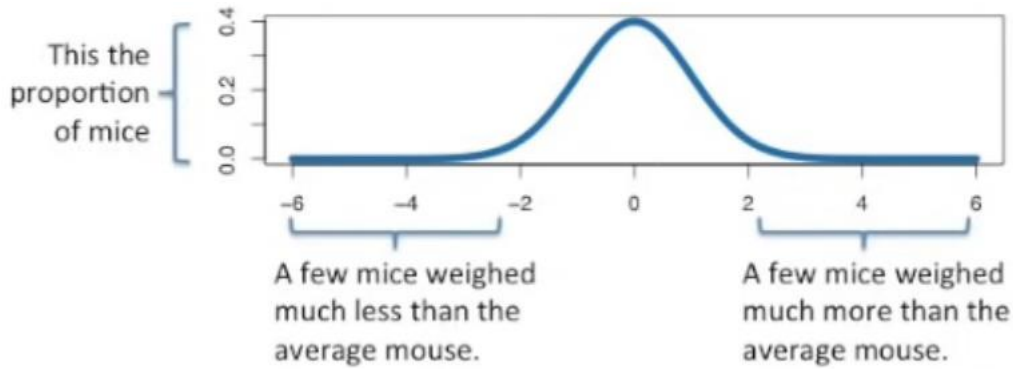
Confidence Intervals

Üçüncü yaygın hata çubuğu türü, güven aralıklarıdır ve bunlar hatalarla ilgilidir.

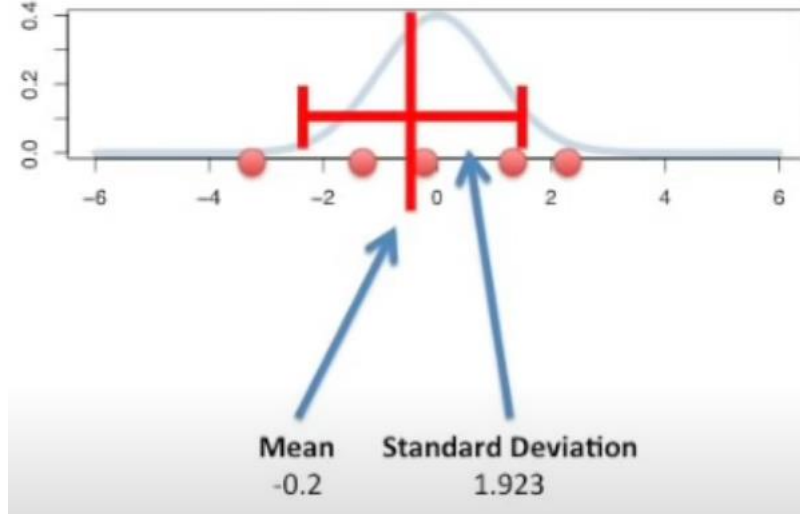
Burada ikinci tip error olan standard error'den bahsedeceğiz.

Standard Errors

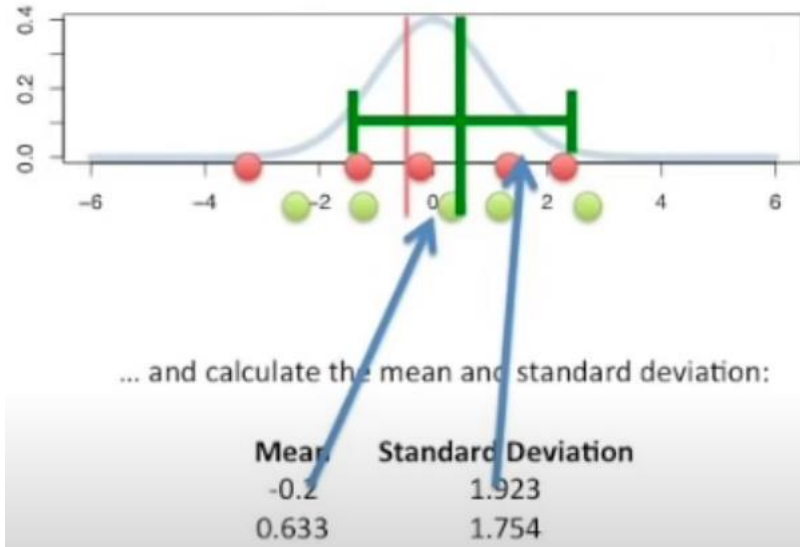
Normal dağılımla başlayalım. Bu case'te, birçok fareyi tarttığımızı ve ortalamadan farklılıkların bir dağılımını çizdiğimiz hayal edelim. y eksenini, tarttığımız farelerin oranıdır. X eksenini, ortalamadan farkıdır. Farelerin çoğu ortalamaya yakın ağırlığa sahipti, birkaç fare ortalama fareden çok daha az ağırlığa sahipti ve diğer birkaç fare ortalama fareden çok daha ağırdı.



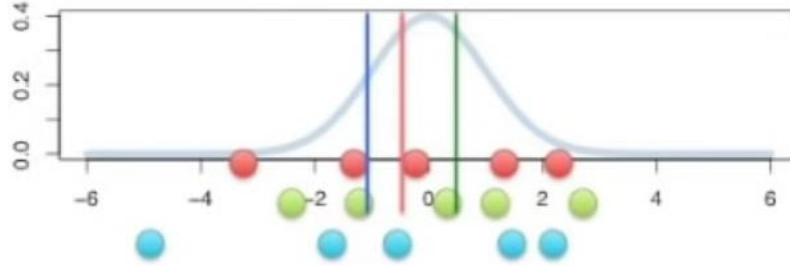
Genellikle tüm farelerin ağırlığını ölçmeye gücümüz yetmez, bu yüzden sadece bir sample alırız. Bu örnekte, tüm fareleri ölçmek yerine popülasyondan beş ölçüm aldığımızı varsayacağız. Çoğu fare ortalama ya yakın ağırlığa sahip olduğundan, örneklerimizin çoğu 0'a yakındır. Bu case'te örneğimizin ortalaması -0.2 ve standart sapma 1.923'tür. Ve grafiğimizde ortalama ve standart sapmayı ortalama artı veya eksi ortalama etrafındaki standart sapma olarak çizebiliriz. Ortalamanın her iki tarafındaki bir standart sapmanın verilerin yaklaşık 68'ini kapsaması gerektiğini, ortalamanın her iki tarafındaki iki standart sapmanın ise verilerin yaklaşık yüzde 95'ini kapsaması gerektiğini unutmayın.



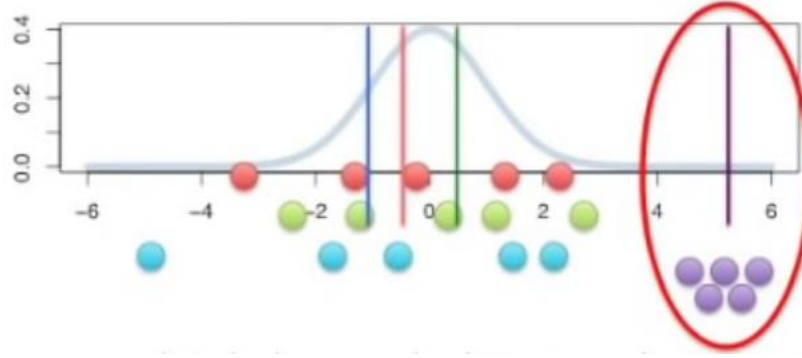
Ortalama artık daha açık bir renk çünkü ek samplelar alacağız ve aynı grafik üzerinde ek ortalamalar ve standart sapmaları üst üste koyacağız. Burada beş ölçüm daha aldık ve bu beş ölçümden ortalamayı ve standart sapmayı hesapladık. Her iki tarafta ortalama artı standart sapma ve ortalama eksi standart sapmayı çizdik.



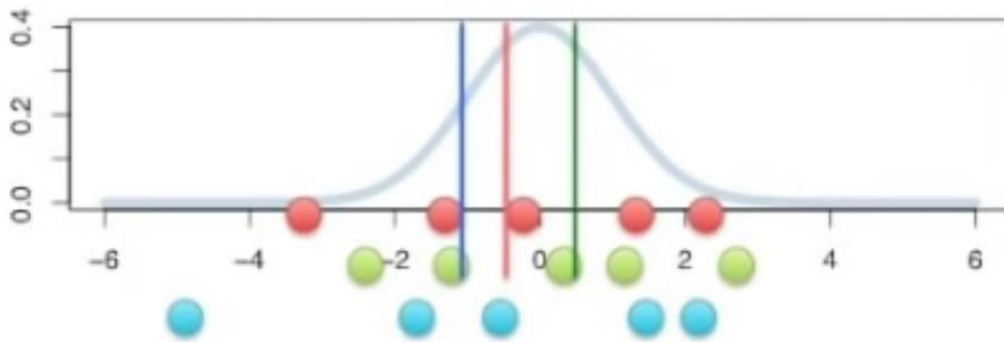
Şimdi beş ölçüm daha alıyoruz, bu ilk örnek veya ölçümlerden biri nispeten aşırı. Ancak bu bir ölçüm, ortalamayı sıfırdan o kadar uzağa sallamıyor. Yani ortalamalar, ham verilere göre nispeten birbirine yakındır.



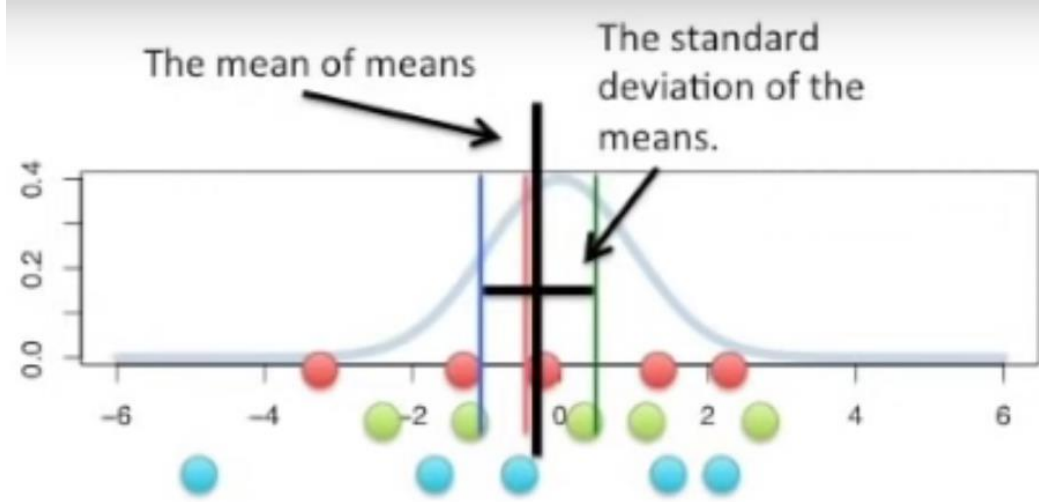
Bunun nedeni, bir ortalamanın en ortadan uzak olması için tüm ham veri noktalarının ortadan uzaktaki tek bir kümede olması gerekmesidir. Örneğin, mor noktalar sample'ının tümü, ortadan uzak bir küme oluşturur. Bu olabilir ama çok nadiren. Daha olası olan, noktaların çoğunun sıfıra yakın olduğu ve yalnızca bir veya ikisinin uzakta olduğu bir örneğe sahip olmaktır.



Şimdiye kadar her sample için standart sapmaları hesaplayabileceğinizi gösterdik, ancak şimdi üç ortalamamız olduğuna göre bu ortalamaların standart sapmasını da hesaplayabiliriz. Çünkü bir standart sapma değerlerin %68'ini, iki adet standart sapma ise değerlerin %95'ini kapsayacaktır. Ortalamaların standart sapması, verilerin standart sapmaları kadar geniş olmayacaktır.

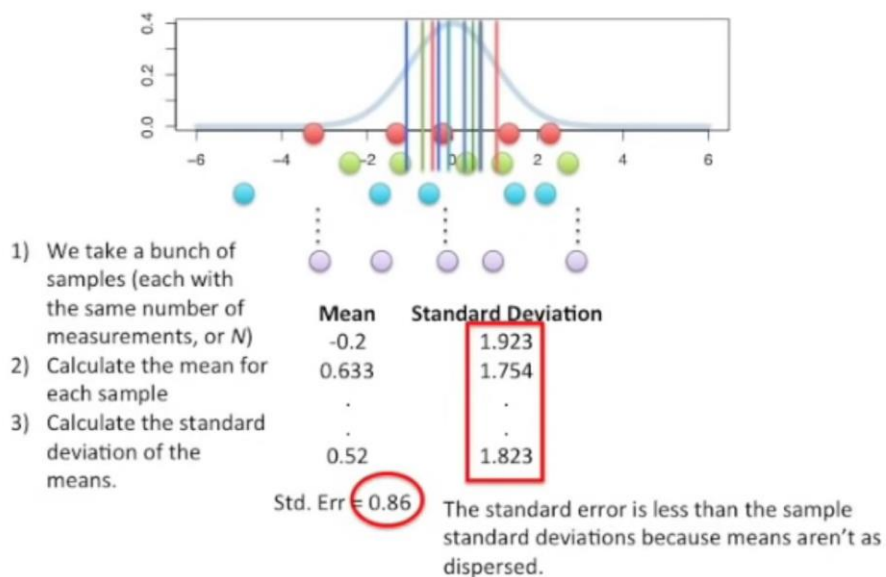


Burada ortalamaların ortalamasını ve artı veya eksi ortalamaların standart sapmasını çizdik. Bu standart sapmanın, tek tek sample'lerden elde ettiğimiz standart sapmalardan çok daha küçük olduğuna dikkat edin. **Ortalamanın standart sapmasına, ortalamanın standart hatası veya daha basit olarak standart hata denir. Standart hata, ortalamalarımızda ne kadar değişiklik bekleyebileceğimiz konusunda bize bir fikir verir.**

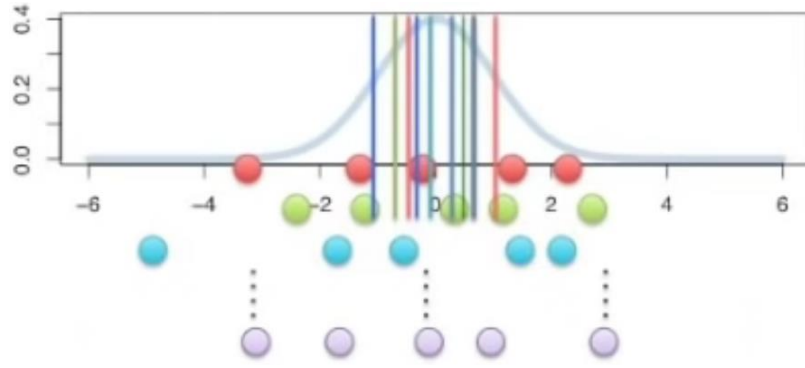


İncelemek için bir grup bağımsız beş ölçüm örneği alırsak, ortalamanın bu standart hatasını şu şekilde hesaplarız:

- 1.) İlk önce, her biri aynı sayıda ölçüm veya N olan bir grup sample alırsak. Bu durumda N, beşe eşittir.
- 2.) İkinci adım, her sample için ortalama hesaplamaktır. Burada her sample için ortalama ve standart sapmayı hesapladık. Standart hatayı bulmak için tek yapmamız gereken ortalama hesaplamak.
- 3.) Her örnek için ortalamaları hesapladığımızda, ortalamaların standart sapmasını hesaplayabiliriz. Bu case'te standart hata 0.86'ya eşittir. Burada standart hatanın standart sapmalardan çok daha az olduğunu fark ederiz çünkü ortalamalar ham veriler kadar geniş bir alana dağılmamıştır.



Standart hatanın nasıl hesaplanacağını gösterdik ancak başka standart hatalar da var. Örneğin, standart sapmaların standart sapmasını da alabiliriz. Buna standart sapmaların standart hatası deniyor sanırım dil bükümünü önlemek için. Birden fazla sample'ın standart sapmalarının nasıl dağıldığını bize anlatır. Herhangi bir istatistiğin standart sapmasını hesaplayabilirsiniz. Örneğin, medyan, mod, yüzdelikler vs. birden fazla sample için hesaplayabileceğimiz herhangi bir şey. Sadece standart sapmayı hesaplırsınız ve sonra o şeyin standart hatasını elde etmiş olursunuz. Dolayısıyla, birçok ortamı hesaplırsak, bu ortamların standart sapmalarını hesaplayabilir ve bu ortamların standart hatasını elde ederiz. Şimdiye kadar bahsettiğimiz her şeyi özetlemek gerekirse, **standart hatanın sadece aynı popülasyondan alınan birden fazla ortalamanın standart sapması** olduğunu bilin. Yani bir popülasyon varsa ve ondan bir sürü farklı sample alabilirsek, standart hatayı elde etmek için tek yapmamız gereken her örneğin ortalamalarının standart sapmasını hesaplamaktır.

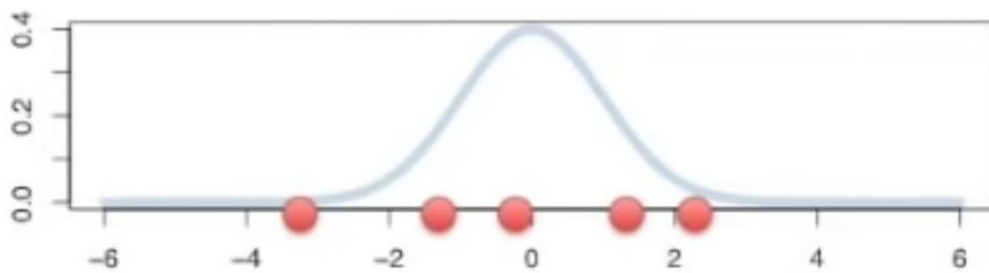


Bu noktada, aynı deneyi birkaç kez yapmak için çok fazla zaman ve para harcamadan standart hatayı hesaplayabilir miyiz diye merak ediyor olabilirsiniz. İyi haber şu ki, cevap evet.

- Nadir durumlarda, onu tahmin etmek için kullanabileceğiniz bir formül vardır. Ortalamanın standart hatasını 1 olarak kabul ederiz. Bunun formülü çok basittir. Bunu, standart sapmayı sample boyutunun kareköküne bölerek elde ediyoruz.

- Yine iyi haber şu ki, geri kalan her şey için bootstrapping adı verilen şeyi kullanabiliriz. Ne zaman basit bir formülümüz yoksa, onu bootstrap yapabiliriz. Bootstrap ile ilgili güzel olan şey, kavramsal olarak çok basit olması ve bu işi bir bilgisayarın yapmasının kolay olmasıdır.

İşte 5 ölçüm yaptığımız bir deneyden önceki gibi bir bootstrap örneği. Genellikle bootstrap için tek bir deneyde 10 veya daha fazla ölçümün olması iyidir. Şimdi aşağıdaki adımlarla verilerimize bootstrap (önyükleme) yapıyoruz.



1. Önce az önce aldığımız sample'dan rastgele bir (1) ölçüm alıyoruz. Bu rastgele ölçüm, daha önce almadığımız yeni bir ölçüm değil. Görmediğimiz yeni bir numara değil. Zaten sahip olduğumuz sample'ın bir parçası.

2. Şimdi sadece bu değeri yazıyoruz. Bu case'te 1.43'tür.

3. Üçüncü adımda, birinci adıma geri dönüyoruz ve yeni bir rastgele ölçüm seçiyoruz ve bu değeri yazıyoruz ve bunu beş kez yapıyoruz. İkinci ölçümümüz -1.38'dir. Üçüncü ölçüm -3.11'dir. Dördüncü ölçümümüz 1.43'tür. Bu ölçümü daha önce seçmiştik ama sorun değil, bootstrap yaparken sample'ınızdan beş ölçüm almanız yeterlidir ve aynısını daha önce seçmiş olmanızın bir önemi yoktur. Son ölçümümüz -0.10'dur.

4. Bootstrapte dördüncü adım, standart hatasını anlamakla ilgilendiğimiz ortalamayı, medyanı, modu veya herhangi bir istatistiği hesaplamaktır ve bunu sample'ımızla hesaplarız. Bu case'te, ortalamamızın standart hatasıyla ilgileniyoruz. Yani tek yaptığımız yeni (bootstrap sonucu elde ettiğimiz) sample'ımızın ortalamasını hesaplamak.

5. Beşinci adım, ilk adıma kadar geri gitmek ve çok fazla ortalamamız veya medyanımız olana veya standart hatayı hesaplamakla ilgilendiğiniz herhangi bir şeye sahip olana kadar bunu tekrar etmek.

6. Bootstrap prosedüründeki altıncı ve son adım, birden beşe kadar olan adımlarda ürettiğimiz tüm ortalamaların standart sapmasını basitçe hesaplamaktır. Hepsi bu kadar. Bu durumda ortalamamızın standart hatasını hesapladık ve bunu grafikte siyah bir çizgi olarak çizdik. Yani standart hatayı hesaplamamıza yardımcı olacak süslü bir formül yoksa, bunu kendimiz sıfırdan yapabiliriz. Sadece bootstrapi kullanabilir ve işi bitirebiliriz.

