

exp distribution simulation and tooth growth data analysis

eric

15/08/2021

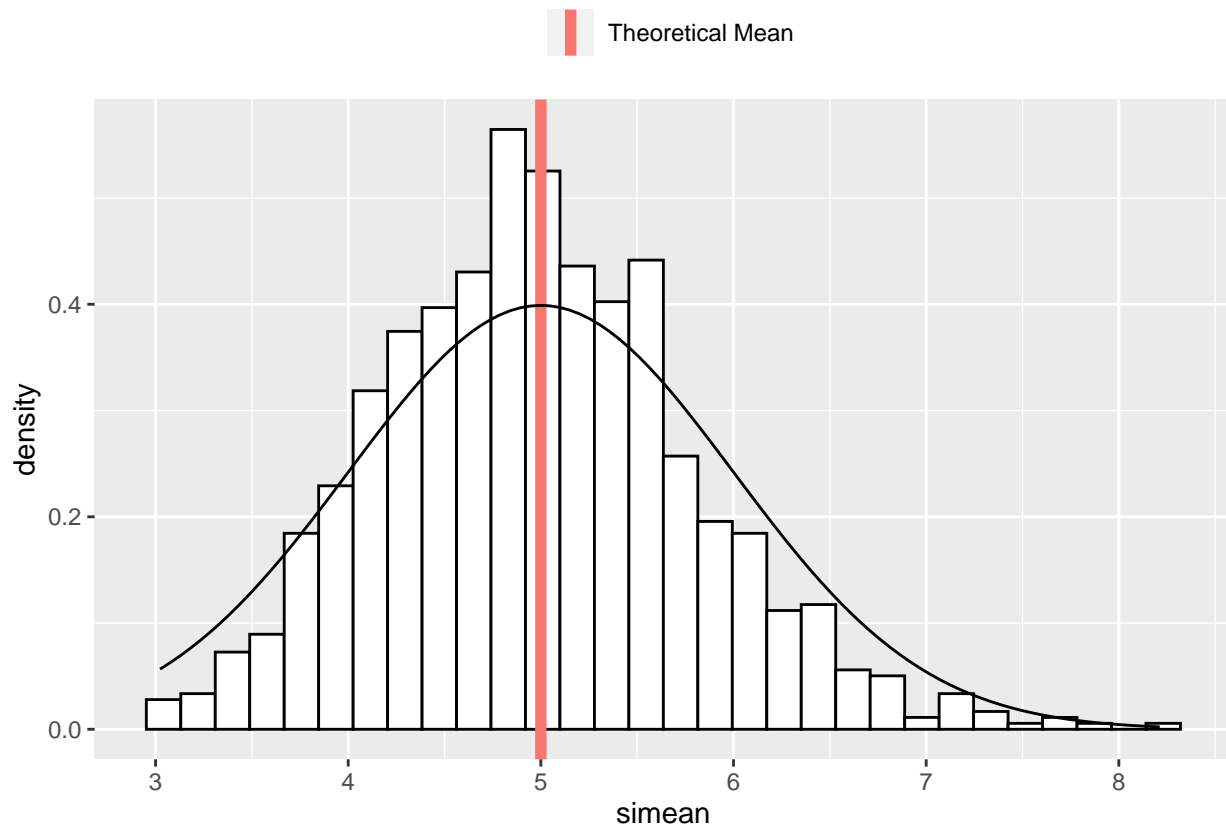
Simulation

We have a pool of exp distribution data. And we use bootstrapping to obtain 40 data, calculate the mean and the variance and repeat this process 1000 times. And then the distribution of the variance and mean data were plotted and a normal distribution with mean=5 was also plotted to see how much the mean distribution fitted .

```
set.seed(1000)
simulate <- matrix(rexp(1000*40, 0.2), nrow = 1000, ncol = 40)
simulatemean <- as.data.frame(apply(simulate, 1, mean))
colnames(simulatemean) <- "simean"
simulatevar <- as.data.frame(apply(simulate, 1, var))
colnames(simulatevar) <- "simvar"

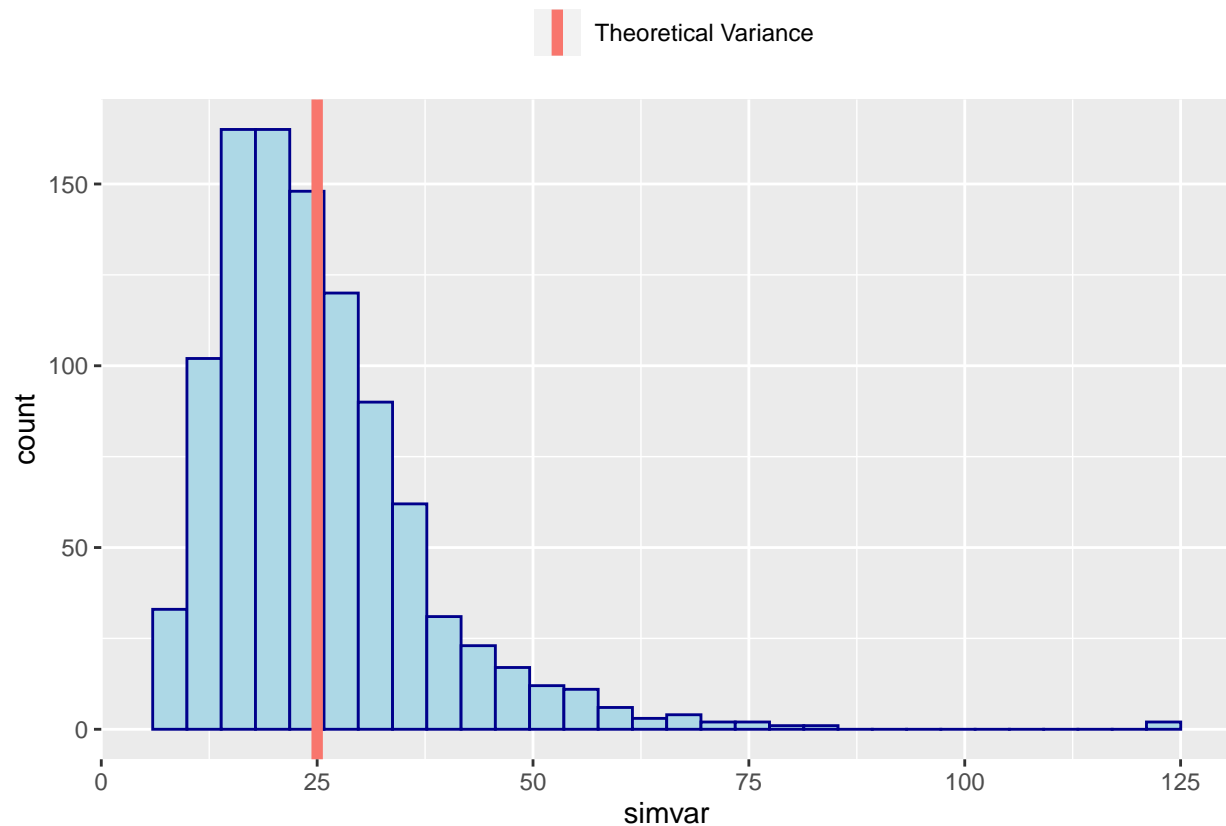
ggplot(simulatemean, aes(x=simean)) +
  geom_histogram(aes(y = ..density..), color="black", fill="white") +
  geom_vline(aes(xintercept=5, color = "Theoretical Mean"), size=2) +
  theme(legend.title=element_blank(), legend.position = "top", legend.background = element_blank()) +
  stat_function(fun = dnorm, args = list(mean = 5, sd = 1))
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



```
ggplot(simulatevar,aes(x=simvar))+
  geom_histogram(color="darkblue", fill="lightblue")+
  geom_vline(aes(xintercept=1/0.04,color = "Theoretical Variance"), size=2)+
  theme(legend.title=element_blank(),legend.position = "top",legend.background=element_blank())
```

'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.



Data summary

we used Dplyr to see the relationship between teeth length, supplement type and supplement dose. we can see with the increase in the dose, two types of supplements can both increase the teeth length.

```
data<-ToothGrowth
data%>%group_by(supp,dose)%>%summarise(mean(len))
```

'summarise()' has grouped output by 'supp'. You can override using the '.groups' argument.

```
## # A tibble: 6 x 3
## # Groups:   supp [2]
##   supp  dose 'mean(len)'
##   <fct> <dbl>     <dbl>
## 1 OJ    0.5      13.2
## 2 OJ    1        22.7
## 3 OJ    2        26.1
## 4 VC    0.5       7.98
## 5 VC    1        16.8
## 6 VC    2        26.1
```

The efficacy of two supplement

T-test was performed to see if there are significant difference between supplements given the same dose.

```
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = ToothGrowth)
```

```
##  
## Welch Two Sample t-test  
##  
## data: len by supp  
## t = 1.9153, df = 55.309, p-value = 0.06063  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.1710156 7.5710156  
## sample estimates:  
## mean in group OJ mean in group VC  
## 20.66333 16.96333
```

```
t.test(len ~ supp, var.equal = F, data = filter(ToothGrowth, dose == 0.5))
```

```
##  
## Welch Two Sample t-test  
##  
## data: len by supp  
## t = 3.1697, df = 14.969, p-value = 0.006359  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 1.719057 8.780943  
## sample estimates:  
## mean in group OJ mean in group VC  
## 13.23 7.98
```

```
t.test(len ~ supp, var.equal = F, data = filter(ToothGrowth, dose == 1))
```

```
##  
## Welch Two Sample t-test  
##  
## data: len by supp  
## t = 4.0328, df = 15.358, p-value = 0.001038  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 2.802148 9.057852  
## sample estimates:  
## mean in group OJ mean in group VC  
## 22.70 16.77
```

```
t.test(len ~ supp, var.equal = F, data = filter(ToothGrowth, dose == 2))
```

```
##  
## Welch Two Sample t-test  
##  
## data: len by supp  
## t = -0.046136, df = 14.04, p-value = 0.9639  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:
```

```
## -3.79807 3.63807
## sample estimates:
## mean in group OJ mean in group VC
##          26.06          26.14
```

Conclusion

the low p value suggested that in dose of 1, two types of supplements had the similar effect on teeth growth. However, taking into consideration of other t-tests, these two supplement were more likely to have different effect on teeth growth. Further data analysis can use permutation to validate the results.