Statistical Interfence Course Project

Peter Sedivec

Saturday, February 21, 2015

Investigation of exponential distribution (rexp)

If we compare the sample mean of mexs with the theoretical mean (1/lambda) we see that with 1000 samples the mean has converged very close to the theoretical mean:

```
mean(mexs) # sample mean
```

```
## [1] 5.013244
```

```
1/0.2 # theoretical mean
```

```
## [1] 5
```

```
mean(mexs) - 1/0.2
```

```
## [1] 0.01324436
```

If we now look at the sample variance and compare it with the theoretical variance 1/lambda^2 we see that the sample variance is tighter (this doesn't really make sense)

```
var(mexs) # sample variance
```

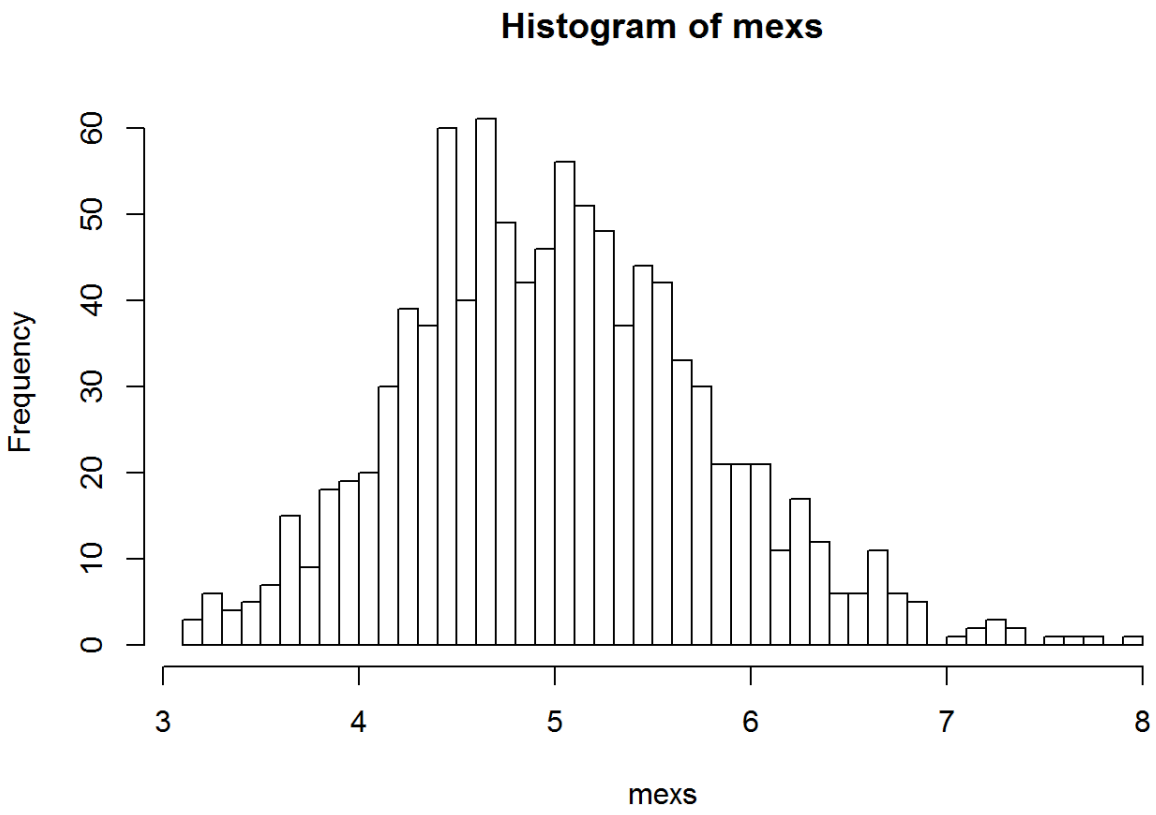
```
## [1] 0.6086785
```

```
1/(0.2^2) # theoretical variance
```

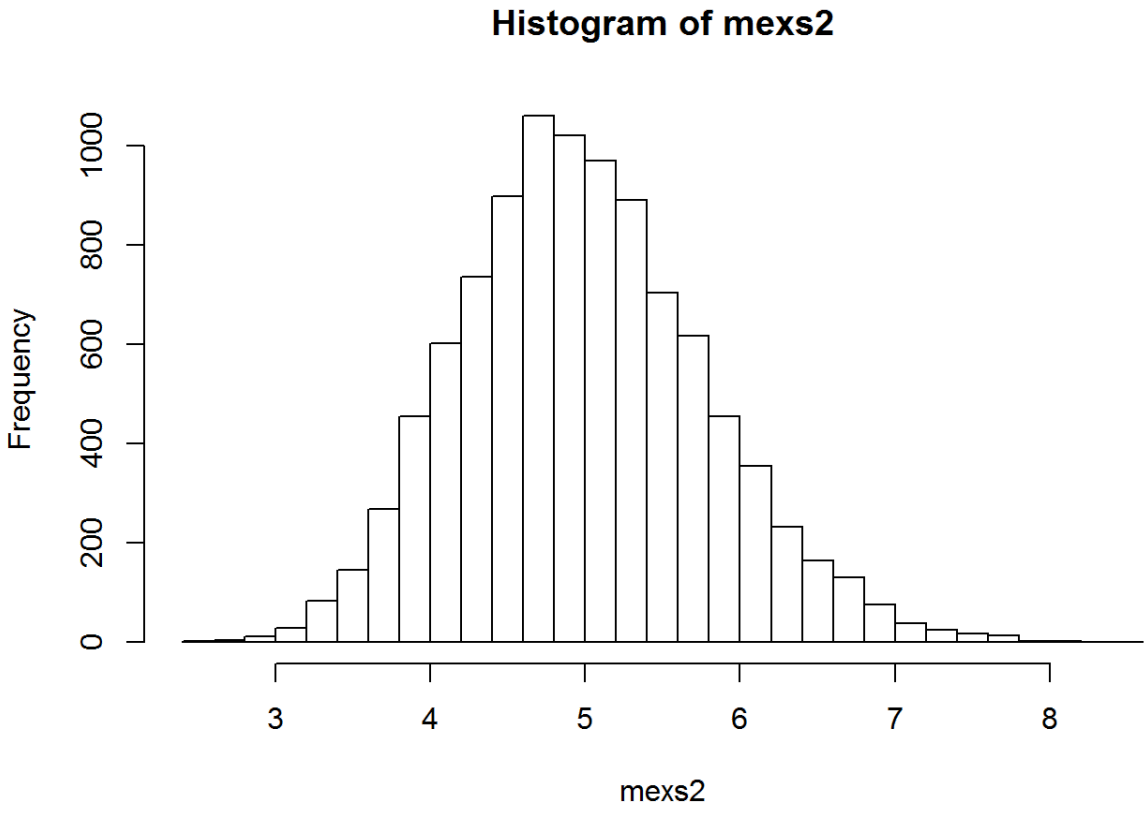
```
## [1] 25
```

If we look at a histogram of the distribution we can see that it is approximately normal

```
hist(mexs, 40)
```



And as we increase the sample size from n=1,000 to n=10,000 we see the histogram converging to a normal distribution as expected per the central limit theorem



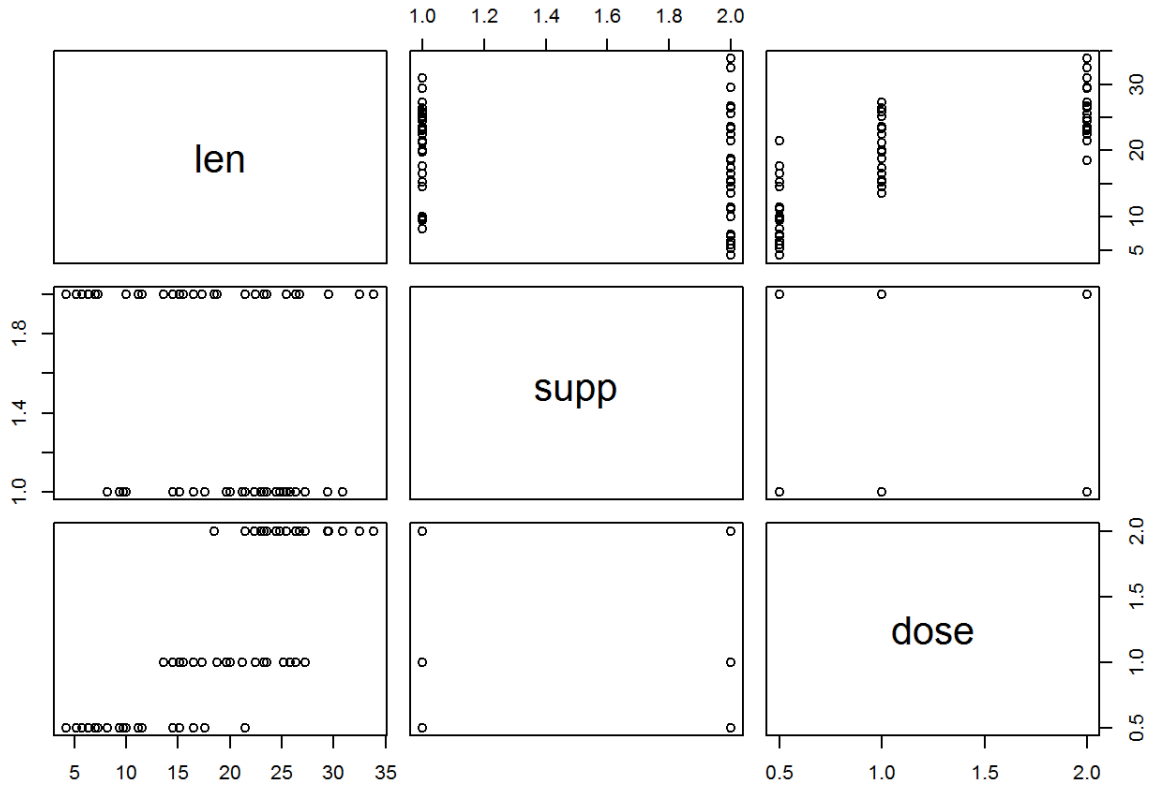
Tooth Growth Investigation

Let’s take a look and explore the ToothGrowth dataset first. Reading the help for the dataset

we learn that: “The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).” Now let’s do a little of our own explanatory analysis

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```




```
## [1] "table of dosages"
```

```
##
## 0.5   1   2
##  20  20  20
```

From looking at the plot, it doesn't seem like there is much correlation between supplement deliveraly method (OJ or vitamin C) and length, however there appears to be a clear pattern of increasing length as dosage increases from 0.5mg to 2mg

Let's take a look at the t-intervals for tooth length based on dosages (0.5, 1, 2) and supplement:

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
##  0.4708204 6.9291796
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15352 -12.83648
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

After looking at the results from the t-tests the conclusions I would draw based on 95% confidence levels are:

1. There is not enough evidence to support the conclusion that delivery of the supplement through Vitamin C is more effective than through absorbic acid. The basis for this assumption is the confidence interval includes 0 indicating that it is possible for there to be no difference between delivery methods.
2. The evidence strongly supports that increasing the dosage of vitamin C supplement has a significant impact on tooth length growth. This result is based on the fact that all confidence intervals (when comparing 0.5 vs 1, 0.5 vs 2 and 1 vs 2mg of supplements) do not contain zero
3. There is diminishing return for the impact of tooth growth as supplement is increased. This conclusion is based on the fact that when looking at the confidence intervals of 0.5 vs 1 and 1 vs 2 we see that the confidence interval is narrower and closer to zero

In using t.test to perform a student t-Test I made the following assumptions:

1. the data is not paired
2. we can assume pooled variance can be used to estimate the variance

Other observations: When looking at the t-test assuming a 90% confidence interval we can in deed claim that the Orange Juice delivery method has a statistically significant effect on tooth growth because the confidence interval does not include 0