Statistical Inference - Part 1

Hayden Roche

December 18, 2017

Synopsis

The examples outline below will be simulation exercises that aim to illustrate the properties of the distribution of the mean of 40 exponentials. Through the comparison of sample and theoretical means and variance, we will show that the distribution is approximately normal.

Part 1

Question 1: Show the sample mean and compare it to the theoretical mean of the distribution

The process seen below is the creating of our variables that will be piped into the Sample mean function. This is a similar example as we've seen in the course Swirl exercises, but I've tailored it to the example in this Course Project. The equation for theoretical mean is (lambda^-1)

```
n <- 40
lambda <- 0.2
Sims <- 1000

TheoMean <- 1/lambda
TheoMean
```

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

```
Sample_mean <- NULL
for(i in 1:Sims) {
  Sample_mean <- c(Sample_mean, mean(rexp(n, lambda)))
}
mean(Sample_mean)
```

```
## [1] 4.963895
```

Question 2: Show how variable the sample is and compare it to the theoretical variance of the distribution

In this part, I'm running the sample variance of the simulated dataset. The sample variance ended up being 0.62. The equation for the theoretical mean of an expontential distribution is 1/(lambda*sqrt(n))^2. The theoretical variance is 0.625, therefore it is incredibly close to the sample variance.

```
Sample_Variance <- var(Sample_mean)
Sample_Variance
```

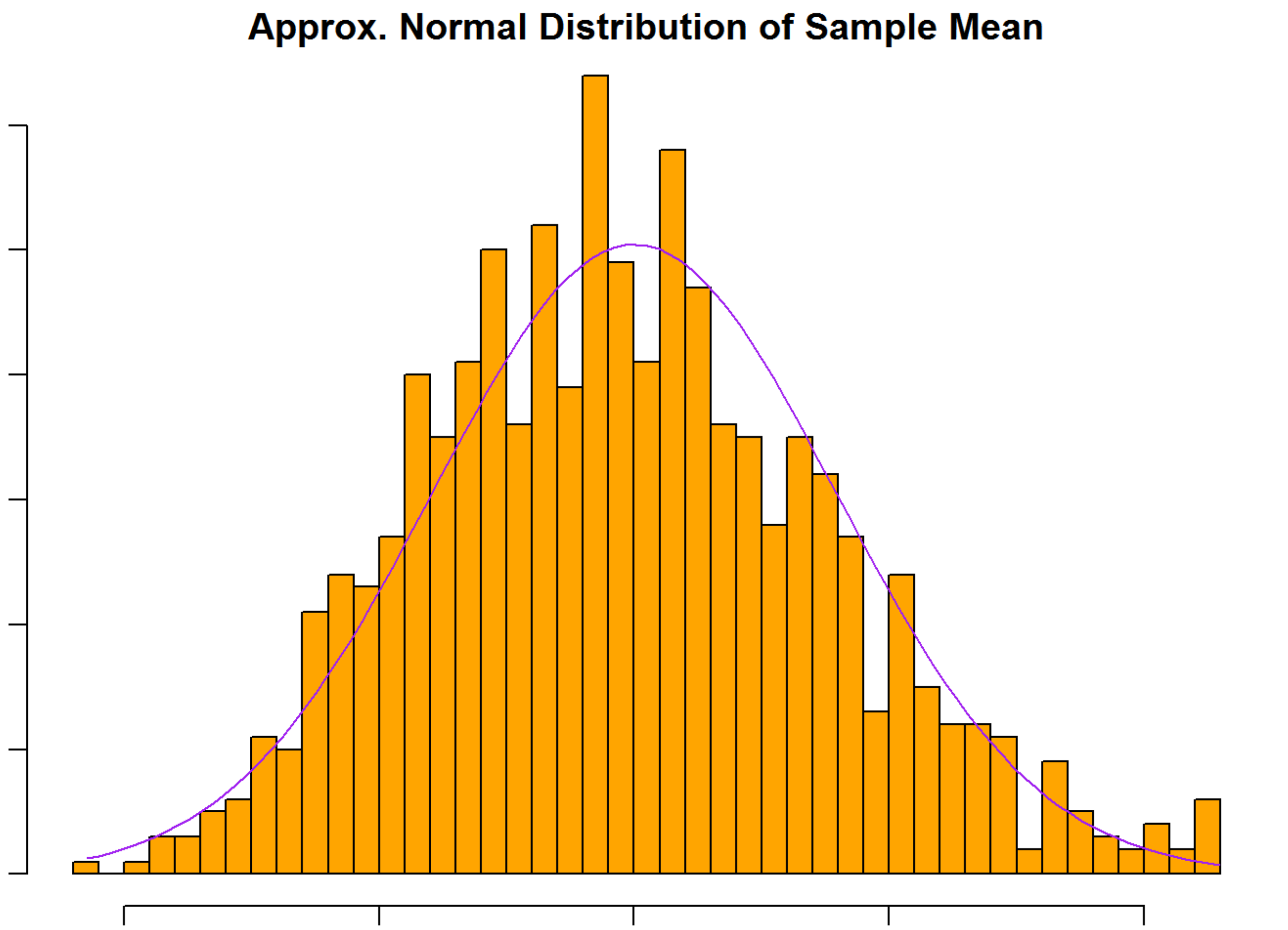
```
## [1] 0.6264656
```

```
TheoVariance <- (lambda*sqrt(n))^-2
TheoVariance
```

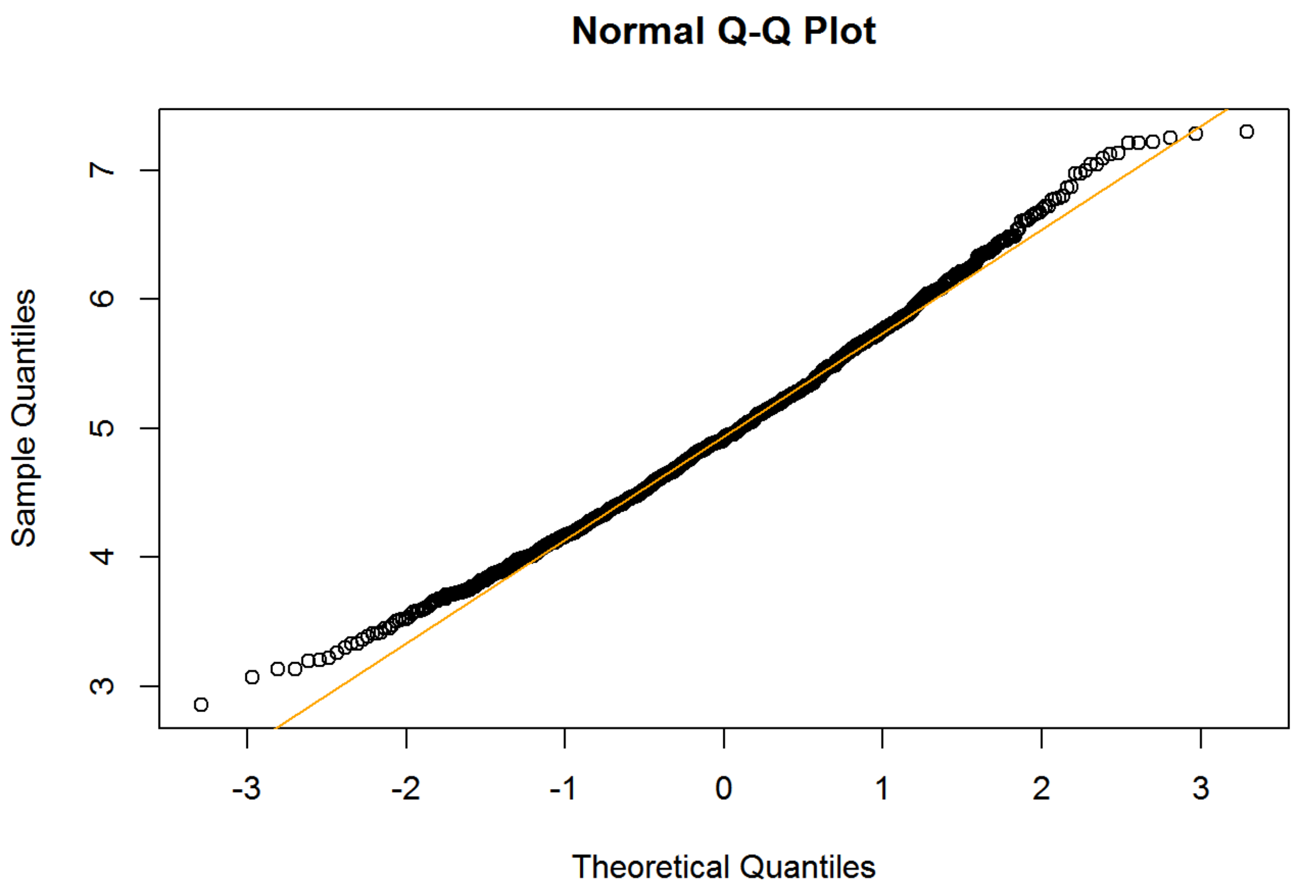
```
## [1] 0.625
```

Question 3: Show that the distribution is approximately normal

```
par(mar=c(1,1,1,1))
hist(Sample_mean, breaks = 50, prob = T, col = "orange", xlab =  "Mean_samples", y
lab = "Density", main = "Approx. Normal Distribution of Sample Mean")
x <- seq(min(Sample_mean), max(Sample_mean), length = 100)
lines(x, dnorm(x, mean = 1/lambda, sd = (1/lambda/sqrt(n))), pch = 25, col = "purp
le")
```




```
qqnorm(Sample_mean)
qqline(Sample_mean, col = "orange")
```



Question 3: Use confidence intervals and/or hypothesis tests to compare growth by supp and dose

The following hypothesis tests will be analysis the null hypothesis of: “OJ and VC deliver the same tooth growth”. The various tests will test the hypothesis in 4 scenarios (entire dataset, 0.5mg dosage, 1.0mg dosage and 2.0mg dosage)

Test – In this test we aren’t subsetting by dosage, therefore we are looking at the entire dataset. The p-value is greater than the 95% confidence interval threshold of 0.05. Therfore, the null hypothesis cannot be rejected.

Test1 – In this test we are subsetting by dosage = 0.5. The confidence interval does not include 0 in this example and the p-value is well below 0.05 at 0.006. Therfore we can reject the null hypothesis and accept the alternative hypothesis that 0.5mg/day of OJ leads to more tooth growth than VC.

Test2 – In this test we are subsetting by dosage = 1.0. The confidence interval does not include 0 in this example and the p-value is well below 0.05 at 0.001. Therfore we can reject the null hypothesis and accept the alternative hypothesis that 1.0mg/day of OJ leads to more tooth growth than VC.

Test3 – In this test we are subsetting by dosage = 2.0. The confidence interval does include 0 in this subset and the p-value is well above 0.05 at 0.96. Therefore, the null hypothesis cannot be rejected.

```
Test <- t.test(len~supp, data = ToothGrowth)
Test$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```
Test$p.value
```

```
## [1] 0.06063451
```

```
Test1 <- t.test(len~supp, data = ToothGrowth, dose == 0.5)
Test1$conf.int
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
Test1$p.value
```

```
## [1] 0.006358607
```

```
Test2 <- t.test(len~supp, data = ToothGrowth, dose == 1)
Test2$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
Test2$p.value
```

```
## [1] 0.001038376
```

```
Test3 <- t.test(len~supp, data = ToothGrowth, dose == 2)
Test3$conf.int
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```
Test3$p.value
```

```
## [1] 0.9638516
```

Conclusions

Below is a breakdown with all of the results: - Entire dataset and 2.0mg/day: we're unable to make any conclusions on whether OJ is more effective than VC - 0.5mg/day and 1.0mg/day: for these two subsets of data, OJ delivers more tooth growth on average than VC