Homework 1

Joaquin Rodriguez

8/22/2017

1

```
v <- seq(from = 1,to = 100,by = 2)
```

2

```
v <- as_tibble(v)
v <-
  v %>%
  filter(., value > 60) %>%
  filter(., value < 100)
```

3

```
v %>% var(.)
```

```
##       value
## value   140
```

4

```
df <- data.frame(c(2.27, 1.98, 1.69, 1.88, 1.64, 2.14),
                 c(8.28, 8.04, 9.06, 8.7, 7.58, 8.34))
names(df) <- c("R", "H")

df$V <- 1/3 * pi * df$R^2 * df$H
```

5

```
x <- c(3.15, -0.55, -0.35, 0.16)
y <- c(2.93, -0.35, -0.25, -0.12)

X <- matrix(
  c(rep(1, 4), x),
  ncol = 2
)

solve((t(X) %*% X)) %*% (t(X) %*% matrix(y))
```

```
##             [,1]
## [1,] 0.001483229
## [2,] 0.914550657
```

```
#summary(lm(y ~ x))
lsfit(x, y)
```

```
## $coefficients
##   Intercept           X
## 0.001483229 0.914550657
##
## $residuals
## [1]  0.0476822  0.1515196  0.0686095 -0.2678113
##
## $intercept
## [1] TRUE
##
## $qr
## $qt
## [1] -1.105000000  2.731595328  0.003732199 -0.318822892
##
## $qr
##      Intercept          X
## [1,]      -2.0 -1.2050000
## [2,]       0.5  2.9868169
## [3,]       0.5  0.6032063
## [4,]       0.5  0.4324559
##
## $qraux
## [1] 1.500000 1.670167
##
## $rank
## [1] 2
##
## $pivot
## [1] 1 2
##
## $tol
## [1] 1e-07
##
## attr(,"class")
## [1] "qr"
```

6

```
library(MASS)

data("cats")
cats <- as_tibble(cats)
```

7

```
cats %>% count(Sex)
```

```
## # A tibble: 2 x 2
##      Sex     n
##   <fctr> <int>
## 1      F    47
## 2      M    97
```

8

```
cats %>%
  group_by(Sex) %>%
  summarise(mean(Bwt))
```

```
## # A tibble: 2 x 2
##      Sex `mean(Bwt)`
##   <fctr>       <dbl>
## 1      F    2.359574
## 2      M    2.900000
```

```
cats %>%
  group_by(Sex) %>%
  summarise(sd(Bwt))
```

```
## # A tibble: 2 x 2
##      Sex `sd(Bwt)`
##   <fctr>     <dbl>
## 1      F 0.2739879
## 2      M 0.4674844
```

9

```
freqover2.5 <-
  cats %>%
  mutate(., Over2.5 = ifelse(Bwt > 2.5, T, F)) %>%
  group_by(Sex) %>%
  count(Over2.5) %>%
  spread(., key = Over2.5, value = n) %>%
  mutate(., Proportion = `TRUE` / sum(`TRUE`, `FALSE`))
```

10

```
catsM <-
  cats %>%
  filter(., Sex == "M") %>%
  list(.)

catsF <-
  cats %>%
```

```
  filter(., Sex == "F") %>%
  list(.)
```

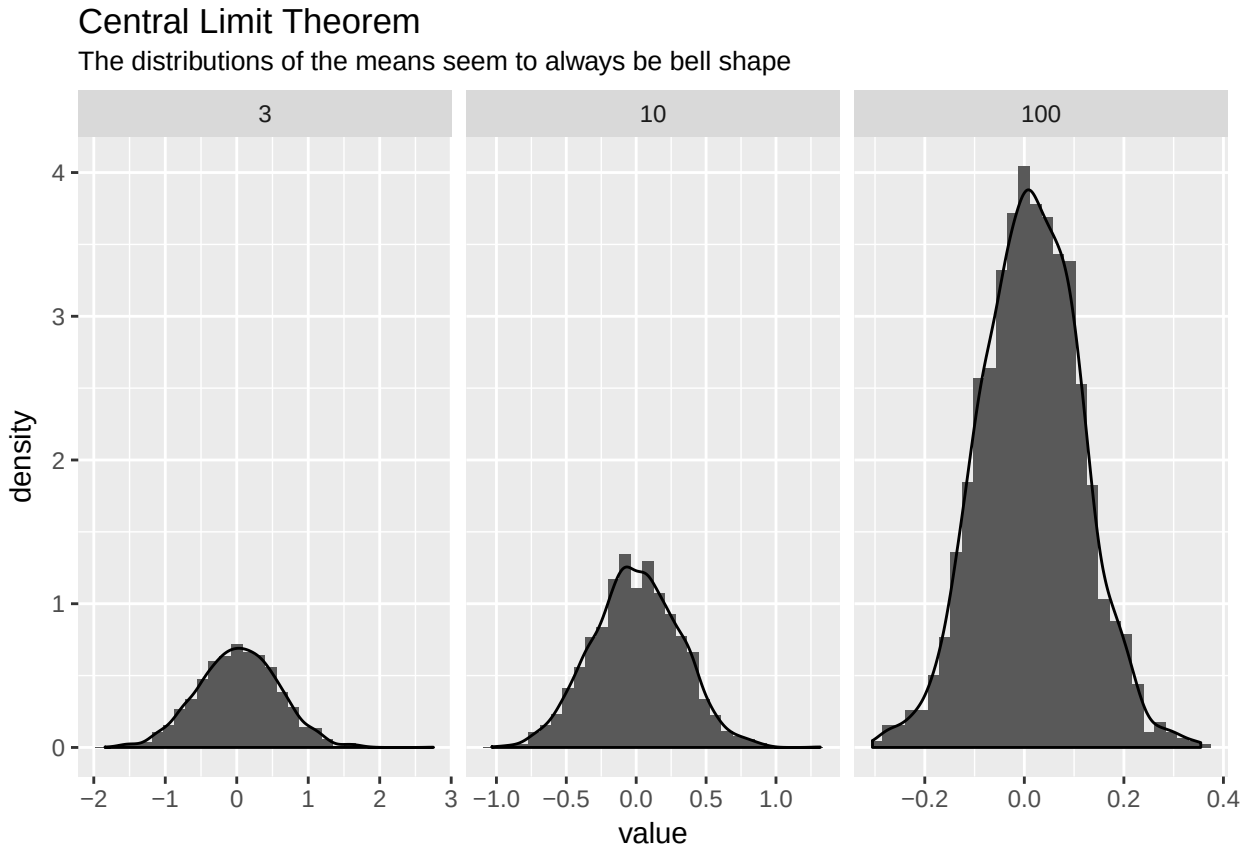
Generate Distributions and Sample Means

```
set.seed(10)
distributions <-
  tribble(
    ~f, ~params,
    "rnorm", list(n = 1000),
    "rexp", list(n = 1000),
    "rbinom", list( n = 1000, size = 1, p = 0.1)
    )

distributions <-
  distributions %>%
  mutate(distr = invoke_map(f, params))

sampling <- function(df, n) {
  result <- vector("double", 2000)
  for(i in 1:2000) {
    result[[i]] <-
      df %>%
      sample(n, replace = T) %>%
      mean
  }
  return(result)
}


distributions <-
distributions %>%
  mutate(samp3 = map(.$distr, sampling, 3),
         samp10 = map(.$distr, sampling, 10),
         samp100 = map(.$distr, sampling, 100)
  )


df <-
  distributions %>%
  .$samp3 %>%
  as.data.frame()

names(df) <- c("rnorm","rexp","rbinom")

df <- gather(df)
df$n <- 3

df1 <-
  distributions %>%
  .$samp10 %>%
  as.data.frame()
```

```r
names(df1) <- c("rnorm","rexp","rbinom")
df1 <- gather(df1)
df1$n <- 10

df <- bind_rows(df,df1)

df1 <-
  distributions %>%
  .$samp100 %>%
  as.data.frame()

names(df1) <- c("rnorm","rexp","rbinom")
df1 <- gather(df1)
df1$n <- 100

df <- bind_rows(df,df1)
```

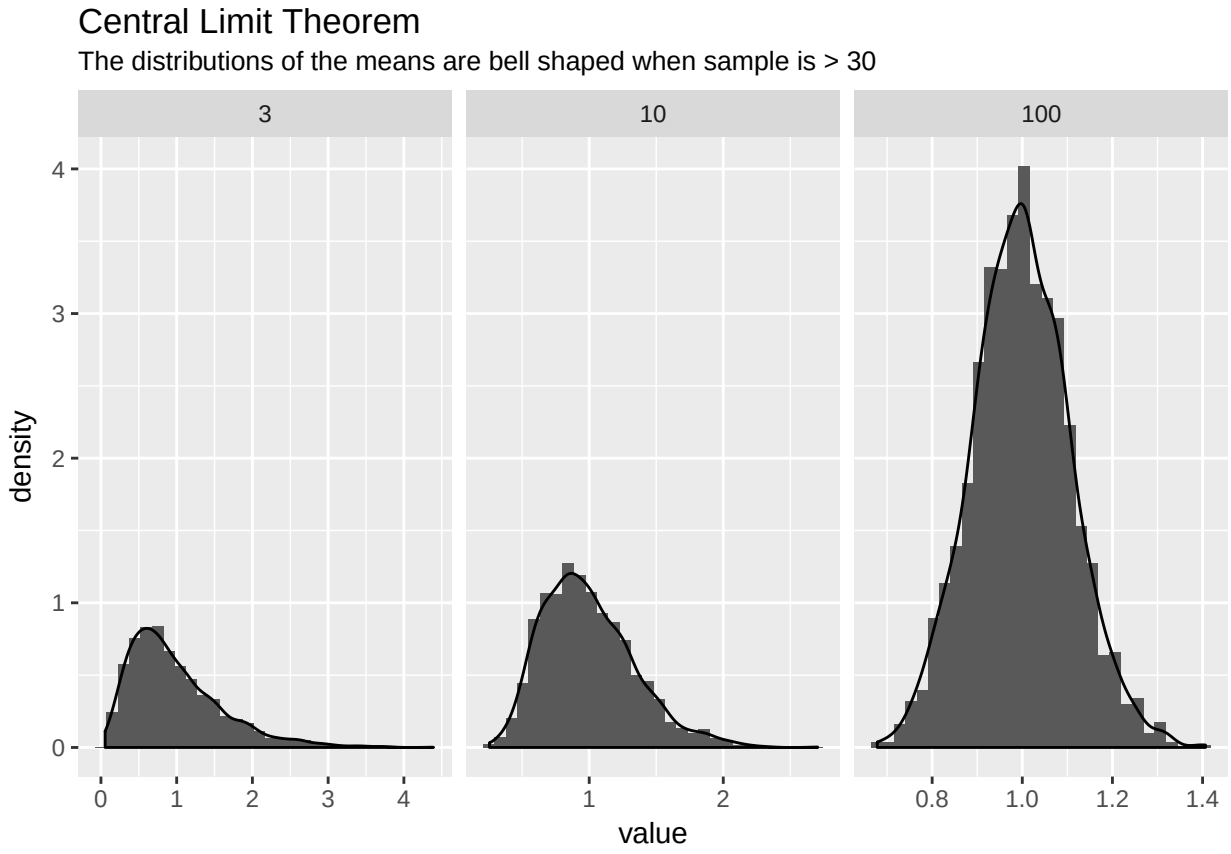
Normal Distribution

```r
df %>%
  filter(key=="rnorm") %>%
  ggplot() +
  geom_histogram(aes(x=value,y=..density..), position="identity") +
  geom_density(aes(x=value,y=..density..)) +
  facet_grid(.~n, scales = "free") +
  ggtitle('Central Limit Theorem',
    subtitle = 'The distributions of the means seem to always be bell shape')
```



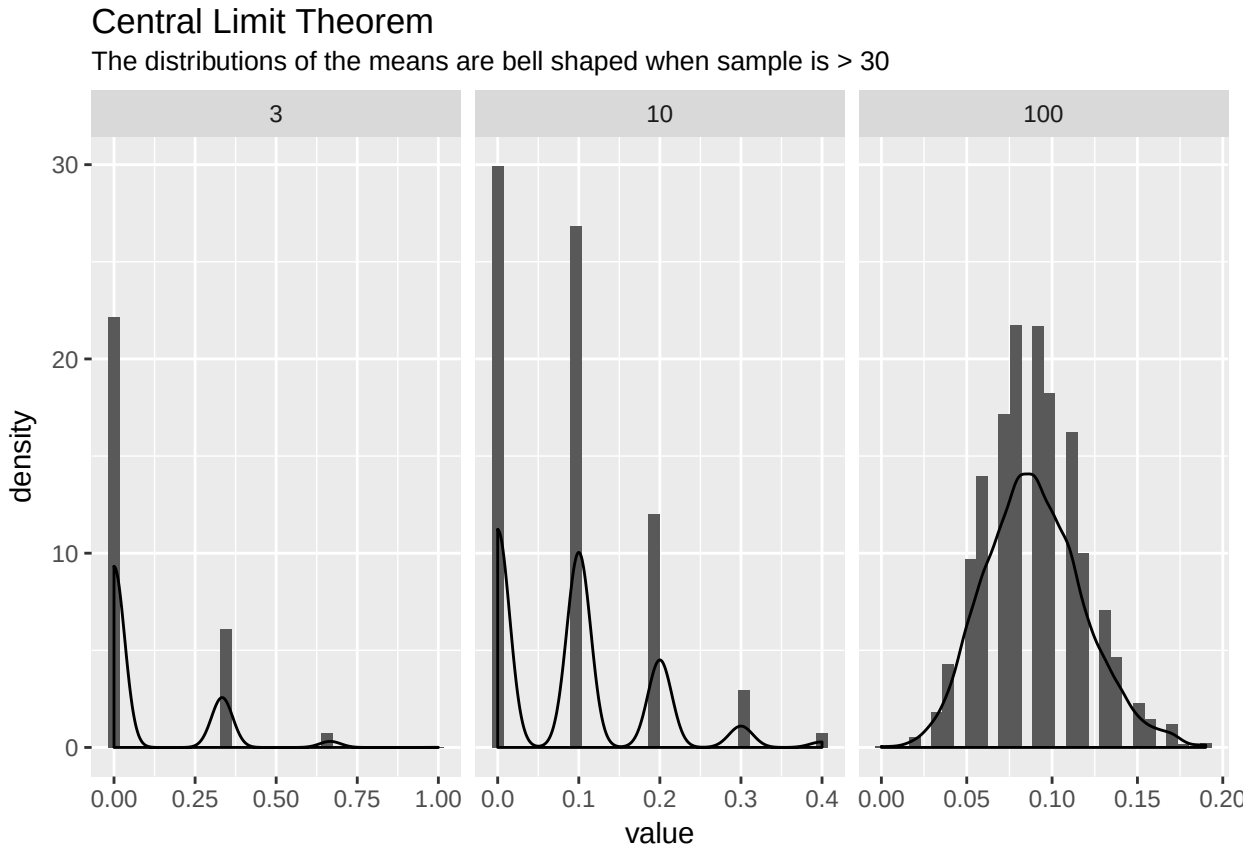
Exponential Distribution

```
df %>%
  filter(key=="rexp") %>%
  ggplot() +
  geom_histogram(aes(x=value,y=..density..), position="identity") +
  geom_density(aes(x=value,y=..density..)) +
  facet_grid(.~n, scales = "free") +
  ggtitle('Central Limit Theorem',
    subtitle = 'The distributions of the means are bell shaped when sample is > 30')
```



Binomial Distribution

```
df %>%
  filter(key=="rbinom") %>%
  ggplot() +
  geom_histogram(aes(x=value,y=..density..), position="identity") +
  geom_density(aes(x=value,y=..density..)) +
  facet_grid(.~n, scales = "free") +
  ggtitle('Central Limit Theorem',
    subtitle = 'The distributions of the means are bell shaped when sample is > 30')
```


Central Limit Theorem
The distributions of the means are bell shaped when sample is > 30
3
10
100
density
30
20
10
0
0.00
0.25
0.50
0.75
1.00
0.0
0.1
0.2
0.3
0.4
0.00
0.05
0.10
0.15
0.20
value