20121865_HypothesisTesting

Louise Branzuela

2022-04-06

Setting my student ID as a seed for the pseudo random number generator:

```
set.seed(20121865)
```

Generating the mean:

```
rnorm(1,10,2)
```

```
## [1] 10.32528
```

Generating the standard deviation

```
sqrt(rchisq(1,df=2))
```

```
## [1] 0.8096723
```

```
mu = 10.32528
sd = 0.8096723
```

Gaussian Distribution

Minimum value: (3 standard deviations to the left of the mean)

```
mu - 3 * sd
```

```
## [1] 7.896263
```

Maximum value: (3 standard deviations to the right of the mean)

```
mu + 3 * sd
```

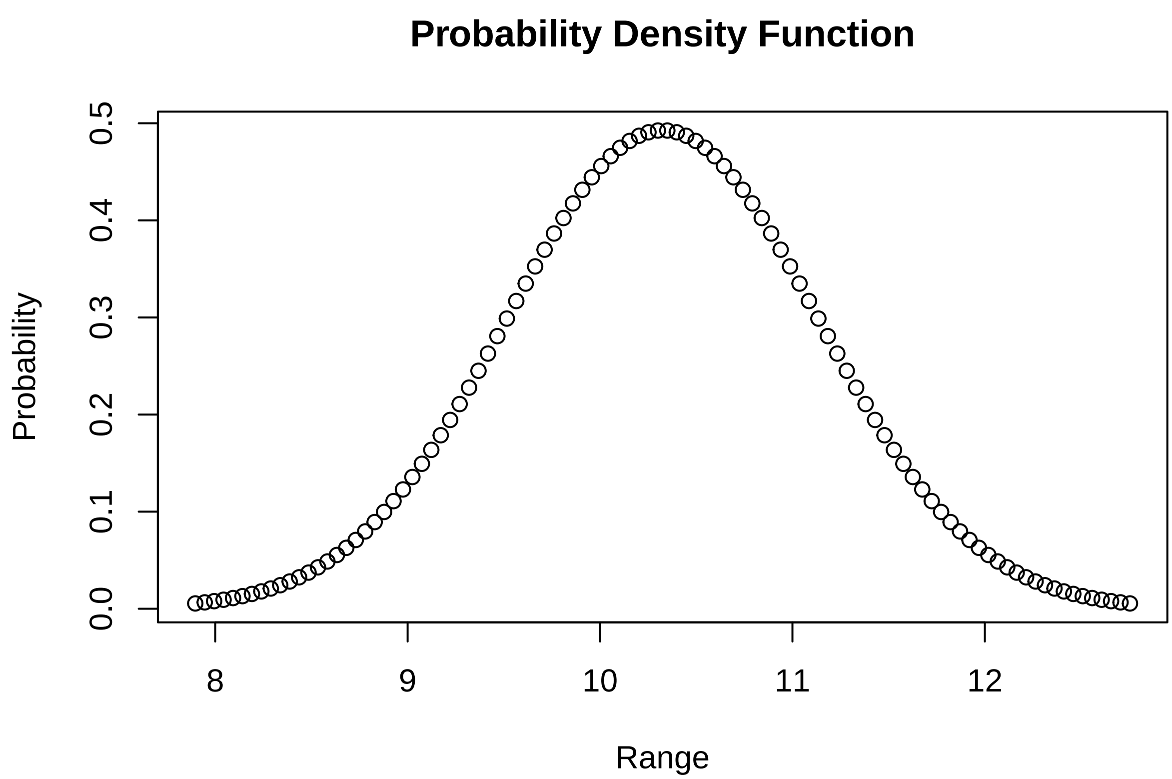
```
## [1] 12.7543
```

Setting the sequence to x

```
minimum = mu - 3 * sd
maximum = mu + 3 * sd
x <- seq(minimum, maximum, len = 100)
```

Probability density function for the distribution:

```
plot(x, dnorm(x, mu, sd),
     ylab = "Probability",
     xlab = "Range",
     main = "Probability Density Function")
```



#

Find the probability that X lies in the range:

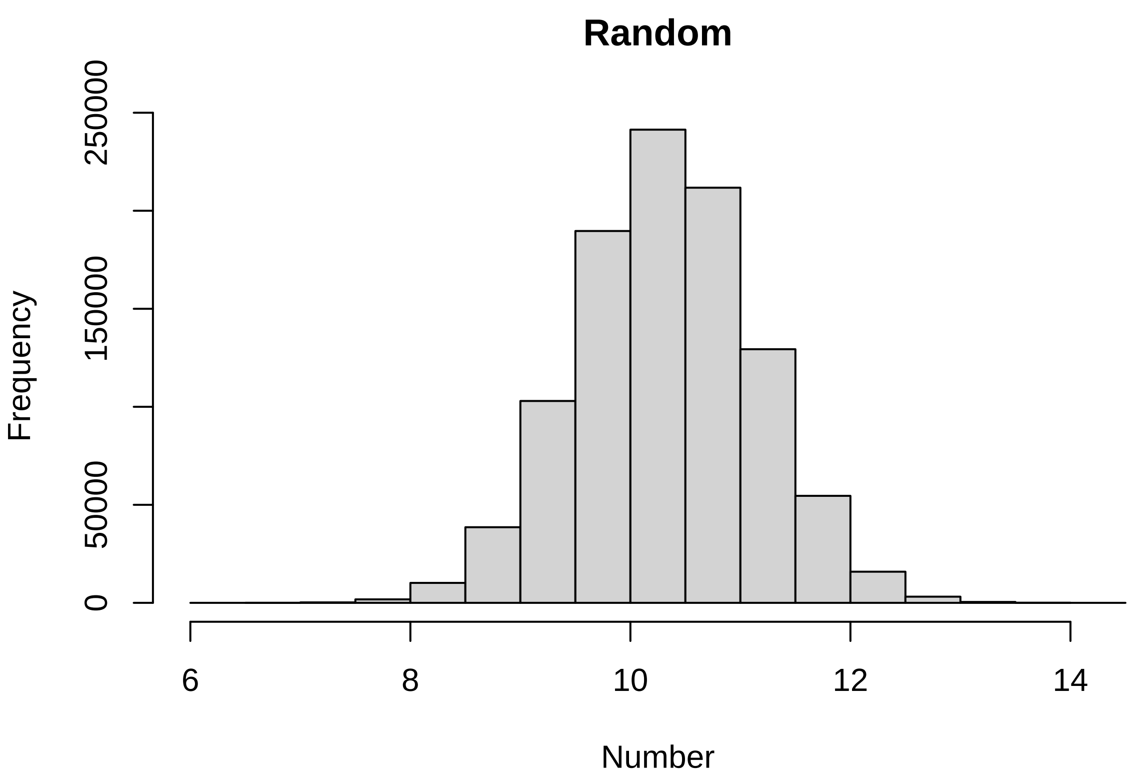
```
pnorm(mu + 2 * sd, mu, sd) - pnorm(mu - sd, mu, sd)
```

```
## [1] 0.8185946
```

Using rnorm() to generate a random sample:

```
r = rnorm(1e6, mu, sd)
```

```
hist(r,
     ylab = "Frequency",
     xlab = "Number",
     main = "Random")
```



Finding the steepest point of the standard normal distribution:

Theoretically, the mean would be the highest steepest point of the normal distribution, since the distribution is perfectly symmetrical, it would mean the middle of the distribution would be the steepest point of the curve which corresponds to the mean.

Student T-Test

```r
w <- c(102.5, 102.9, 99.9, 101.9, 101.3, 98.9, 96.5, 97.7, 101.8,
       100.3)
```

A sensible null hypothesis:

A sensible null hypothesis would be H_0 = the average weight of the bananas which is 100 grams.

P-value and "more extreme" in this context:

```r
n = length(w)
SEM = sd(w)/sqrt(n)
1 - pnorm(mean(w), 100, SEM)
```

```
## [1] 0.2910448
```

The p-value is defined as the likelihood of receiving the observation or a more extreme observation if the null hypothesis is true. The p-value that I got was 0.2910448, in this dataset, "more extreme" means larger than the observed sample mean.

One-sided or two-sided test:

A one-sided test(right) is needed because of the fact that we are testing if the population mean is more than the null mean.

Performing a student t-test:

```r
t.test(w,mu=100,alternative="greater")
```

```
##
##  One Sample t-test
##
## data:  w
## t = 0.55034, df = 9, p-value = 0.2977
## alternative hypothesis: true mean is greater than 100
## 95 percent confidence interval:
##  99.13757      Inf
## sample estimates:
## mean of x
##    100.37
```

The t-test shows us that the p-value is 0.2977, which is different to the 0.2910448 that I got from calculating the p-value myself. This is because of the fact that the sample is too small which causes the uncertainty to be higher.

95% and 99% confidence intervals

```
t.test(w,mu=100,conf.level=0.95,alt="g")
```

```
##
##  One Sample t-test
##
## data:  w
## t = 0.55034, df = 9, p-value = 0.2977
## alternative hypothesis: true mean is greater than 100
## 95 percent confidence interval:
##  99.13757      Inf
## sample estimates:
## mean of x
##    100.37
```

```
t.test(w,mu=100,conf.level=0.99,alt="g")
```

```
##
##  One Sample t-test
##
## data:  w
## t = 0.55034, df = 9, p-value = 0.2977
## alternative hypothesis: true mean is greater than 100
## 99 percent confidence interval:
##  98.4731     Inf
## sample estimates:
## mean of x
##    100.37
```

The 95 percent confidence interval is 99.13757 and the 99 percent confidence interval is 98.4731.

Hypothesis Testing: Binomial Distribution

Probability that one flower is fertilized:

```
dbinom(1, 7, 0.4)
```

```
## [1] 0.1306368
```

Probability that three or more flowers are fertilized:

```
1 - pbinom(2, 7, 0.4)
```

```
## [1] 0.580096
```

A sensible null hypothesis:

A null hypothesis would be that $H_0 = 5$ since the scientists found a plant near a beehive to have 5 fertilized flowers.

P-value and "more extreme" in this context

The p-value is defined as the likelihood of receiving the observation or a more extreme observation if the null hypothesis is true. More extreme in this context means higher or lower than the sample mean.

One-sided or Two-sided:

I think a two-sided test is needed because it's not clear whether we can expect to see an increase or decrease in the actual observed mean.

Calculating the p-value and its significance:

```
mu = 7*0.4
sigma = sqrt(7*0.4*(1 - 0.4))
SEM = sigma/sqrt(7)
1 - pnorm(mu, 5, SEM)
```

```
## [1] 0.9999965
```

Since the p-value that I got was 0.9994665, I don't think the p-value is significant because it's not low enough to reject the null hypothesis.