Chapter 10

The Uniform Distribution

The uniform experiment

Select a number randomly from the interval $[a, b)$.

The uniform distribution is the continuous equivalent of the equiprobable probability model on a discrete sample space.

Probability Density Function

A continuous **uniform** random variable on the interval $[a, b)$ is the random variable with PDF

$$f(x) = \frac{1}{b-a}, \quad a \leq x < b.$$

We will denote this by $X \sim Unif(a, b)$.

Example 10.1

Let $X \sim \mathit{Unif}(0, 10)$.

a. Find $P(3 \leq X < 7)$.

Example 10.1

Let $X \sim \mathit{Unif}(0, 10)$.

b. The median of a continuous random - or 50th percentile - of a continuous random variable is the number q such that

$$P(X < q) = P(X > q) = \frac{1}{2}.$$

Find the median of X.

Uniform PDF calculations in R

```r
dunif(x=3, min=0,max=10)     #f(x)
## [1] 0.1

punif(q = 3, min = 0, max = 10)     #P(X <= 3), P(X < 3)
## [1] 0.3

punif(q = 7, min = 0, max = 10)     #P(X < = 7), P(X < 7)
## [1] 0.7

qunif(p = 0.5, min = 0, max = 10 )    #pth percentile
## [1] 5
```

Fitting a PDF to data

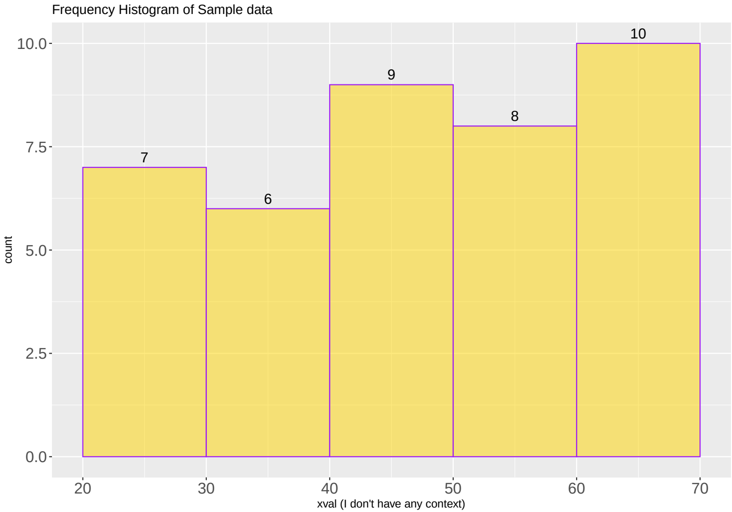
Suppose we have reason to believe that these forty xi 's may be a random sample from a uniform probability function defined over the interval $[20, 70)$ - that is,

$$f(x) = \frac{1}{50} \quad 20 \leq x < 70.$$

```r
sample_data <- tibble(
    xval  = c( 33.8, 62.6, 42.3, 62.9, 32.9, 58.9, 60.8, 49.1,
          42.6, 59.8, 41.6, 54.5, 40.5, 30.3, 22.4, 25.0,
          59.2, 67.5, 64.1, 59.3, 24.9, 22.3, 69.7, 41.2,
          64.5, 33.4, 39.0, 53.1, 21.6, 46.0, 28.1, 68.7,
          27.6, 57.6, 54.8, 48.9, 68.4, 38.4, 69.0, 46.6 ) )
```

How can we appropriately draw the distribution of the xi 's and the uniform probability model on the same graph to assess the *fit*?

We would begin by constructing a histogram of the data.



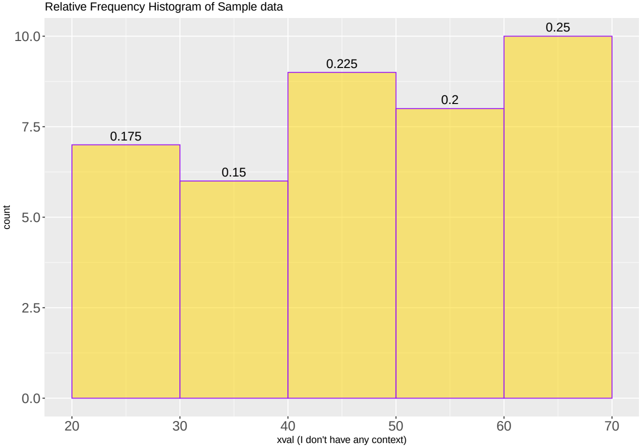
Note, first, that the uniform PDF $f(x)$ and the histogram are not compatible in the sense that the area under $f(x)$ is (necessarily) 1, but the sum of the areas of the bars making up the histogram is 400:

$$7 \times 10 + 6 \times 10 + 9 \times 10 + 8 \times 10 + 10 \times 10$$

A first idea to make them compatible is to re-scale the y axis to be a relative frequency:

$$\text{relative freq} \; = \frac{\text{freq}}{n}$$

where n is the number of observations in the dataset: $n = 40$



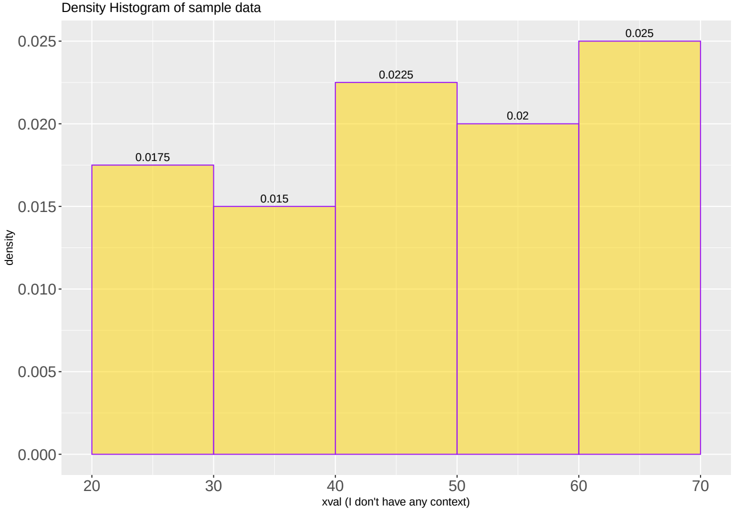
The histogram of the data is still not compatible with a uniform PDF because the total area of the blocks is still not 1.

To make them compatible, we need to use

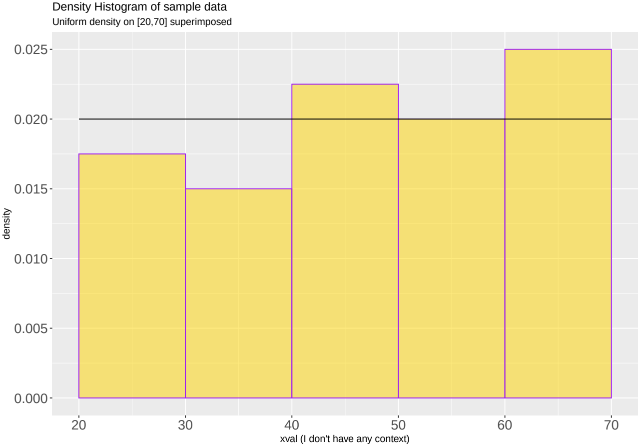
$$\text{density (of a bin)} = \frac{\text{relative frequency}}{\text{width}}$$

on the y axis.

The histogram of the sample data on the density scale is shown below.



And here is the density histogram of the data with the uniform PDF overlaid. Just eyeballing, is the uniform model a good fit?



Code for making the density histogram of the data with uniform density overlaid

```
## Binwidth: A rough rule of thumb for picking the binwidth is to use the range (max - min) of the data
## divided by log_2(n) + 1 where n is number of obs.


ggplot() +
  geom_histogram(data = sample_data,
                 mapping = aes(x = xval,
                               y = after_stat(density)),
                 breaks=seq(20,70,10),
                 alpha = 0.5,
                 color = "purple",
                 fill = "gold" )+
  geom_function(fun = dunif,
                args = list(min = 20, max = 70 ),
                xlim = c(20,70) ) +
  labs(x = "xval (I don't have any context)",
       title = "Density Histogram of sample data ",
       subtitle = "Uniform density on [20,70] superimposed")+
  theme(axis.text=element_text(size=15))
```