Minilab 7a Worksheet

Probability

Probability plays a fundamental part in analysis of data and understanding of statistical models (including probability distributions and linear models). In this minilab, we look at working with probability distributions in R.

R is a language for statistical computing. So R knows all about probability distributions, both *discrete* distributions (such as the Geometric and Binomial distributions) and *continuous* distributions (such as the Uniform and Normal distributions).

Make sure you know the difference between discrete and continuous variables (see https://keydifferences.com/difference-between-discrete-and-continuous-variable.html).

1. Uniform Probability Distribution in R

Suppose you have a circular spinner which gives values as *bearings* (degrees clockwise from North) in the interval $[0,360]$ of the real numbers.

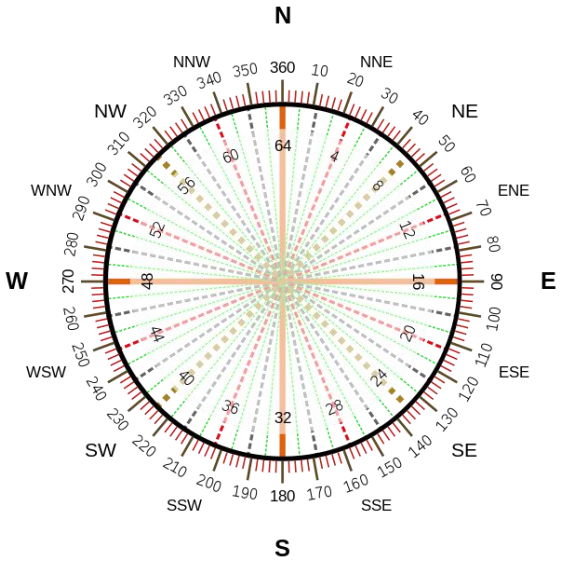
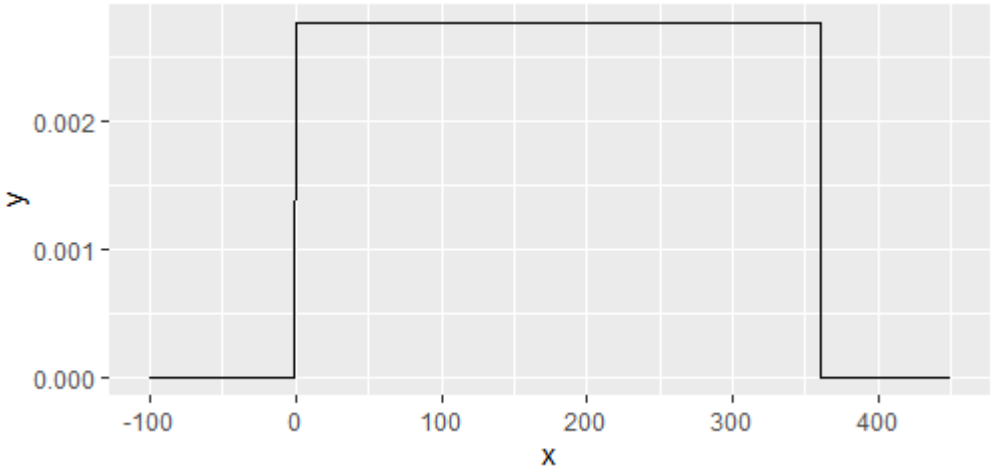


Image from Wikipedia

We say that the bearing (call it X, a random variable) follows a *continuous uniform* distribution, i.e., $X \sim U(0,360)$. We use the `dunif()` function to plot the probability density function (pdf). The function `qplot()` is a shortcut for quickly drawing a plot in R (using ggplot2).

```
library(tidyverse)
# Plot of pdf of U(0,360)
a = 0
b = 360
x = -100:450
y = dunif(x,a,b)
qplot(x,y,geom="line")
```



What is the probability that the spinner lands in the range from 45 to 90 degrees? For a continuous random variable, we are interested in $P(X \leq x)$ for some value of x, so we must be using `punif()` to calculate probabilities.

```
# If X~U(0,360), find P(45<=X<=90)
punif(90,0,360)-punif(45,0,360)
```

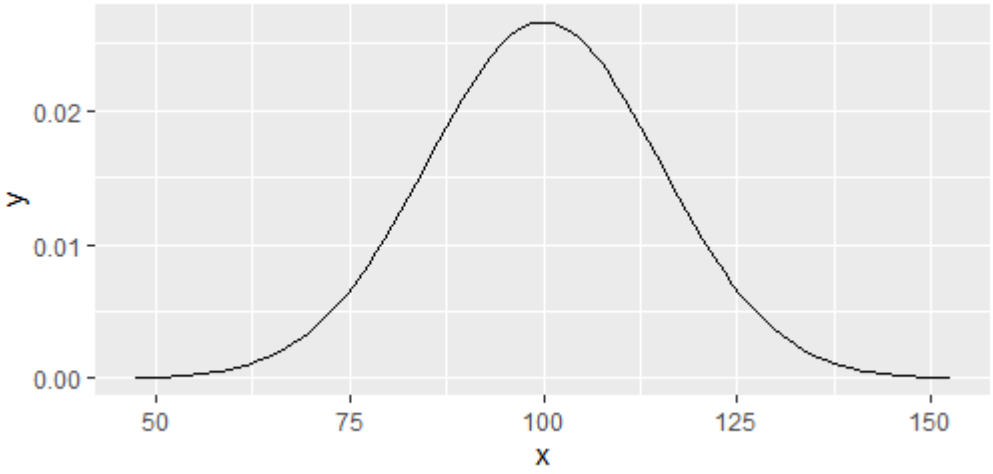
This gives the answer: 0.125.

Exercise. What is the probability that the spinner lands in the range from 315 to 45 degrees? *Be careful.*

2. Normal Uniform Probability Distribution in R

Intelligence Quotient (IQ) is a score derived from a set of standardised tests designed to assess human intelligence. The IQ of people in a population (call it X, a random variable) follows a *Normal distribution* with a mean score of 100 and a standard deviation of 15. We can use `dnorm()` to plot the pdf (the familiar bell-shaped curve) with the peak at 100 (the mean of the distribution).

```
# Normal distribution in R
mu = 100
sigma = 15
x = (mu-3.5*sigma):(mu+3.5*sigma)
y = dnorm(x,mu,sigma)
qplot(x,y,geom="line")
```



(1) What is the probability of randomly selected person having an IQ between 120 and 140? We use `pnorm()` to calculate probabilities.

```
pnorm(140,mu,sigma)-pnorm(120,mu,sigma)
```

This gives the answer: 0.0874 (4dp).

(2) What IQ value is such that 99% of people have IQ less-than-or-equal-to that value? We use `qnorm()` to calculate this value.

```
qnorm(0.99,mu,sigma)
```

This gives the answer: 134.9 (1dp).

(3) It is claimed that a person with an IQ over 140 is a genius. The article "8 People with Higher IQs than Einstein" (see https://www.rd.com/list/highest-iq-in-the-world/) describes a few people with very high IQ scores.

Exercise. Without using any R, what do you think the value of `qnorm(0,100,15)` will be? What about `qnorm(1,100,15)`? Confirm your answers using R.

Challenge. The function `qnorm(p,mu,sigma)` is used to find the value x such that $P(X \leq x) = p$ where the given p is the first argument to the `qnorm` function. Use R and `qplot` to draw a graph where the horizontal axis is values of p and the vertical axis is the corresponding values of the qnorm function for the distribution of IQ scores. You will want to use the `seq` function (to get help type `?seq` in the R console).

Summary

In this minilab, we have looked at the Uniform and Normal continuous probability distributions in R.