Data Science for Doctors – Part 3 : Distributions

Data science enhances people's decision making. Doctors and researchers are making critical decisions every day. Therefore, it is absolutely necessary for those people to have some basic knowledge of data science. This series aims to help people that are around medical field to enhance their data science skills.

This is the third part of the series, it will contain the main distributions that you will use most of the time. This part is created in order to make sure that you have (or will have after solving this set of exercises) the knowledge for the next parts to come. The distributions that we will see are:

1)Binomial Distribution: The binomial distribution fits to repeated trials each with a dichotomous outcome such as success-failure, healthy-disease, heads-tails.

2)Normal Distribution: It is the most famous distribution, it is also assumed for many gene expression values.

3)T-Distribution: The T-distribution has many useful applications for testing hypotheses when the sample size is lower than thirty.

4)Chi-squared Distribution: The chi-squared distribution plays an important role in testing hypotheses about frequencies.

5)F-Distribution: The F-distribution is important for testing the equality of two variances.

Before proceeding, it might be helpful to look over the help pages for the choose, dbinom, pbinom , rbinom, qbinom,pnorm, qnorm, rnorm, dnorm,pchisq, qchisq, dchisq, df, pf, df.

Answers to the exercises are available here.

If you obtained a different (correct) answer than those listed on the solutions page, please feel free to post your answer as a comment on that page.

Exercise 1

Let X be binomially distributed with n = 100 and p = 0.3.Compute the following:
a) $P(X = 34)$, $P(X \geq 34)$, and $P(X \leq 34)$
b) $P(30 \leq X \leq 60)$
c) The quantiles $x_{0.025}$, and $x_{0.975}$

Exercise 2

Let X be normally distributed with mean = 3 and standard deviation = 1.Compute the following:
a) $P(X\ 2)$,$P(2 \leq X \leq 4)$
b) The quantiles $x_{0.025}$, $x_{0.5}$and $x_{0.975}$.

Exercise 3

Let T_8 distribution.Compute the following:
a)$P(T_8 < 1)$, $P(T_8 > 2)$, $P(-1 < T_8 < 1)$.
b)The quantiles $t_{0.025}$, $t_{0.5}$, and $t_{0.975}$. Can you justify the values of the quantiles?

Exercise 4

Compute the following for the chi-squared distribution with 5 degrees of freedom:

a) $P(X^2_5<2)$, $P(X^2_5>4)$,$P(4<X^2_5<6)$.
b) The quantiles $g_{0.025}$, $g_{0.5}$, and $g_{0.975}$.

Exercise 5

Compute the following for the $F_{6,3}$ distribution:

a)$P(F_{6,3} < 2)$, $P(F_{6,3} > 3)$, $P(1 < F_{6,3} < 4)$.

b)The quantiles $f_{0.025}$, $f_{0.5}$, and $f_{0.975}$.

Exercise 6

Generate 100 observations following binomial distribution and plot them(if possible at the same plot):

a) n = 20, p = 0.3

b) n = 20, p = 0.5

c) n = 20, p = 0.7

Exercise 7

Generate 100 observations following normal distribution and plot them(if possible at the same plot):

a) standard normal distribution (N(0,1))

b) mean = 0, s = 3

c) mean = 0, s = 7

Exercise 8

Generate 100 observations following T distribution and plot them(if possible at the same plot):

a) df = 5

b) df = 10

c) df = 25

Exercise 9

Generate 100 observations following chi-squared distribution and plot them(if possible at the same plot):

a) df = 5

b) df = 10

c) df = 25

Exercise 10

Generate 100 observations following F distribution and plot them(if possible at the same plot):

a) $df_1 = 3$, $df_2 = 9$

b) $df_1 = 9$, $df_2 = 3$

c) $df_1 = 15$, $df_2 = 15$