Probability: Random Variables in SPSS

In this document we show how to find probabilities of Binomial and Normal random variables in SPSS.

Binomial Random Variable

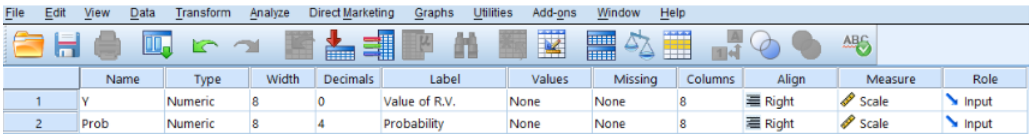
We start with this example: *Suppose in one hospital 10 babies are born each day. The probability that a new born baby is a Boy is 50%. We are interested in the distribution of the number of boys born each day in this hospital.*
Let

Y = Number of Boys born on a day

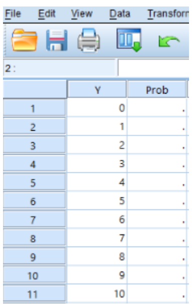
We know that the distribution of Y is Binomial with $n = 10$ and $p = 0.50$

$$Y \sim Bin(10, 0.5)$$

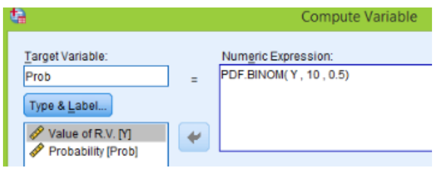
Let's find 'Probability Mass Function' for this random variable. Open a new SPSS data file, go to **Variable View** section, and enter these two variables:



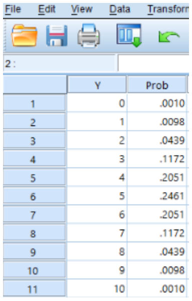
Then in the **Data View** section enter the values Y can take ($Bin(10, 0.5)$ can be $0, 1, 2 \ldots 10$)



To find probabilities we can use a 'calculator' in SPSS, go to **Transform** > **Compute Variable** and enter the following expression:



'PDF.BINOM' calculates probabilities of binomial random variable. Click **OK** and we get probabilities:

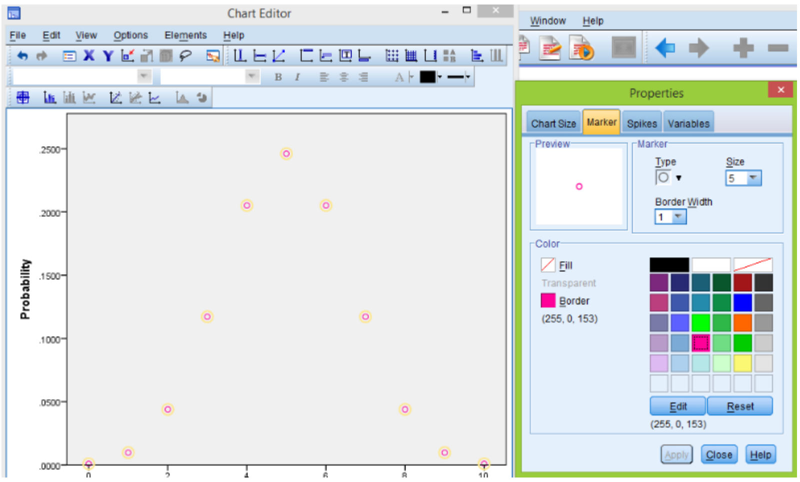


	Y	Prob
1	0	.0010
2	1	.0098
3	2	.0439
4	3	.1172
5	4	.2051
6	5	.2461
7	6	.2051
8	7	.1172
9	8	.0439
10	9	.0098
11	10	.0010

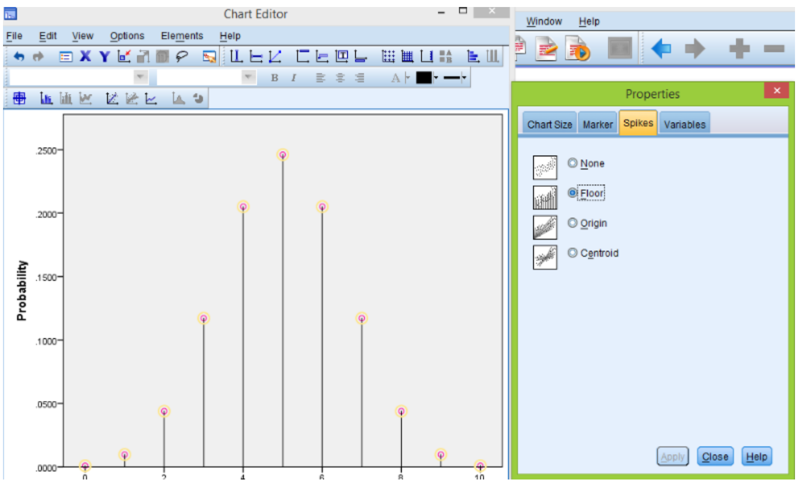
From this result we can say that, for example the probability that 7 boys are born on a given day is 0.1172:

$$P(Y = 7) = 0.1172$$

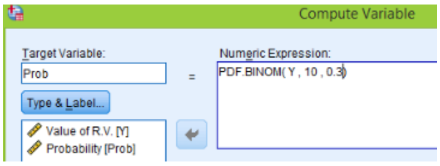
To plot the ‘Probability Mass Function’, go to **Graph > Chart Builder** select **Scatter/Dot > double click on Simple Scatter**. Drag ‘Y’ variable to X-Axis and ‘Prob’ to Y-Axis, click **OK**. The plot is produced, to modify it double click on it to open ‘Chart Editor’, double click on any point on the graph and ‘Properties’ window appears:



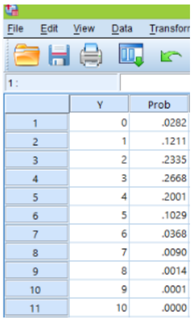
Here we can change color and shapes of the points, and then we click on ‘Spikes’ and select ‘Floor’ to produce vertical bars



If $Y \sim Bin(10, 0.3)$ (probability is different) we can find probabilities by changing 0.5 to 0.3 in ‘Compute Variable’ window:

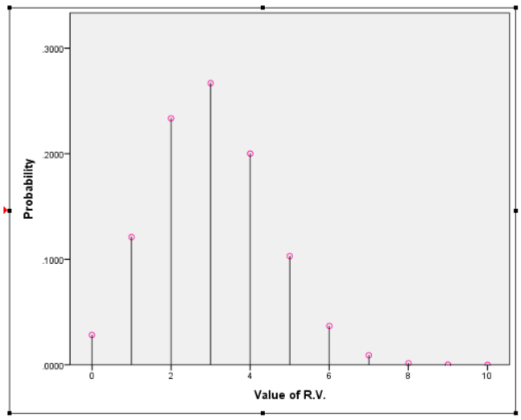


The new probabilities are produced:



	Y	Prob
1	0	.0282
2	1	.1211
3	2	.2335
4	3	.2668
5	4	.2001
6	5	.1029
7	6	.0368
8	7	.0090
9	8	.0014
10	9	.0001
11	10	.0000

If we make plot of 'Prob' versus 'Y' we get:

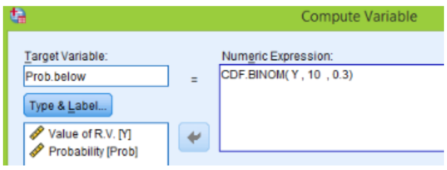


Observe that distribution is no longer symmetric but shifted to the left.

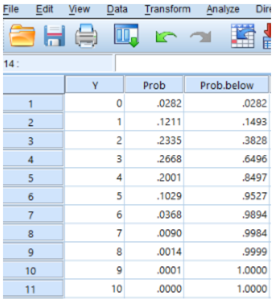
Suppose for the $Y \sim Bin(10, 0.3)$ we want to find probability of being less than or equal to 2. Then we can just add probabilities from the above table:

$$P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2) = 0.0282 + 0.1211 + 0.2335 = 0.3828$$

Instead of finding this probability by adding, we can go to **Transform** > **Compute Variable** and enter the next expression:



The target variable we call 'Prob.below' and 'CDF.BINOM' calculates probability below values of 'Y'. Click **OK** and we get:



	Y	Prob	Prob.below
1	0	.0282	.0282
2	1	.1211	.1493
3	2	.2335	.3828
4	3	.2668	.6496
5	4	.2001	.8497
6	5	.1029	.9527
7	6	.0368	.9894
8	7	.0090	.9984
9	8	.0014	.9999
10	9	.0001	1.0000
11	10	.0000	1.0000

Now we immediately see that $P(Y \leq 2) = 0.3828$ without adding probabilities by hand.

Normal Random Variable

Consider the next problem:

Suppose the birth weights of the infants born at a hospital follow a Normal distribution with mean of 3700 g and a standard deviation of 350 g.

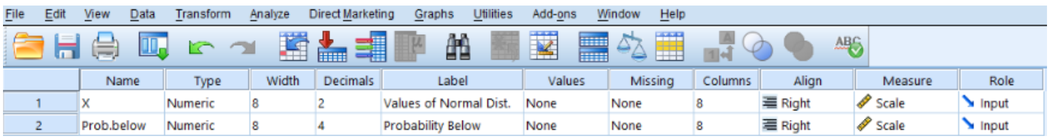
We will work out the answers to these questions:
1. What percent of infants born at the hospital weigh less than 3000 grams?
2. What percent of infants weigh between 4000 and 4500 grams?
3. What birth weight is the first quartile?

Let X be a birth weight of a infant then we know:

$$X \sim N(3700, 350)$$

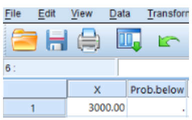
Which means X has normal distribution with mean 3700g and standard deviation 350g.

1. What percent of infants born at the hospital weigh less than 3000 grams?
In this question we need to find $P(X < 3000)$. To do that in SPSS open a new data file and in the **Variable View** section we construct two variables:



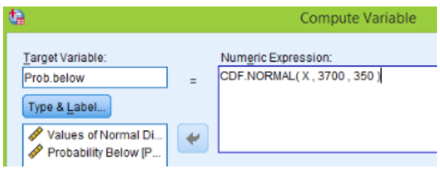
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	X	Numeric	8	2	Values of Normal Dist.	None	None	8	Right	Scale	Input
2	Prob.below	Numeric	8	4	Probability Below	None	None	8	Right	Scale	Input

'X' variable corresponds to the values our normal random variable can be and 'Prob.below' will be probabilities below 'X' values. Now in the **Data View** section enter 3000 in 'X' variable:

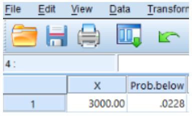


	X	Prob.below
1	3000.00	.

To calculate probability to the left of 3000, go to **Transform** > **Compute Variable** and enter the following expression:



Click **OK** and we get the result:



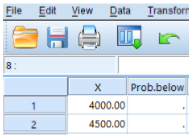
So we have the answer:

$$P(X < 3000) = 0.0228$$

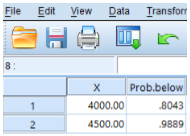
2. What percent of infants weigh between 4000 and 4500 grams?
Here we need $P(4000 \leq X < 4500)$. Observe that we can find this probability by calculating $P(X < 4000)$ and $P(X < 4500)$ since

$$P(4000 \leq X < 4500) = P(X < 4500) - P(X < 4000)$$

We can find these two probabilities as in the first question. First enter 4000 and 4500 to the 'X' variable:



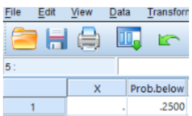
Now as before in the calculator we enter the expression and get:



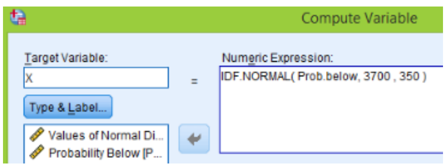
So we get the answer:

$$P(4000 \leq X < 4500) = P(X < 4500) - P(X < 4000) = 0.9889 - 0.8043 = 0.1846$$

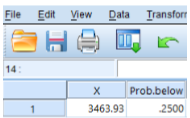
3. What birth weight is the first quartile?
Here we have an opposite problem, we need x such that $P(X < x) = 0.25$.
To find this value, enter 0.25 in 'Prob.below' variable:



As before go to calculator (**Transform** > **Compute Variable**),and type:



Note that instead of 'CDF.NORMAL' we use 'IDF.NORMAL' function. Click **OK**:



So we get

$$P(X < 3463.93) = 0.25$$

And hence the first quartile is about 3464g.