

# **Chapter 32**

## **Simulation Models for Varying Predictors (9,000 Patients)**

### **General Purpose**

In medicine predictors are often varying, like, e.g., the numbers of complications and the days in hospital in patients with various conditions. This chapter is to assess, whether Monte Carlo simulation of the varying predictors can improve the outcome predictions.

### **Specific Scientific Question**

The hospital costs for patients with heart infarction are supposed to be dependent on factors like patients' age, intensive care hours (ichours), numbers of complications. What percentage of patients will cost the hospital over 20,000 Euros, what percentage over 10,000. How will costs develop if the numbers of complications are reduced by 2 and the numbers of ichours by 20.

---

This chapter was previously published in "Machine learning in medicine-cookbook 2" as Chap. 8, 2014.

**Instead of Traditional Means and Standard Deviations, Monte Carlo Simulations of the Input and Outcome Variables are Used to Model the Data. This Enhances Precision, Particularly, With non-Normal Data**

Age Years	complication number	ic hours	costs Euros
48	7	36	5488
66	7	57	8346
75	7	67	6976
72	6	45	5691
60	6	58	3637
84	9	54	16369
74	8	54	11349
42	9	26	10213
71	7	49	6474
73	10	35	30018
53	8	37	7632
79	6	46	6538
50	10	39	13797

Only the first 13 patients of this 9000 patient hypothesized data file is shown. The entire data file is in [extras.springer.com](http://extras.springer.com) and is entitled "simulation1.sav". SPSS 21 or 22 can be used. Start by opening the data file. We will first perform a traditional linear regression with the first three variables as input and the fourth variable as outcome.

**Command:**

click Transform....click Random Number Generators....click Set Starting Point ....click Fixed Value (2000000)....click OK....click Analyze ...Regression.... Linear....Dependent: costs...Independent: age, complication, ichours....click Save....click Browse....Select the desired folder in your computer....File name: enter "exportsimulation"....click Save....click Continue....click OK.

In the output sheets it is observed that all of the predictors are statistically very significant. Also a PMML (predictive model markup language) document, otherwise called XML (eXtended Markup Language) document has been produced and filed in your computer entitled "exportsimulation".

Coefficients<sup>a</sup>

Model	Unstandardized coefficients		Standardized coefficients		Sig.
	B	Std. Error	Beta	t	
(Constant)	-28570,977	254,044		-112,465	,000
age(years)	202,403	2,767	,318	73,136	,000
complications(n)	4022,405	21,661	,807	185,696	,000
ichours(hours)	-111,241	2,124	-,227	-52,374	,000

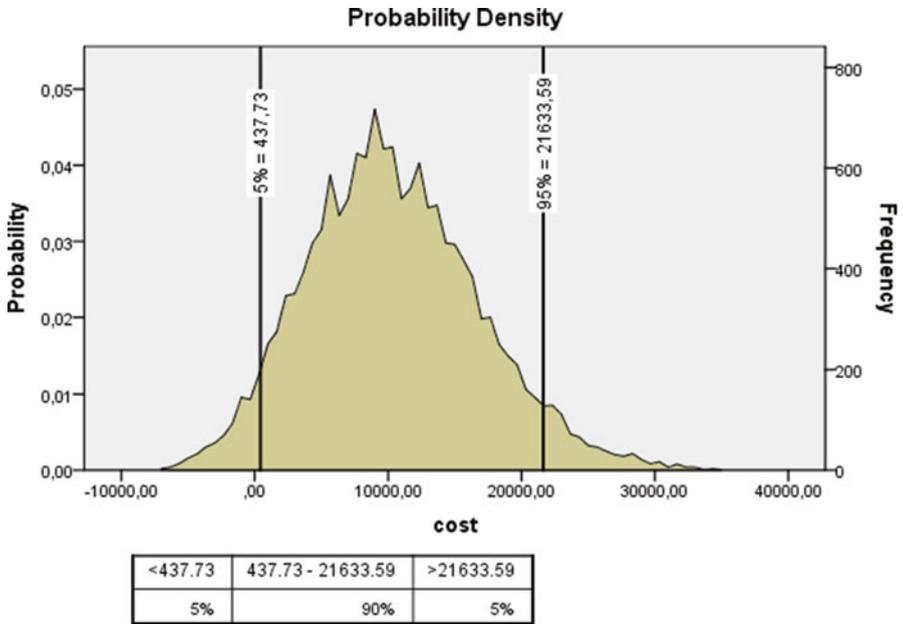
<sup>a</sup>Dependent variable: cost (Euros)

We will now perform the Monte Carlo simulation.

**Command:**

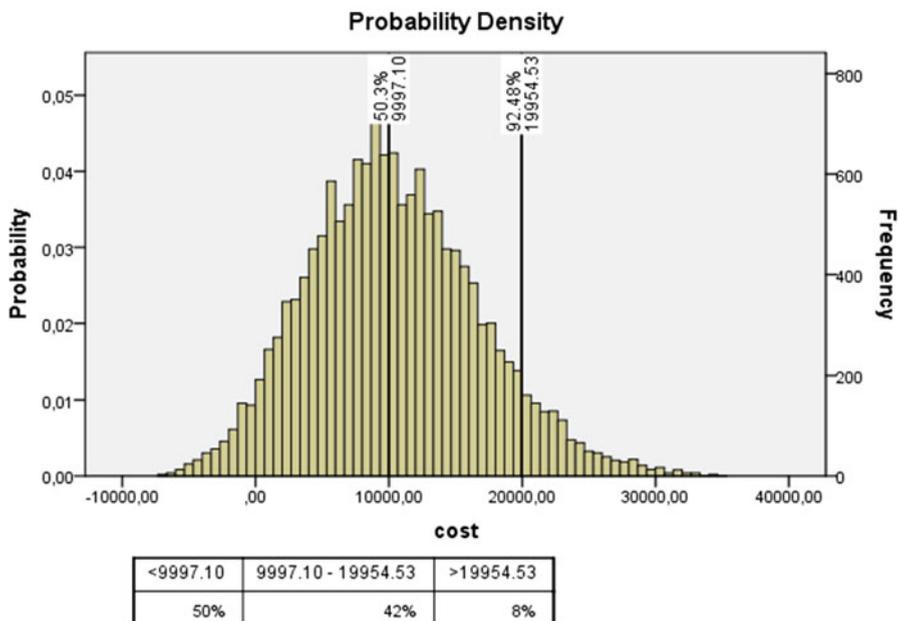
Analyze....Simulation....click Select SPSS Model File....click Continue....in Look in: select folder with "exportsimulation.xml" file....click Open....click Simulation Fields....click Fit All....click Save....mark Save the plan file for this simulation....click Browse....in Look in: select the appropriate folder for storage of a simulation plan document and entitle it, e.g., "splan"....click Save....click Run.

In the output the underneath interactive probability density graph is exhibited. After double-clicking the vertical lines can be moved and corresponding areas under the curve percentages are shown.



Overall 90 % of the heart attacks patients will cost the hospital between 440 and 21.630 Euros. In the graph click Chart Options....in View click Histogram....click Continue.

The histogram below is displayed. Again the vertical lines can be moved as desired. It can, e.g., be observed that, around, 7.5 % of the heart attack patients will cost the hospital over 20.000 Euros, around 50 % of them will cost over 10.000 Euros.



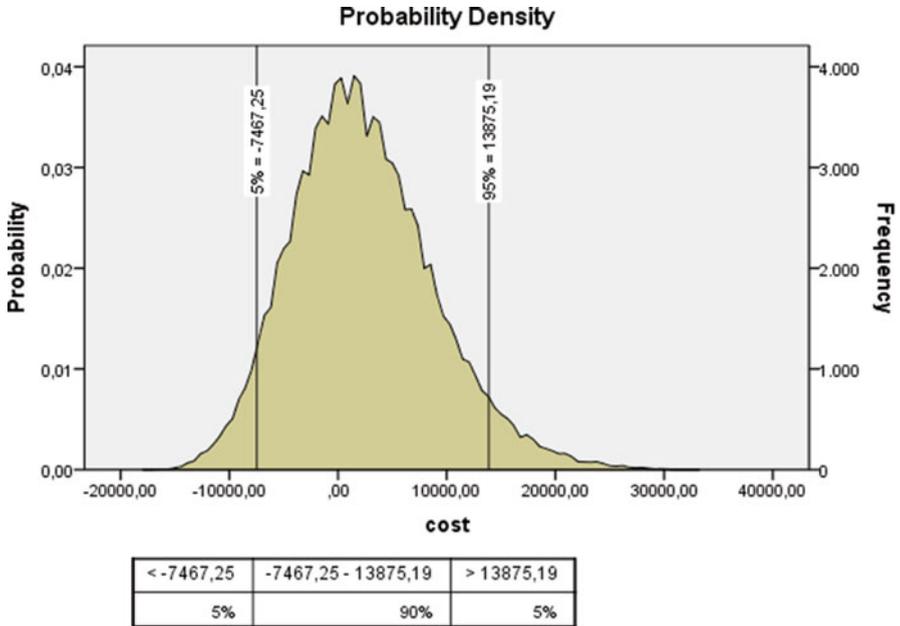
Monte Carlo can also be used to answer questions like "What will happen to costs, if the numbers of complications are reduced by two or the ichours are reduced by 20". For that purpose we will use the original data file entitled "chap8simulation1.sav" again. Also the document entitled "splan" which contains software syntax for performing a simulation is required.

#### **Open "simulation1.sav" and command:**

Transform....Compute Variable....in Numeric Expression enter "complications" from the panel below Numeric Expressions enter "-" and "2"....in Target Variable type complications....click OK....in Change existing variable click OK.

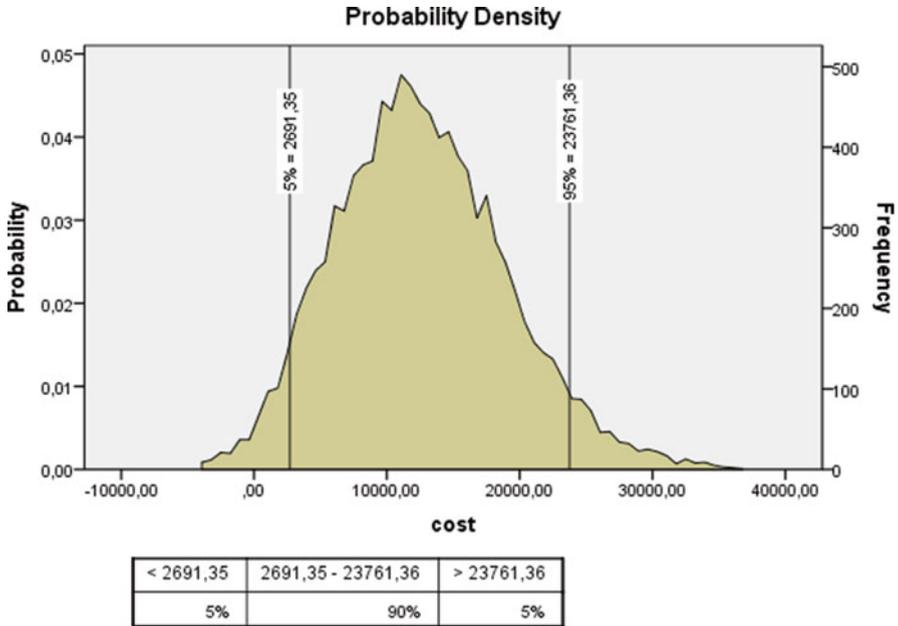
In the data file all of the values of the variable "complications" have now been reduced by 2. This transformed data file is saved in the desired folder and entitled e.g. "simulation2.sav". We will now perform a Monte Carlo simulation of this transformed data file using the simulation plan "splan".

In "simulation2.sav" command: Analyze....Simulation....click Open an Existing Simulation Plan....click Continue....in Look in: find the appropriate folder in your computer....click "splan.splan"....click Open....click Simulation....click Fit All....click Run.



The above graph shows that fewer complications reduces the costs, e.g., 5 % of the patients cost over 13.875 Euros, while the same class costed over 21.633 Euros before.

What about the effect of the hours in the ic unit. For that purpose, in "simulation1.sav" perform the same commands as shown directly above, and transform the ichours variable by -20 hours. The transformed document can be named "simulation3.sav" and saved. The subsequent simulation procedure in this data file using again "splan.splan" produces the underneath output.



It is observed that the costs are now not reduced, but rather somewhat increased with 5 % of the patients costing over 23.761 Euros instead of 21.633. This would make sense, nonetheless, because it is sometimes assumed by hospital managers that the reduction of stay-days in hospital is accompanied with more demanding type of care (Statistics Applied to Clinical Studies 5th edition, Chap. 44, Clinical data where variability is more important than averages, pp 487–498, Springer Heidelberg Germany, 2012).

### Conclusion

Monte Carlo simulations of inputs where variability is more important than means can model outcome distributions with increased precision. This is, particularly, so with non-normal data. Also questions, like “how will hospital costs develop, if the numbers of complications are reduced by 2 or numbers of hours in the intensive care unit reduced by 20”, can be answered.

## Note

More background, theoretical and mathematical information of Monte Carlo simulation is provided in *Statistics applied to clinical studies* 5th edition, Chap. 44, Clinical data where variability is more important than averages, pp 487–498, Springer Heidelberg Germany, 2012, from the same authors as the current publication.