**Sample Design – Exercise 1**

Your client wants you to build a model to forecast the likelihood that a potential credit customer will default in the first 18 months of being offered credit. Default is defined as missing 4 consecutive payments (being 120 days past due). Currently, your client accepts 60% of all applicants, and has a “BAD Rate” (rate of default) of 12%. Someone that has no delinquency is considered a satisfactory customer, while anyone with 1 or 2 months of non-payment is considered to be less than satisfactory, but not a major problem. A customer missing 3 payments is a serious concern, but has not yet reached the level of default on the loan.

**What is the goal of your project? What would be a successful model?**

To build a model to predict the risk of taking on a customer. Specifically, to predict the likelihood of default in the next 18 months. Such a model should predict well enough to improve on the current combination of *acceptance* and *bad* rates of 60% and 12% respectively.

**What is the dependent variable? What values will it take on?**

The dependent variable is the status of the customer (existing customer that is part of the sample).

Values:

0 = Bad (default, missed 4 payments)

1 = Good (current in payments)

2 = Intermediate1 (missed 1 or 2 payments)

3= Intermediate2 (missed 3 payments)

**What is the outcome period?**

The outcome period is 18 months.

**If you want data over a span of 2 years what dates represent the limits of the sample time frame?**

If the meeting occurred on Sep 1st, 2016, then the outcome period takes one back 18 months to March 1, 2015. A span of 2 years for the data would give us the sample time frame of **March 1 2013 to March 1, 2015**. In other words, a person who became a customer within that time frame (and continued being one for at least 18 months) is eligible to be in your sample, since you will have at least 18 months of data on him/her.

**Sample Design – Exercise 2**

Your client wants you to build a model to forecast the likelihood that a customer who is 30 days past due (has missed one payment) will go on to become 60 days past due (or worse) within the next 3 months. You are to build the model on the customers’ payment behavior data with the client (no external credit data will be used). You wish to use data from different months across a span of 2 years to ensure that seasonal variations do not affect your model.

**What is the goal of your project? What would be a successful model?**

To build a model to predict the likelihood of an increase in the level of delinquency for customers who are 30 days past due. The purpose of such predictions is to reduce the incidence of such increases in delinquency through more targeted action, thereby reducing costs.

**What is the dependent variable? What values will it take on?**

The dependent variable is the status of the customer that is 30 days late at a point in time, based on his payments over the 3 months after that point.

Values

0 = Bad (Rolled over to 60 days past due or worse)

1= Good (paid off enough to become current in payments)

2=Intermediate (paid off the minimum each month after to remain at 30 days past due)

The intermediate category may or may not be defined/used, depending on the client’s wishes.

**What is the outcome period?**

Three months.

**What dates represent the limits of the sample time frame?**

If the meeting occurred on Sep 1, 2016, then the outcome period takes us back to June 1, 2016. The sample time frame is thus **June 1st, 2014 – June. 1st, 2016.**In other words, any customer that was 30 days past due during that time period is eligible to be in the sample.