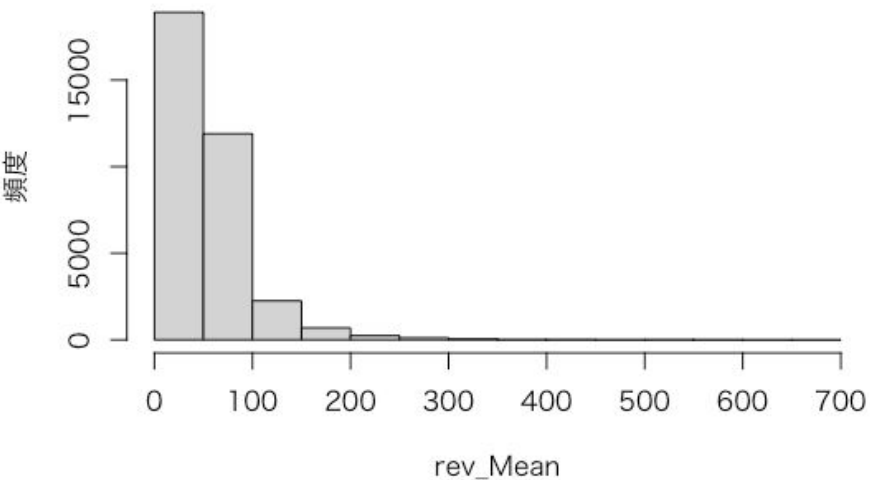


新規事業の提案

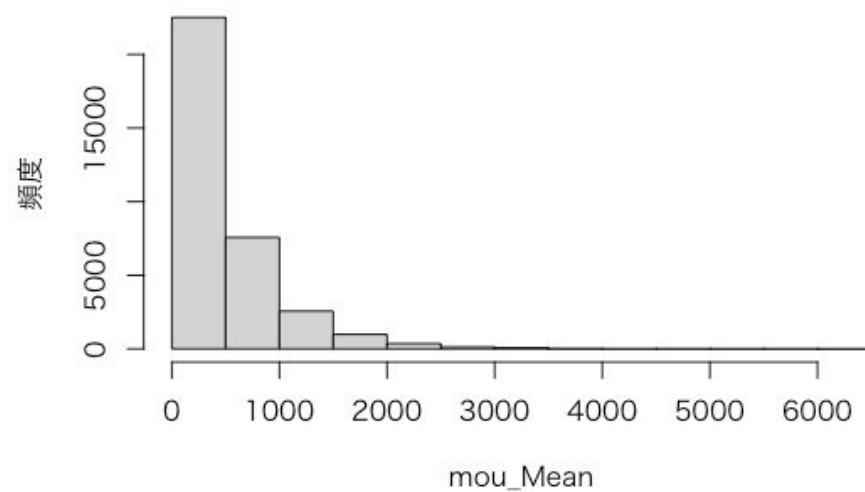
データ可視化

- ヒストグラムを観察すると、以下のように多くの連続変数の分布が偏っている。(以下のヒストグラムはデータの最初の3つの連続変数)
- log変換をして分布の偏りを解消。
- その後、連続変数に関しては標準得点(z値)を計算し、すべての連続変数を同じ単位に統一した。

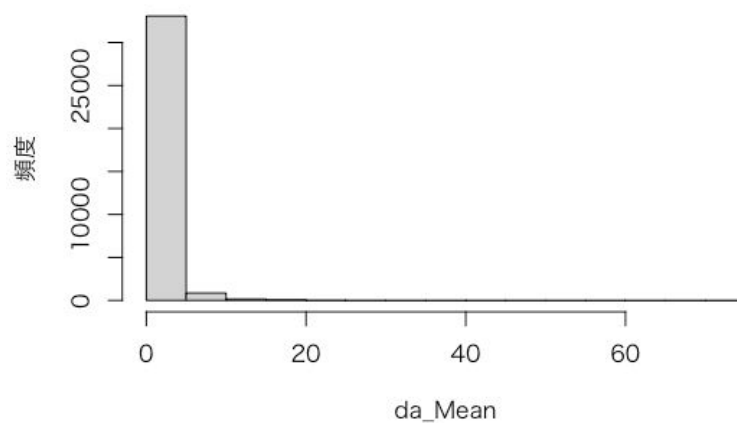
ヒストグラム：rev_Mean

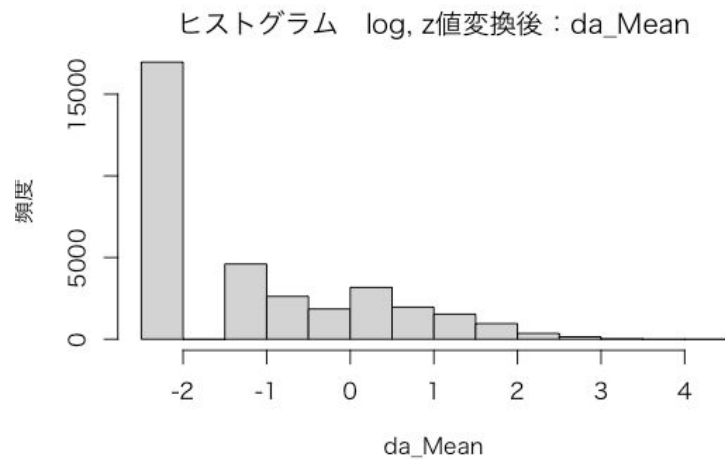
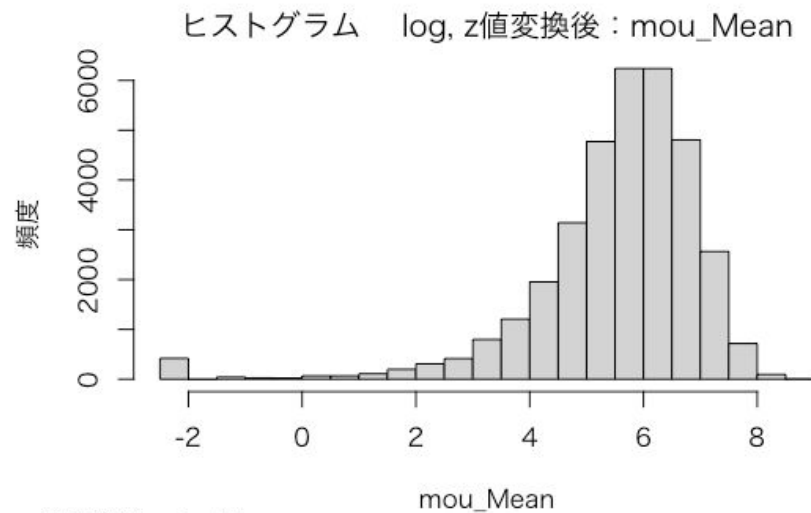
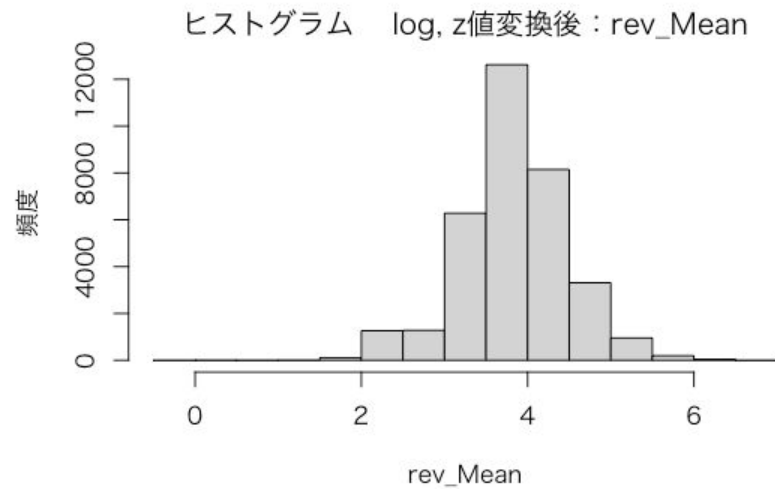


ヒストグラム：mou_Mean



ヒストグラム：da_Mean





解約者を引き止め、収益をあげる

- 顧客が解約しそうであるかどうかを顧客データに基づいて判断する。つまり、顧客データから、データ”churn”が1になるか（解約者）または0になるか（非解約者）を予測する。

解約者であると予測できる場合、彼らを引き止める提案をし、顧客をとどまらせることにより、収益をあげることを提案する。

解約者を予測するモデル

- ロジスティック回帰(ラッソ回帰)

- 通常のロジスティック回帰ではなく、ラッソ回帰を使ったのは、データにある変数の数が多いため。(元のデータは100個、カテゴリ変数をダミー変数に直すため最終的には 202個の変数)。ラッソ回帰を使うことで変数の数を自然と減らすことができる。

- ランダムフォレスト

- 変数が多く、直線的な関係性が解約者とデータの間にあるとは限らないため。例えば、今までサービスを使ってきた月数が多いからと言ってその顧客が解約しないとは限らない。ランダムフォレストのようなツリーモデルではこのような関係性があるかどうか、またそれがどのようなものかを教えてくれる。

データ処理、モデル選択

- データは70%を訓練データ、10%を検証データ、20%をテストデータにランダムに分割した。
- それぞれのモデルにおいて、交差検証(データをランダムに分割してモデルの精度を正確に測る方法)を使ってlog逸脱値(log deviance、モデルの精度の基準の一つ)が最も小さくなるモデルを選択した。

モデルのパフォーマンス

1. ロジスティック回帰

ロジック回帰の値		
	解約しない(実際)	解約した(実際)
解約しない(予測)	2534	1711
解約した(予測)	1039	1557

2. ランダムフォレスト

ランダムフォレストの値		
	解約しない(実際)	解約した(実際)
解約しない(予測)	2425	1485
解約した(予測)	1148	1783

モデルのパフォーマンス

- ランダムフォレストモデルの予測精度(正しく分類した数÷総数)は62%に対し、ロジスティック回帰の予測精度は58%であった。
- 予測精度の高いランダムフォレストモデルを以下の事業に使用する。

事業提案

- モデルが解約すると予測した顧客に対して次の月の20%の割引を提示する。
- これにより、顧客の解約を遅らせることができ、会社の収益を増やすことができる。
- 割引率に関しては、この事業の効果を見つつ変化させる。(顧客を引き止め、かつ収益が増える最適な割引率を今後見つける。)

事業概算

- データによると、一人の顧客が会社にもたらす売上高の月平均は55ドル。
- もし、顧客が解約をしないつもりであったがモデルが解約すると予測した場合、売上高は20%減る。つまり、顧客一人当たり平均で11ドル売上高が減ることになる。
- もし、顧客が解約をするつもりで、モデルがそれを正しく予測した場合、売上高は55ドルの8割、つまり44ドル一人当たり増えることになる。

事業概算

- 仮定:この割引の提示によって2割の顧客が解約を遅らせると期待できる。
- モデルの予測精度を計算する際に使ったテストデータを使用すると、11ドルの売上高減が1148人の顧客に対して起こり、次の月には12,628ドルの売上高減が期待される。
- 一方で、1783人の顧客の2割である、356人が解約を遅らせた場合、次の月は $356 \times 44 = 15,664$ ドルの売上高増が期待できる。

ランダムフォレストによる売上高の変化

	解約しない(実際)	解約する(実際)
解約しない(予測)	0	0.0
解約する(予測)	-12628	15690.4

事業概算

- これらを合わせると、次の月に3062.4ドルの売上高増が期待できる。
- テストデータは実際の顧客の20%なので、実際には約15300ドルの売上高増が期待できる。

まとめ、考察

- 事業を長く行うにつれて、効果が薄くなる可能性がある。(事業を行う最初の月は、顧客が割引を目新しく感じるため、解約をやめる顧客割合が高くなる可能性があるが、同じ20%割引を次の月も得られるとすると、顧客が割引後の値段に慣れてしまい、解約をする顧客が増える可能性がある。)

そのため、数ヶ月後には何か新しいプロモーションを代わりに考える必要があるかもしれない。

- また、実際に何割の顧客が、この割引に影響され解約をやめるかは割引を始めて見ないとわからないので、事業開始後は新しい顧客データを分析をし、割引率を上げ下げする必要がある。