検証列の作成（ホールドアウト法）

モデルを作成するためのサブセット（学習セット）とモデルの予測性能を評価するためのサブセット（検証セット）に、データを分割します。複数のモデルが作成されたときに、検証データで最も性能の高いモデルがしばしば選択されます。また、新しいデータ上でモデルの予測性能を評価するために、3 つ目のサブセット（テストセット）を用いることもあります。テストセットはモデルの作成にも選択にも用いられていないため、モデルの予測性能を評価する上で、この方法がより正確な方法であると考えられています。

データに過適合する傾向のあるモデルを作成する際に、検証列を使用すると特に有用です。JMP のいくつかのモデル作成用のプラットフォームでは、モデルをあてはめる時に検証データの割合を指定するオプションがあるため、検証列の作成は不要です。

JMP Pro での検証列の作成（学習、検証、テスト）

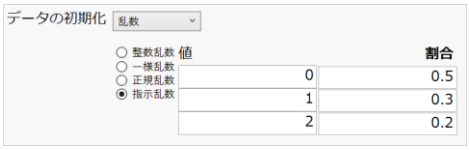
1. データテーブルで、**分析 > 予測モデル > 検証列の作成**を選択します。
2. 層別の列、グループの列、カットポイントの列を指定して、分割手法を選択できます。単純無作為抽出を用いて検証列を作成する場合は、**OK** をクリックします。
3. 表示されるウィンドウで、データを学習セット、検証セット、テストセットへ割り当て方を決めるために値（度数もしくは割合）を入力します。同じ無作為割付を再現する場合は、乱数シード値を指定します。

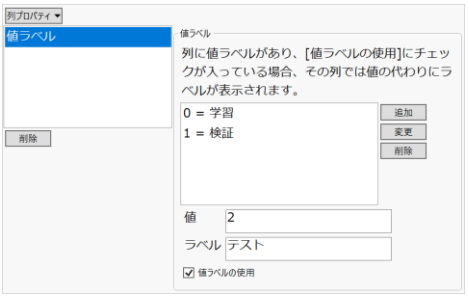
   新しい列が作成され、指定した割合（もしくは度数）で 0、1、2 という値が入力されます。

- 学習セットの 3,576 例（60%）は、モデルを構築（訓練）するのに用いられます。
- 検証セットの 1,788 例（30%）は、最良のモデルを検証して選択するのに用いられます。
- テストセットの 596 例（10%）は、新しいデータを用いて、選択したモデルの性能をテストするのに用いられます。

JMP での検証列の作成

1. データテーブルで、**列**メニューから**列の新規作成**を選択します。
2. **列の新規作成**ウィンドウで、**列名**を**検証**に変更します。
3. **データの初期化**の選択から**乱数**を選択します。
4. **指示乱数**を選択します。希望する割合を入力します。この例では、50%の 0（学習）、30%の 1（検証）、20%の 2（テスト）を選びます。
5. 0、1、2 に対して学習、検証、テストのラベルを割り当てて表示する場合、**列プロパティ**以下の**値ラベル**を選択し、それぞれの値にラベルを割り当てて**追加**をクリックします。
6. **適用**をクリックして、データテーブル上の新しい列を確認します（列が指定したように作成されたかを確認します）。その後 **OK** をクリックします。





注意: 検証列の作成の追加情報に関しては、**基本的な回帰モデル、予測モデルおよび発展的なモデル（ヘルプ > JMP ドキュメンテーションライブラリ**以下）をご覧いただくか、もしくは、JMP のヘルプを「検証」を検索してご確認ください。