

# Эффективность прогнозирования с помощью **metrica**



## ОСНОВЫ

**metrica** – это набор из более чем 80 функций, предназначенных для количественной и визуальной оценки эффективности прогнозирования регрессионных (непрерывных) и классификационных (категориальных) моделей точечного прогноза (например, APSIM, DSSAT, DNG, Supervised Machine Learning).

## Применение функций

Есть два основных аргумента, общих для всех функций **metrica**: (i) **obs** (O; наблюдаемые, они же фактические, измеренные, истинные, целевые, маркированные) и (ii) **pred** (P; предсказанные, они же смоделированные, предполагаемые) значения. Необязательные аргументы включают данные, которые позволяют вызвать существующий фрейм данных, содержащий наблюдаемые и предсказанные векторы, и параметр **tidy**, который управляет типом вывода в виде списка (**tidy = FALSE**) или в виде **data.frame** (**tidy = TRUE**).

## Установка

```
install.packages("metrica")
```

Вы можете установить версию для разработки с помощью **GitHub**:

```
#install.packages("devtools")
devtools::install_github("adriancorrendo/metrica")
```

## Наборы данных

Пакет **metrica** содержит четыре примера наборов данных непрерывных переменных (регрессия) из программного обеспечения APSIM

- **Пшеница**: 137 точек данных по количеству зерен пшеницы IN
- **Ячмень**: 69 точек данных количества зерен ячменя
- **Сорго**: 36 точек данных количества зерен сорго
- **Нут**: 39 точек данных надземной сухой массы нута

Кроме того, **metrica** также предоставляет два собственных примера для категориальных переменных (классификация):

- **land\_cover**: бинарный набор данных о почвенно-растительном покрове с использованием спутниковых снимков. Значения 1 = растительность, 0 = другой тип почвенно-растительного покрова.
- **maize\_phenology**: набор данных о фенологии кукурузы (*Zea mays* L.) (16 стадий развития культуры).

Посмотрите документацию по метрикам, чтобы найти все метрики производительности и их детали **metrica**

## Регрессия

```
R2(data = wheat, obs = obs, pred=pred, tidy = TRUE)
#> R2
#> 1 0.8455538
```

```
RMSE(data = wheat, obs = obs, pred = pred)
#> $RMSE
#> [1] 1.666441
```

```
KGE(data = wheat, obs = obs, pred = pred)
#> $KGE
#> [1] 0.9106471
```

Пользователи также могут рассчитать все (по умолчанию) или выбранный список метрик сразу, используя функцию **metrics\_summary()**:

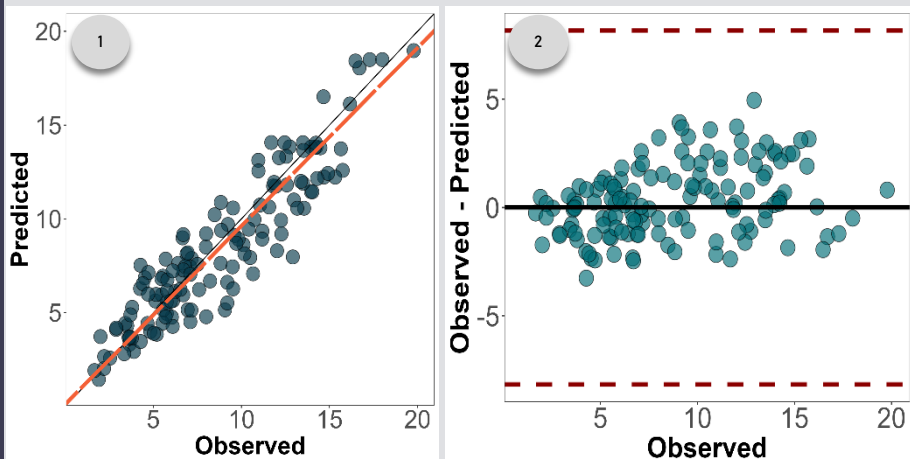
```
sel_r_metrics <- c("R2", "MBE", "RMSE", "RSR", "NSE",
                  "KGE", "CCC")

metrics_summary(data = wheat,
                obs = obs,
                pred = pred,
                type = "regression",
                metrics_list = sel_r_metrics)
```

## Сценарии

```
1. scatter_plot(data = wheat,
                obs = obs, pred = pred)
```

```
2. bland_altman_plot(data = wheat,
                    obs = obs, pred = pred)
```



## Классификация

```
accuracy(data=maize_phenology, obs=actual, pred=predicted)
#> $accuracy
#> [1] 0.8834951
```

```
precision(data=maize_phenology, obs=actual, pred=predicted)
#> $precision
#> [1] 0.8335108
```

```
recall(data = maize_phenology, obs=actual, pred=predicted)
#> $recall
#> 1 0.8405168
```

Для классификации пользователи могут также применять функцию **metrics\_summary()** для получения сразу нескольких метрик

```
sel_c_metrics <- c("accuracy", "precision", "recall",
                  "fscore")
```

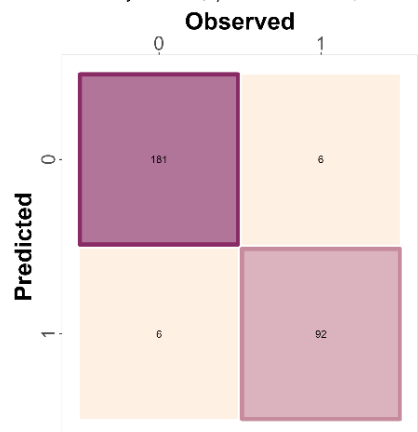
```
metrics_summary(data = landcover,
                obs = actual, pred = predicted,
                type = "classification",
                metrics_list = sel_c_metrics,
                pos_level = 1)
```

## Матрица ошибок

```
confusion_matrix(data = .,
                 obs = labels, pred = predictions,
                 plot = TRUE,
                 unit="count")
```

Биномиальное распределение

Performance metrics:  
accuracy = 0.96; precision = 0.94; recall = 0.94.



Мультиномиальное распределение

Performance metrics:  
accuracy = 0.43; precision = 0.43; recall = 0.43.

