



Statsmodels是一个Python模块，为许多不同统计模型的估计提供类和函数，以及进行统计测试和统计数据探索。

线性回归-模型介绍

统计模型被假定为 $Y = X\beta + \mu$ 且 $\mu \sim \mathcal{N}(0, \Sigma)$.
GLS | 任意协方差的广义最小二乘法 Σ .
OLS | 普通最小二乘法 $\Sigma = I$.
WLS | 异方差误差的加权最小二乘法 $\text{diag}(\Sigma)$.
GLSAR | 具有自相关 $\text{AR}(\rho)$ 误差的可行广义最小二乘法 $\Sigma = \Sigma(\rho)$ ，除了RecursiveLS，GLS是其他回归类的超类。
注：所有回归模型都定义了相同的方法并遵循相同的结构，并且可以以类似的方式使用。

线性回归-模型类

以下方法四个模型均可使用，以OLS为例。
from statsmodels.regression.linear_model.OLS import *
fit | 完全适合模型。
fit_regularized() | 将正则化拟合返回到线性回归模型。
from_formula() | 从公式和数据框中建模。
get_distribution() | 返回预测分布的随机数生成器。
hessian() | 给定点情况下评估Hessian函数。
information() | Fisher模型的信息矩阵。
loglike() | 似然函数。
predict() | 从设计矩阵返回线性预测值。
score() | 给定点的情况下评估分数函数。
whiten(Y) | 返回Y值。

递归最小二乘(RLS)

递归最小二乘(RLS)为OLS的扩展。

from statsmodels.regression.recursive_ls.RecursiveLS import *
initialize_statespace() | 初始化状态空间表示。
observed_information_matrix() | 观测信息矩阵。
opg_information_matrix() | 梯度信息矩阵的外积。
set_conserve_memory() | 设置内存保存。
set_filter_method() | 设置过滤方式。
set_inversion_method() | 设置反演方法。
set_smoother_output() | 设置更平滑输出。
set_stability_method() | 设置数值稳定性。
update() | 更新模型的参数。
transform_params() | 将优化器使用的不受约束的参数转换为约束。
untransform_params() | 变换可能性评估中使用的约束参数。

线性回归-结果类(OLSResults)

拟合OLS模型的结果类（大多数方法和属性继承于RegressionResults）。
from statsmodels.regression.linear_model.OLSResults import *
compare_f_test() | 使用F检验来测试受限模型是否正确。
compare_lm_test() | 使用拉格朗日乘数测试来测试受限模型是否正确。
compare_lr_test() | 似然比检验受限模型是否正确。
condition_number() | 返回外源矩阵的条件数。
conf_int() | 返回拟合参数的置信区间。
cov_params | 返回方差/协方差矩阵。
eigenvals() | 返回按降序排列的特征值。
get_prediction() | 计算预测结果。
remove_data() | 从结果和模型中移除所有nobs数据数组。

load() | 加载pickle格式(类方法)。
save() | 保存这个实例pickle。
summary() | 总结回归结果。
summary2() | 实验总结功能来总结回归结果。
tvalues() | 返回给定参数估计的t统计量。
wald_test() | 联合线性假设的Wald检验。

统计-残差诊断和规范测试

from statsmodels.stats.stattools import *
durbin_watson() | 计算Durbin-Watson统计量。
jarque_bera() | 计算正常Jarque-Bera测试。
omni_normtest() | 计算偏斜的中值稳健度。
robust_skewness() | 计算Kim&White中的四个偏态度量。
robust_kurtosis() | 计算Kim&White中的四个峰度度量。

from statsmodels.stats.diagnostic import *

HetGoldfeldQuandt() | 测试2个子样本中方差是否相同。
het_white() | 白异方差拉格朗日乘子检验。
het_arch() | 自回归条件异方差检验(ARCH)。
breaks_cusumolsresid() | 基于ols残差的参数稳定性检验。
recursive_olsresiduals() | 用残差和点积检验统计量来计算递归ols。
compare_cox() | 测试非嵌套模型。
compare_j() | 用于比较非嵌套模型的J-Test。
lilliefors() | lilliefors测试正常。

统计-异常值和影响因子的措施

from statsmodels.stats.outliers_influence import *
OLSInfluence() | 计算OLS结果的异常值和影响因子VIF，一个外生变量。

统计-拟合优度检验及措施

from statsmodels.stats.gof import *
powerdiscrepancy() | 计算功率差异。
gof_chisquare_discrete() | 执行一个离散分布的随机样本的卡方检验。
gof_binning_discrete() | 获得用于离散分布的卡方gof测试箱。
normal_ad() | 正态分布未知均值和方差的Anderson Darling检验。

统计-多重测试和多重比较

from statsmodels.sandbox.stats.multicomp import *
multipletests() | 测试结果和p值校正。
GroupsStats() | 统计分组。
MultiComparison() | 多重测试比较。
TukeyHSDResults() | TukeyHSD测试的结果。
pairwise_tukeyhsd() | 用TukeyHSD置信区间计算所有两两比较。

from statsmodels.stats.multitest import *

local_fdr() | 计算Z-scores的本地FDR值。
fdrcorrection_twostage() | 迭代两阶段线性升序程序，估计真值。
NullDistribution() | 估计空Z-scores高斯分布。

统计-比例

from statsmodels.stats.proportion import *
proportion_confint() | 二项式比例置信区间。
proportion_effectsize() | 比较两个比例的规模大小。
binom_test() | 做一个成功几率是P的检验。
binom_tost() | 使用二项式分布对一个比例进行TOST检验。
multinomial_proportions_confint() | 多项比例的置信区间。

