```
import pandas as pd
import statsmodels.api as sm


mh_df = pd.read_csv('/content/Mental Health Dataset/mental_health_dataset.csv')
mh_df.head(5)
```

	age	gender	employment_status	work_environment	mental_health_history	seeks_treatment	stress_level	sleep_hours	physical_activity
0	56	Male	Employed	On-site	Yes	Yes	6	6.2	
1	46	Female	Student	On-site	No	Yes	10	9.0	
2	32	Female	Employed	On-site	Yes	No	7	7.7	
3	60	Non-binary	Self-employed	On-site	No	No	4	4.5	
4	25	Female	Self-employed	On-site	Yes	Yes	3	5.4	

Logistic Regression

> This model is built to predict the likelihood of employees seeking mental health treatment

```
# Encoding binary target variable (0 and 1)
mh_df['seeks_treatment'] = mh_df['seeks_treatment'].map({'No': 0, 'Yes': 1})

# Defining predictors and target variables
X = mh_df[['stress_level', 'sleep_hours', 'physical_activity_days',
        'depression_score', 'anxiety_score', 'social_support_score', 'productivity_score']]
y = mh_df['seeks_treatment']

# Adding constant for intercept
X = sm.add_constant(X)

# Fitting logistic regression model
model = sm.Logit(y, X).fit()

# Summary output
print(model.summary())
```

```
Optimization terminated successfully.
         Current function value: 0.672375
         Iterations 4
                           Logit Regression Results
==============================================================================
Dep. Variable:        seeks_treatment   No. Observations:                10000
Model:                          Logit   Df Residuals:                     9992
Method:                           MLE   Df Model:                            7
Date:                Sun, 15 Jun 2025   Pseudo R-squ.:              0.0002191
Time:                        14:37:28   Log-Likelihood:                -6723.7
converged:                       True   LL-Null:                       -6725.2
Covariance Type:            nonrobust   LLR p-value:                    0.8899
==============================================================================================
                                 coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------
const                      -0.5809      0.438     -1.327      0.184      -1.439       0.277
stress_level               -0.0063      0.007     -0.887      0.375      -0.020       0.008
sleep_hours                 0.0027      0.014      0.192      0.848      -0.025       0.030
physical_activity_days     -0.0060      0.009     -0.672      0.502      -0.024       0.012
depression_score            0.0032      0.007      0.487      0.626      -0.010       0.016
anxiety_score               0.0035      0.003      1.087      0.277      -0.003       0.010
social_support_score       -0.0003      0.001     -0.496      0.620      -0.002       0.001
productivity_score          0.0018      0.004      0.431      0.667      -0.006       0.010
==============================================================================================
```

Observation

Model: Logistic regression predicting *"seeks_treatment"*

Report

> Coefficients and p-values were examined for each predictor.
>
> A p-value less than 0.05 indicates that the predictor has a statistically significant effect on the odds of seeking treatment.

Result Interpretation: ***ALL Predictors***

> The results for ALL predictors show that they were not statistical significance predictors for "seeking treatment" (P>0.05). In other words, increase or decrease in the anaylysed predictor variables does not affect the odds of seeking treatment.

General Insight

> The model proved that the anaylsed predictor variables (stress_level sleep_hours, physical_activity_day, depression_score, anxiety_score, social_support_score, and productivity_score) were insignificant predictors of the likelihood of employees to seek treatment for their mental health troubless.
>
> However, although the predictor variables are insignificant, they may still play a role in a multivariate context or interact with others, which could be tested in more advanced models.

Simple Linear Regression

> This model is built to investigate the social and lifestyle determinants of mental health among employees in the workplace.

```
# Defining target (outcome) and predictor variables
target_vars = ['depression_score', 'anxiety_score', 'productivity_score']
predictor_vars = ['sleep_hours', 'physical_activity_days', 'social_support_score']

# Run simple linear regression for each pair
for target in target_vars:
    for predictor in predictor_vars:
        print(f"\nSimple Linear Regression: {target} ~ {predictor}")
        X = sm.add_constant(mh_df[predictor])
        y = mh_df[target]
        model = sm.OLS(y, X).fit()
        print(model.summary())
```

Show hidden output

Observation

Model: OLS - Simple linear regression

Target Variables: *depression_score, anxiety_score, and productivity_score*

Predictor Variables: *sleep_hours, physical_activity_days, social_support_score*

> Result showed no statistical significant correlation between the predictor and target variables

Interpretation

> The model established that sleep hours, physical activity and having social support were insignificant predictors of mental health outcomes such as depression, anxiety, and productivity.

General Insight

> These results suggest that lifestyle factors like sleep and physical activity, and social factors like support, cannot determine or explain variations in mental health status among employees in the work places.

Simple Linear Regression Predicting Productivity Score

> From previously noted significant correlation between depression_score vs productivity_score, we built this model to determine causation

```
# work_environment vs anxiety_score

X = sm.add_constant(mh_df['depression_score'])
y = mh_df['productivity_score']
model = sm.OLS(y, X).fit()
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     productivity_score   R-squared:                       0.882
Model:                            OLS   Adj. R-squared:                  0.882
Method:                 Least Squares   F-statistic:                 7.473e+04
Date:                Sun, 15 Jun 2025   Prob (F-statistic):               0.00
Time:                        15:29:54   Log-Likelihood:                -29938.
No. Observations:               10000   AIC:                         5.988e+04
Df Residuals:                    9998   BIC:                         5.989e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const              99.4025      0.094   1055.855      0.000      99.218      99.587
depression_score   -1.4688      0.005   -273.370      0.000      -1.479      -1.458
==============================================================================
Omnibus:                        4.752   Durbin-Watson:                   2.002
Prob(Omnibus):                  0.093   Jarque-Bera (JB):                4.785
Skew:                          -0.040   Prob(JB):                       0.0914
Kurtosis:                       3.071   Cond. No.                         34.2
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Observation

Model: Simple linear regression

Target Variable: ***Productivity_score***

Predictor Variable: ***Depression_score***

> Coefficient (β) = -1.4688, p-value = 0.000
>
> R-squared = 0.882
>
> There is a significant association between Depression_score and Productivity_score.

Interpretation

> A significant negative β coefficient for depression_score predicting productivity_score (β = -1.4688, p = 0.000, R^2 = 0.882) suggests that as depression increases, productivity decreases. Hence indicating a strong linear relationship.

General Insight

> The result suggest that depression can explain variations in productivity among employees in the work place.

Start coding or generate with AI.

Start coding or generate with AI.