

# Analysis code: part 1

```
In [3]: import csv
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import statsmodels.api as sm
from itertools import combinations
import copy
import collections

df = pd.read_csv("Data.csv")

all_column_names = ["User", "PSQI", "Age", "Gender", "Weight", "Height", "BMI",
                    ,
                    "Cortisol_before", "Cortisol_after", "Melatonin_before", "Mela
tonin_after",
                    "Latency_Efficiency", "Total_minutes_in_bed", "Total_sleep_tim
e",
                    "RR", "Daily_stress",
                    "Activity_medium", "Activity_heavy", "Activity_small_screen_us
age", "Activity_large_screen_usage",
                    "Activity_smoking", "Activity_alcohol_assumption",
                    "Actigraph_X_mean", "Actigraph_Y_mean", "Actigraph_Z_mean",
                    "Steps_mean"]

all_names = [df["User"], df["PSQI"], df["Age"], df["Gender"], df["Weight"], df
["Height"], df["BMI"],
             df["Cortisol_before"], df["Cortisol_after"], df["Melatonin_before"
], df["Melatonin_after"],
             df["Latency_Efficiency"], df["Total_minutes_in_bed"], df["Total_sl
eep_time"],
             df["RR"], df["Daily_stress"], df["Activity_medium"], df["Activity_
heavy"],
             df["Activity_small_screen_usage"], df["Activity_large_screen_usag
e"],
             df["Activity_smoking"], df["Activity_alcohol_assumption"],
             df["Actigraph_X_mean"], df["Actigraph_Y_mean"], df["Actigraph_Z_me
an"],
             df["Steps_mean"]]

print("Done!")
```

Done!

```
In [4]: # Summary statistics  
print(df.describe())
```

	Unnamed: 0	PSQI	Age	Weight	Height	BMI \
count	22.000000	22.000000	22.000000	22.000000	22.000000	22.000000
mean	10.500000	5.318182	27.272727	75.045455	179.909091	23.123541
std	6.493587	1.985336	4.107853	12.789420	8.216760	3.093798
min	0.000000	2.000000	20.000000	60.000000	169.000000	19.409358
25%	5.250000	4.000000	25.000000	67.000000	175.000000	21.686556
50%	10.500000	5.000000	27.000000	70.000000	180.000000	22.848325
75%	15.750000	7.000000	27.750000	80.000000	183.000000	23.388440
max	21.000000	9.000000	40.000000	115.000000	205.000000	33.240837

	Cortisol_before	Cortisol_after	Melatonin_before	Melatonin_after \
count	22.000000	22.000000	2.200000e+01	2.200000e+01
mean	0.028053	0.069860	8.330234e-09	7.281474e-09
std	0.029735	0.051937	6.545962e-09	6.042238e-09
min	0.012017	0.015572	1.629907e-09	8.283802e-10
25%	0.015322	0.042028	3.386563e-09	3.657571e-09
50%	0.020393	0.059983	5.912706e-09	6.856391e-09
75%	0.027532	0.081556	1.098545e-08	8.570786e-09
max	0.155777	0.261252	2.396239e-08	2.853905e-08

	...	Activity_medium	Activity_heavy	Activity_small_screen_usage \
count	...	22.000000	22.000000	22.000000
mean	...	20.909091	93.590909	57.363636
std	...	33.863745	32.854164	54.862204
min	...	0.000000	40.000000	0.000000
25%	...	0.000000	77.750000	0.000000
50%	...	0.000000	92.500000	50.000000
75%	...	33.750000	108.750000	112.250000
max	...	120.000000	180.000000	174.000000

	Activity_large_screen_usage	Activity_smoking \
count	22.000000	22.000000
mean	17.500000	7.272727
std	17.065002	16.292510
min	0.000000	0.000000
25%	0.500000	0.000000
50%	14.500000	0.000000
75%	30.000000	0.000000
max	51.000000	60.000000

	Activity_alcohol_assumption	Actigraph_X_mean	Actigraph_Y_mean \
count	22.000000	22.000000	22.000000
mean	10.318182	16.434876	16.780694
std	11.047027	3.131175	4.245410
min	0.000000	12.151812	10.515468
25%	0.000000	13.905174	14.135690
50%	10.000000	16.338063	16.481773
75%	18.000000	17.824614	19.377193
max	45.000000	24.693821	27.243074

	Actigraph_Z_mean	Steps_mean
count	22.000000	22.000000
mean	19.271554	0.184510
std	4.202566	0.047359
min	13.269682	0.118963
25%	16.264400	0.153240
50%	19.504444	0.178461

75%	21.264420	0.203129
max	29.023668	0.287923

[8 rows x 25 columns]

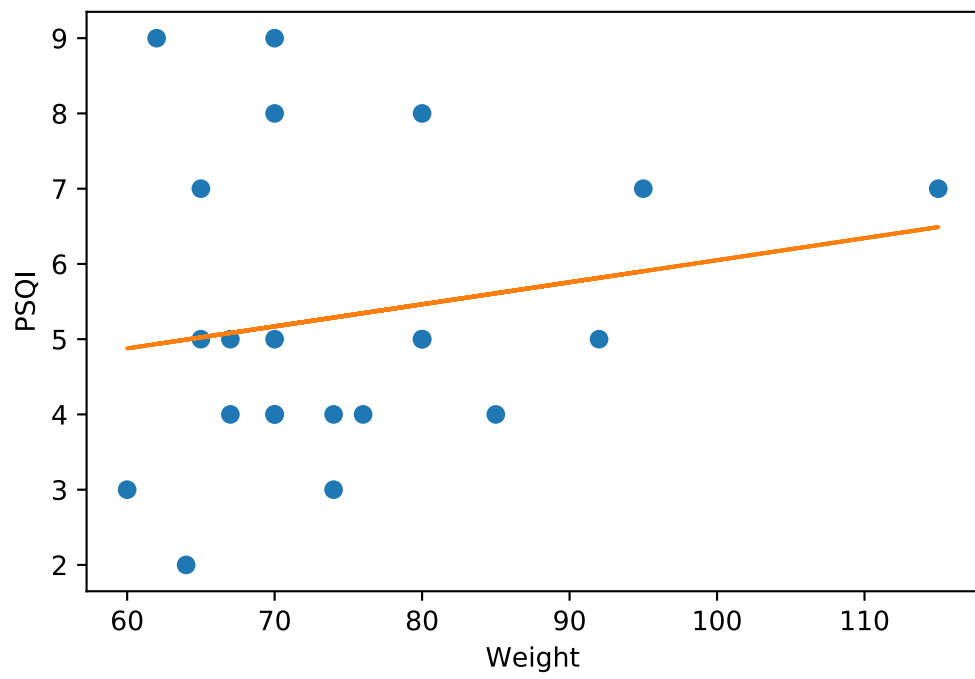
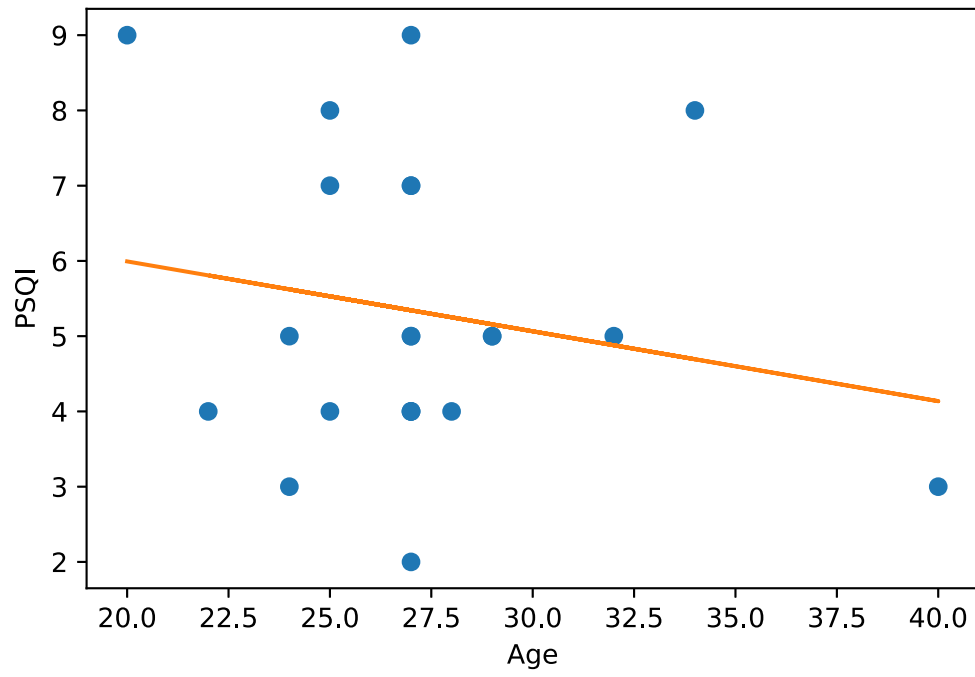
In [26]: *# Gauss Markov Assumptions*

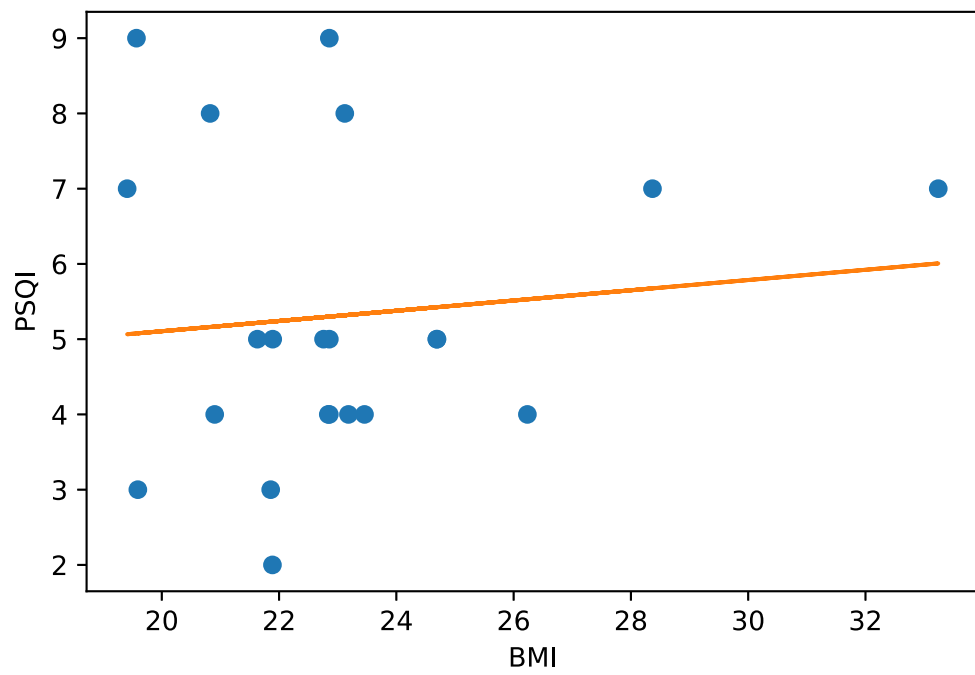
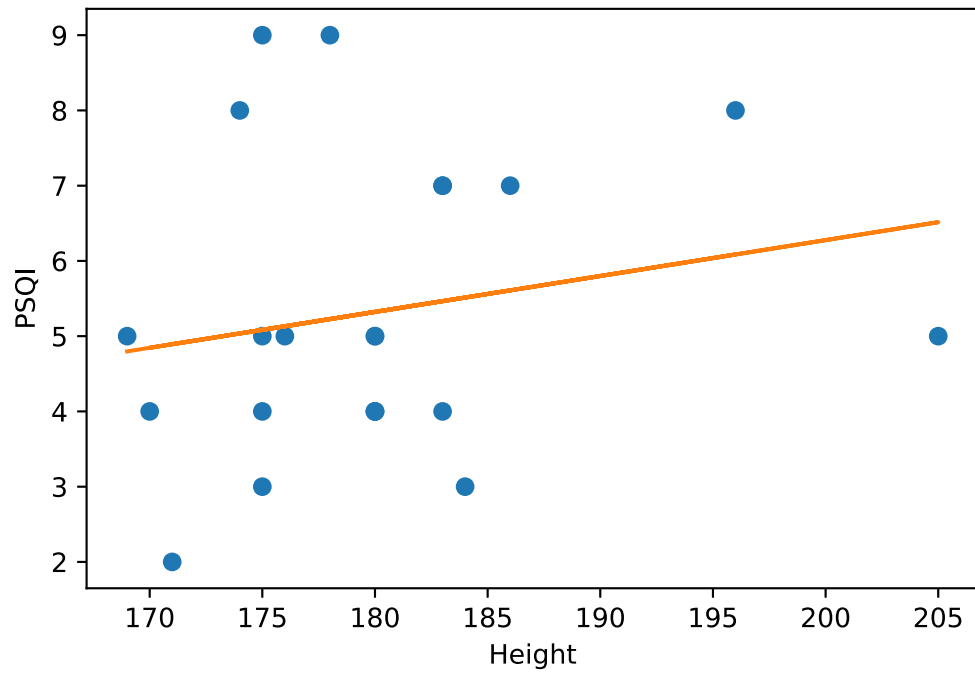
*# 1 Scatter plots*

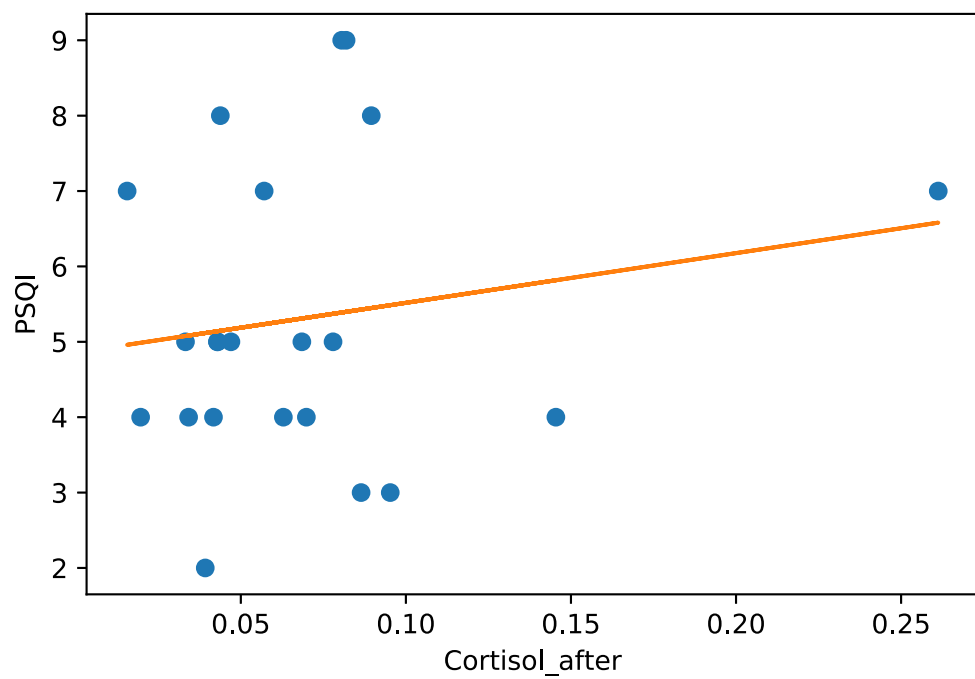
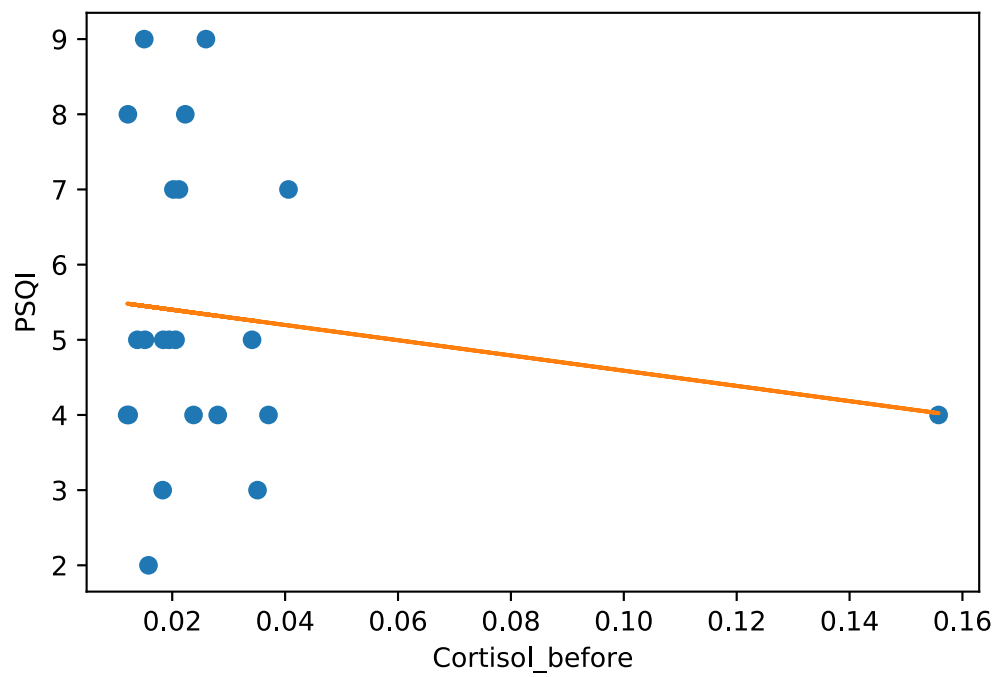
PSQI = df["PSQI"]

```
for i in range(len(all_names)):
    if i != 0 and i != 1 and i != 3:
        X = all_names[i]
        X_name = all_column_names[i]
        plt.figure()
        plt.plot(X, PSQI, 'o')
        plt.xlabel(X_name)
        plt.ylabel("PSQI")

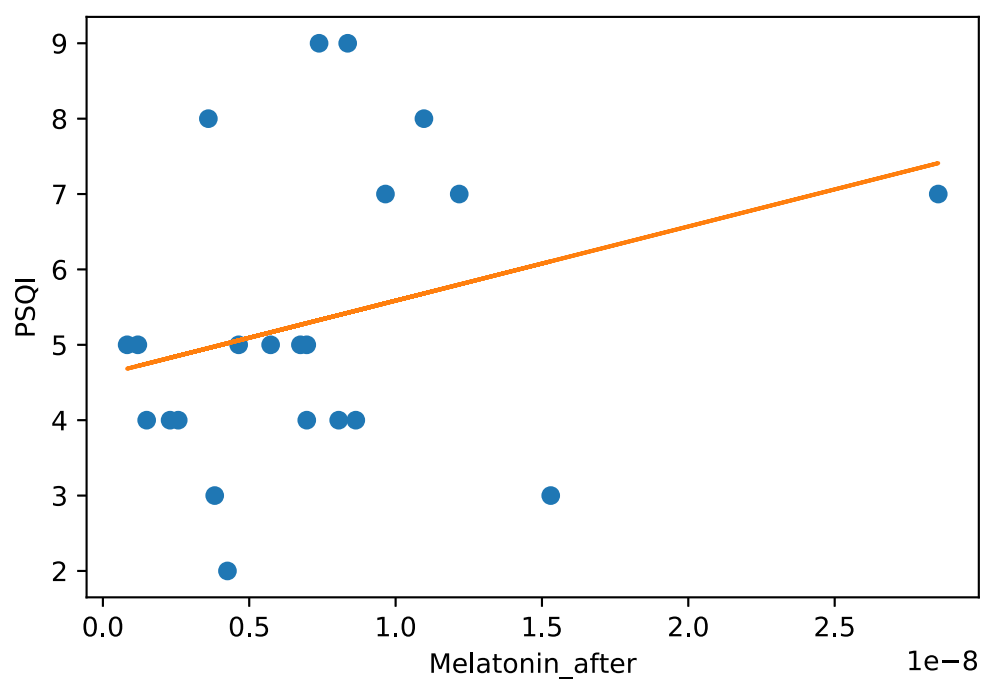
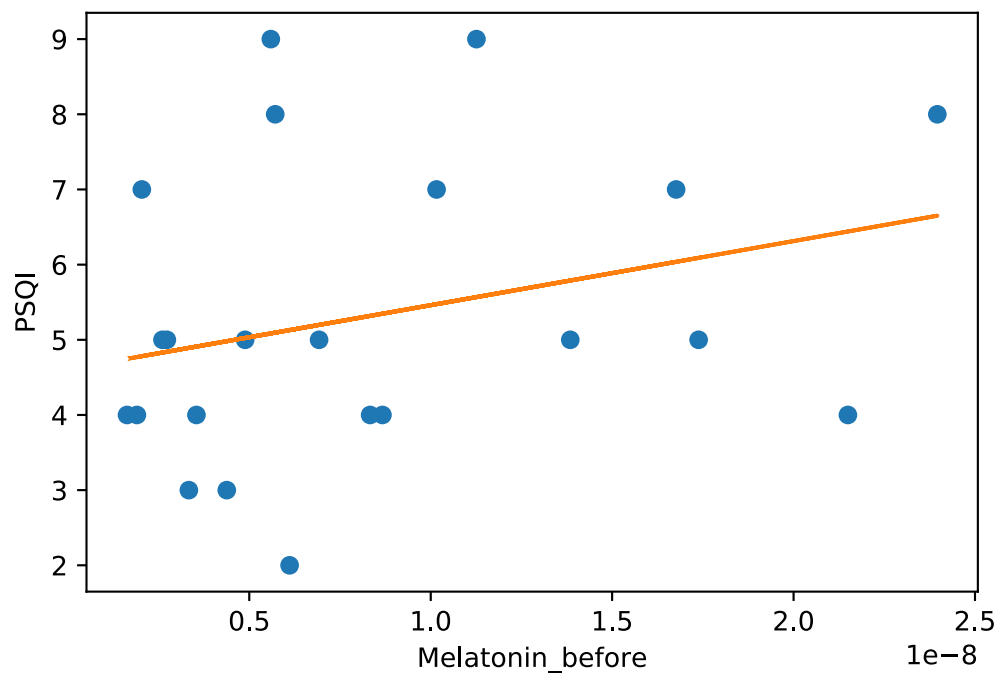
        m, b = np.polyfit(X, PSQI, 1)
        plt.plot(X, m*X + b)
        plt.show()
```

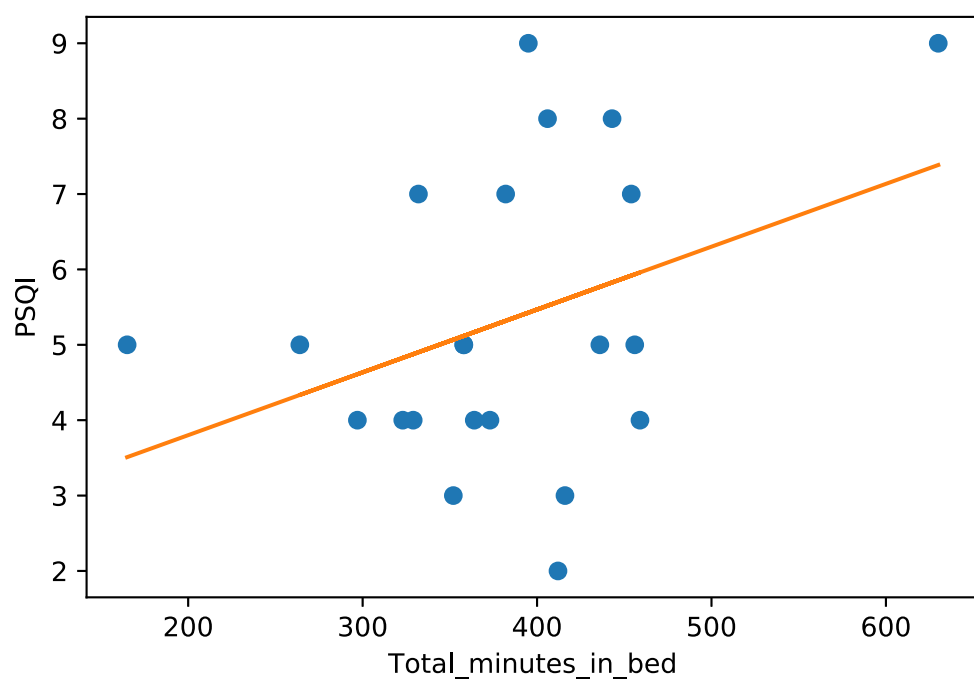
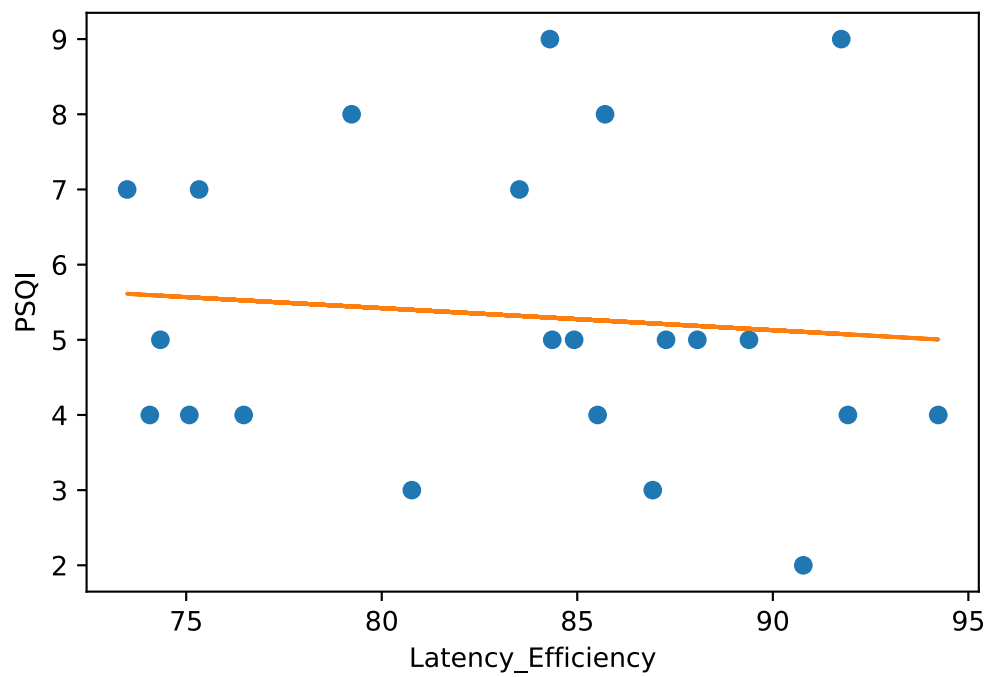


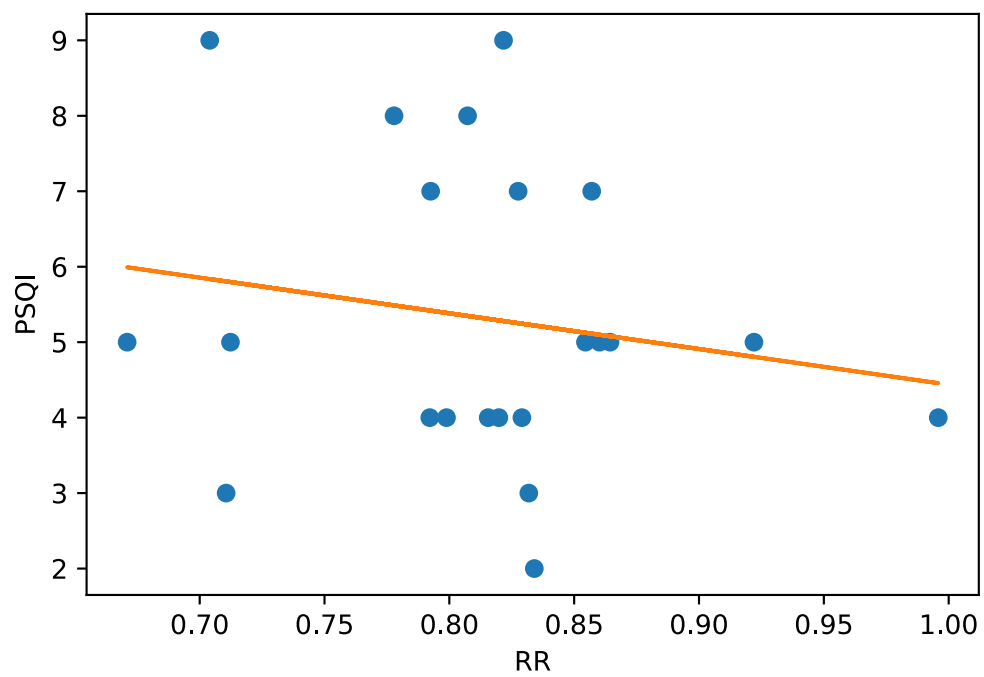
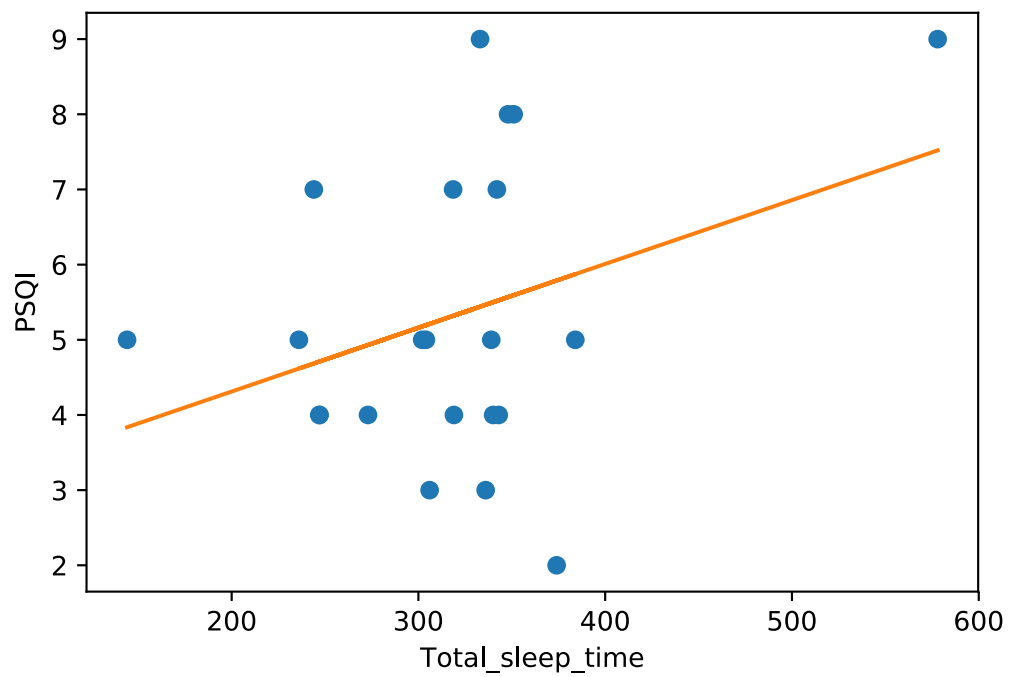


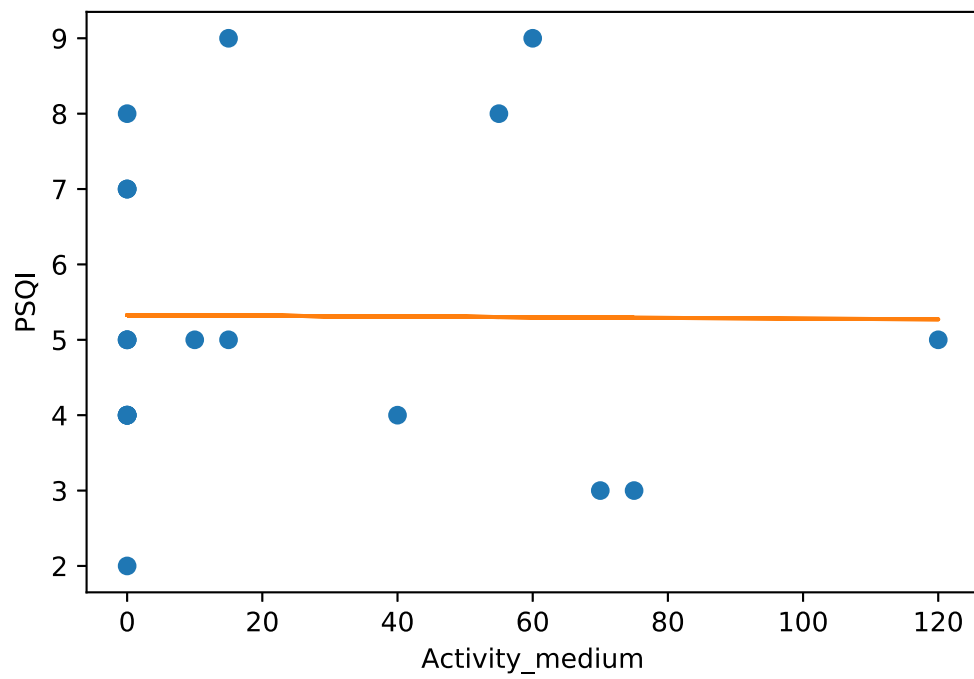
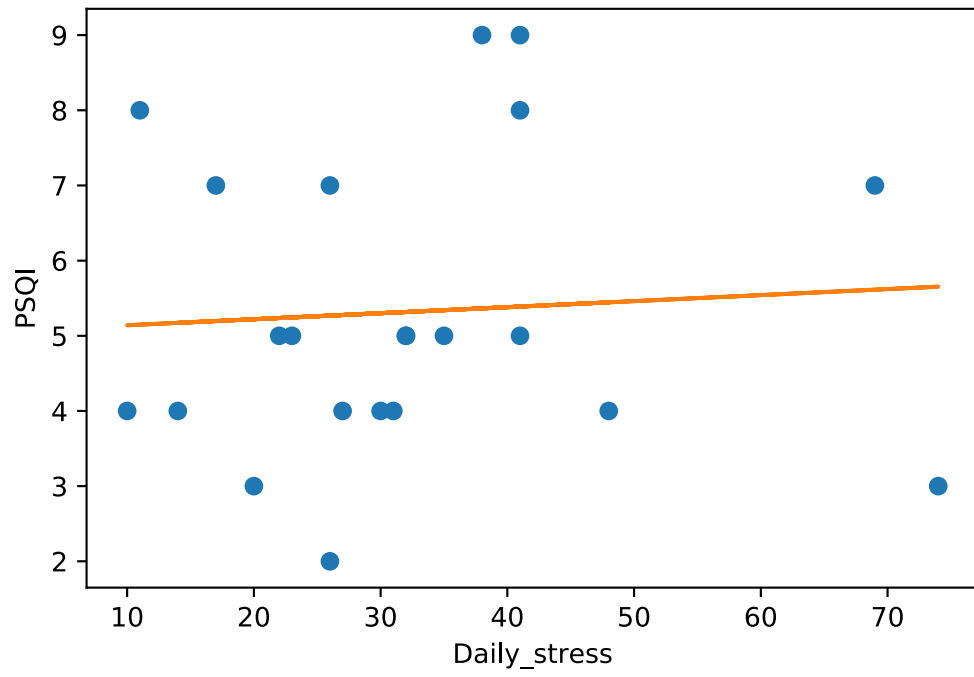


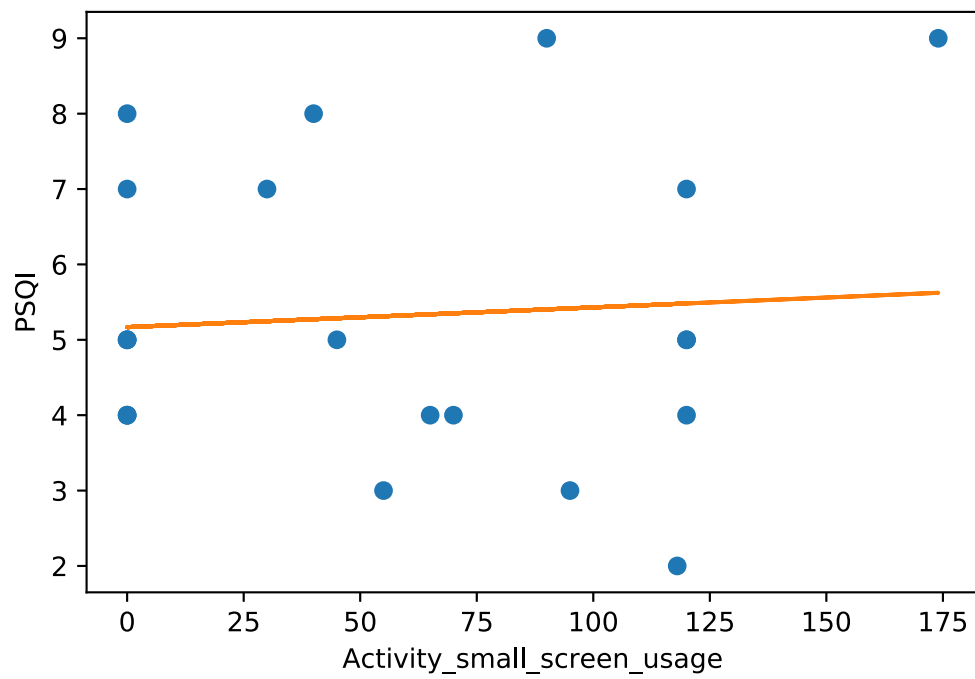
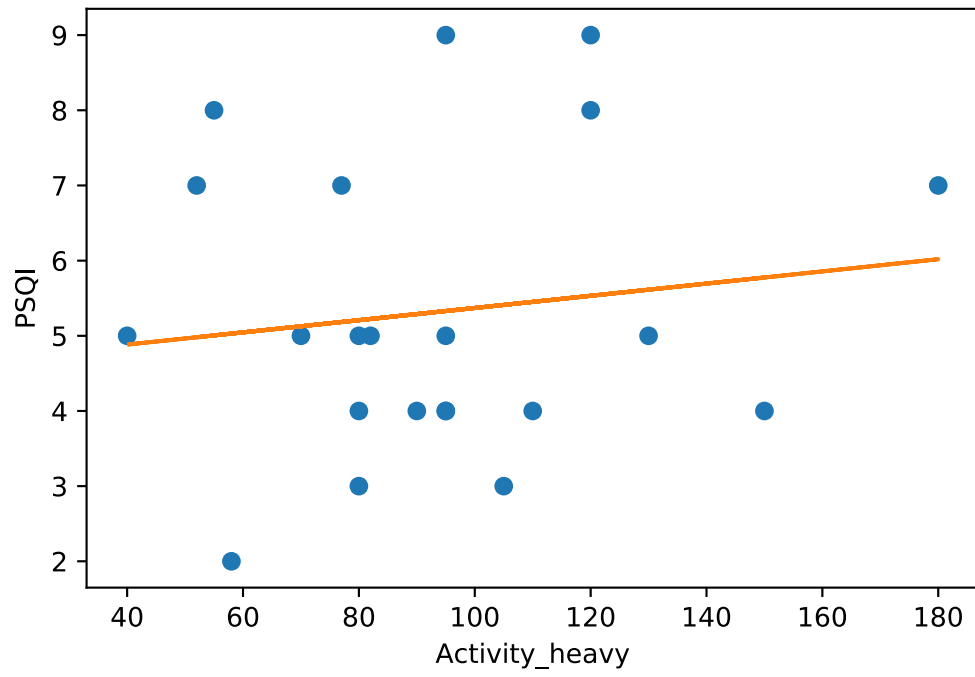


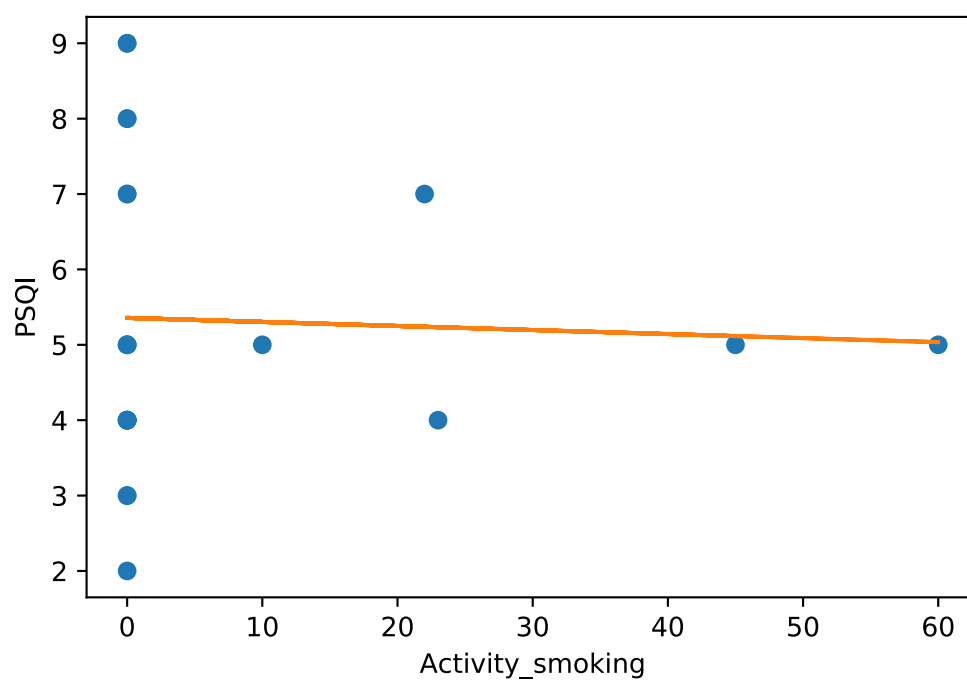
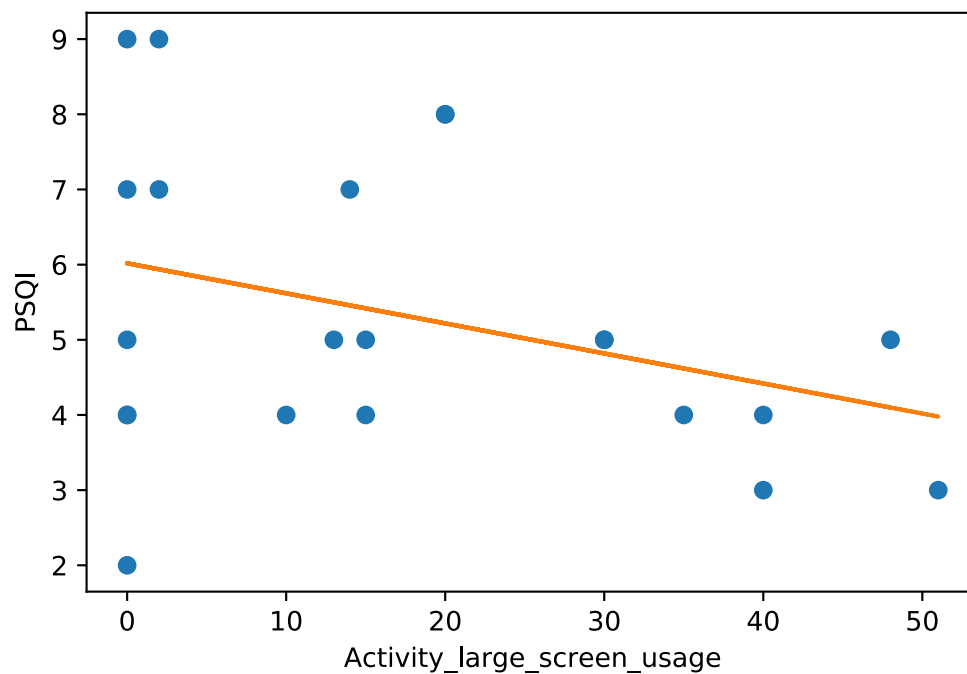


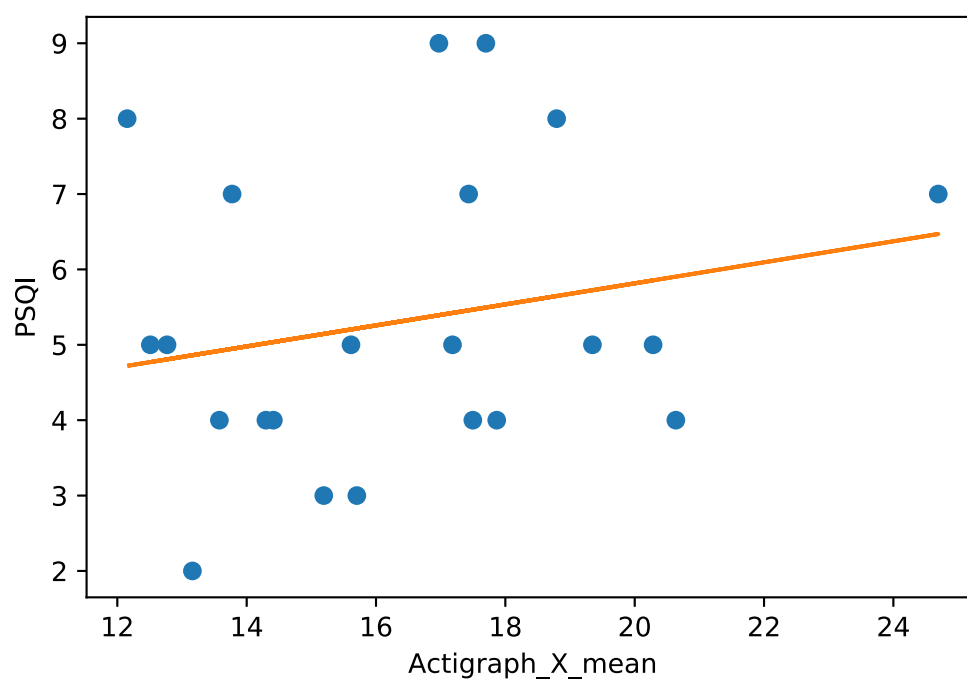
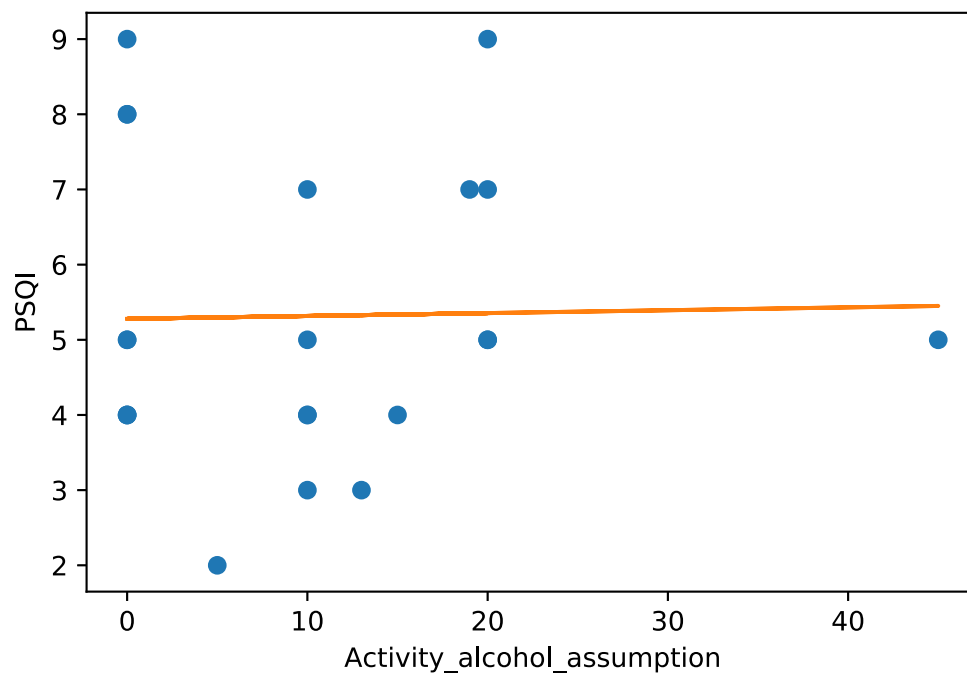


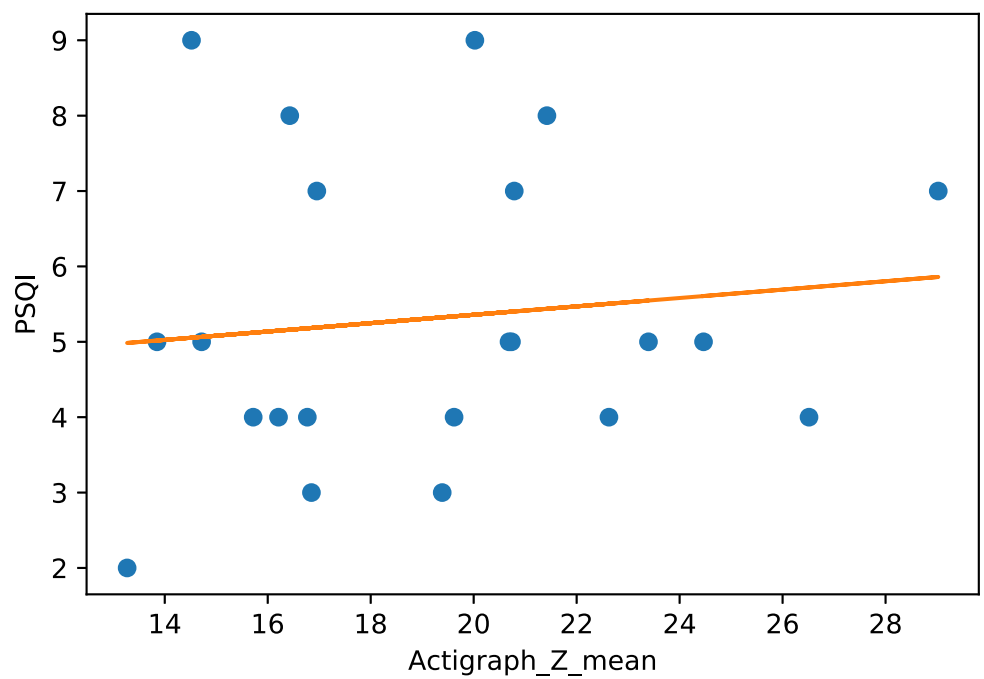
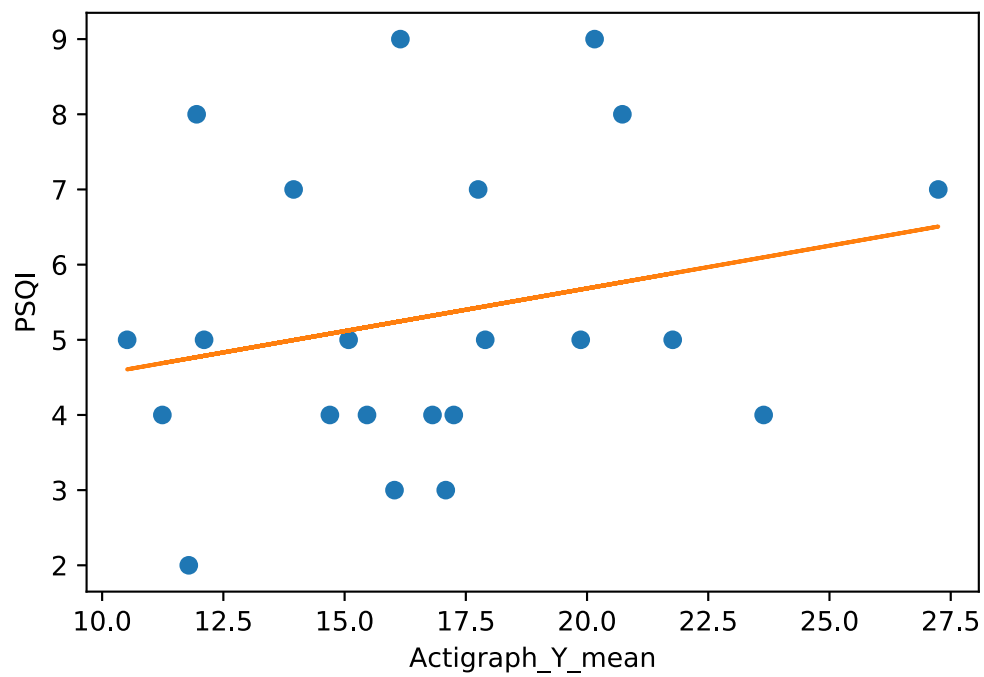




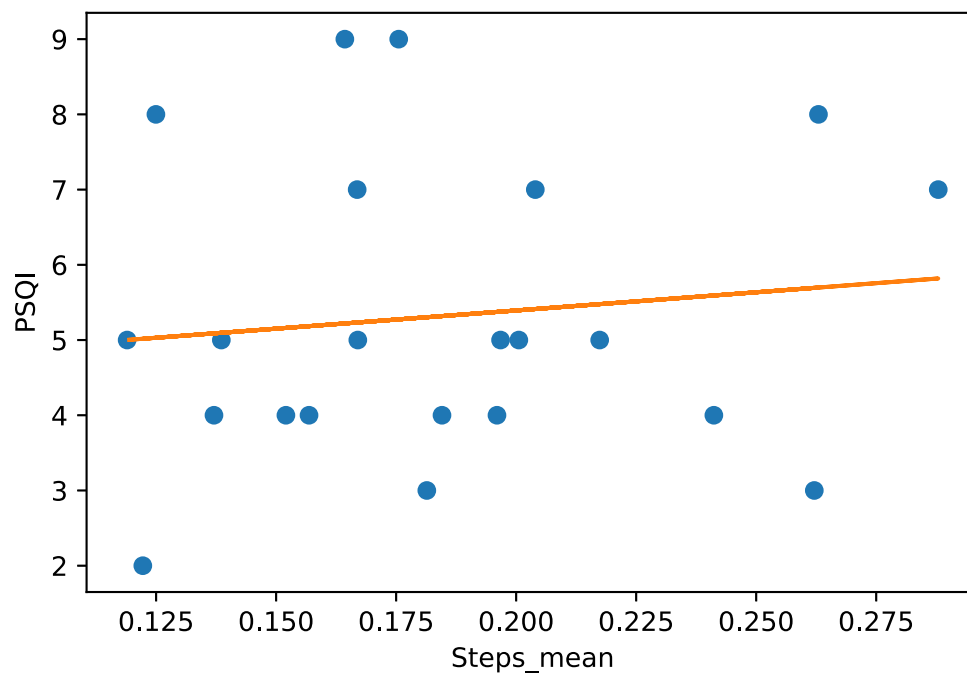












```
In [27]: # 6 No perfect collinearity  
print(df.corr())
```

	Unnamed: 0	PSQI	Age	Weight	\
Unnamed: 0	1.000000	0.459866	0.160666	-0.197531	
PSQI	0.459866	1.000000	-0.192153	0.188820	
Age	0.160666	-0.192153	1.000000	0.094017	
Weight	-0.197531	0.188820	0.094017	1.000000	
Height	0.317721	0.197436	0.134795	0.581416	
BMI	-0.442529	0.106017	0.027300	0.849639	
Cortisol_before	-0.305455	-0.151763	0.047168	-0.009956	
Cortisol_after	0.124927	0.172550	-0.035142	0.011676	
Melatonin_before	0.205865	0.281135	-0.071609	-0.005328	
Melatonin_after	0.115987	0.299470	-0.181221	0.161039	
Latency_Efficiency	-0.003170	-0.095865	-0.081850	-0.201153	
Total_minutes_in_bed	0.596738	0.374383	-0.183670	-0.139196	
Total_sleep_time	0.573445	0.342660	-0.214531	-0.179795	
RR	0.167501	-0.172546	0.182324	0.369637	
Daily_stress	0.472931	0.065829	0.221156	-0.202237	
Activity_medium	0.131014	-0.008049	0.100828	-0.199109	
Activity_heavy	0.139615	0.134231	-0.343151	-0.329966	
Activity_small_screen_usage	0.332963	0.072336	-0.105475	-0.619105	
Activity_large_screen_usage	-0.089168	-0.343652	0.294815	-0.208039	
Activity_smoking	-0.176889	-0.044031	0.068563	0.307995	
Activity_alcohol_assumption	-0.064059	0.021219	-0.231811	0.217285	
Actigraph_X_mean	-0.372321	0.219948	0.063073	0.518704	
Actigraph_Y_mean	-0.330404	0.242861	0.009920	0.515548	
Actigraph_Z_mean	-0.406643	0.117652	0.181214	0.545262	
Steps_mean	-0.150973	0.115394	0.473888	0.474879	

	Height	BMI	Cortisol_before	\
Unnamed: 0	0.317721	-0.442529	-0.305455	
PSQI	0.197436	0.106017	-0.151763	
Age	0.134795	0.027300	0.047168	
Weight	0.581416	0.849639	-0.009956	
Height	1.000000	0.066111	-0.247371	
BMI	0.066111	1.000000	0.149799	
Cortisol_before	-0.247371	0.149799	1.000000	
Cortisol_after	-0.082630	0.073196	0.297433	
Melatonin_before	0.086234	-0.048618	-0.126068	
Melatonin_after	-0.024930	0.209295	0.070775	
Latency_Efficiency	-0.267287	-0.080451	0.378241	
Total_minutes_in_bed	0.057901	-0.208971	-0.111254	
Total_sleep_time	-0.027071	-0.205517	0.010838	
RR	0.354039	0.228185	0.022252	
Daily_stress	0.259016	-0.412626	0.221259	
Activity_medium	-0.214466	-0.103985	0.128193	
Activity_heavy	0.023493	-0.419928	0.027063	
Activity_small_screen_usage	-0.408835	-0.507808	-0.038016	
Activity_large_screen_usage	-0.280853	-0.074093	0.260010	
Activity_smoking	0.061020	0.343779	-0.100976	
Activity_alcohol_assumption	-0.102489	0.336155	0.077173	
Actigraph_X_mean	-0.030709	0.646711	0.261222	
Actigraph_Y_mean	0.029932	0.605799	0.178951	
Actigraph_Z_mean	0.172590	0.547308	0.179709	
Steps_mean	0.080988	0.524388	0.210070	

	Cortisol_after	Melatonin_before	\
Unnamed: 0	0.124927	0.205865	
PSQI	0.172550	0.281135	

Age	-0.035142	-0.071609
Weight	0.011676	-0.005328
Height	-0.082630	0.086234
BMI	0.073196	-0.048618
Cortisol_before	0.297433	-0.126068
Cortisol_after	1.000000	0.317442
Melatonin_before	0.317442	1.000000
Melatonin_after	0.804363	0.452468
Latency_Efficiency	-0.040472	-0.016789
Total_minutes_in_bed	-0.079011	-0.211850
Total_sleep_time	-0.073927	-0.202370
RR	-0.145161	0.146009
Daily_stress	0.187192	-0.125056
Activity_medium	0.052556	-0.376963
Activity_heavy	-0.157818	0.128089
Activity_small_screen_usage	0.029813	0.046978
Activity_large_screen_usage	0.024743	-0.177318
Activity_smoking	-0.215277	-0.004766
Activity_alcohol_assumption	0.219202	-0.023494
Actigraph_X_mean	-0.087814	-0.163126
Actigraph_Y_mean	-0.130411	-0.250316
Actigraph_Z_mean	-0.128553	0.043414
Steps_mean	-0.047327	-0.275441

	Melatonin_after	...	Activity_medium \
Unnamed: 0	0.115987	...	0.131014
PSQI	0.299470	...	-0.008049
Age	-0.181221	...	0.100828
Weight	0.161039	...	-0.199109
Height	-0.024930	...	-0.214466
BMI	0.209295	...	-0.103985
Cortisol_before	0.070775	...	0.128193
Cortisol_after	0.804363	...	0.052556
Melatonin_before	0.452468	...	-0.376963
Melatonin_after	1.000000	...	-0.139459
Latency_Efficiency	-0.154058	...	0.333320
Total_minutes_in_bed	-0.026676	...	0.041627
Total_sleep_time	-0.060334	...	0.145556
RR	-0.058532	...	-0.510024
Daily_stress	-0.004386	...	0.200735
Activity_medium	-0.139459	...	1.000000
Activity_heavy	0.014561	...	-0.142178
Activity_small_screen_usage	0.132500	...	-0.009029
Activity_large_screen_usage	-0.013229	...	0.453213
Activity_smoking	-0.160892	...	-0.230485
Activity_alcohol_assumption	0.229495	...	0.160851
Actigraph_X_mean	0.022427	...	0.237974
Actigraph_Y_mean	-0.016156	...	0.315681
Actigraph_Z_mean	0.076436	...	-0.052154
Steps_mean	-0.003580	...	0.369597

	Activity_heavy	Activity_small_screen_usage \
Unnamed: 0	0.139615	0.332963
PSQI	0.134231	0.072336
Age	-0.343151	-0.105475
Weight	-0.329966	-0.619105
Height	0.023493	-0.408835

BMI	-0.419928	-0.507808
Cortisol_before	0.027063	-0.038016
Cortisol_after	-0.157818	0.029813
Melatonin_before	0.128089	0.046978
Melatonin_after	0.014561	0.132500
Latency_Efficiency	-0.135929	0.058570
Total_minutes_in_bed	0.084562	0.437801
Total_sleep_time	0.036696	0.429204
RR	-0.122141	0.014784
Daily_stress	0.310007	0.136802
Activity_medium	-0.142178	-0.009029
Activity_heavy	1.000000	0.235454
Activity_small_screen_usage	0.235454	1.000000
Activity_large_screen_usage	-0.208472	-0.012055
Activity_smoking	-0.384362	-0.201281
Activity_alcohol_assumption	-0.353217	-0.042943
Actigraph_X_mean	-0.159868	-0.211276
Actigraph_Y_mean	-0.040050	-0.262728
Actigraph_Z_mean	-0.044869	-0.252095
Steps_mean	-0.169991	-0.244500

	Activity_large_screen_usage	Activity_smoking \
Unnamed: 0	-0.089168	-0.176889
PSQI	-0.343652	-0.044031
Age	0.294815	0.068563
Weight	-0.208039	0.307995
Height	-0.280853	0.061020
BMI	-0.074093	0.343779
Cortisol_before	0.260010	-0.100976
Cortisol_after	0.024743	-0.215277
Melatonin_before	-0.177318	-0.004766
Melatonin_after	-0.013229	-0.160892
Latency_Efficiency	0.338350	0.033717
Total_minutes_in_bed	-0.167964	-0.015371
Total_sleep_time	-0.073314	0.005766
RR	-0.186665	0.287556
Daily_stress	0.006335	-0.203215
Activity_medium	0.453213	-0.230485
Activity_heavy	-0.208472	-0.384362
Activity_small_screen_usage	-0.012055	-0.201281
Activity_large_screen_usage	1.000000	0.156371
Activity_smoking	0.156371	1.000000
Activity_alcohol_assumption	0.095103	0.507213
Actigraph_X_mean	-0.065294	-0.107017
Actigraph_Y_mean	-0.109493	-0.128922
Actigraph_Z_mean	-0.036849	-0.075084
Steps_mean	0.095955	-0.141855

	Activity_alcohol_assumption	Actigraph_X_mean \
Unnamed: 0	-0.064059	-0.372321
PSQI	0.021219	0.219948
Age	-0.231811	0.063073
Weight	0.217285	0.518704
Height	-0.102489	-0.030709
BMI	0.336155	0.646711
Cortisol_before	0.077173	0.261222
Cortisol_after	0.219202	-0.087814

Melatonin_before	-0.023494	-0.163126
Melatonin_after	0.229495	0.022427
Latency_Efficiency	0.025460	0.000656
Total_minutes_in_bed	0.182225	-0.255999
Total_sleep_time	0.186612	-0.238652
RR	0.138826	-0.123526
Daily_stress	0.065525	-0.328769
Activity_medium	0.160851	0.237974
Activity_heavy	-0.353217	-0.159868
Activity_small_screen_usage	-0.042943	-0.211276
Activity_large_screen_usage	0.095103	-0.065294
Activity_smoking	0.507213	-0.107017
Activity_alcohol_assumption	1.000000	-0.024932
Actigraph_X_mean	-0.024932	1.000000
Actigraph_Y_mean	-0.048229	0.947543
Actigraph_Z_mean	-0.188746	0.878835
Steps_mean	-0.096015	0.825674

	Actigraph_Y_mean	Actigraph_Z_mean	Steps_mean
Unnamed: 0	-0.330404	-0.406643	-0.150973
PSQI	0.242861	0.117652	0.115394
Age	0.009920	0.181214	0.473888
Weight	0.515548	0.545262	0.474879
Height	0.029932	0.172590	0.080988
BMI	0.605799	0.547308	0.524388
Cortisol_before	0.178951	0.179709	0.210070
Cortisol_after	-0.130411	-0.128553	-0.047327
Melatonin_before	-0.250316	0.043414	-0.275441
Melatonin_after	-0.016156	0.076436	-0.003580
Latency_Efficiency	-0.101548	-0.187780	-0.224716
Total_minutes_in_bed	-0.123400	-0.422397	-0.147154
Total_sleep_time	-0.134615	-0.466152	-0.210776
RR	-0.134948	0.040771	-0.065279
Daily_stress	-0.301151	-0.372133	-0.066857
Activity_medium	0.315681	-0.052154	0.369597
Activity_heavy	-0.040050	-0.044869	-0.169991
Activity_small_screen_usage	-0.262728	-0.252095	-0.244500
Activity_large_screen_usage	-0.109493	-0.036849	0.095955
Activity_smoking	-0.128922	-0.075084	-0.141855
Activity_alcohol_assumption	-0.048229	-0.188746	-0.096015
Actigraph_X_mean	0.947543	0.878835	0.825674
Actigraph_Y_mean	1.000000	0.810613	0.848217
Actigraph_Z_mean	0.810613	1.000000	0.752894
Steps_mean	0.848217	0.752894	1.000000

[25 rows x 25 columns]

```
In [28]: indep_column_names = all_column_names.copy()
indep_column_names.remove("User")
indep_column_names.remove("PSQI")
indep_column_names.remove("Gender")

indep_names = [df["Age"].to_list(), df["Weight"].to_list(), df["Height"].to_list(), df["BMI"].to_list(),
                df["Cortisol_before"].to_list(), df["Cortisol_after"].to_list(),
                df["Melatonin_before"].to_list(), df["Melatonin_after"].to_list(),
                df["Latency_Efficiency"].to_list(), df["Total_minutes_in_bed"].to_list(),
                df["Total_sleep_time"].to_list(),
                df["RR"].to_list(), df["Daily_stress"].to_list(),
                df["Activity_medium"].to_list(), df["Activity_heavy"].to_list(),
                df["Activity_small_screen_usage"].to_list(), df["Activity_large_screen_usage"].to_list(),
                df["Activity_smoking"].to_list(), df["Activity_alcohol_assumption"].to_list(),
                df["Actigraph_X_mean"].to_list(), df["Actigraph_Y_mean"].to_list(),
                df["Actigraph_Z_mean"].to_list(),
                df["Steps_mean"].to_list()]
```

```
In [29]: # Make all possible combinations
comb1 = list(combinations(indep_names, 9))
comb1n = list(combinations(indep_column_names, 9))
```

```
In [31]: # Multiple regressions
def reg_m(y, x):
    ones = np.ones(len(x[0]))
    X = sm.add_constant(np.column_stack((x[0], ones)))
    for ele in x[1:]:
        X = sm.add_constant(np.column_stack((ele, X)))
    results = sm.OLS(y, X).fit()
    return results

y = PSQI
d = dict() # with all combinations
for i in range(len(comb1)):
    x = comb1[i]
    nm = comb1n[i]
    R = reg_m(y, x).rsquared
    d[R] = nm

# Sorting by R^2
sorted_d = collections.OrderedDict(sorted(d.items()))
#print(sorted_d)
print("Done!")
```

Done!

```
In [ ]: # Top combinations with the highest R^2
```

```
## 5
# (0.701168557640911, ('Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_large_screen_usage')),
# (0.7020028563928806, ('Height', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Actigraph_X_mean')),
# (0.7024721510726812, ('Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_alcohol_assumption')),
# (0.7121885543948117, ('Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Actigraph_X_mean')),
# (0.7211803785647528, ('Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Actigraph_Z_mean'))

## 6
# (0.7302144266615292, ('Height', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_smoking', 'Actigraph_X_mean')),
# (0.734410779830915, ('Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_medium', 'Actigraph_Z_mean')),
# (0.735751788866384, ('Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_medium', 'Actigraph_X_mean')),
# (0.7375615120967849, ('Height', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_medium', 'Actigraph_X_mean')),
# (0.7398835465500143, ('Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_large_screen_usage', 'Actigraph_Z_mean'))

## 7
# (0.7518467208889771, ('Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_large_screen_usage', 'Activity_smoking', 'Actigraph_Z_mean')),
# (0.7538309407285697, ('Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_large_screen_usage', 'Actigraph_Y_mean', 'Actigraph_Z_mean')),
# (0.7538755152539959, ('Weight', 'Height', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_medium', 'Actigraph_X_mean')),
# (0.7539598204394766, ('Age', 'Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Actigraph_X_mean', 'Steps_mean')),
# (0.7609064488830729, ('Height', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_medium', 'Activity_smoking', 'Actigraph_X_mean'))

## 8
# (0.7744151685726645, ('Age', 'Weight', 'Melatonin_before', 'Latency_Efficiency',
# 'Total_minutes_in_bed', 'RR', 'Actigraph_X_mean', 'Steps_mean')),
# (0.7749506409363177, ('Age', 'Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Actigraph_X_mean', 'Actigraph_Z_mean', 'Steps_mean')),
# (0.7751386502939965, ('Age', 'Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_large_screen_usage', 'Actigraph_X_mean', 'Steps_mean')),
# (0.7828979872367456, ('Age', 'Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_large_screen_usage', 'Actigraph_Z_mean', 'Steps_mean')),
# (0.812989235803329, ('Age', 'Weight', 'Melatonin_before', 'Total_minutes_in_bed',
# 'RR', 'Activity_heavy', 'Actigraph_X_mean', 'Steps_mean'))
```



```
## 9
# (0.8194901173089273, ('Age', 'Weight', 'Cortisol_after', 'Melatonin_before',
'Total_minutes_in_bed', 'RR', 'Activity_heavy', 'Actigraph_X_mean', 'Steps_mean')),
# (0.823353374745182, ('Age', 'Weight', 'Melatonin_before', 'Total_minutes_in_
bed', 'RR', 'Daily_stress', 'Activity_heavy', 'Actigraph_X_mean', 'Steps_mean')),
# (0.8275561112574359, ('Age', 'Weight', 'Melatonin_before', 'Latency_Efficiency',
'Total_sleep_time', 'RR', 'Activity_heavy', 'Actigraph_X_mean', 'Steps_mean')),
# (0.828635726174611, ('Age', 'Weight', 'Melatonin_before', 'Total_minutes_in_
bed', 'Total_sleep_time', 'RR', 'Activity_heavy', 'Actigraph_X_mean', 'Steps_mean')),
# (0.8300175899574918, ('Age', 'Weight', 'Melatonin_before', 'Latency_Efficiency',
'Total_minutes_in_bed', 'RR', 'Activity_heavy', 'Actigraph_X_mean', 'Steps_mean'))
```