

Sprawozdanie

Zbieranie i Analiza Danych

Labolatorium nr 8 – Test Grubbsa



Oskar Kwidziński

S156013

Fizyka Techniczna

Semestr zimowy 2022/2023

Wstęp

Wykonano zadanie polegające na przeprowadzeniu Testu Grubbsa dla podanego zestawu danych. Test polegał na sprawdzeniu hipotez H_0 oraz H_1 – odpowiednio: nie istnieją wartości odstające i istnieje przynajmniej jedna wartość odstająca. Dane zawierały informacje o temperaturze (temp) oraz dacie i godzinie jej pomiaru (d).

Zgodnie z założeniami testu Grubbsa wyliczono dla trzech podanych zestawów danych:

- wartość krytyczną:

$$G_{critical} = \frac{(N-1)}{\sqrt{N}} \sqrt{\frac{(t_{\alpha(2N), N-2})^2}{N-2 + (t_{\alpha(2N), N-2})^2}}$$

- wartości G_n oraz G_1 , które będą podlegały porównaniu z wartością krytyczną:

$$G_n = \frac{x_n - \bar{x}}{s} \quad G_1 = \frac{\bar{x} - x_1}{s}$$

Wskaźniki te odnoszą się odpowiednio do największej i najmniejszej wartości ze zbioru. W teście Grubbsa przyjmuje się, że gdy któraś z nich przewyższa wartość krytyczną, oznacza to że w rozpatrywanych danych występują wartości odstające.

Rozwiązanie

Po wczytaniu danych z pliku o rozszerzeniu .json, dla każdego z trzech pomiarów temperatury, dane zapisano do zmiennych d oraz temp w postaci serii biblioteki pandas:

```
def read_data(d, temp, filename):
    with open(f'{filename}.json') as f:
        data = json.load(f)
    d = pd.Series(list(itertools.chain.from_iterable([v for k, v in data.items() if f'{d}' in k])))
    temp = pd.Series(list(itertools.chain.from_iterable([v for k, v in data.items() if f'{temp}' in k])))
    return d, temp
```

Następnie dla każdego z pomiarów przeprowadzono obliczenia średniej i odchylenia standardowego temperatury, pozwalające na obliczenie wartości współczynników G_n i G_1 :

```
temp_to_std = pow((temp - temp.mean()), 2)

temp_std = pow(temp_to_std.sum() / n_temp, 0.5)

g_n = float((temp.max() - temp.mean()) / temp_std)
print("Gn:", round(g_n, 2))

g_1 = float((temp.mean() - temp.min()) / temp_std)
print("G1:", round(g_1, 2))
```

A także dla liczebności każdej z prób obliczono wartość krytyczną:

```
def calculate_gcritical_g(n, alpha):  
    t_dist = stats.t.ppf(1 - alpha / (2 * n), n-2)  
    numerator = (n - 1) * np.sqrt(np.square(t_dist))  
    denominator = np.sqrt(n) * np.sqrt(n - 2 + np.square(t_dist))  
    critical_g = numerator / denominator  
    return critical_g
```

Podsumowanie

Dla każdego z trzech rozpatrywanych pomiarów temperatury nie odnaleziono wartości odstających.

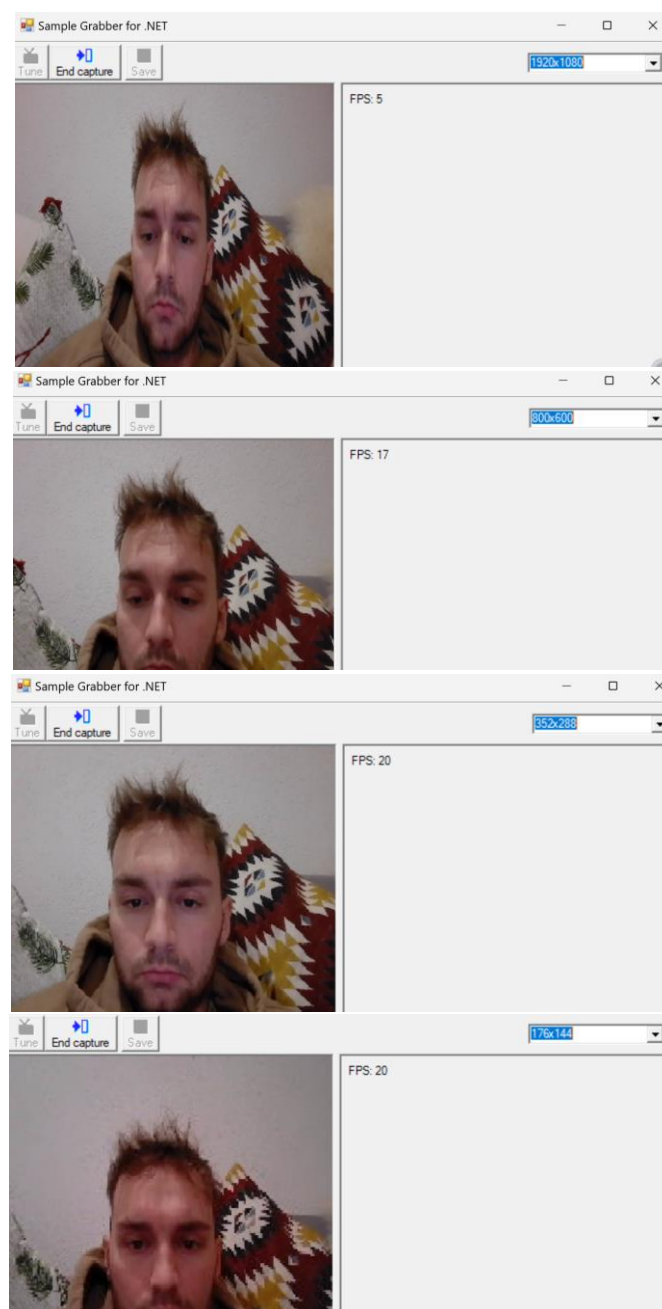
```
-----  
TEST FOR DATASET1  
Gn: 2.25  
G1: 1.39  
Gcritical: 4.1  
HYPOTHESIS H0  
  
-----  
TEST FOR DATASET2  
Gn: 1.99  
G1: 3.13  
Gcritical: 4.1  
HYPOTHESIS H0  
  
-----  
TEST FOR DATASET3  
Gn: 1.97  
G1: 3.17  
Gcritical: 4.1  
HYPOTHESIS H0
```

Przy zmianie współczynnika istotności (alpha) na większą wartość, program również nie odnalazł żadnych wartości odstających.

Wnioskuje się, że test Grubbsa w opisanym sposobie nie jest najlepszym testem jaki można przeprowadzić, aby znaleźć wartości odstające w podanym zestawie danych. Wartości, dla których występuje wyłom nie są aż tak bardzo odległe od reszty, by nazwać je punktami odstającymi. Sugeruje się zastosowanie innego testu na obecność wartości odstających lub przeprowadzenie testu wcześniej dzieląc każdy ze zbiorów danych na mniejsze części.

frame0.bmp	1/12/2023 9:25 PM	BMP File	50 KB
frame0.jpg	1/12/2023 9:25 PM	JPG File	50 KB
frame1.bmp	1/12/2023 9:25 PM	BMP File	50 KB
frame1.jpg	1/12/2023 9:25 PM	JPG File	50 KB
frame2.bmp	1/12/2023 9:25 PM	BMP File	49 KB
frame2.jpg	1/12/2023 9:25 PM	JPG File	49 KB
frame3.bmp	1/12/2023 9:25 PM	BMP File	50 KB
frame3.jpg	1/12/2023 9:25 PM	JPG File	50 KB

Przeprowadzono analizę ilości fps, którą udało się przechwycić dla każdej z rozdzielczości:



Nie udało się zwiększyć ilości fps ponad wyniki wskazane powyżej. Największą wydajność osiągnięto dla niskich rozdzielczości. Nie odnotowano zmian fps przy testach wydajności kamery np. zbliżania do źródła światła lub przy wyłączonym świetle (poza chwilowymi w trakcie zmian oświetlenia). Przypuszcza się, że powodem niskiej ilości fps jest niewydajność dysku twardego, na którym były zapisywane pliki.