

# Míra sebevražd ve světě

Tomáš Jelínek

Červenec 2023

## Abstract

Míra sebevražd je jedním z možných indikátorů duševního zdraví populace, jelikož bývají vyvrcholením závažných psychických<sup>1-4</sup> problémů jedince. Navíc jsou jasně měřitelným ukazatelem oproti prevalenci duševních poruch v populaci. V této statistické práci se zaměřím na analýzu jejich míry dle pohlaví a jak jsou ovlivňovány různými faktory jako je dostupnost duševní péče v dané zemi, GDP nebo štěstí populace.

Data jsou brána z volně dostupného datasetu na kaggle [Mental Health and Suicide Rates](#). Ten vychází z dat WHO - moje analýzy jsou dělány na údajích z roku 2016. Zdrojové kódy jsou volně dostupné na mém [GitHubu](#).

## 1 Sebevraždy u mužů a žen

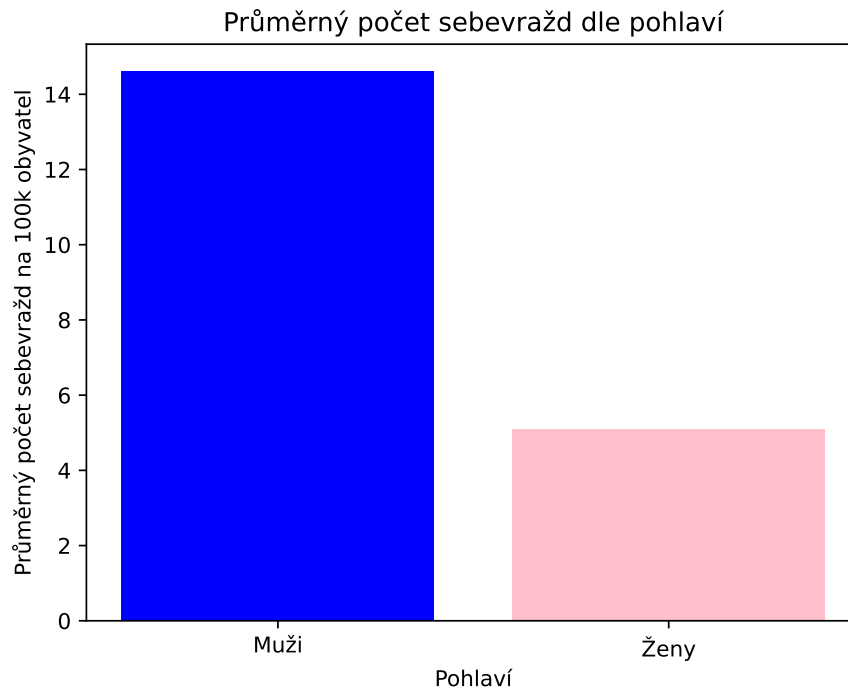
Otázkou je, zda obě dvě skupiny obyvatel jsou pod stejným rizikem. Bud'  $H_0$ , že ženy a muži jsou ve světě pod stejným rizikem a bud'  $H_1$  že jedno pohlaví má větší šanci na sebevraždu. Hypotézu ověřím dvouvýběrovým t-testem a jako hladinu významnosti volím  $\alpha = 0.05$ . T-test má určité předpoklady.

1. Nezávislost jevů - lze předpokládat, že drtivá většina sebevražd bude nezávislá na jiných.
2. Normální rozdělení - dataset je velký ( $n > 30$ ) a tedy díky centrální limitní větě si můžeme dovojit udělat t-test, i když data nemusí sama nutně mít normální rozdělení.
3. Poslední podmínkou k užití je to, že variance budou hodně podobné - vzhledem k tomu, že netuším jaké jsou, tak zvolím Welchův t-test, který tohle nepředpokládá a umožňuje je mít rozdílné.

Zde je analýza oněch dat v pythonu.

```
import numpy as np
import pandas as pd
import scipy.stats as stats
def welch_ttest():
    df = pd.read_csv("data/Age-standardized suicide rates.csv")
    df_male = df.loc[df["Sex"].str.contains("Male")]
    df_female = df.loc[df["Sex"].str.contains("Female")]
    np_males = df_male["2016"].to_numpy()
    np_females = df_female["2016"].to_numpy()
    t_statistic, p_value = stats.ttest_ind(np_males, np_females,
                                           equal_var = False)
    return(t_statistic, p_value)
print(welch_ttest())
```

Jako výstup dostávám:  $t\_statistic = 13.4164773450928$  a  $p\_value = 2.8621990375255226^{-31}$ . Vzhledem k tomu, že  $p\_value$  je výrazně nižší než na začátku zvolená  $\alpha$ , tak  $H_0$  zamítám. Pro zajímavost zde je barplot průměrné míry sebevražd pro obě pohlaví.



A tabulka zemí, kde byla míra sebevražd žen v roce 2016 vyšší než u mužů. Ve zbylých 176 zemích převažovala u mužů.

Country	Female	Male
Antigua and Barbuda	0.9	0.0
Bangladesh	6.7	5.5
China	8.3	7.9
Lesotho	32.6	22.7
Morocco	3.6	2.5
Myanmar	9.8	6.3
Pakistan	3.1	3.0

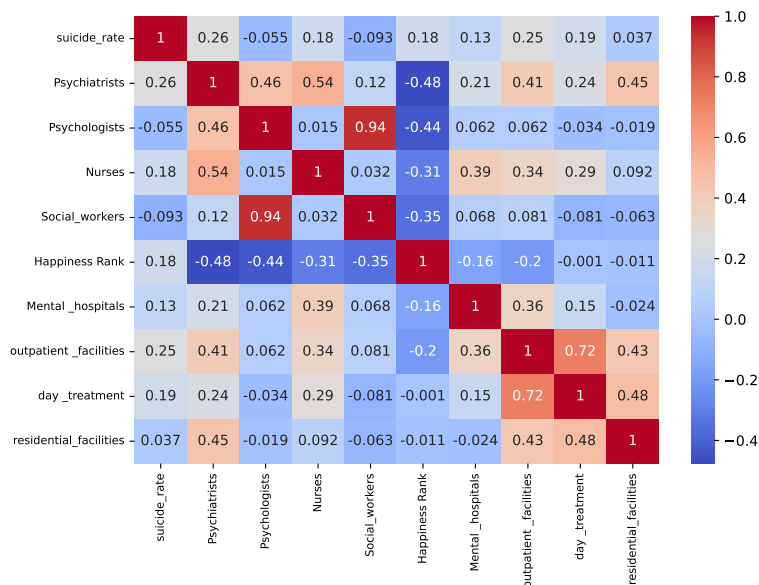
## 2 Léčebné instituce, štěstí a míra sebevražd

### 2.1 Léčebné instituce

Mezi nástroje pro zlepšování mentálního zdraví patří odborná péče od psychiatrů, psychologů, sociálních pracovníků atd. Bude tedy míra sebevražd lepší v zemích, kde je péče dostupnější? Na to vyrobím korelační matici:

```
#Other imports like in previous code snippet
import seaborn as sns
df = load_merged_data() #more about this on my GitHub
df = df.loc[df["Sex"].str.contains("Both sexes")]
df = df.rename(columns={"2016": "suicide_rate"})
df = df.loc[:, ["suicide_rate", "Psychiatrists", "Psychologists",
               "Nurses", "Social_workers",
               "Happiness Rank", "Mental
               _hospitals", "outpatient
               _facilities", "day
               _treatment", "
               residential_facilities"]]

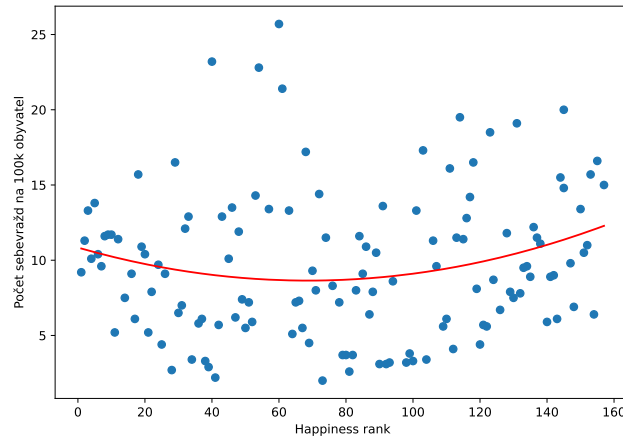
df.to_csv("correlation_matrix.csv")
corr_matrix = df.corr()
fig, ax = plt.subplots(figsize=(8, 6))
sns.heatmap(corr_matrix, annot=True, cmap="coolwarm", ax=ax)
plt.show()
```



Počty psychiatrů/psychologů, mentálních institucí atd. na hlavu nemají samy o sobě téměř žádnou korelaci s mírou sebevražd, i když by člověk na první pohled asi očekával, že budou mít negativní.

## 2.2 Štěstí

Zajímavé je, že ani rank státu v žebříčku šťastnosti obyvatel s tím nekoreluje, byť je v datech patrná mírná polynomiální závislost. (Data štěstí jsou z tohoto [datasetu](#)).



```
def happiness_regression(df, degree):
    df = df.loc[df["Sex"].str.contains("Both sexes")]

    x = df["Happiness Rank"]
    y = df["2016"]

    coeffs = np.polyfit(x, y, degree)
    poly_eqn = np.poly1d(coeffs)

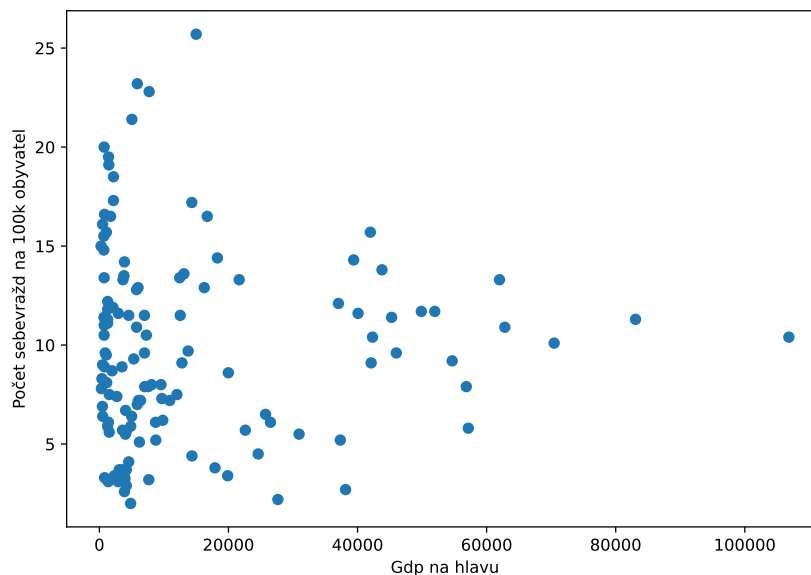
    plt.scatter(x, y)
    plt.plot(x, poly_eqn(x), color="red")

    plt.xlabel("Happiness rank")
    plt.ylabel("Pocet sebevrazd na 100k obyvatel")

    plt.show()
```

## 2.3 GDP

Člověk někdy na [internetu narazí na tvrzení](#), že v bohatších zemích lidé častěji páchají sebevraždy.



```
def gdp_rates_plot(df, region = None):
    df = df.loc[df["Sex"].str.contains("Both sexes")]
    if region is not None:
        df = df.loc[df["Region"].str.contains(region)]
    df = df.dropna(subset=['2016_x', '2016_y'])

    np_rates = df["2016_x"].to_numpy()
    np_gdp = df["2016_y"].to_numpy()

    plt.scatter(np_gdp, np_rates)
    plt.xlabel("Gdp na hlavu")
    plt.ylabel("Pocet sebevrazd na 100k obyvatel")

    plt.show()
```

Z plotu dat je vidět, že u tohoto tvrzení musí být člověk opatrný. Byť je v bohatších zemích míra sebevražd značně vyšší než v některých chudších, existují i chudé oblasti, kde je míra sebevražd vyšší nebo stejná jak v bohatých regionech.

### 3 Slovo závěrem

Pokud vás případně zajímají další statistiky, jako je například vývoj sebevražd v čase, jiné regrese, či analýza dle regionů (Evropa, Afrika...). Můžete si je nechat udělat ze zdrojů na mém [GitHubu](#).

### Reference

- <sup>1</sup> J. M. Bertolote and A. Fleischmann. A global perspective in the epidemiology of suicide. *Suicidologi*, 7(2), 2002.
- <sup>2</sup> J. M. Bertolote, A. Fleischmann, D. De Leo, and D. Wasserman. Psychiatric diagnoses and suicide: revisiting the evidence. *Crisis*, 25(4):147–155, 2004.
- <sup>3</sup> B. Chang, D. Gitlin, and R. Patel. The depressed patient and suicidal patient in the emergency department: evidence-based management and treatment strategies. *Emergency medicine practice*, 13(9):1–23, 2011.
- <sup>4</sup> A. J. Ferrari, R. E. Norman, G. Freedman, A. J. Baxter, J. E. Pirkis, M. G. Harris, A. Page, E. Carnahan, L. Degenhardt, T. Vos, et al. The burden attributable to mental and substance use disorders as risk factors for suicide: findings from the global burden of disease study 2010. *PloS one*, 9(4):e91936, 2014.