

Analyse der Internetnutzung und Schülerleistung

In dieser Analyse untersuchen wir den Zusammenhang zwischen der Internetnutzung und der schulischen Leistung von Schülern. Wir betrachten, wie die Nutzung des Internets das Lernverhalten, die Motivation und die Ergebnisse der Schüler beeinflusst.

Durch die Analyse von Daten zur Internetnutzung und schulischen Leistung können wir wichtige Erkenntnisse gewinnen, die uns helfen, Bildungsstrategien zu verbessern und den Schülern dabei zu helfen, ihr volles Potenzial auszuschöpfen.

Wir werden verschiedene Variablen betrachten, wie z.B. die Zeit, die Schüler online verbringen, die Art der Internetnutzung (z.B. für schulische Zwecke oder Freizeitaktivitäten) und die Auswirkungen auf ihre schulischen Leistungen.

Zu Beginn werden die Daten geladen und eine explorative Datenanalyse durchgeführt, um die Verteilung der Variablen zu untersuchen und mögliche Zusammenhänge zu identifizieren. Anschließend werden wir verschiedene statistische Methoden anwenden, um den Zusammenhang zwischen Internetnutzung und Schülerleistung zu untersuchen.

Laden der Daten

```
In [38]: import pandas as pd
import numpy as np
import plotly.express as px
import statsmodels.api as sm
from statsmodels.formula.api import ols

path_prefix = 'Daten/'

# Load the data
pisa_math = pd.read_csv(path_prefix+'pisaMathe.csv', index_col='Jahr', parse_dates=
pisa_nature = pd.read_csv(path_prefix+'pisaNatur.csv', index_col='Jahr', parse_date
pisa_reading = pd.read_csv(path_prefix+'pisaLese.csv', index_col='Jahr', parse_date

internet_usage = pd.read_csv(path_prefix+'internetNutzung.csv', index_col='Jahr', p
print_usage = pd.read_csv(path_prefix+'printNutzung.csv', index_col='Jahr', parse_d
politics = pd.read_csv(path_prefix+'politischesInteresse.csv', index_col='Jahr', pa
music_school_attendance = pd.read_csv(path_prefix+'schulerMusikschulen.csv', index
number_of_students = pd.read_csv(path_prefix+'anzahlSchuler.csv', index_col='Jahr',

inet_usage_cat3_raw = pd.read_csv(path_prefix+'JIM3.csv')
inet_map = inet_usage_cat3_raw.loc[:,['Kategorie1', 'Kategorie2']].groupby('Kategor
inet_usage_cat3 = inet_usage_cat3_raw.set_index(['Aktivität', 'Kategorie1', 'Kategori
inet_usage_cat = pd.concat([inet_usage_cat3, inet_map], axis=1).groupby('Kategorie2
inet_usage_cat.index.name = 'Jahr'
```

```

if 'Soziales Netzwerk' not in inet_usage_cat.columns:
    inet_usage_cat['Soziales Netzwerk'] = 0
# extend index to 2000
new_idx = [str(i) for i in range(2000, 2024)]
# set 0 to na
inet_usage_cat = inet_usage_cat.replace(0, np.nan)
inet_usage_cat = inet_usage_cat.reindex(new_idx).bfill()
inet_usage_cat.index = pd.to_datetime(inet_usage_cat.index)

# Merge the data
data_raw = pd.concat([pisa_math, pisa_nature, pisa_reading, print_usage, politics,
data = data_raw.sort_index().ffill().bfill()
data = data.loc['2000:']

```

Berechnung der Scores

```

In [40]: data['Internet Gewicht'] = data['Internet Minuten'] / data['Internet Minuten'].loc[
data['Info Score Internet'] = data['Information'] / (data['Kommunikation'] + data['
data['Info Score Print'] = (data['Tageszeitung'] + data['Zeitschriften']) / (data['
data['Hobby Score'] = data['Musikschüler'] / data['Schüler in Deutschland']
data['Bildungs Score'] = (data['Info Score Internet'] + data['Info Score Print'] +
data['PISA Score'] = (data['Mathematik'] + data['Naturwissenschaften'] + data['Lese

```

Explorative Datenanalyse

```

In [41]: fig = px.line(data, y=['Information', 'Kommunikation', 'Soziales Netzwerk', 'Unterha
fig.update_layout(title='Internetnutzung kumulierte Häufigkeiten', autosize=False, w
fig.update_yaxes(title='Kumulierte Häufigkeit')
fig.update_layout(legend=dict(
    orientation='h',
    yanchor='bottom',
    y=-0.2,
    xanchor='right',
    x=0.9,
    title=''
))
# color template
fig.show()

```

```

In [42]: fig = px.line(data, y=['Bücher', 'Zeitschriften', 'Tageszeitung'], template='plotly_w
fig.update_layout(title='Printmedien Nutzung mehrmals die Woche (%)', autosize=False
fig.update_yaxes(title='Prozent')
fig.update_layout(legend=dict(
    orientation='h',
    yanchor='bottom',
    y=-0.2,
    xanchor='right',
    x=0.75,
    title=''
))
fig.show()

```

```
In [43]: fig = px.line(data, y=['Schüler in Deutschland', 'Musikschüler'], template='plotly_
fig.update_layout(title='Anzahl Schüler und Musikschüler',autosize=False, width=800
fig.update_yaxes(title='Anzahl')
fig.update_layout(legend=dict(
    orientation='h',
    yanchor='bottom',
    y=-0.2,
    xanchor='right',
    x=0.75,
    title=''
))
fig.show()
```

```
In [44]: fig = px.line(data, y=['Info Score Internet', 'Info Score Print', 'Hobby Score', 'B
fig.update_layout(title='Scores über die Zeit',autosize=False, width=800,height=600
fig.update_yaxes(title='Score')
fig.update_layout(legend=dict(
    orientation='h',
    yanchor='bottom',
    y=-0.2,
    xanchor='right',
    x=0.9,
    title=''
))
fig.show()
```

```
In [45]: fig = px.imshow(data.loc[:,['Mathematik', 'Naturwissenschaften', 'Lesekompetenz'],'I
fig.update_layout(title='Korrelations Matrix',autosize=False, width=800,height=600)
fig.show()
```

```
In [46]: # plot pisa
fig = px.line(data, y=['Mathematik', 'Naturwissenschaften', 'Lesekompetenz'], templ
fig.update_layout(title='PISA Ergebnisse 2000-2024',autosize=False, width=800,height
fig.update_yaxes(title='Punkte')
fig.update_layout(legend=dict(
    orientation='h',
    yanchor='bottom',
    y=-0.2,
    xanchor='right',
    x=0.9,
    title=''
))
fig.show()
```

Durchführung statistischer Tests

```
In [47]: plt_data = data.loc[:,['PISA Score', 'Bildungs Score']]
plt_data = plt_data / plt_data.loc['2000'].iloc[0]
fig = px.line(plt_data, y=['PISA Score', 'Bildungs Score'], template='plotly_white'
fig.update_layout(title='Score Vergleiche',autosize=False, width=800,height=600)
fig.update_yaxes(title='Normalisierter Score')
fig.update_layout(legend=dict(
```

```

orientation='h',
yanchor='bottom',
y=-0.2,
xanchor='right',
x=0.7,
title=''
))
fig.show()

```

```

In [48]: # fit a linear model with formula
reg_data = data.loc[:,['PISA_Score', 'Bildungs_Score']]
reg_data.columns = ['PISA_Score', 'Bildungs_Score']
formula = 'PISA_Score ~ Bildungs_Score'
model = ols(formula, reg_data).fit()
model.summary()

```

Out[48]:

OLS Regression Results

Dep. Variable:	PISA_Score	R-squared:	0.463
Model:	OLS	Adj. R-squared:	0.438
Method:	Least Squares	F-statistic:	18.93
Date:	Di, 09 Jul 2024	Prob (F-statistic):	0.000256
Time:	20:40:41	Log-Likelihood:	-81.609
No. Observations:	24	AIC:	167.2
Df Residuals:	22	BIC:	169.6
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	417.1504	19.467	21.429	0.000	376.779	457.522
Bildungs_Score	254.6399	58.524	4.351	0.000	133.269	376.011

Omnibus:	1.141	Durbin-Watson:	0.536
Prob(Omnibus):	0.565	Jarque-Bera (JB):	0.867
Skew:	-0.133	Prob(JB):	0.648
Kurtosis:	2.108	Cond. No.	42.0

Notes:

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

Export der Latex Tabelle

```
In [49]: print(model.summary().as_latex())
```

```
\begin{center}
\begin{tabular}{lclcl}
\toprule
\textbf{Dep. Variable:} & & PISA\_Score & & \textbf{R-squared:} & & & & \\
0.463 & & & & & & & & \\
\textbf{Model:} & & OLS & & \textbf{Adj. R-squared:} & & & & \\
0.438 & & & & & & & & \\
\textbf{Method:} & & Least Squares & & \textbf{F-statistic:} & & & & \\
18.93 & & & & & & & & \\
\textbf{Date:} & & Di, 09 Jul 2024 & & \textbf{Prob (F-statistic):} & & 0.0 & & \\
00256 & & & & & & & & \\
\textbf{Time:} & & 20:40:41 & & \textbf{Log-Likelihood:} & & & & -8 \\
1.609 & & & & & & & & \\
\textbf{No. Observations:} & & 24 & & \textbf{AIC:} & & & & \\
167.2 & & & & & & & & \\
\textbf{Df Residuals:} & & 22 & & \textbf{BIC:} & & & & \\
169.6 & & & & & & & & \\
\textbf{Df Model:} & & 1 & & & & & & \\
\\
\textbf{Covariance Type:} & & nonrobust & & & & & & \\
\\
\bottomrule
\end{tabular}
\begin{tabular}{lcccccc}
& & & & \textbf{coef} & & \textbf{std err} & & \textbf{t} & & \textbf{P} \\
$> |$t$|$ & & & & \textbf{[0.025]} & & \textbf{[0.975]} & & & & \\
\midrule
\textbf{Intercept} & & 417.1504 & & 19.467 & & 21.429 & & & & \\
0.000 & & 376.779 & & 457.522 & & & & & & \\
\textbf{Bildungs\_Score} & & 254.6399 & & 58.524 & & 4.351 & & & & \\
0.000 & & 133.269 & & 376.011 & & & & & & \\
\bottomrule
\end{tabular}
\begin{tabular}{lclcl}
\textbf{Omnibus:} & & 1.141 & & \textbf{Durbin-Watson:} & & 0.536 & & \\
\textbf{Prob(Omnibus):} & & 0.565 & & \textbf{Jarque-Bera (JB):} & & 0.867 & & \\
\textbf{Skew:} & & -0.133 & & \textbf{Prob(JB):} & & 0.648 & & \\
\textbf{Kurtosis:} & & 2.108 & & \textbf{Cond. No.} & & 42.0 & & \\
\bottomrule
\end{tabular}
%\caption{OLS Regression Results}
\end{center}
```

Notes: \newline

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