ADP_Red\stats\3.chi-squared.py

```python
# %% Chapter 6. Statistics - Chi-Square Test
# = Cross-Tabulation Test
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

import scipy.stats as stats


# %% 1. 데이터 수집
df = pd.read_csv('../../ADP_Python/data/titanic.csv')

print(df.shape)
print(df.info())

# Check binary variable
for i, var in enumerate(df.columns):
    print(i, var, len(df[var].unique()))

# Check data summary
print(df.describe())

# 2. 데이터 결측치 보정
print(df.isna().sum())


# %% 9. 모델 학습 - Stats
import scipy.stats as stats
import statsmodels.api as sm
import statsmodels.formula.api as smf

# Chi-Square Test for Homogeneity
# 관찰빈도가 기대빈도와 일치하는지 검정한다.
df_value_counts = df[df['survived'] == 1]['sex'].value_counts()
print(stats.chisquare(df_value_counts))

# Chi-Square Test for Independency
# 두 독립변수가 독립인지 관찰도수를 기반으로 검정한다.
df_crosstab = pd.crosstab(df['class'], df['survived'])
print(stats.chi2_contingency(df_crosstab))
```