

Лабораторная работа 11 (итоговая)

Статистическая обработка данных психологического исследования

Гузовская Александра Чеславовна
Б9123-01.03.02сп

16 июня 2025 г.

Структура данных исследования

Данные представляют психологическое исследование. Поле `id` обозначает номер респондента. Рассмотрим каждый лист отдельно.

Социофобия

Лист предоставляет данные опросника о социофобии. Поле `Общий` - сумма баллов по всем вопросам. Поля `Шк1`, `Шк2`, `Шк3`, `Шк4`, `Шк5` - сумма баллов по шкалам.

ПривязанностьМ

Лист предоставляет данные опросника о привязанности к матери. Поле `н` - балл по шкале надёжной привязанности. Поля `т` и `и` - баллы по шкалам тревожной и избегающей привязанности.

Тревожность

Лист предоставляет данные опросника о тревожности. Поле `Общая` - сумма баллов по всем вопросам. Поля `Школьная`, `Самооценочная`, `Межличностная` - сумма баллов по соответствующим шкалам тревожности.

Негатив

Лист предоставляет данные опросника о страхе негативной оценки. Поле `Сумма` - сумма баллов по шкале страха негативной оценки.

Страхи

Лист предоставляет данные опросника о страхах. Поле `Сумма` - сумма баллов по шкале страха. Остальные поля — баллы по каждому из страхов по отдельности.

id	Животные	Темнота	Сумасшествие	Болезнь близких	...	Сумма
1	5	1	2	9	...	130
2	6	2	3	5	...	75
3	3	1	4	10	...	78

Проверка гипотез

Требуется рассмотреть две группы:

надёжно привязанных к матери (secure) и ненадёжно (insecure).

Привязанность к матери считается надёжной, если балл по шкале $n > 5$.

Уровень значимости $\alpha = 0.01$.

Для проверки гипотез о равенстве средних используем `scipy.stats.ttest_ind` с параметром `equal_var`. Перед использованием проверим гипотезу о равенстве дисперсий с помощью критерия Фишера.

Гипотеза (X и Y – две сравниваемые выборки):

- $H_0 : \sigma_X = \sigma_Y$
- $H_1 : \sigma_X \neq \sigma_Y$

Генеральные дисперсии оцениваются на основе выборок, и сам критерий непосредственно рассчитывается как отношение одной выборочной дисперсии к другой

Статистика:

$$F_{n-1, m-1} = \frac{s_X^2}{s_Y^2} \sim F(n-1, m-1)$$

p -значение:

$$p = 2 \min (F_{n-1, m-1}(f), 1 - F_{n-1, m-1}(f))$$

где $F_{n-1, m-1}(f)$ – функция распределения Фишера с $n-1$ и $m-1$ степенями свободы.

Гипотеза о значимости различия уровня социофобии

Проверяем гипотезу о наличии значимых различий в уровне социофобии между генеральными совокупностями лиц с разным типом привязанности к матери.

H_0 : В генеральной совокупности средний уровень социофобии (по шкале **Общий**) у лиц с надёжной и ненадёжной привязанностью к матери статистически не различается

- $H_0 : E[X_{\text{secure}}] = E[X_{\text{insecure}}]$
- $H_1 : E[X_{\text{secure}}] < E[X_{\text{insecure}}]$

Результаты:

- Данные надёжно привязанных: нормальное распределение ($p=0.4241$, критерий Колмогорова-Смирнова)
- Данные ненадёжно привязанных: нормальное распределение ($p=0.7090$, критерий Колмогорова-Смирнова)

- Равенство дисперсий (критерий Фишера): $p=0.7052$
- t-критерий Стьюдента: $p=5.35 \times 10^{-6}$

Вывод: Отвергаем H_0 ($p < 0.01$). Уровень социофобии значимо выше у ненадёжно привязанных.

Гипотеза о значимости различия уровня тревожности

Аналогично проверяем для всех шкал тревожности:

Шкала	Нормальность (secure)	Нормальность (insecure)	p-value	Критерий
Общая	0.6402	0.9566	0.5028	t-критерий
Школьная	0.8024	0.5031	0.8658	t-критерий
Самооценочная	0.8170	0.6383	0.1173	t-критерий
Межличностная	0.8343	0.6918	0.9351	t-критерий

H_0 : В генеральной совокупности средний уровень тревожности (по шкалам **Общая**, **Школьная**, **Самооценочная**, **Межличностная**) у лиц с надёжной и ненадёжной привязанностью к матери статистически не различается

Вывод: Для всех шкал тревожности отвергаем H_0 ($p < 0.01$). У ненадёжно привязанных уровень тревожности значимо выше.

Гипотеза о страхе негативной оценки

H_0 : В генеральной совокупности среднее значение страха негативной оценки (по шкале **Сумма негатива**) у лиц с надёжной и ненадёжной привязанностью к матери статистически не различается

Альтернативная гипотеза: Уровень страха негативной оценки у ненадёжно привязанных к матери выше, чем у надёжно

- Нормальность: secure ($p=0.8171$), insecure ($p=0.1277$)
- Равенство дисперсий: $p=0.8748$
- t-критерий: $p=0.001135$

Вывод: Отвергаем H_0 ($p < 0.01$). Страх негативной оценки значимо выше у ненадёжно привязанных.

Гипотеза об общем уровне страхов

H_0 : В генеральной совокупности среднее значение общего уровня страха (по шкале **Сумма страхов**) у лиц с надёжной и ненадёжной привязанностью к матери статистически не различается

Альтернативная гипотеза: Общий балл страхов у ненадёжно привязанных к матери выше, чем у надёжно

- Нормальность: secure ($p=0.9464$), insecure ($p=0.5083$)
- Равенство дисперсий: $p=0.2488$
- t-критерий: $p=0.000365$

Вывод: Отвергаем H_0 ($p < 0.01$). Общий уровень страхов значимо выше у ненадёжно привязанных.

Корреляционный анализ

H_0 : Количество статистически значимых корреляций между общей тревожностью и конкретными страхами не отличается между группой с надёжной и ненадёжной привязанностью к матери

Сравнение количества значимых корреляций ($p < 0.01$) между общей тревожностью и конкретными страхами:

- Надёжно привязанные:
 - По Пирсону: 11
 - По Спирмену: 11
- Ненадёжно привязанные:
 - По Пирсону: 19
 - По Спирмену: 18

Вывод: У ненадёжно привязанных обнаружено больше значимых корреляционных связей между общей тревожностью и страхами.

Код программы

```
import pandas as pd
import numpy as np
import scipy.stats as sps
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.ExcelFile('C:/Users/AliceWolf13/Documents/mathstats/last_chance_for_4/psyda

socio = pd.read_excel(data, 'Социофобия')
attach = pd.read_excel(data, 'ПривязанностьМ')
anxiety = pd.read_excel(data, 'Тревожность')
```

```

negativity = pd.read_excel(data, 'Негатив')
fears = pd.read_excel(data, 'Страхы')

df = socio.merge(attach, on='id', suffixes=('_socio', '_attach'))
df = df.merge(anxiety, on='id', suffixes=('', '_anxiety'))
df = df.merge(negativity, on='id', suffixes=('', '_neg'))
df = df.merge(fears, on='id', suffixes=('', '_fears'))

df = df.rename(columns={
    'Сумма': 'Сумма негатива',
    'Сумма_fears': 'Сумма страхов'
})

df['attachment'] = np.where(df['Н'] > 5, 'secure', 'insecure')
secure = df[df['attachment'] == 'secure']
insecure = df[df['attachment'] == 'insecure']

def check_normality(data, alpha=0.01):
    if len(data) <= 30:
        stat, p = sps.shapiro(data)
        test_name = "Shapiro-Wilk"
    else:
        stat, p = sps.kstest(data, 'norm', args=(np.mean(data), np.std(data)))
        test_name = "Kolmogorov-Smirnov"
    return p, test_name

def test_hypothesis(data1, data2, col_name, alpha=0.01):
    print(f"\nАнализ переменной: {col_name}")

    p1, test_name1 = check_normality(data1[col_name])
    p2, test_name2 = check_normality(data2[col_name])

    print(f"Тест на нормальность ({test_name1}): secure p-value={p1:.4f}")
    print(f"Тест на нормальность ({test_name2}): insecure p-value={p2:.4f}")

    if p1 > alpha and p2 > alpha:
        if p1 > alpha and p2 > alpha:
            var1 = np.var(data1[col_name], ddof=1)
            var2 = np.var(data2[col_name], ddof=1)
            f_value = max(var1, var2) / min(var1, var2)
            df1 = len(data1[col_name]) - 1
            df2 = len(data2[col_name]) - 1
            p_var = 2 * min(sps.f.cdf(f_value, df1, df2), 1 - sps.f.cdf(f_value, df1,
            equal_var = p_var > alpha
            print(f"Fisher's test p-value: {p_var:.4f}")

        if equal_var:
            t_stat, p_val = sps.ttest_ind(data1[col_name], data2[col_name], equal_var=

```

```

        test_name = "Student's t-test"
    else:
        t_stat, p_val = sps.ttest_ind(data1[col_name], data2[col_name], equal_var=False)
        test_name = "Welch's t-test"
    else:
        u_stat, p_val = sps.mannwhitneyu(data1[col_name], data2[col_name], alternative='two-sided')
        test_name = "Mann-Whitney U"

    h_stat, p_val_kw = sps.kruskal(data1[col_name], data2[col_name])
    print(f"Kruskal-Wallis p-value: {p_val_kw:.4f}")

    print(f"Использован тест: {test_name}")
    print(f"p-value: {p_val:.4f}")

    if p_val < alpha:
        print(f"Отвергаем H0: статистически значимые различия обнаружены (p={p_val:.4f})")
        print(f"Уровень {col_name} значимо выше у ненадёжно привязанных.")
    else:
        print(f"Не отвергаем H0: статистически значимых различий не обнаружено (p={p_val:.4f})")
        print(f"Уровень {col_name} не отличается между группами.")

# 1. Проверка гипотезы о социофобии
# Уровень социофобии у ненадёжно привязанных к матери выше, чем у надёжно
test_hypothesis(insecure, secure, 'Общий')

# 2. Проверка гипотез о тревожности
# Уровень тревожности (общей, школьной, самооценочной, межличностной) у ненадёжно привязанных к матери выше, чем у надёжно
for a_type in ['Общая', 'Школьная', 'Самооценочная', 'Межличностная']:
    test_hypothesis(insecure, secure, a_type)

# 3. Проверка гипотезы о страхе негативной оценки
# Уровень страха негативной оценки у ненадёжно привязанных к матери выше, чем у надёжно
test_hypothesis(insecure, secure, 'Сумма негатива')

# 4. Проверка гипотезы об общем балле страхов
# Общий балл страхов у ненадёжно привязанных к матери выше, чем у надёжно
test_hypothesis(insecure, secure, 'Сумма страхов')

# 5. Корреляционный анализ
# У какой из групп больше корреляционных связей между общей тревожностью и каждым из
fear_columns = [col for col in fears.columns if col not in ['id', 'Сумма']]

def count_correlations(group, columns, alpha=0.01):
    count = 0
    count2 = 0
    for col in columns:
        if col in group.columns:
            try:

```

```

        r, p = sps.pearsonr(group['Общая'], group[col])
        if p < alpha:
            count += 1
        r, p = sps.spearmanr(group['Общая'], group[col])
        if p < alpha:
            count2 += 1
    except:
        continue
return count, count2

secure_corr = count_correlations(secure, fear_columns)
insecure_corr = count_correlations(insecure, fear_columns)

def calculate_correlations(group):
    return {col: sps.spearmanr(group['Общая'], group[col])
            for col in fear_columns if col in group.columns}

def prepare_data(group):
    corrs = calculate_correlations(group)
    return pd.DataFrame.from_dict(
        {k: {'r': v[0], 'p': v[1]} for k, v in corrs.items()}
    ).T

secure_data = prepare_data(secure)
insecure_data = prepare_data(insecure)

corr_data = pd.DataFrame({
    'Надёжная': secure_data['r'],
    'Ненадёжная': insecure_data['r']
})

p_values = pd.DataFrame({
    'Надёжная': secure_data['p'],
    'Ненадёжная': insecure_data['p']
})

plt.figure(figsize=(10, 4))
sns.heatmap(
    corr_data,
    annot=True,
    fmt=".2f",
    cmap="coolwarm",
    center=0,
    mask=p_values > 0.01,
    annot_kws={"fontsize": 9}
)
plt.title("Значимые корреляции между страхами и общей тревожностью у исследуемых групп")
plt.tight_layout()

```

```
plt.show()

print(f"\nКоличество значимых корреляций между общей тревожностью и страхами:")
print(f"Надёжно привязанные: {secure_corr}")
print(f"Ненадёжно привязанные: {insecure_corr}")

if secure_corr > insecure_corr:
    print("Вывод: у надёжно привязанных больше значимых корреляций")
elif insecure_corr > secure_corr:
    print("Вывод: у ненадёжно привязанных больше значимых корреляций")
else:
    print("Вывод: одинаковое количество значимых корреляций")
```

Вывод программы

Анализ переменной: Общий

Тест на нормальность (Kolmogorov-Smirnov): secure p-value=0.4241

Тест на нормальность (Kolmogorov-Smirnov): insecure p-value=0.7090

Fisher's test p-value: 0.5143

Использован тест: Student's t-test

p-value: 0.0000

Отвергаем H0: статистически значимые различия обнаружены (p=0.0000). Уровень Общий зн

Анализ переменной: Общая

Тест на нормальность (Kolmogorov-Smirnov): secure p-value=0.6402

Тест на нормальность (Kolmogorov-Smirnov): insecure p-value=0.9566

Fisher's test p-value: 0.5028

Использован тест: Student's t-test

p-value: 0.0000

Отвергаем H0: статистически значимые различия обнаружены (p=0.0000). Уровень Общая зн

Анализ переменной: Школьная

Тест на нормальность (Kolmogorov-Smirnov): secure p-value=0.8024

Тест на нормальность (Kolmogorov-Smirnov): insecure p-value=0.5031

Fisher's test p-value: 0.8658

Использован тест: Student's t-test

p-value: 0.0009

Отвергаем H0: статистически значимые различия обнаружены (p=0.0009). Уровень Школьная

Анализ переменной: Самооценочная

Тест на нормальность (Kolmogorov-Smirnov): secure p-value=0.8170

Тест на нормальность (Kolmogorov-Smirnov): insecure p-value=0.6383

Fisher's test p-value: 0.1173

Использован тест: Student's t-test

p-value: 0.0001

Отвергаем H0: статистически значимые различия обнаружены (p=0.0001). Уровень Самооцен

Анализ переменной: Межличностная

Тест на нормальность (Kolmogorov-Smirnov): secure p-value=0.8343

Тест на нормальность (Kolmogorov-Smirnov): insecure p-value=0.6918

Fisher's test p-value: 0.9351

Использован тест: Student's t-test

p-value: 0.0000

Отвергаем H0: статистически значимые различия обнаружены (p=0.0000). Уровень Межлично

Анализ переменной: Сумма негатива

Тест на нормальность (Kolmogorov-Smirnov): secure p-value=0.8171

Тест на нормальность (Kolmogorov-Smirnov): insecure p-value=0.1277

Fisher's test p-value: 0.8748

Использован тест: Student's t-test

p-value: 0.0011

Отвергаем H0: статистически значимые различия обнаружены (p=0.0011). Уровень Сумма не

Анализ переменной: Сумма страхов

Тест на нормальность (Kolmogorov-Smirnov): secure p-value=0.9464

Тест на нормальность (Kolmogorov-Smirnov): insecure p-value=0.5083

Fisher's test p-value: 0.2985

Использован тест: Student's t-test

p-value: 0.0004

Отвергаем H0: статистически значимые различия обнаружены (p=0.0004). Уровень Сумма ст

Количество значимых корреляций между общей тревожностью и страхами:

Надёжно привязанные: (11, 11)

Ненадёжно привязанные: (19, 18)

Вывод: у ненадёжно привязанных больше значимых корреляций

