An Analysis of Number of Firms in Metropolitan Areas by Owner Sex

Turgay TURKER

ASRI2022

TABLE OF CONTENTS

INTRODUCTION

Despite many similarities between men and women when it comes to running a business, differences do exist.

In addition, Mahto et al. [1] showed that the gender of the firm owner has a significant effect on the overall performance of the firm.

INTRODUCTION

On the other hand, studies reveal that women and men do not differ in measures of firm survival or profitability when we control for factors such as industry, firm size, and firm age [2].

INTRODUCTION

Although men are ahead in firm ownership, the number of women-owned firms rose 26.8 percent from 2007 to 2012, from 7.8 million to 9.9 million businesses. In contrast, the number of all firms increased 2.0 percent during the same period, from 27.1 million to 27.6 million [3].

Women-owned firms made up only 19.9% of all firms that employed people in the United States in 2018 but their numbers are growing [4].

INTRODUCTION

In this study, firm ownership in 101 metropolitan cities with more than 50,000 firms in the USA was analyzed by gender.

"*Firms by Metropolitan Statistical Area (MSA) and Owner Sex for MSAs with more than 50,000 Firms*" published by the Small Business Administration (SBA) is used as the data set [5].

HYPOTHESIS & METHODS

In this study, it has been tested whether there is a significant difference in terms of the number of firms located in metropolitan areas and the gender of the firm owner using the previously mentioned data set. Hypotheses:

- Null hypothesis (H_0): There is no significant difference between groups
- Alternative hypothesis (H_a): There is a significant difference between groups.

HYPOTHESIS & METHODS

#	Geographical Area	Majority Female-Owned	Majority Male-Owned
1	New York-Newark-Jersey City, NY-NJ-PA Metro Area	831432	1401142
2	Los Angeles-Long Beach-Anaheim, CA Metro Area	626632	946881
3	Miami-Fort Lauderdale-West Palm Beach, FL Metro Area	434496	596885
	...		
	...		
	...		
99	Boulder, CO Metro Area	20003	26092
100	Syracuse, NY Metro Area	17584	29065
101	Springfield, MA Metro Area	17596	29061

HYPOTHESIS & METHODS

An important decision point when working with a sample of data is whether to use parametric or non-parametric statistical methods.

Parametric statistical methods assume that the data has a known and specific distribution, often a Gaussian distribution.

If a data sample is not Gaussian, then the assumptions of parametric statistical tests are violated and nonparametric statistical methods must be used.

HYPOTHESIS & METHODS

There are a range of techniques that you can use to check if your data sample deviates from a Gaussian distribution, called normality tests. So, first I will test whether the data sample is normally distributed.
Hypotheses:

- H_0: Sample looks Gaussian
- H_a: Sample does not look Gaussian

HYPOTHESIS & METHODS

There are many statistical tests that we can use to quantify whether a sample of data looks as though it was drawn from a Gaussian distribution.

Shapiro-Wilk test in Python was used for checking normality.

The Shapiro () SciPy function will calculate the Shapiro-Wilk on a given dataset. The function returns both the W-statistic calculated by the test and the p-value.

HYPOTHESIS & METHODS

As I mentioned before, If a data sample is Gaussian, then we can use parametric statistical methods.

However, If a data sample is not Gaussian, then the assumptions of parametric statistical tests are violated, and nonparametric statistical methods must be used.

As will be seen in the next slides, my data sample is not Gaussian. So, the Mann-Whitney U test, one of the non-parametric statistical methods, was used in this study.

RESULTS

```
import pandas as pd
!pip install researchpy
import researchpy as rp
import scipy.stats as stats
import matplotlib.pyplot as plt
```

```
file = 'dataPYTHON.xlsx'
df = pd.read_excel(file)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 202 entries, 0 to 201
Data columns (total 3 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   sex            202 non-null    object
 1   metro          202 non-null    int64
 2   number_of_firm 202 non-null    int64
dtypes: int64(2), object(1)
memory usage: 4.9+ KB
```

sex	metro	number_of_firm
female	1	831432
female	2	626632
female	3	434496
female	4	367299
female	5	304383
female	6	291563
female	7	300130

```
from scipy.stats import shapiro
stat, p = shapiro(sampling_difference)
print('Statistic=%.4f, p=%.4f' % (stat, p))

alpha = 0.05
if p > alpha:
    print('Sample looks Gaussian (fail to reject H0)')
else:
    print('Sample does not look Gaussian (reject H0)')
```

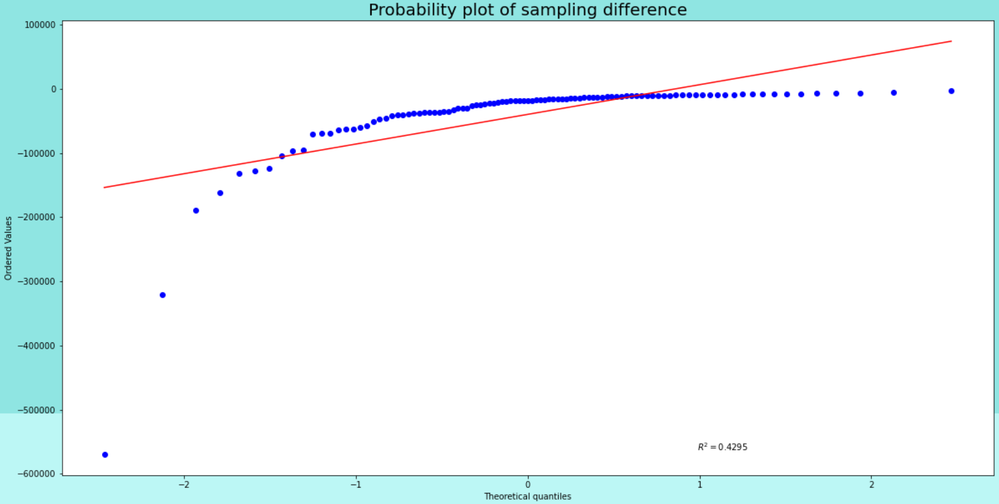
```
Statistic=0.4463, p=0.0000
Sample does not look Gaussian (reject H0)
```

RESULTS

```
fig = plt.figure(figsize= (20, 10))
ax = fig.add_subplot(111)

normality_plot, stat = stats.probplot(sampling_difference, plot= plt, rvalue= True)
ax.set_title("Probability plot of sampling difference", fontsize= 20)
ax.set

plt.show()
```



RESULTS

```python
import pandas as pd
!pip install researchpy
import researchpy as rp
import scipy.stats as stats
import matplotlib.pyplot as plt
```

female	male
831432	1401142
626632	946881
434496	596885
367299	556992
304383	436546
291563	396078
[illegible]	[illegible]

```python
file = 'dataPYTHON_Mann.xlsx'
df = pd.read_excel(file)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 101 entries, 0 to 100
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   female  101 non-null    int64
 1   male    101 non-null    int64
dtypes: int64(2)
memory usage: 1.7 KB
```

RESULTS

```
# code for Mann-Whitney U test
from scipy.stats import mannwhitneyu

# perform mann whitney test
stat, p_value = mannwhitneyu(df['female'], df['male'])
print('Statistics=%.2f, p=%.8f' % (stat, p_value))
# Level of significance
alpha = 0.05
# conclusion
if p_value < alpha:
    print('Reject Null Hypothesis (Significant difference between two samples)')
else:
    print('Do not Reject Null Hypothesis (No significant difference between two samples)')
```

```
Statistics=3496.00, p=0.00011282
Reject Null Hypothesis (Significant difference between two samples)
```

CONCLUSIONS

According to the analysis, there is a significant difference in the number of firms in metropolitan areas in terms of the gender of the firm owners.

THANKS!

Do you have any questions?

turgayturker55@gmail.com

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik** and illustrations by **Stories**

Please keep this slide for attribution

REFERENCES

[1] R. V Mahto, W. C. McDowell, J. Kudlats, and T. C. Dunne, “Learning Orientation and Performance Satisfaction as Predictors of Small Firm Innovation: The Moderating Role of Gender,” *Gr. Decis. Negot.*, vol. 27, no. 3, pp. 375–391, Jun. 2018, doi: 10.1007/s10726-018-9566-x.

[2] S. Coleman, “Gender, Entrepreneurship, and Firm Performance: Recent Research and Considerations of Context,” in Handbook on Well-Being of Working Women, M. L. Connerley and J. Wu, Eds. Dordrecht: Springer Netherlands, 2016, pp. 375–391.

[3] E. H. Becker-Medina, “Women-Owned Businesses on the Rise.” https://www.census.gov/newsroom/blogs/random-samplings/2016/03/women-owned-businesses-on-the-rise.html.

[4] A. W. Hait, “Number of Women-Owned Employer Firms Increased 0.6% From 2017 to 2018.” https://www.census.gov/library/stories/2021/03/women-business-ownership-in-america-on-rise.html.

[5] “Firms by Metropolitan Statistical Area (MSA) and Owner Sex for MSAs with more than 50,000 Firms.” https://data.sba.gov/en/dataset/total-number-of-firms-employer-nonemployer-by-owner-sex-2017.