

## Пример сравнения центров распределения. Использовалась лабораторная работа Андреевой

Версия 2 от 05.04.08

### Описание задачи

Ежегодно журнал "Fortune" публикует список миллиардеров. Перечень за 1992-й год включает 233 индивида или семьи и содержит данные об их состоянии, возрасте и месте жительства.

### Описание переменных

wealth – богатство семьи или индивида в миллиардах долларов,  
age – возраст в годах (для семей – максимальный возраст среди членов семьи),  
region – регион мира (А – Азия, Е – Европа, М – Ближний Восток, U – США и О – другие).

Количество наблюдений: 233

Данные находятся в файле *Billionaires\_1992\_data*.

### Задача

Необходимо сравнить богатство миллиардеров в США и остальном мире.

### Решение

Прежде чем непосредственно приступить к решению задачи, приведем данные файла в удобный для расчетов вид.

Создадим новую переменную *region\_new*: она равна 1, когда *region* = U (т.е. для США), и равна 0, когда *region* = О (т.е. для всех остальных регионов).

Теперь можно приступить к первому этапу решения.

### Проверка нормальности распределения

Поскольку данные о богатстве в США и других регионах содержатся в одном столбце *wealth*, разделим данные на 2 части с помощью команды Split File.

У нас имеется 233 наблюдения, поэтому нельзя однозначно сказать, какой из критериев для определения нормальности распределения богатства в США и других регионах нам использовать (критерий Колмогорова-Смирнова или Шапиро-Уилка). Оба этих критерия проверяют гипотезу о нормальности распределения данных:

**H:** ряд имеет нормальное распределение

**K:** ряд не нормально распределен

Посмотрим на результаты проверки для регионов, отличных от США:

#### Tests of Normality(b)

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
wealth	.312	169	.000	.385	169	.000

*region\_new* = 0

Видим, что оба критерия показывают, что распределение богатства для этой подгруппы регионов не является нормальным:

p-значение = Sig. =  $0 < \alpha = 0,05 \Leftrightarrow$  гипотеза нормальности распределения отвергается.

Теперь посмотрим на результаты проверки для США:

#### Tests of Normality(b)

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
wealth	.293	64	.000	.516	64	.000

region\_new = 1

Видим аналогичный результат:

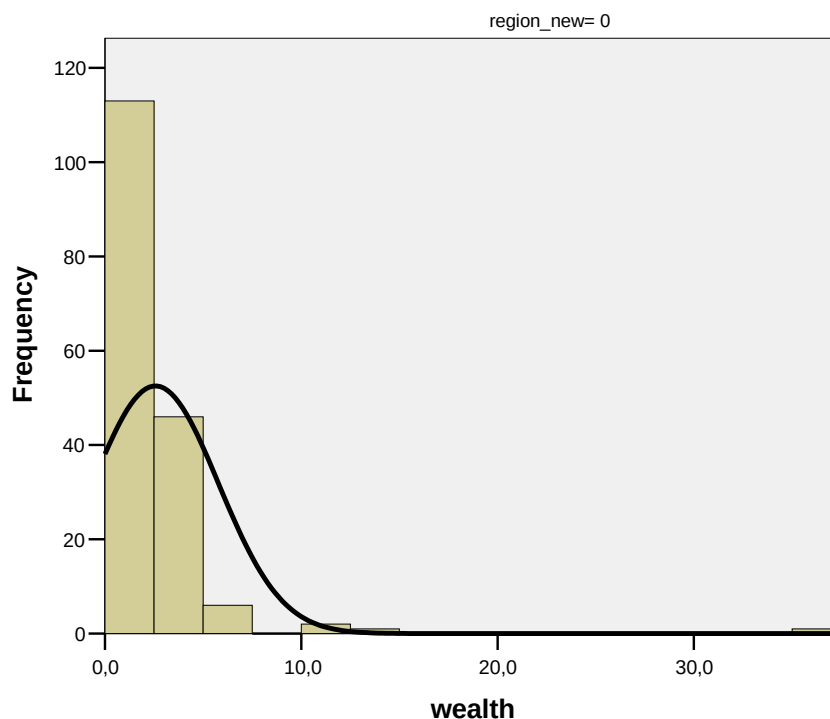
$p_{USA} = \text{Sig.} = 0 < \alpha = 0,05 \Leftrightarrow$  гипотеза нормальности отвергается.

Итак, распределение богатства для обеих подгрупп регионов не является нормальным. Однако, возможно, что распределение несущественно отличается от нормального.

Продолжим анализ: рассмотрим гистограммы и выясним, существенны ли отличия распределений сравниваемых совокупностей от нормального.

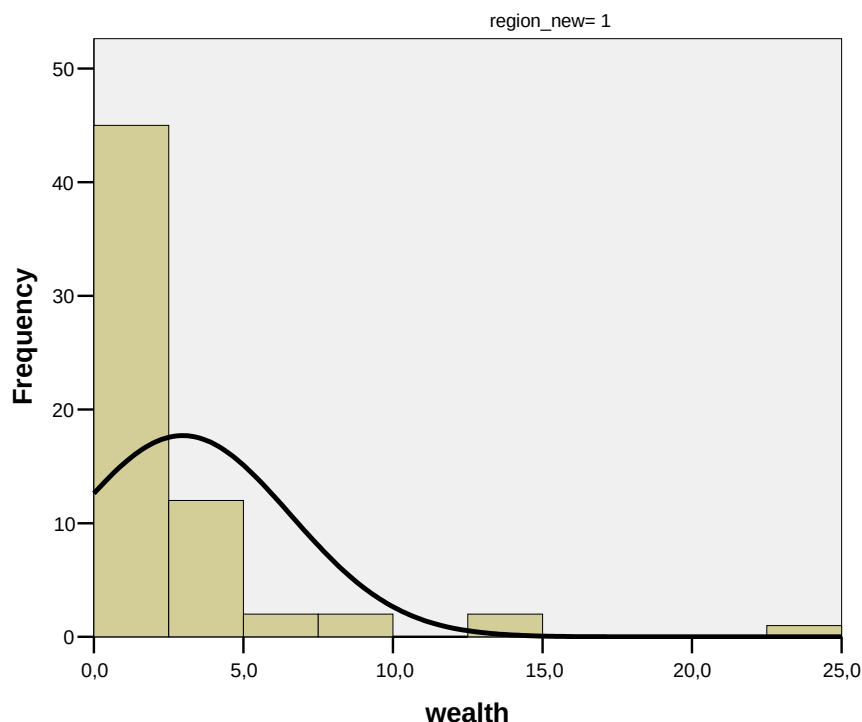
Построим гистограммы.

Для регионов, отличных от США, гистограмма выглядит так:



На представленной гистограмме имеются выбросы (отдельно стоящие наблюдения), ярко выраженная асимметрия. Все это позволяет сделать вывод о том, что распределение богатства в регионах, отличных от США, существенно отличается от нормального. Распределение в одной группе, существенно отличается от нормального, можно сделать вывод, что нельзя сравнивать средние (нельзя применять критерий Стьюдента), а надо сравнивать медианы, применяя критерий Манна-Уитни.

Тем не менее, скорее из любопытства, проведем аналогичную проверку для США. Гистограмма имеет следующий вид:



Для этой подгруппы данных видим аналогичный результат: имеются выбросы, явная асимметричность, следовательно, распределение богатства в регионе США не является нормальным.

Итак, распределение богатства для обеих подгрупп регионов даже является существенно отличным от нормального, следовательно, для сравнения богатства семей и индивидов в этих регионах мы не можем пользоваться сравнением средних этих двух совокупностей.

Итак, для сравнения совокупностей будем использовать сравнение их медиан.

### Проверка равенства медиан

Для проверки гипотезы о равенстве медиан используем критерий Манна-Уитни для независимых выборок. Он проверяет следующую гипотезу:

**H:** медианы равны

**K:** медианы не равны

Для вычисления Критерия Манна-Уитни используем точный способ вычисления p-значений (опция Exact). Получаем следующие результаты:

#### Test Statistics(a)

	wealth
Mann-Whitney U	5217.500
Wilcoxon W	19582.500
Z	-.415
Asymp. Sig. (2-tailed)	.678
Exact Sig. (2-tailed)	.679

Exact Sig. (1-tailed)	.339
Point Probability	.000

Exact Sig. (2-tailed) = 0,679 >  $\alpha = 0,05 \Leftrightarrow$  гипотеза о равенстве медиан не отвергнута.

Интерпретируем результат: богатство миллиардеров в США и богатство миллиардеров из других регионов мира можно считать примерно равным.

Для более наглядного представления результатов анализа приведем ящиковую диаграмму, построенную по сравниваемым выборкам:

