В городе Н. правительство решило начать борьбу с превышениями скорости автомобилей. Для выбора стратегии борьбы оно сначала решило провести исследования касательно того, влияет ли используемый водителем автомобиль на среднюю скорость передвижения.

Для этого было сформировано 3 выборки по 20 человек, в каждой из которой людям выдали одинаковые автомобили марок Mitsubishi, Audi и BMW, соответственно. В течение месяца замерялась средняя скорость каждого из автомобилей (см. файл).

Каждая из пар групп была проверена двувыборочным критерием на равенство распределений, также была проведена поправка на множественность гипотез.

Требуется:

- Описать, в чём недостаток подхода правительства.
- Предложить метод для более корректного решения задачи.
- Записать формальное условие задачи.
- Решить задачу аналитически (все аналитические выкладки должны быть описаны)

In [1]:

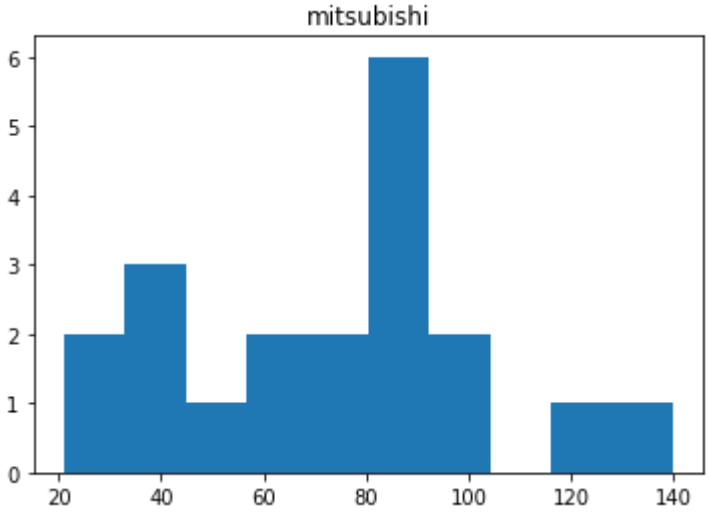
```
import pandas as pd

import matplotlib.pylab as plt

```

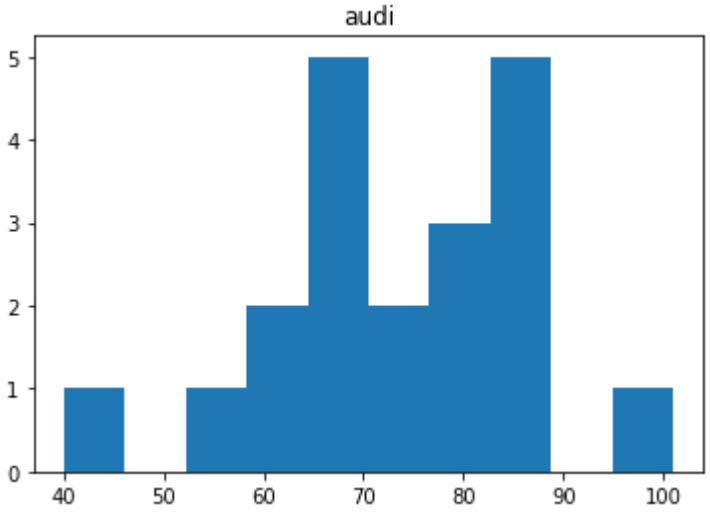
In [2]:

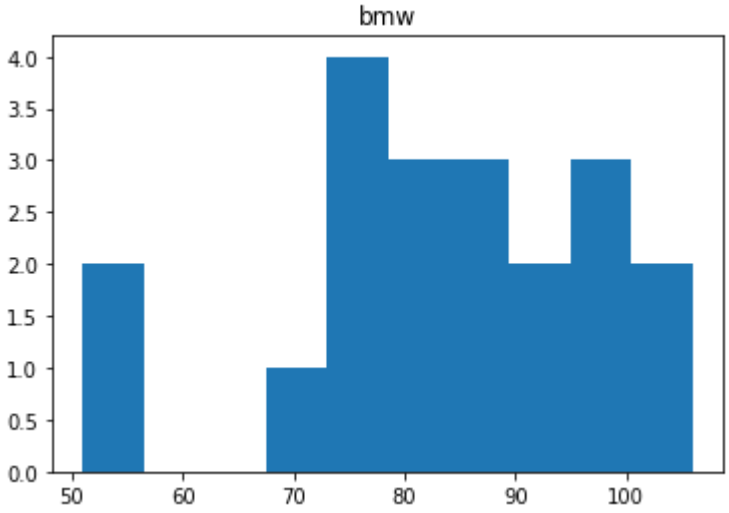
```
data = pd.read_csv('2.4.csv')
```

In [3]:

```
for v in ['mitsubishi', 'audi', 'bmw']:
    _ = plt.hist(data[v])
    plt.title(v)
    plt.show()
```



# Недостаток подхода

А какие выводы можно сделать из того, что распределения разные? Например водители фур наверняка и превышают скорость, но приэтом средняя скорость зачастую мала. Чтото подобное можно увидеть в распределении владельцев `Mitsubishi` .

# Корректный подход

А вообще мы же хотим оценить интервенцию марки авто на среднюю скорость. В таком случае предлагаю для начала оценить `Conditional effect` .

$ACE = P(y| \ do(Viehicle)) - P(...)$

# Формальное условие задачи

Отдельно хотелось бы обсудить, что же писать в скобочках. Вообще, государство по идее хочет ответить на вопрос, запрет езды на `BMW` сократит число штрафов. Тоесть они предполагают что

$avg \ velocity \rightarrow \ fine.$

И хотят проверить первую стрелку в этой причинность

$brand \rightarrow avg \ velocity \rightarrow \ fine.$

Посему предлагаю оценивать те интервалы, которые близки к штрафам. Напимер рассматривать авто со средней скоростью выше 90. Я считаю, что если за городом можно ездить 90, и водитель почти всегда ездит 90, то очень вероятно что он нарушает.

$ACE = P(y > 90| \ do(Viehicle = BMW)) - P(y > 90| \ do(Viehicle \neq BMW))$

Подсчитаем, сколько же таких нарушителей в кажой категории авто.

In [4]:

```
th = 90
for v in ['mitsubishi', 'audi', 'bmw']:
    print(f'for {v} >= {th} \n\t{data[data[v] >= th].shape[0]}/ {data.shape[0]}
```

```
for mitsubishi >= 90
        5/ 20
for audi >= 90
        1/ 20
for bmw >= 90
        7/ 20
```

# Решение задачи аналитически

## (описываем выкладки)

"Ручками" подсчитаем вероятность $P(y > 100| \ do(Viehicle = BMW))$ и $P(y > 100| \ do(Viehicle \neq BMW))$

$P(=)$

$$P(y > 100|\ do(Viehicle = BMW)) = \frac{NumBMW\ drivers{>}90}{NumBMW\ drivers} = \frac{NumBMW\ drivers{>}90}{20}$$

## $P(\neq)$

$$P(y > 100|\ do(Viehicle \neq BMW)) = \frac{(not\,NumBMW\ drivers){>}90}{not\,NumBMW\ drivers} = \frac{alldrivers{>}90 - BMW\ drivers{>}90}{40}$$

In [5]:

```
th = 90
g = {}
s = 0
for v in ['mitsubishi', 'audi', 'bmw']:
    g[v] = data[data[v] >= th].shape[0]
    s += g[v]
eq_p = {}
neq_p = {}
eff = {}
for v in ['mitsubishi', 'audi', 'bmw']:
    print(v)
    eq_p[v] = g[v] / 20
    neq_p[v] = (s - g[v]) / 40
    print(f'\t\tP(y > 100|\  do(Viehicle = BMW)) = {eq_p[v]: .2e}')
    print(f'\t\tP(y > 100|\  do(Viehicle !=BMW)) = {neq_p[v]: .2e}')
    eff[v] = eq_p[v] - neq_p[v]
    print(f'\tEffect = {eff[v]: .2f}')

```

```
mitsubishi
                P(y > 100|\  do(Viehicle = BMW)) =  2.50e-01
                P(y > 100|\  do(Viehicle !=BMW)) =  2.00e-01
        Effect =  0.05
audi
                P(y > 100|\  do(Viehicle = BMW)) =  5.00e-02
                P(y > 100|\  do(Viehicle !=BMW)) =  3.00e-01
        Effect = -0.25
bmw
                P(y > 100|\  do(Viehicle = BMW)) =  3.50e-01
                P(y > 100|\  do(Viehicle !=BMW)) =  1.50e-01
        Effect =  0.20
```

Как и ожидалось, если есть BMW , есть деньги и на штрафы. Водители ауди видимо неплохо зарабатывающие законопослушные граждане