In [1]:
```
#!/usr/bin/python
# -*- coding: utf-8 -*-

%matplotlib inline

import os, sys, math
import matplotlib
import numpy as np
import matplotlib.pyplot as plt
from collections import Counter
from IPython.display import display, HTML
```

Задача 1

Выборка данных

Даны результаты измерений максимальной скорости испытаний спортивного самолета, м/с:

In [2]:
```
# Датасет
a = [431, 398, 423, 401, 423, 404, 389, 428, 402, 404,
     427, 398, 422, 409, 420, 422, 397, 458, 403, 411,
     398, 408, 438, 414, 413, 404, 426, 434, 430, 397,
     383, 415, 418, 438, 394, 417, 412, 404, 389, 398,
     431, 423, 401, 423, 435, 427, 428, 405, 414, 415,
     439, 409, 391, 416, 419, 401, 372, 395, 418, 413,
     407, 445, 428, 420, 429, 395, 433, 406, 402, 398,
     399, 432, 405, 412, 425, 417, 424, 416, 396, 403,
     432, 402, 431, 419, 423, 441, 424, 410, 424, 413,
     393, 412, 302, 408, 437, 416, 436, 415, 421, 407,
     404, 404, 403, 434, 412, 419, 405, 402, 394, 423,
     398, 415, 401, 398, 428, 416, 453, 371, 424, 417]

a.sort()

# Считаем количество уникальных
stat_a = Counter(a)
```

Количество интервалов

По формуле Стержеса:

$k = 1 + 3,32 * \lg n$

Где k - количество интервалов, n - размер выборки

In [3]:
```
icount = int(round(1 + 3.32 * math.log10(len(a))))
```

Размер интервала:

$$\Delta a = \frac{a_{max} - a_{min}}{k}$$

Где k - количество интервалов, a_{max}, a_{min} - максимальный и минимальный элементы выборки

In [4]:
```
da = (max(a) - min(a)) / icount;

print "Количество интервалов:\t", icount
print "Размер интервала:\t", da
print
print "min \t=", min(a), "\nmax \t=", max(a)
print "Медиана =", a[len(a)/2]
```

```
Количество интервалов:  8
Размер интервала:       19

min     = 302
max     = 458
Медиана = 414
```

Статический ряд

In [5]:
```
s = "<style></style><table><tr><th>Значение</th><th>Частота</th></tr>"
for key, val in stat_a.items():
    s += '<tr><td>{}</td><td>{}</td></tr>'.format(key, val)
s += "</table>"
display(HTML(s))
```

Значение	Частота
389	2
391	1
393	1
394	2
395	2
396	1
397	2
398	7
399	1
401	4
402	4
403	3
404	6
405	3

406	1
407	2
408	2
409	2
410	1
411	1
412	4
413	3
414	2
415	4
416	4
417	3
418	2
419	3
420	2
421	1
422	2
423	6
424	4
425	1
426	1
427	2
428	4
429	1
302	1
431	3
432	2
433	1
434	2
435	1
436	1
437	1
438	2
439	1
441	1
445	1
453	1
458	1
371	1
372	1
430	1
383	1

Находим интервалы

In [6]:

```
bins = []

for i in range(icount):
    bins.append(min(a) + i*da)
    print min(a) + i*da, "-", min(a) + (i+1)*da

bins.append(max(a))
```

```
302 - 321
321 - 340
340 - 359
359 - 378
378 - 397
397 - 416
416 - 435
435 - 454
```

Интервальный статический ряд

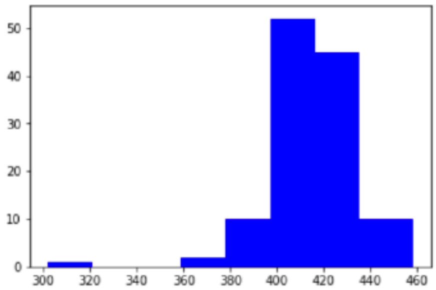
In [7]:
```
hist, bins = np.histogram(a, bins=bins)
bin_counts = zip(bins, bins[1:], hist)

s = "<style></style><table><tr><th>Интервал</th><th>Частота</th></tr>"
for bin_start, bin_end, count in bin_counts:
    s += '<tr><td>{} - {}</td><td>{}</td></tr>'.format(bin_start, bin_end, count)
s += "</table>"
display(HTML(s))
```

Интервал	Частота
302 - 321	1
321 - 340	0
340 - 359	0
359 - 378	2
378 - 397	10
397 - 416	52
416 - 435	45
435 - 458	10

Гистограмма по интервальным значениям

In [8]:
```
plt.hist(a, bins, color="blue")
plt.show()
```



Точечные оценки выборки

Среднее выборочное (оценка мат. ожидания):

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Выборочная дисперсия:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

Где X - выборка, X_i - i-й элемент выборки, n - размер выборки

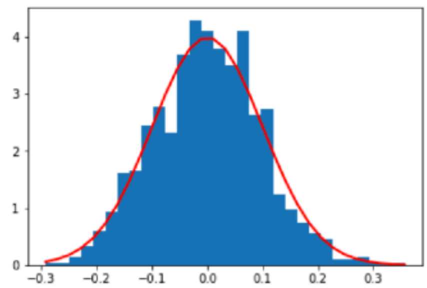
In [9]:
```
print "Среднее выборочное:\t", (0.0+np.sum(a))/len(a)
print "Исправленная дисперсия:\t", np.var(a)
print "Ср. кв. отклонение:\t", np.std(a)
```

```
Среднее выборочное:     412.95
Исправленная дисперсия: 338.6975
Ср. кв. отклонение:     18.4037360338
```

Из метрик, приведенных выше можно сделать вывод:
Данная случайная величина имеет нормальное распределение

Пример нормального распределения

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



Задача 2

В результате 16 испытаний инерционного звена определены значения статистических характеристик случайной величины t:

$$\bar{x} = \widehat{\mu_t} = 120.1\text{c}$$
$$S^2 = \widehat{\sigma_t}^2 = 9.64\text{c}^2$$

Считая закон распределения случайной величины нормальным, построить для параметров $\widehat{\mu_t}$ и $\widehat{\sigma_t}^2$ доверительные интервалы, отвечающие доверительным вероятностям: $\gamma_1 = 0.95$ и $\gamma_2 = 0.90$

1. Найдем доверительную оценку математического ожидания

$$Y = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}}$$

Y соответсвует распределению стьюдента с $(k = n - 1)$ степенями свободы.
По заданным вероятностям и размеру выборки n найдем квантили $t = t(\gamma; n - 1)$

$t(\gamma_1, 16 - 1) = 2.131$
$t(\gamma_2, 16 - 1) = 1.753$

С помощью центральной предельной теоремы получаем оценку величины Y:

$$|Y| < t(\gamma; n - 1)$$

$$\left| \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}} \right| < t(\gamma; n - 1)$$

$$\bar{x} - t(\gamma; n - 1) * \frac{s}{\sqrt{n}} < m < \bar{x} + t(\gamma; n - 1) * \frac{s}{\sqrt{n}}$$

$S = \sqrt{S^2} = \sqrt{9.64} \approx 3.105$

Подставим значения для γ_1:

$$120.1 - 2.131 * \frac{3.105}{\sqrt{16}} < m < 120.1 + 2.131 * \frac{3.105}{\sqrt{16}}$$

$$118.4 < m < 121.8$$

Подставим значения для γ_2:

$$120.1 - 1.753 * \frac{3.105}{\sqrt{16}} < m < 120.1 + 1.753 * \frac{3.105}{\sqrt{16}}$$

$$118.7 < m < 121.5$$

2. Найдем доверительную оценку дисперсии

$$Z = \frac{S^2 * (n - 1)}{\sigma^2}$$

Z соответсвует распределению χ^2 с $(k = n - 1)$ степенями свободы. По таблицам распределения χ^2 найдем u_1 и u_2.

Получаем оценку величины Z:

$$u_1 < Z < u_2$$

$$u_1 < \frac{S^2 * (n - 1)}{\sigma^2} < u_2$$

$$\frac{u_1}{S^2 * (n - 1)} < \frac{1}{\sigma^2} < \frac{u_2}{S^2 * (n - 1)}$$

$$\frac{S^2 * (n - 1)}{u_2} < \sigma^2 < \frac{S^2 * (n - 1)}{u_1}$$

Подставим для γ_1:

$u_1 = 6.24$
$u_2 = 27.47$

$$\frac{9.64 * (16 - 1)}{27.47} < \sigma^2 < \frac{9.64 * (16 - 1)}{6.24}$$
$$5.264 < \sigma^2 < 23.173$$

Подставим для γ_2:

$u_1 = 7.24$
$u_2 = 24.97$

$$\frac{9.64 * (16-1)}{24.97} < \sigma^2 < \frac{9.64 * (16-1)}{7.24}$$
$$5.791 < \sigma^2 < 19.972$$

Ответ

Для $P = \gamma_1 = 0.95$:

$118.4 < m < 121.8$
$5.264 < \sigma^2 < 23.173$

Для $P = \gamma_2 = 0.90$:

$118.7 < m < 121.5$
$5.791 < \sigma^2 < 19.972$

Задача 3

На двух токарных станках изготавливают детали по одному чертежу. Из продукции первого станка было отобрано $n_1 = 9$ деталей, а из продукции второго — $n_2 = 11$ деталей. Выборочные дисперсии контрольного размера, определенные по этим выборкам, равны:
${S_1}^2 = 5.9\text{мкм}^2$ и
${S_2}^2 = 23.3\text{мкм}^2$ соответственно.
Проверить гипотезу о равенстве дисперсий при уровне значимости $\alpha = 0.05$ при конкурирующей гипотезе, утверждающей, что дисперсии не равны.

Решение

Проверяемая гипотеза - $H_0 : {\sigma_1}^2 = {\sigma_2}^2$
Конкурирующая гипотеза - $H_1 : {\sigma_1}^2 \neq {\sigma_2}^2$

Рассмотрим величину (в качестве статистики), соответствующую распределению Фишера с $(n_1 - 1; n_2 - 1)$ степенями свободы:

$$R(\vec{x}, \vec{y}) = \frac{{S_1}^2 * n_2}{{S_2}^2 * n_1}$$

На основании основной и конкурирующей гипотез, критическая область будет задана двусторонним интервалом:
$$[0; c_1] \cup [c_2; \infty)$$

Причем, $P(R(\vec{x}, \vec{y}) \in [0; c_1]) = P(R(\vec{x}, \vec{y}) \in [c_2; \infty)) = \frac{\alpha}{2}$

Таким образом, найдем квантили c_1 и c_2 распределения Фишера с $(n_1 - 1; n_2 - 1)$ степенями свободы по таблицам:

$$c_1 = C_{\frac{\alpha}{2}} = 0.23$$
$$c_2 = C_{1-\frac{\alpha}{2}} = 3.85$$

Вычислим значение величины $R(\vec{x}, \vec{y})$ (статистики):

$$R(\vec{x}, \vec{y}) = \frac{{S_1}^2 * n_2}{{S_2}^2 * n_1} = \frac{5.9 * 11}{23.3 * 9} = 0.31$$

$$R(\vec{x}, \vec{y}) \notin [0; c_1] \cup [c_2; \infty)$$

Таким образом, нет оснований отвергать основную гипотезу H_0.