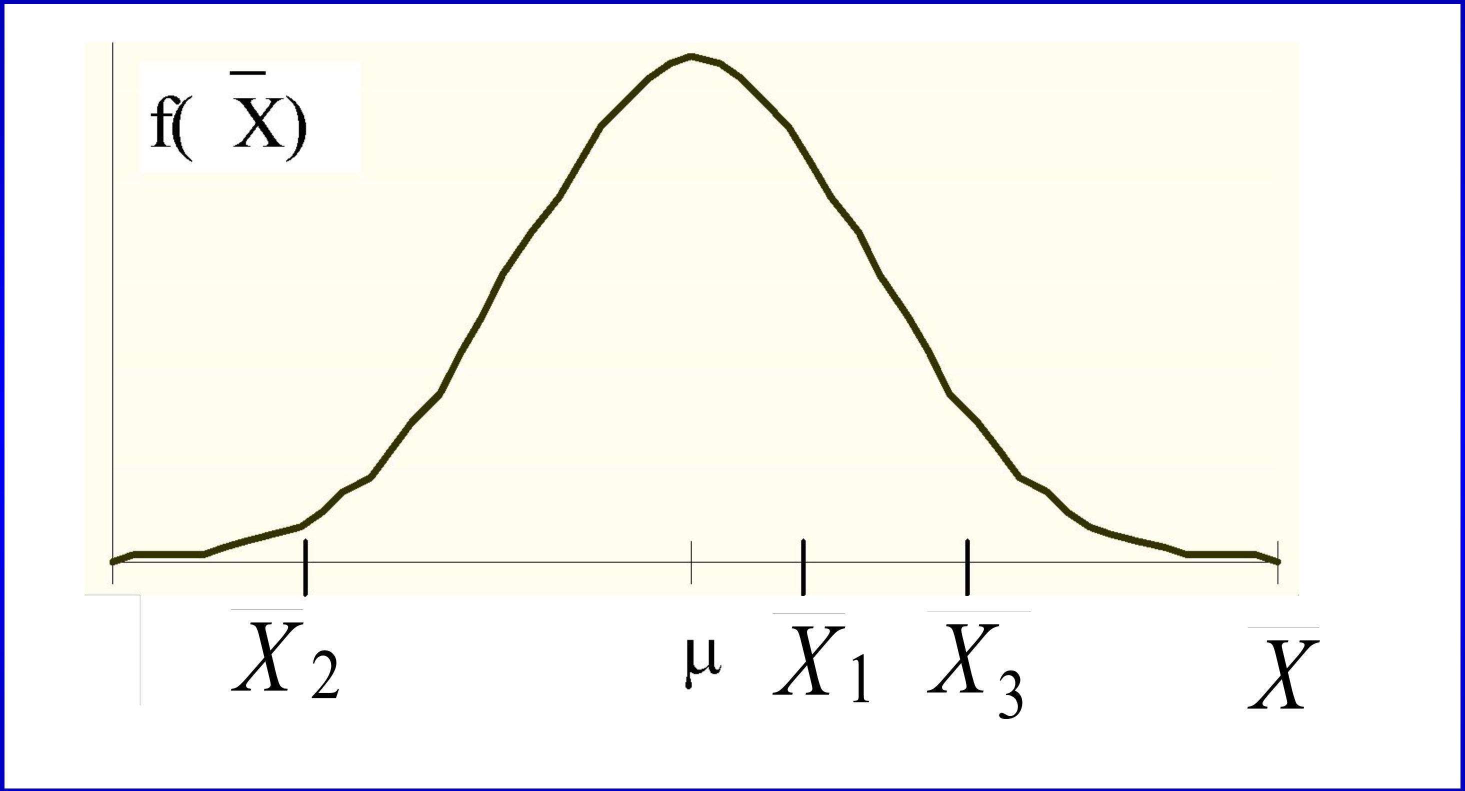
9. Распределения статистик
(Выборочные распределения)

Параметры совокупности θ, например μ – детерминированные величины («так природа захотела»)

Определенные по выборке объема n оценки $\hat{\theta}$, например $\bar{X}$ – случайные, варьируют от выборки к выборке

булки



Пример →

Пример

Процент брака на 5 линиях по производству кирпича – совокупность значений X: 2, 4, 5, 8, 11; $\mu = 6$.

На 2-х линиях 1-го завода это 5 и 11,

$\bar{X}_1 = 8$, ошибка $\bar{X}_1 - \mu = 2$.

На 2-ом заводе 2 линии дают 2 и 8% брака,

$\bar{X}_2 = 5$, ошибка $\bar{X}_2 - \mu = -1$.

→ использование $\bar{X}$ в качестве оценок μ предусматривает наличие ошибки

При рассмотрении точности оценок используется понятие **распределения статистики** (например, $\bar{X}$)

N=5
n=2
$C_5^2=10$
выборок
→
10
значени
й
средних

Распределение статистики включает все возможные значения, которые она (например, среднее) может принять при данном объеме выборки.

ределения статистик называют *выборочными распределениями*

(они происходят из вариации выборок)

У них есть свои параметры μ и σ,

в частности у распределения $\square X$ это $\mu_{\square X}$ и $\sigma_{\square X}$

Стандартное отклонение распределения статистики называют

стандартной ошибкой статистики

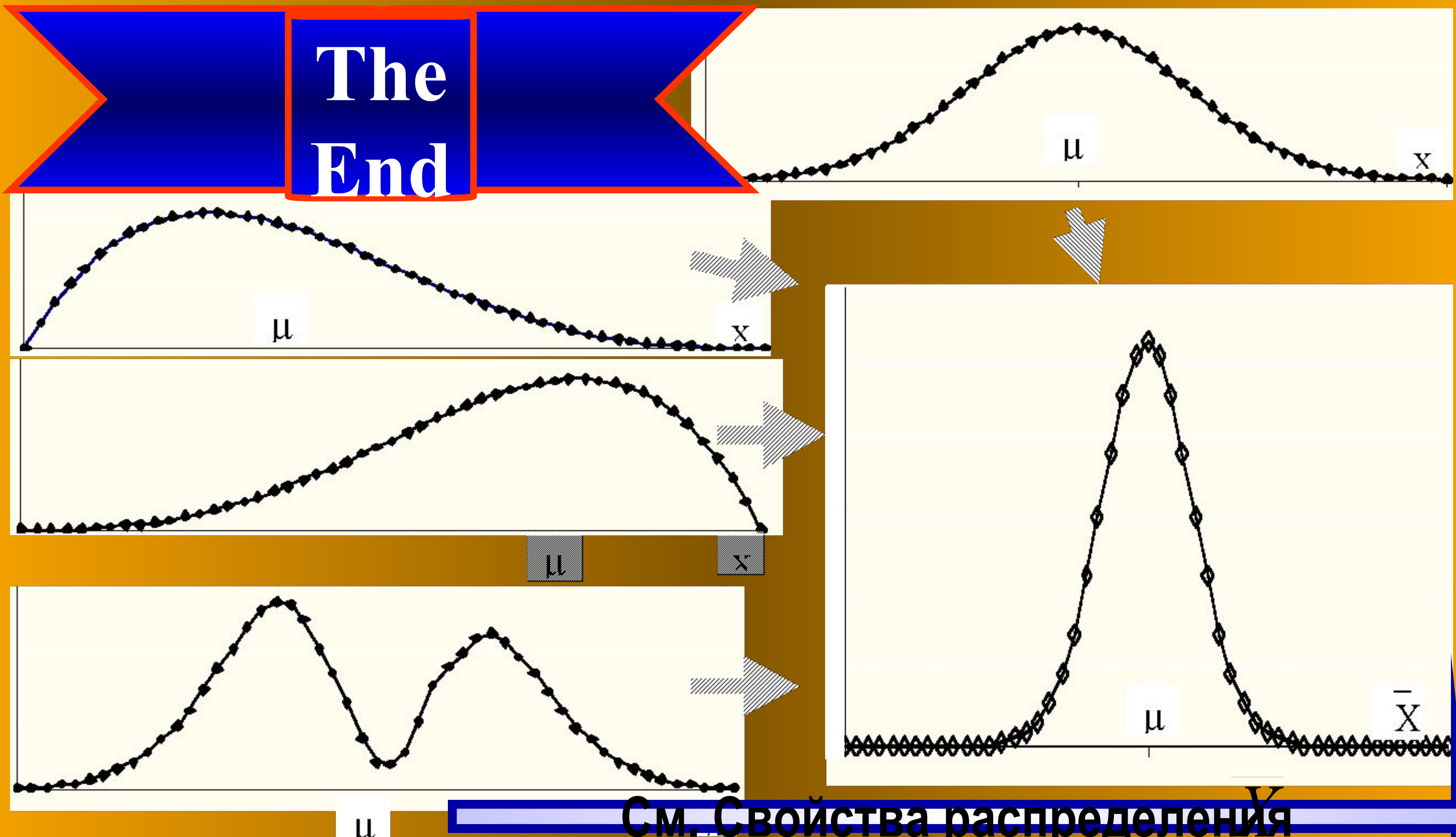
Основа статистических выводов :

На них процедуры оценивания

и процедуры проверки гипотез

Важно математическое описание распределений статистик

Пример с большим значением



См. Свойства распределения

10. Оценивание

Точечн ервальная оценка,

довер ал,

довер ость
(наде точность оценки

Содержат ошибку, неизвестно какую? Насколько можно верить?

То оценка **–** определяемое по выборке единственное значение, используемое как оценка характеристики совокупности

Примеры: $\bar{X}$, **s^2**, **s**, **$\hat{v}$** и др.

Зависит от объема выборки (n) и изменчивости совокупности (σ)

Пример про оценивание возраста

рительная вероятность

Интервал!

Чем длиннее, тем *вероятней* содержит истинное значение, «накрывает» его

Интервальная оценка указывает на точность и доверие к ней, устанавливая интервал, в котором, *вероятно*, лежит характеристика совокупности

Доверительные границы

Интервальная оценка **определяется двумя числами – *границами интервала*, который накрывает оцениваемый параметр**

Доверительный интервал **– который с *заданной вероятностью* накрывает оцениваемый параметр θ**

Записывается → $P(\theta_н < \theta < \theta_в) = \gamma\ (\%)$

где $\theta_н$, $\theta_в$ – **нижняя и верхняя границы интервала**

Означает: $\boxed{\text{с вероятностью (надежностью) } \gamma \text{ параметр } \theta \text{ имеет значение в интервале } (\theta_н, \theta_в)}$

Простейший, наиболее частый случай – **симметричный интервал:**

$$P(\hat{\theta} - \delta < \theta < \hat{\theta} + \delta) = \gamma = (1 - \alpha)$$

здесь $\hat{\theta}$ **– точечная оценка характеристики** θ

δ – **ошибка оценки**

γ – **надежность оценки**

α – **р и с к** (что это не так, уровень значимости)

Важный пример

Доверительный интервал для генерального среднего

Определяется среднее совокупности **μ** по точечной оценке $\overline{X}$

С вероятностью **γ** интервал ($\overline{X}$ **± δ**) будет содержать **μ**

? Используется распределение статистики $\overline{X}$,

при «большом» n **– нормальное** (см. свойства распр. среднего)

$$P(\overline{X} - \delta < \mu < \overline{X} + \delta) = P(\mu = \overline{X} \pm \delta) = \gamma = 1 - \alpha =$$

Определяет:

- ошибку δ при заданной надежности γ
- надежность γ при заданной точности δ

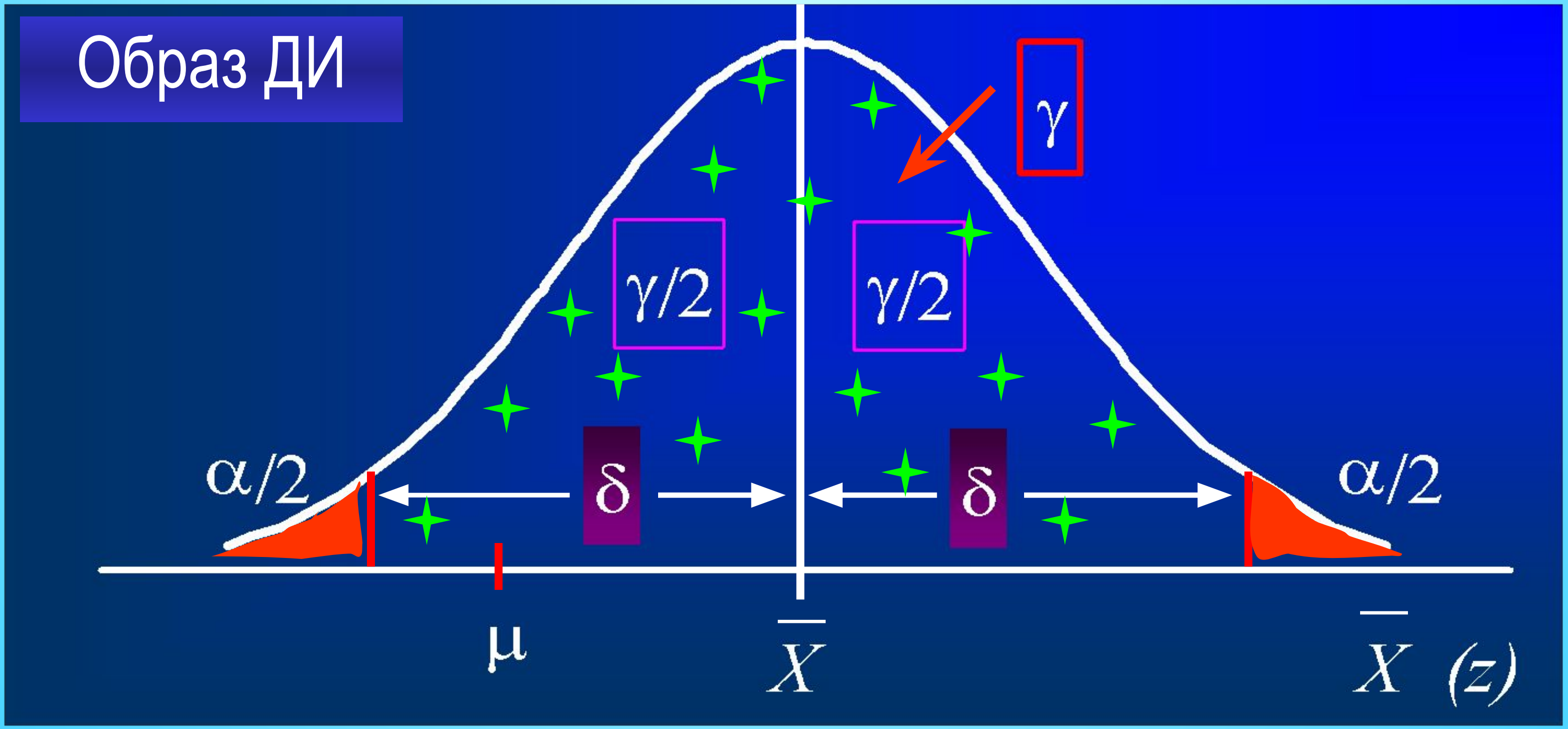
$$P(\mu = \overline{X} \pm \underbrace{z_{\gamma/2}\frac{\sigma}{\sqrt{n}}}_{\delta})$$

- требуемый объем выборки n, обеспечивающий заданную точность δ при заданной надежности γ

В формуле:

z − значение нормированной нормальной величины (аргумент функции Лапласа), **соответствующее** $\gamma/2$;
это число «сигм» в δ − *к а к и х* **«сигм» ? −**

стандартных отклонений выборочного среднего $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$



Как рассчитать :

δ , если задана γ ⇒ $\gamma/2 = \Phi\left(z_{\gamma/2} = \delta \cdot \sqrt{n}/\sigma\right)$

$$\downarrow$$

$$z_{\gamma/2}$$

$$\boxed{\delta = z_{\gamma/2}\frac{\sigma}{\sqrt{n}}}$$

γ , если задана δ ⇒ $z = \delta \cdot \sqrt{n}/\sigma \rightarrow \Phi(z) = \gamma/2 \rightarrow \gamma$

n , если задана γ и δ ⇒

$$\boxed{n = \frac{z_{\gamma/2}^2 \cdot \sigma^2}{\delta^2}}$$

Если σ неизвестна

→ Используется

→ σ²,

выборочное распределением

(Стьюдента - student), при f = n – 1

», ошибка из-за использования

ого распределения допустима

См. «Д... ситуаций при оценивании средних» на стр. 22 «Статистики»)

The End

Интервальная оценка среднего при «малом» n

$$P(\mu = \overline{X} \pm t_\gamma \frac{s}{\sqrt{n}}) = \gamma = 1 - \alpha$$

t – **квантиль** для заданной γ при соответствующем n или f