Пример 1. На токарном станке вытачивают цилиндры. Наладчик станка может добиться нулевой систематической ошибки, однако с.к.о. ошибки станка со временем увеличивается в результате износа и необходимо контролировать находится ли она в заданных границах. На настроенном станке при изготовлении 10 цилиндров диаметра 5 см получены цилиндры диаметров:

x_i	$5,2$	$5,3$	$5,0$	$4,9$	$4,8$	$5,1$	$5,1$	$5,0$	$5,1$	$5,1$

Считая, что ошибка подчиняется нормальному закону, найти доверительные границы для σ с уровнем доверия $\alpha = 0,9$.

Решение

Для $n = 10$ находим квантили $\chi^2(10)$:
$z_{(1-\alpha)/2} = z_{0,05} = 3,94$, $z_{(1+\alpha)/2} = z_{0,95} = 18,3$.

Тогда

$$0,9 = P\left\{\frac{nS_0^2}{z_{(1+\alpha)/2}} < \sigma^2 < \frac{nS_0^2}{z_{(1-\alpha)/2}}\right\}.$$

$$S_0^2(\vec{x}_n) = \frac{1}{n}\sum_{k=1}^{n}(x_k - a)^2 = 0,022.$$

$\sigma \in (0,11; 0,24)$. ◄

8.1. Лемма Фишера.

Распределение χ^2 используется для доверительной оценки параметра σ^2 при неизвестном математическом ожидании.

Это связано со следующим результатом Фишера.

Лемма 1. (Фишера). *Пусть дан вектор независимых случайных величин* $\vec{X}_n = (X_1, \ldots, X_n)$, $X_i \sim N(0,1)$.

C *- ортогональная матрица* ($C^{-1} = C^T$) *размерности* $n \times n$ *и* $\vec{Y} = C\vec{X}_n$.

Тогда для любого $m = 1, \ldots, n-1$ *статистика*

$$g(\vec{X}_n) = \sum_{k=1}^{n} X_k^2 - Y_1^2 - \ldots - Y_m^2 \text{ не зависит от } Y_1, \ldots, Y_m$$

и $g(\vec{X}_n) \sim \chi^2(n-m)$.

Доказательство. В силу ортогональности матрицы C нормы векторов $\vec{X}_n$, $\vec{Y}_n$ совпадают, т.е. $||\vec{X}||^2 = \sum_{k=1}^{n} X_k^2 = \sum_{j=1}^{n} Y_j^2 = ||\vec{Y}||^2$.

Поэтому

$$g(\vec{X}) = \sum_{k=1}^{n} X_k^2 - Y_1^2 - \ldots - Y_m^2 =$$
$$= \sum_{j=1}^{n} Y_j^2 - Y_1^2 - \ldots - Y_m^2 = Y_{m+1}^2 + \ldots + Y_n^2.$$

Ортогональность матрицы C означает, что

$$\sum_{k=1}^{n} c_{ik} c_{jk} = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$$

Поэтому набор случайных величин $Y_1, \ldots, Y_n$ имеет следующие свойства ($k = 1, \ldots, n$):

1. Случайная величина $Y_k = \sum_{s=1}^{n} c_{ks} X_s$ - нормально распределена.

2. $M(Y_k) = M\left(\sum_{j=1}^{n} c_{kj} X_j\right) = \sum_{j=1}^{n} c_{kj} M(X_j) = 0.$

3.
$$M(Y_j\, Y_k) = M\left(\sum_{m=1}^{n} c_{jm} X_m \sum_{s=1}^{n} c_{ks} X_s\right) = \sum_{m=1}^{n} \sum_{s=1}^{n} c_{jm} c_{ks} M(X_m\, X_s) =$$
$$= \sum_{m=1}^{n} c_{jm} c_{km} M(X_m\, X_m) = \sum_{m=1}^{n} c_{jm} c_{km};$$

$\Rightarrow$ $M(Y_j\, Y_k) = 0$ при $j \neq k$ и $M(Y_j^2) = D(Y_j) = 1$ при $j = k$;
$\Rightarrow$ набор попарно независим и $D(Y_j) = 1$.

Отсюда, в силу свойств набора $Y_1, \ldots, Y_n$, следует, что
$g(\vec{X}) = Y_{m+1}^2 + \ldots + Y_n^2 \sim \chi^2(n-m)$ и не зависит от величин $Y_1, \ldots, Y_m$.

□

Следствие 1 (леммы Фишера). *Для* $X_k \sim N(0,1)$, $k = 1, \ldots, n$*:*

1) Величина $\sum_{k=1}^{n} (X_k - \bar{X})^2 \sim \chi^2(n-1)$.

2) Случайные величины $\bar{X}$ *и* $\sum\limits_{k=1}^{n}(X_k-\bar{X})^2$ *независимы.*

Доказательство.

$$\sum_{k=1}^{n}(X_k-\bar{X})^2=\sum_{k=1}^{n}X_k^2-2\bar{X}\sum_{k=1}^{n}X_k+n\bar{X}^2=\sum_{k=1}^{n}X_k^2-n\bar{X}^2=\sum_{k=1}^{n}X_k^2-Y_1^2,$$

где Y_1 определяется выражением $Y_1=\sqrt{n}\bar{X}=\dfrac{X_1}{\sqrt{n}}+\ldots+\dfrac{X_n}{\sqrt{n}}$.

Чтобы применить лемму Фишера к наборам $\vec{X}_n$ и Y_1, нужно найти ортогональную матрицу C, такую, что Y_1 - первая координата вектора $\vec{Y}_n=C\vec{X}_n$.

Возьмем матрицу C с первой строкой $(\frac{1}{\sqrt{n}},\ldots,\frac{1}{\sqrt{n}})$. Норма этого вектора равна 1 и можно дополнить базис, включающий этот вектор, другими векторами так, чтобы в целом матрица C была ортогональной.

Применим лемму Фишера для $m=1$. Получаем по доказанной лемме, что $\sum\limits_{k=1}^{n}X_k^2-Y_1^2\sim\chi^2(n-1)$
и не зависит от Y_1, то есть от $\bar{X}$, что и требовалось доказать. $\square$

8.2. Доверительный интервал для σ^2 при неизвестной a

Пусть $X_k\sim N(a,\sigma^2)$ для $k=1,\ldots,n$. Построим центральную статистику для σ^2 при неизвестном a.

Статистика S^2 (ее смещенный аналог $\hat{\sigma}^2$) являются наилучшей оценкой для σ^2 при неизвестной величине a.

8.3. Центральная статистика для σ^2

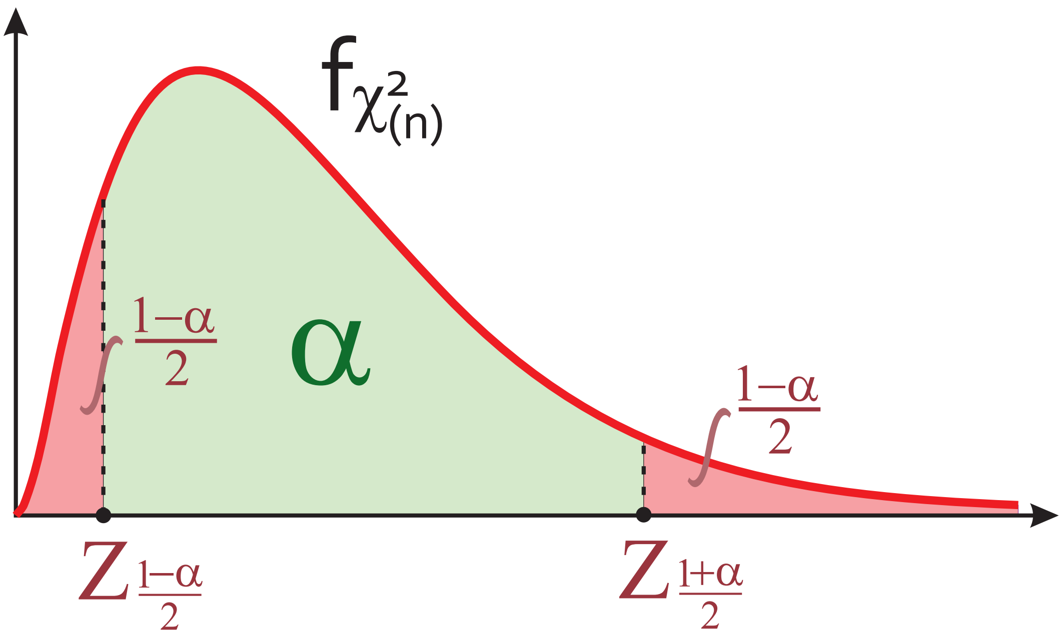
$$\frac{n\,\hat{\sigma}^2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{k=1}^{n}\left(X_k-\frac{1}{n}\sum_{j=1}^{n}X_j\right)^2=$$

$$=\frac{1}{\sigma^2}\sum_{k=1}^{n}\left(X_k-a-\frac{1}{n}\sum_{j=1}^{n}(X_j-a)\right)^2=$$

$$=\sum_{k=1}^{n}\left(\frac{X_k-a}{\sigma}-\frac{1}{n}\sum_{j=1}^{n}\frac{X_j-a}{\sigma}\right)^2\sim\chi^2(n-1).$$

(Использовано следствие леммы Фишера для сл. вел. $N(0,1)$.)

8.4. Доверительный интервал дял σ^2



(График для $\chi^2(n-1)$ аналогичен.)

$Z(\vec{X}_n) = \dfrac{n\hat{\sigma}^2(\vec{X}_n)}{\sigma^2} \sim \chi^2(n-1).$ - центральная статистика (для σ^2).

$$P\{z_{(1-\alpha)/2} < Z(\vec{X}_n) < z_{(1+\alpha)/2}\} = \alpha.$$

$$P\{z_{(1-\alpha)/2} < \frac{n\hat{\sigma}^2}{\sigma^2} < z_{(1+\alpha)/2}\} = \alpha.$$

$$P\left\{\frac{n\hat{\sigma}^2}{z_{(1+\alpha)/2}} < \sigma^2 < \frac{n\hat{\sigma}^2}{z_{(1-\alpha)/2}}\right\} = \alpha.$$

Доверительный интервал $\sigma^2 \in \left(\dfrac{n\hat{\sigma}^2(\vec{x}_n)}{z_{(1+\alpha)/2}}; \dfrac{n\hat{\sigma}^2(\vec{x}_n)}{z_{(1-\alpha)/2}}\right)$.

Пример 2. (Аналог предыдущей задачи с неизвестной a.) На токарном станке вытачивают цилиндры. С.к.о. ошибки станка со временем увеличивается в результате износа и необходимо контролировать находится ли она в заданных границах. На настроенном станке при изготовлении 10 цилиндров (одного диаметра) получены цилиндры диаметров:

x_i	$5,2$	$5,3$	$5,0$	$4,9$	$4,8$	$5,1$	$5,1$	$5,0$	$5,1$	$5,1$

Считая, что ошибка подчиняется нормальному закону, найти доверительные границы для σ с уровнем доверия $\alpha = 0,9$.

◀

8.5. Распределение $t(n)$ (Стьюдента).

Определение 8.1. *Пусть* $X_k \sim N(0,1)$, *для* $k = 0,\dots,n$ *независимые случайные величины (выборка).*

Тогда случайная величина (или по другому статистика)

$$\frac{X_0}{\sqrt{\frac{1}{n}\sum\limits_{j=1}^{n} X_j^2}} \sim t(n)$$

(имеет распределение по закону Стьюдента) n *- параметр, называемый числом степеней свободы.*

Утверждение 1. *Плотность распределения*

$$f_{t(n)}(x) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \; x \in \mathbb{R}.$$

Доказательство. Пусть $x > 0$.

$$
\begin{aligned}
F_{t(n)}(x) &= P\left\{\frac{X_0}{\sqrt{\frac{1}{n}(X_1^2+\ldots+X_n^2)}} < x\right\} = \\
&= 1 - P\left\{\frac{X_0}{\sqrt{\frac{1}{n}(X_1^2+\ldots+X_n^2)}} > x\right\} = \\
&= 1 - P\left\{X_0 > x\sqrt{\frac{1}{n}(X_1^2+\ldots+X_n^2)}\right\} = \\
&- 1 - \frac{1}{2}P\left\{X_0^2 > \frac{x^2}{n}(X_1^2+\ldots+X_n^2)\right\} = \\
&= 1 - \frac{1}{2}\iint\limits_{t > sx^2/n} f_{X_0^2}(t) f_{\chi^2(n)}(s)\,dt\,ds = \\
&= 1 - \frac{1}{2}\int_0^\infty f_{\chi^2(n)}(s)ds \int_{sx^2/n}^\infty f_{X_0^2}(t)dt.
\end{aligned}
$$

$$
\begin{aligned}
f_{t(n)}(x) &= \frac{dF_{t(n)}(x)}{dx} = \frac{1}{2}\int_0^\infty f_{\chi^2(n)}(s) f_{X_0^2}\left(\frac{sx^2}{n}\right)\frac{2sx}{n}ds = \\
&= \frac{x}{n}\int_0^\infty f_{\chi^2(n)}(s) f_{X_0^2}\left(\frac{sx^2}{n}\right) sds = \\
&= \frac{x}{n}\int_0^\infty \frac{(1/2)^{n/2}}{\Gamma(n/2)} s^{\frac{n}{2}-1} e^{-\frac{s}{2}} \frac{(1/2)^{1/2}}{\Gamma(1/2)} \left(\frac{sx^2}{n}\right)^{-1/2} e^{-\frac{sx^2}{2n}} sds = \\
&= \frac{(1/2)^{(n+1)/2}}{\sqrt{\pi n}\,\Gamma(n/2)} \int_0^\infty s^{\frac{n+1}{2}-1} e^{-\frac{s}{2}\left(1+\frac{x^2}{n}\right)} ds = \left\{\begin{matrix} z=\frac{s}{2}\left(1+\frac{x^2}{n}\right),\ s=\frac{2z}{\left(1+\frac{x^2}{n}\right)} \\ ds=\frac{2dz}{\left(1+\frac{x^2}{n}\right)} \end{matrix}\right\} = \\
&= \frac{(1/2)^{(n+1)/2}}{\sqrt{\pi n}\,\Gamma(n/2)} \int_0^\infty \frac{2^{\frac{n+1}{2}-1} z^{\frac{n+1}{2}-1}}{\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}-1}} e^{-z} \frac{2dz}{1+\frac{x^2}{n}} = \\
&= \frac{1}{\sqrt{\pi n}\,\Gamma(n/2)} \left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} \int_0^\infty z^{\frac{n+1}{2}-1} e^{-z} dz = \\
&= \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\,\Gamma(n/2)} \left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}.
\end{aligned}
$$

□

Занятие 9. Доверительная оценка параметра a при неизвестной σ^2

Пусть $X_k \sim N(a, \sigma)$, для $k = 1, \ldots, n$.

9.1. Центральная статистика для a при неизвестной σ^2.

Величины $\hat{a}(\vec{X}_n) = \bar{X}$ и $S^2(\vec{X}_n)$ независимы (по доказанному выше следствию леммы Фишера).

Получаем, что

1. величина
$$\frac{\sqrt{n}(\hat{a} - a)}{\sigma} \sim N(0, 1);$$
2. величина
$$\frac{1}{\sigma^2} \sum_{k=1}^{n} (X_k - \hat{a})^2 \sim \chi^2(n-1).$$
3. Первая и вторая величины независимы (по следствию леммы из Фишера).

Из определения распределения Стьюдента следует, что
$$Z(\vec{X}_n) = \frac{\sqrt{n}(\hat{a} - a)/\sigma}{\sqrt{\frac{1}{n-1}\frac{1}{\sigma^2}\sum\limits_{k=1}^{n}(X_k - \hat{a})^2}} = \frac{\sqrt{n}(\hat{a} - a)}{\sqrt{S^2(\vec{X}_n)}} \sim t(n-1).$$
- центральная статистика.

9.2. Доверительный интервал для a.

$$P\{z_{(1-\alpha)/2} < Z(\vec{X}_n) < z_{(1+\alpha)/2}\} = \alpha$$

Плотность $t(n-1)$ является четной функцией, поэтому $z_{(1-\alpha)/2} = -z_{(1+\alpha)/2}$.

$$P\left\{-z_{(1+\alpha)/2} < \frac{\sqrt{n}(\hat{a} - a)}{\sqrt{S^2(\vec{X}_n)}} < z_{(1+\alpha)/2}\right\} = \alpha$$

$$P\left\{-\frac{\sqrt{S^2(\vec{X}_n)}}{\sqrt{n}}z_{(1+\alpha)/2} < \hat{a}-a < \frac{\sqrt{S^2(\vec{X}_n)}}{\sqrt{n}}z_{(1+\alpha)/2}\right\} = \alpha$$

$$P\left\{\hat{a}-\frac{\sqrt{S^2(\vec{X}_n)}}{\sqrt{n}}z_{(1+\alpha)/2} < a < \hat{a}+\frac{\sqrt{S^2(\vec{X}_n)}}{\sqrt{n}}z_{(1+\alpha)/2}\right\} = \alpha$$

Пример 1. По результатам 9 измерений напряжения батареи получено среднее арифметическое значение $30,6$ В, оценка среднего квадратического отклонения равна $0,2$ В. Требуется найти доверительный интервал для истинного значения напряжения батареи, соответствующий доверительной вероятности $0,95$, предполагая, что контролируемый признак имеет нормальный закон распределения.

◀

Пример 2. Из большой партии диодов взята выборка у которой измерено время восстановления (нс): 51, 62, 53, 52, 63. Считая что случайные величины распределены по нормальному закону, найти 95% доверительный интервал для среднего.

◀