Modelos de Probabilidad más comunes

- X: Variable aleatoria.
- $E(X)$: Esperanza (media) de X
- $V(X)$: Varianza de X
- $f(x)$: Función de densidad
- $F(x) = P(X \leq x)$: Función de Distribución.

1. **Binomial:** $X \sim B(n;p)$ con $n \in \mathbb{N}$, $0 < p < 1$ y $q = 1 - p$.
   - $P(X = k) = \binom{n}{k} p^k\, q^{n-k}, \quad$ para $k = 0, 1, 2, \ldots, n$.
   - $E(X) = n\,p, \quad V(X) = n\,p\,q$.

2. **Poisson:** $X \sim P(\lambda)$ con $\lambda > 0$.
   - $P(X = k) = e^{-\lambda}\,\dfrac{\lambda^k}{k!}, \quad$ para $k = 0, 1, 2, \ldots$.
   - $E(X) = \lambda, \quad V(X) = \lambda$.

3. **Geométrica:** $X \sim \mathrm{Geom}(p)$; $0 < p < 1$.
   - $P(X = k) = (1-p)^{k-1}p, \quad$ para $k = 1, 2, 3, \ldots$.
   - $E(X) = \dfrac{1}{p}, \quad V(X) = \dfrac{1-p}{p^2}$.

4. **Binomial negativa** $X \sim \mathrm{BinNeg}(n,p)$; $n \in \mathbb{N}$, $0 < p < 1$ y $q = 1 - p$.
   - $P(X = n + k) = \binom{n+k-1}{k} p^n\, q^k, \quad$ para $k = 0, 1, 2, \ldots$.
   - $E(X) = \dfrac{n}{p}, \quad V(X) = \dfrac{nq}{p^2}$.

5. **Hipergeométrica** $X \sim \mathrm{HGeom}(n, n_1, r)$; para $n \in \mathbb{N}$; $n_1 = 1, 2, \ldots, n-1$; $r \in \mathbb{N}$.
   - $P(X = k) = \dfrac{\binom{n_1}{k}\binom{n-n_1}{r-k}}{\binom{n}{r}}, \quad k = 0, 1, \ldots, \min\{n_1, r\}$.
   - $E(X) = \dfrac{r}{p}, \quad V(X) = r\dfrac{n_1}{n} \cdot \dfrac{n-n_1}{n} \cdot \dfrac{n-r}{n-1}$.

6. **Multinomial** $(X_1, \ldots, X_m) \sim \mathrm{Mult}(n, p_1, p_2, \ldots, p_{m-1})$; para $n \in \mathbb{N}$; $p_i > 0$, $p_1 + \cdots + p_m = 1$.
   - $P(X_1 = x_1, \ldots, X_m = x_m) = \dfrac{n!}{\prod_{i=1}^m x_i!} \prod_{i=1}^m p_i^{x_i}$, $x_i \in \mathbb{N}$, $x_1 + \cdots + x_m = n$.

7. **Uniforme:** $X \sim U(a,b)$ con $a < b$, $a, b \in \mathbb{R}$.
   - $f(x) = \dfrac{1}{b-a}, \quad$ si $x \in (a,b)$. $\qquad F(x) = \dfrac{x-a}{b-a}, \quad$ si $x \in (a,b)$.
   - $E(X) = \dfrac{a+b}{2}, \quad V(X) = \dfrac{(b-a)^2}{12}$.

8. **Exponencial:** $X \sim \exp(\lambda)$, con $\lambda > 0$.
   - $f(x) = \lambda\, e^{-\lambda x}$, $x > 0$. $\qquad F(x) = 1 - e^{-\lambda x}$, $x > 0$.
   - $E(X) = 1/\lambda, \quad V(X) = 1/\lambda^2$.

9. **Normal:** $X \sim N(\mu, \sigma)$ con $\mu \in \mathbb{R}$ y $\sigma > 0$.
   - $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in \mathbb{R}$.
   - $E(X) = \mu, \quad V(X) = \sigma^2$.

10. χ^2 **de Pearson con** n **grados de libertad:** $X \sim \chi^2_n$.

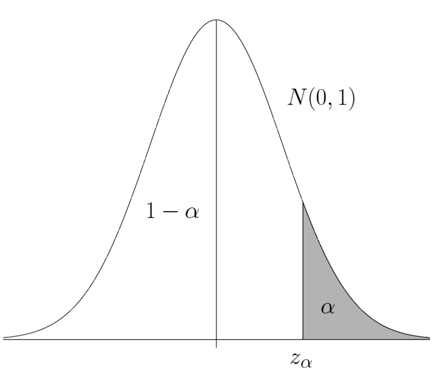
- $f(x) = \dfrac{x^{n/2-1} e^{-x/2}}{\text{constante} \cdot 2^{n/2}}$, $x > 0$.
- $E(X) = n, \quad V(X) = 2\,n$.

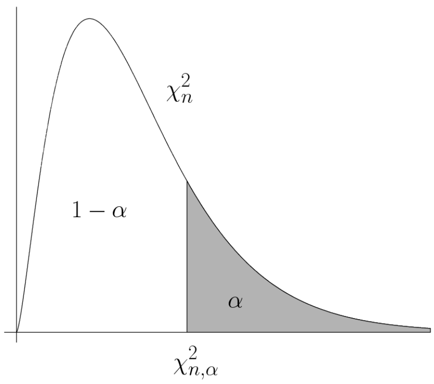
11. **t de Student con n grados de libertad:** $X \sim t_n$.

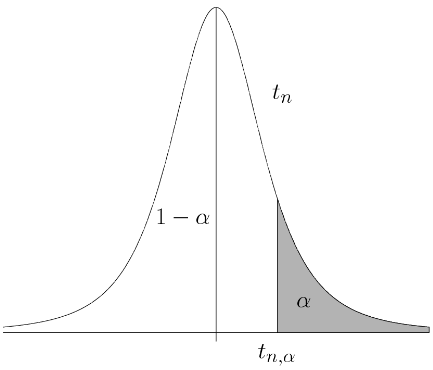
- $f(x) = \dfrac{1}{\sqrt{n\pi}} \dfrac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}$.
- $E(X) = 0; \quad V(X) = \dfrac{n}{n-2}$, $(n > 2)$.

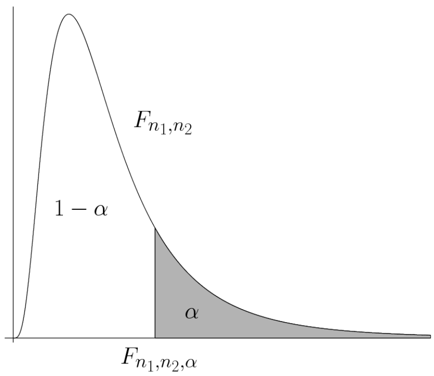
12. **F de Snedecor con m y n grados de libertad:** $X \sim F_{m,n}$

- $f(x) = \dfrac{m}{n} \dfrac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\dfrac{m}{n}x\right)^{\frac{m}{2}-1} \left(1 + \dfrac{m}{n}x\right)^{-\frac{m+n}{2}}$.
- $E(X) = \dfrac{n}{n-2}$, $(n > 2)$; $\quad V(X) = \dfrac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$, $(n > 4)$.









Aproximaciones de una distribución Binomial

- $\boxed{\text{Aproximación de una Binomial por una Normal}}$

  Para n grande en relación a p y $q = 1 - p$ $\quad (np \geq 10q$ y $nq \geq 10p)$:

$$B(n,p) \approx N\Big(\mu = n\,p, \sigma = \sqrt{n\,p\,(1-p)}\Big)$$

- $\boxed{\text{Aproximación de una Binomial por una Poisson}}$

  Para n grande $(n \geq 30)$ y $0 < p < 0.1$:

$$B(n,p) \approx P\,(\lambda = n\,p\,)$$

Estimadores

$(X_1, \ldots, X_n)$ muestra aleatoria simple (m.a.s.) de X.

Media muestral:

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

Cuasi-varianza:

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

$\boxed{\text{Distribución de } \bar{X} \text{ cuando } X \sim N(\mu,\sigma)}$

Si $X \sim N(\mu,\sigma)$ y $(X_1, X_2, \ldots, X_n)$ es una m.a.s. de X, entonces $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$

$\boxed{\text{Distribución de } S^2 \text{ cuando } X \sim N(\mu,\sigma)}$

Si $X \sim N(\mu,\sigma)$ y $(X_1, X_2, \ldots, X_n)$ es una m.a.s. de X, entonces

$$\frac{S^2}{\sigma^2/(n-1)} \sim \chi^2_{n-1}.$$

Intervalos de Confianza y Contrastes de Hipótesis

Intervalos de confianza más usuales

1. $\boxed{X \sim N(\mu,\sigma)}$

Intervalo de confianza $1-\alpha$ para μ: $\begin{cases} I = \left[\bar{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right] & (\sigma \text{ conocida}) \\ I = \left[\bar{x} \pm t_{n-1;\alpha/2}\dfrac{s}{\sqrt{n}}\right] & (\sigma \text{ desconocida}) \end{cases}$

Intervalo de confianza $1-\alpha$ para σ^2: $I = \left[\dfrac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}\ ,\ \dfrac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}\right]$

2. $\boxed{X \sim B(1,p)}$ (muestras grandes).

Intervalo de confianza $1-\alpha$ para p: $I = \left[\bar{x} \pm z_{\alpha/2}\sqrt{\dfrac{\bar{x}(1-\bar{x})}{n}}\right]$

3. $\boxed{X \sim P(\lambda)}$ (muestras grandes).

Intervalo de confianza $1-\alpha$ para λ: $I = \left[\bar{x} \pm z_{\alpha/2}\sqrt{\dfrac{\bar{x}}{n}}\right]$

4. $\boxed{\text{Dos poblaciones Normales independientes}}$

$X \sim N(\mu_1,\sigma_1)$, $Y \sim N(\mu_2,\sigma_2)$ independientes

$(X_1, \ldots, X_{n_1})$ m.a.s. de X; se calcula $\bar{x}$ y s_1^2.

$(Y_1, \ldots, Y_{n_2})$ m.a.s. de Y; se calcula $\bar{y}$ y s_2^2.

$$s_p^2 = \frac{(n_1-1)\,s_1^2 + (n_2-1)\,s_2^2}{n_1+n_2-2}$$

Intervalo de confianza $1-\alpha$ para $\mu_1 - \mu_2$:

$$I = \left[\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\,\right] \qquad \sigma_1,\, \sigma_2 \text{ conocidas}$$

$$I = \left[\bar{x} - \bar{y} \pm t_{n_1+n_2-2;\alpha/2}\ s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\,\right] \qquad \sigma_1,\ \sigma_2 \text{ desconocidas, } \sigma_1 = \sigma_2$$

El caso $\sigma_1,\ \sigma_2$ desconocidas, $\sigma_1 \neq \sigma_2$ es complicado

Intervalo de confianza $1-\alpha$ para σ_1^2/σ_2^2: $I = \left[\dfrac{s_1^2/s_2^2}{F_{n_1-1;n_2-1;\alpha/2}}\ ,\ (s_1^2/s_2^2)\, F_{n_2-1;n_1-1;\alpha/2}\right]$

5. $\boxed{\text{Comparación de proporciones}}$ (muestras grandes e independientes)

$X \sim B(1,p_1)$, $Y \sim B(1,p_2)$, independientes.

$(X_1,\dots,X_{n_1})$ m.a.s. de X; se calcula $\bar{x}$ y s_1^2.

$(Y_1,\dots,Y_{n_2})$ m.a.s. de Y; se calcula $\bar{y}$ y s_2^2.

Intervalo de confianza $1-\alpha$ para $p_1 - p_2$: $I = \left[\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\dfrac{\bar{x}\,(1-\bar{x})}{n_1} + \dfrac{\bar{y}\,(1-\bar{y})}{n_2}}\,\right]$

6. $\boxed{\text{Datos emparejados}}$

$X \sim \mathrm{N}(\mu_1,\sigma_1)$, $Y \sim \mathrm{N}(\mu_2,\sigma_2)$.

$D = X - Y \sim N(\mu = \mu_1 - \mu_2,\ \sigma)$,

donde el cálculo de σ supera el nivel de este curso.

Contrastes de hipótesis más usuales

- α = nivel de significación del contraste.
- n = tamaño de la muestra.
- H_0 = hipótesis nula.
- R = región crítica o de rechazo de H_0.

1.- $\boxed{X \sim N(\mu, \sigma)}$

$H_0 : \mu = \mu_0$ (σ conocida) $\quad R = \left\{ |\bar{x} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$

$H_0 : \mu = \mu_0$ (σ desconocida) $\quad R = \left\{ |\bar{x} - \mu_0| > t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} \right\}$

$H_0 : \mu \leq \mu_0$ (σ conocida) $\quad R = \left\{ \bar{x} - \mu_0 > z_{\alpha} \frac{\sigma}{\sqrt{n}} \right\}$

$H_0 : \mu \leq \mu_0$ (σ desconocida) $\quad R = \left\{ \bar{x} - \mu_0 > t_{n-1;\alpha} \frac{s}{\sqrt{n}} \right\}$

$H_0 : \mu \geq \mu_0$ (σ conocida) $\quad R = \left\{ \bar{x} - \mu_0 < z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$

$H_0 : \mu \geq \mu_0$ (σ desconocida) $\quad R = \left\{ \bar{x} - \mu_0 < t_{n-1;1-\alpha} \frac{s}{\sqrt{n}} \right\}$

$H_0 : \sigma = \sigma_0$ $\quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 \notin \left[\chi^2_{n-1;1-\alpha/2} \, , \, \chi^2_{n-1;\alpha/2} \right] \right\}$

$H_0 : \sigma \leq \sigma_0$ $\quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 > \chi^2_{n-1;\alpha} \right\}$

$H_0 : \sigma \geq \sigma_0$ $\quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 < \chi^2_{n-1;1-\alpha} \right\}$

2.- $\boxed{X \sim B(1, p)}$ (muestras grandes)

$H_0 : p = p_0$ $\quad R = \left\{ |\bar{x} - p_0| > z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$

$H_0 : p \leq p_0$ $\quad R = \left\{ \bar{x} - p_0 > z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$

$H_0 : p \geq p_0$ $\quad R = \left\{ \bar{x} - p_0 < z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$

3.- $\boxed{X \sim P(\lambda)}$ (muestras grandes)

$H_0 : \lambda = \lambda_0$ $\quad R = \left\{ |\bar{x} - \lambda_0| > z_{\alpha/2} \sqrt{\lambda_0/n} \right\}$

$H_0 : \lambda \leq \lambda_0$ $\quad R = \left\{ \bar{x} - \lambda_0 > z_{\alpha} \sqrt{\lambda_0/n} \right\}$

$H_0 : \lambda \geq \lambda_0$ $\quad R = \left\{ \bar{x} - \lambda_0 < z_{1-\alpha} \sqrt{\lambda_0/n} \right\}$

4.- $\boxed{\text{Dos poblaciones Normales independientes}}$

(s_p^2 calculado como en los intervalos de confianza)

$H_0: \mu_1 = \mu_2$ (σ_1, σ_2 conocidas) $\qquad R = \left\{ |\bar{x} - \bar{y}| > z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$

$H_0: \mu_1 = \mu_2$ ($\sigma_1 = \sigma_2$) $\qquad R = \left\{ |\bar{x} - \bar{y}| > t_{n_1+n_2-2;\alpha/2}\ s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$

$H_0: \mu_1 \leq \mu_2$ (σ_1, σ_2 conocidas) $\qquad R = \left\{ \bar{x} - \bar{y} > z_{\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$

$H_0: \mu_1 \leq \mu_2$ ($\sigma_1 = \sigma_2$) $\qquad R = \left\{ \bar{x} - \bar{y} > t_{n_1+n_2-2;\alpha}\ s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$

$H_0: \mu_1 \geq \mu_2$ (σ_1, σ_2 conocidas) $\qquad R = \left\{ \bar{x} - \bar{y} < z_{1-\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$

$H_0: \mu_1 \geq \mu_2$ ($\sigma_1 = \sigma_2$) $\qquad R = \left\{ \bar{x} - \bar{y} < t_{n_1+n_2-2;1-\alpha}\ s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$

$H_0: \sigma_1 = \sigma_2$ $\qquad R = \left\{ s_1^2/s_2^2 \notin \left[F_{n_1-1;n_2-1;1-\alpha/2}\ ,\ F_{n_1-1;n_2-1;\alpha/2}\right] \right\}$

$H_0: \sigma_1 \leq \sigma_2$ $\qquad R = \left\{ s_1^2/s_2^2 > F_{n_1-1;n_2-1;\alpha} \right\}$

$H_0: \sigma_1 \geq \sigma_2$ $\qquad R = \left\{ s_1^2/s_2^2 < F_{n_1-1;n_2-1;1-\alpha} \right\}$

5.- $\boxed{\text{Comparación de proporciones}}$ (muestras grandes e independientes)

$$\left. \begin{array}{ll} X \sim B(1, p_1), & (X_1, \dots X_{n_1}) \text{ m.a.s. de } X \\ Y \sim B(1, p_2), & (Y_1, \dots Y_{n_2}) \text{ m.a.s. de } Y \end{array} \right\} \rightsquigarrow \quad \bar{p} = \frac{\sum_i x_i + \sum_i y_i}{n_1 + n_2} = \frac{n_1\, \bar{x} + n_2\, \bar{y}}{n_1 + n_2}$$

$H_0: p_1 = p_2$ $\qquad R = \left\{ |\bar{x} - \bar{y}| > z_{\alpha/2} \sqrt{\bar{p}\,(1-\bar{p}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \right\}$

$H_0: p_1 \leq p_2$ $\qquad R = \left\{ \bar{x} - \bar{y} > z_{\alpha} \sqrt{\bar{p}\,(1-\bar{p}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \right\}$

$H_0: p_1 \geq p_2$ $\qquad R = \left\{ \bar{x} - \bar{y} < z_{1-\alpha} \sqrt{\bar{p}\,(1-\bar{p}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \right\}$

6.- $\boxed{\text{Datos emparejados}}$

$X \sim \mathrm{N}(\mu_X, \sigma_X)$, $Y \sim \mathrm{N}(\mu_Y, \sigma_Y)$.

$D = X - Y \sim N(\mu = \mu_X - \mu_Y,\ \sigma)$,

se utilizan los métodos para una muestra con σ desconocida.

Contrastes χ^2

- $\alpha =$ nivel de significación del contraste.
- $n =$ tamaño de la muestra.
- $H_0 =$ hipótesis nula.
- $R =$ región crítica o de rechazo de H_0.

1. $\boxed{\text{Contraste de la bondad del ajuste: Primer caso}}$

- H_0 : La población X sigue el modelo P indicado.
- $A_1, A_2, \dots, A_k$: k clases de los posibles valores de X.
- $O_i =$ frecuencia observada en la clase A_i.
- $e_i = n\,P(A_i) =$ frecuencia esperada en la clase A_i, suponiendo que H_0 es cierta.

$$R = \left\{ \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} = \sum_{i=1}^{k} \frac{O_i^2}{e_i} - n > \chi^2_{k-1;\,\alpha} \right\}$$

2. $\boxed{\text{Contraste de la bondad del ajuste: Segundo caso.}}$

- H_0 : La población X sigue algún modelo P_θ de una cierta familia de distribuciones
- $r =$ número de los parámetros desconocidos: $\theta = (\theta_1, \theta_2, \dots, \theta_r)$.
- $A_1, A_2, \dots, A_k$: k clases de los posibles valores de X.
- $O_i =$ frecuencia observada en la clase A_i.
- $e_i = n\,P_{\hat\theta}(A_i) =$ frecuencia esperada en la clase A_i, suponiendo que H_0 es cierta (y usando el estimador de máxima verosimilitud $\hat\theta$ del parámetro θ).

$$R = \left\{ \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} = \sum_{i=1}^{k} \frac{O_i^2}{e_i} - n > \chi^2_{k-1-r;\,\alpha} \right\}$$

3. $\boxed{\text{Contraste de homogeneidad de poblaciones}}$

- H_0 : Las p poblaciones $X_1, X_2, \dots, X_p$ son homogéneas
- $A_1, A_2, \dots, A_k$: k clases de los posibles valores de X.
- $O_{ij} =$ frecuencia observada en la clase A_i con la muestra j-ésima.
- $e_{ij} = n_j\,\hat P(A_i) = \frac{(\sum \text{columna } i\text{-ésima })\cdot(\sum \text{fila } j\text{-ésima })}{n} =$ frecuencia esperada en la clase A_i con la muestra j-ésima, si H_0 es cierta.

$$R = \left\{ \sum_{i=1}^{k} \sum_{j=1}^{p} \frac{(O_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k} \sum_{j=1}^{p} \frac{O_{ij}^2}{e_{ij}} - n > \chi^2_{(k-1)(p-1);\,\alpha} \right\}$$

4. $\boxed{\text{Contraste de independencia}}$

- H_0 : Las características X e Y de la población son independientes.
- $A_1 \times B_1, \dots, A_i \times B_j, \dots, A_k \times B_p$: $k\,p$ clases de los posibles valores de $X \times Y$.
- $O_{ij} =$ frecuencia observada en la clase $A_i \times B_j$.
- $e_{ij} = n\,\hat P(A_i)\,\hat P(B_j) = \frac{(\sum \text{columna } i\text{-ésima })\cdot(\sum \text{fila } j\text{-ésima })}{n} =$ frecuencia esperada en la clase $A_i \times B_j$ suponiendo que H_0 es cierta.

$$R = \left\{ \sum_{i=1}^{k} \sum_{j=1}^{p} \frac{(O_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k} \sum_{j=1}^{p} \frac{O_{ij}^2}{e_{ij}} - n > \chi^2_{(k-1)(p-1);\,\alpha} \right\}$$