Wzory:

- Ilość klas: $k = \sqrt{n}$
- Gęstość liczebności: $f_{ni} = \frac{n_i}{h_i}$
- Gęstość częstości: $f_{\omega i} = \frac{\omega_i}{h_i}$
- Rozpiętość przedziału: $h \approx \dfrac{x_{\max} - x_{\min}}{k} \approx \dfrac{R}{k}$
- Rozstęp: $R = x_{\max} - x_{\min}$
- Średnia arytmetyczna: $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$
- Średnia arytmetyczna ważona: $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{k} x_i n_i$ lub $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \dot{x}_i n_i$
- Średnia arytmetyczna dla grup: $\bar{\bar{x}} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{r} \bar{x}_i n_i$
- Średnia harmoniczna: $\bar{x}_H = \dfrac{n}{\displaystyle\sum_{i=1}^{n}\dfrac{1}{x_i}}$ lub $\bar{x}_H = \dfrac{\displaystyle\sum_{i=1}^{k} n_i}{\displaystyle\sum_{i=1}^{k}\dfrac{n_i}{x_i}}$
- Średnia geometryczna $\bar{x}_G = \sqrt[n]{\displaystyle\prod_{i=1}^{n} x_i}$
- Modalna $M_0 = x_{0m} + \dfrac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} h_m$
- Mediana: $Me = \begin{cases} x_{\frac{n+1}{2}} & \text{gdy } n \text{ jest nieparzyste} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{gdy } n \text{ jest parzyste} \end{cases}$ lub $Me = x_{0m} + \dfrac{N_{Me} - \displaystyle\sum_{i=1}^{m-1} n_i}{n_m} \cdot h_m$
  $N_{Me} = \dfrac{n}{2}$
- Kwantyl pierwszy: $Q_1 = x_{0m} + \dfrac{N_{Q_1} - \sum_{i=1}^{m-1} n_i}{n_m} \cdot h_m \qquad N_{Q_1} = \dfrac{n}{4}$
- Kwantyl trzeci: $Q_3 = x_{0m} + \dfrac{N_{Q_3} - \sum_{i=1}^{m-1} n_i}{n_m} \cdot h_m \qquad N_{Q_3} = \dfrac{3n}{4}$
- Wariancja: $s^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2$ lub $s^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i$
- Równość wariancyjna: $s^2 = \bar{s_i^2} + s^2(\bar{x}_i)$ gdzie $\bar{s_i^2} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{r} s_i^2 \cdot n_i$ oraz $s^2(\bar{x}_i) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{r}(\bar{x}_i - \bar{\bar{x}})^2 \cdot n_i$
- Odchylenie standardowe: $s = \sqrt{s^2}$
- Typowy obszar zmienności: $\bar{x} - s < x_{\text{typ}} < \bar{x} + s$ lub $Me - Q < x_{\text{typ}} < Me + Q$
- Odchylenie przeciętne: $d = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}|x_i - \bar{x}|$ lub $d = \dfrac{1}{n}\displaystyle\sum_{i=1}^{k}|x_i - \bar{x}| \cdot n_i$

- Reguła trzech sigm: $\begin{cases} (\bar{x}-3s, \bar{x}+3s) & 99,7\% \\ (\bar{x}-2s, \bar{x}+2s) & 95\% \\ (\bar{x}-s, \bar{x}+s) & 68\% \end{cases}$
- Odchylenie ćwiartkowe: $Q = \frac{Q_3 - Q_1}{2}$
- Klasyczne współczynniki zmienności: $V_s = \frac{s}{\bar{x}}$ $\qquad V_d = \frac{d}{\bar{x}}$
- Pozycyjne współczynniki zmienności: $V_Q = \frac{Q}{Me}$ $\qquad V_{Q_1,Q_3} = \frac{Q_3 - Q_1}{Q_1 + Q_3}$
- Rozkład symetryczny, gdy $\bar{x} = Me = M_0$
- Asymetria prawostronna: $\bar{x} > Me > M_0$
- Asymetria lewostronna: $\bar{x} < Me < M_0$
- Współczynniki skośnosci: $A_s = \frac{\bar{x} - M_0}{s}$ $\qquad A_d = \frac{\bar{x} - M_0}{d}$ $\qquad A_Q = \frac{Q_3 + Q_1 - 2Me}{2Q}$
- $A = \dfrac{m_3}{s^3} \qquad m_3 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3$ lub $m_3 = \dfrac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^3 n_i$
- Kurtoza: $K = \dfrac{m_4}{s^4} \qquad m_4 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$ lub $m_4 = \dfrac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^4 n_i \qquad K' = \frac{m_4}{s^4} - 3$
- Współczynnik koncentracji Lorentza: $K_L \approx 1 - \dfrac{1}{5000}\sum_{i=1}^{k}\dfrac{z_{isk} + z_{i-1sk}}{2} \cdot \omega_i$ lub $K_L = \frac{a}{5000} = \frac{5000-b}{5000}$

Regresja:

- Wieloraka:
  Y - zmienna zależna (objaśniana) $\qquad X_1 ... X_k$ - zmienne niezależne (objaśniające)
  $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_n X_{ki} + \varepsilon_i \quad i = 1...n$
  Równoważnie:

  $$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{nk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$

  Estymator MNK: $\hat{\beta} = (X^T X)^{-1} X^T Y$

  Współczynnik $R^2 = 1 - \dfrac{Y^T Y - \hat{\beta}^T X^T Y}{Y^T Y - n(\bar{Y})^2}$ $\qquad \bar{Y}$ - średnia arytmetyczna

- Modele regresji nieliniowej i ich linearyzacje

Nazwa	Postać	Transformacja
Regresja wielomianowa	$Y = a_0 + a_1 X + a_2 X^2 + \ldots + a_n X^n$	$X_1 = X, X_2 = X^2, ..., X_n = X^n$ $Y = a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$
Regresja potęgowa	$Y = aX^b$	$\log Y = \log a + b \log X$ $Y' = \log Y \quad a' = \log a \quad X' = \log X$ $Y' = a' + bX'$
Regresja wykładnicza	$Y = ab^X$	$\log Y = \log a + X \log b$ $Y' = \log Y \quad a' = \log a \quad b' = \log b$ $Y' = a' + b'X$
Regresja logarytmiczna	$Y = a + b \log X$	$X' = \log X$ $Y = a + bX'$

Przedział ufności dla wariancji na poziomie ufności $1-\alpha$

$X_1 ... X_n$ - niezależne zmienne losowe o jednakowym rozkładzie $\mathcal{N}(m, \sigma^2)$ gdzie m, σ^2 nieznane

- $n \leqslant 30$

  $$\mathcal{P}\left(\sigma^2 \in \left[\frac{ns^2}{\chi_1}, \frac{ns^2}{\chi_2}\right]\right) = 1-\alpha \text{ gdzie } s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}, \text{ a } \chi_1, \chi_2 \text{ spełniają równania}$$

  $$\mathcal{P}(\chi^2 \leqslant \chi_1) = \frac{\alpha}{2} \qquad \mathcal{P}(\chi^2 \geqslant \chi_2) = \frac{\alpha}{2} \text{ alternatywnie } \mathcal{P}(\chi^2 \leqslant \chi_2) = 1 - \frac{\alpha}{2} \qquad \chi^2 \sim \chi^2_{n-1}$$

- $n > 30$

  $$\mathcal{P}\left(\sigma^2 \in \left[\frac{s}{1+\frac{\mathcal{U}_\alpha}{\sqrt{2n}}}, \frac{s}{1-\frac{\mathcal{U}_\alpha}{\sqrt{2n}}}\right]\right) = 1-\alpha \text{ gdzie } s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}, \text{ a } \mathcal{U}_\alpha \text{ spełnia równanie}$$

  $$\mathcal{P}(|\mathcal{U}| \leqslant \mathcal{U}_\alpha) = 1-\alpha \qquad \mathcal{U} \sim \mathcal{N}(0,1)$$

Przedział ufności dla wskaźnika struktury na poziomie ufności $1-\alpha$

$X_1 ... X_n$ - niezależne zmienne losowe o jednakowym rozkładzie dwupunktowym z parametrem p, $n > 100$, m - liczba elementów wyróżnionych w próbie.

$$\mathcal{P}\left(p \in \left[\frac{m}{n} - \mathcal{U}_\alpha\sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}}, \frac{m}{n} + \mathcal{U}_\alpha\sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}}\right]\right) = 1-\alpha \text{ gdzie } \mathcal{U}_\alpha \text{ spełnia równanie}$$

$$\mathcal{P}(|\mathcal{U}| \leqslant \mathcal{U}_\alpha) = 1-\alpha \qquad \mathcal{U} \sim \mathcal{N}(0,1)$$

TESTOWANIE HIPOTEZ DLA WARIANCJI na poziomie istotności α

Model I: $X_1, \ldots, X_n$ i.i.d. rozkład $\mathrm{N}(m, \sigma^2)$, gdzie m, σ^2 nieznane, $n \leqslant 30$

$$\text{Statystyka testowa } \chi^2 = \frac{(n-1)\tilde{S}^2}{\sigma_0^2},$$

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

Hipotezy	(i) $H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	(ii) $H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	(iii) $H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$
Obszar krytyczny	$B = \{\chi^2 : \chi^2 \leqslant c_1 \vee \chi^2 \geqslant c_2\}$ gdzie $P(\chi^2 \leqslant c_1) = \frac{\alpha}{2}$, $P(\chi^2 \geqslant c_2) = \frac{\alpha}{2}$	$B = \{\chi^2 : \chi^2 \leqslant \chi^2_\alpha\}$ gdzie $P(\chi^2 \leqslant \chi^2_\alpha) = \alpha$	$B = \{\chi^2 : \chi^2 \geqslant \chi^2_\alpha\}$ gdzie $P(\chi^2 \geqslant \chi^2_\alpha) = \alpha$

Przy założeniu hipotezy H_0 zmienna losowa χ^2 ma rozkład chi-kwadrat z $n-1$ stopniami swobody.

Model II: $X_1, \ldots, X_n$ i.i.d. rozkład $\mathrm{N}(m, \sigma^2)$ lub zbliżony do normalnego, $n > 30$

$$\text{Statystyka testowa } U = \sqrt{\frac{2(n-1)\tilde{S}^2}{\sigma_0^2}} - \sqrt{2n-3},$$

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

Hipotezy	(i) $H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	(ii) $H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	(iii) $H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$
Obszar krytyczny	$B = \{u : \lvert u\rvert \geqslant u_\alpha\}$, gdzie $P(\lvert U\rvert \geqslant u_\alpha) = \alpha$	$B = \{u : u \leqslant u_\alpha\}$, gdzie $P(U \leqslant u_\alpha) = \alpha$	$B = \{u : u \geqslant u_\alpha\}$, gdzie $P(U \geqslant u_\alpha) = \alpha$

Przy założeniu hipotezy H_0 statystyka U ma rozkład asymptotycznie normalny $\mathrm{N}(0,1)$.

TESTOWANIE HIPOTEZ DLA WARTOŚCI ŚREDNIEJ NA POZIOMIE ISTOTNOŚCI α

1. **M o d e l I:** $X_1, \ldots, X_n$ i.i.d. rozkład $\mathrm{N}(m, \sigma^2)$, σ^2 - **znana** wariancja populacji

$$\text{Statystyka testowa } U = \frac{\bar{X} - m_0}{\sigma}\sqrt{n},$$

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

Hipotezy	(i) $H_0 : m = m_0$ $H_1 : m \neq m_0$	(ii) $H_0 : m = m_0$ $H_1 : m < m_0$	(iii) $H_0 : m = m_0$ $H_1 : m > m_0$
Obszar krytyczny	$B = \{u : \lvert u\rvert \geqslant u_\alpha\}$, gdzie $P(\lvert U\rvert \geqslant u_\alpha) = \alpha$	$B = \{u : u \leqslant u_\alpha\}$, gdzie $P(U \leqslant u_\alpha) = \alpha$	$B = \{u : u \geqslant u_\alpha\}$, gdzie $P(U \geqslant u_\alpha) = \alpha$

Przy założeniu hipotezy H_0 zmienna losowa U ma rozkład $\mathrm{N}(0, 1)$.

2. **M o d e l II:** $X_1, \ldots, X_n$ i.i.d. rozkład $\mathrm{N}(m, \sigma^2)$, σ^2 - **nieznana** wariancja populacji

$$\text{Statystyka testowa } T = \frac{\bar{X} - m_0}{S}\sqrt{n},$$

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

Hipotezy	(i) $H_0 : m = m_0$ $H_1 : m \neq m_0$	(ii) $H_0 : m = m_0$ $H_1 : m < m_0$	(iii) $H_0 : m = m_0$ $H_1 : m > m_0$
Obszar krytyczny	$B = \{t : \lvert t\rvert \geqslant t_\alpha\}$, gdzie $P(\lvert T\rvert \geqslant t_\alpha) = \alpha$	$B = \{t : t \leqslant t_\alpha\}$, gdzie $P(T \leqslant t_\alpha) = \alpha$	$B = \{t : t \geqslant t_\alpha\}$, gdzie $P(T \geqslant t_\alpha) = \alpha$

Przy założeniu hipotezy H_0 zmienna losowa T ma rozkład t-Studenta z $n-1$ stopniami swobody.

3. **M o d e l III:** $X_1, \ldots, X_n$ i.i.d. dowolny rozkład z wartością średnią m i skończoną wariancją σ^2, ponadto $n \geqslant 30$.

$$\text{Statystyka testowa } U = \frac{\bar{X} - m_0}{S}\sqrt{n},$$

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

Hipotezy	(i) $H_0 : m = m_0$ $H_1 : m \neq m_0$	(ii) $H_0 : m = m_0$ $H_1 : m < m_0$	(iii) $H_0 : m = m_0$ $H_1 : m > m_0$
Obszar krytyczny	$B = \{u : \lvert u\rvert \geqslant u_\alpha\}$, gdzie $P(\lvert U\rvert \geqslant u_\alpha) = \alpha$	$B = \{u : u \leqslant u_\alpha\}$, gdzie $P(U \leqslant u_\alpha) = \alpha$	$B = \{u : u \geqslant u_\alpha\}$, gdzie $P(U \geqslant u_\alpha) = \alpha$

TESTOWANIE HIPOTEZ DLA WSKAŹNIKA STRUKTURY na poziomie istotności α

M o d e l: Populacja generalna ma rozkład dwupunktowy z parametrem p. Liczebność próby równa jest $n \geqslant 100$.

$$\text{Wartość statystyki testowej } U \text{ wynosi: } u = \frac{\frac{m}{n} - p_0}{\sqrt{\frac{p_0 q_0}{n}}},$$

gdzie m jest liczbą elementów wyróżnionych znalezionych w próbie, $q_0 = 1 - p_0$.

Hipotezy	(i) $H_0 : p = p_0$ $H_1 : p \neq p_0$	(ii) $H_0 : p = p_0$ $H_1 : p < p_0$	(iii) $H_0 : p = p_0$ $H_1 : p > p_0$
Obszar krytyczny	$B = \{u : \lvert u\rvert \geqslant u_\alpha\}$, gdzie $P(\lvert U\rvert \geqslant u_\alpha) = \alpha$	$B = \{u : u \leqslant u_\alpha\}$, gdzie $P(U \leqslant u_\alpha) = \alpha$	$B = \{u : u \geqslant u_\alpha\}$, gdzie $P(U \geqslant u_\alpha) = \alpha$

Przy założeniu hipotezy H_0 statystyka U ma rozkład asymptotycznie normalny $\mathrm{N}(0, 1)$.

9.3. Testowanie hipotezy o równości dwóch wartości przeciętnych

Dane są dwie zbiorowości generalne o rozkładach normalnych $N(m_1, \sigma_1)$ i $N(m_2, \sigma_2)$. Chcemy zweryfikować hipotezę $H_0: m_1 = m_2$ wobec hipotezy $H_1: m_1 \neq m_2$ (lub $H_1: m_1 < m_2$, albo $H_1: m_1 > m_2$). Niech n_1, n_2 oznaczają wielkości prób prostych, wylosowanych z każdej zbiorowości, a $\overline{X}_1$, $\overline{X}_2$ oraz $S^2_{(1)}$ i $S^2_{(2)}$ oznaczają odpowiednio średnie arytmetyczne i wariancję S^2 z prób.

W zależności od założeń dotyczących zbiorowości generalnych oraz od liczebności prób — sprawdzian hipotezy H_0 ma różną postać i jest związany z rozkładem normalnym lub rozkładem Studenta.

Jeśli:

* σ_1, σ_2 — znane i $n_1 \leqslant 30$, $n_2 \leqslant 30$,
* σ_1, σ_2 — znane i $n_1 > 30$, $n_2 > 30$,
* σ_1, σ_2 — nieznane i $n_1 > 30$, $n_2 > 30$, to $\sigma_1^2 \approx S^2_{(1)}$, $\sigma_2^2 \approx S^2_{(2)}$,

wówczas sprawdzian hipotezy ma postać:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{9.3}$$

Rozkład tej statystyki, przy założeniu prawdziwości H_0, jest $N(0, 1)$.

W przypadku, gdy:

* σ_1, σ_2 — nieznane, lecz $\sigma_1 = \sigma_2$ i $n_1 \leqslant 30$, $n_2 \leqslant 30$,

to korzystamy ze sprawdzianu:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{n_1 S^2_{(1)} + n_2 S^2_{(2)}}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{9.4}$$

który, przy założeniu prawdziwości H_0, ma rozkład Studenta o $(n_1 + n_2 - 2)$ stopniach swobody.