**Zadanie 1: (1)**

Niech $X_1, \ldots, X_n$ będzie próbą losową z rozkładu normalnego $N(m, 2^2)$. Hipotezę $H_0 : m = 1$ przy alternatywie $H_1 : m = 3$ będziemy weryfikować, wykorzystując test o zbiorze krytycznym postaci $\{(x_1, \ldots, x_n) : \sum_{i=1}^{n} x_i = k_\alpha\}$.

- Wyznacz $k_\alpha$, aby otrzymać test o rozmiarze $0,05$.
- Jak dużą próbę losową należy pobrać, aby uzyskać test o mocy nie mniejszej niż $0,95$?

**Rozwiązanie:**

Przyjmuje że zapis $N(m, 2^2)$ oznacza, że na drugiej współrzędnej stoi wariancja i będę się trzymał tej konwencji.

Chcemy policzyć

$$P(X_1 + \cdots + X_n > k_\alpha | H_0) = P(\frac{X_1 + \cdots + X_n}{n} > \frac{k_\alpha}{n} | m = 1)$$

Weźmy

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} \sim \mathrm{Norm}(1, \frac{4}{n})$$

Wówczas

$$P(\bar{X} > \frac{k_\alpha}{n}) = P(\frac{\bar{X} - 1}{2}\sqrt{n} > \frac{k_\alpha - n}{2\sqrt{n}}) = 0,05$$

$$\frac{\bar{X} - 1}{2}\sqrt{n} \sim N(0, 1)$$

Korzystając z tablic dla rozkładu normalnego obliczam

$$\frac{k_\alpha - n}{2\sqrt{n}} = 1,645 \Rightarrow k_\alpha = 3,29\sqrt{n} + n$$

Policzenie mocy o wartości nie mniejszej niż $0,95$ sprowadza się do rozwiązanie nierówności

$$P(X_1 + \cdots + X_n > k_\alpha | H_1) = P(\frac{X_1 + \cdots + X_n}{n} > \frac{k_\alpha}{n} | m = 3) \geq 0,95$$

Następnie rozumowanie analogiczne jak wcześniej, otrzymujemy

$$\frac{k_\alpha - 3n}{2\sqrt{n}} \leq -1,645 \Rightarrow 0 \leq n - 3,29\sqrt{n}$$

Najmniejsze $n$ dla jakiego to zachodzi wynosi $n = 11$

**Zadanie 1:**

Niech $X_1, ..., X_n$ będzie próbą losową z rozkładu normalnego $N(m, 2^2)$ Hipotezę $H_0 : m = 1$ przy alternatywie $H_1 : m = 3$ 3 będziemy weryfikować, wykorzystując test o zbiorze krytycznym postaci $\{(x_1, x_2, ..., x_n) : \sum_{i=1}^{n} x_i > k_\alpha\}$.

• Wyznacz k $k\alpha$, aby otrzymać test o rozmiarze 0,05.

• Jak dużą próbę losową należy pobrać, aby uzyskać test o mocy nie mniejszej niż 0,95?

**Rozwiązanie:**

a)$\sum_{i=1}^{n} x_i$ ma rozkład $N(nm, n2^2)$

$P(\sum_{i=1}^{n} x_i > k_\alpha \mid H_0) = \alpha$

Jeśli $H_0$ to suma ma rozkład $N(n, n2^2)$

$\sum_{i=1}^{n} x_i = U$

$P(\frac{U-n}{2\sqrt{n}} > \dfrac{k_\alpha - n}{2\sqrt{n}}) = 1 - P(\frac{U-n}{2\sqrt{n}} \leq \dfrac{k_\alpha - n}{2\sqrt{n}})$

$1 - P(\frac{U-n}{2\sqrt{n}} \leq \dfrac{k_\alpha - n}{2\sqrt{n}}) = 0,05$

$P(\frac{U-n}{2\sqrt{n}} \leq \dfrac{k_\alpha - n}{2\sqrt{n}}) = 0,95$

$\dfrac{k_\alpha - n}{2\sqrt{n}} = 1,65$

$k_\alpha = 3,3\sqrt{n} + n$

b)$P(\sum_{i=1}^{n} x_i > k_\alpha \mid H_1) = 1 - \beta$

$P(\sum_{i=1}^{n} x_i > k_\alpha \mid H_1) \geqslant 0,95$

Jeśli $H_1$ to U ma rozkład $N(3n, n2^2)$

$P(\sum_{i=1}^{n} x_i > k_\alpha) = P(U > k_\alpha) = P(\frac{U-3n}{2\sqrt{n}} > \dfrac{k_\alpha - 3n}{2\sqrt{n}}) = 1 - P(\frac{U-3n}{2\sqrt{n}} \leqslant \dfrac{k_\alpha - 3n}{2\sqrt{n}})$

$1 - p(\frac{u-3n}{2\sqrt{n}} \leqslant \dfrac{k_\alpha - 3n}{2\sqrt{n}}) \geqslant 0,95$

$p(\frac{u-3n}{2\sqrt{n}} \leqslant \dfrac{k_\alpha - 3n}{2\sqrt{n}}) \leqslant 0,05$

$\phi(\dfrac{k_\alpha - 3n}{2\sqrt{n}}) \leqslant 0,05$

$1 - \phi(-\dfrac{k_\alpha - 3n}{2\sqrt{n}}) \leqslant 0,05$

$-\dfrac{k_\alpha - 3n}{2\sqrt{n}} \geq 1,65$

$\dfrac{k_\alpha - 3n}{2\sqrt{n}} \leqslant -1,65$

$k_\alpha \leqslant -3,3\sqrt{n} + 3n$

$3,3\sqrt{n} + n \leqslant -3,3\sqrt{n} + 3n$

$6,6\sqrt{n} - 2n \leqslant 0$

$n \geq 10,89$

próba losowa musi być równa co najmniej 11 by moc testu wynosiła 0,95

**Zadanie 2:**

Wyhodowano nową odmianę pewnej rośliny. Hipotezę, że kiełkuje 70% sadzonek, wobec hipotezy alternatywnej, że kiełkuje więcej niż 70%, testowano na podstawie próbki 10 sadzonek. Hipotezę zerową odrzucamy, gdy wykiełkuje 8 lub więcej sadzonek.

- Czy rozmiar tego testu jest mniejszy niż 0,05?

- Jaki jest rozmiar innego testu, który odrzuca hipotezę zerową, jeśli wszystkie sadzonki wykiełkują?

**Rozwiązanie:**

$$H_0 : \mu = 0.7 => X \sim N(0.7, 10)$$
$$H_1 : \mu > 0.7$$

(a)

$$P(x \geq 8) = \binom{10}{8}(0.7)^8(0.3)^2 + \binom{10}{9}(0.7)^9(0.3) + \binom{10}{10}(0.7)^{10}$$
$$\approx 0.38 > 0.05$$

Zatem rozmiar tego testu nie jest mniejszy od 0.05

(b)

$$P(x = 10) = (0.7)^{10} = 0.02 < 0.05$$

Rozmiar tego testu jest równy 0.02

**Zadanie 2:**
Wyhodowano nową odmianę pewnej rośliny. Hipotezę, że kiełkuje 70% sadzonek, wobec hipotezy alternatywnej, że kiełkuje więcej niż 70%, testowano na podstawie próbki 10 sadzonek. Hipotezę zerową odrzucamy, gdy wykiełkuje 8 lub więcej sadzonek.

- Czy rozmiar tego testu jest mniejszy niż 0,05?
- Jaki jest rozmiar innego testu, który odrzuca hipotezę zerową, jeśli wszystkie sadzonki wykiełkują?

**Rozwiązanie:**

a) Rozmiar tego testu jest równy jego poziomowi istotności (hipoteza prosta)

$H_0 : kiekuje = 70\%$

$H_1 : kiekuje > 70\%$

$n = 10$

$P(x \in W \mid H_0) = \alpha$

$H_0 \sim Binom(10, p)$

Zakładamy, że $H_0$ jest prawdziwe czyli p = 0,7 i

$P(x \in W \mid H_0) = \frac{P(X \in W \bigcap H_0)}{1}$

$P(X \in W \bigcap H_0) = \binom{n}{k} p^k (1-p)^{n-k} = \sum_{i=8}^{10} \binom{10}{i} p^i (1-p)^{10-i} = \sum_{i=8}^{10} \binom{10}{i} (\frac{7}{10})^i (\frac{3}{10})^{10-i} \approx 0,38$

0,38 > 0,05

Rozmiar tego testu jest większy niż 0,05

b) $H_0 : wykiekuje 70\%$

$H_1$ : wykiełkują wszystkie

$$\alpha = P(x \in W \mid H_0) = \frac{P(x \in W \cap H_0)}{P(H_0)}$$

Zakładamy, że $H_0$ jest spełnione, czyli $P(H_0) = 1$

$\alpha = P(x \in W \cap H_0)$

$P(x \in W \cap H_0) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{10}{10} (\frac{7}{10})^{10} (\frac{3}{10})^0 \approx 0,028$

Rozmiar tego testu to ok 0,028

**Zadanie 3:**
Populacja ma rozkład opisany funkcją gęstości postaci $f(x) = \beta e(-\beta x)$ dla $x > 0$. Z tej populacji wylosowano dziesięcioelementową próbę:

1 0,8 1,7 5,5 1,9 8,1 2,6 2,5 1,4 2,4

Przetestuj, wykorzystując test najmocniejszy przy poziomie istotności $\alpha = 0,05$ hipotezę zerową, że $\beta = \frac{1}{2}$, przeciw hipotezie alternatywnej, że $\beta = \frac{1}{3}$.

**Rozwiązanie:** Skorzystam z lematu Neymana-Pearsona dla funkcji wiarygodności $L$:

$$\begin{aligned} L(x_1, \ldots, x_n, \beta) &= \Pi_{i=1}^n \beta \exp(-\beta x_i) \\ &= \beta^n \exp\left(-\beta \sum_{i=1}^n x_i\right) \\ &= \beta^n \exp(-\beta n \bar{x}) \end{aligned}$$

Wtedy dostajemy następujące $\Lambda$:

$$\begin{aligned} \Lambda(x_1, \ldots, x_n, \beta_0, \beta_1) &= \frac{L(x_1, \ldots, x_n, \beta_1)}{L(x_1, \ldots, x_n, \beta_0)} \\ &= \frac{\beta_1^n \exp(-\beta_! n \bar{x})}{\beta_0^n \exp(-\beta_0 n \bar{x})} \\ &= \left(\frac{\beta_1}{\beta_0}\right)^n \exp((\beta_0 - \beta_1) n \bar{x}) \end{aligned}$$

Obszar krytyczny wyznaczam z nierówności $\mathbb{P}(\Lambda > k) = \alpha$. Operując na konkretnych liczbach ($x \in \mathbb{R}^{10}, x_i \geq 0$):

$$\Lambda\left(x, \frac{1}{2}, \frac{1}{3}\right) = \left(\frac{1/3}{1/2}\right)^{10} \exp\left(\left(\frac{1}{2} - \frac{1}{3}\right) 10\bar{x}\right) = \left(\frac{2}{3}\right)^{10} \exp\left(\frac{10}{6}\bar{x}\right)$$

Upraszczając dla uzyskania $\bar{x}$:

$$\begin{aligned} \Lambda\left(x, \frac{1}{2}, \frac{1}{3}\right) &> k \\ \left(\frac{2}{3}\right)^{10} \exp\left(\frac{10}{6}\bar{x}\right) &> k \\ \bar{x} &> 6 \cdot \ln\left(\frac{3}{2}\right) + \frac{6}{10} \cdot \ln k \\ \sum_{i=1}^{10} x_i &> 60 \cdot \ln\left(\frac{3}{2}\right) + 6 \ln k \end{aligned}$$

Rozkład $\Gamma$ jest uogólnieniem rozkładu wykładniczego. Jeśli $X \sim \mathrm{Exp}(\beta)$, to $X \sim \Gamma(1, \beta)$, a $\sum_{i=1}^n X_i \sim \Gamma(n, \beta)$. Odczytuję kwantyl z RStudio:

```
> qgamma(c(0.05), shape=10, rate=0.5, lower.tail=FALSE)
31.41043
```

Stąd maksymalna dozwolona wartość $k$ to:

$$k = \exp\left(\frac{1}{6}\left(31.41043 - 60\ln\left(\frac{3}{2}\right)\right)\right) \approx \exp(1.18042) \approx 3.25574$$

Jednocześnie uzyskana $\Lambda$ to:

$$\Lambda\left([1, 0.8, \dots]^T, \frac{1}{2}, \frac{1}{3}\right) = \left(\frac{2}{3}\right)^{10} \exp\left(\frac{10}{6} \cdot 2.79\right) \approx 1.8137$$

Wartość $\Lambda$ naszej próby jest znacznie poniżej progu, więc istnieją przesłanki do przyjęcia hipotezy zerowej.

**Zadanie 3:**
Populacja ma rozkład opisany funkcją gęstości postaci f(x) = $\beta e^{(-\beta x)}$ dla $x > 0$. Z tej populacji wylosowano dziesięcioelementową próbę:

1 0,8 1,7 5,5 1,9 8,1 2,6 2,5 1,4 2,4

Przetestuj, wykorzystując test najmocniejszy przy poziomie istotnośc $\alpha = 0,05$ hipotezę zerową, że $\beta = \frac{1}{2}$, przeciw hipotezie alternatywnej, że $\beta = \frac{1}{3}$.

**Rozwiązanie:** $\alpha = 0,05$
n = 10
$H_0 : \beta = \frac{1}{2}$

$H_1 : \beta = \frac{1}{3}$

$f_1(x) = \prod_{i=1}^n \frac{1}{2} e^{-\frac{1}{2} x_i} = (\frac{1}{2})^{10} e^{-\frac{1}{2}\sum_{i=1}^{10} x_i} = 2^{-10} e^{-\frac{1}{2}\sum_{i=1}^{10} x_i}$

$f_2(x) = \prod_{i=1}^n \frac{1}{3} e^{-\frac{1}{3} x_i} = (\frac{1}{3})^{10} e^{-\frac{1}{3}\sum_{i=1}^{10} x_i} = 3^{-10} e^{-\frac{1}{3}\sum_{i=1}^{10} x_i}$

$\frac{f_2(x)}{f_1(x)} = \frac{3^{-10} e^{-\frac{1}{3}\sum_{i=1}^{10} x_i}}{2^{-10} e^{-\frac{1}{2}\sum_{i=1}^{10} x_i}} = (\frac{3}{2})^{-10} e^{-\frac{1}{3}\sum_{i=1}^n x_i + \frac{1}{2}\sum_{i=1}^n x_i} = (\frac{3}{2})^{-10} e^{\frac{1}{6}\sum_{i=1}^n x_i} \approx 1,81366$

$\alpha = P(\frac{f_2(x)}{f_1(x)} \geqslant k) = P((\frac{3}{2})^{-10} e^{\frac{1}{6}\sum_{i=1}^n x_i} \geqslant k) = P(e^{\frac{1}{6}\sum_{i=1}^n x_i} \geqslant (\frac{3}{2})^{10} k) = P(\frac{1}{6}\sum_{i=1}^n x_i \geqslant \ln((\frac{3}{2})^{10} k))$

$\alpha = P(\sum_{i=1}^n x_i \geqslant 6\ln((\frac{3}{2})^{10} k))$

$0,05 = P(\sum_{i=1}^n x_i \geqslant 6\ln((\frac{3}{2})^{10} k))$

$0,05 = P(\sum_{i=1}^n x_i \geqslant C)$

Sprawdzam wartość kwantyla rozkładu chi-kwadrat dla 20 stopni swobody (2n) i $\alpha = 0,05$ i otrzymuję wartość 31,4104

$0,05 = P(\sum_{i=1}^{n} x_i \geqslant 31,4104)$

$6\ln((\frac{3}{2})^{10}k) = 31,4104$

$\ln((\frac{3}{2})^{10}k) = 5,235$

$\ln((\frac{3}{2})^{10}) + \ln(k) = 5,235$

$\ln(k) = 1,18$

$k = e^{1,18}$

$k = 3,25$

$1,81360 < 3,25$

$\frac{f_2(x)}{f_1(x)} < k$

Czyli nie ma podstaw do odrzucenia $H_0$

**Zadanie 4:**
Skonstruuj test ilorazu wiarogodności dla rozkładu wielomianowego. Przyjmij hipotezy

- Zerową $H_0 : p = [p_1(\theta), \dots, p_k(\theta)]$, $\theta \in \omega_0$
- Alternatywną $H_1 : p \neq [p_1(\theta), \dots, p_k(\theta)]$, $\theta \in \omega_0$, nie czyni żadnych założeń o prawdopodobieństwach $p$, poza $\sum_i p_i = 1$. $\Omega = \{[p_1(\theta), \dots, p_k(\theta)] | \sum_i p_i = 1\}$

**Rozwiązanie:**

$$P_X([n_1, ..., n_k]^T = \binom{n}{n_1 ... n_k} p_1^{n_1} p_2^{n_2} ... p_k^{n_k}$$

$$\sum_i n_i = n$$

$$L(n_1, \dots, n_k, \theta) = n! \prod_{i=1}^{k} \frac{p_i(\theta)^{n_i}}{n_i!}$$

$$\wedge = \frac{\max_{\theta \in \omega_0} L(n_1, \dots, n_k, \theta)}{\max_{\theta \in \Omega} L(n_1, \dots, n_k, \theta})$$

Korzystając z zadania 5 z poprzedniej serii zadań, estymator największej wiarygodności dla rozkładu wielomianowego wynosi

$$\hat{p}_i = \frac{n_i}{n}$$

Obliczamy statystykę ilorazu wiarygodności

$$\wedge = \frac{\prod_{i=1}^{k} \frac{p_i(\theta)^{n_i}}{n_i!}}{\prod_{i=1}^{k} \frac{\hat{p_i}^{n_i}}{n_i!}} = \frac{\prod_{i=1}^{k} p_i(\theta)^{n\hat{p}_i}}{\prod_{i=1}^{k} \hat{p_i}^{n\hat{p}_i}} = \prod_{i=1}^{k} (\frac{p_i(\theta)}{\hat{p}_i})^{n\hat{p}_i}$$

Korzystając z twierdzenia z wykładu mamy

$$-2\ln(\wedge) = -2n \sum_{i=1}^{k} \hat{p}_i \ln(\frac{p_i(\theta)}{\hat{p}_i})$$

Rozkład $-2\ln(\wedge)$ zbiega do rozkładu $\chi^2(k-1)$ przy $n \to \infty$

**Zadanie 5:**
Pokazać, że przy hipotezie zerowej spełnionej, iloraz wiarogodności w teście dla rozkładu wielomianowego i statystyka w teście zgodności Pearsona są asymptotycznie równoważne.
**Rozwiązanie:** Z zadania 4 mamy

$$2\ln(\wedge) = 2n\sum_{i=1}^{k}\hat{p_i}\ln(\frac{\hat{p_i}}{p_i(\theta)})$$

Weźmy funkcję $g(y) = y\ln(\frac{y}{y_0})$
Rozwijamy $g$ w szereg Taylora wokół $y_0$

$$g(y) \sim (y - y_0) + \frac{1}{2x_0}(x - x_0)^2$$

Przyjmujemy $y = \hat{p_i}, y_0 = p_i(\theta)$, wówczas

$$2n\sum_{i=1}^{k}\hat{p_i}\ln(\frac{\hat{p_i}}{p_i(\theta)}) \sim 2n\sum_{i=1}^{n}(\hat{p_i} - p_i(\theta)) + n\sum_{i=1}^{n}\frac{(\hat{p_i} - p_i(\theta))^2}{p_i(\theta)}$$

Hipotez zerowa jest spełniona stąd dddn $\hat{p_i} \sim p_i(\theta)$ Wtedy

$$2n\sum_{i=1}^{k}\hat{p_i}\ln(\frac{\hat{p_i}}{p_i(\theta)}) \sim n\sum_{i=1}^{n}\frac{(\hat{p_i} - p_i(\theta))^2}{p_i(\theta)}$$

A prawa strona jest dokładnie równa statystyce w teście zgodności Pearsona, zatem przy $n \to \infty$ iloraz wiarogodności w teście dla rozkładu wielomianowego i statystyka w teście zgodności Pearsona są asymptotycznie równoważne.

**Zadanie 6:**
Zmienna losowa X ma gęstość $\frac{1}{\theta}\delta_{(0,\theta)}(x)$, gdzie $\delta_{(0,\theta)}(x)$ to funkcja przyjmująca wartość 1 dla $x \in (0,\theta)$ i i 0 dla x spoza tego przedziału, a $\theta$ jest nieznanym parametrem. Niech c będzie ustaloną dodatnią stałą. Test polega na tym, że jeśli $X \geqslant c$, to należy przyjąć hipotezę alternatywną $H_1 : \theta = 4$, a gdy $X < c$, należy przyjąć hipotezę zerową, $H_0 : \theta = 2$. Obliczyć
a) prawdopodobieństwa błędów pierwszego i drugiego rodzaju ($\alpha$ i $\beta$),
b) moc testu,
c) $\beta$, gdy $\alpha$ = 0.05.

**Rozwiązanie:**
a) prawdopodobieństwa błędów pierwszego i drugiego rodzaju ($\alpha$ i $\beta$):
$\alpha(W) = \mathbb{P}(Z \in W|H_0) = \mathbb{P}(X \in [c,\infty)|\theta = 2) = \int_c^{+\infty}\frac{1}{2}1_{(0<x<2)}dx =$
$$= \begin{cases} \int_{-\infty}^{0} 0\ dx + \int_0^2 \frac{1}{2}\times 1\ dx + \int_2^{+\infty} 0\ dx\ ,\ dla\ c < 0 \\ \int_c^2 \frac{1}{2}\times 1\ dx + \int_2^{+\infty} 0\ dx\ ,\ dla\ 0 < c < 2 \\ \int_c^{+\infty} 0\ dx\ ,\ dla\ 2 < c \end{cases} = \begin{cases} 1\ dla\ c < 0 \\ 1 - \frac{c}{2}\ dla\ 0 < c < 2 \\ 0\ dla\ 2 < c \end{cases}$$
$\beta(W) = \mathbb{P}(Z \in W'|H_1) = \mathbb{P}(X \in (-\infty,c)|\theta = 4) = \int_{-\infty}^{c}\frac{1}{4}1_{(0<x<4)}dx =$
$$= \begin{cases} \int_{-\infty}^{c} 0\ dx\ ,\ dla\ c < 0 \\ \int_{-\infty}^{0} 0\ dx + \int_0^c \frac{1}{4}\times 1\ dx\ ,\ dla\ 0 < c < 4 \\ \int_{-\infty}^{0} 0\ dx + \int_0^4 \frac{1}{4}\times 1\ dx + \int_4^c 0\ dx\ ,\ dla\ 4 < c \end{cases} = \begin{cases} 0\ dla\ c < 0 \\ \frac{c}{4}\ dla\ 0 < c < 4 \\ 1\ dla\ 4 < c \end{cases}$$
b) moc testu:
$$\mathbb{P}(Z \in W|H_1) = 1 - \beta(W) = \begin{cases} 1 - 0 = 1\ dla\ c < 0 \\ 1 - \frac{c}{4}\ dla\ 0 < c < 4 \\ 1 - 1 = 0\ dla\ 4 < c \end{cases}$$

c) $\beta$, gdy $\alpha$ = 0.05:
$\alpha = 0.05 \in (0, 1)$
$c \in (0, 2)$
$\alpha = 1 - \frac{c}{2} = 0.05$
$0.95 = \frac{c}{2}$
$c = 1.9$

$c \in (0, 4)$
$\beta(W) = \frac{c}{4} = \frac{1.9}{4} = 0.475$

**Zadanie 7:**
Cena pewnego wyrobu jest różna w zależności od punktu sprzedaży. Przyjęto, że cena w wylosowanym punkcie sprzedaży jest zmienną losową o rozkładzie normalnym. Po obliczeniu średniej i wariancji z 15 wylosowanych punktów sprzedaży otrzymano $\bar{X} = 2.5$, $\hat{S}^2 = 1.8$. Na poziomie istotności $\alpha = 0.05$ zweryfikuj hipotezę, że wahania cen (mierzone wariancją) są równe 1, przeciwko hipotezie, że są większe od 1. Wskazówka. $\chi^2(0.95, 14) = 23.685$.

**Rozwiązanie:**

$$H_0 = \hat{\sigma}^2 = 1$$
$$H_1 = \hat{\sigma}^2 > 1$$
$$\chi^2(0.95, 14) = 23.685$$

$$W = [\chi^2(0.95, 14), \infty]$$
$$W = [23.685, \infty]$$

obliczam statystkę:

$$\chi^2 = \frac{n{S_n}^2}{{\sigma_0}^2} = \frac{15 * 1.8}{1} = 27$$
$$\chi^2 \in W$$

zatem odrzucam $H_0$ i przyjmuję hipotezę, że wariancje są większe od 1.

**Zadanie 7:**
Cena pewnego wyrobu jest różna w zależności od punktu sprzedaży. Przyjęto, że cena w wylosowanym punkcie sprzedaży jest zmienną losową o rozkładzie normalnym. Po obliczeniu średniej i wariancji z 15 wylosowanych punktów sprzedaży otrzymano $\overline{X} = 2.5$, $S\hat{}2 = 1.8$. Na poziomie istotności $\alpha = 0.05$ zweryfikuj hipotezę, że wahania cen (mierzone wariancją) są równe 1, przeciwko hipotezie, że są większe od 1. Wskazówka. $\chi^2(0.95, 14) = 23.685$.

**Rozwiązanie:**
X - zm. los. o rozkładzie normalnym
$H_0 : \sigma^2 = 1$
$H_1 : \sigma^2 > 1$
$\alpha = 0.05,\ S_n^2 = 1.8,\ n = 15$
Przyjmujemy prawdziwość $H_0, \sigma_0^2 = 1$
$chi^2 = \frac{nS_n^2}{1} = \frac{15 \times 1.8}{1} = 27$

Obszar krytyczny dla $H_1$: $W = (\chi^2(1-\alpha, n-1), \infty) = (\chi^2(0.95, 14), \infty) = (23.685, +\infty)$
$chi^2 \in W$, więc możemy przyjmując $H_0$ za prawdziwą.

## Zadanie 8:

Według teorii Profesora Genka, komórki macierzyste pewnego organizmu różnicują się na 5 typów dojrzałych komórek, z prawdopodobieństwami: $p_1 = 7/16, p_2 = 1/4, p_3 = p_4 = 1/8, p_5 = 1/16$. Przeprowadzono 496 niezależnych powtórzeń eksperymentu różnicowania i w 212 powtórzeniach powstała komórka typu 1, w 123 powstała komórka typu 2, w 62 typu 3, w 45 typu 4, oraz w 54 powtórzeniach powstały komórki typu 5. Testem $\chi^2$ na poziomie istotności $\alpha = 0.01$ zweryfikować hipotezę $H_0$, że teoria Genka dobrze opisuje zjawisko zderzeń.

### Rozwiązanie:

$p_1 = \frac{7}{16},\ p_2 = \frac{1}{4}, p_3 = p_4 = \frac{1}{8},\ p_1 = \frac{1}{16}$

$n = 496$

$n_1 = 212,\ n_2 = 123,\ n_3 = 62,\ n_4 = 45,\ n_5 = 54$

$\chi^2 = \sum_{i=1}^{5} \frac{(n_i - np_i)^2}{np_i} = \frac{25}{217} + \frac{1}{124} + \frac{0}{62} + \frac{289}{62} + \frac{529}{31} \approx 21,85$

$\chi^2(0,99;4) < \chi^2$ - wpada do zbioru krytycznego

Odp.: Odrzucam hipotezę - teoria Genka źle opisuje zjawisko zdarzeń.

## Zadanie 9:

10 robotnikom wprowadzono gimnastykę w trakcie pracy. Notowano wyniki pracy przed i w trakcie eksperymentu. Zarząd fabryki chciałby wiedzieć, czy wyniki pracy polepszyły się dzięki gimnastyce. Wyniki pomiarów wydajności pracy i-tego pracownika przed eksperymentem $x_{i1}$ oraz w trakcie eksperymentu $x_{i2}$ podane są w sztukach na godzinę w tabelce poniżej. Przyjmij hipotezę zerową $H_0\ : \mu_R = 0$ (wydajność pracy przed i po jednakowa) oraz hipotezę alternatywną $H_1\ \mu_R \neq 0$. Zweryfikuj hipotezę zerową na poziomie istotności $\alpha = 0.05$.

| $i$ | $x_{i1}$ | $x_{i2}$ | $r_i = x_{i1} - x_{i2}$ | $r_i^2$ |
|---|---|---|---|---|
| 1 | 28 | 32 | -4 | 16 |
| 2 | 27 | 30 | -3 | 9 |
| 3 | 24 | 24 | 0 | 0 |
| 4 | 27 | 28 | -1 | 1 |
| 5 | 26 | 28 | -2 | 4 |
| 6 | 22 | 24 | -2 | 4 |
| 7 | 30 | 29 | 1 | 1 |
| 8 | 26 | 24 | 2 | 4 |
| 9 | 25 | 27 | -2 | 4 |
| 10 | 26 | 29 | -3 | 9 |
| $\Sigma$ | | | -14 | 52 |

### Rozwiązanie:

$n = 10$

$H_0 : \mu_R = 0 = \mu_0$

$H_1 : \mu_R \neq 0$

$\alpha = 0,05$

Gdy $H_0$ prawdziwa, statystyka $T = \frac{\bar{R} - \mu_0}{S_R} \sqrt{n-1} \sim$ t-Studenta $t(n-1)$

$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i = \frac{1}{n}\sum_{i=1}^{n}(X_{i1} - X_{i2})$

$S_R^2 = \frac{1}{n}\sum_{i=1}^{n}(R_i - \bar{R})^2$

$\bar{R} = \frac{1}{10} \cdot (-14) = -1,4$

$S_R^2 = \frac{1}{10} \cdot \frac{162}{5} = \frac{81}{25} = 3,24$

$\sqrt{S_R^2} = \frac{9}{5} = 1,8$

$T = \frac{-1,4-0}{1,8} \cdot \sqrt{10-1} = -\frac{1,4}{1,8} \cdot 3 = -\frac{7}{3} = -2,334$

$W = (-\infty; -t(1-\frac{\alpha}{2}, n-1)] \cup [(1-\frac{\alpha}{2}, n-1); +\infty)$

$t_{kryt} = t_{0,975;9} = 2,26$

$W = (-\infty; -2,26] \cup [2,26; +\infty)$

$T \in W$

$H_0$ wpada w obszar krytyczny $\Rightarrow$ $H_0$ odrzucona.

**Zadanie 12:**
**Rozkład T-Studenta**

Niech $X_1, \ldots, X_n$ będzie próbą prostą z rozkładu normalnego $N(\mu, \sigma^2)$, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ będzie średnią empiryczną oraz $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ będzie estymatorem wariancji największej wiarygodności. Pokaż, że $\bar{X}$ i $S_n^2$ są niezależne oraz że zmienna $\frac{nS_n^2}{\sigma^2}$ ma rozkład $\chi_{n-1}^2$.

**Rozwiązanie:**

$X_1, \ldots, X_n$ - iid

$X_i \sim N(\mu, \sigma^2)$

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

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

a)

**Sposób #1**

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$\bar{X} - X_j = \frac{1}{n}(X_i + \ldots + X_{j-1} + X_{j+1} + \ldots + X_n) - \frac{n-1}{n}X_j \sim$

$\sim N\Big(\frac{(n-1)\mu}{n} - \frac{(n-1)\mu}{n}, \quad (n-1)\frac{\sigma^2}{n} + (n-1)^2\frac{\sigma^2}{n^2}\Big) = N\Big(0, \frac{n-1}{n}\sigma^2\Big)$

$Cov(X_j - \bar{X}, \bar{X}) = Cov(X_j, \bar{X}) - Cov(\bar{X}, \bar{X}) = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$

Wektor $(\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X})^T$ ma wielowymiarowy rozkład normalny z macierzą kowariancji:

$$Cov = \begin{bmatrix} \frac{\sigma^2}{n} & \mathbf{0}^T \\ \mathbf{0}^T & \alpha \end{bmatrix}$$

Przez wielowymiarowy rozkład normalny można wywnioskować, że $\bar{X}$ oraz $\boldsymbol{X} = (X_1 - \bar{X}, \ldots, X_n - \bar{X})^T$ są niezależnymi wektorami o rozkładzie normalnym oraz że $\bar{X}$ jest także niezależny od $\boldsymbol{X^T X} = \sum_{i=1}^{n}(X_i - \bar{X})^2 = nS_n^2$.

$\square$

**Sposób #2**

Udowodnić to można również opisowo. Średnia z próby $\bar{X}$ jest kompletną/zupełną i wystarczającą/dostateczną statystyką.

Kompletna/zupełna statystyka oznacza, że:

$E_\theta f(T) = 0$ dla każdego $\theta \Rightarrow f = 0$ z prawdopodobieństwem = 1

Wystarczająca/dostateczna statystyka oznacza, że:

Statystyka $T = T(X)$ jest wystarczająca dla $X$ (lub rodziny rozkładów $X$ lub $\theta$) jeśli jego warunkowy rozkład $X|T = t$ nie zależy od $\theta$ dla każdego $t$.

Wariancja z próby jest pomocniczą/poboczną statystyką. Oznacza to, że:

Statystyka $V = V(X)$ jest pomocniczą/poboczną statystyką gdy jej rozkład nie zależy od $\theta$.

Z twierdzenia Basu, jeśli:

(1) $X_1, \ldots, X_n$ - iid $\sim N(\mu, \sigma^2)$

(2) $\hat{\mu} = \frac{\sum_{i=1}^n X_i}{n}$ jest średnią z próby oraz jest kompletną i dostateczną statystyką.

(3) $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ jest wariancją z próby oraz jest pomocniczą statystyką - jej rozkład nie zależy od $\mu$.

to można wywnioskować, że te statystyki $(\bar{X}, S_n^2)$ są niezależne.

□

b)

Przypuśćmy, że jest funkcja $W$:

$W = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2$

Teraz można wziąć $W$ i dodać do wyrażenia 0. Takie postępowanie oczywiście nie zmienia wartości $W$:

$W = \sum_{i=1}^n \left(\frac{(X_i - \bar{X}) + (\bar{X} - \mu)}{\sigma}\right)^2$

Jak widać, 0 zostało dodane poprzez dodanie i odjęcie średniej z próbki w liczniku. Teraz można spierwiastkować wyrażenie:

$W = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 + \sum_{i=1}^n \left(\frac{\bar{X} - \mu}{\sigma}\right)^2 + 2\left(\frac{\bar{X} - \mu}{\sigma}\right)\sum_{i=1}^n (X_i - \bar{X})$

Ostatni wyraz:

$2\left(\frac{\bar{X} - \mu}{\sigma}\right)\sum_{i=1}^n (X_i - \bar{X}) = 0$, ponieważ $\sum_{i=1}^n (X_i - \bar{X}) = n\bar{X} - n\bar{X} = 0$

$W$ redukuje się do:

$W = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$

Teraz, gdyby wziąć wariancję z próbki i przemnożyć obie strony przez $n$:

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

$W$ można zapisać za pomocą funkcji wariancji z próbki:

$W = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{nS_n^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$

Lewa strona równania:

$\sum_{i=1}^{n} \left( \frac{X_i - \mu}{\sigma} \right)^2$

jest sumą $n$ niezależnych zmiennych losowych o rozkładzie $\chi_1^2$. Jest tak, ponieważ $X_1, \ldots, X_n$ są obserwacjami losowej próbki o wielkości $n$ z rozkładu normalnego $N(\mu, \sigma^2)$. W związku z tym:

$\frac{X_i - \mu}{\sigma}$

jest zgodny ze standardowym rozkładem normalnym. Gdyby spierwiastkować standardową normalną zmienną losową, wynikiem będzie zmienna losowa o rozkładzie $\chi_1^2$. Więc:

$\sum_{i=1}^{n} \left( \frac{X_i - \mu}{\sigma} \right)^2$

jest sumą $n$ niezależnych zmiennych losowych o rozkładzie $\chi_1^2$. Suma jest zmienną losową o rozkładzie $\chi_n^2$. Dlatego funkcja generująca moment $W$ $(mgf)$ jest taka sama jak funkcja generująca moment zmiennej losowej o rozkładzie $\chi_n^2$, a mianowicie:

$M_W(t) = (1 - 2t)^{-\frac{n}{2}}$ dla $t < \frac{1}{2}$

Teraz drugi wyraz $W$, po prawej stronie znaku równości, czyli:

$\frac{n(\bar{X} - \mu)^2}{\sigma^2}$

jest zmienną losową o rozkładzie $\chi_1^2$. Jest tak, ponieważ średnia z próbki $\sim N(\mu, \frac{\sigma^2}{n})$. W związku z tym:

$Z = \frac{\bar{X} - \mu}{\sigma^2 / \sqrt{n}} \sim N(0, 1)$

jest standardową normalną zmienną losową. Tak więc, gdyby podnieść $Z$ do potęgi, wynikiem będzie zmienna losowa o rozkładzie $\chi_1^2$:

$Z^2 = \frac{n(\bar{X} - \mu)^2}{\sigma^2} \sim \chi_1^2$

Dlatego funkcja generująca moment $Z^2$ to:

$M_{Z^2}(t) = (1 - 2t)^{-\frac{1}{2}}$ dla $t < \frac{1}{2}$

Teraz można zastosować unikatową właściwość funkcji generujących moment. Z definicji funkcja generująca moment $W$ to:

$$M_W(t) = E(e^{tW}) = E\left[ e^{t(\frac{nS_n^2}{\sigma^2} + Z^2)} \right] = E\left[ e^{t(\frac{nS_n^2}{\sigma^2})} \cdot e^{tZ^2} \right] = M_{\frac{nS_n^2}{\sigma^2}}(t) \cdot M_{Z^2}(t)$$

Ostatnia równość w powyższym równaniu pochodzi z niezależności między $\bar{X}$ i $S_n^2$. To znaczy, że jeśli są niezależne, to ich funkcje też są niezależne. Teraz używając funkcji generującej moment $W$ i $Z^2$, wynikiem jest:

$(1 - 2t)^{-\frac{n}{2}} = M_{nS_n^2/\sigma^2}(t) \cdot (1 - 2t)^{-\frac{1}{2}}$

Teraz rozwiązując funkcję generującą moment $nS_n^2/\sigma^2$ wynikiem jest:

$M_{\frac{nS_n^2}{\sigma^2}}(t) = (1 - 2t)^{-\frac{n}{2}} \cdot (1 - 2t)^{-\frac{1}{2}} = (1 - 2t)^{-\frac{(n-1)}{2}}$ dla $t < \frac{1}{2}$

Jest to funkcja generująca moment zmiennej losowej o rozkładzie $\chi_{n-1}^2$. Unikatowa właściwość funkcji generujących moment pozwala stwierdzić, że $\frac{nS_n^2}{\sigma^2}$ musi też być zmienną losową o rozkładzie $\chi_{n-1}^2$.

$\frac{nS_n^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$

□

**Zadanie 13:**

Wiadomo, że jeśli $X$ ma rozkład $N(0, 1)$ oraz $Z$ jest niezależne od $X$ o rozkładzie $\chi^2(k)$ to zmienna losowa $T = \frac{X}{\sqrt{Z/k}}$ ma rozkład t–studenta o k stopniach swobody. Korzystając z powyższego faktu pokaż że dla próby prostej $X1, ..., Xn$ z $N(\mu, \sigma^2)$

$\sqrt{n-1}\frac{\bar{X}-\mu}{S_n}$

ma rozkład t-studenta o n-1 stopniach swobody

**Rozwiązanie:** Wiemy, że:

$$X = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$Z = \frac{n{S_n}^2}{\sigma^2} \sim \chi^2_{n-1}$$

Zatem rozkład t-studenta o n-1 stopniach swobody jest równy:

$$\begin{aligned}
T &= \frac{X}{\sqrt{Z/(n-1)}} = \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\frac{n{S_n}^2}{\sigma^2}}{n-1}}} \\
&= \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \cdot \frac{\sigma\sqrt{n-1}}{S_n\sqrt{n}} \\
&= \frac{\bar{X}-\mu}{S_n}\sqrt{n-1}
\end{aligned}$$

zatem

$$\frac{\bar{X}-\mu}{S_n}\sqrt{n-1} \sim t_{n-1}$$

**Zadanie 14:**

Każda ze sprzedanych suszarek do włosów pewnego typu z nieznanym prawdopodobieństwem $\theta$ zostanie już pierwszego dnia po zakupie zareklamowana przez kupującego z powodu ukrytej wady. Przypuszcza się, że prawdopodobieństwo zgłoszenia reklamacji później, niż pierwszego dnia po zakupie jest takie samo. Zweryfikuj tą hipotezę za pomocą testu $\chi^2$ na poziomie istotności $\alpha = 0.05$ na podstawie danych dotyczących 1000 suszarek, z których 40 reklamowano pierwszego dnia, 60 reklamowano później, a pozostałe nie były reklamowane

**Rozwiązanie:**

$H_0 =$ P-stwo zgłoszenia reklamacji po 1 dniu równe $\theta$
$H_1 =$ P-stwo zgłoszenia reklamacji po 1 dniu nie równe $\theta$

Obliczam statystykę:

$$\begin{aligned}
\chi^2 &= \frac{(40-1000\theta)^2}{1000\theta} + \frac{(60-1000\theta)^2}{1000\theta} + \frac{(900-(1-2\theta)1000)^2}{(1-2\theta)1000} \\
&= \frac{1600-80000\theta+1000000\theta^2}{1000\theta} + \frac{3600-120000\theta+1000000\theta^2}{1000\theta} + \frac{(2000\theta-100)^2}{1000-2000\theta} \\
&= \frac{20000\theta^2-2000\theta+52}{10\theta} + \frac{4000\theta^2-400\theta+10}{1-2\theta} \\
&= \frac{12000\theta^2-1796\theta+52}{10\theta-20\theta^2}
\end{aligned}$$

Zbiór krytyczny na poziomie istotności $\alpha = 0.05$ jest równe

$$
\begin{aligned}
W &= [\chi^2(0.95, 1), \infty] \\
&= [3.84, \infty]
\end{aligned}
$$

$\theta$ może przyjmować wartości od 0 do 0.5

Po narysowaniu wykresu, otrzymuję taki wynik:

$\chi^2 \geq 3.84$ dla $\theta \in (0, 0.037] \cup [0.1141, 0.5)$ (w tym przypadku odrzucam $H_0$)
$\chi^2 < 3.8$ wpp. (w tym przypadku nie ma podstawy aby odrzucić $H_0$)