1. Коцкица се баца док не падне број 2, а највише 5 пута. Вршен је експеримент у 100 покушаја; постигнути резултати су дати табелом. Са нивоом значајности 0.05 испитати да ли је коцкица хомогена.

Број бацања	1	2	3	4	5
Број покушаја	17	14	11	9	49

Решење:

1) H_0 : „Коцкица је хомогена“
H_1 : „Коцкица није хомогена“

2) $$\tau = \sum_{k=1}^{r} \frac{\left(m_k - np_k\right)^2}{n \cdot p_k} : \ \chi^2_{r-1}$$

3) критична област

$$P\left\{\tau > \chi_0^2\right\} = 0.05$$

$$1 - P\left\{\tau < \chi_0^2\right\} = 0.05$$

$$P\left\{\tau < \chi_0^2\right\} = 0.95 \ \underset{\chi_4^2}{\overset{таб}{===>}} \ \chi_0^2 = 9.488$$

$$c \in \left(9.488; +\infty\right)$$

4) $\tau = ?$

Ради лакшег рачунања, користимо следећу табелу

X	1	2	3	4	5
m_k	17	14	11	9	49
p_k	0.167	0.139	0.116	0.096	0.482

Вредности p_k одређујемо, из дефиниције Геометријске расподеле, на следећи начин:

$$p_k = P\left(X = k\right) = q^{k-1} p, k = 1, 2, 3, 4, 5.$$

$$p_1 = \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right)^0 = \frac{1}{6} = 0.167$$

$$p_2 = \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right)^1 = \frac{5}{36} = 0.139$$

$$p_3 = \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right)^2 = \frac{25}{216} = 0.116$$

$$p_4 = \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right)^3 = \frac{125}{1296} = 0.096$$

$$p_5 = 1 - \sum_{k=1}^{4} p_k = 0.482$$

Након тога, рачунамо вредност статистике τ

$$\tau = \sum_{k=1}^{5} \frac{\left(m_k - np_k\right)^2}{n \cdot p_k} =$$

$$= \frac{\left(17 - 100 \cdot 0.167\right)^2}{100 \cdot 0.167} + \frac{\left(14 - 100 \cdot 0.139\right)^2}{100 \cdot 0.139} + \frac{\left(11 - 100 \cdot 0.116\right)^2}{100 \cdot 0.116} +$$

$$+ \frac{\left(9 - 100 \cdot 0.096\right)^2}{100 \cdot 0.096} + \frac{\left(49 - 100 \cdot 0.482\right)^2}{100 \cdot 0.482} = 0.07$$

5) $\tau \notin c \Rightarrow H_0$ прихватамо, тј. закључујемо да је коцкица хомогена.

2. Компанија FedEx је формирала комисију за испитивање проблема учесталих кашњења пошиљки. Добијени су следећи подаци: број кашњења је 163, просечно време кашњења је 34 мин, а стандардна девијација времена кашњења је 25 мин. Одредити 95% двострани интервал поверења за варијансу кашњења (претпоставља се да кашњење има Нормалну расподелу).

Решење:

$$X : N\left(m, \sigma^2\right),\ n = 163,\ \overline{x_n} = 34,\ S_n = 25$$

β=0.95

Користимо статистику τ:

$$\tau = \chi^2_{n-1} = \frac{nS_n^2}{\sigma^2} : \chi^2_{n-1}$$

Двострани интервал поверења добијамо на следећи начин:

$$P\left\{\chi_1^2 < \chi^2_{n-1} < \chi_2^2\right\} = \beta$$

$$P\left\{\chi_1^2 < \frac{n \cdot S_n^2}{\sigma^2} < \chi_2^2\right\} = \beta$$

$$P\left\{\frac{\chi_1^2}{n \cdot S_n^2} < \frac{1}{\sigma^2} < \frac{\chi_2^2}{n \cdot S_n^2}\right\} = \beta$$

$$P\left\{\frac{n \cdot S_n^2}{\chi_1^2} > \sigma^2 > \frac{n \cdot S_n^2}{\chi_2^2}\right\} = \beta$$

$$P\left\{\frac{n\cdot S_n^2}{\chi_2^2}<\sigma^2<\frac{n\cdot S_n^2}{\chi_1^2}\right\}=\beta$$

па је интервал поверења:

$$\left[\frac{n\cdot S_n^2}{\chi_2^2};\frac{n\cdot S_n^2}{\chi_1^2}\right]$$

Прво ћемо израчунати вредност χ_1^2

$$P\left\{\chi_{n-1}^2\le\chi_1^2\right\}=\frac{1-\beta}{2}$$

Увиђамо проблем да је број степени слободе код Хи-квадрат расподеле 162, те није могуће пронаћи вредност у таблици расподела. Стога користимо следеће правило:

За n >30, χ_n^2 расподела се апроксимира N(n,2n) расподелом. За расподелу статистике која се примењује у овом задатку, одговарајућа апроксимација је:

$$\chi_{n-1}^2\approx N(n-1;\ 2\cdot(n-1))$$

$$\frac{\chi_{n-1}^2-(n-1)}{\sqrt{2\cdot(n-1)}}:N(0,1)$$

Сада апроксимирану Хи-кварадрат расподелу стандардизујемо на $N(0,1)$ расподелу

$$P\left\{\frac{\chi_{n-1}^2-(n-1)}{\sqrt{2(n-1)}}<\frac{\chi_1^2-162}{\sqrt{2\cdot162}}\right\}=0.025$$

$$\varphi\left(\frac{\chi_1^2-162}{18}\right)=0.025$$

$$\varphi\left(-\frac{\chi_1^2-162}{18}\right)=0.975\underset{N}{\overset{таб}{\Longrightarrow}}-\frac{\chi_1^2-162}{18}=1.96$$

$$\chi_1^2-162=-35.28$$

$$\chi_1^2=126.72$$

Слично се добија:

$$\chi_2^2-162=35.28$$

$$\chi_2^2=197.28$$

Тражени двострани интервал поверења је:

$$\left[\frac{n\cdot S_n^2}{\chi_2^2};\frac{n\cdot S_n^2}{\chi_1^2}\right]=\left[\frac{163\cdot25^2}{197.28};\frac{163\cdot25^2}{126.72}\right]=\left[516.398;803.938\right]$$

3. У таблици су дате цена половних аутомобила *Застава 101* и њихова старост. Да ли се, са прагом значајности 0.05, може закључити да између година старости и цене постоји таква линеарна веза да ако је аутомобил старији опада му цена (за годину посматрања узети текућу годину)?

Цена у 1000 €	2.45	1.8	2	2	1.7	1.2	1.15	0.69	0.6	0.47
Годиште	2009	2008	2008	2007	2007	2006	2004	2003	2002	2000

Решење:

Проблем који се пред нас поставља је да утврдимо да ли постоји линеарна веза између старости аутомобила и цене. Треба обратити пажњу да су у Табели дата годишта посматраних аутомобила, а не њихова старост. Морамо прво да од године посматрања, тј. текуће 2010 године, одузмемо годиште аутомобила. Пошто је питање да ли постоји таква веза да ако је аутомобил старији опада му цена, можемо закључити да је старост аутомобила независна промењива X а цена зависна Y.

i	X	Y	XY	X^2	Y^2
1	1	2,45	2,45	1	6,002
2	2	1,8	3,6	4	3,24
3	2	2	4	4	4
4	3	2	6	9	4
5	3	1,7	5,1	9	2,89
6	4	1,2	4,8	16	1,44
7	6	1,15	6,9	36	1,322
8	7	0,69	4,83	49	0,476
9	8	0,6	4,8	64	0,36
10	10	0,47	4,7	100	0,221
$\sum$	46	14,06	47,18	292	23,952

$$\hat{Y} = \hat{\alpha} X + \hat{\beta}$$

$$\overline{X} = \frac{\sum X}{N} = \frac{46}{10} = 4.6$$

$$\overline{Y} = \frac{\sum Y}{N} = \frac{14.06}{10} = 1.406$$

$$S_x^2 = \frac{\sum X^2}{N} - \overline{X}^2 = \frac{292}{10} - 21.16 = 8.04$$

$$\hat{\alpha} = \frac{\frac{1}{n}\sum XY - \overline{X}\,\overline{Y}}{S_x^2} = \frac{\frac{47.18}{10} - 6.467}{8.04} = -0.217$$

$$\hat{\beta} = \overline{Y} - \hat{\alpha}\overline{X} = 2.407$$

$$\hat{Y} = -0.217X + 2.407$$

1) $H_0(\alpha = 0)$
$H_1(\alpha < 0)$

2) $\tau = \frac{\alpha - \alpha_0}{\hat{\sigma}}\sqrt{S_x^2(n-2)} : t_{n-2}$

3) критична област

$P\{\tau < t_0\} = \alpha$
$F(t_0) = 0.05$
$F(-t_0) = 0.95 \underset{t_8}{\overset{\text{таб}}{\Rightarrow}} t_0 = -1.9$
$c \in (-\infty; -1.9)$

4) $\tau = ?$

$$\hat{\sigma}^2 = \frac{1}{n}\left(\sum Y_i^2 - \hat{\alpha}\sum X_i \cdot Y_i - \hat{\beta}\sum Y_i\right) = 0.0376$$
$$\tau = \frac{\alpha - \alpha_0}{\hat{\sigma}}\sqrt{S_x^2(n-2)} = \frac{-0.217}{\sqrt{0.0376}}\sqrt{8.04 \cdot 8} = -8.997$$

5) $\tau \in c \Rightarrow H_0$ одбацујемо. Закључујемо да се што је аутомобил старији, то му опада цена.

4. Позната руска скакачица у даљ (троскок) Татјана Лебедева се припрема за наступајућу сезону. Након одрађеног тренинга на отвореном, где је постигла следеће даљине: 14,90 14,22 14,57 15,09 15,15 14,71 14,09 15,00 15,02 14,57 14,77 14,11; дошло је до погоршања времена и Татјана је принуђена да свој тренинг настави у дворани. Током серије од 12 скокова Татјана је постигла следеће даљине 14,40 14,42 14,77 15,19 15,01 14,31 14,29 15,06 15,09 14,17 14,27 14,51 респективно. Тренер је закључио да нема разлике између резултата на отвореном и оних постигнутих у дворани. Да ли је са ризиком грешке од 0.02 у праву? Претпоставља се да су узорци извучени из популације са Нормалном расподелом.

Решење:

Означимо са m_1 и m_2 просечне даљине које је Татјана постигла на отвореном и дворани. Треба да проверимо да ли постоји разлика између резултата постигнутих на отвореном и оних у дворани:

1) $H_0(m_1 = m_2)$
$H_1(m_1 \neq m_2)$

2) Користимо статистику:

$$\tau = \frac{\left(\overline{x_{n1}} - \overline{x_{n2}}\right) - \left(m_1 - m_2\right)}{\sqrt{n_1 S_{n1}^2 + n_2 S_{n2}^2}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}\left(n_1 + n_2 - 2\right)} \ : \ t_{n_1+n_2-2}$$

3) Одређујемо критичну област теста:

$$P\left\{|\tau| > t_0\right\} = 0.02$$
$$P\left\{|\tau| \le t_0\right\} = 0.98$$
$$2F\left(t_0\right) - 1 = 0.98$$
$$F\left(t_0\right) = 0.99 \ \underset{t_{22} \approx N(0,1)}{\overset{\text{таб}}{===}} \text{-------->} \quad t_0 = 2.33$$

Критична област теста је $c \in \left(-\infty; -2.33\right) \cup \left(2.33; +\infty\right)$

4) $\tau = ?$

Из података се добија

$$\overline{x}_{n1} = \frac{1}{n_1} \cdot \sum_i X_{1i} = 14.683$$
$$\overline{x}_{n2} = \frac{1}{n_2} \cdot \sum_i X_{2i} = 14.624$$
$$S_{n1}^2 = \frac{1}{n_1} \cdot \sum_i X_{1i}^2 - \overline{x}_1^2 = 0.131$$
$$S_{n2}^2 = \frac{1}{n_2} \cdot \sum_i X_{2i}^2 - \overline{x}_2^2 = 0.128$$

тако да статистика τ има вредност:

$$\tau = \frac{14.683 - 14,624}{\sqrt{12 \cdot (0.131 + 0.128)}} \cdot \sqrt{\frac{12 \cdot 12}{12 + 12}\left(12 + 12 - 2\right)}$$

$$\tau = 0.385$$

5) $\tau \notin c \Rightarrow$ прихватамо H_0 и закључујемо да нема разлике у постигнутим даљинама на отвореном и оних постигнутих у дворани (тренер је исправно закључио).

5. Вељко је у проблему, помагајте! Задатак који се пред њега поставља је да упореди рејтинг шахиста Русије и Србије. Како би избегао тешку физикалију у виду прикупљању података за све играче обе репрезентације, Вељко се одлучио да случајним избором испита 55 руских и 45 српских играча, и након тога интервалном оценом предвиди разлику математичких очекивања рејтинга ове две репрезентације. Утврђено је да је просечан рејтинг Руса 2630, док њихове колеге из Србије могу да се похвале са просечним рејтингом од 2550. Зна се да је стандардна девијација популације руских играча 120, а српских 100. Решите проблем ако је ниво поверења 0,90.

Решење:

$$\begin{matrix} n_1 = 55 \\ n_2 = 45 \end{matrix}, \begin{matrix} \overline{X_{n1}} = 2630 \\ \overline{X_{n2}} = 2550 \end{matrix}, \begin{matrix} \sigma_1 = 120 \\ \sigma_2 = 100 \end{matrix}, \ \beta=0.90$$

Пошто је познато σ_1, σ_2 и пошто су узорци довољно велики, важи:

$$\overline{X}_1 \approx N\left(m_1, \frac{\sigma_1^2}{n_1}\right); \overline{X}_2 \approx N\left(m_2, \frac{\sigma_2^2}{n_1}\right)$$

Користимо статистику

$$\hat{Z} = \overline{X}_1 - \overline{X}_2$$

која има приближно нормалну расподелу, тј.

$$\hat{Z} : N\left(E(\hat{Z}), \sigma^2(\hat{Z})\right)$$

где је

$$E(\hat{Z}) = m_1 - m_2$$

$$\sigma^2(\hat{Z}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

према томе је

$$\hat{Z} \approx N\left(m_1 - m_2 \ , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Како статистику $\hat{Z}$ стандардизујемо, добићемо статистику $\hat{Z}^*$ која има N(0, 1) расподелу, тј.

$$Z^* = \frac{\hat{Z} - E(\hat{Z})}{\sigma(\hat{Z})} \quad : \quad N(0,1)$$

$$P\left\{-z_0 \le \hat{Z}^* \le z_0\right\} = 0.90$$

$$2\Phi(z_0) - 1 = 0.90 \Rightarrow \Phi(z_0) = 0.95 \underset{N}{\overset{таб}{\Rightarrow}} z_0 = 1.64$$

$$P\left\{-z_0 \le \frac{\hat{Z} - E(\hat{Z})}{\sigma(\hat{Z})} \le z_0\right\} = 0.90$$

$$P\left\{-z_0 \le \frac{\hat{Z} - (m_1 - m_2)}{\sigma(\hat{Z})} \le z_0\right\} = 0.90$$

Пошто је $\sigma^2(\hat{Z}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{120^2}{55} + \frac{100^2}{45} = 484 \Rightarrow \sigma(\hat{Z}) = 22$

$\hat{Z} = \overline{X}_1 - \overline{X}_2 = 2630 - 2550 = 80$, следи

$$P\left(-1.64 \le \frac{80 - (m_1 - m_2)}{22} \le 1.64\right) = 0.90$$

$$P\left\{-1.64 \cdot 22 - 80 \le -(m_1 - m_2) \le 1.64 \cdot 22 - 80\right\} = 0.90$$

$$P\left\{-116.08 \le -(m_1 - m_2) \le -43.92\right\} = 0.90$$

$$P\left\{43.92 \le m_1 - m_2 \le 116.08\right\} = 0.90$$

Интервал је: $\left[43.92\,;116.08\right]$

6. Гранд Кафа је ангажовала тим стручњака који треба да утврди узроке учесталих грешака приликом паковања. Током истраживања, тим је прикупио следеће податке: број дефектних паковања је 129, просечна грешка приликом паковања је 5 гр, док је стандардна девијација 3гр. Одредити 90% двострани интервал поверења за варијансу грешке паковања.

Решење:

$X : N\left(m, \sigma^2\right),\ n = 129\,,\ \overline{x_n} = 5\,,\ S_n = 3$

β=0.90

Користимо статистику τ:

$$\tau = \chi^2_{n-1} = \frac{nS_n^2}{\sigma^2} : \chi^2_{n-1}$$

Двострани интервал поверења је :

$$\left[\frac{n \cdot S_n^2}{\chi_2^2}\,;\frac{n \cdot S_n^2}{\chi_1^2}\right]$$

Прво ћемо израчунати вредност χ_1^2

$$P\left\{\chi_{n-1}^2 \le \chi_1^2\right\} = \frac{1-\beta}{2}$$

За n >30, χ_n^2 расподела се апроксимира N(n,2n) расподелом, тј. у нашем случају

$$\chi_{128}^2 \approx N(128,\ 256)$$

Сада апроксимирану Хи-кварадрат расподелу стандардизујемо на $N(0,1)$ расподелу

$$P\left\{\frac{\chi_{n-1}^2-(n-1)}{\sqrt{2(n-1)}} < \frac{\chi_1^2-128}{\sqrt{2\cdot 128}}\right\} = 0.05$$

$$\varphi\left(\frac{\chi_1^2-128}{16}\right) = 0.05$$

$$\varphi\left(-\frac{\chi_1^2-128}{16}\right) = 0.95 \underset{N}{\overset{\text{таб}}{\Longrightarrow}} -\frac{\chi_1^2-128}{16} = 1.64$$

$$\chi_1^2 = 101.76$$

Слично се добија:

$$\chi_2^2 = 154.24$$

Тражени двострани интервал поверења је:

$$\left[\frac{n\cdot S_n^2}{\chi_2^2};\frac{n\cdot S_n^2}{\chi_1^2}\right] = \left[\frac{129\cdot 9}{154.24};\frac{129\cdot 9}{101.76}\right] = [7.527;11.409]$$

7. Полиција је узела податке о потрошњи цитостатика код појединих малигних болести. Са нивоом значајности 0.05 испитати да ли постоји статистички значајна разлика у количини употребљеног цитостатика код ове три врсте тумора.

Врста тумора	Количина цитостатика			
Нонхочкинсов лимфом	17	19	21	22
Остеосарком	12	9	14	13
Шваном	19	12	13	15

Решење:

1) $H_0 : m_1 = m_2 = m_3$
H_1 : "Бар две вредности математичких очекивања се међусобно разликују"

2) $\mathrm{F} = \frac{S_1^2}{S_2^2} : F_{(k-1),(n-k)}$

3) критична област

$P\{F > F_0\} = \alpha$

$c \in (F_0; +\infty)$

$P\{F > F_0\} = 0.05$

$P\{F < F_0\} = 0.95 \underset{F_{2;9}}{\overset{\text{таб}}{\Rightarrow}} F_0 = 4.26$

$\mathrm{c} \in (4.26; +\infty)$.

4) $\mathrm{F} = ?$

	Збир квадрата одступања	Број степени слободе	Средње квадратно одступање	F
Између група	$q_1=123.5$	k-1=2	$S_1^2 = \frac{q_1}{k-1} = 61.75$	$\mathrm{F} = \frac{S_1^2}{S_2^2} = \frac{61.75}{6.389} = 9.665$
Унутар група	$q_2=57.5$	n-k=9	$S_2^2 = \frac{q_2}{n-k} = 6.389$	
Укупан	q=181	n-1=11	$S^2 = \frac{q}{n-1}$	

5) $\mathrm{F} \in c \Rightarrow$ одбацујемо H_0, закључујемо да постоји статистички значајна разлика у количини употребљеног цитостатика код ове три врсте тумора

8. Венус Вилијамс је у резултатској кризи. Управо због тога је ангажовала новог тренера који је убрзо утврдио да је главни узрочник велики број неизнуђених грешака. Наиме, просечан број грешака које је Венус начинила у својим мечевима је 32. Извршена су мерења и добијени су резултати:

Број грешака	Број мечева
[0-20]	5
(20-28]	15
(28-36]	70
(36-64]	10

Да ли се, са прагом значајности од 0.01, може сматрати да број грешака има Биномну расподелу?

Решење:

1) H_0 : "X има $\mathrm{B}(n;p)$ "
H_1 : "X нема $\mathrm{B}(n;p)$ "

2) $$\tau=\sum_{i=1}^{r}\frac{\left(m_i-np_i\right)^2}{n\cdot p_i}:\ \ \chi^2_{r-1}$$

3) критична област

$$P\left\{\tau>\chi_0^2\right\}=0.01$$

$$1-P\left\{\tau<\chi_0^2\right\}=0.01$$

$$P\left\{\tau<\chi_0^2\right\}=0.99\ \underset{\chi_3^2}{\overset{\text{таб}}{===>}}\ \chi_0^2=11.341$$

$$c\in\left(11.341;+\infty\right)$$

4) $\tau=?$

Из података се види да је $n=64$. Код биномне расподеле просечна вредност је једнака

$$E\left(X\right)=n\cdot p$$

$$64\cdot p=32\Rightarrow p=0.5\Rightarrow q=0.5$$

Треба проверити хипотезу да X има $\mathrm{B}\left(64,0.5\right)$.

Вероватноће p_i ћемо тражити апроксимацијом. Пошто је $np=32>10$, апроксимацију ћемо извршити преко Нормалне расподеле

$$\mathrm{B}\left(n,p\right)\approx N(np,npq)$$

$$\mathrm{B}\left(64,0.5\right)\approx N(32,16)$$

$$p_1=P\left\{0<X<20\right\}=P\left\{\frac{0-32}{4}<X^*<\frac{20-32}{4}\right\}=P\left\{X<-3\right\}$$
$$p_1=1-\Phi(3)=1-0.9987=0.0013$$

Слично се добијају и остале вероватноће:

$$p_2 = P\{20 < X < 28\} = P\left\{\frac{20-32}{4} < X^* < \frac{28-32}{4}\right\} = \Phi(3) - \Phi(1) = 0.1574$$

$$p_3 = P\{28 < X < 36\} = P\left\{\frac{28-32}{4} < X^* < \frac{36-32}{4}\right\} = 2\Phi(1) - 1 = 0.6826$$

$$p_4 = P\{36 < X < 64\} = P\{1 < X^* < 8\} = 1 - 0.8413 = 0.1587$$

Ради лакшег рачунања, користимо следећу табелу

X	0-20	20-28	28-36	36-64
m_k	5	15	70	10
p_i	0.0013	0.1574	0.6828	0.1587
np_i	0.13	15.74	68.26	15.87

Након тога, рачунамо вредност статистике τ

$$\tau = \sum_{i=1}^{4} \frac{(m_i - np_i)^2}{n \cdot p_i}$$

$$\tau = \frac{(5-0.13)^2}{0.13} + \frac{(15-15.74)^2}{15.74} + \frac{(70-68.26)^2}{68.26} + \frac{(10-15.87)^2}{15.87}$$

$$\tau = 184.649$$

5) $\tau \in c \Rightarrow H_0$ одбацујемо, тј. закључујемо да X нема $\mathrm{B}(64, 0.5)$.

9. Позната руска скакачица у даљ (троскок) Татјана Лебедева се припрема за наступајућу сезону. Њен тренер приликом сваког скока проверава јачину ветра. Током тренинг серије од 10 скокова, Татјана је постигла следеће даљине: 14,90 14,22 14,57 15,09 15,15 14,71 14,09 15,00 15,02 14,57. Јачина ветра приликом сваког скока је била 2,1 1,7 1,5 1,9 2,4 0,4 0,9 1,6 1,1 1,9 респективно. Тренер је закључио да су резултати валидни, пошто ветар није имао утицаја на постигнуту даљину. Да ли је са ризиком грешке од 0.05 у праву?

Решење:

i	X	Y	XY	X^2	Y^2
1	2,1	14,9	31,29	4,41	222,01
2	1,7	14,22	24,174	2,89	202,2084
3	1,5	14,57	21,855	2,25	212,2849
4	1,9	15,09	28,671	3,61	227,7081
5	2,4	15,15	36,36	5,76	229,5225
6	0,4	14,71	5,884	0,16	216,3841
7	0,9	14,09	12,681	0,81	198,5281
8	1,6	15	24	2,56	225
9	1,1	15,02	16,522	1,21	225,6004
10	1,9	14,57	27,683	3,61	212,2849
$\sum$	15,5	147,32	229,12	27,27	2171,531

$$\hat{Y} = \hat{\alpha} X + \hat{\beta}$$

$$\overline{X} = \frac{\sum X}{N} = \frac{15.5}{10} = 1.55$$

$$\overline{Y} = \frac{\sum Y}{N} = \frac{147.32}{10} = 14.732$$

$$S_x^2 = \frac{\sum X^2}{N} - \overline{X}^2 = \frac{27.27}{10} - 2.4025 = 0.3245$$

$$\hat{\alpha} = \frac{\frac{1}{n}\sum XY - \overline{X}\overline{Y}}{S_x^2} = \frac{\frac{229.12}{10} - 22.838}{0.3245} = 0.238$$

$$\hat{\beta} = \overline{Y} - \hat{\alpha}\overline{X} = 14.362$$

$$\hat{Y} = 0.238X + 14.362$$

1) $H_0(\alpha = 0)$

$H_1(\alpha \neq 0)$

2) $\tau = \dfrac{\alpha - \alpha_0}{\hat{\sigma}}\sqrt{S_x^2(n-2)} : t_{n-2}$

3) критична област

$$P\{|\tau| \geq t_0\} = 0.05$$

$$P\{|\tau| \leq t_0\} = 0.95$$

$$2F(t_0) - 1 = 0.95$$

$$F(t_0) = 0.975 \underset{t_8}{\overset{таб}{\Rightarrow}} t_0 = 2.3$$

$$c \in (-\infty; -2.3) \cup (2.3; +\infty)$$

4) $\tau = ?$

$$\hat{\sigma}^2 = \frac{1}{n}\left(\sum Y_i^2 - \hat{\alpha}\sum X_i \cdot Y_i - \hat{\beta}\sum Y_i\right) = 0.1028$$

$$\tau = \frac{\alpha - \alpha_0}{\hat{\sigma}}\sqrt{S_x^2(n-2)} = \frac{0.238}{\sqrt{0.1028}}\sqrt{0.3245 \cdot 8} = 1.198$$

5) $\tau \notin c \Rightarrow H_0$ прихватамо. Закључујемо је тренер био у праву, пошто ветар није имао утицаја на постигнуту даљину.

10. На основу података Републичког завода за статистику, просечна зарада у Републици Србији, исплаћена у фебруару 2010. године, износи 44871 динара, са варијансом 343 000 динара. Уколико претпоставимо да је зарада нормално распоређена величина, израчунати вероватноћу да ће варијанса просечне зараде у Републици Србији за случајно одабраних 50 грађана у Републици Србији бити већа од 350 000 динара.

Решење:

$n = 50$, $m = 44871$, $\sigma^2 = 343000$

$P\{S_n^2 > 350000\} = ?$

$$\tau = \frac{n \cdot S_n^2}{\sigma^2} : \chi_{n-1}^2$$

$$P\{\frac{n \cdot S_n^2}{\sigma^2} > \frac{50 \cdot 350000}{343000}\} = P\{\tau > 51.02\} = 1 - P\{\tau < 51.02\} =$$

$$\tau : \chi_{49}^2 \approx N(49, 98)$$

$$= 1 - P\{\frac{\tau - 49}{\sqrt{98}} < \frac{51.02 - 49}{\sqrt{98}}\} = 1 - P\{\tau^* < 0.204\} = 1 - 0.5793 = 0.4207$$

11. Мара Циклама се припрема за летовање. У дилеми је коју туристичку агенцију да одабере па је питала своје колеге преко које агенције су прошле године летовали и која је била цена аранжмана. Са нивоом значајности 0.05 испитати да ли постоји статистички значајна разлика у цени између следећих туристичких агенција.

Агенција	Цена аранжмана		
Аргус	210	220	230
Контики	350	330	340
Сабра	290	270	310

Решење:

1) $H_0 : m_1 = m_2 = m_3$
H_1 : "Бар две вредности математичких очекивања се међусобно разликују"

2) $$F = \frac{S_1^2}{S_2^2} : F_{(k-1),(n-k)}$$

3) критична област

$$P\{F > F_0\} = \alpha$$

$c \in (F_0; +\infty)$

$P\{F > F_0\} = 0.05$

$P\{F < F_0\} = 0.95 \underset{F_{2;6}}{\overset{\text{таб}}{\Rightarrow}} F_0 = 5.14$

$c \in (5.14; +\infty)$.

4) F = ?

	Збир квадрата одступања	Број степени слободе	Средње квадратно одступање	F
Између група	q_1=21800	k-1=2	$S_1^2 = \frac{q_1}{k-1} = 10900$	$F = \frac{S_1^2}{S_2^2} = \frac{10900}{200} = 54.5$
Унутар група	q_2=1200	n-k=6	$S_2^2 = \frac{q_2}{n-k} = 200$	
Укупан	q=23000	n-1	$S^2 = \frac{q}{n-1}$	

5) $F \in c \Rightarrow$ одбацујемо H_0, закључујемо да постоји статистички значајна разлика у цени између туристичких агенција.

12. Вељко је у проблему, помагајте! Задатак који се пред њега поставља је да упореди рејтинг шахиста Русије и Србије. Како би избегао тешку физикалију у виду прикупљању података за све играче обе репрезентације, Вељко се одлучио да случајним избором испита 55 руских и 45 српских играча. Утврђено је да је просечан рејтинг Руса 2630, док њихове колеге из Србије могу да се похвале са просечним рејтингом од 2550. Зна се да је стандардна девијација популације руских играча 120, а српских 100. Вељко је "шац методом" закључио да не постоји статистички значајна разлика између ове две репрезентације (ниво значајности 0.01). Да ли је Вељко у праву?

Решење:

$n_1 = 55$, $\overline{X_{n1}} = 2630$, $\sigma_1 = 120$

$n_2 = 45$, $\overline{X_{n2}} = 2550$, $\sigma_2 = 100$, α=0.01

1) $H_0(m_1 = m_2)$

$H_0(m_1 \neq m_2)$

2) Користимо статистику:

$$\tau = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} : N(0,1)$$

3) Одређујемо критичну област c.

$$P\{|\tau| > z_0\} = 0.01$$

$$c \in (-\infty; -z_0) \cup (z_0; +\infty)$$

$$2\Phi(z_0) - 1 = 0.99$$

$$\Phi(z_0) = 0.995 \overset{таб}{\underset{N}{===>}} \ z_0 = 2.58$$

$$c \in (-\infty; -2.58) \cup (2.58; +\infty)$$

4) $\tau = ?$

$$\tau = \frac{2630 - 2550}{\sqrt{\frac{14400}{55} + \frac{10000}{45}}}$$

$$\tau = \frac{80}{\sqrt{261.81 + 222.22}}$$

$$\tau = \frac{80}{22} = 3.636$$

5) $\tau \in c \Rightarrow$ одбацујемо H_0, закључујемо да је Вељко погрешио, тј. постоји статистички значајна разлика између између ове две репрезентације.

13. Компанија ИНТЕРЕХ треба да прими велику пошиљку Милка чоколаде. Прописана тежина паковања је 300 гр, са стандардним одступањем 12. Компанија ће пошиљку вратити уколико је стандардно одступање пошиљке веће од 12. Испитан је узорак од 61 паковања и утврђено је да је стандардно одступање 13. Са нивоом значајности 0.05 испитати какву ће компанија донети одлуку? (Претпоставља се да је тежина артикла нормално распоређена величина)

Решење:

$m = 300, \sigma = 12$

$n = 61$

$S_n = 13$

1) $H_0 : "\sigma = 12"$

$H_1 : "\sigma > 12"$

2) $\tau = \frac{n \cdot S_n^2}{\sigma^2} : \chi^2_{n-1}$

3) критична област

$$P\{\tau > \chi_0^2\} = 0.05$$
$$P\{\tau < \chi_0^2\} = 0.95$$

$$\chi^2_{60} \approx N(60,120)$$

$$P\{\frac{\tau - 60}{\sqrt{120}} < \frac{\chi_0^2 - 60}{\sqrt{120}}\} = 0.95$$
$$P\{\tau^* < \frac{\chi_0^2 - 60}{\sqrt{120}}\} = 0.95$$
$$\frac{\chi_0^2 - 60}{\sqrt{120}} = 1.64$$
$$\chi_0^2 = 77.96$$

$$c \in (77.96; +\infty)$$

4) $\tau = ?$

$$\tau = \frac{61 \cdot 13^2}{12^2}$$
$$\tau = 71.59$$

5) $\tau \notin c \Rightarrow$ прихватамо H_0, компанија ће прихватити пошиљку.

14. Наша најбоља спортискиња свих времена Јасна Шекарић се припрема за наступајућу сезону. Због неадекватних услова за тренинг, принуђена је да своје припреме одржи на отвореном. Током тренинг серије од 12 метака, Јасна је постигла следеће резултате: 9,9 9,8 9,1 9,2 8,5 9,1 9,7 10 8,9 9,2 9,9 9,7. Након интензивне медијске кампање телевизије Б92, нашој прослављеној спортискињи је омогућено тренирање у стрељани. У новим условима, Јасна је постигла следеће резултате: 9,9 9,9 9,3 9,4 8,7 9,2 9,5 10 8,8 9,6 9,9 9,8 респективно. Тренер је закључио да нема разлике између резултата на отвореном и оних постигнутих у стрељани. Да ли је са ризиком грешке од 0.04 у праву? Претпоставља се да су узорци извучени из популације са Нормалном расподелом.

Решење:

1) $H_0(m_1 = m_2)$
$H_1(m_1 \neq m_2)$

2) Користимо статистику:

$$\tau = \frac{\left(\overline{x_{n1}} - \overline{x_{n2}}\right) - \left(m_1 - m_2\right)}{\sqrt{n_1 S_{n1}^2 + n_2 S_{n2}^2}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}\left(n_1 + n_2 - 2\right)} \ : \ t_{n_1+n_2-2}$$

3) Одређујемо критичну област теста:

$$P\left\{|\tau| > t_0\right\} = 0.04$$
$$P\left\{|\tau| \le t_0\right\} = 0.96$$
$$2F\left(t_0\right) - 1 = 0.96$$
$$F\left(t_0\right) = 0.98 \ \underset{t_{22} \approx N(0,1)}{\overset{\text{таб}}{===}} \text{-------->} \quad t_0 = 2.05$$

Критична област теста је $c \in \left(-\infty; -2.05\right) \cup \left(2.05; +\infty\right)$

4) $\tau = ?$

Из података се добија

$$\overline{x}_1 = \frac{1}{n_1} \cdot \sum_i X_{1i} = 9.416$$
$$\overline{x}_2 = \frac{1}{n_2} \cdot \sum_i X_{2i} = 9.5$$
$$S_1^2 = \frac{1}{n_1} \cdot \sum_i X_{1i}^2 - \overline{x}_1^2 = 0.209$$
$$S_2^2 = \frac{1}{n_2} \cdot \sum_i X_{2i}^2 - \overline{x}_2^2 = 0.175$$

тако да статистика τ има вредност:

$$\tau = \frac{9.416 - 9.5}{\sqrt{12 \cdot (0.209 + 0.175)}} \cdot \sqrt{\frac{12 \cdot 12}{12 + 12}\left(12 + 12 - 2\right)}$$

$$\tau = -0.446$$

5) $\tau \notin c \Rightarrow$ прихватамо H_0 и закључујемо да нема разлике у постигнутим резултатима на отвореном и оних постигнутих у дворани.

15. Светски призната фармацеутска кућа је изазвала револуцију у свету лепоте. Стручњаци ове компаније су створили 3 средства (Бљесни1, Бљесни2 и Бљесни3) која доводе до ефекта "напућене усне". Да би утврдили који од ових производа ће избацити на тржиште, на узорку од 12 мисица су тестирали ефекат производа (за колико мм се "напуће" усне). Сваки од производа је тестиран на по 4 девојке. Компанија сматра да сва три производа доводе до истог ефекта. Да ли су, са ризиком грешке од 0.01, донели исправан закључак?

Бљесни1	5	3	6	3
Бљесни2	4	2	3	4
Бљесни3	7	6	3	3

Решење:

1) $H_0 : m_1 = m_2 = m_3$
H_1 : "Бар две вредности математичких очекивања се међусобно разликују"

2) $F = \frac{S_1^2}{S_2^2} : F_{(k-1),(n-k)}$

3) критична област

$$P\{F > F_0\} = \alpha$$

$$c \in (F_0; +\infty)$$

$$P\{F > F_0\} = 0.01$$

$$P\{F < F_0\} = 0.99 \quad \underset{F_{2;9}}{\overset{\text{таб}}{\Rightarrow}} F_0 = 8.02$$

$$c \in (8.02; +\infty).$$

4) $F = ?$

	Збир квадрата одступања	Број степени слободе	Средње квадратно одступање	F
Између група	q_1=4.667	k-1=2	$S_1^2 = \frac{q_1}{k-1} = 2.333$	$F = \frac{S_1^2}{S_2^2} = \frac{2.333}{2.472} = 0.944$
Унутар група	q_2=22.25	n-k=9	$S_2^2 = \frac{q_2}{n-k} = 2.472$	
Укупан	q=26.917	n-1=11	$S^2 = \frac{q}{n-1}$	

5) $F \notin c \Rightarrow$ прихватамо H_0 , закључујемо да три производа доводе до истог ефекта.

16. Парис Хилтон је у проблему, помозите јој! Она покушава да утврди који бренд (Луј Витон или Хермес) прави квалитетније ташне. Сходно томе, испробала је 12 Луј Витон и 10 Хермес ташни. За сваку од ташни је мерила колико дана је кориштена пре него што је на било који начин оштећена.

Луј Витон (дана)	8,3	16,1	20	18,7	19	16,2	20,1	18,2	15,2	18,4	20	15,7
Хермес (дана)	14	13,3	16,6	19,4	22	16,3	14,5	11	17	14,1		

Парис је закључила да је квалитет ташни ова два бренда подједнак. Да ли је (са ризиком од 5%) у праву?

Решење:

1) H_0 : "Оба узорка су из исте популације"
H_1 : "Узорци нису из исте популације"

2)
$$\tau = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} : \mathrm{N}(0;1)$$

3) Критична област је левострана

$$P\{\tau < w_0\} = \alpha$$

$$\Phi(-w_0) = 1 - \alpha$$

$$c \in (-\infty; w_0)$$

$$\Phi(-w_0) = 0.95 \underset{N}{\overset{tab}{\Rightarrow}} -w_0 = 1.64 \Rightarrow w_0 = -1.64$$

$$c \in (-\infty; -1.64)$$

4) $\tau = ?$

Поређајмо по величини све вредности у оба узорка и рангирати их.

Ранг	Вредност	Узорак
1	8.3	1
2	11	2
3	13.3	2
4	14	2
5	14.1	2
6	14.5	2
7	15.2	1

8	15.7	1
9	16.1	1
10	16.2	1
11	16.3	2
12	16.6	2
13	17	2
14	18.2	1
15	18.4	1
16	18.7	1
17	19	1
18	19.4	2
19	20	1
20	20	1
21	20.1	1
22	22	2

$$R_1 = 1 + 7 + ... + 21 = 157$$

$$R_2 = 2 + 3 + ... + 22 = 96$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 12 \cdot 10 + \frac{12(12+1)}{2} - 157 = 198 - 157 = 41$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 12 \cdot 10 + \frac{10(10+1)}{2} - 96 = 175 - 96 = 79$$

$$U = \min(U_1, U_2) = \min(79, 41) = 41$$

$$\tau = \frac{41 - \frac{10 \cdot 12}{2}}{\sqrt{\frac{10 \cdot 12\left(10+12+1\right)}{12}}}$$

$$\tau = \frac{-19}{\sqrt{230}}$$

$$\tau = -1.253$$

5) τ ∉ c, прихватамо хипотезу H_0 и закључујемо да узорци потичу из исте популације. Дакле, Парис је исправно закључила да је квалитет ташни ова два бренда подједнак.

17. На основу истраживања *AGB Nielsen Media* Research, званичног мерача гледаности у Србији, регистровано је да је у периоду емитовања другог серијала емисије „Фарма”, 6566429 гледалаца погледало бар једно емитовање тог реалиту схоw-а, и то 6566,429 гледалаца дневно са варијансом 1000. Уколико претпоставимо да је гледаност нормално распоређена величина, израчунати вероватноћу да ће у току 51 дана емитовања трећег серијала реалиту схоw-а „Фарма”, варијанса бити мања од 80% варијансе гледаности првог серијала.

Решење:

$n=51, m=6566.429, \sigma^2 \;=1000$

$P\{S_n^2<800\}=?$

$\tau=\frac{n\cdot S_n^2}{\sigma^2}:\chi_{n-1}^2$

$P\{\frac{n\cdot S_n^2}{\sigma^2}<\frac{51\cdot 800}{1000}\}=P\{\tau<40.8\}=$

$\tau:\chi_{50}^2\approx N(50,100)$

$=P\{\frac{\tau-50}{\sqrt{100}}<\frac{40.8-50}{\sqrt{100}}\}=P\{\tau^*<-0.92\}=$
$=1-\Phi(0.92)=1-0.8218=0.1782$

18. Мара Циклама је у дилеми који Факултет да упише. Случајним избором интервјуисала је 65 студената ФОН-а и 55 студената Економског факултета. Анкетирани студенти су оценом од 5 до 10 оценили свој Факултет. Фоновци су свој Факултет оценили са 8,55 док су “Економисти“ свом Факултету дали оцену 8,1. Зна се да су стандардне девијације обе популације исте и износе 0,8. Мара је закључила да између ова два Факултета не постоји статистички значајна разлика. Да ли је са ризиком грешке од 0.05 у праву?

Решење:

$$\begin{matrix} n_1=65 \\ n_2=55 \end{matrix}, \begin{matrix} \overline{X_{n1}}=8.55 \\ \overline{X_{n2}}=8.1 \end{matrix}, \sigma_1=\sigma_2=\sigma=0.8, \beta=0.95$$

1) $H_0(m_1=m_2)$
$H_0(m_1\neq m_2)$

2) Користимо статистику:
$$\tau=\frac{\overline{x}_1-\overline{x}_2}{\sigma}\cdot\sqrt{\frac{n_1\cdot n_2}{n_1+n_2}}:N(0,1)$$

3) Одређујемо критичну област c.

$P\{|\tau| > z_0\} = 0.05$

$c \in (-\infty; -z_0) \cup (z_0; +\infty)$

$2\Phi(z_0) - 1 = 0.95$

$\Phi(z_0) = 0.975 \overset{\text{таб}}{\underset{N}{\Longrightarrow}} z_0 = 1.96$

$c \in (-\infty; -1.96) \cup (1.96; +\infty)$

4) $\tau = ?$

$$\tau = \frac{8.55 - 8.1}{0.8}\sqrt{\frac{65 \cdot 55}{65 + 55}}$$

$$\tau = 3.07$$

5) $\tau \in c \Rightarrow$ одбацујемо H_0, закључујемо да између ова два Факултета постоји статистички значајна разлика, тј. Мара није била у праву..

19. У последње време, Владимир Крамник пружа врло бледе партије у дуелима са врхунским велемајсторима. Као главни проблем се истиче велики број непрецизних потеза. Наиме, просечан број непрецизности је 20. Извршена су мерења и добијени су резултати:

Број непрецизности	[0-16]	(16-20]	(20-28]	(28-100]
Број партија	30	70	95	5

Да ли се ,са прагом значајности од 0.01, може сматрати да број непрецизности има Биномну расподелу?

Решење:

1) H_0 : "X има $B(n; p)$ "
H_1 : "X нема $B(n; p)$ "

2) $\tau = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{n \cdot p_i} : \chi^2_{r-1}$

3) критична област

$P\{\tau > \chi_0^2\} = 0.01$

$1-P\left\{\tau<\chi_0^2\right\}=0.01$

$P\left\{\tau<\chi_0^2\right\}=0.99 \quad \underset{\chi_3^2}{\overset{таб}{===>}} \quad \chi_0^2=11.341$

$c\in(11.341;+\infty)$

4) $\tau=?$

Из података се види да је $n=100$. Код биномне расподеле просечна вредност је једнака

$$E(X)=n\cdot p$$

$$100\cdot p=20\Rightarrow p=0.2\Rightarrow q=0.8$$

Треба проверити хипотезу да X има $\mathrm{B}(100,0.2)$.

Вероватноће p_i ћемо тражити апроксимацијом. Пошто је $np=20>10$, апроксимацију ћемо извршити преко Нормалне расподеле

$$\mathrm{B}(n,p)\approx N(np,npq)$$

$$\mathrm{B}(100,0.2)\approx N(20,16)$$

$$p_1=P\{0<X<16\}=P\left\{\frac{0-20}{4}<X^*<\frac{16-20}{4}\right\}=P\left\{X^*<-1\right\}$$
$$p_1=1-\Phi(1)=1-0.8413=0.1587$$

Слично се добијају и остале вероватноће:

$$p_2=P\{16<X<20\}=P\left\{\frac{16-20}{4}<X^*<\frac{20-20}{4}\right\}=\Phi(1)-\Phi(0)=0.3413$$
$$p_3=P\{20<X<28\}=P\left\{\frac{20-20}{4}<X^*<\frac{28-20}{4}\right\}=\Phi(2)-\Phi(0)=0.4772$$
$$p_4=P\{28<X<100\}=P\left\{\frac{28-20}{4}<X^*<\frac{100-20}{4}\right\}=1-\Phi(2)=1-0.9772=0.0228$$

Ради лакшег рачунања, користимо следећу табелу

X	0-16	16-20	20-28	28-100
m_k	30	70	95	5
p_i	0.1587	0.3413	0.4772	0.0228
np_i	31.74	68.26	95.44	4.56

Након тога, рачунамо вредност статистике τ

$$\tau = \sum_{i=1}^{4} \frac{\left(m_i - np_i\right)^2}{n \cdot p_i}$$

$$\tau = \frac{\left(30-31.73\right)^2}{31.74} + \frac{\left(70-68.26\right)^2}{68.26} + \frac{\left(95-95.44\right)^2}{95.44} + \frac{\left(5-4.56\right)^2}{4.56}$$

$$\tau = 0.18$$

5) $\tau \notin c \Rightarrow H_0$ прихватамо, тј. закључујемо да X има $\mathrm{B}\left(100, 0.2\right)$.

20. Нутрициониста покушава да утврди да ли између исхране Магнуса Карлсена и резултата које постиже постоји одређена веза. У Табели је представљено 10 дана у којима је израчуната калоријска снага унесене хране и перформанс (шаховска снага) који је Магнус постигао тог дана.

Перформанс (10^3)	2.712	2.837	2.755	2.625	2.802	2.811	2.748	2.737	2.748	2.899
Калорија (10^3)	1.8	2.24	2.3	2.1	2.09	2.4	2.02	2.3	2	2.48

Да ли са ризиком грешке од 5% може тврдити да се са порастом калорија побољшава перформанс који Магнус остварује?

Решење:

i	X	Y	XY	X^2	Y^2
1	1,8	2,712	4,8816	3,24	7,354944
2	2,24	2,837	6,35488	5,0176	8,048569
3	2,3	2,755	6,3365	5,29	7,590025
4	2,1	2,625	5,5125	4,41	6,890625
5	2,09	2,802	5,85618	4,3681	7,851204
6	2,4	2,811	6,7464	5,76	7,901721
7	2,02	2,748	5,55096	4,0804	7,551504
8	2,3	2,737	6,2951	5,29	7,491169
9	2	2,748	5,496	4	7,551504
10	2,48	2,899	7,18952	6,1504	8,404201
$\sum$	21,73	27,674	60,21964	47,6065	76,63547

$$\hat{Y} = \hat{\alpha} X + \hat{\beta}$$

$$\overline{X} = \frac{\sum X}{N} = \frac{21.73}{10} = 2.173$$

$$\overline{Y} = \frac{\sum Y}{N} = \frac{27.674}{10} = 2.7674$$

$$S_x^2=\frac{\sum X^2}{N}-\bar{X}^2=\frac{47.606}{10}-4.722=0.0387$$

$$\hat{\alpha}=\frac{\frac{1}{n}\sum XY-\bar{X}\bar{Y}}{S_x^2}=\frac{\frac{60.219}{10}-6.013}{0.0387}=0.217$$

$$\hat{\beta}=\bar{Y}-\hat{\alpha}\bar{X}=2.296$$

$$\hat{Y}=0.217X+2.296$$

1) $H_0(\alpha=0)$
$H_1(\alpha>0)$

2) $\tau=\dfrac{\alpha-\alpha_0}{\hat{\sigma}}\sqrt{S_x^2(n-2)}:t_{n-2}$

3) критична област

$$P\{\tau>t_0\}=0.05$$

$$P\{\tau<t_0\}=0.95$$

$$F(t_0)=0.95\overset{\text{таб}}{\underset{t_8}{\Rightarrow}}t_0=1.9$$

$$c\in(1.9;+\infty)$$

4) $\tau=?$

$$\hat{\sigma}^2=\frac{1}{n}\left(\sum Y_i^2-\hat{\alpha}\sum X_i\cdot Y_i-\hat{\beta}\sum Y_i\right)=0.0032$$

$$\tau=\frac{\alpha-\alpha_0}{\hat{\sigma}}\sqrt{S_x^2(n-2)}=\frac{0.217}{\sqrt{0.0032}}\sqrt{0.0387\cdot 8}=2.129$$

5) $\tau\in c\Rightarrow H_0$ одбацујемо. Закључујемо да се са порастом калорија побољшава перформанс који Магнус остварује.

21. Амерички институт за истраживање јавног мњења испитивао је однос републиканаца, демократа и либерала преам питању затварања "кампа" Гвантанамо. По четири испитаника из сваке политичке опције је попуњавало упитник, који је након тога одговарајуће бодован. Резултати тестова су дати на крају задатка. Утврдити да ли по питању Гвантанама постоји статистички значајна разлика између ове три политичке групације? (узети за праг значајности 0.01)

Републиканци: 55, 69, 75, 68
Демократе: 45, 39, 60, 56
Либерали: 25, 35, 44, 39

Решење:

1) $H_0 : m_1 = m_2 = m_3$
H_1 : "Бар две вредности математичких очекивања се међусобно разликују"

2) $\mathrm{F} = \dfrac{S_1^2}{S_2^2} : F_{(k-1),(n-k)}$

3) критична област
$P\{F > F_0\} = \alpha$

$c \in (F_0; +\infty)$

$P\{F > F_0\} = 0.01$

$P\{F < F_0\} = 0.99 \quad \underset{F_{2;9}}{\overset{\text{таб}}{\Rightarrow}} F_0 = 8.02$

$\mathrm{c} \in (8.02; +\infty)$.

4) $\mathrm{F} = ?$

	Збир квадрата одступања	Број степени слободе	Средње квадратно одступање	F
Између група	q_1=1926.167	k-1=2	$S_1^2 = \dfrac{q_1}{k-1} = 963.083$	$\mathrm{F} = \dfrac{S_1^2}{S_2^2} = \dfrac{963.083}{76.611} = 12.57$
Унутар група	q_2=689.5	n-k=9	$S_2^2 = \dfrac{q_2}{n-k} = 76.611$	
Укупан	q=2615.667	n-1=11	$S^2 = \dfrac{q}{n-1}$	

5) $F \notin c \Rightarrow$ прихватамо H_0 , закључујемо да три производа доводе до истог ефекта.

22. Италијански премијер, Силвио Берлускони, је пре одласка у Њујорк на Генералну Скупштину Уједињених Нација имао подршку 46% бирача. По повратка из САД, Берлускони је председника Барака Обаму назвао "потамнели згодни дечко". Следећег дана, од 200 испитаних Италијана, њих 88 је пружило подршку Берлусконију. Да ли са прагом значајности од 0,05 можемо тврдити да му је ова контроверзна изјава " срозала " рејтинг?

Решење:

$p_0 = 0.46$, $n = 200$, 88 подржава Берлусконија $\Rightarrow \hat{p} = 0.44$, $\alpha = 0.05$

1) $H_0\left(p = 0.46\right)$
$H_1\left(p < 0.46\right)$

2) Користимо статистику

$$\tau = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \ : \ \mathrm{N}(0,1)$$

3) Критичну област c одређујемо на следећи начин:

$$P\{\tau < z_1\} = \alpha$$

па је критична област теста $c \in (-\infty; z_1)$

$$P\{\tau < z_1\} = 0.05$$

$$\Phi(-z_1) = 0.95 \ \overset{\text{таб}}{\underset{N}{==>}} \ z_1 = -1.64$$

$$c \in (-\infty; -1.64)$$

4) Рачунамо вредност статистике τ:

$$\tau = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$$

$$\tau = \frac{0.44 - 0.46}{\sqrt{0.46 \cdot 0.54}}\sqrt{200}$$

$$\tau = -0.5676$$

5) $\tau \notin c \Rightarrow H_0$ прихватамо, тј. закључујемо да није дошло до "срозавања" рејтинга.

23. Дати су подаци о 12 шаховских партија Гарија Каспарова у којима је утврђен број одиграних потеза и број грешака које је Гари начинио.

Број потеза	56	42	72	36	63	47	55	49	38	42	68	60
Број грешака	1	2	1	1	2	4	2	2	1	1	1	3

Са нивоом значајности од 0.01 тестирати хипотезу да су број потеза и грешака независне случајне промењиве (претпоставља се да су подаци узети из популације са Нормалном расподелом).

Решење:

1) $H_0\left(\rho=0\right)$

$H_1\left(\rho\neq 0\right)$

2) $\tau=\dfrac{r}{\sqrt{1-r^2}}\sqrt{n-2}:t_{n-2}$

3) Одређујемо критичну област

$P\left\{|\tau|>t_0\right\}=0.01$

$P\left(|\tau|<t_0\right)=0.99$

$2F(t_0)-1=0.99$

$F(t_0)=0.995\underset{t_8}{\overset{tab}{\Rightarrow}}t_0=3.3$

$c\in\left(-\infty;-3.3\right)\cup\left(3.3;+\infty\right).$

4) $\tau=?$

Из узорка се добија да је

$\sum X_i=628 \quad \sum X_i^2=34416$

$\sum Y_i=21 \quad \sum X_iY_i=1098 \quad \sum Y_i^2=47$

Узорачки коефицијент је:

$$r=\frac{\frac{1}{12}\cdot 1098-\frac{1}{12}\cdot 628\cdot\frac{1}{12}\cdot 21}{\sqrt{\frac{1}{12}\cdot 34416-52.333^2}\cdot\sqrt{\frac{1}{12}\cdot 47-1.75^2}}$$

$r=-0.0079$

Вредност τ статистике износи:

$$\tau=\frac{-0.0079}{\sqrt{1-(0.0079)^2}}\sqrt{10-2}$$

$\tau=-0.0632$

5) $\tau\notin c\Rightarrow H_0$ прихватамо, односно број потеза и грешака су независне случајне промењиве.

24. Једна од најмоћнијих српских компанија жели да се рекламира током трајања тениских мечева Јелене Јанковић или Ане Ивановић. Наравно, желе да то буде тенисерка која привлачи више гледалаца. Да би утврдили број гледалаца који прати њихове мечеве, на случајан начин је изабрано 12 мечева Јелене Јанковић и 10 мечева Ане Ивановић.

Јелена (мил. гледалаца)	0,93	1,71	2,1	1,97	2	1,72	2,11	1,92	1,62	1,94	2,1	1,67
Ана (мил. гледалаца)	1,5	1,43	1,76	2,04	2,3	1,73	1,55	1,2	1,8	1,5		

Компанија је закључила да не постоји статистички значајна ралика између наше две најбоље тенисерке. Да ли су (са ризиком од 5%) у праву?

Решење:

1) H_0 : "Оба узорка су из исте популације"
H_1 : "Узорци нису из исте популације"

2)
$$\tau = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 \left(n_1 + n_2 + 1\right)}{12}}} : \mathrm{N}\left(0;1\right)$$

3) Критична област је левострана

$P\left\{\tau < w_0\right\} = \alpha$

$\Phi\left(-w_0\right) = 1 - \alpha$

$c \in \left(-\infty; w_0\right)$

$$\Phi\left(-w_0\right) = 0.95 \underset{N}{\overset{tab}{\Rightarrow}} -w_0 = 1.64 \Rightarrow w_0 = -1.64$$

$c \in \left(-\infty; -1.64\right)$

4) $\tau = ?$

Поређајмо по величини све вредности у оба узорка и рангирати их.

Ранг	Вредност	Узорак
1	0.93	1
2	1.2	2
3	1.43	2
4	1.5	2
5	1.5	2
6	1.55	2
7	1.62	1
8	1.67	1

9	1.71	1
10	1.72	1
11	1.73	2
12	1.76	2
13	1.8	2
14	1.92	1
15	1.94	1
16	1.97	1
17	2	1
18	2.04	2
19	2.1	1
20	2.1	1
21	2.11	1
22	2.3	2

$R_1 = 1 + 7 + ... + 21 = 157$

$R_2 = 2 + 3 + ... + 22 = 96$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 12 \cdot 10 + \frac{12(12+1)}{2} - 157 = 198 - 157 = 41$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 12 \cdot 10 + \frac{10(10+1)}{2} - 96 = 175 - 96 = 79$$

$$U = \min(U_1, U_2) = \min(79, 41) = 41$$

$$\tau = \frac{41 - \frac{10 \cdot 12}{2}}{\sqrt{\frac{10 \cdot 12 (10 + 12 + 1)}{12}}}$$

$$\tau = \frac{-19}{\sqrt{230}}$$

$$\tau = -1.253$$

5) $\tau \notin c$, прихватамо хипотезу H_0 и закључујемо да узорци потичу из исте популације. Дакле, компанија је исправно закључила да не постоји статистички значајна ралика између наше две најбоље тенисерке.

25. Мара Циклама је у дилеми који Факултет да упише. Случајним избором интервјуисала је 65 студената ФОН-а и 55 студената Економског факултета. Анкетирани студенти су оценом од 5 до 10 оценили свој Факултет. Фоновци су свој Факултет оценили са 8,55 док су “Економисти“ свом Факултету дали оцену 8,1. Зна се да су стандардне девијације обе популације исте и износе 0,8. Интервалном оценом предвиди разлику математичких очекивања оцене Факултета (ниво поверења 0,90).

Решење:

$$n_1 = 55 \quad \overline{X_{n1}} = 8.55$$
$$n_2 = 45 , \overline{X_{n2}} = 8.1 \quad , \sigma_1 = \sigma_2 = \sigma = 0.8 \text{ , } \beta = 0.90$$

Користимо статистику:

$$\tau = \frac{\overline{x}_1 - \overline{x}_2 - (m_1 - m_2)}{\sigma} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} : N(0,1)$$

$$P\{-z_0 \le \tau \le z_0\} = 0.90$$

$$2\Phi(z_0) - 1 = 0.90 \Rightarrow \Phi(z_0) = 0.95 \underset{N}{\overset{таб}{\Rightarrow}} z_0 = 1.64$$

$$P\left\{-z_0 \le \frac{(\overline{x_{n1}} - \overline{x_{n2}}) - (m_1 - m_2)}{\sigma} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \le z_0\right\} = 0.90$$

$$P\left(-1.64 \le \frac{0.45 - (m_1 - m_2)}{0.8} \cdot 5.458 \le 1.64\right) = 0.90$$

$$P\{-0.24 \le 0.45 - (m_1 - m_2) \le 0.24\} = 0.90$$
$$P\{-0.69 \le -(m_1 - m_2) \le -0.21\} = 0.90$$
$$P\{0.21 \le m_1 - m_2 \le 0.69\} = 0.90$$

Интервал је $: [0.21 ; 0.69]$