Instrukcja do wykonania zadania

W pierwszej kolejności należy przygotować tabelę z danymi. W niniejszej instrukcji przyjęto, że do każdego wyniku z tabeli pierwotnej dodano wartość **6**. Zatem tabela wygląda następująco:

Inicjały	Grupa	Płeć	Wiek	Masa ciała	Wys. Ciała	Czas przy tv i komp. [godz/doba]	Skolioza_ Przed	Skolioza_ Po
AA	A	K	38	62	161	14	18	19
BB	A	K	47	65	167	15	19	19
CC	A	K	46	71	166	18	20	18
DD	A	K	41	61	165	14	22	19
EE	A	K	42	76	168	16	18	15
FF	A	M	47	72	172	15	17	14
GG	A	M	48	74	176	16	20	18
HH	A	M	43	81	179	17	18	17
II	A	M	45	86	176	18	20	14
JJ	A	M	44	82	175	15	22	16
KK	B	K	49	64	166	13	23	22
LL	B	K	46	66	168	15	22	21
MM	B	K	45	71	172	14	20	20
NN	B	K	42	68	170	13	24	20
OO	B	K	41	72	168	15	24	21
PP	B	M	39	74	180	13	22	18
QQ	B	M	41	76	183	12	23	19
RR	B	M	47	76	178	14	23	19
SS	B	M	46	71	177	13	22	20
TT	B	M	41	74	179	13	23	18

Następnie dla każdej zmiennej (WIEK, MASA CIAŁA, WYSOKOŚĆ CIAŁA, CZAS PRZY TV I KOMP., SKOLIOZA_PRZED ORAZ SKOLIOZA_PO) obliczamy miary położenia (średnia, mediana, modalna), dyspersji (współczynnik zmienności, odchylenie standardowe), asymetrii (skośność) i koncentracji (kurtoza) oddzielnie dla grupy A oraz dla grupy B.

W tym celu korzystamy z następujących wzorów:

średnia arytmetyczna :

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

mediana (dla parzystej liczby obserwacji):

$$Me = \frac{(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})}{2}$$

modalna (wskazujemy tę wartość, która występuje najczęściej, jeśli jest kilka takich wartości, wówczas orzekamy, że rozkład jest wielomodalny, co jest równoznaczne, że niemożna wskazać modalnej.

współczynnik zmienności: $$V = \frac{s}{\bar{x}} \cdot 100\%$$

odchylenie standardowe: $$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \text{ lub } \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

współczynnik skośności: $$A_s = \frac{m_3}{s^3} \qquad m_3 = \frac{\Sigma(x_i - \bar{x})^3}{n}$$

kurtoza: $$K = \frac{m_4}{s^4} \qquad m_4 = \frac{\Sigma(x_i - \bar{x})^4}{n}$$

Obliczenia:

Żeby wyliczyć **średnią arytmetyczną ($\bar{X}$)** dla wieku w grupie A należy w następujący sposób podstawić dane do wzoru:

$$\bar{x} = \frac{38 + 47 + 46 + 41 + 42 + 47 + 48 + 43 + 45 + 44}{10} = \frac{441}{10} = 44{,}1$$

W celu wyliczenia **mediany** (**Me**) dla **wieku w grupie A** należy uporządkować dane w szyku rosnącym, a następnie podstawić je do wzoru:

$$Me = \frac{(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})}{2}$$

38;41;42;43;**44**;**45**;46;47;47;48

x_1; x_2; x_3; x_4; **x_5; x_6**; x_7; x_8; x_9; x_{10}

wyrażenie z tego wzoru $x_{\frac{n}{2}}$ oznacza, że należy „n” czyli liczbę obserwacji, która wynosi 10 podzielić przez 2; daje nam to wynik równy 5, co oznacza, że piątą z kolei wartość w uszeregowanym rosnąco szyku uwzględnimy przy wyliczaniu mediany; drugie wyrażenie w tym wzorze $x_{\frac{n}{2}+1}$ oznacza, że do wcześniej wyliczonej wartości 5 należy dodać jeszcze 1, w związku z czym do obliczenia mediany oprócz piątej wartości weźmiemy także wartość 6 (zaznaczone na czerwono). Zatem podstawiając do wzoru, otrzymujemy:

$$Me = \frac{x_5 + x_6}{2} = \frac{44 + 45}{2} = 44{,}5$$

Modalna

38;41;42;43;44;45;46;**47;47**;48

Jak widać w powyżej uszeregowanych rosnąco danych, najczęstszą wartości jest wartość 47, która pojawia się dwukrotnie. Oznacza to, że modalna w tym szeregu wynosi **47**.

Żeby obliczyć **odchylenie standardowe**, najlepiej jest posłużyć się tabelą pomocniczą, w której prowadzi się odrębne obliczenia. Tabela ta, będzie także potrzebna, przy wyliczaniu pozostałych parametrów, takich jak współczynnik skośności oraz kurtozy.

We wzorze na **odchylenie standardowe** (oznaczanym symbolem „s" lub „SD") mamy następujące wyrażenie: $(x_i-\bar{x})^2$ co oznacza, że od każdej wartości składającej się na zmienną wiek, należy odjąć wcześniej wyliczoną średnią arytmetyczną. Wyniki zapisujemy w poniższej tabeli.

Wiek	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
38	38-44,1=-6,1	$(-6,1)^2=37,21$
47	41-44,1=-3,1	$(-3,1)^2=9,61$
46	42-44,1=-2,1	$(-2,1)^2=4,41$
41	43-44,1=-1,1	$(-1,1)^2=1,21$
42	44-44,1=-0,1	$(-0,1)^2=0,01$
47	45-44,1=0,9	$(0,9)^2=0,81$
48	46-44,1=1,9	$(1,9)^2=3,61$
43	47-44,1=2,9	$(2,9)^2=8,41$
45	47-44,1=2,9	$(2,9)^2=8,41$
44	48-44,1=3,9	$(3,9)^2=15,21$
SUMA		$\sum_{i=1}^{n}(x_i-\bar{x})^2$ **88,9**

Ponadto, wyżej wspomniane we wzorze na odchylenie standardowe wyrażenie $(x_i-\bar{x})^2$ poprzedzone jest symbolem Σ (sigma), który oznacza sumę tych wszystkich różnic poszczególnych wartości składających się na zmienną wiek od średniej arytmetycznej wieku. Zatem, żeby obliczyć tę część wzoru $\sum_{i=1}^{n}(x_i-\bar{x})^2$ należy zsumować wszystkie wiersze z 3 kolumny tabeli po lewej (88,9).

Tak więc pozostaje tylko podstawić do wzoru:

$$s=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}=\sqrt{\frac{1}{10}*88,9}=\sqrt{\frac{88,9}{10}}=\sqrt{8,89}=2,98$$

Znając wartość średniej arytmetycznej oraz odchylenia standardowego, można teraz wyliczyć wartość współczynnika zmienności „V", wg wzoru:

$$V=\frac{s}{\bar{x}}\cdot 100\%=\frac{2,98}{44,1}\cdot 100=0,07\cdot 100=6,76$$

W celu obliczenia **współczynnika asymetrii** na początek przeanalizujmy poniższy wzór:

$$A_s = \frac{m_3}{s^3} \quad \text{gdzie:} \quad m_3 = \frac{\Sigma(x_i - \overline{x})^3}{n}$$

widzimy, że w liczniku tego wzoru jest to samo wyrażenie, które jest we wzorze na odchylenie standardowe $\sum_{i=1}^{n}(x_i - \overline{x})^3$, z tą jednak różnicą, że obliczoną wartość wynikającą z odejmowania **(x_i-x̄)** podnosimy do 3 potęgi. Z kolei w mianowniku jest odchylenie standardowe podniesione do 3 potęgi. Dla uproszczenia obliczeń wykorzystujemy tabelę pomocniczą (poniżej).

Wiek	**x_i-x̄**	**(x_i-x̄)^2**	**(x_i-x̄)^3**
38	38-44,1=-6,1	$(-6,1)^2$=37,21	$(-6,1)^3$=-226,981
47	41-44,1=-3,1	$(-3,1)^2$=9,61	$(-3,1)^3$=24,389
46	42-44,1=-2,1	$(-2,1)^2$=4,41	$(-2,1)^3$=6,859
41	43-44,1=-1,1	$(-1,1)^2$=1,21	$(-1,1)^3$=-29,791
42	44-44,1=-0,1	$(-0,1)^2$=0,01	$(-0,1)^3$=-9,261
47	45-44,1=0,9	$(0,9)^2$=0,81	$(0,9)^3$=24,389
48	46-44,1=1,9	$(1,9)^2$=3,61	$(1,9)^3$=59,319
43	47-44,1=2,9	$(2,9)^2$=8,41	$(2,9)^3$=-1,331
45	47-44,1=2,9	$(2,9)^2$=8,41	$(2,9)^3$=0,729
44	48-44,1=3,9	$(3,9)^2$=15,21	$(3,9)^3$=-0,001
SUMA		88,9	$\sum_{i=1}^{n}(x_i - \overline{x})^3$ **-151,68**

Zatem podstawiając do wzoru:

$$A_s = \frac{m_3}{s^3} = \frac{\frac{\Sigma(x_i - \overline{x})^3}{n}}{s^3} = \frac{\frac{-151,68}{10}}{2,98^3} = \frac{-15,17}{26,46} = -0,57$$

Jeśli chodzi o wyliczenie kurtozy, to postępujemy analogicznie jak w przypadku współczynnika asymetrii, bowiem wzór na kurtozę różni się tylko wykładnikiem potęgi, do której podnosimy wyrażenie w liczniku jak i odchylenie standardowe znajdujące się w mianowniku. Tabela pomocnicza wygląda następująco:

Wiek	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^3$	$(x_i-\bar{x})^4$
38	38-44,1=-6,1	$(-6,1)^2=37,21$	$(-6,1)^3=-226,981$	1384,584
47	41-44,1=-3,1	$(-3,1)^2=9,61$	$(-3,1)^3=24,389$	70,7281
46	42-44,1=-2,1	$(-2,1)^2=4,41$	$(-2,1)^3=6,859$	13,0321
41	43-44,1=-1,1	$(-1,1)^2=1,21$	$(-1,1)^3=-29,791$	92,3521
42	44-44,1=-0,1	$(-0,1)^2=0,01$	$(-0,1)^3=-9,261$	19,4481
47	45-44,1=0,9	$(0,9)^2=0,81$	$(0,9)^3=24,389$	70,7281
48	46-44,1=1,9	$(1,9)^2=3,61$	$(1,9)^3=59,319$	231,3441
43	47-44,1=2,9	$(2,9)^2=8,41$	$(2,9)^3=-1,331$	1,4641
45	47-44,1=2,9	$(2,9)^2=8,41$	$(2,9)^3=0,729$	0,6561
44	48-44,1=3,9	$(3,9)^2=15,21$	$(3,9)^3=-0,001$	0,0001
SUMA		88,9	-151,68	$\Sigma(x_i-\bar{x})^4$ **1884,34**

$$Ku = \frac{m_4}{s^4} = \frac{\frac{\Sigma(x_i-\bar{x})^4}{n}}{s^4} = \frac{\frac{1884,34}{10}}{2,98^4} = \frac{184,43}{78,86} = 2,34$$

Testy istotności różnic

Kolejnym krokiem analizy jest sprawdzenie czy grupy A i B są jednorodne pod względem wieku oraz parametrów somatycznych (wysokości i masy ciała). **Uwaga, na zaliczenie porównujemy tylko masę ciała!** Żeby to sprawdzić należy za pomocą testu **t-Studenta** (wtedy, kiedy n<30) lub **testu „z"** (wtedy, kiedy n>30) dla danych niezależnych przeprowadzić odpowiednie porównania. Przed przystąpieniem do porównań formułujemy hipotezy: zerową oraz alternatywną:

H_0 ---Osoby z grupy A nie różnią się pod względem wieku od osób z grupy B

H_1 ---Wiek osób z grupy A różni się od wieku osób z grupy B

Ponieważ mamy 10 obserwacji w grupie A oraz 10 w grupie B, zatem n<30, obliczenia będą prowadzone w oparciu o następujący wzór:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

gdzie:

x_1, x_2– średnia wieku w grupie A (x_1) i w grupie B (x_2)

n_1, n_2 – ilość obserwacji w grupie A (n_1) i w grupie B (n_2)

$S(x)_1$, $S(x)_2$ – wariancje w grupie A $S(x)_1$ i w grupie B $S(x)_2$

Wartość średniej wieku w grupie A wynosi 44,1 natomiast w grupie B 43,7. Wariancji co prawda nie mamy jeszcze obliczonej, ale wiedząc, że odchylenie standardowe jest pierwiastkiem kwadratowym z wariancji, wystarczy podnieść odchylenie standardowe do kwadratu (drugiej potęgi), zatem wariancja w grupie A wynosi 8,88 ($2{,}98^2$), natomiast w grupie B równa się 11,09 ($3{,}33^2$). Podstawiając do wzoru otrzymujemy:

$$t=\frac{\overline{x}_1-\overline{x}_2}{\sqrt{\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{44{,}1-43{,}7}{\sqrt{\frac{10\cdot 8{,}88+10\cdot 11{,}09}{10+10-2}\cdot\left(\frac{1}{10}+\frac{1}{10}\right)}}=\frac{0{,}4}{\sqrt{\frac{199{,}7}{18}\cdot\left(\frac{2}{10}\right)}}$$

$$=\frac{0{,}4}{\sqrt{11{,}09\cdot 0{,}2}}=\frac{0{,}4}{\sqrt{2{,}22}}=\frac{0{,}4}{1{,}49}=0{,}27$$

Aby zweryfikować postawioną hipotezę zerową (tzn. podjąć decyzję o jej podtrzymaniu lub odrzuceniu) należy odszukać wartość krytyczną w tablicy zawierającej rozkład testu „t” (poniżej). Wartość tę odszukujemy uwzględniając zakładany poziom istotności α=0,05 oraz liczbę stopni swobody $n_1+n_2-2=10+10-2=18$.

stopnie swobody	**Alfa α**		
	0,01	**0,05**	**0,10**
16	2,9208	2,1199	1,7459
17	2,8982	2,1098	1,7396
18	2,8784	2,1009	1,7341
19	2,8609	2,0930	1,7291
20	2,8453	2,0860	1,7247

Odczytana wartość krytyczna z tablicy wynosi $t_{0,05}$=2,1.

Ponieważ wartość testu t-studenta jest mniejsza od wartości odczytanej z tablicy ($t<t_\alpha$), zatem **nie mamy** podstaw do odrzucenia hipotezy zerowej, stąd też możemy stwierdzić, że grupy A i B **nie różnią się istotnie pod względem wieku.**

Kolejnym elementem analizy jest ocena efektów zastosowanej terapii. Żeby to sprawdzić, należy porównać oddzielnie w każdej z grup wielkość skoliozy przed terapią do wielkości skoliozy po terapii.

Do tego celu wykorzystamy test t-studenta dla danych zależnych, którego wzór jest następujący:

$$t = \frac{\bar{d}}{s_d} \cdot \sqrt{n-1} \qquad S_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n}}$$

gdzie: $\bar{d}$ - średnia z różnicy pomiędzy pomiarami skoliozy_przed i skoliozy_po
n – liczba obserwacji
S_d – odchylenie standardowe z różnicy pomiędzy pomiarami skoliozy_przed i skoliozy_po

Na wstępie należy sformułować hipotezy:

H_0 ---Wielkość skoliozy przed zabiegami **nie różni się** od wielkości skoliozy po zabiegach

H_1 --- Wielkość skoliozy przed zabiegami **różni się** od wielkości skoliozy po zabiegach

Dla ułatwienia obliczeń korzystamy z tabeli pomocniczej (poniżej). W kolumnie 5 (d_i) obliczamy różnicę pomiędzy wielkością skoliozy przed zabiegami i wielkością skoliozy po zabiegach. Różnica ta jest nam potrzebna do obliczenia średniej wartości z tych różnic, którą podstawimy do licznika wzoru testu t-studenta.

Inicjały	Grupa	Skolioza_ Przed	Skolioza_ Po	d_i różnica pomiarów	$(d_i - \bar{d})$	$(d_i - \bar{d})^2$
AA	A	18	19	18-19=-1	-1-2,5=-3,5	12,25
BB	A	19	19	19-19=0	0-2,5=-2,5	6,25
CC	A	20	18	20-18=2	2-2,5=-0,5	0,25
DD	A	22	19	22-19=3	3-2,5=0,5	0,25
EE	A	18	15	18-15=3	3-2,5=0,5	0,25
FF	A	17	14	17-14=3	3-2,5=0,5	0,25
GG	A	20	18	20-18=2	2-2,5=-0,5	0,25
HH	A	18	17	18-17=1	1-2,5=-1,5	2,25
II	A	20	14	20-14=6	6-2,5=3,5	12,25
JJ	A	22	16	22-16=6	6-2,5=3,5	12,25
SUMA				**25**		$\sum_{i=1}^{n}(d_i - \bar{d})^2$ =**46**

Jak wiać suma tych różnic wynosi 25, zatem, żeby obliczyć średnią, dzielimy 25 przez 10, co daje wynik 2,5. Ponieważ w mianowniku wzoru testu t jest odchylenie standardowe z różnicy w „skoliozach przed i po", to w następnej kolumnie $(d_i - \bar{d})$ obliczamy różnicę pomiędzy wartością tych różnic a średnią arytmetyczną tych różnic. Jak widać pod pierwiastkiem wzoru na odchylenie standardowe jest wyrażenie $\sum_{i=1}^{n}(d_i - \bar{d})^2$, co jest równoznaczne z sumą wszystkich wierszy w kolumnie 7. Teraz już wystarczy tylko podstawić do wzoru.

$$S_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n}} = \sqrt{\frac{46}{10}} = \sqrt{4,6} = 2,14$$

$$t = \frac{\bar{d}}{s_d} \cdot \sqrt{n-1} = \frac{2,5}{2,14} \cdot \sqrt{10-1} = 1,17 \cdot 3 = 3,51$$

Po wyliczeniu wartości testu t-studenta, należy z tablic dla rozkładu testu t odczytać wartość krytyczną na poziomie istotności α=0,05 oraz n-1 (10-1=9) liczbach stopni swobody.

stopnie swobody	Alfa α		
	0,01	0,05	0,10
6	3,7074	2,4469	1,9432
7	3,4995	2,3646	1,8946
8	3,3554	2,3060	1,8595
9	3,2498	2,2622	1,8331
10	3,1693	2,2281	1,8125

Ponieważ wartość testu t-studenta jest większa od wartości odczytanej z tablic ($t>t_{\alpha}$), daje nam to podstawę do odrzucenia hipotezy zerowej H_0 i przyjęcia alternatywnej. Na tej podstawie wnioskujemy, iż wykonane zabiegi rehabilitacyjne spowodowały istotną statystycznie poprawę w zakresie zmniejszenia wielkości skoliozy mierzonej po zabiegach.

Ostatnim elementem analizy jest ocena współzależności czyli analiza korelacji. Naszym celem jest poznanie przyczyny wielkości skoliozy w oparciu o znany parametr, jakim jest czas poświęcany na oglądanie TV oraz pracę przy komputerze. Żeby obliczyć tę zależność wykorzystamy wzór na korelację liniową Pearsona:

$$r_{xy} = \frac{\operatorname{cov}(x, y)}{Sd_x \cdot Sd_y} \qquad \operatorname{cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Jak widać na podstawie mianownika wzoru na korelację Pearsona potrzebujemy wartość odchylenia standardowego obydwu cech (czyli czasu poświęconego na oglądanie TV i pracę z komputerem), natomiast do obliczenia licznika tegoż wzoru potrzebna będzie suma iloczynów różnic pomiędzy wartościami średnimi obu cech a poszczególnymi wartościami składającymi się na te cechy $\sum (x_i - \bar{x})(y_i - \bar{y})$.

Dla uproszczenia obliczeń należy utworzyć następującą tabelę pomocniczą:

Czas przy tv i komp. x_i	Skol_Przed y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
14	18	14-15,8=-1,8	18-19,4=-1,4	-1,8*-1,4=2,52	3,24	1,96
15	19	15-15,8=-0,8	19-19,4=-0,4	0,32	0,64	0,16
18	20	18-15,8=2,2	20-19,4=0,6	1,32	4,84	0,36
14	22	14-15,8=-1,8	22-19,4=2,6	-4,68	3,24	6,76
16	18	16-15,8=0,2	18-19,4=-1,4	-0,28	0,04	1,96
15	17	15-15,8=-0,8	17-19,4=-2,4	1,92	0,64	5,76
16	20	16-15,8=0,2	20-19,4=0,6	0,12	0,04	0,36
17	18	17-15,8=1,2	18-19,4=-1,4	-1,68	1,44	1,96
18	20	18-15,8=2,2	20-19,4=0,6	1,32	4,84	0,36
15	22	15-15,8=-0,8	22-19,4=2,6	-2,08	0,64	6,76
Σ=158,00	**Σ=194,00**			**Σ=-1,20**	**Σ=19,60**	**Σ=26,40**

$\bar{x}$=158/10=15,8 **$Sd_x = \sqrt{\frac{19,6}{10}} = \sqrt{1,96} = 1,4$**

$\bar{y}$=194/10=19,4 **$Sd_y = \sqrt{\frac{26,4}{10}} = \sqrt{2,64} = 1,62$**

W oparciu o wcześniejsze instrukcje dotyczące obliczania średniej arytmetycznej oraz odchylenia standardowego, możemy przystąpić do podstawienia do wzoru.

$$r_{xy} = \frac{\operatorname{cov}(x, y)}{Sd_x \cdot Sd_y} = \frac{\frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{n}}{Sd_x \cdot Sd_y} = \frac{\frac{-1,2}{10}}{1,4 \cdot 1,62} = \frac{-0,12}{2,27} = -0,05$$

Jak wynika z obliczeń, bardzo niski współczynnik korelacji wskazuje, iż nie ma związku pomiędzy czasem spędzonym na oglądaniu TV oraz pracą przy komputerze na wielkość skoliozy przed terapią.

Oczywiście proszę z tego nie wyciągać daleko idących wniosków, bo jak pisałem na wstępie dane są zupełnie przypadkowe i stworzone „patrząc w sufit" tylko dla potrzeb ćwiczeń ze statystyki.