

Analiza wyników Hong Kong Marathon 2016 przy pomocy statystyki opisowej

Autorzy

Maciej Karczewski 262282
Karolina Wypych 262333



Politechnika
Wrocławska

Wydział Matematyki
15 maja 2022r.

Spis treści

1	Wprowadzenie	2
2	Transformacja danych	2
3	Rodzaje użytych miar	2
3.0.1	Miary położenia	2
3.0.2	Miary rozproszenia	3
3.0.3	Inne miary	3
4	Analiza miar	3
4.1	Pełen zestaw danych	3
4.2	Rozróżnienie płci	5
4.2.1	Analiza rezultatów osiągniętych przez kobiety	6
4.2.2	Analiza rezultatów osiągniętych przez mężczyzn	8
4.2.3	Porównanie wyników po rozróżnieniu z ogólnymi	9
5	Wizualizacja danych	9
5.1	Czas dotarcia na metę	9
5.2	Średnie tempo na danym odcinku	12
5.3	Średnie tempo na poszczególnych częściach dystansu	14
5.4	Podział na płeć	17
6	Wnioski	20
7	Bibliografia	21

1 Wprowadzenie

Tradycja biegów maratońskich sięga czasów nowożytnych. Jako dyscyplina olimpijska zadebiutowała podczas pierwszych nowożytnych Igrzysk Olimpijskich w Atenach w 1896 roku. Wówczas jego dystans liczył sobie 40000m. Obecnie jednak znamy maraton 42195 metrowy. Skąd wzięła się ta magiczna liczba? Zawdzięczamy ją angielskiej rodzinie królewskiej. W 1908 roku podczas IO w Londynie przesunięto start do pałacu w Windsorze, by królewskie dzieci mogły lepiej śledzić rywalizację. Okazało się, że wydłużyło to trasę o 2195m i tak już pozostało. Maratony uliczne cieszą się popularnością wśród biegaczy i organizowane są w miastach na całym świecie, m. in. w Berlinie, Chicago, czy Londynie. To sprawozdanie natomiast poświęćmy maratonowi mającemu miejsce w Hong Kongu w 2016 roku. Dane niezbędne do przeprowadzenia wnikliwych obserwacji pozyskaliśmy za pomocą platformy Kaggle. Przeanalizujemy rezultaty uzyskane przez biegaczy pod kątem wyników częściowych, tzn. na pierwszych 10 kilometrach, na dystansie półmaratonu (21.0975km), na 30 kilometrach oraz pełnej trasie. Ponieważ udział w biegu wzięli przedstawiciele obu płci, zestawimy ze sobą wyniki uzyskane przez kobiety z rezultatami mężczyzn.

2 Transformacja danych

Na początku musimy obrobić nasze dane. W oryginalnych danych czasy na odpowiednich dystansach są w formacie $hh:mm:ss$, co nie jest zmienną numeryczną, ale możemy ją uzyskać, przekonwertowując ten zapis na liczbę minut. Kolejną edycją będzie zamiana wyniku czasowego na tempo na tym odcinku (ilość minut potrzebnych na pokonanie 1 km). Ponadto, by porównać średnie tempo na poszczególnych odcinkach trasy, obliczymy średnie tempo dla następujących po sobie czterech częściach biegu.

3 Rodzaje użytych miar

3.0.1 Miary położenia

1. Średnia arytmetyczna, bardzo podatna na wartości odstające $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$
2. Średnia harmoniczna $\bar{X}_{harm} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$
3. Średnia geometryczna $\bar{X}_{geo} = \sqrt[n]{x_1 x_2 \cdots x_n}$
4. Średnia ucinana, trymowana, średnia arytmetyczna bez k najmniejszych i największych obserwacji $\bar{X}_{trim} = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} x_{(i)}$

- Średnia winsorowska, średnia arytmetyczna ale k najmniejszych obserwacji jest zamieniana na obserwację minimalną po obcięciu ich, analogicznie z wartościami największymi

$$\bar{X}_{wins} = \frac{1}{n} \left[(k+1)x_{(k+1)} + \sum_{i=k+2}^{n-k-1} x_{(i)} + (k+1)x_{(n-k)} \right]$$

- Mediana, czyli wartość środkowa, drugi kwartył (Q2)

$$x_{med} = \begin{cases} x_{(\frac{n+1}{2})} & n \text{ jest nieparzyste} \\ \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) & n \text{ jest parzyste} \end{cases}$$

- Pierwszy kwartył (Q1), mediana obserwacji mniejszych od Q2

- Trzeci kwartył (Q3), mediana obserwacji większych od Q2

3.0.2 Miary rozproszenia

- Rozstęp międzykwartyłowy, różnica trzeciego kwartyłu i pierwszego $IQR = Q3 - Q1$

- Rozstęp, różnica między maksymalnym a minimalnym elementem $R = x_{(n)} - x_{(1)}$

- Wariancja z próby $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$

- Odchylenie standardowe z próby $S = \sqrt{S^2}$

- Współczynnik zmienności $V = \frac{S}{\bar{X}} \cdot 100\%$

3.0.3 Inne miary

- Miara asymetrii, wsółczynnik skośności $\alpha = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{X}}{S} \right)^3$

- Miara spłaszczenia, kurtoza $K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2 \right)^2}$

4 Analiza miar

4.1 Pełen zestaw danych

W niniejszym sprawozdaniu będziemy posługiwać się nazwami "czas oficjalny" oraz "czas netto". Pierwszy z terminów oznacza czas brutto, obejmujący okres od wystrzału startera do minięcia mety przez zawodnika. Natomiast drugi jest

rzeczywistym czasem pokonania dystansu mierzonym od momentu przekroczenia linii startu do mety. Dla analizowanych danych policzyliśmy różne miary ich wartości umieściliśmy w poniższych tabelach.

czas[<i>min</i>]	oficjalny	netto	po 10 km	po połowie dystansu	po 30km	oficjalny-netto
\bar{X}	269.00	267.10	61.27	126.60	183.07	1.91
$Q1$	233.04	231.83	54.47	112.29	160.49	0.66
x_{med}	268.36	266.28	61.35	126.10	181.70	1.57
$Q3$	305.38	302.97	68.42	141.62	206.32	2.82
X_{harm}	257.02	255.37	58.73	121.48	175.21	0.00
X_{geo}	261.36	259.59	59.56	123.07	177.80	0.00
X_{trim}	265.06	263.12	60.40	124.50	179.79	1.72
X_{wins}	269.00	267.10	61.27	126.60	183.07	1.91

Tabela 1: Wartości miar położenia dla czasów grupy mieszanej.

tempo[<i>min/km</i>]	oficjalne	netto	na odcinku 0-10km	na odcinku 10-21.0975km	na odcinku 21.0975-30km	na odcinku 30-42.195km	na odcinku 0-2.10975km	na odcinku 0-30km
\bar{X}	6.38	6.34	6.13	5.89	6.35	6.90	6.01	6.11
$Q1$	5.53	5.50	5.45	5.20	5.38	5.72	5.33	5.35
x_{med}	6.36	6.32	6.14	5.84	6.19	6.80	5.98	6.06
$Q3$	7.24	7.19	6.85	6.57	7.24	8.00	6.72	6.88
X_{harm}	6.10	6.06	5.88	5.65	6.01	6.53	5.76	5.85
X_{geo}	6.20	6.16	5.96	5.72	6.14	6.68	5.84	5.93
X_{trim}	6.29	6.24	6.04	5.77	6.18	6.79	5.91	6.00
\bar{X}_{wins}	6.38	6.34	6.13	5.89	6.35	6.90	6.01	6.11

Tabela 2: Wartości miar położenia dla tempa grupy mieszanej.

czas[<i>min</i>]	oficjalny	netto	po 10 km	po połowie dystansu	po 30km	oficjalny-netto
S	46.98	46.26	9.8	19.82	30.41	1.52
S^2	2206.51	2139.55	95.91	392.75	924.55	2.31
V	0.18	0.18	0.16	0.16	0.17	0.8
IQR	72.35	71.14	13.95	29.33	45.84	2.17
R	234.87	227.52	65.52	119.02	171.84	10.4

Tabela 3: Wartości miar rozproszenia dla czasów grupy mieszanej.

tempo [<i>min/km</i>]	oficjalne	netto	na odcinku 0-10km	na odcinku 10-21.0975km	na odcinku 21.0975-30km	na odcinku 30-42.195km	na odcinku 0-21.0975km	na odcinku 0-30km
S	1.12	1.1	0.98	0.94	1.28	1.45	0.94	1.02
S^2	1.24	1.21	0.96	0.89	1.64	2.09	0.89	1.03
V	0.18	0.18	0.16	0.16	0.21	0.21	0.16	0.17
IQR	1.72	1.69	1.4	1.38	1.87	2.29	1.4	1.53
R	5.57	5.4	6.56	5.87	9.54	8.75	5.65	5.73

Tabela 4: Wartości miar rozproszenia dla tempa grupy mieszanej.

czas[<i>min</i>]	oficjalny	netto	po 10 km	po połowie dystansu	po 30km	oficjalny-netto
min	132.20	132.19	30.57	64.80	93.60	0.00
max	367.07	359.70	96.09	183.82	265.44	10.40
α	-0.02	-0.01	-0.11	-0.04	0.04	0.89
K	-0.78	-0.78	-0.43	-0.52	-0.63	0.43

Tabela 5: Wartości innych miar dla czasów grupy mieszanej.

tempo [min/km]	oficjalne	netto	na odcinku 0-10km	na odcinku 10-21.0975km	na odcinku 21.0975-30km	na odcinku 30-42.195km	na odcinku 0-21.0975km	na odcinku 0-30km
<i>min</i>	3.14	3.14	3.06	3.09	3.21	3.17	3.08	3.12
<i>max</i>	8.70	8.53	9.61	8.95	12.74	11.91	8.72	8.85
α	-0.02	-0.01	-0.11	0.16	0.43	0.20	-0.04	0.04
K	-0.78	-0.78	-0.43	-0.48	-0.30	-0.69	-0.52	-0.63

Tabela 6: Wartości innych miar dla tempa grupy mieszanej.

Dane umieszczone w tabeli 1 potwierdzają zasadność rozróżniania czasu oficjalnego i netto. Różnica między ich średnimi to około 1.88 min, czyli 1 minuta i 53 sekundy. Oznacza to, że średnio zawodnik potrzebował niemalże dwóch minut na dobiegnięcie od miejsca, w którym rozpoczął bieg do właściwej linii startu.

Rozważania na temat rezultatów uzyskanych przez uczestników Hong Kong Marathon 2016 rozpoczniemy od analizy miar położenia, rozproszenia, asymetrii i spłaszczenia (koncentracji) dla pełnego zestawu obrabianych danych. Odchylenie standardowe wskazuje, że zarówno dla netto jak i brutto wyniki są rozrzucone od średniej o około 18% jej wartości 1. W obu przypadkach drugi kwartył, czyli mediana jest nieco mniejsza od średniej arytmetycznej 1 (implikuje to rozkład prawostronnie skośny), co oznacza, że ponad połowa zawodników zmieściła się w średnim czasie a w pozostałej grupie, choć mniej licznej pojawiły się na tyle długie czasy, że podwyższyły średnią wszystkich obserwacji. Natomiast pierwszy i trzeci kwartył obrazują poniżej jakiej wartości wynik uzyskało odpowiednio 25% i 75% uczestników.

Niezwykle istotne dla biegaczy, ale także bardzo ciekawe pod kątem analizy są informacje o tempie na poszczególnych odcinkach trasy. Wartości miar dla 4 części biegu zostały przedstawione w tabelach 2-4-6. Widać z nich, że zawodnicy najszybciej pokonują drugi odcinek trasy a najwolniej ostatni, co jest spowodowane naturalnym funkcjonowaniem ludzkiego organizmu, który nie posiada wystarczająco energii, by przebiec cały maraton i bardzo często w okolicach 35 kilometra zaczyna się tzw. "ściana", czyli brak sił u biegaczy, co zmusza ich do porzucenia biegu na rzecz marszu. Patrząc na wartość odchylenia standardowego, widać, że tempo zawodników jest najbardziej zróżnicowane w czwartym segmencie (nie każdego dopadł kryzys ściany) 4. Zauważamy, że najwolniejsze tempo biegu wystąpiło w trzeciej ćwiartce i mocno odstaje od trzeciego kwartyłu 8. Natomiast wartość minimum niezależnie od etapu wyścigu oscyluje na poziomie nieco powyżej 3 minut 6. Wszędzie mamy do czynienia z rozkładem prawostronnie skośnym. Wyjątek stanowi jedynie odcinek pierwszy, gdzie mediana jest większa, co implikuje lewostronną skośność rozkładu.

W przypadku tempa na całej trasie jego wartość netto od oficjalnego jest nieco mniejsza, ponieważ liczy się ją względem krótszego odcinka czasu. Z mediany wnioskujemy, że większość pobiegła poniżej średniej 1.

4.2 Rozróżnienie płci

W gronie zawodników Hong Kong Marathon 2016 znaleźli się zarówno mężczyźni jak i kobiety. Interesujący materiał do dalszych badań daje nam rozróżnienie zawodników ze względu na płeć.

4.2.1 Analiza rezultatów osiągniętych przez kobiety

W wyścigu wzięło udział 899 kobiet. Statystyki ich wyników umieściliśmy w poniższych tabelach.

czas[<i>min</i>]	oficjalny	netto	po 10 km	po połowie dystansu	po 30km	oficjalny-netto
\bar{X}	286.62	284.59	66.00	136.88	197.23	2.03
$Q1$	260.95	259.40	60.40	125.79	180.00	0.86
x_{med}	288.85	286.45	67.19	138.27	198.65	1.82
$Q3$	318.59	316.58	72.59	150.03	217.35	2.97
\bar{X}_{harm}	279.71	277.86	64.60	134.14	192.92	0.00
\bar{X}_{geo}	283.32	281.37	65.33	135.57	195.16	0.00
\bar{X}_{trim}	288.71	286.57	66.53	137.79	198.48	1.92
\bar{X}_{wins}	286.62	284.59	66.00	136.88	197.23	2.03

Tabela 7: Wartości miar położenia dla czasów grupy kobiet.

tempo[<i>min/km</i>]	oficjalne	netto	na odcinku 0-10km	na odcinku 10-21.0975km	na odcinku 21.0975-30km	na odcinku 30-42.195km	na odcinku 0-2.10975km	na odcinku 0-30km
\bar{X}	6.80	6.75	6.60	6.39	6.78	7.17	6.49	6.58
$Q1$	6.19	6.15	6.04	5.84	6.06	6.29	5.97	6.00
x_{med}	6.85	6.79	6.72	6.42	6.77	7.24	6.56	6.63
$Q3$	7.56	7.51	7.26	6.99	7.57	8.10	7.12	7.25
\bar{X}_{harm}	6.63	6.59	6.46	6.26	6.59	6.94	6.36	6.44
\bar{X}_{geo}	6.72	6.67	6.54	6.33	6.69	7.06	6.43	6.51
\bar{X}_{trim}	6.85	6.80	6.66	6.42	6.80	7.18	6.54	6.62
\bar{X}_{wins}	6.80	6.75	6.60	6.39	6.78	7.17	6.49	6.58

Tabela 8: Wartości miar położenia dla tempa grupy kobiet.

tempo[<i>min/km</i>]	oficjalne	netto	na odcinku 0-10km	na odcinku 10-21.0975km	na odcinku 21.0975-30km	na odcinku 30-42.195km	na odcinku 0-2.10975km	na odcinku 0-30km
\bar{X}	6.80	6.75	6.60	6.39	6.78	7.17	6.49	6.58
$Q1$	6.19	6.15	6.04	5.84	6.06	6.29	5.97	6.00
x_{med}	6.85	6.79	6.72	6.42	6.77	7.24	6.56	6.63
$Q3$	7.56	7.51	7.26	6.99	7.57	8.10	7.12	7.25
\bar{X}_{harm}	6.63	6.59	6.46	6.26	6.59	6.94	6.36	6.44
\bar{X}_{geo}	6.72	6.67	6.54	6.33	6.69	7.06	6.43	6.51
\bar{X}_{trim}	6.85	6.80	6.66	6.42	6.80	7.18	6.54	6.62
\bar{X}_{wins}	6.80	6.75	6.60	6.39	6.78	7.17	6.49	6.58

Tabela 9: Wartości miar położenia dla tempa grupy kobiet.

czas[<i>min</i>]	oficjalny	netto	po 10 km	po połowie dystansu	po 30km	oficjalny-netto
\bar{S}	41.82	41.23	9.02	18.29	27.55	1.46
\bar{S}^2	1748.42	1699.42	81.22	334.48	758.57	2.11
V	0.15	0.15	0.14	0.14	0.14	0.72
IQR	57.64	57.18	12.19	24.25	37.35	2.12
R	210.22	202.9	47.57	102.64	152.15	8.6

Tabela 10: Wartości miar rozproszenia dla czasów grupy kobiet.

tempo [min/km]	oficjalne	netto	na odcinku 0-10km	na odcinku 10-21.0975km	na odcinku 21.0975-30km	na odcinku 30-42.195km	na odcinku 0-21.0975km	na odcinku 0-30km
S	1.00	0.98	0.91	0.88	1.11	1.24	0.87	0.92
S^2	0.99	0.96	0.82	0.76	1.22	1.54	0.76	0.85
V	0.15	0.15	0.14	0.14	0.17	0.18	0.14	0.14
IQR	1.37	1.36	1.22	1.16	1.51	1.82	1.15	1.25
R	4.99	4.81	4.76	4.97	6.23	7.47	4.87	5.08

Tabela 11: Wartości miar rozproszenia dla tempa grupy kobiet.

Zarówno w przypadku oficjalnego czasu jak i czasu netto wyniki odchylają się od średniej o około 14% jej wartości 10. Tym razem mediana przewyższa średnią (rozkład lewostronnie skośny), więc większość kobiet biegła dłużej niż wskazuje średnia 7. Powodów dla których średnia jest niższa od wyniku ponad połowy zawodniczek jest wiele, ale jednym z nich może być obecność grupy wybitnych zawodniczek lub nieduża liczba kobiet ze znacznie wyższym czasem.

Analizując tempo biegu na czterech odcinkach, wnioskujemy, że najszybciej biegnięto w drugiej a najwolniej w czwartej części wyścigu, w której to wyniki miały największe rozbieżności. Patrząc na to, że jedynie w trzecim segmencie mediana jest niższa od średniej, można powiedzieć, że jako grupa zawodniczeki zaprezenowały się najlepiej właśnie na tym fragmencie⁹.

Zważając na to, że na przeważającej części trasy mediana przewyższa średnią, podejrzewamy, że sytuacja będzie analogiczna dla całego maratonu. Słuszność naszych spekulacji potwierdzają dane zarówno dla czasu netto jak i brutto. Ponadto w obu przypadkach dane są prawie tak samo "odchylone" 9 4.

czas[min]	oficjalny	netto	po 10 km	po połowie dystansu	po 30km	oficjalny-netto
min	156.85	156.80	38.60	80.70	113.29	0.00
max	367.07	359.70	86.17	183.34	265.44	8.60
α	-0.50	-0.48	-0.58	-0.52	-0.47	0.81
K	-0.01	-0.03	0.28	0.26	0.12	0.77

Tabela 12: Wartości innych miar dla czasów grupy kobiet.

tempo [min/km]	oficjalne	netto	na odcinku 0-10km	na odcinku 10-21.0975km	na odcinku 21.0975-30km	na odcinku 30-42.195km	na odcinku 0-21.0975km	na odcinku 0-30km
min	3.72	3.72	3.86	3.80	3.67	3.57	3.83	3.78
max	8.70	8.53	8.62	8.76	9.89	11.04	8.69	8.85
α	-0.50	-0.48	-0.58	-0.32	-0.14	-0.13	-0.52	-0.47
K	-0.01	-0.03	0.28	0.12	-0.13	-0.22	0.26	0.12

Tabela 13: Wartości innych miar dla tempa grupy kobiet.

4.2.2 Analiza rezultatów osiągniętych przez mężczyzn

czas[<i>min</i>]	oficjalny	netto	po 10 km	po połowie dystansu	po 30km	oficjalny-netto
\bar{X}	265.62	263.74	60.36	124.63	180.35	1.88
$Q1$	229.53	227.97	53.45	110.48	157.98	0.62
x_{med}	263.14	260.97	60.27	123.44	177.69	1.52
$Q3$	301.25	298.95	67.50	139.29	203.02	2.75
X_{harm}	257.02	255.37	58.73	121.48	175.21	0.00
X_{geo}	261.36	259.59	59.56	123.07	177.80	0.00
X_{trim}	265.06	263.12	60.40	124.50	179.79	1.72
X_{wins}	265.62	263.74	60.36	124.63	180.35	1.88

Tabela 14: Wartości miar położenia dla czasów grupy mężczyzn.
Wśród zawodników znalazło się 4687 mężczyzn (84% uczestników), dla których statystyki wyglądają następująco.

tempo[<i>min/km</i>]	oficjalne	netto	na odcinku 0-10km	na odcinku 10-21.0975km	na odcinku 21.0975-30km	na odcinku 30-42.195km	na odcinku 0-2.10975km	na odcinku 0-30km
\bar{X}	6.30	6.26	6.04	5.80	6.26	6.84	5.91	6.02
$Q1$	5.44	5.41	5.35	5.10	5.29	5.63	5.24	5.27
x_{med}	6.24	6.19	6.03	5.70	6.05	6.71	5.86	5.93
$Q3$	7.14	7.09	6.75	6.46	7.12	7.99	6.61	6.77
X_{harm}	6.10	6.06	5.88	5.65	6.01	6.53	5.76	5.85
X_{geo}	6.20	6.16	5.96	5.72	6.14	6.68	5.84	5.93
X_{trim}	6.29	6.24	6.04	5.77	6.18	6.79	5.91	6.00
\bar{X}_{wins}	6.30	6.26	6.04	5.80	6.26	6.84	5.91	6.02

Tabela 15: Wartości miar położenia dla tempa grupy mężczyzn.

czas[<i>min</i>]	oficjalny	netto	po 10 km	po połowie dystansu	po 30km	oficjalny-netto
S	47.16	46.42	9.68	19.49	30.18	1.46
S^2	2223.81	2154.38	93.62	379.83	910.74	2.11
V	0.18	0.18	0.17	0.16	0.17	0.72
IQR	71.72	70.99	14.05	28.81	45.05	2.12
R	230.77	227.32	65.52	119.02	163.69	8.6

Tabela 16: Wartości miar rozproszenia dla czasów grupy mężczyzn.

tempo [<i>min/km</i>]	oficjalne	netto	na odcinku 0-10km	na odcinku 10-21.0975km	na odcinku 21.0975-30km	na odcinku 30-42.195km	na odcinku 0-21.0975km	na odcinku 0-30km
S	1.12	1.11	0.97	0.93	1.30	1.48	0.93	1.01
S^2	1.25	1.22	0.94	0.85	1.68	2.18	0.86	1.02
V	0.18	0.18	0.17	0.16	0.21	0.22	0.16	0.17
IQR	1.70	1.69	1.41	1.36	1.84	2.36	1.37	1.51
R	5.47	5.39	6.56	5.87	9.54	8.75	5.65	5.46

Tabela 17: Wartości miar rozproszenia dla tempa grupy mężczyzn.

czas[<i>min</i>]	oficjalny	netto	po 10 km	po połowie dystansu	po 30km	oficjalny-netto
min	132.20	132.19	30.57	64.80	93.60	0.00
max	362.97	359.50	96.09	183.82	257.29	10.40
α	0.09	0.10	-0.03	0.06	0.15	0.91
K	-0.79	-0.79	-0.39	-0.49	-0.60	0.39

Tabela 18: Wartości innych miar dla czasów grupy mężczyzn.

tempo [min/km]	oficjalne	netto	na odcinku 0-10km	na odcinku 10-21.0975km	na odcinku 21.0975-30km	na odcinku 30-42.195km	na odcinku 0-21.0975km	na odcinku 0-30km
<i>min</i>	3.14	3.14	3.06	3.09	3.21	3.17	3.08	3.12
<i>max</i>	8.61	8.52	9.61	8.95	12.74	11.91	8.72	8.58
α	0.09	0.10	-0.03	0.25	0.56	0.27	0.06	0.15
K	-0.79	-0.79	-0.39	-0.40	-0.17	-0.72	-0.49	-0.60

Tabela 19: Wartości innych miar dla tempa grupy mężczyzn.

Mężczyźni podobnie do kobiet w najlepszym tempie przebiegli drugą część biegu a najslabiej ostatnią, charakteryzująca się również największym odchyleniem standardowym. Każdy segment większość biegaczy pokonała w lepszym tempie niż średnia 15 17.

Patrząc na relację większości pomiędzy medianą a średnią tempa na poszczególnych fragmentach trasy, nie jest zakoczeniem, że zachodzi ona dla całości wyścigu.

Wnioski dla mężczyzn są takie same jak dla kategorii open. Dzieje się tak ponieważ mężczyźni stanowią 84 % wśród wszystkich uczestników.

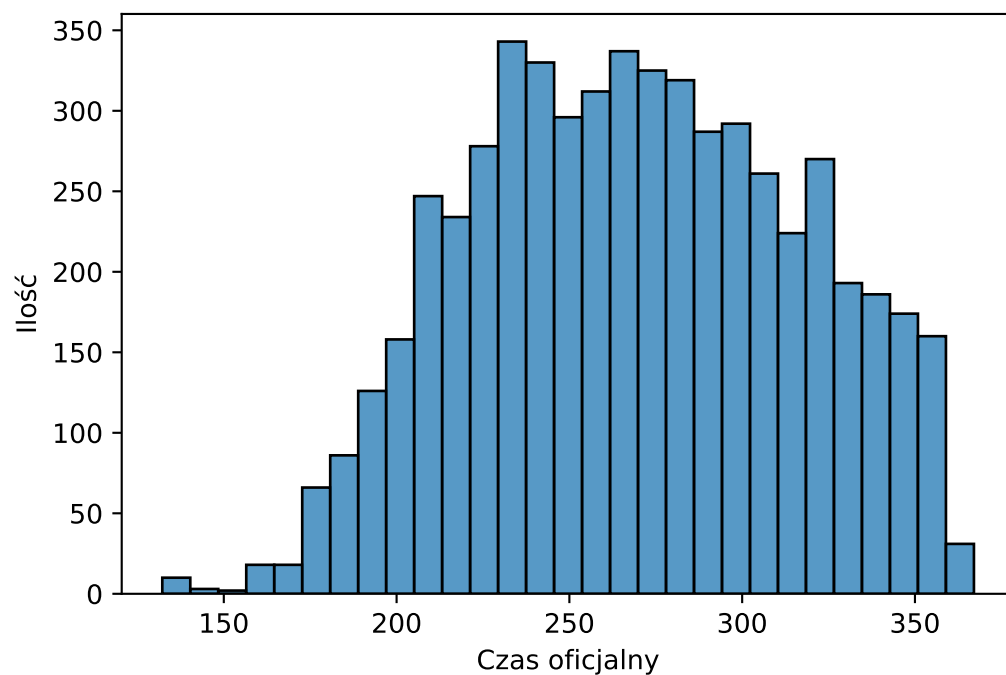
4.2.3 Porównanie wyników po rozróżnieniu z ogólnymi

Na koniec przyjrzyjmy się jak odrębne rezultaty kobiet i mężczyzn mają się do wyników grupy mieszanej. Kobiety pokonały maraton w czasie około 6.5% dłuższym a mężczyźni około 1% krótszym niż pokazuje średnia. Jest to efekt przede wszystkim warunków fizycznych, dzięki którym mężczyźni z natury szybciej biegają. Ponadto mniejsze odchylenie wyniku od ogólnej średniej grupy mężczyzn niż kobiet świadczy o tym, że ci pierwsi stanowili zdecydowaną większość zawodników. Minimalny czas przebiegnięcia maratonu pokrywa się w tabeli męskiej i ogólnej, co oznacza, że to właśnie przedstawiciel tej kategorii wygrał wyścig 5 18. Niewielka różnica czasów netto i brutto oznacza, że rozpoczął wyścig stojąc blisko linii startu. Najdłużej natomiast biegła kobieta, której dotarcie do umownego miejsca rozpoczęcia maratonu zajęło ponad 7 minut.5 12

5 Wizualizacja danych

5.1 Czas dotarcia na mete

Na początku zobaczymy jak wygląda histogram liczności czasu ukończenia maratonu, który pokaże nam jak dużo osób osiągnęło dany wynik.

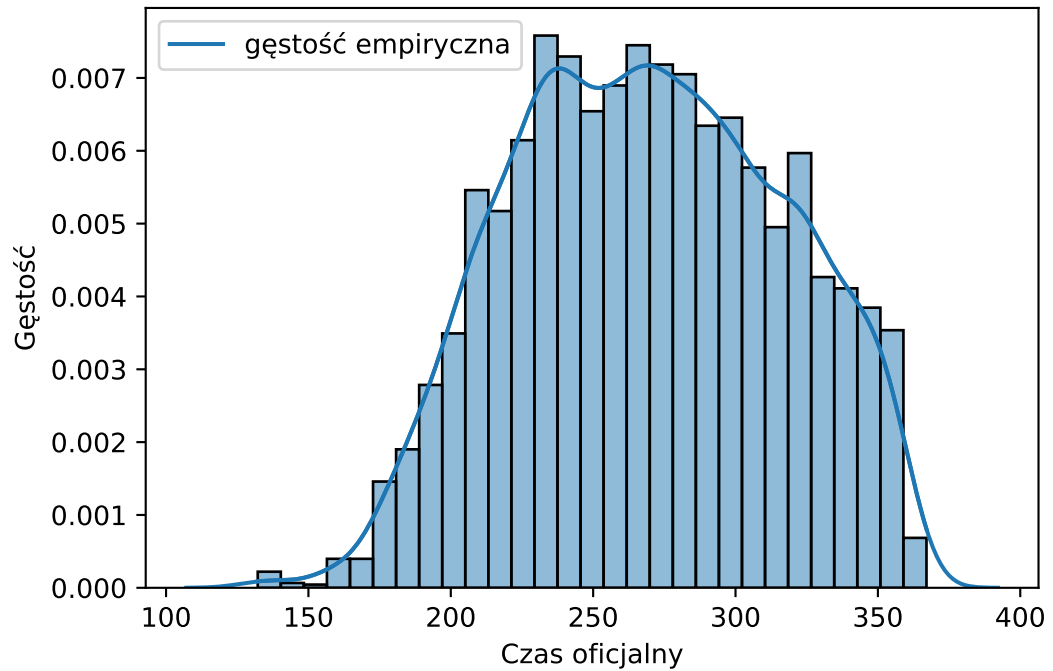


Rysunek 1: Histogram częstości wyniku oficjalnego

Analizując histogram dochodzimy do wniosku, że na biegu startowała elita co sugerują nam wyniki w okolicach 150 minut, których jest zdecydowanie mniej niż innych. Widzimy też, że najwięcej biegaczy ukończyło bieg pomiędzy 225 minut a 320 minut (3:45 - 5:20).

Następnie zobaczymy też histogram ale częstotliwości, który pokaże nam jak wygląda gęstość zmiennej losowej.

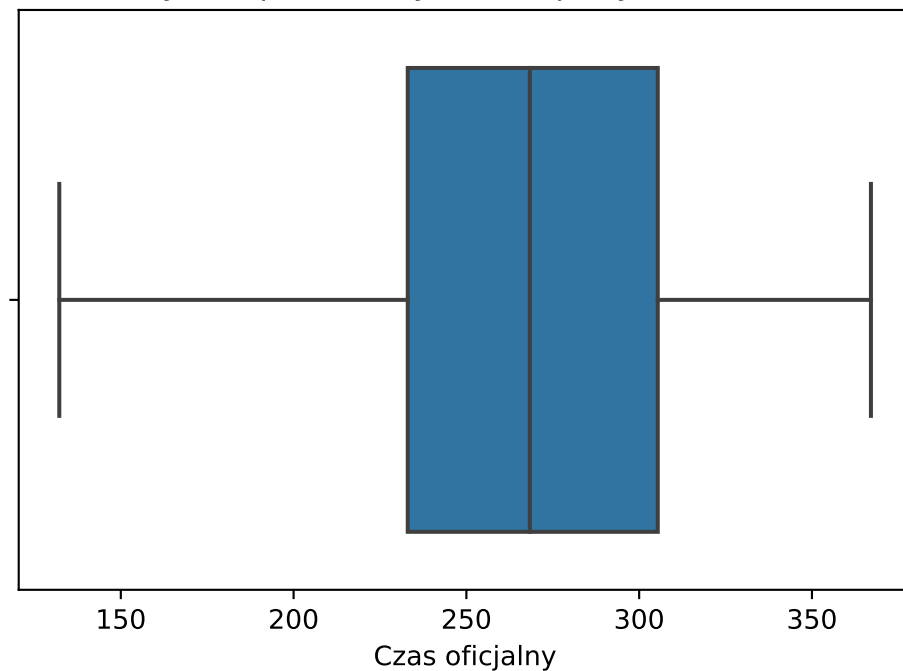
Prawdopodobieństwa i gęstość empiryczna dla oficjalnych czasów



Rysunek 2: Gęstość zmiennej losowej

Analizując ten wykres dochodzimy do identycznych wniosków jak wyżej.
Następnie przejdziemy do wykresu pudełkowego

Wykres pudełkowy dla oficjalnych czasów

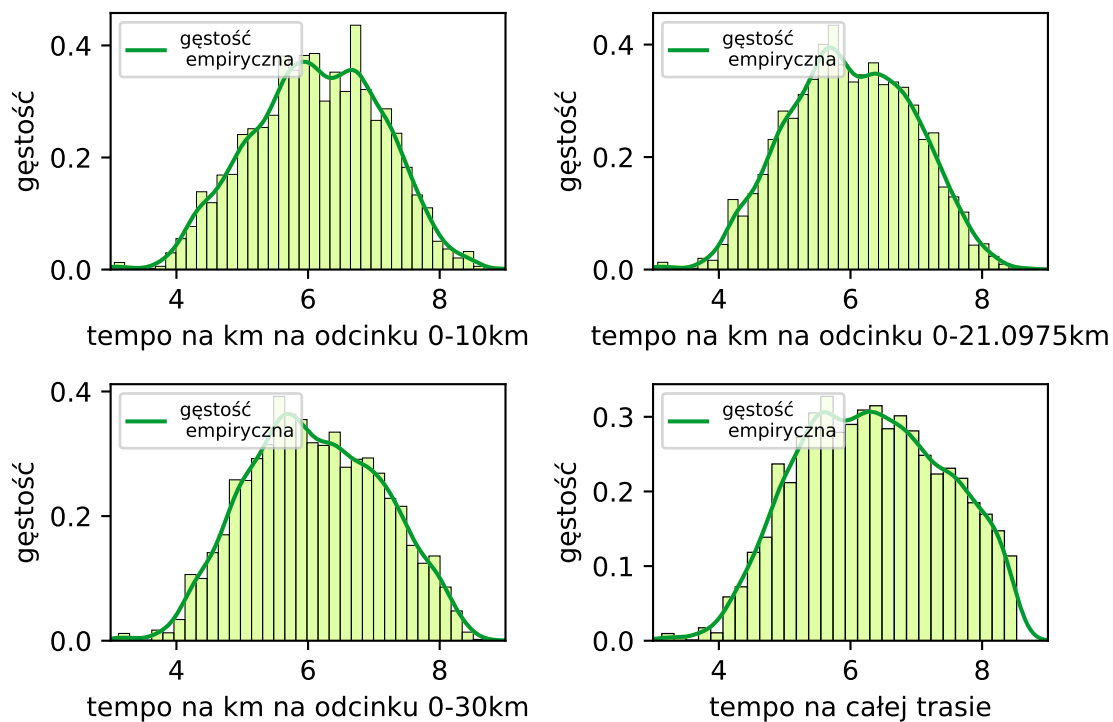


Rysunek 3: Wykres pudełkowy dla czasów ukończenia

Z tego wykresu doskonale widzimy że istnieje grupa szybszych maratończyków, elita biegu oraz, że najbardziej spodziewamy się wyniku pomiędzy 230 a 310 min (3:50 - 5:10).

5.2 Średnie tempo na danym odcinku

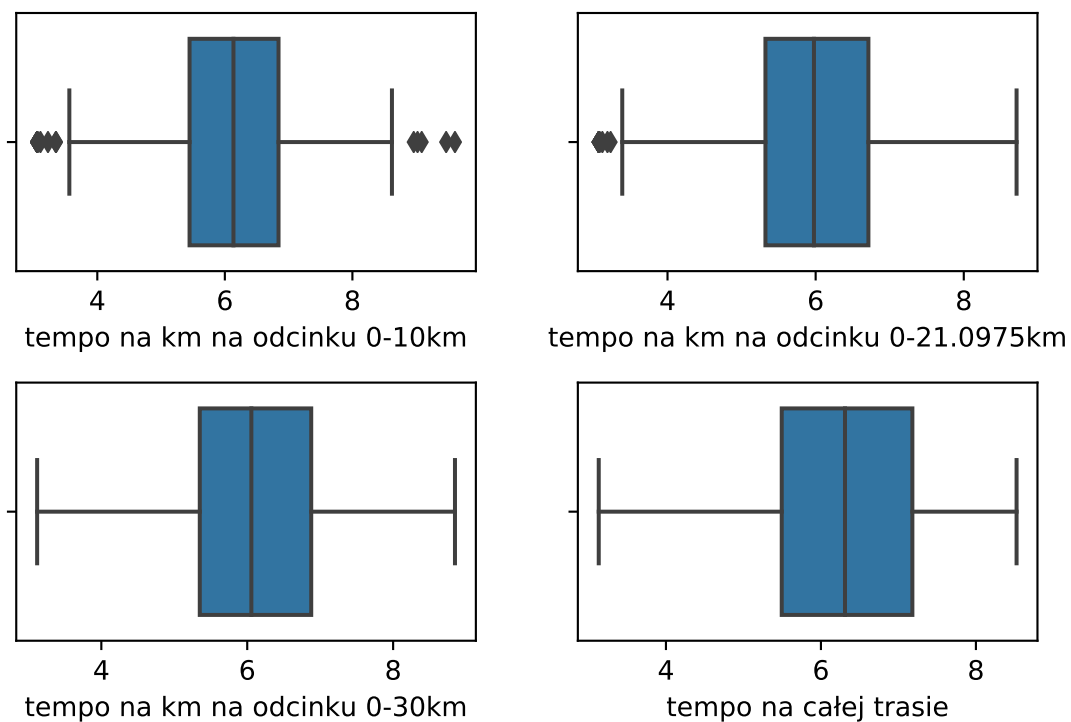
Na początku porównamy jak wyglądało średnie tempo biegaczy na danych odcinkach za pomocą porównania histogramów częstości.



Rysunek 4: Średnie tempo na danym dystansie

Analizując wykresy widzimy, że średnie tempo zmienia się na przestrzeni kilometrów. Lewa strona histogramu staje się coraz bardziej stroma, natomiast prawa coraz łagodniejsza, tym samym dłuższa. To może oznaczać że ludzie wraz z dystansem zwalniają.

Teraz zobaczmy jak te same dane wyglądają na boxplocie.

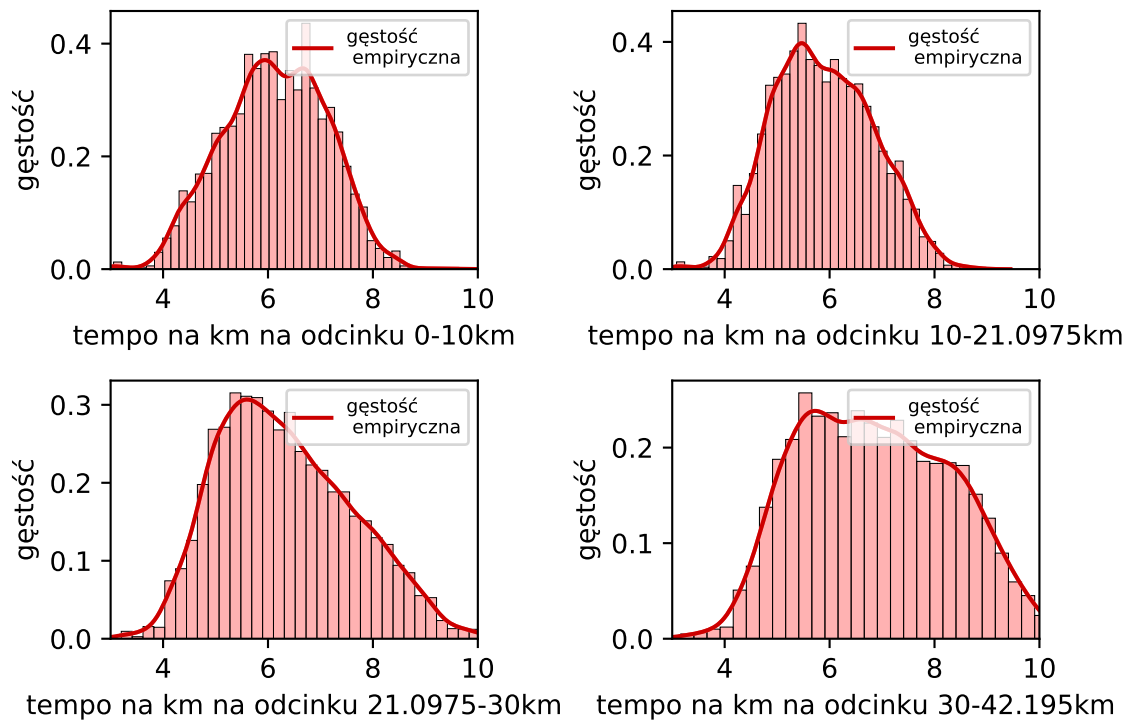


Rysunek 5: Wykres pudełkowy tempa na wybranych odcinkach

Tutaj już dobrze widać, średnie tempo na końcu jest najślabsze, a rozstęp między kwartylowy ponieważ pudełko jest najszersze.

5.3 Średnie tempo na poszczególnych częściach dystansu

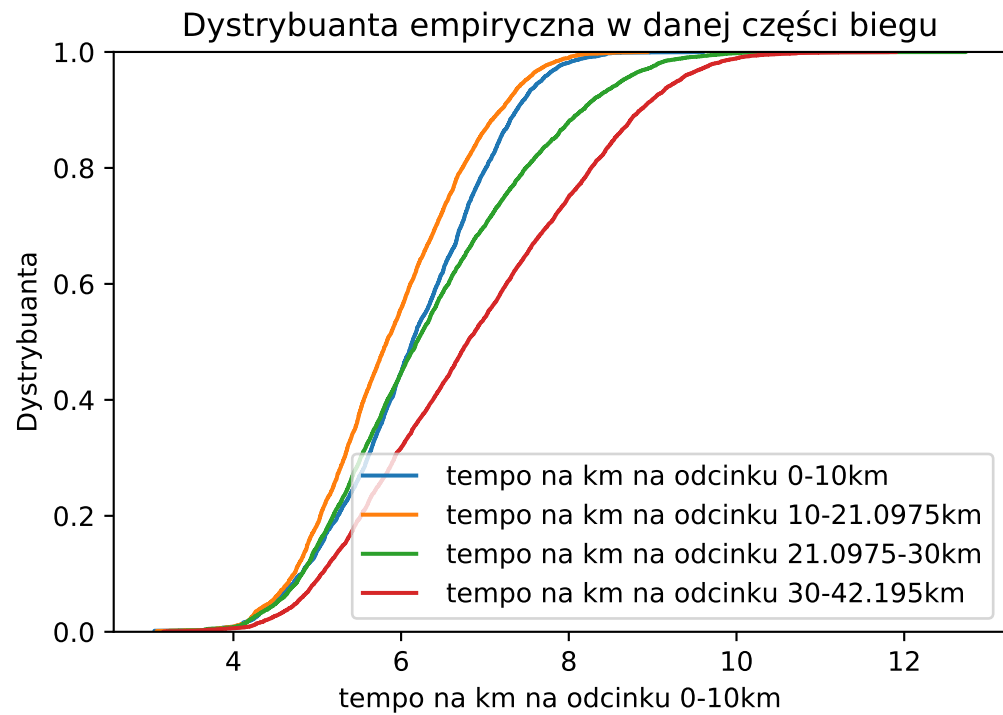
Jak zobaczyliśmy, że średnie tempo zmienia się z czasem to możemy się temu przyjrzeć bliżej analizując średnie tempa z danych odcinków trasy. Najpierw zobaczmy to na histogramie.



Rysunek 6: Tempo na danej części dystansu

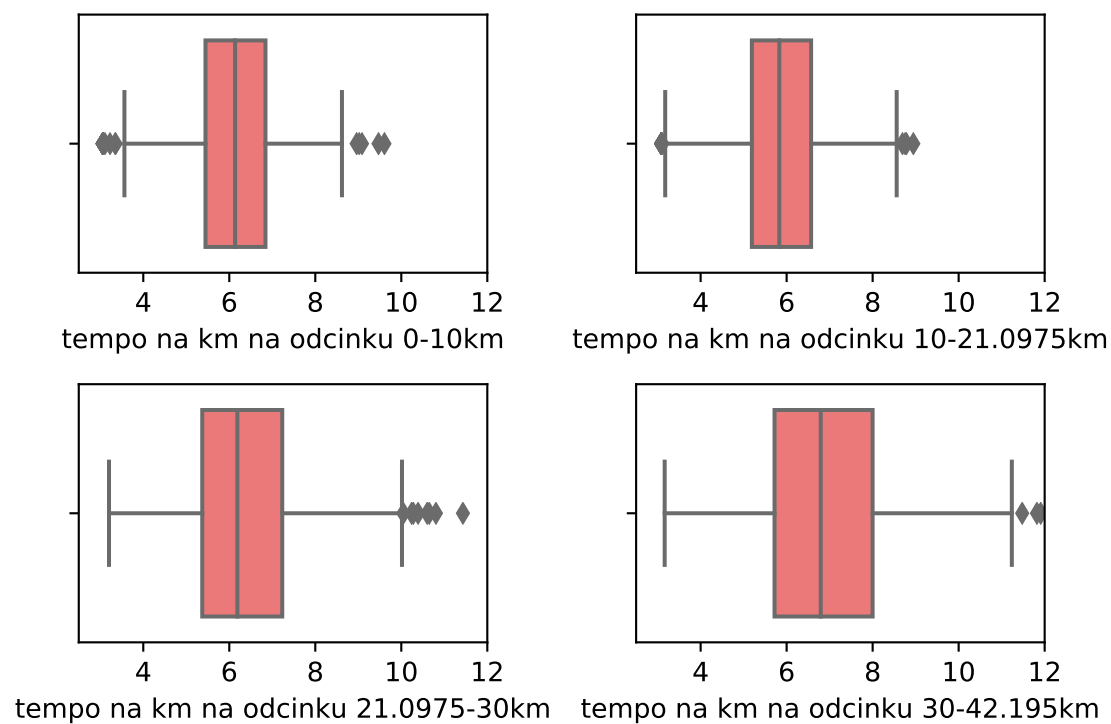
Tutaj bardzo dobrze widać zmęczenie biegaczy, więc i wolniejsze tempo. Wystarczy spojrzeć na tempo na km na odcinku 21.0975-30 km i porównać z pierwszymi 10km. Jeszcze większa różnica jest na odcinku 30-42.195km. Tutaj bardzo dobrze widać mityczną "ścianę" na około 35 km

Następnie zobaczymy jak wyglądają ich dystrybuanty



Rysunek 7: Dystrybuanty tempa na danej części biegu

Tutaj dochodzimy do tych samych wniosków oraz załóżamy, że tempo na odcinku 10-21.0975 było szybsze niż na pierwszych 10km. Teraz zobaczymy jak to wygląda na boxplotach.

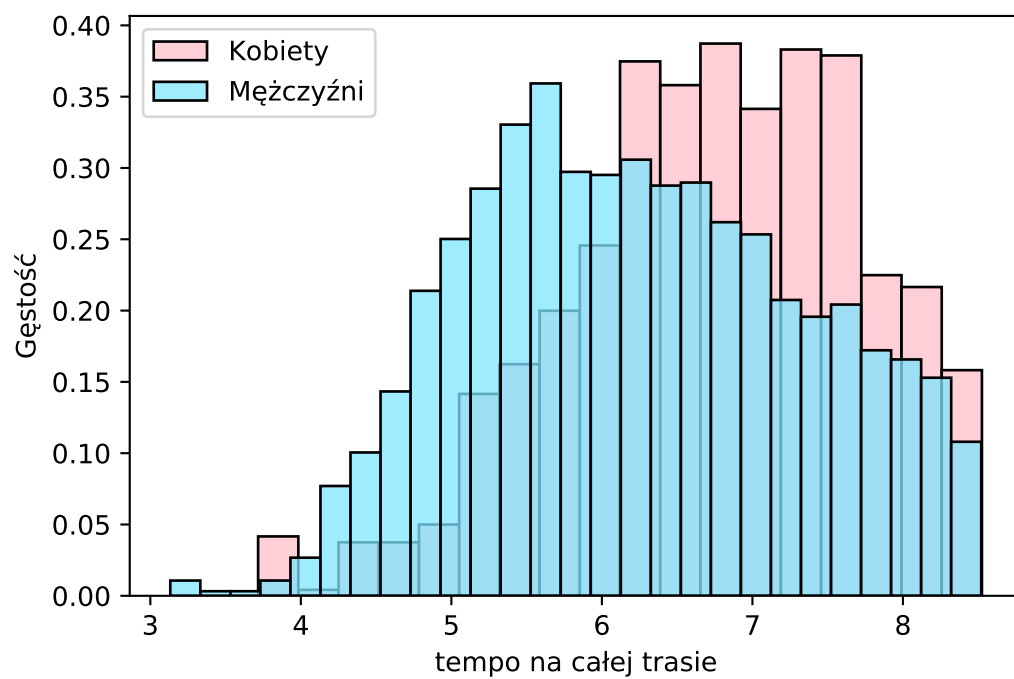


Rysunek 8: Wykres pudełkowy tempa na danych częściach biegu

Tutaj widzimy dodatkowo zwiększające się rozstępy między kwartyłowe , oraz jak bardzo tempo najwolniejszych spada , a najszybszych niekoniecznie.

5.4 Podział na płeć

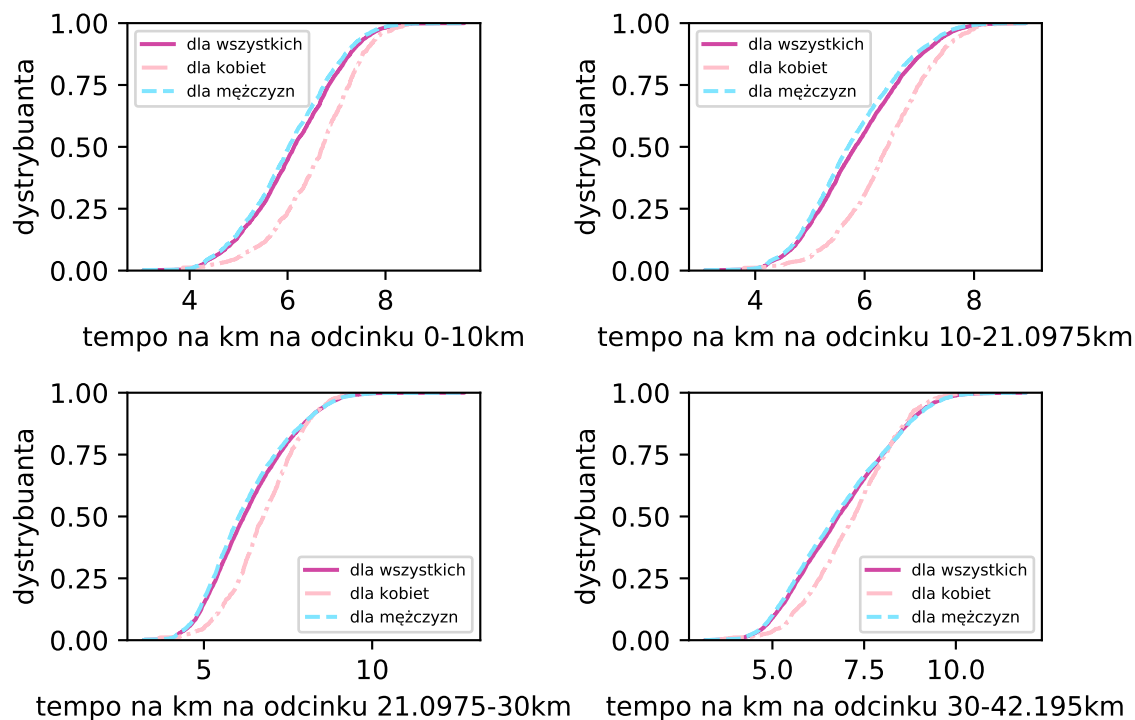
W tej części porównamy czasy w zależności od płci. Na początku użyjemy histogramu.



Rysunek 9: Porównanie tempa w zależności od płci

Tutaj możemy załważyć, że kobiety biegają wolniej od mężczyzn, jednakże wyniki najgorsze mają porównywalne co może być wynikiem małej ilości kobiet w porównaniu do ilości mężczyzn.

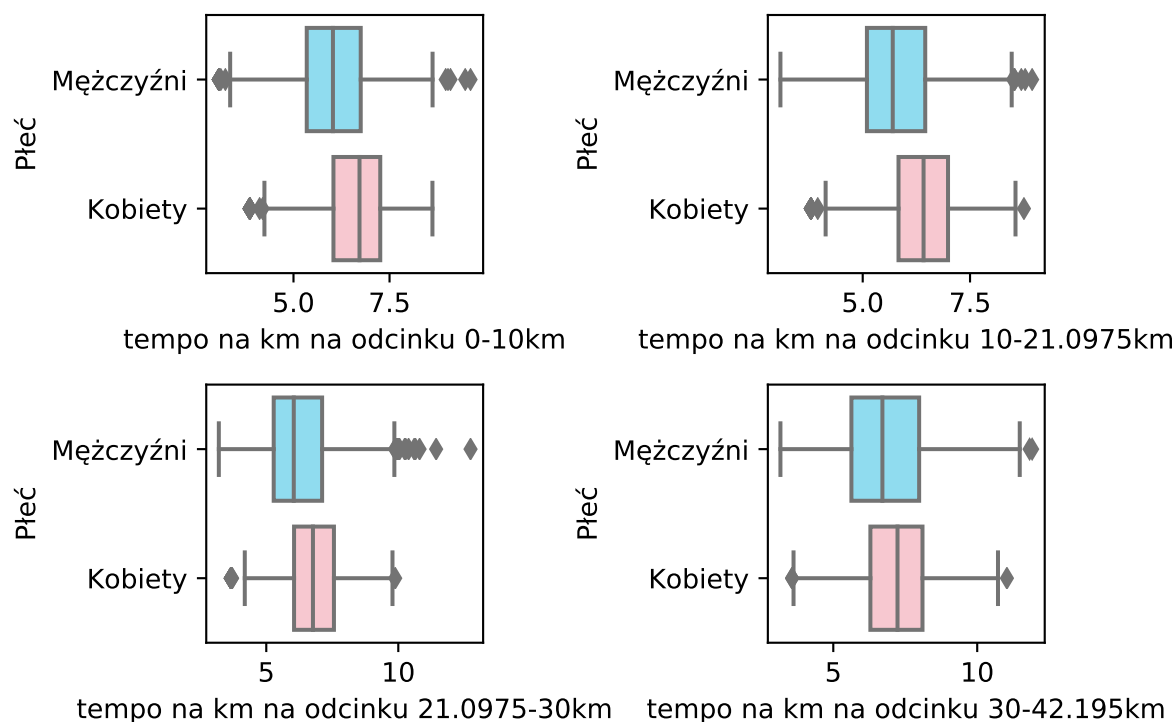
Teraz zobaczymy jak to wygląda na wykresie dystrybuantry



Rysunek 10: Dystrybuanta na danych częściach dystansu w zależności od płci

Z Dystrybuandach możemy odczytać, że kobiety jak i mężczyźni spowolniają z czasem. Ta różnica jest największa dla pierwszej połowy dystansu co może wynika z posiadania sił. Tutaj można także załważyć, że kobiet jest zdecydowanie mniej ponieważ dystrybuanta dla wszystkich prawie się pokrywa z dystrybuantą dla mężczyzn.

Na koniec porównamy dane dla płci na boxplotach.



Rysunek 11: Tempo w każdej części maratonu ze względu na płeć

Tutaj możemy załważyć, że dla kobiet i mężczyzn zachodzą te same własności oraz, że wraz z dystansem maleje różnica 3 kwartyła mężczyzn od 3 kwartyła kobiet.

6 Wnioski

Badany przez nas zestawem danych były oficjalne wyniki Hong Kong Marathon 2016. Dzięki kilku transformacjom rozszerzyliśmy tę bazę o kolejne parametry, które pozwoliły wnikliwie przyjrzeć się rezultatom uzyskanym przez zawodników. Podzieliliśmy i porównaliśmy dane przez wzgląd odcinek trasy i płeć. Efekty wykonanych zaprezentowaliśmy pod postacią wartości liczbowych w przejrzystych tabelach. Ponadto zwizualizowaliśmy dane przy pomocy różnych rodzajów wykresów. Zdołaliśmy w ten sposób wyciągnąć liczne wnioski, m. in. że kobiety biegną wolniej oraz, że amatorzy tracą siłę po połowie dystansu i spotykają legendarną ścianę na ostatnim odcinku biegu. Ta kompleksowa analiza przeprowadzona na danych z Hong Kong Marathon 2016 jest również dowodem jak potężnym narzędziem są elementy statystyki opisowej

7 Bibliografia

1. Wyniki maratonu