

Statystyczna analiza danych procentowych na przykładzie wyników produkcyjnych kurcząt brojlerów

Agata Gruzewska

Akademia Podlaska w Siedlcach

W wielu dyscyplinach naukowych wyniki doświadczeń podawane są w postaci frakcji, czyli stosunku dwóch wartości, i na ogół w ujęciu procentowym. Najczęściej ma to miejsce, gdy doświadczenie uwzględnia obiekt kontrolny. Problem pojawia się wówczas, gdy tak przedstawione wyniki doświadczeń są poddawane analizie statystycznej opartej na testach parametrycznych. Można spotkać opracowania, w których do dalszej analizy wykorzystano dane w postaci empirycznej. Nie jest to jednak podejście właściwe wówczas, gdy wartości znajdują się poza zakresem 20-80%, ponieważ nie spełniają podstawowego warunku stosowalności wnioskowania parametrycznego, którym jest rozkład normalny wartości analizowanej zmiennej. Znaczące odchylenia od normalności rozkładu powodują zawyżenie wartości wariancji błędu, zmieniając wnioski formułowane na podstawie wyników analizy wariancji, a to decyduje bezpośrednio o przebiegu i wnioskowaniu płynącym z analizy ilościowej. Z tego powodu, dla danych procentowych w rozumieniu frakcji zaleca się stosowanie odpowiedniej transformacji.

Z taką prezentacją wartości cechy możemy mieć do czynienia w przypadku badań śmiertelności (lub przeżywalności) organizmów żywych. Początkowe opracowania danych liczbowych z tej tematyki dotyczyły zagadnień demograficznych, a pierwsze próby z tego zakresu przypisuje się żyjącemu w XVII wieku Grantowi, który skonstruował tablice śmiertelności [8]. W 1978 r. Malthus sformułował najgłośniejszą doktrynę ludnościową, ale największy wkład do teorii i metody analizy przeżycia wnieśli Cox i Oakes (1984). Aktualnie w analizie przeżycia wykorzystuje się model regresji logistycznej lub szeroki i bardzo zróżnicowany zakres procedur statystycznych, dla których badaną zmienną niezależną jest czas wystąpienia (pojawienia się) określonego wydarzenia typu śmierć, nawrót choroby itp. [13].

Wyniki badań dotyczące śmiertelności (a znacznie rzadziej przeżywalności) kurcząt brojlerów w czasie odchowu są publikowane najczęściej w postaci frakcji, prezentowane w zestawieniach tabelarycznych jako wartości średnie w badanych grupach doświadczalnych i nie poddawane analizie statystycznej. Takie podejście ogranicza możliwości wnioskowa-

nia co do wpływu czynników na badaną cechę. W tym przypadku można zastosować metody wspomnianej analizy przeżycia lub poddać wartości danych liczbowych odpowiedniej transformacji. Obszerny przegląd różnego rodzaju przekształceń przydatnych w transformacji danych procentowych przedstawili Wójcik i wsp. [17].

Tematyka transformacji poruszana była w wielu pracach. O roli transformacji w normalizowaniu rozkładu i stabilizacji wariancji pisał m.in. Bartlett w 1947 roku. Bardzo słynną i często cytowaną jest publikacja Boxa i Coxa z 1964 roku, w której autorzy zaproponowali transformację potęgową [18].

Celem prezentowanych badań był dobór odpowiedniej transformacji (normalizującej rozkład zmiennej oraz błędów losowych) danych procentowych dotyczących śmiertelności i przeżywalności kurcząt brojlerów. Podejście takie wynikało z założenia, że dla każdej z tych cech skuteczniejszy będzie inny typ przekształcenia wartości, zawierających się przecież w dwóch zupełnie odmiennych przedziałach liczbowych, znajdujących się na dwóch końcach skali od 0 do 100%.

Źródłem danych były informacje o śmiertelności, a więc jednocześnie o przeżywalności, zaobserwowanej w czasie odchowu kurcząt brojlerów w dwóch fermach należących do Spółki „Drosed” S.A. w Siedlcach, uzyskane w Dziale Zaopatrzenia Surowcowego Spółki. Analizie poddano wartości obu cech, dotyczące czterech zestawów towarowych w 299 cyklach produkcyjnych, w latach 1995-2002.

Zbadano skuteczność czterech przekształceń zalecanych dla danych procentowych w postaci frakcji [15]:

1. transformacja Bliss'a $y'_i = \arcsin\sqrt{y_i}$

2. transformacja Freemana i Tukey'a

$$y'_i = 0,5 \left[\arcsin \sqrt{\frac{k_i}{k_i + 1}} + \arcsin \sqrt{\frac{k_i + 1}{n_i + 1}} \right]$$

3. $y'_i = \log(y_i - \log(1 - y_i))$

4. $y'_i = \log[1 + 100(1 - y_i)]$

przy założeniu, że wartość $y_i = \frac{k_i}{n}$ jest frakcją, k_i jest liczbą wartości badanej zmiennej należącej do określonej kategorii, n jest liczbą wszystkich wartości badanej zmiennej, a y'_i jest wartością otrzymaną na skutek transformacji.

Skuteczność powyższych przekształceń sprawdzono, badając zgodność rozkładu wartości uzyskanych na drodze transformacji z rozkładem normalnym oraz rozkładu wartości błędów losowych z rozkładem normalnym o parametrach $(0, \sigma^2)$, korzystając z testu zgodności χ^2 Pearsona dostępnego w programie Statistica [14]. W dalszej analizie przyjęto dwuczynnikową klasyfikację krzyżową danych, zgodnie z modelem:

$$y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk}$$

gdzie:

y_{ijk} – wartość cechy;

μ – średnia populacji;

a_i – stały efekt i -tej fermy ($i = 1, 2$);

b_j – stały efekt j -tego zestawu towarowego kurcząt brojlerów ($j = 1, \dots, 4$);

$(ab)_{ij}$ – stały efekt interakcji obu czynników;

e_{ijk} – błąd losowy.

Oszacowania błędów losowych ($\hat{\epsilon}_{ijk}$) dla wartości transformowanych dokonano zgodnie z poniższymi wzorami, wynikającymi z opisanej klasyfikacji:

$$\hat{\epsilon}_{ijk} = y_{ijk} - \mu - \hat{a}_i - \hat{b}_j - (\hat{ab})_{ij}$$

gdzie:

$$\hat{a}_i = \bar{y}_i - \mu;$$

$$\hat{b}_j = \bar{y}_j - \mu;$$

$$(\hat{ab})_{ij} = \bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \mu$$

\bar{y}_i – średnia dla i -tego poziomu czynnika A;

\bar{y}_j – średnia dla j -tego poziomu czynnika B;

\bar{y}_{ij} – średnia dla i -tego poziomu czynnika A oraz j -tego poziomu czynnika B.

Dane uzyskane na drodze transformacji uznanej za skuteczną poddano analizie wariancji zgodnie z podanym modelem. Na podstawie jej wyników przeprowadzono analizę ilościową, wykorzystując test Tukey'a. Ze względu na znaczącą nieortogonalność układu (silnie zróżnicowane liczebności w poszczególnych podklasach) wartości NIR_{α} obliczono odrębnie dla każdej pary porównywanych średnich, korzystając z testu t-Studenta.

Aby dokładniej opisać strukturę badanych zbiorowości liczbowych obliczono dla wartości empirycznych cechy, a następnie dla wartości poddanych transformacji uznanej za skuteczną i dla wartości oszacowanych błędów losowych:

- współczynnik skośności (asymetrii), według wzoru [11]:

$$g = \frac{\bar{u}_3}{S^3}$$

w którym:

S oznacza odchylenie standardowe wartości cechy,

$$\bar{u}_3 = \frac{\sum_{i=1}^n (y_i - \bar{y})^3}{n}$$

jest momentem centralnym rzędu trzeciego,

y_i jest i -tą obserwacją,

\bar{y} oznacza średnią arytmetyczną, a n to liczba wszystkich danych empirycznych;

- współczynnik spłaszczenia (kurtozę) jako [9]:

$$K = \frac{m_4}{S^4} - 3$$

przy czym

$$m_4 = \frac{\sum_{i=1}^n (y_i - \bar{y})^4}{n}$$

a pozostałe symbole jak wyżej.

Wartość współczynnika skośności wyniosła $-1,40$ dla przeżywalności brojlerów oraz $1,18$ dla ich śmiertelności. Rozkłady empiryczne charakteryzowały się więc znaczną asymetrią, ponieważ wartości kurtozy były równe odpowiednio $2,75$ i $2,06$. Rozkład normalny charakteryzują natomiast zerowe wartości obu omawianych współczynników.

Śmiertelność brojlerów wyniosła średnio 6% , zawierając się w przedziale od $2,4\%$ do 20% , a szczegółowe informacje na temat ubytków stwierdzonych w porównywanych fermach, wraz z wartościami utrzymanymi na skutek transformacji, przedstawiono w tabelach 1 i 2.

Porównanie wartości średnich z odpowiadającymi im wartościami odchyłeń standardowych pozwala stwierdzić, że większą zmiennością charakteryzowała się śmiertelność brojlerów (w porównaniu z przeżywalnością), bowiem współczyn-

Tabela 1

Śmiertelność oraz przeżywalność brojlerów w fermie Chotycze – średnie dla wartości empirycznych i transformowanych z uwzględnieniem zestawu towarowego

Zestaw towarowy	Statystyki	Śmiertelność transformacja		Przeżywalność transformacja	
		%	Freemana i Tukey'a	%	$y'_i = \log(y_i) - \log(1 - y_i)$
L1	\bar{y}	5,6	0,69	94,4	1,28
	S_y	2,9	0,08	2,9	0,25
	n	83		83	
L2	\bar{y}	4,7	0,68	95,4	1,36
	S_y	2,4	0,06	2,4	0,21
	n	95		95	
L3	\bar{y}	6,5	0,73	93,5	1,19
	S_y	2,6	0,06	2,6	0,18
	n	45		45	
L4	\bar{y}	6,3	0,72	93,7	1,21
	S_y	3,7	0,06	3,7	0,19
	n	76		76	

Tabela 2

Śmiertelność oraz przeżywalność brojlerów w fermie Dobrzyniec – średnie dla wartości empirycznych i transformowanych z uwzględnieniem zestawu towarowego

Zestaw towarowy	Statystyki	Śmiertelność transformacja		Przeżywalność transformacja	
		%	Freemana i Tukey'a	%	$y'_i = \log(y_i) - \log(1 - y_i)$
L1	\bar{y}	7,8	0,74	92,2	1,13
	S_y	4,3	0,08	4,3	0,24
	n	79		79	
L2	\bar{y}	4,4	0,66	95,6	1,39
	S_y	2,0	0,07	2,0	0,22
	n	94		94	
L3	\bar{y}	7,7	0,75	92,3	1,11
	S_y	2,7	0,06	2,7	0,17
	n	60		60	
L4	\bar{y}	6,9	0,74	93,1	1,16
	S_y	2,3	0,05	2,3	0,16
	n	34		34	

nik zmienności zawierał się w przedziale od 33% do 58% . Zastosowana transformacja Freemana i Tukey'a znacząco zmniejszyła jego wartość (7% do 12%) i jest to pierwsza korzyść płynąca z przekształcenia danych empirycznych.

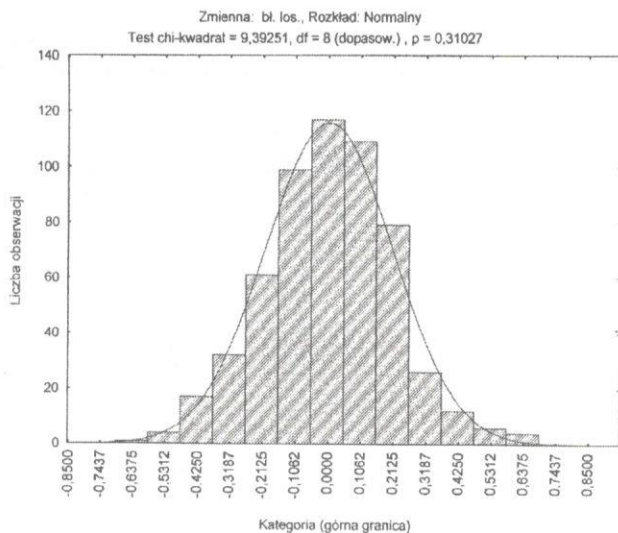
Rozkłady empiryczne obu cech istotnie różniły się od rozkładu normalnego. Wartości empiryczne statystyki χ^2 ($100,83$ dla śmiertelności i $115,91$ dla przeżywalności) znacząco przewyższyły wartość tabelaryczną $\chi^2_{0,05} = 11,070$. Poszukiwanie odpowiedniej transformacji było więc uzasadnione.

Ze względu na zgodność rozkładu empirycznego z rozkładem normalnym, skuteczne okazały się:

- transformacja Freemana i Tukey'a dla śmiertelności brojlerów

$$(\chi_0^2 = 3,75 < \chi_{0,05}^2 = 9,49);$$

- transformacja logarymiczna $y'_i = \log(y_i) - \log(1 - y_i)$ dla ich przeżywalności ($\chi_0^2 = 6,78 < \chi_{0,05}^2 = 9,49$).



Rys. Rozkład błędów losowych oszacowanych dla przeżywalności poddanej transformacji logarytmicznej

Wartości współczynnika skośności dla danych transformowanych wynosiły odpowiednio: $-0,15$ i $0,14$, a wartości kurtozy: $0,30$ i $0,48$, czyli blisko wartości zero.

Kolejnym etapem wyboru najskuteczniejszej transformacji było zbadanie rozkładu błędów losowych, oszacowanych dla obu omawianych przekształceń. Okazało się, że zgodność rozkładu empirycznego z rozkładem normalnym o parametrach $(0, \sigma^2)$ uzyskano jedynie dla wartości błędów losowych, oszacowanych z wartości przeżywalności brojlerów poddanych transformacji logarytmicznej (rys.). Współczynnik skośności wyniósł w tym przypadku $0,04$, a kurtoza $0,32$.

Dwuczynnikowa analiza wariancji wartości transformowanych (zgodnie z modelem omówionym w metodyce) wykazała istotny wpływ obu badanych czynników oraz istotność ich interakcji. Przeprowadzono więc analizę ilościową, porównując średnie w interakcji ferma x zestaw towarowy (tab. 3). Szczegółowe porównanie średnich wartości transformowanych wskazało na zbliżony układ grup jednorodnych w porównywanych fermach. Istotnie najwyższą przeżywalnością charakteryzowały się kurczęta zestawu L2, a średnie wartości przeżywalności ptaków należących do zestawów L3 i L4 tworzyły grupę jednorodną, do której w fermie Dobrzyniec należała również średnia wartość badanej cechy ptaków zestawu L1.

Tabela 3
Średnie przeżywalności dla wartości transformowanych ($y' = \log(y_i) - \log(1 - y_i)$) z uwzględnieniem fermy oraz zestawu towarowego

Zestaw towarowy	Ferma	
	Chotycze	Dobrzyniec
L1	1,28 ^a	1,13 ^a
L2	1,36 ^b	1,39 ^b
L3	1,19 ^c	1,11 ^a
L4	1,21 ^c	1,16 ^a

a, b, c – wartości w wierszach oznaczone różnymi literami różnią się istotnie ($P \leq 0,05$)

Tabela 4

Wartości $NIR_{0,05}$ w porównaniu zestawów towarowych, obliczone dla poszczególnych par porównywanych średnich przeżywalności brojlerów z uwzględnieniem fermy

Zestaw towarowy	Ferma	
	Chotycze	Dobrzyniec
L1 i L2	0,062	0,063
L1 i L3	0,077	0,071
L1 i L4	0,066	0,085
L2 i L3	0,075	0,069
L2 i L4	0,064	0,083
L3 i L4	0,078	0,089

Podsumowanie

Wielkość ubytków stwierdzonych w badanych fermach znacząco przewyższała wyniki podawane przez innych autorów. W warunkach doświadczalnych notowano śmiertelność od $3,8\%$ do $5,2\%$ [12] lub nieco wyższą: $1,7-8,3\%$ [3]. Według Koreleskiego i wsp. [7] istotny wpływ na przeżywalność kurcząt miał sposób ich żywienia w pierwszych dniach odchovu.

Rozkłady empiryczne obu cech charakteryzowały się znaczącą asymetrią oraz znaczącą wartością kurtozy – przekraczającą liczbę dwa, świadcząca o większej niż w przypadku rozkładu normalnego „smukłości” omawianych rozkładów empirycznych.

Analiza zmienności badanego materiału liczbowego wyraźnie wskazała, że ze względu na to kryterium wyboru, śmiertelność ptaków była cechą właściwszą do wykorzystania w analizie statystycznej, po przekształceniu wartości empirycznych zgodnie z transformacją Freemana i Tukey'a. Jednak wartości te nie spełniły kryterium normalności rozkładu błędów losowych, w przeciwieństwie do wartości błędów losowych oszacowanych dla przeżywalności poddanej transformacji logarytmicznej.

Wielu autorów wskazuje na przydatność transformacji logarytmicznej. Okazała się ona skuteczna m.in. w badaniach występowania zarodników różnych gatunków grzybów w glebie [1], liczby komórek somatycznych mleka [4], precyzji doświadczeń polowych [10], agresywności pszczoł [2, 5], nieśności kur [6] oraz w ultrasonograficznych badaniach tkanki tłuszczowej owiec [16].

Przedstawione wyniki badań można podsumować następująco:

- skuteczną transformacją w badaniach przeżywalności kurcząt brojlerów okazała się transformacja logarytmiczna $y'_i = \log(y_i) - \log(1 - y_i)$;
- w badaniach śmiertelności kurcząt brojlerów nie znaleziono funkcji, którą można uznać za skuteczną w transformacji wartości empirycznych tej cechy.

Literatura: 1. Błaszowski J., 1995 – Acta Mycologia, Vol. 30, 2, 213-222. 2. DeGrandi-Hoffman G., Collins A., Martin J.H., Schmidt J.O., Spangler H.G., 1998 – J. Ins. Behav., Vol. 11, 1, 37-45. 3. Górski J., Majczyna S., Koncerewicz A., Olejnik A., 1997 – Zesz. Nauk. Przeglądu Hodowlanego 32, 177-233. 4. Guliński P., Gruzewska A., 1996 – Roczn. Zoot., T. 23, 4, 75-83. 5. Guzman-Novoa E., Page R.E., 1994 – Behav. Ecol., Vol. 5, 1, 91-96. 6. Koerhuis

A.N.M., McKay J.C., 1996 – *Livest. Prod. Sci.*, Vol. 46, 117-127. 7. Koreleski J., Świątkiewicz S., Kubicz M., 1997 – *Zesz. Nauk. Przeglądu Hodowlanego* 32, 235-241. 8. Oktaba W., 1998 – *Sylwetki odkrywców praw przyrody*. Lubelskie Towarzystwo Naukowe, Lublin. 9. Ostasiewicz S., Rusnak Z., Siedlecka U., 1999 – *Statystyka. Elementy teorii i zadania*. Wyd. Akademii Ekonomicznej im. Oskara Langego, Wrocław. 10. Pilarczyk W., 1987 – *Biuletyn Oceny Odmian*, T. XII, (1-2), 161-169. 11. Platt C., 1974 – *Problemy Rachunku Prawdopodobieństwa i Statystyki Matematycznej*. PWN, Warszawa. 12. Sokołowicz Z., Ruda M., 2000 – *Zesz. Nauk. Przeglądu Hodowlanego* 49, 485-491. 13. Stanisław A., 2000 – *Przystępny kurs statystyki*

z wykorzystaniem programu Statistica PI na przykładach z medycyny. T. II, StatSoft, Kraków. 14. StatSoft, Inc., 2001 – *Statistica (data analysis software system)*, version 6. www.statsoft.com. 15. Trętowski J., Wójcik A.R., 1988 – *Metodyka doświadczeń rolniczych*. Wyd. WSRP w Siedlcach. 16. Van Heelsum A.M., Levis R.M., Haresign W., Williams S.P., Davies M.H., 2001 – *Livest. Prod. Sci.* 69, 113-127. 17. Wójcik A.R., Gabriel W., Woźnica W., 1976 – *Biuletyn Instytutu Ziemniaka* 17, 83-93. 18. Zieliński A., 2002 – *Teoria i praktyka stosowania transformacji potęgowej w układach niezależnych i układach zależnych typu split-plot*. Rozprawy Naukowe i Monografie, SGGW, Warszawa.

Grypa ptasia – nowe dane epidemiologiczne, możliwości leczenia (cz. I)

Antoni J. Furowicz,
Danuta Czernomysy-Furowicz,
Magdalena Ferlas, Jacek Borkowski

AR w Szczecinie

W artykule zostaną omówione nowe aspekty epidemiologiczne grypy (H5N1) oraz scharakteryzowane leki hamujące rozwój wirusa grypy u zakażonych ludzi i zwierząt, w tym preparaty pochodzenia roślinnego, które mogą być zastosowane także profilaktycznie. W związku z zagrożeniem epidemią grypy ptasiej populacji ludzkiej oraz innych ssaków, jak również coraz częstszymi masowymi zachorowaniami ptaków, istnieje pilne zapotrzebowanie na takie preparaty. Zasadnicze informacje dotyczące wymienionej formy grypy przedstawiono we wcześniejszych opracowaniach [6, 7]. W związku ze znaczną liczbą nowych danych, których autorzy starali się uporządkować szereg faktów i hipotez, zostaną one krótko omówione.

Epidemiologia i klinika ptasiej grypy – dane „na dzień dzisiejszy”

Występowanie pandemii i większych epidemii grypy w latach 1918-2005 przedstawiono w tabeli 1, natomiast poszczególne etapy rozwoju grypy ptasiej (H5N1) w 2005 roku zaprezentowano w tabeli 2. Choroba ta pojawiła się u ptactwa w Azji w roku 2003; jej pierwsze przypadki u ludzi odnotowano w styczniu roku 2004 w Wietnamie i Tajlandii [8]. W sumie zachorowało około 200 osób, zmarło 70, w tym małe dzieci (tab. 2).

Wirus powoduje epidemie zarówno wśród ptaków hodowlanych (brojlery), jak i wśród ptactwa dzikiego (gatunków wędrownych) związanego z akwenami wodnymi. Choroba ma z reguły ostry przebieg, odnotowuje się wiele zejść śmiertelnych. Istnieje możliwość bezobjawowego nosicielstwa [8]. Odzwierzęce zakażenia ludzi, z reguły związanych z chowem drobiu stwierdza się stosunkowo rzadko. Niekiedy infekcje przypominają małe epidemie rodzinne [7, 8]. Przebieg choroby u człowieka jest bardzo ostry, z wyjątkowo wysoką śmiertelnością. Dotyczy to zwłaszcza małych dzieci. Transmisja wirusa między ludźmi zachodzi bardzo rzadko, jednak zakażenia rodzinne oraz wśród personelu medycznego sprawującego opiekę nad ludźmi chorymi w krajach azjatyckich wskazują, że do takich zakażeń może dochodzić [8]. Wirus H5N1, podobnie jak inne wirusy grypy występujące u drobiu [9, 14, 16], może przenosić się także na inne ssaki (świnie, dzikie kotowate oraz niekiedy walenie i foki). W początkowej fazie choroby, w sposób zbliżony jak ma to miejsce w grypie „klasycznej”, dochodzi szybko do replikacji wirusa (rozmnażania), co aktywuje niespecyficzne mechanizmy obronne organizmu związane z komórkami linii monocytarno-makrofagowej oraz z komórkami NK („naturalnymi zabójcami”). Komórki NK odgrywają zasadniczą rolę w zakażeniach wirusowych; poprzez określone cytokiny pobudzają inne komórki układu odpornościowego odpowiedzialne za mechanizmy obrony swojej, głównie limfocyty T (CD8) cytotoksyczne, znajdujące się w regionalnych węzłach chłonnych, a następnie limfocyty B, co powoduje ograniczenie wirusii [13, 17, 18]. Jednak przy zakażeniu wirusem H5N1 dochodzi często do nadmiernego pobudzenia wymienionych komórek; duża ilość cytokin („kaskada cytokinowa”) powoduje zaostrzenie kursu choroby. Odnotowano, że poza układem oddechowym wirus atakuje także

Tabela 1
Pandemie i większe epidemie grypy w latach 1918-2005, wg Gibbisa i Soares [8], zmodyf. i uzupełnione [7]

1.	Pandemia 1918-1920*, wirus H1N1; około 50 milionów zejść śmiertelnych na całym świecie w wyniku bardzo ostrego przebiegu choroby
2.	Pandemia 1957, wirus H3N2; 1-4 milionów zgonów na świecie**
3.	Pandemia 1968, wirus H3N2; około 1 miliona zejść śmiertelnych na świecie
4.	Rok 2003, epidemia różnych gatunków ptaków hodowlanych i dzikich (w tym wędrownych, związanych z akwenami wodnymi) oraz w roku 2004 pojedyncze zachorowania ludzi (głównie zakażenia odzwierzęce i "epidemie rodzinne"); do chwili obecnej około 200 zachorowań z wysoką śmiertelnością (55-60%). Zachorowania głównie na terenie Azji, wirus H5N1 typowy dla grypy ptaków
5.	Epidemia 2003, wirus H7N7; epicentrum zakażeń w Holandii; chorowało około tysiąca mieszkańców tego kraju

*Według niektórych autorów okresem zakażeń o charakterze pandemii były lata 1918-1919, w 1920 roku zachorowania miały już charakter ograniczony; **Różnice związane z różnymi źródłami informacji