1. Jaka jest różnica między cechą skokową i ciągłą? – podać przykłady każdej z nich.
Cecha ilościowa : skokowa – przyjmująca pewne wartości liczbowe i nie przyjmująca wartości pośrednich cecha ta też jest nazywana dyskretną, przykład: ilość bakterii, pracowników, pasażerów. **ciągła** – przyjmująca wartości z pewnego przedziału liczbowego przykład: wzrost, waga, plon.
2. Wymienić typy cech i podać po jednym przykładzie.
Cechy jakościowe (opisowe, niemierzalne) przyjmujące wartości nie będące liczbami, np.: kolor włosów, płeć, smakowitość, pochodzenie społeczne.
Cechy ilościowe (mierzalne): np.: wzrost (w centymetrach), wiek (w latach), zarobek (w złotówkach)
Cechy skokowe : np.: liczba studentów w grupie
Cechy ciągłe : np.: waga
3. Podać przynajmniej trzy nazwy rozkładów cech i jakiego typu są to cechy.
Rozkłady cech skokowych:
1. Rozkład zero – jedynkowy
2. Rozkład dwumianowy (Bernoulliego)
3. Rozkład Poissona

Rozkłady cech ciągłych:
4. Rozkład normalny jedno i dwu wymiarowy
5. Rozkład jednostajny

4. Podać znane nazwy rozkładu cech i jakiego typu są to cechy.
Rozkład zero-jedynkowy: Podstawą do określania rozkładu zero-jedynkowego jest doświadczenie, którego rezultatem mogą być dwa wzajemnie wykluczające się zdarzenia losowe. Oznaczyć je możemy jako A i zdarzenie przeciwne A' np. strzelając do celu trafiamy (A=1) lub nie (A'=0). Zmienna losowa X ma taki rozkład, jeśli przyjmuje wartość A z prawdopodobieństwem $0<p<1$ oraz wartość A' z prawdopodobieństwem $q = 1-p$. Funkcja prawdopodobieństwa zmiennej losowej ma postać: $P(X=1) = p$, $P(X=0) = p-1$; $p\in(0,1)$. Dystrybuanta zmiennej losowej $F(X) = \{0$, dla $X<0$; $1-p$, dla $0 =< X <1$; 1, dla $X >=1\}$. Wartość oczekiwana $E(X) = 0(1-p) + (1p) = p$. Wariancja $D^2(X) = (0-p)^2(1-p)+(1-p)^2p=p(1-p)$.
Rozkład dwumianowy : Wykonujemy doświadczenie, którego rezultatem może być zdarzenie A z $P(A)=p$ lub A' z $P(A')=1-p$. Jedno z nich przujmuje się za sukces drugie jako porażkę. Liczbę sukcesów zaobserwowanych w "n" próbach może być równa k=1,2,3,...,n. Zdarzenie X=k zachodzi, gdy w wyniku n-krotnego powtarzania doświadczenia zaobserwujemy k-razy zdarzenie A (więc n-k razy zdarzenie A'). Prawdopodobieństwo otrzymania k sukcesów w doświadczeniu powtarzanym n razy (suma prawdopodobieństw takich kombinacji, że występuje k razy A):

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Zmienna losowa X ma taki rozkład, jeśli przyjmuje wartości k=0,1,2,...,n z prawdopodobieństwami określonymi wzorem:

$$F(x) = P(X \le x) = \sum_{k \le x} \binom{n}{k} p^k (1-p)^{n-k}.$$

Wartość oczekiwana $E(X)=np$ -suma wartości oczekiwanych niezależnych zmiennych losowych o rozkładach zerojedynkowych (pojedyncze doświadczenia), $D^2(X)=np(1-p)$
Rozkład Poissona : zmienna losowa X przyjmująca wartości k = 0,1,2,... ma taki rozkład o parametrze λ, jeśli jej funkcja prawdopodobieństwa opisana jest wzorem: $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ k = 0,1,2,..., gdzie λ jest dodatnią stałą ($\lambda > 0$). Dystrybuantę rozkładu Poissona określa wzór: $F(x) = \sum_{k \le x} \frac{\lambda^k}{k!} e^{-\lambda}$. Opierając się na definicji wartości oczekiwanej i wariancji zmiennej losowej skokowej, dla rozkładu Poissona otrzymujemy: $E(X)= \lambda$, $D^2(X)= \lambda$.
Rozkład normalny: Zmienna losowa ma rozkład normalny N-(μ,σ^2) o wartości średniaj μ i wariancji σ^2, jeżeli jej funkcja gęstości wyraża się wzorem (w pytaniu 14).
5. Podać dwa przykłady cech w rozkładzie dwumianowym.
- 5 prób trafienia w tarczę
- 10 prób wyciągnięcia czarnej kuli z urny zawierającej kule czarne i białe (ze zwracaniem)

6. Podać dwa przykłady cech w rozkładzie normalnym.
- Waga oraz wzrost osobników jednorodnych populacji ludzkich lub zwierzęcych.
- Plon na jednakowych poletkach doświadczalnych.
- Wynik osiągany w biegu na 100m

7. Podać dwa przykłady cech w rozkładzie Poissona.
- Liczba usterek w produkowanych urządzeniach
- Liczba skaz na określonej powierzchni materiału
- Liczba błędów drukarskich na jednej stronie.

8. Zmienna losowa X ma rozkład N(10,25). Obliczyć P{|X-10|=<10}
Cecha X-(μ,σ^2) ma rozkkład normalny N-(μ,σ^2). Z prawa trzech sigm:

$P\{|X-\mu|<\sigma\}=0,68$
$P\{|X-\mu|<2\sigma\}=0,95$
$P\{|X-\mu|<3\sigma\}=0,997$
X-N(10,25); $\mu=10$, $\sigma=5$ z prawa trzech sigm:
$P\{|X-10|=<10\}= P\{|X-\mu|<2\sigma\}=0,95$

9. Zmienna losowa X ma rozkład N(10,25). Obliczyć P{|X-10|=<5}

N(10,25), $\mu=10$, $\sigma=5$
$P\{|X-10|=<5\}= P\{|X-\mu|<\sigma\}=0,68$

10. X ~ N(100,100). Ile wynosi P{X ϵ(90,110)}?

Dystrybuanta F(X) dla standardowego rozkładu jest stablicowana. Dla x=<0 zachodzi F(x)=1-F(-x). **Standaryzacja**. Jeżeli X-
N(μ ,σ^2), to Z=(X- μ)/ σ-N(0,1)

$P\{X \in \left(\frac{a-\mu}{\delta},\frac{b-\mu}{\delta}\right)\} = F\left(\frac{b-\mu}{\delta}\right) - F\left(\frac{a-\mu}{\delta}\right)$.Podstawiając $\mu=100$, $\sigma=10$, a=90, b=110 otrzymujemy

$P\{X\in(-1,1)\}=F(1)-F(-1)=F(1)-1+F(1)=2F(1)-1=2(0,84134)-1=0,68=68\%$

11. X ~ N(120,64). Ile wynosi P{X ϵ(104,136)}?

N(120,64); Podstawiając $\mu=120$, $\sigma=8$, a=104, b=136 do wzoru z pyt. 10 otrzymujemy:
$P\{X\in(104,136)\}= P\{X\in(-2,2)\}=F(+2)-F(-2)=F(2)-1+F(2)=2F(2)-1=2x0,97725-1=0,9545=95\%$

12. Cecha X ma rozkład N(12,16). Bez użycia tablic obliczyć P{X ϵ (8,16)}?

$\mu=12$, $\sigma=4$, $P\{X\in(8,16)\}=P\{|X-12|=<4\}=P\{|X-\mu|=<\sigma\}=68\%$

13. Cecha X ma rozkład N(12,16). Bez użycia tablic obliczyć P{X ϵ (4,20)}?

$\mu=12$, $\sigma=4$, $P\{X\in(8,16)\}=P\{|X-12|=<8\}=P\{|X-\mu|=<\sigma\}=96\%$

14. W jaki sposób można sprawdzić założenie o normalności.

Zmienna losowa X ma rozkład normalny, jeżlei jej funkcja gęstości wyraża się wzorem:

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, -\infty < x < \infty$$

15. W jakim celu stosuje się w praktyce uśrednianie wartości pewnej cechy.

Dzięki średniej możemy sprawdzić, czy dana wartość cechy jest względnie większa czy mniejsza niż w reszcie populacji tzn. Jeżeli jakaś wartość jest powyżej średniej to jest mniej wartości większych w populacji, a więcej mniejszych. Średnia pozwala także przewidzieć najbardziej prawdopodabny wynik np. jeśli średnia ilość trafień na 10 wynosi 3, to gdy szacujemy ile będzie trafień, najbardziej prawdopodobną liczbą trafień jest 3.

16. Wymienić rozkłady pojawiające się we wnioskowaniu statystycznym, a związane z rozkładem normalnym.

- Rozkład Piscona
- Rozkład Chi – kwadrat
- Rozkład T - Studenta

17. Co to jest populacja?

Populacja – zbiór obiektów (fizycznych i nie tylko) z wyróżnioną cechą (-ami). Jeśli zbiór elementów populacji jest skończony to określamy ją jako **skończoną** np. zbiorowość mieszkańców Polski, zbiorowość gospodarstw rolnych w danym województwie. Jeśli zbiór elementów populacji jest nieskończony to określamy ją jako **nieskończoną** – dotyczy raczej zjawisk niż obiektów materialnych np. zbiorowość rzutów monetą, zbiorowość możliwych wyników pomiaru wytrzymałości materiału.

18. Co to jest próba reprezentatywna?

Próba – wybrana część populacji podlegająca badaniu (próba), jest reprezentatywna, gdy jej struktura ze względu na interesujące nas cechy statystyczne jest zbliżona do struktury populacji z której ona pochodzi, czyli wnioski wyciągnięte z próby można uogólnić na całą populcje. Próba jest reprezentatywna gdy spełnione są warunki:

- Elementy populacji są pobierane do próby w sposób losowy.
- Próba jest dostatecznie liczna.

19. Co to jest wnioskowanie statystyczne?

Wnioskowanie statystyczne – to możliwość uogólnienia uzyskanych wyników na całą populację elementów oraz oszacowanie wielkości popełnionych przy tym błędów. Wynik wnioskowania musi być użyteczny.

20. Jakie są podstawowe różnice między populacją i próbą?

Próba jest wybraną częścią populacji, na podstawie jej danych wnioskujemy o populacji, czyli próba pozwala scharakteryzować populację, np.: Spośród wszystkich kobiet w Warszawie (Populacja) losujemy jakąś część (Próba) i na tej podstawie charakteryzujemy średni wzrost kobiet w Warszawie.

21. Podać przykład próbki niereprezentatywnej dla oszacowania zróżnicowania zarobków w Polsce?

Próbę przeprowadzamy wśród rolników.

22. Podać przykład próbki niereprezentatywnej dla oszacowania średnich zarobków ludzi w Polsce?

Próbę przeprowadzamy wśród ludności W-wy i ustalamy

23. Podać przykład próbki niereprezentatywnej dla wzrostu wszystkich kobiet w Polsce.
Próbę przeprowadzamy wśród zawodniczek drużyny koszykarskiej.
24. Co wpływa na jakość wnioskowania statystycznego.
Na jakość wnioskowania statystycznego wpływa:

- estymacja (szacowanie) nieznanych wartości parametrów rozkładu cechy w populacji.
- słuszność hipotez dotyczących albo wartości parametrów rozkładu cechy w populacji albo postaci tego rozkładu.
- jakość próby: liczność, losowy wybór.

25 i 26. Jakie są źródła błędów we wnioskowaniu statystycznym? Podać przynajmniej dwa źródła błędów we wnioskowaniu statystycznym.
Źródła błędów: nieliczne lub nielosowo wybrane elementy próby wybór złego rozkładu cechy w populacji,
Estymacja : Z uwagi na to że estymacji pewnego parametru za pomocą określonego jego estymatora Z_n dokonujemy na podstawie wyników próby losowej, istnieje możliwość popełnienia błędu. W celu uzyskania małego błędu estymacji należy dbać o prawidłowe losowanie próby, jak i dobór możliwie najlepszego estymatora Z_n.
W tym celu wprowadza się pewne własności, które powinien posiadać dobry estymator: zgodność, efektywność, dostateczność i nieobciążoność
Testowanie hipotez statystycznych : Z uwagi na to, że testowanie hipotez statystycznych opiera się na wynikach próby losowej, podjęta w wyniku zastosowania danego testu decyzja o przyjęciu lub odrzuceniu hipotezy nie zawsze jest bezbłędna (występują błędy I i II stopnia).
27. Co to jest estymator?
Estymator jest narzędziem wnioskowania statystycznego. Estymator jest to funkcja wyników z próby, czyli statystyka służąca do oszacowania nieznanej wartości parametru populacji. Wartość estymatora z konkretnej próby jest liczbą zwaną *oceną parametru*. Estymatorem może być zatem każda wielkość otrzymana dla wyników próby, czyli: średnia arytmetyczna, dominanta, kolejne kwartyle, rozstęp, odchylenie standardowe i wiele innych. Estymator jako funkcja wyników próby losowej, będących zmiennymi losowymi, jest zmienną losową. Rozkład prawdopodobieństwa estymatora zależy od rozkładu populacji i od sposobu losowania próby (schemat losowania). Szczególnie ważne są dwa parametry rozkładu: a)wartość oczekiwana (momenty), b)wariancja. Jest wiele metod znajdowania estymatora. Najczęściej stosowane to: a)metoda momentów, b)metoda największej wiarygodności, c)metoda kwadratów. Mówimy, że estymator T_n parametru O jest *nieobciążony* gdy spełniona jest relacja: $E(T_n)=O$. Inaczej estymator T_n jest obciążony, a parametr $E(T_n)-O=b(T_n)$ nazywamy obciążeniem estymatora. *Asymptotyczny nieobciążony* tzn. Lim(n->8) $b(T_n)=0$. *Zgodny* spełnia relacje Lim(n->8) $P\{|T_n-O|<\varepsilon\}=1$, dla dowolnego $\varepsilon>0$.
28. Co znaczy, że estymator jest precyzyjny?
Przy wzrastającej do nieskończoności liczebności próby wariancji $D^2(Z_n)$ estymatora Z_n przyjmuje wartości coraz bliższe wariancji najefektywniejszego estymatora. Odwrotność wariancji estymatora nosi nazwę *precyzji*. Estymator najefektywniejszy to taki, który ma największą precyzję.
29. Podać przynajmniej dwa różne oszacowania średniej wartości cechy.
Na podstawie próby 1.1, 1.2, 0.8, 0.9, 1.2, 1.3, 1.0, 0.7, 0.8, 1.0 oszacować wartość średnią rozkładu obserwowanej cechy.
$X_{sr}=(1.1+...+1.0)/10=1$
var $x = (1.1 - 1.0)^2 + ... (1.0 - 1.0)^2 = 0.36$
(suma kwadratów odchyleń)
$s^2 = 0.36/10-1 = 0.04$
$s = 0.2$
poziom ufności $1-\alpha = 0{,}95$, czyli $\alpha = 0.05 = 5\%$
$t(0{,}05 \;\; , 9) = 2{,}2622$
$t(0.05 ,9) * s/\sqrt{n} = 2{,}622 * 0{,}2/\sqrt{10} = 0{,}14$
$1 - 0{,}14 = 0{,}86$
$1+ 0{,}14 = 1{,}14$
<u>ODPOWIEDŹ</u> Średnia wartość cechy jest jakąś liczbą z przedziału (0,86; 1,14)
30. Co to jest przedział ufności.
PRZEDZIAŁ UFNOŚCI – jest przedziałem o końcach zależnych od próby, który z pewnym z góry zadanym prawdopodobieństwem pokrywa nieznaną wartość parametru Õ
$P\{(\tilde{O} \in (O(\underline{x}_1 ,..........x_n), \;\; \tilde{O} \;\; (x_1 ,..........x_n)\} = 1 - \alpha$ (Poziom ufności)
W wyniku pobrania próby losowej z populacji i obliczenia na tej podstawie wartości estymatora szacowanego parametru uzyskuje się tzw. punktową ocenę parametru. Prawdopodobieństwo że estymator przyjmuje wartość równą wartości szacowanego parametru jest równa 0. Oznacza to że przy estymacji punktowej z prawdopodobieństwem równym jeden popełniamy błąd. Jest to jeden ze sposobów dla których stosuje się estymację przedziałową, polegającą na tym, że zamiast jednej oceny wartości parametru podaje się pewien przedział, który z określonym z góry prawdopodobieństwem (>0) pokrywa nieznaną wartość szacowanego parametru.
31. Co to jest poziom ufności.
Jest to prawdopodobieństwo mające opisać nasze przekonanie co do trafności oceny, oznaczone przez $1-\alpha$
32. Jaka jest interpretacja poziomu ufności.
Poziom ufności $1-\alpha$ jest zaufaniem do wystawionych wniosków.

33 i 34. Od jakich czynników zależy długość przedziału ufnośći?
Na długość przedziału wpływa:
1. liczebność próby – gdy zwiększymy ilość obserwacji (rośnie n), to zwiększa się precyzja oceny, co wyraża się skróceniem przedziału. Prowadzący może mieć wpływ na długość przedziału ufności, ponieważ to on decyduje o ilości obserwacji.
2. poziom ufności – aby zwiększyć precyzję oszacowania należy zmniejszyć poziom ufności bowiem nastąpi skrócenie długości przedziału. Aby zwiększyć dokładność należy zwiększyć współczynnik ufności co spowoduje rozszerzenie przedziału.
3. wariancja cechy - im większa tym większy przedział

35. Na podstawie badań uzyskano dla średniej następujący przedział ufności (2,13). Czy można uznać, że średnia w populacji jest równa 7 i dlaczego?
Ponieważ 7 należy do przedziału ufności może być średnią populacji(tak jak wszystkie liczby z tego przedziału), przy czym zaufanie do tego wniosku wynosi 1-α.

36. Uzyskano 95% przedział ufności dla różnicy średnich : (1.23;7.9). Czy na tej podstawie można uznać, że badane średnie nie różnią się?

42. Co to jest hipoteza statystyczna.
Hipotezą statystyczną nazywamy dowolne przypuszczenie dotyczące rozkładu prawdopodobieństwa cechy. Hipotezy statystyczne są formalnym zapisem przypuszczeń merytorycznych sformułowanych w trakcie rozwiązywania problemów naukowych i praktycznych. Testowaną hipotezę statystyczną oznacza się symbolem H_0 i nazywa się hipotezą zerową. Obserwujemy cechę X w pewnej populacji. Hipoteza – to przypuszczenie dotyczące rozkładu prawdopodobieństwa tej cechy. Prawdziwość tego przypuszczenia jest oceniana na podstawie wyników próby losowej. Jest to każdy sąd (przypuszczenie) dotyczące populacji wydany bez przeprowadzenia badania wyczerpującego.

43. Przykłady hipotezy statystycznej i podaj przykład hipotezy niestatystycznej.
1.Hipoteza $H_0 : \mu = 250$, Hipoteza ta orzeka, że średnia wartość cechy w populacji wynosi 250.
2.Hipoteza niestatystyczna „w roku 2010 będzie klęska żywiołowa” – nie ma mowy o postaci rozkładu i jego parametrach.

44. Co to jest błąd pierwszego rodzaju.
Błędem I rodzaju - błąd we wnioskowaniu polegający na odrzuceniu hipotezy, gdy w rzeczywistości jest ona prawdziwa.

45. Co to jest poziom istotności.
Poziomem istotności Jest to prawdopodobieństwo popełnienia błędu I rodzaju (2). Najczęściej przyjmowanymi poziomami istotności są: 0,1; 0,05; 0,01; 0,001.

46. Interpretacja poziomu istotności. (odp. W 45)

47. Co to jest błąd drugiego rodzaju.
Błędem II rodzaju - błąd we wnioskowaniu polegający na nie odrzuceniu hipotezy, gdy w rzeczywistości jest ona fałszywa.

48. Co to jest moc testu.
Mocą testu nazywamy prawdopodobieństwo nieodrzucenia hipotezy nieprawdziwej
Moc testu = prawdopodobieństwo popełnienia błędu II rodzaju

49. Zinterpretować wniosek: odrzucono weryfikowaną hipotezę na poziomie istotności 0,05.
Na 95% była fałszywa i na 5% była prawdziwa.

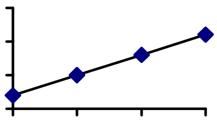
50. Co mierzy współczynnik korelacji.
Współczynnik korelacji jest miernikiem siły zależności między badanymi zmiennymi. Przyjmuje wartości < -1; 1 >.

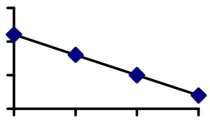
51. Interpretacja współczynnika korelacji.
Współczynnik korelacji jest liczbą niemianowaną, należy do przedziału < -1; 1 >.
Interpretujemy dwa elementy współczynnika korelacji:
1. znak współczynnika korelacji;
2. wartość współczynnika korelacji;

Jeżeli chodzi o znak to:
- jeżeli współczynnik korelacji > 0, to większym wartościom jednej cechy odpowiadają większe wartości drugiej cechy; jest to zależność dodatnia (rosnąca, stymulująca);
- jeżeli współczynnik korelacji < 0, to większym wartościom jednej cechy odpowiadają mniejsze wartości drugiej cechy; jest to zależność ujemna (malejąca, limitująca);
- jeżeli współczynnik korelacji = 0, to bez względu na wartość przyjmowane przez jedna z cech, średnia wartość drugiej cechy jest taka sama; są to cechy nieskolerowane

Jeżeli g= $\pm$1 , to istnieją takie liczby a i b, że Y = aX + b – zależność między cechami jest ściśle liniowa.
Jeżeli g= 1, to a > 0, oraz jeżeli g = -1 to a <0.
W związku z tym współczynnik korelacji traktowany jest jako miernik liniowej zależności między cechami X oraz Y. Wartość współczynnika korelacji interpretowana jest ; że im |g| jest bliższe 1, tym bardziej liniowa jest zależność między cechami. Korelację między X i Y obliczamy ze wzoru $r = \frac{COV(X,Y)}{\sqrt{\operatorname{var} X}\sqrt{\operatorname{var} Y}}$, gdzie COV(X,Y) to kawariancja- suma iloczynów odchyleń od średniej.

52. Jakie wartości może przyjmować współczynnik korelacji.
Współczynnik korelacji przyjmuje wartości z przedziału < -1; 1 >
Im korelacja jest silniejsza (bliższe jedynki), tym linie regresji są położone bliżej siebie.
r=1 r=-1





r=0

53. Co to znaczy, ze współczynnik korelacji między zmiennymi X i Y wynosi 0.
Jeżeli współczynnik korelacji między dwiema zmiennymi wynosi zero, to znaczy, że są to zmienne nieskorelowane. Wartość jednej zmiennej nie zależy od drugiej.
54. Jaką postać ma liniowa funkcja regresji, gdy współczynnik korelacji między zmiennymi X i Y wynosi 0.
Jeżeli współczynnik korelacji wynosi 0 to nie ma zależności pomiędzy dwoma zmiennymi, a wykresem funkcji regresji są wszystkie punkty układu współrzędnych.
55. Na podstawie obliczeń uzyskano współczynnik korelacji równy –0.97. Jak można zinterpretować tę wartość?
Współczynnik korelacji równy –0,97, oznacza, że większym wartościom jednej cech odpowiadają średnio mniejsze wartości drugiej cechy. Taką zależność nazywamy ujemną lub malejącą.
56. Na podstawie obliczeń uzyskano współczynnik korelacji równy 1.09. Jak można zinterpretować tę wartość?
Współczynnik korelacji nie może przyjąć wartości powyżej 1.
57.W badaniu wpływu długości czasu (w latach) pracy (X) pewnego urządzenia na przciętny czas (w miesiacach) bezawaryjnej pracy (Y) tego urządzenia na podstawie obserwacji dziesięciu maszyn uzyskano współczynnik korleacji r=-0,9983. Czy można na tej podstawie przyjąć, że istnieje zależność między długością czasu pracy i przeciętnego czasu pracy bezawaryjnej.
Jeśli próba zaostała dobrana poprawnie (zapewniono reprezentatywność) to można uznać, że istnieje taka zależność, że im dłuższy czas pracy w latach tym krótszy okres (w m-cach) bezawaryjnej pracy. Wynika to z tego, że korelacja jest równa prawie -1.
58. W dwudziestu gospodarstwach wiejskich badano zależność między spożyciem ziemniaków (cecha X) i artykułów zbożowych (cecha Y). Uzyskano współczynnik korelacji r=-0,9983. Czy można na tej podstawie przyjąć, że istnieje zaleznosć między spożyciem ziemniaków i artykułów zbożowych?
Tak jak w 57.
59. Co to jest indeks Fishera zmian cen ?
Indeks Fishera zmian cen jest średnią geometryczną z indeksów wyznaczonych przez Laspeyersa i Paaschego. Można go uważać za dobre przybliżenie indeksu poprawnie mierzącego zmiany cen (z dwóch różnych okresów) , jeśli przyjąć, że indeksy Laspeyersa i Paaschego określają granice przedziału, w którym zawarta jest prawdziwa wartość indeksu.
60. Co to jest indeks Fishera zmian ilości?
Jeśli przyjąć, że indeksy Laspeyersa i Paaschego poprawnie określają granice przedziału, w którym zawarta jest prawdziwa wartość indeksu, to : **Indeks Fishera zmian ilości** uważa się za dobre przybliżenie indeksu właściwie mierzącego zmiany ilości (rozmiarów fizycznych)
61. Co to jest indeks Laspayresa zmian cen ?

Dynamika zjawisk				
Numer artykułu	Ilość		Cena jednostkowa	
	Rok0	Rok1	Rok0	Rok1
1	q_{10}	q_{11}	p_{10}	p_{11}
...	...	...	...	...
k	q_{k0}	q_{1k}	p_{k0}	p_{1k}

Numer	Wartość	Wartość
1	$w_{1,00}=p_{10}q_{10}$	$w_{1,11}=p_{11}q_{11}$
...	...	...
k	$w_{k,00}=p_{k0}q_{k0}$	$w_{k,11}=p_{k1}q_{k1}$
Razem	w_{00}	w_{11}

Indeks Laspayersa zmian cen to indeks określający wpływ zmian cen na dynamikę wartości; informuje o tym , jak zmieniałaby się łączna wartość wszystkich towarów w momencie badanym w stosunku do momentu podstawowego, gdyby ilości poszczególnych towarów były w obu porównywalnych momentach jednakowe oraz takie jak w momencie podstawowym, a zmiana wartości nastąpiłaby tylko na skutek zmian cen. $LI_{pq}=(w_{10}/w_{00})$, gdzie $W_{ij}=p_iq_j$.
62. Co to jest indeks Laspayresa zamian ilości ?
Indeks Laspayersa zmian ilości mówi jak zmieniałaby się całościowo wartość wszystkich towarów w momencie badanym w stosunku do momentu podstawowego, gdyby w obu porównywalnych momentach ceny były niezmienne i

takie jak **w momencie podstawowym,** a zmiana wartości nastąpiłaby tylko i wyłącznie na skutek zmian ilości poszczególnych towarów; co więcej informuje o przeciętnych zmianach ilości poszczególnych towarów w obu porównywalnych momentach. $LI_{qp}=(w_{01}/w_{00})$

63. Co to jest indeks Paaschego zmian cen ?

Indeks Paaschego zmian cen to średnia harmoniczna z indywidualnych indeksów cen, a której wagami są wartości towarów w momencie badanym; Informuje o tym ,jak zmieniałaby się łączna wartość wszystkich towarów w momencie badanym w stosunku do momentu podstawowego, gdyby ilości poszczególnych towarów były w obu porównywalnych momentach jednakowe oraz takie, **jak w momencie badanym** , a zmiana wartości nastąpiłaby wyłącznie na skutek zmian cen. $PI_{pq}=(w_{11}/w_{01})$

64. Co to jest indeks Paaschego zmian ilości?

Indeks Paaschego zmian ilości to średnia harmoniczna indywidualnych indeksów ilości; informuje, jak zmieniałaby się globalna wartość wszystkich towarów w momencie badanym w stosunku do momentu podstawowego, gdyby w obu porównywalnych momentach ceny były niezmienne i takie jak w momencie badanym, a zmiana wartości nastąpiłaby tylko i wyłącznie na skutek zmian ilości poszczególnych towarów. $PI_{qp}=(w_{11}/w_{10})$

65. Co to jest indeks zmian wartości.

Indeks zmian wartości to indeks, który informuje o łącznych zmianach wartości danych produktów (równocześnie) w momencie badanym w stosunku do momentu podstawowego. Zmiany te wynikają zarówno ze zmian ilości , jak i cen tych produktów. $I_w=(w_{11}/w_{00})$

66. Jaka jest zależność między indeksami zmian wartości, ilości oraz cen.

Wartości, ceny i ilości są wielkościami, które mają szczególne znaczenie w badaniu zjawisk ekonomicznych. Indeksy zmian tych wielkości są badane razem w tzw. **indeksach agregatowych (zespołowych),** które w odpowiedni sposób wyrażają łączne zmiany zachodzące w czasie w całej zróżnicowanej zbiorowości. $I_w=LI_{pq}xPI_{qp}=PI_{pq}xLI_{qp}=FI_pxFI_q$

67. W jaki sposób można oszacować przeciętne tempo zmian na przestrzeni kilku lat.

Czas	Zjawisko	Indeksy łańcuchowe		
		absolutne	względne	$i_{t/t-1}$
t_0	y_0			
t_1	y_1	y_1-y_0	$(y_1-y_0)/y_0$	y_1/y_0
t_2	y_2	y_2-y_1	$(y_2-y_1)/ y_1$	y_2/y_1
...	...	...	...	...
t_k	y_k	y_k-y_{k-1}	$(y_k-y_{k-1})/ y_{k-1}$	y_k/y_{k-1}

Średnie tempo zmian -I_g jest średnią geometryczną z indeksów łańcuchowych $i_{t/t-1}$ ($t\in T_1$) ; metoda ta zawodzi, gdy duże są nieregularności w obserwowanej dynamice zjawisk. $i_{t/t-1}$ jest stopą rocznego wzrostu, czyli jeżeli wartość w roku 1 wynosi 2, a wartość w roku 3- 2,5, to $i_{3/2}$=1,25. Średnim tempem zmian w okresie 0-t nazywamy średnią geometryczną z $i_{t/t-1}$.

68. Co to jest indeks łańcuchowy.

Indeks łańcuchowy należy do obszernej klasy mierników dynamiki zjawisk wartości y_t , gdzie y_{t*} oznacza podstawę porównania dla wartości zjawiska y_t w kolejnych momemtach czasu $t\subset T1$. Jeśli ta podstawą jest zawsze moment poprzedni do badanego to indeksy dynamiki są nazwane **indeksami łańcuchowymi.** Wartość indeks łańcuchowego w czasie t/: $i_{t/t-1}=(y_t/y_{t-1})$

69. Co to jest indeks jednopodstawowy.

Indeks jednopodstawowy jest miernikiem dynamiki zjawisk; występuje wtedy, gdy podstawa porównania jest stała dla wszystkich wartości y_t , tzn. y_{t*} = const. Czyli wartość indeksu w czasie t: $i_t=(y_t/y_0)$.

70. Co to jest trend?

Trend składnik szeregu czasowego wyrażający ogólną tendencję systematycznych zmian poziomu danej zmiennej; (tendencja rozwojowa) – funkcja opisująca generalny przebieg zjawiska, zmiany średniego zjawiska w czasie.

Metody wyznaczania trendu: Tendencję rozwojową można wyodrębnić dwiema metodami:

-Metodą mechaniczną która polega na wygładzeniu szeregu czasowego, poprzez „oczyszczenie" go z wszelkiego typu wahań. Wygładzenia dokonuje się przy użyciu średnich ruchomych lub metody najmniejszych kwadratów.

-Metoda analityczna która polega na wyznaczeniu postaci funkcji trendu. Metoda analityczna wyodrębniania tendencji rozwojowej polega na ustaleniu takiej postaci funkcji matematycznej, która najlepiej przybliża trend zjawiska.

71. Co to są wahania okresowe (sezonowe) ?

Powtarzające się regularnie zmiany poziomu zjawiska. Najczęstszym okresem wahań jest rok.

72. i 76. Do czego służy metoda średnich ruchomych. Na czym polega metoda średnich ruchomych.

Szeregi czasowe ze znacznym udziałem wahań okresowych i przypadkowych poddaje się zwykle wyrównaniu którego rezultatem jest nowy szereg eksponujący trend rozwojowy środowiska. Najprostszą metodą eliminacji wahań z szeregu czasowego jest obliczenie tzw. średnich ruchomych i zastąpienie nimi pierwotnych wyrazów szeregu czasowego.

Średnie oblicza się zwykle z nieparzystej (parzystej) liczby sąsiadujących ze sobą wyrazów szeregu, tak aby uzyskany wynik móc podporządkować całkowitej wartości t znajdującej się w środku uwzględnionego w obliczeniach przedziału.

a) r(długość cyklu wahań)-nieparzyste

$$y_m = \frac{1}{r}(y_{m-\frac{r-1}{2}} + \ldots + y_m + \ldots + y_{m+\frac{r-1}{2}}),$$ m=(r-1)/2,....,k-(r-1)/2

b) r - parzyste

$$y_m = \frac{1}{r}(\frac{1}{2} y_{m-\frac{r}{2}} + \ldots + y_m + \ldots + \frac{1}{2} y_{m+\frac{r}{2}}),$$ m=r/2,.....,k-r/2

73. Jak można oszacować wielkość wahań okresowych?

W zależności od tego jaki charakter mają wahania sezonowe, rachunek wskaźników, opisujących zakres działania czynników sezonowych ,przebiega inaczej. Jeżeli rezultatem działania czynników sezonowych jest zmienna amplituda wahań to zakres działania sezonowości opisują relatywne wskaźniki sezonowości. Jeżeli zaś rezultatem działania czynników sezonowych jest stała amplituda odchyleń od trendu, to zakres działania tych czynników opisują absolutne wskaźniki sezonowości.

W pierwszej kolejności wyznacza się tzn. *surowe wskaźniki* sezonowości, które określają przeciętne odchylenia od trendu w kolejnych podokresach cyklu sezonowości $O_{si} = \frac{Y_i}{\overline{Y}_i}$, Y_i=y_i+y_{r+i}+y_{2r+i}..., $\overline{Y}_i$=y_i+y_{r+i}+y_{2r+i}; y_t=dopasowana funkcja trendu

Następnie wyznaczamy wskaźnik korygujący .

$$k = \frac{r}{O_{s1} + O_{s2} + \ldots + O_{sr}}$$

Poprawny wskaźnik sezonowości wynosi Oi = Osi * k , i=0,1,...,r-1 Wskaźnik ten określa zakres względnych, czyli zależnych od poziomu trendu, odchyleń spowodowanych działaniem czynników sezonowości.

Na koniec obliczamy *absolutny wskaźnik okresowości* g_i. g_i = (Oi – 1)*y, i=0,1,...,r-1, $\overline{y} = (\sum_{j=0}^{n-1} y_j)/n$

Wskaźnik ten określa stałe niezależne od poziomu trendu, odchylenia poziomu zjawiska od trendu spowodowane działaniem czynników sezonowych.

74. Jak wykonuje się prognozę w szeregu czasowym w którym występuje zjawisko wahań okresowych?

Obliczamy tak samo jak w 73 do g_i. Potem musimy *oszacować odchylenie standartowe*.

$$S = \sqrt{\frac{1}{n-2}\sum_{j=0}^{n-1}(y_j - \overline{y_j} - g_t)^2}$$, gdzie t jest resztą dzielenia j/r

Obliczamy prognozę w chwili m>n, y_m=$y_{m(sr)}$+g_t+/-S, (gdzie t jest resztą z dzielenia m/r).

75. Jaką metodą można wyznaczyć Trend?

Wyznaczanie trendu:

1. Metoda empiryczna (średnich ruchomych)- w pyt. 72.
2. Metoda analityczna (najmniejszych kwadratów)- metoda aproksymacji funkcji określonego typu, do zbioru punktów empirycznych. Metoda ta polega na takim doborze parametrów aproksymowanej funkcji, by suma kwadratów odchyleń rzędnych punktów empirycznych od wykresu tej funkcji była mniejsza. Sprowadza sie ona do rozwiązania odpowiedniego dla danego typu aproksymowanej funkcji układu równań. Metodą najmniejszych kwadratów wyznacz się najczęściej w staystyce funkcjię regresji II rodzaju.

78. Co to jest *szereg rozdzielczy*.

Jeden z szeregów statystycznych przedstawiający budowę (strukturę) zbiorowości, czyli jej podział na części z określonego, rzeczowego punku widzenia. Cecha statystyczna na podstawie której dokonuje się podziału zbiorowości na mniejsze części, może być cechą niemierzalną lub mierzalną. W szeregu rozdzielczym w jednej kolumnie w sposób uporządkowany przedstawiony jest wykaz klasyfikacyjny, czyli warianty badanej cechy, a w drugiej kolumnie przedstawione są liczebności odpowiadające poszczególnym klasom z wykazu. Jest to więc uporządkowany i pogrupowany zbiór informacji dotyczących badanej cechy określonej zbiorowości. W zależności od rodzaju cechy według której podzielono zbiorowość szeregi dzielimy na dwie grupy:

- szeregi oparte na cesze niemierzalnej:

Poziom wykształcenia	podst.	Zasadnicze zawodowe	średnie	wyższe	ogółem
Liczba pracowników	8	12	30	10	60

Np. Szeregi rozdzielcze cechy niemierzalnej uzyskuje się grupując budynki wg dzielnic miasta. Jeśli przedmiotem badania statystycznego są np. budynki mieszkalne oddane do użytku to punktowy szereg rozdzielczy uzyskuje się grupując budynki wg liczby kondygnacji, natomiast przedziałowy szereg rozdzielczy można uzyskać grupując te same budynki wg trwania budowy.

- Szeregi oparte na cesze mierzalnej np. czas pozostawania bez pracy...

Pyt. 79. Co to jest *histogram?*

Rodzaj wykresu słupkowego oparty na prostokątnym układzie współrzędnych; Histogram składa się z pionowych przylegających do siebie prostokątów (słupków). Długości podstaw tych prostokątów są proporcjonalne do rozpiętości przedziałów klasowych, a wysokość do ich liczebności na jednostkę rozpiętości. Zwykle histogram służy do przedstawiania struktury szeregów rozdzielczych o równych przedziałach klasowych i wówczas wysokość prostokąta jest proporcjonalna do liczebności. Budując histogram na podstawie szeregu o nierównych przedziałach klasowych, należy uprzednio obliczyć liczebności przypadające w danym przedziale na jednostkę jego rozpiętości. Histogram umożliwia poznanie typu rozkładu zbiorowości statystycznych wg badanej cechy.

80. Wymień mierniki położenia próby:

Średnia, Mediana, Dolny kwartyl, Górny kwartyl, Dominanta, Minimum**,** Maksimum

81. Pyt. 82. Co to jest dolny kwartyl. Co to jest górny kwartyl.

Do najczęściej zaliczanych kwartyli zaliczamy kwartyle:

Kwartyl dolny- dzieli on zbiorowość uporządkowaną na dwie części, w ten sposób że 25% jednostek ma wartość cechy niższe, a 75% wyższe od kwartyla dolnego.

Kwartyl górny– dzieli zbiorowość uporządkowaną na dwie części w ten sposób, że 75% jednostek ma wartości cechy niższe a 25% wyższe od kwartyla górnego.

83. Co to jest MEDIANA?

Me – wartość wyrazu środkowego w uporządkowanym szeregu statystycznym; to taki punkt (liczba) która ilościowo rozdziela dane na dwie równe części. Sposób obliczania mediany zależy od rodzaju szeregu statystycznego, w którym przedstawiono informacje o wartości cechy statystycznej, a także od tego czy liczba jednostek statystycznych jest parzysta czy nieparzysta.

$$Me = X_{0,5} + \frac{h_{0,5}}{n_{0,5}}\left(\frac{n}{2} - n_{0,5}\right)$$

84. Co to jest DOMINANTA?

(moda, wartość typowa, wartość modalna) jest to wartość cechy, która najczęściej występuje w danej zbiorowości. W zależności od formy w której przedstawione są informacje o wartości cechy jednostek statyst. Stosuje się różne techniki ustalania dominanty. W *przypadku indywidualnego szeregu wartości cechy* wartość dominanty należy jedynie wskazać – i jest to wartość cechy która najczęściej występuje w badanej zbiorowości statystycznej. W *szeregach z cechą mierzalną ze zmiennością skokową* –wartość dominanty jest to ta wartość dla której liczebność cząstkowa jest największa. W *szeregu z cechą mierzalną ze zmiennością ciągłą –* wartość dominanty liczona jest wg wzoru: Dx = Xd + Hd (Nd – Nd-1) / 2 Nd – Nd+1 – Nd-1) gdzie

Xd – początek przedziału w którym znajduje się dominanta

Hd – szerokość przedziału , Nd – ilość danych w przedziale , Nd –1 ilość danych w przedziale poprzedzającym przedział z dominantą, Nd+1 ilość danych po przedziale zawierającym dominantę.

85. Jaka jest wzajemna relacja między średnią, medianą a dominantą?

Średnia = mediana = dominanta czyli wszystkie tendencje mają taką samą wartość**-**że liczba jednostek statystycznych która posiada wartość cechy wyższe niż średnia arytmetyczna jest taka sama jak liczba jednostek , która posiada wartości cechy niższe niż średnia arytmetyczna. Taki rozkład wartości cechy w zbiorowości określany jest - rozkładem symetrycznym. Wartość średniej jest większa niż wartość mediany i wartość mediany jest większa od wartości dominanty tj. x>Me >D- oznacza że wartość cechy większości jednostek statystycznych jest niższa od średniej arytmetycznej. Taki rozkład nosi nazwę rozkładu o asymetrii prawostronnej. Wartość średniej jest mniejsza niż wartość mediany i wartość mediany jest mniejsza od wartości dominanty tj. x< Me<D- oznacza że wartość cechy większości jednostek statystycznych jest wyższa od średniej arytmetycznej. Jest to rozkład o asymetrii lewostronnej.

86. Wymień mierniki rozproszenia cechy:

-Odchylenie standardowe (δx) jest to pierwiastek kwadratowy z sumy kwadratów odchyleń poszczególnych wartości zmiennej x od średniej arytmetycznej, podzielonej przez liczebność szeregu X+/- S -typowy obszar zmienności ma sens jeśli układ jest symetryczny wokół średniej. $S^2 = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \\ \frac{1}{n}\sum_{i=1}^{n} n_i(\dot{X}_i - \overline{X})^2 \end{cases}$

-Współczynnik zmienności jest to stosunek bezwzględnej miary odchylenia do średniej arytmetycznej, wyrażony w procentach.

V = (S/X_{sr}) 100%. Jeżeli współczynnik jest mały to dane są mniej zróżnicowane.

-Rozstęp – miara ta obrazuje różnice między wartością największą a najmniejszą w badanej zbiorowości, wyznaczamy więc jej wartość odejmując od najwyższej , najniższą wartość cechy: R = Max – Min.

-Odchylenie przeciętne – jest to średnia arytmetyczna bezwzględnych wartości (modułów) odchyleń wartości faktycznych szeregu od średniej arytmetycznej. $d = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} |X_i - \overline{X}| \\ \frac{1}{n}\sum_{i=1}^{n} n_i |\dot{X}_i - \overline{X}| \end{cases}$

Odchylenie ćwiartkowe– $Q=(Q_3-Q_1)/2$; gdzie $Q_{3\ i\ 1}$ odpoweidnio górny i dolny kwartyl.

87. Co mierzy rozstęp?

Określa największą rozbieżność, jaką zaobserwowano wśród wartości badanej cechy. Miara ta określa zróżnicowanie jednostek na podstawie oceny wartości skrajnych cechy statystycznej. Wartościom tym mogą odpowiadać niewielkie lub wręcz znikome liczebności. Dlatego też nie jest to precyzyjna miara zróżnicowania i służy jedynie wstępnej ocenie zmienności zjawiska. Informuje ona jak bardzo różnią się wartości cechy statystycznej w ogóle.

88. Co mierzy odchylenie standardowe?

Odchylenie jest miarą która podobnie jak odchylenie przeciętne, charakteryzuje przeciętny poziom odchyleń faktycznych wartości cechy od średniej arytmetycznej. Jest to miara bardziej precyzyjna niż odchylenie przeciętne**.**

89. Co mierzy wariancja?

Wariancja D2X zmiennej losowej jest liczbą charakteryzującą rozrzut zbioru jej wartości wokół wartości EX. Charakteryzuje zróżnicowanie cechy.

90. Co mierzy współczynnik zmienności?-

W przypadku konieczności porównania rozproszenia dwóch różnych zjawisk należy posłużyć się współczynnikiem zmienności. Współczynnik zmienności to iloraz odchylenia standardowego i średniej w danym rozkładzie $V=(s/X_{sr})$. Im wyższy jest ten procent, tym większe jest względne zróżnicowane cechy w rozkładzie. o iloraz odchylenia standardowego i średniej w danym rozkładzie $V=(s/X_{sr})$. Współczynnik zmienności wyraża się często procentowo, aby określić, jaki procent poziomu średniej stano i odchylenia standardowe w rozkładzie. Tego typu badania są szczególnie przydatne w porównywaniu zróżnicowania takich wielkości jak dochody, wydajność pracy, absencja w pracy w różnych przedsiębiorstwach lub działach jednego przedsiębiorstwa.

92. Co to jest typowy obszar zmienności?

– zwykle przedział w którym leżą wszystkie wartości, cechy mierzalnej jednostki, zbiorowości statystycznej. Obszar zmienności wyznaczany jest przez najmniejszą i największą wartość cechy. Zawiera on podstawowe informacje o zmienności badanej cechy. Średnia arytmetyczna i odchylenie standardowe pozwalają na określenie obszaru wartości typowych dla określonej zbiorowości statystycznej. Ten obszar wyznaczany jest jako przedział liczbowy, którego dolną granicą jest wartość średniej arytmetycznej pomniejszona o odchylenie standardowe, a górną granicą jest wartość średniej arytmetycznej powiększona o odchylenie standardowe. Obszar typowych wartości cechy można zapisać: $(X_{sr}- S, X_{sr} + S)$

94. Jaki procent populacji zawiera się między kwartylami.

Ponieważ dolny kwartyl odcina 25% danych z dołu a górny 25% z góry, to pomiędzy nimi pozostaje 50% danych.

96. Co można powiedzieć o asymetrii cechy, jeżeli mediana jest średnią z pozostałych kwartyli.

Jeżeli mediana jest średnią z pozostałych kwartyli, to środkowe 50% danych jest symetrycznych.

98. Co można powiedzieć o skośności cechy, jeżeli mediana jest większa od średniej.

Jeżeli mediana jest większa od średniej to mamy doczynienia z asymetrią lewostronną.

99-100. Jaki jest zakres zmienności współczynnika koncentracji Lorentza?

Koncentracje ocenia się poprzez porównanie liczby jednostek o określonych wartościach cechy, jaką łącznie jednostki te posiadają. Mała liczebność klasy wartości cechy statystycznej oraz znaczna suma wartości cechy, którą jednostki te łącznie posiadają świadczą o silnej koncentracji rozkładu cechy statystycznej. W przeciwnym wypadku następuje rozdrobnienie rozkładu. Współczynnik przyjmuje wartości z przedziału <0,0 ; 1,0> i im większa jest jego wartość tym koncentracje rozkładu uznaje się za silniejszą. Współczynnik ten przyjmuje wartość 0, gdy rozdział ogólnej sumy wartości cechy przebiega według linii równomiernego rozdziału, zaś 1,0 gdy krzywa Lorenza pokrywa się z osią OX.