MALES.csv
Baza zawiera dane dotyczace pracy i wyksztalcenia młodych mezczyzn w USA

nr – identyfikator; year – rok urodzenia; school – liczba lat nauki; exper – doswiadczenie zawodowe w latach; union – czy jest w zwiazkach zawodowych?; ethn – narodowość; married – czy jest żonaty; wage – logarytm godzinowej stawki; placa- godzinowa stawka; industry – sektor gospodarki; occupation – branża; residence – miejsce zamieszkania

Prawdopodobieństwo (20 pkt):

1. Narysuj funkcję gęstości prawdopodobieństwa dla standardowego rozkładu normalnego, opisz osie. (5 pkt.)
2. Zidentyfikuj obszar zawierający 75% prawdopodobieństwa rozkładu zmiennej dla N~(0,1). (5 pkt.)
3. Oblicz prawdopodobieństwo tego, ze podczas 150 rzutów moneta orzeł wypadnie:
a) 90 razy; (3 pkt.)
b) mniej niz 30 razy. (3 pkt.)
4. Jakie jest prawdopodobieństwo tego, że wylosujemy ze zbioru Male kogoś o narodowości hiszpańskiej (“hisp”)? (4 pkt.)

Wnioskowanie statystyczne (20 pkt):

1. Załóżmy, że zbiór Male to dane dotyczące populacji młodych mężczyzn pewnego miasta. Oblicz średnią płacę godzinową w tej populacji (kolumna „placa”). (2 pkt.)
2. Wylosuj z populacji próbę N=150, oblicz średnią płacę godzinową dla próby (kolumna „placa”). Powtórz losowanie 100 tys. razy, aby otrzymane rezultaty były jak najbardziej reprezentatywne. Za każdym razem zapisz średnią płacę godzinową dla losowanej próby. (8 pkt.)
3. Jakie jest prawdopodobieństwo, że średnia wartość płacy godzinowej obliczonej na podstawie wylosowanej próby odda nam średnią wartość płacy godzinowej w całej populacji? (5 pkt.)
4. Utwórz 95% przedział ufności dla zdarzenia jakim jest napotkanie młodego i żonatego mężczyzny. Co oznaczają otrzymane przez Ciebie rezultaty? (5 pkt.)

Regresja (21 pkt):

 # zbiór AIDS.csv

baza zawiera dane dotyczące diagnostyki i śmiertelności osob z choroba AIDS; zmienna "year" to rok; "diagnosed" to liczba zdiagnozowanych przypadków AIDS w danym roku; zmienna "deaths" to liczba zgonów w danym roku

1. Stwórz model regresji z "deaths" jako zmienną objaśnianą i "diagnosed" jako zmienna objasniającą i oszacuj parametry równania. Jakie jest nachylenie prostej, jaki jest punkt przecięcia z OY? Który z parametrów jest istotny statystycznie? (6 pkt.)
3. Narysuj wykres rozrzutu dla tych zmiennych i dodaj do niego linie regresji. (5 pkt.)
4. Stwórz model regresji ze zmienną "year" jako objaśniającą i "diagnosed" jako objaśnianą. Jakie wartości maja teraz oszacowane parametry (nachylenie i przecięcie)? Czy są one istotne statystycznie? (6 pkt.)
5. Który z modeli lepiej opisuje trend w danych? Dlaczego? (4 pkt.)

max = 61 pkt.; dst.= 30 pkt.