

MWS LAB4

Piotr Mikołajczyk

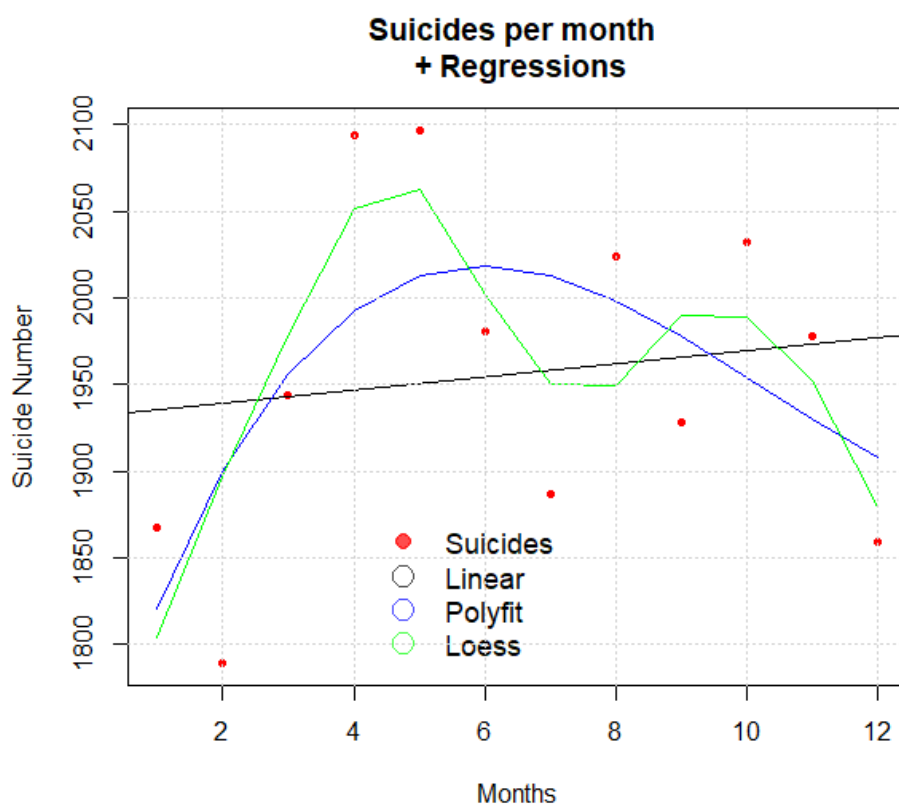
30.05.2021

Zad1

Dane przedstawiają liczbę zanotowanych samobójstw w Stanach Zjednoczonych w 1970 roku z podziałem na poszczególne miesiące. Zbadaj, czy zamieszczone poniżej dane wskazują na sezonową zmienność liczby samobójstw, czy raczej świadczą o stałej intensywności badanego zjawiska.

Tabela 1 – Dane samobójstw z 1970 roku w Stanach Zjednoczonych

Miesiące	L. Samobójstw	Liczba dni w miesiącu
Sty.	1867	31
Lut.	1789	28
Mar.	1944	31
Kwi.	2094	30
Maj.	2097	31
Cze.	1981	30
Lip.	1887	31
Się.	2024	31
Wrz.	1928	30
Paz.	2032	31
Lis.	1978	30
Gru.	1859	31



Rys. 1 – Wykres z tabeli 1 oraz regresja liniowa, wielomianowa 3 rzędu oraz za pomocą funkcji lokalnego dopasowania wielomianowego

W celu zbadania hipotezy wykorzystano twierdzenie Pearsona:

$$T = \sum_{i=1}^r \frac{(v_i - np_i^*)^2}{np_i^*}$$

Gdzie:

V_i – samobójstwa w danym miesiącu,

n – samobójstwa w całym roku

p_i^* – prawdopodobieństwo wystąpienia samobójstwa w konkretnym miesiącu

Zmienna losowa T , zgodnie z twierdzeniem Pearsona, dąży do rozkładu Chi Kwadrat z $r-1$ stopniami swobody. W celu wyznaczenia zmiennej T , wykorzystano wbudowaną funkcję `qchisq()` w oprogramowaniu R. Przyjęto hipotezy H_0 jako stałą intensywność (brak zależności) od miesiąca, oraz H_1 jako sezonowość. W celu ustalenia która hipoteza jest prawdziwa wykorzystano test Chi Kwadrat za pomocą wbudowanej funkcji w oprogramowaniu R `chisq.test()`. Poziom istotności przyjęto na poziomie $\alpha = 0.1$. Przyjęto również że prawdopodobieństwo samobójstwa w ciągu roku jest równe – ponieważ mamy hipotezę H_0 jako hipotezę jednostajną – i opisane rozkładem jednostajnym. Daje to $p_{rok} = 1/365$ dla całego roku oraz $p_{miesiac} = \text{liczba dni w miesiącu} / 365$.

Wnioski:

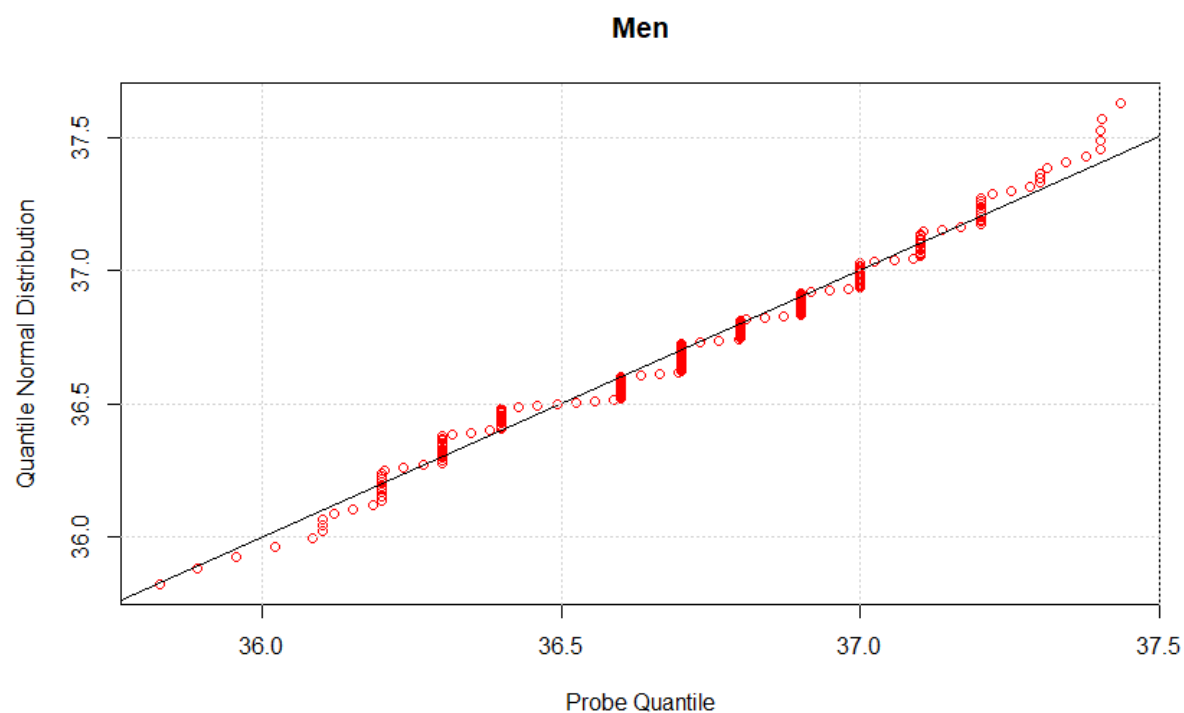
Wykonanie testu wykazało że samobójstwa w istocie mają charakter sezonowy. Wartość P dla tych danych wynosi 0.3472. Na rysunku wrysowano w punkty danych samobójstw z 1970 roku, regresje liniową, wielomianową oraz funkcją lokalnego dopasowania wielomianowego. Druga oraz trzecia funkcja uwypukla i potwierdza hipotezę sezonowości. Regresja liniowa natomiast pokazuje jedynie tendencje wzrostową samobójstw na przestrzeni roku 1970.

Zad2

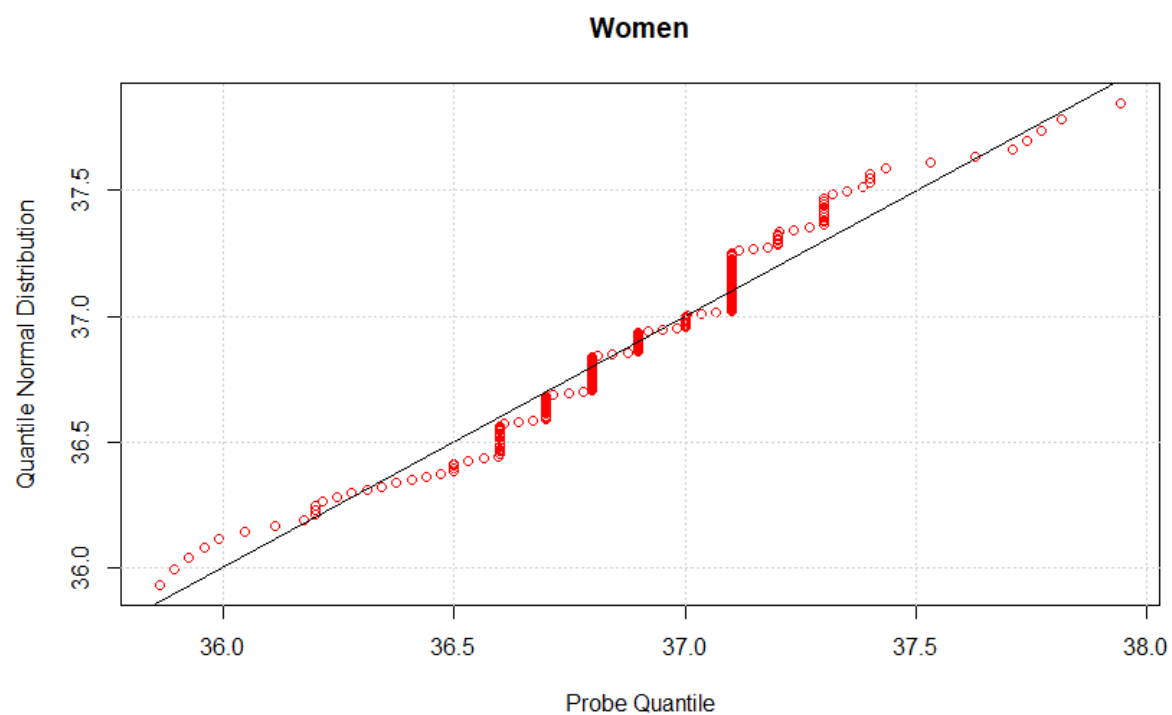
Plik `tempciala.txt` zawiera zarejestrowane wartości tętna oraz temperatury ciała dla 65 mężczyzn (kod 1) oraz 65 kobiet (kod 2).

a) Wy estymuj wartość średnią i odchylenie standardowe temperatury ciała, osobno dla mężczyzn i kobiet, a następnie wykreśl tzw. wykresy kwantyl-quantyl dla rozkładu normalnego (na płaszczyźnie X Y narysuj punkty o współrzędnych $(x; y)$, gdzie x jest kwantylem rzędu α (0; 1) z próby, a y jest kwantylem rzędu α rozkładu normalnego o wy estymowanych parametrach). Co możesz powiedzieć na podstawie otrzymanych wykresów o zgodności rozkładu temperatury ciała mężczyzn/kobiet z rozkładem normalnym? (W celu sprawdzenia, jak taki wykres mogłyby wyglądać gdyby dane pochodziły z rozkładu normalnego można przeprowadzić odpowiednie symulacje.

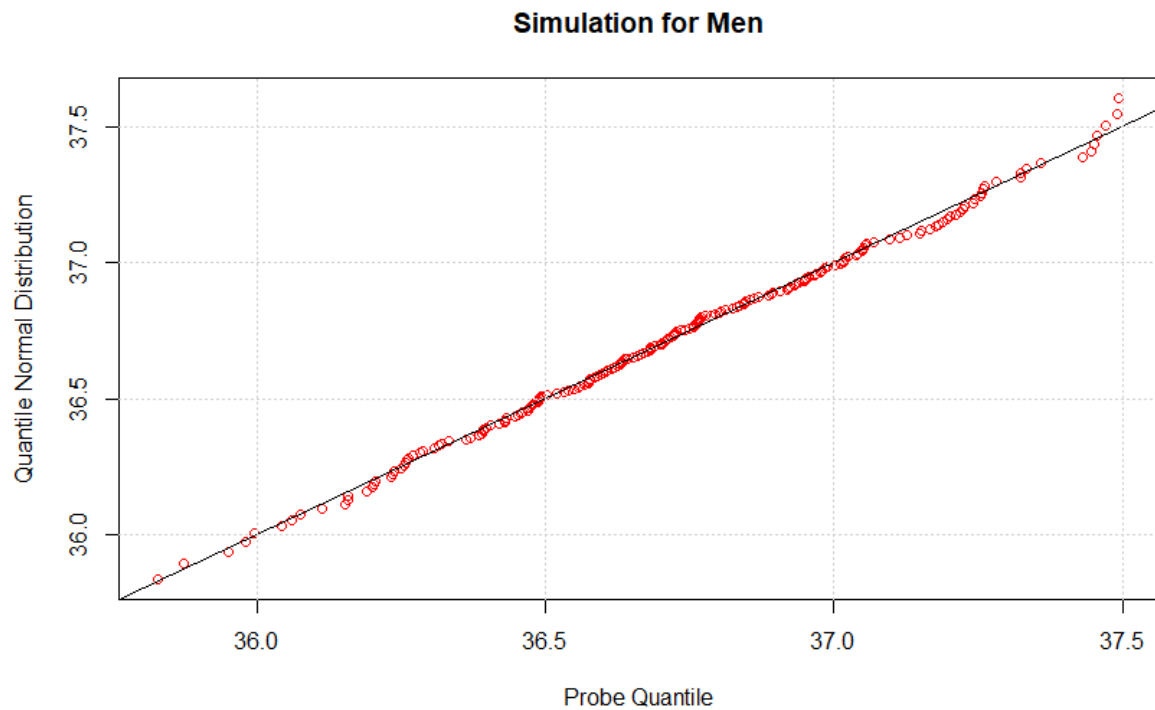
Na rysunkach 2 oraz 3 narysowano wykresy kwantyl – kwantyl estymując uprzednio parametry rozkładu normalnego – jego średnią oraz odchylenie standardowe temperatury ciała dla mężczyzn oraz kobiet, korzystając z wbudowanych funkcji programu R `quantile` oraz `qnorm`.



Rys. 2 – Wykres Kwantyl – Kwantyl dla mężczyzn



Rys. 3 – Wykres Kwantyl – Kwantyl dla kobiet



Rys. 4 – Wykres symulacji kwantyl-kwantyl dla mężczyzn



Rys. 5 – Wykres symulacji kwantyl-kwantyl dla kobiet

Wnioski:

Z rysunków 2 oraz 3 rozróżnić można cechy mężczyzn jako bardziej stałych temperaturowo – przynajmniej dla danej próby. Kobiety posiadają większy rozrzut temperaturowy. Między nimi

natomiast a symulacją trzeba zauważyć że znika kwantyzacja dla symulacji – wynika to z braku ograniczenia jakim są pomiary co 0.1 stopnia Celsjusza. **Wydaje się że dla próbek porównując z symulacją, rozkład temperatur dąży do rozkładu normalnego zarówno dla kobiet jak i mężczyzn.**

b) Przeprowadź testy, osobno dla mężczyzn i kobiet, tego że średnia temperatura ciała jest równa 36.6 stopni wobec hipotezy alternatywnej, że ta średnia temperatura jest jednak inna.

W celu zbadania hipotezy czy jest to zasadne twierdzenie przeprowadzono test t-Studenta oraz posłużono się wartością zmiennej T dla rozkładu t-Studenta. Hipoteza H_0 oznaczała dla kobiet jak i dla mężczyzn że mają oni średnio 36.6 stopni Celsjusza natomiast hipoteza H_1 , że nie mają średnio 36.6 stopni Celsjusza. Przyjęto dla testu poziom istotności $\alpha = 0.01$.

Wnioski:

Wykazał on że dla tego poziomu istotności w istocie średnia temperatura ciała wynosi 36.6 stopnia Celsjusza natomiast średnia temperatura dla kobiet jest inna niż 36.6 stopni Celsjusza. W przybliżeniu jest to zgodne z rysunkami 2,3,4 oraz 5.

Bibliografia:

[1] Slajdy wykładowe do przedmiotu MWS, Rafał Rytel-Andrianik