Opis zbiorów danych

W naszym projekcie użyliśmy dwóch zbiorów danych. Zbiór “BodyWeight” z pakietu MEMSS został użyty w modelu podstawowym ze zmiennymi jakościowymi. Natomiast zbiór “fat” z pakietu faraway został użyty w modelu podstawowym + współliniowość, a także w modelu z moderacją. Zbiór “BodyWeight” opisuje dane dotyczące masy ciała szczurów mierzonej przez 64 dni. Masę ciała szczurów (w gramach) mierzono pierwszego dnia, a następnie co siedem dni aż do dnia 64, z dodatkowym pomiarem dnia 44. Eksperyment rozpoczął się kilka tygodni przed „dniem 1”. Istnieją trzy grupy szczurów, każda na innej diecie (a, b i c). Zbiór ten zawiera ma 176 wierszy i 4 kolumny. Drugim zbiorem danych jest zbiór “fat”, w którym wiek, wagę, wzrost i 10 innych pomiarów obwodu ciała zarejestrowano dla 252 mężczyzn. Procent tkanki tłuszczowej każdego mężczyzny został dokładnie oszacowany za pomocą techniki ważenia pod wodą. Zbiór ten zawiera 252 obserwacje dla 18 zmiennych.

```
Rows: 176
Columns: 4
$ weight <dbl> 240, 250, 255, 260, 262, 258, 266, 266, 265,
$ Time   <dbl> 1, 8, 15, 22, 29, 36, 43, 44, 50, 57, 64, 1,
$ Rat    <fct> A, A, A, A, A, A, A, A, A, A, A, B, B, B, B,
$ Diet   <fct> a, a, a, a, a, a, a, a, a, a, a, a, a, a, a,
```

Wgląd na dane BodyWeight

```
Rows: 252
Columns: 18
$ brozek  <dbl> 12.6, 6.9, 24.6, 10.9, 27.8,
$ siri    <dbl> 12.3, 6.1, 25.3, 10.4, 28.7,
$ density <dbl> 1.0708, 1.0853, 1.0414, 1.0751
$ age     <int> 23, 22, 22, 26, 24, 24, 26,
$ weight  <dbl> 154.25, 173.25, 154.00, 184.75
$ height  <dbl> 67.75, 72.25, 66.25, 72.25,
$ adipos  <dbl> 23.7, 23.4, 24.7, 24.9, 25.6,
$ free    <dbl> 134.9, 161.3, 116.0, 164.7,
$ neck    <dbl> 36.2, 38.5, 34.0, 37.4, 34.4,
$ chest   <dbl> 93.1, 93.6, 95.8, 101.8, 97.3,
$ abdom   <dbl> 85.2, 83.0, 87.9, 86.4, 100.0,
$ hip     <dbl> 94.5, 98.7, 99.2, 101.2, 101.9
$ thigh   <dbl> 59.0, 58.7, 59.6, 60.1, 63.2,
$ knee    <dbl> 37.3, 37.3, 38.9, 37.3, 42.2,
$ ankle   <dbl> 21.9, 23.4, 24.0, 22.8, 24.0,
$ biceps  <dbl> 32.0, 30.5, 28.8, 32.4, 32.2,
$ forearm <dbl> 27.4, 28.9, 25.2, 29.4, 27.7,
$ wrist   <dbl> 17.1, 18.2, 16.6, 18.2, 17.7,
```

Wgląd na dane fat

Model podstawowy ze zmiennymi jakościowymi

Interesującą nas zmienną w “BodyWeight” była zmienna dotycząca diety dla szczurów Niestety, ale jest ona cechą, której nie można zmierzyć ilościowo. Dlatego dodajemy zmienne do modelu tworząc zmienne fikcyjne (0-1) dla każdej kategorii. Ze względu na współliniowość jedna z tych zmiennych, w naszym przypadku była to zmienna “b” musi zostać wyłączona z modelu. Po utworzeniu przeze mnie nowych zmiennych używam ich jako zmiennych niezależnych w modelu regresji liniowej do przewidywania masy (w gramach) ciała szczurów (zmienna zależna).

Uzasadnienie merytoryczne:

Zrozumienie wpływu różnych diet na przyrost i utratę masy ciała może pomóc w sformułowaniu zaleceń dietetycznych i interwencji dla poszczególnych osób i populacji. Wiemy również, że testowanie na zwierzętach jest nieetyczne i jesteśmy temu przeciwni, jednakże dane z tego badania wpływu diety na szczury były z 1996 roku więc myślę, że nadal mogą zostać one wykorzystane w celach analizy różnych hipotez i zbadania mechanizmów leżących u podstaw zmian masy ciała wywołanych dietą. Chociaż wyników badań na zwierzętach nie zawsze można bezpośrednio przenosić na ludzi, mogą one dostarczyć ważnych spostrzeżeń, które można potem dalej testować w badaniach na ludziach (oczywiście jak testowana strona się zgadza). Dlatego też, chociaż zestaw danych BodyWeight może nie mieć bezpośredniego zastosowania w rzeczywistych kontekstach, przyczynia się do naszego ogólnego zrozumienia wpływu diety na wagę i zdrowie.

Analiza modelu:

W modelu założono, że masa szczurów zależy od rodzaju stosowanej diety, przy czym dieta typu a i dieta typu c mają wpływ na masę szczurów. Dieta typu b została wyłączona z modelu tak jak to było powiedziane wcześniej, ze względu na współliniowość.

Współczynnik dla punktu przecięcia wynoszący wartość 484,705 to szacowana masa szczura, który był karmiony dietą podstawową (nie dietą a, b lub c). Oszacowany współczynnik dla zmiennej a wynosi -220,989, co oznacza, że szczur na diecie a ważył średnio o 220,989 g mniej niż szczur nie będący na diecie a, utrzymując dietę c na stałym poziomie. Podobnie oszacowany współczynnik dla zmiennej c wynosi 41,091, co oznacza, że szczur na diecie c ważył średnio 41,091 g więcej niż szczur nie będący na diecie c, utrzymując dietę a na stałym poziomie.

Znaczenie współczynników określają wartości p, które mierzą prawdopodobieństwo zaobserwowania współczynnika tak skrajnego jak ten w modelu, przy założeniu, że hipoteza zerowa (że współczynnik wynosi zero) jest prawdziwa. W tym modelu wartości p dla wszystkich współczynników są bardzo małe (<0,001), co wskazuje na mocne dowody przeciwko hipotezie zerowej.

Wielokrotna wartość R-kwadrat wynosząca 0,9203 wskazuje, że model wyjaśnia bardzo dużą część zmienności zależnej. Oznacza to, że rodzaj diety, jaką im podawano, odpowiada za znaczną różnicę w wadze szczurów. Skorygowana wartość R-kwadrat wynosząca 0,9193 uwzględnia liczbę zmiennych w modelu i jest nieco niższa niż wielokrotność wartości R-kwadrat. Sugeruje to, że włączenie zmiennych do modelu zapewnia dobre dopasowanie do danych, ale model może nie być dobrze uogólniony na nowe dane. Przechodzimy teraz do kolejnej ważnej miary naszego modelu, którą jest statystyka F. Wartość statystyki F wynosząząca 998,2 z wartością p mniejszą niż 0,001 wskazuje, że że ogólny model jest statystycznie istotny.

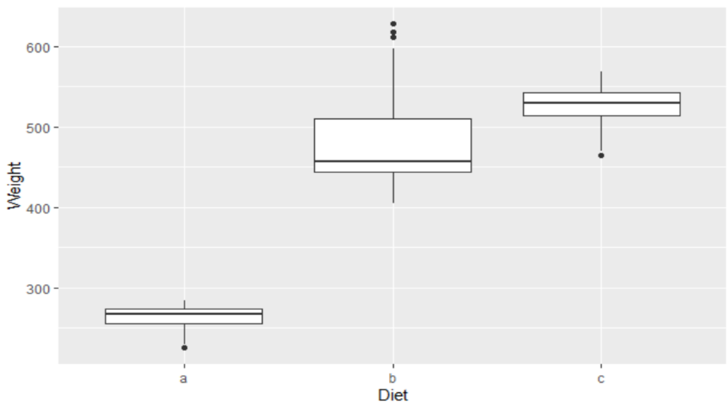
Względny błąd standardowy wynoszący 36,12 zapewnia oszacowanie zmienności masy, której nie wyjaśnia model. Im mniejsza wartość resztkowego błędu standardowego, tym lepsze dopasowanie modelu do danych. W tym przypadku wartość sugeruje, że model zapewnia dobre dopasowanie do danych.

```
Call:
lm(formula = weight ~ a + c, data = BodyWeight)

Residuals:
    Min      1Q  Median      3Q     Max 
-79.705 -19.026   1.284  11.287 143.295 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  484.705      5.445  89.022  < 2e-16 ***
a           -220.989      6.668 -33.139  < 2e-16 ***
c             41.091      7.700   5.336 2.94e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 36.12 on 173 degrees of freedom
Multiple R-squared:  0.9203,	Adjusted R-squared:  0.9193 
F-statistic: 998.2 on 2 and 173 DF,  p-value: < 2.2e-16
```

Analiza modelu ze zmiennymi jakościowymi

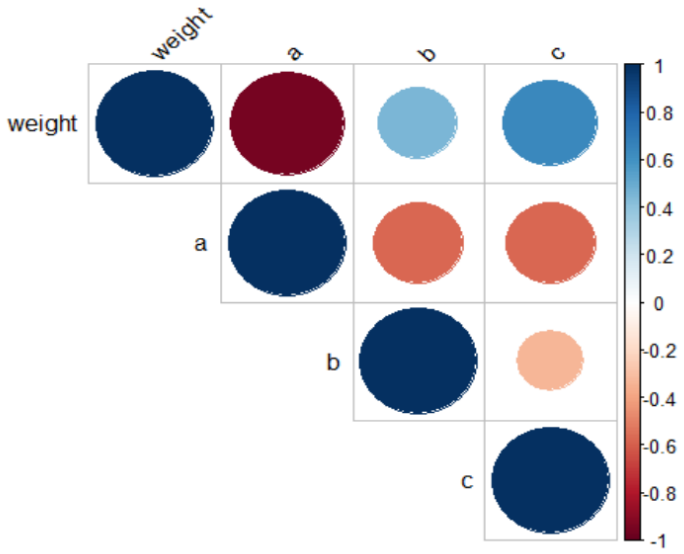
Podsumowanie:

Podsumowując, model ten pokazuje, że na wagę szczurów istotny wpływ ma rodzaj diety, jaką są karmione, przy czym dieta typu a prowadzi do obniżenia masy ciała, a dieta typu c prowadzi do zwiększenia masy ciała. Wartość F-statystyki wskazuje, że model jest statystycznie istotny. Wartości p dla obu zmiennych pokazują, że obie zmienne są istotne statystycznie w modelu. Natomiast współczynnik R^2 mówi nam, że 92,03% zmienności zmiennej zależnej jest wyjaśnione przez zmienne niezależne.



Wykres pudełkowy

Wykres przedstawia wykres pudełkowy rozkładu masy ciała dla każdej grupy dietetycznej. Oś x reprezentuje grupę dietetyczną, a oś y przedstawia masę ciała. Ramka przedstawia rozstęp międzykwartylowy (IQR), który zawiera 50% danych. Linia wewnątrz pudełka reprezentuje medianę wagi, a wąsy rozciągają się do wartości minimalnej i maksymalnej w granicach 1,5-krotności IQR. Kropki reprezentują wartości odstające poza wąsami. Z wykresu widać, że mediana masy ciała jest najniższa w przypadku diety a, najwyższa w przypadku diety c, a pośrednia w przypadku diety b. Dieta b ma najszerszy rozkład wag, z największym zakresem wartości i największą liczbą wartości odstających.

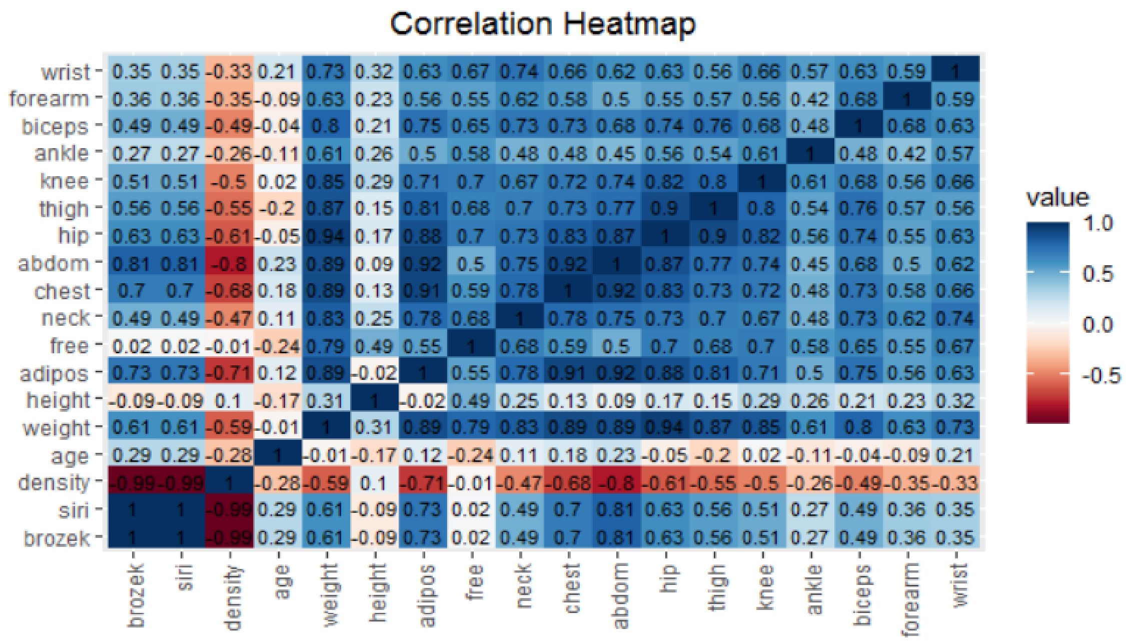


Kołowa macierz korelacji

Wykres przedstawia kołową macierz korelacji. Z tego wykresu widać, że masa ma silną ujemną korelację z a i dodatnią korelację z c. Jest to zgodne z wynikami modelu regresji, które otrzymaliśmy wcześniej, gdzie współczynnik dla a był ujemny, a współczynnik dla c był dodatni. Wykres potwierdza związek między zmiennymi niezależnymi a zmienną zależną w modelu. Dodatkowo widzimy, że istnieje umiarkowana dodatnia korelacja między a i c, co sugeruje, że diety mogą być w jakiś sposób powiązane.

Model podstawowy + współliniowość

Następnymi danymi, które będziemy analizować jest zbiór danych “fat”. Naszą zmienną zależną będzie nadgarstek. Zanim jednak przejdziemy do ostatecznego “modelu” przyjrzyjmy się mapie ciepła korelacji i zobaczmy, jak skorelowane są zmienne.



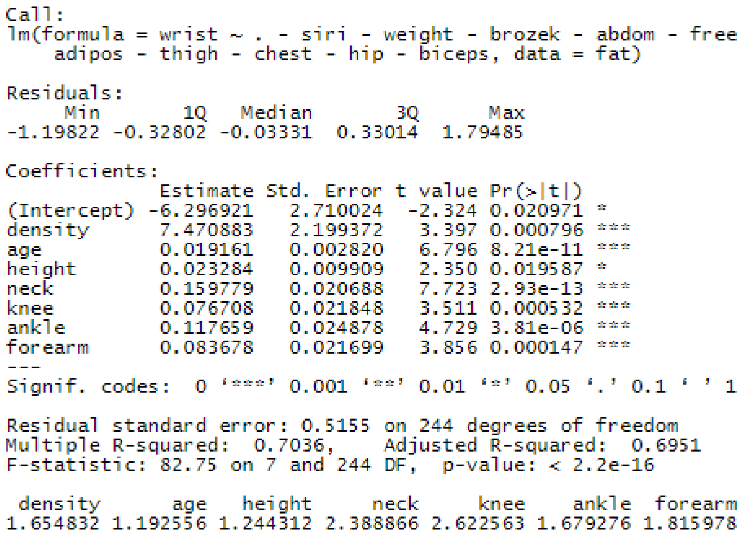
Jak możemy zauważyć korelacja między siri i brozek jest niezwykle wysoka. Wynosi ona wartość 1. Oprócz tego możemy zauważyć również, że wysoce skorelowane są także zmienne thigh i hip, a wartość wynosi 0,9. Ogólnie zmienne adipos, weight mają wysokie korelacje z innymi zmiennymi. Możemy również zauważyć, że między density, a brozek i między density, a siri istnieje również silna zależność liniowa tyle,że ujemna co oznacza, że gdy jedna zmienna rośnie, druga zmienna ma tendencję spadkową. Pokazuje nam to istnienie współliniowości. Może to być problematyczne dla naszego modelu ze względu na to, że może to utrudnić określenie indywidualnych efektów każdej zmiennej niezależnej na zmienną zależną. Dlatego w takich przypadkach konieczne może być rozważenie usunięcia silnie skorelowanych zmiennych. Za chwilę zobaczymy jak wygląda taki model po usunięciu silnie skorelowanych zmiennych. Za naszą zmienną zależną posłuży zmienna “wrist”.

Uzasadnienie merytoryczne

Nasz model bada związek między obwodem nadgarstka a innymi pomiarami ciała, takimi jak wzrost, waga czy procent tkanki tłuszczowej. Obwód nadgarstka jest często używany jako wskaźnik wielkości szkieletu i może być używany do oszacowania rozkładu tkanki tłuszczowej i ryzyka różnych schorzeń. Dlatego zrozumienie czynników związanych z obwodem nadgarstka może dostarczyć cennych informacji do oceny ogólnego stanu zdrowia i ryzyka chorób.

Analiza modelu:

Przeanalizujmy teraz wyniki, które otrzymaliśmy. Jak można zauważyć usunęliśmy aż 10 zmiennych. Jednak na chwilę się tutaj zatrzymajmy, bo tak naprawdę moglibyśmy usunąć 9 zmiennych zamiast 10. Zmienną, którą usunęłam dodatkowo jest zmienna “biceps”. We wcześniejszym modelu „biceps” miał wartość p równą 0,100758, czyli większą niż powszechnie stosowany próg istotności statystycznej wynoszący 0,05. W modelu, który będziemy analizować i w którym usunięto „biceps”, skorygowana wartość R-kwadrat spadła tylko nieznacznie z 0,6972 do 0,6951, a wartości p pozostałych zmiennych nie zmieniły się dramatycznie. W związku z tym zasadne jest według mnie usunięcie z modelu zmiennej „biceps”. Wracając jednak już do analizy przedstawionego modelu



```
Call:
lm(formula = wrist ~ . - siri - weight - brozek - abdom - free -
    adipos - thigh - chest - hip - biceps, data = fat)

Residuals:
     Min       1Q   Median       3Q      Max
-1.19822 -0.32802 -0.03331  0.33014  1.79485

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.296921   2.710024  -2.324 0.020971 *
density      7.470883   2.199372   3.397 0.000796 ***
age          0.019161   0.002820   6.796 8.21e-11 ***
height       0.023284   0.009909   2.350 0.019587 *
neck         0.159779   0.020688   7.723 2.93e-13 ***
knee         0.076708   0.021848   3.511 0.000532 ***
ankle        0.117659   0.024878   4.729 3.81e-06 ***
forearm      0.083678   0.021699   3.856 0.000147 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5155 on 244 degrees of freedom
Multiple R-squared:  0.7036,    Adjusted R-squared:  0.6951
F-statistic: 82.75 on 7 and 244 DF,  p-value: < 2.2e-16

 density      age   height     neck     knee    ankle  forearm
1.654832 1.192556 1.244312 2.388866 2.622563 1.679276 1.815978
```

Analiza modelu podstawowego + współliniowość

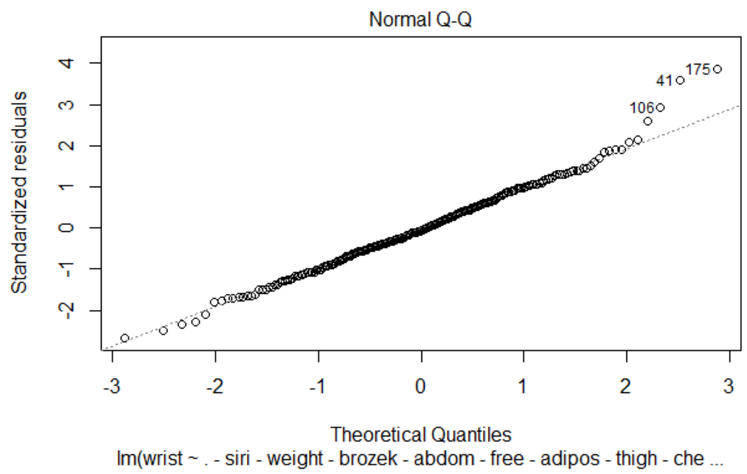
Skupmy się najpierw na analizie reszt. Reszty to różnice między obserwowanym obwodem nadgarstka a obwodem nadgarstka przewidywanym przez model. Minimalna wartość resztkowa wynosi -1,19822, a maksymalna - 1,79485, co wskazuje, że przewidywania modelu mogą się mylić nawet o około 1,8 cm. Mediana reszty jest bliska zeru, co wskazuje, że model jest przeciętnie dość dokładny

Współczynniki pokazują szacunkową zmianę obwodu nadgarstka związaną ze wzrostem o jedną jednostkę w każdej zmiennej niezależnej. Na przykład przy utrzymaniu wszystkich innych zmiennych na stałym poziomie, wzrost obwodu przedramienia o jedną jednostkę wiąże się ze wzrostem obwodu nadgarstka o 0,083678 jednostki. Albo na przykład utrzymując wszystkie inne zmienne na stałym poziomie, wzrost obwodu szyi o jedną jednostkę wiąże się ze wzrostem obwodu nadgarstka o 0,159779 jednostki.

Patrząc na wartości p, możemy zobaczyć, że wszystkie zmienne niezależne zawarte w modelu np. wiek, szyja, wzrost) są statystycznie istotne, ponieważ ich wartości p są mniejsze niż 0,05. Natomiast resztkowy błąd standardowy wynosi 0,5155 cm, co oznacza, że przewidywania modelu są zwykle błędne średnio o około 0,5 cm. Wielokrotna wartość R-kwadrat wynosząca 0,7036 sugeruje, że model wyjaśnia około 70% wariancji zmiennej zależnej, a skorygowany R-kwadrat wynosi 0,6951, co oznacza, że po skorygowaniu o liczbę zmiennych niezależnych w modelu, 69,51% wariancji zmiennej zależnej można wyjaśnić zmiennymi niezależnymi w modelu. Statystyka F wynosi 82,75 z 7 stopniami swobody licznika i 244 stopniami swobody mianownika z wartością p mniejszą niż 0,001, co wskazuje, że model jest wysoce istotny. Ostatnią miarą, którą będziemy analizować jest VIF, który ocenia stopień, w jakim zmienne niezależne w modelu są wzajemnie skorelowane. W przypadku naszego modelu żadna z wartości VIF nie przekracza 5, a nawet maksymalna ma poniżej 3, co sugeruje, że w modelu nie ma poważnej współliniowości.

Podsumowanie

Oszacowania współczynników dla zmiennych w modelu pokazują, że gęstość, wiek, obwód szyi, obwód kolana, obwód kostki i obwód przedramienia mają statystycznie istotny związek z obwodem nadgarstka, podczas gdy wzrost jest marginalnie istotny. Model ma dobre dopasowanie (wielokrotny R-kwadrat: 0,7036), a skorygowany R-kwadrat sugeruje, że model może wyjaśnić 69,51% zmienności obwodu nadgarstka.



Wykres Normal Q-Q

Wykres, który znajduje się powyżej to normalny wykres Q-Q. Pokazuje rozkład reszt w porównaniu z oczekiwanym rozkładem normalnym. Jak widzimy wykres ma w przybliżeniu linię prostą, oznacza to, że reszty mają w przybliżeniu rozkład normalny. Czyli założenie o normalności może być spełnione. Jednak warto zauważyć, że są również obserwacje, które są odstające, ale na szczęście jest ich niewiele.

Model z moderacją

Ostatnim modelem, który będziemy analizować jest model z moderacją. Tak jak w poprzednim modelu będziemy wykorzystywać zbiór danych “fat”. W naszym modelu zmienną Z (moderatorem) będzie zmienna “thigh”, natomiast zmienną X - zmienna “chest” i zmienną Y - zmienna “abdom”.

Merytoryczne uzasadnienie:

Model, który zostanie przedstawiony będzie sugerował, że zarówno obwód klatki piersiowej, jak i obwód uda są pozytywnie związane z tłuszczem brzucha. Może to być przydatne w dziedzinach takich jak zdrowie i sprawność fizyczna, w których badacze lub praktycy są zainteresowani przewidywaniem tkanki tłuszczowej w jamie brzusznej na podstawie pomiarów fizycznych takich obszarów jak klatka piersiowa czy uda. Ponadto model ten można wykorzystać do identyfikacji osób, które mogą być bardziej narażone na pewne schorzenia związane z wysokim poziomem tkanki tłuszczowej w jamie brzusznej.

Analiza modelu:

Zmienną zależną jest „abdom”, która przedstawia obwód brzucha, a zmiennymi niezależnymi są „xc” (scentrowany obwód klatki piersiowej) i „zc” (scentrowany obwód bioder), jak również ich termin interakcji „xc:zc”.

Współczynnik Intercept (Y), który wynosi 92,322948 to szacowana wartość zmiennej zależnej (abdom), gdy obie zmienne niezależne (chest i thigh) mają wartość zero. W tym przypadku zerowa wartość dla każdej ze zmiennych niezależnych nie ma sensu, ale współczynnik ten jest nadal ważny dla modelu, aby dokonywać na nim analiz. Oprócz Intercept mamy współczynnik dla scentrowanego obwodu klatki piersiowej (xc) wynoszący 0,968787. Tak więc, na przykład, jeśli obwód klatki piersiowej wzrośnie o jeden cm, model przewiduje, że obwód brzucha wzrośnie średnio o 0,968787 cm, zakładając, że wszystkie inne zmienne będą na stałym poziomie. Natomiast współczynnik dla scentrowanego obwodu uda (zc), który wynosi 0,405890 możemy interpretować, że jeśli obwód uda wzrośnie o jeden cm, to model przewiduje, że obwód brzucha wzrośnie średnio o 0,405890 cm, zakładając, że wszystkie inne zmienne będą na stałym poziomie. Współczynnik terminu interakcji, „xc:zc”, który wynosi 0,007242 przedstawia zmianę wpływu „xc” (scentrowany obwód klatki piersiowej) na obwód brzucha przy wzroście „zc” (scentrowany obwód uda) o jedną jednostkę. Innymi słowy, przedstawia on zmianę zależności między obwodem klatki piersiowej, a obwodem brzucha, gdy uwzględniony zostanie obwód uda. Znak dodatni wskazuje, że wpływ obwodu klatki piersiowej na obwód brzucha jest silniejszy, gdy obwód uda jest większy.

Wartość R-kwadrat wynosi 0.8618, co oznacza, że zmienne niezależne wyjaśniają około 86% zmienności zmiennej zależnej. Skorygowana wartość R-kwadrat jest niewiele niższa i wynosi ona 0,8601. Patrząc na wartości p, możemy zauważyć, że wszystkie zmienne niezależne zawarte w modelu są statystycznie istotne, ponieważ ich wartości p są mniejsze niż 0,05. Resztkowy błąd standardowy wynosi 4,033. Natomiast statystyka F, która wynosi 515 z 3 stopniami swobody licznika i 248 stopniami swobody mianownika, a także z wartością p, która jest mniejsza niż 0,001, wskazuje nam, że model jest wysoce istotny.

```
Call:
lm(formula = abdom ~ xc + zc + xc * zc, data = fat)

Residuals:
     Min       1Q   Median       3Q      Max 
-11.7076  -2.6530   0.1556   2.7936   9.7125 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 92.322948   0.277933 332.177  < 2e-16 ***
xc           0.968787   0.044285  21.876  < 2e-16 ***
zc           0.405890   0.072019   5.636 4.72e-08 ***
xc:zc        0.007242   0.003501   2.068   0.0396 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.033 on 248 degrees of freedom
Multiple R-squared:  0.8618,	Adjusted R-squared:  0.8601 
F-statistic: 515.3 on 3 and 248 DF,  p-value: < 2.2e-16


Call:
lm(formula = abdom ~ xc + zc + xc * zc, data = fat)

Coefficients:
(Intercept)           xc           zc        xc:zc  
  92.322948     0.968787     0.405890     0.007242  


ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
Level of Significance =  0.05 

Call:
 gvlma(x = fitmod) 
```

Analiza modelu z moderacją

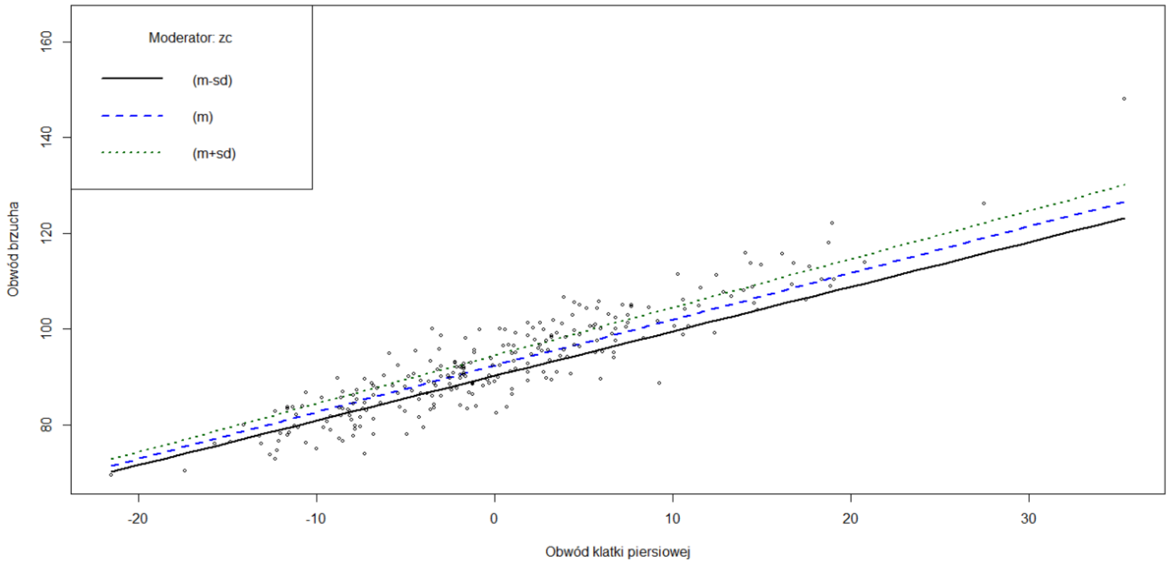
Za pomocą funkcji „gvlma” ocenione zostały założenia modelu liniowego. Założeniami tymi są: (1) normalność reszt (skośność i kurtoza), (2) liniowość związku między zmiennymi niezależnymi a zmienną zależną, (3) homoskedastyczność reszt oraz (4) brak jakichkolwiek pominiętych zmiennych, które mogą być ważne dla prognozowania zmiennej zależnej. Wynik wskazuje, że wszystkie cztery założenia są spełnione, ponieważ global stat i każda indywidualna statystyka testowa mają wartości p większe niż poziom istotności 0,05.

	Value <dbl>	p-value <dbl>	Decision <chr>
Global Stat	3.651498723	0.4552194	Assumptions acceptable.
Skewness	0.385626529	0.5346073	Assumptions acceptable.
Kurtosis	0.001925178	0.9650026	Assumptions acceptable.
Link Function	1.768394047	0.1835810	Assumptions acceptable.
Heteroscedasticity	1.495552969	0.2213569	Assumptions acceptable.

5 rows

Podsumowanie:

Model ten jest modelem regresji liniowej z "abdom" jako zmienną zależną oraz xc (scentrowany obwód klatki piersiowej), zc (scentrowany obwód ud) i ich terminem interakcji xc:zc jako zmiennymi niezależnymi. Model ma na celu przewidywanie obwodu brzucha na podstawie zmiennych xc i zc. Tabela współczynników pokazuje, że wszystkie zmienne są istotne statystycznie na 5% poziomie istotności. Termin przecięcia ma szacunkową wartość 92,322948. Współczynniki dla xc, zc i xc:zc wynoszą odpowiednio 0,968787, 0,405890 i 0,007242. Ocena założeń modelu liniowego za pomocą globalnego testu na 4 stopniach swobody sugeruje, że założenia modelu są akceptowalne. Wielokrotny R-kwadrat modelu wynosi 0,8618, co wskazuje, że około 86,18% zmienności zmiennej zależnej można wyjaśnić zmiennymi niezależnymi.



Wykres pokazuje zależność między obwodem klatki piersiowej (xc) a obwodem brzucha (y) przy uwzględnieniu obwodu uda (zc). Oś x reprezentuje wartości obwodu klatki piersiowej, a oś y reprezentuje wartości obwodu brzucha. Punkty na wykresie przedstawiają zaobserwowane wartości obwodu brzucha dla różnych wartości obwodu klatki piersiowej, przy zachowaniu stałego obwodu uda. Niebieska linia na wykresie pokazuje przewidywane wartości obwodu brzucha dla różnych wartości obwodu klatki piersiowej, w oparciu o model regresji liniowej z xc, zc i xc*zc jako zmiennymi niezależnymi. Wykres wskazuje dodatni związek między obwodami klatki piersiowej i brzucha, kontrolując obwód uda. Oznacza to, że wraz ze wzrostem obwodu klatki piersiowej wzrasta również przewidywana wartość obwodu brzucha. Ale jak widać na wykresie istnieją również wartości odstające, szczególnie wśród osób o większej masie ciała, które według wykresu mają większe obwody ud niż można by przewidzieć na podstawie obwodu klatki piersiowej i brzucha.