Zadania z wykładu 12

Krystian Baran 145000

1 czerwca 2021

Spis treści

1 Zadanie 3

Rozwiązać zadanie 5.7 z Krysickiego.

Z trzech różnych wydziałów pewnej uczelni wylosowano po pięciu studentów z każdego roku studiów i obliczono średnią ocen uzyskaną przez każdego studenta w ostatnim semestrze. Uzyskano rezultaty

Rok studiów	Wydział											
	A				B				C			
I	2.6	4.1	3.1	2.4	3.1	2.5	3.3	3.8	2.7	4.2	2.9	3.7
II	2.8	4.3	3.8	3.0	3.9	2.6	3.2	3.3	3.0	4.4	3.9	3.1
III	3.2	4.1	4.8	4.0	3.4	2.9	4.1	2.8	4.0	3.3	3.4	3.0
IV	3.2	3.9	4.2	3.6	3.6	4.4	2.8	3.9	3.7	5.0	2.6	3.4
V	4.0	4.0	3.5	3.8	4.0	3.0	4.5	3.7	3.0	3.8	4.8	3.5

Zakładając, że średnie uzyskiwanych ocen mają rozkłady normalne o tej samej wariancji na poziomie $\alpha = 0.05$, zweryfikować następujące hipotezy:

a) wartości przeciętne średnich ocen dla studentów różnych wydziałów są jednakowe;

b) wartości przeciętne średnich ocen dla różnych lat studiów są jednakowe;

c) wartości przeciętne ocen średnich dla pierwszych dwóch lat są jednakowe;

Wartości z tabeli przepisano do pliku csv w celu wgrania go do programu R.

data	**wydzial**	**rok**
2.6	"A"	"I"
4.1	"A"	"I"
3.1	"A"	"I"
2.4	"A"	"I"
3.1	"B"	"I"
2.5	"B"	"I"
3.3	"B"	"I"
3.8	"B"	"I"
2.7	"C"	"I"
4.2	"C"	"I"
2.9	"C"	"I"
3.7	"C"	"I"
2.8	"A"	"II"
4.3	"A"	"II"
3.8	"A"	"II"
3.0	"A"	"II"
3.9	"B"	"II"
2.6	"B"	"II"
3.2	"B"	"II"
3.3	"B"	"II"
3.0	"C"	"II"
4.4	"C"	"II"
3.9	"C"	"II"
3.1	"C"	"II"
3.2	"A"	"III"
4.1	"A"	"III"
4.8	"A"	"III"
4.0	"A"	"III"
3.4	"B"	"III"
2.9	"B"	"III"
4.1	"B"	"III"
2.8	"B"	"III"
4.0	"C"	"III"
3.3	"C"	"III"
3.4	"C"	"III"
3.0	"C"	"III"
3.2	"A"	"IV"
3.9	"A"	"IV"
4.2	"A"	"IV"
3.6	"A"	"IV"
3.6	"B"	"IV"
4.4	"B"	"IV"
2.8	"B"	"IV"
3.9	"B"	"IV"
3.7	"C"	"IV"
5.0	"C"	"IV"
2.6	"C"	"IV"
3.4	"C"	"IV"
4.0	"A"	"V"
4.0	"A"	"V"
3.5	"A"	"V"
3.8	"A"	"V"
4.0	"B"	"V"
3.0	"B"	"V"
4.5	"B"	"V"
3.7	"B"	"V"
3.0	"C"	"V"
3.8	"C"	"V"
4.8	"C"	"V"
3.5	"C"	"V"

1.1 a)

Dane z pliku csv wgrano pod zmienną "data" w następujący sposób: *data = read.csv("w12zad3.csv", colClasses = c("numeric", "factor", "factor")*.

Test ANOVA natomiast wykonano w następujący sposób: *model = aov(data∼wydzial, data)*. Otrzymano następujący wynik *summary(model) =*

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
wydzial	2	0.345	0.1727	0.437	0.648
Residuals	57	22.502	0.3948		

Ponieważ *p-value*, obliczone w ostatniej kolumnie, jest większe niż $\alpha = 0.05$ wnioskujemy że wartości przeciętne średnich ocen dla studentów różnych wydziałów są sobie równe.

1.2 b)

Test ANOVA przeprowadzono podobnie jak w poprzednim podpunkcie, tj: *model = aov(data∼rok, data)*. Otrzymano następujące wyniki *summary(model) =*

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
rok	4	2.612	0.6531	1.775	0.147
Residuals	55	20.235	0.3679		

Ponieważ *p-value*, obliczone w ostatniej kolumnie, jest większe niż $\alpha = 0.05$ wnioskujemy że wartości przeciętne średnich ocen dla studentów różnych lat studiów są sobie równe.

1.3 c)

Ponieważ zakładamy że dane mają rozkład normalny możemy zastosować test Tukeya w następujący sposób: *tukey = TukeyHSD(model, conf.level = 0.95)* Otrzymano następujący wynik

	diff	lwr	upr	p adj
II-I	0.2416667	-0.45671717	0.9400505	0.8648729
III-I	0.3833333	-0.31505051	1.0817172	0.5364514
IV-I	0.4916667	-0.20671717	1.1900505	0.2865813
V-I	0.6000000	-0.09838384	1.2983838	0.1244862
III-II	0.1416667	-0.55671717	0.8400505	0.9785925
IV-II	0.2500000	-0.44838384	0.9483838	0.8498736
V-II	0.3583333	-0.34005051	1.0567172	0.6004999
IV-III	0.1083333	-0.59005051	0.8067172	0.9921775
V-III	0.2166667	-0.48171717	0.9150505	0.9048593
V-IV	0.1083333	-0.59005051	0.8067172	0.9921775

Dla tego typu testu dostajemy także *p-value* i widzimy że każda ta wartość jest większa od $\alpha = 0.05$. Zatem wnioskujemy że wartości średnie dla pierwszych dwóch lat są jednakowe.

2 Zadanie 4

Korzystając ze wspomagania komputerowego rozwiązać przykład 3 z wykładu.

Jednym z aspektów jakości samochodów osobowych jest koszt naprawy uszkodzeń spowodowanych drobnymi ulicznymi stłuczkami. Decydujące znaczenie mają tu zderzaki. Producent rozważa wprowadzenie nowego typu zderzaków spośród czterech zaprojektowanych typów. Zainstalowano po siedem zderzaków każdego typu na pojazdach popularnej klasy i poddano je próbom zderzania ze ścianą z prędkością 30 km/h. Następnie oszacowano koszty napraw powstałych uszkodzeń (w j.m.). Wyniki są przedstawione w tablicy.

Typ zderzaka			
1	2	3	4
315	285	269	255
288	292	277	287
293	263	273	265
306	249	252	279
299	275	263	241
310	266	251	312
282	252	272	310

a) Przyjmując 5-procentowy poziom istotności zbadać, czy są istotne różnice w kosztach usuwania uszkodzeń dla badanych czterech typów zderzaków.

b) W przypadku występowania różnic ustalić typy zderzaków różniących się ze względu na koszty usuwania awarii.

2.1 a)

Wartości z tablicy wybrano do pliku csv aby wczytać je w R. Wartości zapisano w kolumnie nazwaną "data" a odpowiadające wartościom typy zderzaka zapisano pod kolumną "type" i uwzględniono w R że to ma być kolumna typu "factor". Dane wgrano korzystając z funkcji *data = read.csv("w12zad4.csv")*. Aby sprawdzić czy są istotne różnice w kosztach usuwania uszkodzeń wykorzystano następującą funkcje: *model = aov(data~type, data=data)* która wykonuje test ANOVA jedno kierunkowy na zależność kosztu od typu zderzaka. Poniżej uzyskane tą funkcją wyniki (*summary(model)*).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
type	3	4805	1601.6	5.197	0.00658
Residuals	24	7396	308.2		

Funkcja ta oddaje *p-value* zapisane pod ostatnią kolumną, zatem, przyjmując $\alpha = 0.05$ stwierdzamy że typ zderzaka ma wpływ na koszt usuwania uszkodzeń.

2.2 b)

Ponieważ w podpunkcie **a)** okazało się że występują różnice w kosztach usuwania uszkodzeń więc możemy poszukać które typy różnią się od siebie. W R istnieje funkcja obliczająca test Tukeya na podstawie wcześniej wyznaczonej analizy wariancji (*aov()*). Ten test został wywołany następująco: *TukeyHSD(model, conf.level = 0.95)*.
Otrzymano następującą tabele:

	diff	lwr	upr	p adj
2-1	-30.142857	-56.02790	-4.257812	0.0182410
3-1	-33.714286	-59.59933	-7.829241	0.0074533
4-1	-20.571429	-46.45647	5.313616	0.1540625
3-2	-3.571429	-29.45647	22.313616	0.9807839
4-2	9.571429	-16.31362	35.456474	0.7394841
4-3	13.142857	-12.74219	39.027902	0.5111022

Funkcja ta także zwraca *p-value* zatem, porównując z $\alpha = 0.05$ widzimy że jedynie typ 2 i typ 1 się różnią a także typ 1 i typ 3.
Obliczenia programem potwierdzają wyniki obliczone w przykładzie z wykładu.

3 Zadanie 6

Korzystając ze wspomagania komputerowego rozwiązać przykład 6 z wykładu.

Wycena prywatyzowanego przedsiębiorstwa państwowego poprzedzona jest szczegółową analizą wartości majątku, potencjału produkcyjnego, możliwości przestawienia produkcji, sposobów zabezpieczenia socjalnego pracowników, itp. Szacowanie wartości majątku przeprowadzają specjalistyczne firmy zajmujące się wyceną. Przeszacowanie wartości przedsiębiorstwa zmniejsza szanse prywatyzacji firmy, natomiast zaniżenie wartości zmniejsza przychód z prywatyzacji.

W celu zmniejszenia ryzyka popełnienia błędu przedstawiciel odpowiedniego ministerstwa zamierza porównać średnie oszacowania wartości trzech niezależnych firm wyceniających majątek zanim zleci jednej z nich dokonanie oszacowania wartości rynkowej prywatyzowanego przedsiębiorstwa.

Przedstawiciel zebrał informacje o wycenie majątku tych samych czterech przedsiębiorstw przez każdą z rozważanych trzech firm wyceniających. Uzyskane dane o wycenach (w mln zł) są podane w tabeli.

Firma	Wycena przedsiębiorstwa			
wyceniająca	1	2	3	4
A	4.6	6.2	5.0	6.6
B	4.9	6.3	5.4	6.8
C	4.4	5.9	5.4	6.3

a) Przeprowadzić analizę wariancji dla przeprowadzonych wycen. Na poziomie istotności $\alpha = 0,05$ sprawdzić, czy są istotne różnice między oczekiwanymi wycenami dla zabiegów i bloków.

b) Wyznaczyć 90-procentowy przedział ufności dla różnic między oczekiwanymi wycenami dla firm wyceniających A i B.

Wartości z tabeli zapisano w pliku csv następującej postaci tak aby można było dokonać obliczenia w R.

data	**firma**	**wycena**
4.6	"A"	"1"
6.2	"A"	"2"
5.0	"A"	"3"
6.6	"A"	"4"
4.9	"B"	"1"
6.3	"B"	"2"
5.4	"B"	"3"
6.8	"B"	"4"
4.4	"C"	"1"
5.9	"C"	"2"
5.4	"C"	"3"
6.3	"C"	"4"

3.1 a)

Jak w poprzednim zadaniu wykorzystamy funkcję R-owską *model = aov(data~firma+wycena)*. Wynik tej funkcji można odczytać za pomocą *summary(model)*.

	Df	Sum Sq	Mean Sq	F value	Pr(¿F)	
firma	2	0.260	0.1300	4.179	0.073	.
wycena	3	6.763	2.2544	72.464	4.2e-05	***
Residuals	6	0.187	0.0311			

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Obliczone zostały przez funkcje *p-value* zatem, porównując z $\alpha = 0.05$ widzimy że nie ma różnicy wyceniania pomiędzy firmami, natomiast istnienie różnica wyceny między przedsiębiorstwami. Potwierdzone jest zatem to co zostało obliczone na wykładzie.

3.2 b)

Aby wyznaczyć przedział ufności wykorzystano funkcje *TukeyHSD(model, conf.level = 0.9)* która oddaje następującą tablice:

$firma				
	diff	lwr	upr	p adj
B-A	0.25	-0.06381888	0.56381888	0.1918699
C-A	-0.10	-0.41381888	0.21381888	0.7156978
C-B	-0.35	-0.66381888	-0.03618112	0.0692699
$wycena				
2-1	1.5000000	1.0860287	1.9139713	0.0001924
3-1	0.6333333	0.2193620	1.0473047	0.0178616
4-1	1.9333333	1.5193620	2.3473047	0.0000445
3-2	-0.8666667	-1.2806380	-0.4526953	0.0038512
4-2	0.4333333	0.0193620	0.8473047	0.0851260
4-3	1.3000000	0.8860287	1.7139713	0.0004312

Przedział ufności ma granice zaznaczone pod "lwr" i "upr". Zatem dla różnicy B-A przedział ufności jest następujący:

$$(-0.06381888; 0.56381888)$$

Wynik ten nie zgadza się z wartościami obliczonymi na wykładzie, może to wynikać z tego że funkcja została źle użyta lub że wynik z wykładu jest nie prawidłowy.