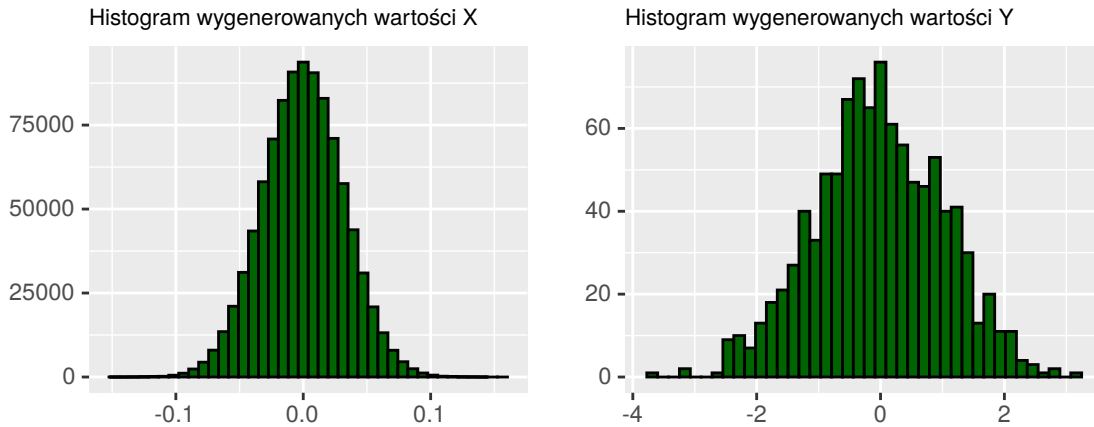


ADZD - lista1

Testowanie wielokrotne

Wstęp

Rozważamy dane wygenerowane w oparciu o teoretyczny model liniowy $Y = X\beta + \epsilon$ z macierzą planu $X_{n \times p}$ z rozkładu $N(0_n, \Sigma = \frac{1}{1000}I_{n \times n})$, $\beta = (3, 3, 3, 3, 3, 0, \dots)^T$ i $\epsilon \in R^n \sim N(0_n, \sigma^2 I_n)$. Gdy traktujemy X jako deterministyczną macierz, Y jako przekształcenie afiniczne wektora ϵ ma rozkład $N(X\beta, \sigma^2 I_n)$. Ustalamy $n = 1000, p = 950$ i $\sigma^2 = 1$.



Dysponując X i Y możemy szacować parametry modelu. Estymator uzyskany metodą najmniejszych kwadratów ma postać

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

i jest wektorem losowym z rozkładu $N(\beta, \sigma^2 (X^T X)^{-1})$. Przedział ufności wyznaczony doświadczalnie dla i -tego elementu jest postaci

$$\hat{\beta}_i \pm t_c s(\beta_i),$$

gdzie $s^2(\hat{\beta}_i) = s^2 (X^T X)^{-1}_{i,i}$ oraz s^2 jest nieobciążonym estymatorem wariancji błędów losowych i wyraża się jako suma kwadratów residuów w utworzonym modelu z p parametrami podzielona przez $n - p$, a t_c to kwantyl rzędu $1 - \frac{\alpha}{2}$ rozkładu Studenta z $n - p$ stopniami swobody. W przypadku, gdy znana jest wariancja błędów losowych, możemy obliczyć **teoretyczny przedział ufności** zamieniając: s^2 na σ^2 ; $X^T X^{-1}_{1,1}$ na wartość oczekiwaną tego wyrazu. **Macierz kowariancji próbkowej** $S := X^T X$ (nie mylić z Σ - teoretyczną kowariancją X) jest rozmiaru $p \times p$ i gdy $n > p - 1$, możemy ją odwrócić. S ma rozkład Wisharta $W_p(\Sigma, n)$ (uogólnienie rozkładu chi-kwadrat na zmienne wielowymiarowe), a S^{-1} **odwrotny rozkład Wisharta** $W_p^{-1}(\Sigma^{-1}, n)$. Wtedy:

$$ES^{-1} = \frac{\Sigma^{-1}}{n - p - 1}$$

Przez testowanie istotności poszczególnych elementów wektora β będziemy rozumieli testowanie hipotez postaci $H_{0,i} : \beta_i = 0$ przeciw $H_{1,i} : \beta_i \neq 0$. Odrzucenie pojedynczej hipotezy zerowej nazwiemy **odkryciem**. W problemie wielokrotnego testowania chcemy kontrolować “sumaryczny” błąd I rodzaju dla hipotez $H_{0,1}, \dots, H_{0,p}$.

FWER (Familywise Error Rate) definiujemy jako prawdopodobieństwo, że zrobimy conajmniej 1 fałszywe odkrycie. Używamy go jako kryterium w przypadku, gdy badanych hipotez jest mało (np. mniej niż 30) i zależy nam na tym, żeby nie robić żadnych fałszywych odkryć. W przypadku większej liczby hipotez próba kontroli FWER może skutkować brakiem jakichkolwiek odkryć. Zamiast tego możemy badać wartość **FDR (False Discovery Rate)**, czyli wartość oczekiwana wyrażenia $FDP = \frac{FD}{\max(1, TD + FD)}$. Gdy wszystkie hipotezy zerowe są prawdziwe, $FDP = FWER$. Ponadto zawsze $FDR \leq FWER$, więc **kontrolując FWER kontrolujemy zawsze FDR**. Podstawową procedurą badania hipotezy na poziomie istotności α jest odrzucenie $H_{0,i}$ jeśli stowarzyszona z nią p-wartość jest mniejsza niż ten poziom - nie gwarantuje ona jednak odpowiedniego ograniczenia FWER i FDR. Popelniamy wówczas średnio $n \cdot \alpha$ błędów I rodzaju. Zakładając, że znamy liczbę prawdziwych hipotez oraz że są one niezależne (w praktyce łamiemy to założenie, ale dla dużych p przybliżenie będzie dobre) możemy obliczyć **teoretyczną wartość FWER**:

$$FWER = P(FD \geq 1) = 1 - P(FD = 0) = 1 - \prod_{i=1}^k (\neg R_i) = 1 - (1 - \alpha)^k,$$

gdzie R_i - odrzucenie $H_{0,i}$ oraz hipotezy $H_{0,1}, \dots, H_{0,k}$ są prawdziwe. W tym przypadku $\max_{k \in \{1, \dots, p\}} FWER = 1 - (1 - \alpha)^p$, dla $p > 1$ mamy $FWER > \alpha$. Istnieją sposoby na ograniczenie błędu, m.in.:

- **korekta Bonferroniego**: p-wartości dzielimy przez liczbę hipotez. Wtedy $FWER = 1 - (1 - \frac{\alpha}{p})^p \leq \alpha$.
- **korekta Benjaminiego-Hochberga**: sortujemy p-wartości rosnąco w ciąg $v_{(1)}, \dots, v_{(p)}$ i znajdujemy $i_0 := \max\{i : v_{(i)} \leq \frac{i}{p}\alpha\}$. Odrzucamy hipotezy H_1, \dots, H_{i_0} - taka procedura gwarantuje $FDR \leq \alpha$ (zazwyczaj nie kontroluje FWER).
- Uwaga: za obiema korektami stoi teoretyczne założenie o niezależności p-wartości i w sytuacji gdy nie jest ono spełnione wyniki mogą odbiegać od teoretycznych ograniczeń. Istnieje twierdzenie mówiące o tym, że w rzeczywistości procedura B-H kontroluje FDR na poziomie $\alpha S(n)$ dla pewnej funkcji $S(n)$.

Zadanie 1

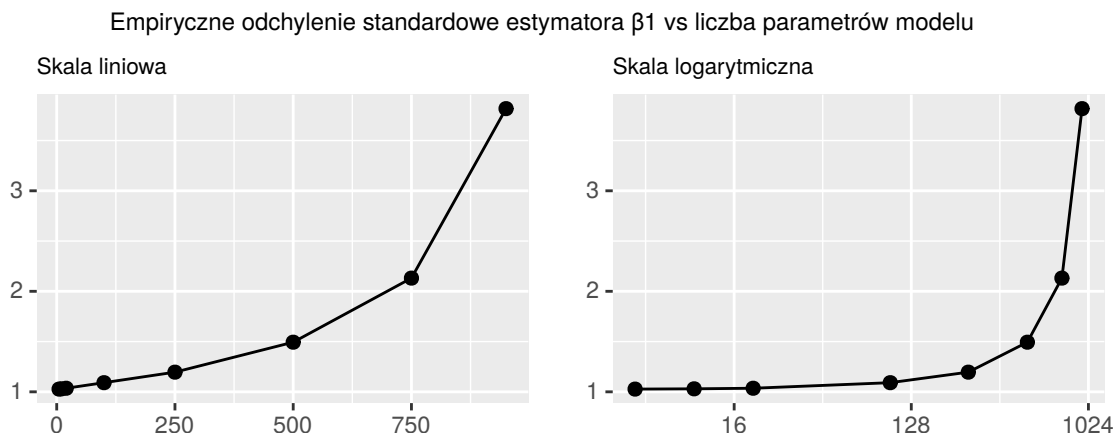
W oparciu o wygenerowane wcześniej dane tworzymy modele uwzględniające pierwsze p kolumn macierzy planu z użyciem funkcji `lm`. Przyjmujemy oznaczenia zmiennych:

- `b1`: estymator $\hat{\beta}_1$
- `s_b1`: empiryczne $s^2(\hat{\beta}_1)$
- `s2`: empiryczne s^2
- `conf_w`: szerokość empirycznego przedziału ufności $\hat{\beta}_1$
- `TD / FD` - prawdziwe / fałszywe odkrycia (`_rodzaj_korekty`)

Table 1: Zadanie 1 - wyniki liczbowe

| p | b1 | s_b1 | s2 | conf_w | TD | FD | TD_Bon | FD_Bon | TD_BH | FD_BH |
|-----|------|------|------|--------|----|-----|--------|--------|-------|-------|
| 5 | 3.08 | 1.03 | 1.05 | 4.03 | 2 | 0 | 2 | 0 | 2 | 0 |
| 10 | 3.10 | 1.03 | 1.05 | 4.04 | 2 | 0 | 2 | 0 | 2 | 0 |
| 20 | 3.37 | 1.04 | 1.05 | 4.06 | 2 | 1 | 2 | 0 | 2 | 0 |
| 100 | 3.55 | 1.09 | 1.05 | 4.28 | 2 | 3 | 0 | 0 | 0 | 0 |
| 250 | 3.38 | 1.20 | 1.04 | 4.69 | 3 | 10 | 0 | 0 | 0 | 0 |
| 500 | 4.56 | 1.49 | 1.05 | 5.87 | 3 | 22 | 0 | 0 | 0 | 0 |
| 750 | 4.81 | 2.13 | 1.01 | 8.40 | 3 | 45 | 0 | 0 | 0 | 0 |
| 950 | 4.00 | 3.82 | 0.58 | 15.34 | 0 | 141 | 0 | 1 | 0 | 1 |

Wykresy: wnioskowanie statystyczne



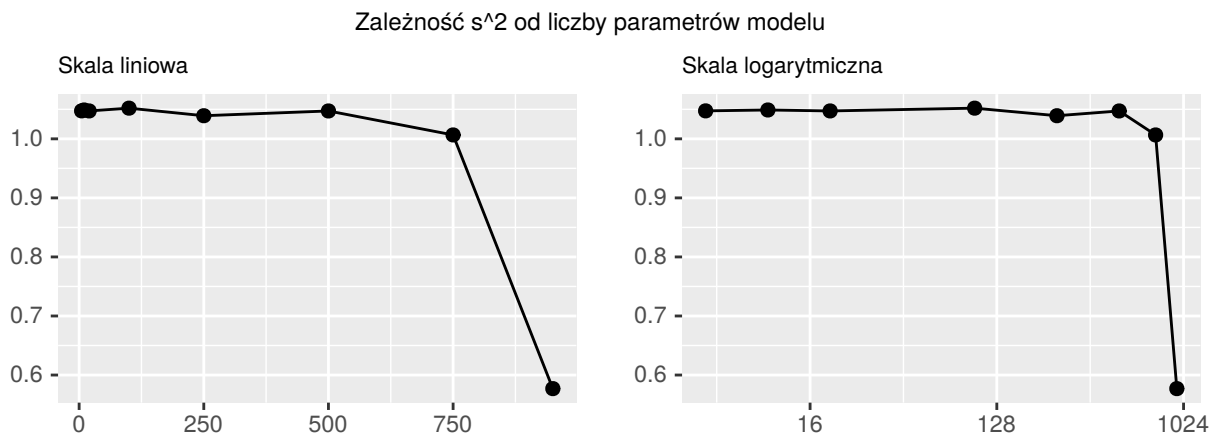
Komentarz: Widać, że wraz ze wzrostem liczby parametrów $s(\hat{\beta}_1)$ rośnie. Przypomnijmy, że liczymy je ze wzoru

$$s^2(\hat{\beta}_i) = s^2(X^T X)^{-1}_{i,i},$$

stąd mamy dwie potencjalne przyczyny takiego zachowania:

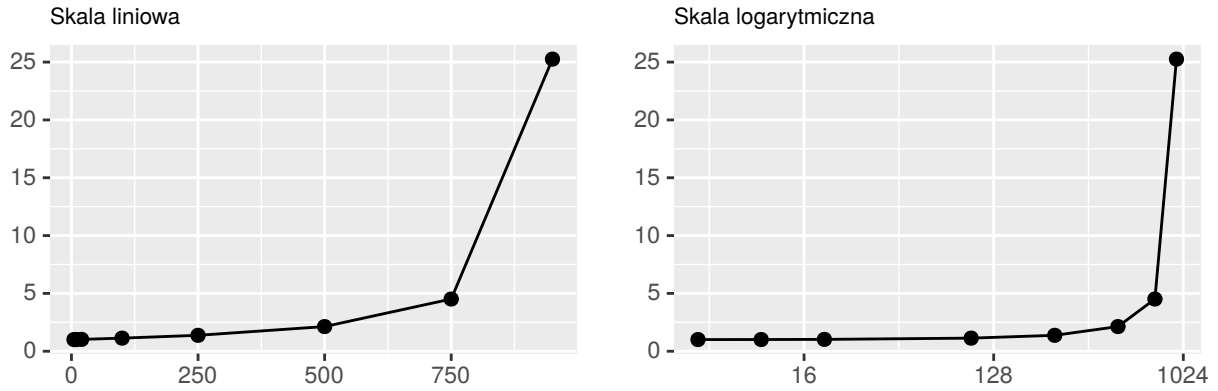
- wraz ze wzrostem liczby parametrów rośnie wartość $s^2 = \frac{1}{n-p} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$, gdzie $\hat{Y} = X\hat{\beta}$ lub
- wyrazy na przekątnej $(X^T X)^{-1}$ są większe dla szerszych macierzy planu

Zależność s^2 od p obrazuje poniższy wykres.



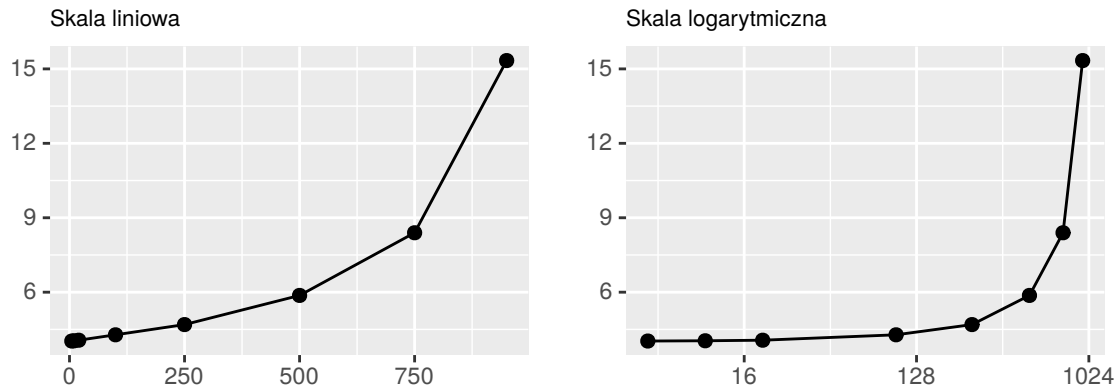
Jak widać, dla $p \leq 750$ estymator jest dobry - wartości utrzymują się na bardzo podobnym poziomie zbliżonym do prawdziwej wartości σ^2 . Można domyślać się, że dla kolejnej wartości $p = 950$ równej liczbie kolumn macierzy X dochodzi do zjawiska overfittingu i błąd staje się bardzo mały. Dodatkowo można zauważyć, że skoro $s^2 \sim 1$ to β_1 jest w przybliżeniu proporcjonalne do wyrazu $X^T X_{1,1}$ dla $p \leq 750$ i to właśnie jego wzrost będzie odpowiadać za wzrost $s^2(\hat{\beta}_1)$.

Zależność $(X^T X)^{-1}_{1,1}$ od liczby parametrów modelu



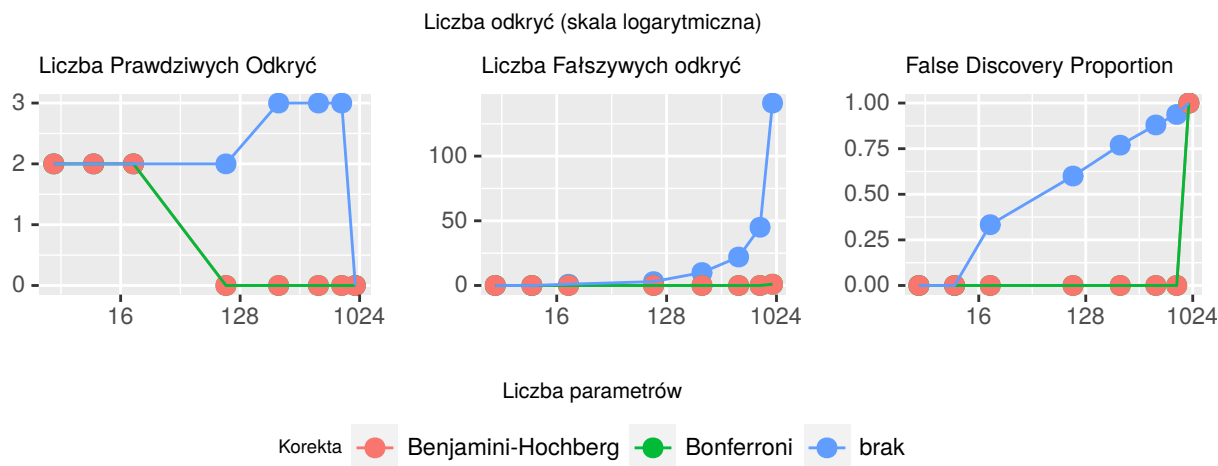
Można uzasadnić to patrząc na wzór wartości oczekiwanej ES^{-1} - jeśli zwiększamy p , to mianownik maleje, a więc wyrazy macierzy S^{-1} są coraz większe.

Szerokość przedziału ufności (95%) dla β_1 vs liczba parametrów modelu



Komentarz: Szerokość przedziału ufności jest równa $2t_c\sigma(\hat{\beta}_1)$, a więc proporcjonalna do odchylenia standardowego estymatora. Z tego powodu wykresy ich zależności od p mają dokładnie ten sam kształt (zmienia się tylko zakres wartości na osi Y).

Wykresy: liczba odkryć



Komentarz: W przypadku pojedynczego eksperymentu nie możemy mówić o FDR ani FWER, dlatego porównujemy tylko liczby odkryć i False Discovery Proportion. Widać, że zastosowanie korekt poskutkowało zmniejszeniem liczby zarówno prawdziwych jak i fałszywych odkryć (w zasadzie nie dochodzi do praktycznie żadnych odkryć). Na trzecim wykresie widać że przy $p = 950$ dla wszystkich procedur wszystkie odkrycia są fałszywe. W przypadku korekt liczba tych odkryć to 1.

Zadanie 2

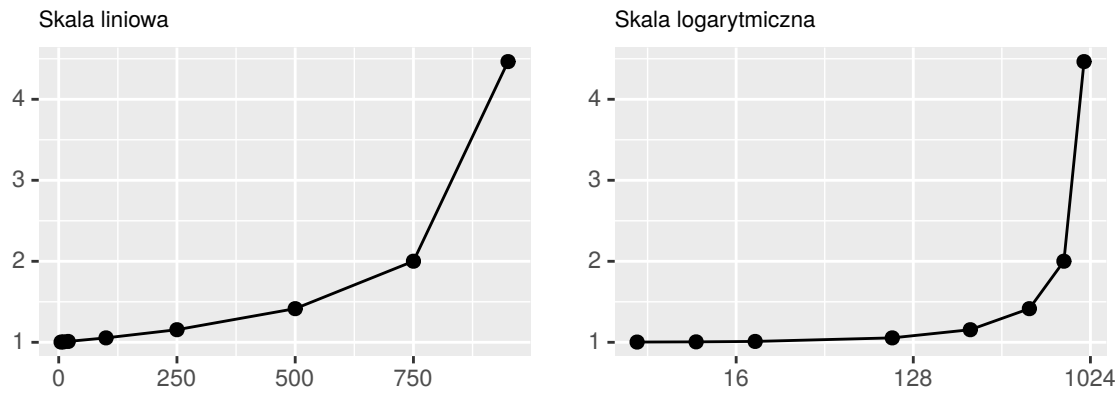
W tym zadaniu powtarzamy 1000-krotnie eksperyment z zadania 1 w celu uzyskania uśrednionych na dużej próbie wyników.

Table 2: Zadanie 2 - wyniki liczbowe uśrednione dla 1000 iteracji

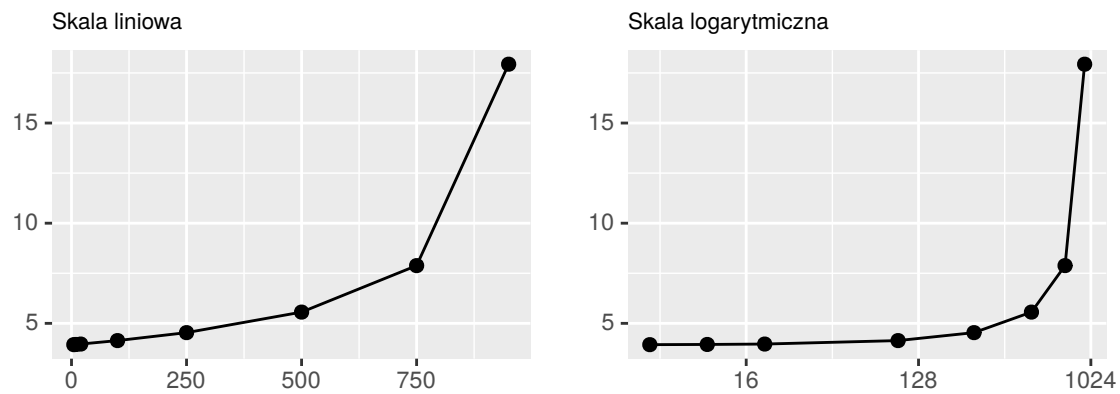
| p | b1 | s_b1 | s2 | conf_w | TD | FD | FWER | FDP | TD_Bon | FD_Bon | FWER_Bon | FDP_Bon | TD_BH | FD_BH | FWER_BH | FDP_BH |
|-----|------|------|------|--------|------|-------|------|------|--------|--------|----------|---------|-------|-------|---------|--------|
| 5 | 2.98 | 1.00 | 1.00 | 3.94 | 4.27 | 0.00 | 0.00 | 0.00 | 3.35 | 0.00 | 0.00 | 0.00 | 4.17 | 0.00 | 0.00 | 0.00 |
| 10 | 2.98 | 1.01 | 1.00 | 3.95 | 4.25 | 0.26 | 0.24 | 0.05 | 2.87 | 0.02 | 0.02 | 0.01 | 3.79 | 0.13 | 0.12 | 0.03 |
| 20 | 2.97 | 1.01 | 1.00 | 3.97 | 4.22 | 0.75 | 0.54 | 0.13 | 2.42 | 0.03 | 0.03 | 0.01 | 3.24 | 0.16 | 0.14 | 0.03 |
| 100 | 2.96 | 1.05 | 1.00 | 4.14 | 4.10 | 4.67 | 0.99 | 0.51 | 1.30 | 0.05 | 0.05 | 0.03 | 1.73 | 0.17 | 0.14 | 0.05 |
| 250 | 2.97 | 1.16 | 1.00 | 4.54 | 3.72 | 12.02 | 1.00 | 0.75 | 0.65 | 0.05 | 0.05 | 0.04 | 0.81 | 0.12 | 0.10 | 0.05 |
| 500 | 2.99 | 1.42 | 1.00 | 5.56 | 2.79 | 24.72 | 1.00 | 0.89 | 0.20 | 0.06 | 0.05 | 0.05 | 0.24 | 0.08 | 0.07 | 0.05 |
| 750 | 2.97 | 2.00 | 1.00 | 7.88 | 1.63 | 37.57 | 1.00 | 0.96 | 0.03 | 0.06 | 0.05 | 0.05 | 0.03 | 0.07 | 0.05 | 0.05 |
| 950 | 3.06 | 4.47 | 0.99 | 17.94 | 0.50 | 46.52 | 1.00 | 0.99 | 0.00 | 0.04 | 0.03 | 0.03 | 0.00 | 0.14 | 0.03 | 0.03 |

Wykresy: wnioskowanie statystyczne

Empiryczne odchylenie standardowe estymatora β_1 vs liczba parametrów modelu



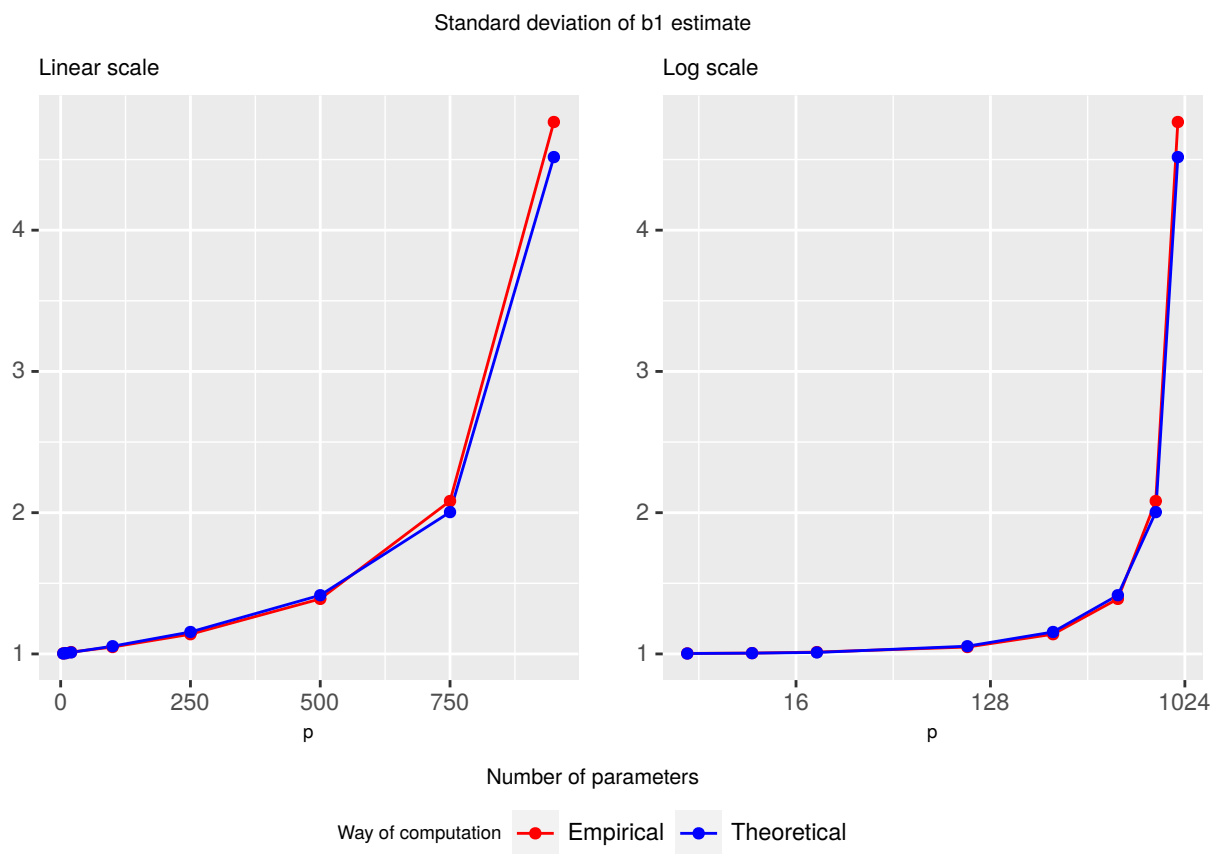
Szerokość przedziału ufności (95%) dla β_1 vs liczba parametrów modelu



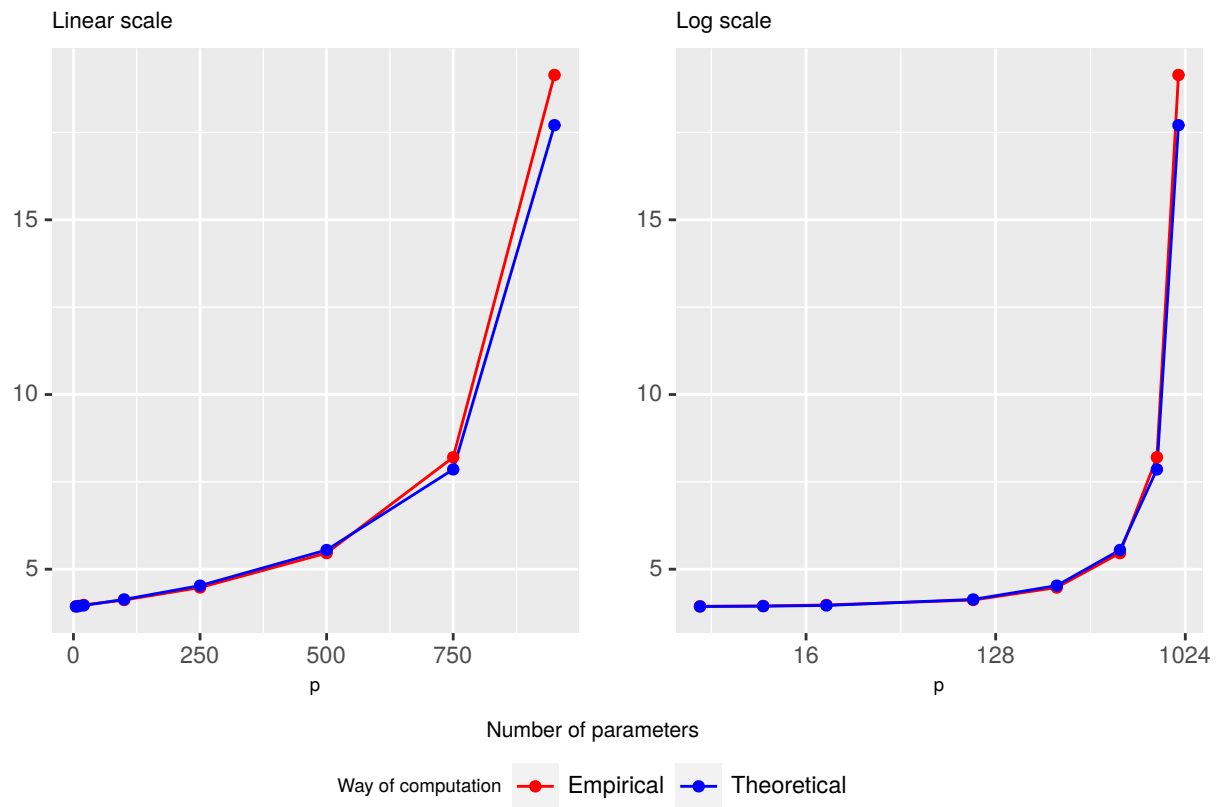
Nie ma znaczących różnic względem zadania 1.

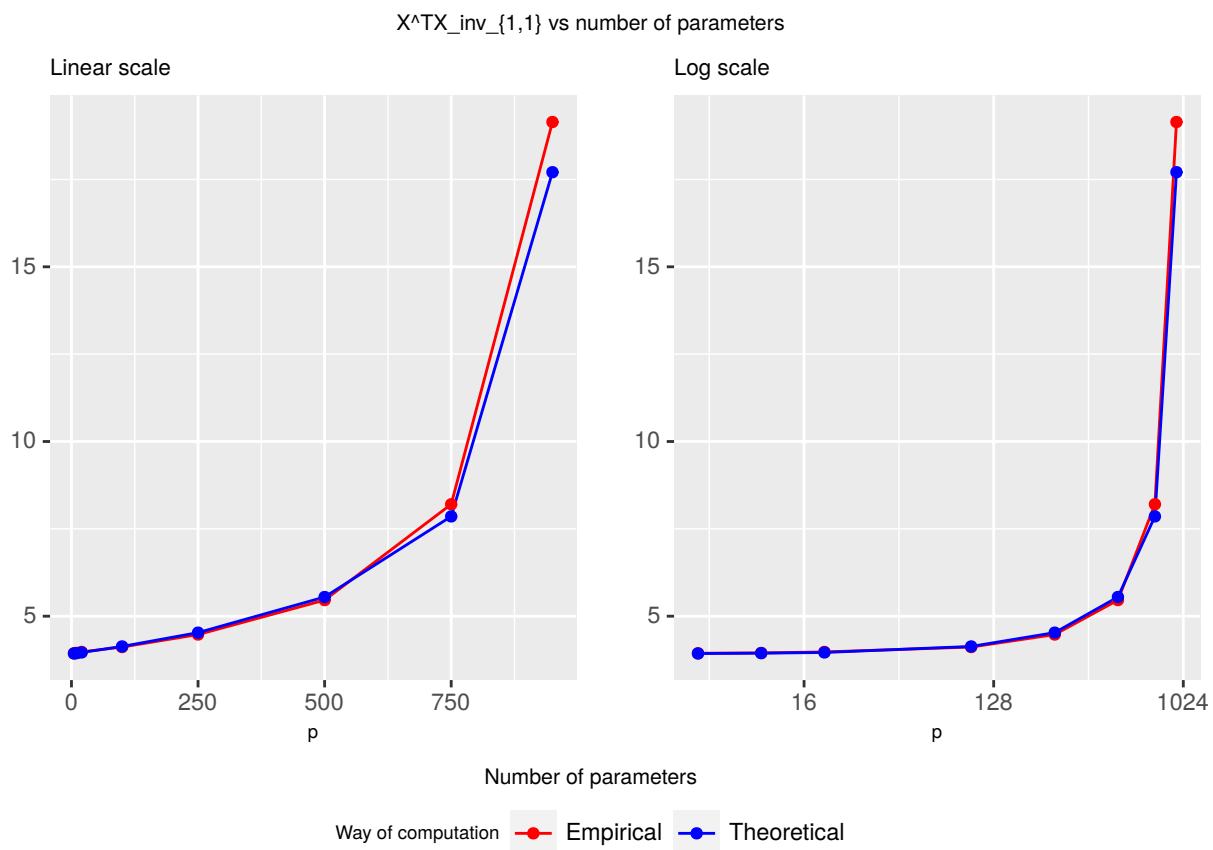
Wartości teoretyczne

Porównamy wartości empiryczne $s(\hat{\beta}_1)$, szerokości przedziału ufności $X^T X_{1,1}$ z teoretycznymi oszacowaniami. Dla 1000 replikacji eksperymentu wyniki są prawie równe.

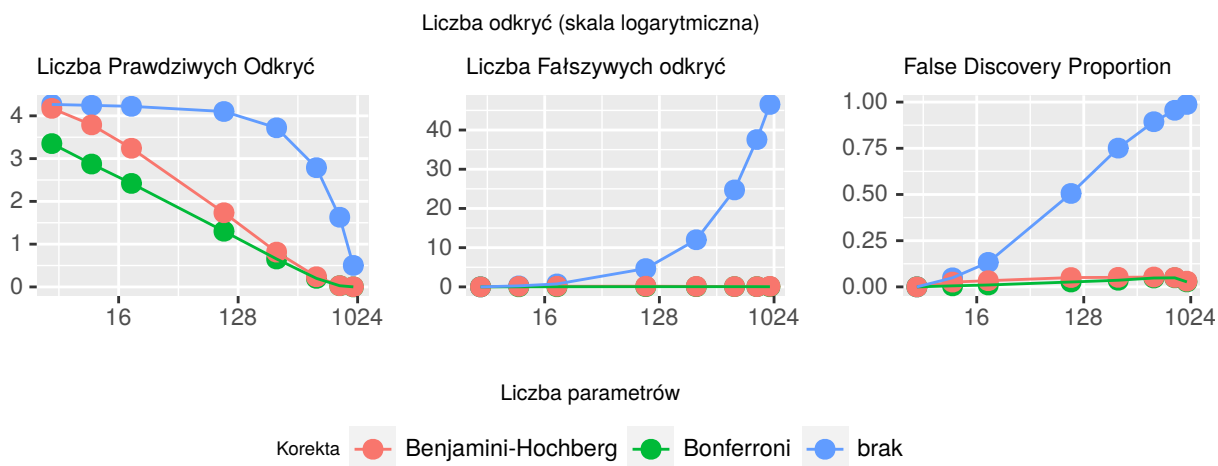


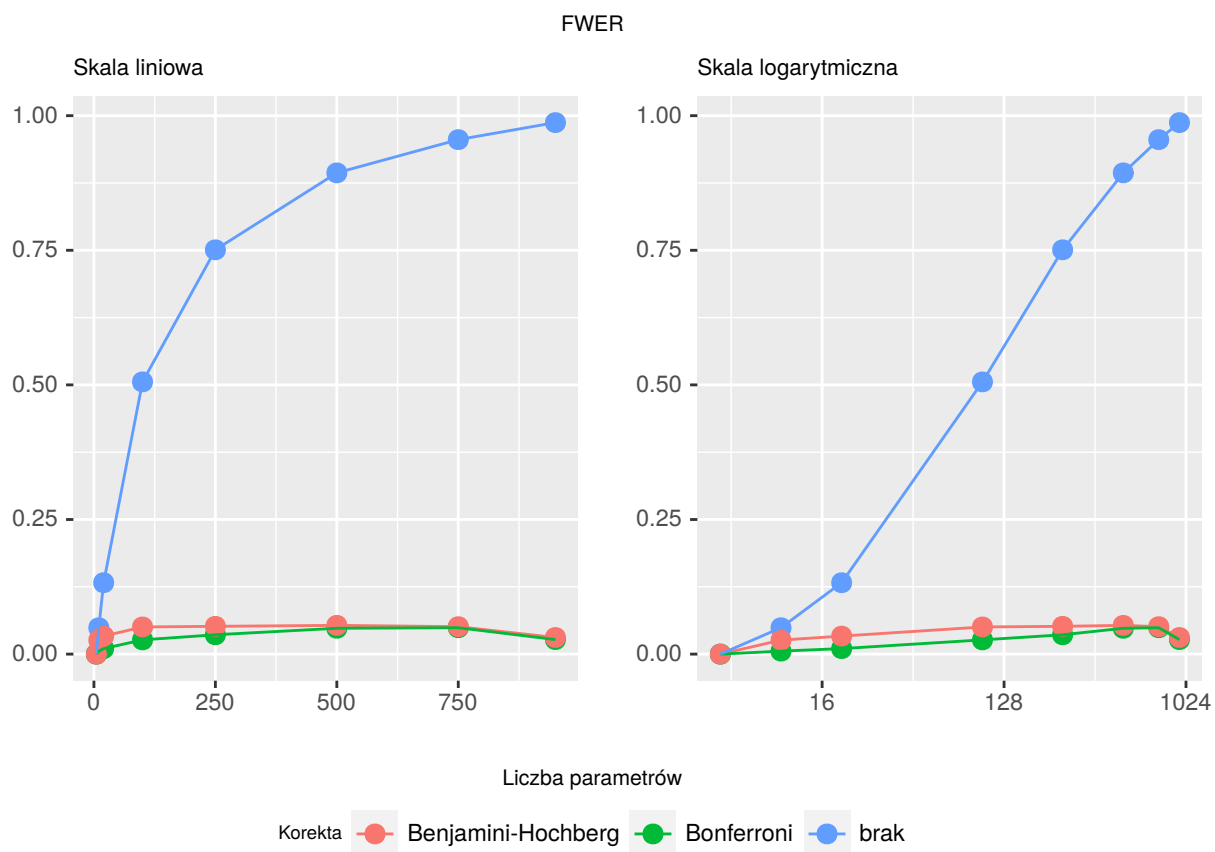
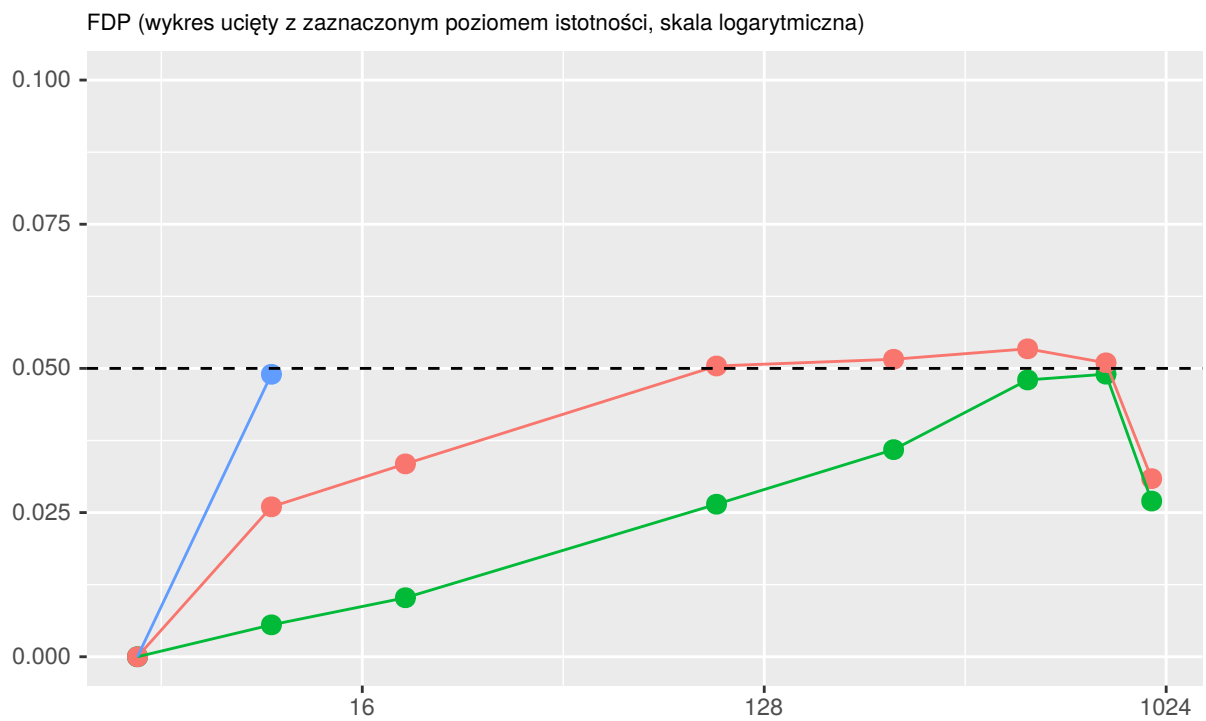
90% confidence interval of b1 estimate

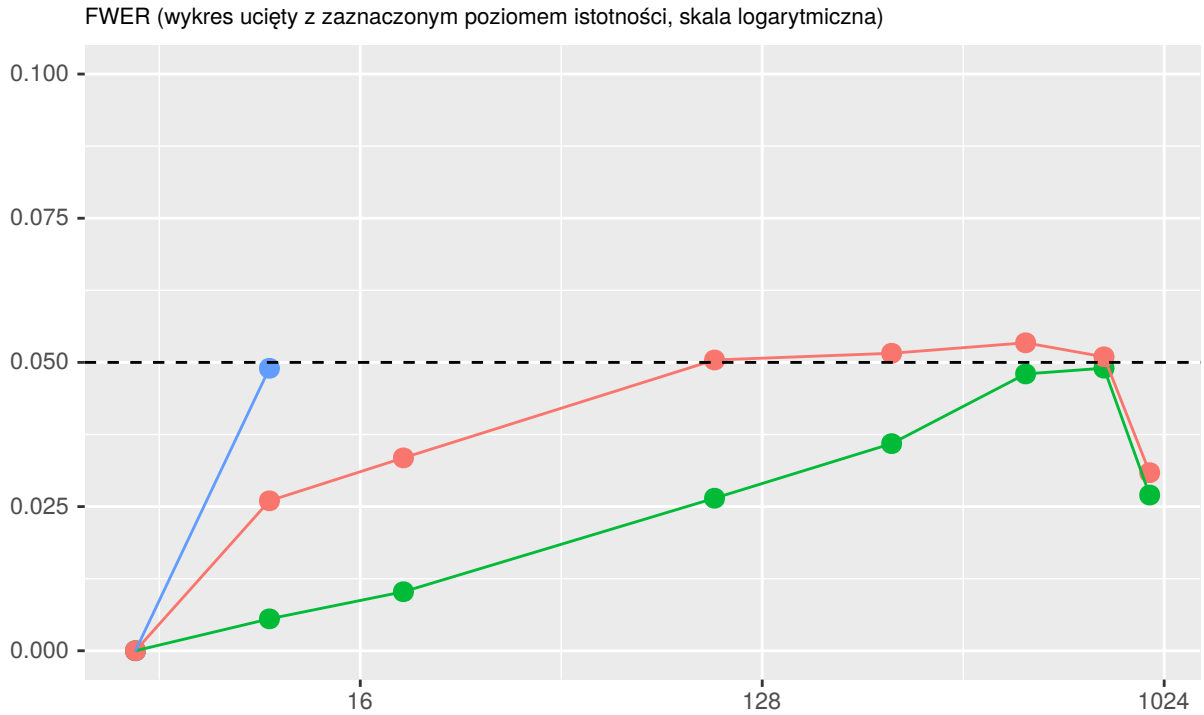




Wykresy: liczba odkryć







Komentarz: Najwięcej prawdziwych odkryć mamy dla procedury bez korekty - zgodnie z oczekiwaniami. Można zauważyć że średnia liczba prawdziwych odkryć dla metody B-H jest większa niż dla metody Bonferroni, bo również zgadza się z teorią - B-H jest mniej rygorystyczna, co tłumaczy również dlaczego na wykresach widać że FWR i FWER są mniejsze dla metody Bonferroni'ego. Wartości dla obu metod trzymają się założonych ograniczeń (w przypadku metody B-H widać lekkie wyjście wykresu poza linię poziomu istotności, ale należy to zignorować, gdyż tak małą różnicę można tłumaczyć błędem losowym eksperymentu i spodziewamy się, że zniknęłaby ona dla większej liczby iteracji). Na wykresach można zauważyć również, że rzeczywiście $FDR \leq FWER$. Wyniki potwierdziły teoretyczne oszacowania.