Modele i Wnioskowanie Statystyczne
Laboratorium 5

Sprawozdanie

Igor Markiewicz

Zadanie 1
Model :

$$X \sim \mathcal{N}(\mu_X, \sigma^2)$$
$$Y \sim \mathcal{N}(\mu_Y, \sigma^2)$$
$$\overline{X} - \overline{Y} \sim \mathcal{N}\left(\mu_X - \mu_Y, \sigma^2\left(n^{-1} + m^{-1}\right)\right)$$

a)

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$\overline{X} \approx 0,146$
$\overline{Y} \approx 0,8884$
$\overline{X} - \overline{Y} = -0,7424$

b)

$$s_i^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\overline{x} - x_i)^2$$

$s_X^2 \approx 1,155962$
$s_Y^2 \approx 0,9695833$

$$s^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{m+n-2}$$

$s^2 = 1,04946$

c) $s\sqrt{n^{-1} + m^{-1}} \approx 0,6872$

d) Dwustronny, ponieważ hipoteza zerowa i alternatywna się dopełniają w przeciwieństwie do testu jednostronnego.

e) Testowana hipoteza ma postać :

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

A statystyka testowa :

$$T = \frac{\overline{X} - \overline{Y}}{s\sqrt{n^{-1} + m^{-1}}}$$

Wykorzystano funkcję $t.test$ z parametrami $two.sided = TRUE$ oraz $var.equal = TRUE$. W efekcie otrzymano $p - value \approx 0,316$.

f) Dla poziomu $\alpha = 0,1$ nie mamy podstaw do odrzucenia hipotezy zerowej.

Zadanie 2

a) Ze względu na brak informacji o równości wariancji :

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2)$$

$$Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

postanowiono zastosować statystykę :

$$T = \frac{\overline{X} - \overline{Y}}{s}$$

$$s^2 = \frac{s_X^2}{n} + \frac{s_Y^2}{m}$$

dla testu :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Statystyka T ma w przybliżeniu rozkład t – Studenta, z liczbą stopni swobody (po zaokrąlgeniu do liczby całkowitej) :

$$d \approx \frac{(s^2)^2}{\frac{(s_X^2/n)^2}{n-1} + \frac{(s_Y^2/m)^2}{m-1}}$$

Wykorzystano funkcję *t.test* z parametrami $two.sided = TRUE$ oraz $var.equal = FALSE$. W efekcie otrzymano $p - value \approx 0,0541$, co oznacza że na poziomie istotności $\alpha = 0,1$ możemy odrzucić hpotezę zerową na rzecz alternatywnej, zaś dla $\alpha = 0,05$ nie mamy przeciwskazań do przyjęcia hipotezy zerowej.

b) Wykonano test *Manna – Whitneya – Wilcoxona (test sumy rang Wilcoxona)* sprawdzający czy rozkłady dwóch zbiorów próbek różnią się o stałą wartość μ (przyjętą tutaj jako 0) (m.in przy założeniu niezależności obserwacji, równej wariancji oraz równości rozkładów) :

$$H_0 - \text{dystrybuanty rozkładów dla dwóch grup są przesunięte o } 0$$

$$H_1 - \text{dystrybuanty rozkładów dla dwóch grup są przesunięte wartość inną niż } 0$$

Zastosowano funkcję *wilcox.test* z parametrem domyślnym *alternative="two.sided"* oraz $paired = FALSE$, w efekcie czego otrzymano $p - value = 0,063$, a więc dla poziomu $\alpha = 0,1$ możemy odrzucić hipotezę zerową na rzecz alternatywnej, zaś dla $\alpha = 0,05$ nie ma przeciwskazań do przyjęcia hipotezy zerowej.

c) Test niepraparametryczny z podpunktu **b)** wydaje się być lepszym wyjściem, z racji na braku konieczności zakładania rozkładów normalnych, jednak posiada dużo dodatkowych obostrzeń do jego stosowania.

d) Wykorzystując bootstrap nieparametryczny, zliczono ilość sytuacji w której czas żywotności łożysk z pierwszego materiału był większy od czasu żywotności łożysk wykonanych z drugiego materiału, a następnie podzielono tę liczbę przez liczbę próbkowań otrzymując estymatę szukanego prawdpodobieństwa ok. $0,750232$.

Zadanie 3

Model - dopasowanie funkcji liniowej w grupę punktów, tak aby minimalizowała błąd kwadratowy. Zakłada się że niepewności mają rozkład $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$:

$$y = \beta_0 + \beta_1 x$$

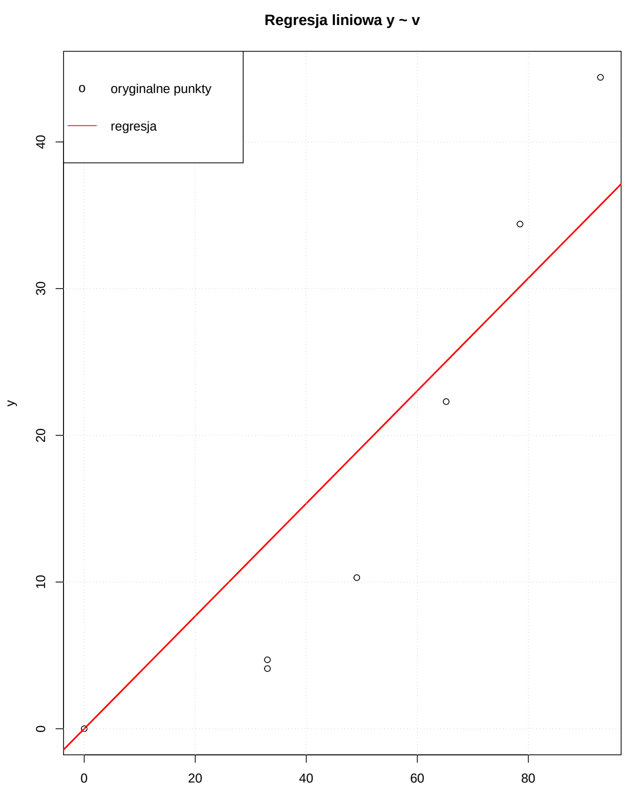
$$E(Y_i|X_i = x) = \beta_0 + \beta_1 x + \epsilon_i$$

$$\hat{\beta}_1 = \frac{\overline{XY} - \overline{X}\overline{Y}}{\overline{X^2} - \overline{X}^2}$$

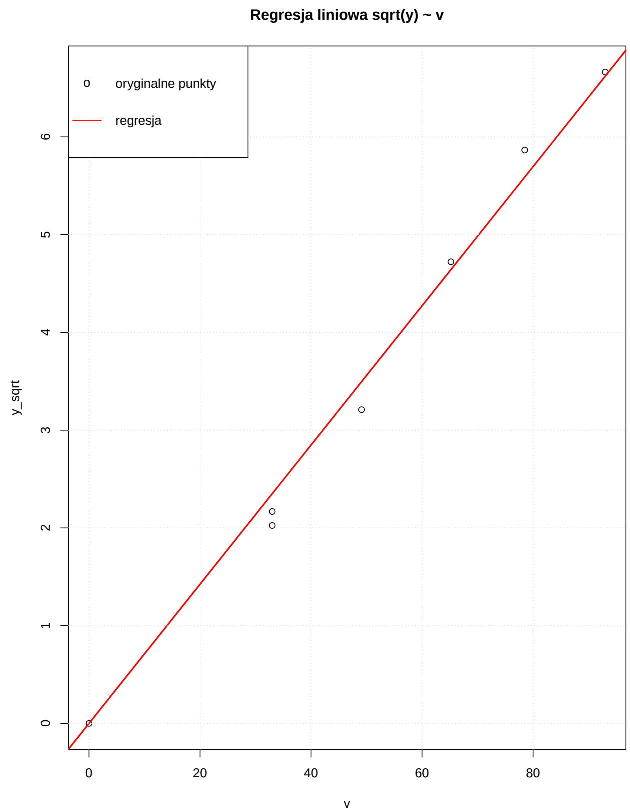
$$\hat{\beta}_0 = \hat{Y} - \hat{\beta}_1 \overline{X}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \right)^2$$

Z racji na założenia fizyczne zjawiska, przyjęto $\beta_0 = 0$. Z takich samych względów dodano punkt pomiarowy $[0, 0]$.



Rys. 1: Regresja liniowa dla $y \sim v$



Rys. 2: Regresja liniowa dla $\sqrt{y} \sim v$

	$\hat{\beta}_1$
$y \sim v$	0,38390
$\sqrt{y} \sim v$	0,071199

Tab. 1: Wyliczone współczynnik $\hat{\beta}_1$

Współczynniki dopasowania modelu do danych:

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - f(x_i))^2}{\sum_{i=1}^{n}(y_i - \overline{y})}$$

$$R^2_{adjusted} = 1 - \left(1 - R^2\right) \frac{n-1}{n-p-1}$$

Gdzie n to liczebność próbki, a p liczba zmienny objaśniających.

	R^2	$R^2_{adjusted}$
$y \sim v$	0,918	0,9044
$\sqrt{y} \sim v$	0,9975	0,997

Tab. 2: Wyliczone współczynniki R^2 oraz $R^2_{adjusted}$

Wnioski :

Możemy zauważyć że graficznie lepiej dopasowan jest zależność $\sqrt{y} \sim v$, co potwierdzają również współczynniki R^2 oraz $R^2_{adjusted}$ (im bliżej jedności tym lepiej model jest dopasowany).