Weryfikacja hipotez statystycznych

Część 1

- Składnik losowy modelu

$$\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$$

- Estymator wektora parametrów

$$b = \beta + (X'X)^{-1}X'\varepsilon$$

- Rozkład estymatora wektora parametrów

$$b \sim \mathcal{N}\left(\beta, \sigma^2 (X'X)^{-1}\right)$$

- Reszty są punktowymi nieobciążonymi oszacowaniami składnika losowego

$$\frac{\varepsilon}{\sigma} \sim \mathcal{N}(0, \mathbb{I})$$

- Macierz M jest nielosowa, więc estymator wariancji składnika losowego

$$S^2 = \frac{e'e}{\sigma^2} = \frac{\varepsilon'}{\sigma} M \frac{\varepsilon}{\sigma}$$

- Zatem rozkład estymatora wariancji składnika losowego

$$S^2 \sim \chi^2_{N-k}$$

- Niezależność estymatora b oraz estymatora e (W. H. Greene (2003))
- Szkic dowodu pełnego (H. Cramer (1958) Metody matematyczne w statystyce)
- Dowód uproszczony

Twierdzenie 4.4, Greene (2003)

Jeżeli ε jest wektorem losowym o rozkład normalnym wówczas rozkład estymatora Metody Najmniejszych Kwadratów dla wektora parametrów b jest niezależny od rozkładu wektora reszt e i wszystkich jego funkcji.

Twierdzenie Fishera (1921)

Jeżeli para zmiennych losowych (X, Y) ma łączny rozkład normalny to

$$f(\bar{x}, \bar{y}, S_x, S_y, R) = f_1(\bar{x}, \bar{y}) f_2(S_x, S_y, R)$$

oraz $f_1(\cdot)$ i $f_2(\cdot)$ są niezależne.

Próba o dużej liczbie obserwacji

$$b \sim \mathcal{N}(\beta, \Sigma_b)$$

$$b_k \sim \mathcal{N}(\beta_k, [\Sigma_b]_{kk})$$

$$\frac{b_k - \beta_k}{se(b_k)} = \frac{b_k - \beta_k}{\sqrt{[\Sigma_b]_{kk}}} \sim \mathcal{N}(0, 1)$$

Próba o małej liczbie obserwacji

$$t = \frac{b_k - \beta_k}{se(b_k)} =$$

$$t = \frac{\frac{b_k - \beta_k}{\sqrt{\sigma^2 (X'X)^{-1}_{kk}}}}{\sqrt{\frac{\frac{e'e}{\sigma^2}}{(N-k)}}}$$

- Zakładamy prawdziwość H_0

stan natury / decyzja	przyjęcie H_0	odrzucenie H_0
H_0 prawdziwa	OK	błąd 1 rodzaju
H_1 prawdziwa	błąd 2 rodzaju	OK

Źródło: Niemiro (1999)

- Wartość p (ang. *p-value*) jest prawdopodobieństwem popełnienia błędu pierwszego rodzaju
- Jeżeli wartość p jest mniejsza od poziomu istotności testu (α) to mamy przesłanki do odrzucenia H_0

Weryfikacja hipotezy prostej

- Hipotezy

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

- Statystyka testowa

$$\frac{b_k}{se(b_k)} \sim t_{\alpha, N-k}$$

lzarobki	Coef.	Std. Err.	t
_Iplec_2	-.2810472	.0057012	-49.30
wiek	.027991	.0019732	14.19
wiek2	-.0002725	.0000254	-10.73
_Iwyksztal~2	-.2687475	.0149665	-17.96
_Iwyksztal~3	-.2706048	.0093954	-28.80
_Iwyksztal~4	-.2394147	.0126095	-18.99
_Iwyksztal~5	-.4051296	.0091924	-44.07
_Iwyksztal~6	-.5260538	.0106914	-49.20
_Iwyksztal~7	-.680285	.0790781	-8.60
_cons	5.689387	.0380066	149.69

$$Pr(\mid t \mid > t_{\frac{\alpha}{2}}) = Pr\left(\left|\frac{b_k - \beta_k}{\hat{se}(b_k)}\right| > t_{\frac{\alpha}{2}}\right)$$

$$Pr\left(\left|\frac{b_k - \beta_k}{\hat{se}(b_k)}\right| > t_{\frac{\alpha}{2}}\right) = 2\left[1 - F_{t,N-k}(t_{\frac{\alpha}{2}})\right]$$

$$Pr\left(b_k - \hat{se}(b_k)t_{\frac{\alpha}{2}} < \beta_k < b_k + \hat{se}(b_k)t_{\frac{\alpha}{2}}\right)$$

```
      Source |       SS       df       MS              Number of obs =   16162
-------------+------------------------------           F(  9, 16152) =  626.18
       Model |  660.681047     9  73.4090052            Prob > F      =  0.0000
    Residual |   1893.5542 16152   .11723342            R-squared     =  0.2587
-------------+------------------------------           Adj R-squared =  0.2582
       Total |  2554.23525 16161  .158049332            Root MSE      =  .34239

------------------------------------------------------------------------------
    lzarobki |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Iplec_2 |  -.2810472   .0057012   -49.30   0.000    -.2922222   -.2698722
        wiek |    .027991   .0019732    14.19   0.000     .0241232    .0318588
       wiek2 |  -.0002725   .0000254   -10.73   0.000    -.0003223   -.0002227
_Iwyksztal~2 |  -.2687475   .0149665   -17.96   0.000    -.2980836   -.2394115
_Iwyksztal~3 |  -.2706048   .0093954   -28.80   0.000    -.2890209   -.2521887
_Iwyksztal~4 |  -.2394147   .0126095   -18.99   0.000    -.2641307   -.2146987
_Iwyksztal~5 |  -.4051296   .0091924   -44.07   0.000    -.4231476   -.3871115
_Iwyksztal~6 |  -.5260538   .0106914   -49.20   0.000    -.5470101   -.5050975
_Iwyksztal~7 |   -.680285   .0790781    -8.60   0.000    -.8352868   -.5252832
       _cons |   5.689387   .0380066   149.69   0.000     5.614889    5.763884
------------------------------------------------------------------------------
```

```
      Source |       SS       df       MS              Number of obs =   16162
-------------+------------------------------           F(  9, 16152) =  626.18
       Model |  660.681047     9  73.4090052            Prob > F      =  0.0000
    Residual |   1893.5542 16152   .11723342            R-squared     =  0.2587
-------------+------------------------------           Adj R-squared =  0.2582
       Total |  2554.23525 16161  .158049332            Root MSE      =  .34239

------------------------------------------------------------------------------
    lzarobki |      Coef.   Std. Err.      t    P>|t|     [99% Conf. Interval]
-------------+----------------------------------------------------------------
    _Iplec_2 |  -.2810472   .0057012   -49.30   0.000    -.2957343   -.2663601
        wiek |    .027991   .0019732    14.19   0.000     .0229077    .0330744
       wiek2 |  -.0002725   .0000254   -10.73   0.000      -.000338   -.0002071
_Iwyksztal~2 |  -.2687475   .0149665   -17.96   0.000    -.3073033   -.2301917
_Iwyksztal~3 |  -.2706048   .0093954   -28.80   0.000    -.2948087   -.2464009
_Iwyksztal~4 |  -.2394147   .0126095   -18.99   0.000    -.2718984    -.206931
_Iwyksztal~5 |  -.4051296   .0091924   -44.07   0.000    -.4288103   -.3814488
_Iwyksztal~6 |  -.5260538   .0106914   -49.20   0.000    -.5535962   -.4985113
_Iwyksztal~7 |   -.680285   .0790781    -8.60   0.000    -.8840007   -.4765693
       _cons |   5.689387   .0380066   149.69   0.000     5.591476    5.787297
------------------------------------------------------------------------------
```

```
      Source |       SS       df       MS              Number of obs =   16162
-------------+------------------------------           F(  9, 16152) =  626.18
       Model |  660.681047     9  73.4090052            Prob > F      =  0.0000
    Residual |   1893.5542 16152   .11723342           R-squared     =  0.2587
-------------+------------------------------           Adj R-squared =  0.2582
       Total |  2554.23525 16161  .158049332           Root MSE      =  .34239

------------------------------------------------------------------------------
    lzarobki |      Coef.   Std. Err.      t    P>|t|     [90% Conf. Interval]
-------------+----------------------------------------------------------------
    _Iplec_2 |  -.2810472   .0057012   -49.30   0.000    -.2904254    -.271669
        wiek |    .027991   .0019732    14.19   0.000     .0247451    .0312369
       wiek2 |  -.0002725   .0000254   -10.73   0.000    -.0003143   -.0002307
_Iwyksztal~2 |  -.2687475   .0149665   -17.96   0.000    -.2933667   -.2441284
_Iwyksztal~3 |  -.2706048   .0093954   -28.80   0.000    -.2860598   -.2551498
_Iwyksztal~4 |  -.2394147   .0126095   -18.99   0.000    -.2601566   -.2186727
_Iwyksztal~5 |  -.4051296   .0091924   -44.07   0.000    -.4202505   -.3900086
_Iwyksztal~6 |  -.5260538   .0106914   -49.20   0.000    -.5436406    -.508467
_Iwyksztal~7 |   -.680285   .0790781    -8.60   0.000    -.8103643   -.5502057
       _cons |   5.689387   .0380066   149.69   0.000     5.626868    5.751905
------------------------------------------------------------------------------
```