

Preizkus hipotez

Milena Kovač

10. januar 2013

Preizkušanje hipotez

Metode **najmanjših kvadratov (OLS)**,
tehtanih najmanjših kvadratov (WLS) in
splošnih najmanjših kvadratov (GLS)

sprejemajo svoje zaključke na osnovi

- **vsote kvadratov**, ki ga pojasnijo posamezni vplivi,
- **vsote kvadratov za ostanek**, ki praviloma služi za primerjavo, in
- **stopinj prostosti** = število ocenljivih parametrov

Dnevni prirast (g/dan) pri mladica

Žival	Pasma	Mesec	DP
1	SL	JAN	540
2	SL	JAN	550
3	SL	FEB	550
4	SL	FEB	580
5	LW	JAN	520
6	LW	FEB	500
7	LW	FEB	490
8	NL	JAN	560
9	NL	JAN	550
10	NL	FEB	600
11	NL	FEB	610
Skupaj			

Preveritev modela: dnevni prirast pri mladica

$$y_i = \mu + e_i$$

$$e_i \sim IIDN(0, \sigma_e^2)$$

$$y_i \sim IIDN(\mu, \sigma_e^2)$$

Za preveritev modela potrebujemo:

- skupno vsoto kvadratov (TSS)
- vsoto kvadratov za model (MSS)
- vsoto kvadratov za ostanek (RSS)

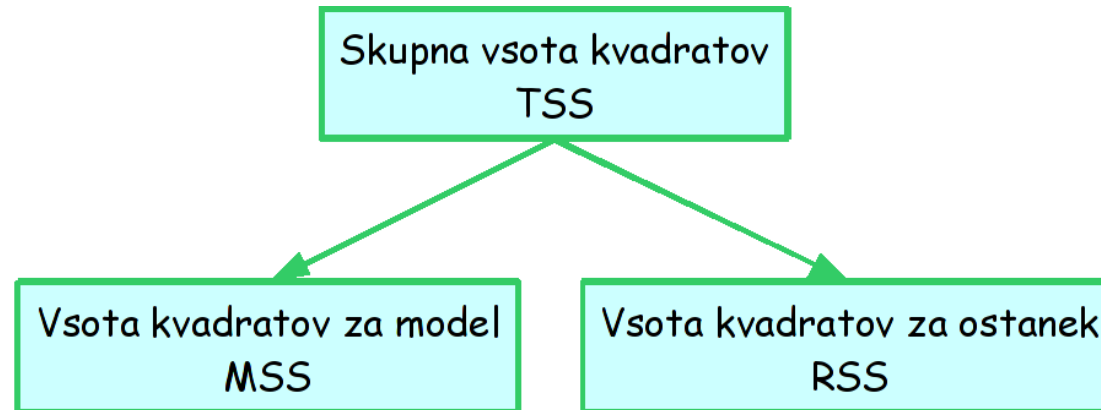
Skupna vsota kvadratov

- Total Sum of Square, TSS
- vsota kvadriranih opazovanj

Skalarna oblika zapisa

$$TSS = \sum y_i^2 = 540^2 + 550^2 + \dots 610^2 = 3341700$$

Razdelitev TSS



- izjema: model vsebuje samo srednjo vrednost

$$\hat{\mu} = (540 + 550 + 550 + \dots + 610) / 11 = 550$$

Vsota kvadratov za model

- model sum of square, MSS
- enaka vsoti kvadratov za pričakovane vrednosti

$$MSS = \sum [E(y_i)]^2 = \sum [\hat{\mu}]^2 = 11 * 550^2 = 3327500$$

- **praviloma** odstranimo vsoto kvadratov za srednjo vrednost

$$MSS = \sum [E(y_i) - \hat{\mu}]^2$$

- naš model ima samo srednjo vrednost, zato bomo uporabili izjemoma prvo enačbo!

Vsota kvadratov za ostanek

- residual sum of square, RSS
- vsota kvadriranih ostankov
- v sistematskem modelu velja:

$$RSS = \sum (y_i - E(y_i))^2$$

Dnevni prirast (g/dan) pri mladich

Žival	P_i	M_j	DP	y_i^2	$\hat{\mu}$	$(\hat{\mu})^2$	\hat{e}_i	\hat{e}_i^2
1	SL	JAN	540		550			
2	SL	JAN	550		550			
3	SL	FEB	550		550			
4	SL	FEB	580		550			
5	LW	JAN	520		550			
6	LW	FEB	500		550			
7	LW	FEB	490		550			
8	NL	JAN	560		550			
9	NL	JAN	550		550			
10	NL	FEB	600		550			
11	NL	FEB	610		550			
Σ								
				TSS		MSS		RSS

Izračun vsote kvadratov za ostanek pri modelu

Žival	P_i	M_j	DP	y_i^2	$\hat{\mu}$	$(\hat{\mu})^2$	\hat{e}_i	\hat{e}_i^2
1	SL	JAN	540	540 ²	550			
2	SL	JAN	550	550 ²	550			
3	SL	FEB	550	550 ²	550			
4	SL	FEB	580	580 ²	550			
5	LW	JAN	520	520 ²	550			
6	LW	FEB	500	500 ²	550			
7	LW	FEB	490	490 ²	550			
8	NL	JAN	560	560 ²	550			
9	NL	JAN	550	550 ²	550			
10	NL	FEB	600	600 ²	550			
11	NL	FEB	610	610 ²	550			
Σ				3341700				

- vsote kvadratov za ostanek

Žival	P_i	M_j	DP	y_i^2	$\hat{\mu}$	$(\hat{\mu})^2$	\hat{e}_i	\hat{e}_i^2
1	SL	JAN	540	540 ²	550	550 ²		
2	SL	JAN	550	550 ²	550	550 ²		
3	SL	FEB	550	550 ²	550	550 ²		
4	SL	FEB	580	580 ²	550	550 ²		
5	LW	JAN	520	520 ²	550	550 ²		
6	LW	FEB	500	500 ²	550	550 ²		
7	LW	FEB	490	490 ²	550	550 ²		
8	NL	JAN	560	560 ²	550	550 ²		
9	NL	JAN	550	550 ²	550	550 ²		
10	NL	FEB	600	600 ²	550	550 ²		
11	NL	FEB	610	610 ²	550	550 ²		
Σ				3341700		332750		

- vsota kvadratov za ostanek

Žival	P_i	M_j	DP	y_i^2	$\hat{\mu}$	$(\hat{\mu})^2$	\hat{e}_i	\hat{e}_i^2
1	SL	JAN	540	540^2	550	550^2	-10	100
2	SL	JAN	550	550^2	550	550^2		
3	SL	FEB	550	550^2	550	550^2		
4	SL	FEB	580	580^2	550	550^2		
5	LW	JAN	520	520^2	550	550^2		
6	LW	FEB	500	500^2	550	550^2		
7	LW	FEB	490	490^2	550	550^2		
8	NL	JAN	560	560^2	550	550^2		
9	NL	JAN	550	550^2	550	550^2		
10	NL	FEB	600	600^2	550	550^2		
11	NL	FEB	610	610^2	550	550^2		
Σ				3341700		3327500		

- vsota kvadratov za ostanek

Žival	P_i	M_j	DP	y_i^2	$\hat{\mu}$	$(\hat{\mu})^2$	\hat{e}_i	\hat{e}_i^2
1	SL	JAN	540	540^2	550	550^2	-10	100
2	SL	JAN	550	550^2	550	550^2	0	0
3	SL	FEB	550	550^2	550	550^2	0	0
4	SL	FEB	580	580^2	550	550^2	30	900
5	LW	JAN	520	520^2	550	550^2	-30	900
6	LW	FEB	500	500^2	550	550^2	-50	2500
7	LW	FEB	490	490^2	550	550^2	-60	3600
8	NL	JAN	560	560^2	550	550^2	10	100
9	NL	JAN	550	550^2	550	550^2	0	0
10	NL	FEB	600	600^2	550	550^2	50	2500
11	NL	FEB	610	610^2	550	550^2	60	3600
Σ				3341700		3327500	0	14200

Viri variabilnosti za dnevni prirast

$$y_i = \mu + e_i$$

Vir variabiln.	d.f.	Vsota kvadratov	Srednji kvadrat	F- vredn.	P- vredn.
Srednja vrednost	1	3327500			
Ostanek	10	14200			
Skupno	11	3341700			

- v modelu sta samo srednja vrednost in ostanek - dve komponenti variabilnosti
- za srednjo vrednost smo porabili eno stopinjo prostosti
- in pojasnili večino variabilnosti

Srednji kvadrat

$$MS_x = \frac{xSS}{d.f.x}$$

Vir variabiln.	d.f.	Vsota kvadratov	Srednji kvadrat	F- vredn.	P- vredn.
Srednja vrednost	1	3327500	3327500		
Ostanek	10	14200	1420	$\leftarrow \hat{\sigma}_e^2$	
Skupno	11	3341700			

F –statistika

$$F = \frac{MS_x}{MS_e} =$$

= *srednji kvadrat za pojasnjeni del modela*
srednji kvadrat za (nepojasnjeni) ostanek

Vir variabiln.	d.f.	Vsota kvadratov	Srednji kvadrat	F- vredn.	P- vredn.
Srednja vrednost	1	3327500	3327500	2343.3	
Ostanek	10	14200	1420		
Skupno	11	3341700			

- F –vrednost je zelo velika, a na osnovi nje se ne odločamo
- preverjamo verjetnost (p –vrednost), da ničelna hipoteza drži

P -vrednost

- verjetnost, da drži ničelna hipoteza $H_0 : \mu = 0$
- preberemo iz tabel za F-porazdelitev ali izračunamo s pomočjo F-porazdelitve
- statistični paketi vrednost vedno izpišejo

Vir variabiln.	d.f.	Vsota kvadratov	Srednji kvadrat	F- vredn.	P- vredn.
Srednja vrednost	1	3327500	3327500	2343.3	<0.0001
Ostanek	10	14200	1420	$\leftarrow \hat{\sigma}_e^2$	
Skupno	11	3341700			

- za μ je zelo majhna verjetnost, da je enaka nič

Preveritev modela: dnevni prirast pri mladichah

- Modelu dodajmo pasmo:

$$y_{ij} = \mu + P_i + e_{ij}$$

$$e_{ij} \sim IIDN(0, \sigma_e^2)$$

$$y_{ij} \sim IIDN(\mu + P_i, \sigma_e^2)$$

- Ker model vsebuje samo pasmo, preverjamo: ■

$$H_0 : \mathbf{K}\boldsymbol{\beta} = \mathbf{0} \blacksquare$$

$$H_1 : \mathbf{K}\boldsymbol{\beta} \neq \mathbf{0} \blacksquare$$

$$\mathbf{K} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \text{ ali } \mathbf{K} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

Poiščimo rešitve za neznane parametre

$$\hat{\mu} = 1/n \sum y_{ij} = 6050/11 = 550.0$$

$$\hat{P}_1 = 1/n_1 \sum (y_{1j} - \hat{\mu}) = (-10 + 0 + 0 + 30) / 4 = +5.0$$

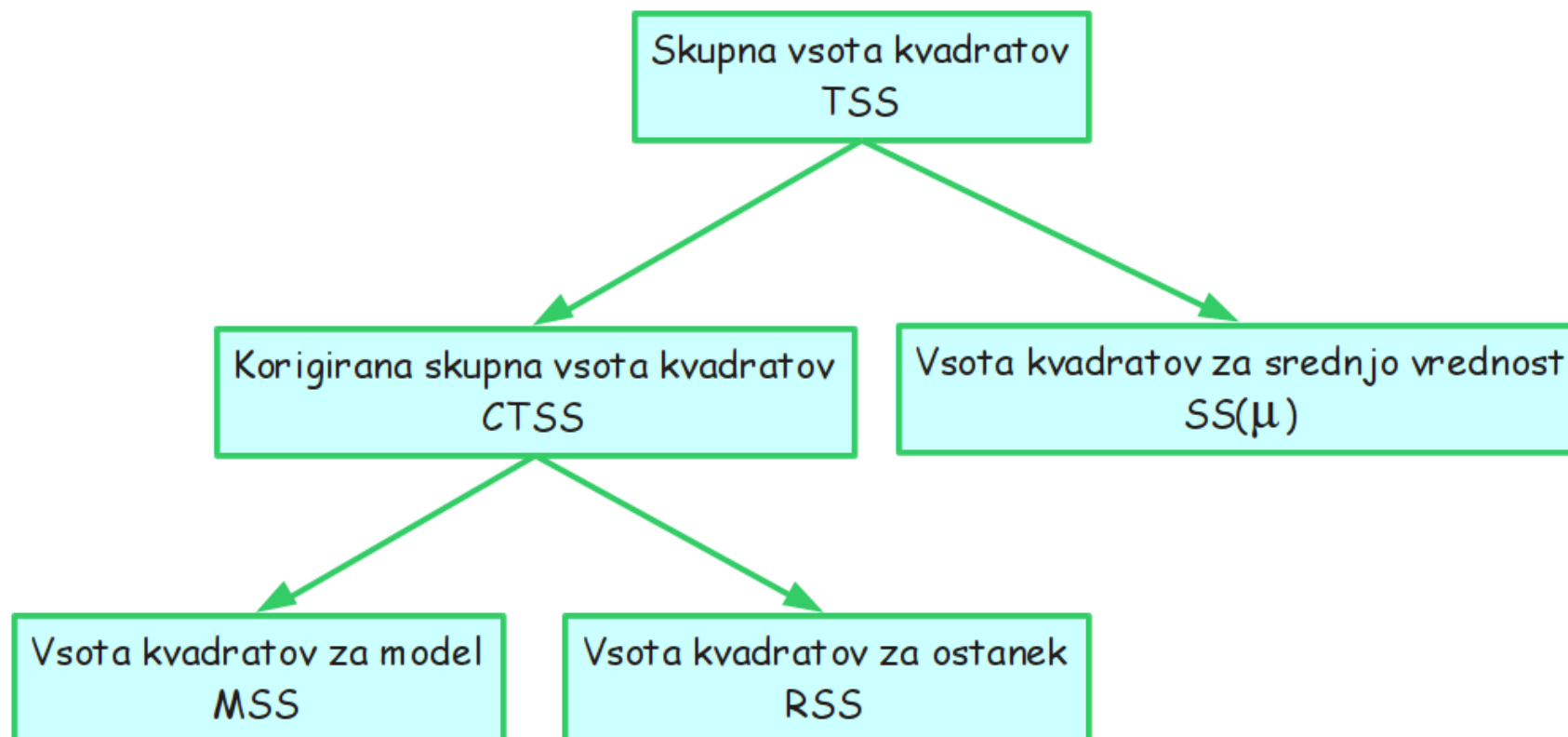
$$\hat{P}_2 = 1/n_2 \sum (y_{2j} - \hat{\mu}) = (-30 - 50 - 60) / 3 = -46.6$$

$$\hat{P}_3 = 1/n_3 \sum (y_{3j} - \hat{\mu}) = (10 + 0 + 50 + 60) / 4 = +30.0$$

Za preveritev modela potrebujemo

- skupno vsoto kvadratov: $TSS = \sum y_{ij}^2$
- vsoto kvadratov za srednjo vrednost: $SS(\mu) = n\hat{\mu}^2$
- korigirana skupna vsota kvadratov: $CTSS = \sum y_{ij}^2 - SS(\mu)$
- vsoto kvadratov za model brez srednje vrednosti:
 $MSS = \sum [E(y_{ij}) - \hat{\mu}]^2$
- vsoto kvadratov za ostanek: $RSS = \sum (y_{ij} - E(y_{ij}))^2$

Razdelitev vsote kvadratov



Izračunajmo potrebne vsote kvadratov

P_i	DP	y_{ij}^2	$\hat{\mu}$	$(\hat{\mu})^2$	\hat{P}_i	$(\hat{P}_i)^2$	\hat{e}_{ij}	\hat{e}_{ij}^2
SL	540							
SL	550							
SL	550							
SL	580							
LW	520							
LW	500							
LW	490							
NL	560							
NL	550							
NL	600							
NL	610							
Σ		TSS		SS(μ)		MSS		RSS

- skupna vsota kvadratov

P_i	DP	y_{ij}^2	$\hat{\mu}$	$(\hat{\mu})^2$	\hat{P}_i	$(\hat{P}_i)^2$	\hat{e}_{ij}	\hat{e}_{ij}^2
SL	540	540^2						
SL	550	550^2						
SL	550	550^2						
SL	580	580^2						
LW	520	520^2						
LW	500	500^2						
LW	490	490^2						
NL	560	560^2						
NL	550	550^2						
NL	600	600^2						
NL	610	610^2						
Σ		3341700						

- Skupna vsota kvadratov je enaka kot prej

- vsota kvadratov za srednjo vrednost

P_i	DP	y_{ij}^2	$\hat{\mu}$	$(\hat{\mu})^2$	\hat{P}_i	$(\hat{P}_i)^2$	\hat{e}_{ij}	\hat{e}_{ij}^2
SL	540	540^2	550	550^2				
SL	550	550^2	550	550^2				
SL	550	550^2	550	550^2				
SL	580	580^2	550	550^2				
LW	520	520^2	550	550^2				
LW	500	500^2	550	550^2				
LW	490	490^2	550	550^2				
NL	560	560^2	550	550^2				
NL	550	550^2	550	550^2				
NL	600	600^2	550	550^2				
NL	610	610^2	550	550^2				
Σ		3341700		3327500				

- $CTSS = TSS - SS(\mu) = 3341700 - 3327500 = 14200$

- vsota kvadratov za model

P_i	DP	y_{ij}^2	$\hat{\mu}$	$(\hat{\mu})^2$	\hat{P}_i	$(\hat{P}_i)^2$	\hat{e}_{ij}	\hat{e}_{ij}^2
SL	540	540^2	550	550^2	+5	5^2		
SL	550	550^2	550	550^2	+5	5^2		
SL	550	550^2	550	550^2	+5	5^2		
SL	580	580^2	550	550^2	+5	5^2		
LW	520	520^2	550	550^2	-46.6	-46.6^2		
LW	500	500^2	550	550^2	-46.6	-46.6^2		
LW	490	490^2	550	550^2	-46.6	-46.6^2		
NL	560	560^2	550	550^2	+30	30^2		
NL	550	550^2	550	550^2	+30	30^2		
NL	600	600^2	550	550^2	+30	30^2		
NL	610	610^2	550	550^2	+30	30^2		
Σ		3341700		332750		10233		

- Vsota kvadratov za model - samo sistematske vplive (brez μ)

- vsota kvadratov za ostanek

P_i	DP	y_{ij}^2	$\hat{\mu}$	$(\hat{\mu})^2$	\hat{P}_i	$(\hat{P}_i)^2$	\hat{e}_{ij}	\hat{e}_{ij}^2
SL	540	540^2	550	550^2	+5	5^2	-15	225
SL	550	550^2	550	550^2	+5	5^2	-5	25
SL	550	550^2	550	550^2	+5	5^2	-5	25
SL	580	580^2	550	550^2	+5	5^2	25	625
LW	520	520^2	550	550^2	-46.6	-46.6^2	16.6	277
LW	500	500^2	550	550^2	-46.6	-46.6^2	-3.3	11
LW	490	490^2	550	550^2	-46.6	-46.6^2	-13.3	177
NL	560	560^2	550	550^2	+30	30^2	-20	400
NL	550	550^2	550	550^2	+30	30^2	-30	900
NL	600	600^2	550	550^2	+30	30^2	20	400
NL	610	610^2	550	550^2	+30	30^2	30	900
Σ		3341700		3327500		10233	0	3966

- vsota kvadratov za ostanek je zmanjšana z dodanimi vplivi

Viri variabilnosti za dnevni prirast

$$y_{ij} = \mu + P_i + e_{ij}$$

Vir variabiln.	d.f.	Vsota kvadratov	Srednji kvadrat	F- vredn.	P- vredn.
Srednja vrednost	1	3327500			
Pasma	2	10233			
Ostanek	8	3966			
CTSS	10	14200			
Skupno	11	3341700			

- korigirana skupna vsota kvadratov se razdeli na dve komponenti
- zanima nas, koliko te vsote kvadratov odpade na eno stopinjo prostosti

Srednji kvadrat

$$MS_x = \frac{xSS}{d.f.x}$$

Vir variabiln.	d.f.	Vsota kvadratov	Srednji kvadrat	F-vredn.	P-vredn.
Srednja vrednost	1	3327500	3327500		
Pasma	2	10233	5116		
Ostane	8	3966	496	$\leftarrow \sigma_e^2$	
CTSS	10	14200			
Skupno	11	3341700			

- pri pasmi odpade 10x več vsote kvadratov na stopinjo prostosti kot pri ostanku
- izračunajmo razmerje bolj natančno

F – statistika

$$F = \frac{MS_x}{MS_e} =$$

= *srednji kvadrat za pojasnjeni del modela*
srednji kvadrat za (nepojasnjeni) ostanek

Vir variabiln.	d.f.	Vsota kvadratov	Srednji kvadrat	F- vredn.	P- vredn.
Srednja vrednost	1	3327500	3327500	6711	
Pasma	2	10233	5116	10.3	
Ostanek	8	3966	496		
CTSS	10	14200			
Skupno	11	3341700			

- določiti moramo še verjetnost, ko velja ničelna hipoteza

Vir variabiln.	d.f.	P-vrednost		F- vredn.	P- vredn.
		Vsota kvadratov	Srednji kvadrat		
Srednja vrednost	1	3327500	3327500	6711	<0.0001
Pasma	2	10233	5116	10.3	0.0061
Ostanek	8	3966	496		
CTSS	10	14200			
Skupno	11	3341700			

- Model je dober. Vsota kvadratov za ostanek je majhna.
- Sistematski vplivi v modelu so pomembni. Ker je vpliv samo eden, vemo, da je značilen.
- Izračunajte delež pojasnjene variance in standardni odklon za ostanek!

Model s pasmo in mesecem

$$y_{ijk} = \mu + P_i + M_j + e_{ijk}$$

$$e_{ijk} \sim IIDN(0, \sigma_e^2)$$

$$y_{ijk} \sim IIDN(\hat{\mu} + \hat{P}_i + \hat{M}_j, \sigma_e^2)$$

- v modelu sta dva sistematska vpliva
- napaka je porazdeljena (D)
 - identično (I): vse meritve enako natančno merjeni
 - neodvisno (I): meritve niso povezane, "sorodne"
 - normalno (N)

- nizi rešitev za sistem enačb

$$\begin{bmatrix} \hat{\mu} \\ \hat{P}_1 \\ \hat{P}_2 \\ \hat{P}_3 \\ \hat{M}_1 \\ \hat{M}_2 \end{bmatrix} = \begin{bmatrix} 297.2\bar{7} \\ 109.0\bar{9} \\ 54.0\bar{9} \\ 134.0\bar{9} \\ 138.6\bar{3} \\ 158.6\bar{3} \end{bmatrix} = \begin{bmatrix} 590 \\ -25 \\ -80 \\ 0 \\ -20 \\ 0 \end{bmatrix} = \begin{bmatrix} 550 \\ 15 \\ -40 \\ 40 \\ -20 \\ 0 \end{bmatrix}$$

- sistem ni polnega ranga in ima več rešitev
- pri preizkušanju hipotez dobimo iste zaključke ne glede na uporabljen niz

- vsota kvadratov za model

P_i	M_j	DP	y_{ijk}^2	$\hat{\mu}$	$(\hat{\mu})^2$	$\hat{P}_i + \hat{M}_j$	$(\hat{P}_i + \hat{M}_j)^2$
SL	JAN	540		550			
SL	JAN	550		550			
SL	FEB	550		550			
SL	FEB	580		550			
LW	JAN	520		550			
LW	FEB	500		550			
LW	FEB	490		550			
NL	JAN	560		550			
NL	JAN	550		550			
NL	FEB	600		550			
NL	FEB	610		550			
Σ			TSS		SS(μ)		MSS

- popolnimo tabelo

- vsota kvadratov za model

P_i	M_j	DP	y_{ijk}^2	$\hat{\mu}$	$(\hat{\mu})^2$	$\hat{P}_i + \hat{M}_j$	$(\hat{P}_i + \hat{M}_j)^2$
SL	JAN	540	540^2	550	550^2	-5	25
SL	JAN	550	550^2	550	550^2	-5	25
SL	FEB	550	550^2	550	550^2	15	225
SL	FEB	580	580^2	550	550^2	15	225
LW	JAN	520	520^2	550	550^2	-60	3600
LW	FEB	500	500^2	550	550^2	-40	1600
LW	FEB	490	490^2	550	550^2	-40	1600
NL	JAN	560	560^2	550	550^2	20	400
NL	JAN	550	550^2	550	550^2	20	400
NL	FEB	600	600^2	550	550^2	40	1600
NL	FEB	610	610^2	550	550^2	40	1600
Σ			3341700		3327500		11300

- vsota kvadratov za model je večja z dodanim vplivom meseca

- vsota kvadratov za ostanek

P_i	M_j	DP	$\hat{P}_i + \hat{M}_j$	$\left(\hat{P}_i + \hat{M}_j\right)^2$	\hat{e}_{ijk}	\hat{e}_{ijk}^2
SL	JAN	540	-5	25		
SL	JAN	550	-5	25		
SL	FEB	550	15	225		
SL	FEB	580	15	225		
LW	JAN	520	-60	3600		
LW	FEB	500	-40	1600		
LW	FEB	490	-40	1600		
NL	JAN	560	20	400		
NL	JAN	550	20	400		
NL	FEB	600	40	1600		
NL	FEB	610	40	1600		
Σ				11300		

- dokončajmo delo!

- vsota kvadratov za ostanek

P_i	M_j	DP	$\hat{P}_i + \hat{M}_j$	$\left(\hat{P}_i + \hat{M}_j\right)^2$	\hat{e}_{ijk}	\hat{e}_{ijk}^2
SL	JAN	540	-5	25	-5	25
SL	JAN	550	-5	25	5	25
SL	FEB	550	15	225	-15	225
SL	FEB	580	15	225	15	225
LW	JAN	520	-60	3600	30	900
LW	FEB	500	-40	1600	-10	100
LW	FEB	490	-40	1600	-20	400
NL	JAN	560	20	400	-10	100
NL	JAN	550	20	400	-20	400
NL	FEB	600	40	1600	10	100
NL	FEB	610	40	1600	20	400
Σ				11300		2900

- vsota kvadratov za ostanek se zmanjša z dodanim vplivom meseca

- analiza variance

$$y_{ijk} = \mu + P_i + M_j + e_{ijk}$$

Vir	d.f.	Vsota	Srednji	F-	P-
variabiln.		kvadratov	kvadrat	vredn.	vredn.
Srednja vrednost		3327500			
Model		11300			
Ostanek		2900			
CTSS		14200			
Skupno		3341700			

- določimo še število stopinj prostosti in izpolnimo tabelo

Model s pasmo in mesecem (nadalj.)

$$y_{ijk} = \mu + P_i + M_j + e_{ijk}$$

Vir variabiln.	d.f.	Vsota kvadratov	Srednji kvadrat	F- vredn.	P- vredn.
Srednja vrednost	1	3327500	3327500	8031.90	<0.0001
Model	3	11300	3767	9.09	0.0082
Ostanek	7	2900	414		
CTSS	10	14200			
Skupno	11	3341700			

- Model je dober. Sistematski vplivi v modelu so pomembni
- Drugi korak: določamo vpliv, ki pomembno prispeva k rezultatom prireje