

Оддий чизиқли регрессион модель

Режа саволлари

- Зарурий тушунчалар
- Иқтисодий ва эконометрик моделлар
- ЭКК усулининг моҳияти
- ЭКК механикаси
- Детерминация коэффициенти
- Чизиқсиз боғланишлар
- Ҳисобланган параметрлар стандарт хатоликлари
- Регрессия стандарт хатолиги

Зарурий тушунчалар (univariate)

- Тасодифий миқдор: дискрет ва узлуксиз
- Бош тўплам ва танлама тўплам
- Тасвирий статистика
 - Ўртача қиймат, математик кутилиш
 - average, mean, expected value, expectation
 - Дисперсия, стандарт четланиш
 - variance, standard deviation
 - Экссесс, асимметрия
 - kurtosis, skewness
 - Гистограмма

Зарурий тушунчалар (bivariate)

- Икки ўзгарувчи орасидаги ихтиёрий боғлиқлик
 - Ковариация
 - covariance [like variance]
 - Корреляция
 - correlation
 - Ковариация
 - Икки танлама тестлари
 - Mean comparison tests [=ttest]

Иқтисодий модель

Иқтисодий модель (мисоллар)

- Модель – реалликни соддалаштирилган формуласи
- Шунинг учун, “тасаввур қилинг!”, “фараз қилайлик!” ва ҳ.к. деб бошланади!
- **Тасаввур қилинг**, y ҳамда $x, z \dots$
[иқтисодий назария ёки бошқа мулоҳазаларга кўра] боғлиқ
$$y = f(x, z, \dots)$$

• Мисол 1 (кўп омилли)

- GM-Ўзбекистон ишлаб чиқарадиган машиналарнинг талаб функцияси

$$Q_{GMUZB}^d = f(P, P_s, P_c, INC, \dots)$$

• Мисол 2 (бир омилли)

$$C = f(INC) = c_0 + c_1 \cdot INC$$

- Бунда C – истеъмол ҳажми, INC – даромад, c_0, c_1 - ўзгармас сонлар
- **Q:** c_0, c_1 интерпретацияси қандай?

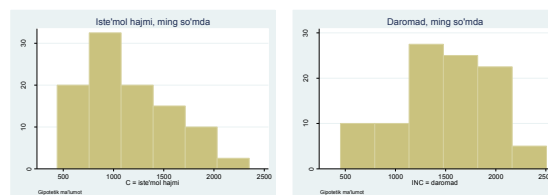
Маълумот (танлама)

Озиқ-овқат истеъмоли, минг сўмда	Даромад, минг сўмда
460.88	553.5
543.92	658.5
477.36	712.5
...	...

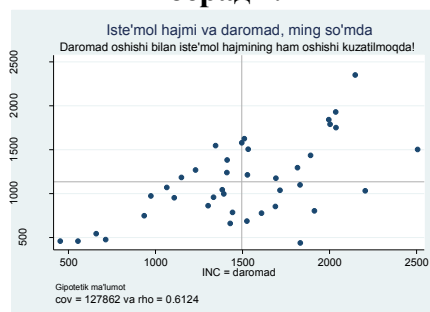
- Q: Бу қайси турдаги маълумот?

Истеъмол ҳажми ва даромадлар

Ўзгарувчи номи, минг сўмда	Кузатиш лар сони	Ўрта часи	Стандарт четланиши	Мини мум	Макси мум
Истеъмол ҳажми	40	1134.294	450.701	438.84	2350.64
Даромадлар	40	1494.413	463.261	452.25	2505



$\rho=0.61$ эканлиги нимани беради?



Эконометрик (регрессион) модель

- Умуман, $y_i = \beta_0 + \beta_1 x_i + u_i$,
Бунда
 - x_i : эрксиз ўзгарувчи, регрессор
 - *independent variable, regressor*
 - y_i : эрки ўзгарувчи, регрессанд
 - *dependent variable, regressand*
 - u_i : тасодифий хатолик
 - *random error*
 - $i = 1, 2, \dots, n$
 - β_0, β_1 : номаълум параметрлар
 - β_0 : *intercept*, β_1 : *slope*

u_i структураси

- Ҳисобга олинмаган омиллар
 - *omitted variables*
- у ўзгарувчисининг ўлчов хатоликлари
 - *measurement error*
- x ва y ўртасида чизикли бўлмаган боғлиқликлар таъсири
- Соф тасодифий таъсирлар

Энг кичик квадратлар (ЭКК) усули

- ЭКК бўйича (β_0, β_1) параметрлари тасодифий хатолик квадратлари йиғиндиси минималлаштирилиб танланади

$$RSS = \min_{\hat{\beta}_0, \hat{\beta}_1} \sum \hat{u}_i^2 = \sum [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

$$\begin{cases} \frac{dRSS}{d\hat{\beta}_0} = 0 \\ \frac{dRSS}{d\hat{\beta}_1} = 0 \end{cases} \Rightarrow \hat{\beta}_0, \hat{\beta}_1$$

ЭКК параметрларини ҳисоблаш

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Ҳисоблаш натижалари

$$\hat{C} = \hat{\beta}_0 + \hat{\beta}_1 INC$$

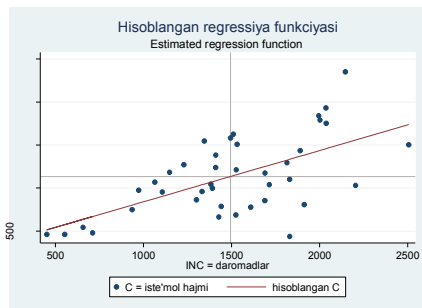
Source	SS	df	MS	Number of obs =
Model	2970936.04	1	2970936.04	F(1, 38) = 22.80
Residual	4951178.41	38	130294.169	Prob > F = 0.0000
Total	7922114.45	39	203131.14	R-squared = 0.3750
				Adj R-squared = 0.3586
				Root MSE = 360.96

food_exp_usz	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
income_usz	0.6	.1247684	4.78	0.000	.3432035 .8483644
_cons	244	194.9949	1.25	0.219	-150.7994 638.6935

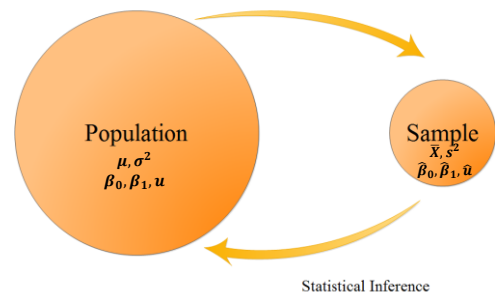
$$\hat{\beta}_0 \approx 244, \hat{\beta}_1 \approx 0.6 \Rightarrow C = 244 + 0.6 \cdot INC$$

- Даромадлар қўшимча 100 000 сўмга ошганида, ўрта ҳисобда истеъмол ҳажми [тирик истеъмолчи] \approx 60 000 сўмга ошади, *ceteris paribus*

$$\hat{C} = 244 + 0.6 \cdot INC$$

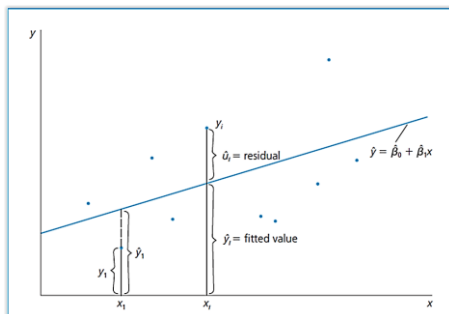


Бош тўплам ва танлама тўплам



16

$$y, \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \text{ ҳамда } \hat{u} = y - \hat{y}$$



ЭКК статистикасининг алгебраик хоссалари

1. $\sum \hat{u}_i = 0$
 2. $\sum x_i \hat{u}_i = 0$
 3. (\bar{x}, \bar{y}) нуктаси $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ чизиғида ётади
- Ҳамда**
4. $\bar{\hat{y}} = \bar{y}$
 5. $cov(\hat{y}_i, \hat{u}_i) = 0$

Детерминация коэффициенти R^2

- $TSS = \sum (y_i - \bar{y})^2$
 - Total sum of squares
 - у ўзгарувчисидаги вариация
- $ESS = \sum (\hat{y}_i - \bar{y})^2$
 - Explained sum of squares
 - \hat{y} ўзгарувчисидаги вариация
- $RSS = \sum \hat{u}_i^2$
 - Residual sum of squares
 - \hat{u} ўзгарувчисидаги вариация
- $TSS = ESS + RSS$
- $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$
 - R^2 : детерминация коэффициенти
 - у ўзгарувчисидаги вариациянинг \hat{y} тушунтирган қисми
 - $R^2 \in [0,1]$

Детерминация коэффициенти R^2

Source	SS	df	MS	Number of obs =
Model	1	2970936.04		40
Residual	38	130294.169		F(1, 38) = 22.80
Total	39	203131.14		Prob > F = 0.0000
				R-squared = 0.3586
				Adj R-squared = 0.3586
				Root MSE = 360.96

Variable	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
income_uzs	.5957839	.1247684	4.78	0.000	.3432035 .8483644
_cons	243.947	194.9949	1.25	0.219	-150.7994 638.6935

$R^2 = 0.375$

- Озиқ-овқат истеъмолидаги вариациянинг 37.5% модель орқали тушунтирилади

Чизиқсиз боғланишлар

- Регрессор ва эркин ўзгарувчининг боғланиши чизиқли бўлмастлиги мумкин.
- Фараз қилинг, $y = e^{\beta_0 + \beta_1 x + u}$ бўлсин.
- У ҳолда, $\ln y = \beta_0 + \beta_1 x + u$ ва $\frac{1}{y} dy = \beta_1 dx \Rightarrow \beta_1 = \frac{dy}{y dx} \Rightarrow 100 \cdot \beta_1 = \frac{\% \Delta y}{\Delta x}$

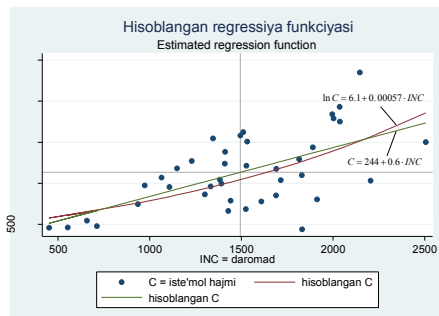
Ҳисоблаш натижалари $\ln \hat{C} = \hat{\beta}_0 + \hat{\beta}_1 INC$

Source	SS	df	MS	Number of obs =
Model	2.7187448	1	2.7187448	40
Residual	4.29475607	38	.113019897	F(1, 38) = 24.06
Total	7.01350088	39	.179833356	Prob > F = 0.0000
				R-squared = 0.3715
				Adj R-squared = 0.3715
				Root MSE = .33618

Variable	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
log_food_e-s	.0001162	.0001162	4.90	0.000	.0003347 .0008052
income_uzs	.1816093	.01816093	33.59	0.000	5.731945 6.467242

- Даромадлар кўшимча 100 000 сўмга ошса, ўрта ҳисобда озиқ-овқат истеъмоли 5.7% га ошадди, *ceteris paribus*
- Ушбу модель аввалгиси ($y = \beta_0 + \beta_1 x + u$) дан бир оз яхшироқ. **Нега?**

Чизиқли ва кўрсаткичли функциялар



Логарифмик боғланишлар

Модель	Эркин (боғлиқ) ўзгарувчи	Регрессор	Интерпретация (талқини)
level – level	y	x	$\Delta y = \beta_1 \Delta x$
level – log	y	$\ln(x)$	$\Delta y = (\beta_1 / 100) \% \Delta x$
log – level	$\ln(y)$	x	$\% \Delta y = (100 \beta_1) \Delta x$
log – log	$\ln(y)$	$\ln(x)$	$\% \Delta y = \beta_1 \% \Delta x$

$\hat{\beta}_0, \hat{\beta}_1$ сифатини баҳолаш

Фаразлар:

- $E(u|x) = 0 \iff cov(u, x) = corr(u, x) = 0$
 - zero conditional mean assumption
- $(x_i, y_i), i = 1, 2, \dots, n$ тасодифий танлама
 - random sample
- Кескин (жуда ҳам юқори ва жуда ҳам кичик) кузатишлар кам ёки йўқ
 - outliers are unlikely
- $var(u_i) = \sigma^2$
 - homoskedasticity assumption
- $cov(u_i, u_j) = 0$
- $y_i = \beta_0 + \beta_1 x_i + u_i$ модели тўғри танланган
- $E(u_i) = 0 \quad \forall i$ учун
- x_i тасодифий эмас ҳамда қийматлари хар хил, $var(x_i) > 0$

$\hat{\beta}_0, \hat{\beta}_1$ статистик хусусиятлари

- $E(\hat{\beta}_0) = \beta_0$ ва $E(\hat{\beta}_1) = \beta_1$
 - Такрорий танловда ҳисобланган $\hat{\beta}_0$ ва $\hat{\beta}_1$, мос равишда, назарий β_0 ва β_1 га тенг бўлади
 - repeated samples

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} y_i = \sum w_i y_i =$$

$$= \sum w_i (\beta_0 + \beta_1 x_i + u_i) = \beta_0 + \sum w_i x_i + \sum w_i u_i = \frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

$$\Rightarrow E(\hat{\beta}_1) = E(\beta_0 + \sum w_i x_i + \sum w_i u_i) = \beta_0 + \sum w_i E(x_i) + \sum w_i E(u_i) = \beta_1$$

- $var(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}, var(\hat{\beta}_0) = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}$

Регрессиянинг стандарт хатолиги

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum \hat{u}_i^2 = \frac{SSR}{n-2}$$

- $n-2$ эркинлик даражаси
- $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ регрессия стандарт хатолиги. Умуман, дисперсиянинг квадрат илдизи стандарт хатоликини беради.
 - SER: standard error of regression
 - Root MSE: root mean squared error

Ҳисобланган параметрлар стандарт хатоликлари

Source	SS	df	MS	Number of obs =
Model	2970936.04	1	2970936.04	F(1, 38) = 22.80
Residual	4951178.41	38	130294.169	Prob > F = 0.0000
Total	7922114.45	39	203131.14	R-squared = 0.3750
				Adj R-squared = 0.3586
				Root MSE = 0.3686

	Coeff.	Std. Err.	t	P> t	[95% Conf. Interval]
income_uzs	.5957839	0.124	4.78	0.000	.3432035 .8483644
_cons	243.947	150.7994	1.25	0.219	-150.7994 638.6935

- Стандарт хатоликлар фаразлар амалга ошганда чизикли функциялар синфида энг кичик

Қўшимча ўқиш

- R. Carter Hill, William E. Griffiths, Guay C.Lim. *Principles of Econometrics* (3, 4th edition). Ch2.
 - сода ва осон тушунтирилган. Уздуксиз ўзгарувчиларнинг моментлари иловаларда бағависил тушунтирилган.
- Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach* (any edition). Ch2.
 - интуиция даражасида яхши тушунтирилган.
- Damodar Gujarati, Dawn Porter. *Basic Econometrics* (any edition). Ch1-2.
 - Тушунтириш даражаси сода, кўпроқ техник тарафлари яхши тушунтирилган.
- Я.Р. Магнус, П.К. Катывшев, А.А.Плесецкий. *Эконометрика*. Начальный курс.
 - киска ва тушунарли берилган, лекин чизикли алгебра бўйича дастлабки билим талаб қилади.

Гаусс-Марков теоремаси

- Стандарт фаразлар амалга ошса (қониктирилса), ЭКК усули бўйича ҳисобланган параметрлар дисперсияси мавжуд чизикли усуллар ичида энг кичик дисперсияга эга бўлади
 - OLS estimator is BLUE
 - Best Linear Unbiased Estimator