MAT 271E: PROBABILITY AND STATISTICS

PROF. DR. CANAN SARICAM

INFERENCE ON THE VARIANCE OF THE NORMAL POPULATION

Suppose that we wish to test the hypothesis that the variance of normal population σ^2 equal to a specified value, say σ_0^2. Let X_1,, X_2,…X_n be a random sample of n observations from this population. To establish hypothesis testing:

$$H_0: \sigma^2 = \sigma_0^2$$
$$H_1: \sigma^2 \neq \sigma_0^2 \qquad (9\text{-}26)$$

we will use the test statistic:

$$X_0^2 = \frac{(n-1)S^2}{\sigma_0^2} \qquad (9\text{-}27)$$

INFERENCE ON THE VARIANCE OF THE NORMAL POPULATION

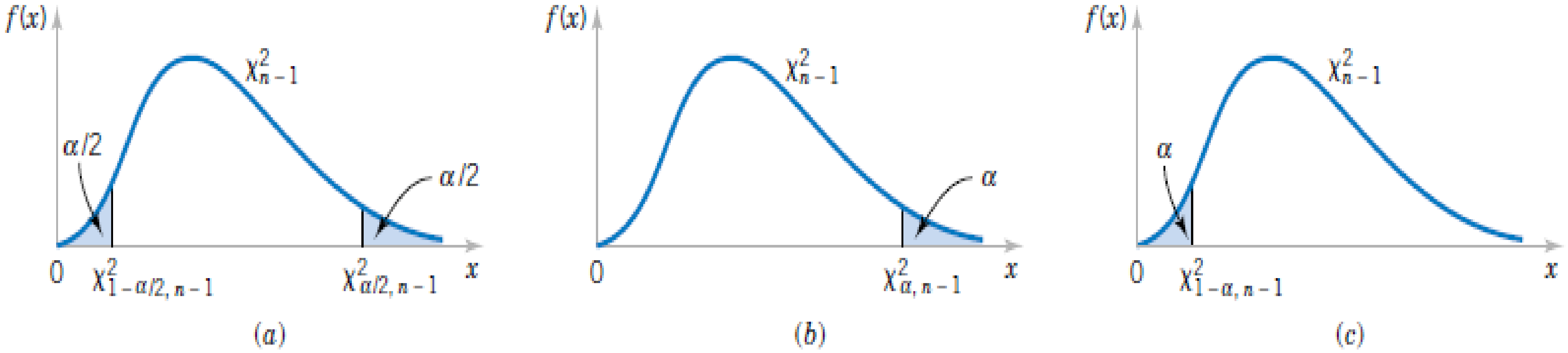


Figure 9-10 Reference distribution for the test of H_0: $\sigma^2 = \sigma_0^2$ with critical region values for (a) H_1: $\sigma^2 \neq \sigma_0^2$, (b) H_1: $\sigma^2 > \sigma_0^2$, and (c) H_1: $\sigma^2 < \sigma_0^2$.

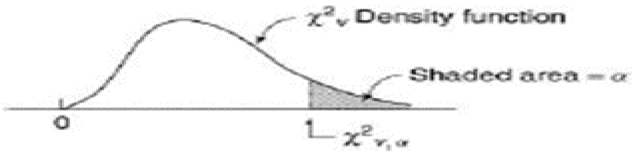
INFERENCE ON THE VARIANCE OF THE NORMAL POPULATION

Example 8.11

An automatic filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153$ (fluid ounces)^2. If the variance of fill volume exceeds 0.01 (fluid ounces)^2, an unacceptable proportion of bottles will be underfilled or overfilled. Is there evidence in the sample data to suggest that the manufacturer has a problem with underfilled or overfilled bottles? Use $\alpha = 0.05$, and assume that fill volume has a normal distribution.

INFERENCE ON THE VARIANCE OF THE NORMAL POPULATION

Table C.3 Critical Values $\chi^2_{\nu,\alpha}$ for Chi-Square Distribution



ν	α = .995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.843	5.025	6.637	7.882
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.344	12.837
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.085	16.748
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.440	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.012	18.474	20.276
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.534	20.090	21.954
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.022	21.665	23.587
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.724	26.755
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.735	27.687	29.817
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.600	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.577	32.799
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.407	7.564	8.682	10.085	24.769	27.587	30.190	33.408	35.716
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.843	7.632	8.906	10.117	11.651	27.203	30.143	32.852	36.190	38.580
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.033	8.897	10.283	11.591	13.240	29.615	32.670	35.478	38.930	41.399
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.195	11.688	13.090	14.848	32.007	35.172	38.075	41.637	44.179
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.519	11.523	13.120	14.611	16.473	34.381	37.652	40.646	44.313	46.925
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.807	12.878	14.573	16.151	18.114	36.741	40.113	43.194	46.962	49.642
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.120	14.256	16.147	17.708	19.768	39.087	42.557	45.772	49.586	52.333
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.457	15.655	17.538	19.280	21.433	41.422	44.985	48.231	52.190	55.000
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.814	17.073	19.046	20.866	23.110	43.745	47.400	50.724	54.774	57.646
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.191	18.508	20.569	22.465	24.796	46.059	49.802	53.203	57.340	60.272
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.584	19.960	22.105	24.075	26.492	48.363	52.192	55.667	59.891	62.880
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.994	21.425	23.654	25.695	28.196	50.660	54.572	58.119	62.420	65.473
40[a]	20.706	22.164	24.433	26.509	29.050	51.805	55.758	59.342	63.691	66.766

[a] For $\nu > 40$, $\chi^2_{\nu,\alpha} \simeq \nu\left(1 - \frac{2}{9\nu} + z_\alpha\sqrt{\frac{2}{9\nu}}\right)^3$.

Source: Reprinted with permission of Pearson Education, Inc.

INFERENCE ON THE VARIANCE OF THE NORMAL POPULATION

Example 8.11

Using the eight-step procedure results in the following:

1. The parameter of interest is the population variance σ^2.
2. H_0: $\sigma^2 = 0.01$
3. H_1: $\sigma^2 > 0.01$
4. $\alpha = 0.05$
5. The test statistic is

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

6. Reject H_0 if $\chi_0^2 > \chi_{0.05,19}^2 = 30.14$.
7. Computations:

$$\chi_0^2 = \frac{19(0.0153)}{0.01} = 29.07$$

8. Conclusions: Since $\chi_0^2 = 29.07 < \chi_{0.05,19}^2 = 30.14$, we conclude that there is no strong evidence that the variance of fill volume exceeds 0.01 (fluid ounces)^2.

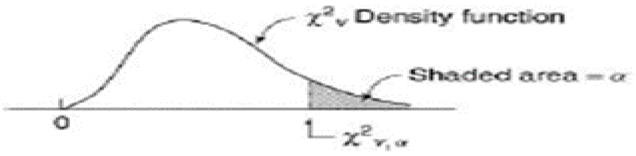
INFERENCE ON THE VARIANCE OF THE NORMAL POPULATION

Extra Exercise

A tire manufacturer claims that the variance of the diameters in a certain tire model is 8.6. A random sample of 10 tires has a variance of 4.3. At $\alpha=0.01$, is there enough evidence to reject the manufacturer's claim? Assume the population is normally distributed.

INFERENCE ON THE VARIANCE OF THE NORMAL POPULATION

Table C.3 Critical Values $\chi^2_{\nu,\alpha}$ for Chi-Square Distribution



	α									
ν	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.843	5.025	6.637	7.882
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.344	12.837
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.085	16.748
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.440	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.012	18.474	20.276
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.534	20.090	21.954
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.022	21.665	23.587
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.724	26.755
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.735	27.687	29.817
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.600	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.577	32.799
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.407	7.564	8.682	10.085	24.769	27.587	30.190	33.408	35.716
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.843	7.632	8.906	10.117	11.651	27.203	30.143	32.852	36.190	38.580
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.033	8.897	10.283	11.591	13.240	29.615	32.670	35.478	38.930	41.399
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.195	11.688	13.090	14.848	32.007	35.172	38.075	41.637	44.179
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.519	11.523	13.120	14.611	16.473	34.381	37.652	40.646	44.313	46.925
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.807	12.878	14.573	16.151	18.114	36.741	40.113	43.194	46.962	49.642
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.120	14.256	16.147	17.708	19.768	39.087	42.557	45.772	49.586	52.333
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.457	15.655	17.538	19.280	21.433	41.422	44.985	48.231	52.190	55.000
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.814	17.073	19.046	20.866	23.110	43.745	47.400	50.724	54.774	57.646
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.191	18.508	20.569	22.465	24.796	46.059	49.802	53.203	57.340	60.272
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.584	19.960	22.105	24.075	26.492	48.363	52.192	55.667	59.891	62.880
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.994	21.425	23.654	25.695	28.196	50.660	54.572	58.119	62.420	65.473
40[a]	20.706	22.164	24.433	26.509	29.050	51.805	55.758	59.342	63.691	66.766

[a] For $\nu > 40$, $x^2_{\nu,\alpha} \simeq \nu\left(1 - \frac{2}{9\nu} + z_\alpha\sqrt{\frac{2}{9\nu}}\right)^3$.

Source: Reprinted with permission of Pearson Education, Inc.

INFERENCE ON THE VARIANCE OF THE NORMAL POPULATION

Extra Exercise

$n=10$, $df=9$, $s^2=4.3$, $\sigma^2=8.6$

$Ho: \sigma^2=8.6$

$H_1: \sigma^2 \neq 8.6$

$\chi_{9,0.995}=1.735$

$\chi_{9,0.005}=23.589$

$\chi_{calc}=\frac{(n-1)s^2}{\sigma^2}=\frac{9.4}{8.6}=4.5$

$1.735 < 4.5 < 23.589$

Fail to reject Ho.

EXTRA EXAMPLES FOR WEEK 13

EKSTRA EXAMPLE 1 FOR 13TH WEEK

- Boys of a certain age are known to have a mean weight of μ=85 pounds. A complaint is made that the boys living in a municipal children's home are underfed. As one bit of evidence, n=25 boys (of the same age) are weighed and found to have a mean weight of $\bar{x}$ = 80.94 pounds. It is known that the population standard deviation σ is 11.6 pounds (the unrealistic part of this example!). Based on the available data, what should be concluded concerning the complaint?

EKSTRA EXAMPLE 1 FOR 13TH WEEK-SOLUTION

The null hypothesis is $H_0 : \mu = 85$, and the alternative hypothesis is $H_A : \mu < 85$. In general, we know that if the weights are normally distributed, then:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

follows the standard normal $N(0, 1)$ distribution. It is actually a bit irrelevant here whether or not the weights are normally distributed, because the same size $n = 25$ is large enough for the Central Limit Theorem to apply. In that case, we know that Z, as defined above, follows at least approximately the standard normal distribution. At any rate, it seems reasonable to use the test statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

for testing the null hypothesis

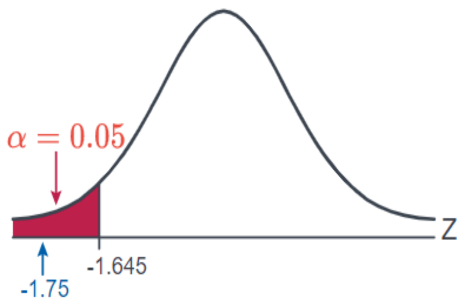
$$H_0 : \mu = \mu_0$$

against any of the possible alternative hypotheses $H_A : \mu \neq \mu_0$, $H_A : \mu < \mu_0$, and $H_A : \mu > \mu_0$.

For the example in hand, the value of the test statistic is:

$$Z = \frac{80.94 - 85}{11.6/\sqrt{25}} = -1.75$$

The critical region approach tells us to reject the null hypothesis at the $\alpha = 0.05$ level if $Z < -1.645$. Therefore, we reject the null hypothesis because $Z = -1.75 < -1.645$, and therefore falls in the rejection region:



EKSTRA EXAMPLE 2 FOR 13TH WEEK

- A particular brand of diet margarine was analyzed to determine the level of polyunsaturated fatty acid (in percent). A sample of 6 packages resulted in the following data: 16.8, 17.2, 17.4, 16.9, 16.5, 17.1.Test the hypothesis H_o: $\mu=17$ against $H_{1:}$ $\mu \neq 17$ using $\alpha=0.01$.

EKSTRA EXAMPLE 2 FOR 13TH WEEK-SOLUTION

- From the sample data X=16.98, s=0.318
- Parameter of interest is μ
- Ho: $\mu=17$
- H1: $\mu\neq 17$
- $\alpha= 0.01$
- The test statistics student t calculated=-0.154
- Reject Ho if t>4.032 and t<-4.032
- Tcalculated is bigger than -4.032 and smaller than 4.032 therefore it is in acceptance reject.
- We accept Ho

EKSTRA EXAMPLE 3 FOR 13TH WEEK

- A dairy processing company claims that the variance of the amount of fat in the whole milk processed by the company is no more than 0.25. You suspect this is wrong and find that a random sample of 41 milk containers has a variance of 0.27. At $\alpha = 0.05$, is there enough evidence to reject the company's claim?

EKSTRA EXAMPLE 3 FOR 13TH WEEK-SOLUTION

- Parameter of interest is variance
- Ho: σ^2=0.25
- H1: σ^2>0.25
- α= 0.05
- The test statistics χ^2calculated=43.2
- Reject Ho if χ^2calculated> 55.758 (From chi-square table)
- X^2calculated is smaller than 55.758 therefore it is in acceptance reject.
- We accept Ho

WEEK 14

SIMPLE LINEAR REGRESSION AND CORRELATION

SIMPLE LINEAR REGRESSION AND CORRELATION

- Regression analysis is a statistical technique for modelling and investigating the relationship between two or more variables.
- For example, in a chemical process suppose that the yield of two products are related to process operating temperature. Regression analysis can be used to build a model to predict yield at a given temperature.
- The regression analysis helps to express the relationship between two variables with a formula.

SIMPLE LINEAR REGRESSION AND CORRELATION

Simple Linear Regression Model

The equation that describes how y is related to x and an error term is called the regression model.

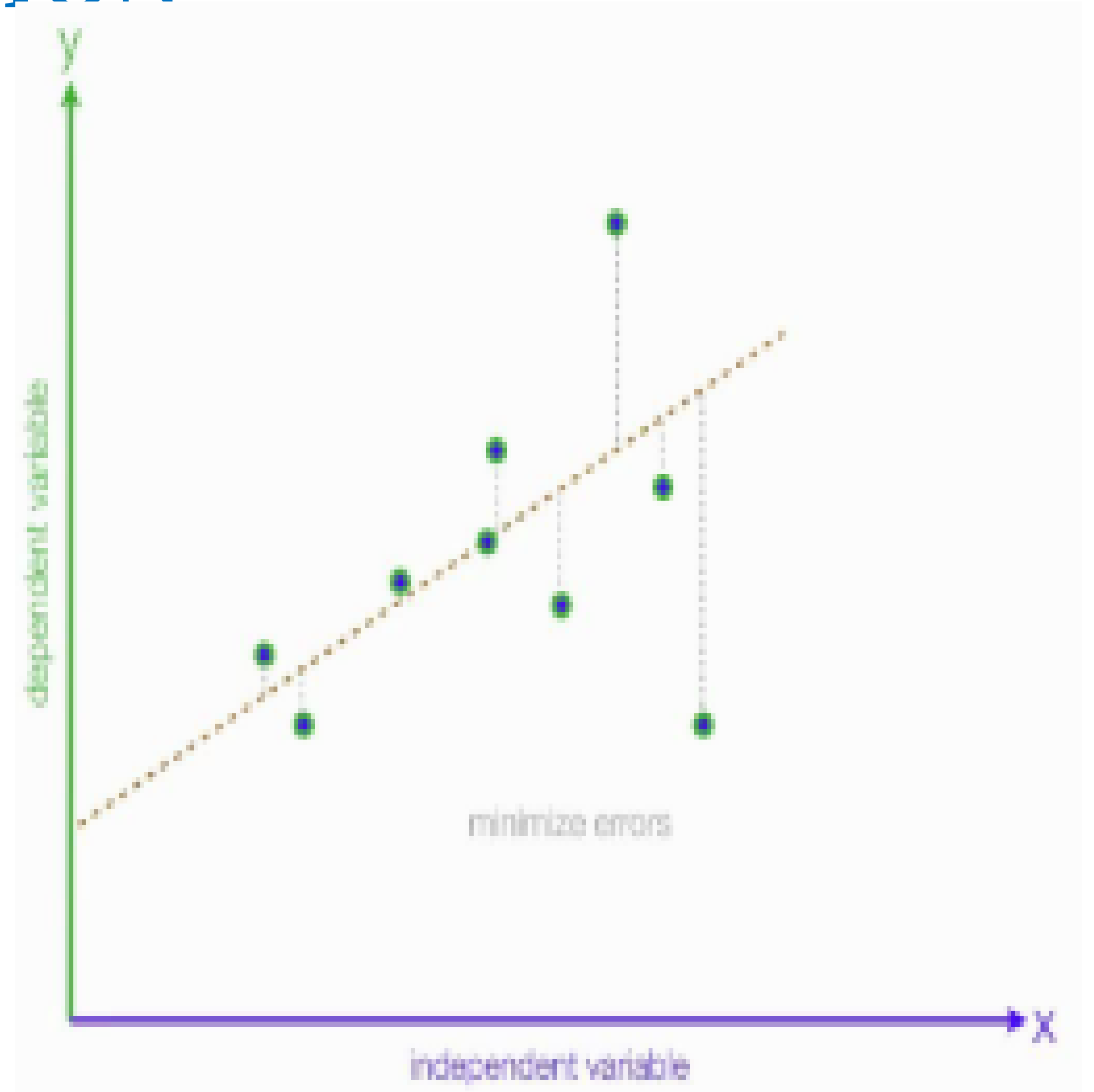
The simple linear regression model is:

$$\boxed{y = \beta_0 + \beta_1 x + \varepsilon}$$

where:

β_0 and β_1 are called parameters of the model,

ε is a random variable called the error term.



LINEAR REGRESSION MODEL FORMULAS

The **least squares estimates** of the intercept and slope in the simple linear regression model are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{11-7}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \tag{11-8}$$

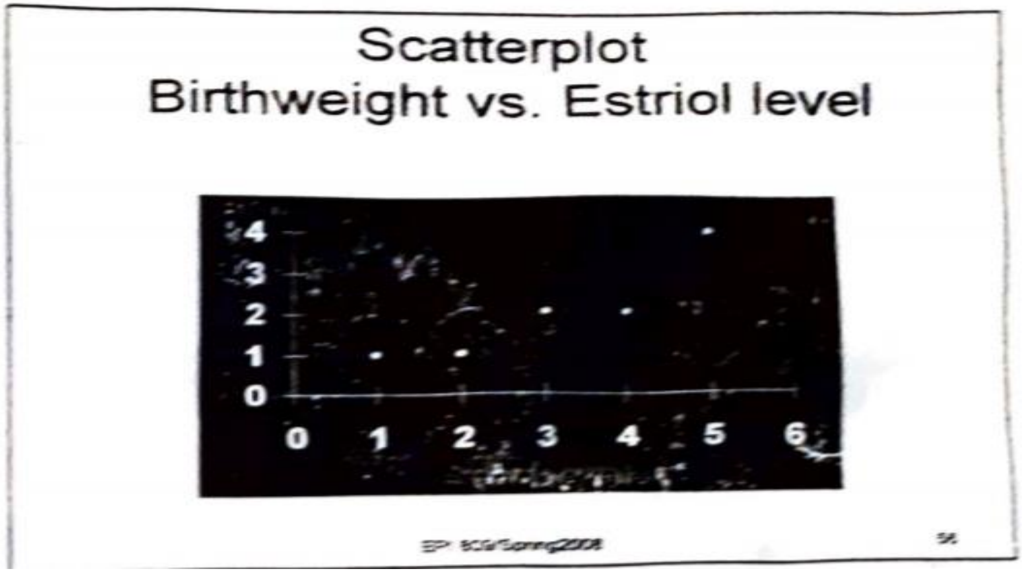
where $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ and $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$.

LINEAR REGRESSION MODEL FORMULAS

Parameter Estimation Example

Obstetrics: What is the **relationship** between Mother's Estriol level & Birthweight using the following data?

Estriol (mg/24h)	**Birthweight** (g/1000)
1	1
2	1
3	2
4	2
5	4



What is the birthweight when estrio level is 6 mg/24h?

LINEAR REGRESSION MODEL FORMULAS

Parameter Estimation Example

Parameter Estimation Solution Table

X_i	Y_i	X_i^2	Y_i^2	X_iY_i
1	1	1	1	1
2	1	4	1	2
3	2	9	4	6
4	2	16	4	8
5	4	25	16	20
15	10	55	26	37

The **least squares estimates** of the intercept and slope in the simple linear regression model are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \qquad (11\text{-}7)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \qquad (11\text{-}8)$$

where $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ and $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$.

LINEAR REGRESSION MODEL FORMULAS

Parameter Estimation Example

Parameter Estimation Solution Table

X_i	Y_i	X_i^2	Y_i^2	X_iY_i
1	1	1	1	1
2	1	4	1	2
3	2	9	4	6
4	2	16	4	8
5	4	25	16	20
15	10	55	26	37

Parameter Estimation Solution

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \frac{37 - \frac{(15)(10)}{5}}{55 - \frac{(15)^2}{5}} = 0.70$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 2 - (0.70)(3) = -0.10$$

Y=0.7x-0.1

Y=0.7*6-0.1=4.2-0.1=4.1

Regression-Extra Example

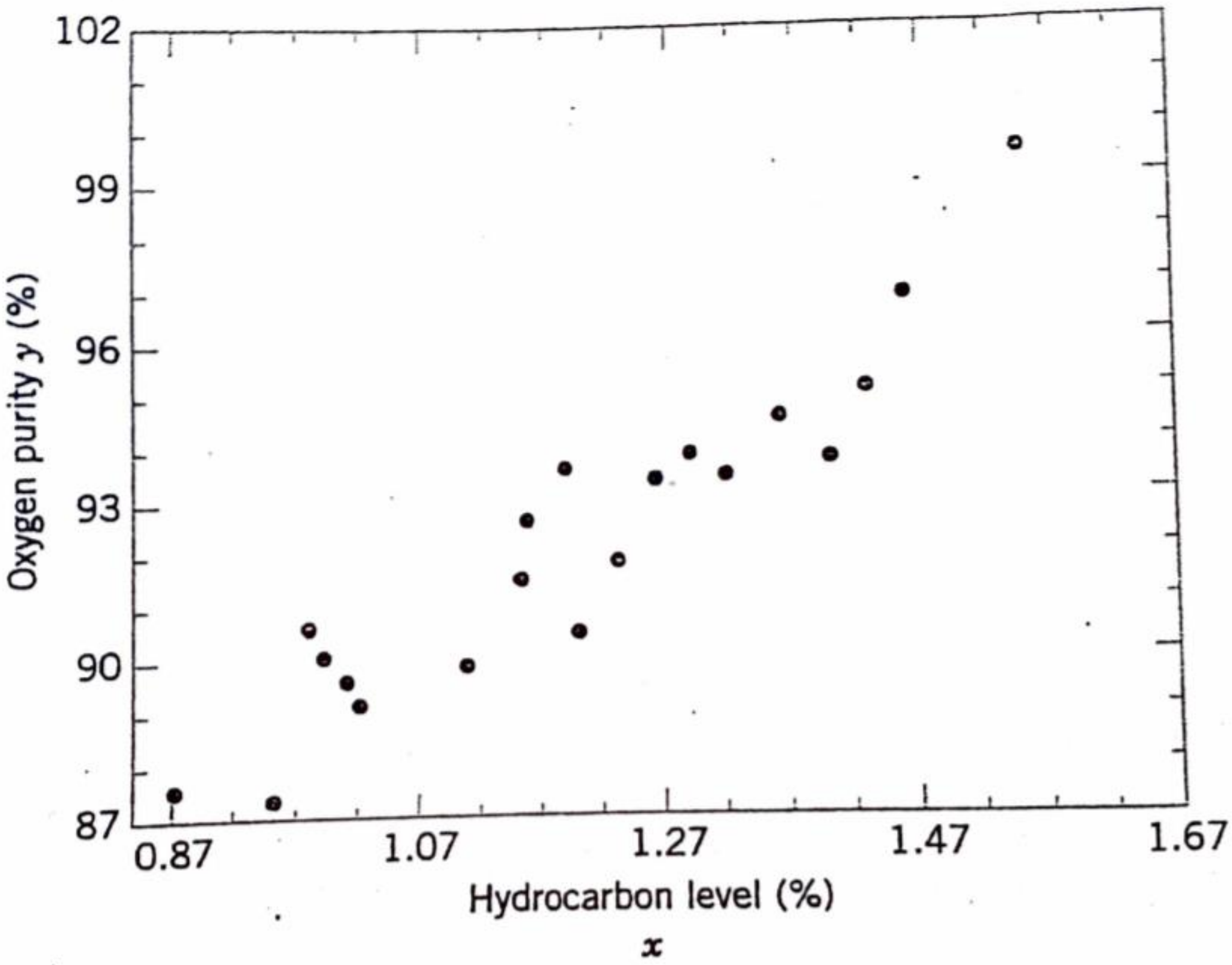
y is the purity of oxygen produced in a chemical distillation process,

and

x is the percentage hydrocarbons that are present in the main condenser of the distillation unit.

Table 10-1 Oxygen and Hydrocarbon Levels

Observation Number	Hydrocarbon Level x (%)	Purity y (%)
1	0.99	90.01
2	1.02	89.05
3	1.15	91.43
4	1.29	93.74
5	1.46	96.73
6	1.36	94.45
7	0.87	87.59
8	1.23	91.77
9	1.55	99.42
10	1.40	93.65
11	1.19	93.54
12	1.15	92.52
13	0.98	90.56
14	1.01	89.54
15	1.11	89.85
16	1.20	90.39
17	1.26	93.25
18	1.32	93.41
19	1.43	94.98
20	0.95	87.33



102
99
96
93
90
87
Oxygen purity y (%)
0.87
1.07
1.27
1.47
1.67
Hydrocarbon level (%)
x

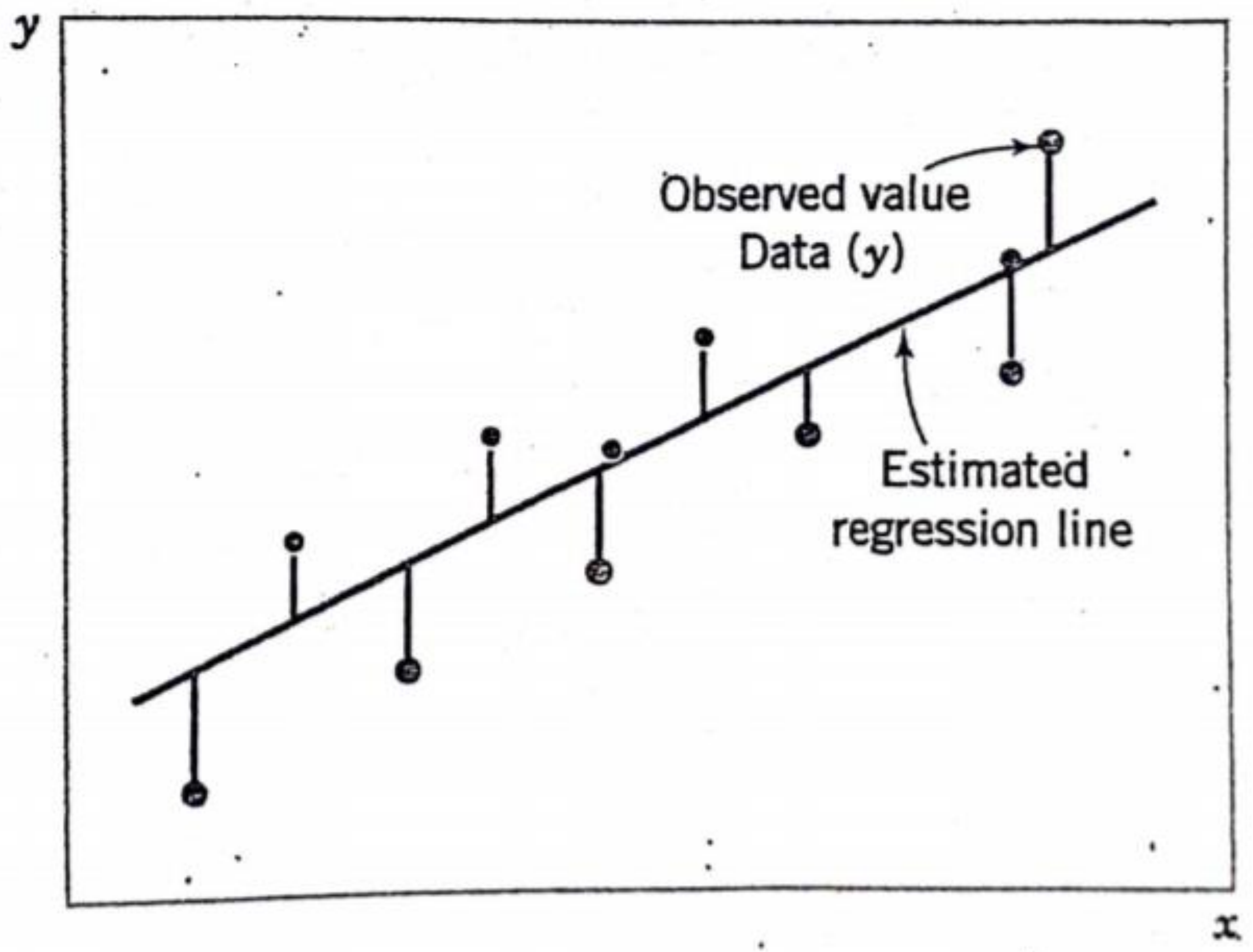


Figure 10-3 Deviations of the data from the estimated regression model.

$$n = 20 \quad \sum_{i=1}^{20} x_i = 23.92 \quad \sum_{i=1}^{20} y_i = 1{,}843.21 \quad \bar{x} = 1.20 \quad \bar{y} = 92.16$$

$$\sum_{i=1}^{20} x_i^2 = 29.29 \quad \sum_{i=1}^{20} x_i y_i = 2{,}214.66$$

refore, the least squares estimates of the slope and intercept are

$$\hat{\beta}_1 = \frac{10.18}{0.68} = 14.97$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 92.16 - (14.97)1.20 = 74.20$$

fitted simple linear regression model is

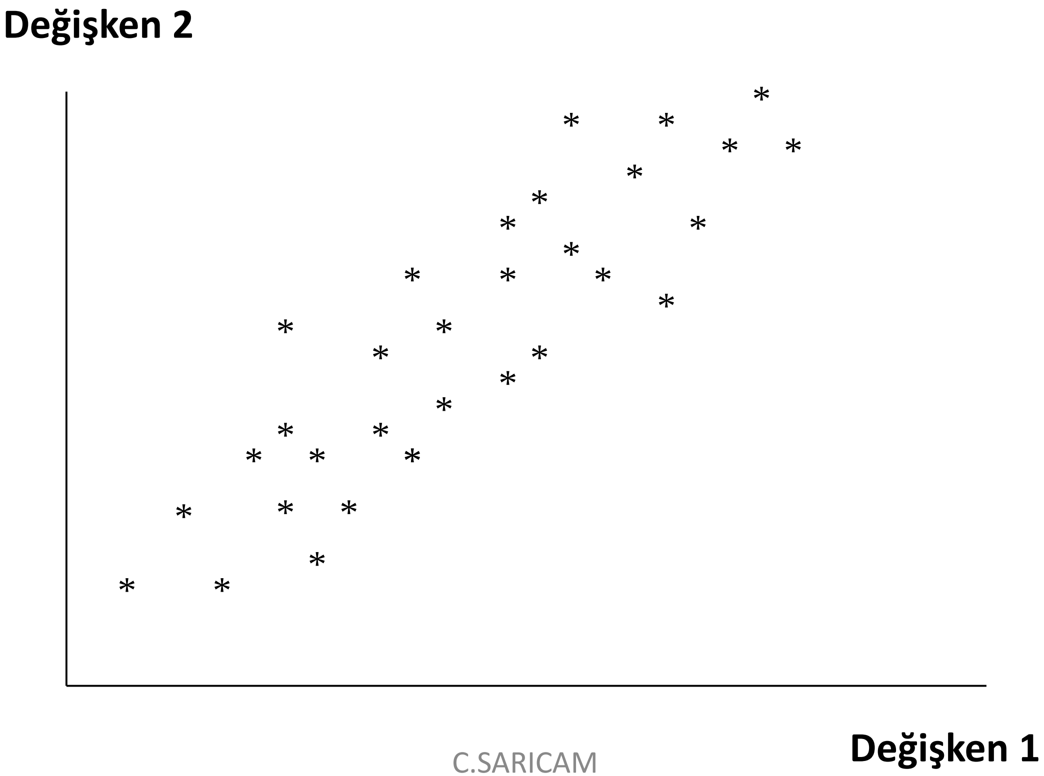
$$\hat{y} = 74.20 + 14.97x$$

CORRELATION

- Correlation quantifies to which two quantitative variables, X and Y go together.
- When high values of X are associated with high values of Y, a positive correlation exists.
- When high values of X are associated with low values of Y, a negative correlation exists.

CORRELATION

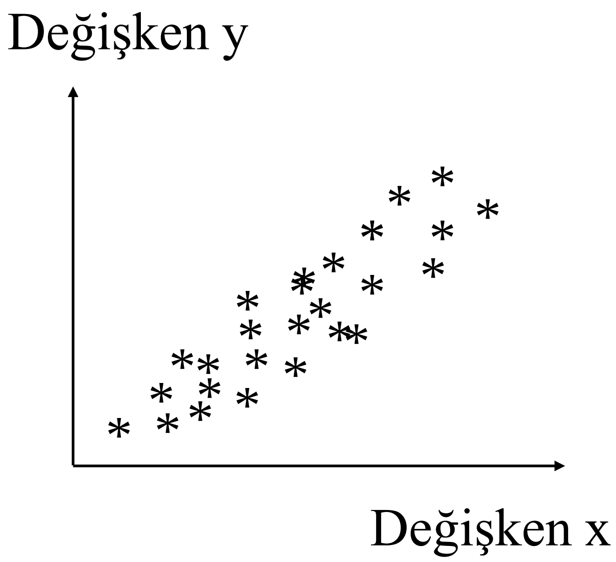
- Scatter diagrams are used to study possible relationships **between 2 variables.**
- These diagrams **can't prove that one variable cau**ses the other, BUT they do indicate the **existence of a relationship**.



CORRELATION

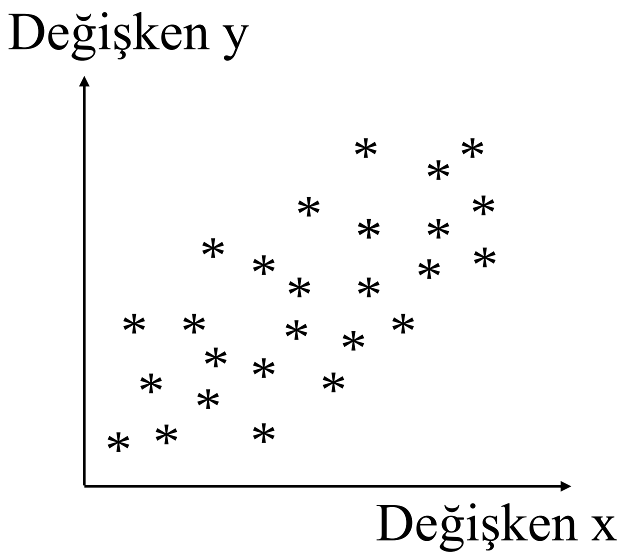
Positive Relation ;

- Y increases as x increases. There is a chance of controlling the variable y by controlling the variable x.



Possible Positive Relation;

- As x increases, y may also increase. But the relationship is not so strong. There may be some other variables that influence the variable y



CORRELATION

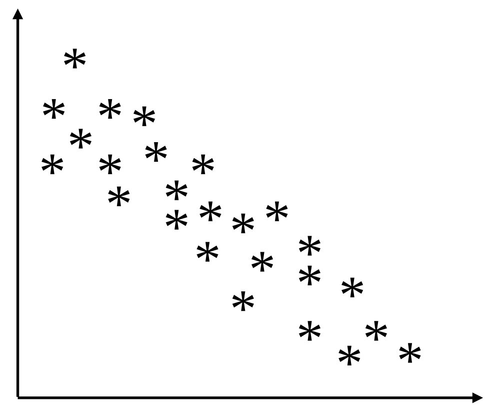
Negative relation ;

- As x increases y decreases. This means that if you can control y, you can control x.

Possible Negative Relation;

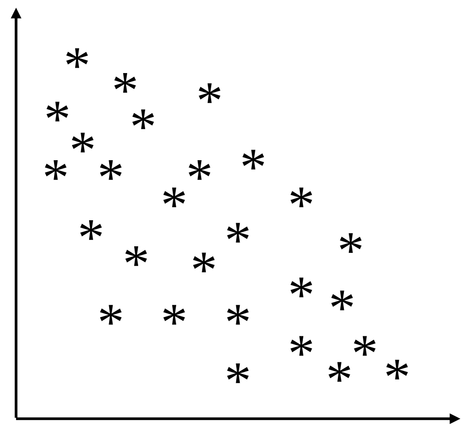
- As x increases, y tends to decrease. But this relatiosnship is not so strong. There may be some other variables effecting the variable y.

Değişken y



Değişken x

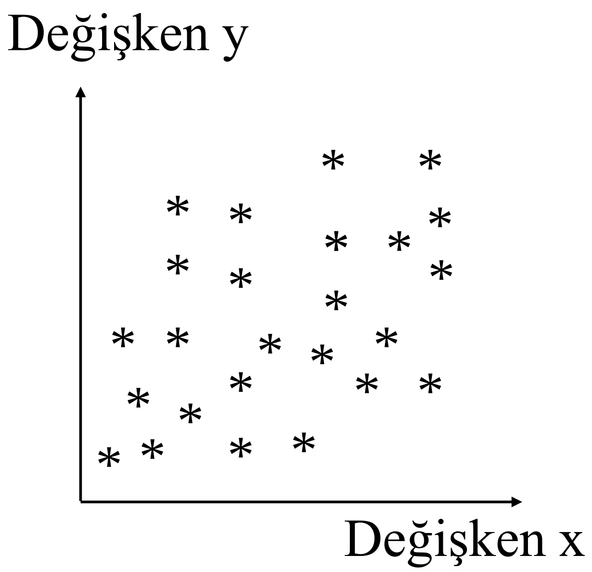
Değişken y



Değişken x

CORRELATION

➢**No relation;** There is no relation between these two variables.



CORRELATION

- If the value of **one variable changes with the change in the value of another variable**, then there may a correlation between these two variables.
- Correlation analysis gives the **numerical value** for the **strength of the relationship.**
- The **correlation coefficient r** measures the strength of relationship between two variables.

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

1. $-1 \leq r \leq 1$
2. When there is a perfect positive correlation $r = 1$
3. When there is a perfect negative correlation $r = -1$
4. When there is no correlation $r = 0$

$$R = \frac{\sum_{i=1}^{n} Y_i(X_i - \overline{X})}{\left[\sum_{i=1}^{n} (X_i - \overline{X})^2 \sum_{i=1}^{n} (Y_i - \overline{Y})^2\right]^{1/2}}$$

R^2 = coefficient of determination
Indicates how well data points fit a statistical model
ranges from 0 to 1.