PSY 121- Statistics in Social Sciences

Covariance: Understanding the Relationship Between Variables

Covariance: Understanding the Relationship Between Variables

- Covariance measures how **two variables change together**.
- It shows whether variables **increase, decrease, or move independently**.

Covariance examines deviations from THE MEAN:

- If deviations multiply positively → variables move in the same direction
- If deviations multiply negatively → variables move in opposite directions
- If covariance ≈ 0 → no systematic relationship

So…

Student	X - M	Y - M	Result (Product)
1	+	–	
2	–	+	
3	–	–	

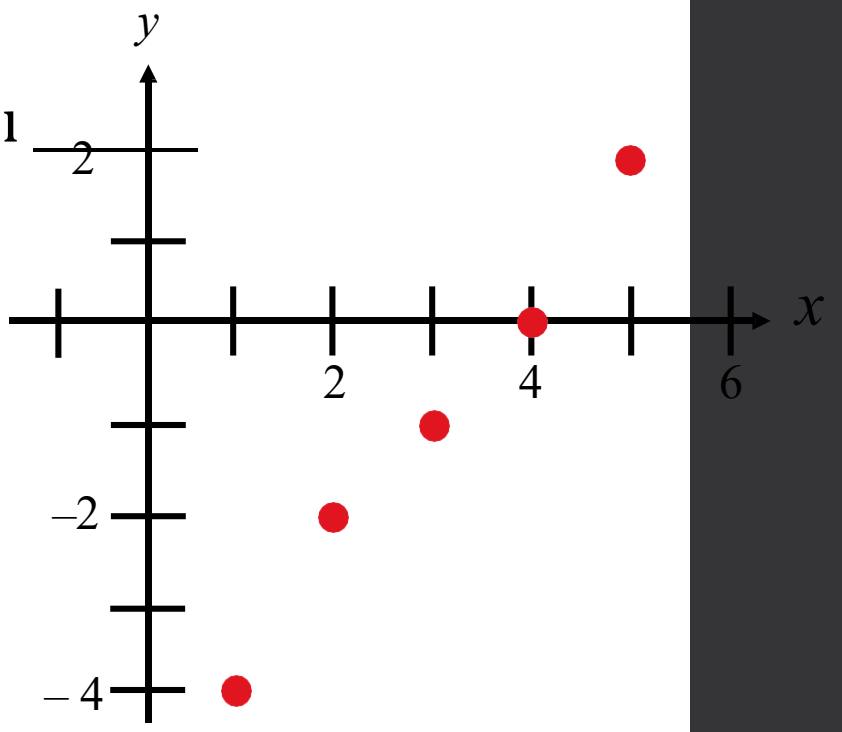
Correlation Analysis

- **Correlation analysis** is a statistical method used to **measure and describe the relationship between two or more variables**.
- In psychology, we use it to understand whether variables **change together** and how **strongly** they are related.
- Correlation does **not** tell us why the relationship exists — **only that it does**.
- Two or more variables are similar or *associated* with each other.
- Correlations are bivariate or multivariate in nature, i.e., relational between variables.

Correlation Analysis

Veriler, x'in bağımsız (açıklayıcı) değişken olduğu ve y'nin bağımlı değişken olduğu, sıralı (x, y) çiftler ile temsil edilebilir.

İki değişken arasında doğrusal (düz çizgi) bir korelasyon olup olmadığını belirlemek için bir saçılım grafiği (scatter plot) kullanılabilir.



Örnek:

x	1	2	3	4	5
y	– 4	– 2	– 1	0	2

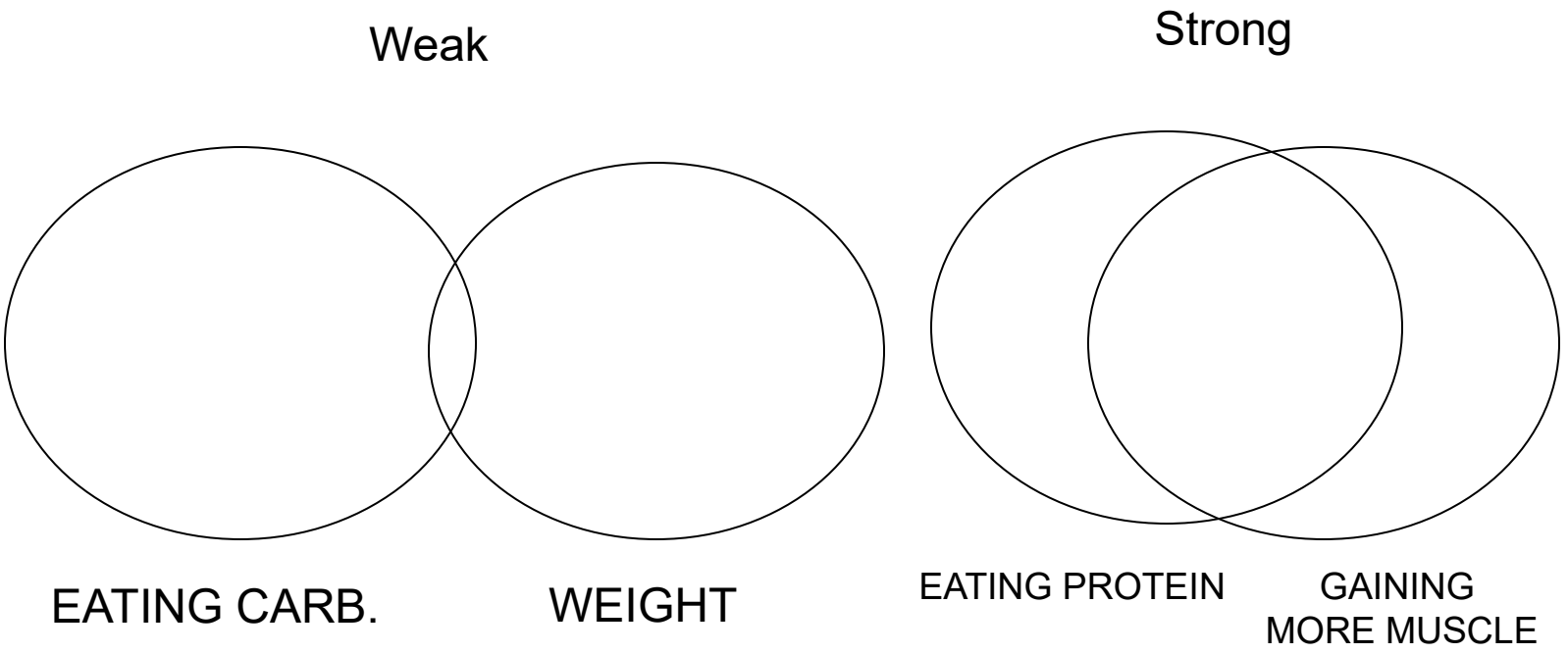
Correlation Analysis

- Correlations measure association by measuring how when one variable changes another changes with it, which means it is measuring dynamic relationships.

Correlation analysis answers questions like:

- Do two variables move in the same direction?
- How strong is the association?
- Is the relationship positive, negative, or zero?
- Can one variable predict another?

Conceptualizing Correlation



Which type of correlation could we say?

Correlation Coefficient (r)

- The strength and direction of the relationship are expressed with a number called the **correlation coefficient**, usually **Pearson's r**.
- It ranges from **–1 to +1**:
- **+1.00 → Perfect positive relationship**
- Both variables increase together.
- **–1.00 → Perfect negative relationship**
- One increases while the other decreases.
- **0 → No relationship**
- Variables do not move together!

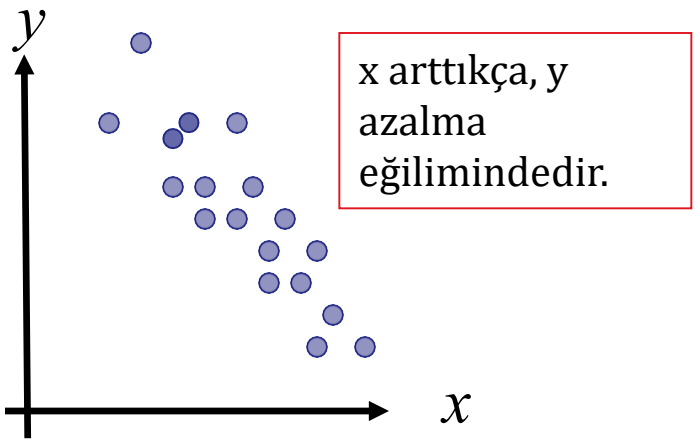
TABLE 5.1 Types of Correlations and the Corresponding Relationship Between Variables

What Happens to Variable X	*What Happens to Variable Y*	*Type of Correlation*	*Value*	*Example*
X increases in value	*Y* increases in value	Direct or positive	Positive, ranging from .00 to +1.00	The more time you spend studying, the higher your test score will be.
X decreases in value	*Y* decreases in value	Direct or positive	Positive, ranging from .00 to +1.00	The less money you put in the bank, the less interest you will earn.
X increases in value	*Y* decreases in value	Indirect or negative	Negative, ranging from –1.00 to .00	The more you exercise, the less you will weigh.
X decreases in value	*Y* increases in value	Indirect or negative	Negative, ranging from –1.00 to .00	The less time you take to complete a test, the more you'll get wrong.

Correlation Coefficient Int.

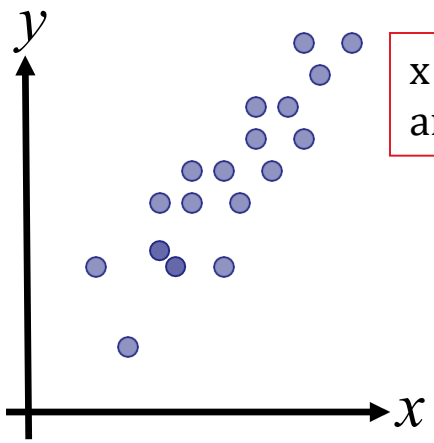
Size of correlation coefficient	General Interpretation
.8 - 1.0	Very Strong
.6 - .8	Strong
.4 - .6	Moderate
.2 - .4	Weak
.0 - .2	Very Weak or no relationship

Linear Correlation (Scatter Plot)



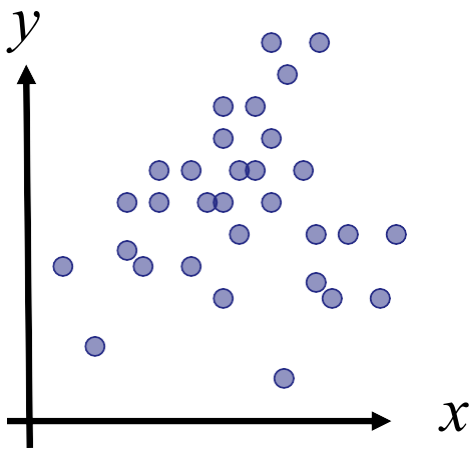
x arttıkça, y azalma eğilimindedir.

Negative Linear Correlation



x arttıkça, y artar.

Positive Linear Correlation



No Correlation

What type of relationship do you see?

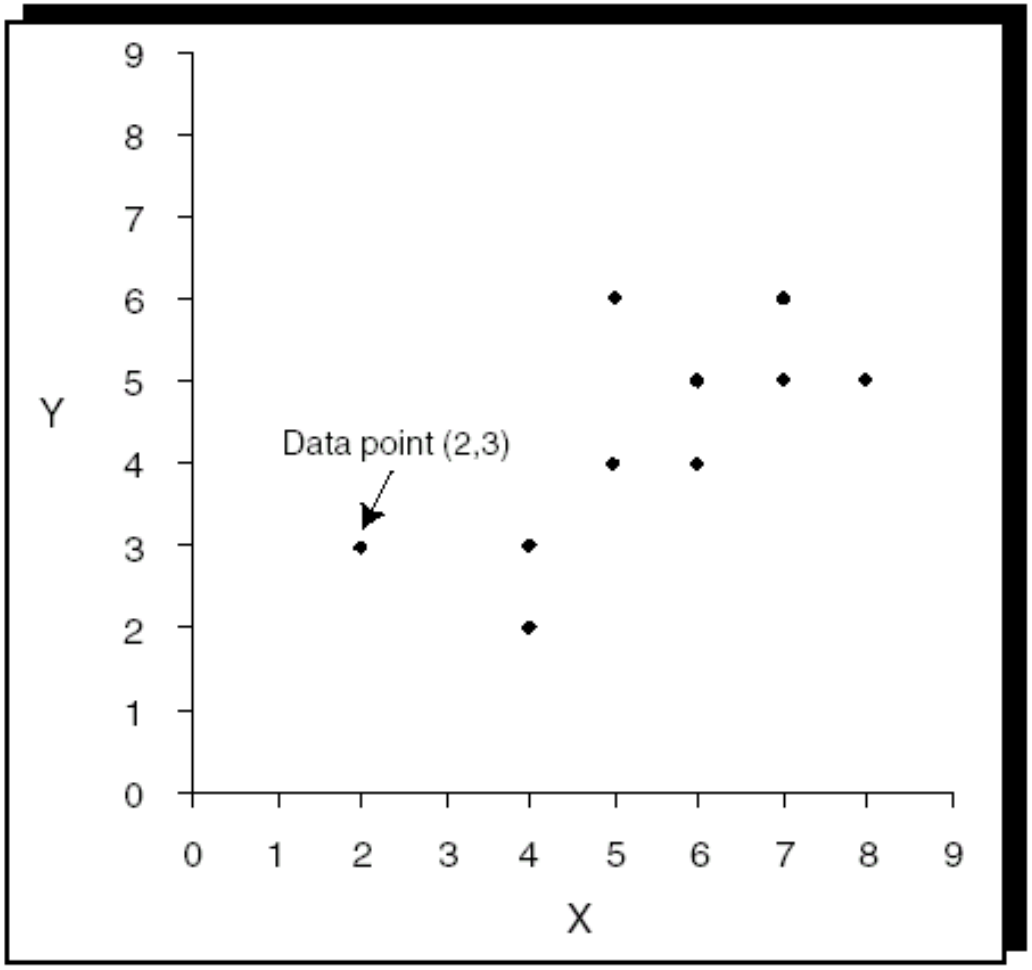


Figure 5.1. A Simple Scatterplot

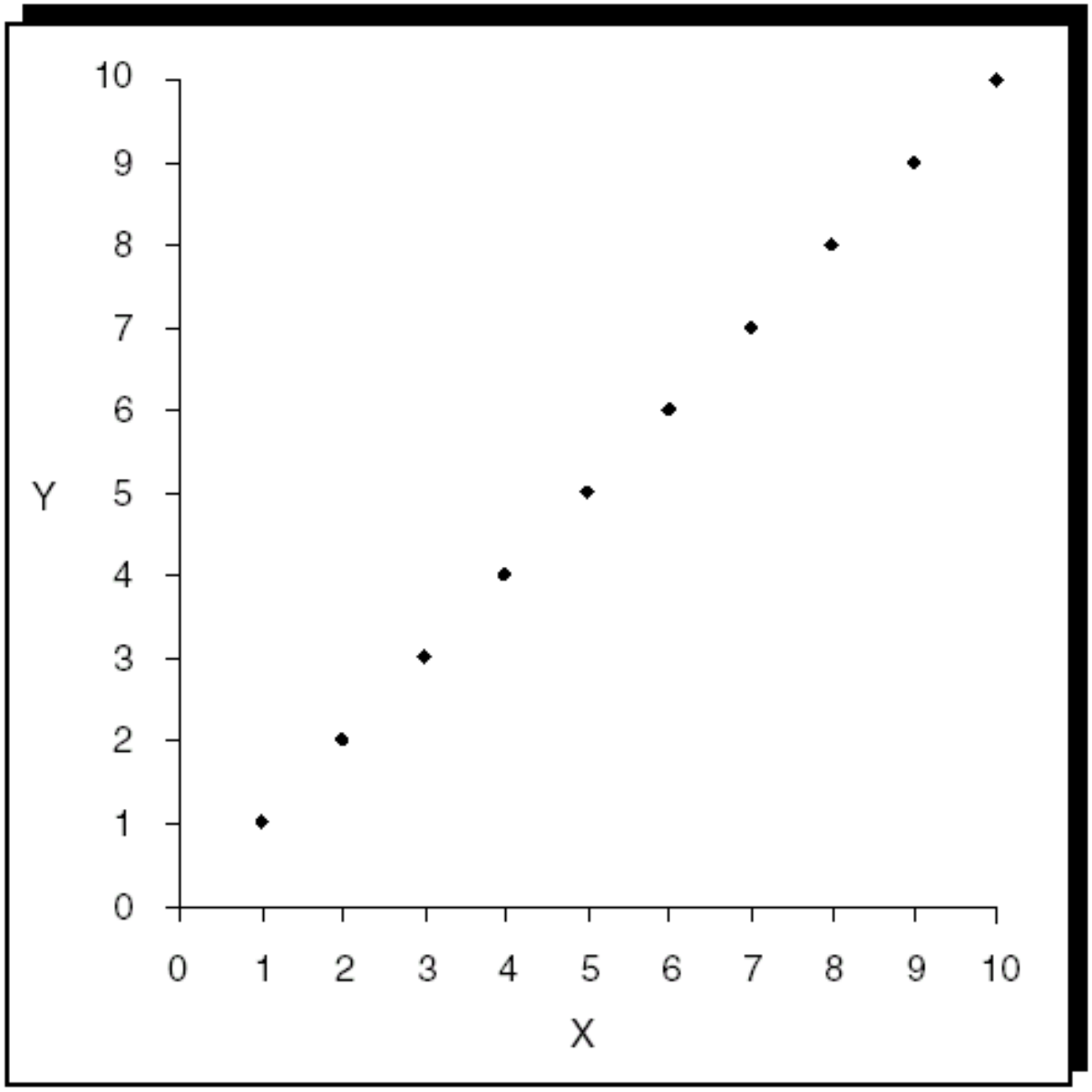
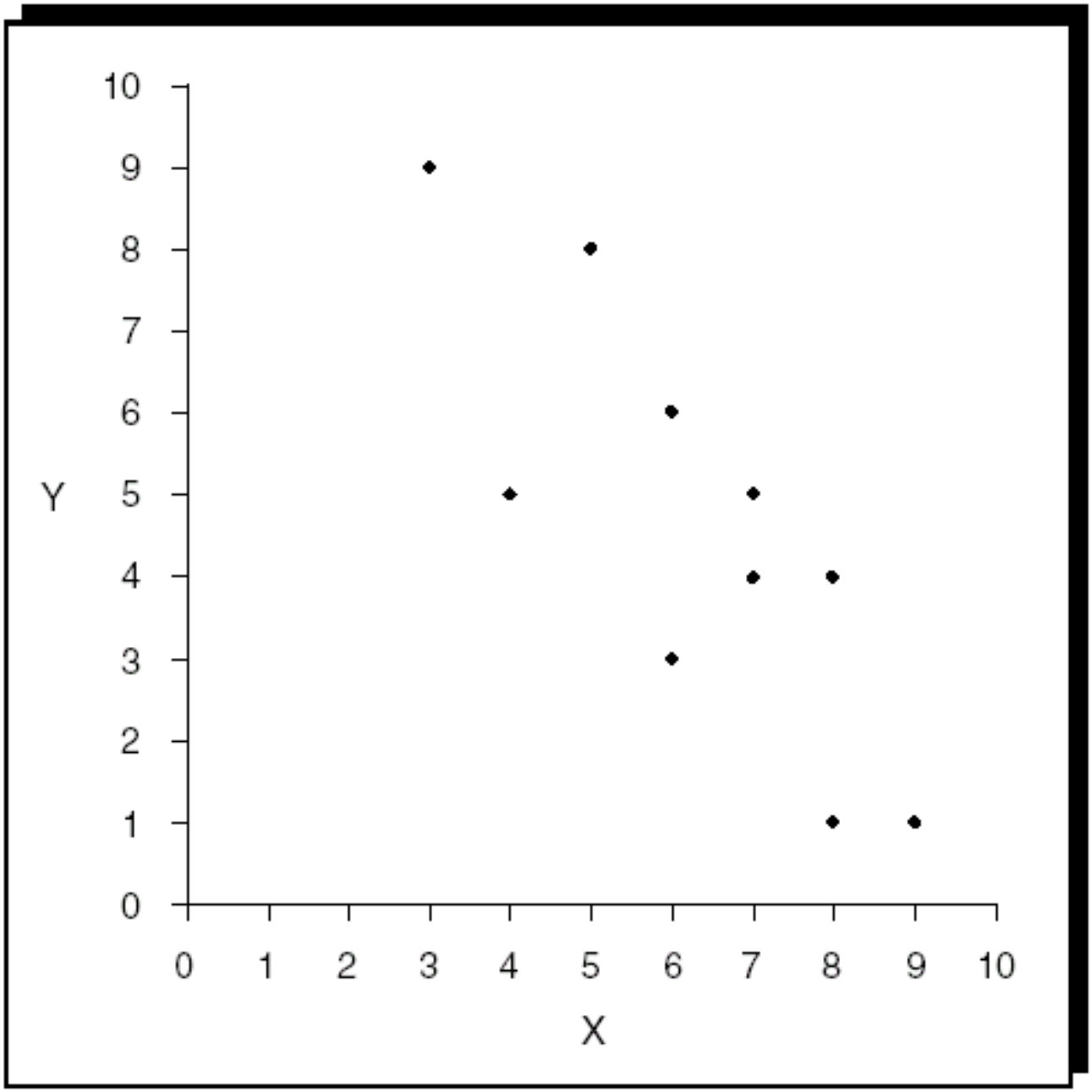


Figure 5.2. A Perfect [redacted] Correlation

The Figure is a perfect linear relationship between X and Y.

In your opinion, is it positive or negative?



What relationship do you see here?

Figure 5.4. A Strong Indirect Relationship

Correlation Analysis
What is Pearson Correlation?

- Pearson correlation (r) measures the **strength and direction of the linear relationship** between two continuous variables.
- It is the **most commonly used correlation coefficient** in psychology.
- Values range from **–1 to +1**:
  - **+1:** Perfect positive linear relationship
  - **–1:** Perfect negative linear relationship
  - **0:** No linear relationship

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \cdot \sum(Y - \bar{Y})^2}}$$

Pearson Correlation Formula
(En temel formül)

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \cdot \sum(Y - \bar{Y})^2}}$$

X, Y: Variables
$\bar{X}$, $\bar{Y}$: Means of X and Y
Σ: Sum across all observations

p-value (Statistical Significance) and Correlation

- In correlation, **r** shows the strength and direction of the relationship between two variables.
- However, **observing a relationship** and **finding it statistically reliable** are different. This is where **p-value!!!**

What is p-value?

- The p-value is the probability that the observed relationship is **due to chance**.
- **Small p-value → relationship is unlikely to be due to chance → significant**
- **Large p-value → relationship may be due to chance → not significant**

p-value (Statistical Significance) and Correlation

p = 0.05 & p = 0.01 & p= 0.001…

p = 0.05 means; There is a **5% chance** that the observed relationship is due to random variation.

- In other words: we can be **95% confident that the relationship is real**.
- This is the most commonly used significance threshold in psychology.

Example:

- r = 0.30, p = 0.05 → moderate positive correlation
- This correlation is likely real (95% confidence), but there is a 5% chance it could be random.
- p= 0.01 → only **1% chance** the observed correlation is random!
- **99% confident** that the correlation is real → very strong statistical evidence
- Often used for larger samples…

The Main Differences Between r and p

Feature	r (Correlation Coefficient)	p-value
What it shows	**Strength and direction** of the relationship	Statistical **significance** of the relationship
Range	–1 … +1	0 … 1
Interpretation	0.1 = weak, 0.3 = moderate, 0.5+ = strong	<0.05 = significant, ≥0.05 = not significant
Example	r = 0.25 → moderate positive correlation	p = 0.03* → statistically significant

Correlation and SPSS

Correlations

		GNPPERCAP	Urbanpop
GNPPERCAP	Pearson Correlation	1	.643**
	Sig. (2-tailed)	.	.000
	N	202	199
Urbanpop	Pearson Correlation	.643**	1
	Sig. (2-tailed)	.000	.
	N	199	260

**. Correlation is significant at the 0.01 level (2-tailed).

Another Example

Correlations

		Democ	Rebellion
Democ	Pearson Correlation	1	-.182*
	Sig. (2-tailed)	.	.032
	N	262	140
Rebellion	Pearson Correlation	-.182*	1
	Sig. (2-tailed)	.032	.
	N	140	156

*. Correlation is significant at the 0.05 level (2-tailed).

Limitation of Correlation Coefficients

Limitation		
Correlation ≠ Causation	Korelasyon, nedensellik anlamına gelmez.	Ortak bir üçüncü değişken ilişkiyi açıklıyor olabilir.
Only Linear Relationships	Sadece doğrusal ilişkileri ölçer.	Eğrisel/non-linear ilişkiler düşük r ile görünür.
Sensitive to Outliers	Aykırı değerler r'yi ciddi şekilde etkiler.	Uç değerler korelasyonu bozabilir.
Restricted Range / Measurement Errors	Dar veri aralığı veya ölçüm hataları korelasyonu azaltır.	Gerçek ilişki olduğundan düşük gözükebilir.
Direction & Mechanism Not Explained	Korelasyon gücü ve yönü gösterir, nedeni açıklamaz.	Yordayıcı değildir.
Assumes Interval/Ratio & Normality	Sürekli ve normal dağılan veriler için uygundur.	

Corelation and Causation

- Two variables are strongly correlated does not mean that there is **a cause–effect** relationship between them.
- If there is a significant relationship between two variables, **you should consider the following possibilities:**

1. **Is there a direct cause–effect relationship between the variables?**
   - Does X cause Y?
2. **Is there a reverse cause–effect relationship?**
   - Does Y cause X?
3. **Could the relationship between the variables be due to a third variable or a combination of other variables?**
4. **Could the relationship between the two variables be due to chance?**

Corelation and Causation

For example;

- **Dondurma tüketimi ↑**
- **Boğulma vakaları ↑**

Corelation and Causation

- Bu iki değişken yüksek korelasyona sahip olabilir ama **dondurma boğulmaya neden olmaz!**

- Her ikisinin de nedeni *yaz aylarında sıcaklık* olabilir → **üçüncü değişken etkisi**.

Any questions??