PRACTICAL EXERCISES - DETAILED ANSWERS
Slides 23-26: Correlation Analysis Exercises

SLIDE 23: Task 1 - Identify Correlation Strength

Question:
For the following correlations:
r = 0.29   r = -0.63   r = 0.15   r = -0.34   r = 0.04

1. Which is the strongest correlation?
2. Which is the weakest correlation?

DETAILED ANSWERS:

Answer 1: Strongest Correlation
r = -0.63 is the strongest correlation

Explanation:
- Correlation strength is determined by the **absolute value** (ignoring the sign)
- Absolute values: |0.29| = 0.29, |-0.63| = 0.63, |0.15| = 0.15, |-0.34| = 0.34, |0.04| = 0.04
- The largest absolute value is 0.63
- The negative sign indicates direction (inverse relationship), not strength
- r = -0.63 shows a **strong negative correlation**
  - As one variable increases, the other strongly decreases
  - Example: Higher medication adherence → Lower disease symptoms

Strength Classification:
- |r| = 0.00 to 0.29: Weak correlation
- |r| = 0.30 to 0.69: Moderate correlation
- |r| = 0.70 to 1.00: Strong correlation

Therefore, -0.63 falls in the **moderate-to-strong** range.

Answer 2: Weakest Correlation
r = 0.04 is the weakest correlation

Explanation:
- Absolute value: |0.04| = 0.04
- This is closest to zero (no correlation)
- r = 0.04 indicates almost no linear relationship between variables
- Example: Patient's shoe size and blood pressure (no meaningful connection)
- In practical terms, this correlation is so weak it's essentially **negligible**

Visual Interpretation:
- A scatterplot of r = 0.04 would show points scattered randomly
- No discernible pattern or trend line
- Knowing one variable tells you almost nothing about the other

---

Summary Table:

Correlation	Absolute Value	Strength	Direction
r = -0.63	0.63	**Strongest** (Moderate-Strong)	Negative

r = -0.34	0.34	Moderate	Negative
r = 0.29	0.29	Weak-Moderate	Positive
r = 0.15	0.15	Weak	Positive
r = 0.04	0.04	**Weakest** (Negligible)	Positive

Healthcare Context Example:
If these correlations represented:
- **r = -0.63**: Exercise frequency vs. Resting heart rate (strong negative - more exercise, lower resting HR)
- **r = -0.34**: Fiber intake vs. Cholesterol (moderate negative)
- **r = 0.29**: Age vs. Recovery time (weak positive)
- **r = 0.15**: Patient height vs. Medication dosage (weak positive)
- **r = 0.04**: Patient birthdate (day of month) vs. Blood pressure (negligible)

SLIDE 24: Task 2 - Visualize the Data (Scatterplot)

Data Provided:
Heart Disease Study: Cholesterol vs. Age

- **Ages (X):** 45, 52, 38, 61, 49, 55, 43, 58, 50, 47
- **Cholesterol (Y):** 180, 210, 165, 245, 195, 225, 175, 235, 200, 185

DETAILED SOLUTION:

Step 1: Create Data Table

Patient	Age (X)	Cholesterol (Y)
1	45	180
2	52	210
3	38	165
4	61	245
5	49	195
6	55	225
7	43	175
8	58	235
9	50	200
10	47	185

Step 2: Plot the Scatterplot

Instructions for creating the scatterplot:

1. **X-axis (Horizontal):** Patient Age
   - Range: 35 to 65 (to accommodate all values with padding)
   - Label: "Patient Age (years)"

2. **Y-axis (Vertical):** Cholesterol Level
   - Range: 160 to 250 (to accommodate all values with padding)
   - Label: "Cholesterol Level (mg/dL)"

3. **Plot each data point:**
   - (45, 180), (52, 210), (38, 165), (61, 245), (49, 195)
   - (55, 225), (43, 175), (58, 235), (50, 200), (47, 185)

4. **Title:** "Age vs. Cholesterol Level in Heart Disease Study"

Step 3: Observe the Pattern

What pattern do you observe?

Answer: POSITIVE CORRELATION

Detailed Observation:
1. **Trend Direction:** As age increases, cholesterol level tends to increase
2. **Pattern Type:** Points roughly follow an upward-sloping line from bottom-left to top-right
3. **Strength:** Moderate to strong positive correlation (we'll calculate exact value in Task 3)
4. **Outliers:** No major outliers; all points follow the general trend

Specific Observations:
- Youngest patient (38 years) has lowest cholesterol (165 mg/dL)
- Oldest patient (61 years) has highest cholesterol (245 mg/dL)
- Middle-aged patients have intermediate cholesterol levels
- The relationship appears roughly linear (not curved)

Visual Description of the Scatterplot:

```
Cholesterol (mg/dL)
250 |                                    ●
    |                              ●
225 |                         ●
    |                    ●
200 |               ●
    |          ●
175 |     ●                ●
    |          ●
165 | ●
    |
    +-----|-----|-----|-----|-----|-----|
         38    43    48    53    58    63
                   Age (years)
```

Trend Line: Would slope upward from left to right

Clinical Interpretation:

What does this pattern mean in healthcare?

1. **Age-Related Changes:** Cholesterol tends to increase with age
2. **Risk Assessment:** Older patients may have higher cardiovascular risk
3. **Screening Importance:** Regular cholesterol monitoring becomes more critical with age
4. **Intervention Timing:** Early lifestyle modifications may prevent age-related increases

Important Note:
Correlation ≠ Causation. Age doesn't *cause* high cholesterol directly, but:
- Metabolic changes occur with aging
- Longer exposure to dietary factors

- Cumulative lifestyle effects
- Natural biological processes

SLIDE 25: Task 3 - Calculate Correlation (15 min)

Formula Provided:
$r = \Sigma(Zx \times Zy) / (n - 1)$

Where:
- $Zx = (X - \bar{X}) / SDx$
- $Zy = (Y - \bar{Y}) / SDy$
- n = number of data points

DETAILED CALCULATION:

Step 1: Calculate Means

Mean Age ($\bar{X}$):
$\bar{X} = (45 + 52 + 38 + 61 + 49 + 55 + 43 + 58 + 50 + 47) / 10$
$\bar{X} = 498 / 10$
$\bar{X} = 49.8$ years

Mean Cholesterol ($\bar{Y}$):
$\bar{Y} = (180 + 210 + 165 + 245 + 195 + 225 + 175 + 235 + 200 + 185) / 10$
$\bar{Y} = 2015 / 10$
$\bar{Y} = 201.5$ mg/dL

Step 2: Calculate Standard Deviations

Standard Deviation of Age (SDx):

First, calculate deviations and squared deviations:

Age (X)	$X - \bar{X}$	$(X - \bar{X})^2$
45	-4.8	23.04
52	2.2	4.84
38	-11.8	139.24
61	11.2	125.44
49	-0.8	0.64
55	5.2	27.04
43	-6.8	46.24
58	8.2	67.24
50	0.2	0.04
47	-2.8	7.84
Sum		**441.60**

Variance = $\Sigma(X - \bar{X})^2 / (n - 1) = 441.60 / 9 = 49.067$

$SDx = \sqrt{49.067} = 7.00$ years

Standard Deviation of Cholesterol (SDy):

Chol (Y)	$Y - \bar{Y}$	$(Y - \bar{Y})^2$

180	-21.5	462.25
210	8.5	72.25
165	-36.5	1332.25
245	43.5	1892.25
195	-6.5	42.25
225	23.5	552.25
175	-26.5	702.25
235	33.5	1122.25
200	-1.5	2.25
185	-16.5	272.25
Sum		**6452.50**

Variance = $\Sigma(Y - \bar{Y})^2 / (n - 1) = 6452.50 / 9 = 716.944$

$SD_y = \sqrt{716.944} = 26.78$ mg/dL

Step 3: Calculate Z-Scores and Products

Patient	Age (X)	Chol (Y)	$Z_x = (X-\bar{X})/SD_x$	$Z_y = (Y-\bar{Y})/SD_y$	$Z_x \times Z_y$
1	45	180	-0.686	-0.803	0.551
2	52	210	0.314	0.317	0.100
3	38	165	-1.686	-1.363	2.298
4	61	245	1.600	1.624	2.598
5	49	195	-0.114	-0.243	0.028
6	55	225	0.743	0.877	0.652
7	43	175	-0.971	-0.990	0.961
8	58	235	1.171	1.251	1.465
9	50	200	0.029	-0.056	-0.002
10	47	185	-0.400	-0.616	0.246
				$\Sigma(Z_x \times Z_y)$	**8.897**

Step 4: Calculate Correlation Coefficient

$r = \Sigma(Z_x \times Z_y) / (n - 1)$

$r = 8.897 / (10 - 1)$

$r = 8.897 / 9$

$r = 0.989$

INTERPRETATION:

$r = 0.989$ indicates:

1. **Very Strong Positive Correlation**
   - This is close to perfect correlation ($r = 1.0$)
   - The relationship between age and cholesterol is nearly linear

2. **Clinical Significance:**
   - Age is an excellent predictor of cholesterol level in this sample
   - For every 1 standard deviation increase in age (~7 years), we expect approximately 1 standard deviation increase in cholesterol (~27 mg/dL)

3. **Coefficient of Determination (r^2):**
   - $r^2 = (0.989)^2 = 0.978$
   - **97.8% of the variance in cholesterol is explained by age**

   - Only 2.2% is due to other factors

4. **Statistical Strength:**
   - $|r| > 0.9$: Considered very strong correlation
   - This correlation is statistically significant ($p < 0.001$ in any standard test)

Verification with Excel:

Excel Formula:
```
=CORREL(A2:A11, B2:B11)
```
Where:
- A2:A11 contains ages: 45, 52, 38, 61, 49, 55, 43, 58, 50, 47
- B2:B11 contains cholesterol: 180, 210, 165, 245, 195, 225, 175, 235, 200, 185

Excel Result: r = 0.989 ✓

SLIDE 26: Calculation Worksheet - COMPLETED

Full Calculation Table:

#	X (Age)	$X - \bar{X}$	Zx	Y (Chol)	$Y - \bar{Y}$	Zy	$(X-\bar{X})(Y-\bar{Y})$	Zx×Zy
1	45	-4.8	-0.69	180	-21.5	-0.80	103.20	0.551
2	52	2.2	0.31	210	8.5	0.32	18.70	0.100
3	38	-11.8	-1.69	165	-36.5	-1.36	430.70	2.298
4	61	11.2	1.60	245	43.5	1.62	487.20	2.598
5	49	-0.8	-0.11	195	-6.5	-0.24	5.20	0.028
6	55	5.2	0.74	225	23.5	0.88	122.20	0.652
7	43	-6.8	-0.97	175	-26.5	-0.99	180.20	0.961
8	58	8.2	1.17	235	33.5	1.25	274.70	1.465
9	50	0.2	0.03	200	-1.5	-0.06	-0.30	-0.002
10	47	-2.8	-0.40	185	-16.5	-0.62	46.20	0.246
Σ	**498**	**0**		**2015**	**0**		**1668.00**	**8.897**
Mean	**49.8**			**201.5**				
SD			**7.00**			**26.78**		

Alternative Formula Verification:

Using Covariance Method:

$$\text{Covariance} = \Sigma(X - \bar{X})(Y - \bar{Y}) / (n - 1)$$
$$\text{Covariance} = 1668.00 / 9 = 185.33$$

$r = \text{Covariance} / (SDx \times SDy)$
$$r = 185.33 / (7.00 \times 26.78)$$
$$r = 185.33 / 187.46$$
$r = 0.989$ ✓

Both methods give the same result!

KEY TAKEAWAYS:

From These Exercises:

1. **Correlation Strength:** Always use absolute value to determine strength
2. **Direction:** Sign indicates positive or negative relationship
3. **Visualization:** Scatterplots reveal patterns before calculation
4. **Calculation Methods:** Multiple approaches (Z-score, covariance) yield same result
5. **Clinical Relevance:** r = 0.989 shows age is highly predictive of cholesterol in this sample

Important Notes:

⚠️ **Sample Size:** This is a small sample (n=10). Larger samples give more reliable correlations.

⚠️ **Correlation ≠ Causation:** High correlation doesn't prove age *causes* high cholesterol; other factors may be involved.

⚠️ **Clinical Context:** Always interpret statistics within medical and biological context.

SUMMARY ANSWERS AT A GLANCE:

Task	Question	Answer
Slide 23	Strongest correlation?	r = -0.63
Slide 23	Weakest correlation?	r = 0.04
Slide 24	Pattern observed?	Strong positive correlation
Slide 25	Correlation coefficient?	r = 0.989 (very strong positive)
Slide 26	Variance explained?	r^2 = 97.8%

End of Answer Key