



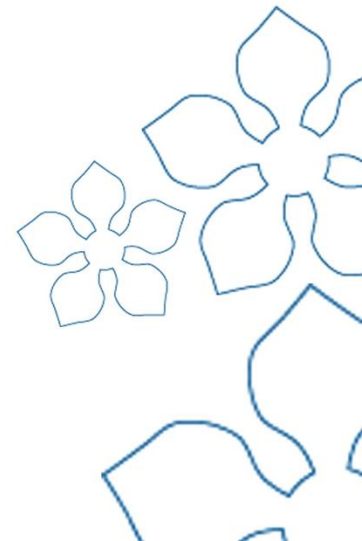
Applied Research, Ethics, and Analysis (AREA)

Understanding quantitative data 2: Introduction to SPSS: T-test and Linear Regression

Dr. Kate Williams



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David



T-test

- Psychological research often involves the investigation of one variable on another:
- This can be achieved by systematically manipulating one variable to look at its effect on the other variable.
 - E.g., the effect of a new teaching style on exam marks.



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

Teaching Style Example

- Option 1: Look at student exam marks before and after new teaching style is implemented.
- Option 2: Compare exam marks of 1 group of students in class using old teaching style against exam marks for another group of students who received new teaching style.

*These are just very basic examples: you would need to take in to consideration a number of factors if these were real studies!



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

Teaching Style Example

- Option 1: For the same participant you would have a mean value of exam marks before new teaching style was implemented and a mean value of exam marks after new style was implemented.
- You would be comparing TWO different means for the SAME person.



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

Teaching Style Example

- Option 2: Compares the means of exam marks from TWO DIFFERENT groups of participants:
- One mean of exam marks from the group who had lessons using the old teaching style and another mean of exam marks from a different group who had lessons using the new style.

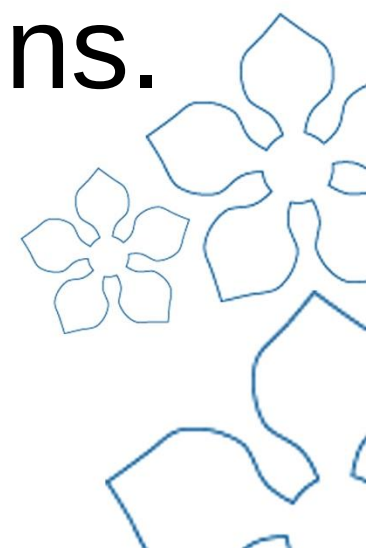


PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David



How can we statistically see if there is a significant difference between the means?

For both option 1 and option 2 of our experimental designs.



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

T-test

- Used to determine whether two means are significantly different to one another.
- A Parametric test: uses actual values
- Data must meet the following assumptions:
 - Interval or ratio data
 - Homogeneity of variance
 - Drawn from population with normal distribution



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

Types of t-test

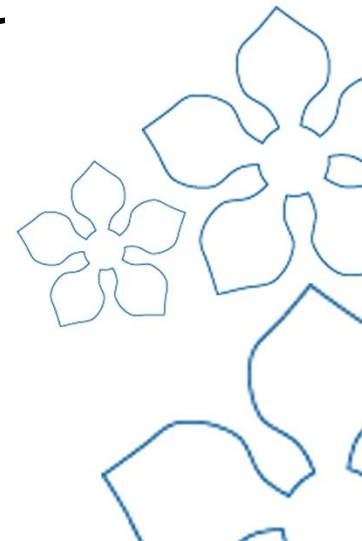
- **Paired samples (*Within*):** compares the mean from two experimental conditions, where the same participant has taken part in both conditions of the experiment.
- **Independent samples (*Between*):** compares the means from two different groups of participants.





How do these different types of t-tests relate to our teaching style examples:

- Option 1 (same P) = Paired samples t-test
- Option 2 (different Ps) = Independent samples t-test



Presenting results: Paired samples

- State the ***t*-value**, **degrees of freedom (df)** and **significance** (less than or more than .05)

$$t(14) = -4.02, p < .05$$

Degrees of freedom: For paired-samples it will be one less than the number of participants. For independent-samples it will be one less from each group, so the df would be two less than the total number of Participants.

The sign of the *t*-value just tells the direction of the difference

T-value from SPSS output

Significance value

Paired samples Write-up

- A paired-samples t-test was conducted in order to evaluate teaching style (old/new) on student exam marks. There was a statistically significant increase in exam scores following the implementation of the new teaching style ($M = 80.2$, $SD = 2.15$) compared to the old teaching style ($M = 60.8$, $SD = 1.73$), $t(14) = -4.02$, $p < .01$.



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

Paired samples t-test worksheet



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

Independent Samples t-test

- Similar to the paired samples t-test
 - BUT!!!!

How you set up the data is critical.

- Data needs to be entered in 2 columns: a **group variable** which has a number as a code to indicate which group a P belongs and a single column of the scores.



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

Presenting results: Independent samples

- Again, state the ***t*-value**, the **degrees of freedom (df)** and the **significance** (less than or more than .05)

$t(df) = t\text{-value}, p (</>).05$



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

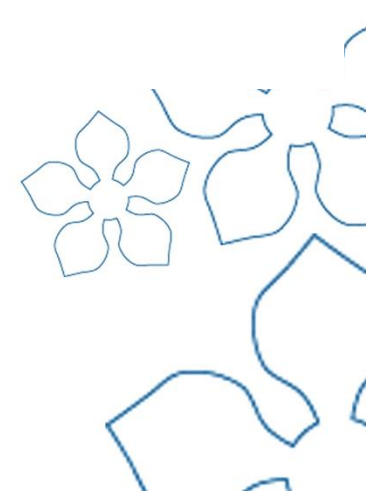


Independent samples Write-up

- An independent-samples t-test was conducted in order to compare the neuroticism scores between male and female participants. Neuroticism scores were significantly higher for females ($M = 17.20$, $SD = 4.42$) compared to male scores ($M = 12.90$, $SD = 2.81$), $t(18) = -2.59$, $p < .05$.



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David



Independent samples t-test worksheet



PRIFYSGOL CYMRU

Y Drindod Dewi Sant

UNIVERSITY OF WALES

Trinity Saint David

T-test analysis example from the CARIAD Trial



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

The Positive and Negative Affect Schedule PANAS

Watson, D., Clarke, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS Scales. *Journal of Personality and Social Psychology*, 54, 1063-70.

Scores for each scale are derived by adding item scores (1-5) for the ten PA adjectives for the PA scale and for the ten NA adjectives for the NA scale.

1=Very slightly/not at all;

2= A little;

3= Moderately;

4= Quite a bit;

5= Extremely.

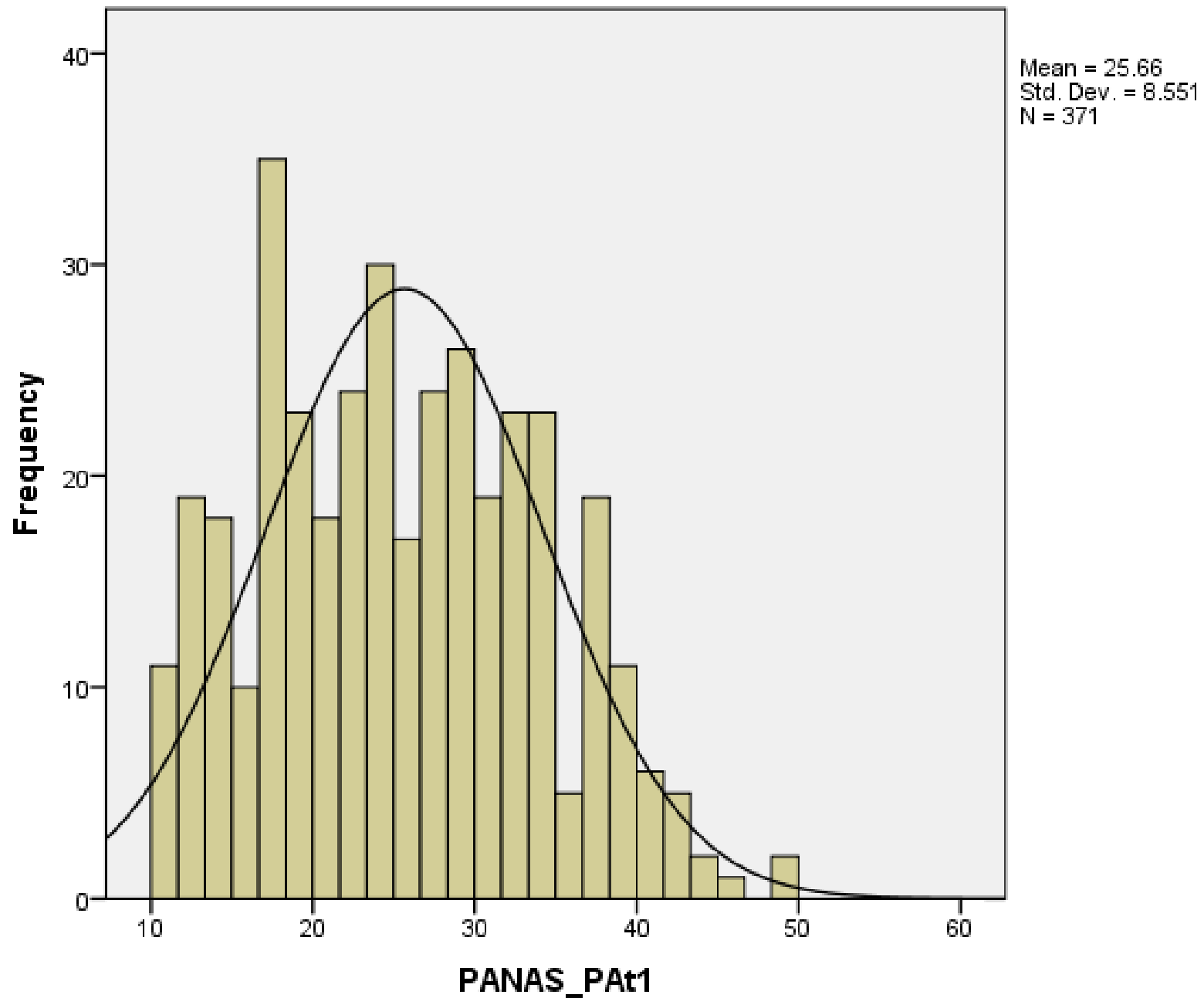
Total NA and PA scores can range from 10 to 50.

High scores indicate high levels of positive or negative affect.

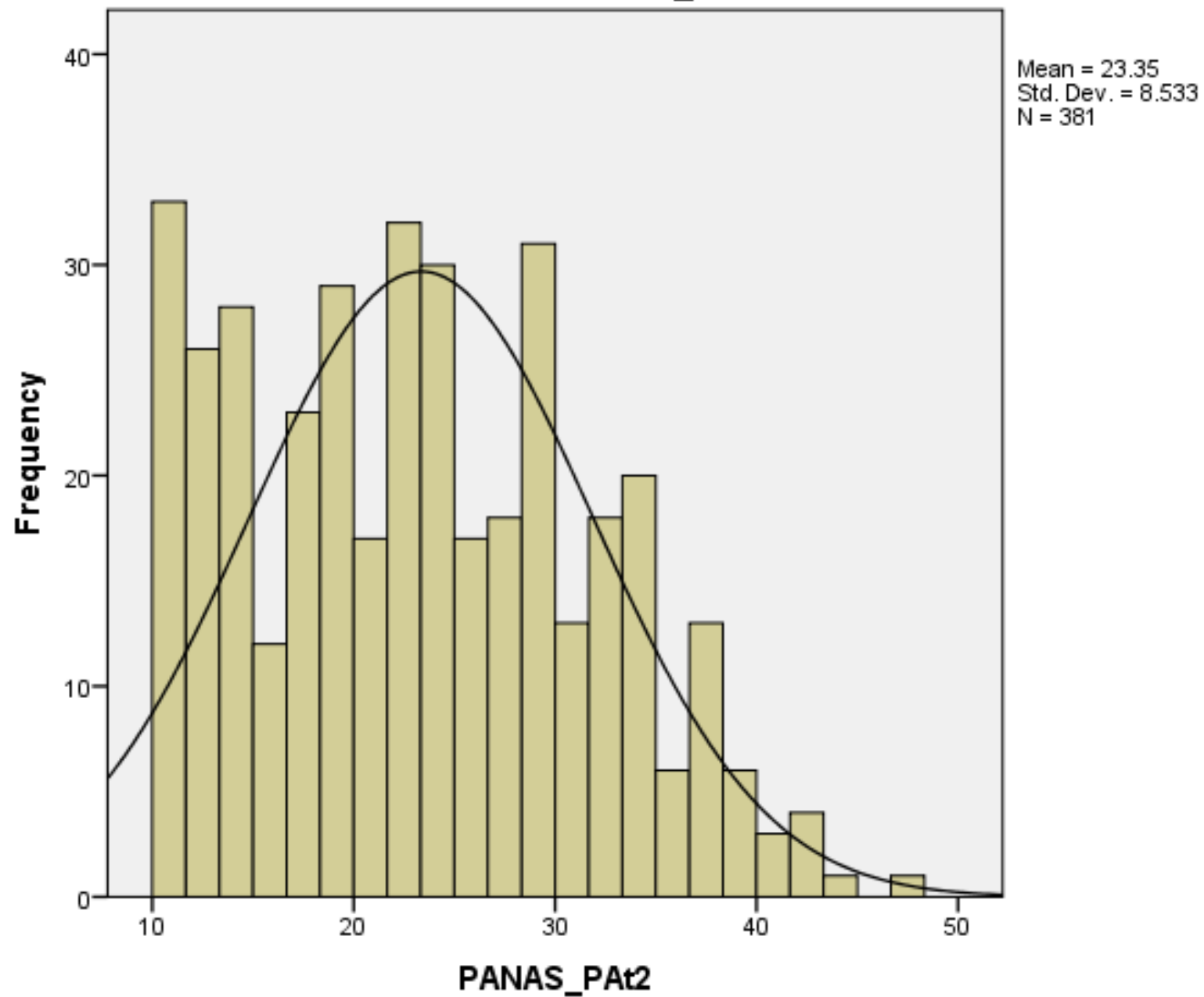
Statistics

		PANAS_PAt1	PANAS_PAt2
N	Valid	371	381
	Missing	15	5
Mean		25.66	23.35
Median		25.00	23.00
Mode		25	10
Std. Deviation		8.551	8.533
Variance		73.117	72.811
Skewness		.211	.265
Std. Error of Skewness		.127	.125
Kurtosis		-.687	-.764
Std. Error of Kurtosis		.253	.249
Range		40	37
Minimum		10	10
Maximum		50	47

PANAS_PAt1



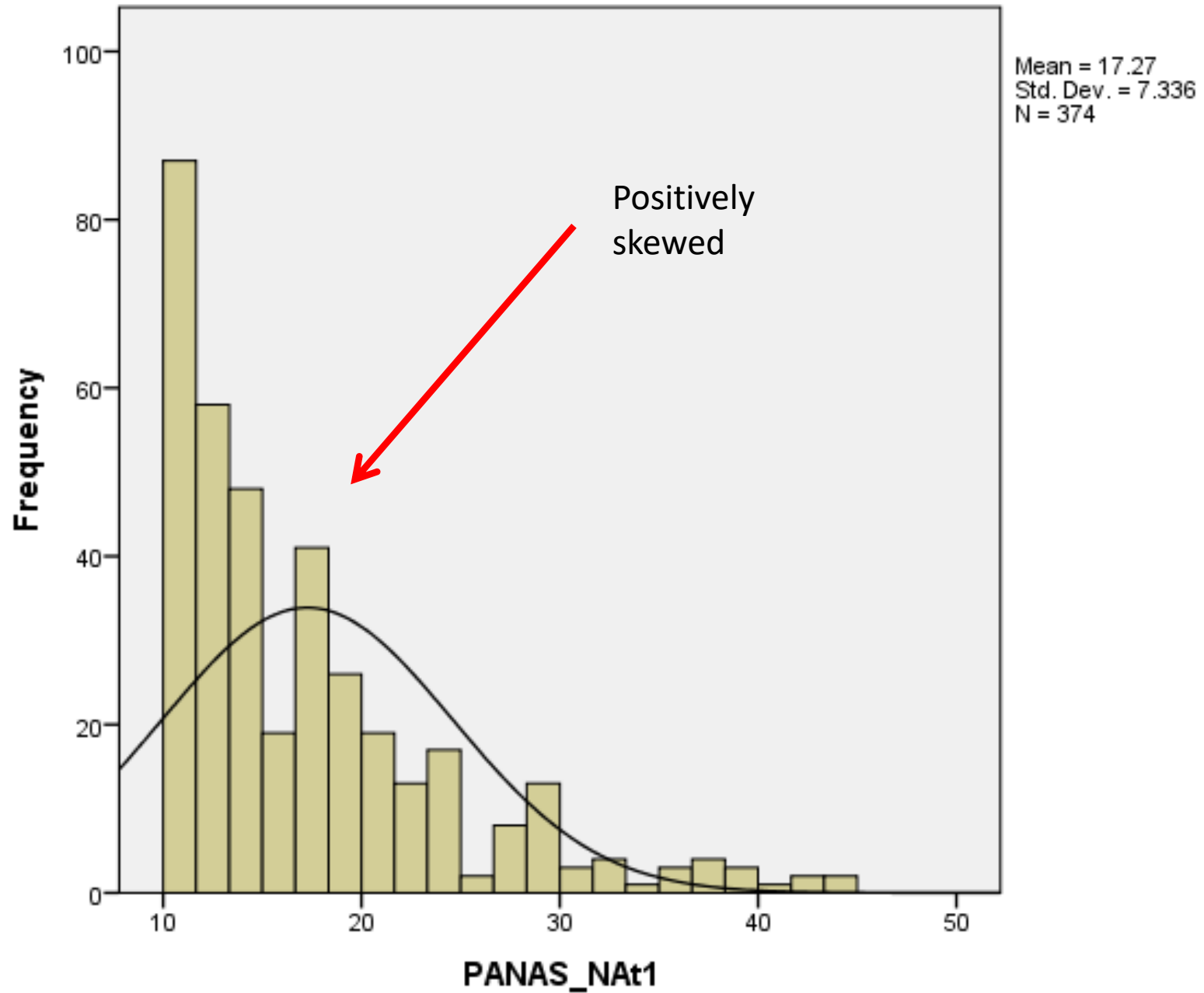
PANAS_PAt2



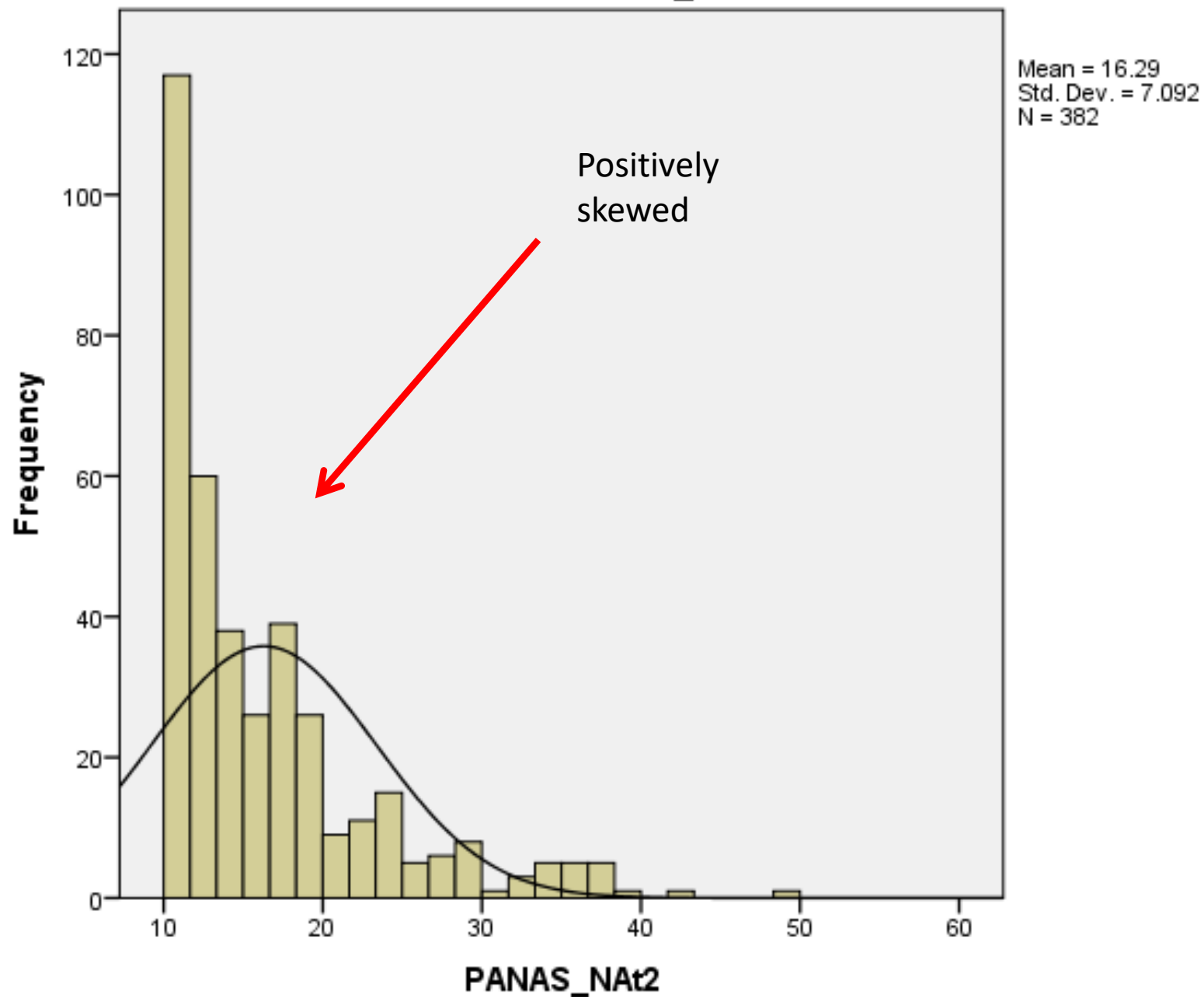
Statistics

		PANAS_NAt1	PANAS_NAt2
N	Valid	374	382
	Missing	12	4
Mean		17.27	16.29
Median		15.00	14.00
Mode		10	10
Std. Deviation		7.336	7.092
Variance		53.818	50.295
Skewness		1.441	1.578
Std. Error of Skewness		.126	.125
Kurtosis		1.923	2.480
Std. Error of Kurtosis		.252	.249
Range		35	40
Minimum		10	10
Maximum		45	50

PANAS_NAt1



PANAS_NAt2



Do participants levels of positive affect change over time?

Hypothesis 1: Participants will report significantly lower levels of positivity at Time 2 than they did at Time 1

How do we explore this?

What test?

Hypothesis 1: Participants will report significantly lower levels of positivity at Time 2 than they did at Time 1

- **Within-groups test**

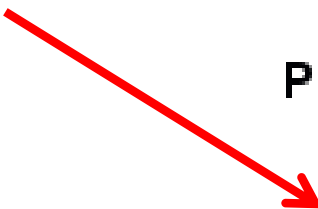
- **Paired Samples t-test (parametric)**

Or

- **Wilcoxon U test (non-parametric)**

- **Conduct the appropriate test on the data provided in the Week One Workshop materials folder on Moodle.**
- **Write up the results as you would for a report.**

Paired Samples Statistics



		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PANAS_PAt1	25.59	369	8.522	.444
	PANAS_PAt2	23.13	369	8.485	.442

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	PANAS_PAt1 & PANAS_PAt2	369	.626	.000

Paired Samples Test

		Paired Differences							
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	
					Lower	Upper			
Pair 1	PANAS_PAt1 - PANAS_PAt2	2.458	7.359	.383	1.704	3.211	6.415	368	Sig. (2-tailed)

Hypothesis 2: There will be a significant difference in positive affect between participants in the control and intervention groups at time 2.

- **Between-groups**

- **Independent T-test (parametric)**

- **Or**

- **Mann Whitney T-test (non-parametric)**

- **Conduct the appropriate test on the data provided in the Week One Workshop materials folder on Moodle.**
- **Write up the results as you would for a report.**

Group Statistics



	group	N	Mean	Std. Deviation	Std. Error Mean
PANAS_PAt2	control	199	23.74	8.547	.606
	intervention	182	22.92	8.520	.632

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
PANAS_PAt2	Equal variances assumed	.006	.940	.942	379	.347	.825	.875	-.896	2.546
	Equal variances not assumed			.942	376.193	.347	.825	.875	-.896	2.546

Linear Regression

- Used to discover effect of one variable (x) on another variable (y).
- Similar to the correlation: **However!**
 - While a correlation allows us to conclude how strongly two variables relate to each other (both in terms of magnitude and direction),
 - Linear Regression also tells us **by HOW MUCH** one variable will change if the other variable changes.
 - It aims to ***predict*** scores on one variable based on scores from another variable.



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

Linear Regression

- Basically – if x changes by a certain amount we will be able to estimate the amount in which y will change.
- Imagine you are working as a marketing Psychologist and your job is to see whether the price of your company's soft drink relates to amount of cans being bought.



Linear Regression

- You would suspect that as the price increases then sales would decrease.
- A simple correlation would be able to show us this – i.e., a negative correlation.
- What linear regression does is show us the amount of the change.
 - E.g., if the price of the drink increases by 50%, sales will fall by 40%.



Linear Regression

- If we know by how much sales decrease, as a result of every penny increase, then we are able to predict sales from price.
 - E.g., for every 2p increase in price, sales decrease by 100 cans.
- Conducting a linear regression analysis to examine this would allow us to suggest that a score on one variable influenced the score on the other variable.
- In this way we can try to infer causal relationships (but only if the variables have been manipulated).



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

Linear Regression

- Proceed in the same way as with a correlation:
 - i.e., Two variables are examined to see if there's a linear relationship between them.
- The variable you want to predict is called the *Criterion* variable (Dependent variable)
 - This is also called y .
- The variable that is predicting the DV is called the *Predictor* variable (or explanatory) variable.
 - This is also called x
- **With only 1 predictor variable & 1 criterion variable you would perform a bivariate regression.**



Important terminology!

- ***b*** = The amount by which *y* changes, when *x* changes by 1 unit – this is represented by the slope or tilt of our straight line in a scatterplot.
 - Also known as the regression coefficient (regression weight)
 - The larger the value of *b*, the steeper the slope.
 - The steeper the slope the more *y* changes as a result of *x*.
 - Using a straight line allows us to predict a persons score on *y* from *x*.
 - SPSS will draw the line in the right place for you.



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

Important Terminology!

- **a** = the intercept (also called the Constant)
 - It is the value of Y' when X is 0.
 - The symbol Y' (pronounced y prime) is used in an equation when we want to predict values of Y from our observed values of X .
 - When the regression line is added to the scatterplot, a is the value of Y at the point where the line intercepts the Y -axis.
 - Note: As a default SPSS applies axis scales around values in the dataset so the scatterplot axis may not include a 0. In such cases a cannot be read from the chart – SPSS provides this value for a anyway.



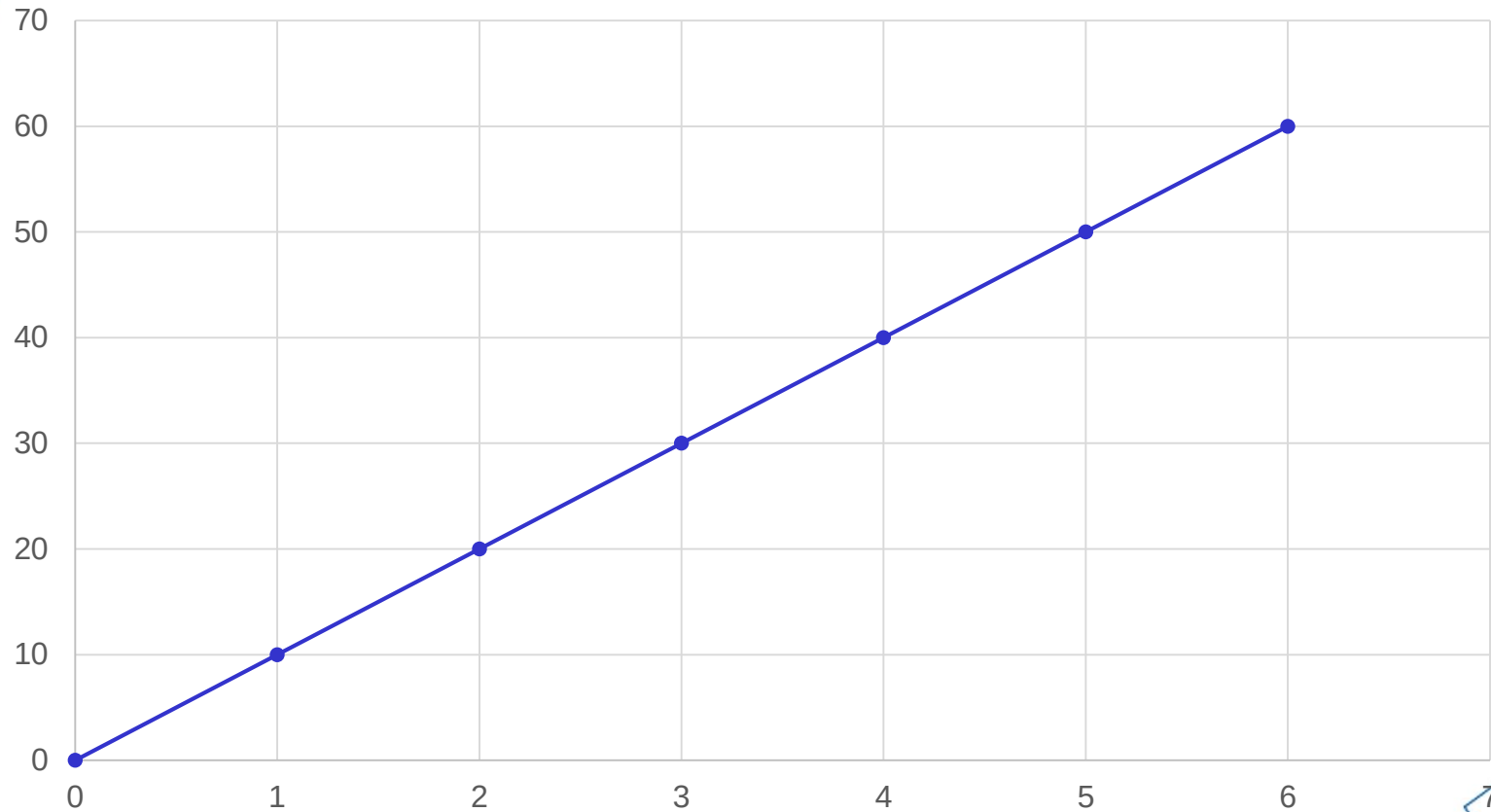
Perfect Relationships - Positive

- A perfect positive correlation is when it is possible for the straight line to pass through every data point.
- E.g., if every time x increased by 1, y increased by 10.



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David

Perfect Relationships – Positive



The point where the line intersects the y axis is called a. Here $a = 0$.

The Regression Equation

- When we perform a linear regression analysis, we obtain a regression equation, which shows the way in which y changes as a result of a change in x .
- From the formula we can calculate someone's score on y from their score on x .
- The general formula is:
 - $y = (b \cdot x) + a$ or $y = a + (b \cdot x)$



The Regression Equation

- Say the slope of a line is 10.
 - This means that, every time x increases by 1 unit, y changes by 10.
- The equation shows that y is able to be predicted by multiplying a person's score on x by b and then adding the constant, which is a .
- So, in our example:
 - $a = 0$ (starting point on the chart, intercept, constant)
 - $b = 10$
 - $y = a + 10 \cdot x$



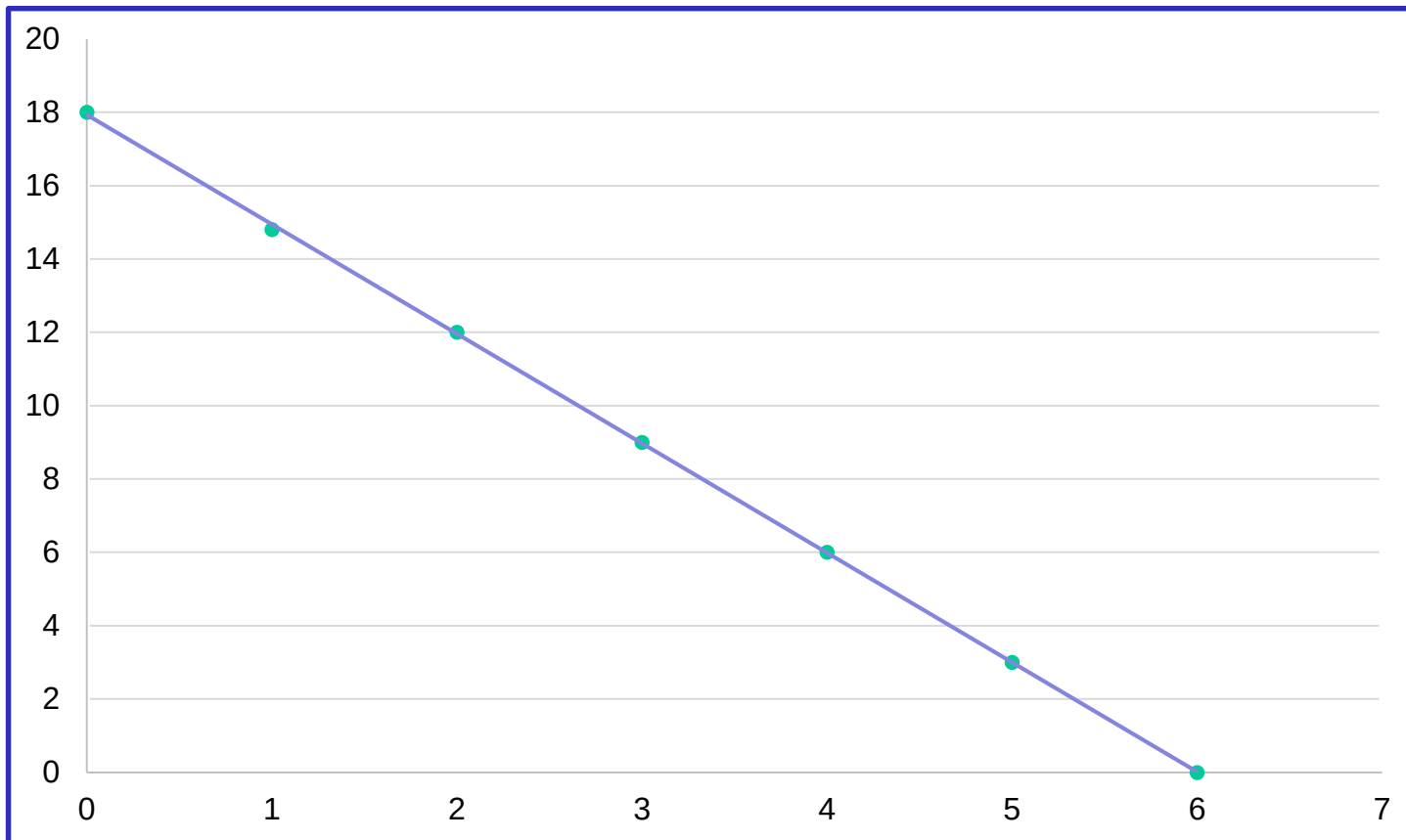
The Regression Equation

- So, if we know that a person has a score of 7 on x, then we are able to predict his or her score on y:
 - $y = a + (b \cdot x)$
 - $y = 0 + (10 \cdot 7) = 70$
- Of course its easy to predict a person's score when all of the points fall on a straight line.
- Activity = What is someone's predicted score on y when their score on x is 20?
 - Assume a is 5 and b is 2.



Perfect Relationships - Negative

- We can also predict negative relationships:
 - $y = a - (b \cdot x)$ The minus sign = a negative relationship
 - Using data from the chart below, what is y if a person watched 3.5 hours of television a night,

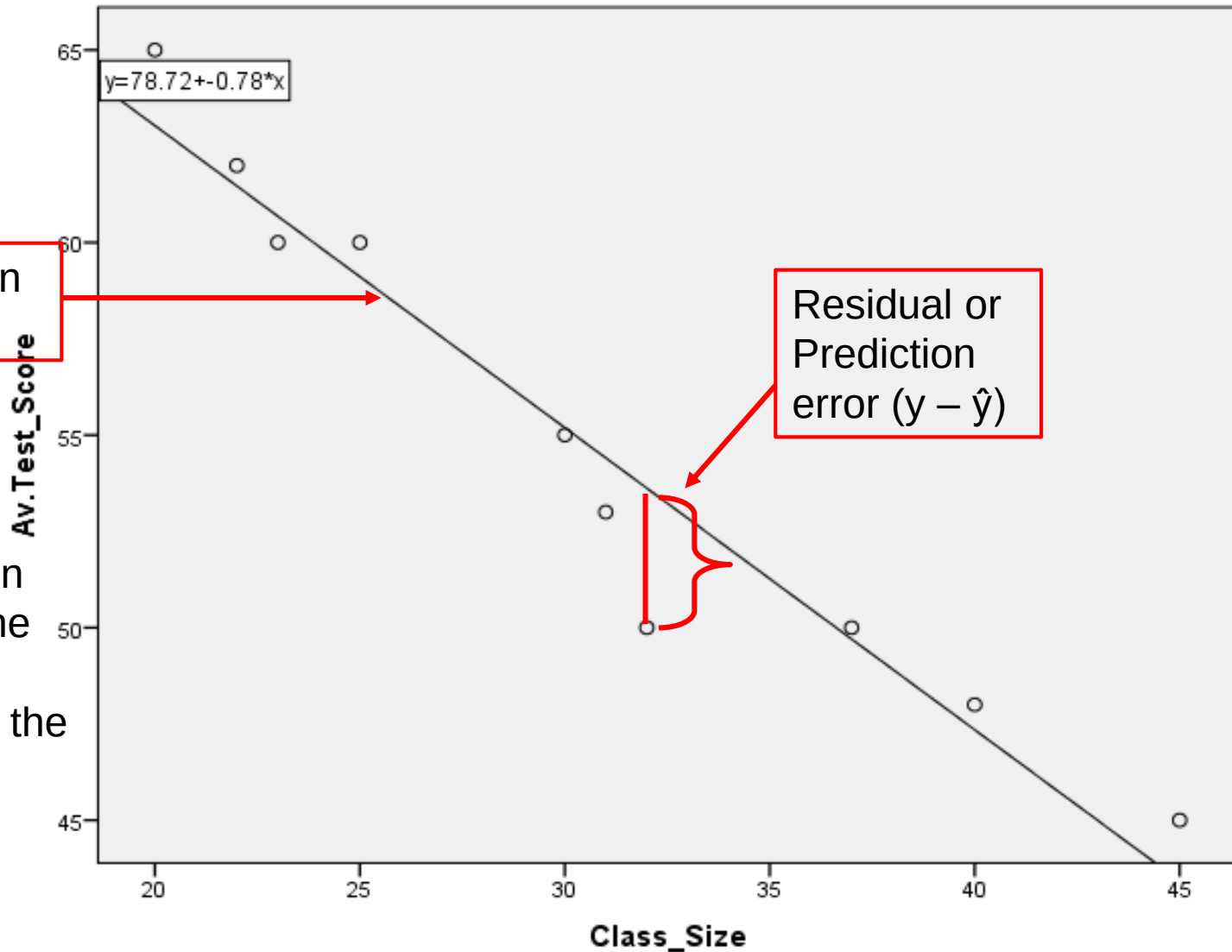


Residuals ($Y - Y'$)

- If we use our line of best fit to predict a person's score, unless we have a perfect relationship we will always make some errors.
- Data points that are off the line represent errors of prediction.
- If we use the line of best fit to predict sales for each price of soft drink for example, then we will obtain a predicted sales figure.
- The difference between the actual scores (Y) and predicted scores (Y') are called residuals.
- They can be positive or negative.



Residuals



Regression
Line

Residual or
Prediction
error ($y - \hat{y}$)

Regression
line is a line
that
minimises the
residuals.



UNIVERSITY OF WALES
Trinity Saint David

Non-perfect relationships

- Its easy to predict from a perfect relationship, but if all your points don't fit exactly on the line then what can you do?
 - You can still get SPSS to draw a line of best fit – that is a line that gives the best fit through all the data points.
 - The way SPSS calculates this is by minimising the vertical distances from the data points and the line.
 - This is called the least-squared regression line.



Linear Regression Worksheet



PRIFYSGOL CYMRU

Y Drindod Dewi Sant

UNIVERSITY OF WALES

Trinity Saint David