

# The t-test

# Data files

- SATpassage.sav
- selfhelpBook.sav

Homework:

- Word\_recall.sav
- LuckyCharm.sav

# Comparing means

- Correlation and regression look for relationships between variables
- Tests comparing means look for differences between the means of two or more samples
- General null hypothesis: the samples come from the same population
- General research hypothesis: the samples come from different populations

# Choosing a statistical test: normality

- There are two sets of statistical tests for comparing means:
  - Parametric tests: work with normally distributed scale data
  - Non-parametric tests: used with not normally distributed scale data and with ordinal data

# Choosing statistical tests: how many means?

- Within parametric and non-parametric tests
  - some work with just two means
  - others work with more than two means

# Choosing statistical tests: research design

- Most statistical tests have at least two versions:
  - one for related (paired, within-subjects, repeated measures) design
  - and one for unrelated (independent, between subjects) designs

# The *t*-test

- Parametric: works on normally distributed scale data
- Compares TWO means
- There are different versions for different designs

# Types of t-test

- One sample:
  - compare the mean of a sample to a predefined value
- Dependent (related) samples:
  - compare the means of two conditions in which the same (or closely matched) participants participated
- Independent (unrelated) samples:
  - compare the means of two groups of participants



# William Gosset again: Student's t-distribution



William Sealy Gosset (1876–1937)



Hired by Guinness to find a cheap way of monitoring stout quality

# **THE ONE-SAMPLE T-TEST**

# The One Sample $t$ test

The One-sample  $t$  test is used to compare a sample mean to a specific value (e.g., a population parameter; a neutral point on a Likert-type scale, chance performance, etc.).

## Examples:

1. A study investigating whether stock brokers differ from the general population on some rating scale where the mean for the general population is known.
2. An observational study to investigate whether scores differ from chance.

Calculation of  $t$ :

$$t = \frac{\text{mean} - \text{comparison value}}{\text{Standard Error}}$$

# Standard Error $SE = \frac{SD}{\sqrt{N}}$

- The standard error (SE) is the standard deviation of the sampling distribution:
  - An estimate of the “error” in the estimation of the population mean
    - We use the sample mean to estimate the population mean
    - This estimate may be more or less accurate
    - If we have a large number of observations and all of these observations are close to the sample mean (large N, small SD), we can be confident that our estimate of the population mean (i.e., that it equals the sample mean) is fairly accurate => small SE
    - If we have a small number of observations and they vary a lot (small N, large SD), our estimate of the population is likely to be quite inaccurate => large SE

# The one-sample t formula

$$t = \frac{M - X}{SE}$$

M: the mean of our sample

the population mean (or any parameter we want to compare our mean to)

SD: the sample Standard Deviation

N: sample size

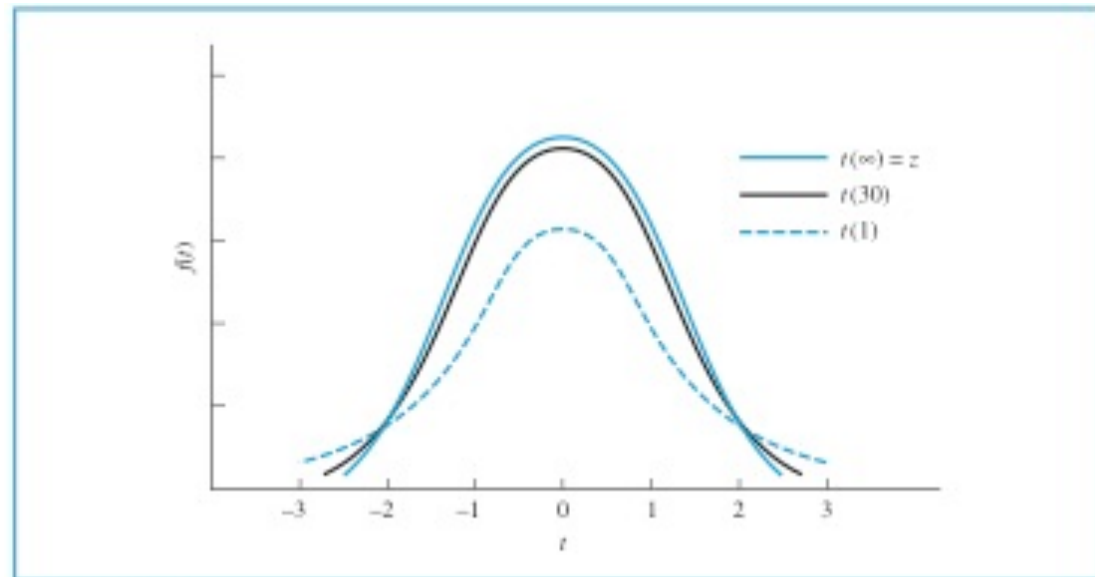
SE: the Standard Error

# Hypothesis testing with $t$

- We can draw a sampling distribution of  $t$ -values (the Student  $t$ -distribution)
  - this shows the likelihood of each  $t$ -value if the null hypothesis is true
  - <http://www.distributome.org/V3/calc/index.html>
- The distribution will be affected by sample size (or more precisely, by degrees of freedom)
- We evaluate the likelihood of obtaining our  $t$ -value given the  $t$ -distribution

Degrees of freedom (df):  
The number of values  
that are free to vary if  
the mean of the total is given

With one sample,  $df = N - 1$



**Figure 12.5**

$t$  distribution for 1, 30, and  $\infty$  degrees of freedom.

# Assumptions

The one-sample  $t$  test requires the following statistical assumptions:

1. Random and Independent sampling.
2. Data are from normally distributed populations.

Note: The one-sample  $t$  test is generally considered robust against violation of this assumption once  $N \geq 30$ .

# An example: Katz et al 1990

- SAT scores without reading the passage (SATpassage.sav)
- Research hypothesis: students do better than chance
- Null hypothesis: students perform at chance level
- Chance: 20 (100 questions with 5 choices each)
- Run the test



# SPSS Output

$t = \text{Diff between means} / \text{Std Error}$

## One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Score	28	46.57	6.828	1.290

The value to which you compare your sample mean

degrees of freedom =  $N-1$

## One-Sample Test

	Test Value = 20					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Score	20.591	27	.000	26.571	23.92	29.22

# Writing up the results

Katz et al. (1990) presented students with exam questions similar to those on the SAT, which required them to answer 100 five-choice multiple-choice questions about a passage that they had presumably read. One of the groups ( $N = 28$ ) was given the questions without being presented with the passage, but they were asked to answer them anyway. A second group was allowed to read the passage, but they are not of interest to us here.

If participants perform purely at random, those in the NoPassage condition would be expected to get 20 items correct just by chance. On the other hand, if participants read the test items carefully, they might be able to reject certain answers as unlikely regardless of what the passage said. A one-sample  $t$  test revealed that participants in the NoPassage condition scored significantly above chance ( $M = 46.57$ ,  $t(27) = 20.59$ ,  $p < .001$ ). That is, it is possible to answer some of the questions correctly without even a single glance at the text to which they refer.

# **PAIRED SAMPLE T-TEST**

# Repeated Measures Designs

- Also called a within-subject or related or paired design
  - Entire experiment is conducted “within” each subject/participant
  - Each individual participates in each condition of the experiment
    - Repeated measurements on dependent variable

# Repeated Measures Designs

- Why Use This Design:
  - Don't have to worry about balancing individual differences across conditions of the experiment (because all participants are in each condition)
  - Require fewer participants
  - Convenient and efficient
  - More *sensitive*

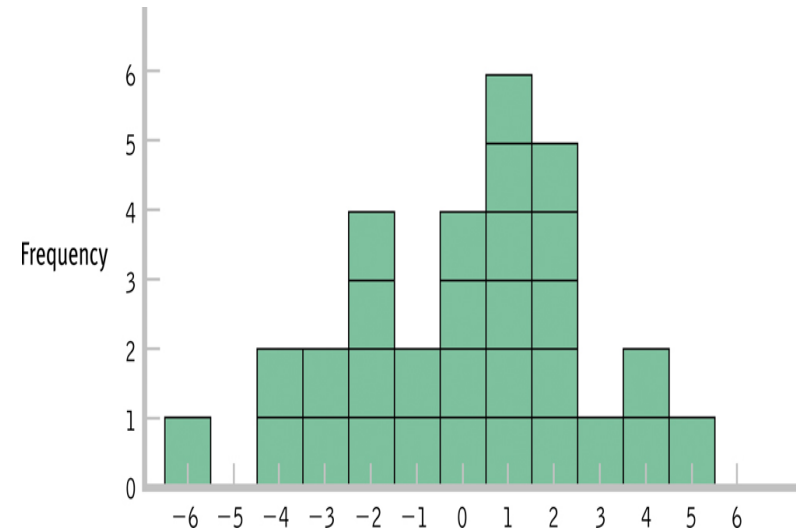
# Paired-sample t test

$$t = \frac{\text{mean difference} - \text{comparison value}}{\text{Standard Error}}$$

- Mean difference: mean difference between scores in the two conditions
- Comparison value: 0, because we expect no difference if the IV manipulation has no effect
- Standard Error: standard error **of the differences**
  - i.e.: estimate of accuracy of the mean difference measured in the sample when it is used to estimate the mean difference in the population

# “Sampling distribution” of mean differences

- Central Limit Theorem Revisited. We can plot the mean difference between two scores of a random sample. The plot will approach a normal distribution. Its standard deviation will be the SS of the deviation of each difference score from the mean difference divided by  $N-1$ .



A distribution of mean differences between scores.

# Standard Error of mean differences

- If mean differences can have a distribution
  - The distribution will have a Standard Error

$$SD_D = \sqrt{\frac{\Sigma(D - M_D)^2}{N - 1}} = \sqrt{\frac{SS_D}{N - 1}}$$

$$SE_D = \frac{SD_D}{\sqrt{N}}$$



# Calculating One-Sample $t$ and Paired-Sample $t$ Statistics

## Single Sample $t$ Statistic

$$SD = \sqrt{\frac{\Sigma(X - M)^2}{N - 1}} = \sqrt{\frac{SS}{N - 1}}$$

Standard Deviation of a Sample

$$SE = \frac{SD}{\sqrt{N}}$$

Standard Error of a Sample

$$t = \frac{(M - \mu)}{SE}$$

$t$  Statistic for Single-Sample  $t$  Test

## Paired Sample $t$ Statistic

$$SD_D = \sqrt{\frac{\Sigma(D - M_D)^2}{N - 1}} = \sqrt{\frac{SS_D}{N - 1}}$$

Standard Deviation of Sample Differences

$$SE_D = \frac{SD_D}{\sqrt{N}}$$

Standard Error of Sample Differences

$$t = \frac{(M_D - \mu_D)}{SE_D}$$

$T$  Statistic for Paired-Sample  $t$  Test I (mean difference divided by SE)

# Paired Sample $t$ Test Example

- We want to know if there is a difference in the salary for the same job in Boise, ID, and LA, CA. The salary of 6 employees in the 25<sup>th</sup> percentile in the two cities is given.



Profession	Boise	Los Angeles
Executive Chef	53,047	62,490
Genetics Counselor	49,958	58,850
Grants Writer	41,974	49,445
Librarian	44,366	52,263
School teacher	40,470	47,674
Social Worker	36,963	43,542

# Paired Sample $t$ Test Example

- We need to know if there is a difference in the salary for the same job in Boise, ID, and LA, CA.
- **Step 1: Define Pops. Distribution and Comparison**  
**Distribution and Assumptions**
  - Pop. 1. Jobs in Boise
  - Pop. 2.. Jobs in LA
  - Comparison distribution will be a distribution of mean differences, it will be a paired-samples test because every job sampled contributes two scores, one in each condition.
  - Assumptions: the dependent variable is scale, we do not know if the distribution is normal, we must proceed with caution; the jobs are not randomly selected, so we must proceed with caution

# Paired Sample $t$ Test Example

- **Step 2: Determine the Characteristics of Comparison Distribution (mean difference, standard deviation of difference, standard error)**
- **M difference = 7914.333      Sum of Squares (SS) = 5,777,187.333**

$$SD_D = \sqrt{\frac{SS_D}{N-1}} = \sqrt{\frac{5,777,186.333}{5}} = 1074.93 \quad SE_D = \frac{SD_D}{\sqrt{N}} = \frac{1074.93}{\sqrt{6}} = 438.83$$

<b>Profession</b>	<b>Boise</b>	<b>Los Angeles</b>	<b>X-Y</b>	<b>D (X-Y)-M</b> <b>M = 7914.33</b>	<b>D<sup>2</sup></b>
Executive Chef	53,047	62,490	-9,443	-1,528.67	<b>2,336,821.78</b>
Genetic Counselor	49,958	58,850	-8,892	-977.67	<b>955,832.11</b>
Grants Writer	41,974	49,445	-7,471	443.33	<b>196,544.44</b>
Librarian	44,366	52,263	-7,897	17.33	<b>300.44</b>
School teacher	40,470	47,674	-7,204	710.33	<b>504,573.44</b>
Social Worker	36,963	43,542	-6,579	1,335.33	<b>1,783,115.11</b>

# Paired Sample $t$ Test Example

- **Step 3: Determine Critical Cutoff**
- **$df = N-1 = 6-1= 5$**
- **$t$  statistic for 5 df ,  $p < .05$ , two-tailed, are -2.571 and 2.571**
- **Step 5: Calculate  $t$  Statistic**

$$t = \frac{(M_D - \mu_D)}{SE_D} = \frac{(7914.333 - 0)}{438.333} = 18.04$$

- **Step 6 Decide**

# Effect sizes

## Cohen's d

$$d = \frac{M_2 - M_1}{SD_{\text{of differences scores}}} = \frac{t}{\sqrt{N}}$$

SD: SD of the distribution of difference scores

(There is quite a bit of disagreement between authors as to the correct formula. It's safer to use  $r^2$ .)

## r (coefficient of determination)

$$r^2 = \frac{t^2}{t^2 + df}$$

Can be misleading because it's influenced by sample size.

But it's easy to interpret because the values will be between 0 and 1.

# Exercise: selfhelpBook.sav

A researcher is interested in assessing the effectiveness of a self-help book (“Men are from Mars, Women are from Venus”) designed to increase relationship happiness. 500 participants read both the self-help book and a neutral book (statistics book). Relationship happiness is measured after reading each book. The order in which the books are read is counterbalanced and there is a six month delay between the two. (selfhelpBook.sav)

# Running paired-samples t-test on SPSS

**Paired Samples Statistics**

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 Self-help book	20.0180	500	9.98123	.44637
Stats book	18.4900	500	8.99153	.40211

**Paired Samples Correlations**

	N	Correlation	Sig.
Pair 1 Self-help book & Stats book	500	.117	.009

**Paired Samples Test**

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 Self-help book - Stats book	1.52800	12.62807	.56474	.41843	2.63757	2.706	499	.007

$t = \text{Mean difference} / \text{SE of mean difference}$



# Reporting the results

The relationship happiness of 500 participants was measured after reading *Men are from Mars, Women are from Venus* and after reading a statistics book. On average, the reported relationship happiness after reading the self-help book ( $M = 20.2$ ,  $SD = 9.98$ ) was significantly higher than in the control condition ( $M = 18.49$ ,  $SD = 8.99$ ),  $t(499) = 2.71$ ,  $p = .007$ . However, the small effect size estimate ( $r^2 = .01$ ) indicates that this difference was not of practical significance.

You could report SE or confidence intervals instead of SD

# The Independent Groups *t* test: Between-subjects designs

Participants contributing to the two means come from different groups; therefore, each person contributes only one score to the data.

$$t = \frac{(M_2 - M_1) - \textit{comparisonvalue}}{SE}$$

M2 – M1: the difference between the means of the two groups

Comparison value: the expected difference between the two means,  
normally 0 under the null hypothesis

SE: standard error of the distribution of differences between means

# Independent $t$ Test

## Paired-Sample

- Two observations from each participant
- The second observation is **dependent** upon the first since they come from the same person.
- Comparing the **mean of differences** to a distribution of mean difference scores

## Independent $t$ Test

- Single observation from each participant from **two independent groups**
- The observation from the second group is **independent** from the first since they come from different subjects.
- Comparing the **difference between two means** to a distribution of differences between mean scores.

# Sampling distribution of differences between means

- Take two samples from a population, calculate the difference between their means
- Take another two samples, calculate the difference between their means
- Do this a few hundred times and then plot the frequency distribution of the differences: you get a sampling distribution of differences between means

# Standard Error: Applied to Differences

We can extend the concept of standard error to situations in which we're examining differences between means.

The standard error of the differences estimates the extent to which we'd expect sample means to differ by chance alone--it is a measure of the unsystematic variance, or variance not caused by the experiment.

$$SE_{DM} = \sqrt{\left( \frac{SD_1^2}{N_1} + \frac{SD_2^2}{N_2} \right)}$$

# The formula

- When sample sizes are equal:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{SD_1^2}{N_1} + \frac{SD_2^2}{N_2}}}$$

- When sample sizes are unequal:

$$SD_p^2 = \frac{(N_1 - 1)SD_1^2 + (N_2 - 1)SD_2^2}{N_1 + N_2 - 2}$$

$$t = \frac{M_1 - M_2}{\sqrt{\frac{SD_p^2}{N_1} + \frac{SD_p^2}{N_2}}}$$

# Equal variances

- Because SE for the independent t-test is calculated from the variances of the two groups, the t-distribution is distorted when the two variances are very different.
- Testing the equality (homogeneity) of variances: Levene's test (SPSS automatically calculates it)
  - Equal variances assumed: when the variances of the two sets of scores are approximately equal
  - Equal variances not assumed: when the two variances are quite different – a correction is applied to the df

- Degrees of freedom:  $N_1 - 1 + N_2 - 1$   
(the degrees of freedom of the two groups added together)



# Effect sizes

<http://www.socscistatistics.com/effectsize/Default3.aspx>

## Cohen's d

$$d = \frac{M_2 - M_1}{SD_{pooled}}$$

$$SD = \sqrt{\frac{(N_1 - 1)SD_1^2 + (N_2 - 1)SD_2^2}{N_1 + N_2 - 2}}$$

There's some disagreement over the formula here.

## $r^2$

$$r^2 = \frac{t^2}{t^2 + df}$$

Can be misleading because it's influenced by sample size.

But it's easy to interpret because the values will be between 0 and 1.

# Example: SATpassage.sav

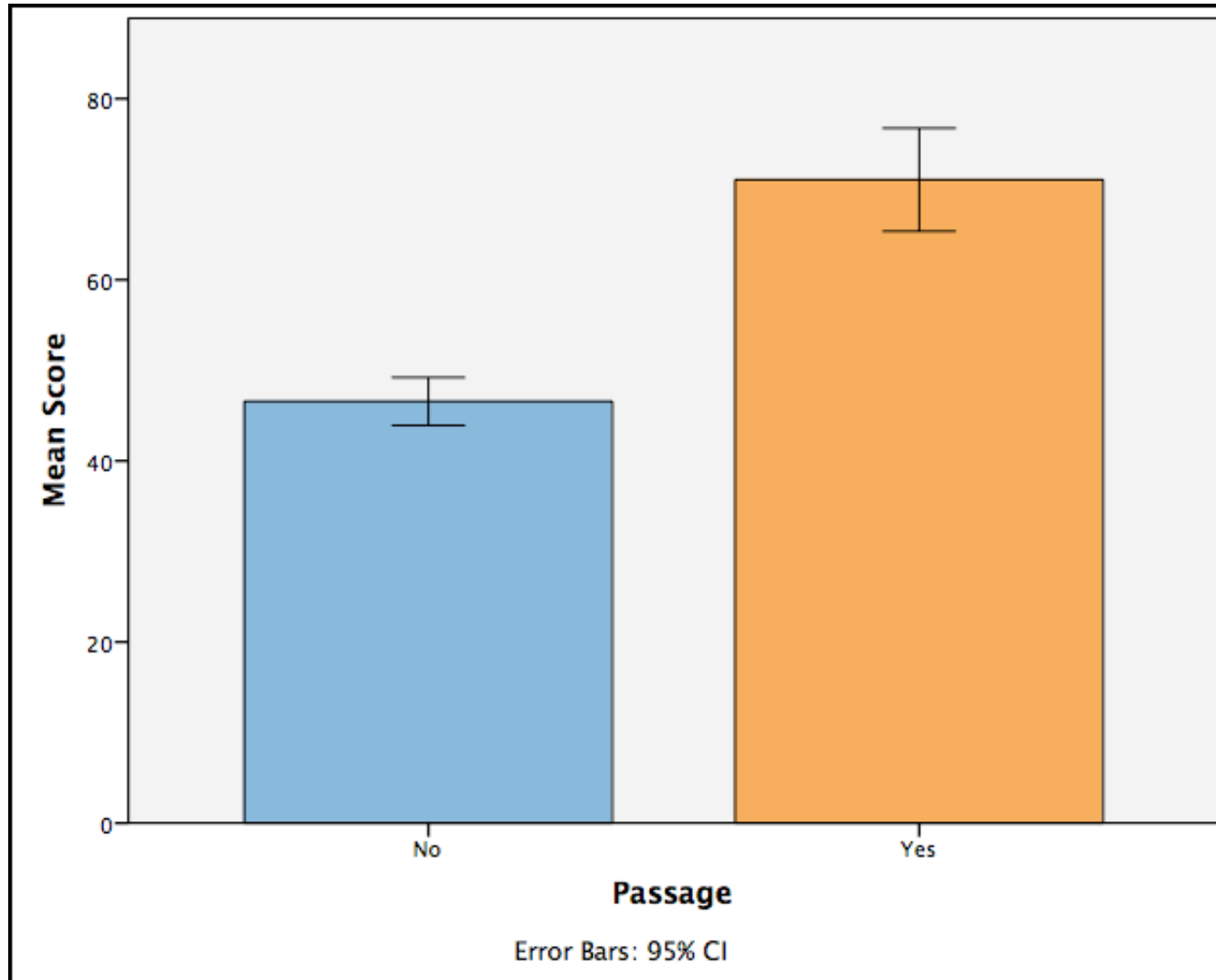
The SAT performance of the group of students who did not read the passage was compared to the performance of a group of students who did read the passage.

Look at distributions.

Outliers?

Draw graph.

# Plotting the results



# SPSS Output: Independent-Groups *t* test

Group Statistics					
	Passage	N	Mean	Std. Deviation	Std. Error Mean
Score	No	28	46.57	6.828	1.290
	Yes	17	71.06	11.065	2.684

Look at Levene's test first:

If significant ( $p < .05$ ), the variances of the two samples are not equal and a correction must be applied.

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
Score	Equal variances assumed	2.291	.137	-9.207	43	.000	-24.487	2.660	-29.851	-19.123
	Equal variances not assumed			-8.223	23.509	.000	-24.487	2.978	-30.640	-18.335

What would it mean if the confidence interval included 0?

# Reporting the Results

The SAT performance of the group of students who did not read the passage ( $N = 28$ ) was compared to the performance of a group of students who did read the passage ( $N = 17$ ). An independent samples  $t$  test revealed that the students who read the passage had significantly higher scores ( $M = 71.06$ ,  $SD = 11.06$ ) than the No-passage group ( $M = 46.57$ ,  $SD = 6.83$ ), ( $t(43) = -9.21$ ,  $p < .001$ ). While people may show better than chance performance without reading the passage, their scores will not approximate the scores of those who read the passage. This effect was very large (Cohen's  $d = 2.66$ ).

You could report SE or confidence intervals instead of SD

# A note on problems

- If the data
  - Are not quite normally distributed or
  - Have outliers
- The simplest solution is to use bootstrapped confidence intervals
  - Look at the confidence intervals of the difference between the groups/conditions (this should not include 0) and
  - The confidence intervals of the means of the two groups/conditions (these should not overlap)

# Homework 1: Word\_recall.sav

- Participants recalled words presented with or without pictures. The researchers used counterbalancing: each participant saw words both with and without pictures.
- Run descriptives, check normality, check for outliers
- Draw graph
- Run appropriate t-test
- Calculate effect size
- Write up the results

# Homework 2: LuckyCharm.sav

- Damish et al (2010): Does a lucky charm work?
- Participants bring lucky charm and play a memory game: picture cards are turned face down, participant has to find the pairs turning one card at a time.
  - One group has the lucky charm with them when doing the test.
  - Other group has to leave the charm in another room.
- Dependent variable: how fast participants finished relative to a baseline (low score = better performance)
- Hypothesis? Significance level? Tails? Descriptives, outliers, graphs
- Run appropriate t-test
- Calculate effect size
- Write it up