# Psychological Statistics

## The t-test



## What We Will Cover in This Section

- Introduction
- One-sample t-test.
- Independent samples t-test.
- Dependent samples t-test.
- Power and effect size.
- Key learning points.

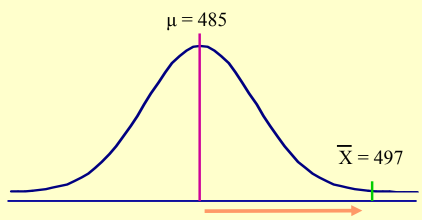


## A Research Situation

A high school wants to know if a special SAT preparation program has helped students raise their scores. They got scores of a group of 25 students. Historically the mean verbal score for all of their graduating seniors is $\mu = 485$, but they don't have the standard deviation. The sample has a mean SAT score of 497 with a standard deviation of 10.

1. What is the research hypothesis?
2. What is $H_o$?
3. What is the statistical hypothesis?
4. Is this a one-tailed or two-tailed test?

## DTP






## z-test and the Single Sample t-test

| Known statistics | z-test | Single sample t-test |
|---|---|---|
| μ | Yes | Yes |
| σ | Yes | **No** |
| X̄ | Yes | Yes |
| s | Yes* | Yes |
| N | Yes | Yes |



## Comparing the Formulas

| | z- test | Single Sample t-test |
|---|---|---|
| Standard Error of the Mean | $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}$ | $\sigma_{\overline{X}} = \frac{s}{\sqrt{N}}$ |
| Test of the Difference Between Means | $Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$ | $t_{(N-1)} = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$ |

## Degrees of Freedom (df)

- Developed from the notion that when you know that a group of N numbers sum to S, and if you know N-1 of the numbers, the Nth number is fixed.
- Example.
  If a group of 4 numbers add up to 15 and three of the numbers are 5, 6, and 2, what is the fourth number?

  In this case you have N-1 degrees of freedom.



## Back to the Example: Computation

A high school wants to know if a special SAT preparation program has program has helped students raise their scores. They got scores from a group of 25 participants. Historically the mean verbal score for all of their graduating seniors is $\mu$ = 485, but they don't have the standard deviation. The sample has a mean SAT score of 497 with a standard deviation of 10.

$$\sigma_{\overline{X}} = \frac{10}{\sqrt{25}}$$

$$\sigma_{\overline{X}} = 2$$

$$t_{(N-1)} = \frac{497 - 485}{2}$$

$$t_{(24)} = 6.00$$



α, 1 tail

α, 2 tail

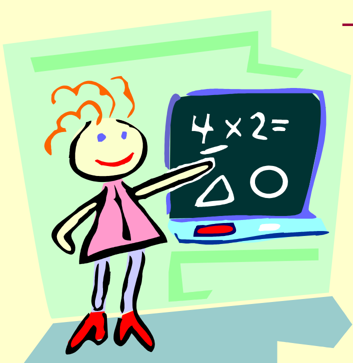
| | PROPORTION IN ONE TAIL | | | | | |
|---|---|---|---|---|---|---|
| | 0.25 | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 |
| | PROPORTION IN TWO TAILS | | | | | |
| *df* | 0.50 | 0.20 | 0.10 | 0.05 | 0.02 | 0.01 |
| 1 | 1.000 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 |
| 2 | 0.816 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 |
| 3 | 0.765 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 |
| 4 | 0.741 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 |
| 5 | 0.727 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 |
| 6 | 0.718 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 |
| 7 | 0.711 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 |
| 8 | 0.706 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 |
| 9 | 0.703 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 |
| 10 | 0.700 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 |
| 11 | 0.697 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 |
| 12 | 0.695 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 |
| 13 | 0.694 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 |
| 14 | 0.692 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 |
| 15 | 0.691 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 |
| 16 | 0.690 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 |
| 17 | 0.689 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 |
| 18 | 0.688 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 |
| 19 | 0.688 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 |
| 20 | 0.687 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 |
| 21 | 0.686 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 |
| 22 | 0.686 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 |
| 23 | 0.685 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 |
| 24 | 0.685 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 |
| 25 | 0.684 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 |
| 26 | 0.684 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 |
| 27 | 0.684 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 |
| 28 | 0.683 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 |
| 29 | 0.683 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 |
| 30 | 0.683 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 |
| 40 | 0.681 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 |
| 60 | 0.679 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 |
| 120 | 0.677 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 |
| ∞ | 0.674 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 |

Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research,*

## Assumptions of Single Sample t-test

1. The population mean is available.
2. The population distribution is normal.
3. The observations are *independent*.
4. Measurement is done on an interval or ratio scale.
5. You have the sample
   - Mean
   - Standard Deviation (s)
   - Sample Size (N)






## Independent Sample t-test



## Independent Sample t-Test

The grades for a group of CUP soccer players tend to be somewhat below average. This might be a result of bouncing balls off their heads. To deal with this, twenty first-year players are equipped with helmets. A control group of 20 players play without helmets. At the end of the school year their grades are compared.

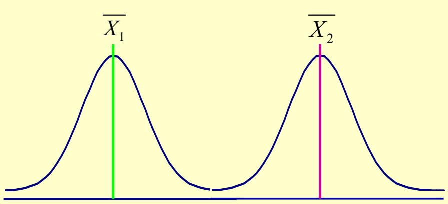
1. What is the research hypothesis?
2. What is $H_o$?
3. What is the statistical hypothesis?
4. Is this a one-tailed or two-tailed test?

## Single Sample vs Independent Sample t-test

| Known statistics | Single Sample t | Independent Sample t |
|---|---|---|
| μ | Yes | |
| σ | No | |
| $M_1$ | Yes | |
| $s_1$ | Yes | |
| N | Yes | |
| $M_2$ | | |
| $S_2$ | | |
| N | | |



## First: DTP






## Model

$$t_{(df)} = \frac{\bar{X}_1 - \bar{X}_2}{\text{standard error}}$$

## Problem!

$$\frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}}}$$

**How do you get the Standard Error?** →



## Second: Calculate the Average Variance

You calculate the sum of the two sample variances.

$$\frac{SS_1}{df_1} + \frac{SS_2}{df_2}$$



## Pooled Error Variance ($s^2_p$)

- Represents the mean variance estimate from the two samples.
- Weighted by the sample size.
- Two estimates mean *df* = ($N_1$+$N_2$-2).

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

## Second: Standard Error of Difference Between Two Means

- Represents the *Standard Error* when sampling the difference between two means.

$$S_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{S_P^2}{N_1} + \frac{S_P^2}{N_2}}$$



## Soccer Study Results, Step 1

| Helmet | No-Helmet |
|---|---|
| M = 2.50 | M = 1.90 |
| SS = 4.75 | SS = 6.84 |
| N = 20 | N = 20 |

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

$$s_p^2 = \frac{4.75+6.84}{19+19}$$

$$s_p^2 = \frac{11.59}{38}$$

$$s_p^2 = .305$$



## Soccer Study Results, Step 2

Compute the Standard Error of the Difference

$$S_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{S_P^2}{N_1} + \frac{S_P^2}{N_2}}$$

$$S_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{.305}{20} + \frac{.305}{20}}$$

$$S_{\overline{X}_1-\overline{X}_2} = \sqrt{.0305}$$

$$S_{\overline{X}_1-\overline{X}_2} = .1746$$

## Soccer Study Results, Step 3

Compute t.

$$t_{(N_1+N_2-2)} = \frac{\overline{X}_1 - \overline{X}_2}{S_{\overline{X}_1-\overline{X}_2}}$$

$$t_{(38)} = \frac{2.50-1.90}{.1746}$$

$$t_{(38)} = \frac{.60}{.1746}$$

$$t_{(38)} = 3.436$$

QUESTION: Is this a statistically significant result?



## Assumptions

1. The observations within each sample are independent.
2. The populations from which the samples are drawn are normally distributed.
3. The populations from which the samples are drawn have equal variances.



## Repeated Measures (Within Groups) t-test

## Issues

- A significant component of the standard error in the independent groups t-test is random error generated by two separate samples.
- This random error masks any treatment effect.
- One way to control for this is to use the same subjects in both treatment conditions.



## Repeated Measures t-test

*A study in which participants are measured more than once on the same dependent variable. The same subjects are used in all treatment conditions.*



## Example

A group of ten freshmen football players were told that they had to bulk up to make the first string squad. They were weighed in September and then put on a strict weight training program with a food supplement designed to increase muscle mass. They were then weighed again in December.

1. What is the research hypothesis?
2. What is $H_o$?
3. What is the statistical hypothesis?
4. Is this a one-tailed or two-tailed test?

| Before | After | Difference |
|---|---|---|
| 182 | 177 | -5 |
| 184 | 186 | 2 |
| 184 | 192 | 8 |
| 181 | 180 | -1 |
| 187 | 187 | 0 |
| 180 | 189 | 9 |
| 179 | 183 | 4 |
| 171 | 182 | 11 |
| 184 | 186 | 2 |
| 180 | 184 | 4 |
| | Mean | 3.4 |

Statistical question.

Could these difference scores have happened by chance?



## Model

$$t_{(N-1)} = \frac{\overline{D}}{S_{\overline{D}}} \qquad S_{\overline{D}} = \sqrt{\frac{S_D^2}{N}}$$



## Weight Gain Study, Step 1

| Before | After | Difference | $(D - M_D)^2$ |
|---|---|---|---|
| 182 | 177 | -5 | 70.56 |
| 184 | 186 | 2 | 1.96 |
| 184 | 192 | 8 | 21.16 |
| 181 | 180 | -1 | 19.36 |
| 187 | 187 | 0 | 11.56 |
| 180 | 189 | 9 | 31.36 |
| 179 | 183 | 4 | .36 |
| 171 | 182 | 11 | 57.76 |
| 184 | 186 | 2 | 1.96 |
| 180 | 184 | 4 | .36 |
| | | | $SS_D$ = 216.40 |

## Weight Gain Study, Step 2

Compute the Standard Error of D.

$$S_{\overline{D}} = \sqrt{\frac{S_D^2}{N}}$$

$$S_{\overline{D}} = \sqrt{\frac{21.64}{10}}$$

$$S_{\overline{D}} = \sqrt{2.164}$$

$$S_{\overline{D}} = 1.471$$



## Weight Gain Study, Step 3

Compute t.

$$t_{(N-1)} = \frac{\overline{D}}{S_{\overline{D}}}$$

$$t_{(9)} = \frac{3.44}{1.471}$$

$$t_{(9)} = 2.34$$

QUESTION: What is your statistical conclusion?



## Assumptions

1. Observations within each treatment condition are independent.
2. The population distribution of the difference scores is normal.

## Issues with Repeated Measures Designs

1. Carryover effect.
2. Practice.
3. Fatigue.



## Power and Effect Size



## Power

- Can the test detect a treatment difference when the difference exists?
- POWER is the probability that the test will correctly reject a false null hypothesis.
- A weak statistical test will raise the probability of making a Type II error

## Things That Influence Power

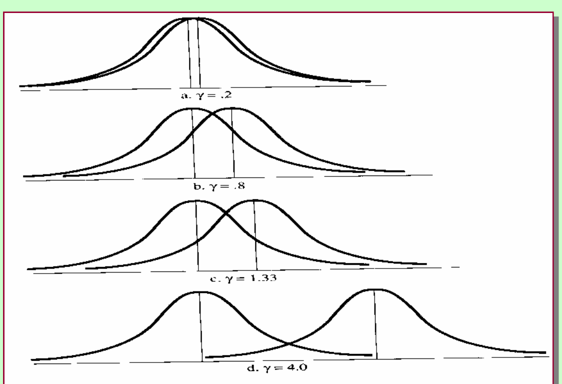
1. Alpha level.
2. One vs. two-tailed test.
3. Sample size.



## Effect Size

- The magnitude or influence of the independent variable on the dependent variable.





## Power and Effect Size

1. A powerful (*sensitive*) statistical test will detect a weak effect.
2. A weak test will fail to detect a small effect (Type II error).



## Statistical vs. Practical Significance

- Large sample sizes increase the power of a test, make it more sensitive.
- Powerful tests detect relatively small differences.



## eta$^2$

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

Interpreted in terms of the amount of variability accounted for in the dependent variable when one knows the level of the independent variable.

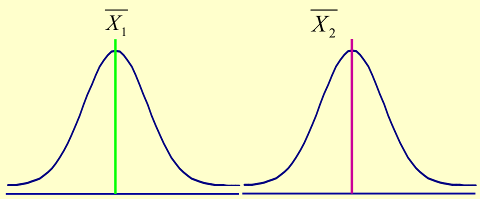
**Soccer Study**

$$eta^2 = \frac{3.46^2}{3.46^2 + 18}$$

$$eta^2 = \frac{11.9716}{29.9716}$$

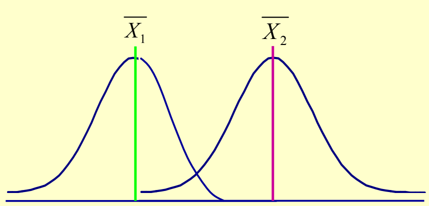
$$eta^2 = .40$$

## Strong Effect






## Weaker Effect






# THE END