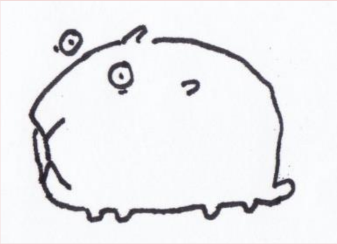
Effect size

Is the difference that I measure important/practically relevant?

1-sample t-test



Statistical & practical significance

- Statistical significance:

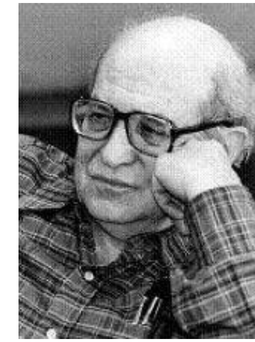
Is the difference between sample average and the reference value convincing? I.e., is the difference (much) larger than the error in this difference?

Significant if: $\left| t = \frac{\bar{y} - \mu}{s_{\bar{y}}} \right| > t_{\text{crit.2}}$

- Practical significance:

Is the difference between sample average and the reference value important? I.e., is the difference (much) larger than the noise in the data? Practical significance = effect size: Cohen's $d_s = \frac{\bar{y} - \mu}{s}$

Effect size: Cohen's d


Jacob Cohen

The effect size for a t-test is usually expressed as the value of Cohen's d:

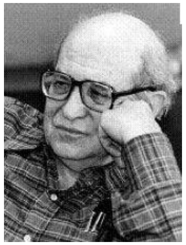
$$d_s = \frac{\bar{y} - \mu}{s}$$

- Cohen's d measures the difference $\bar{y} - \mu$ in "units" of the noise = standard deviation s
- Cohen's d is dimensionless, $\pm$ independent from n!
- What is a "large" difference, what is "small"? Depends on the context, but general guidelines*:

Value Cohen's | d_s |	Interpretation
0.2	Small difference
0.5	Medium difference
0.8	Large difference

*) Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Effect size: Cohen's d



Cohen's d: $d_s = \frac{\bar{y}-\mu}{s}$

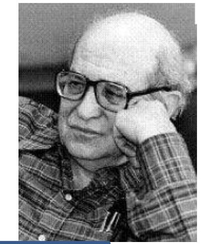
Calculation:

$$d_s = \frac{(\bar{y} - \mu)}{s} = \frac{t}{\sqrt{n}}$$

because

$$t = \frac{(\bar{y} - \mu)}{s/\sqrt{n}} = \frac{(\bar{y} - \mu)}{s} \cdot \sqrt{n} = d_s \cdot \sqrt{n}$$

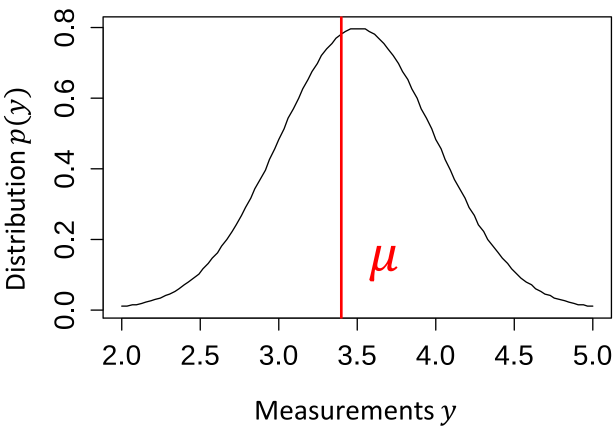
Effect size: Cohen's d

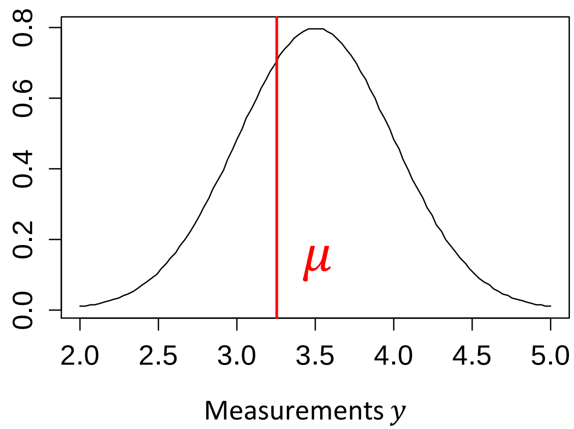


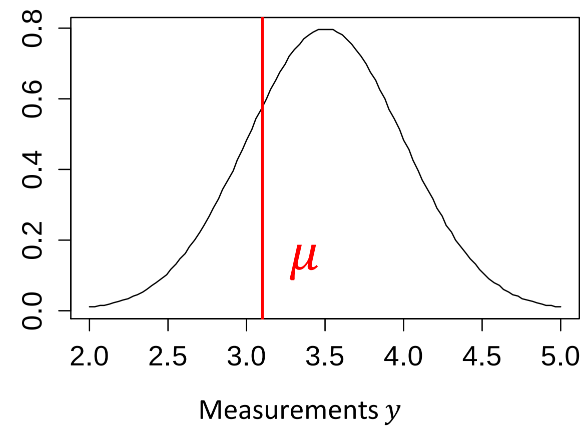
Cohen's d: $d_s = \frac{\bar{y} - \mu}{s}$

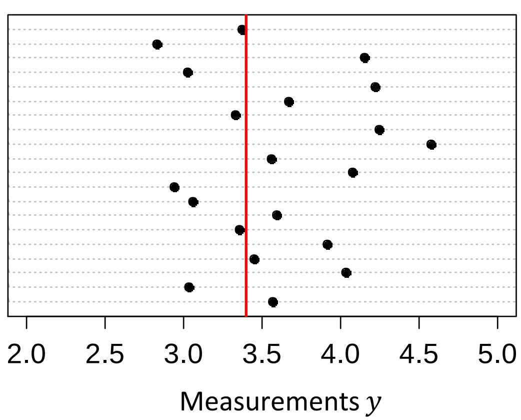
Example: $\bar{y} = 3.50$, $s = 0.50$

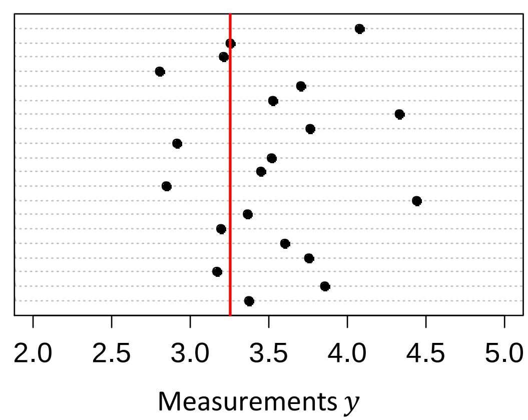
$d_s = 0.20$: small diff.	$d_s = 0.50$: medium diff.	$d_s = 0.80$: large diff.

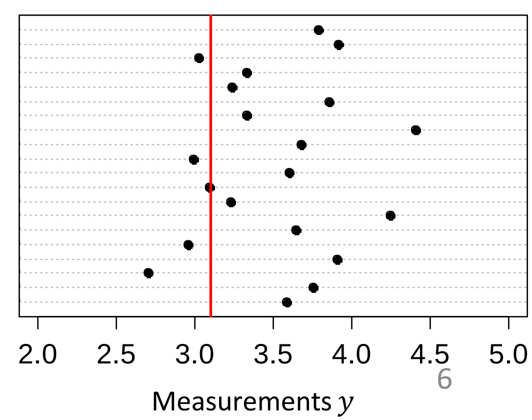




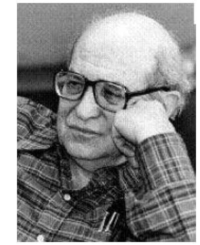






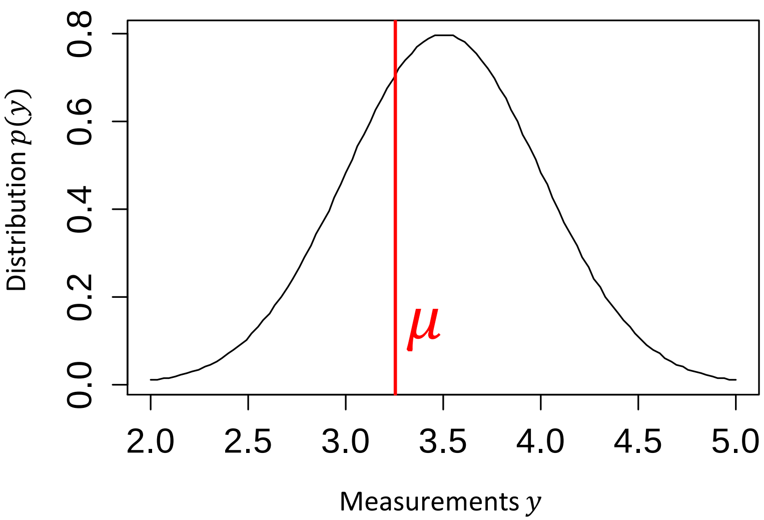


Effectsterkte: Cohen's d

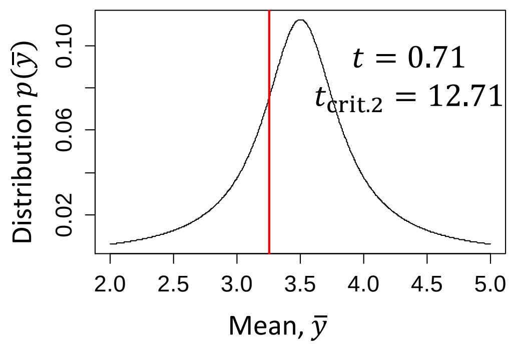


Cohen's d: $d_s = \frac{\bar{y} - \mu}{s}$

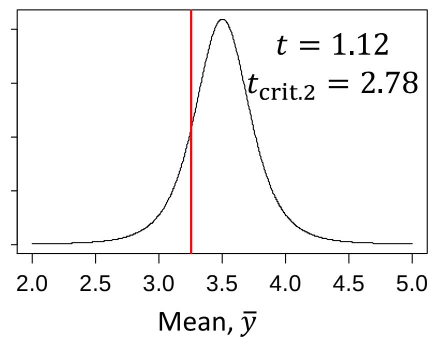
Example: $\bar{y} = 3.50$, $s = 0.50$, $\mu = 3.25$



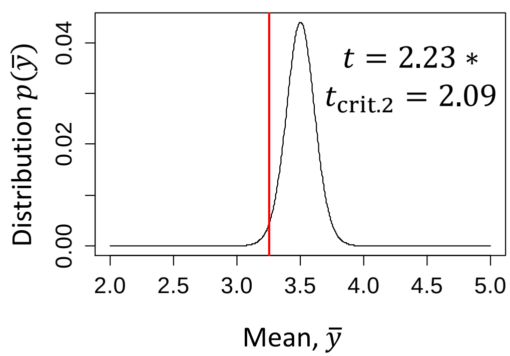
$n = 2$ data



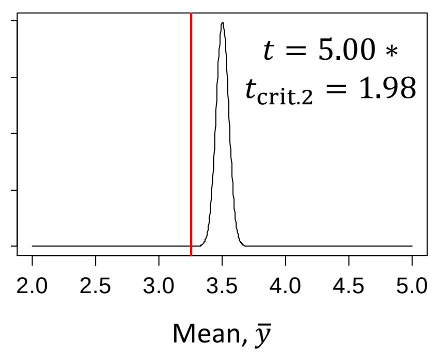
$n = 5$ data



$n = 20$ data



$n = 100$ data



Effect size: $d_s = \frac{3.50 - 3.25}{0.50} = +0.50$, so a medium difference!

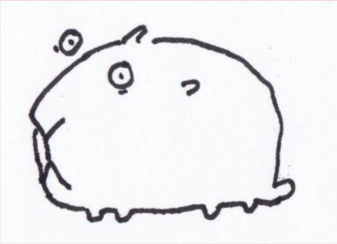
Nr. of measurements n & effect size

- To have a "small" difference significant, you need a large number of measurements…
- To have a "large" difference significant, you need less measurements…
- We have $t = d_s \cdot \sqrt{n}$, so $t_{\text{crit.2}} = d_s \cdot \sqrt{n}$, so

$$\text{tinv}(0.05; n-1) = d_s * \text{sqrt}(n)$$

Value Cohen's $\lvert d_s \rvert$	Interpretation	Minimum n to get significant difference
0.2	Small difference	99
0.5	Medium difference	18
0.8	Large difference	9

Welch t-test



Statistical & practical significance

- Statistical significance:

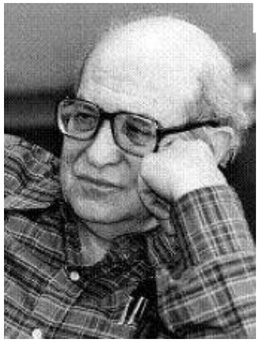
Is the difference between sample average and the reference value convincing? I.e. is the difference (much) larger than the error in this difference?

Significant if: $\left| t = \frac{\bar{y}-\mu}{s_{\bar{y}}} \right| > t_{\mathrm{crit.2}}$

- Practical significance:

Is the difference between sample average and the reference value important? I.e. is the difference (much) larger than the noise in the data? Practical significance = effect size: Cohen's $d_{\mathrm{av}} = \frac{\bar{y}_1-\bar{y}_2}{s_p}$

Effect size: Cohen's d_{av}



Jacob Cohen

The effect size for a t-test is usually expressed as the value of Cohen's d :

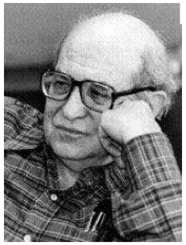
$$d_{\text{av}} = \frac{\bar{y}_1 - \bar{y}_2}{s_p}$$

- Cohen's d measures the difference $\bar{y} - \mu$ in "units" of the noise = average standard deviation s_p
- Cohen's d is dimensionless, $\pm$ independent from n!
- What is a "large" difference, what is "small"? Depends on the context, but general guidelines*:

Value Cohen's | d_{av} |	Interpretation
0.2	Small difference
0.5	Medium difference
0.8	Large difference

*) Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Effect size: Cohen's d_{av}



Cohen's d: $d_{\mathrm{av}} = \frac{\bar{y}_1 - \bar{y}_2}{s_p}$

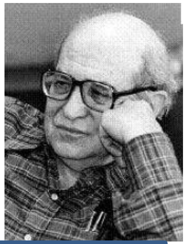
Calculation:

$$d_{\mathrm{av}} = \frac{\bar{y}_1 - \bar{y}_2}{s_p} = t \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

because

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\bar{y}_1 - \bar{y}_2)}{s_p} \cdot \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = d_{\mathrm{av}} \cdot \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

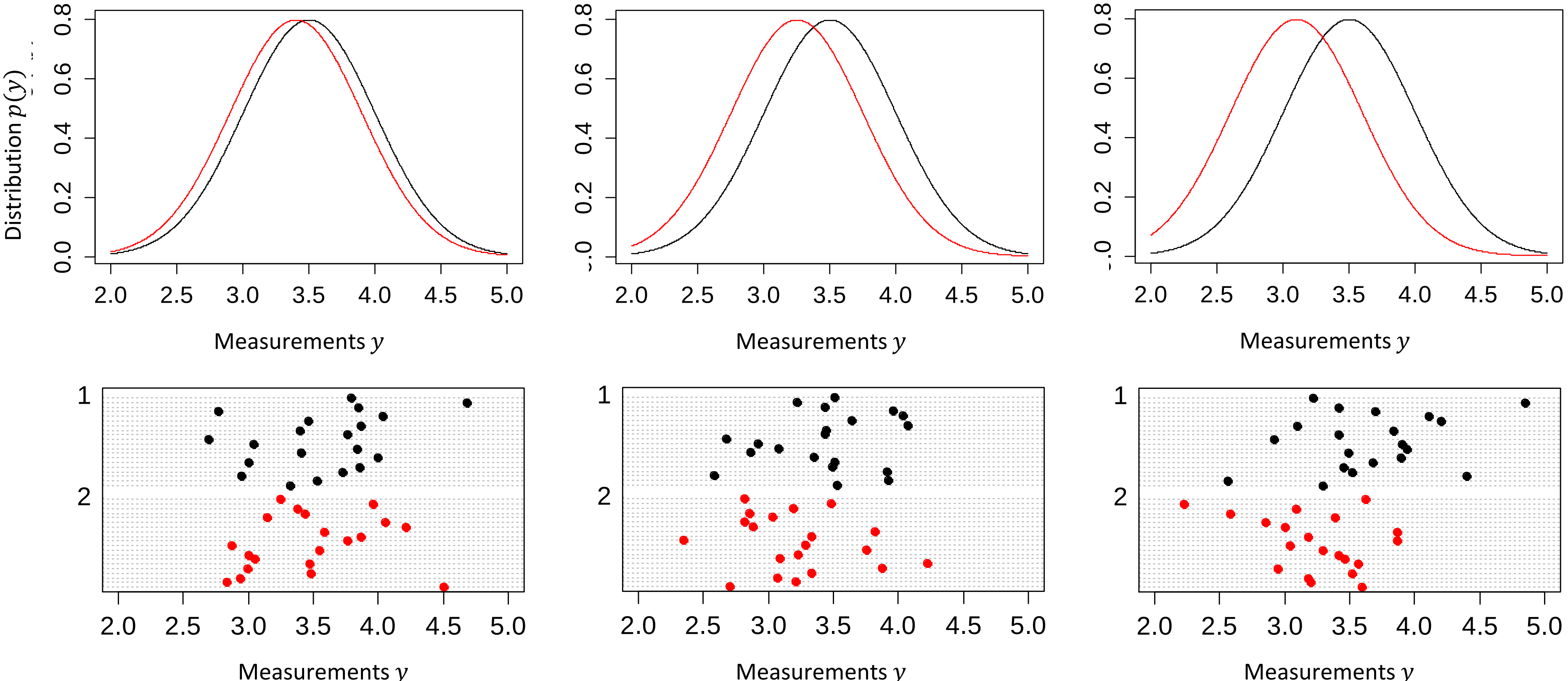
Effect size: Cohen's d_{av}



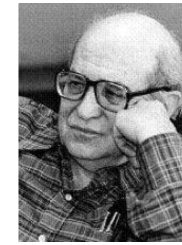
Cohen's d: $d_{\mathrm{av}} = \dfrac{\bar{y}_1 - \bar{y}_2}{s_p}$

Example: $\bar{y}_1 = 3.50$, $s_p = 0.50$

$d_{\mathrm{av}} = 0.20$: small diff. | $d_{\mathrm{av}} = 0.50$: medium diff. | $d_{\mathrm{av}} = 0.80$: large diff.

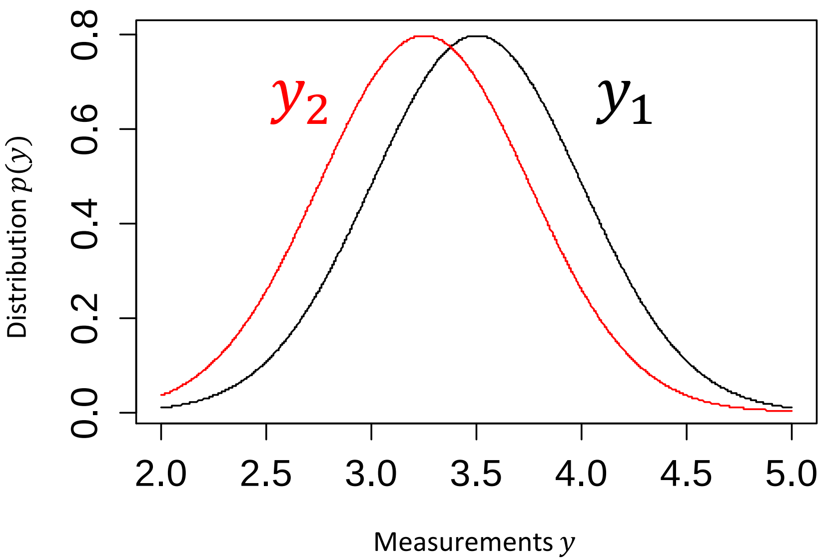


Effect size: Cohen's d_{av}

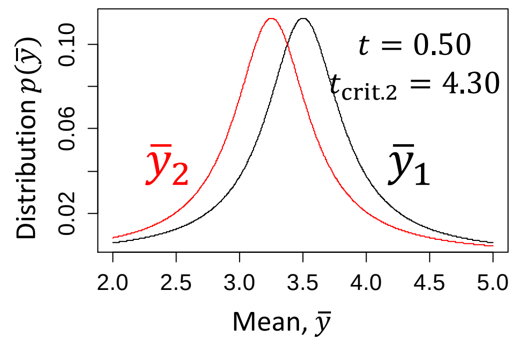


Cohen's d: $d_{\text{av}} = \dfrac{\bar{y}_1 - \bar{y}_2}{s_p}$

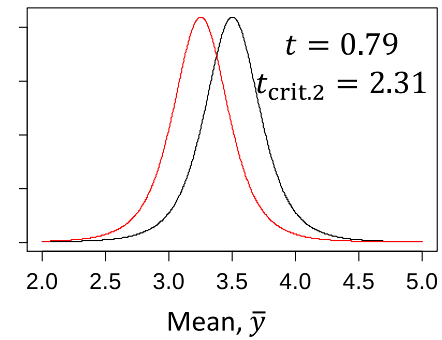
Example: $\bar{y}_1 = 3.50$, $\bar{y}_2 = 3.25$, $s_p = 0.50$



$n = 2$ data*

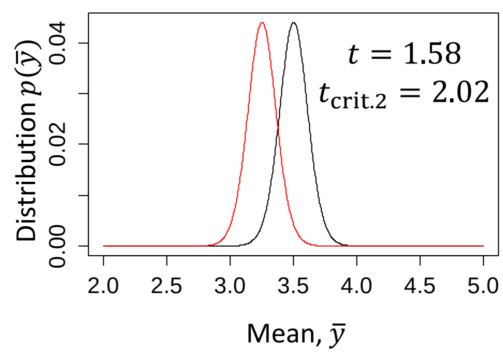


$n = 5$ data

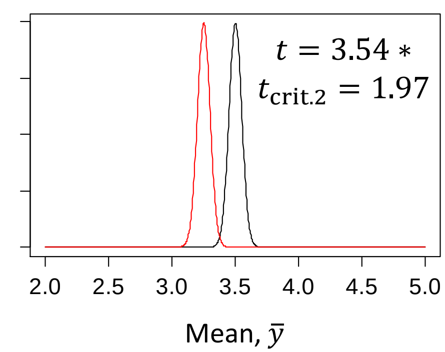


Effect size: $d_{\text{av}} = \dfrac{3.50 - 3.25}{0.50} = +0.50$, so a medium difference!

$n = 20$ data



$n = 100$ data



* Per dataset n, in total $2n$ data points

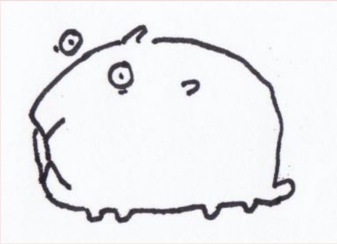
Nr. of measurements n & effect size

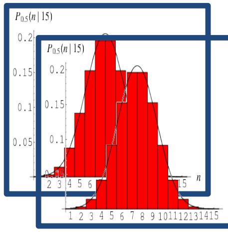
- To have a "small" difference significant, you need a large number of measurements…
- To have a "large" difference significant, you need less measurements…
- We have $t = d_{\mathrm{av}} / \sqrt{\frac{1}{n} + \frac{1}{n}}$, so $t_{\mathrm{crit.2}} = d_{\mathrm{av}} \cdot \sqrt{n/2}$, so

$$\mathrm{tinv}(0.05; 2n-2) = d_{\mathrm{av}} * \mathrm{sqrt}(n/2)$$

Value Cohen's | d_{av} |	Interpretation	Minimum n per set to get significant difference
0.2	Small difference	194
0.5	Medium difference	32
0.8	Large difference	14

1-sample z-test





1-sample z-test: effect size

- Cramér's V^2 (cf. R^2 at regression):

$$V^2 = \frac{\chi^2}{N} = \frac{z^2}{N}$$

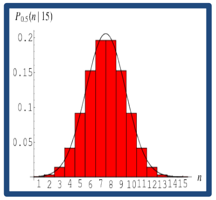
- Interpretation*:

Harald Cramér

Value Cramer's V^2:	Interpretation
0.00	No difference w.r.t. expected
0.01	Small difference w.r.t. expected
0.09	Medium difference w.r.t. expected
0.25	Large difference w.r.t. expected
1.00	Maximum difference w.r.t. expected

* J. Cohen (1988) Statistical Power Analysis for the Behavioral Sciences (2nd ed.)

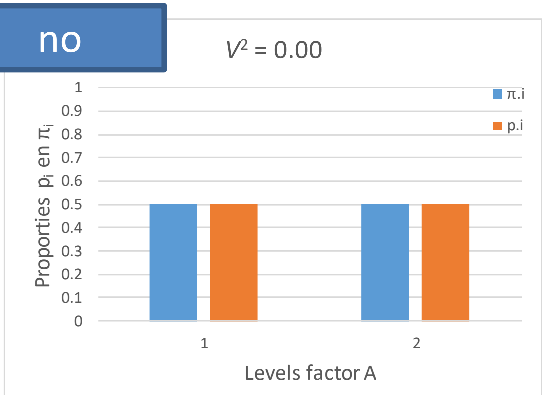
1-sample z-test: effect size



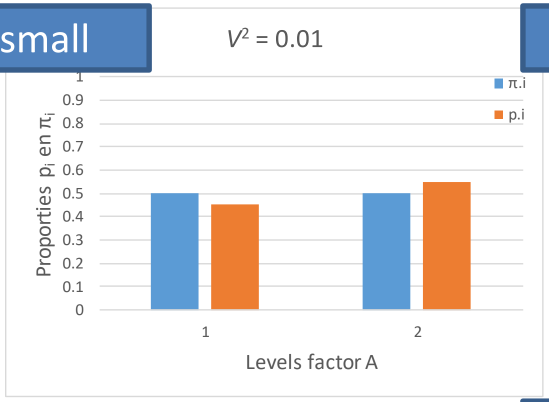
V^2 measures strength *deviation* from expected

π_i: expected
p_i: observed

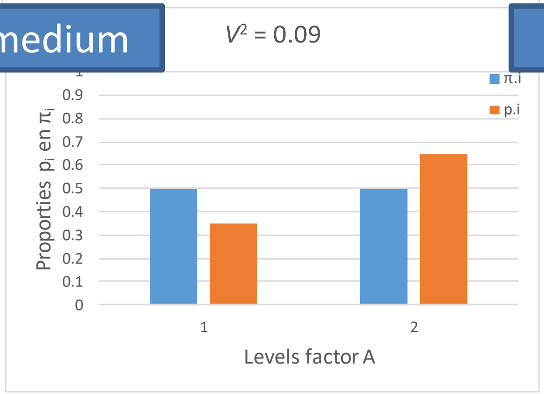
no



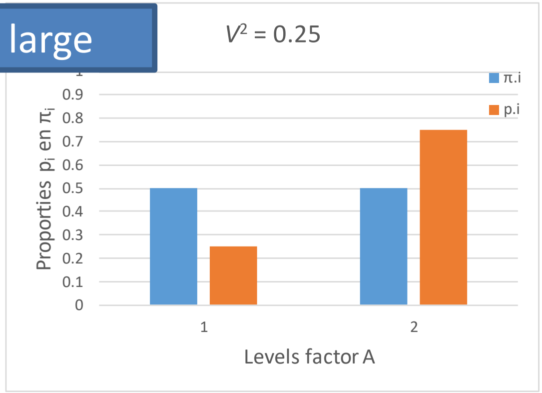
small



medium



large



maximal