11.2 Comparing Two Means AKA "two-sample problems" (page 648)

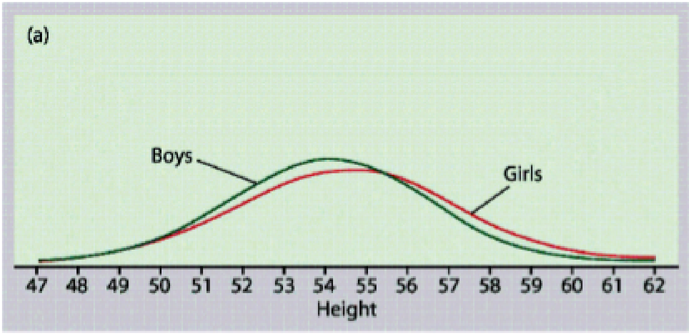
Compare

- responses to two treatments.
- characteristics of two populations.

Assumptions/conditions (page 650):

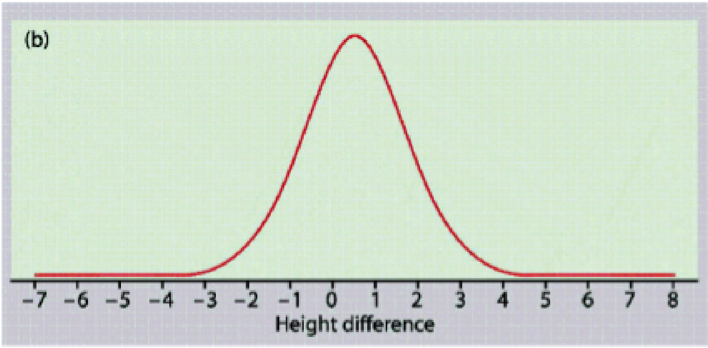
- 2 random samples from distinct populations
- Responses are independent, & measure same variable
- Ten % of populations are taken, or less.
- A large enough n to treat each sample mean as having a normal sampling distribution.

Suppose boys' heights follow N(μ_b, σ_b) & girls' heights follow N(μ_g, σ_g).



The differences in heights follow

$$N\left(\mu_b - \mu_g, \sqrt{\sigma_b^2 + \sigma_g^2}\right)$$



The two-sample t procedures (page 652-654)

To estimate $\mu_1 - \mu_2$ with $\bar{x}_1 - \bar{x}_2$, use $\left(\bar{x}_1 - \bar{x}_2\right) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

To test a hypothesis about the difference of means,

use $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ → 0

(page 653) distribution of two-sample t statistic
$\approx t_k$ distribution.

conservative degrees of freedom:
use k equal to the smaller of $n_1 - 1$ and $n_2 - 1$

more accurate degrees of freedom:
use calculator or software to find k
(page 659)

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

Robustness for the two-sample t procedures (page 656)

- more robust (resistant to violations of assumptions) than the one-sample t methods.

- most robust when $n_1=n_2$ & the populations have similar shapes.

ex. With 95% confidence, estimate the difference in mean cholesterol levels between dogs owned as pets and those residing in veterinary research clinics. Data come from a clinic's dogs and others brought there to be neutered.

group	n	mean	s
Pets	26	193	68
Clinic	23	174	44

ex. Fifty adult volunteers did a standard step-aerobics workout, but were randomly divided into two groups so that the 25 in group 1 used the low step height & the 25 in group 2 used the high step height. Assume the two groups are independent and the data are approximately normal. Let μ_1 and μ_2 represent the mean heart rates we would observe for the entire population represented by the volunteers if all members of this population did the workout using the low or high step height, respectively. Is there evidence of a higher heart rate at the end of the workout when a higher step height is used?

group	n	mean	s
1 low	25	90.00	9
2 high	25	95.08	12

You're now ready to do 11.37, 11.39, 11.41, 11.43, 11.45, 11.47, 11.49