One-sample

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ indpt

$\bar{X} = \frac{1}{n} \sum_i X_i$

- $E(\bar{X}) = \mu$
- $Var(\bar{X}) = \sigma^2/n$
- $\bar{X} \sim$ normal.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

if σ^2 known, done.

<u>But</u>: if σ^2 not known,

estimate σ^2 by $S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim ?$$

- def: if $z_1 \dots z_k \sim N(0,1)$
  then $\sum_{j=1}^{k} z_j^2 \sim \chi^2_k$
- def: if $z \sim N(0,1)$
  $u \sim \chi^2_k$
  then $\frac{z}{\sqrt{u/k}} \sim t_k$
- fact: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

$$t = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{(\bar{X}-\mu)/\sigma/\sqrt{n}}{S/\sqrt{n} \,/\, \sigma/\sqrt{n}}$$

$$= \frac{z}{S/\sigma} = \frac{z}{\sqrt{S^2/\sigma^2}} = \frac{z}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \cdot \frac{1}{n-1}}}$$

$$= \frac{z}{\sqrt{u/(n-1)}} \sim t_{n-1}$$

- if X_i not normal, this <u>fails</u>
- but if n large, CLT

and approx normality of Z

hence $s^2 \simeq \sigma^2 \Rightarrow$ approx normality of T

- if n not large, ???

Two-sample

$X_1, \ldots X_{n_1} \sim N(\mu_1, \sigma_1^2)$
$Y_1, \ldots Y_{n_2} \sim N(\mu_2, \sigma_2^2)$ indpt

$E(\bar{X}) = \mu_1 \quad Var(\bar{X}) = \sigma_1^2/n_1$

$E(\bar{Y}) = \mu_2 \quad Var(\bar{Y}) = \sigma_2^2/n_2$

$E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2$

$$Var(\bar{X} - \bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

if σ_i^2 unknown, estimate by s_i^2

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim ?$$

- assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\frac{(n_1-1)s_1^2}{\sigma^2} + \frac{(n_2-1)s_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$$

$$T \sim t_{n_1+n_2-2}$$

pooled t

$$\frac{(n_i-1)S_i^2}{\sigma_i^2} \sim \chi^2_{n_i-1}$$

So its mean is $n_i - 1$ and variance is $2(n_i-1)$

$$\Rightarrow E(S_i^2) = \sigma_i^2$$

$$Var(S_i^2) = \frac{2\sigma_i^4}{n_i-1}$$

$$E\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$Var\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right) = \frac{2\sigma_1^4}{n_1-1} + \frac{2\sigma_2^4}{n_2-1}$$

if $\sigma_1^2 \neq \sigma_2^2$, $\quad \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \nsim \chi^2$

Not exactly χ^2 but $W = a\chi^2_b$

$E(W) = ab \quad Var(W) = 2a^2b$

$$ab = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$2a^2b = \frac{2\sigma_1^4}{n_1-1} + \frac{2\sigma_2^4}{n_2-1}$$

$$a = \frac{\frac{2\sigma_1^4}{n_1-1} + \frac{2\sigma_2^4}{n_2-1}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$b = \frac{\left[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right]^2}{\frac{2\sigma_1^4}{n_1-1} + \frac{2\sigma_2^4}{n_2-1}}$$

<u>pretend</u> that $\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \sim \chi_f^2$

pretend that $T \sim t_f$ even though $\sigma_1^2 \neq \sigma_2^2$

finally in formula for t,
est. σ_i^2 by S_i^2

Welch-Satterthwaite t

use when don't assume $\sigma_1^2 = \sigma_2^2$

in practice, lose little by using even if $\sigma_1^2 = \sigma_2^2$