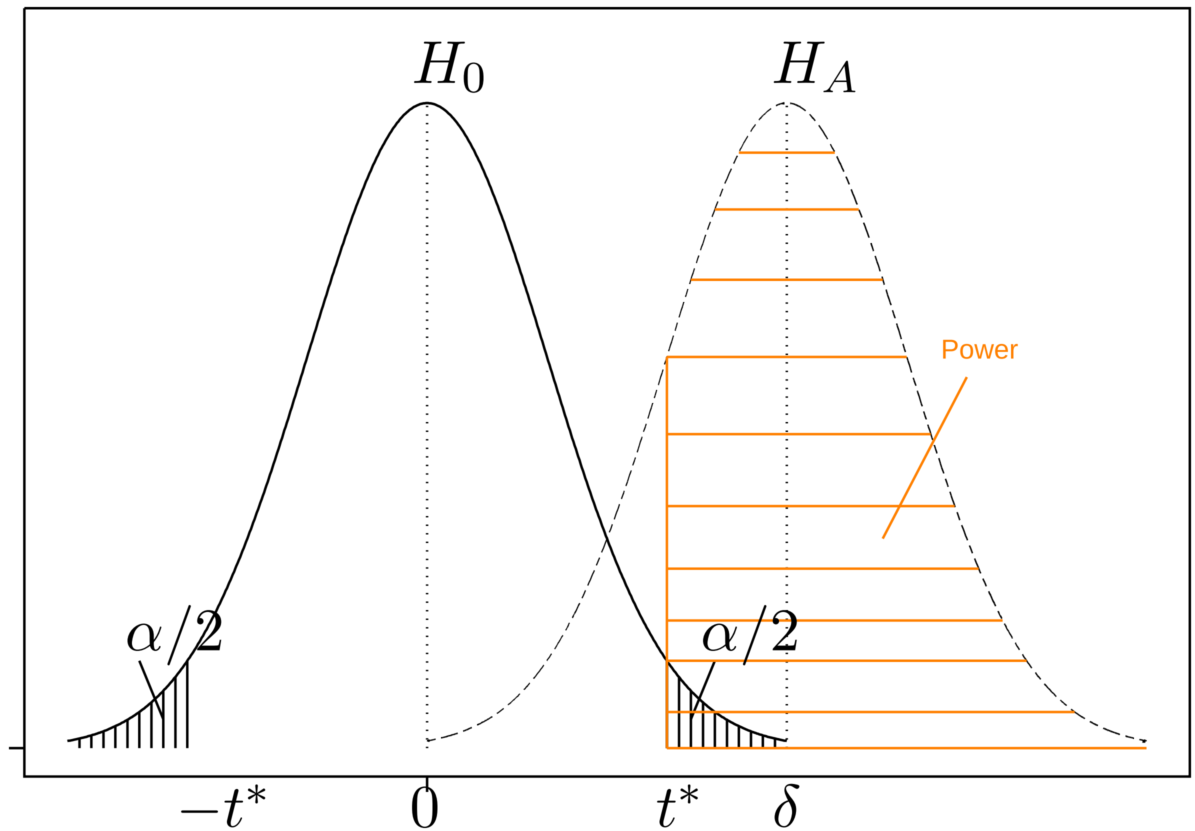
Power, Type I and II Error

- Type I error = reject H_0 when H_0 is true. The probability of a Type I error is called the significance level of the test, denoted by α.
- Type II error= fail to reject H_0 when H_0 is false. The probability of a type II error is denoted by β.
- The power of a test is

$$\text{power} = P(\text{ reject } H_0 | \text{ } H_0 \text{ is false }) = 1 - \beta$$

Test statistic under H_0 *and* H_A



($t^* = t_{1-\alpha/2}$)

The power depends on α, δ, σ and n

Power calculation in general

- Prospective: want a power of $\geq 80\%$, determine the necessary sample size.
- Retrospective: sample size was given, test not significant, how much power did we have?

2-sample t test

Let $X_{11}, \ldots, X_{1n}$ iid and $X_{21}, \ldots, X_{2n}$ iid independent.

$H_0 : X_{1i} \sim \mathcal{N}(\mu_1, \sigma^2), X_{2j} \sim \mathcal{N}(\mu_2, \sigma^2)$ with $\mu_1 = \mu_2$
$H_A : X_{1i} \sim \mathcal{N}(\mu_1, \sigma^2), X_{2j} \sim \mathcal{N}(\mu_2, \sigma^2)$ with $\mu_1 \neq \mu_2$

Under H_0:
$\bar{X}_1 - \bar{X}_2 \sim \mathcal{N}(0, \sigma^2(\frac{1}{n} + \frac{1}{n})) \Rightarrow \frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{2/n}} \sim \mathcal{N}(0, 1)$

Estimate σ^2 by $S_p^2 = \frac{S_1^2 + S_2^2}{2}$
$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{2/n}}$ follows a t distribution with $2n - 2$ df

Power calculation

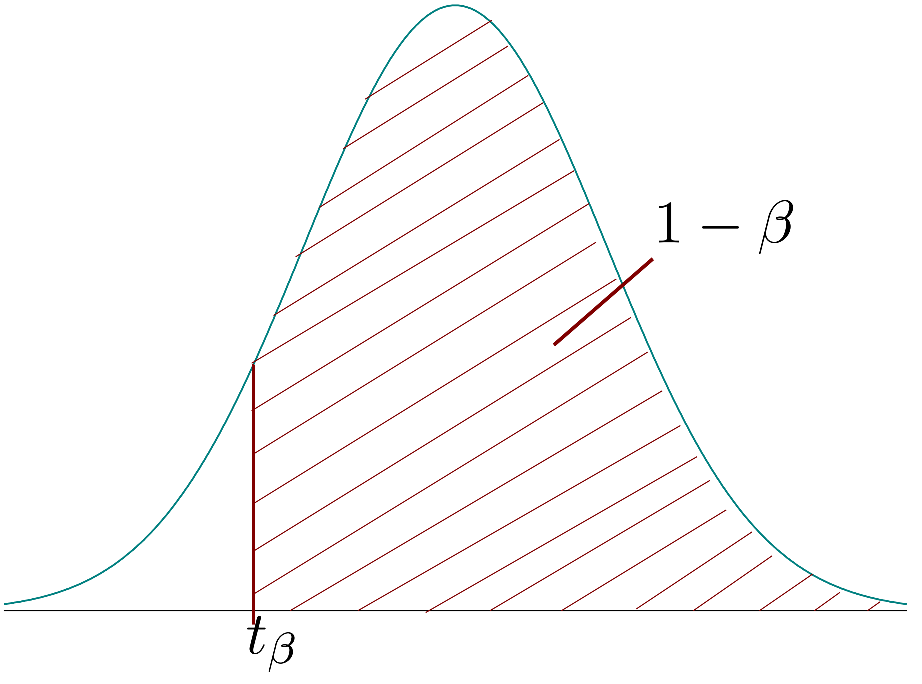
We reject H_0 if $t = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p\sqrt{2/n}} > t_{1-\alpha/2, 2n-2}$.

$$1-\beta = P(\frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{2/n}} < -t_{1-\alpha/2,2n-2}|H_A) + P(\frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{2/n}} > t_{1-\alpha/2,2n-2}|H_A).$$

Under H_A $\frac{\bar{X}_1 - \bar{X}_2 - \delta}{S_p\sqrt{2/n}}$ follows a t distribution with $2n-2$ df.
This implies

$$1-\beta = P(\frac{\bar{X}_1 - \bar{X}_2 - \delta}{S_p\sqrt{2/n}} > t_{1-\alpha/2} - \frac{\delta}{S_p\sqrt{2/n}}) + \underbrace{P(\frac{\bar{X}_1 - \bar{X}_2 - \delta}{S_p\sqrt{2/n}} < t_{\alpha/2} - \frac{\delta}{S_p\sqrt{2/n}})}_{\text{Prob } \approx 0 \quad (\text{for } \delta > 0)}.$$

Quantiles of the t distribution



It follows that $t_\beta = t_{1-\alpha/2} - \frac{\delta\sqrt{n}}{S_p\sqrt{2}}$

Equations for power calculation

For any $\delta \neq 0$, the following equations hold.

$$t_\beta = t_{1-\alpha/2} - \frac{|\delta|\sqrt{n}}{s_p\sqrt{2}} \quad (1)$$

$$n = 2(t_{1-\alpha/2} - t_\beta)^2 \cdot \frac{s_p^2}{\delta^2} \quad (2)$$

One-way anova

- The power of the F test for $H_0 : \mu_1 = \mu_2 = \ldots = \mu_I$ is

  $$1-\beta = P_{H_A}(\text{Test significant}) = P(F > F_{1-\alpha,I-1,N-I}|H_A).$$

- The distribution of F under H_A follows a *noncentral* F distribution with non-centrality parameter $\delta^2 = \frac{J\sum A_i^2}{\sigma^2}$ and $I-1$ and $N-I$ degrees of freedom.
- There are tables, graphs and software (e.g. GPower) which determine the power given $I-1, N-I, \alpha$ and δ.
- Use $\Delta = \frac{maxA_i - minA_i}{\sigma}$.

Daily weight gains

Average daily weight gains are to be compared among pigs receiving 4 levels of vitamin B_{12} in their diet.

We estimate σ with $\hat{\sigma} = 0.015$ lbs./day and we would like to detect a difference $maxA_i - minA_i = 0.03$ lbs/day. We set $\alpha = 0.05$ and want a power of 0.90 at least for a balanced design.

This implies $\Delta = 2$ and leads to a minimum of $n = 9$ pigs per group.