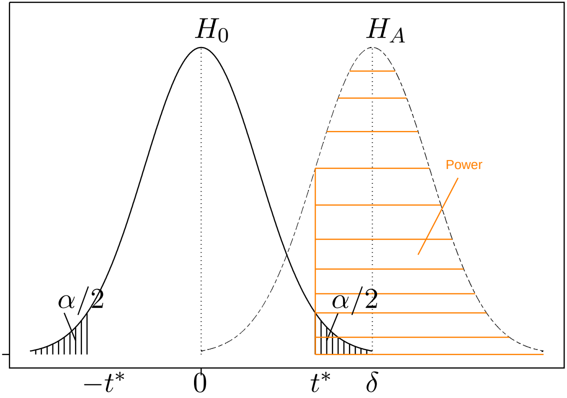
Power, Type I and II Error

- 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

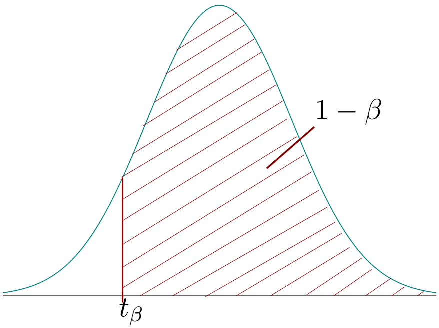
We reject H_0 if $t = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p\sqrt{2/n}} > t_{1-\alpha/2, 2n-2}$.

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

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

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

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.

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

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

- The power of the F test for $\mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_I$ 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 $\Delta = \frac{maxA_i - minA_i}{\sigma}$.

Detectable differences Δ for $\alpha = 5\%$ and $1 - \beta = 90\%$

	Number of groups I				
J	2	3	4	5	6
2	6.796	6.548	6.395	6.333	6.317
3	3.589	3.838	3.967	4.065	4.149
4	2.767	3.010	3.148	3.251	3.337
5	2.348	2.568	2.698	2.795	2.876
6	2.081	2.280	2.401	2.492	2.567
7	1.890	2.073	2.186	2.271	2.341
8	1.745	1.915	2.020	2.100	2.166
10	1.534	1.684	1.778	1.850	1.910
12	1.385	1.521	1.607	1.673	1.727
14	1.273	1.398	1.478	1.539	1.589
16	1.185	1.301	1.375	1.432	1.479
18	1.112	1.222	1.292	1.345	1.390
20	1.052	1.155	1.222	1.273	1.315
22	1.000	1.099	1.162	1.210	1.251
24	0.956	1.050	1.110	1.157	1.195
26	0.917	1.007	1.065	1.109	1.146
28	0.882	0.969	1.025	1.068	1.103
30	0.851	0.935	0.989	1.030	1.065
40	0.734	0.806	0.852	0.888	0.918
60	0.597	0.655	0.693	0.722	0.747
80	0.516	0.566	0.599	0.624	0.645
100	0.461	0.506	0.535	0.558	0.577
200	0.325	0.357	0.377	0.393	0.407
500	0.205	0.225	0.238	0.248	0.257
1000	0.145	0.159	0.168	0.176	0.181

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.