Power and Sample Size

power = $1-\beta$ = 1-P(fail to reject H_0 | H_0 false)
= P(reject H_0 | H_0 false)
= P(reject H_0 | H_A true)

Suppose I want to know how often the following test will reject H_0: $\mu_0 = 12$ when the true mean is $\mu = 10$. Suppose also that I can only afford to select 25 subjects.

<u>Model</u>
$X_1,\ldots,X_{n=25} \sim N(\mu,36)$ and $\alpha=0.05$

<u>Hypotheses</u>
H_0: $\mu = 12$ versus H_A: μ ? 12

Whether or not we reject the null hypothesis depends on the form of the alternative hypothesis. Hence we have the following three cases in which to calculate the power:

Case 1: H_A: $\mu < 12$
Case 2: H_A: $\mu > 12$
Case 3: H_A: $\mu \neq 12$

Case 1: H_A: $\mu < 12$

In this case, we reject when $T_{obs}^* < -Z_\alpha = -1.645$ ($\alpha = 0.05$)

power = 1-β = P(reject H_0 | H_A true)

$$P\left(T^* < -Z_\alpha \mid H_A\right) = P\left(T^* < -1.645 \mid \mu = 10\right)$$

$$= P\left(\frac{\sqrt{n}\left(\overline{X} - \mu_0\right)}{\sigma} < -1.645 \mid \mu = 10\right)$$

$$= P\left(\overline{X} - \mu_0 < -1.645\,\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right)$$

$$= P\left(\overline{X} < \mu_0 - 1.645\,\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right)$$

To calculate we need to standardize, so :

$$1 - \beta = P\left(\overline{X} < \mu_0 - 1.645\,\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right)$$

$$= P\left(\frac{\sqrt{n}\left(\overline{X} - \mu_A\right)}{\sigma} < \frac{\sqrt{n}\left(\mu_0 - 1.645\,\frac{\sigma}{\sqrt{n}} - \mu_A\right)}{\sigma} \mid \mu = 10\right)$$

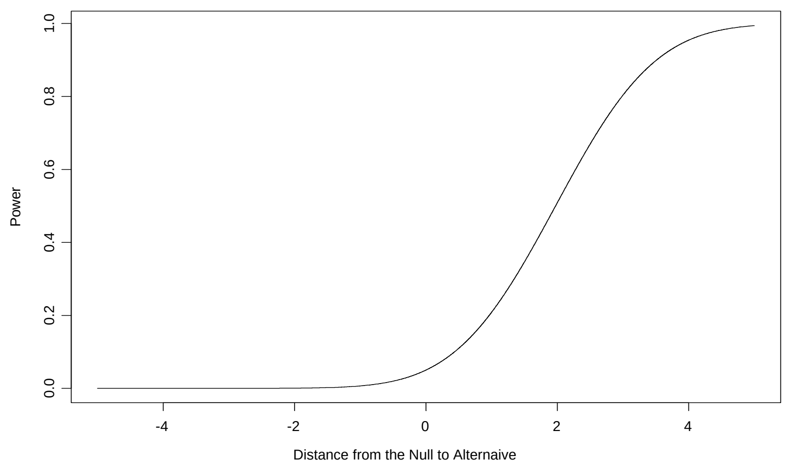
$$= P\left(Z < \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} - 1.645 \mid \mu = 10\right)$$

$$= P\left(Z < \frac{5\left(12 - 10\right)}{6} - 1.645\right) = P\left(Z < 0.0217\right) = 1 - 0.492 = 0.5180$$

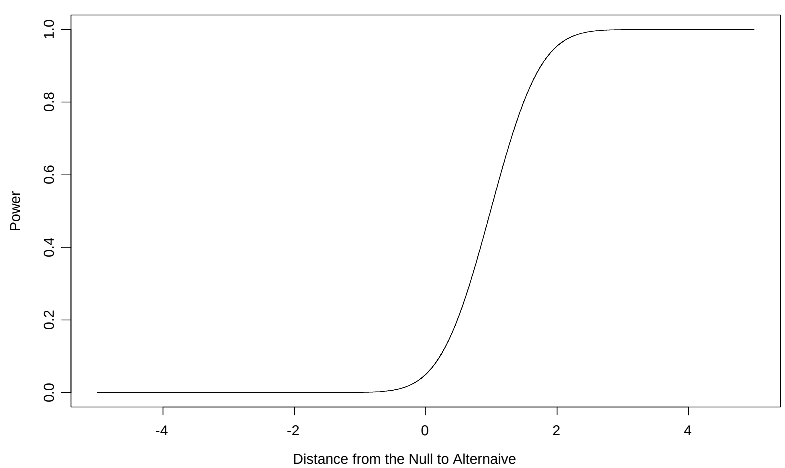
So we have that, for $X_1,\ldots,X_n \sim N(\mu,\sigma^2)$, the power to test H_0: $\mu = \mu_0$ versus H_A: $\mu < \mu_0$ at μ_A is given by

$$1-\beta = P\left(Z < \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} - Z_\alpha \right)$$

for n=25 and σ=6



for n=100 and σ=6



Both the variance and the sample size affect the slope of this curve.

Notice also that

$$1-\beta = P\left(Z > Z_{1-\beta}\right) \quad \text{and} \quad 1-\beta = P\left(Z < \frac{\sqrt{n}\left(\mu_0-\mu_A\right)}{\sigma} - Z_\alpha\right)$$

together imply that

$$\frac{\sqrt{n}\left(\mu_0-\mu_A\right)}{\sigma} - Z_\alpha = -Z_{1-\beta} \qquad \text{or}$$

$$n = \frac{\left(Z_\beta + Z_\alpha\right)^2 \sigma^2}{\left(\mu_0-\mu_A\right)^2}$$

In fact, we can use this formula to calculate sample size no matter what the alternative specifies. In the two sided case we simply replace α with $\alpha/2$.

Case 2: H_A: $\mu > 12$

In this case, we reject when $T_{obs}^* > Z_\alpha = 1.645$ ($\alpha=0.05$)

power = $1-\beta$ = P(reject H_0 | H_A true)

$$P\left(T^* > Z_\alpha \mid H_A\right) = P\left(T^* > 1.645 \mid \mu = 10\right)$$

$$= P\left(\frac{\sqrt{n}\left(\overline{X} - \mu_0\right)}{\sigma} > 1.645 \mid \mu = 10\right)$$

$$= P\left(\overline{X} < \mu_0 + 1.645\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right)$$

To calculate we need to standardize:

$$1 - \beta = P\left(\overline{X} > \mu_0 + 1.645\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right)$$

$$= P\left(\frac{\sqrt{n}\left(\overline{X} - \mu_A\right)}{\sigma} > \frac{\sqrt{n}\left(\mu_0 + 1.645\frac{\sigma}{\sqrt{n}} - \mu_A\right)}{\sigma} \mid \mu = 10\right)$$

$$= P\left(Z > \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} + 1.645 \mid \mu = 10\right)$$

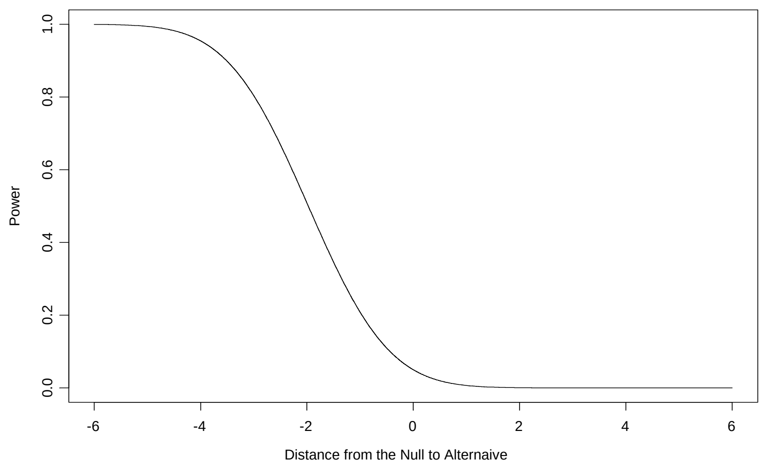
$$= P\left(Z > \frac{5\left(12 - 10\right)}{6} + 1.645\right) = P\left(Z > 3.3117\right) \approx 0$$

There is almost no power. Does this make sense? Why? (Hint: Draw a picture.)

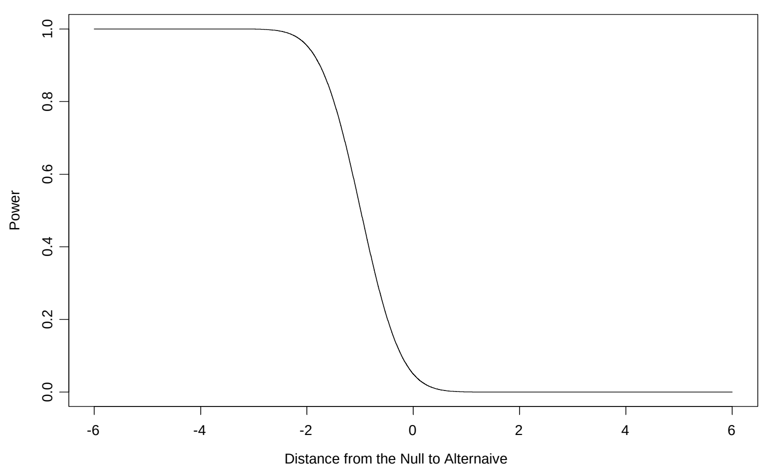
So we have that, for $X_1,\ldots,X_n \sim N(\mu,\sigma^2)$, the power to test H_0: $\mu = \mu_0$ versus H_A: $\mu > \mu_0$ at μ_A is given by

$$1-\beta = P\left(Z > \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} + Z_\alpha \right)$$

for n=25 and σ=6



for n=100 and σ=6



Both the variance and the sample size affect the slope of this curve.

Notice also that

$1-\beta = P\left(Z > Z_{1-\beta}\right)$ and $1-\beta = P\left(Z > \dfrac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} + Z_\alpha\right)$

together imply that

$\dfrac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} + Z_\alpha = Z_{1-\beta}$ or

$$n = \frac{\left(Z_{1-\beta} - Z_\alpha\right)^2 \sigma^2}{\left(\mu_0 - \mu_A\right)^2} = \frac{\left(Z_\beta + Z_\alpha\right)^2 \sigma^2}{\left(\mu_0 - \mu_A\right)^2}$$

This is the same formula we saw before.

Case 3: H_A: $\mu \neq 12$

In this case, we reject when $|T_{obs}^*| > Z_{\alpha/2} = 1.96$ (α=0.05). So to calculate the power we need to consider both rejection regions.

power = 1-β = P(reject H_0 | H_A true)

$$1-\beta = P\left(\left|T^*\right| > Z_{\alpha/2} \mid H_A\right)$$

$$= P\left(T^* < -Z_{\alpha/2} \quad \text{or} \quad T^* > Z_{\alpha/2} \mid H_A\right)$$

$$= P\left(T^* < -1.96 \mid \mu = 10\right) + P\left(T^* > 1.96 \mid \mu = 10\right)$$

$$= P\left(\frac{\sqrt{n}\left(\overline{X} - \mu_0\right)}{\sigma} < -1.96 \mid \mu = 10\right) + P\left(\frac{\sqrt{n}\left(\overline{X} - \mu_0\right)}{\sigma} > 1.96 \mid \mu = 10\right)$$

$$= P\left(\overline{X} < \mu_0 - 1.96\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right) + P\left(\overline{X} > \mu_0 + 1.96\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right)$$

$$= P\left(\frac{\sqrt{n}\left(\overline{X} - \mu_A\right)}{\sigma} < \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} - 1.96 \mid \mu = 10\right)$$

$$+ P\left(\frac{\sqrt{n}\left(\overline{X} - \mu_A\right)}{\sigma} > \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} + 1.96 \mid \mu = 10\right)$$

$$= P\left(Z < \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} - 1.96 \mid \mu = 10\right) + P\left(Z > \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} + 1.96 \mid \mu = 10\right)$$

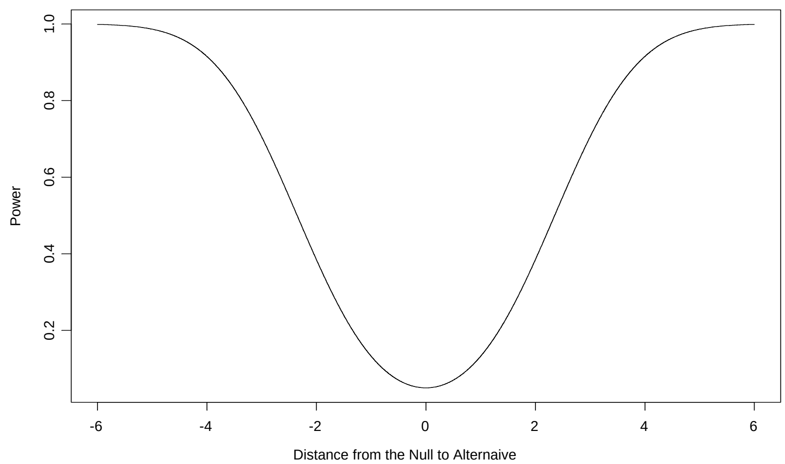
$$= P\left(Z < -0.2933 \mid \mu = 10\right) + P\left(Z > 3.6267 \mid \mu = 10\right)$$

$$= 0.386 + 0 = 0.386$$

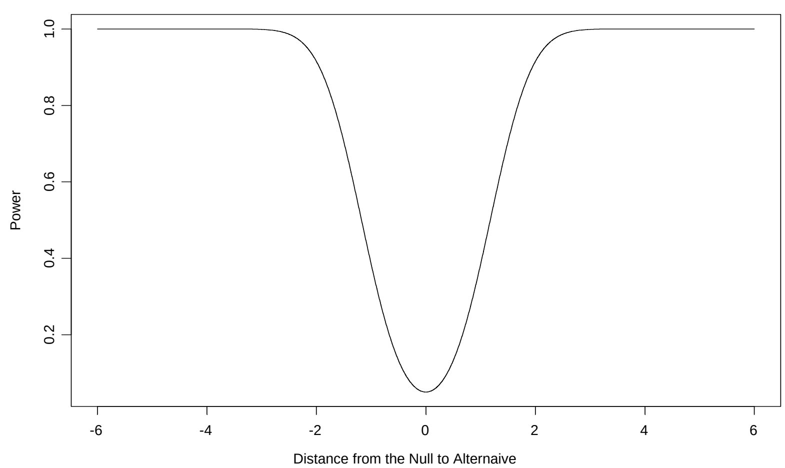
So we have that, for $X_1,\ldots,X_n \sim N(\mu,\sigma^2)$, the power to test H_0: $\mu = \mu_0$ versus H_A: $\mu \neq \mu_0$ at μ_A is given by

$$1-\beta = P\left(Z < \frac{\sqrt{n}(\mu_0 - \mu_A)}{\sigma} - Z_{\alpha/2} \right) + P\left(Z > \frac{\sqrt{n}(\mu_0 - \mu_A)}{\sigma} + Z_{\alpha/2} \right)$$

for n=25 and σ=6



for n=100 and σ=6



Both the variance and the sample size affect the curvature of this parabola.

Because the power in one tail will be close to zero we can ignore that probability and calculate the sample size with the following formula.

$$n = \frac{\left(Z_{\beta} + Z_{\alpha/2}\right)^2 \sigma^2}{\left(\mu_0 - \mu_A\right)^2}$$

This is the same formula we saw before and is technically conservative (why?).

Preliminaries about proportions

In the one-sample case:

$X_1,\ldots,X_n \sim \text{Ber}(\theta)$

H_0: $\theta=\theta_0$

Use either test statistic

$$T^* = \frac{\sqrt{n}\left(\hat{\theta}-\theta_0\right)}{\sqrt{\hat{\theta}\left(1-\hat{\theta}\right)}} \sim N(0,1) \text{ under } H_0$$

or

$$T^* = \frac{\sqrt{n}\left(\hat{\theta}-\theta_0\right)}{\sqrt{\theta_0\left(1-\theta_0\right)}} \sim N(0,1) \text{ under } H_0$$

For power calculations we use the latter.

The first is `robust', while the second will provide increased power under the null hypothesis.

Similarly for the two-sample case we have:

$X_1,\ldots,X \sim \text{Ber}(\theta_X)$
$Y_1,\ldots,Y \sim \text{Ber}(\theta_Y)$

When

H_0: $\theta_X = \theta_Y$ or H_0: $\theta_X - \theta_Y = 0$

Use either tests statistic

$$T^* = \frac{\hat{\theta}_X - \hat{\theta}_Y - (\theta_X - \theta_Y)}{\sqrt{\frac{\hat{\theta}_X(1-\hat{\theta}_X)}{n} + \frac{\hat{\theta}_Y(1-\hat{\theta}_Y)}{m}}} \sim \text{ Z in large samples}$$

or

$$T^* = \frac{\hat{\theta}_X - \hat{\theta}_Y - (\theta_X - \theta_Y)}{\sqrt{\hat{\theta}_p(1-\hat{\theta}_p)\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim \text{ Z in large samples}$$

Where $\hat{\theta}_p = \frac{n\hat{\theta}_X + m\hat{\theta}_Y}{n+m} = \frac{\sum X + \sum Y}{n+m}$ is the pooled estimate of $\theta_X = \theta_Y$

The first is used more often, but the second is actually a statistics that we come across again when we consider 2x2 tables.

Power for tests of proportions

Model

$X_1,\ldots,X_n \sim \text{Ber}(\theta)$ and $\alpha = 0.05$

Hypotheses

H_0: $\theta = \theta_0$ versus H_A: $\theta = \theta_A > \theta_0$

power = $1-\beta$ = P(reject H_0 | H_A true)

$$
\begin{aligned}
1-\beta &= P\left(T^* > Z_{\alpha/2} \mid H_A\right) \\
&= P\left(T^* > 1.96 \mid \theta = \theta_A\right) \\
&= P\left(\frac{\sqrt{n}\left(\hat{\theta}-\theta_0\right)}{\sqrt{\theta_0\left(1-\theta_0\right)}} > 1.96 \mid \theta = \theta_A\right) \\
&= P\left(\frac{\sqrt{n}\left(\hat{\theta}-\theta_A\right)}{\sqrt{\theta_A\left(1-\theta_A\right)}} > \frac{\sqrt{n}\left(\theta_0-\theta_A\right)}{\sqrt{\theta_A\left(1-\theta_A\right)}} + 1.96\frac{\sqrt{\theta_0\left(1-\theta_0\right)}}{\sqrt{\theta_A\left(1-\theta_A\right)}} \mid \theta = \theta_A\right) \\
&= P\left(Z > \frac{\sqrt{n}\left(\theta_0-\theta_A\right)}{\sqrt{\theta_A\left(1-\theta_A\right)}} + 1.96\sqrt{\frac{\theta_0\left(1-\theta_0\right)}{\theta_A\left(1-\theta_A\right)}} \mid \theta = \theta_A\right) \\
&= P\left(Z > Z_{1-\beta}\right) = P\left(Z < Z_\beta\right)
\end{aligned}
$$

Which implies that $\frac{\sqrt{n}\left(\theta_0-\theta_A\right)}{\sqrt{\theta_A\left(1-\theta_A\right)}} + Z_\alpha\sqrt{\frac{\theta_0\left(1-\theta_0\right)}{\theta_A\left(1-\theta_A\right)}} = Z_{1-\beta}$ or

$$
n = \left[\frac{Z_\beta\sqrt{\theta_A\left(1-\theta_A\right)} + Z_\alpha\sqrt{\theta_0\left(1-\theta_0\right)}}{\left(\theta_0-\theta_A\right)}\right]^2
$$

Power for tests of the difference between two means

Model
$X_1,\dots,X_n \sim N(\mu_X, \sigma_X^2)$
$Y_1,\dots,Y_m \sim N(\mu_Y, \sigma_Y^2)$

Hypotheses
H_0: $\mu_X - \mu_Y = \mu_0$ versus H_A: $\mu_X - \mu_Y = \mu_A$

power = $1-\beta$ = P(reject H_0 | H_A true)

$$
\begin{aligned}
1-\beta &= P\left(T^* > Z_{\alpha/2} \mid H_A\right) \\
&= P\left(T^* > Z_{\alpha} \mid \theta = \theta_A\right) \\
&= P\left(\frac{\overline{X} - \overline{Y} - \mu_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} > Z_\alpha \mid \mu = \mu_A\right) \\
&= P\left(Z > \frac{(\mu_0 - \mu_A)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} + Z_\alpha \mid \mu = \mu_A\right)
\end{aligned}
$$

To solve for n, let r=n/m be the ratio of sample sizes or the `allocation ratio'.

$$1-\beta = P\left(Z > \frac{(\mu_0-\mu_A)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{r\sigma_Y^2}{n}}} + Z_\alpha \right)$$

$$= P\left(Z > \frac{\sqrt{n}(\mu_0-\mu_A)}{\sqrt{\sigma_X^2 + r\sigma_Y^2}} + Z_\alpha \right)$$

which implies that

$$\frac{\sqrt{n}(\mu_0-\mu_A)}{\sqrt{\sigma_X^2 + r\sigma_Y^2}} + Z_\alpha = Z_{1-\beta}$$

and

$$n = \frac{(Z_\beta + Z_\alpha)^2}{(\mu_0-\mu_A)^2}\left(\sigma_X^2 + r\sigma_Y^2\right)$$

where r=n/m is the allocation ratio.

Consider the general formula

$$n = \frac{\left(Z_{\beta} + Z_{\alpha}\right)^2 \sigma^2}{\left(\Delta\right)^2}$$

Where σ^2 is the variance and Δ is the difference between the null and alternative hypotheses. Δ is often called the effect size.

Holding all thing constant what happens as

Δ increases $\Rightarrow$ n

σ^2 decreases $\Rightarrow$ n

$1-\beta$ increases $\Rightarrow$ n

α increases $\Rightarrow$ n

and

Δ increases $\Rightarrow$ $1-\beta$

σ^2 decreases $\Rightarrow$ $1-\beta$

n increases $\Rightarrow$ $1-\beta$

α increases $\Rightarrow$ $1-\beta$

A note on paired samples

Suppose I have paired observations (say, pre & post tests on an individual).

<u>Model</u>
$X_1,\ldots,X_n \sim N(\mu_X, \sigma_X^2)$
$Y_1,\ldots,Y_n \sim N(\mu_Y, \sigma_Y^2)$

<u>Hypotheses</u>
H_0: $\mu_X - \mu_Y = \mu_0$ versus H_A: $\mu_X - \mu_Y \neq \mu_0$

Here I might be tempted to use the following test statistic (or something similar to it)

$$T^* = \frac{\overline{X} - \overline{Y} - \mu_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$$

But this would wrong because

$$Var\left(\overline{X} - \overline{Y}\right) = Var\left(\overline{X}\right) + Var\left(\overline{Y}\right) - 2\operatorname{cov}(\overline{X}, \overline{Y})$$

The covariance in this case in not zero, because the X_i and Y_i observations are no long independent (they are paired). Hence the variance estimate in T^* is the wrong estimate.

Instead of going to all this trouble estimating the covariance, it is easier to simply analyze the differences.

Hence we define $X_i - Y_i = d_i$, Now

Model

$d_1,\ldots,d_n \sim N(\mu_{X}-\mu_Y, \sigma_d^2)$

Hypotheses

H_0: $\mu_d = \mu_0$ versus H_A: $\mu_d \neq \mu_0$

So that the problem is simply a one-sample problem, and since we do not know σ_d^2 we use:

$$T^* = \frac{\bar{d} - \mu_0}{\sqrt{\frac{S_d^2}{n}}} \sim \text{t with df=n-1}$$

where S_d^2 is the sample variance of the differences.

Note that we have avoided the problem of estimating the covariance because

$$S_d^2 \to Var\left(\bar{X} - \bar{Y}\right) = Var\left(\bar{X}\right) + Var\left(\bar{Y}\right) - 2\,\text{cov}(\bar{X},\bar{Y})$$