

# Contents

3.5.4 Paired Two Sample  $t$  Tests . . . . . 3

## Power Calculations

We now consider power calculations for two-sample tests. First, consider what influences power.

- **The signal.** We are trying to detect a signal that we use to reject  $H_0$ , namely  $|\mu_x - \mu_y|$ , and the larger this signal, the easier it is to detect.
- **The noise.** The noise that drowns the signal is the sampling variance  $\sigma^2$  or  $\sigma_x^2$  and  $\sigma_y^2$ . If the variance is very high, then the estimator  $|\bar{X} - \bar{Y}|$  will vary more than the signal  $|\mu_x - \mu_y|$ .
- **Sample sizes.** Since the variance of  $\bar{X}$  and  $\bar{Y}$  are inversely proportional to sample size, it should be clear that increasing sample size reduces the noise, allowing the signal to shine through!
- **Significance level.** The significance level  $\alpha$  is set to control type I error, but it also affects power. Trivially, if we set  $\alpha = 1$ , then we always reject  $H_0$  and have 100% power.

Rigorous power calculations require the sampling distribution of a  $t$  statistic under the specific alternative  $H_a$ . This distribution is a non-central  $t$ , under assumptions of normality. However, rather than trouble ourselves with this new distribution, we will make extra assumptions. We feel that there will be relatively little cost since power calculations always start with assumptions, namely the specific alternative  $H_a$ , that may or may not prove to be relevant once data are available. In other words, power calculations give guidance, but you probably won't want to bet the farm on their predictions.

Recall the steps to a power calculation for  $H_0 : \mu_x = \mu_y$ .

1. **Specify  $H_a$ .** Let's assume  $|\mu_x - \mu_y| = \Delta$ ,  $\sigma$  is known, and  $\alpha$  is given. Further, we will assume equal sample sizes  $n = m$ .
2. **Find critical region under  $H_0$ .** The test statistic is a  $Z$  statistic, since  $\sigma$  is known:

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{2}/\sqrt{n}}$$

The critical region is defined as  $|Z| > \phi_{\frac{\alpha+1}{2}}$ .

3. **Compute power.** We now compute the probability that  $Z$  lands in the critical regions given  $H_a$ .

$$\begin{aligned} P\left(|Z| > \phi_{\frac{1+\alpha}{2}} \mid \Delta\right) &= P\left(|\bar{X} - \bar{Y}| > \phi_{\frac{1+\alpha}{2}} \sigma\sqrt{2/n} \mid \Delta\right) \\ &= P\left(\bar{X} - \bar{Y} > \phi_{\frac{1+\alpha}{2}} \sigma\sqrt{2/n}\right) + P\left(\bar{X} - \bar{Y} < -\phi_{\frac{1+\alpha}{2}} \sigma\sqrt{2/n}\right) \\ &= P\left(\frac{\bar{X} - \bar{Y} - \Delta}{\sigma\sqrt{2/n}} > \phi_{\frac{1+\alpha}{2}} - \Delta/\sigma\sqrt{n/2}\right) \\ &\quad + P\left(\frac{\bar{X} - \bar{Y} - \Delta}{\sigma\sqrt{2/n}} < -\phi_{\frac{1+\alpha}{2}} - \Delta/\sigma\sqrt{n/2}\right) \\ &= 1 - \Phi\left(\phi_{\frac{1+\alpha}{2}} - \Delta/\sigma\sqrt{n/2}\right) + \Phi\left(\phi_{-\frac{1+\alpha}{2}} - \Delta/\sigma\sqrt{n/2}\right) \end{aligned}$$

Standardization in these derivations occurs under the  $H_a$  assumptions.

Now we have a functional relationship between power  $1 - \beta$  and significance  $\alpha$ , sample size  $n$ , and noise  $\sigma$

$$1 - \beta = 1 - \Phi\left(\phi_{\frac{1+\alpha}{2}} - \Delta/\sigma\sqrt{n/2}\right) + \Phi\left(\phi_{-\frac{1+\alpha}{2}} - \Delta/\sigma\sqrt{n/2}\right)$$

We can use this relationship to plot any vs. the other or to solve for one given all the others.

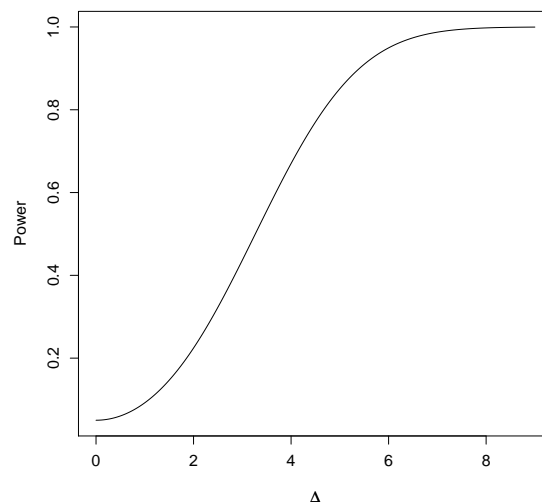
To make life a bit simpler, we also recognize that one of these terms, either  $1 - \Phi\left(\phi_{\frac{1+\alpha}{2}} - \Delta/\sigma\sqrt{n/2}\right)$  or  $\Phi\left(\phi_{\frac{1+\alpha}{2}} - \Delta/\sigma\sqrt{n/2}\right)$  will be much larger depending on the sign of  $\Delta$ . If  $\Delta > 0$ , then the right tail rejection region is the most likely to be hit under  $H_a$ , whereas the other tail becomes likely when  $\Delta < 0$ .

Example:

Returning to the iron retention example, suppose  $n = m = 18$ ,  $\sigma_x = \sigma_y = 5$ ,  $\alpha = 0.05$ . (1) Plot the power as a function of  $\Delta$  and (2) compute the sample size needed to detect  $\Delta = 1$  with power  $1 - \beta = 0.9$ .

(1)

```
delta <- seq(from=0, to=9, length.out=1000)
plot(x=delta, y=1-pnorm(qnorm(0.975) - delta*sqrt(9)/5) + pnorm(-qnorm(0.975)-delta*sqrt(9)/5),
     ylab="Power", xlab=expression(Delta), type="l")
```



(2) Since  $\Delta = 1 > 0$ , we recognize that the right tail term will dominate, so we seek  $n$  such that

$$0.9 = 1 - \Phi\left(\phi_{\frac{1+\alpha}{2}} - \Delta/\sigma\sqrt{n/2}\right)$$

but this is obtained by setting the 0.1 quantile of a standard normal to the expression inside,

$$\phi_{0.10} = \phi_{\frac{1+\alpha}{2}} - \Delta/\sigma\sqrt{n/2}$$

yielding

$$n = 2 \left[ 5(\phi_{0.1} - \phi_{\frac{1+\alpha}{2}}) \right]^2 = 525.4$$

computed as `R> 2*(5*(qnorm(0.1) - qnorm(0.975)))^2`. That's a lot of mice! If this many mice is infeasible, then changes in the experimental design should be considered that could reduce experimental noise.

### 3.5.4 Paired Two Sample $t$ Tests

#### Pairing Data

Often the  $X$  and  $Y$  samples are not independent. Some examples include

- Sampled individuals are matched on additional criteria before being randomly assigned to one of two treatments.
- Sampled individuals may be naturally paired, e.g. siblings or parent/offspring pairs.
- Before/after measurements on the same individual.

Pairing is one way to reduce the noise and increase power.

The data are now

$$(X_i, Y_i) \stackrel{\text{iid}}{\sim} N((\mu_x, \mu_y), \Sigma), \quad (3)$$

a bivariate normal data with mean vector  $(\mu_x, \mu_y)$  and some covariance matrix  $\Sigma$ . The  $X_i$  and  $Y_i$  are *not* independent of each other, but the the pairs are independent of each other, e.g.  $(X_i, Y_i)$  is independent of  $(X_j, Y_j)$ . Paired two-sample data are particularly easy to work with. In fact, it just reduces to the one-sample case. Define the difference

$$D_i = X_i - Y_i$$

Then the statistic is

$$T = \frac{\bar{D} - (\mu_x - \mu_y)}{S_{\bar{D}}} \sim t_{n-1}$$

where  $\bar{D}$  is the sample mean of differences and  $S_{\bar{D}}$  is the sample standard deviation. The confidence interval for  $\mu_x - \mu_y$  is

$$\bar{D} \pm t_{n-1} \left( \frac{1 + \alpha}{2} \right) S_{\bar{D}}$$

and the rejection region includes all statistics  $\bar{D}$  such that

$$|\bar{D}| > t_{n-1} \left( \frac{1 + \alpha}{2} \right) S_{\bar{D}}.$$

The  $p$ -value for observed statistic  $d$  is  $P(|D| > d)$ , computed as `R> 2*pt(-d, df=n-1)`.