Experimental Design and Statistical Methods

Workshop

COMPARISON OF MEANS OF TWO SAMPLES. PARAMETRIC TESTS

Jesús Piedrafita Arilla

jesus.piedrafita@uab.cat

Departament de Ciència Animal i dels Aliments

UAB

Items

- Comparison of two means
  - Independent samples
    - Comparison of two variances
  - Paired data
- Power of a test
  - One sided
  - Two sided
- Determination of sample size

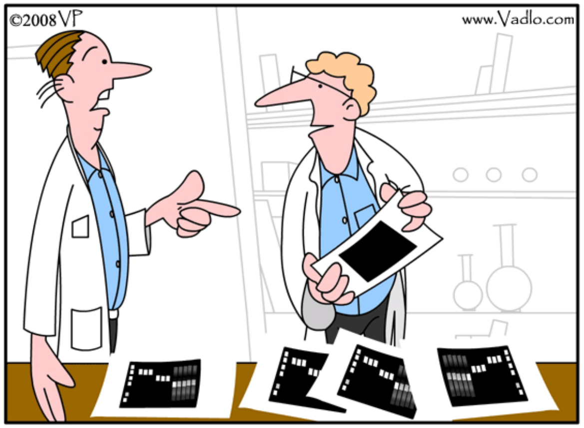
- Basic commands
  - t.test, var.test
  - t.test – paired
  - qt
  - pt
  - qnorm

Inferences on two samples

This session is devoted to present methods to compare location parameters of two samples through **parametric procedures**, which assume that the **samples follow a normal distribution** ($y \sim N(\mu,\sigma)$).

Non-parametric procedures will be presented later.

	Parametric	**Non parametric**
Two independent samples	t-test	Wilcoxon Sum-Rank test
Paired data	Paired t-test	Wilcoxon Signed-Rank test



Data don't make any sense, we will have to resort to ***statistics****.*

Test for the comparison of two means, unknown but equal variance in both groups (T-test), independent samples

Test statistic:

Pooled variance

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}; \qquad S_p^2 = \frac{(n_1 - 1)S_1^2 + \overbrace{(n_2 - 1)S_2^2}^{\text{Sum of squares of sample 2}}}{n_1 + n_2 - 2}$$

Alternative hypothesis		**Accepted if:**	
$H_1 : \mu_1 \neq \mu_2$	$\rightarrow$	$t > t_{1-\frac{\alpha}{2}}^{n_1+n_2-2}$ or $t < t_{\frac{\alpha}{2}}^{n_1+n_2-2}$	**Two-sided test**
$H_1 : \mu_1 > \mu_2$	$\rightarrow$	$t > t_{1-\alpha}^{n_1+n_2-2}$	**One-sided test**
$H_1 : \mu_1 < \mu_2$	$\rightarrow$	$t < t_{\alpha}^{n_1+n_2-2}$	

This test is **little sensitive to the normality assumption**, specially if sample size is large

Test for the comparison of two means, unknown but unequal variances in both groups (T-test), independent samples

Test statistic:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_\nu; \qquad \frac{1}{\nu} = \frac{1}{n_1 - 1}\left(\frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{\frac{s_2^2}{n_2}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)^2$$

Alternative hypothesis		**Accepted if:**	
$H_1 : \mu_1 \neq \mu_2$	$\rightarrow$	$t > t^{\nu}_{1-\frac{\alpha}{2}}$ or $t < t^{\nu}_{\frac{\alpha}{2}}$	**Two-sided test**
$H_1 : \mu_1 > \mu_2$	$\rightarrow$	$t > t^{\nu}_{1-\alpha}$	**One-sided test**
$H_1 : \mu_1 < \mu_2$	$\rightarrow$	$t < t^{\nu}_{\alpha}$	

If Y_1 and Y_2 are not normal, but sample size is large, this statistic is distributed N(0, 1) and then we can use z instead of t.

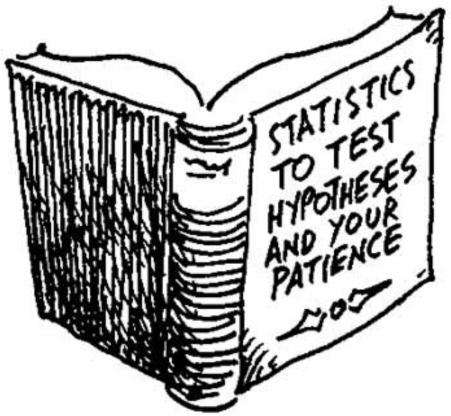
The Behrens-Fisher problem

The Behrens-Fisher problem arises in testing the difference between two means with a t test when the ratio of variances of the two populations from which the data were sampled is not equal to one. This condition is known as heteroscedasticity, which is a violation of one of the underlying assumptions of the t test. The resulting statistic is not distributed as t, and therefore the associated p values based on the entries found in standard t tables are incorrect. Use of tabulated critical values may lead to increased false positives, which are known as Type I errors, or a conservative test that lacks statistical power to detect significant treatment effects.

The typical solution in statistics packages for solving the two sample problem ($k = 2$) is the separate variances test, with modified degrees of freedom (Satterthwaite) as in the previous slide.

More details can be found in Sawilowsky (2002).

Test for the comparison of two variances, independent samples



Test statistic:

$$F = \frac{s_1^2}{s_2^2}$$

Alternative hypothesis | **Accepted if:**

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \quad \rightarrow \quad F > F_{1-\frac{\alpha}{2}}^{n_1-1, n_2-1} \text{ or } F < F_{\frac{\alpha}{2}}^{n_1-1, n_2-1}$$

Two-sided test

$$H_1 : \sigma_1^2 > \sigma_2^2 \quad \rightarrow \quad F > F_{1-\alpha}^{n_1-1, n_2-1}$$

$$H_1 : \sigma_1^2 < \sigma_2^2 \quad \rightarrow \quad F < F_{\alpha}^{n_1-1, n_2-1}$$

One-sided test

This test is **very sensitive to the non normality** of distributions. Small deviations to normality lead to accept the alternative hypothesis.

F distribution

The F-distribution becomes relevant when we try to calculate the ratios of variances of normally distributed statistics. Suppose we have two samples with n_1 and n_2 observations, the ratio

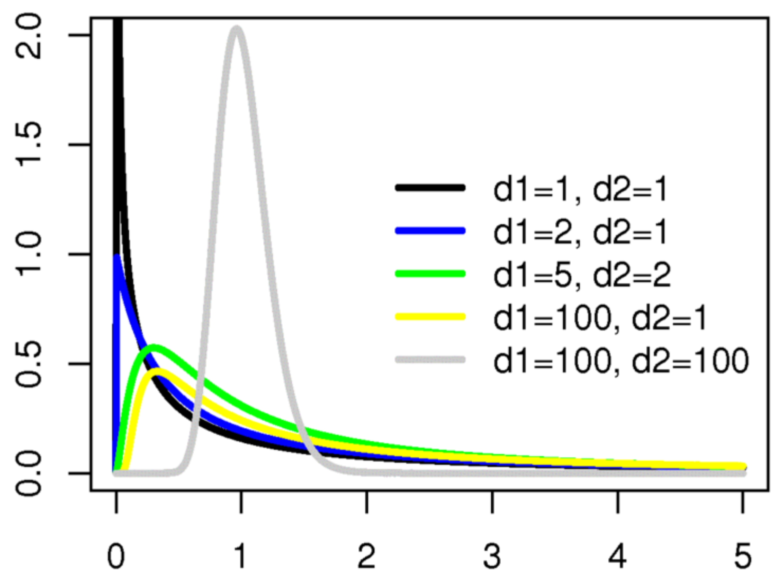
$$F = \frac{s_1^2}{s_2^2}$$

is distributed according to an F distribution (named after R.A. Fisher) with $\mathrm{df}_1 = n_1 - 1$ numerator degrees of freedom, and $\mathrm{df}_2 = n_2 - 1$ denominator degrees of freedom, with p.d.f.:

$$f(x) = \frac{\sqrt{\frac{(d_1\,x)^{d_1}\,d_2^{d_2}}{(d_1\,x+d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\!\left(\frac{d_1}{2},\frac{d_2}{2}\right)}$$

B is the beta-distribution

The F-distribution is skewed to the right, and the F-values can be only positive.



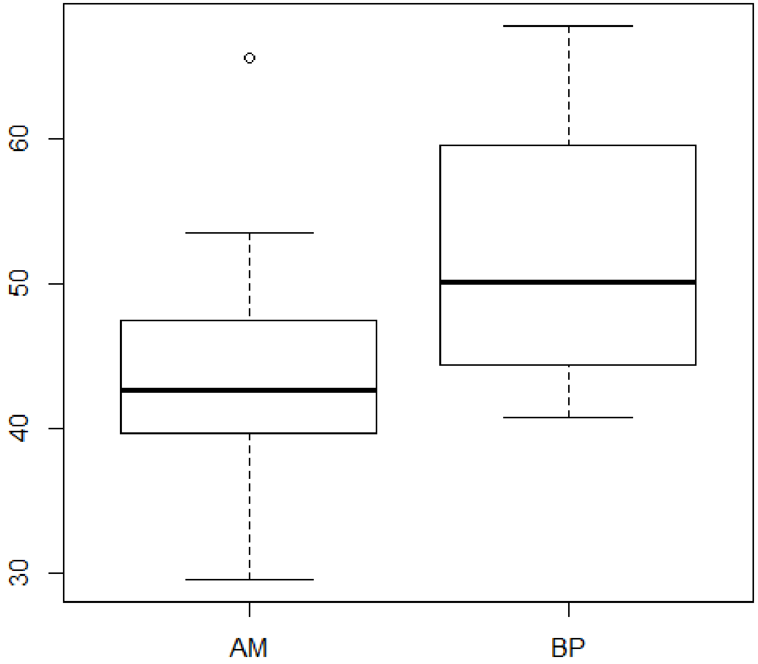
(http://en.wikipedia.org/wiki/F-distribution)

Comparing two sample means with R (1)

We want to compare the loin muscle area (cm^2) of two Spanish beef breeds: *Asturiana de la Montaña* and *Bruna dels Pirineus*.
Data (lmareaAMBP.txt) of the two samples are assumed to be independent.

A good starting point of the statistical analysis is to have an idea of the distributions:

```
> boxplot(LMAREA~BREED)
```



We can see that the means are different at sight and also the IQR. Are the variances similar? Are the means the same?

Comparing two sample means with R (2)

For the statistical analysis we need to test first if the variances are equal to make the best choice of the *t*-test method.

```
> var.test(LMAREA~BREED)

        F test to compare two variances

data:  LMAREA by BREED
F = 0.629, num df = 34, denom df = 35, p-value = 0.1794
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3196259 1.2418408
sample estimates:
ratio of variances
           0.62896
```

H_0: variances are equal (their ratio = 1), is not rejected

```
> 2*pf(.629,34,35)
[1] 0.1794193
```

Comparing two sample means with R (3)

```
> t.test(LMAREA~BREED, var.equal=T)

        Two Sample t-test

data:  LMAREA by BREED
t = -4.4878, df = 69, p-value = 2.804e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.637893  -4.475218
sample estimates:
mean in group AM mean in group BP
        43.92400         51.98056
```

The null hypothesis (means do not differ) is rejected

Observe that the confidence interval does not contain 0

```
> 2*pt(-4.4878,69)
[1] 2.803767e-05
```

In this case, this is the appropriate test, as the variances in the two groups did not differ (see previous slide)

Comparing two sample means with R (4)

```
> t.test(LMAREA~BREED)

        Welch Two Sample t-test

data:  LMAREA by BREED
t = -4.5024, df = 66.348, p-value = 2.784e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.628804  -4.484308
sample estimates:
mean in group AM mean in group BP
        43.92400         51.98056
```

The null hypothesis is rejected

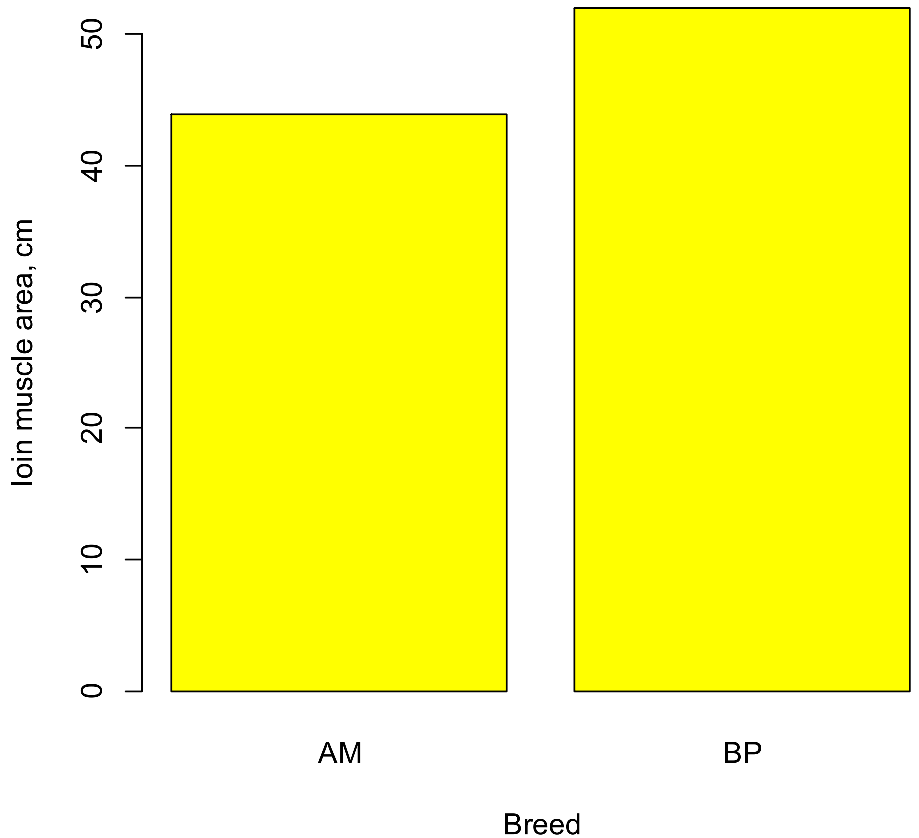
```
> 2*pt(-4.5024,66.348)
[1] 2.784752e-05
```

Note that:

1. **This is not the correct test**, as the variances did not differ.
2. The degrees of freedom are not an entire number, they have been calculated according to the Satterthwaite formula of a previous slide.

A graphical representation of the means

```
> MEANS<-tapply(LMAREA, BREED, mean)
> barplot(MEANS, xlab="Breed", ylab="loin muscle area, cm",
  + col="yellow")
```



Test for the comparison of two means, paired data

Test statistic:

$$T = \frac{\bar{d}}{S_d / \sqrt{n}} \sim t_{n-1}$$

Alternative hypothesis		**Accepted if:**	
$H_1 : \mu_1 \neq \mu_2$	$\rightarrow$	$t > t_{1-\frac{\alpha}{2}}^{n-1}$ or $t < t_{\frac{\alpha}{2}}^{n-1}$	**Two-sided test**
$H_1 : \mu_1 > \mu_2$	$\rightarrow$	$t > t_{1-\alpha}^{n-1}$	**One-sided test**
$H_1 : \mu_1 < \mu_2$	$\rightarrow$	$t < t_{\alpha}^{n-1}$	

An example of comparison of two means, paired data

Data of **hearth rate of hamsters** before and after an explosion.

	A	B	C
1	HAMSTER	BASAL	SCARED
2	1	70	115
3	2	84	148
4	3	88	176
5	4	110	191
6	5	105	158
7	6	100	178
8	7	110	179
9	8	67	140
10	9	79	161
11	10	86	157

```
> t.test(BASAL, SCARED, paired=T)

        Paired t-test

data:  BASAL and SCARED
t = -16.6325, df = 9, p-value = 4.585e-08
alternative hypothesis: true difference in
means is not equal to 0
95 percent confidence interval:
 -79.97498 -60.82502
sample estimates:
mean of the differences
                  -70.4
```

We reject H_0

The student is supposed to arrive by hand to the mean difference, t-value and the confidence intervals.

Power (1-β) of a one-sided test (1)

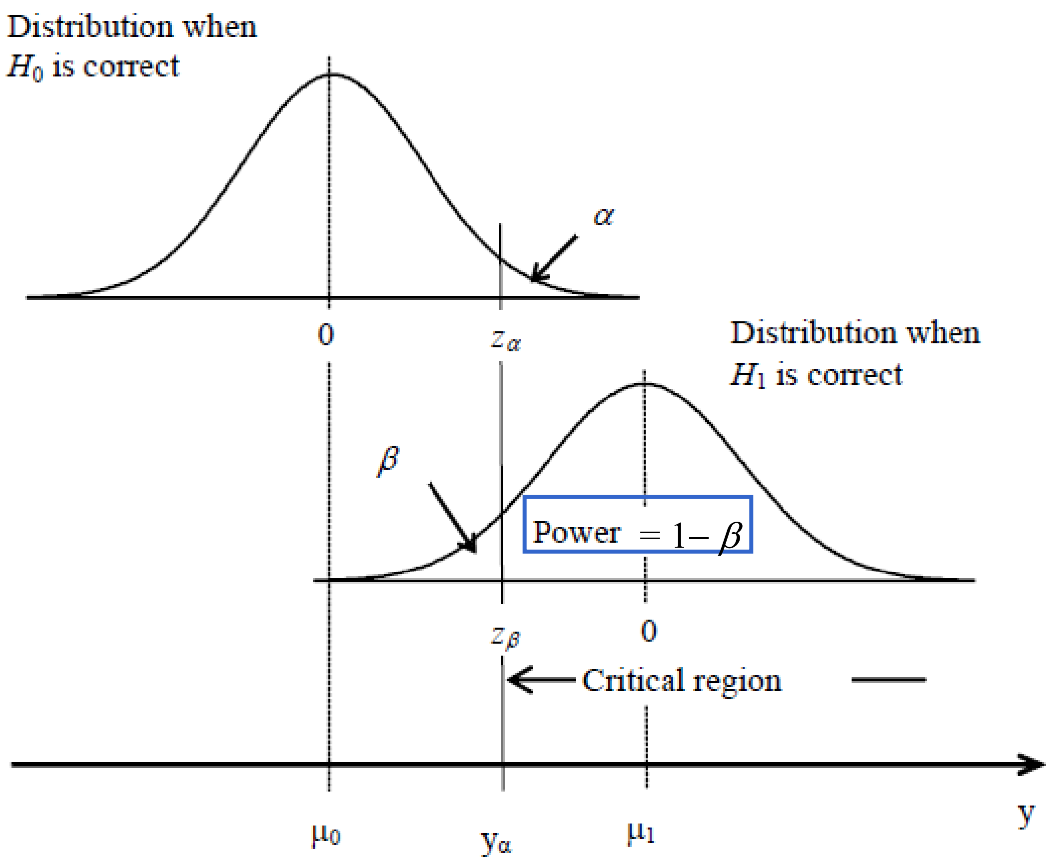


Figure 6.9 Standard normal distributions for H_0 and H_1. The power, type I error (α) and type II error (β) for the one-sided test are shown. On the bottom is the original scale of variable y

$$z_\beta = \frac{y_\alpha - \mu_1}{\sigma_{D1}}$$

$$y_\alpha = (\mu_0 + z_\alpha \sigma_{D0})$$

$$z_\beta = \frac{(\mu_0 + z_\alpha \sigma_{D0}) - \mu_1}{\sigma_{D1}}$$

z_β is expressed in terms of z_α and the means and SD of the distributions

Power

$\mu_0 < \mu_1 : P[z > z_\beta]$

$\mu_0 > \mu_1 : P[z < z_\beta]$

(Kaps and Lamberson, 2004)

Power of a one-sided test (2)

For specific tests, the **appropriate standard deviation** must be specified.

For the test hypothesis $\mu_1 > \mu_0$:

$$z_\beta = \frac{(\mu_0 + z_\alpha \, \sigma_0 / \sqrt{n}) - \mu_1}{\sigma_1 / \sqrt{n}}$$

$$= \frac{(\mu_0 - \mu_1)}{\sigma / \sqrt{n}} + z_\alpha$$

$$\sigma_0 / \sqrt{n} \, , \sigma_1 / \sqrt{n}$$ Standard errors

$$\sigma_0 = \sigma_1 = \sigma$$

Testing the hypothesis of the **difference of two means**:

$$z_\beta = \frac{(\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} + z_\alpha$$

We shall see later that from these formulas we can arrive at calculating the sample size of an experiment

A first example

Example: The arithmetic mean of milk yield from a sample of 30 cows is 4100 kg. Is that value significantly greater than 4000 kg? The variance is 250000. Calculate the power of the test.

$$\mu_0 = 4000 \text{ kg (if } H_0)$$
$$\bar{y} = 4100 \text{ kg } (= \mu_1 \text{ if } H_1)$$
$$\sigma^2 = 250000, \text{ and the standard deviation is } \sigma = 500 \text{ kg}$$

$$z = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}} = \frac{4100 - 4000}{500 / \sqrt{30}} = 1.095$$

For $\alpha = 0.05$, $z_\alpha = 1.65$, since the calculated value $z = 1.095$ is not more extreme than the critical value $z_\alpha = 1.65$, H_0 is not rejected with $\alpha = 0.05$. The sample mean is not significantly different than 4000 kg. The power of the test is:

$$z_\beta = \frac{(\mu_0 - \mu_1)}{\sigma / \sqrt{n}} + z_\alpha = \frac{4000 - 4100}{500 / \sqrt{30}} + 1.65 = 0.55$$

Using the H_1 distribution, the power is $P(z > z_\beta) = P(z > 0.55) = 0.29$. The type II error, that is, the probability that H_0 is incorrectly accepted is $1 - 0.29 = 0.71$. The high probability of error is because the difference between means for H_0 and H_1 is relatively small compared to the variability.

(Kaps and Lamberson, 2004)

Power of a two-sided test

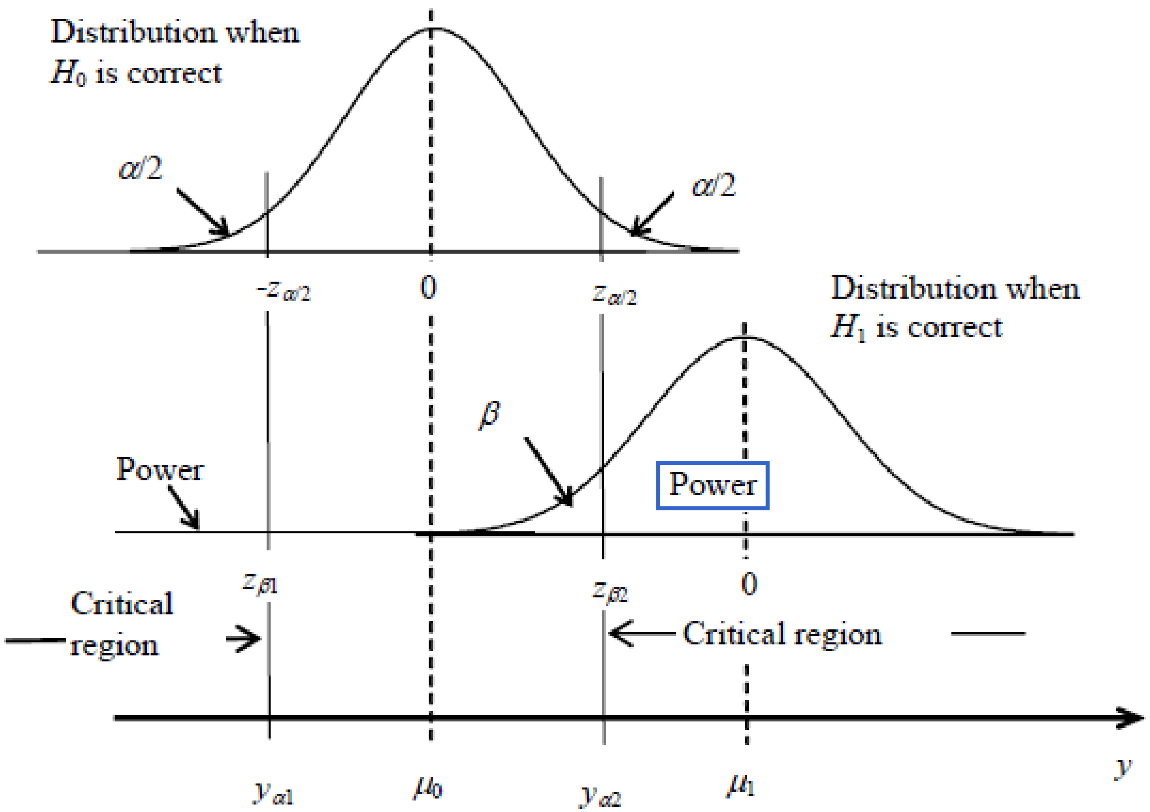


Figure 6.10 Standard normal distributions for H_0 and H_1. The power, type I error (α) and type II error (β) for the two-sided test are shown. On the bottom is the original scale of the variable *y*

$$z_{\beta 1} = \frac{(\mu_0 - z_{\frac{\alpha}{2}} \sigma_{D0}) - \mu_1}{\sigma_{D1}}$$

$$z_{\beta 2} = \frac{(\mu_0 + z_{\frac{\alpha}{2}} \sigma_{D0}) - \mu_1}{\sigma_{D1}}$$

Power:

$P[z < z_{\beta 1}] + P[z > z_{\beta 2}]$, using the H_1 distribution

(Kaps and Lamberson, 2004)

Power with unknown variances

1. Calculate it by using a t distribution:
   - If H_0 holds, the test statistic t has a central t distribution
   - If H_1 holds, the t statistic has a noncentral t distribution
2. For a given critical value t_α the probability (area) is calculated from the noncentral t distribution.
3. Power
   - One-sided: $P[t > t_\alpha = t_\beta]$
   - Two-sided: $P[t < -t_{\alpha/2} = t_{\beta 1}] + P[t > t_{\alpha/2} = t_{\beta 2}]$
4. Other data needed:

Hypothesis	**One-sided / Two-sided**	**Noncentrality parameter**	**Degrees of freedom**
H_0: $\mu = \mu_0$ H_1: $\mu = \mu_1$	One-sided and Two-sided	$\lambda = \frac{\lvert\mu_1 - \mu_0\rvert}{s}\sqrt{n}$	n - 1
H_0: μ_1 - μ_2 = 0 H_1: μ_1 - μ_2 = δ	One-sided	$\lambda = \frac{\lvert\mu_1 - \mu_2\rvert}{s_p}\sqrt{\frac{n}{2}}$	$2n$ - 2
	Two-sided	$\lambda = \frac{\lvert\mu_1 - \mu_2\rvert}{s_p}\sqrt{\frac{n}{2}}$	$2n$ - 1

A second example (1)

Another example: Two groups of eight cows were fed two different diets (A and B) in order to test the difference in milk yield. From the samples the following was calculated:

	Diet A	Diet B
Mean ($\bar{y}$), *kg*	21.8	26.4
Std. deviation (s)	4.1	5.9
Number of cows (n)	7	7

Test the null hypothesis, H_0: $\mu_2 - \mu_1 = 0$, and calculate the power of test by defining the alternative hypothesis H_1: $\mu_2 - \mu_1 = \bar{y}_2 - \bar{y}_1 = 4.6$ kg.

The test statistic is:

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{s_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

The standard deviation is:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(7-1)(4.1^2) + (7-1)(5.9^2)}{7+7-2}} = 5.080$$

(Kaps and Lamberson, 2004)

A second example (2)

The calculated t value is:

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{s_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(26.4 - 21.8) - 0}{5.080\sqrt{\left(\frac{1}{7} + \frac{1}{7}\right)}} = 1.694$$

For $\alpha = 0.05$ and degrees of freedom $(n_1 + n_2 - 2) = 12$, the critical value for the two-sided test is $t_{0.25} = 2.179$. The calculated $t = 1.694$ is not more extreme than the critical value and H_0 is not rejected.

The power for this test is:

$$Power = (1 - \beta) = P[t < -t_{\alpha/2} = t_{\beta 1}] + \quad P[t > t_{\alpha/2} = t_{\beta 2}]$$

Using a t distribution for H_1 with the noncentrality parameter

$$\lambda = \frac{|\mu_1 - \mu_2|}{s_p}\sqrt{\frac{n}{2}} = \frac{|21.8 - 26.4|}{5.080}\sqrt{\frac{7}{2}} = 1.694$$ and degrees of freedom $df = 12$, the

power is:

$$Power = (1 - \beta) = P[t < -2.179] + P[t > 2.179] = 0.000207324 + 0.34429 = 0.34450$$

(Kaps and Lamberson, 2004)

Power calculations in a *t*-distribution – R script-

```
> ###Power of a two sample comparison of means, two sided test
> #qt(p,df): quantile at the critical point p of a t distribution,
> #with df degrees of freedom
> #pt(qt(p,df),df,ncp): cumulated probability until the quantile in a
> #noncentral t distribution with df degrees of freedom and noncentrality
> #parameter ncp
>
> ###An example:
> POWERL <- pt(qt(0.025,12),12,1.694); POWERL
[1] 0.0002072797
> POWERR <- 1-pt(qt(0.975,12),12,1.694); POWERR
[1] 0.3443096
> POWER2t <- POWERL + POWERR; POWER2t
[1] 0.3445168
```

We have a probability of 0.3445 (or 34.45%) of rejecting the null hypothesis (H_0) when false.

This is a value lower than the standard required power for a test: 0.8.

Power calculations in a *t*-distribution – simpler-

```
> power.t.test(n=7, delta=4.6, sd=5.08,
type="two.sample", alternative="two.sided")

     Two-sample t test power calculation

              n = 7
          delta = 4.6
             sd = 5.08
      sig.level = 0.05
          power = 0.344329
    alternative = two.sided

NOTE: n is number in *each* group
```

We have a probability of 0.3443 (or 34.43%) of rejecting the null hypothesis (H_0) when false.

This function is useful for comparing means with equal sample size.

`sd` is the pooled standard deviation

Determination of Sample size through power analysis (1)

- The effect size (minimum difference) of biological interest (relevance), δ
- The standard deviation, σ
- The significance level, α (5%)
- The desired power of the experiment, 1-β (80%)
- The alternative hypothesis (i.e. a one- or two-sided test)
- The sample size

Fix any five of these variables and a mathematical relationship can be used to estimate the sixth

Determination of Sample size through power analysis (2)

After some algebra from the formulas of power, it can be arrived at:

One-sided test of a population mean

$$n = \frac{(z_\alpha - z_\beta)^2 \sigma^2}{\delta^2}$$

One-sided test of the difference of two population means

$$n = \frac{(z_\alpha - z_\beta)^2}{\delta^2} 2\sigma^2$$

For a **two-sided test**, replace z_α with $z_{\alpha/2}$ in the above expressions.

Remember that z_α and z_β are obtained from a table of percentile of a standard normal distribution:

$$z_{\alpha/2}(0.05) = 1.96;\ \ z_\alpha(0.05) = 1.65$$
$$z_\beta(0.8) = -0.85;\ \ z_\beta(0.9) = -1.28$$

```
> qnorm(0.975)
[1] 1.959964
```

Determination of Sample size. An example

Example: What is the sample size required in order to show that a sample mean of 4100 kg for milk yield is significantly larger than 4000 kg? It is known that the standard deviation is 800 kg. The desired level of significance is 0.05 and power is 0.80.

$\mu_0 = 4000$ kg and $\bar{y} = 4100$ kg, thus the difference is $\delta = 100$ kg, and $\sigma^2 = 640000$

For $\alpha = 0.05$, $z_\alpha = 1.65$
For power $(1 - \beta) = 0.80$, $z_\beta = -0.84$
Then:

$$n = \frac{(z_\alpha - z_\beta)^2}{\delta^2}\sigma^2 = \frac{(1.65 + 0.84)^2}{100^2} 640000 = 396.8$$

Thus, 397 cows are needed to have an 80% chance of proving that a difference as small as 100 kg is significant with $\alpha = 0.05$.

For observations drawn from a normal population and when the variance is unknown, the required sample size can be determined by using a noncentral t distribution. If the variance is unknown, the difference and variability are estimated.

(Kaps and Lamberson, 2004)

A final remark

By increasing the sample size of the experiment (number of replicates of each group) we can always arrive at an STATISTICALLY SIGNIFICANT difference.

This difference, however, must be also RELEVANT in any sense (biological, economical, etc.).

This reasoning must be applied to any other statistic test we apply in our research.

References

Kaps M., Lamberson W.R. 2004. *Biostatistics for the Animal Sciences*. CABI, Wallinford.