Estimation of Statistical Parameters
OUTLINE

- Estimation & Confidence Intervals:
  - Normality analysis based on statistical terms
  - Confidence intervals for means
  - Confidence intervals for frequencies

- Hypothesis testing:
  - Concept and Practice
Normal distribution:
- Gaussian distribution
- Symmetric
- Not skewed
- Unimodal
- Described by two parameters:
  - Probability density function:
  - $\mu$ & $\sigma$ are parameters
  - $\mu =$ mean
  - $\sigma =$ standard deviation
  - $\pi$, $e =$ constants

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2}$$
Normal distribution: Why do we use it!
- Many biological variables follow a normal distribution
- The normal distribution is well-understood, mathematically

Punctual estimation
- Is a value for estimated theoretical parameter
  - \( m \) (sample mean) is a punctual estimation of \( \mu \) (population mean)
- Is influenced by the fluctuations from sampling
- Could be very far away from the real value of the estimated parameter
Why Confidence Intervals?

- It is recommended to estimate a theoretical parameter by using a range of value not a single value
  - It is called confidence intervals
  - The estimated parameter belong to the confidence intervals with a high probability.
Definitions

- A range around the sample estimate in which the population estimate is expected to fall with a specified degree of confidence, usually 95% of the time at a significance level of 5%.
  - \( P[\text{lower critical value} < \text{estimator} < \text{higher critical value}] = 1-\alpha \)
  - \( \alpha = \text{significance level} \)
- The range defined by the critical values will contains the population estimator with a probability of \( 1-\alpha \)
- It is applied when variables are normal distributed!
Confidence Intervals: Interpretation

- If 0 is contains by the confidence intervals for the distance between an observed and theoretical mean, the difference between the two investigated means is 0.
- If 0 is NOT contains by the confidence intervals for the distance between an observed and theoretical mean, the difference between the two investigated means is NOT 0.
Confidence Intervals: Interpretation

- Were this procedure to be repeated on multiple samples, the calculated confidence interval (which would differ for each sample) would encompass the true population parameter 95% of the time.
- The confidence interval represents values for the population parameter for which the difference between the parameter and the observed estimate is not statistically significant at the 5% level.
Confidence Intervals

- It is calculated taking into consideration:
  - The sample or population size
  - The type of investigated variable (qualitative OR quantitative)
- Formula of calculus comprised two parts
  - One estimator of the quality of sample based on which the population estimator was computed (standard error)
    - Standard error: is a measure of how good our best guess is.
    - Standard error: the bigger the sample, the smaller the standard error.
    - Standard error: is always smaller than the standard deviation
  - Degree of confidence ($Z_\alpha$ score)
- It is possible to be calculated for any estimator but is most frequent used for mean
Confidence Intervals for Means

- Standard error of mean is equal to standard deviation divided by square root of number of observations:
  - If standard deviation is high, the chance of error in estimator is high
  - If sample size is large, the chance of error in estimator is small

\[
\left[ \bar{X} - Z_\alpha \frac{s}{\sqrt{n}}, \bar{X} + Z_\alpha \frac{s}{\sqrt{n}} \right] \quad \left[ m - Z_\alpha \frac{s}{\sqrt{n}}, m + Z_\alpha \frac{s}{\sqrt{n}} \right]
\]
The mean of blood sugar concentration of a sample of 121 patients is equal to 105 and the variance is equal to 36.

Which is the confidence levels of blood sugar concentration of the population from which the sample was extracted?

Use a significance level of 5% ($Z = 1.96$). It is considered that the blood sugar concentration is normal distributed.

- $n = 121$
- $s^2 = 36$
- $s = 6$
- $m = 105$

$$\left[105 - 1.96 \frac{6}{\sqrt{121}}; 105 + 1.96 \frac{6}{\sqrt{121}}\right]$$

- $[105 - 1.07; 105 + 1.07]$
- $[103.93; 106.07]$
- $[104; 106]$
Comparing Means by using Confidence Levels

SBP (mmHg)

Treatament A

Treatament B

Treatament C

100

200
Confidence Intervals for Frequencies

Could be computed if:
- \( n \times f > 10 \), where \( n \) = sample size, \( f \) = frequency

\[
\left[ f - Z_\alpha \sqrt{\frac{f(1-f)}{n}} ; f + Z_\alpha \sqrt{\frac{f(1-f)}{n}} \right]
\]
We are interested in estimating the frequency of breast cancer in women between 50 and 54 years with positive family history. In a randomized trial involving 10,000 women with positive history of breast cancer were found 400 women diagnosed with breast cancer.

What is the 95% confidence interval associated frequently observed?

\[ f = \frac{400}{10000} = 0.04 \]

\[ [0.04 - 1.96 \sqrt{\frac{0.04 \cdot 0.96}{10000}}; 0.04 + 1.96 \sqrt{\frac{0.04 \cdot 0.96}{10000}}] \]

\[ [0.04 - 0.004; 0.04 + 0.004] \]

\[ [0.036; 0.044] \]
Correct estimation of a statistical parameter is done with confidence intervals.

Confidence intervals depend by the sample size and standard error.

The confidence intervals is larger for:

- High value of standard error
- Small sample sizes
Objective:
- Understand the principles of hypothesis-testing
- To be able to interpret P values correctly
- To know the steps needed in application of a statistical test
Definitions

- **Statistical hypothesis test** = a method of making statistical decisions using experimental data.
- A result is called **statistically significant** if it is unlikely to have occurred by chance.
- Statistical hypothesis = an assumption about a population parameter. This assumption may or may not be true.
Clinical hypothesis = a single explanatory idea that helps to structure data about a given client in a way that leads to better understanding, decision-making, and treatment choice.

Clinical hypothesis:

- A proposition, or set of propositions, set forth as an explanation for the occurrence of some specified group of phenomena, either asserted merely as a provisional conjecture to guide investigation (working hypothesis) or accepted as highly probable in the light of established facts.

- A tentative explanation for an observation, phenomenon, or scientific problem that can be tested by further investigation.

- Something taken to be true for the purpose of argument or investigation; an assumption.
Population:

The set of all individuals of interest (e.g. all women, all college students)

Sample:

A subset of individuals selected from the population from whom data is collected
What we Learned from Probability

1) The mean of a sample can be treated as a random variable.

2) By the central limit theorem, sample means will have a normal distribution (for n > 30) with $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

3) Because of this, we can find the probability that a given population might randomly produce a particular range of sample means.

$$ P(\bar{X} > \text{something}) = P(Z > \text{something}) = \text{Use standard table} $$
Inferential Statistics

Population:
The set of all individuals of interest
(e.g. all women, all college students)

Sample:
A subset of individuals selected from
the population from whom data is collected
- Once we have got our sample
- The key question in statistical inference:
  - Could random chance alone have produced a sample like ours?
- Distinguishing between 2 interpretations of patterns in the data:
  - Random Causes: Fluctuations of chance
  - Systematic Causes Plus Random Causes:
    - True differences in the population
    - Bias in the design of the study
**Reasoning of Hypothesis Testing**

1. Make a statement (the null hypothesis) about some unknown population parameter.

2. Collect some data.

3. Assuming the null hypothesis is TRUE, what is the probability of obtaining data such as ours? (this is the “p-value”).

4. If this probability is small, then reject the null hypothesis.
State the research question in terms of a statistical hypothesis

- **Null hypothesis** (the hypothesis that is to be tested): abbreviated as $H_0$
  - Straw man: “Nothing interesting is happening”

- **Alternative hypothesis** (the hypothesis that in some sense contradicts the null hypothesis): abbreviated as $H_a$ or $H_1$
  - What a researcher thinks is happening
  - May be one- or two-sided
Hypothesis Testing: Step 1

- Hypotheses are in terms of population parameters

<table>
<thead>
<tr>
<th>One-sided</th>
<th>Two-sided</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0: \mu = 110$</td>
<td>$H_0: \mu = 110$</td>
</tr>
<tr>
<td>$H_{1/a}: \mu &lt; 110$</td>
<td>$H_{1/a}: \mu \neq 110$</td>
</tr>
<tr>
<td>OR</td>
<td>OR</td>
</tr>
<tr>
<td>$H_{1/a}: \mu &gt; 110$</td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis Testing: Step 2

- Set decision criterion:
  - Decide what p-value would be “too unlikely”
  - This threshold is called the alpha level.
  - When a sample statistic surpasses this level, the result is said to be significant.
  - Typical **alpha levels** are **0.05** and **0.01**.

- Alpha levels (level of significance) = probability of a type I error (the probability of rejecting the null hypothesis even that H₀ is true)

- The probability of a type II error is the probability of accepting the null hypothesis given that H₁ is true. The probability of a Type II error is usually denoted by β.
Setting the rejection region:

- The range of sample mean values that are “likely” if $H_0$ is true.

- If your sample mean is in this region, retain the null hypothesis.

- The range of sample mean values that are “unlikely” if $H_0$ is true.

- If your sample mean is in this region, reject the null hypothesis.
Hypothesis Testing: Step 3

Null hypothesis

Accept $H_0$

Reject $H_0$

$Z_{crit}$  $\mu_{H_0}$  $Z_{crit}$

29
Compute sample statistics

A test statistic (e.g. $Z_{\text{test}}$, $T_{\text{test}}$, or $F_{\text{test}}$) is information we get from the sample that we use to make the decision to reject or keep the null hypothesis.

A test statistic converts the original measurement (e.g. a sample mean) into units of the null distribution (e.g. a z-score), so that we can look up probabilities in a table.
If we want to know where our sample mean lies in the null distribution, we convert $X$-bar to our test statistic $Z_{test}$.

If an observed sample mean were lower than $z = -1.65$ then it would be in a critical region where it was more extreme than 95% of all sample means that might be drawn from that population.
State the test conclusion:

- If our sample mean turns out to be extremely unlikely under the null distribution, maybe we should revise our notion of $\mu_{Ho}$

- We never really “accept” the null. We either reject it, or fail to reject it.
Steps in Hypothesis Testing

1. Step 1: State hypothesis ($H_0$ and $H_1/H_a$)
2. Step 2: Choose significance level
3. Step 3: Setting the regression region
4. Step 4: Compute test statistic ($Z_{test}$) and get a $p$-value
5. Step 5: Make a decision
One- vs. two-tailed tests

- In theory, should use one-tailed when
  1. Change in opposite direction would be meaningless
  2. Change in opposite direction would be uninteresting
  3. No rival theory predicts change in opposite direction

- By convention/default in the social sciences, two-tailed is standard
- Why? Because it is a more stringent criterion (as we will see). A more conservative test.
One- vs. two-tailed tests

- $H_a$ is that $\mu$ is *either* greater or less than $\mu_{H_0}$
  - HA: $\mu \neq \mu_{H_0}$

- $\alpha$ is divided equally between the two tails of the critical region
Two-Tailed Hypothesis Testing

$H_0: \mu = 100$

$H_1: \mu \neq 100$

Reject $H_0$

Fail to reject $H_0$

Values that differ significantly from 100

$Z_{\text{crit}}$

$\alpha$

Values that differ significantly from 100
Values that differ "significantly" from 100

One tail

Reject $H_0$  
Fail to reject $H_0$

Two tail

Reject $H_0$  
Fail to reject $H_0$  
Reject $H_0$

Values that differ significantly from 100
Difference between P values and Confidence Intervals

- A P value measures the strength of evidence against the null hypothesis.
- A P value is the probability of getting a result as, or more, extreme if the null hypothesis were true.
- It is easy to compare results across studies using P values.
- P values are measures of **statistical significance**.
- Confidence intervals give a plausible range of values in clinically interpretable units.
- Confidence intervals enable easy assessment of **clinical significance**.
A general purpose approach to constructing confidence intervals is to define a $100(1-\alpha)\%$ confidence interval to consist of all those values $\theta_0$ for which a test of the hypothesis $\theta=\theta_0$ is not rejected at a significance level of $100\alpha\%$.

- Such an approach may not always be available since it presupposes the practical availability of an appropriate significance test.
- Naturally, any assumptions required for the significance test would carry over to the confidence intervals.
It may be convenient to make the general correspondence that parameter values within a confidence interval are equivalent to those values that would not be rejected by an hypothesis test, but this would be dangerous.

In many instances the confidence intervals that are quoted are only approximately valid, perhaps derived from "plus or minus twice the standard error", and the implications of this for the supposedly corresponding hypothesis tests are usually unknown.
Estimation of statistical parameters:
  - Confidence intervals for means
  - Confidence intervals for frequencies

Hypothesis testing:
  - Concept and Practice
Examples

- [http://www.biomedcentral.com/1746-6148/8/127](http://www.biomedcentral.com/1746-6148/8/127)
- *BMC Veterinary Research* 2012, 8:127 doi:10.1186/1746-6148-8-127

Background

Enzyme treatment is the mainstay for management of exocrine pancreatic insufficiency (EPI) in dogs. ‘Enteric-coated’ preparations have been developed to protect the enzyme from degradation in the stomach, but their efficacy has not been critically evaluated. The hypothesis of the current study was that enteric coating would have no effect on the efficacy of pancreatic enzyme treatment for dogs with EPI.

Thirty-eight client-owned dogs with naturally occurring EPI were included in this multicentre, blinded, randomised controlled trial. Dogs received either an enteric-coated enzyme preparation (test treatment) or an identical preparation without the enteric coating (control treatment) over a period of 56 days.
Examples

- http://www.biomedcentral.com/1746-6148/8/127
- BMC Veterinary Research 2012, 8:127 doi:10.1186/1746-6148-8-127

Results

There were no significant differences in either signalment or cobalamin status (where cobalamin deficient or not) between the dogs on the test and control treatments. Body weight and body condition score increased in both groups during the trial ($P<0.001$) but the magnitude of increase was greater for the test treatment compared with the control treatment ($P<0.001$). By day 56, mean body weight increase was 17% (95% confidence interval 11-23%) in the test treatment group and 9% (95% confidence interval 4-15%) in the control treatment group. The dose of enzyme required increased over time ($P<0.001$) but there was no significant difference between treatments at any time point ($P=0.225$). Clinical disease severity score decreased over time for both groups ($P=0.011$) and no difference was noted between groups ($P=0.869$). No significant adverse effects were reported, for either treatment, for the duration of the trial.
Probabilities by example

- [http://www.biomedcentral.com/1746-6148/8/68](http://www.biomedcentral.com/1746-6148/8/68)
- *BMC Veterinary Research* 2012, 8:68 doi:10.1186/1746-6148-8-68

**Results**

After optimising the cut-off values in order to avoid doubtful results without deteriorating the concordance between the results of the two tests, the I-ELISA appeared to be slightly more sensitive than CFT ($Se_{I-ELISA} = 0.917 [0.822; 0.992]$, 95% Credibility Interval (CrI)) compared to $Se_{CFT} = 0.860 [0.740; 0.967]$, 95% CrI). However, CFT was slightly more specific than I-ELISA ($Sp_{CFT} = 0.988 [0.947; 1.0]$, 95% CrI) compared to $Sp_{I-ELISA} = 0.952 [0.901; 1.0]$, 95% CrI).

The tests were then associated with two different interpretation schemes. The series association increased the specificity of screening and could be used for pre-movement testing in rams from uninfected flocks. The parallel association increased sequence sensitivity, thus appearing more suitable for eradicating the disease in infected flocks.
Sample Size and Power of a Test

How large a sample should be? Statistically speaking, the larger the the better!!!
Outline

- Concept & Power analysis
- Ways to choose sample sizes
- Effect of increasing the sample size
Clinical significance – Clinical relevance

- The method of estimating a minimum sample size is to specify the differences between means and the required changes of error and to find the sample size that satisfied these specifications.

Power Analysis

- A power analysis is used to reveal the minimum sample size which is required compared to the significance level and expected effects.
- Symbols: $1 - \beta$ (POWER) where $\beta =$ the chance of falsely accepting the null hypothesis
Choosing sample size

- Expedience:
  - items readily available or convenient to collect
  - small sample sizes $\rightarrow$ wide confidence intervals or risks of errors in statistical hypothesis testing.

- Using a target variance for an estimate to be derived from the sample eventually obtained (pilot study)

- Using a target for the power of a statistical test to be applied once the sample is collected.

- Using tables
Sample size estimation

- Significance level ($\alpha = \text{the chance of erroneously accepting the alternative hypothesis}$): $\alpha = 5\%$
- Power = 80% ($\beta = 20\%$)
- Clinical relevance ($d$ or $\delta$)
- One- or two-sided hypothesis:
  - Two-tailed: $z_{5\%} = 1.96$
  - One-tailed: $z_{5\%} = 1.645$
Sample size in mean testing (normal distribution)

- $H_0: m_s = \mu$ vs. $H_1: m_s \neq \mu$
- $\alpha = 5\%$
- $\beta = 20\% \rightarrow \text{Power} = 80\%$
- Critical value two-tailed test: $z_{1-5\%} = 1.960$
- $z_{1-\beta} = 0.842$

$$n = \frac{(z_{1-\alpha} - z_{1-\beta})^2 \sigma^2}{(m_s - \mu)^2}$$
Sample size estimation for a test of two means (normal distribution)

- $H_0: m_1 = m_2$ vs. $H_1: m_1 \neq m_2$
- $\alpha = 5\%$
- $\beta = 20\% \rightarrow \text{Power} = 80\%$

\[
n_1 = n_2 = \frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2 (\sigma_1^2 + \sigma_2^2)}{d^2}
\]
Sample size estimation for a test of two means

- Compare two artificial knees for range of motion (measured in degrees).
- **Hypothesis:** $H_0: m_1 = m_2$ vs. $H_1: m_1 \neq m_2$
- **1st type:** $m_1 = 112^\circ$ with $s_1 = 13^\circ$
- **2nd type:** $m_1 = 118^\circ$ with $s_1 = 11^\circ$

- If we want to perform a prospective, randomized clinical trial to decide whether a $6^\circ$ is statistically significant, what is the minimum number of patients receiving each knee we must record?
Sample size estimation for a test of two means

- If we want to perform a prospective, randomized clinical trial to decide whether a $6^\circ$ is statistically significant, what is the minimum number of patients receiving each knee we must record?

- $d = 112^\circ - 118^\circ = -6^\circ$
- $s_1^2 = 13^2 = 169 - s_2^2 = 11^2 = 121$
- $z_{1-2.5\%} = 1.95 - z_{1-\beta} = 0.842$

$$n_1 = n_2 = \frac{(1.95 + 0.842)^2(169 + 121)}{6^2} = 63.16$$
Suppose we could take a set of International Normalized Ratio (INR) readings from the clinic and another set from the laboratory (not paired data). We want to test, for a two-sided difference between means, when a difference of 0.25 INR is clinically relevant. We take $\alpha = 0.05$ and power = 0.90, $\sigma_{\text{clinic}} = 0.54$ INR and $\sigma_{\text{lab}} = 0.63$. How many readings are needed?

$$n_1 = n_2 = (1.96 + 1.28)^2 \times (0.54^2 + 0.63^2) / (0.25^2) = 115.6$$

→ A minimum of 116 readings in each group for a total of 232 readings are required.
Choose $k =$ the difference you want to detect between the sample mean and population mean (a clinical choice), expressed as the number of standard deviations of distance between them.

\[ n = \frac{\sigma^2}{\alpha \cdot k^2} \]
Sample size estimation for mean (nonnormal distribution)

- Suppose we want to know the required minimum sample size to find a difference between mean intraocular pressure (IOP in mmHg) between patients who have been treated with a new drug and those who have not. The standard deviation is 4 mmHg. We decide as a clinical judgment that we want to detect a 2 mm Hg decrease in IOP. We choose $\alpha = 0.05$ (we are willing to take a 5% risk of being wrong in our choice).
  - $\sigma = 4$ mmHg
  - $\alpha = 0.05$
  - $k = 2$ mmHg
  - $n = (4^2)/(0.05*2^2) = 80$
Sample size calculation: No objective Prior Data

- We have neither data not experience with a phenomenon being studied
  - From experience, guess the smallest and largest values of the variable of interest
  - Take the difference between them as a guess of the interval: mean ± 2 standard deviations (you assumed that the data are normal and your experience cover about 95%)
  - Estimate $\sigma = 0.25(\text{largest} - \text{smallest})$

\[
n = \frac{(z_{1-\alpha} - z_{1-\beta})^2 \sigma^2}{(m_s - \mu)^2}
\]
Effectiveness of an Herbal Remedy in Treating Colds

The wife of an internist has been treating her common colds with an herbal remedy for 3 years and claims it reduces the number of days to disappearance of symptoms. Her best time was 8 days, and the worst was 15 days. The internist decided to conduct a prospective, randomized, double-masked study to evaluate the remedy’s efficacy. How many data should he take? He decides that a reduction of $d = 1$ day would be clinically meaningful. He chooses $\alpha = 0.05$ and $\beta$ (power) = 80%.

- One-tailed z-values: $z_{1-\alpha} = 1.645$ and $z_{1-\beta} = 0.85$
- Estimate $\sigma = 0.25$ (largest - smallest) = $0.25(15-8) = 1.75$
- $n = ((1.645+0.85)^2 \times 1.75^2)/(1^2) = 18.91 \rightarrow 19$ data in each group (the experimental and the placebo) is the minimum
One proportion:

- Tested against a theoretical proportion or a previously established proportion
  - Central proportion (is not near 0 or 1): binomial distribution
  - Extreme proportion (is near 0 or 1): Poisson distribution
  - Moderately sample size: normal distribution
- \( \pi = \text{theoretical proportion} \)
- \( p = \text{desired proportion} \)

\[
n = \left[ \frac{z_{1-\alpha/2} \sqrt{\pi (1-\pi)}}{p - \pi} + z_{1-\beta} \sqrt{p (1-p)} \right]^2
\]

\[
n = \left[ \frac{z_{1-\alpha/2} \sqrt{\pi} + z_{1-\beta} \sqrt{p}}{p - \pi} \right]^2
\]
Sample size for tests on rates (proportions)

- **One proportion**
- We want to compare the positive result biopsy rate (30%) from a sample of 10 patients with a theoretical positive biopsy result rate defined as 25% of patients older than 50 years presenting a urology clinic.
- **Is the sample of 10 sufficient?**
- $\pi = 0.25$ & $p = 0.30$ & $\alpha = 0.05$ and $\beta$ (power) = 80%.
- One-tailed z-values: $z_{1-\alpha} = 1.645$ and $z_{1-\beta} = 0.85$
- $n = ((1.645\sqrt{(0.025*0.075)}+0.84 \sqrt{0.30*0.70})/0.05)^2 \sim 482$
- $\rightarrow$ at least 482 biopsies are required
Two proportions

- We assume normal theory
- Samples proportions $p_1$ and $p_2$
- Proportion mean: $p_m = (p_1 + p_2) / 2$
- If $p_m$ is central (not near 0 or 1):
  \[
  n_1 = n_2 = \left[ \frac{z_{1-\alpha/2}\sqrt{2p_m(1-p_m)} + z_{1-\beta}\sqrt{p_1(1-p_1) + p_2(1-p_2)}}{p_1 - p_2} \right]^2
  \]
- If $p_m$ is extreme (near 0 or 1):
  \[
  n_1 = n_2 = \left[ \frac{(z_{1-\alpha/2} + z_{1-\beta})\sqrt{p_1 + p_2}}{p_1 - p_2} \right]^2
  \]
Two proportions

A psychiatrist wants to know if the proportion of people having a personality disorder is the same for those committing violent crimes ($p_1$) and those committing non-violent crimes ($p_2$). She examines a few of her past records to serve as a pilot study survey and estimates $p_1$ as 0.06 and $p_2$ as 0.02. How many patients are needed to detect a difference of 0.04, significant at two-tailed $\alpha = 0.05$ and power = 0.08?
Sample size for tests on rates (proportions)

- Two proportions
  - \( p_1 = 0.06 \) & \( p_2 = 0.02 \) & \( \alpha = 0.05 \) & power = 0.08
  - \( z_{1-\alpha} = 1.96 \) & \( z_{1-\beta} = 0.84 \)
  - \( p_m = (0.06+0.02)/2 = 0.04 \)
  - \( n_1 = n_2 = \{[(1.96+0.84)\times\sqrt{(0.06+0.04)}]/0.05\}^2 \)
  - \( n_1 = n_2 = 392 \)
  - → the psychiatrist will need a minimum of 392 patients in each group
Factors that affect sample size calculation

<table>
<thead>
<tr>
<th>Factor</th>
<th>Magnitude</th>
<th>Impact on identification of effect</th>
<th>Required sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$ value</td>
<td>Small</td>
<td>Stringent criterion; difficult to achieve ‘significance’</td>
<td>Large</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>Relaxed criterion; ‘significance’ easier to attain</td>
<td>Small</td>
</tr>
<tr>
<td>Power</td>
<td>Low</td>
<td>Identification unlikely</td>
<td>Small</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>Identification more probable</td>
<td>Large</td>
</tr>
<tr>
<td>Effect</td>
<td>Small</td>
<td>Difficult to identify</td>
<td>Large</td>
</tr>
<tr>
<td></td>
<td>Large</td>
<td>Easy to identify</td>
<td>Small</td>
</tr>
</tbody>
</table>
The probability that the test will reject the null hypothesis when the null hypothesis is actually false (i.e. the probability of not committing a Type II error, or making a false negative decision)

As the power increases, the chances of a Type II error occurring decrease.

The probability of a Type II error occurring is referred to as the false negative rate (abbreviated as $\beta$). Therefore power is equal to $1 - \beta$, which is also known as the *sensitivity*. 
Factors influencing power:

- statistical significance
  - 0.05 (5%, 1 in 20)
  - 0.01 (1%, 1 in 100)
  - 0.001 (0.1%, 1 in 1000).

- magnitude of the effect of interest in the population
  - As the difference in outcomes means of two samples

- sample size used to detect the effect
  - determines the amount of sampling error inherent in a test result
The assumption of normality could be tested

Normality is needed in order to correctly apply some summaries or tests (e.g. mean and standard deviation/error, test to compare means, Pearson correlation coefficient, regression analysis, etc.)

The minimum sample size needed to answer to a research question could be calculated based on some previously known data or assumptions.