Design and Analysis of Experiments

02 - Inference about the difference in means

Ahmed A.(Msc)

DIRE DAWA UNIVERSITY
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES
DEPARTMENT OF SATATISTICS

October 1, 2024

Outline

1. **Inference about the difference in means,Randomized Designs**
   - Hypothesis Testing
   - Confidence Intervals
   - Choice of Sample Size
   - The Case Where $\sigma_1 \neq \sigma_2$
2. **Inferences About the Differences in Means, Paired Comparison Designs**

Inference about the difference in means,Randomized Designs

- The concepts of comparison between two populations based on information obtained from their samples follow the same principles used for testing hypotheses about a single population;
- eatment) against a *control group*: placebo, classical technique, random search, etc;

Usual questions involve:

- Comparison of means;
- Comparison of variances;
- Comparison of proportions;

Inference about the difference in means

- let see the Portland cement mortar problem.
- Recall that two different formulations of mortar were being investigated to determine if they differ in tension bond strength.
- In this section, we discuss how the data from this simple comparative experiment can be analyzed using **hypothesis testing** and **confidence interval** procedures for comparing two treatment means.

Assumptions:

- Completely randomized experimental design.
- Data is viewed as a random sample from a normal distribution.

Inference about the difference in means

■ **TABLE 2.1**

Tension Bond Strength Data for the Portland Cement Formulation Experiment

j	Modified Mortar y_{1j}	Unmodified Mortar y_{2j}
1	16.85	16.62
2	16.40	16.75
3	17.21	17.37
4	16.35	17.12
5	16.52	16.98
6	17.04	16.87
7	16.96	17.34
8	17.15	17.02
9	16.59	17.08
10	16.57	17.27

Hypothesis Testing

- Now consider the Portland cement experiment
- Recall that we are interested in comparing the strength of two different formulations: an unmodified mortar and a modified mortar.
- In general, we can think of these two formulations as two levels of the factor **"formulations."**
- Let $y_{11}, y_{12}, \ldots, y_{1n_1}$ represent the n_1 observations from the first factor level, and $y_{21}, y_{22}, \ldots, y_{2n_2}$ represent the n_2 observations from the second factor level.
- We assume that the samples are drawn at random from two independent normal populations.

Hypothesis Testing

A possible statistical model for this kind of data would be:

$$y_{ij} = \mu_i + \epsilon_{ij} \begin{cases} i = 1, 2 \\ j = 1, \ldots, n_i \end{cases}$$

Lets initially assume that the residuals ϵ_{ij} are iid $\mathcal{N}\left(0, \sigma_i^2\right)$, which implies:

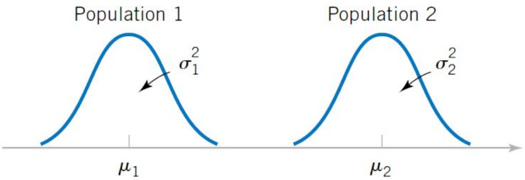


Figure: Image: D.C.Montgomery,G.C. Runger, *Applied Statistics and Probability for Engineers*,Wiley 2003.

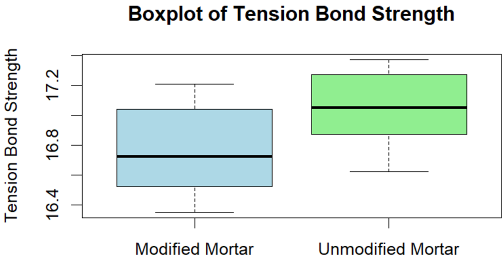


Figure: Box Plot.

Statistical Hypotheses

- A **statistical hypothesis** is a statement either about the parameters of a probability distribution or the parameters of a model.
- The hypothesis reflects some conjecture about the problem situation. For example, in the Portland cement experiment, we may think that the mean tension bond strengths of the two mortar formulations are equal.
- This may be stated formally as The statistical hypotheses can be stated as:

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases} \quad \textbf{or, equivalently,} \quad \begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

Hypothesis Testing

- where
  μ_1 is the mean tension bond strength of the modified mortar and
  μ_2 is the mean tension bond strength of the unmodified mortar.
- The statement
  H0:$\mu_1 = \mu_2$ is called the null hypothesis and
  H1:$\mu_1 \neq \mu_2$ is called the alternative hypothesis.
- The alternative hypothesis specified here is called a two-sided alternative hypothesis because it would be true if
  $\mu_1 < \mu_2$ or if $\mu_1 > \mu_2$.

Testing the Hypothesis

To test a hypothesis, we advise a procedure for:

- Taking a random sample,
- Computing an appropriate test statistic,
- Then rejecting or failing to reject the null hypothesis H_0 based on the computed value of the test statistic.

Part of this procedure involves specifying the set of values for the test statistic that lead to rejection of H_0. This set is called the **critical region** or **rejection region** for the test.

Errors in Hypothesis Testing

Two kinds of errors may be committed when testing hypotheses:

- If the null hypothesis is rejected when it is true, a **Type I error** has occurred.
- If the null hypothesis is not rejected when it is false, a **Type II error** has been made.

The probabilities of these two errors are given special symbols:

- $\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0 \text{ is true})$
- $\beta = P(\text{Type II error}) = P(\text{Fail to reject } H_0 \mid H_0 \text{ is false})$
- The general procedure in hypothesis testing is to specify a value of the probability of type I error α, often called the **significance level** of the test, and then design the test procedure so that the probability of type II error β has a suitably small value.

The Two-Sample t-Test

Suppose that we could assume that the variances of tension bond strengths were identical for both mortar formulations. Then the appropriate test statistic to use for comparing two treatment means in the completely randomized design is

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (1)$$

where $\bar{y}_1$ and $\bar{y}_2$ are the sample means, n_1 and n_2 are the sample sizes, S_p^2 is an estimate of the common variance $\sigma_1^2 = \sigma_2^2 = \sigma^2$ computed from

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (2)$$

Two-Sample t-Test (cont.)

- S_1^2 and S_2^2 are the two individual sample variances.
- The quality $S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ in the denominator of Equation 1 is often called the **standard error** of the difference in means in the numerator, abbreviated $se(\bar{y}_1 - \bar{y}_2)$.
- To determine whether to reject $H_0 : \mu_1 = \mu_2$, we would compare t_0 to the t distribution with $n_1 + n_2 - 2$ degrees of freedom.
- If $|t_0| > t_{\alpha/2, n_1+n_2-2}$, where $t_{\alpha/2, n_1+n_2-2}$ is the upper $\alpha/2$ percentage point of the t distribution with $n_1 + n_2 - 2$ degrees of freedom,
- we would reject H_0 and conclude that the mean strengths of the two formulations of Portland cement mortar differ. This test procedure is usually called the **two-sample t-test**.

Justification of the Two-Sample t-Test

This procedure may be justified as follows. If we are sampling from independent normal distributions, then the distribution of $\bar{y}_1 - \bar{y}_2$ is $N(\mu_1 - \mu_2, \sigma^2(1/n_1 + 1/n_2))$.
Thus, if σ^2 were known, and if $H_0 : \mu_1 = \mu_2$ were true, the distribution of

$$Z_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

would follow a standard normal distribution.

One-Sided Hypothesis Testing

- In some problems, one may wish to reject H_0 only if one mean is larger than the other.
- Thus, one would specify a one-sided alternative hypothesis $H_1 : \mu_1 < \mu_2$ and would reject H_0 only if $t_0 < t_{\alpha, n_1+n_2-2}$.
- If one wants to reject H_0 only if μ_1 is less than μ_2, then the alternative hypothesis is $H_1 : \mu_1 > \mu_2$, and one would reject H_0 if $t_0 > t_{\alpha, n_1+n_2-2}$.

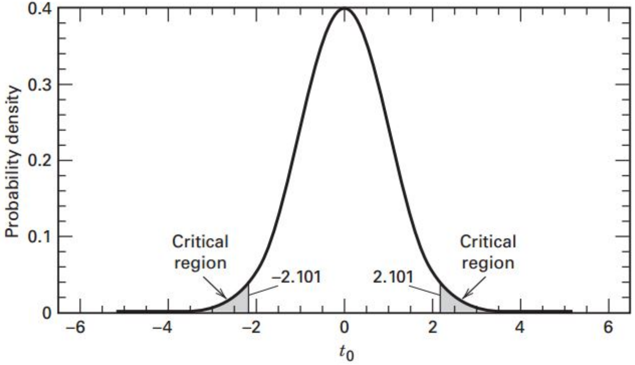
Portland Cement Data Example

To illustrate the procedure, consider the Portland cement data in Table 2.1. For these data, we find that:

Modified Mortar	**Unmodified Mortar**
$\bar{y}_1 = 16.76$ kgf/cm^2	$\bar{y}_2 = 17.04$ kgf/cm^2
$S_1 = 0.316$	$S_2 = 0.248$
$n_1 = 10$	$n_2 = 10$

The test statistic t_0 can be computed for these data, and based on the critical value of the t distribution for $n_1 + n_2 - 2$ degrees of freedom, we can make a decision to either reject or fail to reject H_0.

Portland Cement Data Example



■ **FIGURE 2.10** The t distribution with 18 degrees of freedom with the critical region $\pm t_{0.025,18} = \pm 2.101$

Solution

Furthermore, $n_1 + n_2 - 2 = 10 + 10 - 2 = 18$, and if we choose $\alpha = 0.05$, then we would reject $H_0 : \mu_1 = \mu_2$ if the numerical value of the test statistic $t_0 < -t_{0.025,18} = -2.101$, or if $t_0 > t_{0.025,18} = 2.101$.
These boundaries of the critical region are shown on the reference distribution (a t-distribution with 18 degrees of freedom).

Test Statistic Calculation

Using Equation (2), we find that the pooled variance S_p^2 is given by:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Substitute the values:

$$S_p^2 = \frac{9(0.100) + 9(0.061)}{10 + 10 - 2} = 0.081$$

Therefore, the pooled standard deviation is:

$$S_p = \sqrt{0.081} = 0.284$$

Final Test Statistic

Now, calculate the test statistic t_0:

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Substitute the values:

$$t_0 = \frac{16.76 - 17.04}{0.284\sqrt{\frac{1}{10} + \frac{1}{10}}}$$

$$t_0 = \frac{-0.28}{0.127} = -2.20$$

Conclusion

Since $t_0 = -2.20 < -t_{0.025,18} = -2.101$, we reject H_0 and conclude that the mean tension bond strengths of the two formulations of Portland cement mortar are different. This is an important engineering finding.

The Use of P-Values in Hypothesis Testing

One way to report the results of a hypothesis test is to state whether the null hypothesis was rejected at a specified α-value or level of significance. This is called fixed significance level testing.

For example, in the Portland cement mortar formulation test above, we can say that $H_0 : \mu_1 = \mu_2$ was rejected at the 0.05 level of significance.

Limitations of Fixed Significance Level Testing

- This method gives no indication if the test statistic was just barely in the rejection region or very far into it. - It imposes a predefined level of significance on other decision makers, which may not be suitable for everyone. Some might prefer a significance level different from $\alpha = 0.05$.

The P-Value Approach

To avoid these issues, the P-value approach is often used.
Definition: The P-value is the probability that the test statistic will take a value at least as extreme as the observed value, assuming the null hypothesis H_0 is true.
Interpretation: The P-value conveys the weight of evidence against H_0, and allows decision makers to make a conclusion at any specified level of significance.

P-Value in the Portland Cement Example

In the Portland cement experiment:
- Test statistic: $t_0 = -2.20$ - From the t-distribution with 18 degrees of freedom, we know $t_{0.025,18} = -2.101$, meaning the P-value is less than 0.05. - From Appendix Table II, $t_{0.01,18} = -2.552$, so the P-value is greater than 0.01. - Therefore, the P-value must be between 0.02 and 0.05.

Exact P-Value Calculation

Using a handheld calculator like the HP-48, we can calculate the exact P-value for $t_0 = -2.20$ in the Portland cement experiment as:

$$P = 0.0411$$

Thus, we would reject the null hypothesis $H_0 : \mu_1 = \mu_2$ at any significance level $\alpha < 0.0411$.

Conclusion

The P-value provides flexibility, allowing decision makers to assess the significance of the results at any level of significance, instead of being constrained by a fixed α.
In this case, since $P = 0.0411$, the null hypothesis would be rejected at the 0.05 significance level but not at the 0.01 level.

R Code for Two-Sample T-Test

The R code used for the analysis is shown below:

```
tension_bond_data <- data.frame(
j = 1:10,
Modified_Mortar = c(16.85, 16.40, 17.21, 16.35, 16.52, 17.04,

16.96, 17.15, 16.59, 16.57),

Unmodified_Mortar = c(16.62, 16.75, 17.37, 17.12, 16.98,

16.87, 17.34, 17.02, 17.08, 17.27)
)
t_test_result <- t.test(tension_bond_data$Modified_Mortar,

tension_bond_data$Unmodified_Mortar,var.equal = TRUE)
print(t_test_result)
```

Two Sample t-test with equal variance

```
        Two Sample t-test

data:  tension_bond_data$Modified_Mortar and
tension_bond_data$Unmodified_Mortar
t = -2.1869, df = 18,
p-value = 0.0422
alternative hypothesis: true difference in m
eans is not equal to 0
95 percent confidence interval:
 -0.54507339 -0.01092661
sample estimates:
mean of x mean of y
   16.764    17.042
```

Checking Assumptions in the t-Test

In using the t-test procedure, we make the following assumptions:

- Both samples are random samples drawn from independent populations.
- These populations can be described by a normal distribution.
- The standard deviation or variances of both populations are equal.

Assumption of Independence

The assumption of independence is critical. This assumption will usually be satisfied if:

- The run order is randomized.
- Other experimental units and materials are selected at random (if appropriate).

Assumptions of Equal Variance and Normality

The assumptions of equal variance and normality can be checked using a **normal probability plot**. These assumptions are easy to verify with most statistics software packages.

Conclusion

The t-test requires assumptions of normality, equal variance, and independence of samples. Probability plots and other graphical techniques help in verifying these assumptions. Always use appropriate tools to ensure these assumptions are met before relying on the results of the t-test.

Introduction

In this presentation, we will check the assumptions for conducting a t-test:

- Normality of the data using the Shapiro-Wilk test.
- Visual inspection with Q-Q plots.
- Equality of variances using Levene's and Bartlett's tests.

R Code for Checking Assumptions

this R code used to check the assumptions for the t-test:

```
[language=R]
# Load necessary libraries
library(car)

# Combine data into long format for easier plotting and analys
long_data <- stack(tension_bond_data)
```

Normality Test

```
[language=R]
# 1. Check normality using the Shapiro-Wilk test
shapiro_test_modified <- shapiro.test(tension_bond_data$Modifi
shapiro_test_unmodified <- shapiro.test(tension_bond_data$Unmo
cat("Shapiro-Wilk Test for Normality:\n")
print(shapiro_test_modified)
print(shapiro_test_unmodified)
```

Q-Q Plots for Visual Inspection

```
[language=R]
# Plot Q-Q plots for both samples
par(mfrow = c(1, 2))  # Plot side by side
qqnorm(tension_bond_data$Modified_Mortar, main = "Q-Q Plot: Mo
qqline(tension_bond_data$Modified_Mortar, col = "blue")

qqnorm(tension_bond_data$Unmodified_Mortar, main = "Q-Q Plot:
qqline(tension_bond_data$Unmodified_Mortar, col = "blue")
```

Variance Tests

```
[language=R]
# 2. Check equal variance using Levene's Test
levene_test <- leveneTest(values ~ ind, data = long_data)
cat("Levene's Test for Equal Variance:\n")
print(levene_test)

# 3. Perform Bartlett's Test for Equal Variance
bartlett_test <- bartlett.test(values ~ ind, data = long_data)
cat("Bartlett's Test for Equal Variance:\n")
print(bartlett_test)
```

Output Interpretation

- **Shapiro-Wilk Test**: Checks for normality. A p-value greater than 0.05 suggests the data is normally distributed.
- **Q-Q Plots**: Provides a visual inspection of the data's normality.
- **Levene's and Bartlett's Tests**: Both tests check for equality of variances. A p-value greater than 0.05 indicates equal variances.

Confidence Intervals

- Hypothesis testing is useful but may not tell the entire story
- Confidence intervals provide a range for parameter values
- Often more informative in engineering and industrial experiments
- Especially useful when means are known to differ

Definition of Confidence Intervals

- For an unknown parameter θ, find statistics L and U such that:

$$P(L \leq \theta \leq U) = 1 - \alpha$$

- The interval $[L, U]$ is called a 100(1 - α)% confidence interval for θ
- L and U are lower and upper confidence limits
- $1 - \alpha$ is the confidence coefficient

Interpretation of Confidence Intervals

- In repeated sampling, 100(1 - α)% of intervals will contain the true value
- For $\alpha = 0.05$, we have a 95% confidence interval
- Frequency interpretation: method is correct 100(1 - α)% of the time
- Not a statement about a specific sample

Example: Difference in Means

Consider finding a 100(1 - α)% confidence interval on $\mu_1 - \mu_2$:

- The statistic:

$$\frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

  is distributed as $t_{n_1+n_2-2}$

- This forms the basis for constructing the confidence interval

Confidence Interval for $\mu_1 - \mu_2$

$$P\left(-t_{\alpha/2,n_1+n_2-2} \leq \frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \leq t_{\alpha/2,n_1+n_2-2}\right) = 1-\alpha$$

1. let $\mu_1 - \mu_2 = \gamma$
2. $t_o = t_{\alpha/2,n_1+n_2-2}$

$$P\left(\bar{y}_1 - \bar{y}_2 - t_o S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \leq \gamma \leq \bar{y}_1 - \bar{y}_2 + t_o S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right) = 1-\alpha$$

Confidence Interval Derivation

$$\bar{y}_1 - \bar{y}_2 - t_o S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \quad \leq \bar{y}_1 - \bar{y}_2 + t_o S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Interpreting the Confidence Interval

- Lower bound: $\bar{y}_1 - \bar{y}_2 - t_o S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- Upper bound: $\bar{y}_1 - \bar{y}_2 + t_o S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- Where:
  - $\bar{y}_1, \bar{y}_2$: sample means
  - $t_{\alpha/2, n_1+n_2-2}$: critical value from t-distribution
  - S_p: pooled standard deviation
  - n_1, n_2: sample sizes
- This interval estimates $\mu_1 - \mu_2$ with $(1-\alpha)$ confidence level
- Find the CI for the Portland Cement Example

Importance of Sample Size

- Selection of an appropriate sample size is crucial in experimental design
- It impacts the estimate of the difference in two means
- We'll consider the effect on the confidence interval for this difference

Confidence Interval and Precision

- The 100(1 - α)% confidence interval on the difference in two means measures the precision of estimation
- From Equation 2.30, the length of this interval is determined by:

$$t_{\alpha/2,n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- Where:
  - $t_{\alpha/2,n_1+n_2-2}$ is the critical value from t-distribution
  - S_p is the pooled standard deviation
  - n_1 and n_2 are sample sizes from the two populations

Equal Sample Sizes Case

- Consider the case where sample sizes from both populations are equal:

$$n_1 = n_2 = n$$

- In this case, the length of the CI is determined by:

$$t_{\alpha/2,2n-2} S_p \sqrt{\frac{2}{n}}$$

Impact of Sample Size

- Larger sample sizes generally lead to:
  - Narrower confidence intervals
  - More precise estimates
  - Increased statistical power
- However, there are trade-offs:
  - Cost and time constraints
  - Diminishing returns as sample size increases
- Careful consideration of sample size is essential for balancing precision and practicality

Hypothesis Testing, The Case Where $\sigma_1 \neq \sigma_2$

When testing:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

If we cannot reasonably assume that the variances σ_1^2 and σ_2^2 are equal, the two-sample t-test must be modified.

Modified Test Statistic

The test statistic becomes:

- $$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{3}$$
- This is known as Welch's t-statistic.

Distribution of t_0

- This statistic is not distributed exactly as t.
- However, the distribution of t_0 is well approximated by t if we use an adjusted number of degrees of freedom.

Degrees of Freedom

The degrees of freedom v is calculated as:

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}} \tag{4}$$

This is known as the Welch–Satterthwaite equation.

When to Use This Test

- Use when there's a strong indication of unequal variances on a normal probability plot.
- This version of the t-test is more robust when sample sizes or variances are unequal.
- It's generally recommended to use this test by default, as it performs well even when variances are equal.

Confidence Interval

- You should be able to develop an equation for finding the confidence interval on the difference in means for the unequal variances case.
- Hint: Use the t-statistic and degrees of freedom from this modified test in the general form of a confidence interval.

The Case Where $\sigma_1 \neq \sigma_2$

The R code used for the analysis is shown below:

```
tension_bond_data <- data.frame(
j = 1:10,
Modified_Mortar = c(16.85, 16.40, 17.21, 16.35, 16.52, 17.04,

16.96, 17.15, 16.59, 16.57),

Unmodified_Mortar = c(16.62, 16.75, 17.37, 17.12, 16.98,

16.87, 17.34, 17.02, 17.08, 17.27)
)

t_test_result <- t.test(tension_bond_data$Modified_Mortar,

tension_bond_data$Unmodified_Mortar)
print(t_test_result)
```

Welch Two Sample t-test

```
        Welch Two Sample t-test

data:  tension_bond_data$Modified_Mortar and
tension_bond_data$Unmodified_Mortar
t = -2.1869, df = 17.025,
p-value = 0.043
alternative hypothesis: true difference in m
eans is not equal to 0
95 percent confidence interval:
 -0.546174139 -0.009825861
sample estimates:
mean of x mean of y
   16.764    17.042
```

Reading Assignment

1. Read Inferences About the Differences in Means, Paired Comparison Designs
2. what is the problems of Paired Comparison Design
3. Advantages of the Paired Comparison Design
4. Inferences About the Variances of Normal Distributions

Bibliography

Required reading

- D.C. Montgomery, G.C. Runger, *Applied Statistics and Probability for Engineers*, Ch. 10. 5th ed., Wiley, 2010.; **OR**
- D.C. Montgomery, *Design and Analysis of Experiments*, Ch. 2. 5th ed., Wiley, 2005;
- F. Campelo, F. Takahashi, *Sample size estimation for power and accuracy in the experimental comparison of algorithms*. J. Heuristics, 2018 - `https://goo.gl/TrC2x3`

Recommended reading

- P. Mathews, *Sample Size Calculations: Practical Methods for Engineers and Scientists*, Ch. 1-2, 1st ed., MMB, 2010.
- R. Nuzzo, *Scientific method: Statistical errors*, Nature 506(7487) - `http://goo.gl/Kbq6Rc`