



Fundamentals of Quantitative Biology

STATS 03: Hypothesis Testing

Winter semester 2023/2024

Statistical Inference

What is Statistical Inference?

- ▶ **Descriptive Statistics** summarize data and identifies relationships and patterns.
- ▶ **Interferential Statistics** allow us to go beyond pure description and make conclusions to the population.
- ▶ **Statistical Inference** is the process of using data from a sample to learn about a larger population or process.
- ▶ Inference allows us to:
 1. Estimate **unknown quantities** (called *parameters*) that describe the population.
 2. Test **hypotheses** to verify claims about the population or process.

Statistical Inference

► What is a parameter?

- A parameter is an unknown quantity that describes some aspect of a population or a process.
- Examples:
 - The proportion of cells expressing a specific gene under stress conditions.
 - The rate of mutation in a gene across generations.
 - Air temperature from rocket sound delay at New Year's.

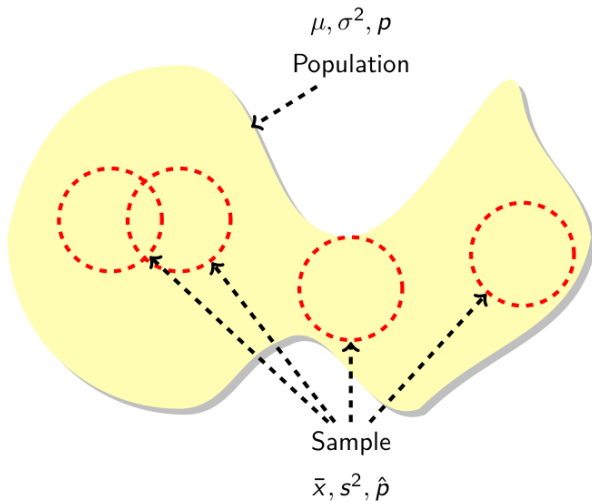
Goals of Statistical Inference

We observe a sample and want to draw conclusions about the parameter θ based on the sample.

1. Estimation: approximate determination of the parameter θ
2. Hypothesis testing: verifying claims about the parameter θ

Populations and samples

Population parameters typically not observable



Populations and samples

The *population* is the entire set of quantities we are interested in.

Population → parameters

Numerical measures of this population are called parameters. Parameters are fixed but unknown.

Inference about the parameters is made by taking a subset of the population: a *sample*.

Sample → statistics

Numerical measures of the sample are called statistics. Statistics can vary from sample to sample but are known.

Statistical modeling

- ▶ We repeat n times independently the experiment \mathcal{E} .
- ▶ The sample variable X_i is a random variable that, with its distribution, describes the probability that a specific characteristic occurs in the i th drawing from the population.
- ▶ x_i is a realization of X_i .

Examples: sample mean and sample variance

Consider data

$$x_1, x_2, \dots, x_n.$$

- ▶ The sample mean is defined as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

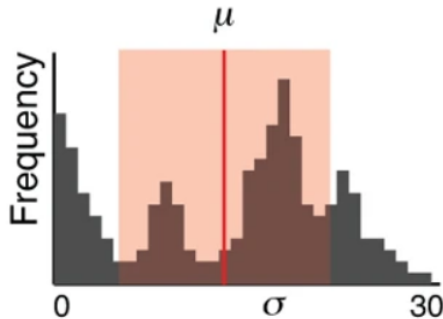
- ▶ The sample variance is defined as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- ▶ \bar{x} and s^2 describe the distribution of data the same way as μ and σ^2 describe probability distributions.

Population parameters are estimated by sampling

True distribution of values between (0, 30)
in a population, with mean μ and standard deviation σ



Nature Methods (Nat Methods) ISSN 1548-7105 (online) ISSN 1548-7091 (print)

Random samples taken from the population

Samples

$$X_1 = [1, 19, 17, 20, 26]$$

$$X_2 = [8, 11, 16, 24, 25]$$

$$X_3 = [16, 17, 18, 20, 24]$$

...

$$\xrightarrow{\bar{X}}$$

Sample means

$$\bar{X}_1 = 14.6$$

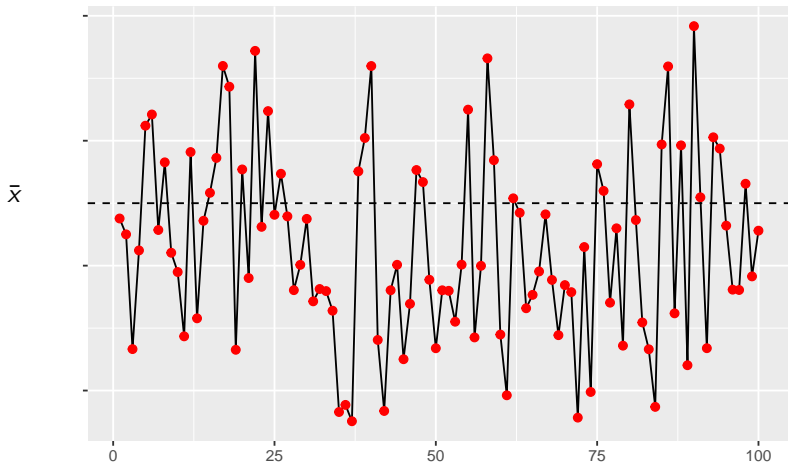
$$\bar{X}_2 = 16.8$$

$$\bar{X}_3 = 19.0$$

...

Example

Repeat 100 times.



#

Sampling distribution for \bar{X}

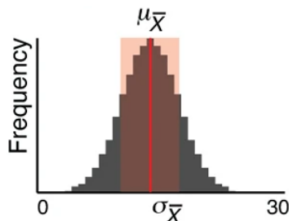
Let X_1, X_2, \dots, X_n be independent and identically distributed random variables with $\mathbb{E}[X_i] = \mu < \infty$ and $\text{Var}(X_i) = \sigma^2 < \infty$.

Let \bar{X} be the sample mean calculated by

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

then the sampling distribution of \bar{X} is

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$



The standard error of \bar{X} is the square-root of $\text{Var}(\bar{X})$: $\text{SE}(\bar{X}) = \sigma/\sqrt{n}$.

The central limit theorem

Theorem

Let Y_1, Y_2, \dots, Y_n be independent and identically distributed random variables with $\mathbb{E}[Y_i] = \mu < \infty$ and $\text{Var}(Y_i) = \sigma^2 < \infty$. Define

$$U_n = \frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}, \quad \text{where } \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

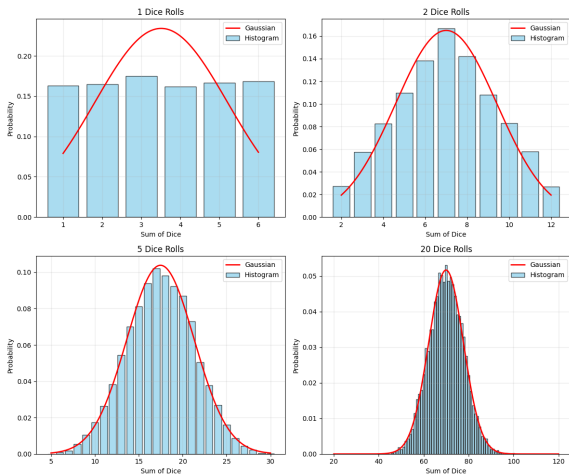
Then the distribution of U_n converges to the standard normal distribution function as $n \rightarrow \infty$. That is,

$$\lim_{n \rightarrow \infty} \mathbb{P}(U_n \leq u) = \int_{-\infty}^u e^{-t^2/2} dt \quad \text{for all } u.$$

3Blue1Brown: <https://youtu.be/zeJD6dqJ51o>

The central limit theorem

Example: Rolling n dice and sum up the result. Repeat $m = 1000$ times.



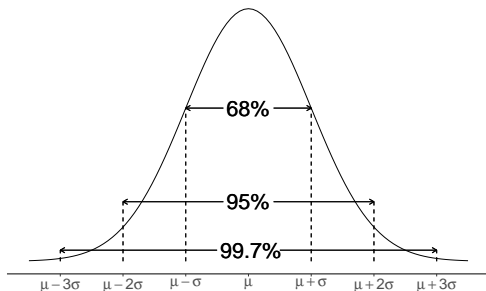
The central limit theorem

- ▶ **Universality Across Distributions:** Regardless of the original distribution, the sum or average of independent random variables tends toward a normal distribution.
- ▶ **Foundation of Inferential Statistics:** Enables confidence intervals, hypothesis testing, and regression by justifying the normality of sampling distributions.
- ▶ **Predictability in Real-World Processes:** Explains why natural phenomena (e.g., heights, measurement errors) often produce normal-like outcomes.
- ▶ **Simplifies Complex Problems:** Allows approximation of probabilities and outcomes using the normal distribution, even in non-normal scenarios.
- ▶ **Practical Applicability:** Works well with finite sample sizes, enabling predictions in polling, surveys, and experiments.

Confidence interval

Confidence interval

- ▶ When estimating a parameter θ uncertainties regarding the accuracy of θ occur, because of uncertainties of the sample, e.g. due to (finite) sampling.
- ▶ These uncertainties are expressed by a confidence interval $[L, U]$, which includes the unknown parameter θ of interest with probability $1 - \alpha$.
- ▶ α is the so-called significance level.



Interval Estimates

- ▶ **average over multiple samples**

- ▶ The sample mean \bar{X} follows a normal distribution:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right),$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

- ▶ **confidence interval derivation:**

$$1 - \alpha = \mathbb{P}\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right).$$

Rearranging gives:

$$\mathbb{P}\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Interval estimates

Thus, the confidence interval for μ is:

$$\mu \in \left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right].$$

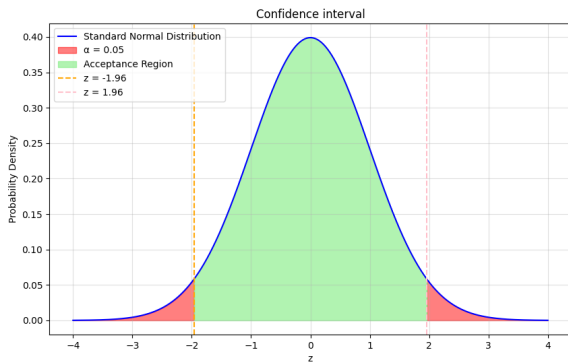
Statement of confidence interval for the mean

We are 95% confident that the mean of <variable> from <population> lies between <lower> and <upper> <units>.

Illustration of CIs

For $1 - \alpha = .95$ we get $z_{\pm} = \pm 1.96$ again by using

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{\alpha}{2}.$$



Hypothesis test

Introduction to Hypothesis Testing

What is Hypothesis Testing?

We test assumptions about data by evaluating whether the observed results are likely under a given hypothesis.

- ▶ **Rolling a Die:** I rolled a fair-looking die 20 times and got the following sequence: 1, 2, 6, 4, 6, 6, 3, 6, 5, 6, 2, 6, 6, 4, 6, 6, 5, 6, 1, 6. *How likely is it that the die is biased towards 6?*
- ▶ **Drug Effectiveness:** A new medication was tested on 100 patients. 65 showed improvement. *Is the drug significantly better than random chance?*
- ▶ **Coin Toss:** A coin was tossed 50 times, and it landed heads 40 times. *Is the coin fair, or is it biased?*
- ▶ **Sports Performance:** A basketball player makes 8 successful shots out of 10. *Is the player performing better than average, or could this be due to luck?*
- ▶ **Website Click-Through Rates:** Two versions of a website were tested. Version A had a click rate of 12%, and Version B had 15%. *Is the difference in performance statistically significant?*

Key Question: How likely are the observed results under the assumption that the null hypothesis is true?

The Null Hypothesis

What is the Null Hypothesis (H_0)?

The Null Hypothesis is a starting assumption in hypothesis testing that there is **no effect**, **no difference**, or that things happen by **random chance**.

Examples of Null Hypotheses:

- ▶ Rolling a Die: The die is fair, with each side equally likely ($p = 1/6$).
- ▶ Drug Effectiveness: The drug has no effect ($p = 0.5$, same as random chance).
- ▶ Coin Toss: The coin is fair ($p = 0.5$).
- ▶ Sports Performance: The basketball player's success rate matches the league average.
- ▶ Website Click-Through Rates: The click rates of Version A and Version B are identical.

Key Idea: Hypothesis testing evaluates how likely the observed data is under H_0 . If unlikely, we reject H_0 .

Hypothesis Testing: Discrete vs. Continuous Variables

Discrete (mutation count, expression level,...)

- ▶ Binomial Test
- ▶ Chi-Square Test
- ▶ Fisher's Exact Test

Continuous (protein concentration, enzyme activity, ...)

- ▶ t-Test
- ▶ ANOVA
- ▶ Z-Test
- ▶ Kolmogorov-Smirnov Test

Example for continuous data

Consider the weights in g of ten mice:

8.87	9.07	9.46	9.57	10.23	9.08	8.14	9.96	7.79	8.13
------	------	------	------	-------	------	------	------	------	------

Question: Is $\bar{x} = 9.03g$ significantly different from $\mu_0 = 10g$, $\sigma = 1g$?



State the appropriate hypotheses

The appropriate hypotheses are

$$\text{Null-hypothesis} \quad H_0 : \mu = \mu_0 \quad \implies \quad H_0 : \mu = 10$$

$$\text{Alternative hypothesis} \quad H_a : \mu \neq \mu_0 \quad \implies \quad H_a : \mu \neq 10$$

Alternatives

► two-sided test

$$H_0 : \mu = \mu_0 \implies H_0 : \mu = 10$$

$$H_a : \mu \neq \mu_0 \implies H_a : \mu \neq 10$$

► one-sided test

$$H_0 : \mu \geq \mu_0 \implies H_0 : \mu \geq 10$$

$$H_a : \mu < \mu_0 \implies H_a : \mu < 10$$

► one-sided test

$$H_0 : \mu \leq \mu_0 \implies H_0 : \mu \leq 10$$

$$H_a : \mu > \mu_0 \implies H_a : \mu > 10$$

The Z-test statistics

For making inferences about the mean of a normal distribution with known variance:

- ▶ the test statistic Z is calculated as

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}},$$

- ▶ where μ_0 is the value under H_0 ($\mu_0 = 10$)
- ▶ If the null hypothesis is true what is $\mathbb{E}[Z]$ and $\text{Var}(Z)$?

Calculate the observed value of the test statistic

- ▶ We assume that the weights are normally distributed

$$X_i \sim \mathcal{N}(\mu_0, \sigma),$$

- ▶ hence,

$$\mathbb{E}[\bar{X}] = \mu_0 \quad \text{and} \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

- ▶ and since σ^2 and n are known, we can rewrite

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \implies Z = a\bar{X} + b$$
$$a = \frac{\sqrt{n}}{\sigma}, \quad b = -\frac{\mu_0\sqrt{n}}{\sigma}.$$

Calculate the observed value of the test statistic

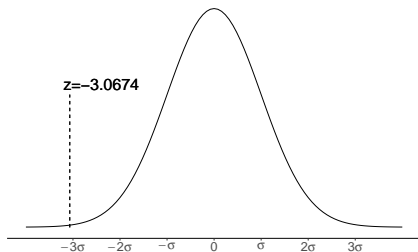
- For the mice experiment (i.e. $\mu_0 = 10$)

$$Z = \frac{\bar{X} - 10}{1/\sqrt{10}}$$

- The observed value of Z for the weight data is:

$$z = \frac{9.03 - 10}{1/\sqrt{10}} \approx -3.0674$$

This is more than 3σ away
and thus very unlikely!



p -values

The p -value is the probability of observing the measured value or a more extreme value, assuming that the null hypothesis and the test statistic are correct.

For z - tests this probability can be calculated directly from the z - value.

$$\mathbb{P}(Z \leq z_-) = \int_{-\infty}^{z_-} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad \leftarrow \text{left handed p-value}$$

$$\mathbb{P}(Z \geq z_+) = \int_{z_+}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad \leftarrow \text{right handed p-value}$$

We found for our weight data $z_{\pm} = \pm 3.0674$.

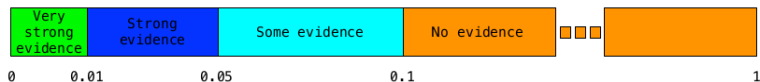
$$\mathbb{P}(Z \leq -3.0674) + \mathbb{P}(Z \geq 3.0674) = 0.000430 \quad \leftarrow \text{p-value}$$

Conclusion about H_0 ?

On the basis of the p -value, decide to retain or reject H_0

Let α denote the significance level and p denote the p -value, then

- ▶ if $p \geq \alpha$: retain H_0
- ▶ if $p < \alpha$: reject H_0 .



Using a critical value

Fix significance level α and compute the critical value $z_{\alpha/2}$ as

$$\mathbb{P}(Z \geq z_{\alpha/2}) = \alpha/2,$$

where $Z \sim \mathcal{N}(0, 1)$.

The decision process is then:

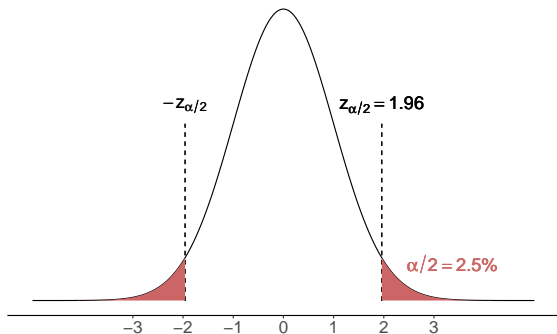
- ▶ if $|z| > z_{\alpha/2}$: reject H_0 ,
- ▶ if $|z| \leq z_{\alpha/2}$: retain H_0 ,

where z is the observed value of the test statistic.

Null Distribution for z-Tests

Model: $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma)$, $H_0 : \mu = \mu_0$

Test Statistic: $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$, and $\alpha = 95\%$



Reject H_0 if z lies within **critical area** (red)

Using a confidence region

For a significance level of α , a $(1 - \alpha)100\%$ confidence interval for the mean is calculated, denoted (L, U) , and then the decision protocol is

- ▶ if $\mu_0 \in (L, U)$: retain H_0 ,
- ▶ if $\mu_0 \notin (L, U)$: reject H_0 ,

where μ_0 is the hypothesised value.

Example: Mice weight data for $\alpha = 95\%$

$$\begin{aligned} 10g &\notin \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \\ &= \left(9.03g - 1.96 \frac{1g}{\sqrt{10}}, \bar{x} + 1.96 \frac{1g}{\sqrt{10}} \right) = (8.41g, 9.65g) \end{aligned}$$

Check the assumptions

- ▶ For inferring the mean of a normal distribution with known variance, the assumption is that the observation is a random sample from a normal distribution:

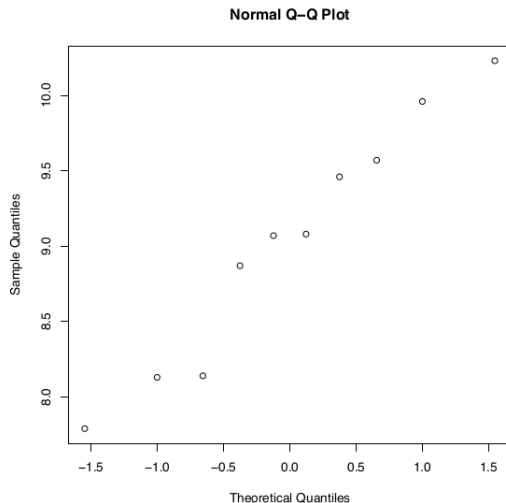
$$X_i \sim \mathcal{N}(\mu, \sigma), \quad i = 1, 2, \dots, n.$$

- ▶ For $n > 30$, we need the random variables $X_i, i = 1, \dots, n$ to be independent and identical distributed. (Central limit theorem)

Checking assumptions for mice data

1. Normality (Q-Q Plot)
2. Identically distributed
3. Known variance (external information)

Normal quantile plot for mice data



Sampling from same distribution

Split measured data into two parts: Randomly select n_1 items and interpret them as part of treatment group 1. Do the same for the remaining items $n_2 = n - n_1$.

Now, we want to know if the means are identical.

- We are interested in the difference $\mu_1 - \mu_2$.

Point estimate

If

- ▶ \bar{X}_1 is the sample mean of $X_{1,i}$, $i = 1, \dots, n_1$
- ▶ \bar{X}_2 is the sample mean of $X_{2,i}$, $i = 1, \dots, n_2$

and for $j = 1, 2$,

$$\bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{j,i},$$

- ▶ then a point estimate of $\mu_1 - \mu_2$ is $\bar{X}_1 - \bar{X}_2$.

Sample distribution

If

$$X_{1,i} \sim \mathcal{N}(\mu_1, \sigma_1), \quad i = 1, \dots, n_1,$$

and

$$X_{2,i} \sim \mathcal{N}(\mu_2, \sigma_2), \quad i = 1, \dots, n_2,$$

- ▶ What is $\mathbb{E}[\bar{X}_1 - \bar{X}_2]$? $\mathbb{E}[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2$
- ▶ What is $\text{Var}(\bar{X}_1 - \bar{X}_2)$? $\text{Var}(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
- ▶ What is the sampling distribution of $\bar{X}_1 - \bar{X}_2$?

Sample distribution

Since the $X_{j,i}$ are all normally distributed, and $\bar{X}_1 - \bar{X}_2$ is just a linear combination of the $X_{j,i}$,

$$(\bar{X}_1 - \bar{X}_2) \sim \mathcal{N}\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

For now, we will use the sample variances in place of the unknown population variances, meaning (for now) our test is only approximate.

Hypothesis testing I (Two-sided Test)

The steps for hypothesis testing are as follows:

1. State the appropriate hypotheses

$$H_0 : \quad \mu = \mu_0 \quad (\text{equivalently: } \mu_1 - \mu_2 = 0)$$

$$H_a : \quad \mu \neq \mu_0 \quad (\text{equivalently: } \mu_1 - \mu_2 \neq 0)$$

2. Choose the required significance level α

3. Check the assumptions:

3.1 The group 1 data are a random sample from $\mathcal{N}(\mu_1, \sigma_1)$.

3.2 The group 2 data are a random sample from $\mathcal{N}(\mu_2, \sigma_2)$.

3.3 The two sets of observations are independent.

Hypothesis testing II (Two-sided Test)

4. Calculate the observed value of the test statistic

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

5. Calculate the p -value

If H_0 is true, then the test statistic z is (approximately) standard normally distributed,

$$z \sim \mathcal{N}(0, 1)$$

and the p -value is

$$\mathbb{P}(|Z| \geq |z|)$$

where z is the observed value of the test statistic.

6. On the basis of the p -value, decide to retain or reject H_0 .

Confidence interval

A $(1 - \alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is (approximately)

$$\left(\bar{x}_1 - \bar{x}_2 - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right),$$

where

- ▶ $z_{\alpha/2}$ is the critical value such that $\mathbb{P}(|Z| \geq z_{\alpha/2}) = \alpha/2$,
 - ▶ where Z has a standard normal distribution.

Randomly Splitting Mice Data

- ▶ **Group 1:** [9.46, 8.87, 9.57, 10.23, 9.08]
 - ▶ Mean: 9.44
 - ▶ Variance: 0.27
- ▶ **Group 2:** [8.14, 9.07, 8.13, 7.79, 9.96]
 - ▶ Mean: 8.62
 - ▶ Variance: 0.79

$$x_1 - \bar{x}_2 = 0.824$$

Confidence interval:

$$(-0.32, 1.97)$$

- ▶ Since 0 lies within the confidence interval, we cannot conclude that the means are significantly different.

The Student's t-Test

Inference for the population mean μ when the population variance is unknown

So far we assumed a normal distribution with known variance.

- ▶ Consider the problem when we have n observations from a normal distribution with unknown variance.
- ▶ Now, we would like to estimate and make inferences about the population mean.

Point estimates for μ and σ^2

- ▶ As before a point estimate for the population mean μ is the sample mean \bar{X} calculated by

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

where X_i , $i = 1, \dots, n$ are the n observations.

- ▶ For the population variance, the sample variance can be used as a point estimate of the population variance, i.e.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

can be used as a point estimate of σ^2 .

T -distribution

Consider n independent observations X_i , $i = 1, \dots, n$, from a normal distribution with mean μ and standard deviation σ , i.e.,

$$X_i \sim \mathcal{N}(\mu, \sigma),$$

then the random variable, denoted T , calculated by

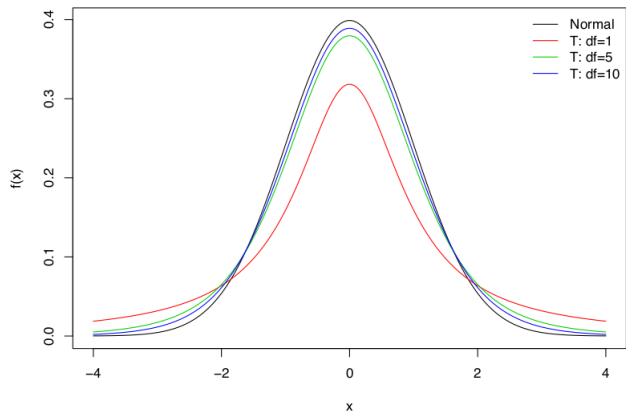
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

where

- ▶ \bar{X} is the estimator for the mean, and
- ▶ S is the estimator for the standard deviation

has a T -distribution with $n - 1$ degrees of freedom.

T-distribution



df: degrees of freedom

Power analysis

Retaining H_0

If we retain H_0 this can be because either

- ▶ H_0 is true, or
- ▶ there is not enough evidence to show that H_0 is false.

Therefore, we use the term retain but make no claims about the validity of H_0 .

Choose the required significance level α

Hypothesis testing results in deciding to retain or reject the null hypothesis. This gives us four possibilities:

		Decision retain H_0	reject H_0
H_0	true	correct $1 - \alpha$	Type I error α (false positive)
	false	Type II error β (false negative)	correct $1 - \beta$

The significance level of a hypothesis testing procedure is the probability of a Type I error.

The power of a test is given by $1 - \beta$. The higher the power the better the test. Calculating β involves knowledge about the specific statistical quantities.

The power of a statistical test

If we reject the null hypothesis if $p \leq \alpha$, we reject true null hypothesis with probability α .

But how likely are we to not reject the false null hypothesis?

To answer that, we must assume a specific alternative model

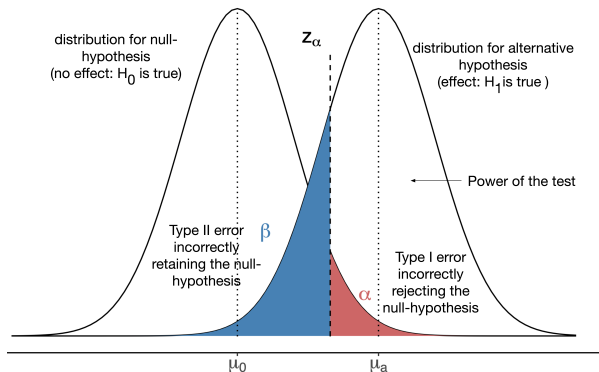
The *false negative* (or type-II error) rate at significance level α is

$$\beta = \mathbb{P}(\{p \geq \alpha | \text{alternative hypothesis is true}\})$$

and the power against that alternative at level α is $1 - \beta$.

- ▶ $p \leq \alpha$ rejects the null hypothesis, it does not prove the alternative to be true!

The power of a statistical test



- ▶ variance increases the power decreases (fatter curves)
- ▶ number of samples increase the power increases (thinner curves)

Power analysis

1. Pick specific alternative
2. Find distribution D_{alt} of test statistic under H_a
3. Pick significance level α
4. Find acceptance region A , i.e., values of test statistic where $p \geq \alpha$
5. Intersect D_{alt} with A to find the false negative rate

$$\beta = \mathbb{P}(\{p \geq \alpha | \text{alternative model is true}\})$$

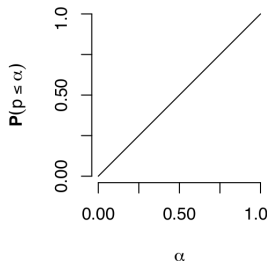
6. The power of the test against the alternative model is $1 - \beta$.
This is tedious! G^* Power implements this for many common tests
<http://www.gpower.hhu.de/>

The multiple testing problem

Multiple Tests

p -value distribution:

If we test a *true* null hypothesis, the p -value distribution is uniform



This means that for $n = 1000$ and $\alpha = 0.05$, about $0.05 \times 1000 = 50$ tests will show significant results purely by chance, regardless of the shape of the null-distribution.

Running multiple tests

Example: scRNA-seq expression levels

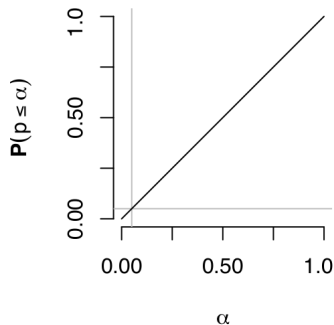
	Control				Treatment			
	cell 1	...	cell n	mean	cell 1	...	cell n	mean
gene 1	c_{11}	...	c_{1n}	$\overline{c_1}$	t_{11}	...	t_{1n}	$\overline{t_1}$
\vdots			\ddots				\ddots	
gene m	c_{m1}	...	c_{mn}	$\overline{c_m}$	t_{m1}	...	t_{mn}	$\overline{t_m}$

Q: Which genes show significant differences in average expression levels?

Running multiple tests

Probability that at least one p -value in m (independent) tests is $\leq \alpha$?

$$\mathbb{P}(p_1 \leq \alpha) = 1 - (1 - \alpha)^1$$

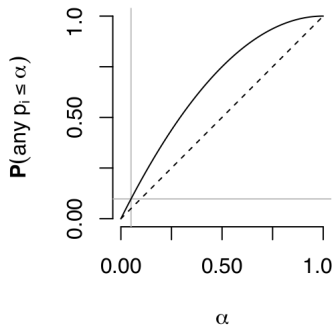
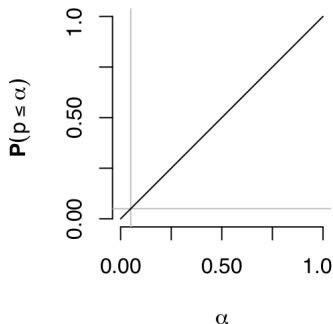


Running multiple tests

Probability that at least one p -value in m (independent) tests is $\leq \alpha$?

$$\mathbb{P}(p_1 \leq \alpha) = 1 - (1 - \alpha)^1$$

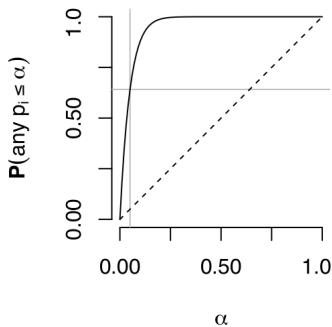
$$\mathbb{P}(p_1 \leq \alpha \text{ or } p_2 \leq \alpha) = 1 - (1 - \alpha)^2$$



Running multiple tests

Probability that at least one p -value in m (independent) tests is $\leq \alpha$?

$$\mathbb{P}(p_1 \leq \alpha \text{ or } \dots \text{ or } p_{20} \leq \alpha) = 1 - (1 - \alpha)^{20}$$



Running multiple tests

Probability that at least one p -value in m (independent) tests is $\leq \alpha$?

$$\mathbb{P}(p_1 \leq \alpha \text{ or } \dots \text{ or } p_{20} \leq \alpha) = 1 - (1 - \alpha)^{20}$$

Example: testing 20 drugs with none of them having an effect. The null-hypothesis is true for all.

For each drug we have a Type I error of $\alpha = 0.05$ which is the probability of incorrectly rejecting H_0 .

Considering all 20 drugs the probability of at least one false positive is

$$1 - (1 - \alpha)^{20} = 1 - (1 - 0.05)^{20} \approx 0.64$$

The multiple testing problem

The probability of making at least one Type I error increases with the number of tests!

If we test m hypotheses, we call the probability

$$\mathbb{P}(p_1 \leq \alpha \text{ or } \dots \text{ or } p_m \leq \alpha)$$

(assuming the hypotheses are true) the **family-wise error rate**. (FWER)

= Probability of making one or more false discovery

- ▶ Without correction, $\text{FWER} > \alpha$ if $m > 1$

To ensure $\text{FWER} \leq \alpha$, we adjust (correct) the individual p -values. (or equivalently, we adjust the individual α 's)

Methods to control the FWER

Assume: $p_1 \leq p_2 \leq \dots \leq p_m$ with $H_0(1), \dots, H_0(m)$.

Method of **Bonferroni**

- ▶ Reject $H_0(i)$ if $p_i < \frac{\alpha}{m}$ for $i = 1, \dots, m$

Method of **Holm-Bonferroni**

- ▶ Sort p-values in ascending order p_1, p_2, \dots, p_m and compare them to $p_k < \frac{\alpha}{m-k+1} \Rightarrow$ Reject H_0 for the smallest p-values until one fails the condition.

Method of **Benjamini-Hochberg**

- ▶ Sort p-values in ascending order p_1, p_2, \dots, p_m . Find the largest p-value such that $p_k \leq \frac{k}{m}\alpha \Rightarrow$ Reject all H_0 for p_1, p_2, \dots, p_k .