

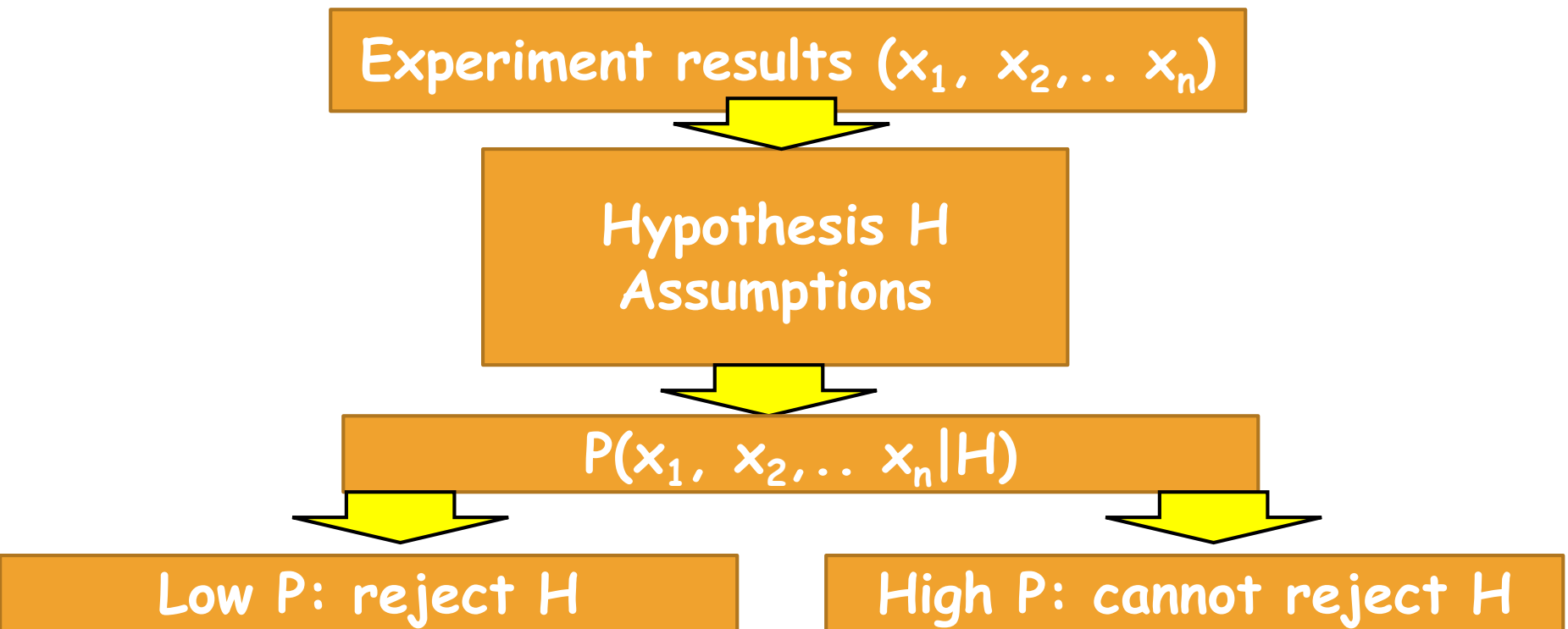
# TESTING HYPOTHESES

# Hypotheses

- When performing an experiment or analysing its results, hypotheses and assumptions are formulated.
- In some cases it is possible to evaluate the agreement between the results and the hypotheses that have been formulated, comparing the theoretical expectation with the actual results

# Rejecting hypotheses

- The high agreement doesn't proof the hypotheses. A low agreement can indicate that the hypothesis has to be rejected



# General procedure

- State the **relevant null and alternative hypotheses** to be tested.
- Formulate the **assumptions**: for example about the statistical independence or about the shape of the distributions of the observations.
- Compute the **relevant test statistic on the basis of experimental data**.
- The **distribution of such statistic under the null hypothesis** can be derived from the assumptions.
- **Compare the test-statistic ( $S$ ) to the relevant critical values ( $CV$ )** determined from the distribution, for a particular significance level.
- **Decide** to either fail to reject the null hypothesis or reject it in favor of the alternative. The decision rule is (usually) to **reject the null hypothesis if  $S > CV$** .

# CAVEAT

## CORRECT STATEMENT

- The "**fail to reject**" terminology highlights the fact that the null hypothesis is assumed to be true from the start of the test; if there is a lack of evidence against it, it simply continues to be assumed true.

## WRONG STATEMENT

- The phrase "**accept the null hypothesis**" may suggest it has been proved simply because it has not been disproved, a logical fallacy known as the argument from ignorance.

# The first significance test (1935): Fisher against a Lady tasting tea

Fisher thoroughly explained his method in a proposed experiment to test a Lady's claimed ability to determine the means of tea preparation by taste (whether the milk or the infusion was added to the cup first)



# The first significance test (1935): Fisher against a Lady tasting tea



- The null hypothesis was that the Lady had **no** such ability.
- The test statistic was a simple count of the number of successes in 8 trials.

# Fisher against a Lady tasting tea

- Suppose you prepare 8 cups (4 for each treatment) and ask the lady to taste them in random order and to divide them into two groups of 4 elements following her judgement on the treatment.
- How many successes the lady must obtain in order to be trusted?
- What is the probability of obtaining the same number of successes if the null hypothesis holds?



# Fisher against a Lady tasting tea

- How many possibilities are there to select 4 cups out of 8 (e.g. in how many ways the lady can indicate 4 cups were she think the milk has been poured before the infusion)?
- How many of them are correct ?
- What is the number of configurations scoring 6,4,2 or 0 correct classifications? (
- (successes are always even because an error in one class always carries an error in the complementary class)

# Fisher against a Lady tasting tea

- There are 70 possibilities for selecting 4 cups out of 8 ( $8!/4!4!$ ).
- Possibilities of having  $k$  successes (*cups where the milk has in fact been poured before the infusion*) out of 4 choices in a set of 4 correct and 4 wrong possibilities:

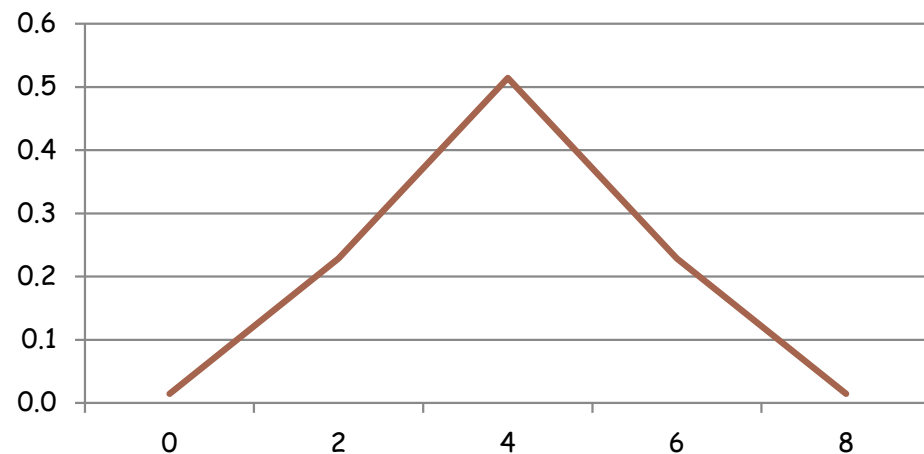
Select  $k$  out of 4 correct cups :  $4 \cdot (4-1) \cdot \dots \cdot (4-k+1)/k!$  choices AND  
( $4-k$ ) out of 4 wrong cups :  $4 \cdot (4-1) \cdot \dots \cdot (k+1)/(4-k)! \rightarrow [4!/k!(4-k)!]^2$

- $k = 4 \rightarrow 1$  possibility(perfect assignment)
- $k = 3 \rightarrow 16$  possibilities(2 errors, on the two classes)
- $k = 2 \rightarrow 36$  possibilities(4 errors, on the two classes)
- $k = 1 \rightarrow 16$  possibilities(6 errors on the two classes)
- $k = 0 \rightarrow 1$  possibility(all assignments are wrong)

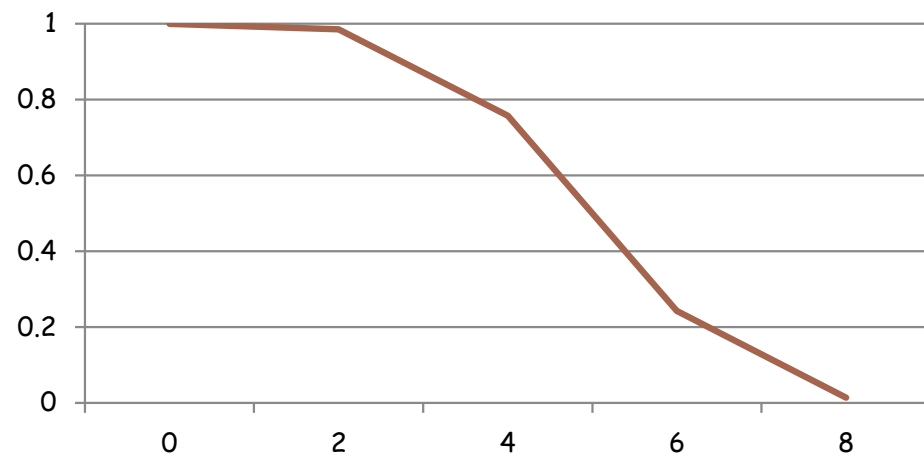
# Fisher against a Lady tasting tea

- Following the null hypothesis (no ability in discrimination  $\rightarrow$  random choice of 4 cups out of 8), the probability of correct assignment is

$P(\text{n.of.success} = k)$

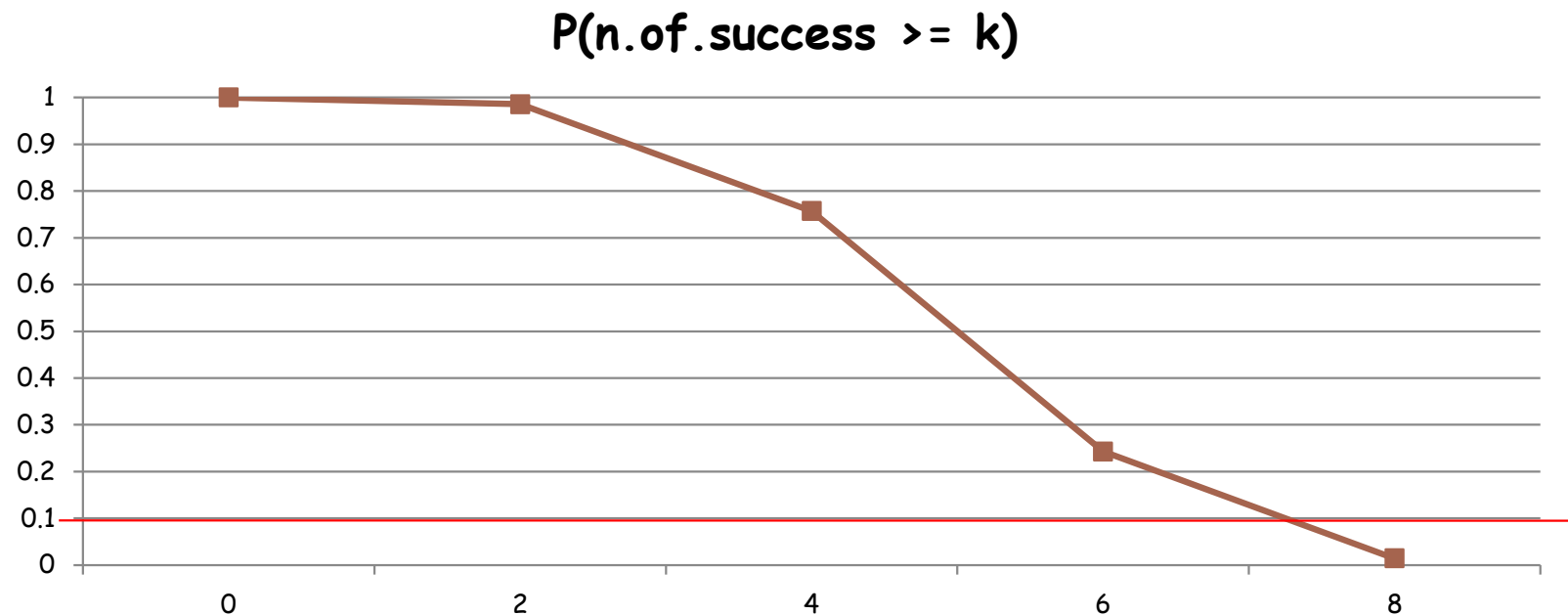


$P(\text{n.of.success} \geq k)$



# Fisher against a Lady tasting tea

- A significance level has to be fixed (e.g 0.05) and if the actual probability is below we can reject the null hypothesis



# Fisher against a Lady tasting tea

- The critical region was the single case of 8 successes in 8 trials based on a conventional probability criterion ( $< 5\%$ ).
- If and only if the 8 trials produced 8 successes was Fisher willing to reject the null hypothesis - effectively acknowledging the Lady's ability with  $> 98\%$  confidence (but without quantifying her ability).

# Glossary

- **p-value** is the conditional probability of observing a result at least as extreme as the test statistic (computed from the samples) *GIVEN* that the null hypothesis is true.
- **Significance level of a test ( $\alpha$ )**: the probability that the test will reject a null hypothesis that is known to hold ("to be true"). It determines the maximum p-value for rejecting the hypothesis.
- **Power of a test ( $1 - \beta$ )**: the probability that the test will reject a null hypothesis that doesn't hold ("is false").  $\beta$  is the probability of a false "non rejection"

# Hypothesis Testing

Test Result –		$H_0$ NOT Rejected	$H_0$ Rejected
True State	$H_0$ True	Correct Decision	Type I Error
	$H_0$ False	Type II Error	Correct Decision

$$\alpha = P(\text{Type I Error}) \quad \beta = P(\text{Type II Error})$$

- Goal: Keep  $\alpha, \beta$  reasonably small

# CAVEAT

- The p-value is NOT the probability that the null hypothesis is true.  
In fact, frequentist statistics does not attach probabilities to hypotheses. Comparison of Bayesian and classical approaches shows that a p-value can be very close to zero while the posterior probability of the null is very close to unity.
- $1 - (\text{p-value})$  is *not* the probability of the alternative hypothesis being true .



# Bayesian approach to significance (not very rigorous intro....)

- Interpreting p-value as  $P(\text{Data} | H_0)$ , the question we'd like to answer is related to the computation of  $P(H_0 | \text{Data})$

$$P(H_0 | D) = \frac{P(D | H_0) \cdot P(H_0)}{P(D | H_0) \cdot P(H_0) + P(D | H_{Alt}) \cdot P(H_{Alt})}$$

Or to compare two hypotheses

$$\underbrace{\frac{P(H_0 | D)}{P(H_1 | D)}}_{\text{A-posteriori odds}} = \underbrace{\frac{P(D | H_0)}{P(D | H_1)}}_{\text{Bayesian factor}} \cdot \underbrace{\frac{P(H_0)}{P(H_1)}}_{\text{A-priori odds}}$$

In both cases we would need to estimate the a-priori probabilities for the hypotheses

□ Under some assumption:

If  $p$  is the p-value, the Bayesian factor lower bound can be estimated as

$$\text{Bayesian\_factor} \geq B(p) = -e \cdot p \cdot \ln(p)$$

$$\frac{P(H_0 | D)}{1 - P(H_0 | D)} \geq B(p) \frac{P(H_0)}{P(H_{Alt})}$$

$$P(H_0 | D) \leq \left[ 1 + \frac{1}{B(p) \frac{P(H_0)}{P(H_{Alt})}} \right]^{-1}$$

# A Dirty Dozen: Twelve *P*-Value Misconceptions

Steven Goodman

The *P* value is a measure of statistical evidence that appears in virtually all medical research papers. Its interpretation is made extraordinarily difficult because it is not part of any formal system of statistical inference. As a result, the *P* value's inferential meaning is widely and often wildly misconstrued, a fact that has been pointed out in innumerable papers and books appearing since at least the 1940s. This commentary reviews a dozen of these common misinterpretations and explains why each is wrong. It also reviews the possible consequences of these improper understandings or representations of its meaning. Finally, it contrasts the *P* value with its Bayesian counterpart, the Bayes' factor, which has virtually all of the desirable properties of an evidential measure that the *P* value lacks, most notably interpretability. The most serious consequence of this array of *P*-value misconceptions is the false belief that the probability of a conclusion being in error can be calculated from the data in a single experiment without reference to external evidence or the plausibility of the underlying mechanism.

Semin Hematol 45:135-140 © 2008 Elsevier Inc. All rights reserved.

---

**Table 1** Twelve *P*-Value Misconceptions

1	<i>If <math>P = .05</math>, the null hypothesis has only a 5% chance of being true.</i>
2	<i>A nonsignificant difference (eg, <math>P \geq .05</math>) means there is no difference between groups.</i>
3	<i>A statistically significant finding is clinically important.</i>
4	<i>Studies with <math>P</math> values on opposite sides of .05 are conflicting.</i>
5	<i>Studies with the same <math>P</math> value provide the same evidence against the null hypothesis.</i>
6	<i><math>P = .05</math> means that we have observed data that would occur only 5% of the time under the null hypothesis.</i>
7	<i><math>P = .05</math> and <math>P \leq .05</math> mean the same thing.</i>
8	<i><math>P</math> values are properly written as inequalities (eg, "<math>P \leq .02</math>" when <math>P = .015</math>)</i>
9	<i><math>P = .05</math> means that if you reject the null hypothesis, the probability of a type I error is only 5%.</i>
10	<i>With a <math>P = .05</math> threshold for significance, the chance of a type I error will be 5%.</i>
11	<i>You should use a one-sided <math>P</math> value when you don't care about a result in one direction, or a difference in that direction is impossible.</i>
12	<i>A scientific conclusion or treatment policy should be based on whether or not the <math>P</math> value is significant.</i>

# STATISTICAL ERRORS:

P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume

NATURE | VOL 506 | 13 FEBRUARY 2014

## PROBABLE CAUSE

A P value measures whether an observed result can be attributed to chance. But it cannot answer a researcher's real question: what are the odds that a hypothesis is correct? Those odds depend on how strong the result was and, most importantly, on how plausible the hypothesis is in the first place.

■ Chance of real effect  
■ Chance of no real effect

### Before the experiment

The plausibility of the hypothesis — the odds of it being true — can be estimated from previous experiments, conjectured mechanisms and other expert knowledge. Three examples are shown here.

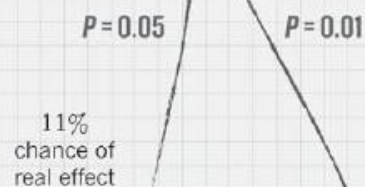
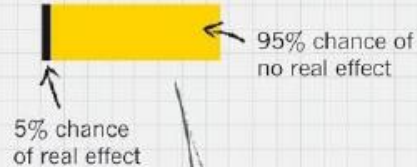
### The measured P value

A value of 0.05 is conventionally deemed 'statistically significant'; a value of 0.01 is considered 'very significant'.

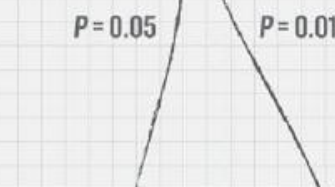
### After the experiment

A small P value can make a hypothesis more plausible, but the difference may not be dramatic.

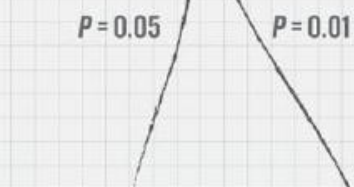
#### THE LONG SHOT 19-to-1 odds against



#### THE TOSS-UP 1-to-1 odds



#### THE GOOD BET 9-to-1 odds in favour



$$P(H_{alt}) : P(H_0)$$

$$P(H_{alt}|D) : P(H_0|D)$$



# CAVEAT 2

- The p-value is *not* the probability that a finding is "merely a fluke"  
As the calculation of a p-value is based on the *assumption* that a finding is the product of chance alone, it patently cannot also be used to gauge the probability of that assumption being true.
- The p-value is *not* the probability of falsely rejecting the null hypothesis.
- The p-value is *not* the probability that a replicating experiment would not yield the same conclusion.

# TESTING HYPOTHESES

Testing independence

# Testing independence

FISHER

- A sample of teenagers might be divided into male and female on the one hand, and those that are and are not currently studying for a statistics exam on the other.
- We hypothesize, for example, that the proportion of studying individuals is higher among the women than among the men, and we want to test whether any difference of proportions that we observe is significant.

CONTINGENCY TABLE	Men	Women	Row total
Studying	1	9	10
Not-studying	11	3	14
Column total	12	12	24



# Testing independence

more extreme  
table: higher  
disproportion

	Men	Women	Row total
Studying	1	9	10
Not-studying	11	3	14
Column total	12	12	24

- The question we ask about these data is: knowing that 10 of these 24 teenagers are studiers, and that 12 of the 24 are female, and **assuming the null hypothesis that men and women are equally likely to study**,
  - 1) what is the probability that these 10 studiers would be so unevenly distributed between the women and the men?
  - 2) If we were to choose 10 of the teenagers at random, what is the probability that 9 or more of them would be among the 12 women, and only 1 or fewer from among the 12 men?

Wikipedia: Fisher test

# Contingency table

	Men	Women	Row Total
Studying	$a$	$b$	$a + b$
Non-studying	$c$	$d$	$c + d$
Column Total	$a + c$	$b + d$	$a + b + c + d (=n)$

# Fisher's exact test

$$p = \binom{a+b}{a}_{\text{or B}} \binom{c+d}{c}_{\text{or D}} / \binom{n}{a+c}$$
$$= \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

IT GIVES US THE P VALUE

- Is the probability of obtaining the samples contingency table in the case of independence

exist the STEARLYNG aproximation for the factorial

# Testing independence

	Men	Women	Row total
Studying	1	9	10
Not-studying	11	3	14
Column total	12	12	24

- 1) what is the probability that these 10 students would be so unevenly distributed between the women and the men?

$$p = \frac{10! \times 14! \times 12! \times 12!}{24! \times 1! \times 9! \times 11! \times 3!} = 0.001346$$

# Testing independence

	Men	Women	Row total
Studying	1	9	10
Not-studying	11	3	14
Column total	12	12	24

2) If we were to choose 10 of the teenagers at random, what is the probability **that 9 or more of them would be among the 12 women, and only 1 or fewer from among the 12 men?**

We have to consider also the table (with equal margin values):

	Men	Women	Row total
Studying	0	10	10
Not-studying	12	2	14
Column total	12	12	24

- For testing extreme data, a sum over the more extreme table is required to: e.g., suppose “b” being very high

	Men	Women	Row Total
Studying	$a$	$b$	$a + b$
Non-studying	$c$	$d$	$c + d$
Column Total	$a + c$	$b + d$	$a + b + c + d$

$$p = \sum p(a', b, c', d)$$

$$b' > b$$

$$a' + b' = a + b$$

$$d' + b' = d + b$$

$$a' + c' = a + c$$

$$d' + c' = d + c$$

you can reject  
the hypothesis  
that there is no  
correlation

# Multiple testing

García-Arenzana et al. (2014) tested associations of 25 dietary variables with mammographic density, an important risk factor for breast cancer, in Spanish women. They found the following results:

Dietary variable	P value	Dietary variable	P value
Total calories	<0.001	Fruit	0.269
Olive oil	0.008	Eggs	0.275
Whole milk	0.039	Blue fish	0.34
White meat	0.041	Legumes	0.341
Proteins	0.042	Carbohydrates	0.384
Nuts	0.06	Potatoes	0.569
Cereals and pasta	0.074	Bread	0.594
White fish	0.205	Fats	0.696
Butter	0.212	Sweets	0.762
Vegetables	0.216	Dairy products	0.94
Skimmed milk	0.222	Semi-skimmed milk	0.942
Red meat	0.251	Total meat	0.975
		Processed meat	0.986

- Considering a significance equal to 5%, five variable would be considered as significantly associated to the mammographic density.

Dietary variable	<i>P</i> value
Total calories	<0.001
Olive oil	0.008
Whole milk	0.039
White meat	0.041
Proteins	0.042

- BUT, remember what is the meaning of the significance!



# Multiple testings

- We are testing several hypotheses
- Significance ( $\alpha$ ) : the probability that the test will reject a null hypothesis that is known to hold ("to be true") → FALSE DISCOVERY.
- If the significance  $\alpha$  is set to 0.05, how many FALSE DISCOVERIES do you expect?
- In 1-2 cases the null hypothesis is rejected by chance

# Correction for multiple testing

- The **Bonferroni correction** is a method used to address this problem.
- If an experimenter is testing  $n$  dependent or independent hypotheses on a set of data, then one way of maintaining the familywise error rate is to **test each individual hypothesis at a statistical significance level of  $1/n$  times what it would be if only one hypothesis were tested**. So if you want the significance level for the whole family of tests to be (at most)  $\alpha$ , then the Bonferroni correction would be to test each of the individual tests at a significance level of  $\alpha/n$ .

- Considering a significance equal to 5%, and 25 different hypotheses, you must test each variable at  $0.05/25 = 0.002$  significance: only one variable can be considered as significantly associated to the mammographic density.

Dietary variable	<i>P</i> value
Total calories	<0.001
Olive oil	0.008
Whole milk	0.039
White meat	0.041
Proteins	0.042

- Bonferroni correction controls the probability of a False Discoveries only. The correction ordinarily comes at the cost of increasing the probability of a False non Rejections.

alternatives BENJAMINI-HOCHBERG procedure

# Enrichment tests

- Different procedure lead to the identification of groups of genes or proteins (e.g, genes differentially expressed in some condition). One important analysis is the identifications of features (eg. Gene Ontology terms) that are overrepresented in a particular group

	GO term j	Other GO terms	Total
Group	a	b	a + b
Other genes	c	d	c + d
total	a + c	b + d	n

# Multiple testings

- We are testing several GO terms (8351)
- Significance ( $\alpha$ ) : the probability that the test will reject a null hypothesis that is known to hold ("to be true") → FALSE DISCOVERY.
- If the significance  $\alpha$  is set to 0.01, how many FALSE DISCOVERIES do you expect?
- Some 84 significant by chance associations are expected
- Each hypothesis must be tested with a significance equal to  $0.01/8351=1.2 \times 10^{-6}$

# TESTING HYPOTHESES

Testing mean and variance

# Testing the mean and the variance of a sample against a known distribution

- Given a random variable  $X$ , that you suppose normally distributed with mean  $\mu$  and variance  $\sigma^2$  known a priori
- Suppose to have a sample  $X_1, X_2, \dots, X_n$

Compute the sample mean

$$M = \frac{1}{n} \sum_{i=1}^n X_i$$

used to understand if a difference between the mean of two distribution is enough to justify the difference between the two distribution

and the sample (unbiased) variance

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



# Distribution of the sample mean

- Consider different sets of sampled data  $X^k$  (e.g. different levels of expression of genes in  $n$  different populations of individuals)

A sample mean for each set can be defined

$$M^{(k)} = \frac{1}{n^{(k)}} \sum_{i=1}^{n^{(k)}} x_i^{(k)}$$

How are they distributed?

it is a NORMAL distribution of the MEAN

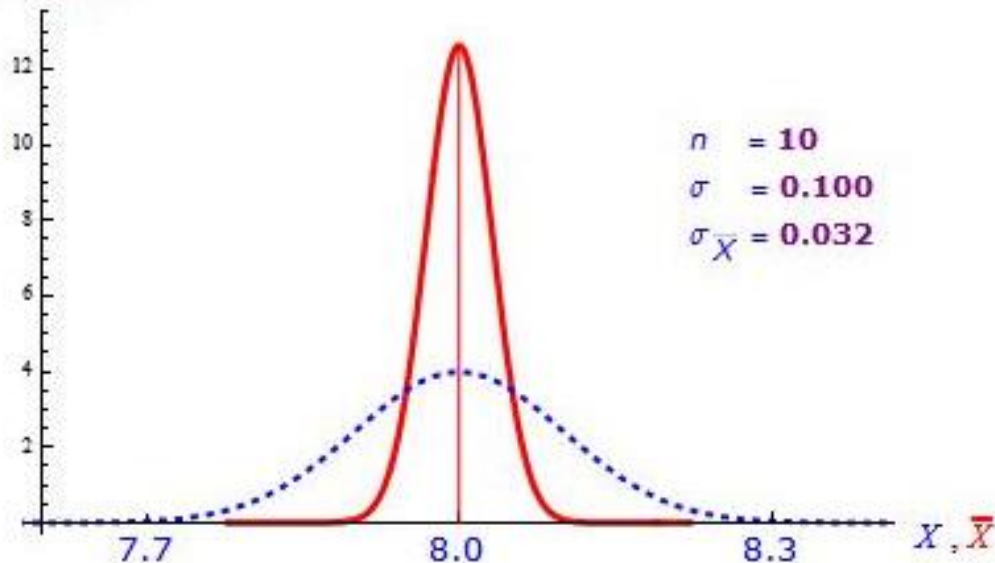
# Distribution of the sample means

$$M^{(k)} = \frac{1}{n} \sum_{i=1}^n x_i^{(k)}$$

M is a linear combination of independent identically distributed normal variables: it is normally distributed

PDFs of  $X$  and  $\bar{X}$

$f(X), g(\bar{X})$



# Distribution of the sample means

$$E[M] = \mu$$

$$\text{var}[M] = E[M^2] - \mu^2$$

$$E[M^2] = E\left[\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n x_i x_j\right] = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E[x_i x_j] =$$

$$= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (\mu^2 +) =$$

$$= \mu^2 + \frac{1}{n} \sigma^2$$

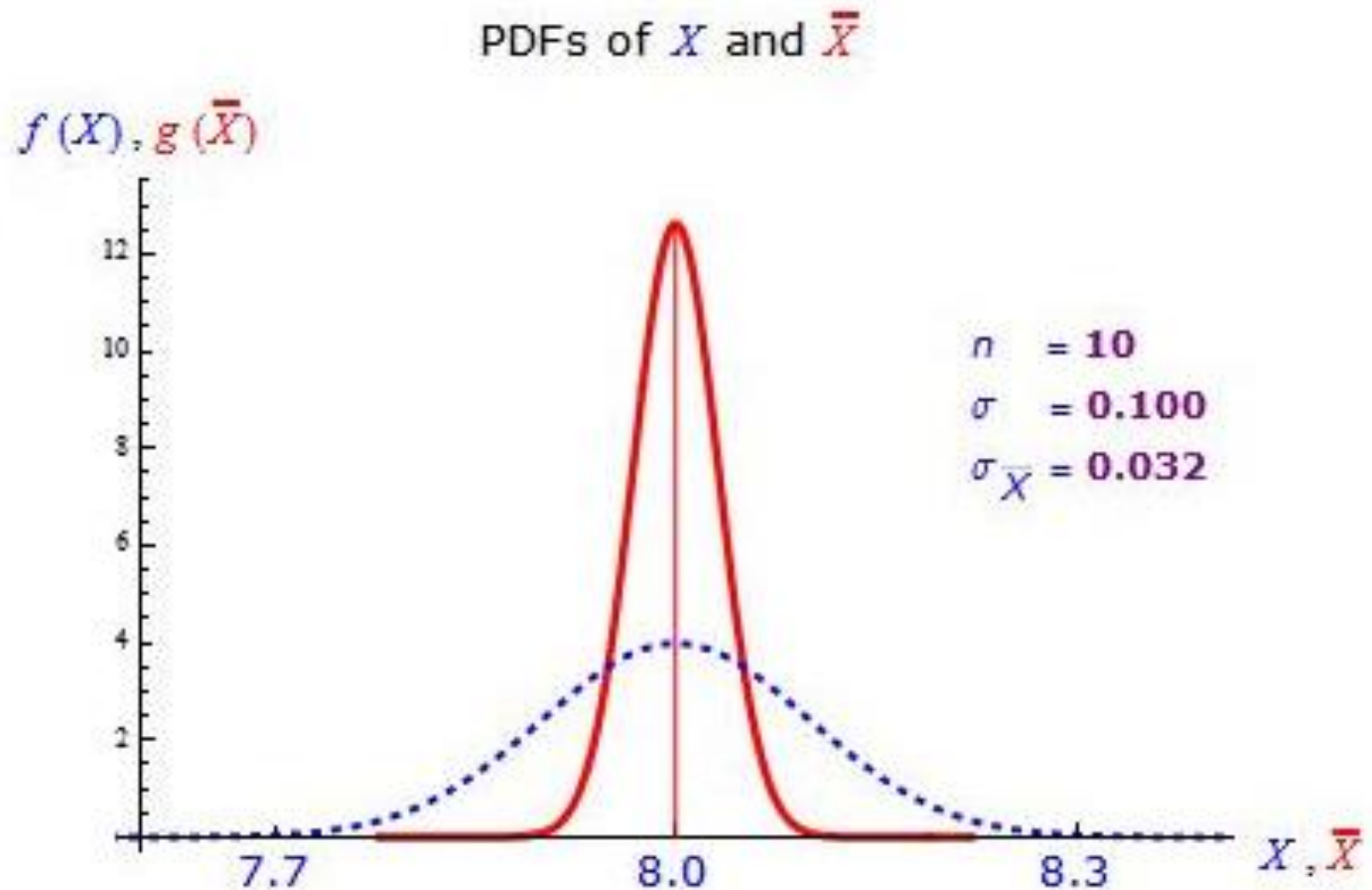
$$\text{var}[M] = \frac{1}{n} \sigma^2$$

Since

$$E[x_i x_i] = \sigma^2 + \mu^2$$

the MEAN is the mean of the mean  
The variance is  $\sigma^2/N$  (number of mean)  
so you can have the error on the mean (standard deviation =  $\sqrt{\text{variance}}$ )

# Distribution of the sample means



# Does $M$ significantly differ from $\mu$ ?

- Null hypothesis  $H_0: M = \mu$
- Alternative hypothesis  $H_a: M \neq \mu$

If  $H_0$  is true, then  $M$  (the sample mean) is normally distributed with mean  $\mu$  and variance  $\sigma^2/n$ . Then:

$$Z = \frac{M - \mu}{\sigma / \sqrt{n}}$$

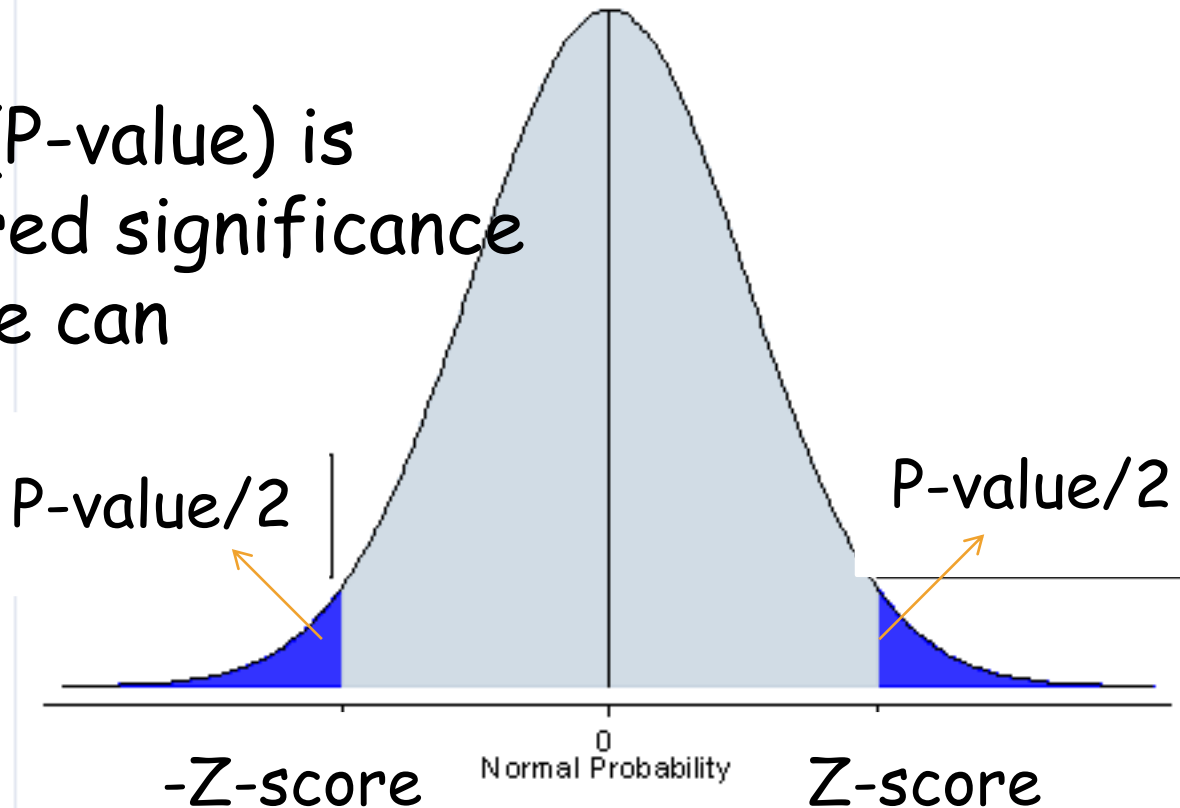
then you have to look for the area under a two tailed normal curve and it is the P value

is normally distributed, with mean 0 and variance 1

# Two-tailed Z-test

Given the sampled **absolute** value of  $Z$  we can compute the probability of obtaining a value equal or higher, on the normal distribution.

If that probability (P-value) is lower than the desired significance P-value (e.g. 0.05) we can reject  $H_0$



# Is $M$ significantly higher than $\mu$ ?

- Null hypothesis  $H_0: M \leq \mu$
- Alternative hypothesis  $H_a: M > \mu$

If  $H_0$  is true,  $M \leq \mu$ . The extreme case for our testing is  $M=0$ : in that case, again,

$$Z = \frac{M - \mu}{\sigma / \sqrt{n}}$$

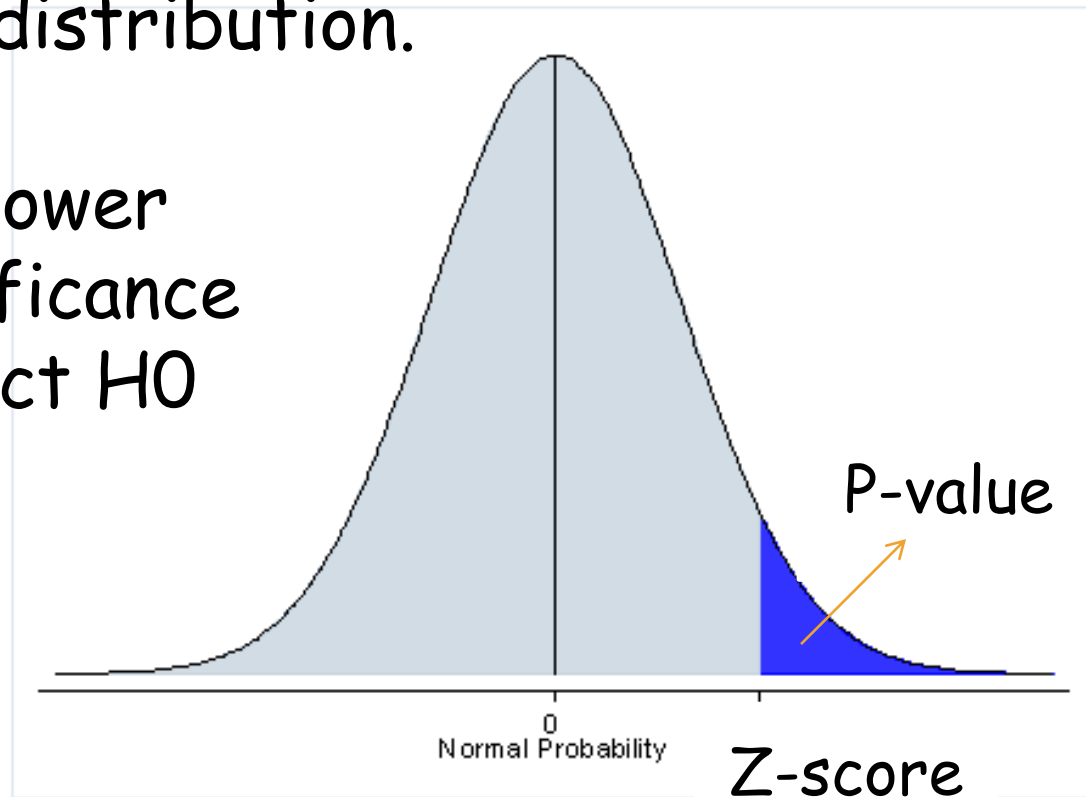
is normally distributed, with mean 0 and variance 1

# One-tailed Z-test

AVOID USING  
THE ONE  
TAILED

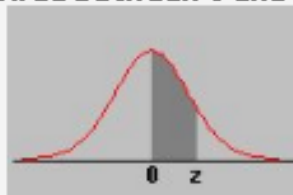
Given the Z-score computed from the sample we consider only the positive values of Z-score and compute the probability of obtaining a value **equal or higher**, on the normal distribution.

If that probability is lower than the desired significance (e.g. 0.05) we can reject  $H_0$





Area between 0 and z



AREA between 0 and x



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817

# Example

- If  $Z$  is equal to 2, can we reject the null hypothesis at significance 5%? And at significance 2%?

# Does $M$ significantly differ from $\mu$ ?

## Second round

- Null hypothesis  $H_0$ :  $M = \mu$
- Alternative hypothesis  $H_a$ :  $M \neq \mu$

If  $H_0$  is true,

$$Z = \frac{M - \mu}{\sigma / \sqrt{n}}$$

is normally distributed, with mean 0 and variance 1  
**BUT IT DEPENDS ON  $\sigma^2$  that in principle is UNKNOWN**

# T-test

William Sealy Gosset (aka Student) proved that the variable :

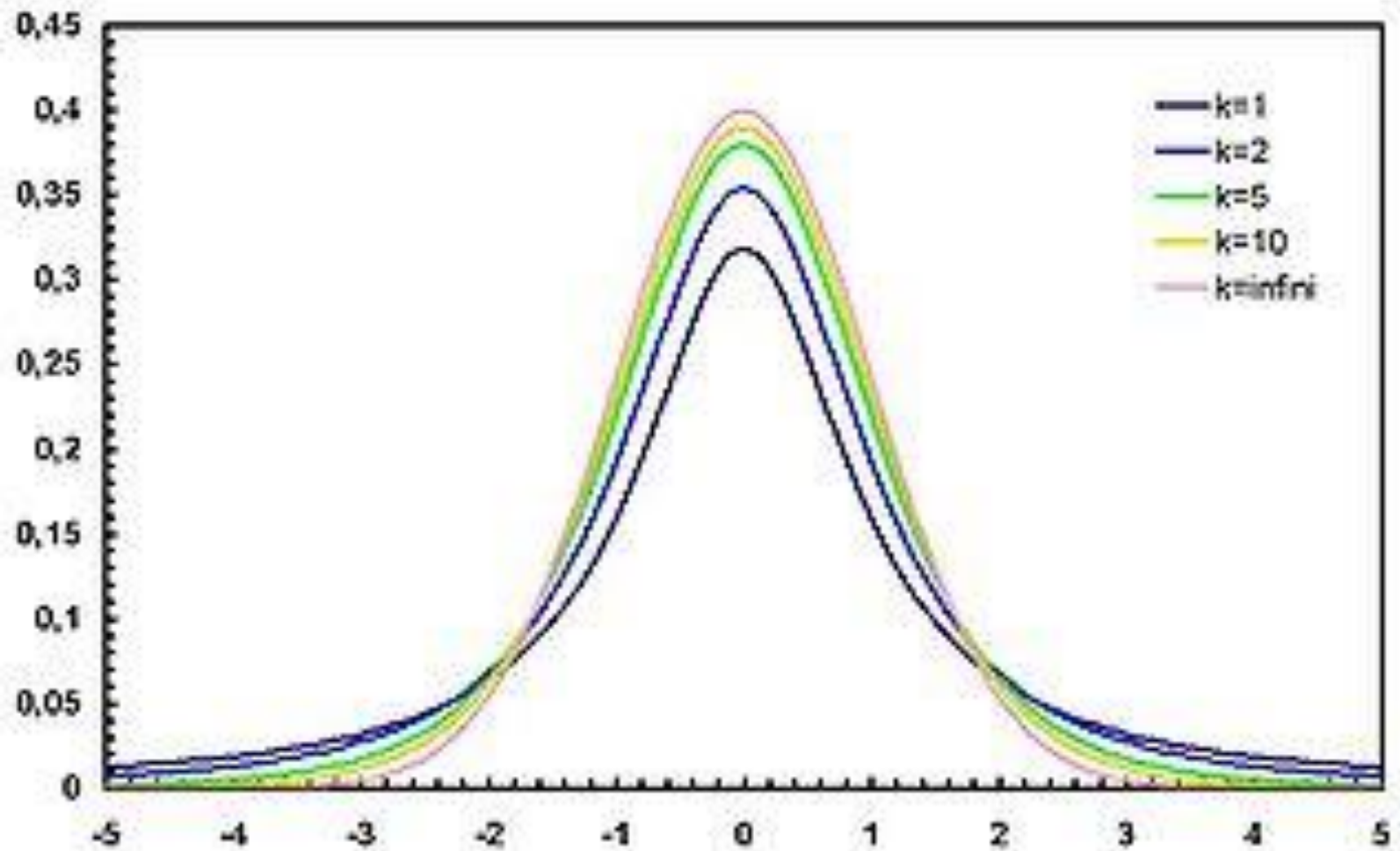
$$t = \frac{M - \mu}{S / \sqrt{n}}$$

it is the Z when  
you cant know  
the sigma and it  
is not normally  
distributed

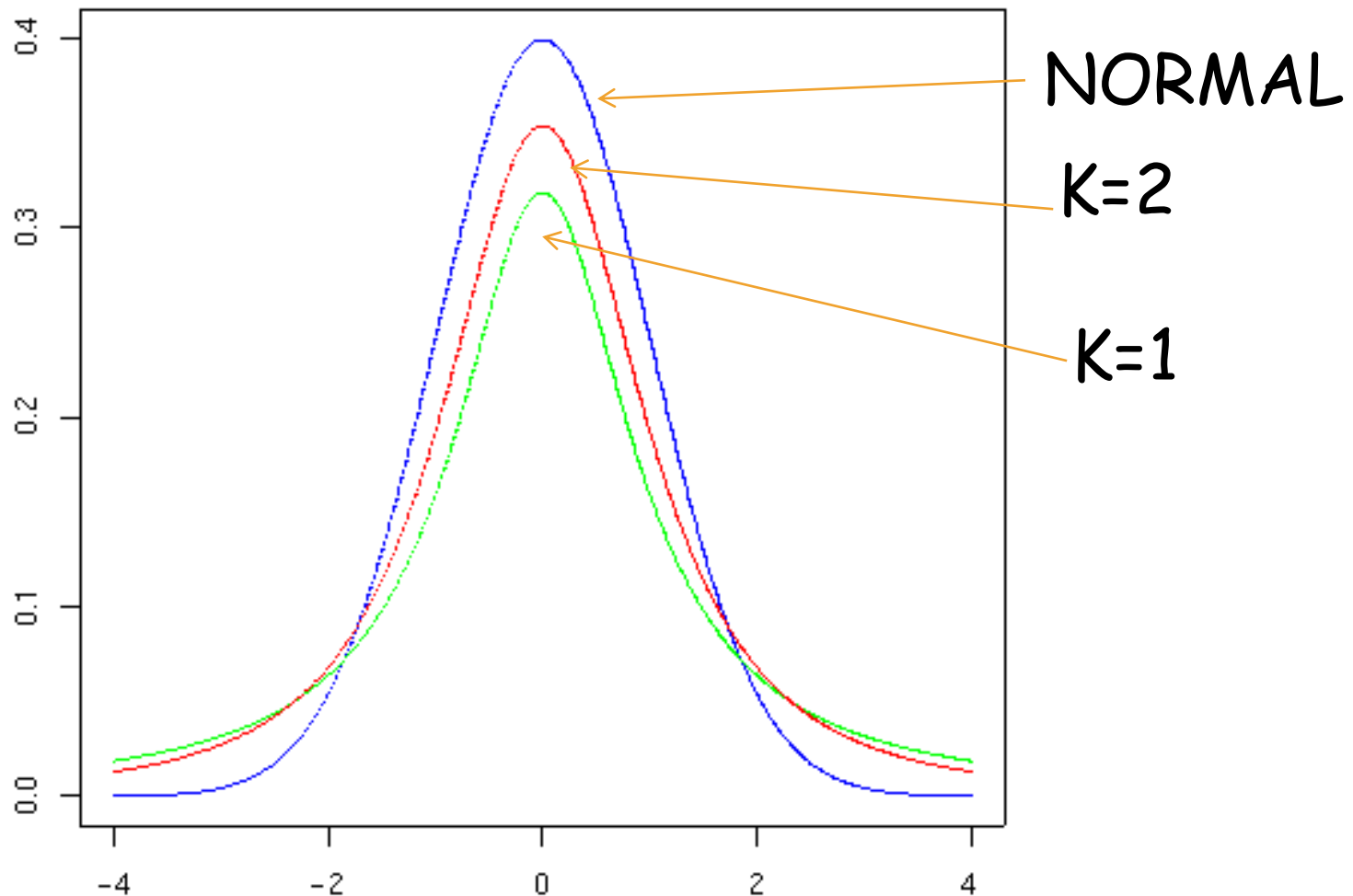
that depends only on the computed sample mean ( $M$ ) and computed sample standard deviation ( $S$ ), follows the Student's t-distribution with  $n-1$  degrees of freedom.

$$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

$\nu$  : degrees of freedom



# Normal Vs t- distributions



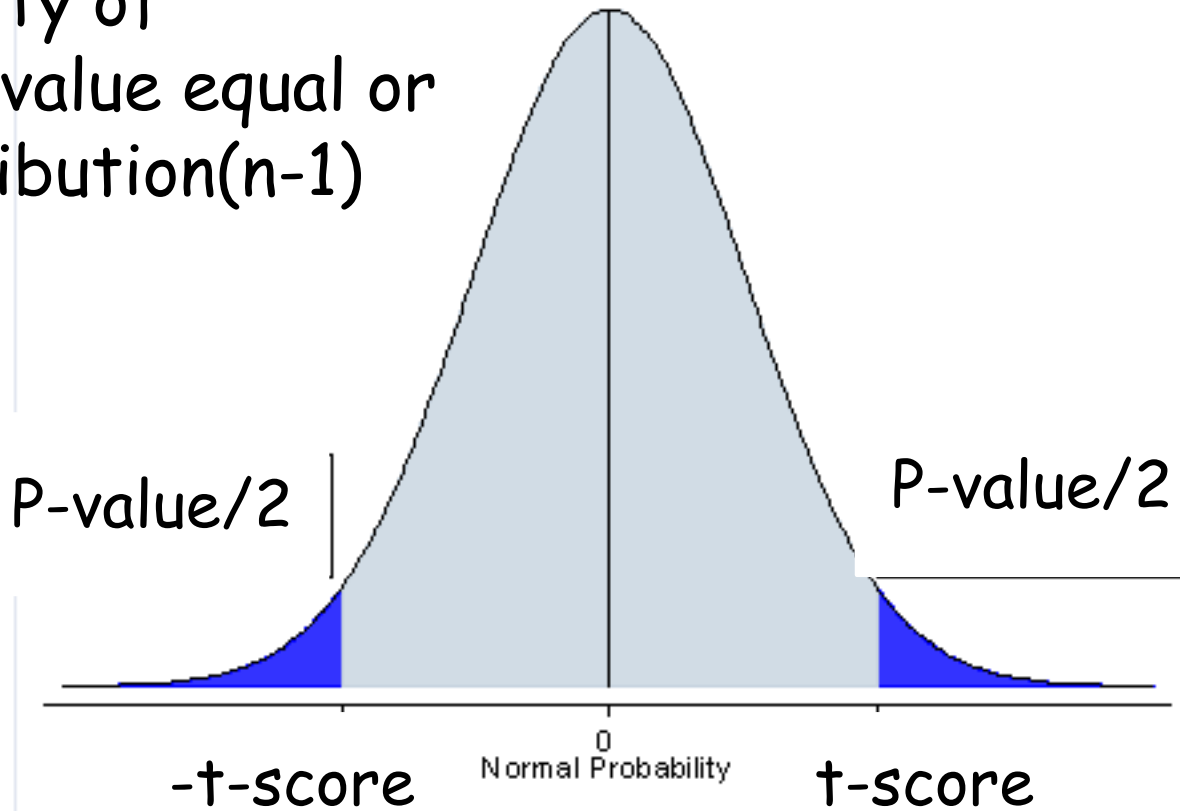
# Student's T-test (two-tailed)

□ Compute

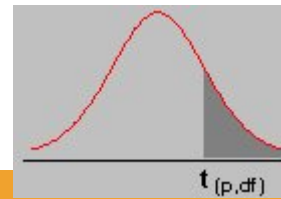
$$t = \frac{M - \mu}{S / \sqrt{n}}$$

compute the probability of obtaining an absolute value equal or higher, on the t-distribution(n-1)

□  
If that probability is lower than the desired critical P-value (e.g. 0.05) we can reject H0



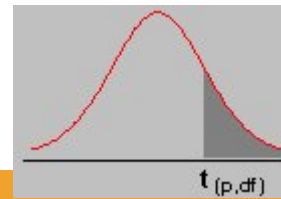
# Critical values



df \ p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869



# Critical values



df \ p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
<b>11</b>	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
<b>12</b>	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	4.3178
<b>13</b>	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
<b>14</b>	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
<b>15</b>	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
<b>16</b>	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
<b>17</b>	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
<b>18</b>	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
<b>19</b>	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
<b>20</b>	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495

# Comparing the means of two samples

*Equal sample size, assumption: same variance*

The t-score to test whether the means are different can be calculated as follows:

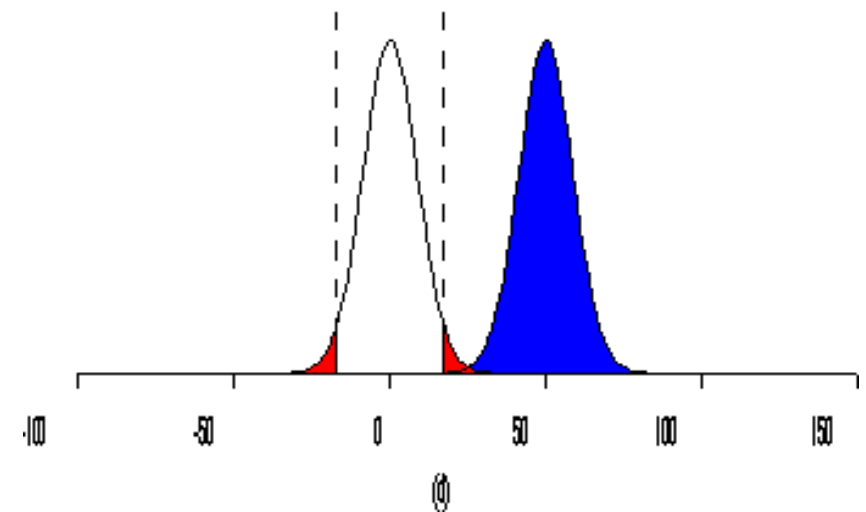
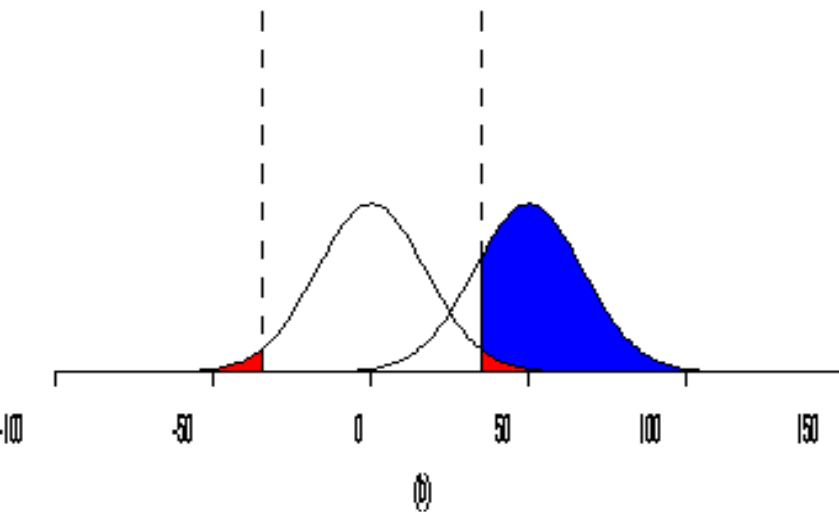
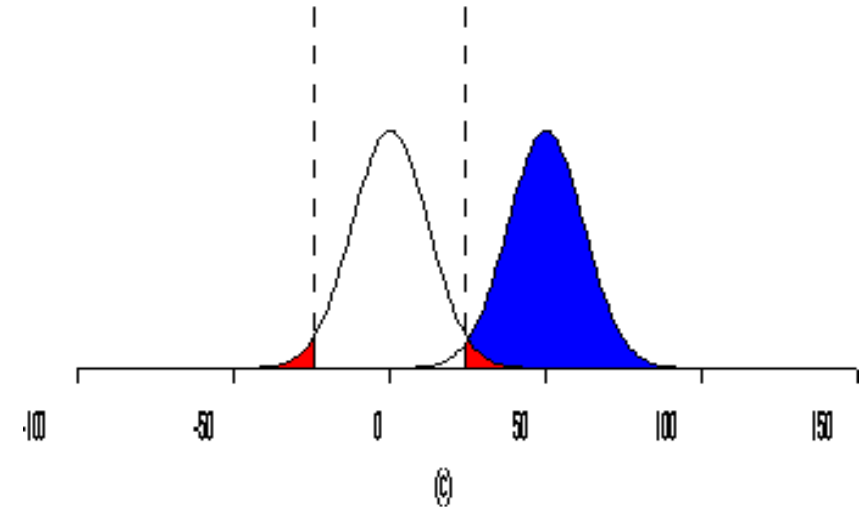
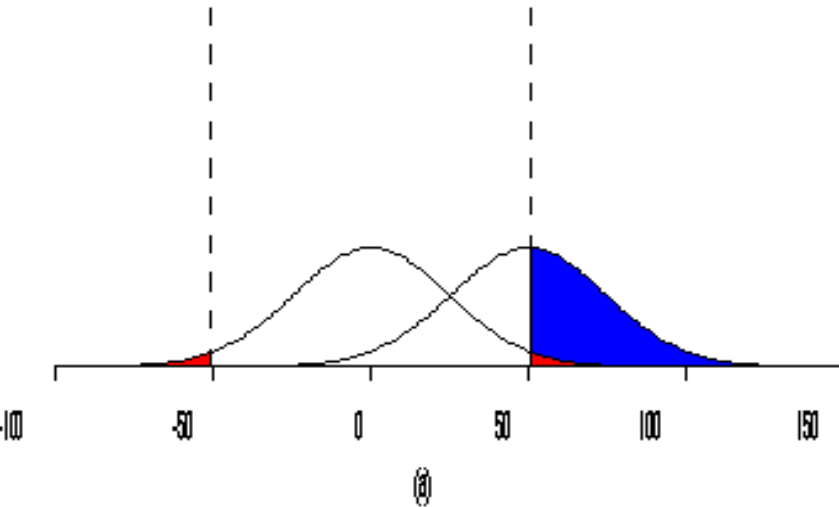
$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1 X_2} \cdot \sqrt{\frac{2}{n}}}$$

where

$$S_{X_1 X_2} = \sqrt{\frac{S_{X_1}^2 + S_{X_2}^2}{2}}.$$

is the or pooled standard deviation. The t-score, if  $H_0$  holds, is distributed as a t-distribution with  $2n-2$  degrees of freedom, where  $n$  is the sample size.

# Mean difference Vs deviation



# Example: Comparing two predictors

Set	Error rate 1	Error rate 2
1	0.2	0.4
2	0.18	0.25
3	0.24	0.38
4	0.14	0.27
5	0.05	0.31
6	0.26	0.21
7	0.15	0.32
8	0.29	0.38
9	0.31	0.28
10	0.35	0.33

# Example: Comparing two methods

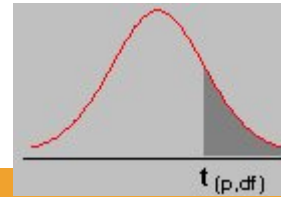
- $X_1=0.217$        $S_1^2=0.0082$
- $X_2=0.313$        $S_2^2=0.0038$

$$S_{12}=0.077$$

$$T=-2.76$$

Is the difference significant at 10%? At 1%?

# Critical values



df \ p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869

# Comparing the means of two samples

*Different sample size, assumption: same variance*

The t-score to test whether the means are different can be calculated as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1 X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$S_{X_1 X_2} = \sqrt{\frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{n_1 + n_2 - 2}}.$$

is the or pooled standard deviation. The t-score, if  $H_0$  holds, is distributed as a t-distribution with  $n_1 + n_2 - 2$  degrees of freedom, where  $n_1$  and  $n_2$  are the sample sizes.

# Comparing the means of two samples

*Different sample size, assumption: different variances*

The t-score to test whether the means are different can be calculated as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$$

where

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

is the or pooled standard deviation. The t-score, if  $H_0$  holds, is distributed as a t-distribution with approximately DF degrees of freedom

$$\text{D.F.} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}.$$



# Comparing more than 2 means: ANOVA

- One-way ANOVA is used to test for differences among two or more independent groups.
- When there are only two means to compare, the T-test and the F-test are equivalent; the relation between ANOVA and  $t$  is given by  $F = t^2$ .

# Does $S^2$ significantly differ from $\sigma^2$ ?

- Null hypothesis  $H_0: S = \sigma$
- Alternative hypothesis  $H_a: S \neq \sigma$

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

We need to find the distribution of  $S^2$  when  $H_0$  is true

# Distribution of the sample deviations

- If  $X_i$  are  $n$  independent, normally distributed random variables with mean  $\mu$  and variance  $\sigma^2$ , then the random variable

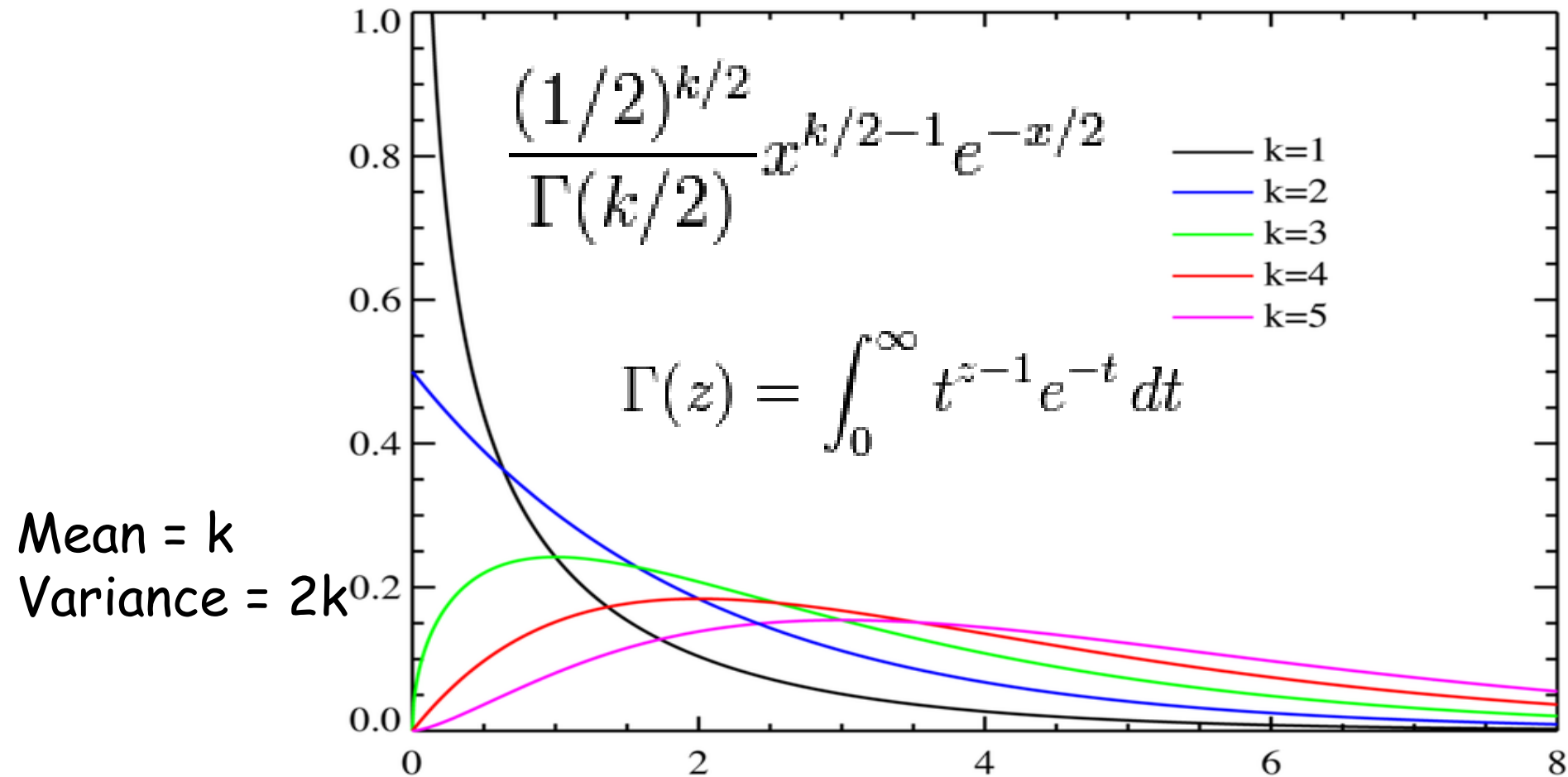
$$Q = \sum_{i=1}^N (X_i - \mu)^2$$

is distributed according to the chi-square distribution with  $n-1$  degrees of freedom. This is usually written as:

$$Q \approx \sigma^2 \chi_{n-1}^2$$

The chi-square distribution has one parameter:  $(n-1)$  that is a positive integer that specifies the number of degrees of freedom (i.e. the number of independent  $X_i - \mu$ )

# Chi-square distribution



# Distribution of the sample deviations

- So, being the unbiased sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - M)^2$$

is distributed according to the chi-square distribution with  $k$  degrees of freedom.

$$S^2 \approx \frac{\sigma^2}{n-1} \chi_{n-1}^2$$

It follows that

$$E[S^2] \approx \frac{\sigma^2}{n-1} E[\chi_{n-1}^2] = \frac{\sigma^2}{n-1} (n-1) = \sigma^2$$

Unbiased

# Does $S^2$ significantly differ from $\sigma^2$ ?

- If the samples  $X_i$  are normally distributed with mean  $M$  and variance  $\sigma^2$  (Null hypothesis)

$$\chi_i = \frac{(X_i - M)}{\sigma}$$

is normally distributed with mean 0 and variance 1

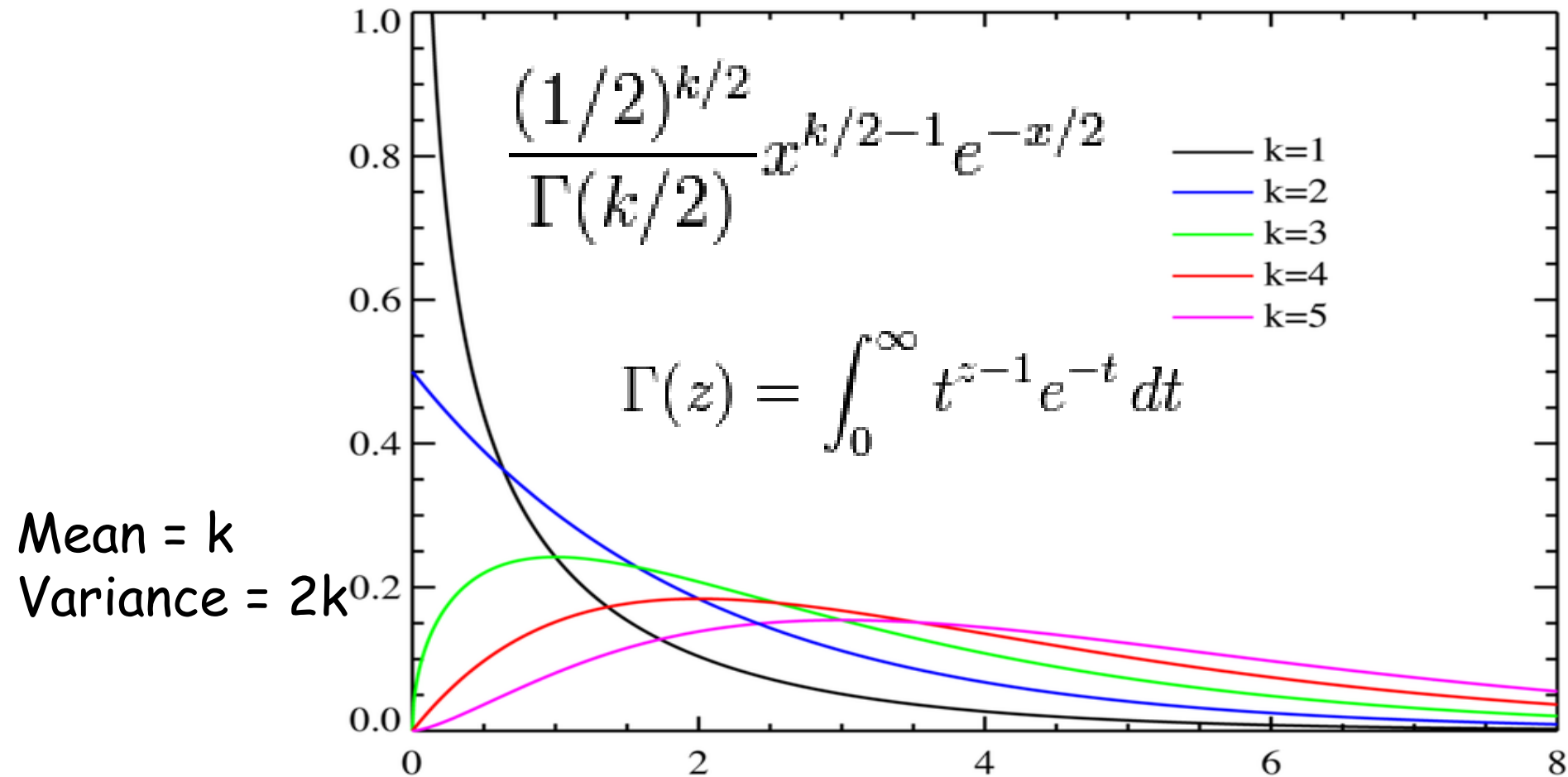
$$\sum_{i=1}^n \chi_i^2 = \sum_{i=1}^n \frac{(X_i - M)^2}{\sigma^2} = \frac{n-1}{\sigma^2} S^2$$

is distributed as a **chi-square distribution** with  **$n-1$  degrees of freedom.**

1 constraint holds:

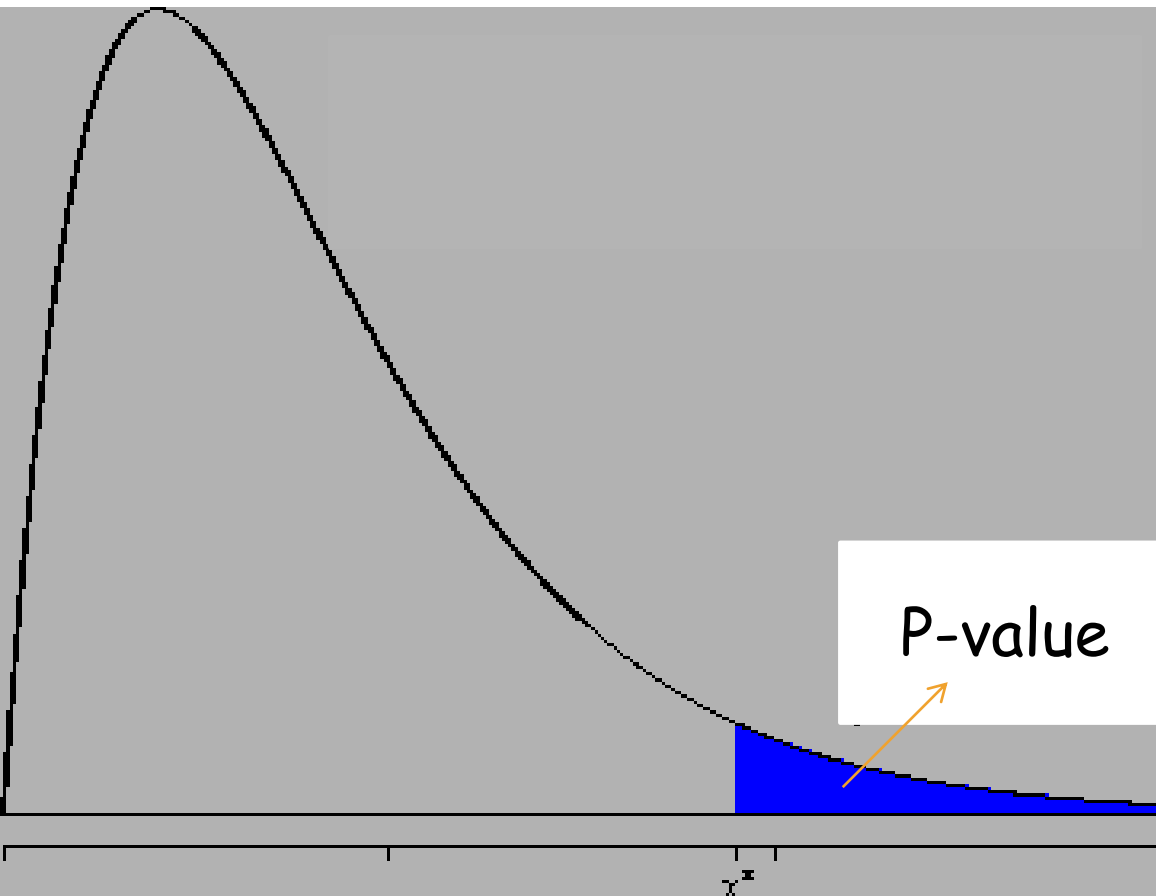
$$\sum_{i=1}^n \chi_i = 0$$

# Chi-square distribution



# Chi-square test

Compute  $\frac{n-1}{\sigma^2} S^2 = \sum_{i=1}^n \frac{(X_i - M)^2}{\sigma^2}$



and compute the probability of obtaining a value equal or higher on the chi-square distribution with  $n-1$  degrees of freedom. If that probability is lower than the desired critical P-value reject  $H_0$ .



# Critical values

$\chi^2$

df\area	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00004	0.00016	0.00098	0.00393	0.01579	0.10153	0.45494	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944
2	0.01003	0.02010	0.05064	0.10259	0.21072	0.57536	1.38629	2.77259	4.60517	5.99146	7.37776	9.21034	10.59663
3	0.07172	0.11483	0.21580	0.35185	0.58437	1.21253	2.36597	4.10834	6.25139	7.81473	9.34840	11.34487	12.83816
4	0.20699	0.29711	0.48442	0.71072	1.06362	1.92256	3.35669	5.38527	7.77944	9.48773	11.14329	13.27670	14.86026
5	0.41174	0.55430	0.83121	1.14548	1.61031	2.67460	4.35146	6.62568	9.23636	11.07050	12.83250	15.08627	16.74960
6	0.67573	0.87209	1.23734	1.63538	2.20413	3.45460	5.34812	7.84080	10.64464	12.59159	14.44938	16.81189	18.54758
7	0.98926	1.23904	1.68987	2.16735	2.83311	4.25485	6.34581	9.03715	12.01704	14.06714	16.01276	18.47531	20.27774
8	1.34441	1.64650	2.17973	2.73264	3.48954	5.07064	7.34412	10.21885	13.36157	15.50731	17.53455	20.09024	21.95495
9	1.73493	2.08790	2.70039	3.32511	4.16816	5.89883	8.34283	11.38875	14.68366	16.91898	19.02277	21.66599	23.58935
10	2.15586	2.55821	3.24697	3.94030	4.86518	6.73720	9.34182	12.54886	15.98718	18.30704	20.48318	23.20925	25.18818

If  $\frac{n-1}{\sigma^2} S^2 = \sum_{i=1}^n \frac{(X_i - M)^2}{\sigma^2} = 10$  and  $n = 5$ , can we say that  $S^2 \neq \sigma^2$  at significance 5%?

If  $\frac{n-1}{\sigma^2} S^2 = \sum_{i=1}^n \frac{(X_i - M)^2}{\sigma^2} = 10$  and  $n = 5$ , can we say that  $S^2 \neq \sigma^2$  at significance 0,5%?

If  $\frac{n-1}{\sigma^2} S^2 = \sum_{i=1}^n \frac{(X_i - M)^2}{\sigma^2} = 20$  and  $n = 10$ , can we say that  $S^2 \neq \sigma^2$  at significance 0,5%?

# TESTING HYPOTHESES

Testing distribution

# Testing distributions

- Is a frequency distribution of certain events observed in a sample consistent with a particular theoretical distribution?
- Null hypothesis: the distributions do not differ

A simple example is the hypothesis that an ordinary six-sided die is "fair", i.e., all six outcomes are equally likely to occur.

# Pearson's chi-square test

- The observed values ( $O_i$ ) and the expected values ( $E_i$ ) are compared over all the  $n$  classes (bins):

$$X^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i},$$

- This test variable approximately follows a chi-square statistics: degrees of freedoms are:  $n-1-s$ , where  $s$  is the number of parameters describing the theoretical distribution (e.g. 2 for normal distribution, 1 for binomial distribution)

# TESTING HYPOTHESES

Comparing more than tw means: ANOVA

# Comparing more than 2 means: ANOVA

- One-way ANOVA is used to test for differences among two or more independent groups.
- When there are only two means to compare, the T-test and the ANOVA (F-test) are equivalent; the relation between ANOVA and t is given by  $F = t^2$ .

# Experiment

---

- Investigator observes the effect of different treatments (each one containing two or more levels ) on the dependent variable
- Dependencies of the observed variable on the treatment are searched.
- Null hypothesis: no dependency

# Examples of Experiments

- Level of expression of a gene (dependent variable) is investigated in mice of different strains (treatment 1), or of different ages (treatment 2) or differently fed (treatment 3). Are the treatments correlated to the expression?



# One-Way ANOVA F-Test

- Tests the equality of  $p$  population means
  - ▣ One Treatment: One nominal independent variable
    - 2 or more ( $p$ ) treatment levels or classifications (e.g Diet A, Diet B, Diet C)
  - ▣ One interval or ratio scaled dependent variable

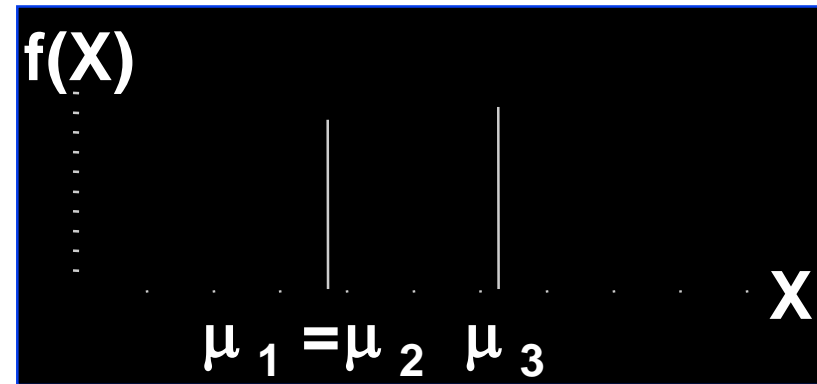
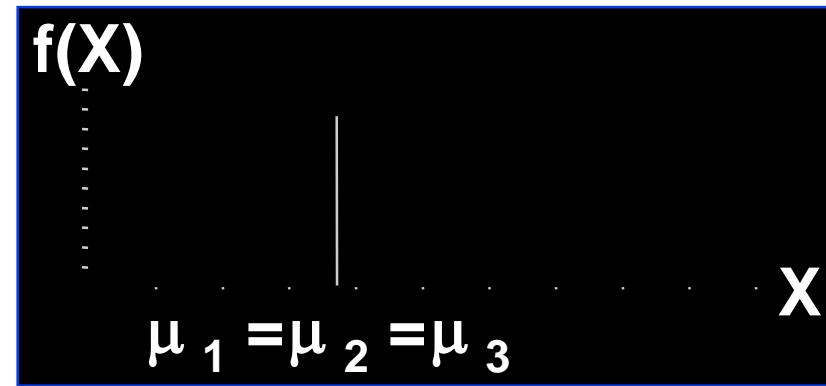
# One-Way ANOVA F-Test Assumptions

- Randomness & independence of errors
  - ▣ Independent random samples are drawn for each condition
- Normality
  - ▣ Populations (for each condition) are normally distributed
- Homogeneity of variance
  - ▣ Populations (for each condition) have equal variances

# One-Way ANOVA F-Test Hypotheses

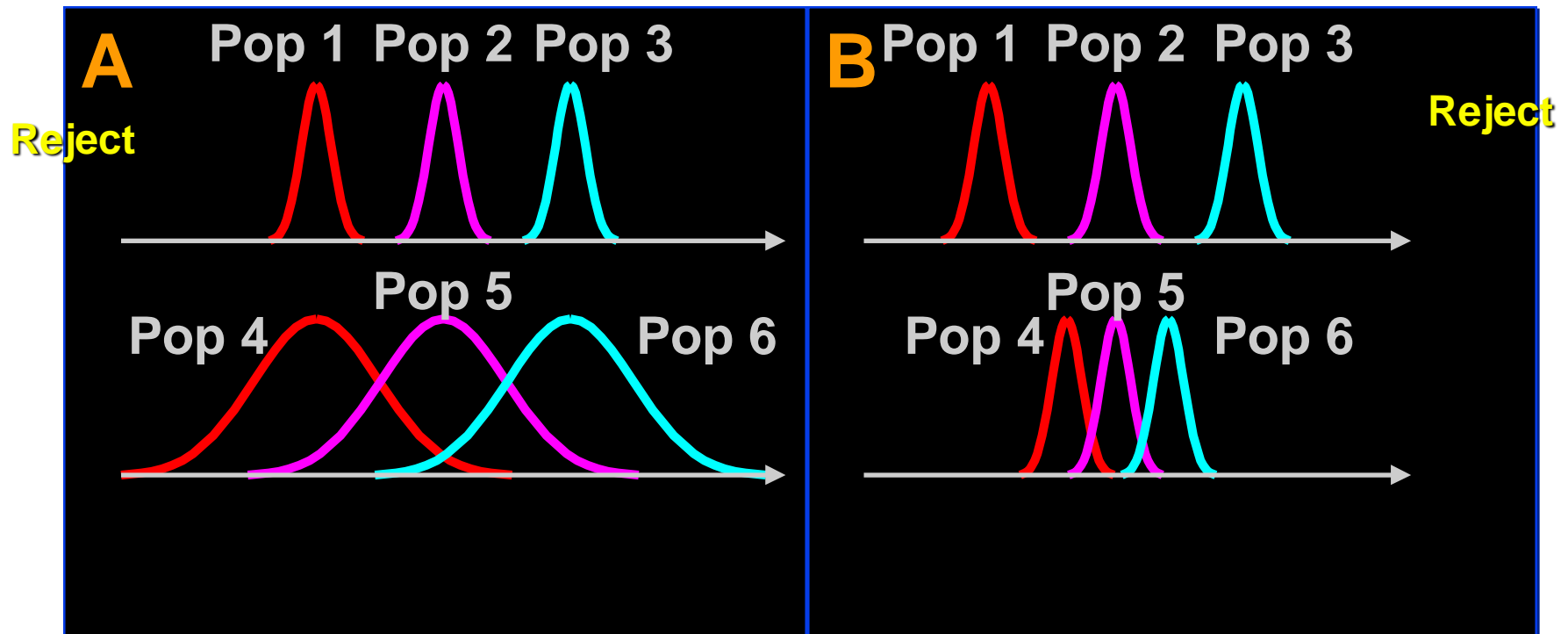
rejecting the null hypothesis means rejecting that ALL THE MEANS ARE EQUAL

- $H_0: \mu_1 = \mu_2 = \dots = \mu_p$ 
  - ▣ All population means are equal
  - ▣ No treatment effect
- $H_a: \text{not all } \mu_j \text{ are equal}$ 
  - ▣ At least 1 pop. Mean is different
  - ▣ Treatment effect
  - ▣ NOT  $\mu_1 \neq \mu_2 \neq \dots \neq \mu_p$



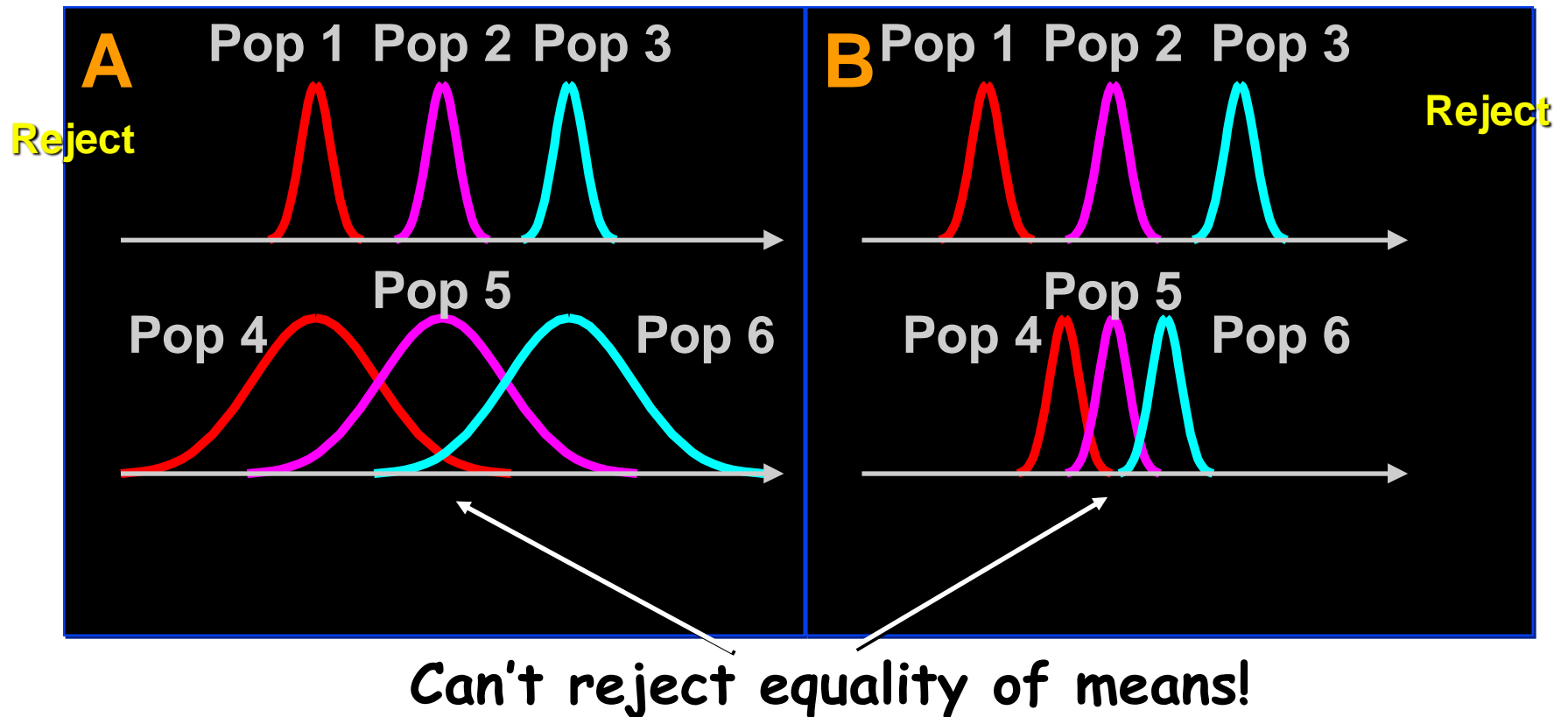
# Comparing means with respect to variance

- Same treatment variation      Different treatment variation
- Different random variation      Same random variation



# Comparing means with respect to variance

- Same treatment variation      Different treatment variation
- Different random variation      Same random variation



# One-Way ANOVA

## Basic Idea

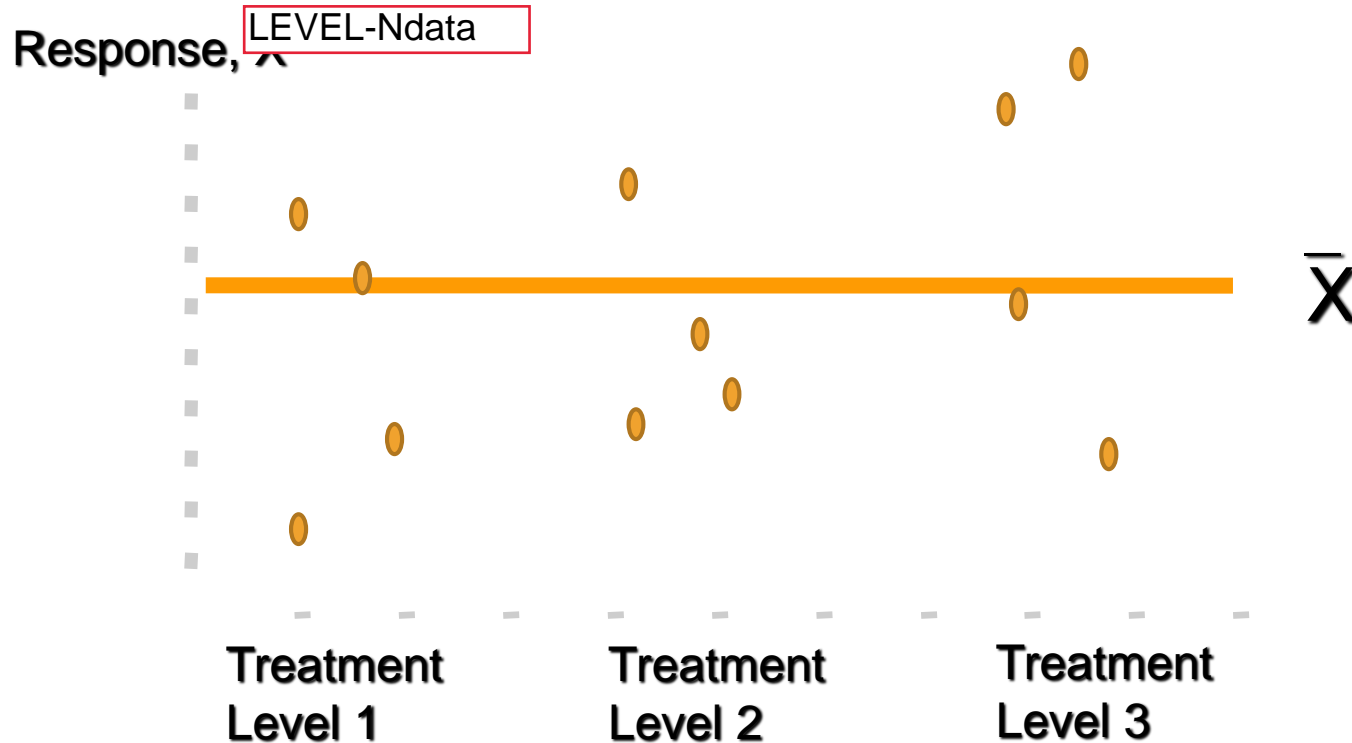
---

1. Compares 2 types of variation (variations between and within treatment levels) to test equality of means
2. Comparison based on ratio of variances
3. If variation between treatment levels is significantly greater than variation within levels then means are not equal
4. Variation measures are obtained by 'partitioning' total variation

# Total Variation

Each value is compared to the global mean of the data, irrespectively of the treatment

$$SS(Total) = (X_{11} - \bar{X})^2 + (X_{21} - \bar{X})^2 + \dots + (X_{ij} - \bar{X})^2$$



# One-Way ANOVA

## Partitions Total Variation



Total variation

Variation due to  
treatment

- Sum of Squares Between  
**SSB**
- Sum of Squares  
Treatment
- Among Groups Variation

Variation due to random  
sampling

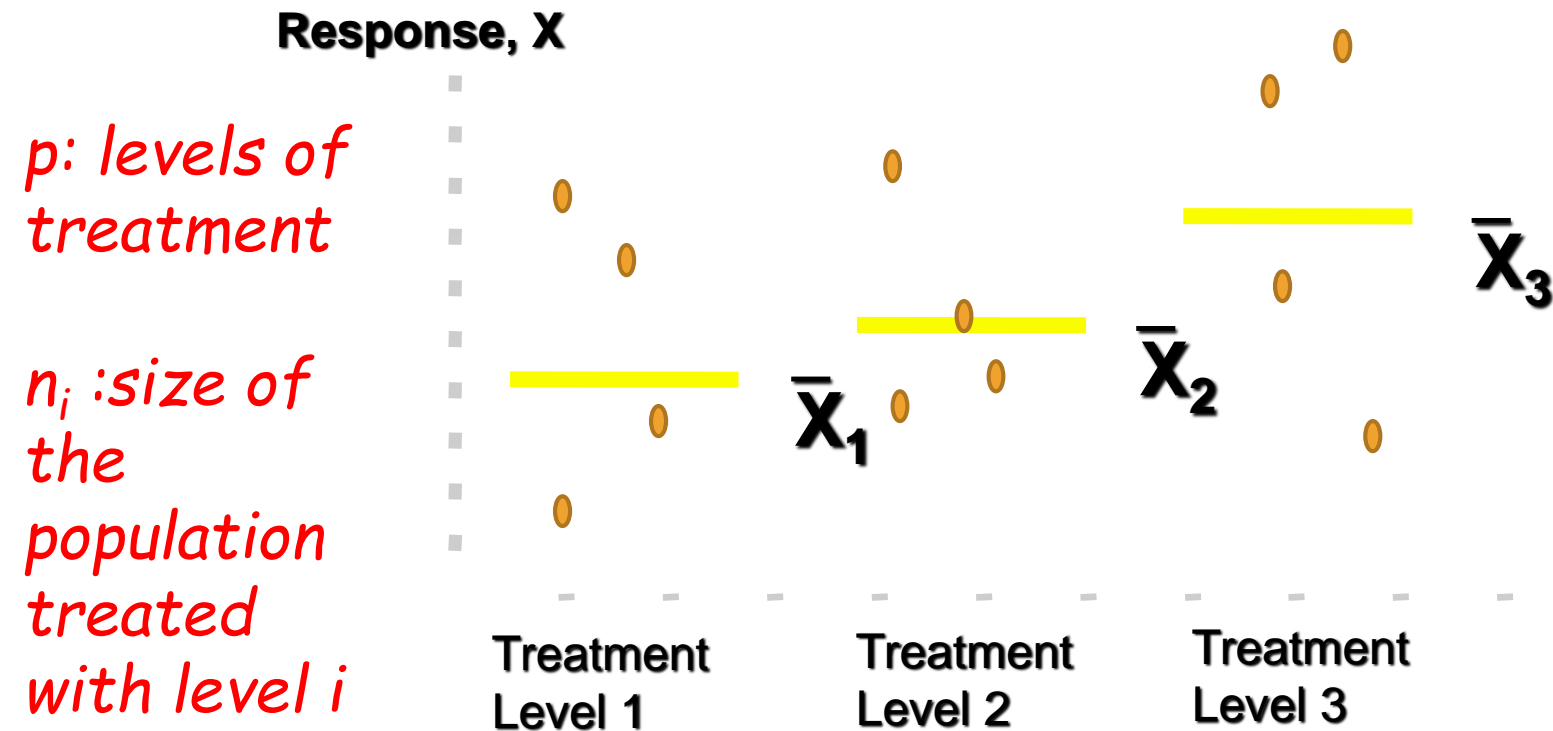
- Sum of Squares Within  
**SSW**
- Sum of Squares Error
- Within Groups Variation



# Random Variation or Variation Within

Each value is compared to the mean of its treatment level

$$SSW = (X_{11} - \bar{X}_1)^2 + (X_{12} - \bar{X}_1)^2 + \dots + (X_{pj} - \bar{X}_p)^2$$



# Treatment Variation or Variation between

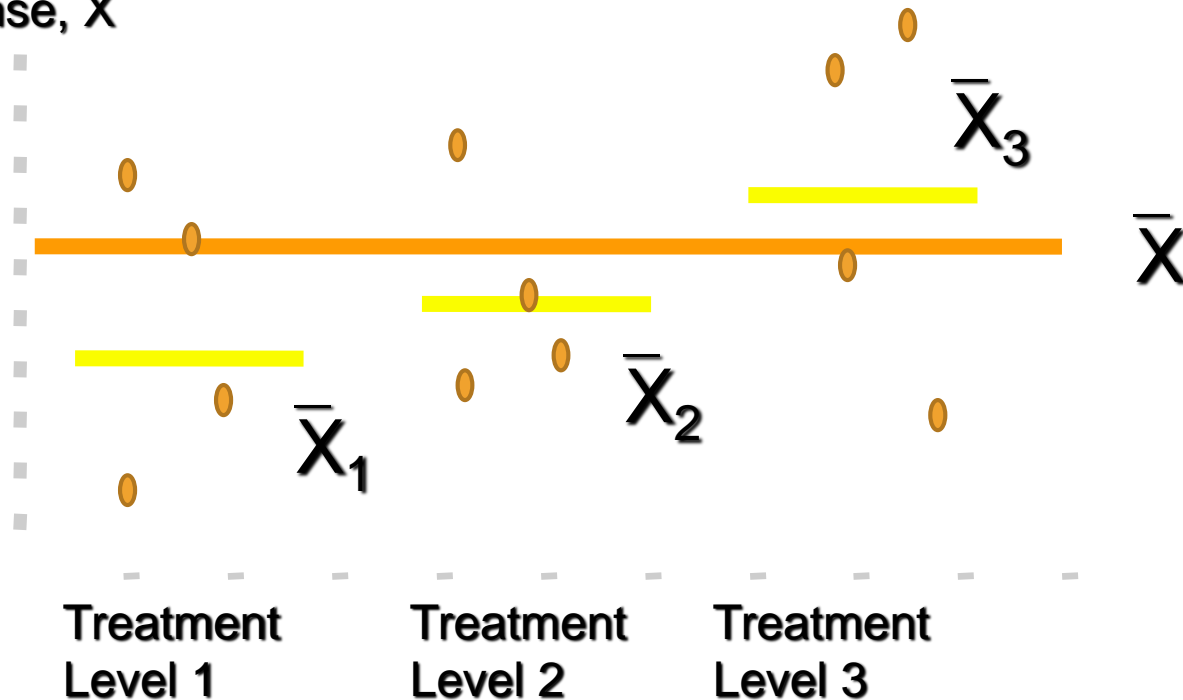
The means on each treatment level are compared to the global mean (weighting by the number of values in each treatment level)

$$SSB = n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 + \dots + n_p(\bar{X}_p - \bar{X})^2$$

Response, X

*p: levels of treatment*

*n<sub>i</sub>: size of the population treated with level i*



# Theorem: $SS = SSB + SSW$

$$SS(Total) = \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$$

**p: levels of treatment**  
 **$n_i$ : size of the population treated with level i**

$$= \sum_{i=1}^p \sum_{j=1}^{n_i} \left[ (X_{ij} - \bar{X}_i) + (\bar{X}_i - \bar{X}) \right]^2 =$$

$$= \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 + \sum_{i=1}^p \sum_{j=1}^{n_i} (\bar{X}_i - \bar{X})^2$$

$$+ 2 \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(\bar{X}_i - \bar{X})$$

# But

$$\begin{aligned} & \sum_{i=1}^p \sum_{j=1}^{n_i} (\bar{X}_i - \bar{X}) (X_{ij} - \bar{X}_i) \\ &= \sum_{i=1}^p \left[ (\bar{X}_i - \bar{X}) \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i) \right] \\ &= \sum_{i=1}^p (\bar{X}_i - \bar{X}) (n\bar{X}_i - n\bar{X}_i) \\ &= 0 \end{aligned}$$

**Thus,  $SS = SSB + SSW$**

$$\begin{aligned} SS(Total) &= \sum_{i=1}^p \sum_{j=1}^{n_i} \left( X_{ij} - \bar{X}_i \right)^2 + \sum_{i=1}^p \sum_{j=1}^{n_i} \left( \bar{X}_i - \bar{X} \right)^2 \\ &= \sum_{i=1}^p \sum_{j=1}^{n_i} \left( X_{ij} - \bar{X}_i \right)^2 + \sum_{i=1}^p n_i \left( \bar{X}_i - \bar{X} \right)^2 \\ &= SSW + SSB \end{aligned}$$

# F-test



As in the case of t-test we want to evaluate the variability of the means in different treatment levels with the random variability in order to evaluate the significance of the null hypothesis (no difference between means)

# F-test

- 1) Variability of the means is evaluated with the mean squared deviation Between groups

$$MSB = \frac{SSB}{Dof(Between)} = \frac{SSB}{p-1}$$

P=number of level

- 2) Random variability is evaluated as the mean squared deviation Within

$$MSW = \frac{SSW}{Dof(Within)} = \frac{SSW}{\sum_{i=1}^p (n_i - 1)} = \frac{SSW}{n_{tot} - p}$$

**Dof: Degrees of freedom**

# F-Test : Test Statistic

- The Test variable F is defined

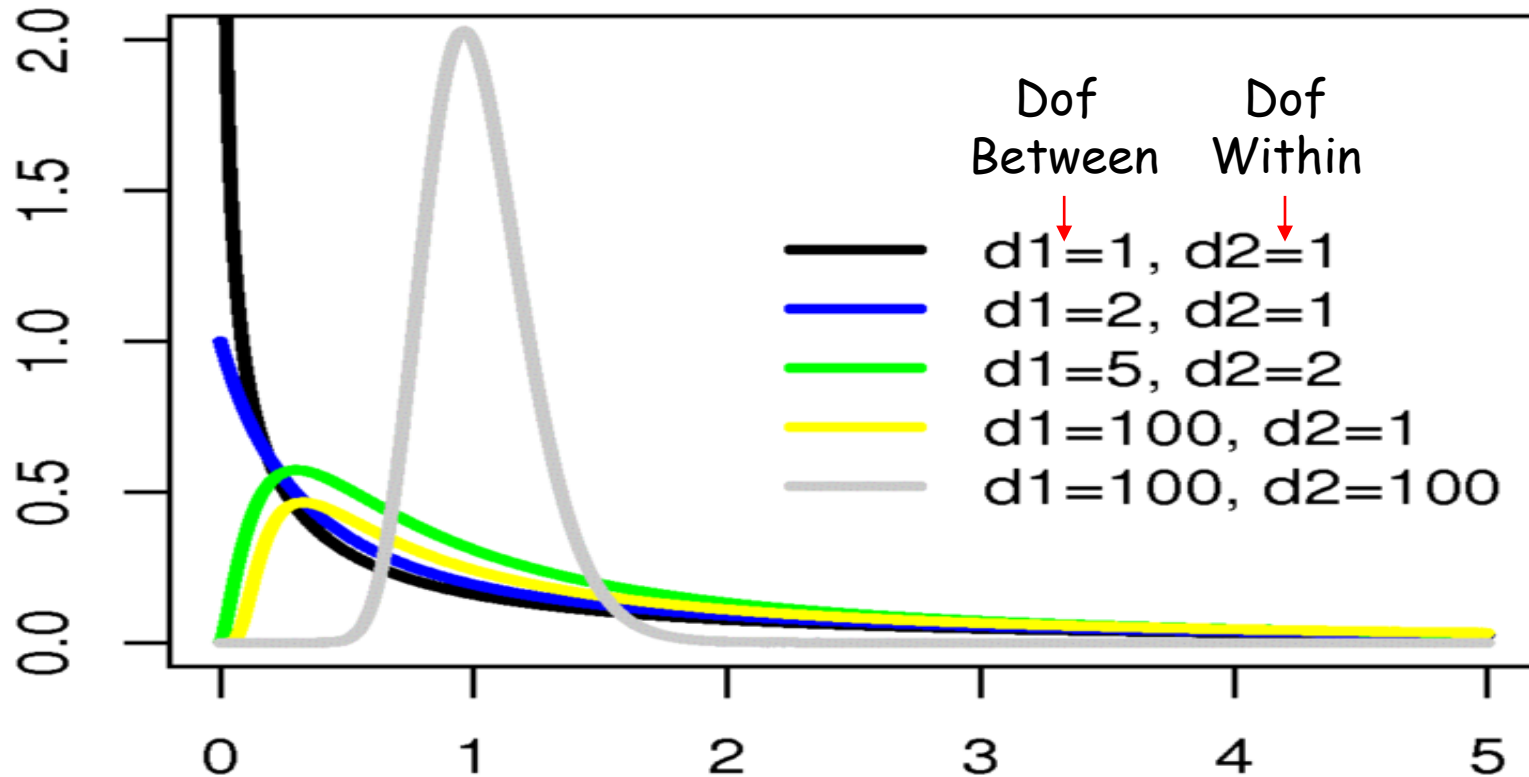
$$F = \frac{MSB}{MSW}$$

with p-1 degrees of freedom Between and n-p degrees of freedom Within

It follows a Fisher-Snedecor F-distribution



# F-distribution



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

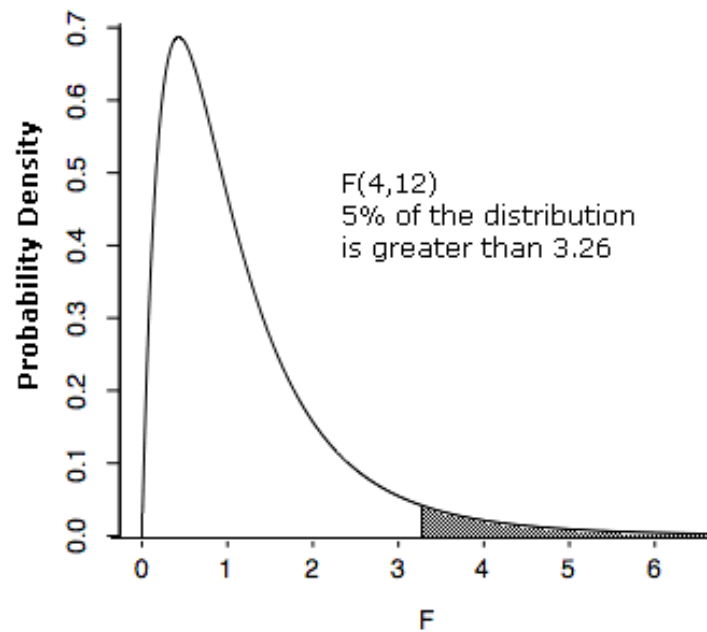
$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$$

# F-test

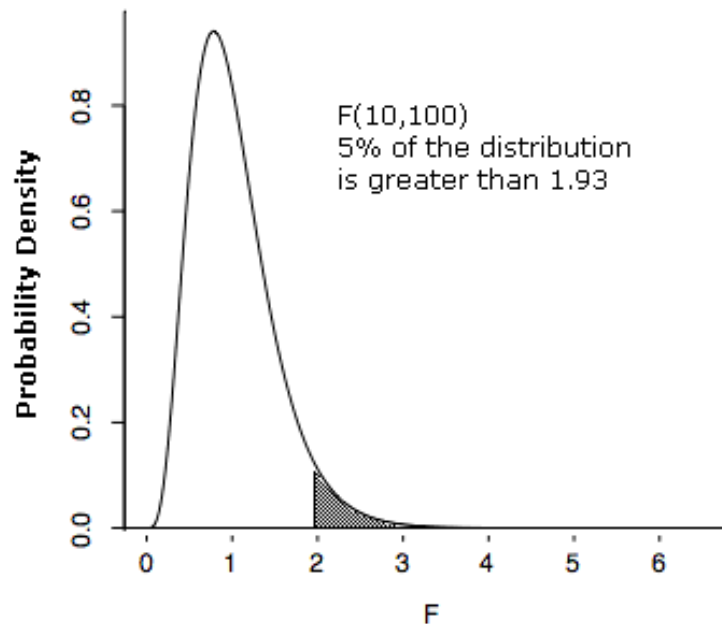
- If the value of  $F$  is 1, or lower, the variability of the means between groups is equal to, or lower than, the random variability:

The null hypothesis cannot be rejected

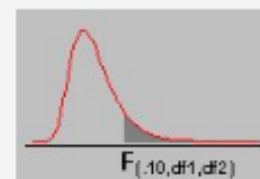
NB



The null hypothesis can be rejected only if  $F$  overcomes the critical value than depends on the desired significance (e.g. 0.05)

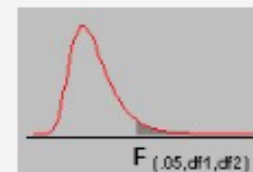


F Table for alpha=.10



df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20
1	39.86346	49.50000	53.59324	55.83296	57.24008	58.20442	58.90595	59.43898	59.85759	60.19498	60.70521	61.22034	61.74029
2	8.52632	9.00000	9.16179	9.24342	9.29263	9.32553	9.34908	9.36677	9.38054	9.39157	9.40813	9.42471	9.44131
3	5.53832	5.46238	5.39077	5.34264	5.30916	5.28473	5.26619	5.25167	5.24000	5.23041	5.21562	5.20031	5.18448
4	4.54477	4.32456	4.19086	4.10725	4.05058	4.00975	3.97897	3.95494	3.93567	3.91988	3.89553	3.87036	3.84434
5	4.06042	3.77972	3.61948	3.52020	3.45298	3.40451	3.36790	3.33928	3.31628	3.29740	3.26824	3.23801	3.20665
6	3.77595	3.46330	3.28876	3.18076	3.10751	3.05455	3.01446	2.98304	2.95774	2.93693	2.90472	2.87122	2.83634
7	3.58943	3.25744	3.07407	2.96053	2.88334	2.82739	2.78493	2.75158	2.72468	2.70251	2.66811	2.63223	2.59473
8	3.45792	3.11312	2.92380	2.80643	2.72645	2.66833	2.62413	2.58935	2.56124	2.53804	2.50196	2.46422	2.42464
9	3.36030	3.00645	2.81286	2.69268	2.61061	2.55086	2.50531	2.46941	2.44034	2.41632	2.37888	2.33962	2.29832
10	3.28502	2.92447	2.72767	2.60534	2.52164	2.46058	2.41397	2.37715	2.34731	2.32260	2.28405	2.24351	2.20074

F Table for  $\alpha=.05$



df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20
1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	243.9060	245.9499	248.0131
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4125	19.4291	19.4458
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740

# One-Way ANOVA Summary Table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square (Variance)	F
Treatment	$p - 1$	SSB	$MSB = SSB / (p - 1)$	<u>MSB/MSW</u>
Noise	$n - p$	SSW	$MSW = SSW / (n - p)$	
Total	$n - 1$	$SS(\text{Total}) = SSB + SSW$		

# Two-Way ANOVA

- Tests the equality of the means of an observed variable when two independent variables (treatments) are under investigation
- Same results as separate one-way ANOVA on each treatment
- Interaction between treatments can be also tested

# Interaction

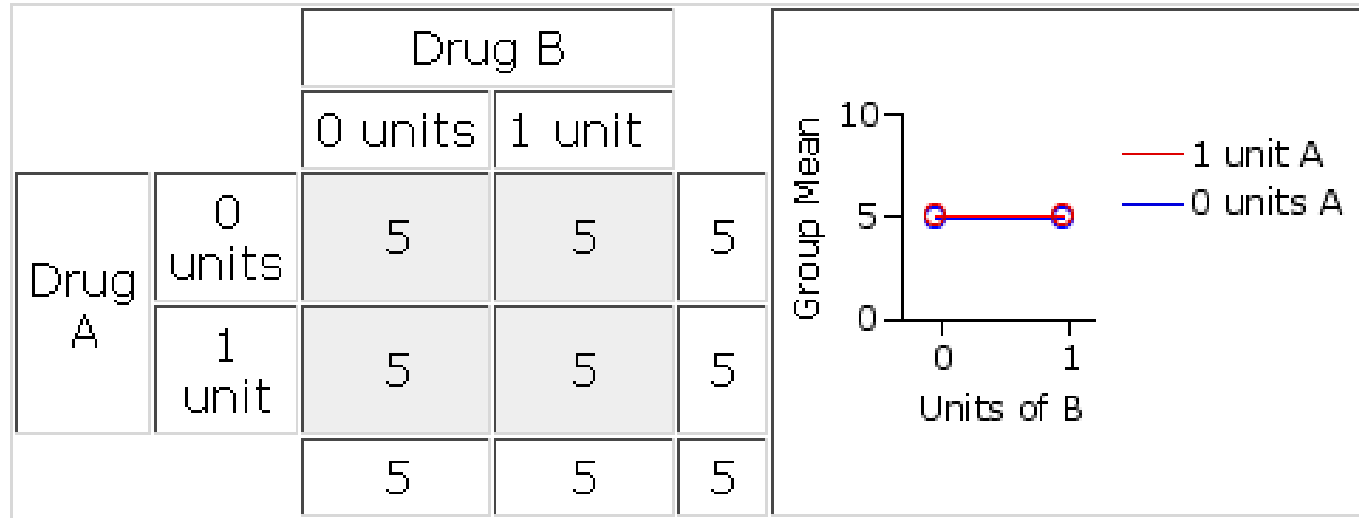
---

- 1. Occurs when effects of one factor vary according to levels of other factor
- 2. When significant, interpretation of main treatments(A & B) is complicated

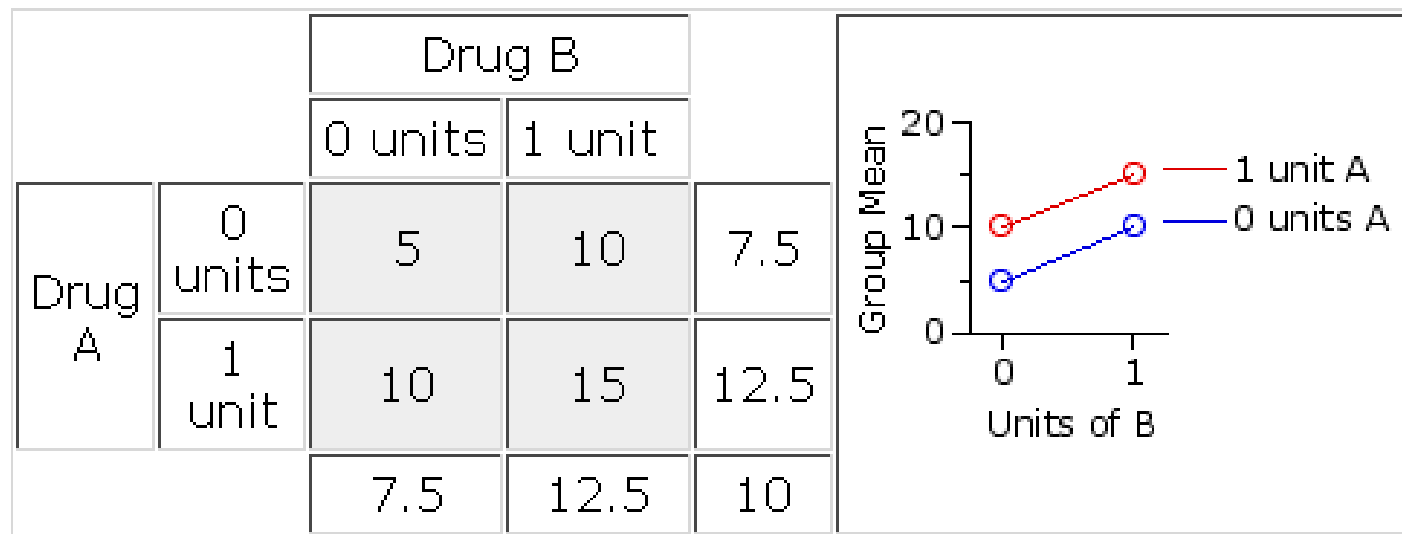


# Two-drugs experiment: no interaction

## Scenario 1



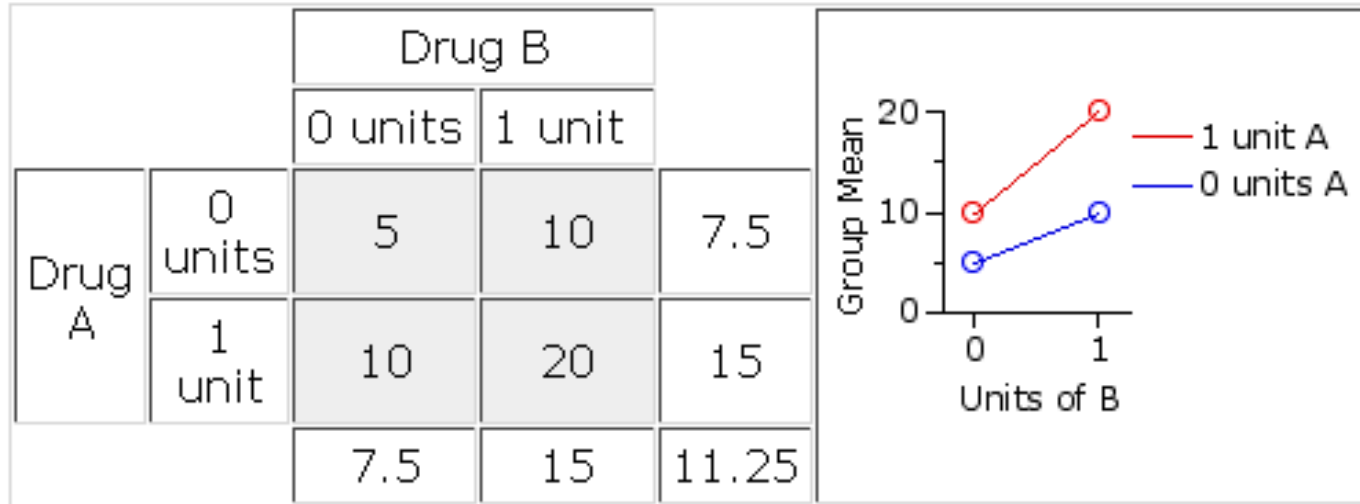
## Scenario 2



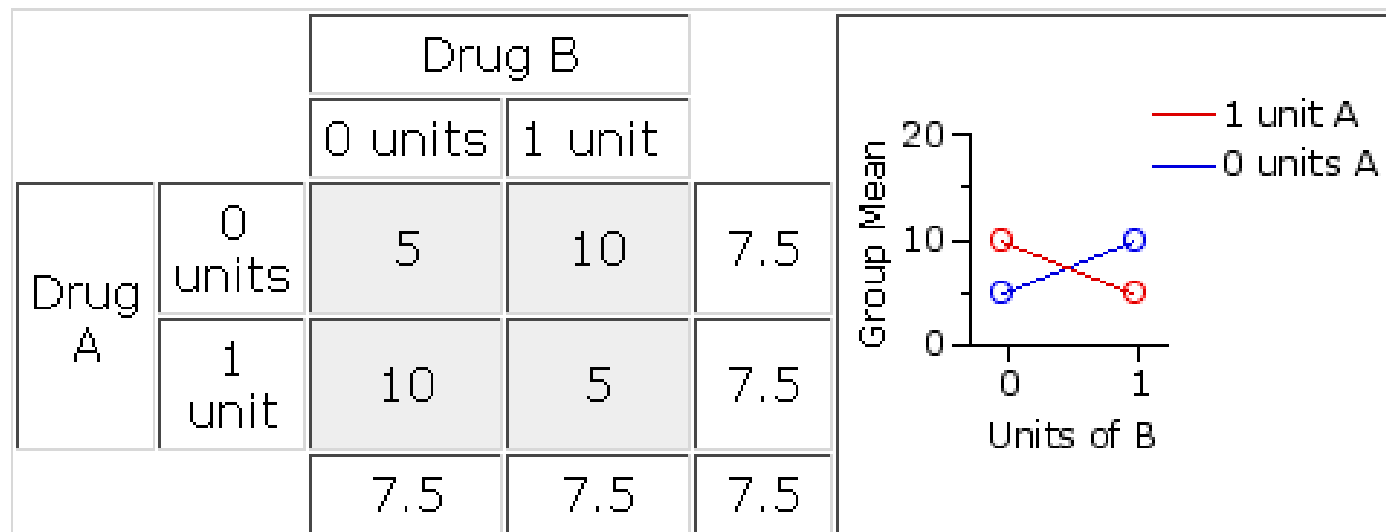
NB additive

# Two-drugs experiment: interaction

## Scenario 3



## Scenario 4



# Two-Way ANOVA Assumptions

- 1. Normality
  - ▣ Populations are normally distributed
- 2. Homogeneity of variance
  - ▣ Populations have equal variances
- 3. Independence of errors
  - ▣ Independent random samples are drawn

# Two-Way ANOVA

## Null Hypotheses

- 1. No difference in means due to factor A
  - $H_0: \mu_{1..} = \mu_{2..} = \dots = \mu_{A..}$
- 2. No difference in means due to factor B
  - $H_0: \mu_{.1.} = \mu_{.2.} = \dots = \mu_{.B.}$
- 3. No interaction of factors A & B
  - $H_0: ab_{ij} = 0$

ALL THE MEANS ARE EQUAL AND THERE IS NO INTERACTION

# NB Total Variation Partitioning

Sum of Squares Between,  
Variable A

Total Sum of Squares

$$SS(Total) = \sum_{i=1}^{p_A} \sum_j^{p_B} \sum_{k=1}^{n_{ij}} (X_{ijk} - \bar{X})^2$$

$$SSB(A) = \sum_{i=1}^{p_A} n_{A-level=i} (\bar{X}_{i-} - \bar{X})^2$$

Sum of Squares Between,  
Variable B

$$SSB(B) = \sum_{j=1}^{p_B} n_{B-level=j} (\bar{X}_{-j} - \bar{X})^2$$

Sum of Squares Within

$$SSW = \sum_{i=1}^{p_A} \sum_{j=1}^{p_B} \sum_{k=1}^{n_{ij}} (X_{ijk} - \bar{X}_{ij})^2$$

Sum of Squares Between,  
Interaction

$$SSB(AB) = \sum_{i=1}^{p_A} \sum_{j=1}^{p_B} n_{A-level=i, B-level=j} (\bar{X}_{ij} - \bar{X})^2$$

# Two-Way ANOVA Summary Table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
A (Row)	$a - 1$	SSB(A)	MSB(A)	<u>MSB(A)</u> /MSW
B (Column)	$b - 1$	SSB(B)	MSB(B)	<u>MSB(B)</u> /MSW
AB (Interaction)	$(a-1)(b-1)$	SSB(AB)	MSB(AB)	<u>MSB(AB)</u> /MSW
Noise	$n - ab$	SSW	MSW	
Total	$n - 1$	SS(Total)		

# The effect of many treatments can be tested

Folio: 2 Level Full Factorial Design (Analysis)

A1 ANOVA Table

	A	B	C	D	E	F
1	ANOVA Table					
2						
3	Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
4						
5	Model	15	10.555	0.7037		
6	A:Susceptor-Rotation Method	1	0.144	0.144	0.6062	0.4385
7	B:Nozzle Position	1	0.7145	0.7145	3.0078	0.0867
8	C:Deposition Temperature	1	0.1438	0.1438	0.6055	0.4388
9	D:Deposition Time	1	5.7634	5.7634	24.262	4.45E-06
10	AB	1	0.0013	0.0013	0.0056	0.9408
11	AC	1	0.2052	0.2052	0.8637	0.3555
12	AD	1	0.06	0.06	0.2526	0.6166
13	BC	1	0.0795	0.0795	0.3345	0.5646
14	BD	1	0.0217	0.0217	0.0914	0.7631
15	CD	1	2.8573	2.8573	12.0283	0.0008
16	ABC	1	0.2281	0.2281	0.9604	0.33
17	ABD	1	0.0151	0.0151	0.0634	0.8019
18	ACD	1	0.0216	0.0216	0.0909	0.7638
19	BCD	1	0.29	0.29	1.2206	0.2725
20	ABCD	1	0.0096	0.0096	0.0402	0.8415
21	Residual	80	19.0038	0.2375		
22	Pure Error	80	19.0038	0.2375		
23	Total	95	29.5588			

Main Options Other

Response  
Thickness

Transformation  
 $Y' = Y$

Risk Level (Alpha)  
0.1

Analysis Settings  
Partial SS Individual  
Calculated  
Observations = 96

Analysis Summary  
P()=... ...

Design Analysis Plot

# Analysis of expression data

Tabular Viewer | Color Mosaic | Display Preference

Marker Name	P-Value	F-statistic	non-GC B-ce...	non-GC B-ce...	GC B-cell_M...	GC B-cell_Std	non-GC Tum...	non-GC Tum...	GC-Tumor_...	GC-Tumor_Std
IGL@	0.00E+00	46.19	10.81	0.54	10.39	0.52	7.01	1.50	7.28	1.41
FCGRT	0.00E+00	44.13	8.57	0.60	6.16	0.98	9.47	0.84	7.89	1.10
TMSB10	0.00E+00	44.16	12.26	0.45	11.31	0.30	12.34	0.37	11.34	0.51
TXNIP	0.00E+00	73.74	12.01	0.32	9.40	0.65	12.11	0.64	9.91	1.07
TUBB2C	0.00E+00	51.76	10.22	0.59	12.13	0.41	9.38	0.95	11.25	0.69
TUBB2C	0.00E+00	57.97	10.33	0.36	11.68	0.24	9.86	0.55	11.07	0.52
OGG1	0.00E+00	57.01	8.72	0.22	9.52	0.18	7.99	0.34	8.61	0.49
KIF14	0.00E+00	51.11	6.45	0.92	9.69	0.25	5.52	1.58	8.49	0.99
RPL39L	0.00E+00	51.58	4.63	0.49	7.18	0.38	4.22	0.90	6.14	1.01
CR1	0.00E+00	58.19	8.37							
CCNF	0.00E+00	52.78	7.52							
HNRNPA3 //...	0.00E+00	88.68	10.36							
CDC45L	0.00E+00	59.53	5.02							
CD1C	0.00E+00	46.22	10.94							
ESPL1	0.00E+00	57.31	7.54							
KIFC1	0.00E+00	44.46	5.84							
SPC25	0.00E+00	44.29	3.43							
NAGPA	0.00E+00	44.01	7.61	0.50	6.49	0.67	8.33	0.32	7.62	0.62
RFC3	0.00E+00	58.54	5.66	1.00	7.80	0.30	5.22	0.73	7.04	0.66

**Display Preference**

Select the columns to display in the Tabular View

☒ F-Statistic ☒ P-Value

☒ Mean ☒ Std



# TESTING HYPOTHESES

Non Parametric tests

# And when the variables are not normally distributed?

- *non-parametric statistics* refers to a statistics (a function on a sample) whose interpretation does not depend on the population fitting any parametrized distributions.
- Statistics based on the ranks of observations are one example of such statistics and these play a central role in many non-parametric approaches.
- As non-parametric methods make fewer assumptions, their applicability is much wider than the corresponding parametric methods

# Example: Comparing two predictors

Set	Error rate 1	Error rate 2
1	0.2	0.4
2	0.18	0.25
3	0.24	0.38
4	0.14	0.27
5	0.05	0.31
6	0.26	0.21
7	0.15	0.32
8	0.29	0.38
9	0.31	0.28
10	0.35	0.33

How to deal when the distributions are not normal?

# Sign test

- $H_0$ : there is "no difference" between the continuous distributions of two random variables  $X$  and  $Y$ , in the situation when we can draw paired samples  $(x_i, y_i)$ , from  $X$  and  $Y$ .
- Defining  $p = \Pr(X > Y)$ , the null hypothesis  $H_0$  states that  $p = 0.50$ .
- Let  $w$  be the number of pairs for which  $y_i - x_i > 0$ . Assuming that  $H_0$  is true, then  $W$  follows a binomial distribution  $W \sim b(m, 0.5)$ .
- The test statistic is expected to follow a binomial distribution
- Left-tail value:  $\Pr(W \leq w) \rightarrow$  p-value for the alternative  $H_1: p < 0.50$ .
- Right-tail value:  $\Pr(W \geq w) \rightarrow$  p-value for the alternative  $H_1: p > 0.50$ .
- For a two-sided alternative  $H_1$  the p-value is twice the smaller tail-value.

# Sign test

- Very few assumptions about the nature of the distributions under test - this means that it has very general applicability but may lack the statistical power of other tests such as the paired-samples T-test.

# Example: Comparing two predictors

SEE IF THE DIFFERENCE IS POSITIVE OR NEGATIVE

Set	Error rate 1	Error rate 2	Sign
1	0.2	0.4	-
2	0.18	0.25	-
3	0.24	0.38	-
4	0.14	0.27	-
5	0.05	0.31	-
6	0.26	0.21	+
7	0.15	0.32	-
8	0.29	0.38	-
9	0.31	0.28	+
10	0.35	0.33	+

No. of "+" = 3; No. of "-" = 7;

- Left-tail value:  $\Pr(W \leq 3)$

PROBABILITY TO HAVE K MINUS OR PLUS

$$\sum_{k=0}^3 \binom{10}{k} 0.5^k (1-0.5)^{10-k} = 0.171$$

- Right-tail value:  $\Pr(W \geq 7)$

$$\sum_{k=7}^{10} \binom{10}{k} 0.5^k (1-0.5)^{10-k} = 0.171$$

The null hypothesis cannot be rejected

# Wilcoxon signed-rank

- The sign test discards all the information about the module of the difference
- Let  $Z_i = Y_i - X_i$  for  $i = 1, \dots, n$ , the differences.  $Z_i$  are assumed to be independent.
- Each  $Z_i$  comes from a continuous population (they must be identical) and is symmetric about a common median  $\theta$ .
- The null hypothesis tested is  $H_0: \theta = 0$ .
- The Wilcoxon signed rank statistic  $W_+$  is computed by ordering the absolute values  $|Z_1|, \dots, |Z_n|$ .
- $W_+$  is the sum of ranks of all positive differences.  $W_-$  is the sum of ranks of all negative differences.
- Let  $W$  be the minimum between  $W_+$  and  $W_-$ . If  $W=0$ , all ranks are positive (negative)  $\rightarrow$  the difference between the two set is significant
- *$W$  must be lower than critical value.*



THE MINIMUM BETWEEN W+ AND W- MUST BE  
INFERIOR TO THE NUMBER FOR Ndata

	One Tailed Significance levels:		
	0.025	0.01	0.005
	Two Tailed significance levels:		
N	0.05	0.02	0.01
6	0	-	-
7	2	0	-
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

THE SMALLER THE DIFFERENCE THE HIGHER THE RANK

# Example: Comparing two predictors

Set	Error rate 1	Error rate 2	Difference	Rank
1	0.2	0.4	+0.2	9
2	0.18	0.25	+0.7	4
3	0.24	0.38	+0.14	7
4	0.14	0.27	+0.13	6
5	0.05	0.31	+0.26	10
6	0.26	0.21	-0.05	3
7	0.15	0.32	+0.17	8
8	0.29	0.38	+0.09	5
9	0.31	0.28	-0.03	2
10	0.35	0.33	-0.02	1

$W_+ = 49$ ;  $W_- = 6$ ;  $W = 6 \rightarrow$  data significantly different at 5%