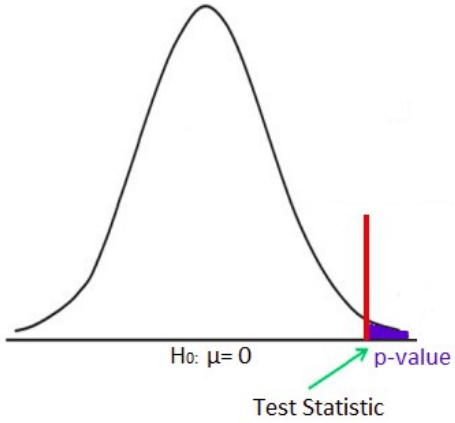
6. DATA ANALYSIS

6.1. *Hypothesis Testing*

- Your ideas have to be formulated in to a *clear* question.
- It is not *acceptable* / *efficient* to collect data and then fish around for low ***p-values*** [1].
- Start with *well-formulated* question before running an exp.



[1] helps to determine the ***significance*** of your ***results***: a small p-value (**≤ 0.05**) indicates strong evidence against the **null hypothesis**, so you ***reject*** the null hypothesis; a large p-value (**> 0.05**) indicates weak evidence against the **null hypothesis**, so you fail to ***reject*** the null hypothesis;

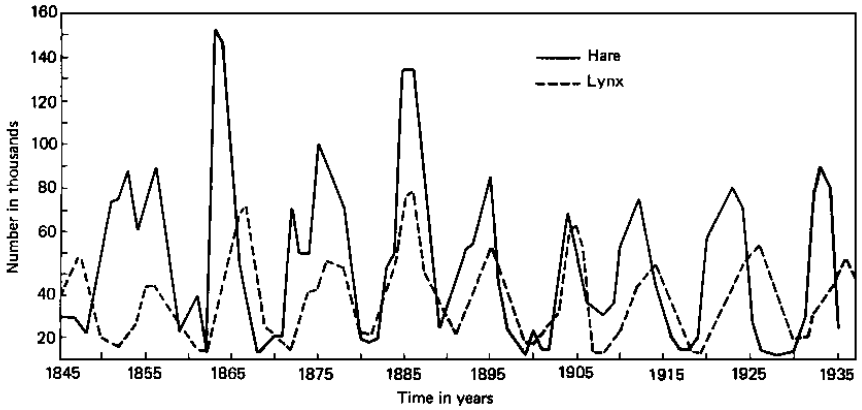
If the **p-value** < α, then this represents a statistically significant **p-value:** we can ***reject*** the claimed hypothesis.

If the **p-value** ≥ α, we cannot ***reject*** the claimed hypothesis.

- A *clear* question will stimulate relevant exp. *manipulations* and *statistical* analyses [1].

- Statistics will allow you to *evaluate* whether the differences *caused* by you *treatments* are likely to be *real* one **or** are likely to reflect *random* noise.

Photo source: Rudolfo's Usenet Animal Pictures Gallery (link no longer exists)



Population cycles in Lynx & its prey (MacLulich after Elton 1925)
https://www.mun.ca/biology

[1] e. g., you observed that the population of the ***species A*** (lynx) reduces the population size of ***species B*** (hare) /hɛ:/

You testable working ***hypothesis*** is that ***population*** of hare will be lower in the presence of lynx then in the absence of lynx.

How to study? You can conduct a manipulation, ***removing*** lynx from half of you plots and keeping the other half as unmanipulated control.

By measuring ***difference*** in hare population you will find that you null ***hypothesis*** can be rejected.

6.2. *Statistical & biological significance*

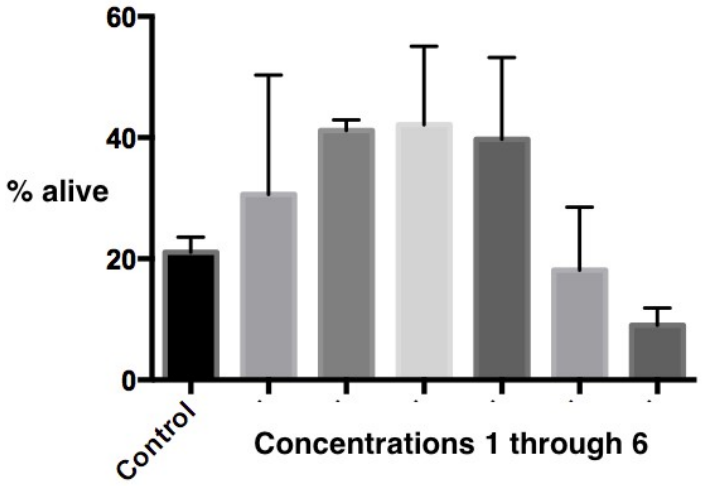
- **SS** is indicated by a *probability* level [1].
- **SS** at **p < 0.05** means that the null hypothesis (H_0)[2] is *rejected*.
- There is a 95% *chance* that there is a difference between these two conditions.
- **BS** is indicated by the *size* of the effect.

[1] not necessarily ***large*** or ***important*** when you perform a ***hypothesis*** test (the validity of a claim that is made about something) in statistics ***p*-value** (probability level) helps determine the *significance* of your results.

The ***p-value*** is a number between 0 and 1 and interpreted in the following way:
≤ 0.05 - strong evidence against the H_0, so you ***reject*** the H_0;
> 0.05 - weak evidence against the H_0, so you fail to reject the H_0;
very close to **0.05** - are considered to be marginal (could go either way).

[2] that there is no difference between the ***treatment*** and ***control***.

- Very *small* differences can be **SS** but not produce *consequences* that are important.
- In fact, we want to know if effects are "**BS**", although we often use "**SS**".
- We need to report both, if the difference is **SS** and the effect *size* [1].

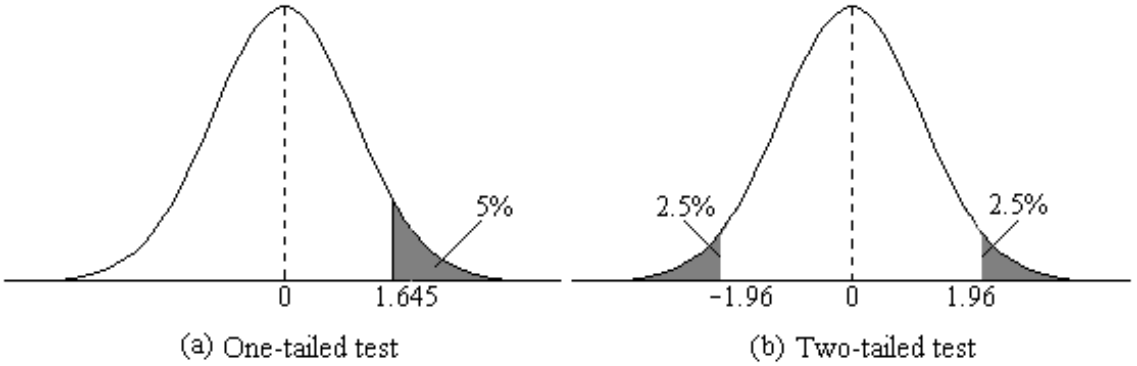


A number of cells with six different concentrations, n=3
https://stats.stackexchange.com/questions

[1] e.g. with a ***histogram*** (mean±SE) or by reporting that hares were 30% more numerous in plots without lynx.

6.3. *How to evaluate statistical significance?*

- Steps to assess **SS**:



https://towardsdatascience.com/one-tailed-or-two-tailed-test-that-is-the-question-1283387f631c

- define your **hypotheses** [1];
- set the significance **level** (**α-level**) [2];
- decide to use a one-tailed or two-tailed **test** [3];
- determine sample **size** [4];
- calculate **SD** [5];

[1] i.e. the question you want to ***answer*** and stating your ***hypothesis***;

[2] the ***threshold*** that you set to determine significance: if your ***p-value*** is ≤ to the set significance level, the data is considered SS;

[3] a one-tailed (**однобічний, is more powerful**): testing for the possibility of the relationship in ***one*** direction, e.g. a new variety "A" you think is "more" productive than already existing variety "B"; two-tailed test: you are testing for the possibility of the relationship in ***both*** directions, e.g. a new variety "A" you think is "more" or "less" productive than already existing variety "B";

[4] practical part question, where 0.01 is confidence interval for μ;

$$1.96 \pm \sqrt{\frac{\sigma^2}{n}} \leq 0.10$$

[5] the formula is $SD = \sqrt{\sum((x_i - \mu)^2/(n-1))}$.

- calculate the **variance** between 2 sample groups [1];
- calculate the **t-score** [2];
- determine the **df** [3];
- use a **t-table** to evaluate significance [4].

E.g. if **t-score** = 2.61 gives *p-value* between 0.01 and 0.025, which is ≤ 0.05, our data is **SS**, we reject the null hypothesis.

[1] the formula for variance is $s_d = \sqrt{((s_1/N_1) + (s_2/N_2))}$;

[2] t-scores allow you to perform a t-test that lets you calculate the probability of two groups being significantly different from each other: $t=(\mu_1-\mu_2)/s_d$;

[3] e.g. if we have 5 samples in each of two groups, df=8 i.e. ((5+5)−2=8);

[4] in a statistics book or online.

6.4. *How to calculate effect size?*

- E.g. you studied how removal of large herbivores (goats) from some large scale plot *effect* the *population* of small herbivores (grasshopper).

- You found that in those without *mammalian* herbivores *grasshopper* population was *higher* and **SS** (**p=0.014**).

- What was the *size* of the effect?

- **One way**: the *absolute difference* (**AD**) in grasshoppers between two treatments:

  $|\text{mean}_{\text{tretm.(no goats)}} - \text{mean}_{\text{ctrl.}}| = \text{AD}$

  |14.8 – 6.75| = 8.05

Another way: AD is not as useful as *relative difference* (RD):

$$|\text{mean}_{\text{tretm.(no goats)}} - \text{mean}_{\text{ctrl.}}| / \text{mean}_{\text{ctrl.}} = \text{RD}$$

$$|14.8 - 6.75| / 6.75 = 119\ \%$$

By removing goats population of grasshoppers *increased* by **119 %.**

Third way: to report standardized effect *sizes* scaled by a measure of the *variance* or *noise* involved in measuring them [1,2]:

$$|\text{mean}_{\text{tretm.(no goats)}} - \text{mean}_{\text{ctrl.}}| / \text{SD}_{\text{tretm.(no goats)} + \text{ctrl.}}$$

i.e. (14.8 – 6.75) / 4.43 = 1.82.

[1] the difference between the means divided by their standard deviation, SD;

[2] gives **unitless** (having no units of measurement) **value:** help evaluate how big or small an effect is when the units of measurement aren't intuitive (інтуїтивно зрозумілий). E.g. 2.3 ° C is a meaningful difference while a 2.3–point difference on an anxiety scale that runs from 7 to 49 is not so meaningful difference.

- Such effect sizes are *unit-less* and allows us to compare it with the effects found in other studies.

- No two populations are *exactly* identical, just as no two people are.

- By testing a $\mathbf{H_0}$[1] we are not calculating the *probability* that they are *identical*, but rater the *probability* that we can detect a *difference* between them.

- We are saying that two populations are:

- truly different if **p=0.049**;

- not different if **p=0.051**

[1] two populations are the same

- These numbers are a bit *arbitrary*, since in both cases we will be wrong appr. 5 times of 100.

- If **p=0.001** we can be more *confident* that the result was not caused by chance than if e.g. **p=0.05**.

- It is good to report *calculated* value rather than reporting p > or < than 0.05.

6.5. *Alternative hypotheses*

https://keydifferences.com/

- Many ecol. *hypotheses* are not simply *true* or *false* [1].
- How can we study this *phenomenon*?
- Instead of rejecting $\boldsymbol{H_0}$ let us think what is the *size* of the effect caused by *competition*?
- Even if *competition* is found to be important, *predation* may also be important.
- By developing a set of *alternatives*, you do not become "attached" to the hypotheses you selected initially.

[1] e.g., we want to understand the role of ***competition*** in communities structuring. We can not conduct a simple experiment (it can be *predation*, *parasitism disturbance* etc.).

- If you start with a *list* of $\boldsymbol{H_A}$, you likely will end up with something *interesting*.

- If you are *focusing* on only a single *hypothesis*, you will have to say *something* only if your results come out one particular way.

- How to *generate* $\boldsymbol{H_A}$?

- Once you have identified a *pattern* that is interesting to you *consider* the other possible *factors* as alternatives that could also produce this pattern[1].

[1] E.g.:

abiotic factors (e.g. precipitation, temperature etc.);

predators, ***parasitism***, disease;

mating factors (e.g. sexual selection, nest-site availability);

microhabitatats (shelter from abiotic conditions);

disturbance (human or natural causes);

genetic or ***developmental*** influence etc.

- Should you test all of your *alternatives*?
- No, you *should* not.
- Although *getting* them all down on the paper for a *consideration* is a first step.

6.6. *Negative results*

- Assume, we *fail* to *reject* a H_0 [1], should we *conclude* that two populations really are the *same*?
- No.
- What can we *say*?
- Only that we *failed* to show ***difference*** or ***effect*** we have hypothesized.
- Why?

[1] i.e. p > 0.05

Reasons:

- statistical tests give us far *more* power to *reject* hypotheses than to *accept* “negative” results (1).

- in most cases we have very *weak* power to *evaluate* whether two populations are *similar* (2).

- we rarely use *relevant* *statistics* to address this issue (3).

- many papers with ”*negative*” results in ecology never get published (4).

- our ability to *accept* a negative result also depend on the *effect* *size* we are looking for [1].

[1] e.g., it is generally accepted that:

a ***small*** effect size as a ***difference*** of 0.10 (**10%**) or <;
a ***large*** effect size as a ***difference*** of 0.40 (**40%**) or >.

- How to interpret results?

- If we expect a *really* large effect, we can be more *confident* that one truly didn't exist than if we were only *expecting* a small effect.

Figure 1: The most common approach taken by journals, in which only those experiments yielding positive results end up as publication material.

Figure 2: A more neutral approach, in which all results are published, as long as they are generated by well-carried out experiments based on sound hypotheses.

https://www.elsevier.com/

6.7. *The numbers of data and test statistics*

- The numbers of the data itself is not very *meaningful*, because it's not *standardized*.
- You can obtain a *lot* of data points, but you have to extract *meaningful* things from it.
- Why test *statistics* [1] is important?
- It shows how *far* or *close* actual results are from claimed data in terms of *standard errors* (**SE**) of the sample mean [2].
- The *sample size* is another variable we need to calculate the *p-value* [3].

[1] a statistic used in ***statistical*** hypothesis ***testing***.

[2] depends on **SD** and the sample size (**n**): **SE = SD/√n**: **SE** is a measure of the ***dispersion*** of sample ***means*** around the ***population mean***.

So, if we want to say how ***widely scattered*** some measurements are, we use the **SD**; if we want to indicate the ***uncertainty*** around the estimate of the mean measurement, we use **SE**.

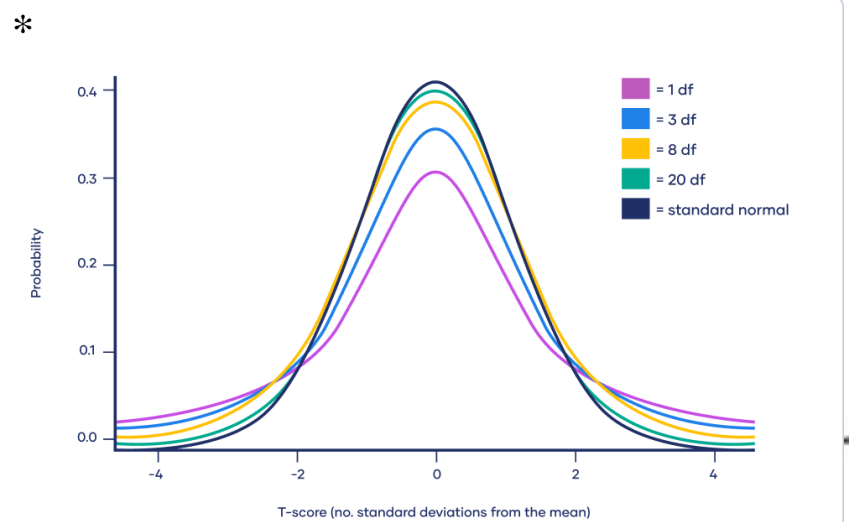
[3] important because it determines whether we use the standard normal ***distribution*** (Z-distribution) to look up the ***p-value***, or we use the ***t-distribution*** to look up the ***p-value***. When you know the population **SD** you should use the **z-test**, when you estimate the sample **SD** you should use the **t-test.** Usually, we don't have the population **SD**, so we use the ***t-test***.

If the sample size is **< 30** (n < 30) [1], we use the *t-distribution** to calculate the ***p-value***.

We calculate the **df** (df=n-1) and use it to calculate the *p-value*.

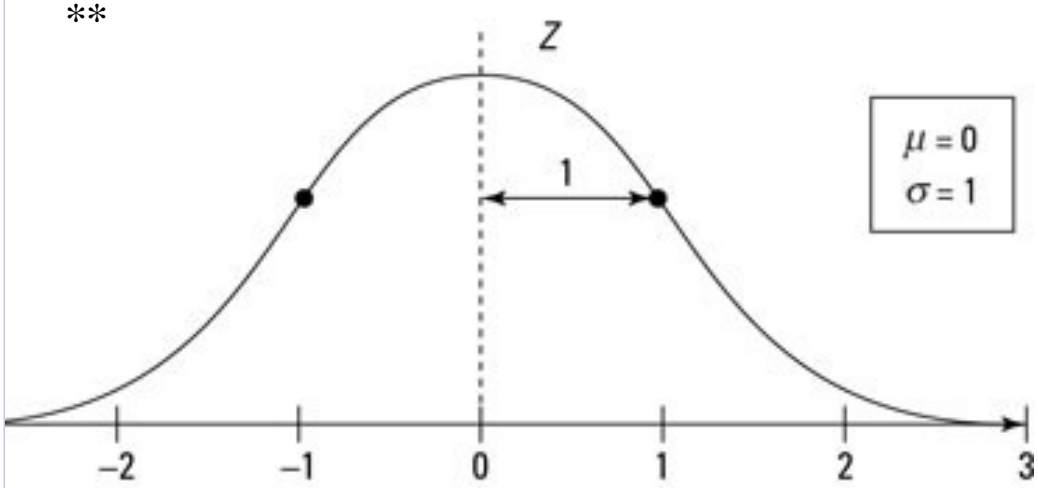
If the sample is > 30 (**n > 30**) [2], we use the *Z-distribution*** to calculate the ***p-value***.

*



t-distributions with different **df** and the standard normal distribution
http://onlinestatbook.com

**



The **Z-distribution**
https://www.dummies.com

[1] this is a ***small*** *sample* ***size;*** * Student's probability distribution that is used to estimate population parameters when the sample size is ***small*** and/or when the population ***variance*** is ***unknown***;

[2] this is a ***large*** sample ***size;*** ** is a normal distribution with mean zero and standard deviation 1.