Central Logic of Statistics & Hypothesis Testing

The whole point of statistics is to prove something scientifically, so the fundamental questions are ① what are you trying to prove? and ② How do you prove what you are trying to?

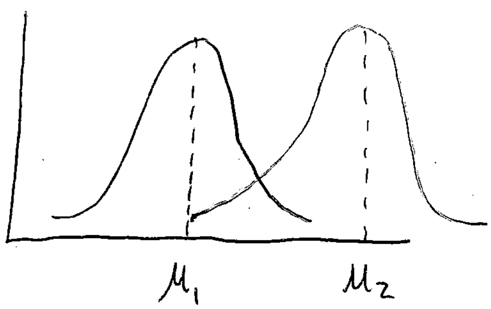
① What are you trying to prove?

In psych research you are generally trying to show a difference in means between two populations

In equations you are trying to show...

$\mu_1 \neq \mu_2$ where the null hypothesis (H_0) would be $\mu_1 = \mu_2$

Graphically you are trying to show...

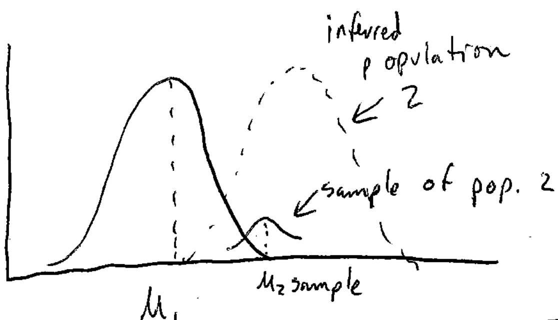


If you could survey all people in both populations this would be easy to show, but that is usually not possible so statistics must be used to infer this difference which leads us to our next question ②

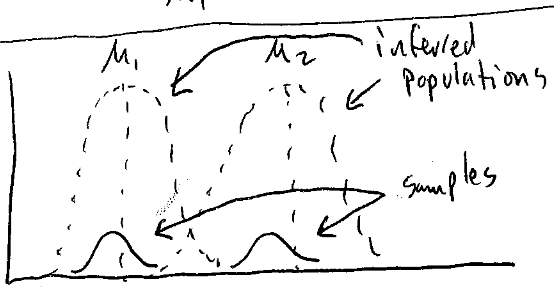
② How Do You Prove What You Are Trying To?

Researchers rarely have access to entire populations and must use samples from the populations and then run statistics to approximate what entire populations look like. This Is the reason for z-tests, t-tests, ANOVAs, etc.

If population 1 is known a sample is taken from population 2 & compared.



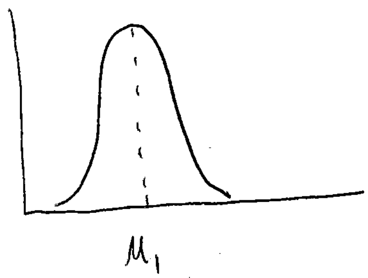
Statistics are run to determine how likely the sample is to have come from population 1. If it is extremely unlikely (5% or 1%) to have come from Population 1, we assume the existence of population 2 (in dotted lines) to prove the research hypothesis.



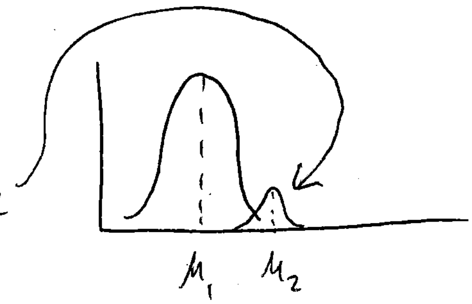
If neither are known two samples must be taken, one to infer population 1 (a control) & one to infer population 2 (your group of interest)

Central Logic Page 2

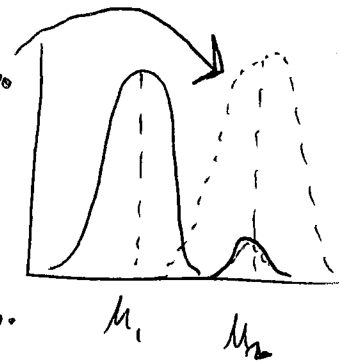
The fundamental reasoning behind p values is this: if this is your population 1



and this is your sample



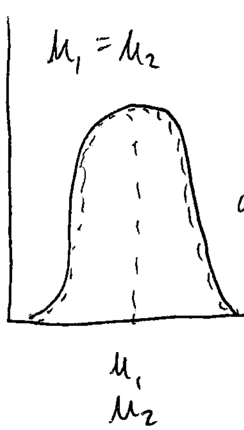
It is extremely unlikely your sample came from population 1, <u>but</u> <u>it</u> <u>would</u> <u>not</u> <u>be</u> <u>unlikely</u> to get a sample like this from a <u>second</u> <u>population</u> if the 2nd pop. looked like this.



So <u>we</u> <u>assume</u> if it is <u>unlikely</u> <u>enough</u> (5% or 1% chance, that is $p<.05$, $p<.01$) to have come from pop. 1 that <u>population 2</u> <u>exists</u>, proving your research hypothesis.

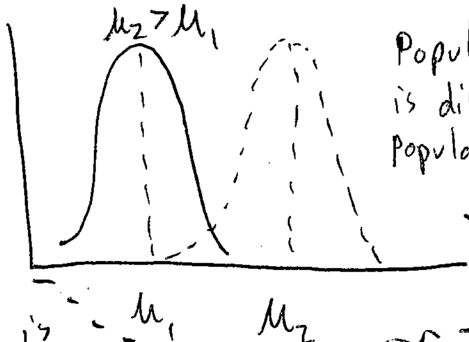
For a given research problem where population parameters are known (this is simplest case, usually pop. 1 parameters must be inferred from control group) graphs for both scenarios would like like this:

<u>Accept Ho</u>



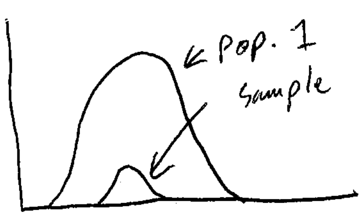
Both populations have same distribution, therefore there is no difference

<u>Reject Ho</u>

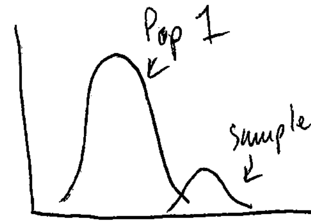


Population 2 is different than Population 1

Population 2 distribution is dotted lines because it is inferred from the statistics.



This would support Ho because it appears sample came from pop. 1



This would be an unlikely scenario if $\mu_1 = \mu_2$ (1 in 20 times at 5%) but very probable if $\mu_2 > \mu_1$: if unlikely enough we assume

Z- Scores & The Normal Curve

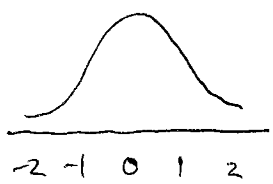
Mathematicians understand general features of Normal Curves just as they understand general features of geometrical figures.

For Example with circles

The area is always πr^2 no matter what r is
and the circumference is always $2\pi r$ no matter what r is

This is a fundamental definition of a circle and is true for any circle no matter what the numerical value of r actually is.

This reasoning also applies to the normal curve and z-scores



Z scores, like r, tell you fundamental features of the normal curve no matter what the raw scores are.

Just as the area of a circle is always πr^2, the .05 probability level always falls at $Z = 1.65$

So raw scores are converted to Z-scores because the Z-score tells us where the raw score falls in the distribution.

The normal curve is a mathematical abstraction just like a circle is; there are no perfect circles in nature just as there are no perfect normal curves, but that does not make the concepts & mathematics of circles & normal curves any less useful

---

A Standard Deviation is essentially the average amount that scores vary from the mean, which gives us a basis for comparison for any one score.

A Z score is $Z = \frac{X - M}{SD}$ or $\frac{\text{The amount score varies from mean}}{\text{The average amount scores vary from mean}}$

A ratio between individual variance and average variance from mean

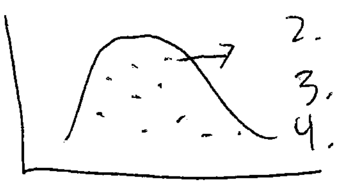
This is how much something varies not s^2 specifically

Distribution of Means & Z-test

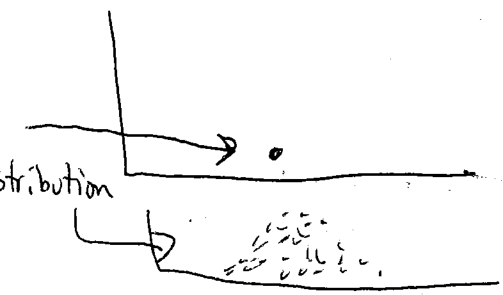
In general in research we are using means of samples rather than individual scores. So we must compare these means to distributions of means rather than to distributions of individual scores. See p. 150-156 for specific questions, this will only be very general!

To make a distribution of means

1. Take normal distribution & pull out "n" scores



2. average them to get mean
3. plot mean as one score
4. do this many times to get distribution of scores



Distributions of means are skinnier than regular distributions because the more scores incorporated in "n", the less the new mean will vary from mean of population

$n = 5$ $n = 10$ $n = 20$ $n =$ # in original distribution

If all scores are included in figuring new mean, distribution is a line

This is because at .05 level there is a 5% chance of pulling out 1 score but there is a $(.05)(.05) = .0025$ or .25% chance of pulling two in a row and $(.05)(.05)(.05) = .000125$ or .0125% chance of pulling three in a row

---

A <u>Z-test</u> is just like doing a <u>Z-score</u> but with a <u>distribution of means</u> rather than a <u>distribution of individual scores</u>

$$Z = \frac{M - \mu_m}{\sigma_m}$$

where μ_m is mean of distribution of means & σ_m is SD of distribution of means

Effect Size & Power

Effect size(d) is exactly that, the size of the effect you have measured.
The equation $d = \frac{\mu_1 - \mu_2}{\sigma}$ is the difference between your two means divided by the standard deviation of scores in general. This equation artificially resembles the equation for a z-score $\left(Z = \frac{X - \mu}{\sigma}\right)$ this is because both are ratios of $\frac{\text{score difference}}{\text{average expected difference}}$.
They both say how many standard deviations there are between the top 2 numbers (the numerator).

Power is the probability of finding an effect if there is one; or in other words, the chance that your sample studied reflects the objective population. Anything that helps the sample more resemble the population (like bigger sample size) or reduces random variation (noise) to more clearly show effect (systematic variation) increases the power.

In hypothesis testing you are trying to compare 2 populations usually (or more). z-tests are used when the parameters (μ, & SD) are known for the population as a whole. This is usually not the case. Usually both population parameters must be estimated or inferred from samples. This is when t-tests are used.

A single sample t-test is used when population mean (μ_p) is known but standard deviation (σ) is not. The standard deviation for the population must be estimated, that is what "S_m" is for, it is an estimate of population standard deviation.

The formula for a single sample t-test is

$$t = \frac{M - \mu}{S_m} \quad \text{where} \quad S_m = \sqrt{S_m^2} = \sqrt{\frac{SS/df}{N}}$$

This is how σ_M is estimated

notice the similarity between this & Z-test

$$Z = \frac{M - \mu_m}{\sigma_m}$$

← this is the only difference. In a Z-test σ_m is known in t-test for single sample it is estimated with S_m

Therefore t-scores are almost exactly like Z scores

A t-test for dependent means is essentially a t-test for a single sample where scores are replaced by difference scores. A difference score (D) is just that; a score that is the result of a difference between 2 other scores such as a "before treatment score" minus an "after treatment score". In a formula that is $D = X_1 - X_2$.

A t-test for independent means is carried out when there are no known population parameters (you don't know μ_p or σ_p). The population parameters must therefore be inferred from a sample (usually called a control group). Just as in single sample t-test we used S_m to infer σ_p, in this t-test we will be estimating not only population standard deviation (σ_p) but also the population mean (μ_p) from our control group & then compare the experimental sample to our "inferred population."

The population mean is estimated to be same as the control mean; that is $\mu_p = M_1$. The standard deviation is a little more difficult and is estimated as $S_{diff} = \sigma_p$ where $S_{diff} = \sqrt{S_{m_1}^2 + S_{m_2}^2}$ which is $\sqrt{\frac{S^2_{pooled}}{N_1} + \frac{S^2_{pooled}}{N_2}}$

S^2_{pooled} is average variance within both samples added together

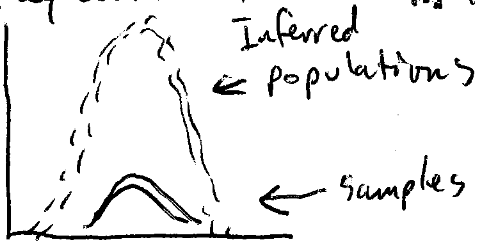
So $t_{ind} = \frac{M_1 - M_2}{S_{diff}}$ ← estimate for σ_p based on σ in both samples

t-tests & Distribution of Differences Between Means

A single sample t-test compares a known population to an experimental sample & so compares those curves, but a dependent t-test and independent t-test are scores based on differences of means, of two samples (rather than with 1 sample & 1 population). These tests therefore give you a score that would be plotted on a Distribution of differences between means.

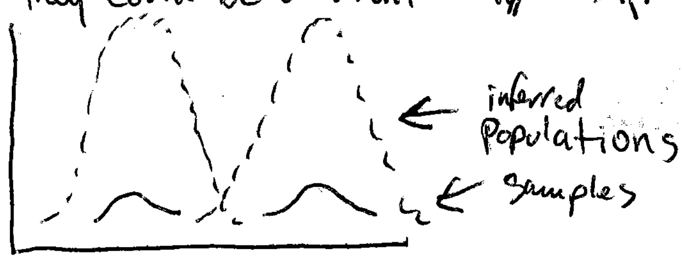
Think of the two possibilities for 2 samples:

They could be the same... $H_0 \rightarrow \mu_1 = \mu_2$



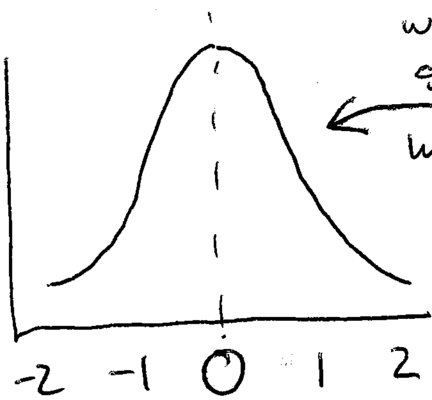
Scenario 1

or They could be different $H_1 \rightarrow \mu_1 \neq \mu_2$



Scenario 2

In Scenario 1 above the difference between means is 0 because if $\mu_1 = \mu_2$ then $\mu_1 - \mu_2 = 0$. In Scenario 2 the difference between means would not be 0. This would be a "difference between means score," & so must be compared on a distribution of differences between means. This distribution looks like this. It is a t-distribution with $\mu = 0$ & σ based on variance within samples. If H_0 is true $\mu_1 - \mu_2 = 0$ that's why $\mu = 0$ for this distribution.



t → this is t score, like a z-score

0 depends on study X → this is diff. of means score & is based on variance

ANOVA

Up to this point all the logic has been essentially the same: a (experimental) sample population is compared to a known or control population to see if their means differ, & by how many standard deviations they differ. This tells you the probability of your experimental population being different from the control population.

remember $z = \frac{X - \mu}{\sigma}$ $z_{means} = \frac{M - \mu_m}{\sigma_m}$ $t_{ss} = \frac{M - \mu}{S_m}$ $t_{ind} = \frac{M_1 - M_2}{S_{diff}}$

all scores essentially mean the same thing: where that difference of means falls on a normal distribution or t-distribution.

ANOVAs instead compare two different populations or more and are based on F-ratios rather than Z-scores or t-scores.

F ratio is a ratio of $\frac{\text{Variance between sample means}}{\text{Variance in general}}$

in other words an F-ratio measures how much the sample means vary vs. how much they would be expected to vary if there was no difference in populations.

$$F = \frac{S^2_{between}}{S^2_{within}}$$ ← Variance between sample means / ← Variance in general

$$S^2_{between} = (S^2_m)(n_{\text{number of scores per group}}) = \left(\frac{\Sigma(\overset{\text{group means}}{\text{Mean}} - \overset{\text{mean of all group means}}{\text{grand Mean}})^2}{df_{between}}\right)(n)$$ ← $N_{groups} - 1$

$$S^2_{within} = \frac{S_1^2 + S_2^2 + \ldots}{N_{groups}}$$ or $\frac{\text{Variances of all groups}}{\#\text{ of groups}}$ = average variance in general