HYPOTHESIS TESTING & NULL HYPOTHESIS SIGNIFICANCE TESTING (NHST)

RESEARCH PROBLEM

Want to provide evidence that our hypothesis, research question, or theory is correct

RESEARCH PROBLEM

Specifically, we're concerned with investigating whether there is a

- Difference, on some outcome variable, between groups in our independent variable
- Relationship between our independent and dependent variables

HYPOTHESIS TESTING

Demonstrate relationship between variables in our sample reflects a true relationship that occurs in the population and is not simply *due to chance (sampling error)*

HYPOTHESIS TESTING

Start with assumption that no relationship exists, and look for evidence that "*no relationship" DOES NOT exist*…

Without a breathalyzer, to determine if someone is drunk, you'd need find evidence ***against*** being sober

Condition	Yes (Probably Sober)	No (Probably NOT Sober)
Walks straight	✅	
No slur in speech		✅
Attentive gaze		✅
Responsive to questions		✅
No smell of alcohol		✅
Can maintain standing balance		✅

HYPOTHESIS TESTING

Develop two types of hypotheses that represent opposite takes on the relationship between variables:

1. Null Hypothesis (H_0)
   - A hypothesis that *nothing/no relationship between variables/no difference exists*, to *disprove* or ***reject*** it

2. Alternative Hypothesis / Research Hypothesis (H_1 / H_a)
   - A contradictory hypothesis that *something/a relationship between variables/a mean difference does exist*
     - Evidence against H_0 lends support for H_1

HYPOTHESIS TESTING

To demonstrate that a relationship does exist…

- We look for extreme evidence AGAINST the null hypothesis (H_0) that “*no relationship/no mean difference” DOES NOT exist*
  - Which means lower likelihood that null is true
  - Allows us to reject the null.

But what counts as extreme?

HYPOTHESIS TESTING

Relationships between variables can be *bivariate* or *multivariate*.

HYPOTHESIS TESTING AND BIVARIATE RELATIONSHIPS

Chi Square Test of Independence (X^2)

- IV: categorical/discrete (e.g. nominal, ordinal)
- DV: categorical/discrete (e.g. nominal, ordinal)

Independent Samples T-test (t)

- IV: categorical/discrete (e.g. nominal, ordinal)
  - Only two groups/samples compared
- DV: continuous (e.g. interval-ratio)

Analysis of Variance / ANOVA (F)

- IV: categorical/discrete (e.g. nominal, ordinal)
  - Three or more groups/samples compared
- DV: continuous (e.g. interval-ratio)

Correlation (r)

- IV: continuous (e.g. interval-ratio)
- DV: continuous (e.g. interval-ratio)

	Dependent Variable	
Independent Variable	**Nominal/Ordinal**	**Interval-Ratio**
Nominal/Ordinal	Chi-Square	T-Test or ANOVA
Interval-Ratio	N/A	Correlation

EXAMPLE 1: HYPOTHESIS TESTING FOR MEAN DIFFERENCES

Gerontologist tests two methods (A and B) for increasing memory of nursing home residents.

- Randomly selects 10 residents, divides them into Method A or Method B.
- Method A has participants engage in memory games on an iPad. Method B is a control group that does daily journals.
- Asks them all to take a recall test after exposure to method
- Mean for group A is 82; Mean for group B is 77

Method A	*Method B*
82	78
83	77
82	76
80	78
83	76
Mean = 82	Mean = 77

How do we know if the difference is real or due to chance (sampling error)?

EXAMPLE 1: HYPOTHESIS TESTING FOR MEAN DIFFERENCES

Specifically, we're concerned with investigating whether there is a

- Difference, on some outcome variable, between groups in our independent variable
- Relationship between our independent and dependent variables

Method A	*Method B*
82	78
83	77
82	76
80	78
83	76
Mean = 82	Mean = 77

EXAMPLE 1: HYPOTHESIS TESTING FOR MEAN DIFFERENCES

Method A: consistently in low 80s

Method B: consistently in high 70s

Method A	*Method B*
82	78
83	77
82	76
80	78
83	76
Mean = 82	Mean = 77

Difference b/w groups probably not sampling error

- Closeness (homogeneity) of scores in each group
- No overlap in scores b/w groups

Method A *likely* better

- But we have to do statistical hypothesis tests to see if this difference is actually the result of chance or sampling error, or a true difference between two groups in population

EXAMPLE 2: HYPOTHESIS TESTING FOR MEAN DIFFERENCES

Do conservatives and liberals differ in the severity of their discipline styles?

- We know from this statement that this is a bivariate relationship (1IV and 1DV)
- We know that one variable (IV) is the grouping variable
  - Political Orientation (categories, including liberal versus conservative)
- We know that the other variable (DV) is the outcome variable
  - Severity (a scale, ranging from 0 to 100)
- We can compare groups by comparing severities – using the *mean* severity for each group.

EXAMPLE 2: HYPOTHESIS TESTING FOR MEAN DIFFERENCES

Specifically, we're concerned with investigating whether there is a

- Difference, on some outcome variable, between groups in our independent variable
- Relationship between our independent and dependent variables

AKA: *Is there a difference in mean severity between liberals and conservatives?*

EXAMPLE 2: HYPOTHESIS TESTING FOR MEAN DIFFERENCES

Do conservatives and liberals differ in the severity of their discipline styles?

$$\bar{X}_{liberals} = 54$$

$$\bar{X}_{conservatives} = 58$$

HYPOTHESIS TESTING FOR MEAN DIFFERENCES

To demonstrate that a **mean difference** *does* exist…

- We look for extreme evidence **AGAINST** the null hypothesis (H_0) that "*no mean difference" exists*
  - Or, put another way, we're looking for evidence that *"no mean difference" DOES NOT exist"*
- We examine the difference to see if it is *extreme enough*

H_0: NULL HYPOTHESIS FOR MEAN DIFFERENCES

There is *no mean difference between the two groups*, in the population. The mean of one group is equal to the mean of the other, in the population.

- H_0: $\mu_1 = \mu_2$

- H_0 attributes mean differences to sampling error.
  - Any observed difference between samples is due to chance/sampling error and doesn't represent a true difference between groups in the population

H_1: ALTERNATIVE HYPOTHESIS FOR MEAN DIFFERENCES

There *<u>is a mean difference between two groups</u>*, in the population. The mean of one group does not equal the mean of the other, in the population

- H_1: $\mu_1 \neq \mu_2$

- H_1 attributes mean differences to true population differences between groups.
  - Any observed difference between samples is not due to chance (sampling error) and represents a true difference in means of each group, in the population.
  - The difference between groups is TOO BIG to be the result of sampling error

HYPOTHESIS TESTING FOR MEAN DIFFERENCES

Questions when examining mean differences between groups:

- Is difference zero?
  - Each group has the same mean
- Is difference not zero?
  - *Is difference small?*
    - May be due to chance (sampling error). Groups just happened to have similar means
  - *Is difference large?*
    - May be due to *true* difference between groups in population

HYPOTHESIS TESTING FOR MEAN DIFFERENCES

We're concerned with determining if there is **extreme evidence AGAINST the null hypothesis (H_0)** – that "*no mean difference" DOES NOT exist* – in favor of our research hypothesis (H_1)

- This means we need a large or extreme mean difference

But what counts as large or extreme enough of a mean difference?

HYPOTHESIS TESTING FOR MEAN DIFFERENCES

We can apply knowledge of the normal distribution and the logic Z-scores (for raw scores) to mean differences, to *create Z-scores for mean differences*.

Large Z scores would indicate large or extreme deviations from mean differences of 0.

Then we can ask:

- *What is the probability of getting these extreme of differences (or greater differences) between groups, given that the null hypothesis is true?*

DISTRIBUTIONS AND Z-SCORES FOR T-TEST

EXAMPLE 2: DIFFERENCES BETWEEN MEANS

Researcher wants to compare differences in child-rearing discipline severity by political orientation

- X = Political Orientation (conservative, liberal)
- Y = Severity (scale of 1 (less severe) to 100 (more severe))

- Gives it to 1 random sample of two groups
  - 30 conservative and 30 liberal

- Finds conservative mean = 58.0; liberal mean = 54.0
  - Difference between (conservative and liberal) means = +4.0

Is this a true difference that occurs in the population or the result of sampling error?

SAMPLING DISTRIBUTION OF DIFFERENCES BETWEEN MEANS

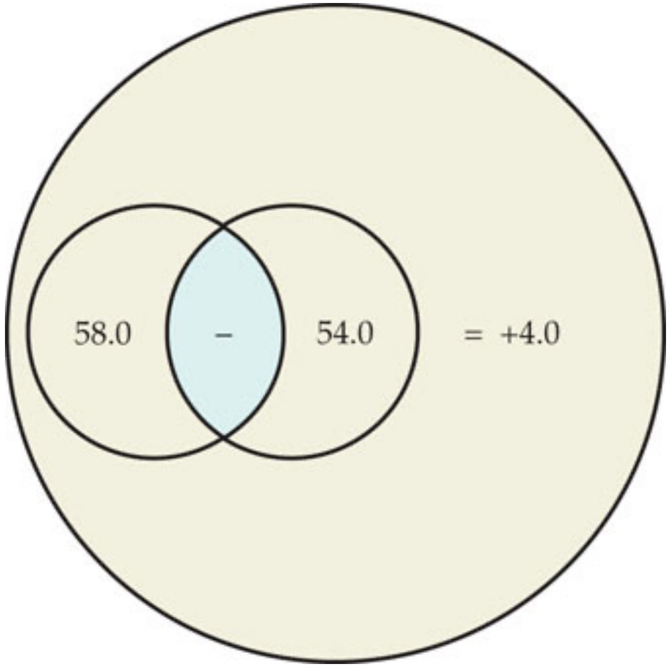
We can use logic of normal distributions to construct a *sampling distribution of differences between means*

- Shows us how likely **our mean difference** (58.0-54.0 = +4.0) is, in relation to all possible samples.

SAMPLING DISTRIBUTION OF DIFFERENCES BETWEEN MEANS

Before, he took 1 pair of random samples (30 conservative and 30 liberal)

- Found a mean difference of +4.0 points

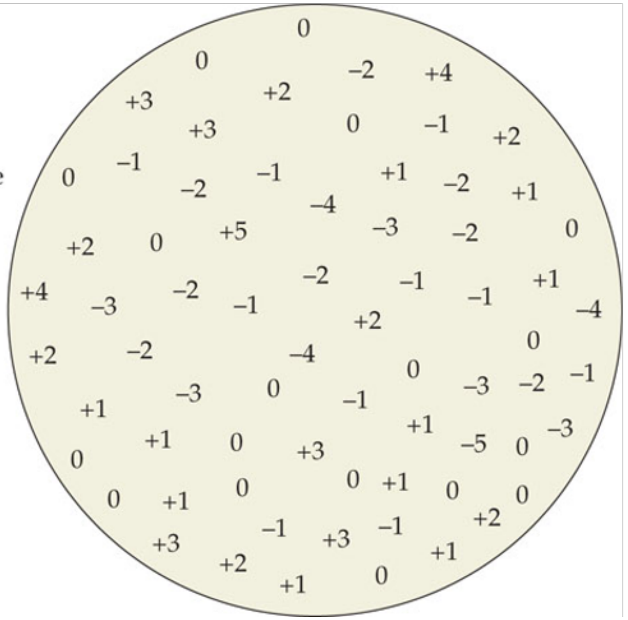


SAMPLING DISTRIBUTION OF DIFFERENCES BETWEEN MEANS

Now, assume he took 70 pairs of random samples (each with 30 conservative and 30 liberal)

- We could calculate a difference-between-means score for each pair of samples

Note: Each score represents the mean difference between a sample of 30 liberal and a sample of 30 conservative

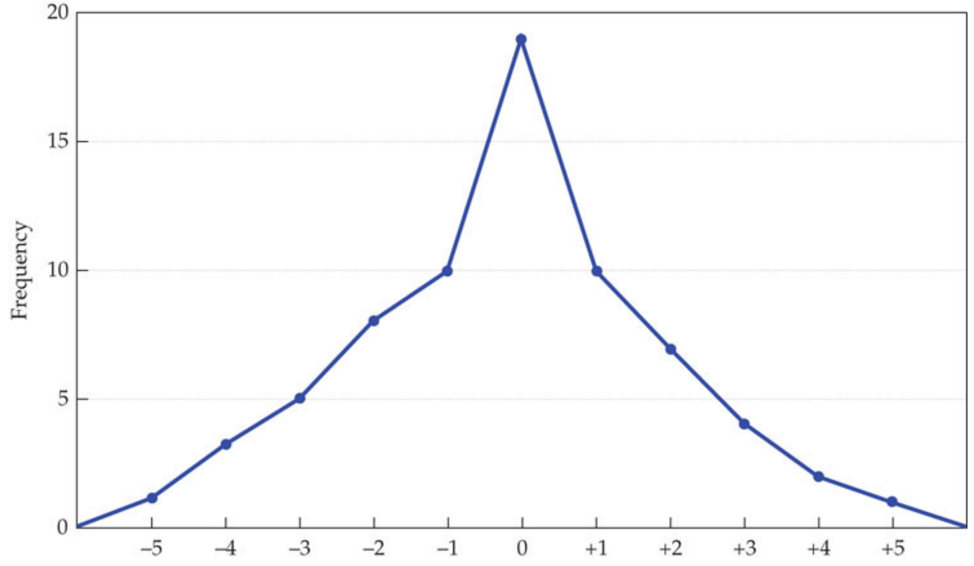


SAMPLING DISTRIBUTION OF DIFFERENCES BETWEEN MEANS

If we plotted these mean difference values, it would be approximately a normal curve

Curve approximates the null hypothesis – NO MEAN DIFFERENCE between each pair of samples

- More mean difference scores around 0



SAMPLING DISTRIBUTION OF DIFFERENCES BETWEEN MEANS

Because it approximates the normal curve, we can identify which percentage of cases (differences between sample means) fall between

- -1 and +1 standard deviation units:
  - 68.26%
- -2 and +2 standard deviation units:
  - 95.44%
- -1.96 and +1.96 standard deviation units:
  - 95.00%
- -3 and +3 standard deviation units:
  - 99.74%
- -2.58 and +2.58 standard deviation units:
  - 99.00%

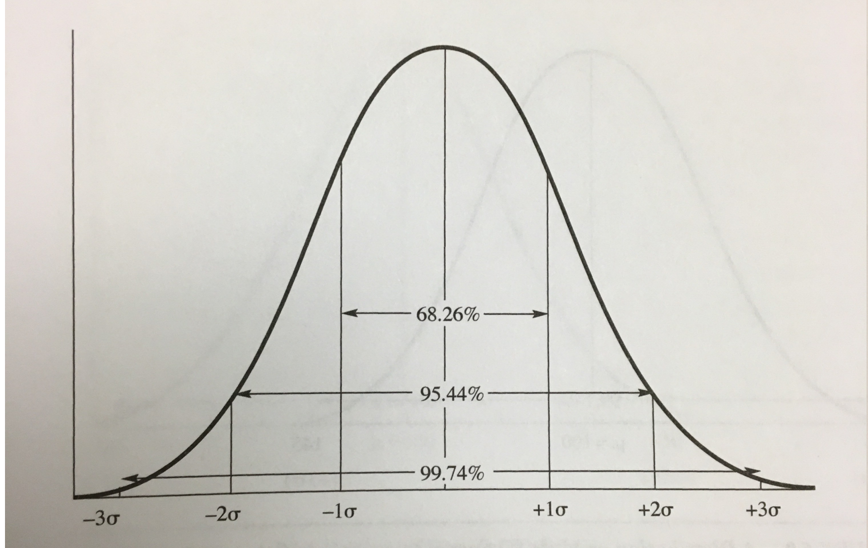


FIGURE 5.8 *The Percent of Total Area under the Normal Curve between* -1σ *and* $+1\sigma$, -2σ *and* $+2\sigma$, *and* -3σ *and* $+3\sigma$

Z-SCORES AS (EXTREME) EVIDENCE AGAINST NULL FOR T-TEST

HYPOTHESIS TESTING FOR MEAN DIFFERENCES

Looking for evidence AGAINST the null

- Extreme mean differences = evidence AGAINST NULL
- Extreme mean differences = further away from null
- Extreme mean differences = extreme Z-scores
  - Lower likelihood that null is true

But what counts as extreme?

EVIDENCE AGAINST NULL = EXTREME Z-SCORE

Use Z-score to figure out where *OUR mean difference falls on a distribution of all possible mean differences* (from all possible pairs of samples)

SAMPLING DISTRIBUTION OF DIFFERENCES BETWEEN MEANS

Start with Z-score for a *raw score* relative to the mean and standard deviation *population/probability distribution*

- Z = $\frac{X - \mu}{\sigma}$

…And since we're talking about means, we can *replace raw scores with sample means*, relative to mean and standard deviation of the theoretical sampling distribution of (all possible) means

- Z = $\frac{\bar{X} - \mu}{\sigma_X}$

But because we're calculating mean differences, we *replace sample means* (and standard deviation of the sampling distribution) *with mean differences*, relative to mean and standard deviation of the theoretical sampling distribution of (all possible) differences between means

- Z = $\frac{(\bar{X}_1 - \bar{X}_2) - \mu}{\sigma_{X_1 - X_2}}$

SAMPLING DISTRIBUTION OF DIFFERENCES BETWEEN MEANS

Because the mean (μ) of the *sampling distribution of the difference between means* is most likely zero (the null hypothesis of no mean difference), plug in zero as our mean

- $Z = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sigma_{\bar{X}_1 - \bar{X}_2}}$

But the zero cancels out, so we're left with…

- $Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}}$

SAMPLING DISTRIBUTION OF DIFFERENCES BETWEEN MEANS

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}}$$

Still an unknown parameter

we don't have information on the sampling distribution of the differences between means, so how would we calculate its standard deviation

EXAMPLE 2: Z-SCORE FOR MEAN DIFFERENCE COMPARED TO SAMPLING DISTRIBUTION OF MEAN DIFFERENCES

If we take our example of disciplinary severity scale differences between liberals and conservatives

- Conservative mean ($\bar{X}_1$) = 58.0
- Liberal mean ($\bar{X}_2$) = 54.0
- Difference-between-means ($\bar{X}_1 - \bar{X}_2$) = +4.0
- Assume that the *__standard deviation__ __of the sampling distribution of differences between means__* ($\sigma_{\bar{X}_1 - \bar{X}_2}$) is 2.00
  - Again, unknown parameter, so here I provided a number

We can calculate the Z-score for our difference between means as

- Z = $\frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{(58 - 54)}{2.00} = \frac{4}{2.00} = 2.00$

EXAMPLE 2: Z-SCORE FOR MEAN DIFFERENCE COMPARED TO SAMPLING DISTRIBUTION OF MEAN DIFFERENCES

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = 2.00$$

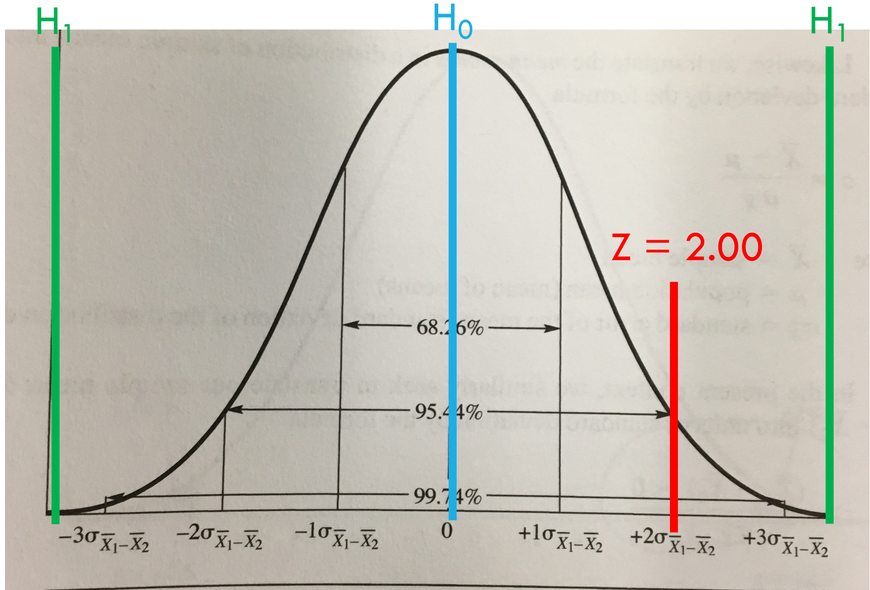


FIGURE 7.4 ***The Sampling Distribution of Differences between Means as a Probability Distribution***

EXAMPLE 2: Z-SCORE FOR MEAN DIFFERENCE COMPARED TO SAMPLING DISTRIBUTION OF MEAN DIFFERENCES

But what if we switched groups so that liberals were group 1?

EXAMPLE 2: Z-SCORE FOR MEAN DIFFERENCE COMPARED TO SAMPLING DISTRIBUTION OF MEAN DIFFERENCES

Liberal mean ($\bar{X}_1$) = 54.0

Conservative mean ($\bar{X}_2$) = 58.0

Difference-between-means would be negative ($\bar{X}_1 - \bar{X}_2$) = -4.0

So the z-score would also be negative

- Z = $\frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{(54 - 58)}{2.00} = \frac{-4}{2.04}$ = -2.00

EXAMPLE 2: Z-SCORE FOR MEAN DIFFERENCE COMPARED TO SAMPLING DISTRIBUTION OF MEAN DIFFERENCES

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = -2.00$$

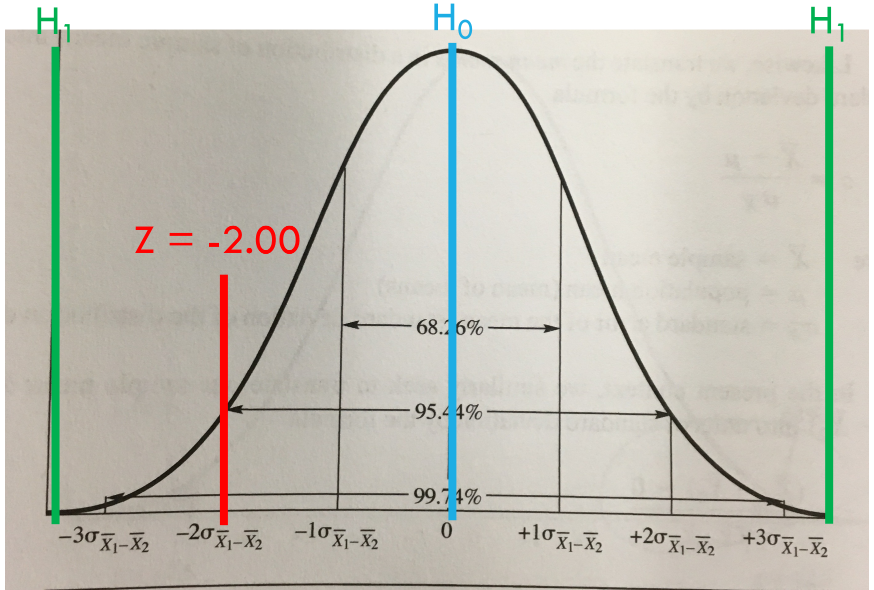


FIGURE 7.4 ***The Sampling Distribution of Differences between Means as a Probability Distribution***

EXAMPLE 2: Z-SCORE FOR MEAN DIFFERENCE COMPARED TO SAMPLING DISTRIBUTION OF MEAN DIFFERENCES

We have two z-scores… (one positive, one negative)

This means that the z-scores are tested based on the two tails of the distribution

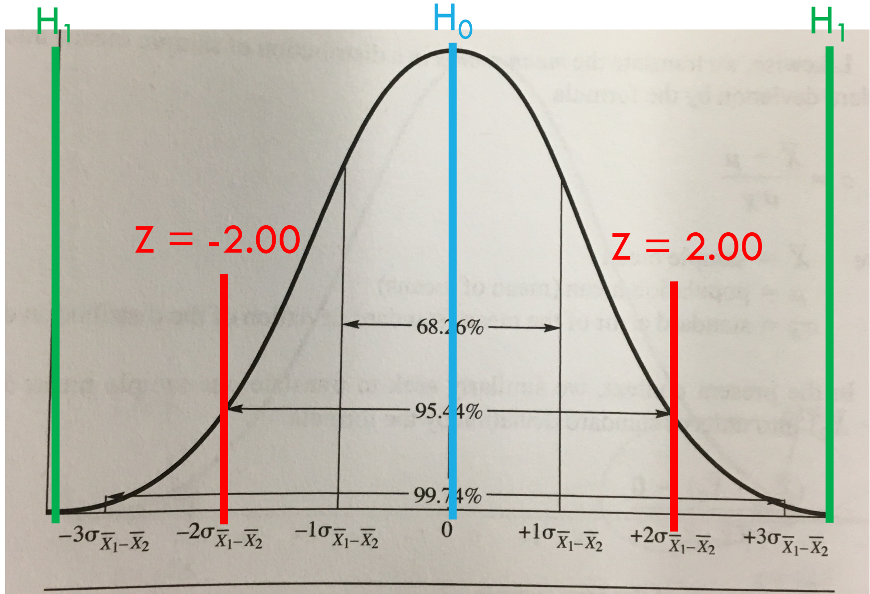


FIGURE 7.4 *The Sampling Distribution of Differences between Means as a Probability Distribution*

$Z = \pm 2.00$

BUT WHAT IS AN EXTREME Z-SCORE?

Is our calculated Z score ($Z_{calculated}$) of ± 2.00 *extreme enough* to be extreme evidence AGAINST the null hypothesis?

- Is it extreme enough to reject the null hypothesis?

Need a cutoff Z score to compare with our calculated Z-score

BUT WHAT IS AN EXTREME Z-SCORE?

$Z_{calculated}$

- Our calculated Z for our mean difference

$Z_{critical}$

- Our cutoff Z score, that we compare with our $Z_{calculated}$
- Stays the same, selected before analyzing differences
- Cuts off an area of the Normal Distribution that represents a percent likelihood (or probability) that the null hypothesis is TRUE
  - The area beyond the $Z_{critical}$ (on both sides/tails of the distribution) represents the likelihood that the null hypothesis is TRUE
  - We want a low percent chance that the null is true

BUT WHAT IS AN EXTREME Z-SCORE?

Extreme mean difference (extreme Z-score) if:

$$|Z_{calculated}| \geq |Z_{critical}|$$

CRITICAL Z-SCORES

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

How is $Z_{critical}$ selected?

$Z_{critical}$ is based on α, or the maximum allowed probability that the null will be true

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

Alpha (α)

- Cutoff for rejection region (tails)

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

Alpha (α) = .05

- 5% chance null hypothesis is true
- 5% of mean differences/cases considered extreme (rejection region)
- 5% cases left in tails of distribution
- .05 maximum allowed probability that the null is true

- Associated Z score ($Z_{critical}$) = ± 1.96

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

α	% cases in tails/extremes/rejection region	Maximum probability H_0 is true	$Z_{critical}$
.05	5%	.05	± 1.96

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

If $|Z_{calculated}| \geq |Z_{critical}|$

- Calculated mean difference is *more extreme* than our cutoff for "extremeness"
- Calculated mean difference is extremely different from what is expected by the null hypothesis

- Statistically significant (extreme) mean difference from what what was expected by the null hypothesis

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

For $\alpha = .05$

- If $| Z_{calculated} | \geq | \pm 1.96 |$
  - Calculated mean difference is ***more extreme*** than our cutoff for “extremeness”
  - Calculated mean difference is extremely different from what is expected by the null hypothesis

  - Statistically significant (extreme) mean difference from what what was expected by the null hypothesis

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

Z = $\pm$ 1.96 spans 95% of all cases around the mean

- Total of 5% of cases are between the Z-scores and their closest tails, Z score and more extreme

- <u>2.5%</u> of all possible mean differences are <u>in the top end of the distribution</u>
- <u>2.5%</u> of all mean differences <u>in the bottom end of the distribution</u>

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

For = $\pm$ 1.96

- There is a only a 5% chance, or .05 maximum probability that our mean difference or *more* extreme occurred because of sampling error
  - Very unlikely that this mean difference was because of sampling error

- Therefore, if its not because of sampling error… we <u>REJECT THE NULL HYPOTHESIS</u> that there is no mean difference between the two populations.

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

If alpha is very small, it is harder and harder to reject the null hypothesis

- Smaller rejection regions
- More stringent requirements for rejecting the null

What if you set alpha (the maximum probability that the null hypothesis is allowed to be true) to a smaller number

- α = .01
- α = .001

STATISTICAL SIGNIFICANCE (α AND P)

α	% cases in tails/extremes/rejection region	Maximum probability H_0 is true	$Z_{critical}$
.05	5%	.05	± 1.96
.01	1%	.01	± 2.58
.001	.1%	.001	± 3.29

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

For $\alpha = .01$

- If $|Z_{calculated}| \geq |\pm 2.58|$

  - Calculated mean difference is *more extreme* than our cutoff for “extremeness”
  - Calculated mean difference is extremely different from what is expected by the null hypothesis

  - Statistically significant (extreme) mean difference from what what was expected by the null hypothesis

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

For $\alpha = .001$

- If $| Z_{calculated} | \geq | \pm 3.29 |$
  - Calculated mean difference is *more extreme* than our cutoff for "extremeness"
  - Calculated mean difference is extremely different from what is expected by the null hypothesis

  - Statistically significant (extreme) mean difference from what what was expected by the null hypothesis

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

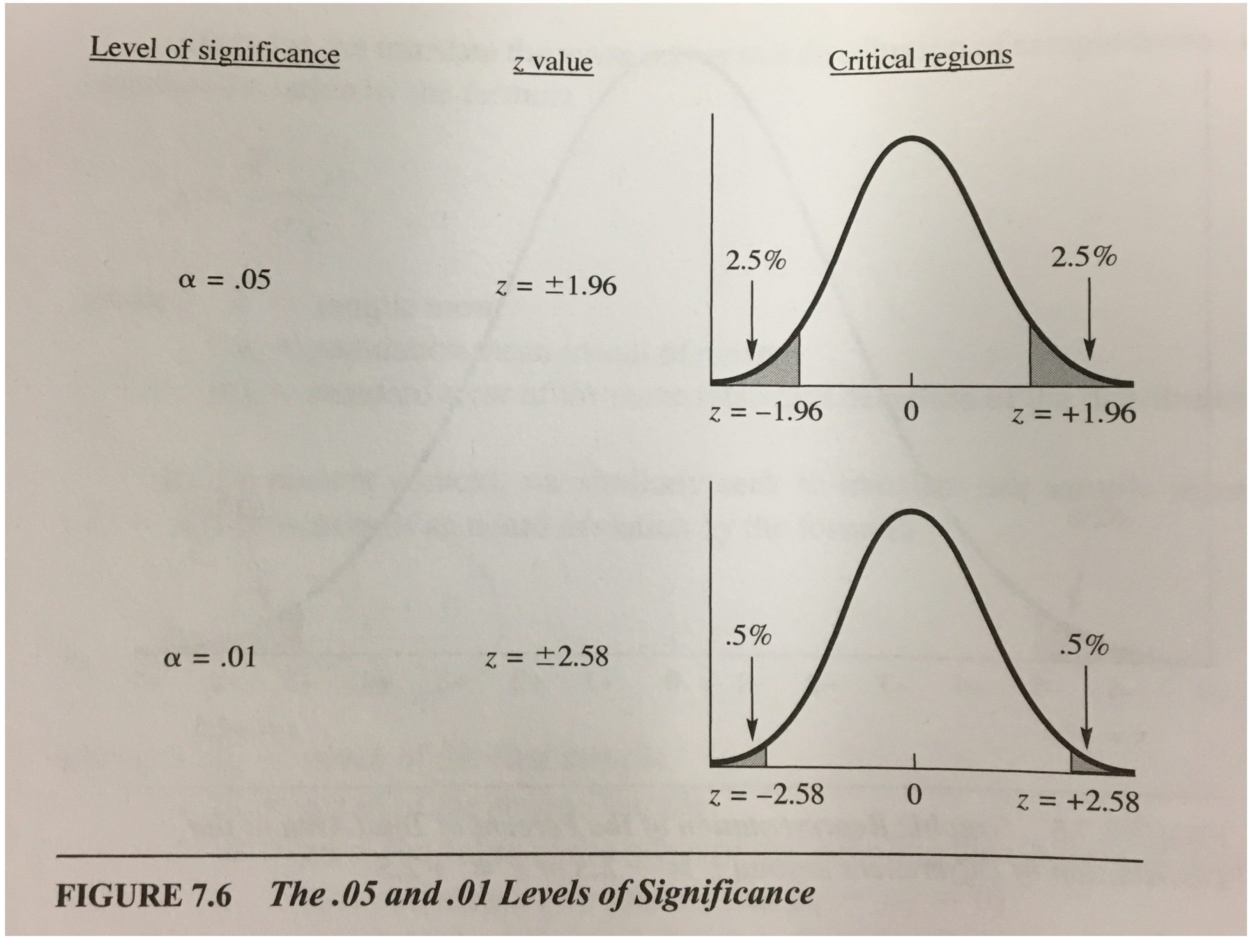


FIGURE 7.6 *The .05 and .01 Levels of Significance*

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

While α is the maximum probability allowed for the null to be true (a cutoff)…

- *p* is different

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

Probability (p)

- Actual probability of finding this extreme of a difference if the null hypothesis is true
  - Probability that the null hypothesis is true, given this extreme of a difference
- Area from that Z score to the closest tails.

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

Difference between p and α

- α
  - A threshold selected beforehand
  - Has associated $Z_{critical}$

- p
  - Based on the data
  - Exact probability of finding this difference, if the null hypothesis is true
  - Based on $Z_{calculated}$

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

For difference to be extreme, we need:

- $|Z_{calculated}| \geq |Z_{critical}|$

If $|Z_{calculated}| \geq |Z_{critical}|$, then..

- the area in the extremes of $Z_{calculated}$ (p, or the exact probability that the null is true for THIS difference) should be smaller than the area in the extremes of the cutoff of $Z_{critical}$ (α, or the maximum probability that the null hypothesis is true)

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

If $|Z_{calculated}| \geq |Z_{critical}|$, then…

- Difference between mean of group 1 and mean of group 2 is so extreme that we can't blame it on sampling error, therefore… H_0 is probably not true, so…
  - $Z_{calculated}$ is in rejection region
  - Reject H_0
  - $p \leq \alpha$

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

If $|Z_{calculated}| \geq |\pm 1.96|$, then..

- Reject H_0
- $p \leq .05$

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

If $|Z_{calculated}| \geq |\pm 2.58|$, then..

- Reject H_0
- $p \leq .01$

CUTOFF Z-SCORE ($Z_{CRITICAL}$), AND STATISTICAL SIGNIFICANCE (α AND P)

Difference between p and α

- If p is less than or equal to α, the null hypothesis is considered to be false, so we reject it:
  - Statistically Significant Difference: $p \leq \alpha$ OR $p \leq .05$

BUT WHAT IS AN EXTREME Z-SCORE?

In our example, our $Z = \pm 2.00$

- What is α?
- What is p?

- Do we reject or fail to reject?

AGAIN, HYPOTHESIS TESTING FOR MEAN DIFFERENCES

Remember Z-score formula:

- $Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}}$

But we never truly know $\boldsymbol{\sigma_{\bar{X}_1 - \bar{X}_2}}$, the standard deviation of the sampling distribution of the difference between means

- So we have to estimate it

AGAIN, HYPOTHESIS TESTING FOR MEAN DIFFERENCES

Similar to Standard Error of the Mean

- Never truly know standard deviation of the population, or the standard deviation of the sampling distribution, since we don't actually take multiple samples

So we estimate the *standard deviation* of the *sampling distribution of difference between means* from our one (paired) sample

- Standard Error of the Difference between Means

AGAIN, HYPOTHESIS TESTING FOR MEAN DIFFERENCES

Estimating the Standard Deviation of the Sampling Distribution of the Difference between Means

- Standard Error of the Difference

$$\sigma_{\bar{X}_1 - \bar{X}_2} \approx SD_{\bar{X}_1 - \bar{X}_2} = \sqrt{\left(\frac{SD_1^2}{N_1}\right) + \left(\frac{SD_2^2}{N_2}\right)}$$

AGAIN, HYPOTHESIS TESTING FOR MEAN DIFFERENCES

So our Z-score formula becomes:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{SD_{\bar{X}_1 - \bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\left(\frac{SD_1^2}{N_1}\right) + \left(\frac{SD_2^2}{N_2}\right)}}$$

HYPOTHESIS TESTING FOR MEAN DIFFERENCES: ADJUSTING Z TO OUR DATA

But when we substitute Standard Error of the Difference, the distribution becomes a t-distribution, making this a *t*-test:

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{SD_{\bar{X}_1 - \bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\left(\frac{SD_1^2}{N_1}\right) + \left(\frac{SD_2^2}{N_2}\right)}}$$