

Inferential Statistics

Inferential statistics allows us to infer trends about a larger population (= collection of all elements possessing common characteristics that comprise universe) based on a study of a sample (= subgroup of the members of population chosen for participation in the study) taken from it.

We use inferential statistics to examine the relationships between variables within a sample and then make predictions about how those variables will relate to a larger population.

The conclusion of a statistical inference is a statistical proposition, which can include any of the following (we will cover these concepts later):

- point estimate (a particular value that best approximates some parameter of interest)
- an interval estimate (e.g. a confidence interval, constructed in a way so that under repeated sampling such intervals would contain the true parameter value)
- a credible interval (set of values containing, e.g. 95% of posterior belief)
- rejection of hypothesis
- clustering or classification of data points into groups

Sampling methods

The population can be very broad or quite narrow: maybe we want to make inferences about the whole adult population of a country, maybe our research focuses on customers of a certain company, patients with a specific health condition, or students in a single school.

If the population is very large, demographically mixed and geographically dispersed, it might be difficult to gain access to a representative sample.

In general, the larger the sample size, the more accurately and confidently we can make inferences about the whole population.

Probability Sampling methods

- every member of the population has a chance of being selected
- used for quantitative research

1. simple random sample

→ every member of the population has an equal chance of being selected
(tools: random number generator.)

2. systematic sampling

→ every member of the population is listed with a number, but instead of randomly generating numbers, individuals are chosen at regular intervals

3. stratified sampling

→ when the population has mixed characteristics divide population into subgroups (strata) based on relevant characteristic and from overall proportions of the population, calculate how many people should be sampled from each subgroup.
Then use random or systematic sampling to select a sample from each subgroup

4. cluster sampling

→ divide population into subgroups, but each subgroup should have similar characteristics to the whole sample. Instead of sampling individuals from each subgroup, randomly choose entire subgroups

e.g. The company has offices in 10 cities (all with roughly the same number of employees in similar roles). You don't have the capacity to travel to every office to collect your data, so you select 3 random offices - these are your clusters

Non-probability Sampling methods

→ individuals/elements are selected based on non-random criteria

→ easier but higher risk of sampling bias, appropriate for qualitative research, where the aim is not to test a hypothesis about a broad population, but to develop an initial understanding of a small under-researched population

1. convenience sample

→ includes individuals who are most accessible

2. voluntary response sampling

→ people volunteer (e.g. public survey) → biased

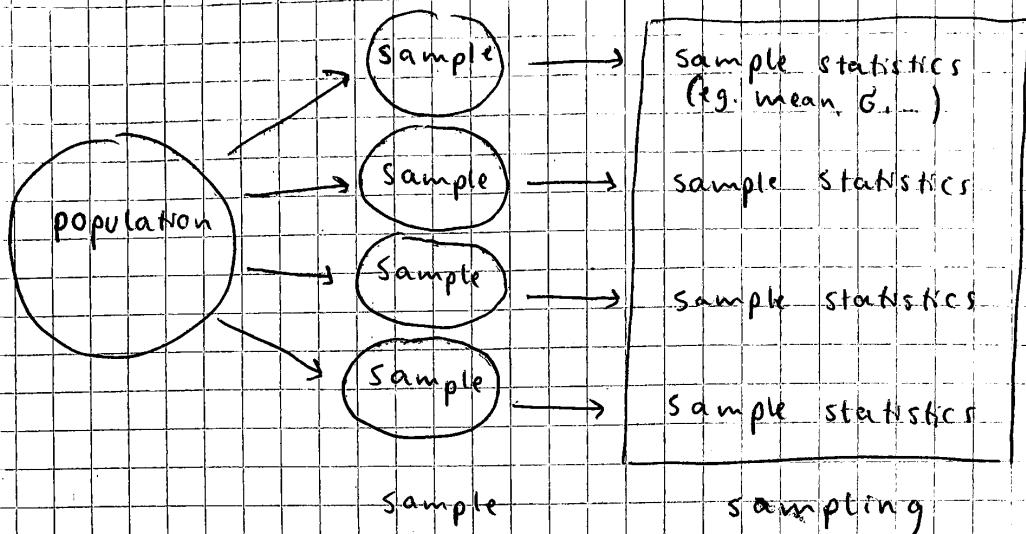
3. purposive sampling

→ researcher uses their judgement to select a sample that is most useful to the purposes of the research

4. snowball sampling

→ if population is hard to access, s.s. can be used to recruit participants via other participants

Sampling distribution



We have a population of x values whose histogram is the probability distribution of x .

Select a sample of size n from this population and calculate a sample statistic e.g. the mean $E(X)$ and independently

This procedure can be repeated indefinitely and generates a population of values for the sample statistic and the histogram is the sampling distribution of the sample statistics.

population distribution			
	5, 6, 7, 8		
some sample distributions	✓	↓	→
	sample 1: 5, 6, 7	sample 2: 5, 8, 8	sample 3: 7, 8, 5
	↓	↓	↓
sampling distribution of means	6.0	7.0	6.66

Every statistic has a sampling distribution. (standard deviation, sample proportion, variance, range)
The sampling distribution of the mean is denoted as μ_x

The standard deviation of the sampling distribution of the mean is called the standard error of the mean and is symbolized by σ_x

The standard deviation and variance measure the variability of the sampling distribution, where the number of observations in a population, the number of observations in a sample and the method determine the variability of a sampling distribution.

Knowing how spread apart the mean of each of the sample sets are from each other and from the population mean will give an indication of how close the sample mean is to the population mean. The standard error of the sampling distribution decreases as the sample size increases *

A population or one sample set of numbers will have a normal distribution.

However, because a sampling distribution includes multiple sets of observations, it will not necessarily have a bell-shaped curve.

* The standard error of the mean is calculated like the following:

$$S_e = \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation of the population distribution and n is the sample size.

→ The sample size must be quadrupled to achieve half the measurement error. This may have a role in understanding cost-benefit tradeoffs.

Normal Distribution

Normal distribution also known as the Gaussian distribution, is a continuous probability distribution that is symmetric about the mean, showing that data near the mean are more frequent than data far from the mean. However not all symmetric distributions are normal.

In graph form, the normal distribution will appear as a bell curve. The general form of its pdf is

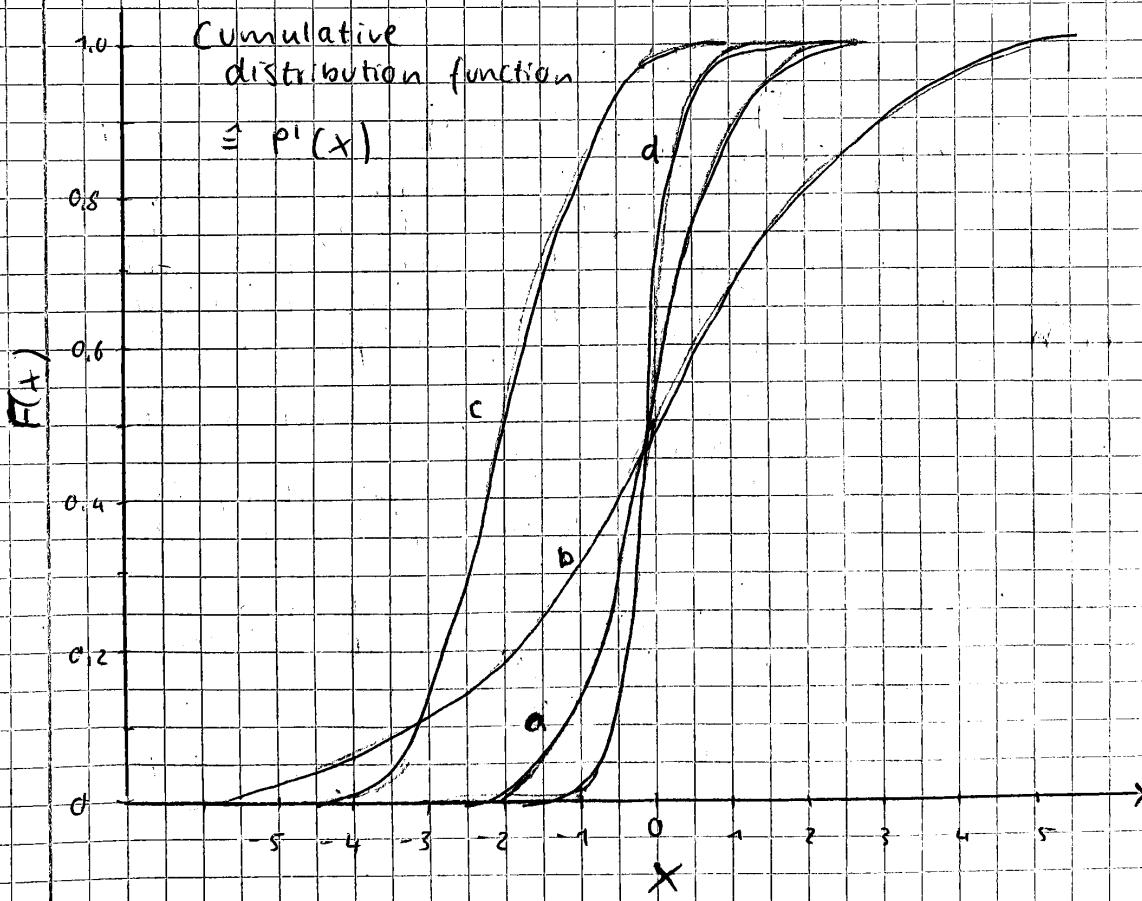
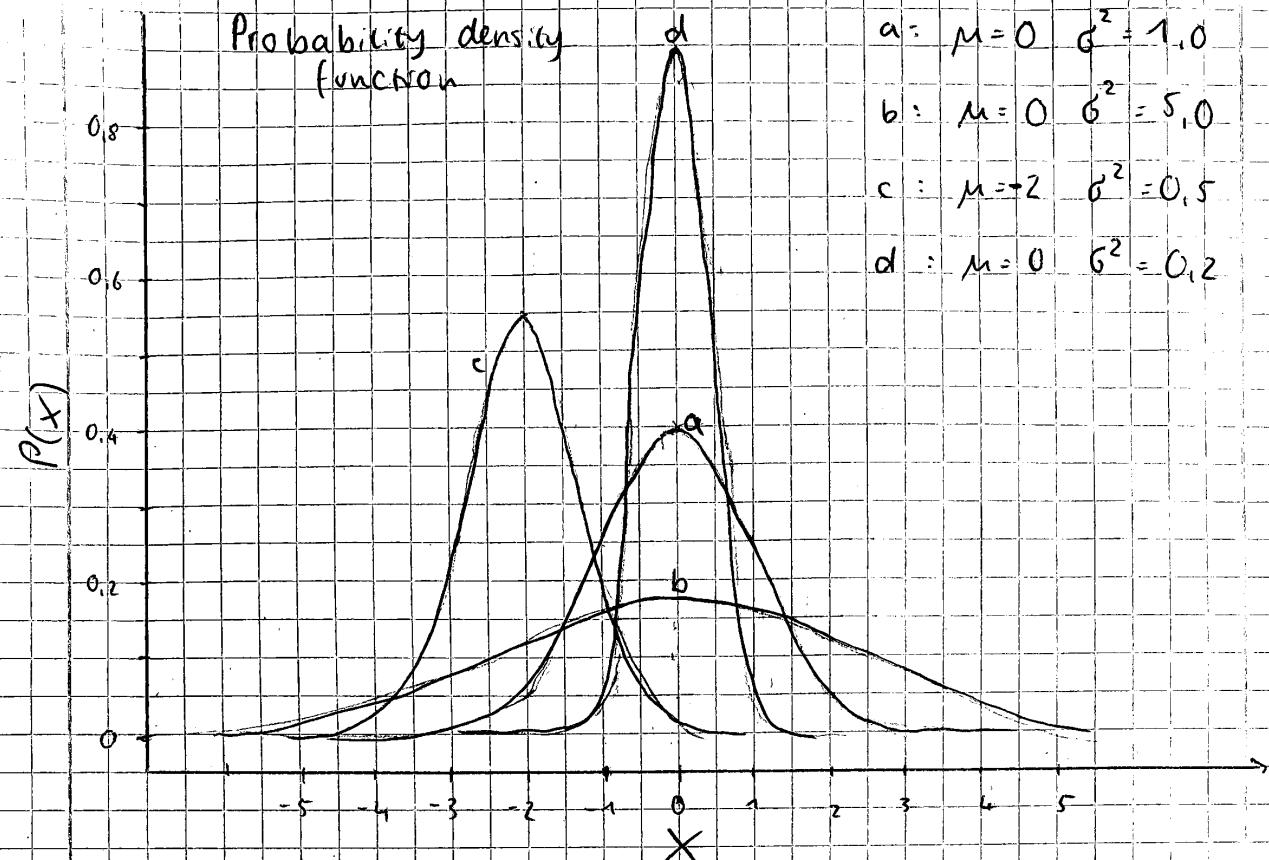
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

The normal distribution is the most used type of distribution assumed in stock market analysis and other.

68% of the observations are within +/- one standard deviation of the mean, 95% are within +/- two standard deviation and 99.7% are within +/- three standard deviation.

Mean, median ("middle" value) and mode (most frequent value) of the normal distribution are the same.

2 The basic advantage of the median to the mean is that it is not skewed by a small proportion of extremely large or small values, so it may give a better idea of a "typical" value.



The normal distribution model is motivated by the
Central Limit Theorem.



Normal Approximation of Binomial

$$X \sim B_{n,p} \quad \mu = np \quad \sigma = \sqrt{npq}$$

$$Y \sim N(np, npq)$$

$$P(X=k) \approx P(k - \frac{1}{2} \leq Y \leq k + \frac{1}{2})$$

$$X \sim B_{100, 0.5} \quad np = 50 \quad \sqrt{npq} = 5$$

$$Y \sim N(50, 25)$$

$$Z = \frac{Y - 50}{5} \quad [1]$$

standardize

$$= P\left(1.9 \leq Z \leq 2.1\right) = \frac{59.5 - 50}{5} \leq Z \leq \frac{60.5 - 50}{5} = 2.1$$

$$= \Phi(2.1) - \Phi(1.9) \approx 0.9821 - 0.9713 = 0.0108$$

$$P(X=60) = \binom{100}{60} 0.5^{100} \approx 0.0108 \quad \checkmark$$

Central Limit Theorem

This theory states that averages calculated from independent, identically distributed random variables have approximately normal distributions, regardless of the type of distribution from which the variables are sampled (given it has finite variance) (although only for reasonable sample sizes, say $n > 30$)

Suppose we are sampling from a population with mean μ and standard deviation σ . Let \bar{X} be a random variable representing the sample mean of n independently drawn observations.

Then:

- The mean of the sampling distribution of the sample mean is equal to the population mean $\mu_{\bar{X}} = \mu$

- The standard deviation of the sampling distribution of \bar{X} is equal to $(\pm \text{standard error})$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

The Z-score [1] standardizing normally distributed random variables tells us the distance between the mean and a data point

standard deviations

$$[1] \text{ Standardizing normally distributed random variables from } X \sim N(\mu, \sigma^2) \text{ to } Z \sim N(0, 1)$$

$X - \mu \rightarrow$ this quantity has a mean of zero

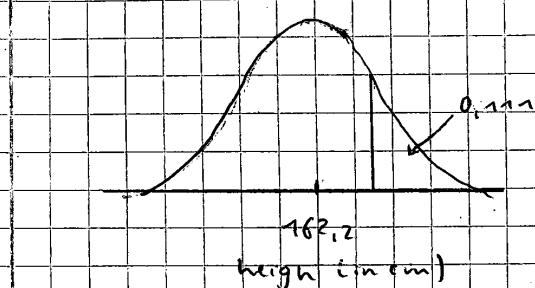
$$Z = \frac{X - \mu}{\sigma} \rightarrow \text{we're forcing this quantity to have } \sigma = 1$$

$\rightarrow Z$ is now a random variable that has the standard normal distribution $* Z \sim N(0, 1) \rightarrow 61$

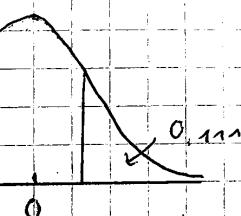
e.g. What is the probability a randomly selected adult American female is taller than 170.5 cm?
 $(\mu = 162.2, \sigma = 6.8)$

$$\begin{aligned} P(X > 170.5) &= P\left(\frac{X-\mu}{\sigma} > \frac{170.5-\mu}{\sigma}\right) \\ &= P\left(Z > \frac{170.5 - 162.2}{6.8}\right) \\ &= P(Z > 1.22) \quad \text{look up in Z-score table} \\ &= 1 - P(Z \leq 1.22) \quad \rightarrow = 1 - 0.8888 = 0.111 \end{aligned}$$

the distribution of heights



the standard normal distribution



The Z value of 1.22 means that 170.5 is 1.22 standard deviations above the mean of 162.2

A little more formally,

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\text{d}} N(0, 1) \text{ as } n \rightarrow \infty$$

(provided μ and σ are finite)

e.g. Suppose salaries at a very large corporation have a mean of \$62 000 and a standard deviation of \$32 000.

$$\mu = 62\ 000 \quad \sigma = 32\ 000$$

If 100 employees are randomly selected what is the probability their salary exceeds \$66 000?

$$P(\bar{X} > 66\ 000)$$

$$\rightarrow P\left(Z > \frac{66\ 000 - 62\ 000}{\frac{32\ 000}{\sqrt{100}}}\right) = P(Z > 1.25) = 0.106$$

Z must also have approximately a normal distribution

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow \bar{X} \text{ (normally distributed)}$$

Exercise A girl has a road-side stand and is selling boxes of 36 apples. The mean weight of an apple is 5.4 oz and $\sigma = 0.5$ oz. What proportion of boxes contain at least 200 ounces of apples?

$$\mu = 5.4 \quad \sigma = 0.5 \quad n = 36$$

$$X \equiv \text{weight of apple} \quad \frac{200}{36} = 5.5 = X$$

$$Z = \frac{X - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{5.5 - 5.4}{\frac{0.5}{\sqrt{36}}} = 0.083$$

$$= \frac{5.5 - 5.4}{0.083} \approx 1.93$$

$$P(Z \geq 1.93) = 0.0268$$

Confidence intervals

A sample will never be a perfect representation of the respective population. Different samples of the same population will give different results.

This is called sampling error or variation due to sampling.

When we express an estimate of a population parameter it is good practice to give it a confidence interval.

X "The mean for the population equals..." •

✓ "The mean for the population lies between..." ←

A confidence interval communicates how accurate our estimate is likely to be.

What affects the width of confidence intervals?

1. variation within the population of interest
2. size of the sample

1. If all the values in the population were almost the same, then our sample will also have little variation. Any sample we take is likely to be pretty similar to any other sample. → We would have a small confidence interval.

2. If we take a small sample we don't have much information on which to base our inference. Small samples will vary more from each other. Large samples on the other hand are more similar to each other which leads to narrower confidence intervals.

Deriving a Confidence Interval for the Mean

(when sampling from a normally distributed population with a known value of σ)

The confidence interval for μ will be of the form:

$$\bar{X} \pm \text{margin of error}$$

Suppose we draw a random sample of n independent observations from a normally distributed population with mean μ and standard deviation σ .

\bar{X} is a normally distributed random variable with a mean of μ and a standard deviation of $\frac{\sigma}{\sqrt{n}}$

$$\rightarrow \text{standardized: } Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

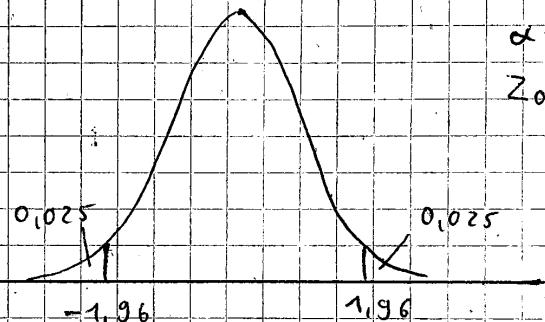
$$P(-1,96 < Z < 1,96) = 0,95$$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\alpha = 0,05$$

$$Z_{0,025} = 1,96$$

The area under the standard normal curve between $-1,96$ and $1,96$ is $0,95$



(95% confidence level is the most commonly chosen one)

To generalize:

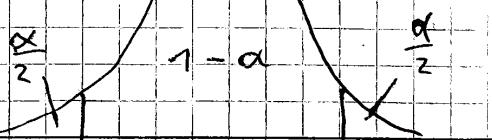
$$P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} + z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

reminder: we're trying to get a confidence interval for μ
isolate μ

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

lower bound fixed, unknown quantity

upper bound



$z_{\frac{\alpha}{2}}$ is the value of z that yields an area to the right of $\frac{\alpha}{2}$

\Rightarrow A $(1 - \alpha) 100\%$ confidence interval for

μ is given by: $\bar{X} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

$$\text{e.g. } 95\% \quad \alpha = 0,05 \rightarrow z_{0,05} = 1,96$$

$$\rightarrow \bar{X} \pm 1,96 \cdot \frac{\sigma}{\sqrt{n}}$$

\hookrightarrow we use the standard deviation of the sample as a measure of variance in the population

Side note on the standard error $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

- Dividing by the square root of the sample size reflects the phenomenon that the more information the less new information we get from one more observation in the sample.

e.g. It makes sense that increasing our sample size from 10 to 20 will give more information and thus decrease our confidence interval more than increasing the sample size from 90 to 100.

The confidence interval gives us an estimated range of values that seem reasonable based on what we've observed. Its' center is still the sample mean, but we've got some room on either side for our uncertainty.

The 95% in a 95% confidence interval tells us that if we calculated a confidence interval from 100 different samples, about 95 of them would contain the population parameter.

Example (Sigma known)

The ratio of the length of the second digit (index finger) to that of the fourth digit finger (ring finger), known as the 2D:4D ratio, is often studied by researchers.

A study investigated the 2D:4D ratio in women of European descent at a large college.

A sample of 135 women yielded an average 2D:4D ratio of 0.988

? What is a 95% confidence interval for the population mean?

(Suppose it is known that $\sigma = 0.028$)

$$\bar{X} = 0.988 \quad \sigma = 0.028 \quad n = 135 \quad \alpha = 0.05$$

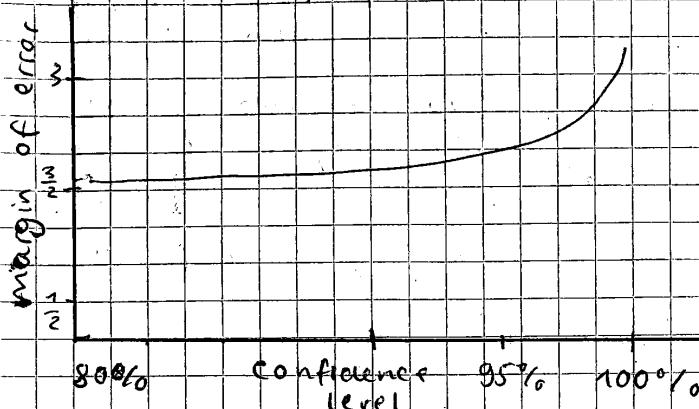
$$\rightarrow 0.988 \pm 1.96 \cdot \frac{0.028}{\sqrt{135}} \\ (\approx 0.025)$$

$$0.988 \pm 0.0047$$

$$\rightarrow (0.983, 0.993)$$

We can be 95% confident that the population mean lies in this interval.
(2D:4D)

- ! Anytime our sample is not a random sample from the population of interest, there may be biases present



$Z_{\alpha/2}$
95% 1.96
99% 2.576
→ beyond 95% this curve goes up rather quickly

In most practical situations we feel that the choice of a 95% confidence level provides a good balance between high confidence and a reasonable margin of error.

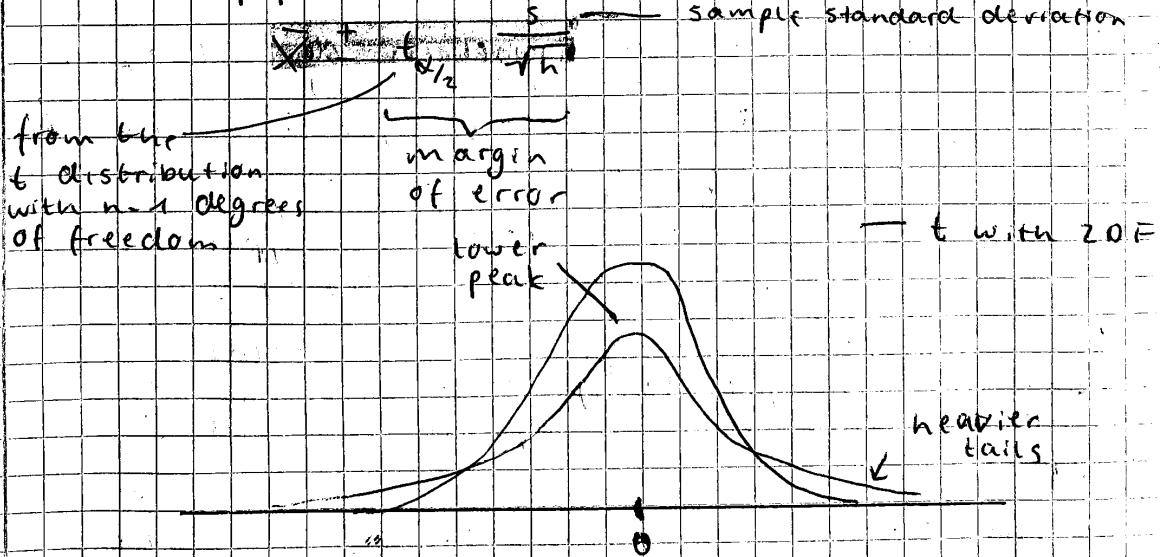
Calculation of Confidence Intervals for one mean

- sigma not known (t method)

For these methods to be reasonable, we require

- a simple random sample from the population of interest
- a normally distributed population

If the population standard deviation σ is not known:



The precise shape of the t distribution depends on the degrees of freedom, which are related to the sample size. As DF increase t tends towards the standard normal distribution.

Example: A cereal producer makes cereal boxes with a stated weight of 750 grams

From a large lot of these boxes a random sample of 7 boxes yielded $\bar{x} = 795.3 \text{ g}$ and $s = 17.8 \text{ grams}$.

Construct a 95% confidence interval for μ , the mean weight of boxes in this lot.

Look up $t_{\alpha/2}$ score for $n-1=6$ degrees of freedom

$$\alpha = 0.05 \quad \rightarrow t_{0.025} = 2.447$$

$$\rightarrow \bar{X} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \\ = 795.3 \pm 2.447 \cdot \frac{17.8}{\sqrt{7}} \\ = 795.3 \pm 16.46$$

$$(778.8, 811.8)$$

→ We can be 95% confident that the true mean weight of cereal in boxes from this lot lies between 778.8 and 811.8 grams.

Hypothesis testing

In hypothesis testing, we turn a question of interest into hypotheses about the value of a parameter or parameters.

We create:

- a null hypothesis, denoted by H_0
("status quo hypothesis")
- an alternative hypothesis, denoted by H_a
("research hypothesis")

→ It is often the hypothesis the researcher is trying to show

We calculate an appropriate test statistic (based on sample data) and determine how much evidence there is against the null hypothesis.

If the evidence is strong enough (needs to meet a certain significance level), we can reject the null hypothesis in favour of the alternative hypothesis.

example: Do men and women have different average salaries after graduating university?

Z-tests for Population mean μ

- two scenarios:
- σ is known \rightarrow Z-test (rare in practice)
 - σ is not known
 \rightarrow more common
hypothesized value

$$H_0: \mu = \mu_0$$

We choose one of the possible alternatives:

- $H_a: \mu < \mu_0$ } one-sided alternatives
- $H_a: \mu > \mu_0$ } (one-tailed tests)
- $H_a: \mu \neq \mu_0$ ← two-sided alternative
(two-tailed test)

! The appropriate choice of alternative hypothesis depends on the ^{given} problem, and should not be based on the current sample's data.

You should not use the same data that suggests a hypothesis to test that hypothesis.

Suppose we have a simple random sample of n observations from a normally distributed population where σ is known

(The normality assumption is very important for a small sample size, but as the sample size gets larger and larger that normality assumption is less and less important due to the central limit theorem.)

To test $H_0: \mu = \mu_0$ we use the test statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma_x} \text{ where } \sigma_x = \frac{\sigma}{\sqrt{n}}$$

If H_0 is true, the Z test statistic will have the standard normal distribution.

The rejection region approach:

> choose a value for α , the significance level of the test

(α is the probability of rejecting the null hypothesis if it is true)

often $\alpha = 0.05$

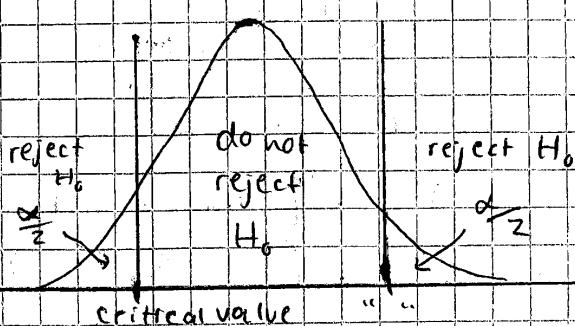
> find appropriate rejection region

→

> reject the null hypothesis if the test statistic falls in the rejection region

Suppose $H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}$$



If this null hypothesis is true, then it's going to be very unlikely to get Z values that are far out in the tails of the distribution

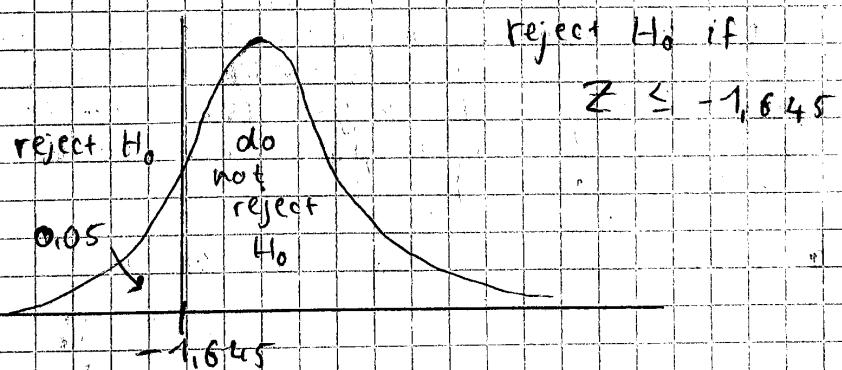
→ Z values far out in the tails gives strong evidence against the null hypothesis,

e.g. $\alpha = 0,05$

$$Z_{0,025} = 1,96 \rightarrow \text{Reject } H_0 \text{ if } Z \geq 1,96 \text{ or } Z \leq -1,96$$

Suppose $H_0: \mu = \mu_0$ $H_a: \mu < \mu_0$ $\alpha = 0,05$

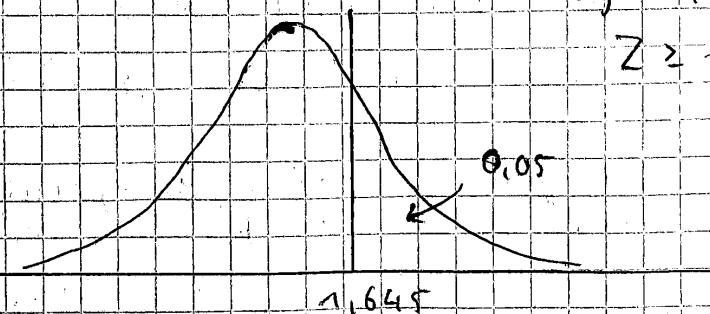
$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}$$



Suppose $H_0: \mu = \mu_0$ $H_a: \mu > \mu_0$ $\alpha = 0,05$

reject H_0 if

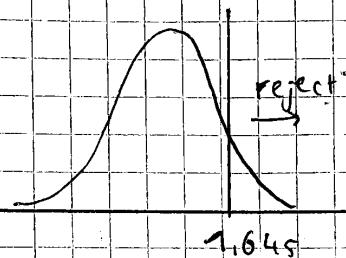
$$Z \geq 1,645$$



But in some ways the rejection region approach is a bit silly.

Suppose $H_0: \mu = \mu_0$, $H_a: \mu > \mu_0$, $\alpha = 0.05$

- 1 $Z_1 = 1.6448$ do not reject H_0
- 2 $Z_2 = 1.6451$ reject H_0
- 3 $Z_3 = 17.819$ reject H_0



- 1 With these observed Z values the conclusions would be very different, even though these two Z values are very, very close.
- 2 Here we have the same conclusions even though Z_3 has by a great deal more evidence against the null hypothesis.

→ reason for p-value approach

The p-value approach

The p-value is a measure of the strength of the evidence against the null hypothesis.

Precisely, the p-value is the probability of getting the observed value of the test statistic, or a value with even greater evidence against H_0 , if the null hypothesis is true.

Example

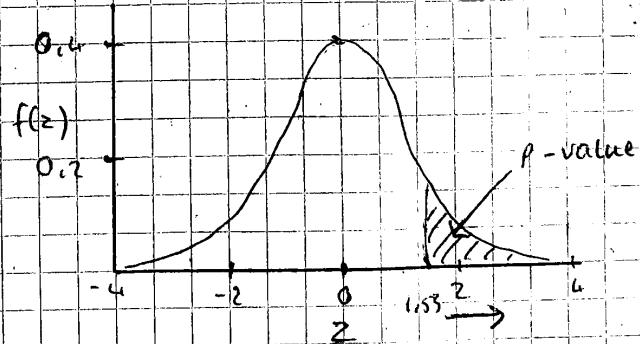
distribution of Z if H_0 is true

$$H_0: \mu = \mu_0$$

$$H_a: \mu > \mu_0$$

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

$$\text{Suppose } Z = 1.53$$



Large values of the test statistic Z are going to give evidence against H_0 in favor of the alternative.

So the p-value is the area out to the right of our test statistic under the standard normal curve

$$p\text{-value} = P(Z > 1.53)$$

$$= 0.063$$

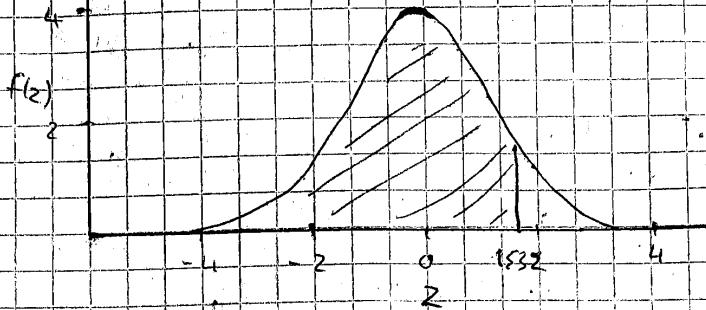
$H_0: \mu = \mu_0$ distribution of Z if H_0 is true

$H_a: \mu < \mu_0$

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

suppose

$$\bar{x} = 1,53$$



$$p\text{-value} = P(Z < 1.53)$$

$$= 0,937$$

$H_0: \mu = \mu_0$

$H_a: \mu \neq \mu_0$

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$Z = 1,53$$

p-value is double this area

$$p\text{-value} = P(Z > 1.53) +$$

$$P(Z < -1.53)$$

$$= 2P(Z > 1.53) = 0,126$$

The smaller the p-value, the greater the evidence against the null hypothesis.

If we have a given significance level α , then:

reject H_0 if p-value $\leq \alpha$

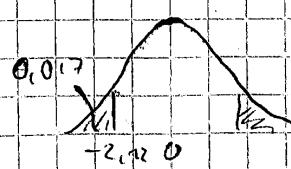
(If p-value $\leq \alpha$, the evidence against H_0 is significant at the α level of significance.)

example:

Suppose $H_0: \mu = 10$ $H_a: \mu \neq 10$ $\alpha = 0,05$

$$Z = -2,12 \quad p\text{-value} = 2 \cdot 0,017 \\ = 0,034$$

→ The p-value is less than α , so the evidence against H_0 is significant at $\alpha = 0,05$. We reject H_0 in favor of the alternative



Exercise: A producer of cereal is producing boxes of cereal with a stated weight of 250 grams.

It is known from a large body of past experience that the standard deviation of the weights in their filling process is approximately 16 grams.

In order to ensure that not many boxes are underfilled, the producer sets the mean fill amount at 780 g.

As part of the quality control process, they periodically draw a random sample of 25 boxes, measure the weights and test the null hypothesis that the mean is 780 g.

For a sample of 25 boxes, $\bar{X} = 776$. Does this give strong evidence the true mean weight differs from 780 g?

$$\alpha = 0,05$$

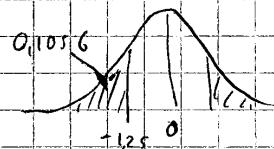
$H_0: \mu = 780$ $H_1: \mu \neq 780$ (we care about difference in either direction)

$$\bar{X} = 776 \quad \sigma = 16 \quad n = 25$$

→ two-sided

$$\rightarrow Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{-4}{\frac{16}{\sqrt{25}}} = -1,25$$

$$p\text{-value} = 2 \cdot 0,1056 = 0,211$$



We can say that the evidence against the null hypothesis is significant if p-value is less than or equal 0,05

→ The evidence against H_0 is not significant at the $0,05$ level of significance.

We do not have significant evidence that the population mean differs from 780g.

Note! This does not imply there is strong evidence the mean is 780!

One-Sided or Two-Sided Test?

We might strongly suspect that a newly developed drug decreases blood pressure.

We may wish to test: $H_0: \mu = 0$

But what is the appropriate alternative?

$H_0: \mu \neq 0$ or $H_0: \mu < 0$?

Because we use the word decrease, we're suspecting a decrease, we might be tempted to choose the one-sided alternative.

What we're gaining if we use the one-sided approach is greater power in that direction, BUT if the population mean is greater than 0 we will not be able to detect that difference.

A two-sided alternative allows us to see a difference from the hypothesized value in either direction.

Choose a one-sided alternative if you care about a difference in only one direction.

Type I & II Errors and the Power of the Test

- A Type I error is rejecting H_0 when, in reality, it is true.
- A Type II error is failing to reject H_0 when, in reality, it is false.

In practice we will not know for certain if we made the correct decision or if we made one of these two errors.

		Unknown underlying reality	
		H_0 is false	H_0 is true
Conclusion from test	reject H_0	correct decision	type I error
	do not reject H_0	type II error	Correct decision

↑
known

- The probability of a Type I error, given H_0 is true, is called the significance level of the test.

$$P(\text{Type I error} | H_0 \text{ is true}) = \alpha$$

We get to pick the value of alpha that we think is appropriate for any given problem.

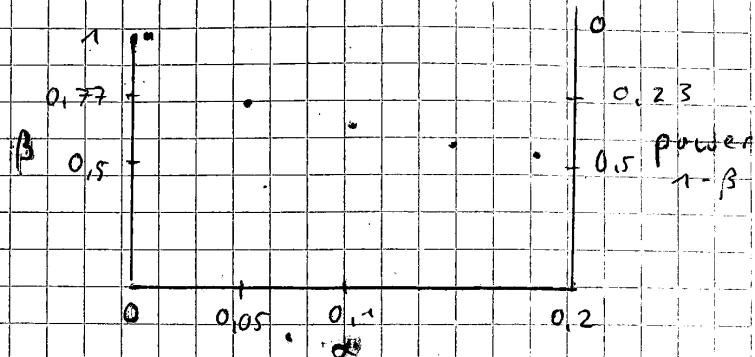
- The probability of a Type II error is represented by β ; it depends on the choice of α , the sample size and the true value of the population parameter.

The power of a test is the probability of rejecting the null hypothesis, given it is false.

$$\text{Power} = 1 - P(\text{Type II error}) = 1 - \beta$$

Why not choose $\alpha = 0.0000001$ to avoid Type I error?

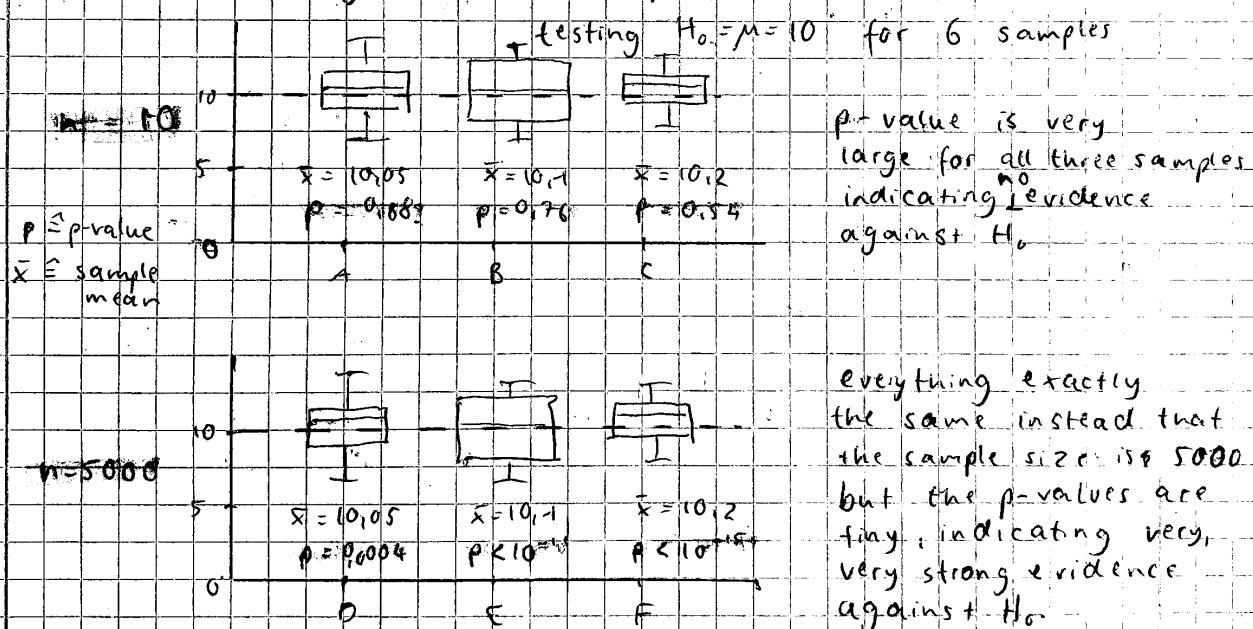
It is because of the relationship between α and β . If we decrease alpha, then beta will increase and thus our test loses power, and type II errors will be common.



Hypothesis testing tests for statistical significance.

Statistical significance means the effect observed in the sample was unlikely to have occurred due to chance alone.

Statistical Significance vs. practical significance



That much larger sample size has found those differences to be highly statistically significant.

But whether these differences from the hypothesized value have any practical meaning is a different question.

A difference from the sample mean to the hypothesized value of 0.1 may result in a more effective drug or better marketing campaign or maybe in the context of a problem the difference is just meaningless.

The relationship between Confidence Intervals and Hypothesis Tests

Suppose we find a $(1-\alpha)$ 100% confidence interval for μ using:

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

and we wish to carry out a test "of"

$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0$$

with a significance level of α

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

We would typically go ahead and use our Z test statistic, come up with a p-value and draw a conclusion.

But you can tell what that conclusion would be by looking at the confidence interval because there is a relationship between confidence intervals and hypothesis tests.

The confidence interval will be made up of all values of μ_0 for which we would not reject the null hypothesis.

Example: $\alpha = 0.1$

Suppose a 90% confidence interval for μ is found to be $(13.1, 26.4)$. The hypothesized value of 25 falls within the 90% interval.

What would be the results of a test of

$$H_0: \mu = 25 \text{ against } H_a: \mu \neq 25 \text{ at } \alpha = 0.1?$$

→ The null hypothesis would not be rejected at a significance level of 0.1

The p-value of the test would be greater than 0.1.

Why is this the case?

We would reject H_0 at $\alpha = 0.05$ if

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq -1.96 \quad \text{or} \quad \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq 1.96$$

Solving μ_0 we would reject H_0 if

$$\mu_0 \geq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \mu_0 \leq \bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}$$

Upper bound of the
95% confidence interval

lower bound of the
95% confidence interval

Calculating Power and the Probability of a Type II Error

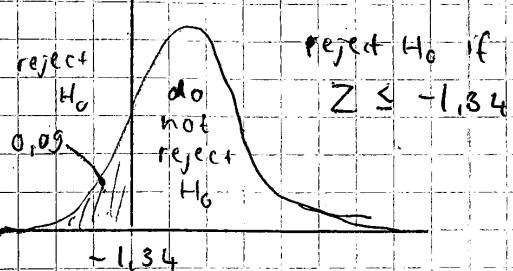
Suppose we are about to randomly sample 36 values from a normally distributed population, where $\sigma = 21$ but μ is unknown. We are going to test:

$$H_0: \mu = 50$$

$$H_a: \mu < 50$$

at $\alpha = 0.09$

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

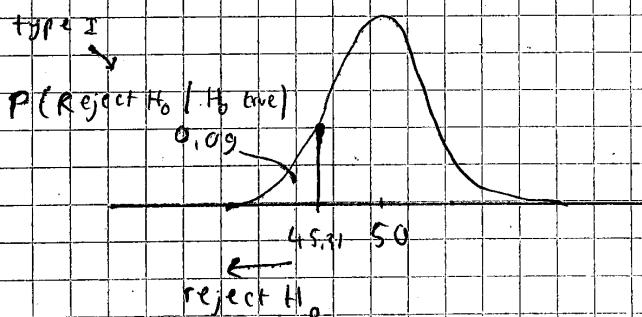


For what values of \bar{X} will we reject H_0 ?

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}, \quad \sigma = 21, \quad n = 36$$

$$\begin{aligned} \bar{X} &= \mu_0 + \frac{\sigma}{\sqrt{n}} \cdot Z \\ &= 50 + \frac{21}{6} \cdot (-1.34) = 45.31 \Rightarrow \text{reject } H_0 \text{ if } \end{aligned}$$

distribution of \bar{X} if $\mu = 50 \quad \bar{X} \leq 45.31$



If the true value of the population mean $\mu = 43$, what is $P(\text{Type II error})$?

$$\begin{aligned} P(\text{Do not reject } H_0 \mid \mu = 43) &= P(\bar{X} > 45.31 \mid \mu = 43) \quad \text{reject } H_0 \text{ if } \\ &\quad \bar{X} \leq 45.31 \end{aligned}$$



$$P(\bar{X} > 45.31 \mid \mu = 43)$$

$$= P\left(Z > \frac{45.31 - 43}{21/\sqrt{36}}\right) = P(Z > 0.60)$$

$$= 0.2785 = \beta$$

Probability of Type II error
(of not rejecting hypothesis when it is wrong)

Two-sample tests for difference of means

Common points of interest:

- testing if there is a significant difference between the groups.
(Is there strong evidence that $\mu_1 \neq \mu_2$?)
- estimating $\mu_1 - \mu_2$ with a confidence interval

Let \bar{X}_1 represent the mean of n_1 independent observations from a normally distributed population with mean μ_1 and variance σ_1^2 .

Let \bar{X}_2 represent the mean of n_2 independent observations from a normally distributed population with mean μ_2 and variance σ_2^2 .

$$\bar{X}_1 \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$$

$$\bar{X}_2 \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$$

What is the mean of $\bar{X}_1 - \bar{X}_2$?

$$\begin{aligned} E(\bar{X}_1 - \bar{X}_2) &= E(\bar{X}_1) - E(\bar{X}_2) \\ &= \mu_1 - \mu_2 \end{aligned}$$

The difference in the sample means is an unbiased estimator of the difference in the population means.

What is the variance of $\bar{X}_1 - \bar{X}_2$?

$$\begin{aligned} \text{Var}(\bar{X}_1 - \bar{X}_2) &\quad (\text{if we have independent samples}) \\ &= \text{Var}(\bar{X}_1) + \text{Var}(\bar{X}_2) \\ &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \end{aligned}$$

What is the distribution of $\bar{X}_1 - \bar{X}_2$?

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

Suppose that for Canadians between 20-39 years of age:

Male heights are approximately normally distributed with $\mu_M = 177.7 \text{ cm}$ and $\sigma_M = 5.6 \text{ cm}$ $N(177.7, 5.6^2)$

Female heights are approximately normally distributed with $\mu_F = 163.0 \text{ cm}$ and $\sigma_F = 5.1 \text{ cm}$ $N(163.0, 5.1^2)$

If 20 males and 15 females in this age group are randomly selected, what is the sampling distribution of $\bar{X}_M - \bar{X}_F$?

$$\bar{X}_M \sim N(177.7, \frac{5.6^2}{20}), \quad \bar{X}_F \sim N(163.0, \frac{5.1^2}{15})$$

$$\bar{X}_M - \bar{X}_F \sim N(177.7 - 163.0, \frac{5.6^2}{20} + \frac{5.1^2}{15})$$

$$\rightarrow \bar{X}_M - \bar{X}_F \sim N(14.7, 3.302)$$

What is the probability the average height of the males is at least 10 cm greater than the average height of the females?

$$\bar{X}_M - \bar{X}_F \sim N(14.7, 3.302)$$

$$P(\bar{X}_M \geq \bar{X}_F + 10)$$

$$= P(\bar{X}_M - \bar{X}_F \geq 10) \rightarrow \text{we standardize}$$

$$= P\left(\frac{\bar{X}_M - \bar{X}_F - 14.7}{\sqrt{3.302}} \geq \frac{10 - 14.7}{\sqrt{3.302}}\right)$$

random variable
with non-standard
normal distribution

$$= P(Z \geq -2.586) = 0.995$$

Pooled-Variance t-Tests:

- assumptions:
- independent simple random samples
 - normally distributed populations
 - equal population variances ($\sigma_1^2 = \sigma_2^2$)
(\rightarrow if not available: Welch t-test) $\rightarrow s_p^2$

Step 1: Estimate the common population variance σ^2 with the pooled sample variance:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

estimates σ^2 - weighted average of the two sample variances

Step 2: Calculate the standard error of the difference in sample means

$$SE(\bar{X}_1 - \bar{X}_2) = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

estimates the true standard deviation of the sampling distribution of the difference in sample means

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

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2} \cdot SE(\bar{X}_1 - \bar{X}_2)$$

Degrees of freedom = $n_1 + n_2 - 2$

We often want to test: $H_0: \mu_1 = \mu_2$

against one of the 3 possible alternatives:

- $H_a: \mu_1 > \mu_2$
- $H_a: \mu_1 < \mu_2$
- $H_a: \mu_1 \neq \mu_2$

$$H_0: \mu_1 = \mu_2 \equiv H_0: \mu_1 - \mu_2 = 0$$

estimator $\bar{x}_1 - \bar{x}_2$ - hypothesized value

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{SE(\bar{x}_1 - \bar{x}_2)}$$

If H_0 is true this t statistic has a t distribution with $n_1 + n_2 - 2$ degrees of freedom

example: Inference for two means

In 2001 in Croatia, there was a large increase in the proportion of veterans claiming PTSD

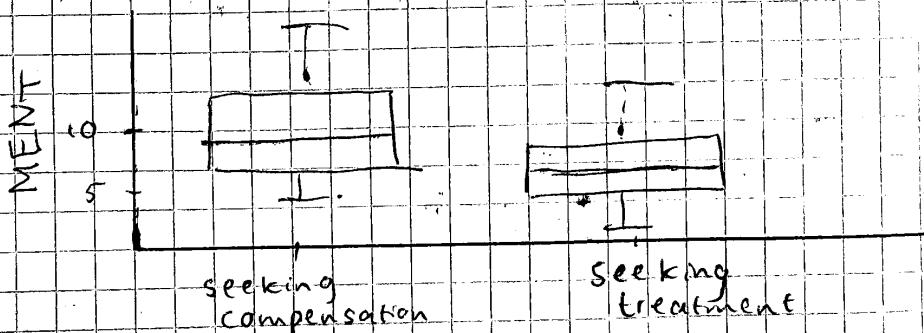
Studies showed that high proportion were malingering (exaggerated symptoms)

The Morel Emotional Numbing Test was designed as a diagnostic tool to help differentiate between those who truly had PTSD and those who were malingering.

μ_1 - true mean of those seeking compensation

μ_2 - true mean of those seeking treatment

scores on the MENT for 49 veterans seeking compensation and 70 veterans seeking treatment:



→ independent sample random samples
equal population variances

Compensation treatment	$\bar{X}_1 = 9,76$	$s_1 = 4,9$	$n_1 = 49$
	$\bar{X}_2 = 6,48$	$s_2 = 3,49$	$n_2 = 70$

Step 1:

$$\text{pooled variance: } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(49 - 1)4,9^2 + (70 - 1)3,49^2}{49 + 70 - 2}$$

$$= 17,033 \rightarrow \text{estimates the common population variance}$$

$$\text{Step 2: } SE(\bar{X}_1 - \bar{X}_2) = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= \sqrt{17,033} \cdot \sqrt{\frac{1}{49} + \frac{1}{70}}$$

$$= 0,7687$$

Suppose we want a 95% confidence interval for the difference of two population means

$$\rightarrow \alpha = 0,05$$

$$DF = 117$$

A 95% confidence interval for $\mu_1 - \mu_2$:

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2} \times SE(\bar{X}_1 - \bar{X}_2)$$

$$t_{0,025} = 1,98$$

$$9,76 - 6,48 \pm 1,98 \cdot 0,7687$$

$$3,28 \pm 1,522$$

$$\Rightarrow (1,758, 4,802) \text{ (indicates that those seeking compensation score higher.)}$$

Test the null hypothesis that the populations have equal mean scores on the MENT.

$$H_0: \mu_1 = \mu_2 \quad H_a: \mu_1 \neq \mu_2$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)} = \frac{9,76 - 6,48}{0,7687} = 4,267$$

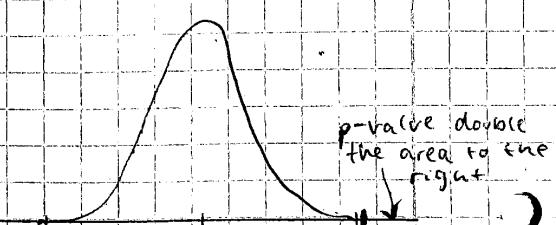
t distribution with 117 DF

$$DF = 117 \rightarrow p\text{-value} = 0,00004$$

\rightarrow There is very strong evidence of a difference in population mean scores on the MENT.

\rightarrow A 95% confidence interval for $\mu_1 - \mu_2$ is $(1,8, 4,8)$.

\rightarrow We can be 95% confident that, on average, those seeking compensation score between 1,8 and 4,8 units higher on the MENT than those seeking treatment.



Welch (Unpooled Variance) t tests for 2 means

The Welch t procedures are similar to the pooled-variance t procedures, but do not require the assumption of equal population variances, so that it is more applicable for a wider variety of situations.

The Welch procedure is an approximate (not exact) procedure

Step 1: Calculate standard error of the difference in sample means:

$$SE_w(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

estimates the standard deviation of the sampling distribution of the difference in sample means

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

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2} \cdot SE_w(\bar{X}_1 - \bar{X}_2)$$

DF = ? → Welch-Satterthwaite approximation

example Does a certain rattlesnake antivenom given after a rattlesnake bite have an effect on swelling?

Researchers ran an experiment to investigate this, randomly assigning pigs to one of two groups:

1. A treatment group, which received an antivenom $n_1 = 9$

2. A placebo group, which received a saline solution $n_2 = 8$

The pigs were then injected with rattlesnake venom, and either antivenom or placebo

Change in leg volume (mL) after 8 hours were measured

$\mu_1 - \mu_2$ is also called the treatment effect

Antivenom	$\bar{X}_1 = 203.33$	$s_1 = 56.18$	$n_1 = 9$
Saline	$\bar{X}_2 = 201.25$	$s_2 = 77.62$	$n_2 = 8$

$$SE_w(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= \sqrt{\frac{56,18^2}{9} + \frac{12,62^2}{8}}$$

$$= 44$$

A $\approx 95\%$ confidence interval for $\mu_1 - \mu_2$:

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2} \cdot SE_w(\bar{X}_1 - \bar{X}_2) \quad (\alpha = 0,05)$$

$$203,33 - 201,25 \pm 2,228 \cdot 44$$

$$2,08 \pm 98,03 \rightarrow (-95,95, 100,01)$$

↓

Hypothesis test

Populations have equal mean
Change in leg volume

○ is contained
so it's plausible that

$$\mu_1 = \mu_2$$

$$H_0: \mu_1 = \mu_2 \quad H_a: \mu_1 \neq \mu_2$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{SE_w(\bar{X}_1 - \bar{X}_2)} = 0,047 \rightarrow p\text{-value is double}$$

the area under the t-distribution to the right of 0,047

$$p\text{-value} = 0,96$$

→ There is no evidence of a difference in the population means of the groups

There is no evidence of a treatment effect.

Paired-Difference Procedures

Example: In an experiment investigating reaction times in 6 volunteers

participant	sober group		alcohol group		→ methods based on independent samples are appropriate here		
	Josie	Jen	Linda	Deb	Nancy	Jul	
time (msec)	332	366	401	384	334	442	

in a different experiment:

participant	Mary	Helen	Sylvia	Amy
sober	363	349	416	338
after drinking	386	367	430	358

not independent → paired-difference procedure to analyse data

↳ Why set up an experiment this way?

- taking two measurements on the same individual can serve to reduce the variability