Introduction to inference

PSLS chapters 14 and 15

Conditions for parametric inference on a mean

Assumptions:

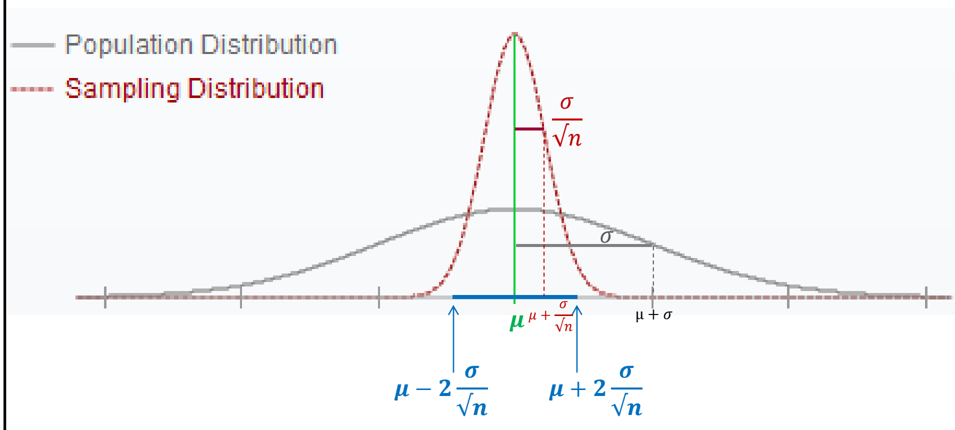
1. The data used for the estimate are a **random sample** (or unbiased representative sample in a randomized experiment) from the population of interest.

2. The population is much larger than the sample (at least 20 times).

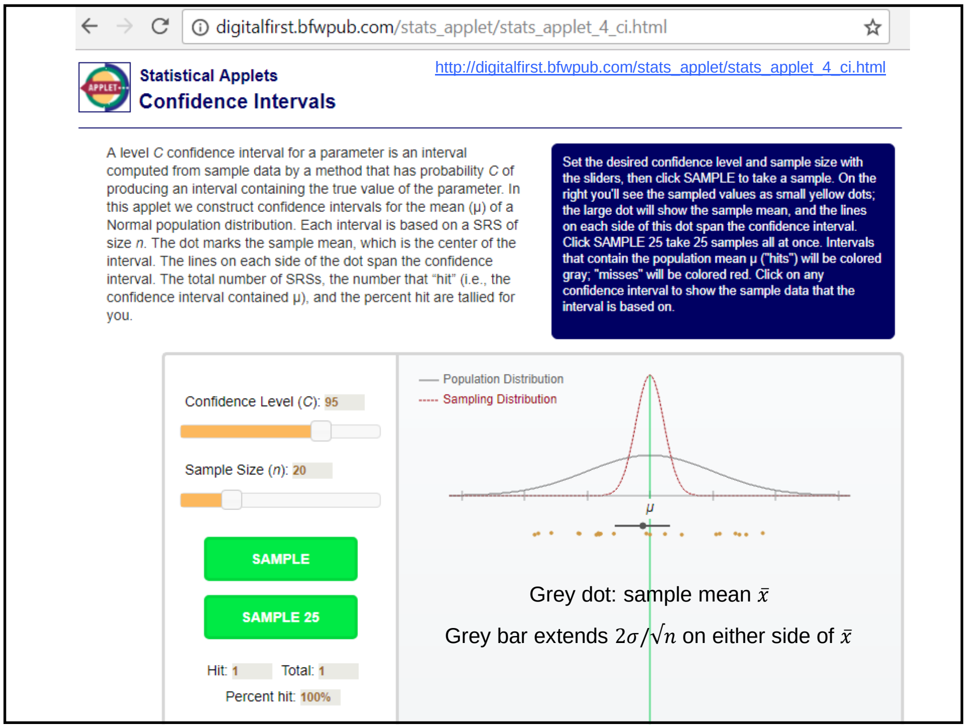
3. The population is Normal **<u>or</u>** we can rely on the central limit theorem for an approximately Normal sampling distribution.

Estimating the unknown value of a population parameter

- We would like to find out the value of a **population parameter** but we can't collect data on the entire population.
- We collect sample data and use the **sample statistic** as our best guess for the value of the parameter, along with an assessment of the **uncertainty** around this guess.



For ~95% of all random samples,
$\bar{x}$ falls in this blue interval between $\mu - 2\frac{\sigma}{\sqrt{n}}$ and $\mu + 2\frac{\sigma}{\sqrt{n}}$
and $2\frac{\sigma}{\sqrt{n}}$ is the maximum distance between $\bar{x}$ and μ.



Confidence intervals

A **confidence interval ("CI"):**
an interval, calculated around a sample statistic, which should contain the unknown value of a population parameter, with some confidence level C

A **confidence level C:**
the overall success rate of the method used to obtain the CI
(based on the probability that the CI captures the true parameter value in repeated samples)

https://pubs.usgs.gov/sir/2018/5076/sir20185076.pdf

A 95% confidence interval for the mean copper concentration in sediment cores at the shoreline of an urban growth area in western Washington is reported as (4.5, 5.6) mg/kg.

What is the correct interpretation of this interval?

A. We know that 95% of copper concentrations in this sample of sediment cores are values between 4.5 and 5.6 mg/kg.

B. We are confident that 95% of copper concentrations in all possible sediment cores at this shoreline are values between 4.5 and 5.6 mg/kg.

C. We are 95% confident that the mean copper concentration in all possible sediment cores at this shoreline is a value between 4.5 and 5.6 mg/kg.

D. We are 95% confident that the mean copper concentration of any sample of sediment cores from this shoreline is a value between 4.5 and 5.6 mg/kg.

CI for a population mean (σ known)

Level C confidence interval for μ:

$$\boxed{\bar{x} \pm z^* \sigma/\sqrt{n}} \quad \text{or} \quad \bar{x} \pm m$$

C is the area under the N(0,1) between $-z^*$ and z^*

The margin of error, *m*, is the maximum distance between the statistic and the parameter in C% of all random samples of size *n* that could be taken from that population.

A 95% confidence interval for the mean copper concentration in sediment cores at the shoreline of an urban growth area in western Washington is reported as (4.5, 5.6) mg/kg.

The value of the margin of error *m* is

A) 0.275 B) 0.55 C) 1.1 D) 3.025 E) 6.05 mg/kg

A 90% confidence interval on the same data would be:

A. wider **B.** narrower **C.** exactly the same **D.** possibly wider or narrower

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

Confidence level C	90%	95%	99%
Critical value z^*	1.645	1.960	2.576

Confidence intervals in practice

The margin of error does not cover all errors: The margin of error in a confidence interval covers only ***random sampling*** error.

Undercoverage, nonresponse, or other forms of bias are often more serious than random sampling error. The margin of error does not take these into account at all.

Commercial surveys typically have ~ 90% nonresponse!
People aren't always truthful with self-reported answers.
…

Beware of reports that cite the sample statistic and make it sound as if it were the actual population parameter (barely mentioning a margin of error, if even!).

Testing a hypothesis about the unknown value of a parameter

- We make a rational claim (reflecting a legitimate inquiry) about the unknown value of a **population parameter**.

- We evaluate the strength of the evidence (**sample data**) against this claim, to help us decide whether or not to reject it.

Hypothesis tests

A null hypothesis:
a hypothesized model of the population distribution, with a specific parameter value

A test statistic:
a measure of how different the sample statistic and the hypothesized value of the parameter are ("the observed effect size"), relative to the expected variability of the statistic

A test *P*-value:
the probability of obtaining a test statistic at least as extreme as that computed, if the null hypothesis was true

related measures

Null and alternative hypotheses

The **null hypothesis,** H_0: [Typically a statement of "no effect."]
A specific claim about the unknown value of a population parameter.

We assess the strength of the evidence *against* the null hypothesis.

The **alternative hypothesis,** H_a:
A more general claim about the unknown value of the parameter based on theory or a legitimate question [not based on the sample data!].

We reject H_0 when the evidence *supports* the alternative hypothesis.

The direction of H_a should reflect the study's objective

H_a: $\mu \neq$ a specific value μ_0 **two-tailed (two-sided) H_a**

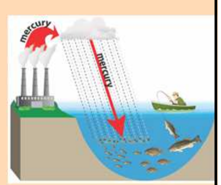
H_a: $\mu <$ a specific value μ_0 **one-tailed (one-sided) H_a**

H_a: $\mu >$ a specific value μ_0 **one-tailed (one-sided) H_a**

Do children's Tylenol bottles contain the amount stated on the label (120 ml)?

Start by identifying:
-Individuals = Tylenol bottles
-Variable = medication amount (quantitative)
-Parameter of interest = mean amount for entire production
-Objective = no deviation from target amount, on average
And label any value cited : 120 ml = target amount

Does the concentration of mercury in fish found in the stream near a factory exceed the FDA action level of 1 part per million?



Start by identifying:
-Individuals
-Variable
-Parameter of interest
-Objective
And label any value cited

IQ test scores follow a Normal distribution with standard deviation σ = 15. In the general population, the mean IQ score is 100.

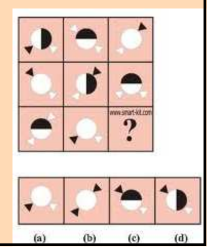
A school district wants to know if the IQ scores of its high-schoolers (HS) differ from that of the general population, on average. The superintendent selects a random sample of 40 HS and finds that their mean IQ score is 104.2.

Define the corresponding null and alternative hypotheses.

-Individuals
-Variable
-Parameter of interest
-Objective
-*Values cited*

The appropriate alternative hypothesis is

A) $H_a\colon \bar{x}_{districtHS} > 100$
B) $H_a\colon \bar{x}_{districtHS} = 104.2$
C) $H_a\colon \mu_{districtHS} > 100$
D) $H_a\colon \mu_{districtHS} = 104.2$
E) $H_a\colon \mu_{districtHS} \neq 100$

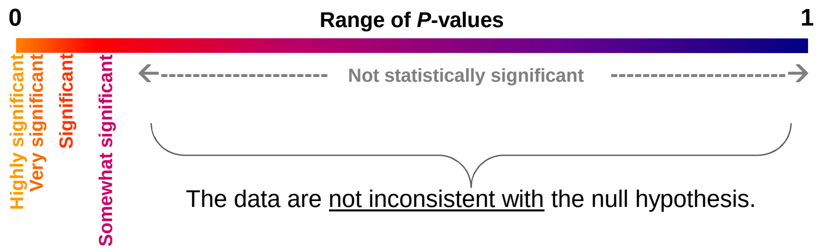


P-value

***P*-value:** **The probability, computed assuming that H_0 is true, that the test statistic would take a value at least as extreme (in the direction of H_a) as that actually observed.**

- *P*-values that are <u>not small</u> don't give enough evidence against H_0 and we **fail to reject H_0**. The data are consistent with H_0, so H_0 could be true. But **we can never "prove H_0."**
- <u>Small *P*-values</u> are strong evidence AGAINST H_0 and we reject H_0. The findings are **"statistically significant."**

The *P*-value of a test evaluates the strength of the evidence **against** the null hypothesis (smaller *P*-values provide more evidence against H_0).



The **significance level, *α* :** a <u>chosen</u> cut-off value such that:

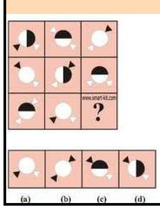
P*-value ≤ *α, we **reject H_0**. ***P*-value > *α***, we **fail to reject H_0**.

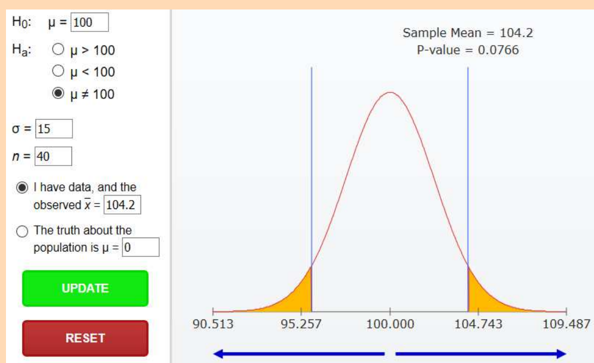
USE α SPARINGLY !!!

IQ test scores follow a Normal distribution with standard deviation σ = 15. In the general population, the mean IQ score is 100.

A school district wants to know if the IQ scores of its High-Schoolers differ from that of the general population, on average. The superintendent selects a random sample of 40 HS and finds that their mean IQ score is 104.2. What can we conclude?

The data failed to provide evidence (*P* = 0.08) that the mean IQ score of HS in this school district differs from 100, the mean IQ in the general population.





People typically think of a healthy body temperature as 98.6 ºF, a value that was cited in a study published in 1868. More than a century later, a study of a random sample130 healthy adults found a mean body temperature of 98.25 ºF, which is significantly different from the historical value (P = 3E-11). What can we conclude?

A) There is very strong, significant evidence ($P \approx 0$) that the mean body temperature of healthy adults is not 98.6 ºF.

B) There is very strong, significant evidence ($P \approx 0$) that the mean body temperature of healthy adults is 98.6 ºF.

C) There is significant evidence that the mean body temperature of healthy adults is very different from 98.6 ºF ($P \approx 0$).

D) There is significant evidence that the mean body temperature of healthy adults is not different from 98.6 ºF ($P \approx 0$).

E) The new study failed to find significant evidence that the mean body temperature of healthy adults differs from 98.6 ºF ($P \approx 0$).

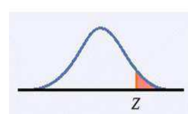
Test for a population mean (σ known)

To test H_0: $\mu = \mu_0$ using a random sample of size n from a normal population with known standard deviation σ, we use the null sampling distribution $N(\mu_0, \sigma\sqrt{n})$.

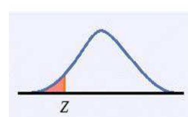
The ***P*-value** is the area under $N(\mu_0, \sigma\sqrt{n})$ for values of $\bar{x}$ at least as extreme in the direction of H_a as that of our random sample.

$$\boxed{z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}}$$

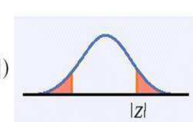
H_a: $\mu > \mu_0$ is $P(Z \geq z)$



H_a: $\mu < \mu_0$ is $P(Z \leq z)$



H_a: $\mu \neq \mu_0$ is $2P(Z \geq |z|)$



Hypothesis tests in practice

Statistical significance only says whether the effect observed is likely to be due to chance alone (random sampling) if H_0 was true.

- Statistical significance doesn't tell about the **magnitude** of the effect.
- Statistical significance may not be practically important.
- Studies based on very large sample sizes can produce results that are statistically significant but substantively trivial.
- Keep in mind that the *P*-value computation assumes that there were no problems with the data collection process.

Read:
ASA Statement on Statistical Significance and *P*-values (2016)
http://dx.doi.org/10.1080/00031305.2016.1154108

1. *P*-values can indicate how incompatible the data are with a specified statistical model.
2. *P*-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a *p*-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A *p*-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a *p*-value does not provide a good measure of evidence regarding a model or hypothesis.

Facebook conducted an experiment in which they varied how many positive and negative feeds 689,003 Facebook users were allowed to see.

"Experimental evidence of ***massive-scale emotional contagion*** *through social networks"* (Kramer et al., 2014, doi:10.1073/pnas.132004011, www.pnas.org/content/111/24/8788.full.pdf)

In an editorial, the study authors acknowledged that, even with their huge sample, they did not find a particularly large effect. The result was that "people produced an average of one fewer emotional word, per thousand words, over the following week."

> "[The work] was consistent with Facebook's Data Use Policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research."

Testing a hypothesis with a CI

When the CI is entirely consistent with H_a, reject H_0 in favor of H_a. Otherwise, fail to reject H_0

A study of a random sample of 130 healthy adults found a mean body temperature of 98.25 ºF, which is significantly different from the historical value of 98.6 ºF (P = 3E-11).

- H_0: $\mu = 98.6$ versus H_a: $\mu \neq 98.6$ ºF
- A 95% confidence interval for μ is (98.15, 98.35) ºF

We are 95% confident that the mean body temperature of all healthy adults is a value lower than 98.6 ºF. The mean body temperature of all healthy adults is *significantly lower* than 98.6 ºF.

Why are type I and type II errors relevant for statistical inference hypothesis tests?

A) Because it is important to keep in mind that conclusions based on a *P*-value are sometimes the wrong conclusion.

B) Because the computations required for a P-value are complex and therefore prone to errors.

C) Because it is important to keep in mind that conclusions based on observed data do not necessarily imply causation.

D) Because it is important to keep in mind that sample data is inherently biased.

E) Because sometimes there are outliers in the sample data.

		Truth about the population	
		H_0 true	H_a true
Decision based on sample	Reject H_0	Type I error (probability, α)	Correct decision
	Fail to reject H_0	Correct decision	Type II error (probability, β)