

## Statistische Signifikanz vs. Relevanz

- Statistische Tests werden in der Praxis oft «missbraucht» und falsch angewendet → schlechter Ruf der Statistik.
- Das Problem ist: Je grösser unsere Stichprobe ist, desto eher werden wir signifikante Effekte finden, denn die Nullhypothese stimmt in der Regel nie **exakt**. (Zur Erinnerung: Je grösser die Stichprobe, desto grösser die Macht.)
- Wenn wir z.B.  $H_0: \mu = 400$  testen und in Tat und Wahrheit gilt aber  $\mu = 401$ , so werden wir bei genügend grosser Stichprobe  $n$  mit hoher Wahrscheinlichkeit ein signifikantes Testresultat erhalten.
- Ob etwas signifikant ist, ist also unter anderem eine Frage des **Aufwands** (\$).

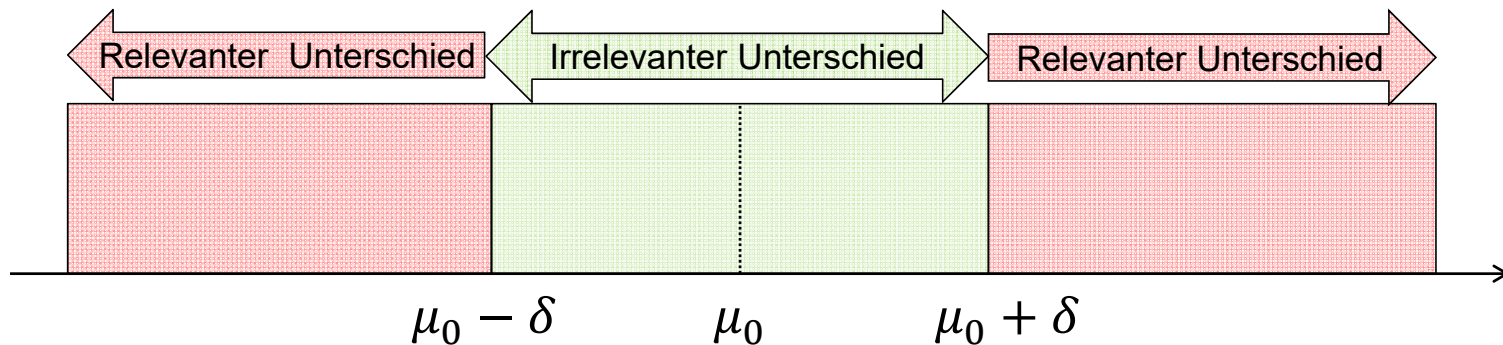
## Statistische Signifikanz vs. Relevanz

- Die (wichtigere) Frage ist: Wann haben wir ein **relevantes** Resultat?
- Wir müssen vorher definieren, was «Relevanz» bedeutet.
- Was ein **relevanter Unterschied** ist, hängt ab vom Fachgebiet/Fachwissen. Die Statistik hat hier **keine Antwort!**
- Bsp: Durchmesser von Zylinderscheiben:

Mit was für Abweichungen vom Sollwert kann man leben?

## Statistische Signifikanz vs. Relevanz

- Wir müssen also eine **Differenz**  $\delta$  angeben, ab der man sagt, dass ein Unterschied **relevant** ist für eine entsprechende Anwendung.

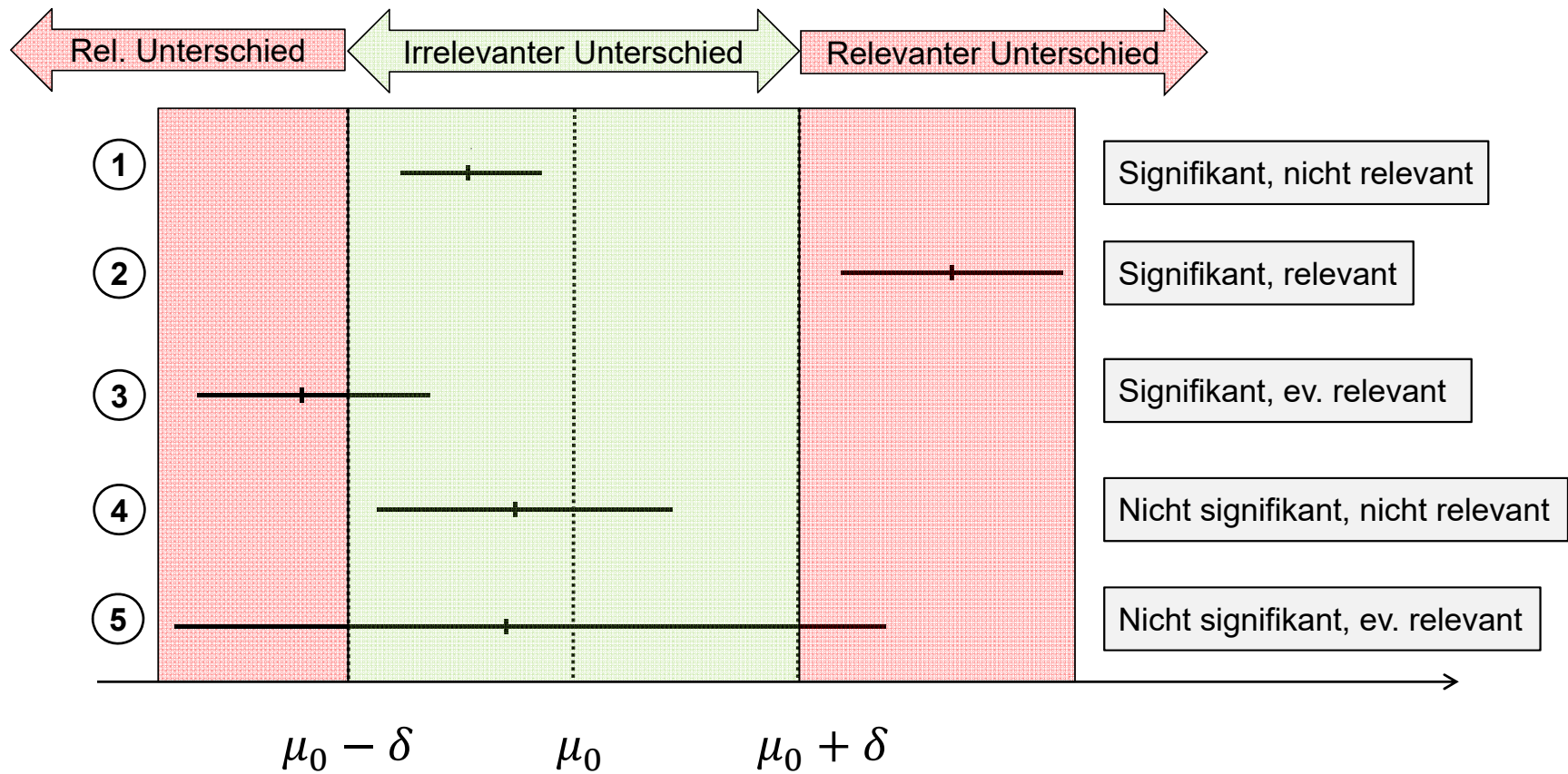


- Basierend auf unseren Daten berechnen wir dann ein Vertrauensintervall für den Parameter von Interesse.
- Die Idee besteht nun darin, dass man schaut, wo das **Vertrauensintervall** bezüglich der obigen Bereiche liegt.

## Statistische Signifikanz vs. Relevanz

- Liegt das Vertrauensintervall ganz im «relevanten Bereich», so spricht man von einem relevanten Effekt.
- Ist zwar der Test signifikant (d.h. KI enthält  $\mu_0$  nicht), aber das KI liegt ganz im «irrelevanten Bereich», so hat man zwar ein signifikantes, aber **kein** relevantes Resultat.
- Siehe auch Beispiel nächste Folie.

# Statistische Signifikanz vs. Relevanz



## Statistische Signifikanz vs. Relevanz

- Man kombiniert also «das Beste aus beiden Welten»:  
Das Fachwissen und die Statistik, die einem hilft, die Unsicherheit zu quantifizieren (durch das KI).
- Es reicht in der Regel also nicht, sich nur «blind» auf die statistische Signifikanz zu verlassen (obwohl dies vielerorts so gemacht wird).
- Wir müssen uns zusätzlich auch immer fragen: «Ist das auch ein **relevantes** Resultat?».

## Beispiel Zylinderscheiben

- Sollwert  $\mu_0 = 20.2$
- Vertrauensintervall für  $\mu$  war:  $KI = [20.32, 20.88]$
- Die Abweichung von  $\mu$  zum Sollwert liegt also im Bereich  $[0.12, 0.68]$
- Falls wir eine Abweichung von 0.3 mm als relevant einstufen, so müssen wir hier **potentiell beunruhigt** sein.