Ordlista för biostatistik

**Variabel:** En mätbar parameter som kan variera.

**Population:** Mängden av alla individer av intresse.

* **Målpopulation:** De individer som man vill få reda på något om.
* **Studiepopulation:** De individer som undersöks.

**’Statistic’:** Ett värde som beräknas utifrån ett ‘sample', exempelvis ett medelvärde eller standardavvikelse.

**Inferens:** Fakta om ett ‘sample' överförs till hela populationen.

**Sample:** Ett urval av populationen, en delmängd av populationen.

* **Random sample:** Ett urval av populationen som skett helt slumpmässigt.
* **Convenience sample:** Ett urval som skett pga det passar förutsättningarna. **Ex**.:
	+ **Systematic sampling:** Urval utifrån en viss princip/regel. **Ex**.: Alla födda ett udda datum.
	+ **Cluster sampling:** Urval utifrån gemensam faktor. **Ex**.: Alla individer i ett geografiskt område. Alla på ett visst sjukhus.

**Sampling frame:** Grupp/referens ur vilken urvalet sker. **Ex**.: databas, register.

**Sampling fraction:** Den andel av hela populationen/substratet som urvalet görs ur. **Ex**.: Urval från alla läkarstudenter i Sverige. Leverbiopsi av en enskild lob.

**Uppskattning (estimate):** Beräknat värde från ett ‘sample’ som antas stämma överens med hela populationen, dvs. $\overline{X}$ ≈ µ eller s ≈ σ (se nedan). Kan uttryckas som med “hatt” över symbolen (t.ex. ŝ).

**Bias:** Systematiskt fel

* **Selection bias:** Systematiskt fel i urvalet (sampling). Kan inte korrigeras i efterhand.
* **Measurement bias:** Systematiskt fel i inhämtning av information/data. **Ex**.: Olika personer mäter i varsina grupper (läkare/ssk.).
* **Migration bias:** Bortfall (drop-outs) av en specifik grupp. **Ex**.: Personer med narkotikamissbruk tenderar att flytta till större städer vilket medför lägre prevalens i småstäder.

**Confounding:** En okänd faktor med koppling till både oberoende och beroende variabel. **Ex**.: De som äter mycket grönsaker (oberoende variabel) tenderar att även motionera (confounder) mer, båda kan påverka kardiovaskulär risk (beroende variabel).

**Genomsnitt (mean):** Summan delat på antal individer.

**Spridning:** Beskrivning av avvikelser från genomsnittet.

**Percentil:** Det värde som motsvarande procentsats av alla värden kommer vara under. **Ex**.: 20e percentilen är det värde som 20% av datan (övriga värden) ligger under och 80% ligger över.

**Quartil:** 25:e, 50:e och 75:e percentilerna (1:a, 2:a respektive 3:e quartilen).

**Median:** Värdet hos den mittersta (50:e) percentilen.

**Mode:** Det vanligaste värdet.

**Inter-quartile range:** Intervallet av värden inom de två mellersta kvartilerna, dvs mellan 25:e och 75:e percentilen eller 1:a och 3:e quartilen.

**Epidemi:** Sjukdom begränsad i tid. Relativt stor prevalensökning (incidens) under relativt kort tid.

**Endemi:** Sjukdom begränsad till plats. Relativt hög prevalens inom ett avgränsat geografiskt område.

**Intervalldata:** Data med konstant intervall mellan varje punkt.

* **Kontinuerliga data:** Data som kan anta alla värden i ett spektrum. Kan anges med decimaler. **Ex**.: Vikt, längd.
* **Diskreta data:** Data som endast kan anges med heltal. **Ex**.: Antal födslar.

**Nominala data:** Diskreta data utan given ordning. **Ex**.: Blodgrupper A, AB, B, O.

* **Dikotoma data:** Nominala data med två möjligheter. **Ex**.: Levande/död.

**Ordinala data:** Diskreta data med given ordning. **Ex**.: Antal födslar. Grad av blåsljud.

**Range:** Intervall från det minsta till det största värdet.

**Normalfördelning:** Symmetrisk fördelning av värden med ett särskilt förhållande mellan medelvärde, median, mode och spridning:

* Medelvärde (mean) = Median = Mode → Medelvärdet är i mitten
* Spridning: “Midja” vid +/- 1 Standardavvikelse
* Ingen “skevhet” (skew)

**µ:** Medelvärde (mean) i en hel population

$\overline{X}$**:** Medelvärde (mean) i ett ”sample”

**Avvikelse:** Skillnad från medelvärdet.

**Standardavvikelse (SD):** Mått på spridning i en population.

**σ:** Standardavvikelse i en hel population

***s*:** Standardavvikelse i ett ”sample”

**Z-score:** Antal standardavvikelser som ett värde ligger inom.

**Standard error (SE):** Beräkning av hur mycket samples skiljer sig från varandra vid upprepade samples. Ger en uppskattning av hur nära medelvärdet av ett sample är hela populationens medelvärde. Beror på populationens SD och storleken på urvalen (samples) enligt ekvationen:
$SE=SD÷\sqrt{n}$ ,
där *n* är antalet individer i urvalen (samples).

**Sampling error:** Imperfektion i ett ‘sample’, urvalet återspeglar inte fullt populationen.

**Sampling variation:** Skillnaden i medelvärde mellan olika ‘samples’.

**Sampling distribution:** Fördelningen av urvalens medelvärden.

**Central limit theorem:** Om man tar ett stort antal ‘samples’ från vilken population som helst kommer medelvärdena av dessa ‘samples’ vara nästan normalfördelade. Ju fler ‘samples’ desto mer normalfördelade blir medelvärdena.
Detta betyder att ju fler ‘samples’ som tas från en population desto närmare kommer uppskattningar som beräknas från dessa samples att återspegla hela populationen.

**Variance:** Kvadraten av avvikelsen från medelvärdet (σ2). Kvadreras för att ta bort negativa värden vid avvikelse under medelvärdet.

**Regression to the mean:** Med upprepade mätningar minskar antalet extrema värden.

**Biologisk variation:** Variation i tid och mellan individer som beror på biologiska system. **Ex**.: Blodtryck, blodglukos.

**Coefficient of variation (COV):** Förhållandet mellan SD och genomsnittet (mean) angivet i procent, dvs: $(SD÷mean) ×100\%$

**Alternativ hypotes (HA):** Den hypotes som prövas.

* **Ex**.:
* En signifikant skillnad/effekt finns. *(Superiority-design)*
* Specialfall*; Non-inferiority-design* (RCT): Interventionen är minst lika bra som kontroll .

**Nollhypotes (H0):** Motsatsen till HA (det som prövas).

* **Ex**.:
* Ingen effekt/skillnad föreligger. *(Superiority-design)*
* *Non-inferiority-design*: Interventionen är underlägsen.

**Statistisk signifikans:** Det vore mycket osannolikt att resultatet uppkommit om nollhypotesen vore sann. Dvs. om nollhypotesen är sann (ingen skillnad finns) så vore det mycket osannolikt att ändå uppmäta en skillnad. *I princip*: Det är osannolikt att resultatet är inkorrekt.

**P-värde:** Sannolikhet att ett uppmätt samband har uppstått pga slumpen.

**Power:** Förmåga hos en undersökning att påvisa en existerande effekt/skillnad.

**Typ 1-fel:** H0 förkastas trots att den stämmer, dvs man tror sig ha påvisat en effekt/skillnad som inte finns. Kallas även alfa.

**Typ 2-fel:** H0 godtas trots att den är inkorrekt, dvs man har inte lyckats påvisa en effekt/skillnad som faktiskt finns. Kallas även beta.

**Konfidensintervall:** Intervallet som en viss proportion (oftast 95%) av värdena kommer ligga inom då försöket upprepas. *I princip*: Det intervall som man med en viss säkerhet (oftast 95%) kan säga att ett värde ligger inom.

**T-test:** Test för att beräkna p-värde vid jämförelse av två uppsättningar normalfördelade värden.

**Chi2-test:** Metod för att beräkna p-värde vid jämförelse av proportioner, dvs ett sätt att avgöra om en skillnad mellan proportioner (t.ex. procentsatser) är signifikant.

**Fisher’s exact test:**

**Wilcoxon signed-rank test:** Metod för att beräkna p-värde vid jämförelse av medianer.

**Validitet:** Förmåga att mäta det man vill mäta (Överensstämmelse med hur det faktiskt är). **Ex**.: En våg visar ett sanningsenligt värde.

**Extern validitet:** Giltighet av ett visst samband i studiepopulationen för en annan grupp (oftast målpopulation), dvs generaliserbarhet.

**Intern validitet:** Giltighet av ett samband inom studiepopulationen, dvs hur väl resultatet stämmer överens med sanningen för studiepopulationen.

**Reliabilitet:** Förmåga att ge samma resultat vid upprepade mätningar. **Ex**.: En våg visar samma resultat varje gång (behöver inte vara rätt resultat).

**Responsiveness:** Förmåga att ge utslag vid förändrade förhållanden. **Ex**.: En våg som ändrar utslag vid små förändringar i vikt har hög ‘responsiveness’.

**Dosrespons:** Effekten/responsen korrelerar med dosen.

**Prevalens:** Förekomst av något.

* **Point prevalence:** Förekomst vid en viss tidpunkt.
* **Period prevalence:** Förekomst under en viss period.

**Incidens:** Andel icke-drabbade som drabbas under en viss period (oftast angivet som antal /100 000/år).

* **Kumulativ incidens:** Incidens fram till en viss tidpunkt (ofta angivet som antal/personår

**Överlevnad:** Prevalens$÷$Incidens
(eftersom Incidens $×$(antal år med sjukdom) = Prevalens)

**Personår:** Antal personer $×$ antal år de följts.

**Naturalförlopp:** Förloppet om inga åtgärder vidtas.

**Sannolikhet:** Andelen händelser av intresse vid stort antal upprepade observationer. Dvs. Proportionen av alla händelser som beskriver förekomsten av händelserna av intresse.

**Betingad sannolikhet (conditional probability):** Sannolikhet för en händelse givet en annan händelse. **Ex**.: Sannolikhet att uppnå 70 års ålder *givet* att man fyllt 40.

**Sannolikhetsfördelning (Probability distribution):** Alla möjligheter med angivna sannolikheter för var och en. Summan av sannolikheterna blir 1.

**Risk:** Sannolikhet att drabbas av något. Proportionen mellan fall och hela populationen (fall + icke-fall).

**Absolut risk:** Procentsats eller proportion som beskriver sannolikhet att drabbas.

**Relativ risk:** Risk i förhållande till en annan population.

* Risk hos exponerad $÷$ Risk hos icke-exponerad

**Attributable risk:** Risk som kan tillskrivas en särskild faktor (exponering).

* Risk hos exponerad - Risk hos icke-exponerad

**Population-attributable risk:** Attributable risk (risk från exponering) hos exponerad population, dvs. andel av exponerade som drabbas pga exponering (man kan drabbas av andra skäl än exponering).

* Attributable risk $×$ Prevalens av exponering

**Population-attributable fraction:** Andel av de som drabbas (andel av incidens) som kan tillskrivas exponering.

* Population-attributable risk $÷$ Incidens av utfall

**Riskfaktor:** Faktor som ökar risk.

* **Direkta orsaker:** En faktor som orsakar utfallet.
* **Indirekta orsaker:** En faktor som orsakar något som i sin tur eller i flera led orsakar utfallet.
* **Markör:** En faktor som inte orsakar utfallet varken direkt eller indirekt men är associerad med utfallet. **Ex**.: Klamydia är en markör för risk för HPV-infektion.

**Riskstratifiering:** Indelning i olika grupper med olika risk utifrån riskfaktorer.

**Population at risk:** Population som kan drabbas. **Ex.**: Barn drabbas generellt inte av demens. Män drabbas generellt inte av ovarialcancer.

**Odds:** Förhållande mellan antal positiva och negativa fall.

* **Prevalensodds:** Förhållande mellan fall och icke-fall vid en viss tidpunkt.
* **Kumulativa odds:** Förhållande mellan fall och icke-fall under en viss period.
* **Odds ratio:** Förhållande mellan två odds.

**Rate:** Hur ofta något sker.

* **Incidence rate:** Hur ofta insjuknande sker. Oftast räknat per personår.
* **Incidence rate ratio (IRR):** Förhållande mellan hur ofta insjuknande sker i olika grupper.

**Hazard ratio:** Förhållandet mellan antalet händelser i två olika grupper. Kan användas för att redovisa överlevnadsdata.
**Ex**.: Antalet dödsfall grupp 1 $÷$ Antalet dödsfall grupp 2.

**Exponering:** En individ eller population utsätts för något.

**Latens:** Fördröjning mellan exponering och utfall.

**Numbers needed to treat:** Antalet som måste behandlas under en given tid för att *en* händelse ska undvikas. 1$÷$Absolut risk; vid riskreduktion.

**Numbers needed to harm:** Antalet som måste behandlas under en given tid för att *en* händelse ska inträffa. 1$÷$Absolut risk; vid riskökning.

**Sensitivitet:** Förmåga hos ett test att upptäcka sant positiva, dvs påvisa sjukdom som faktiskt finns.
 = Sannolikhet att ett test blir positivt om sjukdom finns.

**Specificitet:** Förmåga hos ett test att upptäcka sant negativa, dvs inte påvisa sjukdom hos friska.

 = Sannolikhet att ett test blir negativt om sjukdom inte finns.

**Positivt prediktivt värde:** Sannolikhet att sjukdom föreligger vid positivt test.

**Negativt prediktivt värde:** Sannolikhet att sjukdom inte föreligger vid negativt test.

**Gold standard:** Metod för att med största säkerhet ställa diagnos eller avgöra prognos. Referens vid beräkning av sensitivitet och specificitet.

**Accuracy:** Andel sanna värden.

**Likelihood ratio:** Förhållandet mellan resultaten hos sjuka och resultaten hos friska. Säger något om användbarheten av ett test.

* *Positiva resultat*: Antal sjuka med positivt resultat$÷$Antal friska med positivt resultat. (Högt värde är bra)
* *Negativt resultat*: Antal sjuka med negativt resultat$÷$Antal friska med negativt resultat. (Lågt värde är bra)

**Parallell testing:** Många tester samtidigt. Ökar sensitivitet.

**Serial testing:** Fortsatta enskilda tester vid positivt resultat. Ökar specificitet.

**Clinical prediction rules:** Regler som används kliniskt för att bedöma prognos och sannolikhet för sjukdom.

**Observationell studie:** Undersökning av förhållanden som är givna på förhand, ingen intervention. **Ex**.: Kohortstudie, Fall-kontrollstudie, Fallserier/fallrapporter.

**Kohort:** Definierad grupp utifrån exponering eller annan gemensam faktor.

**Kohortstudie:** Studie som utgår från en definierad kohort.

* **Prospektiv:** Upplägg vari kohorten definieras innan utfallet skett och studiepopulationen följs framåt i tiden i syfte att studera incidens.
* **Historisk (retrospektiv):** Upplägg vari utfallet redan skett och studiepopulationen studeras bakåt i tiden.

**Longitudinell studie:** Studie som mäter effekt hos varje person vid flera tidpunkter.

**Cross-sectional studie:** Studie som mäter effekt/skillnad mellan flera personer vid en och samma tidpunkt.

**Propensity score:** Poängsystem som används vid observationella studier för att kompensera för ‘confounding factors’ och utgår från vetskapen om olika faktorers inverkan. **Ex**.: Då rökning studeras kan man behöva kompensera för utbildningsnivå.

**Effektmodifiering:** Modifiering av effektstorlek utifrån påverkan från en tredje faktor. **Ex**.: Modifiering för ålder vid jämförelse av cancerrisk hos olika grupper.

**Överlevnadsdata (survival data):** Data om tid till händelse. Oftast är händelsen död men kan **t.ex.** även vara utskrivning och återfall av sjukdom.

**Censoring:** Överlevnadsdata om händelser faller bort för att händelsen kan ha inträffat efter sista mätningen/rapporteringen, dvs vissa data kan inte inhämtas.
Pga detta kan överlevnadsdata inte redovisas som vanliga nominala data.

**Kaplan-Meier plot:** Graf som illustrerar överlevnad (tid till händelse) med andel överlevande på y-axeln och tid på x-axeln.

**Meta-analys:** Metod för att kombinera resultat från individuella studier.

**Systematisk reveiw:** Artikel som sammanställer originalstudier som selekterats enligt förbestämda villkor för att få svar på en specifik fråga.

**Forrest plot:** Diagram som sammanställer och summerar konfidensintervall och signifikans för ett flertal originalstudier.

**Funnel plot:** Diagram som relaterar effektstorlek till precision på uppskattningen. Vanligtvis pyramidformad. Illustrerar publikationsbias (asymmetri → bias).

**Histogram:** Stapeldiagram som redovisar diskreta data eller intervaller.

**Inklusionskriterier:** Kriterier för att en individ ska övervägas att inkluderas i studien.

**Exklusionskriterier:** Kriterier för att inkluderade individer ska uteslutas.

**Klinisk relevans:** Bedömning av huruvida en uppmätt effekt är eftersträvansvärd i vården. En gräns för hur stor en effekt ska vara för att betraktas som kliniskt relevant bör anges innan studien genomförs.

**Randomisering:** Slumpmässig uppdelning av en population för att undvika ‘confounding factors’.

**Blindning/maskering:** Försöksperson, försöksledare eller den som analyserar resultatet (Enkel-, dubbel- eller trippelblindning) är ovetandes om vem som får vilken intervention.

**CONSORT-diagram:** Flödesschema vari inklusion, exklusion, randomisering, intervention, bortfall och utfall redovisas.

**Bortfall:** Deltagare i studie avbryter deltagande. Bortfall efter randomisering kan leda till selektionsbias och confounding.

* **Withdrawals:** Bortfall efter randomisering men innan intervention.
* **Drop-outs:** Bortfall efter intervention, innan uppföljning.

**Intention-to-treat:** Alla individer som genomgått randomisering ingår i analysen, även de som inte fullföljt interventionen, dvs. ‘Withdrawals’ och ‘drop-outs’ inkluderas.

**Per protocol:** Endast de som fullföljt interventionen ingår i analysen, dvs ‘withdrawals’ exkluderas.

**Multipel signifikansprövning:** (Data-dredging) - många olika samband testas utifrån samma data. Sannolikheten för att hitta minst en signifikant effekt ökar för varje samband som testas.

**Composite measure:** Utfallsmått sammansatt av flera enskilda utfall.