# Statistiek
# Hoorcollege 3, 23-11-2015

Voorbeeld tentamenvragen: **<u>Zie slide 7 t/m 10</u>**

## The normal distribution
De normale distributie is waarschijnlijk de belangrijkste van alle waarschijnlijkheidsdistributies. De functie van een normale random variabele wordt gegeven door een formule:
$F(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})}$, $-\infty < x < \infty$

Als je het gemiddelde verplaatst, verplaatst het figuur zich naar rechts of naar links. Als de standaarddeviatie groter of kleiner wordt, wordt ook de figuur groter of kleiner. Voor voorbeelden: **<u>zie slide 12</u>**.

## Standard normal distibution
Een normale distributie met μ=0 en σ=1 wordt de standaard normale distributie genoemd. De formule hiervan is:
$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, $-\infty < x < \infty$

## Standardizing normally distributed random variables
Ga ervan uit dat X normaal gedistribueerd is met gemiddelde μ en standaarddeviatie σ. Dan geldt de volgende formule:
$$Z = \frac{X - \pi}{\sigma}$$
De formule is normaal gedistibueerd met gemiddelde 0 en standaardafwijking 1. Elke normale distributie kan geconverteerd worden tot een standaard normale distributie met behulp van deze methode. Hierdoor worden berekeningen gemakkelijker. Probabilities van Z kunnen gevonden worden in Appendix Table 1: **<u>zie slide 15</u>**. Deze tabel staat ook achter in je boek.
Voor voorbeelden: **<u>zie slide 16 t/m 18</u>**.

## Sampling
Een probleem is vaak dat de populatie erg groot is. Om dit op te lossen kan er een steekproef genomen worden. We kunnen zo conclusies trekken uit de populatieparameters op basis van deze steekproef, die genomen wordt uit de populatie. Om een populatieparameter te schatten kunnen we een steekproef statistic berekenen. Deze wordt beschreven in **<u>slide 21</u>**.
Simple Random Sample: een steekproef geselecteerd op een manier dat elke observatie dezelfde kans heeft om gekozen te worden.
Een steekproef wordt weergegeven als $\bar{x}$. Dit is een random variabele met een probability distribution.
Voor voorbeelden: **<u>zie slide 24 t/m 26</u>**.

Vervolgens kun je deze gegevens afleiden:
- Gemiddelde van $\bar{x}$: E($\bar{x}$) =μ$\bar{x}$ = μ
- Variantie van $\bar{x}$: $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$
- Standaarddeviatie van $\bar{x}$: $\sqrt{\sigma_{\bar{x}}^2} = \frac{\sigma}{\sqrt{n}}$

Voor voorbeelden: **<u>zie slide 28 t/m 30</u>**.

### Two major theorems:

- **Law of Large Numbers:** Hoe groter de steekproef is, hoe meer waarschijnlijk deze is en hoe dichter het bij het populatiegemiddelde zal liggen. Voorbeeld: roulette Wheel bij het casino.
- **Central Limit Theorem**: Wanneer je statistiek uitrekent op basis van (steekproef)gemiddelde, is deze normaal verdeeld, onafhankelijk van de distributie waaruit je de sample trekt.

De estimater van de population proportion of successes p is de sample proportion $\hat{P}$. Dit is het aantal successen X in een steekproef van steekproefgrootte n: $\hat{P} = \frac{\mathrm{X}}{\mathrm{n}}$.
Het aantal successen X is binominaal verdeeld met parameters n en p (ga uit van een onafhankelijke steekproef).
**Zie slide 35** voor de formules van gemiddelde en standaardafwijking.
Voor voorbeeld: **zie slide 39**.

## Estimination

### Accuratie vs preciezie

Je kunt heel precies zijn, maar heel verkeerd schatten. De schatter die gemiddeld in de roos zit met weinig variatie wordt geprefereerd.

Voorwaarden van een steekproef

- **Unbiased**: verwachte waarde is gelijk aan de populatie parameter
- **Consistent**: als steekproef afneemt neemt het verschil ook af
- **Efficient**: een kleine variatie is beter dan een grote variatie

De betrouwbaarheidsinterval betekent dat als het 95% is, de populatie in 95% van de gevallen hier in zit. Bij 99% zit de populatieparameter hier 99% in.