

# Distribucions de variables a la població

---

# Objectius

---

Entendre què és una distribució de valors a la població

Calcular quantils i intervals de referència

Calcular amb la distribució normal

Dissenyar un criteri diagnòstic en funció d'un biomarcador

# Distribució de valors d'una variable

---

En general, una determinada característica (pes, colesterol, pressió sanguínia, etc.) es distribueix amb una certa variabilitat en cada població.

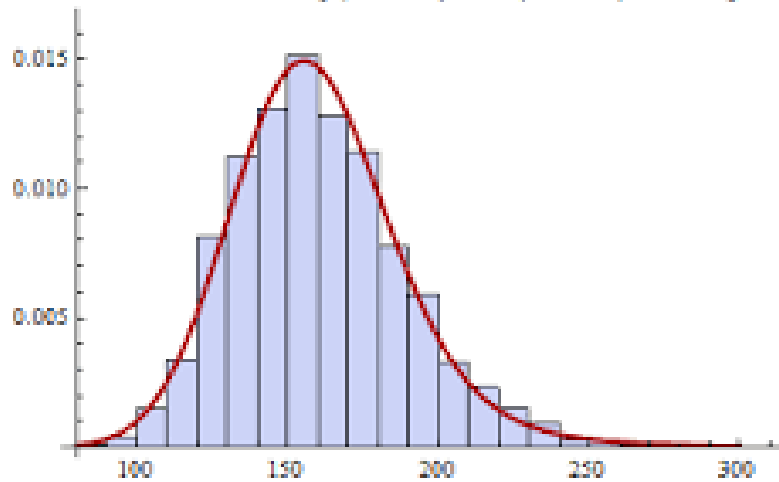
Cada individu presenta valors característics en un moment determinat que conformen en conjunt la distribució a la població.

Un objectiu és elaborar una bona descripció de la distribució i fer-la servir per a predir resultats, com ara intervals de normalitat clínics o criteris diagnòstic.

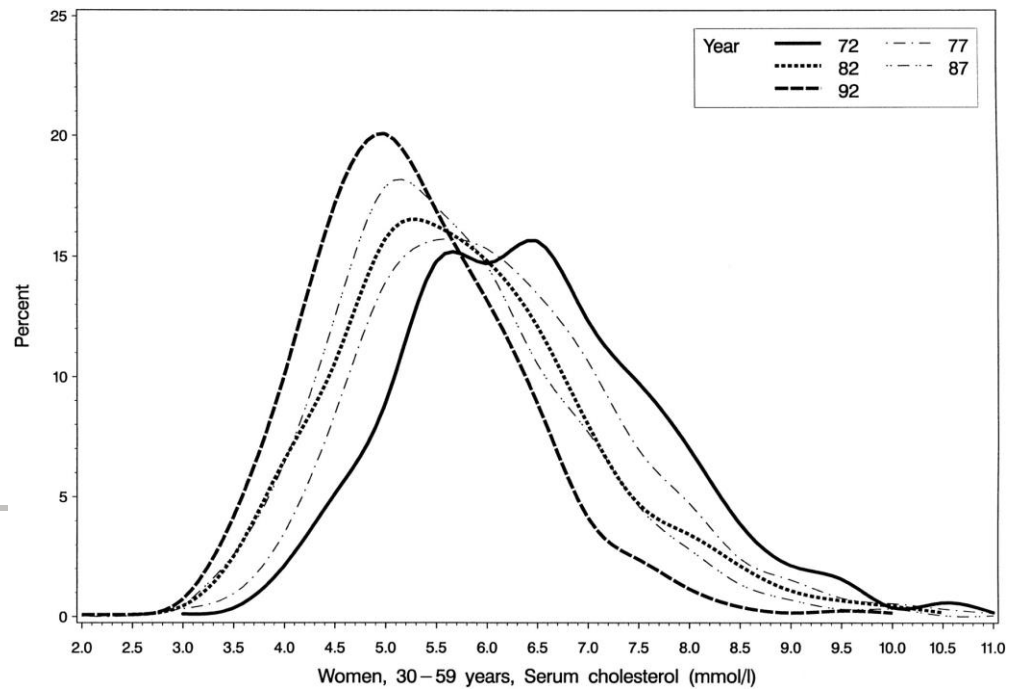
# Exemple

Total Cholesterol Distribution - Age ≤ 20

StableDistribution[1, 1.8259, 0.999, 161.49, 18.8693]



|                  | Statistic | P-Value  |
|------------------|-----------|----------|
| Anderson-Darling | 1.00838   | 0.352878 |
| Cramér-von Mises | 0.142654  | 0.412758 |



# Explorar la distribució

---

Dades d'una característica (quantitativa)

Mètodes:

- Boxplot
- Histograma
- Densitat
- Mitjana, desviació estàndard, etc.

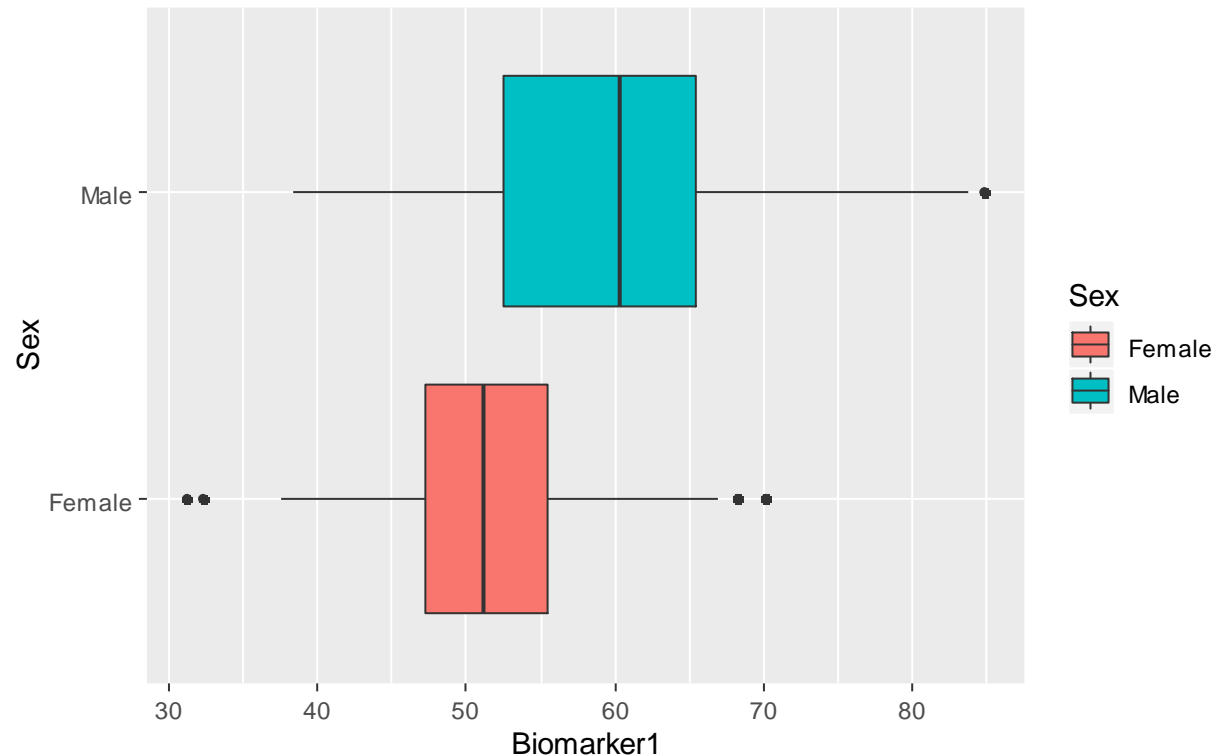
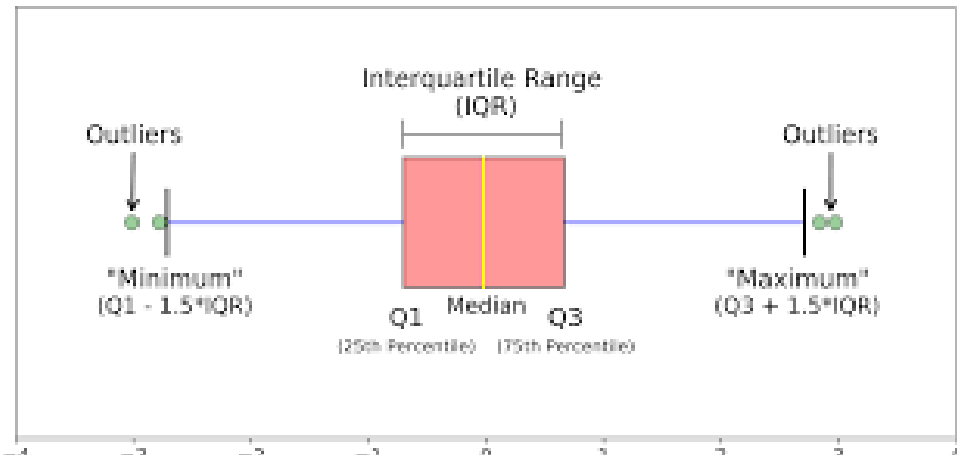
Assumir un model, p.e. Normal

**Exemple: Farem servir les dades de Task1**

# Exemple: Boxplot

Distribució dels valors del  
Biomarcador 1 segons sexe  
(Task1)

La distribució observada en les  
dones té valors inferiors als  
homes, suggerint una  
diferència que cal explorar.



```
ggplot(Task1, aes(x=Sex, y=Biomarker1, fill=Sex))+  
  geom_boxplot()+  
  coord_flip()
```

# Exemple: quantils

---

El quantil q% és el valor de la variable pel qual el q% dels individus tenen valors iguals o inferiors. P.e. Si el quantil 95% és 12.3, vol dir que a la població el 95% del individus tenen valors iguals o inferiors a 12.3.

En l'exemple de Task1:

```
Task1 %>% group_by(Sex) %>%  
  summarise(q05=quantile(Biomarker1,0.05),  
            q95=quantile(Biomarker1,0.95))
```

```
# A tibble: 2 x 3  
  Sex      q05    q95  
  <fct> <dbl> <dbl>  
1 Female 39.8  62.3  
2 Male  45.2  75.7
```

El 5% de del dones tenen valors inferiors a 39.8, mentre que els homes tenen valors inferiors a 45.2

# La funció summary

---

```
Task1 %>% dplyr::select(Biomarker1) %>% summary()
```

```
  Biomarker1  
Min.      :31.20  
1st Qu.   :48.70  
Median    :54.50  
Mean      :55.41  
3rd Qu.   :61.35  
Max.      :84.80
```

Podeu fer servir la funció summary per a obtenir una descriptiva de quantils.



# Filtrar i seleccionar

---

```
> Task1 %>% filter(Sex=='Male') %>%  
+   dplyr::select(Biomarker1) %>% summary()
```

```
  Biomarker1  
Min.   :38.30  
1st Qu.:52.50  
Median :60.30  
Mean   :59.49  
3rd Qu.:65.40  
Max.   :84.80
```

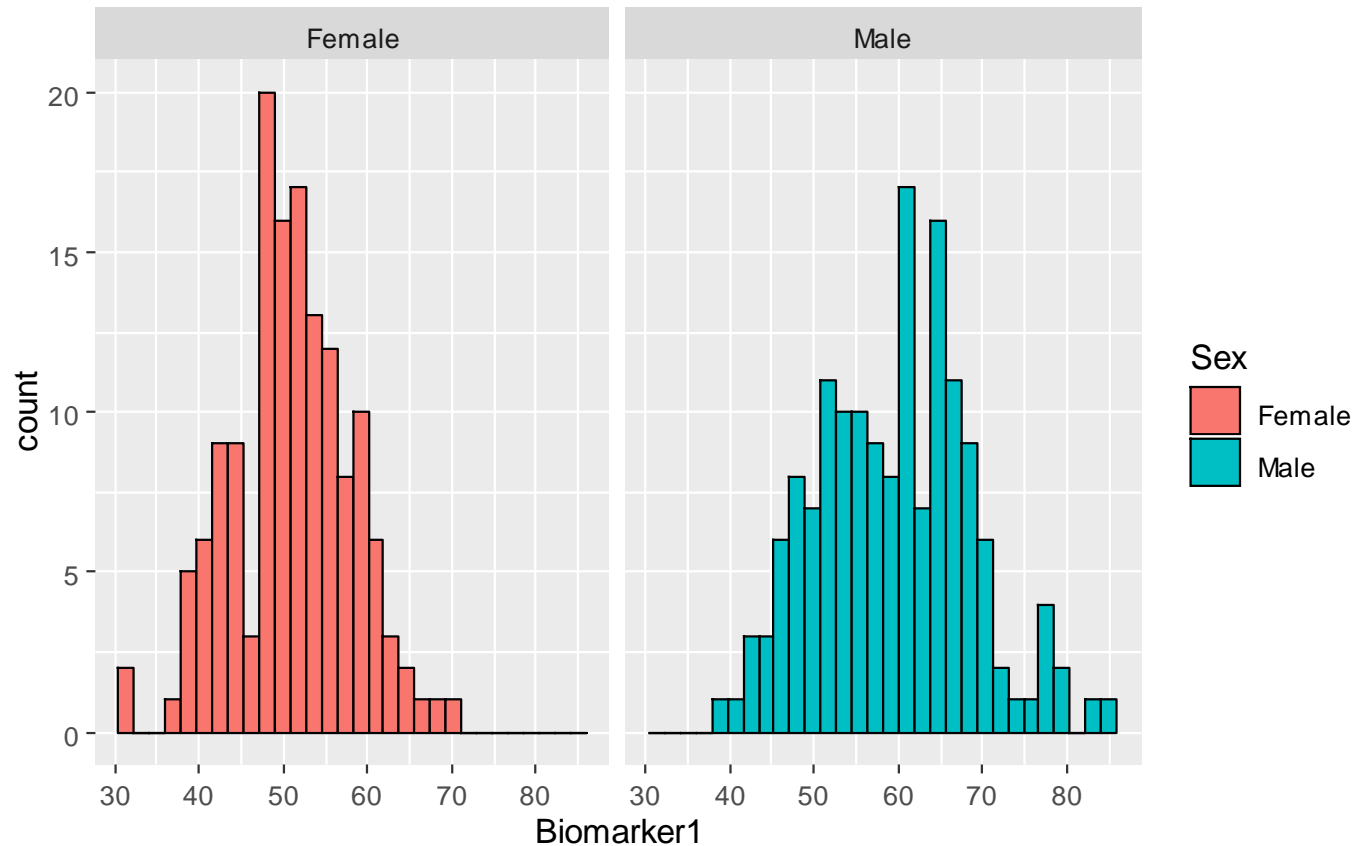
```
> Task1 %>% filter(Sex=='Female') %>%  
+   dplyr::select(Biomarker1) %>% summary()
```

```
  Biomarker1  
Min.   :31.20  
1st Qu.:47.20  
Median :51.20  
Mean   :51.06  
3rd Qu.:55.50  
Max.   :70.10
```

Podem filtrar les dades per a obtenir homes o dones i després seleccionar una variable per aplicar summary

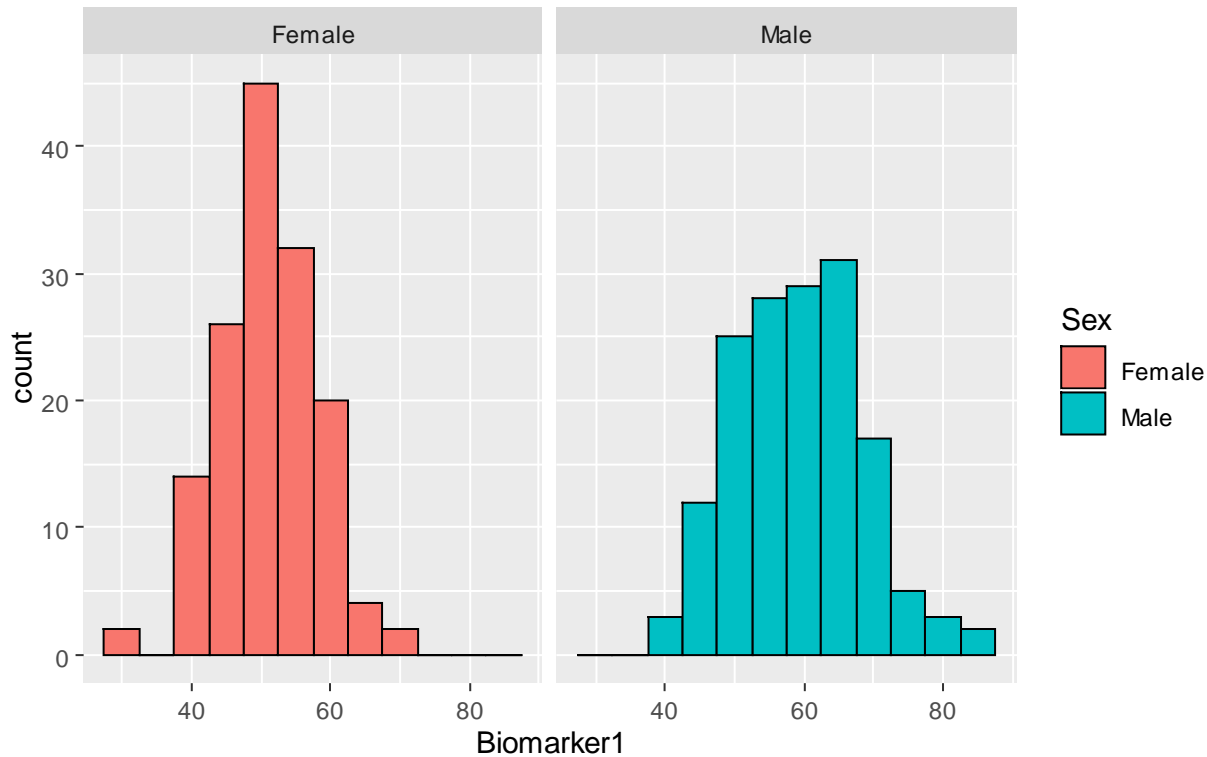
```
ggplot(Task1, aes(x=Biomarker1))+  
  geom_histogram(color='black', aes(fill=Sex))+  
  facet_wrap(~Sex)
```

# Exemple: histogrammes



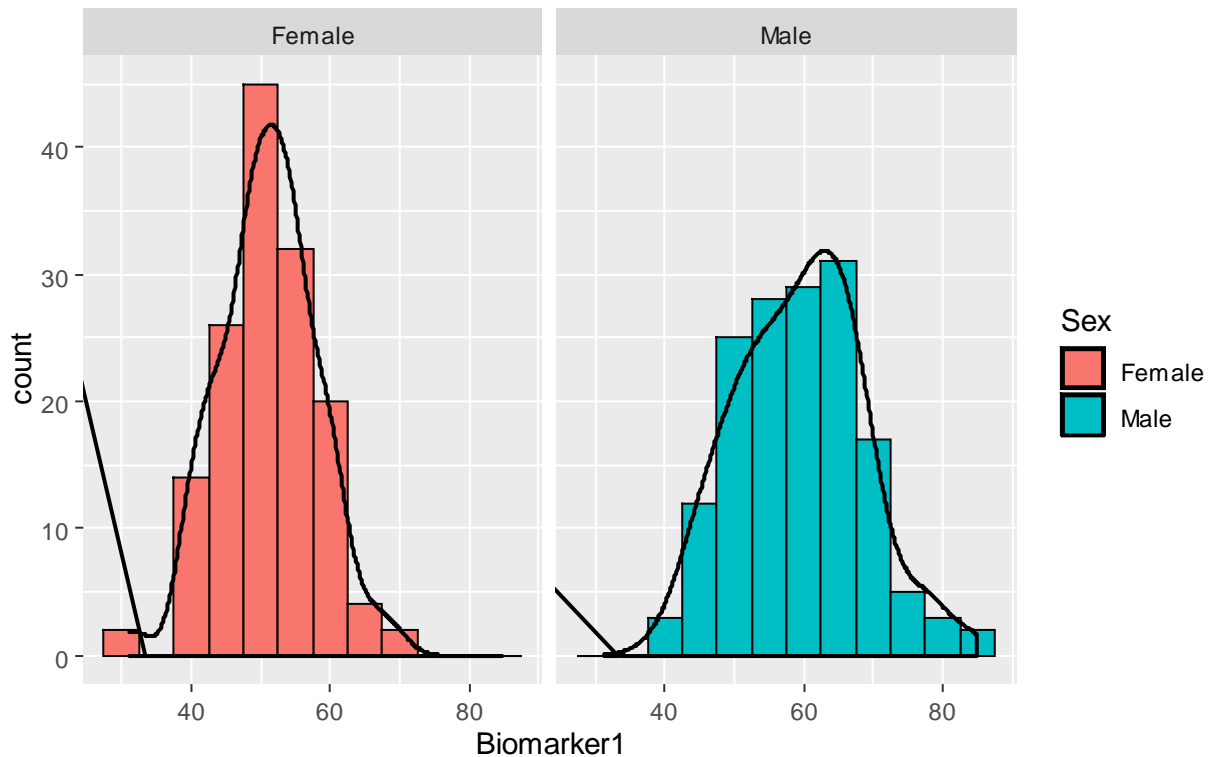
L'elecció de l'amplada de banda (bandwith) de l'histograma és important.

```
ggplot(Task1, aes(x=Biomarker1))+  
  geom_histogram(color='black', aes(fill=Sex), binwidth = 5)+  
  facet_wrap(~Sex)
```



# Example: Density

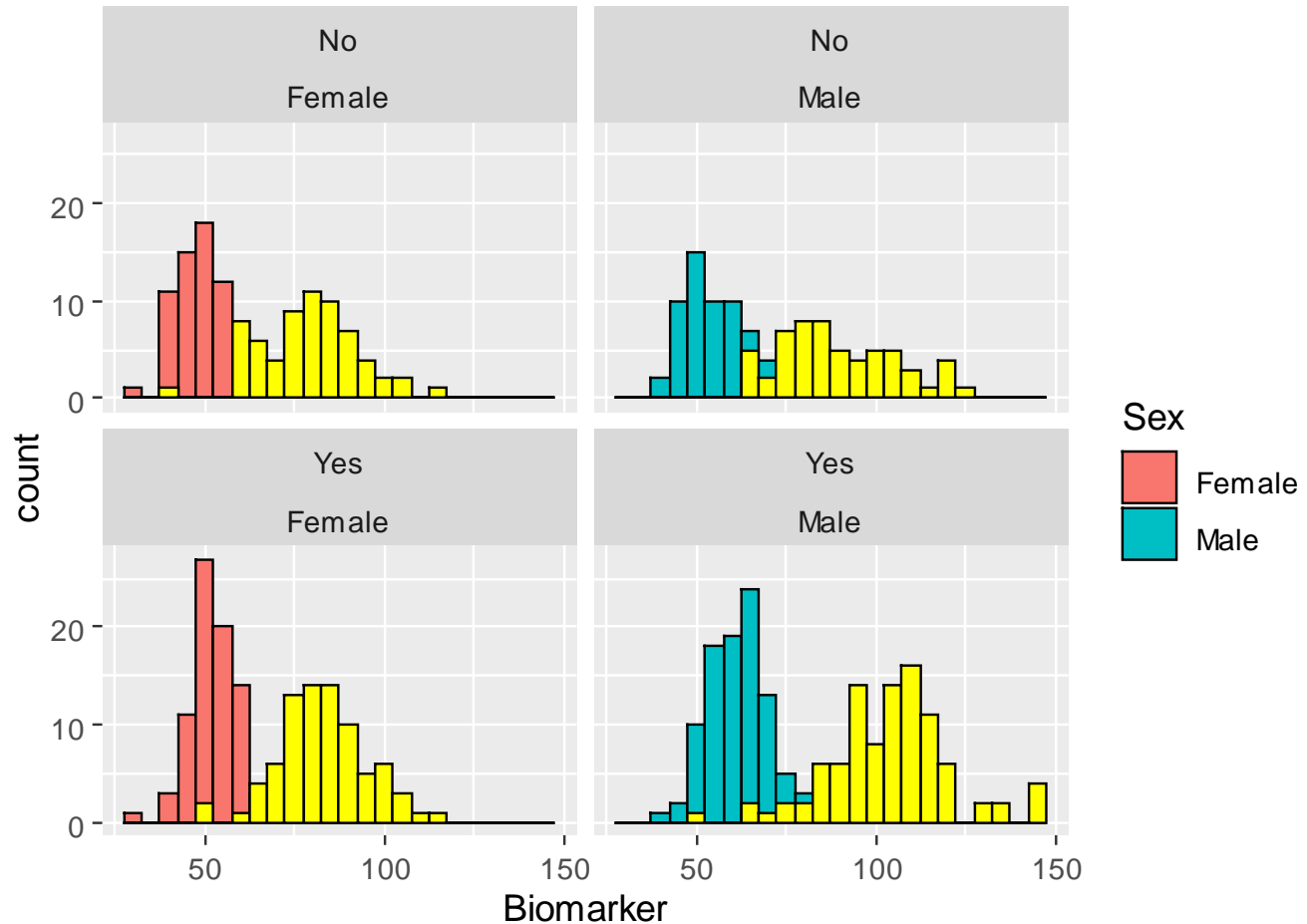
```
ggplot(Task1, aes(x=Biomarker1)) +  
  geom_histogram(color='black', aes(fill=Sex), binwidth = 5) +  
  geom_density(aes(y=5 * ..count..), size=1) +  
  facet_wrap(~Sex)
```



```

ggplot(Task1, aes(x=Biomarker1))+
  geom_histogram(color='black', aes(fill=Sex), binwidth = 5)+
  geom_histogram(aes(x=Biomarker2), color='black', fill='yellow', binwidth = 5)+
  facet_wrap(Disease~Sex)+
  xlab('Biomarker')

```



# El concepte d'interval de referència

L'interval de referència és un rang de valors on es troba un determinat % dels valors de la població.

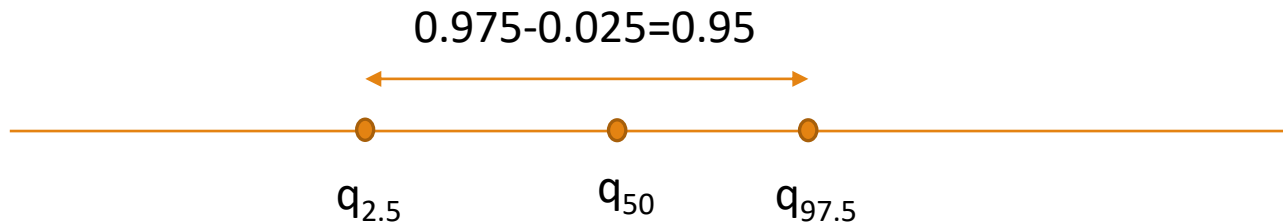
En general, es calculen intervals del 95% (centrats en la mitjana).

Els intervals de referència es fan servir com a valors de normalitat d'un biomarcador

| Healthy Cholesterol Range    |        |         |              |       |
|------------------------------|--------|---------|--------------|-------|
|                              | Unit   | Optimal | Intermediate | High  |
| Total Cholesterol            | mg/dL  | <200    | 200 - 239    | >239  |
|                              | mmol/L | <5.2    | 5.3 - 6.2    | >6.2  |
| LDL Cholesterol (calculated) | mg/dL  | <130    | 130 - 159    | >159  |
|                              | mmol/L | <3.36   | 3.36 - 4.11  | >4.11 |
| HDL Cholesterol              | mg/dL  | >60     | 40 - 60      | <40   |
|                              | mmol/L | >1.55   | 1.03 - 1.55  | <1.03 |
| Triglycerides                | mg/dL  | <150    | 150 - 199    | >199  |
|                              | mmol/L | <1.69   | 1.69 - 2.25  | >2.25 |
| Non-HDL-C (calculated)       | mg/dL  | <130    | 130 - 159    | >159  |
|                              | mmol/L | <3.3    | 3.4 - 4.1    | >4.1  |
| TG to HDL ratio (calculated) | mg/dL  | <3      | 3.1 - 3.8    | >3.8  |
|                              | mmol/L | <1.33   | 1.34 - 1.68  | >1.68 |

# Interval de referència

Per a un 95%, calcularem els quantils 0.025 i 0.975



```
quantile(Task1$Biomarker1,c(0.025,0.975))
```

|  | 2.5%    | 97.5%   |
|--|---------|---------|
|  | 39.6475 | 76.0875 |

```
Task1 %>% group_by(Sex) %>%  
  summarise(`2.5%`=quantile(Biomarker1,0.025),  
            `97.5%`=quantile(Biomarker1,0.975),  
            `50%`=quantile(Biomarker1,0.5),  
            n=n())
```

|   | Sex                | `2.5%`             | `97.5%`            | `50%`              | n                  |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|
|   | <i>&lt;fct&gt;</i> | <i>&lt;dbl&gt;</i> | <i>&lt;dbl&gt;</i> | <i>&lt;dbl&gt;</i> | <i>&lt;int&gt;</i> |
| 1 | Female             | 39.2               | 65.0               | 51.2               | 145                |
| 2 | Male               | 43.3               | 78.0               | 60.3               | 155                |

# Interval de referència

---

Els intervals de referència els hem calculat sobre una mostra, per tant són aproximacions que depenen de la qualitat de la mostra.

En general, un interval serveix només per a la població i les condicions en que s'ha obtingut la mostra.

Els intervals de referència canvien, en genarl, amb l'edat, el sex, i altres variables. Per tant, no és trivial obtenir valors vàlids que serveixin en clínica.

Els laboratoris clínics defineixen el seus estàndards tenint en compte la població i la metodologia de les analítiques (poden canviar de laboratori a laboratori). L'ús d'estàndards és important per a mantenir la qualitat dels resultats.

Valors fora de l'IR no indiquen necessàriament patologia!



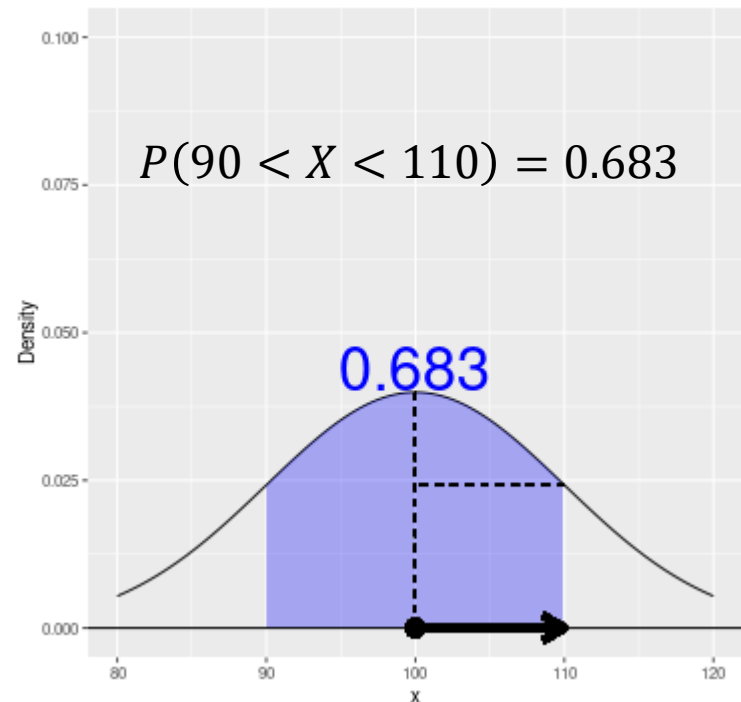
# Càlculs amb distribució normal

La distribució normal és un model matemàtic per a representar les probabilitats de resultats d'observacions d'una característica contínua.

L'àrea sota la corba normal entre dos punts correspon a la probabilitat de trobar individus amb valors dins de l'interval.

Els valors es distribueixen simètricament al voltant de la mitjana  $\mu$ .

La dispersió de valors depèn del paràmetre  $\sigma$ .



# Càlculs amb distribució normal

---

a

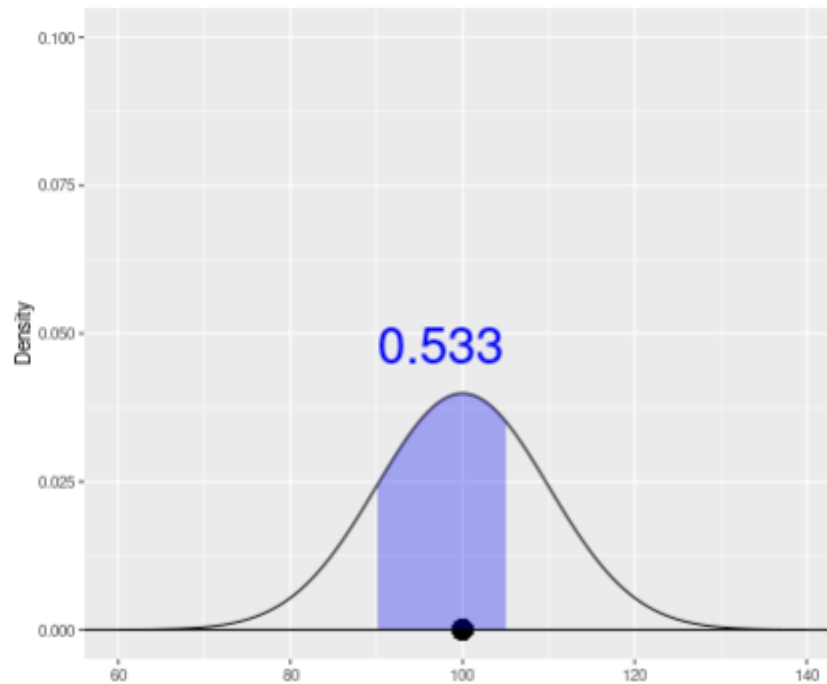
90

b

105

Select an option:

- $P(a < X < b)$
- $P(X < a)$
- $P(X < b)$
- $P(X > a)$
- $P(X > b)$
- $P(X < \mu)$



a

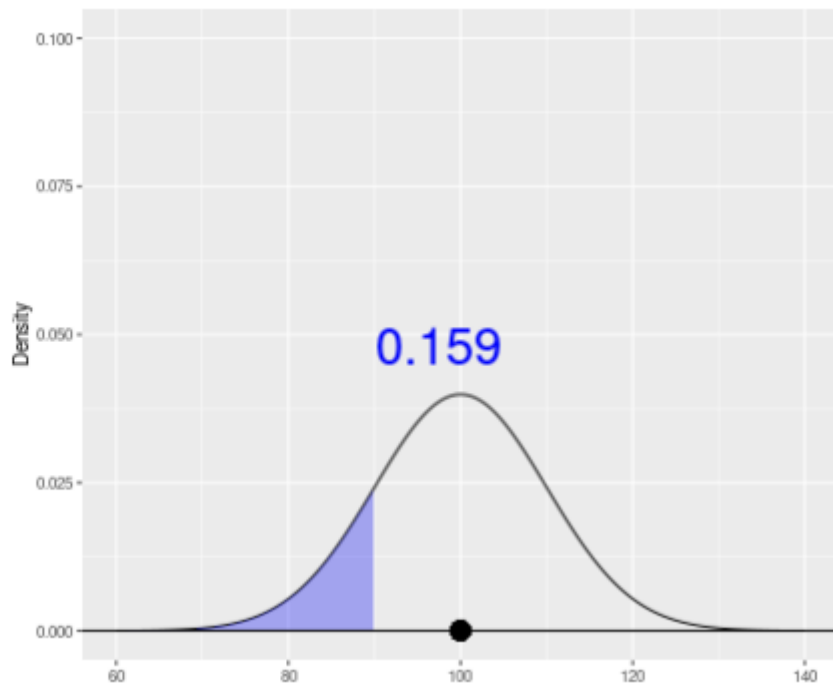
90

b

105

Select an option:

- $P(a < X < b)$
- $P(X < a)$
- $P(X < b)$
- $P(X > a)$
- $P(X > b)$
- $P(X < \mu)$



$$P(X < a) = 1 - P(X > a)$$

$$P(X < 90) = 0.159$$

$$P(X > 90) = 1 - 0.159 = 0.841$$

a

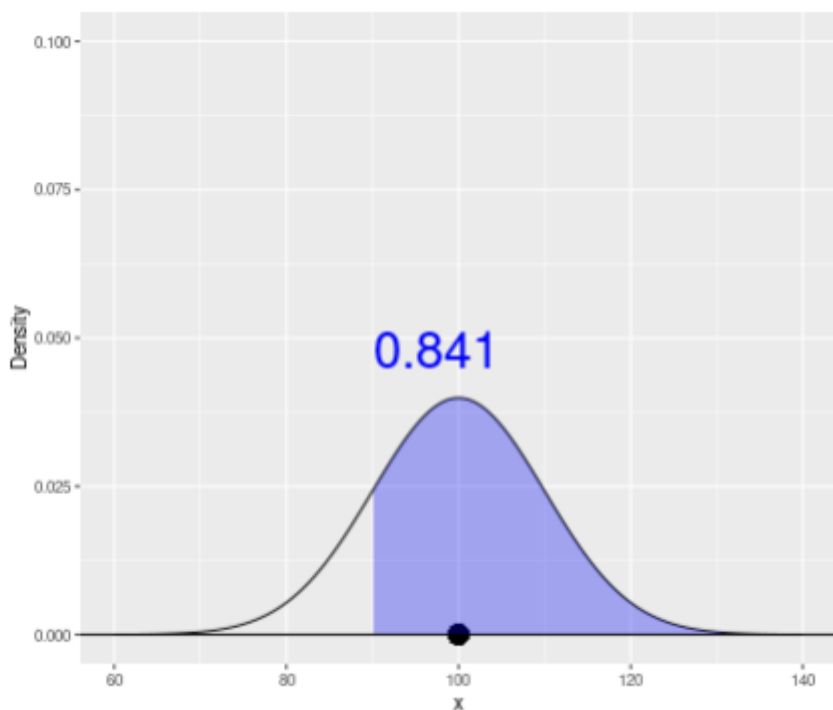
90

b

105

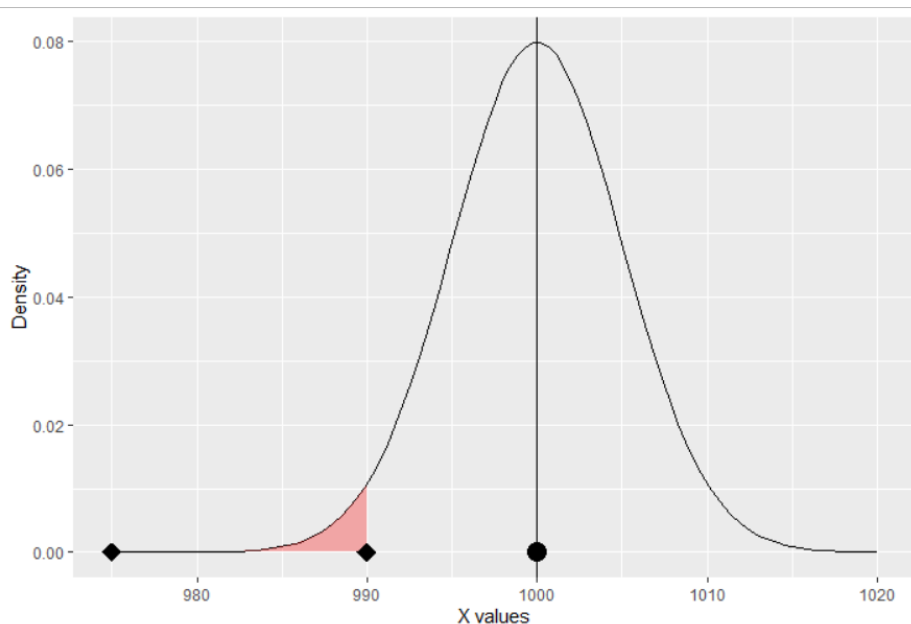
Select an option:

- $P(a < X < b)$
- $P(X < a)$
- $P(X < b)$
- $P(X > a)$
- $P(X > b)$
- $P(X < \mu)$



# Càlculs amb distribució normal

Un procés de producció d'un producte alimentari produeix paquets amb una mitjana de 1000 grs i una desviació estàndard de 5. Assumim una distribució normal  $N(1000,5)$ . La legislació demana un mínim de 990grs per paquet. Quin percentatge de paquets tindran un pes inferior a 990 grs?



$$P(X < 990) = 0.023$$

```
round(pnorm(990,1000,5),3)
```

La funció `pnorm(a,mu,sigma)` permet calcular la probabilitat d'observar resultats inferiors al valor `a` per a una  $N(\mu,\sigma)$

# Càlculs amb distribució normal

---

Un biomarcador segueix una distribució  $N(100,5)$  (mg/ml). Quin percentatge d'individus tindran valors inferiors a 93.5 mg/ml.? Quin percentatge estarà entre 95 i 105 mg/ml. Per sota de quin valor es troba el 90% dels individus (quantil 90)

$$P(X < 93.5) = 0.097 \quad \text{round(pnorm(93.5,100,5),3)}$$

$$P(95 < X < 105) = P(X < 105) - P(X < 95) = 0.6837$$

$$\text{round(pnorm(105,100,5)-pnorm(95,100,5),3)}$$

$$P(X < x_{0.90}) = 106.4 \quad \text{round(qnorm(0.90,100,5),1)}$$

La funció `qnorm` permet calcular un quantil d'una distribució normal.

# Càlculs amb distribució normal

Un biomarcador segueix una distribució  $N(100,5)$  (mg/ml). Calcula l'interval de referència que inclou el 95% de valors al voltant de la mitjana.

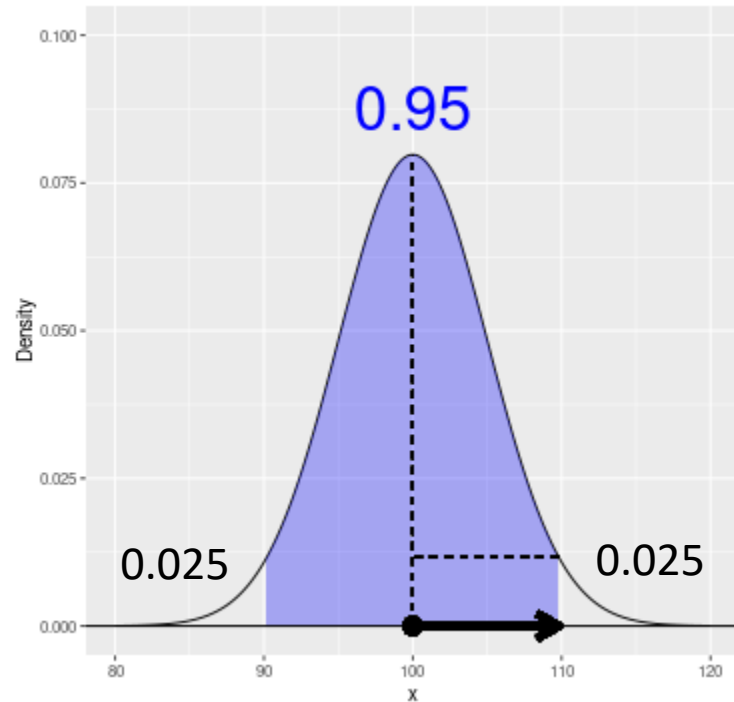
$$P(a < X < b) = 0.95$$

$$P(X < a) = 0.025$$

$$P(X < b) = 0.975$$

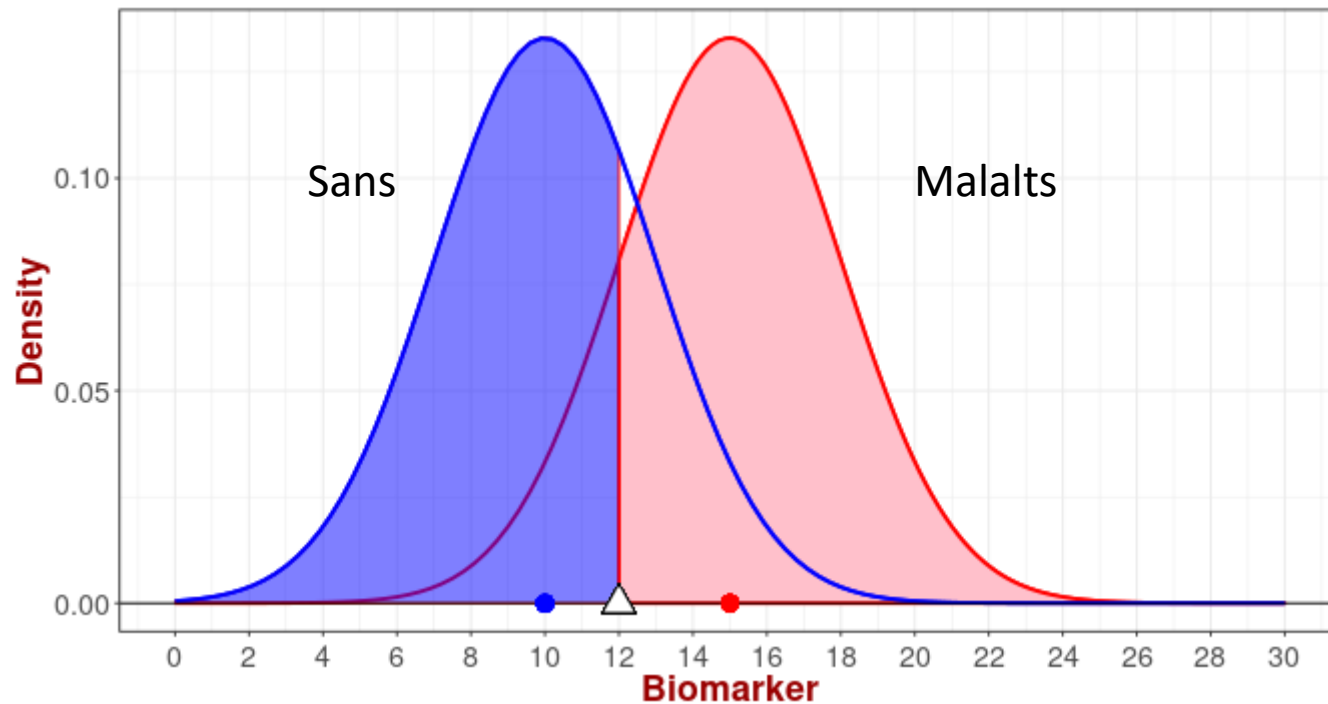
```
round(qnorm(c(0.025,0.975),100,5),1)
```

90.2 109.8



# Distribució en funció del grup

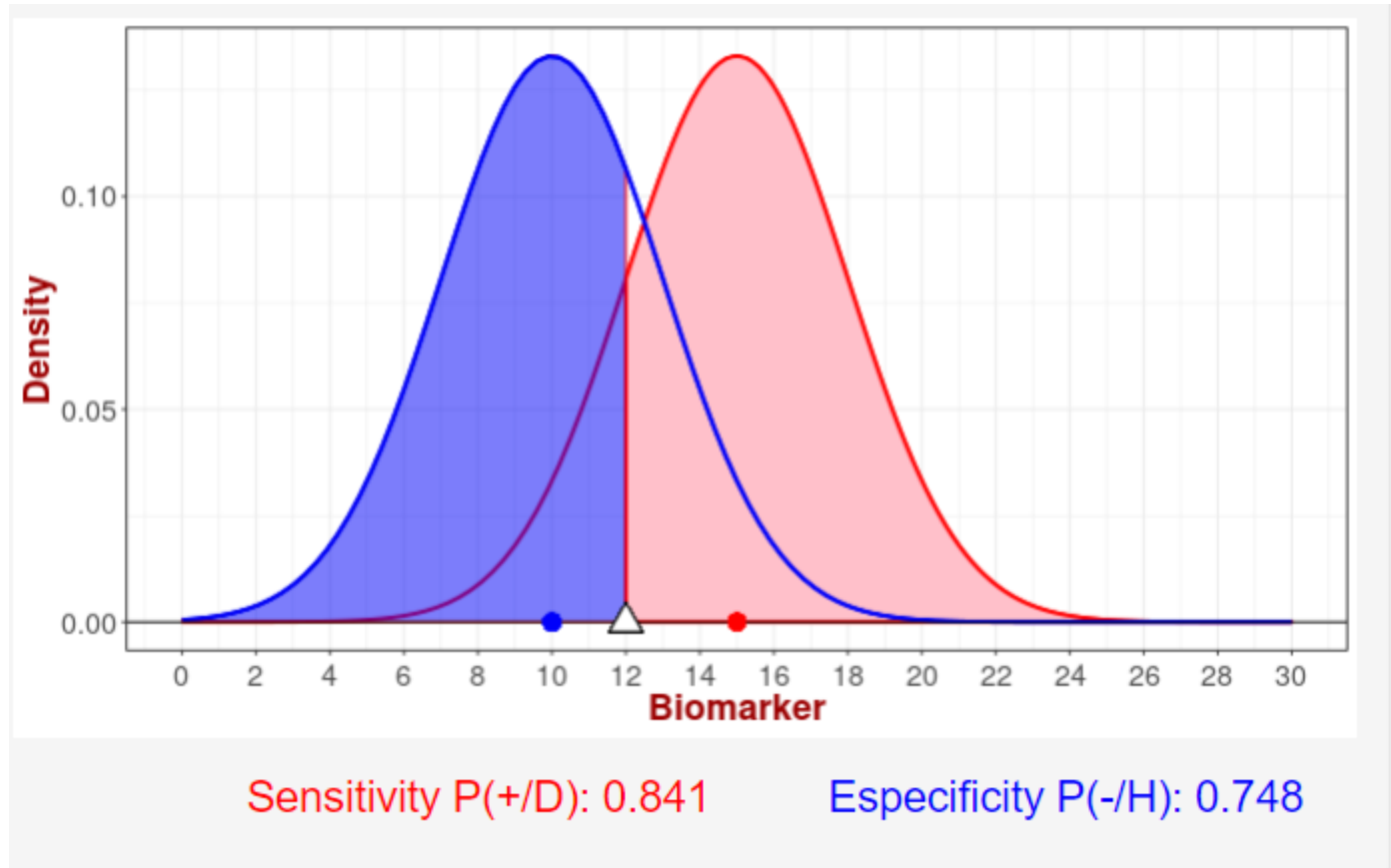
Suposem que les persones sense patologies tenen una distribució  $N(10,3)$ , mentre que les afectades d'una patologia segueixen una  $N(15,3)$ .



# Distribució en funció del grup

Suposem que les persones sense patologies tenen una distribució  $N(10,3)$ , mentre que les afectades d'una patologia segueixen una  $N(15,3)$ .

Si utilitzem el punt diagnòstic 12, la sensibilitat i especificitat resultant són:





# Distribució en funció del grup

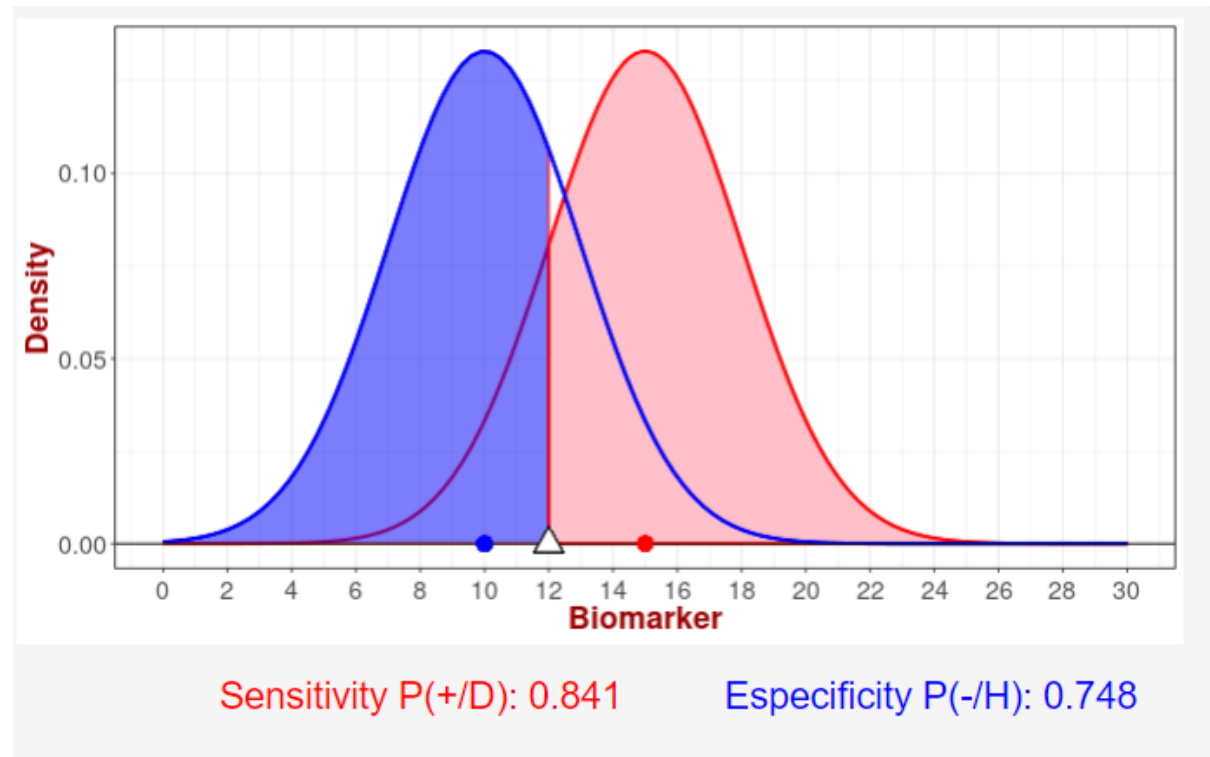
Suposem que les persones sense patologies (H) tenen una distribució  $N(10,3)$ , mentre que les afectades d'una patologia (D) segueixen una  $N(15,3)$ .

$$P(-|H) = P(X_H < 12)$$

`round(pnorm(12,10,3),3)`  
0.748

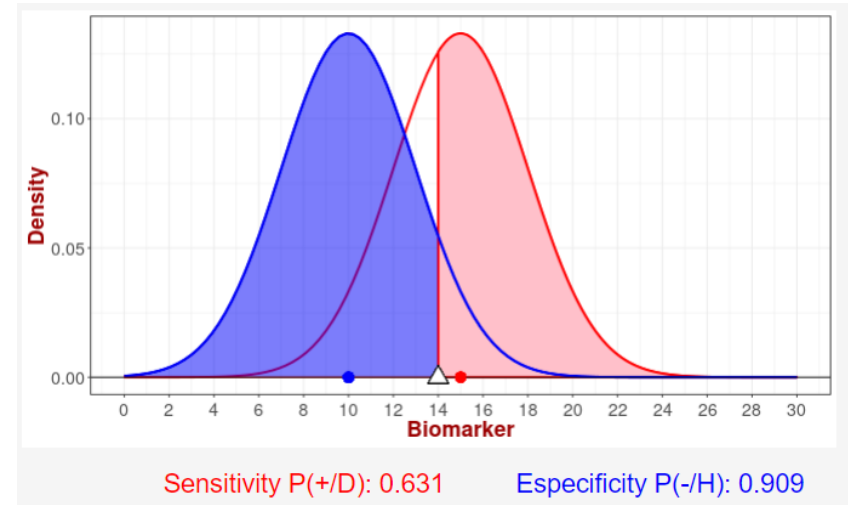
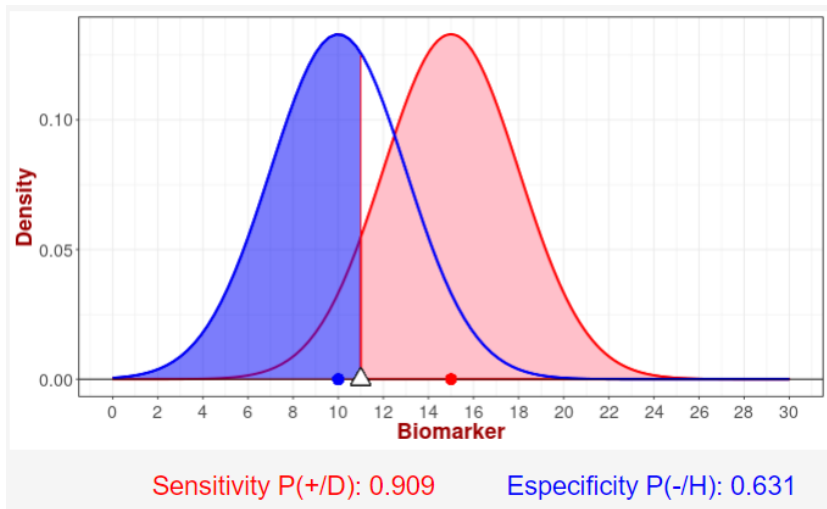
$$P(+|D) = P(X_D > 12)$$

`round(1-pnorm(12,15,3),3)`  
0.841



La sensitivitat i l'especificitat depenen del punt diagnòstic.

Si augmentem la sensibilitat, disminueix l'especificitat, i viceversa.



# Anàlisi en funció de mostres

## Healthy population

Mean:

sd:

Sample size:


## Disease population

Mean:

sd:

Sample size:

Value for criteria

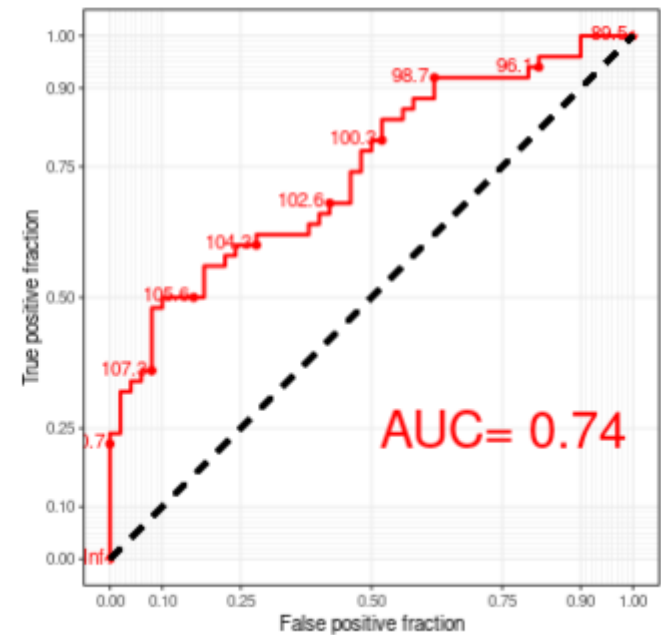
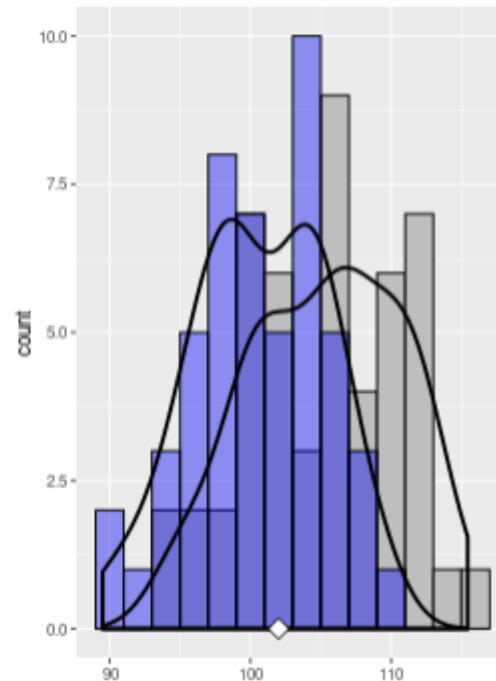
 Generate a new sample

Based on the data, the (+) and (-) results of the test are computed according to the diagnostic point defined.

| Test | H  | D  |
|------|----|----|
| (+)  | 22 | 34 |
| (-)  | 28 | 16 |

Sensitivity: 0.68

Specificity: 0.56



# Anàlisi en funció de mostres

## Healthy population

Mean:

sd:

Sample size:

## Disease population

Mean:

sd:

Sample size:

Value for criteria

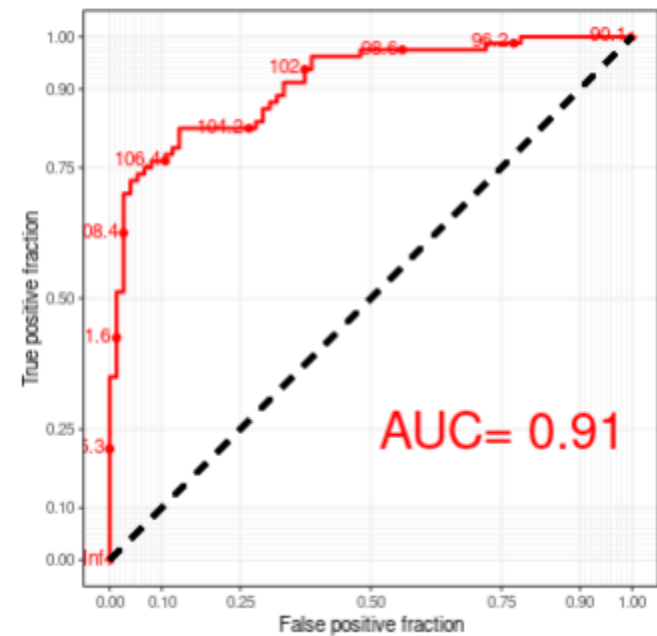
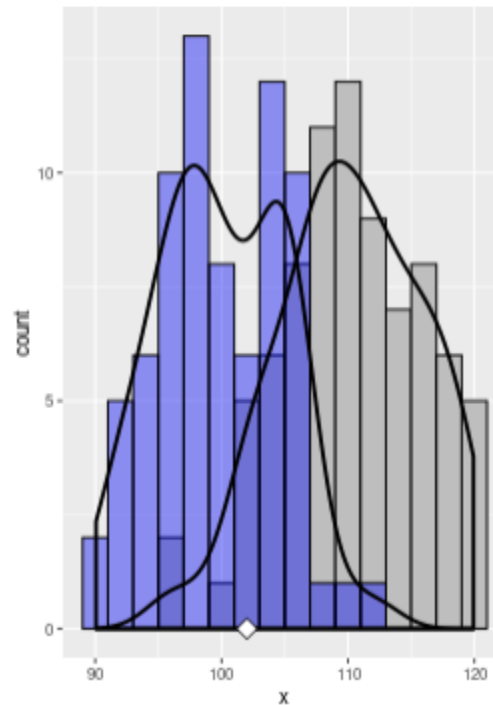
 Generate a new sample

Based on the data, the (+) and (-) results of the test are computed according to the diagnostic point defined.

| Test | H  | D  |
|------|----|----|
| (+)  | 28 | 74 |
| (-)  | 47 | 6  |

Sensitivity: 0.92

Specificity: 0.63



# Anàlisi en funció de mostres

## Healthy population

Mean:

sd:

Sample size:

## Disease population

Mean:

sd:

Sample size:

Value for criteria

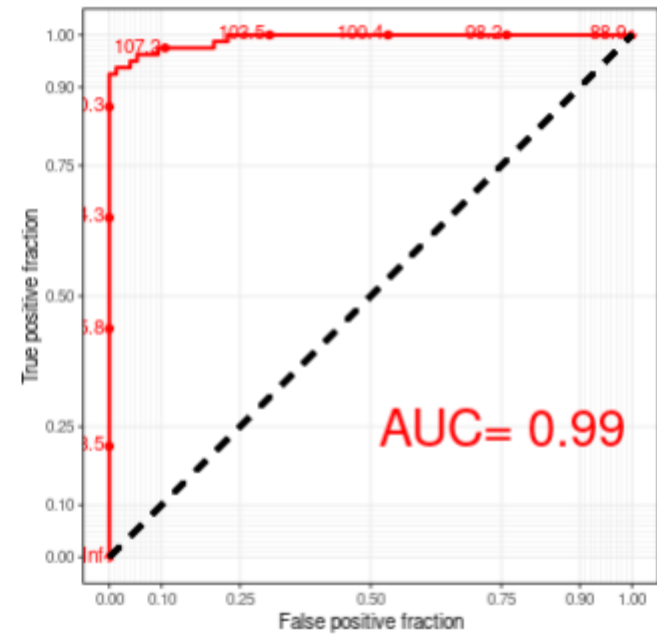
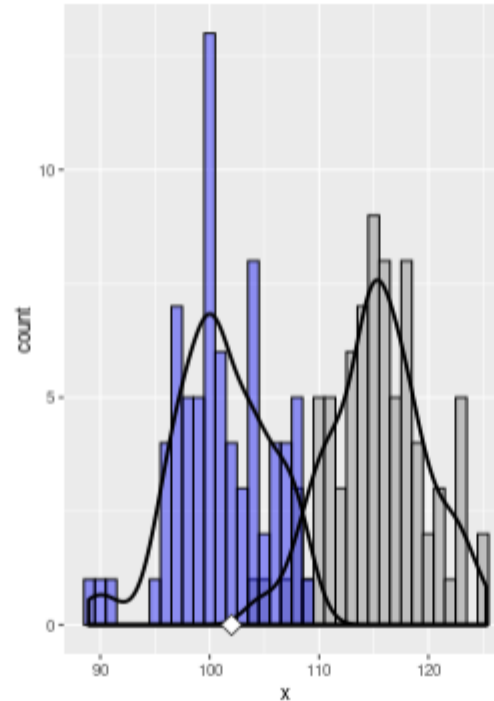
 Generate a new sample

Based on the data, the (+) and (-) results of the test are computed according to the diagnostic point defined.

| Test | H  | D  |
|------|----|----|
| (+)  | 31 | 80 |
| (-)  | 44 | 0  |

Sensitivity: 1

Specificity: 0.59



# Anàlisi en funció de mostres

## Healthy population

Mean:

sd:

Sample size:

## Disease population

Mean:

sd:

Sample size:

Value for criteria

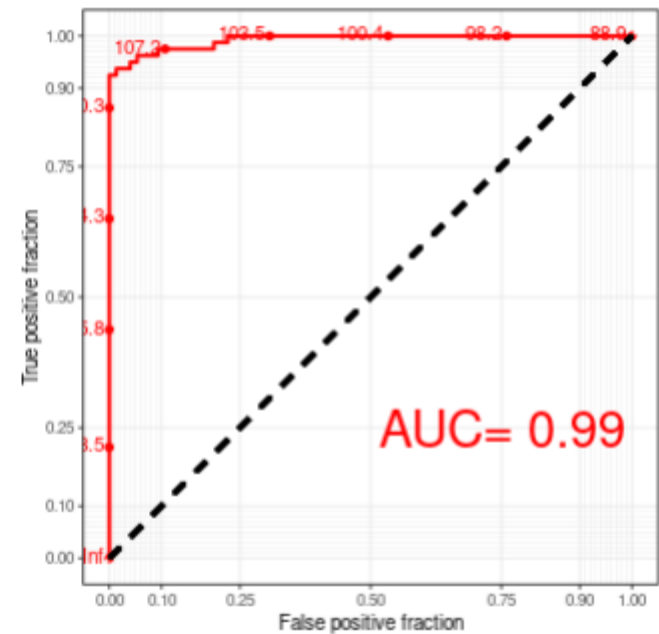
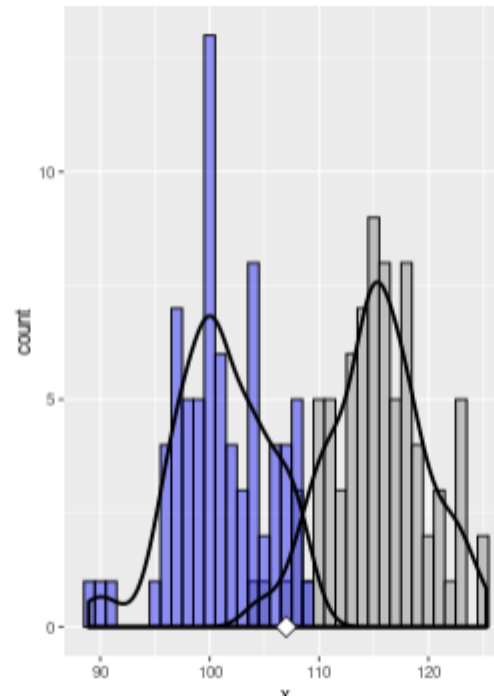
 Generate a new sample

Based on the data, the (+) and (-) results of the test are computed according to the diagnostic point defined.

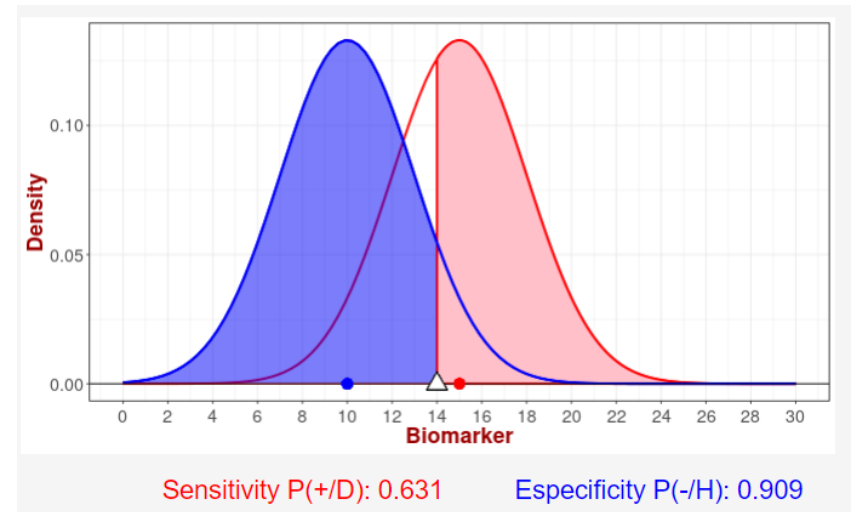
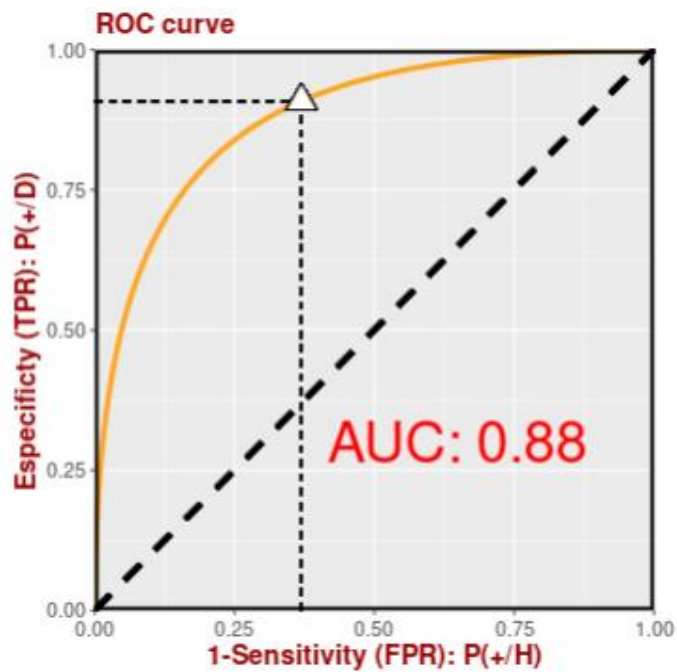
| Test | H  | D  |
|------|----|----|
| (+)  | 9  | 78 |
| (-)  | 66 | 2  |

Sensitivity: 0.98

Specificity: 0.88



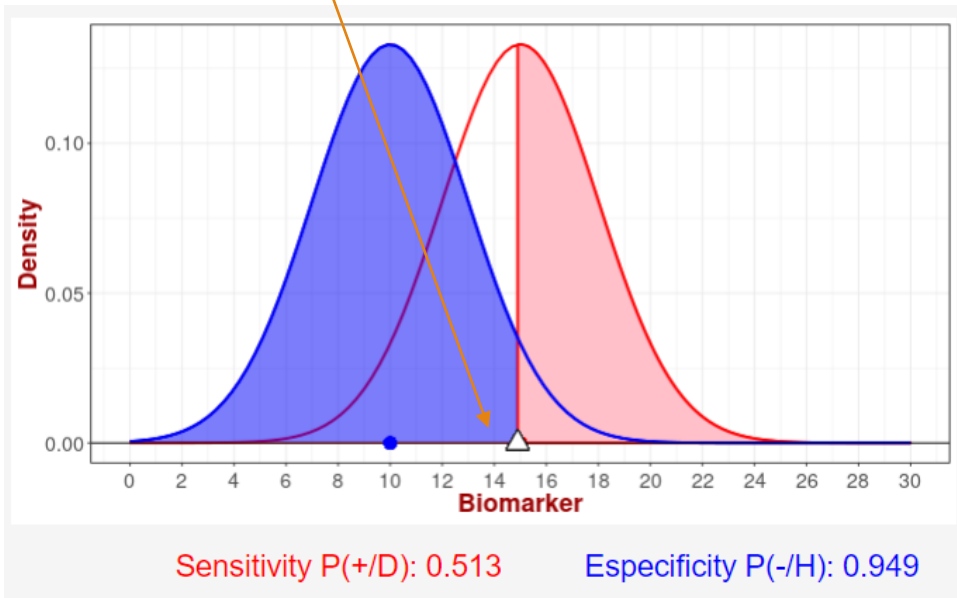
# Corba ROC



# Càlcul d'un punt diagnòstic

Quin punt diagnòstic cal fer servir si volem una especificitat del 95%?  
Quina sensibilitat en resulta?

$$P(-|H) = P(X_H < x_d) = 0.95$$



$$P(-|H) = P(X_H < x_d) = 0.95$$

```
round(qnorm(0.95,10,3),1)
```

14.9

$$P(+|D) = P(X_D > 14.9) = 0.513$$

```
round(1-pnorm(14.9,15,3),3)
```

0.513



# Aplicació

---

Les persones afectades d'una patologia tenen una distribució d'un biomarcador que segueix una  $N(50.1, 4.3)$ . En les persones sense patologia, aquesta distribució és  $N(53.4, 3.2)$ .

Quin punt diagnòstic faríem servir per a tenir una sensibilitat de 0.9?

Quina especificitat tindríem amb aquest punt?

Si fem servir aquest punt diagnòstic, quina proporció de falsos positius faríem si la prevalença de la malaltia és del 5%?

# Aplicació

---

Amb les dades de Task1, dissenyeu un criteri diagnòstic per a classificar sans i malalts en funció del Biomarcador2.

Estima la mitjana i la desviació estàndard del biomarcador per a cada grup i assumeix distribució normal en els dos grups.

Quina sensibilitat i especificitat màximes podries obtenir?

És un bon biomarcador?

## Diagnostic value and dynamic variance of serum antibody in coronavirus disease 2019.

Jin Y<sup>1</sup>, Wang M<sup>1</sup>, Zuo Z<sup>1</sup>, Fan C<sup>2</sup>, Ye F<sup>3</sup>, Cai Z<sup>4</sup>, Wang Y<sup>1</sup>, Cui H<sup>1</sup>, Pan K<sup>1</sup>, Xu A<sup>5</sup>.

### Author information

#### Abstract

**OBJECTIVE:** To investigate the diagnostic value of serological test and dynamic variance of serum antibody in coronavirus disease 2019 (COVID-19).

**METHODS:** We retrospectively included 43 patients with a laboratory-confirmed infection and 33 patients with suspected infection who were finally excluded. The IgM/IgG titer of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was measured by chemiluminescence immunoassay analysis.

**RESULTS:** Compared to molecular detection, the sensitivity of serum IgM and IgG antibodies to diagnose COVID-19 were 48.1% and 88.9%, and the specificity were 100% and 90.9%, respectively. In COVID-19 group, IgM positive rate almost increased first then decreased over time, however, IgG positive rate increased till 100% and was higher than IgM all the time. IgM positive rate and titer were not significantly different before and after viral negative. IgG positive rate was up to 90% and not significantly different before and after viral negative. However, the median titer of IgG after viral negative was twice than that before with significant difference.

**CONCLUSION:** Viral serological test is an effective diagnostic means for SARS-CoV-2 infection. The positive rate and titer variance of IgG are higher than that of IgM in COVID-19.

# Diagnostic performance and usability of the VISITECT CD4 semi-quantitative test for advanced HIV disease screening.

[Ndlovu Z<sup>1</sup>](#), [Massaquoi L<sup>2</sup>](#), [Bangwen NE<sup>2</sup>](#), [Batumba JN<sup>2</sup>](#), [Bora RU<sup>2</sup>](#), [Mbuaya J<sup>2</sup>](#), [Nzadi R<sup>2</sup>](#), [Ntabugi N<sup>2</sup>](#), [Kisaka P<sup>2</sup>](#), [Manciya G<sup>2</sup>](#), [Moudashirou R<sup>3</sup>](#), [Pangani H<sup>4</sup>](#), [Mangochi P<sup>4</sup>](#), [Makoko R<sup>4</sup>](#), [Van Laeken D<sup>4</sup>](#), [Kwitonda C<sup>4</sup>](#), [Ronoh Y<sup>5</sup>](#), [Kuwenyi K<sup>5</sup>](#), [Ortuno R<sup>5</sup>](#), [Mangwanya D<sup>6</sup>](#), [Zvidzai E<sup>6</sup>](#), [Mupepe T<sup>6</sup>](#), [Zinyowera S<sup>7</sup>](#), [Fajardo E<sup>8</sup>](#), [Ellman T<sup>1</sup>](#).

## ⊕ Author information

### Abstract

**BACKGROUND:** In sub-Saharan Africa, a third of people starting antiretroviral therapy and majority of patients returning to HIV-care after disengagement, present with advanced HIV disease (ADH), and are at high risk of mortality. Simplified and more affordable point-of-care (POC) diagnostics are required to increase access to prompt CD4 cell count screening for ambulatory and asymptomatic patients. The Visitect CD4 Lateral Flow Assay (LFA) is a disposable POC test, providing a visually interpreted result of above or below 200 CD4cells/mm<sup>3</sup>. This study evaluated the diagnostic performance of this index test.

**METHODS:** Consenting patients above 18years of age and eligible for CD4 testing were enrolled in Nsanje district hospital (Malawi), Gutu mission hospital (Zimbabwe) and Centre hospitalier de Kabinda (DRC). A total of 708 venous blood samples were tested in the index test and in the BD FACSCount assay (reference test method) in the laboratories (Phase 1) to determine diagnostic accuracy. A total of 433 finger-prick (FP) samples were tested on the index test at POC by clinicians (Phase 2) and a self-completed questionnaire was administered to all testers to explore usability of the index test.

**RESULTS:** Among 708 patients, 67.2% were female and median CD4 was 297cells/mm<sup>3</sup>. The sensitivity of the Visitect CD4 LFA using venous blood in the laboratory was 95.0% [95% CI: 91.3-97.5] and specificity was 81.9% [95% CI: 78.2-85.2%]. Using FP samples, the sensitivity of the Visitect CD4 LFA was 98.3% [95% CI: 95.0-99.6] and specificity was 77.2% [95% CI: 71.6-82.2%]. Usability of the Visitect CD4 LFA was high across the study sites with 97% successfully completed tests. Due to the required specific multiple incubation and procedural steps during the Visitect CD4 LFA testing, few health workers (7/26) were not confident to manage testing whilst multi-tasking in their clinical work.

**CONCLUSIONS:** Visitect CD4 LFA is a promising test for decentralized CD4 screening in resource-limited settings, without access to CD4 testing and it can trigger prompt management of patients with AHD. Lay health cadres should be considered to conduct Visitect CD4 LFA testing in PHCs as well as coordinating all other POC quality assurance.

# Diagnostic value of pentraxin 3 in respiratory tract infections

## A meta-analysis

Wu Ye, MD\*, Qing-Dong Huang, MD, Ting-Yu Tang, MD, Guang-Yue Qin, MD

### Abstract

**Background:** Pentraxin 3 is an acute inflammatory protein of the long pentraxin subfamily. A meta-analysis was performed to assess diagnostic accuracy of pentraxin 3 for respiratory tract infections.

**Methods:** We identify studies examining diagnostic value of pentraxin 3 for respiratory tract infections by searching Pubmed, Web of Knowledge, and Cochrane Library. The sensitivity, specificity, negative likelihood ratio (LR), positive LR, and diagnostic odds ratio were pooled. The area under the summary receiver operator characteristic (SROC) curve and Q point value (Q\*) were calculated.

**Results:** A total of 8 studies with 961 individuals were eligible for this meta-analysis. The pooled sensitivity of pentraxin 3 in diagnosis of respiratory tract infections was 0.78, the pooled specificity was 0.73, the area under the SROC curve was 0.84, and the Q\* was 0.77. The area under the SROC curve of serum and bronchoalveolar lavage fluid (BALF) pentraxin 3 was 0.85 and 0.89, respectively. Meta-regression analysis revealed that cutoff value was the source of heterogeneity among the included studies. The Deek funnel plot test suggested no evidence of publication bias. Subgroup analyses showed that the area under the SROC curve of pentraxin 3 in diagnosis of ventilator-associated pneumonia (VAP) was 0.89.

**Conclusion:** Pentraxin 3 has a moderate accuracy for diagnosing respiratory tract infections and VAP. The overall diagnostic value of BALF level of pentraxin 3 is superior to its serum concentration.

**Abbreviations:** BALF = bronchoalveolar lavage fluid, COPD = chronic obstructive pulmonary disease, DOR = diagnostic odds ratio, ELISA = enzyme-linked immunosorbent assay, ESR = erythrocyte sedimentation rate, FN = false negative, FP = false positive, LR = likelihood ratio, Q\* = Q point value, QUADAS = Quality Assessment of Diagnostic Accuracy Studies, SROC = summary receiver operator characteristic, TN = true negative, TP = true positive, VAP = ventilator-associated pneumonia.

**Keywords:** diagnostic test, meta-analysis, pentraxin 3, respiratory tract infections, ventilator-associated pneumonia

[https://journals.lww.com/md-journal/FullText/2020/04030/Diagnostic\\_value\\_of\\_pentraxin\\_3\\_in\\_respiratory.15.aspx](https://journals.lww.com/md-journal/FullText/2020/04030/Diagnostic_value_of_pentraxin_3_in_respiratory.15.aspx)