

Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in

STATISTICA PER LE TECNOLOGIE E LE SCIENZE



RELAZIONE FINALE

**Uno studio statistico di confronto di misurazioni  
acquisite con TC dual-energy addominale senza  
mezzo di contrasto in ambito veterinario**

Relatrice: Prof.ssa Laura Ventura  
Dipartimento di Scienze Statistiche

Laureanda: Sofia Ferrari  
Matricola n. 1229916

Anno Accademico 2022/2023

# Indice

<b>1</b>	<b>Introduzione.....</b>	<b>1</b>
<b>2</b>	<b>Analisi esplorativa dei dati.....</b>	<b>6</b>
2.1	Il campione .....	6
2.2	Analisi esplorativa .....	8
2.3	Conclusioni.....	23
<b>3</b>	<b>Misure di agreement .....</b>	<b>24</b>
3.1	Bland-Altman.....	24
3.2	Coefficiente di concordanza di Lin .....	33
<b>4</b>	<b>Equivalence Test .....</b>	<b>38</b>
4.1	Test di equivalenza e non-inferiorità .....	38
4.2	Margine di equivalenza.....	42
4.3	TOST .....	44
4.4	TOST e NHST per le differenze di HU tra PC e VNC .....	47
<b>5</b>	<b>Conclusione.....</b>	<b>49</b>
	<b>Appendice.....</b>	<b>54</b>
	<b>Riferimenti bibliografici .....</b>	<b>57</b>

# Capitolo 1

## Introduzione

Le seguenti analisi fanno riferimento a uno studio svolto presso la clinica veterinaria San Marco, a Veggiano (PD). Il campione preso in esame è composto da 45 cani, ognuno dei quali è stato sottoposto, tra Agosto 2017 e Marzo 2022, a due tipologie differenti di tomografia computerizzata dual-energy (DECT) addomino-pelvica senza mezzo di contrasto, una secondo l'approccio tradizionale (PC) e una virtuale (VNC). Con entrambe le TC (o CT in inglese) sono stati rilevati i livelli di radiodensità tramite la cosiddetta scala di Hounsfield su nove regioni d'interesse (ROI, Regions Of Interest), quali ossa, muscoli, grasso, corticale renale, aorta, vena porta, milza, cistifellea e fegato.

Obiettivo dello studio è valutare le immagini virtuali dell'addome dei soggetti, frutto di una tomografia computerizzata dual-energy senza mezzo di contrasto, e confrontarle, secondo i valori HU rilevati, con quelle reali analogamente scannerizzate. Si studiano, inoltre, come la generazione di DECT, l'età, il peso e lo stato sessuale dei soggetti presi in esame, influenzino le differenze tra tomografia tradizionale e virtuale.

Le variabili prese in esame sono:

- **Bone – PC/VNC:** variabili quantitative continue che indicano il valore HU rilevato sulle ossa addominali con DECT tradizionale e virtuale, rispettivamente;
- **Muscle – PC/VNC:** variabili quantitative continue che indicano il valore HU rilevato sui muscoli addominali con DECT tradizionale e virtuale, rispettivamente;
- **Fat – PC/VNC:** variabili quantitative continue che indicano il valore HU rilevato sul grasso addominale sottocutaneo con DECT tradizionale e virtuale, rispettivamente;
- **Renal Cortex – PC/VNC:** variabili quantitative continue che indicano il valore HU rilevato sulla corticale renale con DECT tradizionale e virtuale, rispettivamente;
- **Aorta – PC/VNC:** variabili quantitative continue che indicano il valore HU rilevato sulla aorta addominale con DECT tradizionale e virtuale, rispettivamente;
- **Portal Vein – PC/VNC:** variabili quantitative continue che indicano il valore HU rilevato sulla vena porta con DECT tradizionale e virtuale, rispettivamente;
- **Spleen – PC/VNC:** variabili quantitative continue che indicano il valore HU rilevato sulla milza con DECT tradizionale e virtuale, rispettivamente;
- **Gallbladder – PC/VNC:** variabili quantitative continue che indicano il valore HU rilevato sulla cistifellea con DECT tradizionale e virtuale, rispettivamente;

- **Liver – PC/VNC:** variabili quantitative continue che indicano il valore HU rilevato sul fegato con DECT tradizionale e virtuale, rispettivamente;
- **Age:** variabile quantitativa continua che descrive l'età del soggetto in anni;
- **Weight:** variabile quantitativa continua che descrive il peso del soggetto in chilogrammi;
- **Sexual Status:** variabile categoriale a 4 livelli che descrive il sesso del soggetto e se è stato sottoposto a sterilizzazione (*Non-Neutered Male, Neutered Male, Non-Spayed Female, Spayed Female*);
- **CT:** variabile qualitativa dicotomica che indica la generazione di DECT che si è utilizzata per il soggetto (*Flash, Force*).

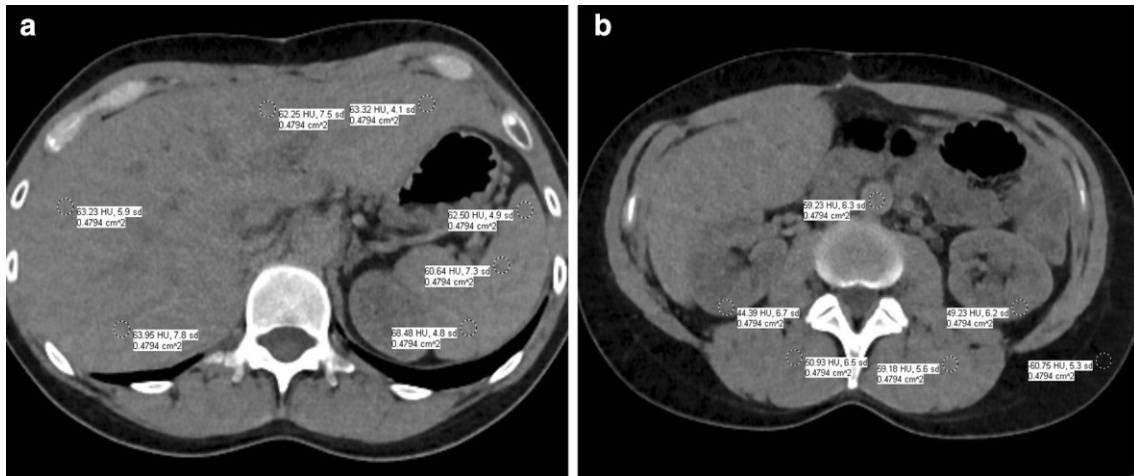
La tomografia computerizzata a doppia energia (DECT), detta anche a doppia immagine, inizialmente concepita da Alvarez e Macovski negli anni '70, è una particolare tecnologia che permette di acquisire le immagini fornendosi di due livelli di energia differenti. Ciò permette di sfruttare la decomposizione materiale, grazie alla quale è possibile generare immagini sulla base di determinati materiali, come ad esempio costruire mappe iodio-specifiche (Siegel *et al.*, 2021). Questo concetto è alla base della tomografia computerizzata virtuale. Infatti, la VNC è una tecnica post processing che consiste, in primo luogo, di acquisire immagini attraverso una DECT con mezzo di contrasto e, in un secondo momento, sottrarre da queste il segnale di iodio, elemento caratterizzante il liquido di contrasto, in modo tale da ottenere scansioni simili a quelle ottenute con la tomografia senza mezzo di contrasto tradizionale (Javadi *et al.*, 2020).

La VNC, nonostante si tratti di una tecnologia in via di sperimentazione che dimostra ancora esiti differenti a seconda del campo di applicazione, presenta svariati vantaggi in molti scenari clinici. Infatti, sopra a tutti, permetterebbe di ottenere le medesime immagini della TC senza contrasto tradizionale riducendo però l'esposizione radioattiva del paziente consentendogli inoltre di eseguire sessioni di radiografia notevolmente più ridotte in termini di durata, caratteristica interessante soprattutto per i soggetti che si devono sottoporre spesso a questo genere di analisi, come i pazienti oncologici (Virarkar *et al.*, 2022).

Strumento utilizzato per valutare le immagini acquisite con la tomografia è la scala di Hounsfield. Anche questa si basa sui principi della decomposizione materiale, per cui la densità fisica dei corpi contribuisce all'assorbimento o all'attenuazione dei raggi X (Javadi *et al.*, 2020). Infatti, l'unità di Hounsfield (HU), scoperta da Sir Godfrey Hounsfield (1979), viene utilizzata per quantificare il livello di radiodensità dei tessuti in esame, ovvero si tratta di una misura relativa che valuta l'assorbimento, o l'attenuazione, delle radiazioni da parte dei tessuti corporei. Nello specifico, l'HU rappresenta il discostamento, in termini di radiodensità, dal livello base rappresentato dall'acqua, distillata e a livelli standard di temperatura e pressione, che assume un valore di 0 HU. Gli altri tipi di materia assumono valori negativi se si presentano meno densi e quindi con un assorbimento minore di raggi X, come ad esempio l'aria con un valore pari a -1000 HU (valore minimo). Mentre valori positivi della scala di Hounsfield, come per le ossa, con un range da 700 a 2000 unità, o per i metalli che assumono valori intorno ai 3000 HU, indicano la presenza di corpi più densi e quindi con un

assorbimento maggiore di raggi X. La scala Hounsfield viene utilizzata dai radiologi per interpretare le immagini rilevate con tomografia computerizzata poiché a livelli di HU minori corrispondono scale di grigio più scure, mentre a valori più alti di radiodensità equivalgono grigi più chiari. Infatti, nelle immagini ricavate attraverso CT, il bianco solitamente corrisponde alle ossa più dense, come la coclea, mentre il nero rappresenta l'aria (DenOtter e Schubert, 2022), come si può osservare nella figura 1.1, dove sono presentate le immagini risultanti da una TC addominale senza contrasto tradizionale, a sinistra, e virtuale, a destra.

Figura 1.1. Scansioni addominali effettuate tramite tomografia computerizzata senza contrasto tradizionale (a) e virtuale (b), con relativi valori di HU per alcune regioni d'interesse (Javadi *et al.*, 2020).



## Capitolo 2

# Analisi esplorativa dei dati

Le analisi e i grafici presentati in questo studio sono stati effettuati principalmente per mezzo del software statistico R (versione 4.2.2), con il supporto dei software SAS® OnDemand for Academics e Microsoft® Excel®. Il livello di significatività è fissato al 5% per tutte le seguenti analisi.

### 2.1 Il campione

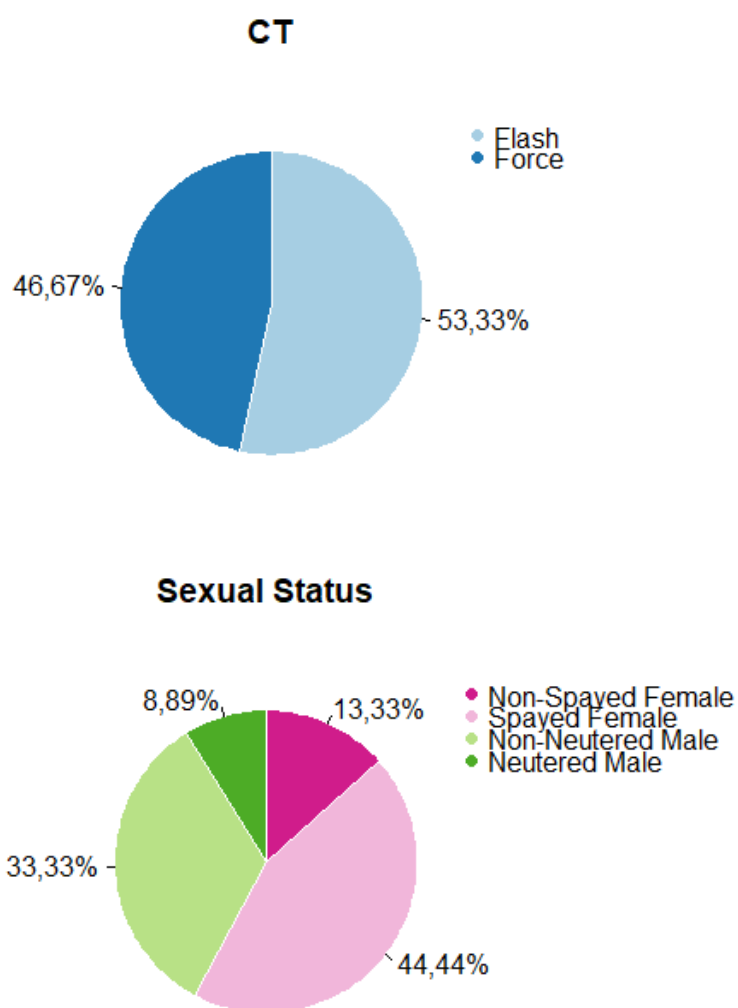
Il campione preso in esame, come già anticipato, è composto da 45 soggetti di razza canina che nell'arco temporale dello studio, ovvero tra agosto 2017 e marzo 2022, possedevano un'età media pari a 8 anni circa (range: 0.08–16.00) e un peso medio uguale a 19.00 chilogrammi (range: 2.20 – 41.40), come viene mostrato nella tabella 2.1. Oltre a peso ed età, sono stati raccolti anche i dati relativi allo stato sessuale dei soggetti. Il campione preso in esame è composto quasi per il 58% da femmine, di cui il 77% (44% del totale) è sterilizzato (*spayed* in inglese), mentre gli individui di sesso maschile costituiscono pressoché il restante 42% del campione, di cui circa il 79% (33% del totale) è sterilizzato (*neutered* in inglese). I soggetti, come precedentemente esposto, sono stati sottoposti a due tomografie, per le quali sono stati utilizzate tecnologie di seconda generazione (*Flash*) per circa il 53% del campione, mentre per il restante 47% si è ricorso a tecniche di terza generazione (*Force*).



Tabella 2.1. Principali statistiche di sintesi delle variabili **Age** e **Weight**.

	Minimo	Media	Mediana	Massimo	IQR	Deviazione Standard
<i>Age</i>	0.08	<b>8.22</b>	9.00	16.00	4.00	3.81
<i>Weight</i>	2.20	<b>19.00</b>	19.70	41.40	21.80	11.70

Figura 2.1. Grafici a torta con distribuzione percentuale delle variabili **CT** e **Sexual Status**.



## 2.2 Analisi esplorativa

Le analisi esplorative si focalizzano sulle variabili di maggiore interesse, ovvero quelle relative alle unità di Hounsfield rilevate attraverso le due tomografie in esame, la PC e la VNC, nelle nove regioni d'interesse. Tali variabili, vengono studiate prima singolarmente attraverso analisi univariate, poi, in un secondo momento, si effettuano le analisi bivariate in relazione alle altre variabili incluse nello studio, studiandone principalmente la forza delle associazioni e il grado di accordo.

Come già anticipato, nel presente caso di studio si dispone di dati appaiati, ovvero significa che gli stessi soggetti vengono osservati in due situazioni cliniche differenti, e sui quali viene rilevato il valore che la medesima variabile assume in entrambe le situazioni. Si tratta, pertanto, di uno scenario in cui sulla stessa unità statistica vengono effettuate delle misure ripetute (Racugno e Ventura, 2017). Infatti, come già precedentemente esposto, per il campione preso in considerazione vengono rilevate le HU per entrambe le tomografie eseguite, su ognuna delle nove regioni d'interesse, su tutti e 45 i cani facenti parte il campione. Trattandosi di dati appaiati, si ritiene opportuno studiare il comportamento delle differenze tra TC tradizionale e virtuale su ogni regione presa in esame. Si riportano quindi le principali statistiche di sintesi nella tabella 2.2 e si studia, tramite il test di Shapiro-Wilk, la normalità delle suddette differenze.

Per completezza, nella tabella A.1 e nelle figure A.1 e A.2 dell'*Appendice*, vengono presentate le analisi esplorative, ovvero le principali statistiche di sintesi e lo studio delle distribuzioni delle variabili PC e VNC per ogni ROI separatamente, quindi senza considerare la loro differenza.

Tabella 2.2. Principali statistiche di sintesi per le differenze dei valori di HU tra PC e VNC, per ogni ROI.

Variabili	Minimo	Media	Mediana	Massimo	IQR	Deviazione Standard	Asimmetria	Test di Shapiro-Wilk	
								Valore osservato	p-value
<i>Bone</i>	-101.6	<b>205.8</b>	200.0	470.3	153.7	125.5	0.19	0.98	0.570
<i>Muscle</i>	-1.5	<b>9.0</b>	9.4	25.4	8.1	5.8	0.31	0.98	0.512
<i>Fat</i>	-21.9	<b>-6.6</b>	-7.7	14.6	8.1	7.3	0.69	0.96	0.104
<i>Renal Cortex</i>	-16.2	<b>3.2</b>	3.4	23.6	14.0	8.2	-0.02	0.98	0.706
<i>Aorta</i>	-17.1	<b>1.5</b>	1.7	16.5	12.7	9.0	-0.21	0.97	0.328
<i>Portal Vein</i>	-16.6	<b>0.7</b>	-0.3	16.7	9.9	7.4	0.16	0.98	0.808
<i>Spleen</i>	-11.4	<b>1.5</b>	0.3	17.3	9.3	6.6	0.14	0.98	0.818
<i>Gallbladder</i>	-11.5	<b>5.4</b>	5.5	18.5	6.5	5.8	-0.71	0.95	<b>0.040</b>
<i>Liver</i>	-12.9	<b>1.4</b>	1.0	14.9	8.1	5.7	0.06	0.98	0.740

Si può notare come le differenze medie siano tutte positive, ad eccezione della variabile **Fat** con un valore pari a -6.6, dimostrando che il VNC presenta valori in media minori rispetto all'approccio tradizionale. In particolare, per la variabile **Bone**, la differenza media pari a 205.8 si presenta distante dallo zero, portando a pensare ci sia una differenza significativa tra le due tomografie, ipotesi che verrà approfondita nel capitolo 4. Le altre variabili, invece, non si discostano molto dal valore nullo. Infatti, il range per le differenze medie, escludendo le ossa, si estende da un valore minimo pari a 0.7 a un valore massimo uguale a 9.0. Osservando il comportamento delle distribuzioni, si nota come le differenze siano abbastanza centrate con valori dell'indice di asimmetria prossimi a zero, e la normalità venga ampiamente accettata per tutte le regioni d'interesse, dati i p-value del test di Shapiro-Wilk superiori a 0.1, raggiungendo anche valori elevati intorno allo 0.8, come nel caso delle variabili **Portal Vein** e **Spleen**.

Costituisce un'eccezione la variabile **Gallbladder**, la quale presenta un p-value inferiore al livello di significatività fissato. Ciononostante, si è deciso di non rifiutare l'ipotesi di normalità per questo sito d'interesse, poiché, in prima istanza, un valore di 0.04 di poco minore del livello fissato al 5%, in secondo luogo, la numerosità campionaria sufficiente permette di concludere in questa direzione, e infine, la cistifellea è l'unica ROI per cui la normalità è stata dubitata.

Le precedenti osservazioni, possono essere verificate anche attraverso lo studio dei boxplot nella figura 2.2 e dei QQ-plot nella figura 2.3.

Figura 2.2. Boxplot delle differenze di HU tra PC e VNC, per ogni ROI.

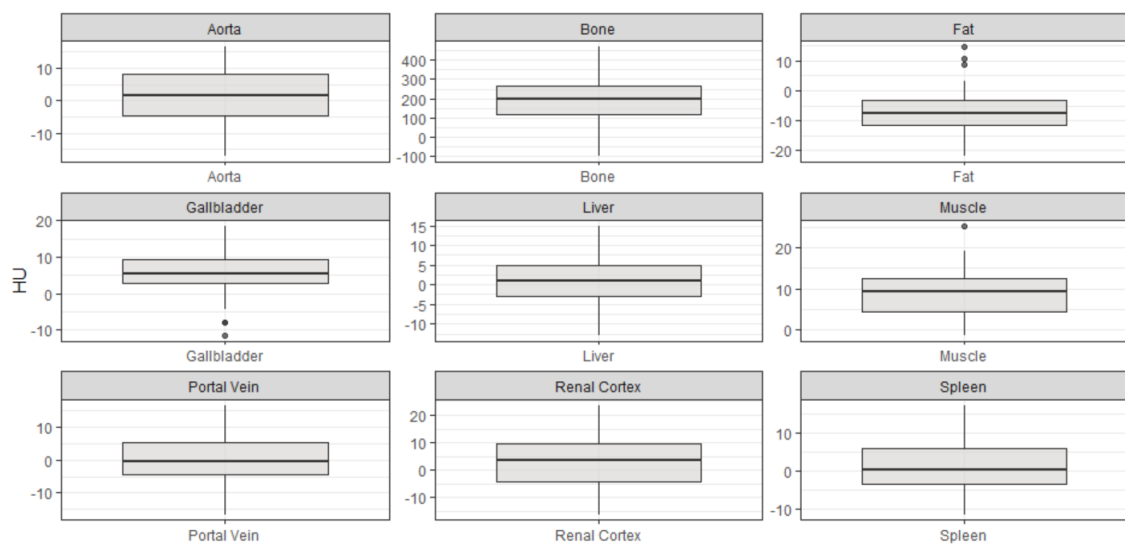
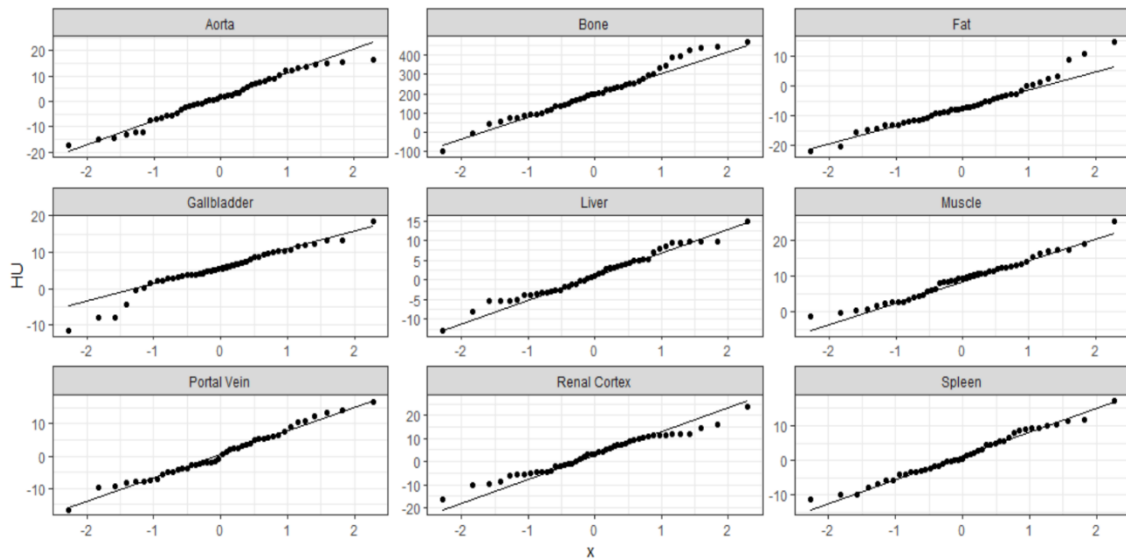


Figura 2.3. QQ-plot delle differenze di HU tra PC e VNC, per ogni ROI.

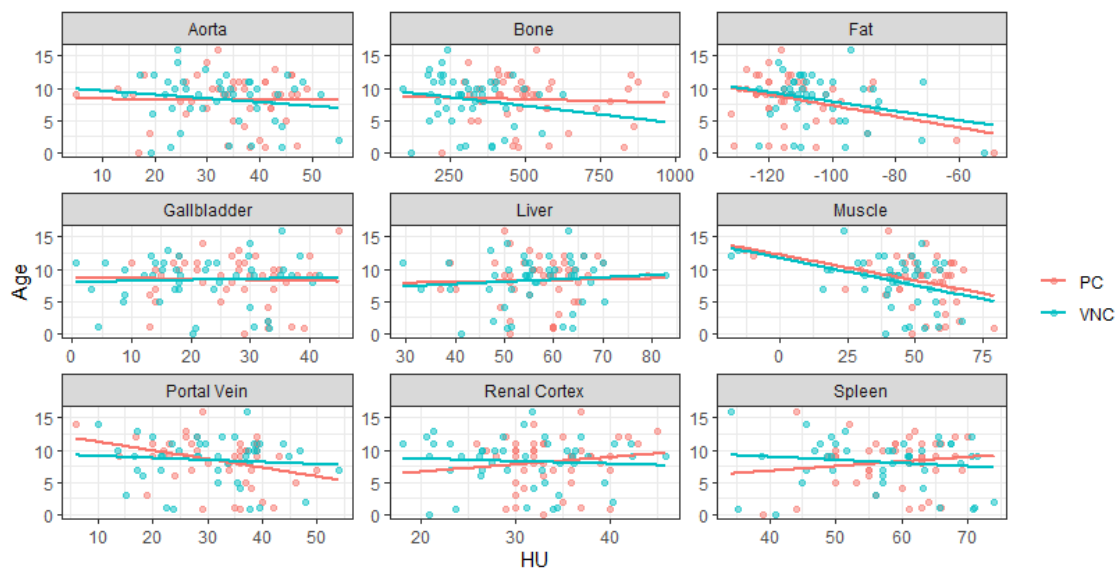


Si analizzano ora le variabili relative agli HU delle regioni d'interesse, in relazione alle altre variabili incluse nello studio. Si osservano inizialmente le variabili quantitative, ovvero **Age** e **Weight**. Per queste si sono innanzitutto studiati i diagrammi di dispersione e, in seguito, si sono calcolati i coefficienti di correlazione di Spearman, con il corrispondente p-value per valutarne la significatività (Hollander *et al.*, 1973). In questo caso, non si sono utilizzate le differenze, ma sono stati analizzati i valori HU per PC e VNC separatamente in modo tale da poter rilevare più facilmente il senso dell'associazione tra le due variabili prese in considerazione.

Nelle figure 2.4 e 2.5 sono presentati i scatterplot per ogni regione d'interesse, ovvero vengono rappresentati i valori HU osservati con tomografia tradizionale in rosso, e con quella virtuale in blu. Inoltre, vengono stimati i rispettivi trend lineari, rappresentati con delle rette ai minimi quadrati, per avere un'idea più precisa della forza e, soprattutto, della direzione della correlazione tra l'età o il peso dei soggetti in relazione alle unità di Hounsfield rilevate.

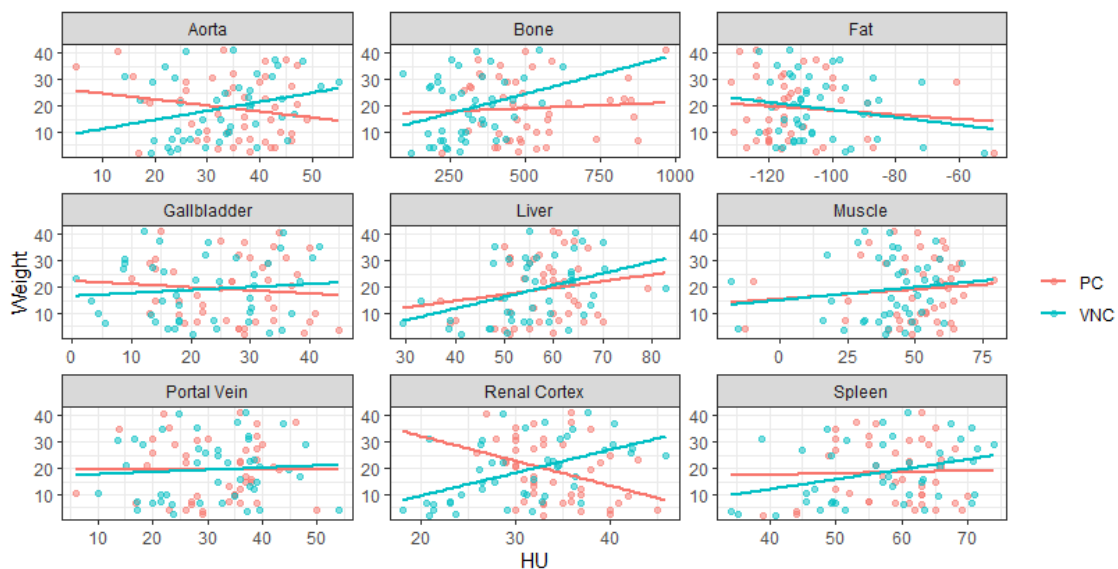
Per quanto riguarda l'età, si nota che non esiste una forte correlazione con i livelli di radiodensità. Infatti, le rette sono pressoché tutte orizzontali e non dimostrano un andamento specifico. Le uniche variabili per cui si coglie una leggera correlazione negativa, ovvero per le quali a soggetti più giovani corrispondono valori di HU maggiori e viceversa, sono quelle relative a muscoli e grasso addominali. Inoltre, non si colgono differenze rilevanti tra PC e VNC in nessuna delle regioni d'interesse.

Figura 2.4. Scatterplot e rette ai minimi quadrati tra HU ed età rilevati per tutte le ROI, divise per tipologia di CT.



Per il peso, invece, si evidenzia come le variabili **Bone**, **Renal Cortex** e **Aorta**, mostrino correlazione differenti condizionatamente alla modalità di rilevazione degli HU. Infatti, se le unità statistiche sono state sottoposte a TC tradizionale, la radiodensità è correlata negativamente al peso, presentando valori maggiori a pesi minori; mentre gli HU per la tomografia virtuale risultano maggiori se il relativo soggetto presenta un peso più elevato. Le altre variabili non solo non mostrano sostanziali differenze tra le due tecnologie, ma anche non sembrano essere particolarmente correlate con il peso, tranne per leggere inclinazioni delle rette lineari come nel caso di **Liver** e **Fat**, le quali presentano una lieve associazione positiva e negativa, rispettivamente.

Figura 2.5. Scatterplot e rette ai minimi quadrati tra HU e peso rilevati per tutte le ROI, divise per tipologia di CT.



Di seguito, nella tabella 2.3, vengono presentati i coefficienti di correlazione  $\rho$  di Spearman, affiancati dai p-value corrispondenti, sia per **Age** che per **Weight**.

Tabella 2.3. Coefficiente di correlazione  $\rho$  di Spearman tra le HU rilevate su ogni ROI da entrambe le tomografie, e le variabili **Age** e **Weight**, con il corrispondente p-value.

Variabili		Correlazione di Spearman			
		Age		Weight	
		$\rho$	p-value	$\rho$	p-value
<i>Bone</i>	PC	-0.09	0.56	-0.01	0.97
	VNC	-0.18	0.19	0.26	0.09
<i>Muscle</i>	PC	-0.32	0.03	0.09	0.55
	VNC	-0.30	0.05	0.13	0.38
<i>Fat</i>	PC	-0.23	0.14	-0.07	0.66
	VNC	-0.15	0.32	-0.12	0.40
<i>Renal Cortex</i>	PC	0.09	0.57	-0.38	0.01
	VNC	-0.10	0.53	0.51	< 0.001
<i>Aorta</i>	PC	-0.02	0.90	-0.12	0.42
	VNC	-0.14	0.37	0.31	0.04
<i>Portal Vein</i>	PC	-0.30	0.05	0.05	0.74
	VNC	-0.07	0.63	0.06	0.71
<i>Spleen</i>	PC	0.11	0.47	-0.04	0.81
	VNC	-0.09	0.55	0.32	0.03
<i>Gallbladder</i>	PC	-0.07	0.64	-0.08	0.59
	VNC	0.01	0.95	0.07	0.65
<i>Liver</i>	PC	0.05	0.72	0.15	0.32
	VNC	0.08	0.59	0.42	0.01



Dalla tabella 2.3 si può concludere che i livelli di HU in tutte le aree d'interesse addominali non dipendano dall'età del soggetto, ad eccezione della variabile **Muscle**, come era già stato osservato analizzando i grafici di dispersione della figura 2.4. Infatti, i muscoli hanno una correlazione negativa in confronto alla scala Hounsfield, sia per PC che VNC, con valori del  $\rho$  pari a -0.3 e p-value uguali a 0.03 e 0.05, rispettivamente. Dalla figura 2.4, si era ipotizzato che anche per la variabile **Fat** l'età potesse essere correlata con la radiodensità in modo significativo, ma osservando la tabella 2.3, tale deduzione viene confutata, portando alla conclusione che i valori di HU non dipendano dall'età per quanto concerne il grasso addominale.

Invece, lo studio della correlazione tra HU e **Weight** nelle varie ROI fa emergere come per la corticale renale quest'associazione sia fortemente significativa per entrambe le tomografie, con p-value minori o uguali a 0.01. L'esito interessante, come si era già osservato dalla figura 2.5, è il comportamento inverso tra PC e VNC in questa regione d'interesse, per la quale soggetti di peso maggiore presentano valori della scala di Hounsfield più elevati se rilevati con l'approccio tradizionale, mentre quello virtuale acquisisce HU maggiori per unità statistiche di peso minore. Interessante notare, per giunta, come i p-value significativi siano tutti in corrispondenza della tomografia virtuale come nel caso delle variabili relative all'aorta, milza e fegato che hanno p-value pari a 0.04, 0.03 e 0.04. Ma anche tutte le altre ROI presentano il p-value concernente la tomografia sperimentale inferiore a quello della TC tradizionale. Ciò porta a pensare che la VNC sia più sensibile a variazioni di peso, per tutte i siti d'interesse dello studio.

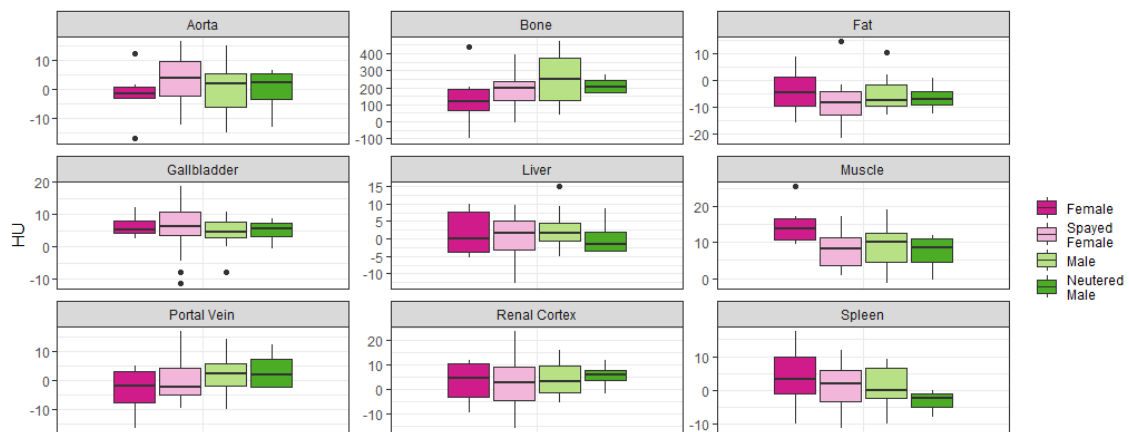
Si procede, quindi, con lo studio delle differenze tra tipologia di tomografia condizionatamente allo stato sessuale del soggetto in esame. Pertanto, vengono presentate le relative medie e deviazioni standard nella tabella 2.4 accompagnate da un'analisi grafica dei boxplot della figura 2.6.

Studiando la tabella 2.4 e i boxplot in figura 2.6, si può evincere come per tutte le regioni d'interesse le differenze di HU tra i livelli della variabile **Sexual Status** non siano elevate. Infatti, i boxplot sembrano essere tutti e quattro intorno al medesimo valore HU. Appaiono però opinabili i risultati relativi alle variabili concernenti le ossa, i muscoli e la vena porta. Mentre per la prima, il gruppo dei maschi con una media pari a 250 circa risulta molto elevato rispetto alle altre modalità, specialmente se confrontato con le femmine non sterilizzate che presentano un valore medio pari a 140 rispetto agli altri tre, per quanto riguarda, invece, la variabile relativa ai muscoli, il gruppo delle femmine non sterilizzate, con un valore medio pari a circa 15, è maggiore degli altri stati sessuali in modo rilevante, i quali presentano medie con un range tra 7 e 9 HU. Infine, la vena porta presenta anch'essa un gruppo di soggetti femminili non sterilizzati su cui vengono rilevati livelli della scala di Hounsfield nettamente minori rispetto agli altri gruppi. Infatti, se per la modalità *Non-Spayed Female* la media pari a quasi -6, gli altri stati sessuali hanno medie che si aggirano tra lo 0 e il 3.5.

Tabella 2.4. Medie e deviazioni standard per le differenze di HU tra PC e VNC, per tutte le ROI, suddivise per stato sessuale del soggetto.

	Non-Spayed Female		Spayed Female		Non-Neutered Male		Neutered Male	
	Media	Deviazione Standard	Media	Deviazione Standard	Media	Deviazione Standard	Media	Deviazione Standard
<i>Bone</i>	140.38	179.45	191.94	95.76	248.79	144.21	211.90	54.52
<i>Muscle</i>	14.92	5.86	7.82	5.05	8.83	5.92	7.10	5.58
<i>Fat</i>	-4.03	9.01	-8.37	7.79	-5.47	6.37	-6.30	5.64
<i>Renal Cortex</i>	3.05	8.91	2.31	9.71	3.74	6.76	5.55	5.57
<i>Aorta</i>	-1.57	9.49	3.67	8.93	0.31	9.11	-0.40	8.79
<i>Portal Vein</i>	-3.58	8.81	0.09	7.62	2.13	6.64	3.38	7.15
<i>Spleen</i>	3.90	9.61	1.56	6.51	1.35	5.88	-3.37	4.08
<i>Gallbladder</i>	6.22	3.63	5.84	7.36	4.54	4.80	4.78	3.95
<i>Liver</i>	1.62	6.78	0.76	6.13	2.38	5.00	0.33	5.76

Figura 2.6. Boxplot delle differenze di HU tra le due tomografie per le ROI, divise per stato sessuale.



Si procede con un'analisi della varianza a una via (ANOVA), ovvero vengono confrontate tramite un test F le medie tra i quattro gruppi relativi allo stato sessuale del soggetto e si verifica se queste siano significativamente differenti tra loro. Per poter procedere con tale analisi, è necessario prima verificare la normalità e l'omoschedasticità delle variabili in esame. La normalità è già stata precedentemente accettata attraverso un test di Shapiro-Wilk, i cui risultati sono presentati nella tabella 2.2. Lo studio dell'omoschedasticità, ovvero la verifica dell'uguaglianza delle varianze, è stata eseguita attraverso un test di Bartlett e l'ipotesi nulla non è stata rifiutata per nessuna delle nove variabili in esame. Tali risultati, cioè il valore osservato del test e il relativo p-value, sono presentati nella tabella A.2 dell'*Appendice*. Verificate normalità e omoschedasticità, si è potuto proseguire con l'ANOVA, i cui esiti sono esposti nella tabella 2.5.

Rispetto alle altre variabili osservate precedentemente, quella relativa allo stato sessuale del soggetto, non mostra differenze tra le relative modalità, portando alla conclusione che i valori delle differenze di HU tra le due tomografie, non differiscano in modo statisticamente rilevante per nessun stato sessuale. Unicamente la variabile **Muscle** presenta il p-value, pari a 0.053, prossimo al livello di significatività fissato, seppur maggiore del 5%. Per tale motivo, si è deciso di procedere comunque con un'analisi post-hoc, per identificare quali coppie di modalità portassero ad avere un valore del p-value così inferiore rispetto a quelli relativi alle altre variabili, tutti superiori a 0.30. Di conseguenza, attraverso un confronto a coppie tramite test t, si scopre che i gruppi di **Sexual Status** che più differiscono tra loro sono quelli delle femmine. Infatti, il t-test produce un p-value pari allo 0.05 in corrispondenza della coppia *Non-Spayed Female* e *Spayed Female*.

Tabella 2.5. Valore osservato del test ANOVA per le differenze di HU tra le due tomografie su ogni ROI, condizionatamente alla variabile **Sexual Status**, con relativo p-value. Gradi di libertà pari a 3.

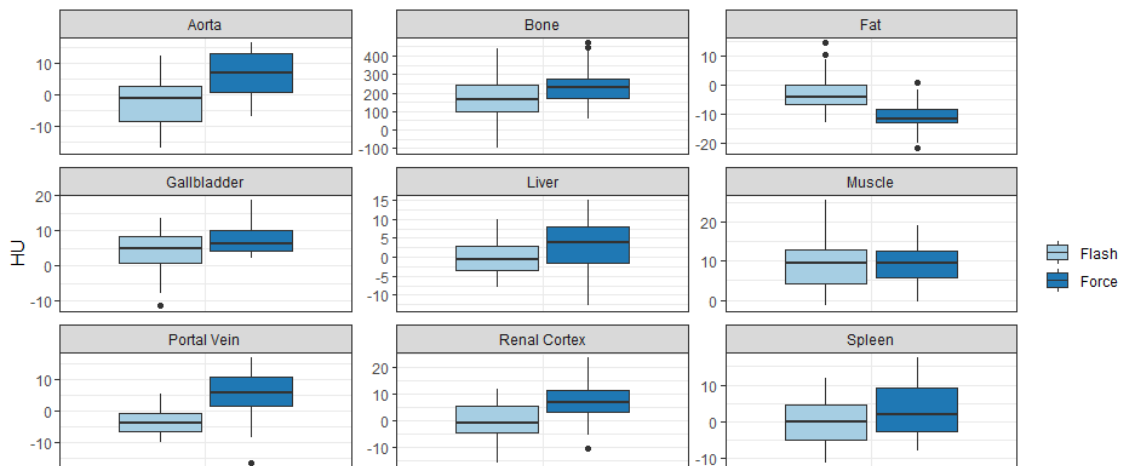
Variabili	ANOVA	
	Valore osservato F	p-value
<i>Bone</i>	1.23	0.310
<i>Muscle</i>	2.78	0.053
<i>Fat</i>	0.75	0.530
<i>Renal Cortex</i>	0.20	0.896
<i>Aorta</i>	0.76	0.525
<i>Portal Vein</i>	0.97	0.418
<i>Spleen</i>	0.79	0.507
<i>Gallbladder</i>	0.18	0.907
<i>Liver</i>	0.27	0.847

Si prosegue, infine, con lo studio degli HU relativamente alla generazione di TC utilizzato per effettuare le scansioni. In questo caso, si ritorna a utilizzare la differenza tra le due tomografie per ogni regione d'interesse e vengono presentate nella tabella 2.6 le relative media e deviazione standard, sia per le tecnologie di seconda generazione, chiamate *Flash*, e quelle di terza, dette *Force*. In aggiunta, nella figura 2.7 sono presentati i boxplot per studiare visivamente le differenze tra *Flash* (in azzurro) e *Force* (in blu). Da questi ultimi, si può notare come per tutte le regioni d'interesse, i valori della scala Hounsfield rilevati dalla tomografia di terza generazione siano maggiori rispetto a quella di seconda. In particolare, tale differenza sembra ampia per la corticale renale e per i vasi sanguigni, quali l'aorta e la vena porta. Unica eccezione è costituita dalla variabile **Fat**, per la quale le TC Force rilevano unità di Hounsfield nettamente inferiori a quelle analogamente acquisite dalla tecnologia Flash.

Tabella 2.6. Medie e deviazioni standard per le differenze di HU tra PC e VNC, per tutte le ROI, suddivise per generazione di CT.

Variabili	Flash		Force	
	Media	Deviazione Standard	Media	Deviazione Standard
<i>Bone</i>	180.21	130.10	235.02	116.13
<i>Muscle</i>	9.26	6.42	8.78	5.20
<i>Fat</i>	-2.77	6.67	-11.06	5.24
<i>Renal Cortex</i>	0.18	7.49	6.60	7.79
<i>Aorta</i>	-2.70	8.18	6.27	7.44
<i>Portal Vein</i>	-2.99	4.55	4.66	7.89
<i>Spleen</i>	-0.30	6.06	3.32	6.85
<i>Gallbladder</i>	3.85	6.81	7.07	4.04
<i>Liver</i>	-0.14	4.81	3.10	6.20

Figura 2.7. Boxplot delle differenze di HU tra le due tomografie per le ROI, divise per generazione di CT.



Per studiare in modo più approfondito se le differenze tra le due generazioni siano significativamente differenti, si procede con un test t (o *t-test*) per due campioni indipendenti, si tratta cioè di un test parametrico, verificata la normalità per le differenze, che consiste nel confrontare, per ogni ROI, le medie delle differenze nei due gruppi di **CT** e quindi studiare se siano differenti in modo significativo. Il valore osservato e il p-value del test t sono presentati nella tabella 2.7.

Tabella 2.7. Valore osservato del test t a due campioni indipendenti tra le differenze di HU tra le due tomografie su ogni ROI, condizionatamente alla variabile **CT**, con relativo p-value.

Variabili	Test t	
	Valore osservato	p-value
<i>Bone</i>	-1.49	0.143
<i>Muscle</i>	0.28	0.783
<i>Fat</i>	4.66	< 0.001
<i>Renal Cortex</i>	-2.81	0.008
<i>Aorta</i>	-3.85	< 0.001
<i>Portal Vein</i>	-3.89	< 0.001
<i>Spleen</i>	-1.83	0.074
<i>Gallbladder</i>	-1.93	0.062
<i>Liver</i>	-1.94	0.060

Osservando la tabella 2.7, viene confermato quanto precedentemente ipotizzato attraverso lo studio dei boxplot della figura 2.7. Infatti, si nota come le misurazioni effettuate siano significativamente differenti per le variabili **Fat, Aorta e Portal Vein** con p-value inferiori a 0.001 e per **Renal Cortex** con un p-value pari a 0.008. Come visto in precedenza, per le ultime tre regioni, la generazione Force presenta HU inferiori a Flash, mentre per **Fat** accade il contrario. Infatti, il valore osservato del t-test è positivo per quanto concerne il grasso, mentre risulta negativo per la corticale renale e specialmente per i vasi sanguigni considerati. I p-value del presente test, per le variabili **Spleen, Gallbladder e Liver**, seppur maggiori del 5%, sono prossimi a tale valore, indicando, dati i relativi valori osservati vicini al valore -2, che anche per queste regioni d'interesse la terza generazione produce valori di radiodensità più elevati rispetto alla seconda. Infine, non sembrano esserci differenze tra generazioni di TC per quanto riguarda i muscoli e le ossa.



## 2.3 Conclusioni

Attraverso le analisi eseguite nel presente capitolo, si sono studiate le relazioni tra i valori di HU rilevati con le due tomografie in esame e specifiche caratteristiche del soggetto in studio. Si è cercato di capire, pertanto, se una data variabile potesse comportare cambiamenti significativi nella scala di Hounsfield.

Verificate in primo luogo normalità e omoschedasticità per le differenze di radiodensità tra PC e VNC, si è evinto che l'età del soggetto influisce esclusivamente sulla densità dei muscoli, senza portare variazioni in tal senso significative sulle altre regioni d'interesse. Si osservato, inoltre, che il peso condiziona i valori di HU sulla corticale renale, per la quale sono presenti anche differenze importanti tra le due TC, una delle quali, quella virtuale, dimostra di essere particolarmente sensibile a cambiamenti di peso dell'individuo, non solo per i reni, ma anche per le altre regioni d'interesse. Proseguendo con le analisi, studiando quindi le relazioni con la generazione di tomografia computerizzata utilizzata, si è visto che con la tecnologia Force, la PC rileva valori di HU significativamente superiori a VNC per i vasi sanguigni studiati, aorta e vena porta, e per la corticale renale. Allo stesso modo, PC acquisisce sul grasso addominale unità di Hounsfield maggiori se viene utilizzata la tecnologia Flash. Infine, si è visto come lo stato sessuale del soggetto non porti a differenze rilevanti tra la tomografia tradizionale e quella virtuale.

## Capitolo 3

# Misure di agreement

Nel presente capitolo s'intende studiare il grado di accordo tra le due tomografie, centro di questo studio, quella tradizionale (PC) e quella virtuale (VNC). Ovvero, una volta osservati nel capitolo precedente i loro comportamenti condizionatamente ad altre variabili, ora si vuole analizzare, a parità di caratteristiche dell'unità statistica, se la tomografia virtuale possa essere un valido sostituto a quella tradizionale misurandone l'accordo delle rilevazioni. S'intende pertanto, attraverso opportune verifiche grafiche, come il grafico di Bland-Altman, e specifiche quantità calcolate, come il coefficiente di concordanza di Lin, studiare la performance della TC virtuale in relazione al cosiddetto gold standard, costituito dalla PC, ovvero verificare la loro correlazione aggiungendo un'ulteriore informazione riguardo il loro grado di accordo.

### 3.1 Bland-Altman

Per valutare l'accordo tra le misurazioni effettuate con le due tomografie tramite verifica grafica, si può utilizzare il grafico di Bland-Altman. Questo diagramma consente di verificare se le due tecniche di rilevazione forniscano misure tra esse comparabili o, in altri termini, se le differenze osservate, comunque inevitabili, siano casuali o frutto di un errore sistematico.

Allo stesso tempo, consente anche di valutare se la tecnica sperimentale, in questo caso la TC virtuale, sottostimi o sovrastimi le misure rispetto alla tomografia tradizionale.

Il grafico di Bland-Altman, proposto circa trent'anni fa da M. J. Bland e D. G. Altman, nel corso degli anni è diventato il metodo più utilizzato (e appropriato) in campo medico-scientifico per valutare l'accordo, insieme ai relativi limiti, tra due tecniche di misurazione (Doğan, 2018). Si tratta, infatti, di un diagramma di dispersione dove in ascissa sono riportate la media aritmetica dei livelli di HU per ogni soggetto del campione, mentre in ordinata le differenze tra le misurazioni effettuate con le due diverse strumentazioni. Inoltre, viene calcolato il bias, ovvero la media aritmetica di tutte le suddette differenze, assieme al suo intervallo di confidenza al livello 95%. Il bias risulta significativo, è presente cioè un errore sistematico, quando il valore 0 non cade all'interno del relativo intervallo di confidenza. In assenza, invece, di errore sistematico, i punti all'interno del diagramma si dispongono intorno alla linea dello zero (Racugno e Ventura, 2017).

L'ampiezza degli intervalli di confidenza delle differenze (*LOA: Limits of Agreement*, in inglese) è pari a due volte il quantile della distribuzione normale ( $z_{1-\frac{\alpha}{2}}$ ) moltiplicato per le deviazioni standard  $\sigma$  delle differenze tra le misurazioni. Il limite inferiore degli intervalli di confidenza sarà quindi uguale a ( $bias - z_{1-\frac{\alpha}{2}} * \sigma$ ), mentre il limite superiore risulterà pari a ( $bias + z_{1-\frac{\alpha}{2}} * \sigma$ ). Da qui, si evince che il grafico di Bland-Altman si basa sull'assunzione di normalità delle informazioni che rappresenta.

Pertanto, non è necessario che le misurazioni siano distribuite normalmente, ma affinché tale diagramma risulti affidabile, le differenze, invece,

lo devono essere (Bland e Altman, 1986). Nel presente caso di studio, quindi, risulta non solo opportuno utilizzare il grafico di Bland-Altman, verificata l'assunzione di normalità delle differenze nel capitolo precedente, ma costituisce anche un ottimo strumento grafico per studiare il grado di accordo tra le tomografie.

Nello specifico, per questo caso di studio, si fa uso della soglia zero e degli intervalli di confidenza, e, in aggiunta, vengono anche stabiliti degli ulteriori limiti di accordo (*LOA*) pari a -10 e 10. Questi vengono chiamati limiti di equivalenza, il cui studio più approfondito si rimanda al capitolo successivo, e si fa presente che tale soglia è stata stabilita da studi precedenti (Javadi *et al.*, 2020). Nel grafico di Bland-Altman, si osserverà pertanto, sia che le osservazioni siano prossime a zero per misurare l'accordo tra le due tomografie, sia che rientrino nei *LOA* per valutare la loro equivalenza.

Nelle figure dalla 3.1 alla 3.9 sono presentati i grafici di Bland-Altman per ogni regione d'interesse. Nello specifico, le linee tratteggiate rappresentano il bias con gli estremi del relativo intervallo di confidenza, in nero è segnata la linea dello zero, mentre sono marcati in rosso i margini di equivalenza sopra descritti. Inoltre, vengono evidenziate le regioni di confidenza per ogni misura, quali estremo superiore in verde, differenza media (o bias) in blu e estremo inferiore in verde.

Figura 3.1. Bland-Altman Plot per la variabile **Bone**.

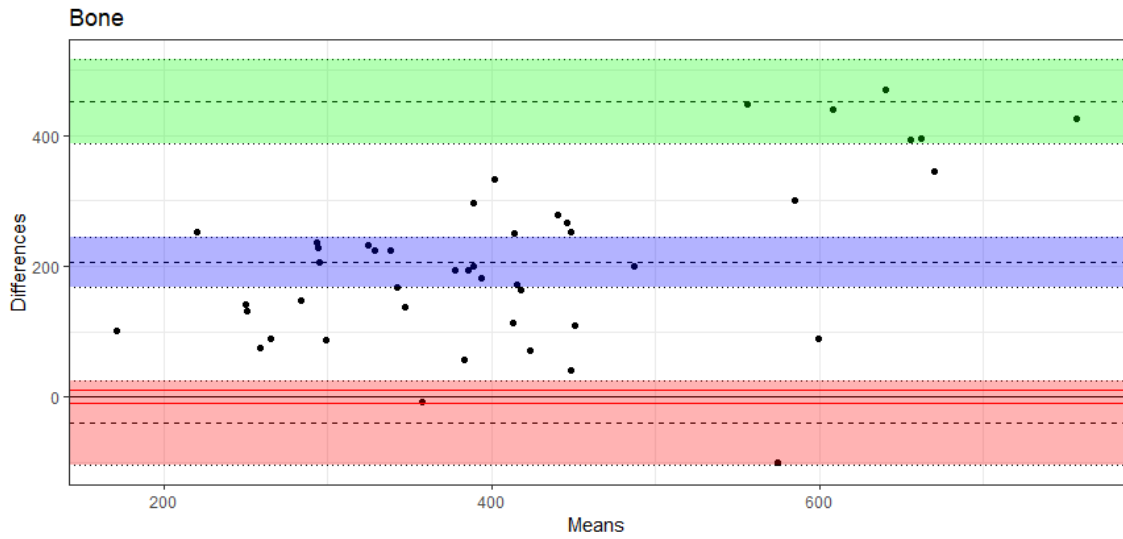


Figura 3.2. Bland-Altman Plot per la variabile **Muscle**.

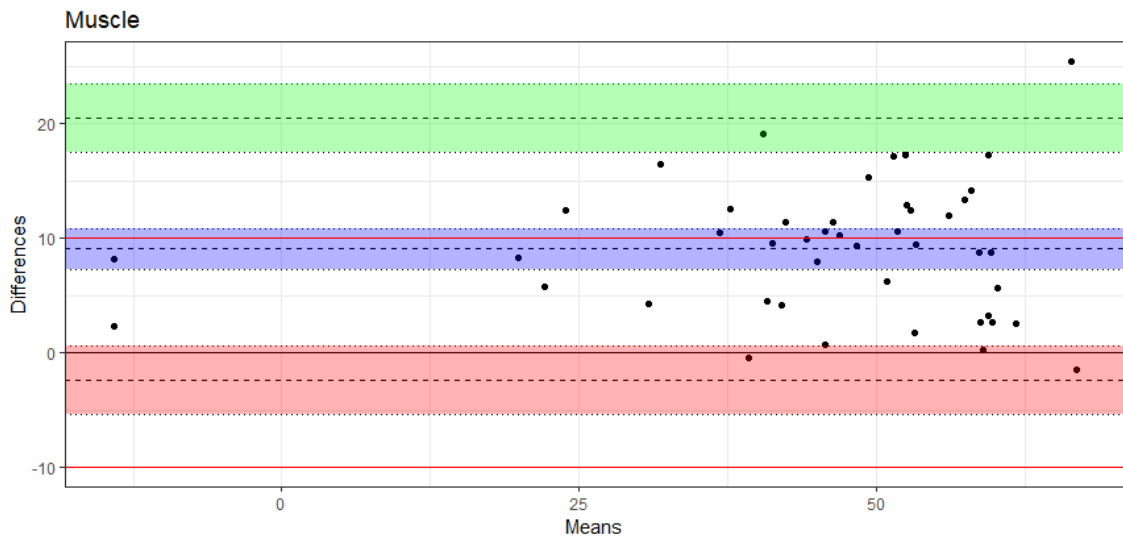


Figura 3.3. Bland-Altman Plot per la variabile **Fat**.

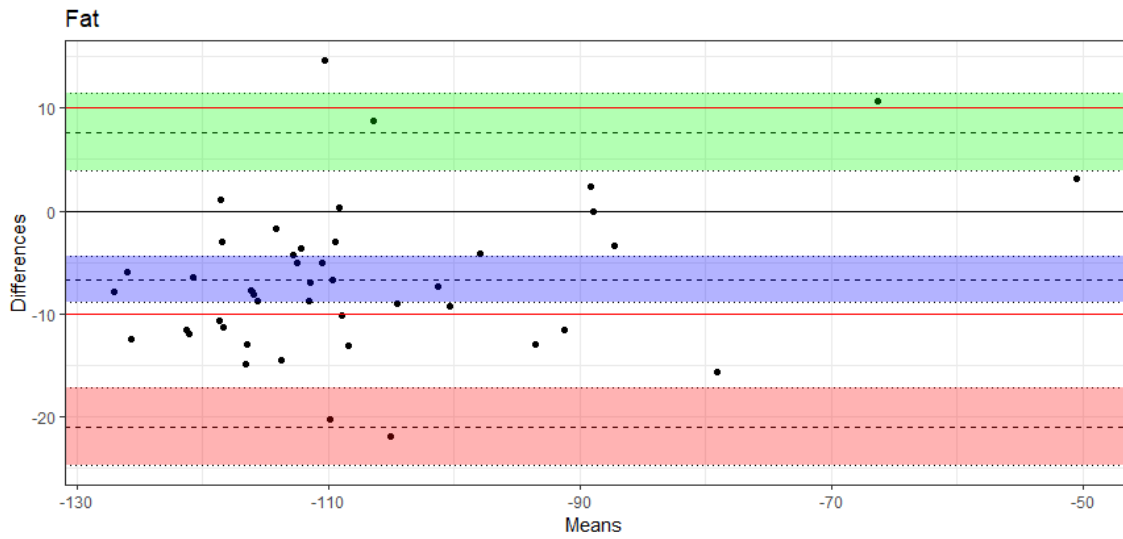


Figura 3.4. Bland-Altman Plot per la variabile **Renal Cortex**.

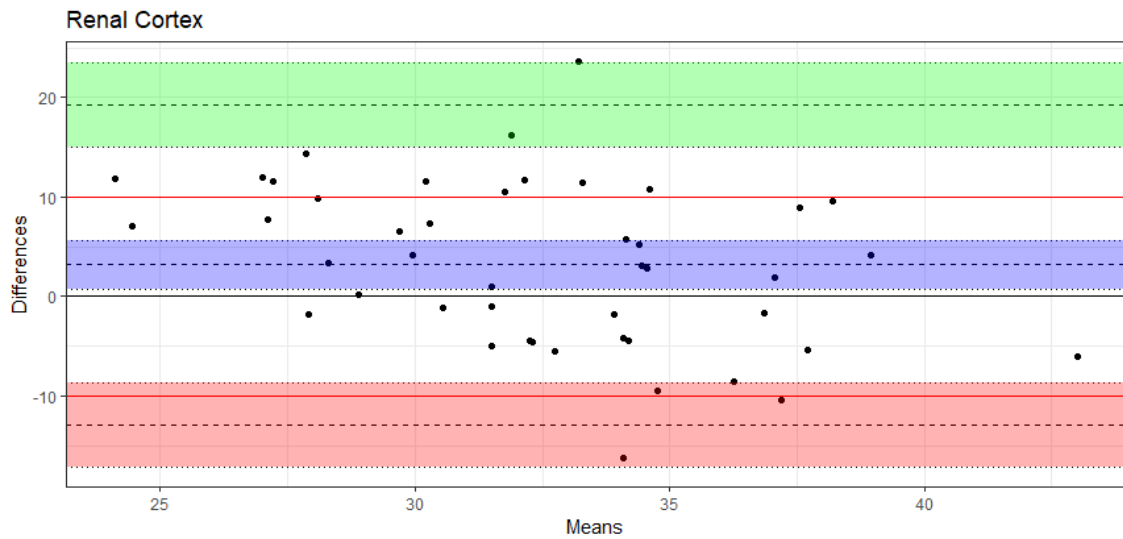


Figura 3.5. Bland-Altman Plot per la variabile **Aorta**.

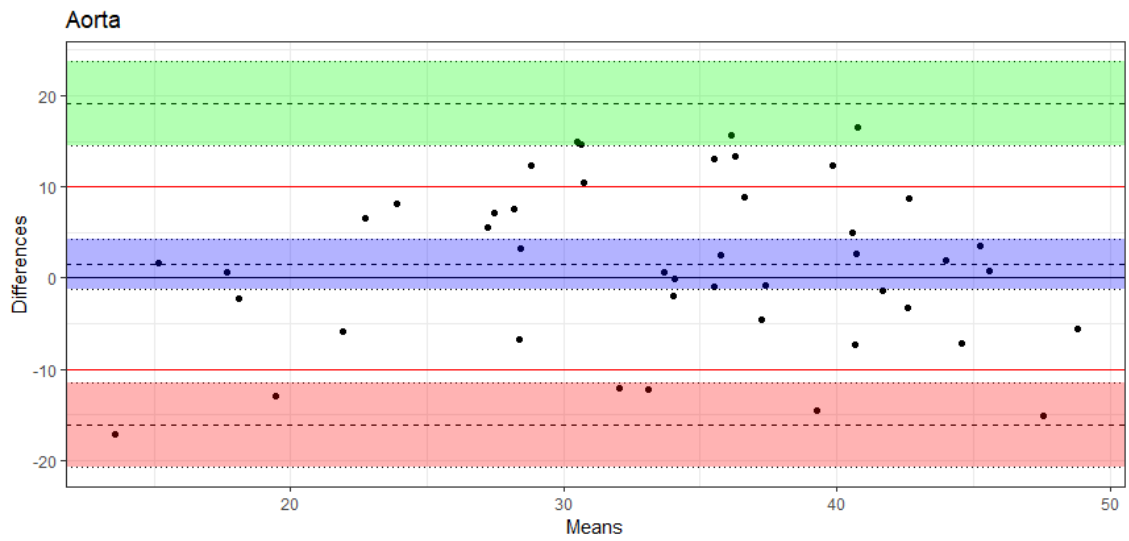


Figura 3.6. Bland-Altman Plot per la variabile **Portal Vein**.

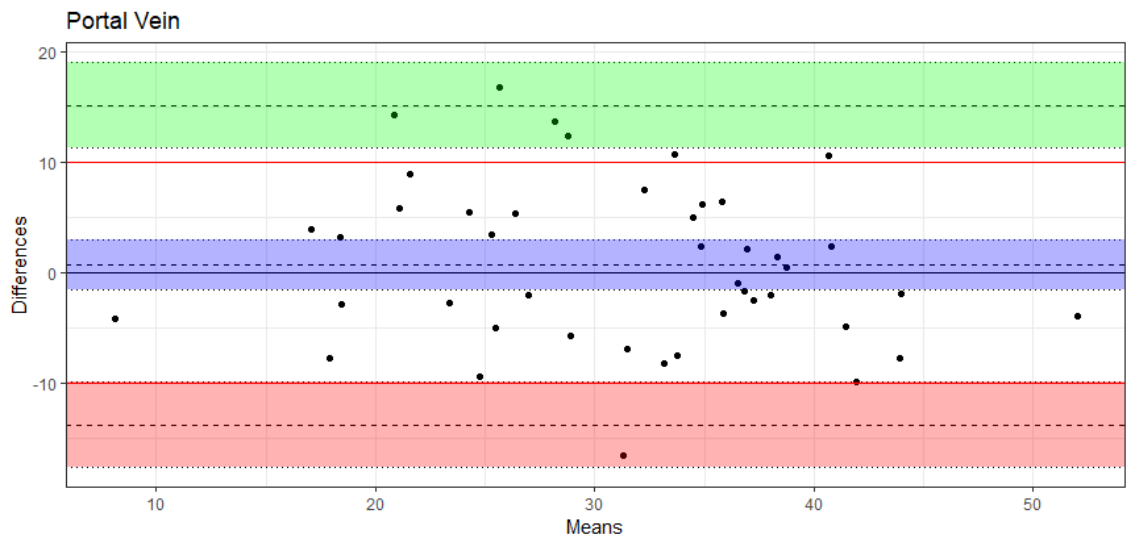


Figura 3.7. Bland-Altman Plot per la variabile **Spleen**.

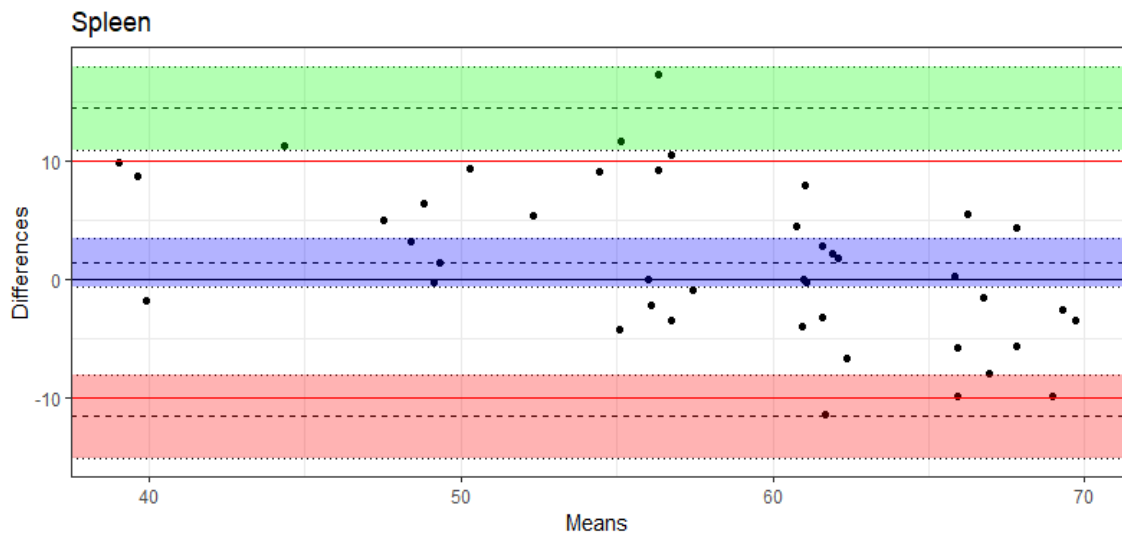


Figura 3.8. Bland-Altman Plot per la variabile **Gallbladder**.

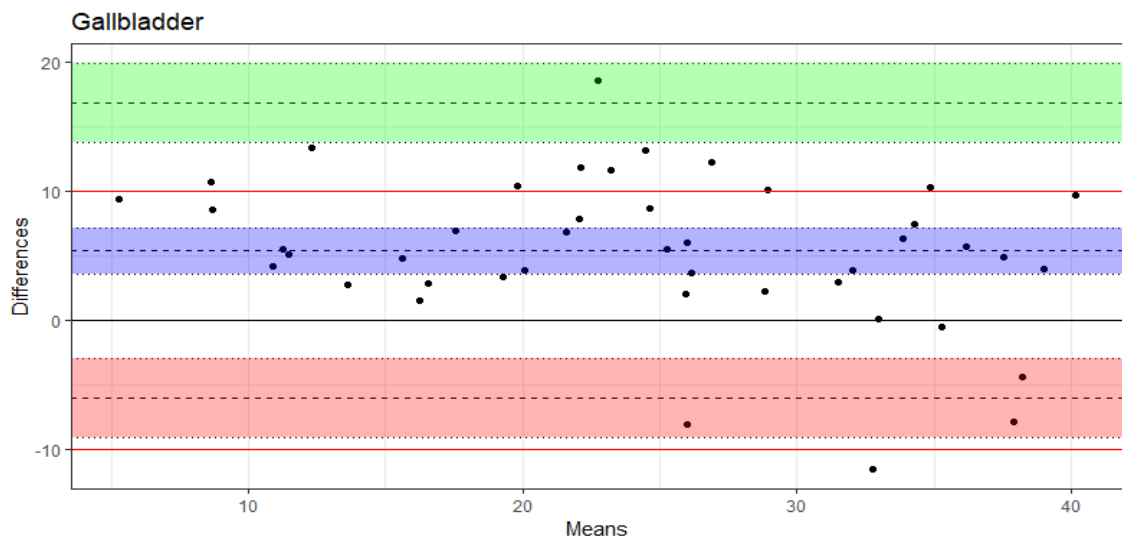
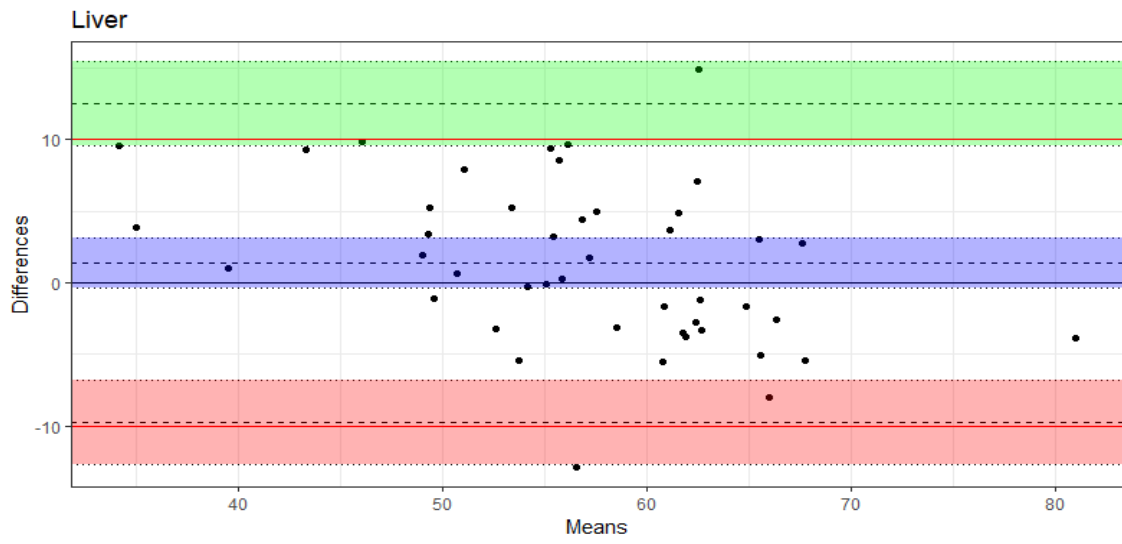




Figura 3.9. Bland-Altman Plot per la variabile **Liver**.



Nelle figure dalla 3.1 alla 3.9 si nota come per le ossa e i muscoli il bias risulti significativo, infatti la linea dello zero si trova nella regione del limite inferiore di confidenza, portando all'ipotesi che le due misurazioni siano significativamente differenti tra loro. Infatti, per i muscoli circa la metà dei punti si trova all'interno dell'intervallo di equivalenza, e per le ossa solo uno è compreso tra -10 e 10, indicando che la tomografia virtuale sovrastima quella tradizionale per entrambi le regioni d'interesse, poiché le restanti differenze assumono valori maggiori di 10 HU. Analogamente, per le variabili **Fat**, **Renal Cortex** e **Gallbladder** si dubita dell'accordo tra le tomografie poiché i grafici di Bland-Altman presentano le linee dello zero lontane dalla differenza media, specialmente per grasso e cistifellea. Per quanto concerne, invece, la soglia di equivalenza, si nota che la maggioranza delle differenze si trovano all'interno dei relativi limiti, per tutti e tre le aree. Infine, per aorta, vena porta, milza e fegato, si osserva che la linea dello zero è quasi coincidente con il bias e che i punti sono prossimi a esso, ma anche che la maggior parte delle differenze, specialmente per la variabile **Liver**, sono comprese tra -10 e 10 HU. Questo permette di dire che per queste ultime variabili, l'accordo tra PC e VNC appare elevato, facendo aumentare la fiducia nel fatto che la tomografia virtuale possa sostituire il gold standard, almeno per tali regioni d'interesse.

## 3.2 Coefficiente di concordanza di Lin

Obiettivo del presente paragrafo è considerare un indice che permetta di quantificare la correlazione delle misurazioni per valutarne ancora una volta l'accordo, analogamente al grafico di Bland-Altman precedentemente studiato.

A tale scopo, si è scelto di ricorrere al coefficiente di concordanza di Lin (CCC: *Concordance Correlation Coefficient*, in inglese). Tale quantità, proposto da Lin nel 1989, viene spesso utilizzata negli ambiti di biostatistica ed epidemiologia e serve per misurare il grado di accordo tra due trattamenti, ovvero vedere se questi due forniscano misure uguali tra loro, importante soprattutto nel caso in cui si stia valutando se sostituire il cosiddetto gold standard con una nuova tecnologia di misurazioni che garantirebbe maggiori vantaggi (Steichen e Cox, 2002). In generale però, viene utilizzato per misurare la concordanza tra due variabili quantitative  $x$  e  $y$ , per le quali il CCC di Lin è così definito:

$$\rho_c = \frac{2\rho\sigma_x\sigma_y}{(\mu_x - \mu_y)^2 + \sigma_x^2 + \sigma_y^2} = \frac{2\sigma_{xy}}{(\mu_x - \mu_y)^2 + \sigma_x^2 + \sigma_y^2} ,$$

dove  $\rho$  indica il coefficiente di correlazione tra  $x$  e  $y$ ,  $\sigma_x$  e  $\sigma_y$ , e  $\mu_x$  e  $\mu_y$  indicano rispettivamente le deviazioni standard e le medie delle popolazioni  $x$  e  $y$ .

Tale coefficiente, per l'obiettivo del presente studio, è preferibile al coefficiente  $\rho$  di Pearson, il quale, nonostante venga utilizzato più ampiamente in molteplici studi, non risulta offrire una giusta misura della riproducibilità delle osservazioni con VNC rispetto a PC. Infatti, mentre il coefficiente di Lin valuta l'*agreement*, ovvero in uno scatterplot dove sono rappresentate le osservazioni di due differenti misurazioni, quantifica la distanza dei punti dalla bisettrice del grafico che indica la perfetta concordanza, il  $\rho$  di Pearson, invece, quantifica la correlazione lineare, cioè la distanza dei punti del diagramma di dispersione da una qualsiasi retta. Di conseguenza, un'ulteriore differenza risiede nell'interpretazione del valore del coefficiente. Entrambi hanno un range che si estende da -1 a 1, ma mentre  $\rho_c$  indica un grado di accordo proporzionale al valore che assume, cioè un perfetto accordo corrisponde al valore +1 e perfetto disaccordo a -1, il coefficiente di Pearson, valuta la correlazione in base alla distanza dallo zero di tale quantità, ovvero +1 e -1 indicano perfetta correlazione, positiva e negativa, rispettivamente (Lin, 1989).

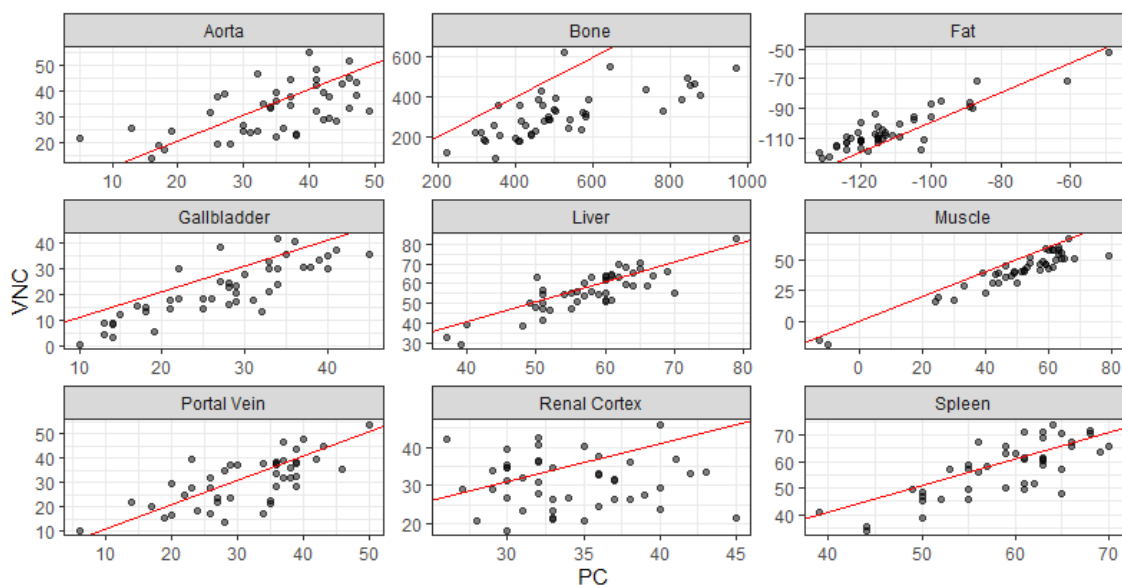
Uno strumento grafico di supporto, come già anticipato, consiste nel rappresentare in uno scatterplot le osservazioni effettuate con le due tecnologie e osservare quanto queste si discostino dalla retta bisettrice che rappresenta l'uguaglianza tra le due misurazioni (Racugno e Ventura, 2017). Ci si aspetta di trovare valori del CCC bassi se le osservazioni sono lontane dalla retta, mentre a valori alti del coefficiente di concordanza di Lin dovrebbero corrispondere punti del grafico prossimi alla bisettrice.

Nella tabella 3.1 e nella figura 3.10 sono presentati rispettivamente i coefficienti di Lin con relativi intervalli di confidenza e i diagrammi di dispersione, per ogni regione d'interesse.

Tabella 3.1. Coefficiente di Concordanza di Lin  $\rho_c$  per ogni ROI,  
con relativo intervallo di confidenza.

Variabili	Coefficiente di Lin		
	$\rho_c$	Intervallo di confidenza	
		<i>lower</i>	<i>upper</i>
<i>Bone</i>	0.332	0.196	0.455
<i>Muscle</i>	0.829	0.738	0.891
<i>Fat</i>	0.817	0.711	0.887
<i>Renal Cortex</i>	-0.027	-0.260	0.208
<i>Aorta</i>	0.589	0.363	0.750
<i>Portal Vein</i>	0.704	0.520	0.825
<i>Spleen</i>	0.716	0.564	0.821
<i>Gallbladder</i>	0.718	0.573	0.820
<i>Liver</i>	0.803	0.684	0.881

Figura 3.10. Scatterplot per i valori di HU misurati con PC e VNC,  
divisi per ROI, con relativa bisettrice.



Osservando la tabella 3.1 e la figura 3.10, si evince che il grado di accordo tra PC e VNC per le variabili **Fat**, **Liver** e **Muscle** appare molto alto, indicando che le due tomografie forniscono valori uguali tra loro. Infatti, il valore di  $\rho_c$  per tali ROI risulta superiore a 0.8 e le osservazioni nei diagrammi sono prossime alla bisettrice. Invece, per le ossa e soprattutto per la renale corticale, i valori rilevati dalla PC non si mostrano uguali a quelli della VNC. Infatti, le osservazioni si presentano distanti dalla bisettrice e il CCC calcolato risulta pari a 0.33 per **Bone** e -0.03 per **Renal Cortex**. Questi risultati possono essere provocati da quanto visto nel capitolo precedente, dove si sono studiate le correlazioni per le differenze di HU per le due tomografie. Infatti, la renale corticale, che presenta un  $\rho_c$  basso, è la stessa regione per cui le misurazioni delle CT erano fortemente correlate con il peso, inversamente tra le due tomografie, e le loro differenze producevano un esito significativo del t-test, relativamente alla generazione di TC utilizzata. Pertanto, poiché il coefficiente di Lin valuta l'accordo tra due misurazioni valutandone non tanto la differenza come per Bland-Altman, ma studiandone l'uguaglianza, i suddetti risultati possono essere provocati dal fatto che le rilevazioni delle due tomografie siano influenzate dall'intervento di altre variabili, le quali producono, per quelle regioni in cui è presente una forte correlazione con una determinata variabile, HU molto diversi tra le due tomografie computerizzate.

In conclusione, attraverso i grafici di Bland-Altman, i coefficienti di Lin e i diagrammi di dispersione per le tomografie si è valutato l'accordo delle misurazioni per verificare quanto la tecnologia sperimentale, in questo caso la VNC, sia sostituibile alla tomografia tradizionale. I CCC di Lin servono per offrire una misura quantitativa della concordanza delle due tecnologie di misurazione in termini assoluti, mentre il grafico di Bland-Altman aiuta a osservare se le differenze di HU rientrano in un certo intervallo, detto margine di equivalenza, oppure siano prossime a zero per indicare l'ampia probabilità di accettare l'ipotesi di uguaglianza tra le tomografie. Si è pertanto evinto che per il grasso addominale, i muscoli e il fegato, la concordanza risulta essere ampia, mentre le relative differenze non si presentano prossime allo zero, tranne per il fegato, i cui punti, oltre ad essere tutti entro i margini di equivalenza, sono anche distribuiti uniformemente intorno alla retta del bias. D'altra parte, per le ossa e la corticale renale, il cui CCC si presenta inferiore a 0.4, il rispettivo bias è risultato essere lontano dal valore nullo.

Pertanto, viste delle misure che permettono di figurare l'accordo tra le tomografie, occorre ora procedere con delle verifiche d'ipotesi per valutare l'effettiva uguaglianza ed equivalenza tra le due tecnologie di misurazione in esame.

# Capitolo 4

## Equivalence Test

Nel capitolo precedente si è cercato di studiare l'accordo tra le due tomografie tramite un'analisi preliminare delle differenze tra PC e VNC e delle loro rilevazioni sulle regioni d'interesse. Ora, lo scopo del presente capitolo consiste nel valutare quantitativamente l'accordo tra le misurazioni attraverso uno studio di verifica d'ipotesi.

### 4.1 Test di equivalenza e non-inferiorità

Per valutare l'*agreement* si è scelto di ricorrere a uno studio di equivalenza, ideato e sviluppato da Hauck e Anderson nel 1984. Questi tipi di test, sempre più popolari negli ultimi trent'anni in svariati campi applicativi, in special modo in quelli medico-scientifici, vengono utilizzati qualora l'obiettivo dello studio sia quello di concludere in favore della somiglianza tra due misure, anziché la loro non differenza. Tale scopo è ciò che rende interessanti e sempre più comuni i test di equivalenza, chiamati anche di somiglianza o di parità, poiché i test d'ipotesi sulle differenze più diffusi in letteratura non permettono di concludere in favore dell'equivalenza, ma solamente di rifiutare o meno l'ipotesi nulla di uguaglianza, potendo al massimo stabilire che due quantità siano tra esse non diverse (Meyners, 2012).



In particolar modo, i test di equivalenza, o di non-inferiorità, solitamente vengono utilizzati nel caso in cui sia necessario verificare che una nuova tecnologia o terapia, preferibile in termini di costi, facilità d'applicazione, salute o altri aspetti, abbia un'efficacia pari o non inferiore a quella già in uso, il cosiddetto gold standard.

Pertanto, in questo caso, occorre costruire un opportuno studio di verifica d'ipotesi che sia il contrario di quelli generalmente diffusi in letteratura. Infatti, tradizionalmente, l'ipotesi nulla costituisce l'opposto dell'obiettivo dello studio, ovvero ciò che la ricerca vuole rifiutare. Pertanto, l'ipotesi nulla viene rifiutata solamente qualora ci sia una forte evidenza a favore dell'ipotesi alternativa, altrimenti  $H_0$  non può essere rifiutata. Nel caso di uno studio di equivalenza, l'ipotesi alternativa non corrisponde più alla differenza, come nel caso di studi di confronto tradizionali, ma all'equivalenza delle misure, mentre  $H_0$  indica il contrario. Quindi, se l'evidenza a favore di  $H_1$  non risulta sufficiente, non si può concludere che le quantità non siano equivalenti. Analogamente, gli studi di non inferiorità indicano come ipotesi alternativa l'equivalenza o la superiorità della nuova tecnologia o terapia rispetto al gold standard (Walker e Nowacki, 2011). Il riepilogo delle ipotesi per tipi di studio precedentemente descritti sono presentati nella tabella 4.1.

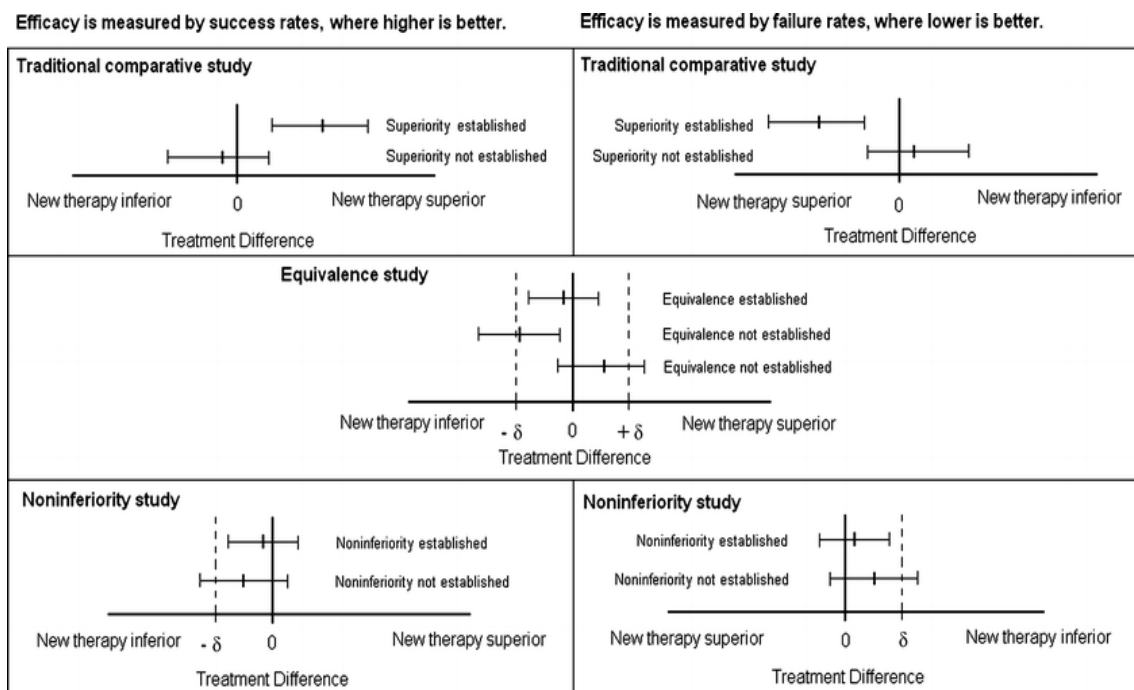
Tabella 4.1. Ipotesi nulla e alternativa per lo studio di verifica d'ipotesi di confronto, di equivalenza e di non-inferiorità (Walker e Nowacki, 2011).

<b>Tipo di studio</b>	<b>H0</b>	<b>H1</b>
<i>Confronto (tradizionale)</i>	Le terapie non sono differenti	Le terapie sono differenti
<i>Equivalenza</i>	Le terapie non sono equivalenti	La nuova terapia è equivalente a quella già in uso
<i>Non inferiorità</i>	La nuova terapia è inferiore a quella già in uso	La nuova terapia non è inferiore a quella già in uso

Si tenga presente che in questo caso il termine “*equivalenza*” non è inteso in senso stretto in quanto risulta quasi impossibile verificare la perfetta uguaglianza tra due terapie. Infatti, l’obiettivo degli studi di non inferiorità, in realtà, è dimostrare che le relative misure siano abbastanza vicine da essere considerate equivalenti. Tale vicinanza viene quantificata da una costante detta margine di equivalenza e viene indicata con  $\delta$ . In pratica, quest’ultimo indica la differenza massima accettabile tra le misurazioni affinché la nuova terapia possa sostituire quella già in uso, ovvero una nuova tecnologia viene considerata equivalente nel momento in cui le relative osservazioni portano a concludere che il suo effetto dista meno di  $\delta$  unità da quella già in uso (Walker e Nowacki, 2011). La quantità  $\delta$  serve quindi per definire i margini di equivalenza pari a  $-\delta_L$  quello inferiore e  $+\delta_U$  quello superiore, come mostrato nella figura 4.1.

Si ricorda che, analogamente a qualsiasi costruzione d'intervalli, anche quelli di equivalenza possono essere asimmetrici, ovvero, nonostante non sia di uso diffuso, non necessariamente  $\delta_L$  e  $\delta_U$  devono essere uguali. Poiché gli intervalli di equivalenza simmetrici sono maggiormente utilizzati e soprattutto più comuni in letteratura, si proseguirà facendo riferimento solamente a questi. Ciononostante, le future considerazioni saranno valide anche per limiti asimmetrici.

Figura 4.1. Esiti per gli studi di verifica d'ipotesi di confronto, di equivalenza e di non inferiorità in relazione ai margini di equivalenza (Walker e Nowacki, 2011).



## 4.2 Margine di equivalenza

Ora, è tanto importante quanto difficile capire che valore attribuire a  $\delta$  nel caso di test di equivalenza o non inferiorità. Scegliere tale valore adeguatamente non solo influenza l'esito dell'analisi, ma ne indica anche la credibilità scientifica. Ovviamente, minore è il valore del margine, più risulterà ristretto il relativo intervallo, rendendo più complicato accettare  $H_1$  in favore dell'equivalenza (o non inferiorità). Si ricorda che il rifiuto dell'ipotesi alternativa non implica che non ci sia nessun effetto, ma anzi significa che viene respinta la possibilità che esista un effetto maggiore rispetto al più piccolo effect size d'interesse (*SESOI: Smallest Effect Size Of Interest*, in inglese). Il SESOI rappresenta una quantità importante nella scelta di  $\delta$  poiché aiuta a definire l'ampiezza dell'intervallo di equivalenza entro il quale ci si aspetta che giaccia la differenza delle terapie (Lakens, 2017). In pratica, il valore del margine di equivalenza  $\delta$  risulta pari alla quantità del SESOI trovata.

A tal proposito, si sono diffuse varie teorie sulla scelta più corretta di SESOI, alcune più diffuse e accettate di altre. Tale quantità può essere determinata in modo oggettivo, ad esempio basandosi sulle differenze delle terapie interessate (Burriss *et al.*, 2015). Più diffuso è l'approccio soggettivo che prevede che siano i ricercatori stessi a impostare un valore del SESOI basandosi su determinate quantità non direttamente rilevabili sui dati a disposizione. Il primo metodo soggettivo può essere trovare una quantità del più piccolo effect size partendo da dei benchmark, che possono essere  $d=0.5$  oppure utilizzare il  $d$  di Cohen pari a 0.1 (Lakens *et al.*, 2018). Altri benchmark considerati nella letteratura sono derivati da meta-analisi come quelli studiati da Weber e Popova (2012).

Un secondo approccio, consiste nel fissare il SESOI a partire da ricerche correlate e analoghe, ovvero fissare tale quantità pari a quella utilizzata in analisi precedenti, come nel presente caso di studio come si vedrà nel prossimo capitolo. Tale approccio, anche se già stato utilizzato (Kordsmeyer e Penke, 2017), non è ancora un metodo molto diffuso. Una variante di questo metodo è data dalla scelta di SESOI pari al minor effect size osservato che avrebbe potuto essere significativo in uno studio precedente (Lakens *et al.*, 2018). Infine, il terzo approccio si basa sulla quantità di risorse in possesso, ovvero fissato il livello di significatività e il piano di campionamento per lo studio, i ricercatori possono calcolare il SESOI opportuno basandosi sulla numerosità campionaria, ricordando che tale quantità influisce sull'inferenza statistica (Meehl, 1978).

## 4.3 TOST

Per procedere con uno studio di equivalenza esistono vari approcci che sono stati oggetto di svariati dibattiti riguardo i loro pro e contro. Infatti, si è discusso a lungo se i criteri di selezione di un test “ottimale” dovessero essere basati solamente sulla sua potenza, oppure anche su altri criteri ad esempio soggettivi. Tra i vari approcci ci sono quello di potenza, usato come standard negli anni '80, quello di Anderson e Hauck, e infine, l'approccio più diffuso e maggiormente utilizzato, nonché quello più semplice, è la procedura denominata TOST (*Two One-Sided Test*) (Schuirmann, 1987). Quest'ultima prevede di valutare l'ipotesi nulla di non equivalenza, come esposto nel primo paragrafo del presente capitolo, attraverso due test unilaterali. Infatti, prevede che venga stabilita l'equivalenza qualora l'intervallo di confidenza  $(1 - 2\alpha) * 100\%$ , anziché quello  $(1 - \alpha) * 100\%$ , per le differenze delle quantità interessate sia contenuto nell'intervallo  $(-\delta ; \delta)$ , costruito come descritto precedentemente. Spesso, nonostante gli intervalli di confidenza siano molto esplicativi e ampiamente utilizzati, è utile affiancare a questi una misura che quantifichi la forza del rifiuto dell'ipotesi alternativa, ovvero calcolare il p-value del test eseguito. In particolare, per verificare che  $-\delta_L < \delta < +\delta_U$ , si procede con due test unilaterali, uno per verificare che la differenza delle terapie non sia minore del limite di equivalenza inferiore,  $\delta \leq -\delta_L$ , e il secondo che la differenza non sia maggiore del margine di equivalenza,  $\delta \geq +\delta_U$ . Nonostante si tratti di due test eseguiti singolarmente, bisogna tenere presente che si stanno eseguendo dei confronti multipli poiché entrambi i test devono risultare significativi affinché venga accettata l'ipotesi alternativa di equivalenza.

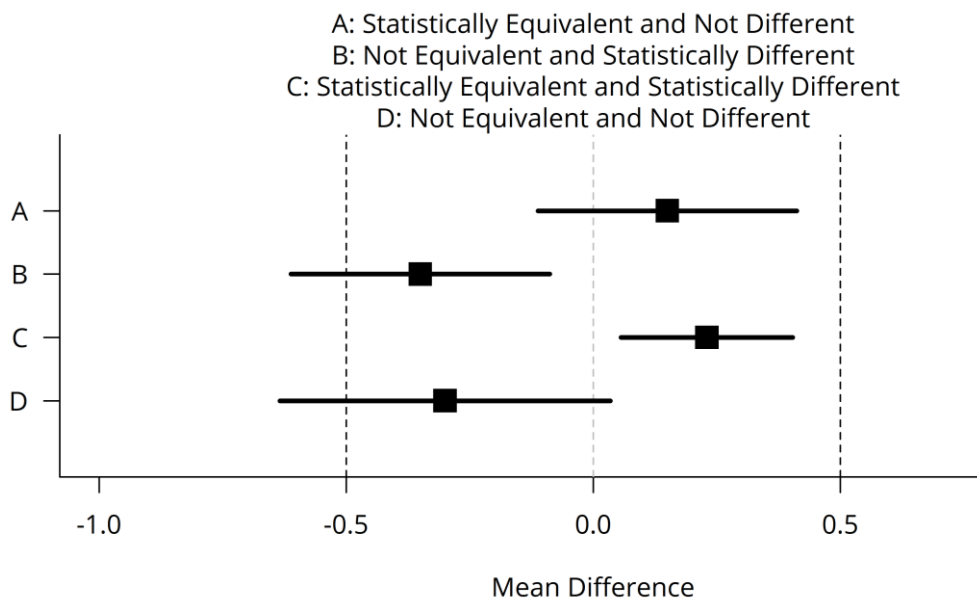
Pertanto, il valore osservato e il p-value di un TOST sono pari a quelli del test unilaterale con p-value maggiore, ovvero con valore osservato minore. In sostanza, due terapie vengono determinate equivalenti se il p-value più grande tra i due derivanti dai test unilaterali, è maggiore del livello di significatività  $\alpha$  fissato (Meyners, 2012).

Come ripetutamente esposto, i test di equivalenza, specificatamente la procedura TOST, si basano su una verifica d'ipotesi dove, contrariamente agli studi tradizionali, l'ipotesi alternativa rappresenta l'obiettivo dello studio e prevedono di eseguire due test unilaterali. Questo studio può essere affiancato a un tradizionale test bilaterale per verificare l'ipotesi nulla di uguaglianza tra le terapie, la quale viene comunemente chiamata con la sigla NHST (*Null Hypothesis Significance Test*, in inglese) (Lakens, 2017).

Quando vengono utilizzati entrambi i test, come viene anche mostrato graficamente nella figura 4.2, nella quale il quadratino indica la differenza media, le linee continue orizzontali il relativo intervallo di confidenza, e infine le linee tratteggiate più spesse rappresentano i limiti di equivalenza, gli esiti possibili sono quattro:

- A) le misurazioni sono significativamente equivalenti e non differiscono;
- B) le misurazioni non sono equivalenti, ma sono significativamente diverse;
- C) le misurazioni sono significativamente sia equivalenti sia diverse;
- D) le misurazioni non sono equivalenti, ma non sono diverse.

Figura 4.2. Differenze medie e relativi intervalli di confidenza al 95% con margini di equivalenza pari a -0.5 e 0.5 per gli esiti di quattro possibili combinazioni di test (Lakens, 2017).



Nella figura 4.2, per l'esito A si osserva che viene rifiutata  $H_0$  per il TOST, mentre lo stesso non accade per NHST. Ciò è indicato dal fatto che gli intervalli di confidenza per la differenza si trovano sia all'interno dei limiti di equivalenza, sia comprendano il valore 0 pari all'ipotesi nulla per NHST. Analogamente l'esito B prevede l'inverso del primo, poiché i relativi intervalli di confidenza sono esterni sia al valore nullo sia al limite di equivalenza inferiore. Osservando sempre la posizione degli intervalli di confidenza per la differenza media, si evince che l'esito C rifiuta entrambe le ipotesi nulle, ovvero vengono stabilite contemporaneamente equivalenza e non differenza delle terapie. Infine, contrariamente al precedente, l'esito D rappresenta il caso meno desiderabile, cioè quello per cui si osserva che le due terapie non sono né equivalenti e né non differenti.



## 4.4 TOST e NHST per le differenze di HU tra PC e VNC

In pratica, nel caso presente, si è deciso di procedere con uno studio di equivalenza per verificare con un approccio quantitativo l'accordo tra le due tomografie e, di conseguenza, accettare o rifiutare la loro equivalenza. Si è eseguita pertanto una procedura TOST, per ognuna delle regioni d'interesse, dopo aver definito un valore del SESOI pari a 10 HU, quantità utilizzata in studi analoghi precedenti come quello presentato da Javadi *et al.* nel 2020, generando così un intervallo di equivalenza pari a 20 HU. Di conseguenza, verificata la normalità delle differenze tra PC e VNC, si è proseguito effettuando due test t per campioni indipendenti. Si ricorda che i risultati di TOST riportati nella tabella 4.2 sono pari all'esito del test unilaterale che presenta valore osservato minore e quindi p-value maggiore. Nella medesima tabella sono contenuti anche le quantità relative al test NHST.

Tabella 4.2. Valore osservato del t-test, con relativi p-value, per le procedure di verifica d'ipotesi TOST e NHST.

Variabili	TOST		NHST	
	Valore osservato	p-value	Valore osservato	p-value
<i>Bone</i>	10.47	1	11.00	< 0.001
<i>Muscle</i>	-1.11	0.137	10.41	< 0.001
<i>Fat</i>	3.09	0.002	-6.11	< 0.001
<i>Renal Cortex</i>	-5.58	< 0.001	2.59	0.016
<i>Aorta</i>	-6.36	< 0.001	1.11	0.272
<i>Portal Vein</i>	-8.39	< 0.001	0.60	0.554
<i>Spleen</i>	-8.43	< 0.001	1.45	0.154
<i>Gallbladder</i>	-5.26	< 0.001	6.13	< 0.001
<i>Liver</i>	-10.19	< 0.001	1.63	0.111

Come si era già inizialmente ipotizzato dalla tabella 2.1, si può notare nella tabella 4.2 che per le variabili **Bone** e **Muscle** sia l'ipotesi di equivalenza che quella di uguaglianza vengono rifiutate, perciò si può concludere che la nuova tomografia VNC, a parità di condizioni, non possa sostituire la tecnologia PC. Per quanto riguarda le altre variabili, si nota come l'equivalenza delle medie possa essere ampiamente accettata per **Aorta**, **Portal Vein**, **Spleen** e **Liver**, poiché sia il TOST è risultato significativo, ma anche NHST non significativo, indicando, pertanto, che non solo viene accettata l'ipotesi che non ci sia differenza, ma viene anche rifiutata l'ipotesi di disuguaglianza. Infine, viene osservato che le misurazioni relative alle variabili **Fat**, **Renal Cortex** e **Gallbladder**, possono essere considerate equivalenti, ma allo stesso tempo, sono significativamente differenti.

# Capitolo 5

## Conclusione

Negli ultimi anni si è diffusa una nuova tecnologia per eseguire delle tomografie computerizzate senza mezzo di contrasto. Consiste nell'eseguire una DECT con contrasto per poi eliminare dalle immagini ottenute il segnale del materiale caratterizzante il liquido di contrasto, per poter ottenere scansioni simili a quelle che si sarebbero rilevate con una tradizionale tomografia computerizzata senza mezzo di contrasto. Questo nuovo approccio viene chiamata tomografia computerizzata senza mezzo di contrasto oppure in breve VNC (Virtual Non Contrast). Tale tecnologia è stata presa in considerazione in vari studi (Javadi *et al.*, 2020) poiché permetterebbe, se idonea a sostituire la tomografia tradizionale, di sottoporre il paziente a sessioni tomografiche più brevi e, di conseguenza, ridurre l'assorbimento di radioattività emesso dalla macchina stessa. Pertanto, l'obiettivo del presente studio è verificare che effettivamente questa nuova tomografia nel caso specifico in esame, ovvero in ambito veterinario e relativamente alla zona addominale, sia equivalente con quella attualmente utilizzata per poter proporre una sostituzione effettiva, nonché vantaggiosa, tra le due tecnologie.

Nel presente studio è stata eseguita preliminarmente un'analisi esplorativa del campione, ponendo particolare attenzione, poiché in presenza di dati appaiati, alla differenza dei valori di HU tra le due tomografie, specificatamente PC – VNC, per la quale si è verificata normalità e omoschedasticità. Inizialmente, si sono osservate differenze medie prossime allo zero, tranne per le ossa dove lo scostamento dal valore nullo si presentava particolarmente elevato indicando valori della scala di Hounsfield della tomografia tradizionale nettamente maggiori rispetto a quella virtuale. Anche la variabile **Muscle** presentava valori leggermente discostati da zero, ma senza suscitare dubbi rilevanti. Dopodiché, si è proseguito con lo studio analizzando se gli esiti delle tomografie fossero influenzati da particolari caratteristiche, come l'età, il peso o lo stato sessuale delle unità statistiche, oppure la generazione di tecnologia utilizzata per le scansioni. Per la variabile **Age** si è notata una correlazione positiva elevata con i muscoli per entrambe le tomografie, mentre per **Weight** il valore del  $\rho$  di Spearman per la corticale renale è risultato elevato, positivamente per PC e negativamente per VNC. Inoltre, si è osservata una particolare sensibilità da parte del peso per i valori rilevati con la tomografia virtuale.

Per quanto riguarda la variabile **Sexual Status**, invece, non si è notata nessuna particolare associazione con i valori HU nelle varie regioni d'interesse. Infine, per la variabile **CT** si è osservato attraverso un test t che i valori rilevati da PC e VNC erano significativamente diversi tra loro per il grasso addominale, la corticale renale, e i vasi sanguigni considerati, l'aorta e la vena porta.

Dopodiché, si è deciso di studiare il livello di accordo tra le due tomografie in esame per figurare preliminarmente se le due tecnologie fossero equivalenti tra loro oppure no. In primo luogo, si sono rappresentate le differenze osservate su un grafico di Bland-Altman e si è studiato se queste si disponessero prossime al bias e se fossero contenute nei margini di equivalenza pari a -10 HU e +10 HU. Attraverso questi diagrammi, si sono ricavate molteplici informazioni, ad esempio per la variabile **Bone**, come si era già visto nel capitolo precedente, le differenze erano risultate sia lontane da zero sia esterne ai margini di equivalenza; mentre per la variabile **Muscle**, nonostante le differenze si fossero avvicinate al bias e in parte rientrassero nell'intervallo di equivalenza, molte di queste risultavano ancora lontane dall'ipotesi di accordo tra le TC. Tutte le altre variabili, invece, mostravano delle differenze per lo più contenute tra  $\delta_L$  e  $\delta_U$  e prossime alla linea della media delle differenze, tranne per **Renal Cortex** e **Gallbladder**, per le quali l'ipotesi di uguaglianza risultava opinabile, poiché il valore nullo si presentava esterno dall'intervallo di confidenza del bias. In aggiunta, per misurare il livello di accordo tra PC e VNC, si è calcolato, per ognuna delle regioni d'interesse, il coefficiente di concordanza di Lin affiancato dagli scatterplot delle differenze. Attraverso quest'analisi si è evinto che la corticale renale presentava valori HU molto diversi tra tomografia tradizionale e virtuale, probabilmente dovuto al fatto che nel caso del peso, le relative osservazioni fossero correlate inversamente, portando quindi a un valore basso per il CCC, ma quantità più che accettabili in altri test dove venivano considerate le differenze tra le due tecnologie.

Infine, si è proseguito con uno studio di verifica d'ipotesi per accettare o rifiutare l'ipotesi di equivalenza tra le due tomografie.

Per fare ciò, si è eseguito un equivalence test con approccio TOST, affiancato da una tradizionale verifica d'ipotesi NHST. Si è concluso che l'equivalenza e la non differenza sono stati rifiutati per le variabili **Muscle** e **Bone**, mentre sono state ampiamente accettate per tutte le altre variabili ad esclusione della cistifellea, della renale corticale e del grasso addominale. Tale risultato è confermato dai grafici di Bland-Altman osservati nel capitolo 3.

In definitiva, in base alle analisi eseguite e sopra descritte, la tomografia virtuale si pone come ottimo sostituto della PC, ad eccezione di qualche regione d'interesse, verso le quali va prestata particolare attenzione. La tomografia computerizzata virtuale senza contrasto sulle ossa addominali rivela valori della scala Hounsfield significativamente differenti dalla DECT senza contrasto tradizionale. Per tale motivo, la VNC non sembra offrire risultati validi e affidabili per tal regione d'interesse. Per i muscoli addominali e la corticale renale, invece, occorre procedere con prudenza. Infatti, per la prima ROI tutte le analisi, specialmente quelle grafiche, offrono risultati limite, nel senso che appaiono difficilmente interpretabili se osservati assolutamente, mentre lo studio di verifica d'ipotesi rifiuta l'equivalenza e la non differenza delle misurazioni tra le due tomografie. Tale esito potrebbe essere provocato dall'età la quale influisce fortemente sulla radiodensità muscolare. Infatti, svariati studi confermano che, nel corpo umano, la sarcopenia, ovvero la perdita della massa muscolare dovuta all'invecchiamento, aumenta del 3-8% ogni decade a partire dai trent'anni, per poi aumentare dopo i 60 (Volpe *et al.*, 2003).

Lo stesso fenomeno è stato studiato anche in ambito veterinario e ne è stata confermata la presenza anche nei cani, soggetti del nostro studio (Freeman, 2011). Pertanto, qualora si dovesse utilizzare una tomografia virtuale senza contrasto, è opportuno tenere presente l'effetto dell'età e la possibile presenza di sarcopenia. Infine, le analisi suggeriscono di utilizzare la TC virtuale senza mezzo di contrasto ponendo particolare attenzione nel caso di pesi corporei estremi, poiché si è riscontrato, specialmente nel caso della renale corticale, per la quale comunque è stata accettata l'equivalenza delle due tecnologie, che la VNC è particolarmente sensibile a cambiamenti di peso nei soggetti in esame.

In conclusione, la tomografia computerizzata virtuale senza mezzo di contrasto è un'alternativa promettente alla TC addominale tradizionale portando a svariati vantaggi, tra tutti quello di poter sostituire le immagini tradizionali nel caso in cui queste non siano disponibili, per l'aorta addominale, la vena porta, la cistifellea, il grasso addominale, la milza, il fegato e la renale corticale, prestando particolare attenzione a quest'ultima regione d'interesse. La VNC non risulta un valido sostituto della PC, invece, per i muscoli e le ossa addominali.

# Appendice

Tabella A.1. Principali statistiche di sintesi per i valori di HU misurati nelle ROI in esame sia con PC sia con VNC.

Variabili		Minimo	Media	Mediana	Massimo	IQR	Deviazione Standard	Asimmetria	Test di Shapiro-Wilk	
									Valore osservato	p-value
<i>Bone</i>	PC	222.00	<b>519.70</b>	483.00	969.00	168.00	177.18	0.85	0.91	<b>0.002</b>
	VNC	94.20	<b>313.90</b>	300.9	625.60	166.30	116.86	0.54	0.97	0.368
<i>Muscle</i>	PC	-13.00	<b>50.09</b>	54.00	79.00	17.00	17.65	<b>-1.85</b>	0.82	<b>&lt; 0.001</b>
	VNC	-18.20	<b>41.05</b>	43.70	67.50	14.30	17.13	<b>-1.65</b>	0.85	<b>&lt; 0.001</b>
<i>Fat</i>	PC	-132.00	<b>-110.90</b>	-115.00	-49.00	15.00	16.77	<b>1.72</b>	0.84	<b>&lt; 0.001</b>
	VNC	-123.20	<b>-104.20</b>	-109.10	-52.10	15.00	14.56	<b>1.47</b>	0.85	<b>&lt; 0.001</b>
<i>Renal Cortex</i>	PC	26.00	<b>34.07</b>	33.00	45.00	7.00	4.48	0.46	0.96	0.168
	VNC	18.20	<b>30.90</b>	31.30	46.00	9.10	6.73	0.09	0.98	0.647
<i>Aorta</i>	PC	5.00	<b>34.40</b>	36.00	49.00	12.00	10.13	-0.86	0.93	<b>0.013</b>
	VNC	14.30	<b>32.91</b>	33.40	55.10	15.00	9.82	0.17	0.98	0.599
<i>Portal Vein</i>	PC	6.00	<b>31.48</b>	34.50	50.00	12.00	9.04	-0.51	0.96	0.167
	VNC	10.20	<b>30.81</b>	31.90	54.00	15.30	10.16	-0.03	0.98	0.555
<i>Spleen</i>	PC	39.00	<b>58.56</b>	61.00	70.00	8.00	7.20	-0.68	0.95	0.057
	VNC	34.10	<b>57.09</b>	58.50	73.90	16.35	10.43	-0.37	0.96	0.147
<i>Gall-bladder</i>	PC	10.00	<b>27.20</b>	28.00	45.00	13.50	8.99	-0.14	0.97	0.280
	VNC	0.60	<b>21.82</b>	20.55	41.80	15.75	10.75	0.01	0.97	0.421
<i>Liver</i>	PC	37.00	<b>57.49</b>	59.00	62.00	11.00	8.12	-0.30	0.97	0.208
	VNC	29.40	<b>56.12</b>	55.70	82.90	12.90	10.30	-0.30	0.97	0.384



Figura A.1. Boxplot dei valori HU per le ROI in considerazione  
divise per tipologia di CT.

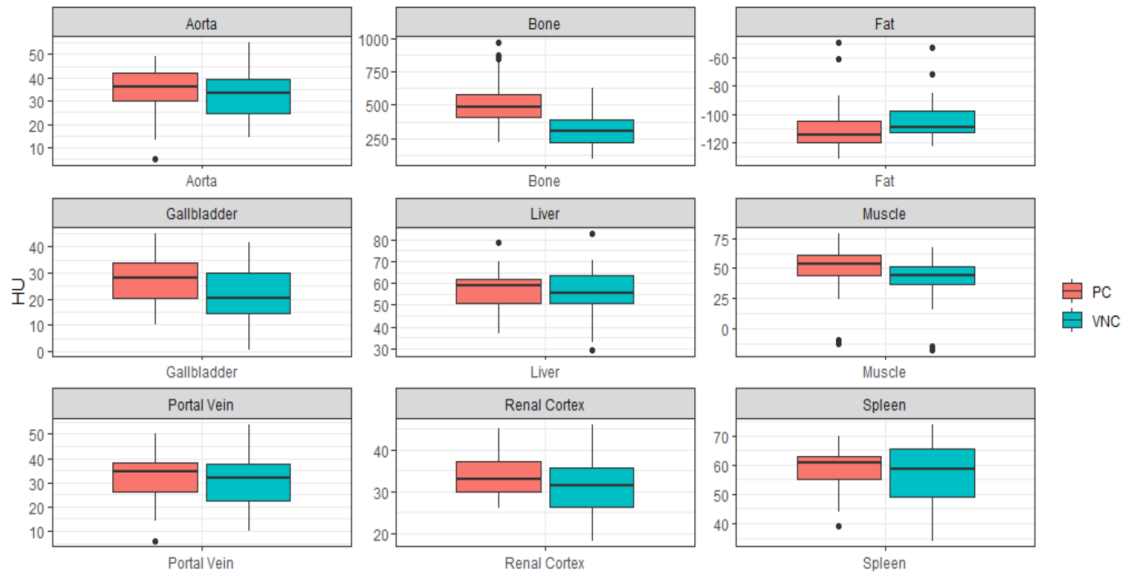


Figura A.2. QQ-plot dei valori HU per tutte le ROI in considerazione  
divise per tipologia di CT.

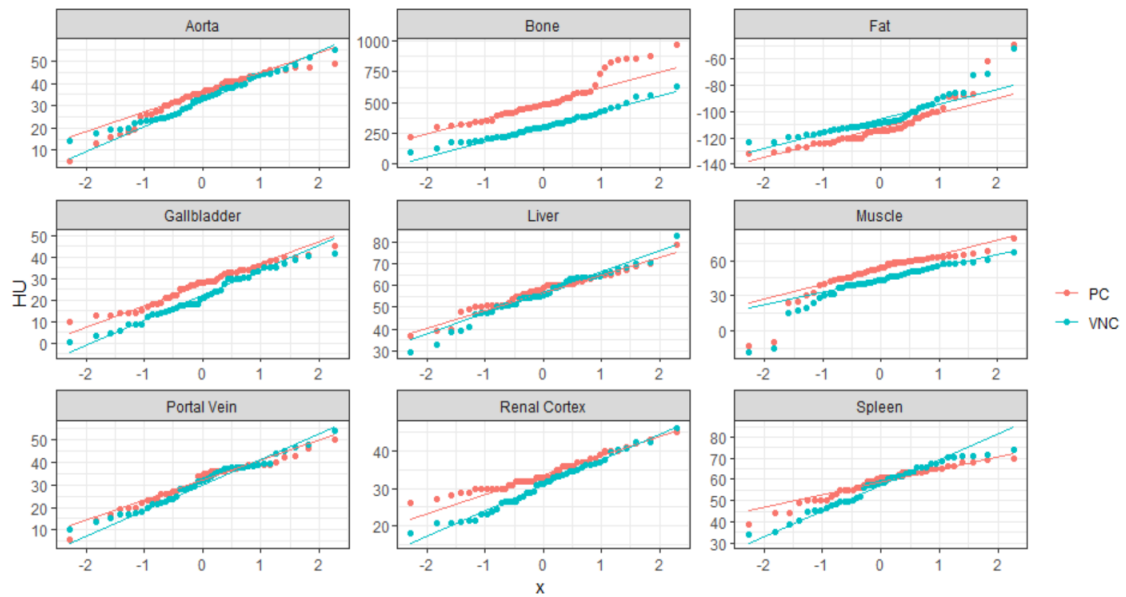


Tabella A.2. Valore osservato  $K^2$  del test di Bartlett per le differenze di HU tra le due tomografie su ogni ROI, condizionatamente alla variabile **Sexual Status**, con relativo p-value.

Variabili	Test di Bartlett	
	Valore osservato $K^2$	p-value
<i>Bone</i>	6.67	0.083
<i>Muscle</i>	0.44	0.930
<i>Fat</i>	1.39	0.708
<i>Renal Cortex</i>	2.66	0.446
<i>Aorta</i>	0.03	0.990
<i>Portal Vein</i>	0.58	0.902
<i>Spleen</i>	2.64	0.451
<i>Gallbladder</i>	5.23	0.156
<i>Liver</i>	0.89	0.829

# Riferimenti bibliografici

- Alvarez, R.E., Macovski, A. (1976), Energy-selective reconstructions in X-ray computerized tomography. *Physics in Medicine and Biology*, **21(5)**, 733-44 (<https://doi.org/10.1088/0031-9155/21/5/002>).
- Bland, J.M., Altman, D.G. (1986), Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, **327(8476)**, 307-310 ([https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)).
- Burriss, R.P., Troscianko, J., Lovell, P.G., Fulford, A.J.C., Stevens, M., Quigley, R., Jenny, P., Saxton, T.K., Rowland, H.M. (2015), Changes in women's facial skin color over the ovulatory cycle are not detectable by the human visual system. *Plos One*, **10(7)** (<https://doi.org/10.1371/journal.pone.0130093>).
- DenOtter, T.D., Schubert, J. (2022), *Hounsfeld Unit*. StatPearls Publishing.
- Doğan, N.O. (2018), Bland-Altman analysis: A paradigm to understand correlation and agreement. *Turkish Journal of Emergency Medicine*, **18(4)**, 139-141 (<https://doi.org/10.1016/j.tjem.2018.09.001>).
- Freeman, L.M. (2011), Cachexia and Sarcopenia: Emerging Syndromes of Importance in Dogs and Cats. *Journal of Veterinary Internal Medicine*, **26(1)**, 3-17 (<https://doi.org/10.1111/j.1939-1676.2011.00838.x>).
- Hauck, W.W., Anderson, S. (1984), A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *Journal of Pharmacokinetics and Biopharmaceutics* **12**, 83-91 (<https://doi.org/10.1007/BF01063612>).

- Hollander, M., Wolfe, D.A., Chicken, E. (2013). *Nonparametric statistical methods*.
- Javadi, S., Elsherif, S., Bhosale, P., Jensen, C.T., Layman, R.R., Jacobsen, M.C., Le, O., Jia, S., Parikh, R.J., Tamm, E.P. (2020), Quantitative attenuation accuracy of virtual non-enhanced imaging compared to that of true non-enhanced imaging on dual-source dual-energy CT. *Abdominal Radiology*, **45**, 1100–1109 (<https://doi.org/10.1007/s00261-020-02415-8>).
- Kordsmeyer, T.L., Penke, L. (2017). The association of three indicators of developmental instability with mating success in humans. *Evolution & Human Behavior*, **38(6)**, 704–713 (<https://doi.org/10.1016/j.evolhumbehav.2017.08.002>).
- Lakens, D. (2017), Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Social Psychological and Personality Science*, **8(4)**, 355–362 (<https://doi.org/10.1177/1948550617697177>).
- Lakens, D., Scheel, A.M., Isager, P.M. (2018), Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practises in Psychological Science*, **1(2)**, 259-269 (<https://doi.org/10.1177/2515245918770963>).
- Lee, D.K. (2016), Alternatives to P value: confidence interval and effect size. *Korean Journal of Anesthesiology*, **69(6)**, 555-562 (<https://doi.org/10.4097/kjae.2016.69.6.555>).
- Lin, L.I. (1989), A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, **45**, 255–268 (<https://doi.org/10.2307/2532051>).

- Maxwell, S.E., Lau, M.Y., Howard, G.S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, **70(6)**, 487–498 (<https://doi.org/10.1037/a0039400>).
- Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, **46**, 806–834 (<https://doi.org/10.1037/0022-006X.46.4.806>).
- Meyners, M. (2012), Equivalence Test – A Review. *Food Quality and Preference*, **26(2)**, 231-245 (<https://doi.org/10.1016/j.foodqual.2012.05.003>).
- Racugno, W., Ventura, L. (2017). *Biostatistica. Casi di studio in R*. Egea Editore.
- Schuirman, D.J. (1987), A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, **15**, 657–680 (<https://doi.org/10.1007/BF01068419>).
- Siegel, M.J., Bhalla, S., Cullinane, M. (2021), Dual-Energy CT Material Decomposition in Pediatric Thoracic Oncology. *Radiology: Imaging Cancer*, **3(1)** (<https://doi.org/10.1148/rycan.2021200097>).
- Steichen, T. J., Cox, N. J. (1998-2002), Concordance correlation coefficient. *Stata Technical Bulletin*, **43**, 35-39 (<https://doi.org/10.1177/1536867X0200200206>).

- Virarkar, M.K., Swarupa, S.S.R., Gupta, A.V., Gopireddy, D.R., Kumar, S., Hernandez, M., Lall, C., Bhosale, P. (2022), Virtual Non-contrast Imaging in The Abdomen and The Pelvis: An Overview. *Seminars in Ultrasound, CT and MRI*, **43(4)**, 293-310 (<https://doi.org/10.1053/j.sult.2022.03.004>).
- Volpi, E., Nazemi, R., Fujita, S., (2004), Muscle tissue changes with aging. *Current Opinion in Clinical Nutrition and Metabolic Care*, **7(4)**, 405-10 (<https://doi.org/10.1097/01.mco.0000134362.76653.b2>).
- Walker, E., Nowacki, A.S., (2011), Understanding equivalence and noninferiority testing. *Journal of Internal General Medicine*, **26(2)**, 192-196 (<https://doi.org/10.1007/s11606-010-1513-8>).
- Weber, R., Popova, L., (2012). Testing equivalence in communication research: Theory and application. *Communication Methods and Measures*, **6**, 190–213 (<https://doi.org/10.1080/19312458.2012.703834>).