4. HOW TO MEASURE THE RESULTS

mercoledì 30 ottobre 2024 23:32

How to measure if the results are significant

PAIRED T- TEST

Definizione: Il t-test per campioni appaiati confronta due insiemi di dati, dove ogni elemento dell'insieme è associato a una controparte nell'altro insieme, formando quindi delle coppie di osservazioni. È comune utilizzarlo quando la stessa entità è misurata due volte, prima e dopo una certa condizione o evento.

Questo tipo di test è utile per vedere se esistono differenze significative tra due misurazioni dello stesso soggetto (come il prima e dopo un trattamento) o tra misurazioni provenienti da due sistemi diversi.

3. **Esempio nell'immagine**: Nel caso in questione, le coppie di osservazioni sono valori di accuratezza calcolati da due sistemi (denominati y^a e y^b) che variano la soglia. I valori $y^a = \{y^a_1, y^a_2, \ldots, y^a_n\}$ e $y^b = \{y^b_1, y^b_2, \ldots, y^b_n\}$ rappresentano i risultati delle due serie di misurazioni per i due sistemi.

Ipotesi Nulla (Null Hypothesis): L'ipotesi nulla afferma che i due sistemi di apprendimento hanno la stessa accuratezza. In altre parole, non c'è una differenza significativa tra le prestazioni dei due sistemi.

2. Ipotesi Alternativa (Alternative Hypothesis): L'ipotesi alternativa suggerisce che uno dei due sistemi sia più accurato dell'altro, indicando che esiste una differenza significativa nelle loro prestazioni.
3. Sotto l'ipotesi nulla: Se assumiamo che l'ipotesi nulla sia vera, qualsiasi differenza osservata tra le due serie di dati è attribuibile alla variazione casuale (random variation), non a una differenza reale nelle prestazioni.
4. Test d'Ipotesi:
   - Si utilizza il t-test per campioni appaiati per calcolare la probabilità (p-value) che la differenza media osservata sia compatibile con l'ipotesi nulla.
   - Se il valore di p è sufficientemente piccolo (tipicamente inferiore a 0,05), si rigetta l'ipotesi nulla, concludendo che esiste una differenza significativa tra i due sistemi.

$\vec{\boldsymbol{y}}^a = \{y^a_1, y^a_2, \ldots, y^a_n\}$ $\quad$ $\vec{\boldsymbol{y}}^b = \{y^b_1, y^b_2, \ldots, y^b_n\}$ $\quad$ $\vec{\boldsymbol{\delta}} = \{y^a_1 - y^b_1, y^a_2 - y^b_2, \ldots, y^a_n - y^b_n\}$

1 PASSO $\quad \overline{\delta} = \frac{1}{M} \sum_{i=1}^{M} \delta_i$ $\quad$ N = num osservazioni

2 T-STATISTIC

$$t = \frac{\overline{\delta}}{\sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} (\delta_i - \overline{\delta})^2}}$$

3

l p-value corrispondente indica la probabilità che la differenza osservata sia dovuta al caso. Se il p-value è inferiore a una soglia predefinita (ad esempio, 0.05), si rigetta l'ipotesi nulla, concludendo che esiste una differenza significativa tra i due sistemi.

Vengono mostrati i diversi modi di interpretare i p-value a seconda del tipo di test d'ipotesi:

Questo test verifica se c'è una differenza significativa tra i due sistemi, indipendentemente dalla direzione della differenza. Il p-value a due code si calcola raddoppiando la probabilità che T (la statistica del test) sia maggiore del valore assoluto di t (la statistica osservata). Questo approccio si usa quando si vuole verificare la presenza di una differenza significativa, senza specificare se un sistema è migliore o peggiore dell'altro.

$$p = 2 \cdot P_r(T > |t|)$$

2 TEST A UNA CODA SUPERIORE

$$p = P_r(T > t)$$

Questo test verifica se la media del primo sistema è significativamente maggiore della media del secondo sistema. Il p-value rappresenta la probabilità che T sia maggiore di t. Questo approccio si usa quando si ha l'ipotesi specifica che un sistema sia migliore dell'altro.

3 Test a una coda inferiore:

$$p = P_r(T < t)$$

Questo test verifica se la media del primo sistema è significativamente inferiore alla media del secondo sistema. Il p-value rappresenta la probabilità che T sia minore di t. Questo approccio si usa quando si ipotizza che un sistema sia peggiore dell'altro.

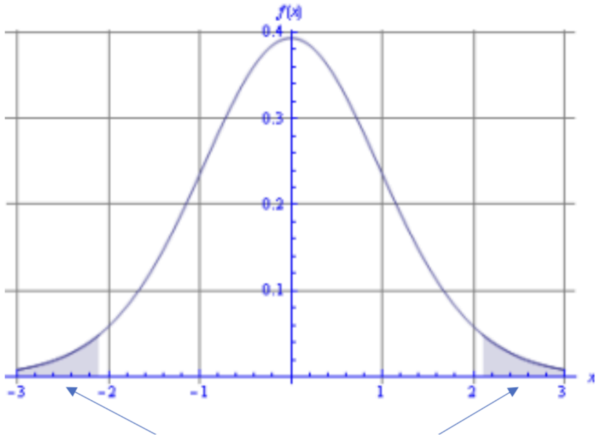
Test t accoppiati che possono essere utilizzati per confrontare la precisione di due sistemi.

Test a due code (Two-tailed test): Questo test verifica se esiste una differenza significativa nella precisione tra i due sistemi. La domanda alla quale risponde è: "La precisione dei due sistemi è diversa?" Un test a due code esplora entrambe le direzioni della differenza, ossia, se uno dei due sistemi ha una precisione superiore o inferiore rispetto all'altro senza presupporre in anticipo quale sistema potrebbe essere migliore.

Test a una coda (One-tailed test): Questo test valuta se un sistema è significativamente migliore dell'altro. La domanda alla quale risponde è: "Il sistema A è migliore del sistema B?" Un test a una coda esplora una sola direzione, quindi si utilizza quando si ha un'ipotesi specifica su quale dei due sistemi possa essere migliore.

In pratica, la scelta tra test a una o due code dipende dall'obiettivo del confronto:

Se si vuole solo verificare se esiste una differenza senza fare ipotesi su quale sistema possa essere superiore, si utilizza un test a due code.
Se si ha una chiara ipotesi su quale sistema potrebbe essere migliore, si utilizza un test a una coda.



Distribuzione nulla: La curva rappresenta la distribuzione della statistica t sotto l'ipotesi nulla (ossia, assumendo che non ci sia una differenza significativa tra i due sistemi). La distribuzione è simmetrica, centrata attorno a zero.

Valore p (p-value): Indica quanto è distante il valore della statistica t osservata dalle regioni centrali della distribuzione. In altre parole, rappresenta la probabilità di ottenere un valore della statistica t così estremo (o più) sotto l'ipotesi nulla.

Decisione sull'ipotesi nulla: Se il valore p è sufficientemente piccolo (tipicamente inferiore a una soglia come 0.05), si conclude che è improbabile ottenere un valore della statistica t così estremo per caso, e si rifiuta quindi l'ipotesi nulla. Questo suggerisce che la differenza osservata tra i due sistemi potrebbe essere significativa.

Test a due code: Per un test a due code, il valore p rappresenta la massa di probabilità nelle due code della distribuzione (a sinistra e a destra), come mostrato in figura. Se la statistica t si trova in una delle due code estreme (valori molto positivi o negativi), suggerisce che esiste una differenza significativa tra i sistemi.

Def :

COEFFICIENTE DI DETERMINAZIONE
COEFFICIENT OF DETERMINATION

Il coefficiente di determinazione (R^2), una misura statistica utilizzata per valutare quanto bene una variabile indipendente può predire una variabile dipendente in un modello di regressione.

Significato di R^2

R^2 rappresenta la proporzione della varianza della variabile dipendente che può essere predetta dalla variabile indipendente. In altre parole, R^2 indica quanta parte della variazione totale può essere spiegata dalla regressione. Un R^2 più alto significa che il modello spiega una maggiore percentuale della varianza totale della variabile dipendente.

1 DEV(T) = DEVIAZIONE TOTALE
2 DEV(R) = DEVAZIONE DELLA REGRESSIONE
3 DEV(E) = DEVIAZIONE RESIDUA

$$(1)\quad Dev(T) = \sum_{i=1}^{m} \left(y^{(i)} - \bar{y}\right)^2$$

$y^{(i)}$ = Valori osservati

$\bar{y}$ = media

$$(2)\quad Dev(R) = \sum_{i=1}^{m} \left(y^{(i)*} - \bar{y}\right)^2$$

$y^{(i)*}$ = VALORI PREDETTI

$$(3)\quad Dev(E) = \sum_{i=1}^{m} \left(y^{(i)} - y^{(i)*}\right)^2$$

IN DEFINITIVA:

$$R^2 = \frac{Dev(R)}{Dev(T)} = 1 - \frac{Dev(E)}{Dev(T)} =$$

$$= \frac{\mathrm{cov}(x,y)^2}{Dev(x)\cdot Dev(y)}$$

Il valore di R^2 (coefficiente di determinazione) varia nell’intervallo [0, 1] e indica quanto bene i risultati osservati sono rappresentati dal modello.

- Valore vicino a 1: Un R^2 prossimo a 1 indica che il modello riesce a spiegare la maggior parte della variazione nei dati osservati, il che significa che il modello si adatta bene ai dati.
- Interpretazione: Il valore di R^2 rappresenta la proporzione della variazione totale nei risultati osservati che viene spiegata dal modello. Ad esempio, un R^2 di 0,8 indica che l'80% della variazione dei dati è spiegato dal modello, mentre il restante 20% è attribuibile a variazioni non spiegate o errori casuali.

.