Questi esempi vi potranno essere utili come riferimento nella ricerca di intervalli di confidenza e test di ipotesi statistiche. Per gli aggiornamenti potete visitare i siti www.boch.net o www.feanor.com.
Per dubbi o assistenza potete inviare una e-mail a info@boch.net o info@feanor.com

Indice degli argomenti

1. Distribuzioni campionarie legate alla distribuzione normale

Il concetto di distribuzione campionaria è il seguente:
dalla popolazione di partenza, supposta gaussiana per semplicità, estraiamo campioni di determinata ampiezza n.
Per ognuno di questi campioni possiamo calcolare, come visto in precedenza, il valore medio $\overline{X}$ e la deviazione standard campionaria s (in italiano la 'deviazione standard' viene chiamata '*scarto tipo*') .
Con il procedere dell'estrazione e l'accumularsi dei campioni, noteremo che anche i valori di $\overline{X}$ ed s calcolati si distribuiscono in un certo modo (sappiamo già, ad esempio, che la distribuzione dei valori medi è anch'essa gaussiana ma con una deviazione standard σ *inferiore* a quella della popolazione e pari, per la precisione, a $\sigma/\sqrt{n}$).
Queste distribuzioni sono le distribuzioni campionarie.

Le formule del valore medio e della deviazione standard campionaria sono tra le statistiche più semplici ed intuitive. Per la prima si sommano i dati estratti e si divide per n, per l'altra si calcola una sommatoria di quadrati di differenze, si divide per n-1 e si prende la radice. Queste operazioni (sommare, elevare al quadrato, etc.) sono due esempi di funzione applicata ai dati del campione, ma ce ne possono essere infinite altre.

Due distribuzioni campionarie molto importanti sono:

- la statistica t di Student
- la statistica χ^2 (Chi quadro)

Cerchiamo di capire come si ottengono queste statistiche ed a cosa possono essere utili. Diamone innanzitutto la definizione.

Se un campione casuale di n osservazioni y1, y2, y3, ... è estratto da una distribuzione normale con media μ e varianza σ^2, allora la distribuzione campionaria della statistica

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

ha una densità di distribuzione chi-quadro con (n-1) gradi di libertà.

In parole: Se si moltiplica il quadrato della deviazione standard campionaria per (n-1) e si divide per la varianza della popolazione, si ottiene una variabile distribuita con una determinata densità , chiamata "densità di distribuzione chi-quadro"

Sia z una variabile casuale distribuita secondo la normale standard e χ^2 una variabile casuale distribuita secondo la chi-quadro con ν = n gradi di libertà.
Se z e χ^2 sono indipendenti, allora la statistica

$$t = \frac{z}{\sqrt{\chi^2/\nu}}$$

ha una densità di distribuzione chiamata t di Student (o semplicemente distribuzione t) con ν gradi di libertà.

Si può dimostrare che vale la relazione:

$$t = \frac{z}{\sqrt{\chi^2 / \nu}} = \frac{\bar{y} - \mu}{s / \sqrt{n}}$$

Questa relazione è molto importante, perché fa vedere che la statistica t dipende solo dalla media della distribuzione da cui si è estratto il campione.

In parole: Se si sottrae al valore medio campionario la media m della popolazione e si divide per il rapporto tra la deviazione standard campionaria e la radice di n, si ottiene una variabile distribuita con una determinata densità , chiamata "densità di distribuzione t (o di Student)"

Analogamente, la statistica χ^2 è legata alla sola varianza della distribuzione da cui si è estratto il campione (e lo si vede direttamente dalla formula)

In sostanza, le statistiche ora viste sono utili proprio perché ciascuna è legata ad uno ed uno solo dei parametri che identificano la gaussiana della popolazione.

Quando parleremo degli intervalli di confidenza vedremo che per definire intervalli di confidenza per la media utilizzeremo la distribuzione di Student, mentre per gli intervalli di confidenza della varianza utilizzeremo la statistica χ^2.

2. La stima dei parametri di una distribuzione

Abbiamo visto, durante il corso relativo al Controllo Statistico di Processo, come le misure di caratteristiche industriali (lunghezza, peso, etc.) siano dotate (e così il processo che le genera) di una distribuzione di probabilità.

Quasi sempre nell'ambiente produttivo si ha a che fare con una approssimazione della distribuzione di probabilità gaussiana o normale, definita dai due parametri media e deviazione standard.

Per conoscere il valore esatto di tali parametri occorrerebbe conoscere tutti i valori costituenti la popolazione. Non essendo ciò possibile, occorre "stimare" i valori di questi parametri basandosi sui pochi dati disponibili.

La stima dei parametri può essere "puntuale", nel senso che ci porta ad assegnare un valore definito, in base ai dati del campione estratto, al parametro incognito, oppure in termini di "intervallo fiduciario".
In tal caso si troverà un intervallo, la cui ampiezza dipende dal livello di rischio prescelto, che conterrà il vero valore del parametro incognito con una certa probabilità (definizione imprecisa ma intuitiva, diventerà più chiara quando prenderemo in esame gli esempi presentati nel seguito).

3. Intervallo di confidenza bilatero per la frazione p di una popolazione

Spesso è necessario determinare un intervallo bilatero di confidenza per la frazione o percentuale p di una popolazione, come ad esempio nel caso del livello di difettosità p di un lotto (rapporto tra pezzi non conformi e pezzi costituenti il lotto).
Vediamo un esempio (ricavato dal Journal of Production Engineering, Jan. 1986)

☟ ESEMPIO:

Un progetto comune di ricerca tra USA e Giappone prevedeva l'analisi di un nuovo tipo di struttura in cemento armato. La struttura venne sottoposta a prove e venne chiesto ai tecnici americani di valutare le proprietà della nuova struttura. Su 48 tecnici interpellati, 36 ritennero che la struttura fosse troppo leggera. Trovare un intervallo di confidenza al 95% per la vera frazione di tecnici che ritenne che la struttura fosse troppo leggera.

La frazione è data da

$$\bar{p} = \frac{36}{48} = 0.75$$

L'intervallo di confidenza bilatero al 95% è dato da

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}\bar{q}}{n}}$$

Dalla tabella della distribuzione normale vediamo che l'area pari a 0.025 è alla destra del punto 1.96.
L'intervallo cercato è:

$$0.75 \pm 1.96\sqrt{\frac{0.75 \cdot 0.25}{48}} = 0.75 \pm 0.122 = (0.628,\ 0.872)$$

L'interpretazione dell'intervallo è la seguente: se chiedessimo a numerosi tecnici il loro parere sulla struttura in esame, il 95% delle volte troveremmo una percentuale compresa tra il 62.8% e l'87.2% che afferma che la struttura è troppo leggera.

4. Intervallo di confidenza per la varianza σ^2 di una popolazione

ESEMPIO:

Un responsabile del controllo in un impianto di riempimento barattoli sa che la quantità contenuta nei singoli barattoli può variare, a causa di fattori incontrollabili del processo. La quantità media contenuta è importante, ma altrettanto importante può essere la sua variabilità, data da σ^2. Se σ^2 è troppo grande, alcuni barattoli conterranno troppo ed altri troppo poco prodotto.
Per valutare tale variabilità, vengono selezionati e pesati accuratamente 10 barattoli, ricavando:

$\bar{y} = 7.98 \text{ g}$

s = 0.04 g

Costruire un intervallo di confidenza al 90% per la varianza del processo di riempimento.

L'intervallo cercato è dato dalla relazione seguente:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$$

Occorre fare l'ipotesi che i dati provengano da una popolazione distribuita secondo la densità di probabilità gaussiana.
Si ha inoltre α=0.10 e quindi α/2=0.05 e (1-α/2) =0.95. I gradi di libertà sono (n-1) = 10 -1= 9.
Dalla tavola dei valori di χ^2 si ricava $\chi^2_{0.05}$ = 16.9190 e $\chi^2_{0.95}$ = 3.32511.
Sostituendo i valori nella formula otteniamo:

$$\frac{(10-1)(0.04)^2}{16.9190} \le \sigma^2 \le \frac{(10-1)(0.04)^2}{3.32511}$$

da cui:

$$0.000851 \le \sigma^2 \le 0.004331$$

Questo intervallo va interpretato nel seguente modo: se effettuassimo numerose misure della varianza del processo di riempimento, noteremmo che il 90% delle volte il valore della varianza misurato cade nell'intervallo trovato.
Tale intervallo può dunque essere utilizzato per controllare se la varianza del processo si discosta dai valori impostati.

5. Intervallo di confidenza per la media μ di una popolazione nel caso siano disponibili numerosi dati (n>30)

Nel caso siano disponibili più di 30 misure possiamo, in virtù del teorema centrale limite, assumere che la distribuzione campionaria approssimi una distribuzione normale.
In tal caso un intervallo di confidenza è dato dalla formula seguente:

$$\bar{y} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \approx \bar{y} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

dove $\bar{y}$ è il valore medio campionario e $z_{\alpha/2}$ il valore, preso dalle tabelle della distribuzione normale, che lascia nella coda superiore (ed in quella inferiore) un valore apri ad $\alpha/2$.
Notare l'approssimazione effettuata sostituendo la deviazione standard della popolazione con quella campionaria.

Non è necessario fare l'ipotesi che la distribuzione della popolazione sia normale.

☝ ESEMPIO:

Supponiamo che un centro di calcolo voglia valutare l'affidabilità delle memorie di massa disponibili. Un indice dell'affidabilità è dato dall'intervallo medio di tempo che intercorre tra due errori successivi di lettura. Per stimare questo valore, il centro ha registrato gli intervalli il tempo tra 45 errori successivi, ricavando come tempo medio $\bar{y}$ *= 1762 e s = 215.*
Stimare la vera media del tempo tra errori successivi con una confidenza del 90%.

Abbiamo come livello di confidenza 1-α = 0.9 e quindi α= 0.1, dunque α/2=0.05.
Dalle tabelle ricaviamo il valore $z_{\alpha/2}$ = **1.645**.
Inserendo i valori numerici nella formula, troviamo:

$$\bar{y} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) = 1.762 \pm 1.645 \ \left(\frac{215}{\sqrt{45}}\right) = 1.762 \ \pm \ 52.7$$

Siamo confidenti al 90% che il vero valore della media della popolazione da cui abbiamo estratto il campione sia contenuta tra 1709.3 e 1814.7).

6. Intervallo di confidenza per la media μ di una popolazione nel caso siano disponibili pochi dati (n<30)

Purtroppo non è sempre possibile estrarre numerosi campioni ed utilizzare la formula precedente, basata sulla densità di distribuzione normale.
Come regola pratica, se le misure raccolte sono meno di 30 non si può invocare il teorema centrale limite e sfruttare la formula precedente, ma occorre utilizzare la distribuzione campionaria t di Student.

In questo caso occorre verificare che la distribuzione della popolazione sia normale.
Se questa ipotesi non è soddisfatta il test può non essere significativo.

Un intervallo di confidenza è pertanto dalla formula:

$$\bar{y} \pm t_{\alpha/2}(\frac{s}{\sqrt{n}})$$

dove il valore $t_{\alpha/2}$, preso dalle tabelle della distribuzione t, è quel valore che lascia alla propria destra un'area pari ad $\alpha/2$.
E' importante ricordare, utilizzando la tabella, che il numero di gradi di libertà della t è pari a (n-1).

☟ ESEMPIO:

Si è scoperto che l'aggiunta di un determinato composto al fluido trasportato da una certa condotta permette di ridurre la quantità di calcare depositata sulle pareti.
Un esperimento viene condotto, introducendo il composto e misurando poi la percentuale di calcare presente in 5 soluzioni, dopo averle fatte decantare per 24 ore.
I risultati sono i seguenti:

229 255 280 203 229

Stimare un intervallo per la media del calcare presente nelle 5 soluzioni esaminate, con una confidenza del 99%.

Come prima cosa occorre calcolare il valore medio del campione e la deviazione standard campionaria, trovando $\bar{y}$ = 239.2 e s = 29.3.
Per un livello di confidenza del 99% troviamo 1-α = 0.99 e quindi α= 0.01, dunque $\alpha/2$=0.005.
Poiché il nostro campione è composto solamente da 5 misure, è necessario fare l'ipotesi che la quantità media di calcare contenuta nel fluido segua una distribuzione gaussiana.
Dalle tabelle ricaviamo il valore $t_{\alpha/2}$ = **4.604**.

Sostituendo i valori numerici ricaviamo:

$$\bar{y} \pm t_{\alpha/2}(\frac{s}{\sqrt{n}}) = \bar{y} \pm t_{0.005}(\frac{s}{\sqrt{n}}) = 239.2 \ \pm \ (\frac{29.3}{\sqrt{5}}) = 239.2 \ \pm \ 60.3$$

L'intervallo di confidenza cercato è dunque (178.9, 299.5).
Se è vera l'ipotesi fatta, che il calcare sia distribuito secondo una gaussiana, allora possiamo essere confidenti al 99% che la media sia compresa tra i due valori sopra indicati.

7. Test di ipotesi statistiche

Un test di ipotesi statistiche permette di verificare se i pochi dati estratti (campione) dalla popolazione supportano (o contraddicono) una certa ipotesi, fatta su uno dei parametri che identificano la popolazione incognita.

I test vengono solitamente formalizzati come nell'esempio seguente:

ipotesi nulla Ho : la media è pari a x

ipotesi alternativa Ha : la media è inferiore al valore x

L'ipotesi alternativa potrebbe anche essere la seguente:

ipotesi alternativa Ha : la media è superiore al valore x

oppure

ipotesi alternativa Ha : la media ha un valore diverso da x (è maggiore o minore)

Oltre a fissare le ipotesi, occorre indicare la probabilità di errore del I° tipo α (già vista quando abbiamo studiato le carte di controllo per variabili).

*La **probabilità di errore del I° tipo (α)** è la probabilità di rifiutare (erroneamente) l'ipotesi nulla, quando questa è vera.*

Avendo posto, ad esempio, che Ho : $\mu=\mu_O$, se i dati mi spingono a rifiutare l'ipotesi che la media valga μ_O, (mentre ciò è vero) allora commetto un errore del primo tipo.

In modo analogo, la probabilità di accettare un'ipotesi, quando quest'ultima è falsa, viene indicata come **probabilità di errore del 2° tipo** (errore ß).

Avendo posto, come ipotesi nulla, che Ho : $\mu=\mu_O$, se i dati mi spingono ad accettare questa l'ipotesi, (mentre la media ha in realtà un valore diverso da μ_O) allora commetto un errore del secondo tipo.

La **regione critica** (CR) è la regione (semiretta o segmento) di rifiuto dell'ipotesi nulla Ho. la sua dimensione dipende dalla probabilità di errore del 1° tipo.
Se il valore calcolato dai dati in nostro possesso cade in questa regione, rifiuteremo l'ipotesi nulla, cioè accetteremo l'ipotesi alternativa Ha.

Spesso si nota una certa difficoltà di comprensione su questo punto, dovuta alle frequenti negazioni (es. non accettare un'ipotesi falsa, etc..) ed al diverso modo di formulare le ipotesi iniziali, che capovolge le convenzioni comunemente adottate.

Per facilitare la comprensione della procedura, occorre esaminare con attenzione gli esempi proposti nel seguito e svolgere gli esercizi (lo svolgimento di alcuni esercizi è presentato alla fine delle dispense). E' molto utile risolvere gli esercizi graficamente.

L'ultima definizione da ricordare è la seguente:

Si definisce **potenza del test** la probabilità di rifiutare l'ipotesi nulla Ho, quando questa è, in effetti, falsa. La potenza del test è data dal valore (1-ß).

La potenza del test è un modo per indicare la capacità del test (come dice anche il nome) di farci prendere la decisione giusta (cioè rifiutare l'ipotesi rivelatasi falsa).

Dopo aver visto le formule di calcolo necessarie per effettuare un test di ipotesi statistiche vediamo qualche esempio reale che possa servire da guida per le future applicazioni:

7.1 Test di ipotesi sulla media di una popolazione con campione grande (ampiezza campione n > 30)

ESEMPIO:

La legge impone che il livello massimo di una sostanza dannosa per l'organismo, contenuta nell'acqua potabile, sia pari a 5 ppm (parti per milione). Un impianto produttivo per la realizzazione di circuiti stampati scarica piccole quantità di tale sostanza nelle acque di un fiume. La direzione dell'impianto ordina il blocco dell'impianto nel caso in cui la concentrazione della sostanza dannosa superi 3 ppm.

50 prelievi casuali permettono di ricavare $\bar{y}$ = 3.1 ppm e s = 0.5 ppm.
Questi risultati provano con sufficiente evidenza che è necessario fermare l'impianto produttivo? Verificare al livello di significatività pari a 0.01.

Abbiamo:

Ho : y = 3 ppm
Ha : y > 3 ppm

Poiché abbiamo 50 campioni (n>30) possiamo usare la statistica

$$z = \frac{\bar{y} - 3}{s / \sqrt{n}} = \frac{3.1 - 3}{\frac{0.5}{\sqrt{50}}} = 1.414$$

Cerchiamo la regione critica CR per α =0.01 dalle tabelle della densità normale. Troviamo z_α= **2.326** e la regione critica per il nostro test è data dai valori z superiori a 2.326.

Avendo ricavato z = **1.414**, non abbiamo ragione di rifiutare l'ipotesi nulla Ho. Non abbiamo pertanto motivo di rifutare l'idea che la media sia pari a 3 ppm e non un livello superiore.

7.2 Test di ipotesi sulla media di una popolazione con campione piccolo (ampiezza campione n < 30)

ESEMPIO:

L'effetto di fermi macchina sulle prestazioni di sistemi di produzione automatici viene spesso simulato tramite appositi software. Uno di questi studi era orientato all'analisi di un sistema costituito da una sola macchina utensile con tempo di arrivo medio dei job ad intervalli di 1.25 minuti, una lavorazione di durata fissa e pari ad 1 minuto ed un tempo di fermo macchina pari al 10% del totale.
Dopo 7 simulazioni complete, ciascuna di lunghezza 160 ore (le simulazioni devono essere sufficientemente lunghe per fare in modo che il sistema vada "a regime"), la produzione media per settimana (40 ore lavorative) é $\bar{y}$ = 1908.8 pezzi (il numero frazionario è un risultato tipico della simulazione).
Per un sistema di produzione senza fermi macchina, la produzione media sarebbe di 1920 pezzi.
Sapendo che per le 7 prove si ha s = 20 pezzi per settimana, verificare l'ipotesi che la vera produzione media per settimana sia inferiore a 1920 pezzi. Adottare un livello di significatività pari a 0.05.

Ho : y = 1920 pezzi

Ha : y < 1920 pezzi

Poiché abbiamo effettuato solo 7 simulazioni complete (n<30) dobbiamo usare la statistica t (Student), con riferimento all' α assegnato.
Abbiamo:

$$t = \frac{\overline{y} - y_o}{s / \sqrt{n}} = \frac{1908.8 - 1920}{\frac{20}{\sqrt{7}}} = -1.482$$

mentre dalle tabelle della distribuzione t troviamo, per 6 gradi di libertà (n-1) il valore $t_\alpha = t_{0.05} = 1.943$. Ricordiamo che la distribuzione di Student è simmetrica rispetto allo zero.
L'area a destra del punto **1.943** è pari a quella a sinistra del punto **-1.943**.
Essendo l'ipotesi alternativa del tipo y **<** 1920 la regione critica CR è data dai valori a sinistra del punto **-1.943**.
Il valore da noi ricavato era t = **- 1.482**, che non appartiene alla regione critica.
Non abbiamo pertanto ragione di rifiutare l'ipotesi nulla Ho e concludiamo che la produzione settimanale è, a quel livello di significatività, effettivamente di 1920 pezzi.

7.3 Test di ipotesi unilatero sulla differenza tra le medie μ di due popolazioni, osservazioni non accoppiate (campioni indipendenti ed entrambi con n > 30)

Questo tipo di test unilatero viene svolto per verificare se esista una differenza tra le medie di due popolazioni, formalizzando così le ipotesi nulla ed alternativa:

ipotesi nulla $H_0: (\mu_1 - \mu_2) = D_o$

ipotesi alternativa $H_a: (\mu_1 - \mu_2) > D_o$ {oppure $H_a: (\mu_1 - \mu_2) < D_o$}

e si svolge calcolando il valore della statistica:

$$z \approx \frac{(\overline{y}_1 - \overline{y}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

e confrontandolo con lo z_α determinato in base al livello di confidenza richiesto.
La regione critica dipende dal modo in cui è stata formulata l'ipotesi alternativa.
Se $z > z_\alpha$ {oppure $z < -z_\alpha$} rifiuterò l'ipotesi nulla.

☝ ESEMPIO:

Per ridurre i costi è stato sviluppato un nuovo processo di panificazione. E' di interesse stabilire se il nuovo processo permetta di diminuire il numero di calorie contenute.
I dati sono i seguenti (1 =nuovo processo, 2 = processo precedente):

$n_1 = 50$	$n_2 = 30$
$\overline{y}_1 = 1255$ cal	$\overline{y}_2 = 1330$ cal
$s_1 = 215$ cal	$s_s = 238$ cal

Poniamo:

ipotesi nulla $H_0:(\mu_1-\mu_2)=0$

ipotesi alternativa $H_a:(\mu_1-\mu_2)<0$

La regione critica, al livello 0.05, è data da z_α= **- 1.645.**

Inserendo i dati numerici nella statistica otteniamo:

$$z \approx \frac{(\overline{y_1}-\overline{y_2})-D_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{(1255-1330)}{\sqrt{\frac{(215)^2}{50}+\frac{(238)^2}{30}}} = -1.41$$

Il valore z ricavato NON cade nella regione critica, essendo $z > -z_\alpha$. Concludo che il campione esaminato NON fornisce con sufficiente evidenza la prova che il nuovo processo permetta di diminuire il contenuto calorico del prodotto.

7.4 Test di ipotesi sulla differenza tra le medie μ di due popolazioni, osservazioni accoppiate

☟ ESEMPIO:

Vengono analizzati 4 esemplari femmina di rana trovate in una zona contaminata da TCDD, un composto nocivo presente negli scarichi di acque industriali.
In particolare viene misurata la quantità (in ppm) di contaminante presente nel fegato e nelle ovaie, compilando la tabella seguente. I ricercatori affermano che "il livello medio di TCDD presente nelle ovaie delle rane è superiore a quello presente nel fegato".
Verificare questa affermazione al 5%.

rana N°	*1*	*2*	*3*	*4*
fegato	*11.0*	*14.6*	*14.3*	*12.2*
ovaie	*34.2*	*41.2*	*32.5*	*26.2*

E' chiaro il motivo per cui si debba effettuare il test "per osservazioni accoppiate".
Dobbiamo verificare se le popolazioni "fegato" ed "ovaie" differiscono, o sono caratterizzate dalla stessa media.

7.5 Test di ipotesi sulla varianza σ^2 di una popolazione

Questo test viene effettuato nei casi in cui si possa ritenere che la variabilità di un processo produttivo sia cresciuta.

☟ ESEMPIO:

L'esperienza mostra che la variabilità di una linea di produzione di sfere per cuscinetti forniti da un certo produttore è pari a 0.00156 (varianza della popolazione).

Per ridurre i costi il fornitore ha installato un sistema di produzione più economico. La varianza campionaria ricavata misurando 100 sfere estratte casualmente da un lotto di produzione consegnato all'utilizzatore è pari a $s^2 = 0.00211$.
I dati disponibili forniscono sufficiente evidenza di una maggiore variabilità nei diametri delle sfere prodotte con il nuovo metodo?. Verificare ad un livello di confidenza pari ad $\alpha = 0.05$.

Formalizziamo il problema nel seguente modo:

ipotesi nulla $H_0: \sigma_2^2 = \sigma_1^2$ la varianza è rimasta costante
ipotesi alternativa $H_a: \sigma_2^2 > \sigma_1^2$ la varianza è aumentata

cioè:

ipotesi nulla $H_0: \sigma^2 = 0.00156$
ipotesi alternativa $H_a: \sigma^2 > 0.00156$

Dovremo esaminare i dati sfruttando una certa statistica, che dipenda solo dalla varianza. Per quanto visto nelle pagine seguenti, sappiamo che la statistica

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

dipende solo dalla varianza ed è distribuita secondo la chi-quadro.

Avendo posto come ipotesi alternativa una varianza superiore, saremo portati a rifiutare l'ipotesi nulla se il valore di χ^2 calcolato dai dati supera (cioè si trova più a destra) di un certo valore di soglia. La regione di rifiuto dell'ipotesi nulla è detta anche **regione critica** CR.

Questo livello di soglia χ^2_α lo leggeremo dalla tabella in funzione del valore di α. Nel nostro caso è $\alpha = 0.05$.

Riassumendo, rifiuteremo l'ipotesi Ho se si verifica la relazione $\chi^2 > \chi^2_{0.05}$ = **123.225**.

Passando ai numeri otteniamo:

$$\chi^2 = \frac{(100-1)(0.00211)}{0.00156} = 133.90$$

Essendo 133.90 > 123.225 , il valore 133.90 cade nella regione critica, in cui si deve rifiutare l'ipotesi nulla.
Dobbiamo dunque concludere che, con confidenza al 95%, la varianza del processo produttivo è cresciuta rispetto al valore 0.00156.