

Verifica di ipotesi

Cicchitelli cap. 18

A.A. 2010/2011

Stima vs Test di ipotesi

- Nell'inferenza statistica l'oggetto di interesse è un parametro (= quantità definita su una certa popolazione)
- Per fare inferenza si utilizza un modello probabilistico che spesso consiste in una famiglia parametrica di v.a.: la distribuzione del carattere nella popolazione viene rappresentata da una v.a. X la cui distribuzione ha un parametro incognito di interesse primario θ , oltre ad eventuali altri parametri di disturbo:
 - es. se $X \sim N(\mu, \sigma^2)$ e l'interesse verte sulla media della popolazione, allora μ è il parametro di interesse primario, mentre σ^2 è un parametro di disturbo

Stima vs Test di ipotesi (continua)

- Quando si effettua una **stima** dei parametri, puntuale o per intervallo, l'informazione campionaria viene utilizzata per stimare un parametro incognito del modello scelto per X
 - Prima leggiamo i dati, poi facciamo un'affermazione sul parametro
- Quando invece si effettua un **test delle ipotesi** sui parametri, l'informazione campionaria viene utilizzata per decidere se accettare o rifiutare una certa ipotesi concernente un parametro incognito del modello scelto per X
 - Prima facciamo un'affermazione sul parametro, poi leggiamo i dati (per vedere se supportano l'affermazione)

Paradigma

- Tutte le procedure inferenziali che consideriamo in questo corso (fra cui i test di ipotesi) fanno riferimento al paradigma dell'**inferenza parametrica** basata su **campione casuale**:

$$X \sim f(x; \theta) \quad e \quad X_1, \dots, X_n \sim iid - X$$

Inferenza parametrica: f è una funzione (di massa o di densità di probabilità) nota a meno di un parametro θ (che può essere uno scalare, es. nella Bernoulli $\theta = p$, oppure un vettore, es. nella Normale $\theta = (\mu, \sigma)$)

Campione casuale: gli elementi campionari sono v.a. indipendenti e identicamente distribuite come X

Ipotesi statistica parametrica

- Un'**ipotesi statistica parametrica** è un'affermazione relativa ad un parametro (= quantità definita sulla popolazione di interesse) o, in modo equivalente, un'affermazione relativa ad un parametro che caratterizza la distribuzione della v.a. adottata come modello
- L'ipotesi statistica in oggetto è detta **ipotesi nulla** (o ipotesi di lavoro) ed è indicata con H_0
 - Se $X \sim \text{Be}(p)$ possibili ipotesi sono $H_0: p = 0.25$ $H_0: p \leq 0.25$ $H_0: p \geq 0.25$
 - Se $X \sim N(\mu, \sigma^2)$ possibili ipotesi sono $H_0: \mu = 360$ $H_0: \mu \leq 360$ $H_0: \mu \geq 360$ (analogamente, si possono formulare ipotesi su σ^2)

Nell'ipotesi nulla non si può mettere il segno di disuguaglianza stretta: è possibile $\leq$ ma non $<$, è possibile $\geq$ ma non $>$

Ipotesi alternativa

- Dato il modello probabilistico per X , una volta definita l'ipotesi nulla H_0 risulta automaticamente definita anche un'**ipotesi alternativa**, indicata con H_1 , che usualmente è il complemento di H_0 nello spazio parametrico in questione
 - Se $X \sim N(\mu, \sigma^2)$ lo spazio parametrico (= insieme dei possibili valori del parametro μ) è l'insieme dei numeri reali per cui
 - $H_0: \mu = 360 \rightarrow H_1: \mu \neq 360$
 - $H_0: \mu \leq 360 \rightarrow H_1: \mu > 360$
 - $H_0: \mu \geq 360 \rightarrow H_1: \mu < 360$
 - Se $X \sim \text{Be}(p)$ lo spazio parametrico (= insieme dei possibili valori del parametro p) è l'intervallo $(0,1)$ per cui
 - $H_0: p = 0.25 \rightarrow H_1: p \neq 0.25$
 - $H_0: p \leq 0.25 \rightarrow H_1: p > 0.25$
 - $H_0: p \geq 0.25 \rightarrow H_1: p < 0.25$è sottinteso che comunque $0 < p < 1$

Statistica 2010/2011

6

Ipotesi nulla e ipotesi alternativa

- Si assume che le due ipotesi H_0 e H_1 siano **esaustive** (= una delle due è necessariamente vera) e **mutuamente esclusive** (= non possono essere vere contemporaneamente)
- Talvolta l'unione delle due ipotesi H_0 e H_1 non coincide con l'intero spazio parametrico. Ciò può corrispondere a
 - una effettiva restrizione dello spazio parametrico, perché si è deciso di escludere a priori un sottoinsieme di valori del parametro; ad es. se si contrastano $H_0: \mu = 360$ vs $H_1: \mu = 370$ significa che a priori sono stati esclusi tutti i valori di μ diversi da 360 e 370
 - una apparente restrizione dello spazio parametrico, perché vedremo in seguito che ad es. $H_0: \mu \leq 360$ vs $H_1: \mu > 360$ conduce allo stesso test di $H_0: \mu = 360$ vs $H_1: \mu > 360$ per cui in questo contesto H_0 si può scrivere indifferentemente con $\leq$ o con $=$

7

Tipi di ipotesi

- In generale le ipotesi si distinguono in
 - **Puntuali** (segno di $=$)
 - **Unidirezionali** (segni di $<$, $\leq$, $>$, $\geq$)
 - **Bidirezionali** (segno di $\neq$)

Statistica 2010/2011

8

Test di ipotesi /1

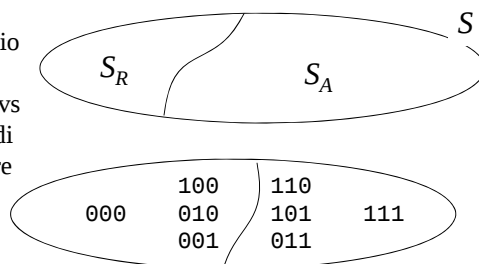
- Un **test di ipotesi** è una *regola* attraverso la quale decidere se **accettare o rifiutare l'ipotesi nulla H_0 sulla base di un campione casuale** ($X_1, \dots, X_n$)
- Concretamente, tale *regola* è attuata nel modo seguente. Indicato con S l'**universo dei campioni** (cioè l'insieme di tutti i possibili campioni ($x_1, \dots, x_n$) di dimensione fissata n che si possono estrarre da X), un *test delle ipotesi* consiste nel suddividere l'insieme S in due sottoinsiemi disgiunti S_A e S_R in modo tale che:
 - se ($x_1, \dots, x_n$) cade in $S_A \rightarrow$ si accetta l'ipotesi H_0
 - se ($x_1, \dots, x_n$) cade in $S_R \rightarrow$ si rifiuta l'ipotesi H_0 (e quindi si accetta H_1)

Test di ipotesi /2

- S_A è detto **regione di accettazione**
- S_R è detto **regione di rifiuto** o **regione critica**
- In sostanza, si tratta di stabilire *a priori* (cioè prima di osservare il campione)
 - quali campioni portano ad accettare H_0
 - e quali portano a rifiutare H_0 (e quindi ad accettare H_1)

Bipartizione dello spazio dei campioni

Ad es. se $X \sim \text{Be}(p)$ e $n=3$ lo spazio dei campioni contiene $2^3=8$ elementi. Per testare $H_0: p \geq 0.5$ vs $H_1: p < 0.5$ un possibile criterio di separazione è quello di assegnare a S_R tutti i campioni con un numero di successi minore o uguale a 1



- Il problema è: come tracciare la linea che separa S_R da S_A ?
- Specificare un test di ipotesi significa specificare un criterio che consenta di assegnare ogni possibile campione ad una delle due categorie:
 - S_A : campioni che supportano H_0 e quindi portano ad accettarla
 - S_R : campioni che non supportano H_0 e quindi portano a rifiutarla

Tipi di errore /1

- Stabilire un criterio per assegnare i campioni a S_R o S_A è un problema decisionale in condizioni di incertezza e quindi il criterio ottimale è quello che **minimizza i rischi di errore**
- Poiché si contrastano due ipotesi, H_0 e H_1 , si possono commettere due tipi di errore:

Decisione	Verità	
	H_0	H_1
$(x_1, \dots, x_n) \in S_A \rightarrow H_0$	ok	errore di II tipo
$(x_1, \dots, x_n) \in S_R \rightarrow H_1$	errore di I tipo	ok

Tipi di errore /2

- I due tipi di errore sono concettualmente diversi ed hanno conseguenze diverse
- Ad esempio, in un test di laboratorio per l'individuazione di un certo virus se si pone
 - $H_0 = \{\text{negativo (= soggetto sano)}\}$
 - $H_1 = \{\text{positivo (= soggetto malato)}\}$

si hanno i seguenti possibili errori:

- **I tipo** (vero H_0 ma si decide H_1): il soggetto è sano ma il test dice che è malato (falso positivo)
- **II tipo** (vero H_1 ma si decide H_0): il soggetto è malato ma il test dice che è sano (falso negativo)

Probabilità degli errori

- Consideriamo innanzitutto il caso in cui sia H_0 che H_1 siano ipotesi puntuali (es. $H_0: \mu=10$ vs $H_1: \mu=7$)
- Per ogni possibile test, cioè per ogni possibile criterio di assegnazione dei campioni a S_R e S_A , risultano definite anche le probabilità di commettere un errore di I e II tipo
 - **Probabilità errore I tipo** (vero H_0 ma si decide H_1):
 $\alpha = P(\text{rifiutare } H_0 \mid H_0) = P((X_1, \dots, X_n) \in S_R \mid H_0)$
 - **Probabilità errore II tipo** (vero H_1 ma si decide H_0):
 $\beta = P(\text{accettare } H_0 \mid H_1) = P((X_1, \dots, X_n) \in S_A \mid H_1)$
- Inoltre, si definisce **potenza del test** la probabilità di rifiutare correttamente H_0
 $P(\text{rifiutare } H_0 \mid H_1) = 1 - P(\text{accettare } H_0 \mid H_1) = 1 - \beta$

Probabilità degli errori: a priori

A priori, cioè prima di estrarre il campione, si hanno le seguenti probabilità:

- Caso H_0 vera:
 - Se esce un campione appartenente a S_R il test sbaglia (**errore I tipo**) $\rightarrow$ prob. α
 - Se esce un campione appartenente a S_A il test indovina $\rightarrow$ prob. $1-\alpha$
- Caso H_1 vera:
 - Se esce un campione appartenente a S_R il test indovina $\rightarrow$ prob. $1-\beta$
 - Se esce un campione appartenente a S_A il test sbaglia (**errore II tipo**) $\rightarrow$ prob. β

Probabilità degli errori: a posteriori

A posteriori, cioè una volta osservato il campione, la situazione è la seguente:

- campione $\in S_R \rightarrow$ si rifiuta $H_0 \rightarrow$ o la decisione è corretta o si è commesso un **errore di I tipo** (ma non lo si può sapere)
- campione $\in S_A \rightarrow$ si accetta $H_0 \rightarrow$ o la decisione è corretta o si è commesso un **errore di II tipo** (ma non lo si può sapere)
- A posteriori non vi sono probabilità, analogamente a quanto visto per gli intervalli di confidenza: o la procedura ha fatto la cosa giusta oppure ha fallito, ma non si può sapere

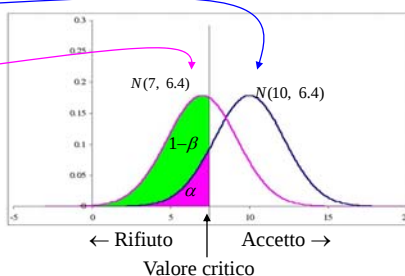
Test per $H_0:\mu=10$ vs $H_1:\mu=7$

- Supponiamo che il carattere X abbia distribuzione Normale con varianza nota, poniamo $\sigma^2 = 64$, per cui l'unico parametro incognito è la media $\mu: X \sim N(\mu, 64)$
- Consideriamo il seguente problema di test: $H_0:\mu=10$ vs $H_1:\mu=7$; questo significa che vi sono due possibilità: o $X \sim N(10, 64)$ o $X \sim N(7, 64)$
- Se si potesse osservare l'intera popolazione sarebbe banale stabilire quale delle due ipotesi è vera; ma disponendo solo di un campione la situazione è caratterizzata da incertezza e la risposta è soggetta ad errore
- Effettuare un test di ipotesi significa:
 - stabilire un criterio di decisione (assegnare i campioni a S_R e S_A)
 - estrarre un campione, vedere se appartiene a S_R o S_A e decidere di conseguenza

Test per $H_0:\mu=10$ vs $H_1:\mu=7$ (continua)

- Siccome le ipotesi riguardano la media μ e poiché la media campionaria è lo stimatore più efficiente di μ , il criterio di assegnazione dei campioni a S_R e S_A dovrebbe basarsi sulla media campionaria:
 - i campioni la cui media è prossima a 10 parlano in favore di H_0 e quindi vanno messi in S_A
 - viceversa, i campioni la cui media è prossima a 7 parlano in favore di H_1 e quindi vanno messi in S_R
- Il problema è stabilire il **valore critico**, cioè il valore della media campionaria per il quale a sinistra si rifiuta H_0 e a destra si accetta H_0
 - Una possibile soluzione è scegliere come valore critico il punto equidistante tra H_0 e H_1 (nell'esempio sarebbe 8.5); tuttavia questa soluzione non consente di controllare le probabilità di errore e quindi non viene adottata

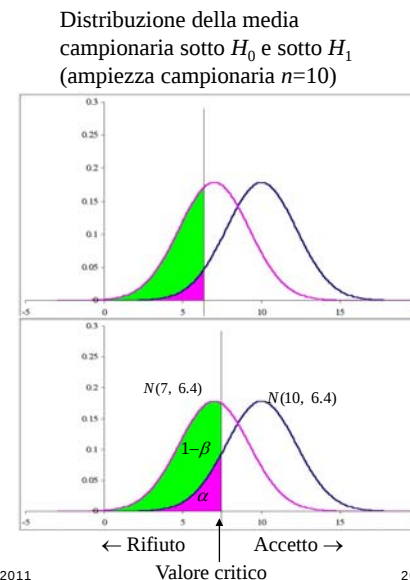
Test per $H_0:\mu=10$ vs $H_1:\mu=7$ (continua)

- Nell'esempio poniamo $n=10$
- $X \sim N(\mu, 64) \rightarrow \bar{X} \sim N(\mu, \frac{64}{n})$
 - Se è vera $H_0:\mu=10 \rightarrow \bar{X} \sim N(10, 6.4)$
 - Se è vera $H_1:\mu=7 \rightarrow \bar{X} \sim N(7, 6.4)$
- La media sotto H_1 è a sinistra di quella sotto $H_0 \rightarrow$ la regione di rifiuto è a sinistra del valore critico
- 

α = area sotto la curva $H_0:\mu=10$ a sinistra del valore critico
 $1-\beta$ = area sotto la curva $H_1:\mu=7$ a sinistra del valore critico
 β = area sotto la curva $H_1:\mu=7$ a destra del valore critico

Test per $H_0:\mu=10$ vs $H_1:\mu=7$ (continua)

- Spostando il valore critico verso sinistra α si riduce, ma contemporaneamente anche $1-\beta$ si riduce (ovvero β aumenta)
- Viceversa se si sposta il valore critico verso destra
- Esiste dunque un trade-off tra α e β per cui quando uno si riduce l'altro aumenta



Statistica test

- L'assegnazione dei campioni a S_R e S_A avviene per mezzo di una **statistica test**, cioè una v.a. T_0 con le seguenti caratteristiche:
 - T_0 è funzione del campione e eventualmente anche del valore del parametro ipotizzato in H_0
 - T_0 ha distribuzione nota **sotto H_0** (cioè se si assume vera H_0 allora la distribuzione di T_0 non ha parametri incogniti)

Statistica test: media campionaria

- Nell'esempio precedente la statistica test impiegata è la media campionaria:
 - è funzione del campione
 - sotto H_0 ha distribuzione $N(10,6.4)$, quindi senza parametri incogniti
- La media campionaria è stata usata per assegnare i campioni a S_R e S_A nel seguente modo:
 - Si assegnano a S_A i campioni con media superiore (o =) al valore critico **c**

$$(x_1, \dots, x_n) \in S_A \Leftrightarrow \bar{X} \geq c$$

- Si assegnano a S_R i campioni con media inferiore al valore critico **c**

$$(x_1, \dots, x_n) \in S_R \Leftrightarrow \bar{X} < c$$

Media campionaria standardizzata

- Nell'esempio precedente una *statistica test equivalente* (nel senso che alla fine porta alla stessa assegnazione dei campioni a S_R e S_A) è la **media campionaria standardizzata basata sulla media della popolazione ipotizzata in H_0**

$$Z_0 = \frac{\bar{X} - 10}{\sqrt{6.4}} \stackrel{H_0}{\sim} N(0,1)$$

- è funzione del campione e del valore del parametro in H_0 ($\mu = 10$)
- sotto H_0 ha distribuzione $N(0,1)$, quindi senza parametri incogniti
- La notazione $\boxed{\sim}^{H_0}$ indica che Z_0 ha distribuzione $N(0,1)$ se H_0 è vera, ma non se è falsa (questo fatto viene segnalato anche dallo 0 al pedice di Z)
- Nell'esempio precedente, **se** H_0 è falsa e quindi è vera $H_1: \mu = 7$ **allora** Z_0 non ha distribuzione $N(0,1)$ perché in Z_0 invece di sottrarre la media giusta (7) si sottrae la media sbagliata (10)

Approccio classico /1

- Alla luce del trade-off tra α e β l'approccio classico prevede di **controllare l'errore di I tipo**, fissando il livello α ad un valore piccolo, e facendo il possibile per contenere l'errore di II tipo, cioè minimizzando β
 1. si fissa la probabilità α ad un valore arbitrario piccolo (di solito 0.01 o 0.05)
 2. tra tutte le regioni di rifiuto S_R che portano al livello α desiderato, si sceglie quella ottimale, cioè quella a cui è associato il livello β più piccolo (o, in modo equivalente, il livello $1-\beta$ più grande: da qui il nome di "**criterio della massima potenza**")

Approccio classico /2

- I passi per determinare la regione di rifiuto sono:
 - Scelta di una statistica test $\rightarrow$ a questo proposito vi sono dei risultati teorici che suggeriscono qual è la migliore
 - Scelta della forma della regione di rifiuto (a sinistra del valore critico, a destra del valore critico, ...) $\rightarrow$ la forma della regione di rifiuto viene suggerita dall'ipotesi alternativa H_1
 - Determinazione del valore critico c che porta al livello α desiderato

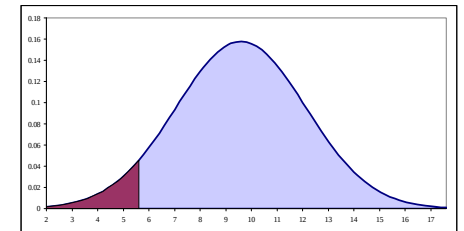
Statistica 2010/2011

25

Test per $H_0:\mu=10$ vs $H_1:\mu=7$ (continua)

- La migliore statistica test è la media campionaria (questo fatto, che è intuitivo, è dimostrabile analiticamente)
- Poiché $\alpha = P(\text{rifiutare } H_0 \mid H_0)$, per determinare la regione di rifiuto si deve fare riferimento allo scenario H_0 , cioè considerare la distribuzione della media campionaria sotto H_0 , che nell'esempio (assumendo campioni di ampiezza $n=10$) è $N(10, 6.4)$
- La media ipotizzata sotto H_1 ($\mu=7$) è più piccola della media ipotizzata sotto H_0 ($\mu=10$) $\rightarrow$ la migliore regione di rifiuto è costituita dai valori della media campionaria inferiori ad un certo valore critico (questo fatto, che è intuitivo, è dimostrabile analiticamente)

Supponiamo di scegliere $\alpha = 0.05 \rightarrow$ il calcolo del valore critico consiste nel determinare quel valore che lascia un'area 0.05 sulla coda sinistra della distribuzione $N(10, 6.4)$. Sulla normale standard $z_{0.05} = -1.65 \rightarrow$ il valore critico è $10 + 2.53(-1.65) = 5.83$



Statistica 2010/2011

26

Test per $H_0:\mu=10$ vs $H_1:\mu=7$ (continua)

- In termini di campioni
 - Si assegnano a S_A i campioni con media superiore (o =) al valore critico 5.83

$$(x_1, \dots, x_{10}) \in S_A \Leftrightarrow \bar{X} \geq 5.83$$
 - Si assegnano a S_R i campioni con media inferiore al valore critico 5.83

$$(x_1, \dots, x_{10}) \in S_R \Leftrightarrow \bar{X} < 5.83$$
- In questo modo, se H_0 è vera, vi è una probabilità del 5% che estraendo un campione questo porti a rifiutarla; in altri termini, a priori vi è una probabilità del 5% di commettere un errore di I tipo
- Per calcolare la probabilità di errore di II tipo, β , occorre considerare lo scenario alternativo H_1 , ovvero valutare quello che accade quando la media campionaria ha distribuzione $N(7, 6.4)$

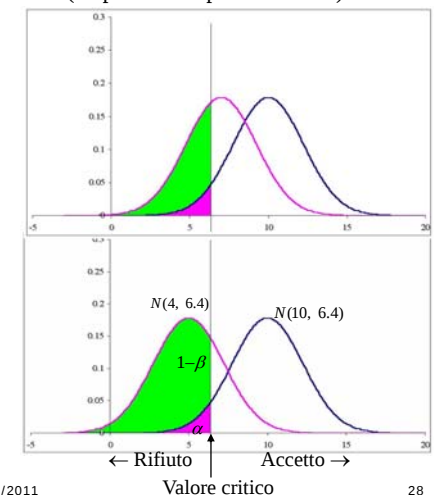
$$\begin{aligned} \beta &= P(\text{accettare } H_0 \mid H_1) = P_{\mu=7}(\bar{X} \geq 5.83) \\ &= P_{\mu=7}\left(\frac{\bar{X} - 7}{\sqrt{6.4}} \geq \frac{5.83 - 7}{\sqrt{6.4}}\right) = P(Z \geq -0.462) = 0.68 \end{aligned}$$

27

β in funzione di H_1

- Quando H_1 si allontana da H_0 diventa più facile discriminare le due ipotesi: α resta invariato ma β diminuisce (ovvero la potenza aumenta)
- Nella figura in basso si ha $H_1: \mu=4$ che porta a $\beta = 0.23$
- Quando H_1 è abbastanza lontana da H_0 la probabilità β diventa approssimativamente zero (e la potenza uno)

Distribuzione della media campionaria sotto H_0 e sotto H_1 (ampiezza campionaria $n=10$)



Statistica 2010/2011

28

Due statistiche test equivalenti

- Nel test per $H_0: \mu=10$ vs $H_1: \mu=7$ si è usata la statistica test

$$\bar{X} \stackrel{H_0}{\sim} N(10, 6.4)$$

che porta a rifiutare H_0 per i campioni con $\bar{X} < 5.83$

- Di solito però si usa la versione standardizzata

$$Z_0 = \frac{\bar{X} - 10}{\sqrt{6.4}} \stackrel{H_0}{\sim} N(0, 1)$$

rifiutando H_0 per i campioni con $z_0 < -1.65$. I due test sono equivalenti poiché hanno la **stessa regione di rifiuto**. Infatti

$$z_0 < -1.65 \Leftrightarrow \frac{\bar{x} - 10}{\sqrt{6.4}} < -1.65 \Leftrightarrow \bar{x} < 10 - 1.65\sqrt{6.4} \Leftrightarrow \bar{x} < 5.83$$

La caratteristica essenziale del metodo scientifico è uno strettissimo legame tra ipotesi e osservazione. I greci erano prolifici nelle ipotesi, ma carenti nell'osservazione. Aristotele, per esempio, pensava che le donne avessero meno denti degli uomini, cosa che non avrebbe pensato se avesse avuto il giusto rispetto per l'osservazione.

Francesco Bacone, al contrario, sopravvalutava la semplice raccolta di dati, supponendo che questa operazione, portata avanti quanto più possibile, avrebbe dato di per sé origine a ipotesi feconde.

Ma esistono tanti dati dell'osservazione e tanti modi di organizzarli, che non si possono raccogliere elementi concreti utilmente, se non si viene precedentemente stimolati da qualche ipotesi ad essi relativa.

Nel corso di un'indagine scientifica, sin dall'inizio vero e proprio, nella mente del ricercatore devono esistere ipotesi generali che determinino la direzione della sua osservazione.

Le ipotesi, comunque, devono modificarsi e svilupparsi continuamente, nel momento in cui nuovi fatti dimostrano che le vecchie ipotesi non sono più adeguate.

Bertrand Russell

Ipotesi alternativa H_1 composta

- Un'ipotesi statistica si dice **composta** quando non specifica completamente la distribuzione di probabilità $\rightarrow$ non esiste una singola distribuzione sotto l'ipotesi, ma più distribuzioni
- Le due tipologie più comuni di H_1 sono la **unidirezionale** (es. $\mu < 10$ oppure $\mu > 10$) e la **bidirezionale** (es. $\mu \neq 10$)
 - entrambe le tipologie sono ipotesi composte, ad es. $H_1: \mu < 10$ significa che H_1 include $\mu = 9.99$, $\mu = 7.5$, $\mu = -2.1$ ecc.

Attenzione: se $X \sim N(\mu, \sigma^2)$ l'ipotesi $H_0: \mu=10$ è

- semplice se la varianza è nota a priori
- composta se la varianza non è nota (in tal caso l'ipotesi identifica l'insieme delle v.a. Normali di media 10)

Ipotesi alternativa H_1 composta (continua)

- Quando H_1 è composta non esiste un singolo valore β della probabilità di errore di II tipo: il valore di tale probabilità dipende dallo specifico punto in H_1 che si considera (è tanto più basso quanto più il punto in esame è lontano da H_0)
 - Nel test $H_0: \mu=10$ vs $H_1: \mu=7$ abbiamo visto che $\beta = 0.68$
 - Nel test $H_0: \mu=10$ vs $H_1: \mu=4$ abbiamo visto che $\beta = 0.23$
 - Pertanto nel test $H_0: \mu=10$ vs $H_1: \mu < 10$ il valore β dipende dallo specifico valore μ che si considera: è 0.68 per $\mu=7$, è 0.23 per $\mu=4$ ecc.
 - Si può dunque calcolare il valore β per ogni $\mu < 10$ (in termini matematici si scrive $\beta(\mu)$ perché β è una funzione di μ)

Ipotesi nulla H_0 composta

- La regione di rifiuto viene determinata usando la distribuzione della statistica test sotto H_0
- Negli esempi precedenti poiché H_0 è puntuale ($\mu=10$) la distribuzione è unica e quindi “sotto H_0 ” implica l'uso della distribuzione con $\mu=10$
- Cosa accade quando H_0 è composta, ad es. in $H_0: \mu \geq 10$ vs $H_1: \mu < 10$? In tal caso “sotto H_0 ” fa riferimento ad una moltitudine di distribuzioni, tutte quelle con media maggiore o uguale a 10

Ipotesi nulla H_0 composta (continua)

- $H_0: \mu \geq 10$ vs $H_1: \mu < 10$
 $\rightarrow H_0$ = tutte le distribuzioni con media maggiore o uguale a 10
- Tuttavia per determinare la regione di rifiuto la distribuzione da considerare è ancora una volta quella con $\mu=10$ perché tale valore sta sulla frontiera: tra i valori in H_0 è quello più vicino a H_1 e quindi è il più sfavorevole perché è quello per il quale le distribuzioni della statistica test sono maggiormente sovrapposte e quindi è più difficile separare H_0 da H_1
- Di conseguenza

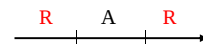
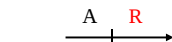
$$H_0: \mu \geq 10 \text{ vs } H_1: \mu < 10 \quad \text{e}$$

$$H_0: \mu = 10 \text{ vs } H_1: \mu < 10$$

portano alla stessa regione di rifiuto e quindi il test è identico

Forma della regione di rifiuto

- θ = parametro oggetto di ipotesi
- θ_0 = valore del parametro alla frontiera tra H_0 e H_1
- T_0 = statistica test adottata, assumendo per semplicità che sia standardizzata, e quindi che abbia media nulla sotto H_0
- Tipicamente T_0 è in relazione diretta con θ , nel senso che al crescere del valore di θ la statistica test T_0 tende ad assumere valori più grandi, e viceversa $\rightarrow$ la forma della regione di rifiuto è suggerita dalla forma di H_1
 - $H_0: \theta \leq \theta_0$ vs $H_1: \theta > \theta_0$
 $\rightarrow$ si rifiuta H_0 quando T_0 è *superiore* ad un certo valore
 - $H_0: \theta \geq \theta_0$ vs $H_1: \theta < \theta_0$
 $\rightarrow$ si rifiuta H_0 quando T_0 è *inferiore* ad un certo valore
 - $H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$
 $\rightarrow$ si rifiuta H_0 quando T_0 è *esterno* ad un certo intervallo



Quale ipotesi in H_0 ?

- L'approccio classico ai test di ipotesi prevede di controllare l'errore di I tipo, scegliendo il valore di α e prendendo il valore di β che ne consegue
- Pertanto in H_0 si dovrebbe mettere l'ipotesi il cui errato rifiuto è più dannoso, in modo da controllare tale rischio
 - Ad esempio, un amico prepara una cena a base di funghi che ha raccolto; essendo l'amico poco esperto, hai il dubbio che i funghi possano essere velenosi. Le ipotesi in gioco sono due, “funghi commestibili” e “funghi velenosi”: dovendo fare un test, quale mettere in H_0 ? Risposta: si dovrebbe porre H_0 : “funghi velenosi”, perché è più grave decidere che sono commestibili quando in realtà sono velenosi piuttosto che decidere che sono velenosi quando in realtà sono commestibili

Quale ipotesi in H_0 ?

(continua)

- Nella ricerca scientifica H_0 è l'ipotesi che si vuole falsificare, cioè l'ipotesi al momento più accreditata e che si vuole sottoporre a verifica empirica
- Nelle valutazioni di efficacia H_0 è l'ipotesi che il prodotto o la procedura in esame non siano efficaci: es. il farmaco non ha effetto, il fertilizzante non ha effetto, l'incentivo economico non ha effetto, il nuovo metodo di produzione non produce miglioramenti.
 - Il principio è il seguente: fino a prova contraria il nuovo non è migliore del vecchio, ovvero, prima di rimpiazzare il vecchio con il nuovo occorre una sostanziale evidenza empirica

Statistica 2010/2011

37

Quale ipotesi in H_0 ?

(continua)

- Nella scrittura delle ipotesi occorre scegliere tra due strutture:
 1. Una puntuale vs una bidirezionale
 2. Due unidirezionali
- Nel primo caso necessariamente si mette in H_0 l'ipotesi puntuale

$$H_0: \theta = \theta_0 \text{ vs } H_1: \theta \neq \theta_0$$

con questa specificazione si rilevano deviazioni da H_0 in entrambe le direzioni. Ad es. $H_0 = \{\text{il risultato del nuovo metodo di produzione è identico al quello del vecchio}\}$, con la specificazione bidirezionale H_0 viene rifiutata sia che il nuovo metodo funzioni meglio sia che funzioni peggio

Quando l'alternativa H_1 è bidirezionale il test si dice **a due code**

Statistica 2010/2011

38

Quale ipotesi in H_0 ?

(continua)

- Nel caso di due ipotesi unidirezionali si hanno due opzioni

$$\text{A) } H_0: \theta \leq \theta_0 \text{ vs } H_1: \theta > \theta_0 \quad \text{B) } H_0: \theta \geq \theta_0 \text{ vs } H_1: \theta < \theta_0$$

in A e B le ipotesi sono scambiate salvo il fatto che il valore θ_0 rimane sempre in H_0 (in questo modo H_0 contiene sempre il punto di frontiera θ_0 : ciò è necessario perché il valore critico viene determinato in modo da produrre il livello α desiderato in corrispondenza del punto più sfavorevole di H_0 , appunto θ_0)

La scelta tra le opzioni A e B va operata in modo che H_0 rappresenti lo status quo, la situazione di inefficacia del nuovo ecc. (quindi dipende dal significato di θ). Ad es. una struttura ricettiva vuole valutare se un certo servizio aggiuntivo aumenta la soddisfazione dei clienti; a tal fine $\theta = \{\text{proporzione di clienti soddisfatti con il servizio aggiuntivo}\}$, $\theta_0 = \{\text{proporzione di clienti soddisfatti senza il servizio aggiuntivo}\} \rightarrow$ si sceglie l'opzione A perché così H_0 significa che il servizio aggiuntivo è inefficace (non aumenta la soddisfazione)

Quando l'alternativa H_1 è unidirezionale il test si dice **a una coda**

Il test Z per la media /1

- Consideriamo un carattere continuo con distribuzione Normale, $X \sim N(\mu, \sigma^2)$, in cui la varianza è nota a priori, e supponiamo di disporre di un campione casuale di ampiezza n
- Siamo interessati a sottoporre a verifica empirica un'ipotesi riguardante la media della popolazione μ , in una delle forme
 - $H_0: \mu \leq \mu_0 \text{ vs } H_1: \mu > \mu_0$
 - $H_0: \mu \geq \mu_0 \text{ vs } H_1: \mu < \mu_0$
 - $H_0: \mu = \mu_0 \text{ vs } H_1: \mu \neq \mu_0$
- La migliore statistica test è la media campionaria, che nella versione standardizzata è

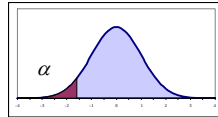
$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \stackrel{H_0}{\sim} N(0,1)$$

40

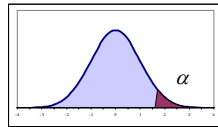
Il test Z per la media /2

- La forma della regione di rifiuto dipende dalla forma dell'alternativa H_1 , mentre il valore critico è determinato dal livello α desiderato:

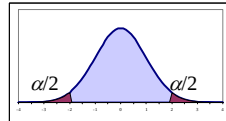
- $H_1: \mu < \mu_0 \rightarrow$ si rifiuta sulla coda sinistra della $N(0,1)$



- $H_1: \mu > \mu_0 \rightarrow$ si rifiuta sulla coda destra della $N(0,1)$



- $H_1: \mu \neq \mu_0 \rightarrow$ si rifiuta su entrambe le code della $N(0,1)$



Statistica 2010/2011

41

Il test Z per la media – Esempio

- Processo produttivo per mettere in scatola i cereali
- Si assume che la quantità in grammi di cereali per scatola abbia distribuzione Normale con deviazione std nota a priori $\sigma = 15$
- Si dispone di un campione casuale di ampiezza $n = 25$
- Poiché la quantità nominale è 368 grammi, si è interessati a verificare ipotesi con $\mu_0 = 368$
- La statistica test da usare è

$$Z_0 = \frac{\bar{X} - 368}{15/\sqrt{25}} \stackrel{H_0}{\sim} N(0,1)$$

Statistica 2010/2011

42

Il test Z per la media – Esempio

(continua)

- Consideriamo innanzitutto il test $H_0: \mu \leq 368$ vs $H_1: \mu > 368$ al livello $\alpha = 0.05$

- Si rifiuta sulla coda destra della $N(0,1)$: il valore critico che lascia 0.05 sulla coda destra è 1.65 per cui la regione di rifiuto è $(1.65, \infty)$

$$(x_1, \dots, x_{25}) \in S_R \Leftrightarrow z_0 > 1.65$$

- A questo punto si guardano i dati: tutto ciò che serve è la media campionaria. Supponiamo che nelle 25 scatole esaminate il peso medio sia pari a 372.5 grammi: $\bar{x} = 372.5$

- Il valore campionario della statistica test Z_0 è

$$z_0 = \frac{372.5 - 368}{15/\sqrt{25}} = +1.5$$

Statistica 2010/2011

43

Il test Z per la media – Esempio

(continua)

- Decisione: poiché $1.5 < 1.65$ il campione estratto non appartiene alla regione di rifiuto, quindi la decisione è di non rifiutare (accettare) H_0
- Questo significa che non vi è abbastanza evidenza empirica contro H_0 e in favore di H_1 .
- In altri termini: alla luce della deviazione std del carattere (15) e dell'ampiezza campionaria (25) una differenza di 4.5 tra media campionaria e media ipotizzata non è abbastanza grande per essere una prova convincente contro H_0 (cioè se la vera media è 368 non è sorprendente osservare una media campionaria di 372.5)

Statistica 2010/2011

44

Il test Z per la media – Esempio (continua)

- Consideriamo adesso il test a due code $H_0: \mu = 368$ vs $H_1: \mu \neq 368$ al livello $\alpha=0.05$
 - Si rifiuta su entrambe le code della $N(0,1)$ per cui 0.05 va diviso a metà: il valore critico che lascia 0.025 sulla coda sinistra è -1.96 e il valore critico che lascia 0.025 sulla coda destra è $+1.96$ per cui la regione di rifiuto è $(-\infty, -1.96) \cup (1.96, \infty)$

$$(x_1, \dots, x_{25}) \in S_R \Leftrightarrow z_0 < -1.96 \text{ oppure } z_0 > 1.96$$

- A questo punto si guardano i dati: tutto ciò che serve è la media campionaria. Supponiamo che nelle 25 scatole esaminate il peso medio sia pari a 372.5 grammi, per cui il valore campionario della statistica test Z_0 è

$$z_0 = \frac{372.5 - 368}{15/\sqrt{25}} = +1.5$$

- Decisione: poiché $-1.96 < 1.5 < 1.96$ il campione estratto non appartiene alla regione di rifiuto, quindi la decisione è di non rifiutare (accettare) H_0

Interpretazione della statistica test Z_0

- La statistica test Z_0 è una misura standardizzata della distanza tra i dati campionari e l'ipotesi nulla

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\text{valore campionario} - \text{valore ipotizzato}}{\text{unità di misura}}$$

- L'unità di misura è l'errore standard della media campionaria: una data differenza tra media campionaria e media ipotizzata è tanto più significativa (contro H_0) quanto più piccolo è l'errore standard

- Supponiamo che nell'esempio precedente la media 372.5 sia stata osservata su un campione di ampiezza $n=100$

$$z_0 = \frac{372.5 - 368}{15/\sqrt{100}} = +3.0 > 1.96$$

Si ha ancora una differenza di 4.5 tra media campionaria e media ipotizzata, ma tale differenza viene giudicata significativa, cioè porta al rifiuto di H_0

Fasi della verifica di ipotesi

(approccio del valore critico)

- Fase 0:** specificare il modello (distribuzione del carattere nella popolazione, tipo di campione)
- Fase 1:** specificare H_0 e H_1
- Fase 2:** scegliere α (e se possibile n così da ottenere il β desiderato)
- Fase 3:** individuare la statistica test (e la sua distribuzione sotto H_0)
- Fase 4:** individuare la forma della regione di rifiuto e determinare i valori critici che portano all' α desiderato
- Fase 5:** raccogliere i dati e calcolare il valore campionario della statistica test
- Fase 6:** prendere la decisione statistica e interpretare il risultato

Approccio del p -value /1

- p -value** (o **livello di significatività osservato**): probabilità, calcolata sotto H_0 , di ottenere un valore della statistica test uguale a o più estremo di quello realmente osservato (detto valore campionario)
- “più estremo” significa “in direzione della regione di rifiuto”
- Nell'esempio precedente il valore campionario della statistica test Z_0 è 1.5 ma il significato di “più estremo” e quindi il p -value dipende dalla forma della regione di rifiuto, la quale dipende da H_1
 - $H_1: \mu > 368 \rightarrow$ si rifiuta sulla coda destra $\rightarrow$ il p -value è l'area a destra di 1.5 sulla $N(0,1)$ e quindi 0.067
 - $H_1: \mu < 368 \rightarrow$ si rifiuta sulla coda sinistra $\rightarrow$ il p -value è l'area a sinistra di 1.5 sulla $N(0,1)$ e quindi 0.933
 - $H_1: \mu \neq 368 \rightarrow$ si rifiuta su entrambe le code $\rightarrow$ il p -value è l'area a destra di 1.5 + l'area a sinistra di -1.5 e quindi 0.134

Approccio del p -value /2

- Il p -value è un indice di sintesi dell'evidenza empirica contro H_0 rispetto a H_1 , nel senso che quanto più il p -value è piccolo tanto maggiore è l'evidenza empirica per rifiutare H_0 in favore di H_1
 - Ad es. $p\text{-value} = 0.245$ significa che, se H_0 fosse vera, estraendo a caso un campione vi sarebbe una probabilità del 24.5% di osservare un valore della statistica test ugualmente o più estremo di quello effettivamente osservato nei dati $\rightarrow$ il valore osservato nei dati non è anomalo e quindi è probabilisticamente compatibile con $H_0 \rightarrow$ i dati non parlano contro H_0 e quindi H_0 non può essere rifiutata
 - Ad es. $p\text{-value} = 0.001$ significa che, se H_0 fosse vera, estraendo a caso un campione vi sarebbe una probabilità di 1 su 1000 di osservare un valore della statistica test ugualmente o più estremo di quello effettivamente osservato nei dati $\rightarrow$ il valore osservato nei dati è decisamente anomalo e quindi è probabilisticamente incompatibile con $H_0 \rightarrow$ i dati parlano contro H_0 e quindi H_0 deve essere rifiutata

Statistica 2010/2011

49

Approccio del p -value /3

- Quanto deve essere piccolo il p -value per rifiutare H_0 in favore di H_1 ? 1 su 10, 1 su 100, 1 su 1000 ...?
- Il livello di soglia è la **probabilità di errore di I tipo**: ad es. se si adotta il criterio di rifiutare H_0 quando il p -value è inferiore a 0.05 allora si ha una probabilità 0.05 di rifiutare H_0 quando è vera (ricorda: il p -value è calcolato usando la distribuzione della statistica test sotto H_0) $\rightarrow$ test al livello $\alpha = 0.05$
- In generale, per effettuare un test di livello α arbitrario si procede nel seguente modo:
 - Calcolare il valore della statistica test usando i dati a disposizione
 - Calcolare il p -value
 - Prendere la decisione statistica confrontando il p -value con la soglia α :
 - $p\text{-value} \geq \alpha \rightarrow H_0$ non è rifiutata
 - $p\text{-value} < \alpha \rightarrow H_0$ è rifiutata

Statistica 2010/2011

50

Approccio del p -value - Esempio

- Nell'esempio relativo al peso dei cereali nelle scatole, il valore campionario della statistica test Z_0 è 1.5 per cui

Ipotesi	p -value	Decisione se $\alpha = 0.15$	Decisione se $\alpha = 0.05$	Decisione se $\alpha = 0.01$
$H_0: \mu \leq 368$ vs $H_1: \mu > 368$	0.067	Rif	Non Rif	Non Rif
$H_0: \mu \geq 368$ vs $H_1: \mu < 368$	0.933	Non Rif	Non Rif	Non Rif
$H_0: \mu = 368$ vs $H_1: \mu \neq 368$	0.134	Rif	Non Rif	Non Rif

Statistica 2010/2011

51

Approccio del p -value – Vantaggi /1

- Un test di ipotesi di livello α può essere effettuato con due approcci alternativi ma equivalenti (= stessa decisione):
 - Approccio del valore critico
 - Approccio del p -value
- In pratica l'approccio del p -value è quello più utilizzato perché il risultato non è solo una decisione rifiuto/non rifiuto, ma anche una **misura su scala continua** (il p -value appunto) **dell'evidenza empirica contro H_0** che fa capire quanto la decisione presa sia supportata dai dati
 - Ad es. in un test si ottiene un p -value di 0.0003, mentre in un altro test si ottiene un p -value di 0.047; adottando un livello $\alpha=0.05$ in entrambi i casi H_0 viene rifiutata; tuttavia, nel primo caso si tratta di un rifiuto con forte supporto empirico, mentre nel secondo caso si rifiuta per poco. In altri termini, nel primo caso si rifiuta per qualunque livello α ragionevole (cioè tra 0.10 e 0.01), mentre nel secondo caso se si fosse adottato $\alpha=0.01$ invece di $\alpha=0.05$ l'ipotesi nulla H_0 non sarebbe stata rifiutata, quindi il risultato dipende fortemente dal livello α scelto

Statistica 2010/2011

52

Approccio del p -value – Vantaggi /2

- Inoltre, il p -value consente di separare la fase (**oggettiva**) di sintesi dell'evidenza empirica contro H_0 dalla fase (**soggettiva**) di scelta del livello di significatività α
 - Dato il modello e le ipotesi da verificare il calcolo del p -value è oggettivo $\rightarrow$ tutti gli analisti giungono allo stesso valore
 - La scelta del livello di significatività α è invece soggettiva perché dipende dalle utilità o costi associati agli errori e dai livelli di rischio che si è disposti ad accettare $\rightarrow$ due analisti possono specificare livelli diversi di α
- Con l'approccio del p -value ogni analista che legge il risultato di una ricerca può fare il test applicando il livello α desiderato

Ogni teoria scientifica "valida" è una proibizione: essa preclude l'accadimento di certe cose. Quante più cose preclude, tanto migliore essa risulta.

Una teoria che non può essere confutata da alcun evento concepibile, non è scientifica. L'inconfutabilità di una teoria non è (come spesso si crede) un pregio, bensì un difetto.

Ogni controllo genuino di una teoria è un tentativo di falsificarla, o di confutarla. La controllabilità coincide con la falsificabilità; vi sono tuttavia dei gradi di controllabilità: alcune teorie sono controllabili, o esposte alla confutazione, più di altre; esse, per così dire, corrono rischi maggiori.

K.R. Popper

[Congetture e confutazioni, Il Mulino, Bologna, 1972, pp. 66-67]

Test e intervalli di confidenza /1

- Ogni test a due code ha un intervallo di confidenza (IC) equivalente
- Pertanto se si è già calcolato un IC per un parametro θ si può usare quell'intervallo per fare un test a due code su θ senza dover ripetere i calcoli
- Ad esempio, se $X \sim N(\mu, \sigma^2)$, con σ^2 nota, per sottoporre a test $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$ al livello di significatività α si può definire la regione di rifiuto nel seguente modo: si rifiuta H_0 quando il valore ipotizzato μ_0 non è incluso nell'IC per μ al livello di confidenza $1-\alpha$

$$(x_1, x_2, \dots, x_n) \in S_R \Leftrightarrow \mu_0 \notin \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

Esempio delle scatole di cereali: il peso medio campionario è 372.5 $\rightarrow$ l'IC per μ al 95% è [366.62, 378.38] $\rightarrow$ il valore 368 è incluso in tale IC e quindi il test $H_0: \mu = 368$ vs $H_1: \mu \neq 368$ al 5% porta a non rifiutare H_0

Test e intervalli di confidenza /2

- Un test eseguito tramite un IC al livello di confidenza $1-\alpha$ ha un livello di significatività α (es. IC al 95% $\rightarrow$ test al 5%, IC al 99% $\rightarrow$ test all'1%)
- Infatti la regione di rifiuto è

$$(x_1, x_2, \dots, x_n) \in S_R \Leftrightarrow \mu_0 \notin \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

e se H_0 è vera, cioè μ_0 è la vera media, per costruzione tale IC include μ_0 con probabilità $1-\alpha$, ovvero non include μ_0 con probabilità α

$$P \left(\mu_0 \notin \left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \middle| H_0 \right) = \alpha$$

Rifiutare H_0 quando è vera (errore di I tipo)

Nonostante questa equivalenza, in un'ottica decisionale è preferibile sottoporre a verifica le ipotesi usando la metodologia dei test di ipotesi, in quanto la teoria dei test fornisce gli strumenti per valutare e limitare entrambi i tipi di errore (I e II tipo)

Il test T per la media /1

- Sia X un carattere continuo con distribuzione Normale, $X \sim N(\mu, \sigma^2)$, in cui anche la varianza è ignota, e supponiamo di disporre di un campione casuale di ampiezza n
- Siamo interessati a sottoporre a verifica empirica un'ipotesi riguardante la media della popolazione μ , in una delle forme
 - $H_0: \mu \leq \mu_0$ vs $H_1: \mu > \mu_0$
 - $H_0: \mu \geq \mu_0$ vs $H_1: \mu < \mu_0$
 - $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$
- Come nel caso di σ nota a priori, anche in questo contesto la migliore statistica test è la media campionaria; tuttavia, proprio perché σ non è nota la versione standardizzata Z_0 è inutilizzabile
- Sostituendo σ con la deviazione std campionaria S si ottiene una statistica test T_0 con struttura analoga a Z_0 ma che sotto H_0 ha distribuzione t di Student con $n-1$ gdl

$$T_0 = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} \stackrel{H_0}{\sim} t_{n-1}$$

57

Il test T per la media /2

- Ai fini del test per la media, passare dal caso di σ nota al caso di σ ignota comporta una sola modifica: la distribuzione della statistica test sotto H_0 non è più Normale standard ma è t_{n-1}
 - con l'approccio del valore critico significa usare la tavola della t invece della tavola della Normale standard per determinare il valore critico
 - invece l'approccio del p -value non è fattibile con la tavola della t , perché tale tavola non consente di determinare la probabilità a sinistra o a destra di un punto arbitrario; una soluzione approssimata è quella di usare la tavola della Normale standard (l'approssimazione è tanto migliore quanto più grande è n ; altrimenti se si dispone di un computer si può fare il calcolo esatto con la t)

Statistica 2010/2011

58

Il test T per la media - Esempio

- Per controllare il processo produttivo di una falegnameria vengono esaminate 10 tavole, il cui spessore medio è di 6.05 mm con deviazione standard campionaria di 0.08 mm.
- Si assume che lo spessore abbia distribuzione Normale e che le 10 tavole siano un campione casuale
- Supponiamo che lo spessore nominale delle tavole sia di 6.00 mm, per cui il responsabile del controllo di qualità vuole sottoporre a verifica $H_0: \mu = 6$ vs $H_1: \mu \neq 6$, fissando $\alpha = 0.05$
- La statistica test T_0 ha distribuzione t_9 se H_0 è vera; poiché H_1 è bidirezionale la regione di rifiuto sta sulle due code della t_9 e i valori critici sono -2.2622 e 2.2622
- Il valore campionario della statistica test è $t_0 = \frac{6.05 - 6.00}{0.08 / \sqrt{10}} = 1.9764$

Pertanto non si rifiuta H_0 : alla luce dei dati raccolti non si può affermare che il processo produttivo sia fuori controllo (nota: se si fosse usata erroneamente la tavola della Normale standard si sarebbe ottenuto il valore critico 1.96, rifiutando quindi H_0)

59

Il test Z per la proporzione /1

- Quando il carattere di interesse è *dicotomico* ($X = 1$ presenza della caratteristica in oggetto, $X = 0$ assenza della caratteristica in oggetto) la sua distribuzione è (necessariamente) di tipo Bernoulli e l'unico parametro è la *probabilità o proporzione di successo* p : $X \sim Be(p)$
- Uno stimatore corretto di p è il suo corrispondente nel campione, cioè la **proporzione campionaria**

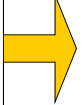
$$\hat{P} = \frac{\text{numero unità campionarie con la caratteristica di interesse}}{\text{ampiezza campionaria}} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

Statistica 2010/2011

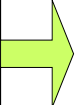
60

Il test Z per la proporzione /2

Se si dispone di un campione casuale abbastanza grande (cioè se il criterio $np(1-p) > 9$ è soddisfatto), per il TLC la distribuzione della proporzione campionaria è approssimativamente Normale


$$\frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \stackrel{\text{approx}}{\sim} N(0,1)$$

Sostituendo il vero p con quello ipotizzato p_0 si ottiene la statistica test Z_0^p che sotto H_0 ha distribuzione approssimativamente Normale standard


$$Z_0^p = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \stackrel{H_0, \text{approx}}{\sim} N(0,1)$$

Test Z per la proporzione – Esempio

- Una struttura ricettiva vuole valutare se un certo servizio aggiuntivo aumenta la soddisfazione dei clienti; a tal fine
 - $p = \{\text{proporzione di clienti soddisfatti con il servizio aggiuntivo}\}$,
 - $p_0 = 0.60 = \{\text{proporzione di clienti soddisfatti senza il servizio aggiuntivo}\}$
- Si formulano le ipotesi in modo che H_0 indichi che il servizio aggiuntivo è inefficace (non aumenta la soddisfazione):
 - $H_0: p \leq 0.60$ vs $H_1: p > 0.60$
- I dati a disposizione consistono in un campione casuale di 200 clienti che hanno ricevuto il servizio aggiuntivo, 130 dei quali si sono dichiarati soddisfatti → la proporzione campionaria è 0.65
- Poiché 0.65 è maggiore di 0.60, *nel campione osservato* il servizio aggiuntivo aumenta la proporzione di soddisfatti.
 - Problema: questa evidenza empirica è sufficiente per affermare che il servizio aggiuntivo aumenta la proporzione di soddisfatti *nella popolazione*, cioè nell'insieme di tutti i clienti?

Test Z per la proporzione – Esempio (continua)

- In altri termini: una proporzione di 0.65 su un campione di 200 unità è o non è probabilisticamente compatibile con una vera proporzione di 0.60? In caso di risposta affermativa si mantiene H_0 , altrimenti si rifiuta

$$n\hat{p}(1-\hat{p}) = 200 \times 0.65 \times 0.35 = 45.5 > 9 \quad \rightarrow \text{si può usare la statistica test } Z_0^p \text{ che sotto } H_0 \text{ ha distribuzione approssimativamente Normale std}$$

- Per il test si sceglie un livello di significatività $\alpha = 0.01$ (così si rende molto bassa, 1% appunto, la probabilità di dire che il servizio è efficace quando in realtà è inefficace)
- Il valore campionario della statistica test è

$$z_0^p = \frac{0.65 - 0.60}{\sqrt{0.60(1-0.60)/200}} = 1.44$$

Test Z per la proporzione – Esempio (continua)

- **APPROCCIO DEL VALORE CRITICO:**
 - poiché H_1 contiene il segno di $>$ la regione di rifiuto sta a destra del valore critico, che è 2.33 (è il valore che sulla Normale std lascia alla sua destra un'area pari a 0.01) → la regola decisionale è di rifiutare H_0 quando il valore campionario della statistica test è maggiore di 2.33 → poiché $1.44 < 2.33$ non si rifiuta H_0
- **APPROCCIO DEL p -value:**
 - poiché H_1 contiene il segno di $>$ la regione di rifiuto sta sulla coda destra → il p -value è la probabilità di ottenere, sulla Normale std, un valore maggiore o uguale a quello osservato: poiché l'area a destra di 1.44 è 0.0749 si ha $p\text{-value} = 0.0749$ → poiché $0.0749 > 0.01$ non si rifiuta H_0
- Conclusione: non si hanno prove sufficienti per affermare che il servizio aggiuntivo aumenta la proporzione di clienti soddisfatti

Test per la varianza (X~Normale) /1

- Si vuole sottoporre a verifica empirica un'ipotesi riguardante la varianza di una popolazione Normale, in una delle forme
 - $H_0: \sigma^2 \leq \sigma_0^2$ vs $H_1: \sigma^2 > \sigma_0^2$
 - $H_0: \sigma^2 \geq \sigma_0^2$ vs $H_1: \sigma^2 < \sigma_0^2$
 - $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$
- A tal fine si usa la varianza campionaria, che ha una distribuzione campionaria legata alla Chi-quadrato

$$X \sim N(\mu, \sigma^2) \quad X_1, \dots, X_n \text{ iid-} X$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \Rightarrow \quad V = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

per cui la statistica test è

$$V_0 = \frac{(n-1)S^2}{\sigma_0^2} \stackrel{H_0}{\sim} \chi_{n-1}^2$$

Statistica 2010/2011

65

Test per la varianza (X~Normale) /2

- Poiché S^2 è uno stimatore di σ^2
 - $H_0: \sigma^2 \leq \sigma_0^2$ vs $H_1: \sigma^2 > \sigma_0^2$ → si rifiuta sulla coda destra
 - $H_0: \sigma^2 \geq \sigma_0^2$ vs $H_1: \sigma^2 < \sigma_0^2$ → si rifiuta sulla coda sinistra
 - $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$ → si rifiuta sulle code

Ad es. se l'ipotesi alternativa è bidirezionale la regione di rifiuto per il test di livello α con un campione di dimensione n è

$$R = \{v_0 : v_0 \notin (\chi_{n-1, 1-\alpha/2}^2, \chi_{n-1, \alpha/2}^2)\}$$

$$\text{dove } v_0 = \frac{(n-1)s^2}{\sigma_0^2}$$

Attenzione: a differenza della Normale e della t di Student, il valore critico inferiore della Chi-quadrato non è il negativo del valore critico superiore

Statistica 2010/2011

66

Calcolare la potenza /1

- Ricordiamo che
 - Probabilità errore I tipo** (vero H_0 ma si decide H_1):
 $P(\text{rifiutare } H_0 \mid H_0) = \alpha$
 - Probabilità errore II tipo** (vero H_1 ma si decide H_0):
 $P(\text{accettare } H_0 \mid H_1) = \beta$
 - Potenza del test** (vero H_1 e si decide giustamente H_1):
 $P(\text{rifiutare } H_0 \mid H_1) = 1 - \beta = \pi$
- La probabilità α viene fissata dall'analista, occorre calcolare il valore β o, in modo equivalente, il valore π (si passa dall'uno all'altro calcolando il complemento a 1, qui illustriamo il calcolo di π)
- Attenzione: la potenza π è una proprietà del test (regola decisionale) → i dati campionari non servono!

Statistica 2010/2011

67

Calcolare la potenza /2

- Per calcolare $\pi = P(\text{rifiutare } H_0 \mid H_1)$ occorrono due passi:

- Esprimere l'evento "rifiutare H_0 " in termini della statistica test non standardizzata, es.

per il test sulla media μ : $\{\bar{X} < 5.83\}, \{\bar{X} < 55.2 \cup \bar{X} > 62.4\}, \dots$

per il test sulla prop. p : $\{\hat{P} > 0.6807\}, \{\hat{P} < 0.123 \cup \hat{P} > 0.152\}, \dots$

- Calcolare la probabilità dell'evento "rifiutare H_0 " ipotizzando che sia vero uno specifico valore in H_1
 - H_1 semplice → un solo valore del parametro → un solo valore di π
 - H_1 composta → più valori del parametro → più valori di π (in questo caso bisogna scegliere uno o più valori del parametro da usare per calcolare la potenza)

Statistica 2010/2011

68

Calcolare la potenza – Esempio

- Riprendiamo l'esempio del test sulla proporzione di clienti soddisfatti a seguito del servizio aggiuntivo:

1. Esprimere l'evento "rifiutare H_0 " in termini della statistica test non standardizzata:

$$\text{Rifiutare } H_0 = \{Z_0^p > 2.33\} = \left\{ \frac{\hat{P} - 0.60}{\sqrt{0.60(1-0.60)/200}} > 2.33 \right\} = \{\hat{P} > 0.6807\}$$

2. Calcolare la probabilità di "rifiutare H_0 " ipotizzando che sia vero uno specifico valore in H_1 , ad es. 0.65

$$P(\text{Rifiutare } H_0 \mid p = 0.65) = P(\hat{P} > 0.6807 \mid p = 0.65)$$

$$\approx P\left(Z > \frac{0.6807 - 0.65}{\sqrt{0.65(1-0.65)/200}}\right) = P(Z > 0.9103) = 0.1813$$

Due nozioni di significatività

- Per la corretta interpretazione del risultato di un test di ipotesi è cruciale comprendere la distinzione tra

- **significatività sostanziale**

- **significatività statistica**

*Come il fuoco,
il test di ipotesi è
un buon servitore
e un cattivo padrone*

Significatività sostanziale

- Un effetto si dice **sostanzialmente significativo** quando è abbastanza grande da essere rilevante, degno di interesse nel contesto applicativo in esame
 - L'effetto di un farmaco per la riduzione della pressione sanguigna può essere misurato dalla riduzione del valore medio della pressione diastolica (in mm di mercurio); da un punto di vista medico, una riduzione di 2 non è sostanzialmente significativa (per cui il farmaco viene giudicato inutile), mentre una riduzione di 20 sì (per cui il farmaco viene preso in considerazione – naturalmente questo beneficio va poi valutato assieme agli effetti collaterali)
 - L'effetto di un servizio aggiuntivo sulla soddisfazione dei clienti può essere misurato dall'aumento della proporzione di clienti soddisfatti; per il management della struttura ricettiva, un aumento di 3% non è sostanzialmente significativo, mentre un aumento di 10% sì (per cui il servizio aggiuntivo viene preso in considerazione – naturalmente la decisione sull'effettiva implementazione dipende anche dal costo)

Significatività statistica

- Un effetto si dice **statisticamente significativo** quando un test di ipotesi porta ad affermare che esiste, cioè non è nullo: posto H_0 : "effetto nullo" vs H_1 : "effetto non nullo" si rifiuta H_0 :
 - L'effetto del farmaco per la riduzione della pressione sanguigna è statisticamente significativo se il test con i dati della sperimentazione porta a rifiutare l'ipotesi che l'effetto sia inesistente
 - L'effetto del servizio aggiuntivo sulla soddisfazione dei clienti è statisticamente significativo se il test con i dati raccolti presso i clienti porta a rifiutare l'ipotesi che l'effetto sia inesistente

Significatività sostanziale vs statistica

- La **significatività sostanziale** è un concetto **soggettivo**: due esperti del settore possono avere opinioni diverse sulla significatività sostanziale di un certo effetto (anche se in molti campi scientifici vi sono delle linee guida condivise)
- Al contrario, dato il livello di significatività α , la **significatività statistica** di un effetto è un fatto **oggettivo**: due analisti (che usano lo stesso modello) giungono allo stessa decisione statistica e quindi necessariamente concordano sulla significatività statistica
- Tuttavia la significatività statistica dipende in modo cruciale dall'ampiezza campionaria n :
 - se n è piccolo anche un forte effetto (sostanzialmente significativo) può risultare non statisticamente significativo
 - viceversa se n è grande possono risultare statisticamente significativi anche effetti trascurabili (sostanzialmente non significativi)

Statistica 2010/2011

73

Significatività statistica e sostanziale?

Il ruolo dell'ampiezza campionaria n si può comprendere guardando l'espressione della statistica test Z per la media

$$Z_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$$

- Se n è piccolo è possibile che una notevole differenza tra valore campionario e valore ipotizzato (sintomo di un forte effetto) porti a mantenere l'ipotesi di effetto nullo $\rightarrow$ è possibile che un effetto sostanzialmente significativo sia statisticamente non significativo
- Se n è grande, al contrario, è possibile che una piccola differenza tra valore campionario e valore ipotizzato (sintomo di un effetto trascurabile) porti a rifiutare l'ipotesi di effetto nullo $\rightarrow$ è possibile che un effetto sostanzialmente non significativo sia statisticamente significativo
- In fase di predisposizione dell'indagine o sperimentazione si dovrebbe scegliere un'ampiezza campionaria n per la quale la significatività statistica sia il più possibile simile alla significatività sostanziale

Statistica 2010/2011

74

Test multipli

- Anche se il livello α è basso, effettuando molti test è probabile incorrere in un errore di I tipo
- Esempio:
 - H_0 : farmaco inefficace vs H_1 : farmaco efficace
 - Livello di significatività: $\alpha = 0.05$
 - Sperimentazione in $k=20$ ospedali, ognuno con il suo campione di pazienti (campioni indipendenti)
- Supponiamo che il farmaco sia inefficace: qual è la probabilità che almeno un ospedale giunga alla conclusione che invece è efficace (errore I tipo)?
 - $P(\text{almeno uno}) = 1 - P(\text{nessuno}) = 1 - (1 - \alpha)^k = 1 - (1 - 0.05)^{20} = 0.642$
- D'altra parte $\alpha = 0.05 = 1/20$, cioè *in media* si compie un errore di I tipo ogni 20 test
- Ripetendo il test più volte, prima o poi qualunque ipotesi vera viene rifiutata!

Statistica 2010/2011

75

Data snooping

- Le ipotesi vanno formulate PRIMA di osservare i dati
- Formulare le ipotesi sulla base dei dati contraddice l'impianto logico del test: è un imbroglio!
- Esempio: X = "km percorsi da un'automobile con un litro di carburante"; si ottiene una media campionaria di 21 km; tutte le seguenti ipotesi nulle vengono accettate
 - $H_0: \mu = 21$
 - $H_0: \mu \geq 21$
 - $H_0: \mu \leq 21$
 - In generale si accettano tutte le ipotesi che distano dalla media campionaria meno di

$$t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}$$

Statistica 2010/2011

76