

Il valore p (in inglese p value) di un test calcolato per una determinata *osservazione* è la probabilità che, *nell'ipotesi* H_0 , si abbia un'osservazione (risultato del test per quel campione) che contraddica l'ipotesi H_0 almeno tanto quanto l'osservazione che abbiamo fatto.

Matematicamente: $p := \text{prob}(T \geq t_{\text{oss}} | H_0)$

Esempio 1. Studio di un nuovo antidepressivo.

H_0 = non ha effetto.

T = quanto fa ridere.

$p = 0.7$... Il 70% degli antidepressivi *inefficaci* fa ridere (almeno) tanto quanto il farmaco esaminato; quindi accettiamo H_0 (cioè l'ipotesi che non abbia effetto).

$p = 0.001$... Solo *uno su mille* degli antidepressivi inefficaci fa ridere tanto quanto il nuovo farmaco. Cominciamo a credere che abbia effetto.

Esempio 2. Interrogazione di un imputato.

H_0 = è innocente.

T = quanto trema.

$p = 0.7$... Il 70% *degli innocenti* trema quanto lui; quindi accettiamo H_0 (non ce la sentiamo di condannarlo).

$p = 0.001$... Solo *uno su mille degli innocenti* trema quanto trema lui. Non accettiamo facilmente H_0 e lo condanniamo.

Attenzione: Quando p è molto piccolo, si dice spesso che il test è molto *significativo*. Ma significativa è solo la probabilità che rileviamo, il valore p non dice nulla sull'*entità* della differenza che (quando il valore p è molto basso) il test "dimostra" che esiste. Quindi nel primo esempio può essere che il nuovo farmaco (magari molto più costoso) faccia ridere di più, ma in modo praticamente impercettibile.

Nel secondo esempio nel caso $p = 0.001$ è vero che ciò è un forte indizio per la colpa dell'imputato, ma può essere che abbia solo rubato una mela.

Quindi *significativo non si riferisce mai all'effetto*.

S. Goodman: A dirty dozen - twelve p-value misconceptions. Sem. Hematology ... (2008), 135-140.

R. Nuzzo: Statistical errors. P values, the gold standard of statistical validity, are not as reliable as many scientists assume. Nature 13 February 2014, 150-152.