Consider a coin being tossed 20 times.

There are $2^{20} = 1{,}048{,}576$ possible outcomes ("samples").

Suppose we wish to test whether the coin is fair or not.

We can "rank" all these outcomes ("samples") from ${}_{20}C_{10} = 184{,}756$ "best" outcomes ("samples") with 10 Hs and 10 Ts down to the "worst" outcomes ("samples"): ${}_{20}C_1 = 20$ outcomes with 1 H and 19 Ts and ${}_{20}C_{19} = 20$ outcomes with 19 Hs and 1 T, followed by ${}_{20}C_0 = 1$ outcome with no Hs and 20 Ts and ${}_{20}C_{20} = 1$ outcome with 20 Hs and no Ts:

"best"	10 H & 10 T			${}_{20}C_{10}$	184,756	184,756
	9 H & 11 T	or	11 H & 9 T	${}_{20}C_9 + {}_{20}C_{11}$	167,960 + 167,960	335,920
	8 H & 12 T	or	12 H & 8 T	${}_{20}C_8 + {}_{20}C_{12}$	125,970 + 125,970	251,940
	7 H & 13 T	or	13 H & 7 T	${}_{20}C_7 + {}_{20}C_{13}$	77,520 + 77,520	155,040
	6 H & 14 T	or	14 H & 6 T	${}_{20}C_6 + {}_{20}C_{14}$	38,760 + 38,760	77,520
	5 H & 15 T	or	15 H & 5 T	${}_{20}C_5 + {}_{20}C_{15}$	15,504 + 15,504	31,008
	4 H & 16 T	or	16 H & 4 T	${}_{20}C_4 + {}_{20}C_{16}$	4,845 + 4,845	9,690
	3 H & 17 T	or	17 H & 3 T	${}_{20}C_3 + {}_{20}C_{17}$	1,140 + 1,140	2,280
	2 H & 18 T	or	18 H & 2 T	${}_{20}C_2 + {}_{20}C_{18}$	190 + 190	380
	1 H & 19 T	or	19 H & 1 T	${}_{20}C_1 + {}_{20}C_{19}$	20 + 20	40
"worst"	0 H & 20 T	or	20 H & 0 T	${}_{20}C_0 + {}_{20}C_{20}$	1 + 1	2

Note that even if the coin is fair, it is possible to end up with 20 Hs and no Ts. Also, even if the coin is not fair, even if P(H) = 0.99 and P(T) = 0.01, it is possible to end up with 10 Hs and 10 Ts. Therefore, we will not be able to determine "for sure" whether the coin is fair or not based on 20 coin-tosses.

Since we do not know if the coin is fair or not, we will give the coin "the benefit of the doubt" and assume (initially) that it is fair. Then we will perform the 20 coin-tosses.

Suppose we encountered 10 Hs and 10 Ts. Even though we observed one of the best possible outcomes ("samples"), it does not prove ("for sure") the coin to be fair. However, we have absolutely no reason to doubt our "benefit of the doubt" assumption that the coin is fair.

Suppose we encountered 13 Hs and 7 Ts. Even though it is not one of the “best” possible outcomes (“samples”), 275,960 possible outcomes (26.32%) are as “bad” or “worse” (as extreme or more extreme) than the one we observed. If the coin is fair, if $P(H) = P(T) = 0.50$, we could observe, just by chance, an outcome as extreme or more extreme than the one we just observed with probability 0.2632. Thus, it is not unusual to observe 13 Hs and 7 Ts in 20 tosses of a fair coin, and it should not raise our suspicions that the coin may not be fair. It does not prove the coin to be fair, but it also does not give us a reason to doubt it.

Suppose we encountered 17 Hs and 3 Ts. There are only 2,702 possible outcomes (0.26%) that are as “bad” or “worse” (as extreme or more extreme) than the one we just observed. If the coin is really fair, if $P(H) = P(T) = 0.50$, then the probability of seeing an outcome as “surprising” as this one, just by chance, is only 0.0026. Thus, it is very unusual to observe 17 Hs and 3 Ts in 20 tosses of a fair coin. Even though this would not prove that the coin is not fair, we now have a very good reason to believe that it is not.

		number of outcomes	as extreme or more extreme	
“best”	10 H & 10 T	184,756	1048576	1.000000
	9 H & 11 T or 11 H & 9 T	335,920	863820	0.823803
	8 H & 12 T or 12 H & 8 T	251,940	527900	0.503445
	7 H & 13 T or 13 H & 7 T	155,040	275960	0.263176
	6 H & 14 T or 14 H & 6 T	77,520	120920	0.115318
	5 H & 15 T or 15 H & 5 T	31,008	43400	0.041389
	4 H & 16 T or 16 H & 4 T	9,690	12392	0.011818
	3 H & 17 T or 17 H & 3 T	2,280	2702	0.002577
	2 H & 18 T or 18 H & 2 T	380	422	0.000402
	1 H & 19 T or 19 H & 1 T	40	42	0.000040
“worst”	0 H & 20 T or 20 H & 0 T	2	2	0.000002

In this scenario, we have two competing assumptions (hypotheses) – the coin is fair ($p = P(H) = 0.50$) and the coin is not fair ($p \neq 0.50$). Under the “benefit of the doubt” assumption that the coin is fair, we examine the outcome (“sample”) of 20 coin-tosses. If the outcome is “typical” and “nothing out of the ordinary” (the probability of obtaining an outcome as extreme or more extreme is not small), then we do not reject the assumption (hypothesis) that the coin is fair. If the outcome is “unusual” and “surprising” (the probability of obtaining an outcome as extreme or more extreme is small), then we reject the assumption (hypothesis) that the coin is fair.