Power of test and $\boldsymbol{\beta}$:

The probability that the null hypothesis (H_0) will be rejected when, in fact, the alternative hypothesis (H_1) is true is called the power of the test.

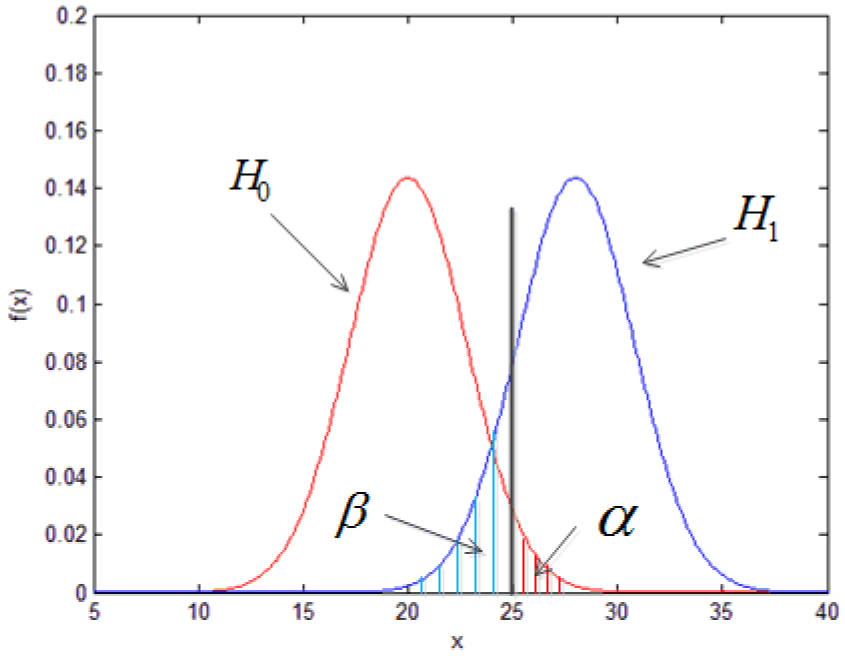
$$power = 1 - \beta$$

An example to explain α *and* β

Example: A sample of size 9 from a normal distribution with $\sigma^2 = 25$ is used to test

$$H_0\colon \mu = 20, \qquad H_1\colon \mu = 28 \ (\mu > 20)$$

The test statistic is the sample mean $\bar{X}$. Let us agree to reject H_0 in favor of H_1 if the observed value of $\bar{X}$ is greater than 25.



$$\alpha = P[H_0 \ is \ rejected | H_0 \ is \ true], \beta = P[H_0 \ is \ not \ rejected | H_1 \ is \ true]$$

(a) If H_0 is true, what is the distribution of $\bar{X}$? $\bar{X} \sim N(20, \frac{25}{9})$

(b) If H_1 is true, what is the distribution of $\bar{X}$? $\bar{X} \sim N(28, \frac{25}{9})$

(c) Create the graphs of both distributions and visualize α *and* β

(d) How to reduce α *and* β at the same time?

Problem with the critical region method:

This method create a rigid critical region with preset α value.

Example: Suppose we want to test $H_0: p \leq 0.1, H_1: p > 0.1$ base on a sample of size 20. The test statistic is X, the number of "success" observed in the 20 trials. When $p = 0.1$, X follows a binomial distribution with $E[X] = np = 2$.

Suppose that we want α to be very small. So we define the critical region to be $c = \{9,10, \dots, 19,20\}$.

Then, the value of α can be found as:

$$\alpha = P[Type\ I\ error] = P[reject\ H_0|H_0\ is\ true]$$
$$= P[X \geq 9|p = 0.1] = 1 - P[X < 9|p = 0.1] = 1 - P[X \leq 8|p = 0.1]$$
$$= 1 - 0.9999 = 0.0001$$

If we change the critical region to $c = \{8,9,10, \dots, 19,20\}$. Then the value of α can be found as:

$$\alpha = P[X \geq 8|p = 0.1] = 1 - P[X < 8|p = 0.1] = 1 - P[X \leq 7|p = 0.1]$$
$$= 1 - 0.9996 = 0.0004$$

The change of the critical region from $c = \{9,10, \dots, 19,20\}$ to $c = \{8,9,10, \dots, 19,20\}$ cause very small change of α

Example: For a particular model of car, the number of miles per gallon has a mean of 26 mpg with a standard deviation of 5 mpg. It is hoped that a new design will increase the mean mileage rating. Assuming σ is not affected by the design. Suppose we obtain a sample of 36 and calculated the sample mean $\bar{x} = 28.04\ mpg$. Carry out the test at significance level of $\alpha = 0.05$

We test:

$$H_0\colon\ \mu \le 26; \quad H_1\colon\ \mu > 26$$

The rejection region is $[x_0, \infty)$. Let's find the value of x_0

Assuming $H_0\ is\ true$, the test statistic $\bar{X} \sim N(26, \frac{25}{36})$. We find the value of x_0 such that

$$P[\bar{X} \ge x_0] = 0.05$$

$$\Rightarrow P\left[\frac{\bar{X} - 26}{5/6} \ge \frac{x_0 - 26}{5/6}\right] = 0.05 \Rightarrow P\left[Z \ge \frac{x_0 - 26}{5/6}\right] = 0.05 \Rightarrow P\left[Z \le \frac{x_0 - 26}{5/6}\right] = 0.95$$

From the Z probability table, we find out that

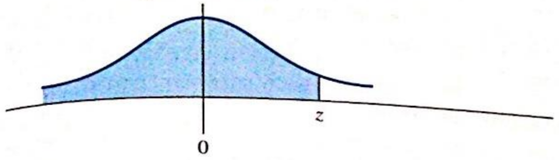
$$\frac{x_0 - 26}{5/6} = 1.645$$

Therefore, we have, $x_0 = 27.37$. i.e., the rejection region is $[27.37, \infty)$.

Since the sample mean is $\bar{x} = 28.04$ that is inside the rejection region, H_0 should be rejected.

TABLE V
Cumulative distribution: Standard normal

$F_Z(z) = P[Z \le z]$



	$F_Z(z) = P[Z \le z]$									
z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1921	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

The p-value method (significance test method):

Example: For a particular model of car, the number of miles per gallon has a mean of 26mpg with a standard deviation of 5mpg. It is hoped that a new design will increase the mean mileage rating. Assuming σ is not affected by the design. We test:

$$H_0: \mu \leq 26; \quad H_1: \mu > 26$$

In p-value method , to see if there is enough difference to cause us to reject H_0, we find the p-value of the test, i.e., *we compute the probability of observing a sample mean of 28.04 or larger if H_0 is true.*

We know that $\bar{X}$ is approximately normally distributed with mean $\mu_{\bar{X}} = 26 \; and \; \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = 5/6$. Then,

$$P[\bar{X} \geq 28.04] = P\left[\frac{\bar{X} - 26}{(5/6)} \geq \frac{28.04 - 26}{(5/6)}\right]$$
$$\approx P[Z \geq 2.45] = 1 - P[Z < 2.45] = 1 - 0.9929 = 0.0071 < \alpha = 0.05$$

This p-value is very small and it can be caused by: (1) the null hypothesis is true, we observed a very rare sample; or (2) the null hypothesis is not true, the sample is from a distribution with higher mean mileage.

We prefer the second explanation! i.e., if the p-value is smaller than the significance level, H_0 will be rejected.

P-value:

The probability of obtaining a result equal to or "more extreme" than what was actually observed, assuming $\boldsymbol{H_0}$ is true.

Three forms of Hypothesis tests:

I. H_0: $p \leq p_0$, H_1: $p > p_0$ (***right-tailed test***) (the rejection region is a right-tailed region of the distribution)
II. H_0: $p \geq p_0$, H_1: $p < p_0$ (***left-tailed test***) (the rejection region is a left-tailed region of the distribution)
III. H_0: $p = p_0$, H_1: $p \neq p_0$ (***two-tailed test***) (the rejection region is a two-tailed region of the distribution)

Let X be the test statistic we choose. Let x_{ob} be the observed value of X, calculated from the sample.

Here is how p-value is calculated: (the meaning of "***more extreme***": to the direction of the alternative hypothesis)

If we have a right-tailed test, p-value $= P[X \geq x_{ob} | H_0 \textit{ is true}]$

If we have a left-tailed test, p-value $= P[X \leq x_{ob} | H_0 \textit{ is true}]$

If we have a two-tailed test, p-value $= 2\min\{P[X \leq x_{ob} | H_0 \textit{ is true}],\ P[X \geq x_{ob} | H_0 \textit{ is true}]\}$

Example (left-tailed test): Highway engineers have found that the proper alignment of the vehicle's headlights is a major factor affecting the performance of reflective highway signs. It is thought that less than 50% of the vehicle on the road have misaimed headlights. 20 vehicles were randomly selected from the roads and headlights inspected and found 8 with misaimed headlights. Design a hypothesis test with the significance level $\alpha = 0.05$.

Let p denote the proportion of vehicle in operation that have misaimed headlights. Since we wish to support the statement that $p < 0.5$, we have the two hypothesis:

$$H_0 : p \geq 0.5; \quad H_1 : p < 0.5$$

The critical region method: Let X be the number of cars in the sample with misaimed headlights. X is the test statistic.

If H_0 is true ($p \geq 0.5$), then X is binomial with $n = 20, p = 0.5. (we\ choose\ p = 0.5)$.

If the observed value of X is less than a critical value x_0, we tend to reject H_0.

According to the definition of α,

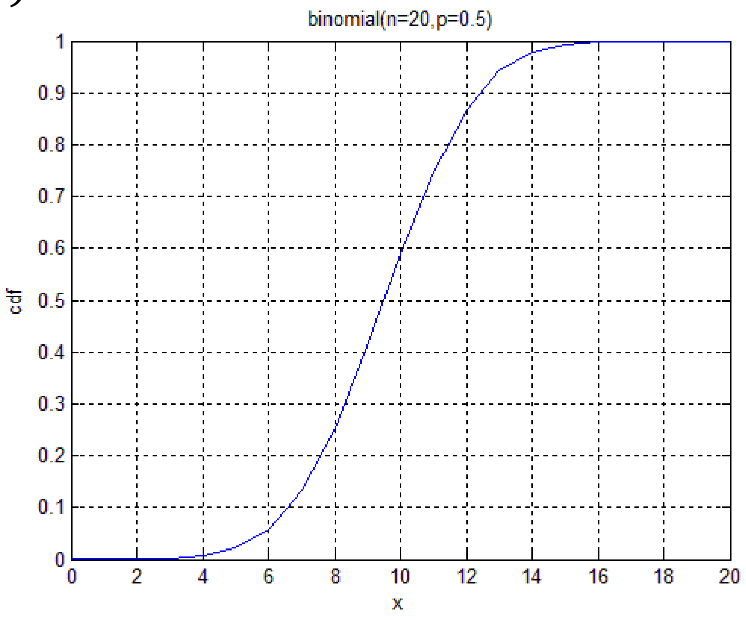
$$P[type\ I\ error] = P[X \leq x_0 | p = 0.5] \approx 0.05$$

From the binomial probability table, we can read out:

$$P[X \leq 6 | p = 0.5] = 0.0577$$

Hence, the rejection region is $[0,1,2,3,4,5,6]$

Since $X = 8$ is not inside the rejection region, H_0 can not be rejected.



CUMULATIVE BINOMIAL PROBABILITIES

Tabulated values are P(<= k)
(Computations are rounded at the third decimal place.)

N = 20

k \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.122	0.012	0.001	0.000	0.000	0.000	0.000	0.000	0.000
1	0.392	0.069	0.008	0.001	0.000	0.000	0.000	0.000	0.000
2	0.677	0.206	0.035	0.004	0.000	0.000	0.000	0.000	0.000
3	0.867	0.411	0.107	0.016	0.001	0.000	0.000	0.000	0.000
4	0.957	0.630	0.238	0.051	0.006	0.000	0.000	0.000	0.000
5	0.989	0.804	0.416	0.126	0.021	0.002	0.000	0.000	0.000
6	0.998	0.913	0.608	0.250	0.058	0.006	0.000	0.000	0.000
7	1.000	0.968	0.772	0.416	0.132	0.021	0.001	0.000	0.000
8	1.000	0.990	0.887	0.596	0.252	0.057	0.005	0.000	0.000
9	1.000	0.997	0.952	0.755	0.412	0.128	0.017	0.001	0.000
10	1.000	0.999	0.983	0.872	0.588	0.245	0.048	0.003	0.000
11	1.000	1.000	0.995	0.943	0.748	0.404	0.113	0.010	0.000
12	1.000	1.000	0.999	0.979	0.868	0.584	0.228	0.032	0.000
13	1.000	1.000	1.000	0.994	0.942	0.750	0.392	0.087	0.002
14	1.000	1.000	1.000	0.998	0.979	0.874	0.584	0.196	0.011
15	1.000	1.000	1.000	1.000	0.994	0.949	0.762	0.370	0.043
16	1.000	1.000	1.000	1.000	0.999	0.984	0.893	0.589	0.133
17	1.000	1.000	1.000	1.000	1.000	0.996	0.965	0.794	0.323
18	1.000	1.000	1.000	1.000	1.000	0.999	0.992	0.931	0.608
19	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.988	0.878
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

k: Number of success
p: Probability of success of one trial

Example (left-tailed test): Highway engineers have found that the proper alignment of the vehicle's headlights is a major factor affecting the performance of reflective highway signs. It is thought that less than 50% of the vehicle on the road have misaimed headlights. 20 vehicles were randomly selected from the roads and headlights inspected and found 8 with misaimed headlights. Design a hypothesis test with the significance level $\alpha = 0.05$.

Let p denote the proportion of vehicle in operation that have misaimed headlights. Since we wish to support the statement that $p < 0.5$, we have the two hypothesis:

$$H_0: p \geq 0.5; \quad H_1: p < 0.5$$

The p-value method: Let X be the number of cars in the sample with misaimed headlights. X is the test statistic.
If H_0 is true ($p \geq 0.5$), then X is binomial with $n = 20, p = 0.5$.

Since this is a left-tailed test, the p-value can be calculated as:

$$p\_value = P[X \leq 8 | p = 0.5]$$

This probability can be read out from the binomial probability table:

$$p\_value = P[X \leq 8 | p = 0.5] = 0.2517$$

Since the p-value (0.2517) is greater than the significance level (0.05), H_0 will not be rejected.