

Hypothesis Testing 2

Day 12 (2/13/20)

4.6 Two-sided tests

Let X_1, X_2, \dots, X_n be a random sample from a normal distribution with mean μ and standard deviation $\sigma = 5000$. Suppose that we want to test

$$H_0 : \mu = 30,000 \text{ versus } H_1 : \mu \neq 30,000$$

It makes sense that we should reject H_0 if the sample mean is too small or too large, i.e.

$$\text{Reject } H_0 \text{ in favor of } H_1 \text{ if } \bar{X} \leq h \text{ or } \bar{X} \geq k$$

The benchmarks h and k are chosen so that the test achieves the desired size

$$\alpha = P_{H_0}(\bar{X} \leq h \text{ or } \bar{X} \geq k) = P_{H_0}(\bar{X} \leq h) + P_{H_0}(\bar{X} \geq k)$$

Suppose we choose to allocate α equally, then

$$P_{H_0}(\bar{X} \leq h) = \alpha/2 \text{ and } P_{H_0}(\bar{X} \geq k) = \alpha/2$$

Under H_0 , the statistic $(\bar{X} - 30000)/(5000/\sqrt{n})$ is $N(0, 1)$. Since

$$\alpha/2 = P\left(\frac{\bar{X} - 30000}{5000/\sqrt{n}} \leq -z_{\alpha/2}\right) = P(\bar{X} \leq 30000 - z_{\alpha/2}5000/\sqrt{n}) \quad (1)$$

then $h = 30000 - z_{\alpha/2}5000/\sqrt{n}$. Similarly,

$$\alpha/2 = P\left(\frac{\bar{X} - 30000}{5000/\sqrt{n}} \geq z_{\alpha/2}\right) = P(\bar{X} \geq 30000 + z_{\alpha/2}5000/\sqrt{n}) \quad (2)$$

so that $k = 30000 + z_{\alpha/2}5000/\sqrt{n}$. For illustration, suppose $\alpha = .05$ and $n = 25$, then $h = 30000 - 1.96(5000)/\sqrt{25} = 28040$ and $k = 30000 + 1.96(5000)/\sqrt{25} = 31960$. We can write equations (1) and (2) together as

$$\alpha = P\left(\left|\frac{\bar{X} - 30000}{5000/\sqrt{n}}\right| \geq z_{\alpha/2}\right)$$

which gives a more compact expression for the decision rule:

$$\text{Reject } H_0 \text{ in favor of } H_1 \text{ if } |\bar{X} - 30000| \geq z_{\alpha/2}5000/\sqrt{n}$$

More generally, let X_1, X_2, \dots, X_n be a random sample from a distribution f with mean μ and standard deviation σ . A large-sample size- α test for

$$H_0 : \mu = \mu_0 \text{ versus } H_1 : \mu \neq \mu_0$$

is given by the decision rule

$$\text{Reject } H_0 \text{ in favor of } H_1 \text{ if } \left|\frac{\bar{X} - \mu_0}{S/\sqrt{n}}\right| \geq z_{\alpha/2} \quad (4.6.4)$$

Power function of the two-sided test

Now, suppose that the true value of μ is in H_1 . Recall that the power of the test is the probability that the test rejects at that value of μ . From the critical regions in equations (1) and (2), the power function is

$$\begin{aligned}\gamma(\mu) &= P_\mu(\bar{X} \leq \mu_0 - z_{\alpha/2}\sigma/\sqrt{n}) + P_\mu(\bar{X} \geq \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}) \\ &= P_\mu\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu - z_{\alpha/2}\sigma/\sqrt{n}}{\sigma/\sqrt{n}}\right) + P_\mu\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{\mu_0 - \mu + z_{\alpha/2}\sigma/\sqrt{n}}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu)}{\sigma} - z_{\alpha/2}\right) + 1 - \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu)}{\sigma} + z_{\alpha/2}\right)\end{aligned}$$

where $\Phi(\cdot)$ is the standard normal cdf.

Relationship between two-sided test and confidence interval

Example: The rejection region $\left|\frac{\bar{X}-30000}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}$ is the same as $|\bar{X} - 30000| > z_{\alpha/2}\sigma/\sqrt{n}$ so the test rejects if and only if

$$\text{the distance } |\bar{X} - 30000| \text{ exceeds } z_{\alpha/2}\sigma/\sqrt{n}$$

or

$$\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n} \text{ does not contain } 30000$$

In general, suppose we test

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

then the rejection region

$$\left|\frac{\hat{\theta} - \theta_0}{\text{SE}}\right| > c_\alpha$$

is equivalent to

$$\hat{\theta} \pm c_\alpha(\text{SE}) \text{ does not contain } \theta_0.$$

Example 4.6.2 (Pooled t-test) For testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$, the pooled-t test

$$\frac{\bar{X} - \bar{Y} - 0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{.025, n_1+n_2-2}$$

is equivalent to

$$\bar{X} - \bar{Y} \pm t_{.025, n_1+n_2-2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \text{ does not contain } 0$$

Example 4.6.3a (One proportion) For testing $H_0 : p = p_0$ versus $H_1 : p \neq p_0$, the z-test

$$\left|\frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}\right| \geq z_{\alpha/2}$$

is the same as $\left(\hat{p} - z_{\alpha/2} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + z_{\alpha/2} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right)$ does not contain 0.

Example 4.6.3b (Two proportions) Suppose we want to compare two population proportions p_1 and p_2 . The difference between two proportions $p_1 - p_2$ is estimated by $\hat{p}_1 - \hat{p}_2$. Since the two sample proportions are independent,

$$\text{Var}(\hat{p}_1 - \hat{p}_2) = \text{Var}(\hat{p}_1) + \text{Var}(\hat{p}_2) = \frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}$$

so that

$$\text{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

A $(1 - \alpha)100\%$ confidence interval for $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

A test for $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$ is

$$\left| \frac{\hat{p}_1 - \hat{p}_2}{\text{SE}} \right| \geq z_{\alpha/2}$$

which is equivalent to the confidence interval $\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}(\text{SE})$ not containing 0.

In the Maternal Flu Vaccine study, positive influenza tests occurred in 16 infants out of 168 in the control group and 6 infants out of 172 in the vaccine group. Using size $\alpha = .05$, does this provide enough evidence that the vaccine is effective?

For $\alpha = .05$, the z-test rejects if $|(\hat{p}_1 - \hat{p}_2)/\text{SE}| \geq 1.96$. We have $\hat{p}_1 = 16/168 = .0952$, $\hat{p}_2 = 6/172 = .0349$, and $\text{SE} = .0277$. Since

$$\frac{\hat{p}_1 - \hat{p}_2}{\text{SE}} = \frac{.0952 - .0349}{.0277} = \frac{.0604}{.0277} = 2.175$$

we reject H_0 and conclude significant difference between the two proportions. Similarly, the 95% confidence interval is

$$\begin{aligned} &\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}(\text{SE}) \\ &.0604 \pm 1.96(.0277) \\ &(.0060, .1147) \end{aligned}$$

Example 4.6.4 (Poisson) Let X_1, \dots, X_{10} be a random sample from Poisson with mean θ . A test for $H_0 : \theta = .1$ against $H_1 : \theta > .1$ would reject for large values of \bar{X} , or equivalently,

$$Y = \sum_{i=1}^{10} X_i \geq k$$

Under the null, $\sum X_i$ has a Poisson distribution with mean $10\theta = 10(.1) = 1$ so that

k	1	2	3	4	5	6
$P(\sum X_i \geq k)$.6321	.2642	.0803	.0190	.0037	.0006

We either use the test $\sum X_i \geq 4$ with size .0190 or $\sum X_i \geq 3$ with size .0803.

Now suppose we have x_1, \dots, X_{100} from Poisson with mean θ and we want a two-sided test $H_0 : \theta = .1$ against $H_1 : \theta \neq .1$. Under the null, $Y = \sum X_i$ has a Poisson distribution with mean 10. Following is a table of k , $P(Y \leq k)$, and $P(Y \geq k)$.

	k	ppois(k,10)	1-ppois(k-1,10)
[1,]	1	0.0004993992	0.999954600
[2,]	2	0.0027693957	0.999500601
[3,]	3	0.0103360507	0.997230604
[4,]	4	0.0292526881	0.989663949
[5,]	5	0.0670859629	0.970747312
[6,]	6	0.1301414209	0.932914037
[7,]	7	0.2202206466	0.869858579
[8,]	8	0.3328196788	0.779779353
[9,]	9	0.4579297145	0.667180321
[10,]	10	0.5830397502	0.542070286
[11,]	11	0.6967761463	0.416960250
[12,]	12	0.7915564764	0.303223854
[13,]	13	0.8644644226	0.208443524
[14,]	14	0.9165415271	0.135535577
[15,]	15	0.9512595967	0.083458473
[16,]	16	0.9729583902	0.048740403
[17,]	17	0.9857223864	0.027041610
[18,]	18	0.9928134954	0.014277614
[19,]	19	0.9965456580	0.007186505
[20,]	20	0.9984117393	0.003454342

Suppose we want type I error rate less than .05. The following gives tests and their size:

RR	Size
$(Y \leq 4) \cup (Y \geq 17)$	$.029 + .027 = .056$
$(Y \leq 4) \cup (Y \geq 18)$	$.029 + .014 = .043$
$(Y \leq 3) \cup (Y \geq 17)$	$.010 + .027 = .037$

The first option exceeds the desired size. The third option is symmetric and may be conveniently written as: Reject $H_0 : \theta = .10$ if

$$\left| \sum_{i=1}^{100} X_i - 10 \right| \geq 7 \quad \text{or} \quad |\bar{X} - .10| \geq .07$$

4.6.1 The p-value

While a test concludes whether to reject the null hypothesis or not, the p-value provides a measure of evidence towards rejection.

Definition 4.6.1. *The p-value of an observed value of the test statistic is the probability of observing a value as extreme as, or more extreme than, the one observed.*

Example (Maternal Flu Vaccine) Respiratory illness with fever occurred in 110 infants out of 153 in the vaccine group, and 153 infants out of 172 in the control group. Calculate a p-value for testing equality of proportions of the two groups.

Solution: The Z-test statistic is

$$\left| \frac{\hat{p}_1 - \hat{p}_2}{SE} \right| = \left| \frac{.8895 - .6547}{.0438} \right| = 5.36$$

A standard normal variable Z exceeds 5.36 in absolute value with probability smaller than .0001, so p-value < .0001. We reject the null and conclude significant difference between the two proportions.

Example Suppose we toss a coin 100 times, and observe 65 heads and 35 tails. Does this allow us to reject $H_0 : p = .5$? Under $H_0 : p = .5$, how often does one observe an imbalance as extreme as 65 percent? Here the p-value is the probability of getting 65 percent or more heads, plus the the probability of getting 35 or fewer heads, i.e.

$$\begin{aligned} \text{p-value} &= P_{p=.50}(65 \text{ or more heads}) + P_{p=.50}(35 \text{ or fewer heads}) \\ &= .0015 + .0015 \\ &= .003 \end{aligned}$$

Comment:

H_0 is rejected if p-value $\leq .05$ so ‘acceptance’ consists of the 95% region of expected behavior, while ‘rejection’ takes the extreme 5%. In the example, 65 heads is in the extreme 5% (and extreme 1%) rather than the middle 95%. This definition of the rejection region provides the following conditional test for comparison of two proportions.

Example (Fisher Exact Test) In a smaller vaccine study, suppose that 5 out of 5 subjects in the control group develop illness, and 1 out of 5 in the treatment group. Is this enough evidence to conclude significant difference?

Soln: The data may be arranged in a contingency table as follows

	Yes	No	
Control	5	0	5
Treatment	1	4	5
	6	4	10

Conditional on the marginals, there are 5 possible arrangements of the cells

5 0	4 1	3 2	2 3	1 4
1 4	2 3	3 2	4 1	5 0

which occur under the hypergeometric distribution with respective probabilities .0238, .2381, .4762, .2381, .0238. None of the other configurations are as extreme as or more extreme than the observed configuration, so p-value=.0238.

Table 1: Summary of tests

Parameter	Estimate	SE	c_α	Comment
μ	\bar{X}	s/\sqrt{n}	$z_{\alpha/2}$	Large sample
	\bar{X}	s/\sqrt{n}	$t_{\alpha/2, n-1}$	Exact but needs normal population
p	\hat{p}	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$z_{\alpha/2}$	Large sample
$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$	$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$	$z_{\alpha/2}$	Large sample
		$\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$	$t_{\alpha/2, n_1+n_2-2}$	Needs normal, $\sigma_1^2 = \sigma_2^2$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$	$t_{\alpha/2, n^*}$	Needs normal
		$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$z_{\alpha/2}$	Large sample
		$\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$		For $H_0 : p_1 = p_2$ where $\hat{p} = \frac{X_1+X_2}{n_1+n_2}$