February 11, 2014

## P-value

- 1. One can perform a number of significance tests at different  $\alpha$  values.
- 2. Tedious and not necessary.
- 3. p-value: the  $\alpha$  level at which we would be indifferent between accepting or rejecting  $H_0$  given the sample data at hand.
- 4. the level at which the given value of the test statistic is on the borderline between the acceptance and rejection regions.  $p = P(t_{n-1} \le t)$



- 5. Compute the p-value for the birthweight data in the previous example The p-value is  $P(t_{99} \le -2.08) = pt(-2.08, 99) = .020$ .
- 6. An alternative definition of p-value: the probability of obtaining a test statistics as extreme as or more extreme than the actual test statistic obtained, given that the null hypothesis is true.
- 7. Guidelines for judging the significance of a p-value.
  - (a) If  $0.01 \le p < 0.05$ , then results are significant.
  - (b)  $0.001 \le p < 0.01$ , then the results are highly significant.
  - (c) If p < 0.001, then the results are very highly significant.

- (d) If p > 0.05, then the results are considered not statistically significant.
- (e) If  $0.05 \le p < 0.10$ , the a trend toward statistical significance is sometimes noted.

## Two sided alternative

We want to compare fasting serum-cholesterol levels among recent Asian immigrants to the US with typical levels found in the general US population. Assuming cholesterol levels in women aged 21-40 in US are approximately normally distributed with mean 190 mg/dL. It is unknown whether cholesterol levels among recent Asian immigrants are higher or lower than those in the general US population. Assuming that levels among recent female Asian immigrants are normally distributed with unknown mean  $\mu$ . We wish to test the null hypothesis  $H_0: \mu = \mu_0 = 190$  versus the alternative hypothesis  $H_1: \mu \neq \mu_0$ . Blood tests are performed on 100 female Asian immigrants aged 21-40, and the mean level is 181.52 mg/dL with standard deviation = 40 mg/dL. What conclusion can we draw?

- 1. The values of the parameter being studied under the alternative hypothesis are allowed to be either greater or less than the values of the parameter under the null hypothesis.
- 2. A reasonable decision rule to test for alternatives on either side of the null mean is, reject H0 if t is either too small or too large.
- 3.  $H_0$  will be rejected if t is either  $\langle c_1 \text{ or } \rangle c_2$  for some constants  $c_1, c_2$  and  $H_0$  will be accepted if  $c_1 \leq t \leq c_2$ .
- 4. P(reject  $H_0 \mid H_0$  true) =  $P(t < c_1 ort > c_2 \mid H_0 true) = P(t < c_1 \mid H_0 true) + P(t > c_2 \mid H_0 true) = \alpha$ .
- 5. To test the hypothesis  $H_0: \mu = \mu_0$  vs.  $H_1: \mu \neq \mu_0$ , with a significance level of  $\alpha$ , the best test is based on  $t = \frac{\bar{x} \mu_0}{s/\sqrt{n}}$ .
- 6. If  $|t| > t_{n-1,1-\alpha/2}$ , then  $H_0$  is rejected.
- 7. if  $|t| \leq t_{n-1,1-\alpha/2}$ , then  $H_0$  is accepted.
- 8. Test the hypothesis that the mean cholesterol level of recent female Asian immigrants is different from the mean in the general US population using the data in Example 7.20, where mean is 181.52, standard deviation is 40 mg/dL, and sample size is 100.  $t = \frac{\bar{x} \mu_0}{s/\sqrt{n}} = \frac{181.52 190}{40/\sqrt{100}} = -8.48/4 = -2.12.$
- 9. The critical values for the two-sided test with  $\alpha = 0.05$  are  $c_1 = t_{99,0.025}, c_2 = t_{99,0.975}$ .

10. 
$$c_1 = qt(0.025, 99) = -1.984, c_2 = qt(0.975, 99) = 1.984$$

Figure 7.3 One-sample t test for the mean of a normal distribution (two-sided alternative)



- 11. Because  $t = -2.12 < -1.984 = c_1$ , we reject  $H_0$  at 5% level of significance.
- 12. Compute the p-value for the hypothesis test in the previous example, where mean is 181.52, standard deviation is 40 mg/dL, and sample size is 100.  $p = 2 \times P(t_{99} < -2.12) = 0.037$ .

**Figure 7.4** Illustration of the *p*-value for a one-sample *t* test for the mean of a normal distribution (two-sided alternative)



## Power of a test

1. Assuming standard deviation is known. Calculate power based on one-sample z test. A new drug is proposed for people with high intraocular pressure (IOP), to prevent the development of glaucoma. A pilot study is conducted with the drug among 10 patients. Their mean IOP decreases by 5 mm Hg with a sd of 10 mm Hg after 1 month of using the drug. The investigator propose to study 100 participants in the main study. Is this a sufficient sample size for the study?

 $H_0: \mu = \mu_0$  vs.  $H_1: \mu < \mu_0$  When the distribution is normal and variance is known.  $H_0$  is rejected if  $Z < Z_{\alpha}$  and  $H_0$  is accepted otherwise. The best test does not depend on the alternative mean  $\mu_1$  What is the difference?

Power = 1 - P(type II error) = P(reject  $H_0 \mid H_0$  false) =  $P(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < z_\alpha \mid \mu = \mu_1)$ Under  $H_1, \bar{X} \sim N(\mu_1, \sigma^2/n)$ . Hence, Power =  $P(\frac{\bar{X}-\mu_1}{\sigma/\sqrt{n}} < (\mu_0 + Z_\alpha \sigma/\sqrt{n} - \mu_1)/(\sigma/\sqrt{n})) = P(Z < Z_\alpha + \frac{(\mu_0 - \mu_1)\sigma}{\sqrt{n}}).$ 





2. Compute the power of the test for the birthweight data with an alternative mean of 115 oz and  $\alpha = 0.05$ , assuming the true standard deviation = 24 oz. We have  $\mu_0 = 120oz, \ \mu_1 = 115oz, \ \alpha = 0.05, \ \sigma = 24, n = 100.$ Power =  $P(Z < Z_{\alpha} + \frac{(\mu_0 - \mu_1)\sqrt{n}}{\sigma})$ P(Z < -1.645 + 5(10)/24) = P(Z < 0.438) = 0.669.

## Alternative is greater than Null

1. The best test is  $H_0$  rejected if  $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > Z_{1-\alpha}$  and  $H_0$  is accepted if  $Z \leq Z_{1-\alpha}$ . The power of the test is given by

$$P(\bar{X} > \mu_0 + Z_{1-\alpha}\sigma/\sqrt{n} \mid \mu = \mu_1) = 1 - P(Z < Z_{1-\alpha} + \frac{(\mu_0 - \mu_1)\sqrt{n}}{\sigma})$$
$$= P(Z < Z_\alpha + \frac{(\mu_1 - \mu_0)\sqrt{n}}{\sigma})$$





2. Using a 5% level of significance and a sample of size 10, compute the power of the test for the cholesterol data with an alternative mean of 190 mg/dL, a null mean of 175 mg/dL, and a standard deviation of 50 mg/dL. We have  $\mu_0 = 175$ ,  $\mu_1 = 190$ ,  $\alpha = 0.05$ ,  $\sigma = 50$ , n = 10.

Power = 
$$P(Z < Z_{\alpha} + \frac{(\mu_0 - \mu_1)\sqrt{n}}{\sigma})$$
  
=  $P(Z < -1.645 + 15\sqrt{10}/50)) = P(Z < -0.696) = 0.243$ 

3. Hence power of a one sample Z test for the mean of a normal distribution with known variance (one-sided alternative) for the hypothesis

$$P(Z < Z_{\alpha} + \frac{(|\mu_0 - \mu_1|)\sqrt{n}}{\sigma})$$

- 4. Factors affecting the power:
  - (a) If the significance level is made small ( $\alpha$  decreases),  $Z_{\alpha}$  decreases and hence the power decreases.
  - (b) If the alternative mean is shifted further away from the null mean  $(|\mu_0 \mu_1|$  increases), then the power increases.
  - (c) If the standard deviation of the distribution of individual observation increases ( $\sigma$  increases), then the power decreases.
  - (d) If the sample size increases, then the power increases.
- 5. Power curve for the birthweight data:



6. Two sided alternative: Reject if  $\frac{(\bar{X}-\mu_0}{\sigma/\sqrt{n}} > Z_{1-\alpha/2}$  or  $\frac{(\bar{X}-\mu_0}{\sigma/\sqrt{n}} < Z_{\alpha/2}$ . Hence power

$$\Phi(Z_{\alpha/2} + \frac{(\mu_0 - \mu_1)\sqrt{n}}{\sigma}) + \Phi(Z_{\alpha/2} + \frac{(\mu_1 - \mu_0)\sqrt{n}}{\sigma})$$

and is approximated by

$$\Phi(Z_{\alpha/2} + \frac{|\mu_0 - \mu_1|\sqrt{n}}{\sigma})$$