Hypothesis testing

Two sample inference

Obstetrics: Mothers with low socioeconomic status (SES) deliver babies with lower than "normal" birthweights. The mean birthweights in US is 120 oz. A list of birthweights is obtained from 100 consecutive, full-term, live-born deliveries from the maternity ward of a hospital in a low-SES area. Two hypotheses:

1. The mean birthweights from this hospital is 120 oz.
2. The mean birthweights from this hospital is lower than 120 oz.

Hypothesis testing provides an objective framework for making decisions using probabilistic methods.

1. **Null hypothesis:** The mean birthweight in the low-SES-area hospital (μ) is equal to the mean birthweights in the US, $\mu_0 = 120$ oz. It is denoted as H_0, the hypothesis to be tested.
2. **Alternative hypothesis:** The mean birthweight is lower than 120 oz. This is denoted by H_1, which contradicts the null hypothesis.

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu < \mu_0$$

3. Four possible outcomes

   (a) Accept H_0, and H_0 is in fact true.

   (b) Accept H_0, and H_1 is in fact true.

   (c) Reject H_0, and H_0 is in fact true.

   (d) Reject H_0, and H_1 is in fact true.

   Practically, it is impossible to prove the null hypothesis is true. Accepting H_0 means we have failed to reject H_0 based on the data.

4. **Type I error:** Rejecting the null hypothesis when it is true. In terms of the previous example, a type I error is deciding the mean birthweight in the hospital was lower than 120 oz when in fact it was 120 oz. Probability of type I error is denoted by α and referred to as the **significance level of a test.**

5. **Type II error:** Accepting the null hypothesis when it is false. Deciding that the mean birthweight was 120 oz when in fact it was lower than 120 oz. Probability of type II error is denoted by β.

6. **Power of a test:** $1 - \beta = 1$ - probability of a type II error = P(rejecting $H_0 \mid H_1$ true).

7. Suppose a new drug is to be tested for pain relief among patients with osteoarthritis (OA). The measure of pain relief will be the percent change in pain level as reported by the patient after taking the medication for 1 month (positive means reduced pain). Fifty OA patients will participate in the study. What hypotheses are to be tested? What do type I error, type II error, and power mean in this situation?
   Hypothesis to be tested: $H_0 : \mu = 0$ versus $H_1 : \mu > 0$, μ : % change in level of pain over a 1-month period. Type I error: P(reject $H_0 \mid H_0$ is true) = Probability of deciding the drug is effective pain reliever, given that the drug has no effect on pain relief. Type II error: P(accept $H_0 \mid H_0$ is false) = Probability of deciding the drug has no effect on pain relief given that the drug is effective. Power of the test: Probability of deciding the drug is effective when it is effective.

8. General aim: Make α and β as small as possible. We can make α small by minimizing P(reject $H_0 \mid H_0$ is true) which leads to rejecting H_0 less often. We can make β small by minimizing P(accept $H_0 \mid H_0$ is false) which leads to accepting H_0 less often. General strategy: Fix α at some specific level and minimize β (maximize the power).
   In the birthweights example, it is intuitive to use the mean of the birthweights as a test. Decision can be made depending the value of the mean.
   Acceptance region: The range of values of the mean for which H_0 is accepted.
   Rejection region: The range of values of the mean for which H_0 is rejected.
   One-sided test (one-tailed test): A test in which the values of the parameters under the alternative hypothesis are allowed to be either greater than or less than the values of the parameter under the null hypothesis.

9. Birthweights example: Rejection region consists of small values. How small should the mean be for H_0 to be rejected?
   The significance level of the test is set at α. If H_0 is rejected or all values of $\bar{x} < c$, the value c should satisfy Probability of type I error = α.
   P(reject $H_0 \mid H_0$ is true) = P($\bar{x} < c \mid H_0$ is true) = α Define test criteria using

standardized values:

$$\begin{aligned} P(\bar{x} < c \mid H_0 \text{ is true}) &= \alpha \\ P(\frac{\bar{x}-\mu_0}{s/\sqrt{n}} < \frac{c-\mu_0}{s/\sqrt{n}}) &= \alpha \\ P(t_{n-1} < \frac{c-\mu_0}{s/\sqrt{n}}) &= \alpha \\ P(t_{n-1} < t_{n-1,\alpha}) &= \alpha. \end{aligned}$$

One sample t test for the mean of a normal distribution with an unknown variance (alternative mean < null mean). To test the hypothesis $H_0 : \mu = \mu_0$, σ unknown vs. $H_1 : \mu < \mu_0$, σ known with a significance level α. Test statistic:

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}.$$

If $t < t_{n-1,\alpha}$, then we reject H_0. If $t > t_{n-1,\alpha}$ then we accept H_0. $t_{n-1,\alpha}$ is called a critical value.

10. Critical-value method of hypothesis testing:

    (a) Select a model assuming null hypothesis is true.

    (b) Decide the significance level.

    (c) Compute the test statistic.

    (d) Compare the test statistic to a critical value determined by the type I error.

    Use the one-sample t test to test the hypothesis $H_0 : \mu = 120$ versus $H_1 : \mu < 120$ based on the birthweight data using a significance level of .05. Sample size is 100, the mean is 115, and sample standard deviation is 24. How about a significance level of .01?

    $$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{115-120}{24/\sqrt{100}} = -2.08$$

    $t_{99,.05} = qt(.05, 99) = -1.66$
    Because $t = -2.08 < -1.66$, we reject H_0 at a significance level of 0.05.
    $t_{99,.01} = qt(.01, 99) = -2.36$ Because $t = -2.08 > -2.36$, we accept H_0 at a significance level of 0.01

P-value

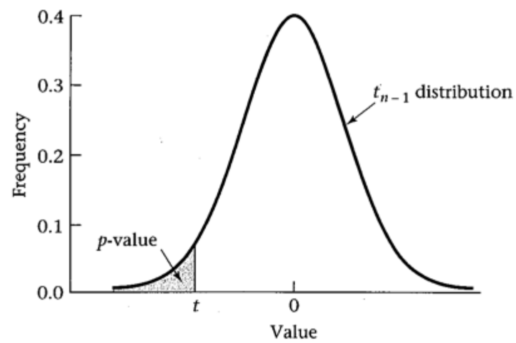
1. One can perform a number of significance tests at different α values.

2. Tedious and not necessary.

3. $p-$value: the α 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. $p = P(t_{n-1} \leq t)$

Figure 7.1 **Graphic display of a *p*-value**



5. Compute the p-value for the birthweight data in the previous example The p-value is $P(t_{99} \leq -2.08) = pt(-2.08, 99) = .020$.

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 $0.01 \leq p < 0.05$, then results are significant.

   (b) $0.001 \leq p < 0.01$, 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 $0.05 \leq p < 0.10$, 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 μ. 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 $< c_1$ or $> c_2$ 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 | H_0 true) $= P(t < c_1 or t > c_2 | H0\, true) = P(t < c_1 \mid H_0\, true) + P(t > c2 \mid H_0 true) = \alpha$.

5. To test the hypothesis $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$, with a significance level of α, 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. $c_1 = qt(0.025, 99) = -1.984, c_2 = qt(0.975, 99) = 1.984$

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

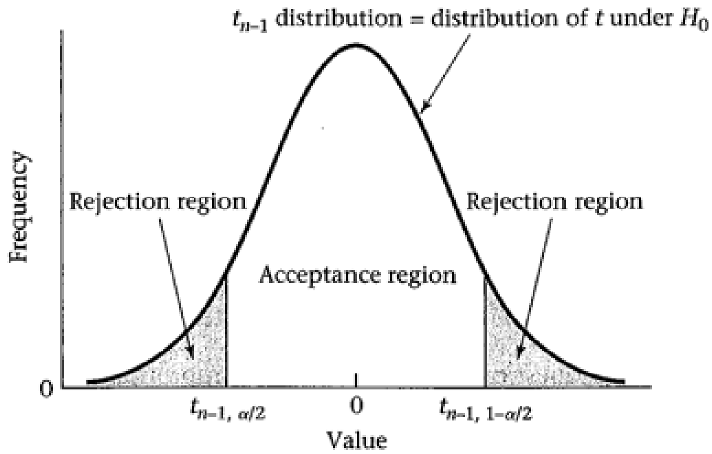


Figure 7.4 Illustration of the p-value for a one-sample t test for the mean of a normal distribution (two-sided alternative)