

## Statistical testing vs. interval estimation

Chp. 8 focused on estimating a parameter ( $\mu$ ,  $p$ , etc.) and placing some bounds on the error of estimation.

- Can describe the population more accurately - knowing the mean income of a population gives us an idea of the average person's economic position
- Can report on the status of a situation - the difference in the percentage of voters supporting Bush and those supporting Gore ( $p_1 - p_2$ ) may be of interest just before an election?

Chp. 10 focuses on statistical testing.

- We have hypotheses about the population, and we use a random sample to help us decide which is accurate
- We may be testing a new medication/treatment, a new scientific theory, etc.
- Is the value of a parameter equal to a particular value, or is it larger? smaller? just different?

## What are the hypotheses?

- Generally, begin with a research hypothesis that we wish to prove/support
- More often, this is called the *alternative hypothesis* and denoted  $H_A$
- Alternative hypotheses usually take one of the three forms:  $\theta \neq \theta_0$ ,  $\theta > \theta_0$ ,  $\theta < \theta_0$ , where  $\theta$  is a population parameter and  $\theta_0$  is a specific value hypothesized for that parameter
- The *null hypothesis* is the hypothesis we fall back on if we don't have enough evidence to support the alternative
- The null hypothesis is denoted as  $H_0$  and typically takes the form  $\theta = \theta_0$

## What is a test statistic?

- As when estimating intervals, our procedure depends on a random sample
- We use test statistics to summarize the information in the sample
- Often our test statistics will be the same estimators that we've used before
- Based on the information from the sample, summarized through the test statistic, we decide whether we have enough evidence to reject the null hypothesis
- The *rejection region* is made up of the values of the test statistic which we consider sufficient evidence to reject the null hypothesis

## Types of errors

- Type I error occurs when  $H_0$  is rejected when it's really true
- Probability of making a type I error is denoted  $\alpha$  and called the *level* of the test
- Type II error occurs when  $H_0$  is not rejected when it should be (when  $H_A$  is true)
- Probability of making a type II error is denoted  $\beta$
- It isn't possible to reduce the probability of both error types to 0 at the same time

## Example: Calculate prob. of type I error

Experience has shown that 30% of people with a certain illness recover. Ten people are selected at random and given an experimental medicine; nine recover shortly thereafter.

1. What are the hypotheses that we want to test?
2. Calculate  $\alpha$ , the probability of type I error, if we decide to reject the null hypothesis when  $Y \geq 8$ ?

## Establishing the rejection region

- In general, we are given  $\alpha$  (the significance level of the test)
- We have to establish a rejection region (RR) based on  $\alpha$
- This means we want to find a RR such that the probability of getting a test statistic in that region, given that the null hypothesis is true, is  $\alpha$
- Once again, we need to know about the sampling distribution of the test statistic

## Example: Two-tailed test

A bottling factory fills thousands of 20oz bottles daily with soda, but not all the bottles are filled to the same level. A random sample of bottles was taken from the factory line, containing the following amounts of soda (in oz):

19.8 20.1 19.7 19.2 19.9 20.0 19.8 19.9 19.7

Assuming that the distribution of amounts of soda is approximately normal, is there evidence to lead us to believe that the mean level of filling is not 20 oz? Test with significance level  $\alpha = 0.05$ .

## Attained levels of significance (p-values)

- If we assume that  $H_0$  is true, we can calculate the probability of seeing a test statistic that is this far away from what we would expect given  $H_0$ , or further
- This probability is the *p-value* or *attained level of significance*
- If the p-value is less than  $\alpha$ , this is equivalent to saying that the test statistic falls in the rejection region
- Reporting the p-value lets each reader know whether the null hypothesis would be rejected using his/her own choice of  $\alpha$ .



## Example: Comparing mean incomes

We want to compare the mean family income in two states. For state 1, we had a random sample of  $n_1 = 100$  families with a sample mean of  $\bar{x}_1 = 35000$ . For state 2, we had a random sample of  $n_2 = 144$  families with a sample mean of  $\bar{x}_2 = 36000$ . Past studies have shown that for both states  $\sigma = 4000$ .

Is the mean income in state 1 lower than that in state 2? Test at significance level  $\alpha = 0.01$ . Give the p-value.

## Summary: Hypothesis testing methodology

1. Establish the null hypothesis.
2. Determine the alternative hypothesis → one-tailed or two-tailed test
3. Define assumptions needed for the test, make sure they are met. (Ex: paired/unpaired, equal variances, etc.)
4. Calculate the test statistic:  
$$Z/T = \frac{\text{estimator for parameter} - \text{value for parameter given by null hypothesis}}{\text{standard error of estimator}}$$
5. Based on  $\alpha$  and  $H_A$ , determine the rejection region. This is the portion of the tail(s) (based on  $H_A$ ) of the estimator's sampling distribution which has area  $\alpha$ .
6. Determine the p-value (attained level of significance). This is the probability of seeing a result as extreme or more extreme than your test statistic under the null hypothesis, where “extreme” is defined by  $H_A$
7. Reject  $H_0$  if  $p < \alpha$  or the test statistic falls in the rejection region

## Example: Diff. in means (small samples)

The strength of concrete depends, to some extent, on the method used for drying it. Two different drying methods yielded the results below for independently tested specimens (in psi). Do the methods appear to produce concrete with a statistically significant difference in mean strengths? Use  $\alpha = 0.05$ . Find the attained significance level.

Method 1     $n_1 = 7$      $\bar{y}_1 = 3250$      $s_1 = 210$

Method 2     $n_2 = 10$      $\bar{y}_2 = 3240$      $s_2 = 190$

## Confidence intervals vs. hypothesis tests

Imagine a two-tailed test with hypotheses  $H_0 : \theta = \theta_0$  and  $H_A : \theta \neq \theta_0$

- For level of significance  $\alpha$ , reject if test statistic (could be  $Z$  or  $T$ ) falls in  $Z < -z_{\frac{\alpha}{2}}$  or  $Z > z_{\frac{\alpha}{2}}$
- This means we fail to reject  $H_0$  when  $-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}$
- We can re-write this:
$$-z_{\frac{\alpha}{2}} < \frac{\hat{\theta} - \theta_0}{SE_{\hat{\theta}}} < z_{\frac{\alpha}{2}}$$
$$-\hat{\theta} - z_{\frac{\alpha}{2}} SE_{\hat{\theta}} < -\theta_0 < -\hat{\theta} + z_{\frac{\alpha}{2}} SE_{\hat{\theta}}$$
$$\hat{\theta} + z_{\frac{\alpha}{2}} SE_{\hat{\theta}} > \theta_0 > \hat{\theta} - z_{\frac{\alpha}{2}} SE_{\hat{\theta}}$$
$$\hat{\theta} - z_{\frac{\alpha}{2}} SE_{\hat{\theta}} < \theta_0 < \hat{\theta} + z_{\frac{\alpha}{2}} SE_{\hat{\theta}}$$
- As long as  $\theta_0$  lies within bounds  $(\hat{\theta} - z_{\frac{\alpha}{2}} SE_{\hat{\theta}}, \hat{\theta} + z_{\frac{\alpha}{2}} SE_{\hat{\theta}})$ , we fail to reject
- These are confidence interval boundaries! A confidence interval can suffice for a 2-tailed test. If interval doesn't include  $\theta_0$ , then we can reject  $H_0$

Can show similar relationship for one-sided CIs and one-tailed hypothesis tests.

## Ex: Compare CIs and hypothesis tests

We have a random sample of checking account balances at each of two banks, and we're interested in testing whether the mean checking account balances are different at the two banks. Use  $\alpha = 0.10$ .

Bank 1     $n_1 = 12$      $\bar{x}_1 = 1000$      $s_1 = 150$

Bank 2     $n_2 = 10$      $\bar{x}_2 = 920$      $s_2 = 120$

## More about type II error

- We've discussed the probability of type I error,  $\alpha$ , associated with hypothesis tests, but we may also be interested in the probability of type II error,  $\beta$
- This is the probability of failing to reject  $H_0$ , when  $H_A$  is really correct
- Type II error is the probability that test statistic does not fall in rejection region, given that the parameter has a value categorized by  $H_A$
- In order to calculate this value, we need to calculate the test statistic given a specific value of the parameter under  $H_A$
- Power of the test =  $1-\beta$

## Example: Probability of type II error

Experience has shown that 30% of people with a certain illness recover. Ten people are selected at random and given an experimental medicine; nine recover shortly thereafter.

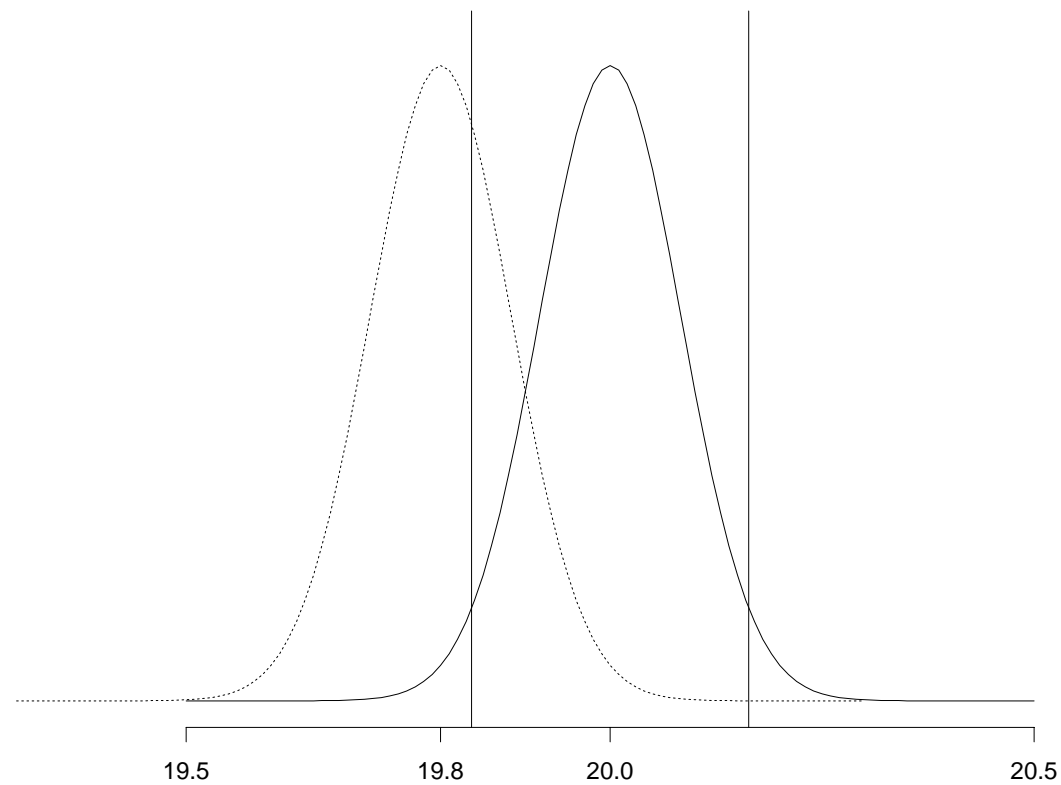
Calculate  $\beta$ , the probability of type II error, with rejection region of  $Y \geq 8$  and  $H_A: p = 0.6$ .

## Example: Probability of type II error

A bottling factory fills thousands of 20oz bottles daily with soda, but not all the bottles are filled to the same level. A random sample of  $n = 9$  bottles was taken from the factory line to try to determine whether the mean filling level is 20oz or not. If we test these hypotheses with  $\alpha = 0.05$ , find the probability of type II error if  $H_A = 19.8$ . Assume that the distribution of amounts of soda is approximately normal and that  $\sigma = 0.25$  (known from previous experience).



## Figure illustrating regions in bottling example



## Example: Probability of type II error

A manufacturer claims that at least 20% of the public prefers her product. A sample of 500 people is taken to check her claim. If we test her claim with  $\alpha = 0.05$ , what is  $\beta$  if the percentage of the public preferring her product is really 22%?

## Answer: Probability of type II error

A manufacturer claims that at least 20% of the public prefers her product. A sample of 500 people is taken to check her claim. If we test her claim with  $\alpha = 0.05$ , what is  $\beta$  if the percentage of the public preferring her product is really 22%?

$$H_0 : p = 0.20 \quad H_A : p > 0.20 \quad \alpha = 0.05$$

$$\text{Test stat: } Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\hat{p} - 0.20}{\sqrt{\frac{(0.20)(0.80)}{500}}}$$

$$\text{RR: reject if } Z > 1.65$$

Find values for  $\hat{p}$  that would cause us to reject  $H_0$ .

$$\frac{\hat{p} - 0.20}{\sqrt{\frac{(0.20)(0.80)}{500}}} > 1.65$$
$$\hat{p} > 0.2295$$

What's probability of getting  $\hat{p} < 0.2295$  (not falling in RR) if  $H_A$  ( $p = 0.22$ ) is true?

$$\begin{aligned} \beta &= P(\hat{p} < 0.2295 | p_A = 0.22) \\ &= P\left(\frac{\hat{p} - p_A}{\sqrt{\frac{p_A(1-p_A)}{n}}} < \frac{0.2295 - 0.22}{\sqrt{\frac{(0.22)(0.78)}{500}}}\right) \\ &= P(Z < 0.51) \\ &= 0.6950 \end{aligned}$$