## **AP Stats Chapter 11 Notes: Significance Tests**

A significance test is a formal procedure for comparing observed data with a hypothesis whose truth we want to assess. The hypothesis is a statement about a population parameter, like the population mean  $\mu$  or a population proportion p. The results of a test are expressed in terms of a probability that measures how well the data and the hypothesis agree.

Example: Vehicle accidents can result in serious injuries to drivers and passengers. When they do, someone usually calls 911. Police, firefighters, and paramedics respond to these emergencies as quickly as possible. Slow response times can have serious consequences for accident victims. In case of life threatening injuries, victims need attention within 8 minutes of the crash.

Several cities have begun to monitor paramedic response time. In one such city, the mean response time to all accidents involving life threatening injuries last year was  $\mu = 6.7$  minutes with a standard deviation of 2 minutes. The city manager shares this information with emergency personnel and encourages them to do better next year. At the end of the next year the city manager selects a simple random sample (SRS) of 400 calls involving life threatening injuries and examines the response times. For this sample, the mean response time was  $\bar{x} = 6.48$  minutes. Do these data provide good evidence that response times have decreased since last year?



**Stating Hypotheses** Null Hypothesis:

Alternative Hypothesis:

## **Conditions for Significance Tests**

- 1. SRS from population of interest
- 2. Normality
- 3. Independent observations

For checking normality of means:

For checking normality of proportions:

#### **Test Statistics**

A significance test uses data in the form of a test statistic. Here are some principles that apply to most tests:

a.

b.

c.

In the example with the response times the null hypothesis was \_\_\_\_\_\_, and the estimate was \_\_\_\_\_\_. Since we are using  $\sigma = 2$  minutes for the distribution of response times this year, our test statistic is

The test statistic z says how far  $\overline{x}$  is from  $\mu_0$  in standard deviation units.

Because the sample result is over two standard deviations below the hypothesized mean 6.7, it gives good evidence that the mean response time this year is not 6.7 minutes but, rather, less than 6.7 minutes.

### **P-Values**

The null hypothesis states the claim we are seeking evidence against. The test statistic measures how much the sample data diverge from the null hypothesis. If the test statistic is large and is in the direction suggested by the alternative hypothesis, we have data that would be unlikely if the

null hypothesis were true. We make "unlikely" precise by calculating a probability called a p-value.

P-Value:

Small p values are evidence against  $H_0$  because they say that the observed result is unlikely to occur when  $H_0$  is true. Large p values fail to give evidence against  $H_0$ .

Example: Suppose we know that differences in job satisfaction scores follow a Normal distribution with standard deviation  $\sigma = 60$ . If there is no difference in job satisfaction between the two work environments, the mean is  $\mu = 0$ . This is the null hypothesis. The alternative hypothesis says simply "there is a difference." This is the alternative hypothesis \_\_\_\_\_\_.

Data from 18 workers gave  $\overline{x} = 17$ . That is, these workers preferred the self paced environment on average. The test statistic is

Because the alternative is two sided, the p value is the probability of getting a z at least as far from 0 in either direction as the observed z = 1.20. As always, calculate the p value taking  $H_0$  to be true. When  $H_0$  is true,  $\mu = 0$ , and z has the standard Normal distribution. The p value is then

Value as far from 0 as  $\bar{x} = 17$  would happen 23% of the time when the true population mean is  $\mu = 0$ . An outcome that would occur so often when  $H_0$  is true is not good evidence against  $H_0$ . **Assignment**: p. 698-699 11.7 to 11.12

## Statistical Significance

We can compare the P-value with a fixed value that we regard as decisive. This amounts to

announcing in advance how much evidence against  $H_0$  we will insist on. The decisive value of P is called the **significance level** ( $\alpha$ )

 $\alpha$  = 0.05 means:

 $\alpha$  = 0.01 means:

## Statistical Significance:

If the p-value is as small as or smaller than alpha, we say that the data are statistically significant at level  $^{\alpha}$  .

\*\*\*Significant in the statistical sense foes not mean "important." It means simply not likely to happen just by chance. And the significance level  $\alpha$  makes not likely more exact.

Think back to the paramedic response time and what level would the 6.48 minutes be statistically significant. The p-value in this case was 0.0139.

 $\alpha = 0.05?$ 

 $\alpha = 0.01?$ 

In practice, the most commonly used significance level is 0.05.

## Interpreting results in context

The final step in a significance test is to draw a conclusion about the competing claims you were testing.

Fail to reject  $H_0$ :

Reject  $H_0$ :

Again with the paramedics, we calculated the p-value for the city manager's study of paramedic

response times as p = 0.0139. If we were using the 0.05 significance level, we would reject  $H_0$ :  $\mu = 6.7$  minutes since our p-value, 0.0139, is less than  $\alpha = 0.05$ . It appears that the mean response time to all life threatening calls this year is less than last year's average o 6.7 minutes.

## z Test for a Population Mean

To test the hypothesis \_\_\_\_\_ based on an SRS of size n from a population with unknown mean  $^{\mu}$  and known standard deviation  $^{\sigma}$ , compute the **one sample z statistic** 

In terms of a variable Z having the standard Normal distribution, the P-value for a test of  $H_0$  against:

These p-values are exact if the population distribution is Normal and are approximately correct for large n in other cases.

Example: The medical director of a large company is concerned about the effects of stress on the company's younger executives. According to the National Center for Health Statistics, the mean systolic blood pressure for males 35 to 44 years of age is 128, and the standard deviation in this population is 15. The medical director examines the medical records of 72 male executives in this age group and finds that their mean systolic blood pressure is  $\bar{x} = 129.93$ . Is this evidence that the mean blood pressure for all the company's younger male executives is different from the national average?

Step 1: Hypotheses

Since  $\sigma$  is known, we will use a one sample z test for a population mean.

- SRS: The 72 records came from a random sample of all executives' annual physicals.
- Normality: We do not know that the population distribution of blood pressures among the company's executives is Normally distributed. But the large sample size is sufficient to guarantee an approximately normal distribution by the central limit theorem.
- Independence: Since the executives are chosen without replacement there must be at least 10\*72 or 720 executives.

Step 3: Calculations

Test Statistic: The one sample z statistic is

P-Value:

Step 4: Interpretation

# Failing to find evidence against $H_0$ means only that the data are consistent with $H_0$ not that we have clear evidence that $H_0$ is true.

The medical director in the example institutes a health promotion campaign to encourage employees to exercise more and eat healthier diet. One measure of the effectiveness of such a

program is a drop in blood pressure. The director chooses a random sample of 50 employees and compares their blood pressure from physical exams given before the campaign and again a year later. The mean change in systolic blood pressure for these n = 50 employees is  $\bar{x}$  = -6. We take the population standard deviation to be  $\sigma = 20$ . The director decides to use an  $\alpha = 0.05$  significance level.

Step 1: Hypotheses

Step 2: Conditions

Since  $\sigma$  is known, we will use a one sample z test for a population mean.

- SRS: Random sample
- Normality: The large sample guarantees approximately normal distribution even if the population distribution is not normal.
- Independence: There must be at least 10\*50 employees in this large company since they were chosen without replacement.

Step 3: Calculations

Test Statistic

P-Value:

Step 4: Interpretation

### Assignment: p. 709-710 11.27 to 11.30 Tests from Confidence Intervals

Confidence Intervals and Two Sided Tests:

Example: The Deely Laboratory analyzes specimens of a drug to determine the concentration of the active ingredient. Such chemical analyses are not perfectly precise. Repeated measurements on the same specimen will give slightly different results. The results of repeated measurements follow a Normal distribution quite closely. The analyses procedure has no bias, so the mean  $^{\mu}$  of the population of all measurements is the true concentration of the specimen. The standard deviation of this distribution is a property of the analysis method and is known to be  $^{\sigma} = 0.0085$  grams per liter. The laboratory analyzes each specimen three times and reports the mean result.

A client sends a specimen for which the concentration of active ingredients is supposed to be 0.86%. Deely's three analyses give concentrations

0.8403 0.8363 0.8447

Is there significant evidence at the 1% level that the true concentration is not 0.86%? This calls for a test of hypotheses

We will carry out the test twice, first with the usual significance test and then form a 99% confidence interval.

What is the mean of the three concentrations?

Test statistic:

Because the alternative hypothesis is two sided, the p-value is

The 99% confidence interval for  $\,^{\mu}\,$  is

Conclusion:

**Assignment:** p. 712-715 11.31 to 11.35, 11.37, 11.40

## Type I and Type II Errors

Type I:

Type II:

Example: A potato chip producer and a supplier of potatoes agree that each shipment must meet certain standards. If less than 8% of the potatoes in the shipment have blemishes, the producer will accept the entire truckload. Otherwise, the truck will be sent away to get another load. The producer inspects a sample instead of the entire truckload. On the basis of the sample results the chip producer uses a significance test to decide whether to accept or reject the shipment. What would the hypotheses be in this case?

Where p is the actual proportion of potatoes with blemishes in a given truckload.

Type I Error:

Type II Error:

Which is more serious, a type I or type II error? It depends...

### Significance and Type I Error:

The significance level  $\alpha$  of any fixed level test is the probability of a Type I error. That is,  $\alpha$  is the probability that the test will reject the null hypothesis when it is in fact true.

**Assignment:** p. 727-728 11.49 a-d, 11.52