Statistical Inference: Review

In statistics, one of the important concepts is that of **statistical inference**: using data from a sample to draw conclusion(s) about the population from which the sample was drawn. Two commonly used methods of statistical inference are (1) confidence intervals and (2) hypothesis testing.

To remind us of what the two methods do, we'll consider the following example.

Ex: A restaurant owner wants to estimate the average dollar purchase per customer. A sample of 81 customers spent an average of $4, and the owner knows that the standard deviation of all the expenditures is $0.63. (a) Construct a 95% confidence interval for the true average expenditure per customer.

Confidence Intervals (C.I.'s)

In this example we want to say something about the mean of a population (all the restaurant customers). The population mean is represented by the Greek letter μ (mu). It is called a **parameter** because it describes a characteristic of the population.

We are given some numbers to work with from the sample. We are told that the mean of the sample is $4. This is represented by $\bar{x}$. It is called a **statistic**, since it comes from sample data. Typically, a statistic is used to **estimate** a parameter. We are also told the standard deviation of the entire population, which is denoted by σ (sigma).

Question: Is σ a parameter or a statistic?

The formula for the $100(1-\alpha)\%$ C.I. for μ in this case is:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \tag{1}$$

where $z_{\alpha/2}$ is a value from the normal distribution. What this value says is, from the normal distribution,

$$P(z > z_{\alpha/2}) = \alpha/2.$$

For a 95% C.I., $z_{\alpha/2} = 1.96$ since $P(z > 1.96) = .05/2 = 0.025$.

In general, confidence intervals take a form like (1):

$$\text{estimate } \pm (\text{critical value}) \times \text{ (standard error of estimate)}$$

Hypothesis Testing

Ex (cont'd): (b) Is there evidence to suggest that the average spending by a customer exceeds \$3.80?

The basic ingredients in hypothesis testing are the hypotheses and the test statistic.

There are 2 hypotheses: the **null hypothesis** (H_o) and the **alternative hypothesis** (H_a). H_o is seen as the status quo, or the present state of things. It is basically what is happening now, and I won't change my mind to believe something else unless you give me very strong evidence to support your claim.

If we are doing a hypthesis test about the population mean μ, we always have a null hypothesis that states that μ is equal to some specified value: $H_o : \mu = \mu_o$.

Whatever is being claimed goes in H_a. This is where the burden of proof lies; we continue to believe that H_o is true unless there is very strong evidence in the data to refute it.

In testing about μ, H_a would have one of 3 possible forms:
(i) $H_a : \mu > \mu_o$ (**upper-tailed test**);
(ii) $H_a : \mu < \mu_o$ (**lower-tailed test**);
(iii) $H_a : \mu \neq \mu_o$ (**two-tailed test**).

Note that the hypotheses make statements about parameters (population terms). We would **never** write $H_o : \bar{x} = 10$, since $\bar{x}$ comes from a sample, not the population.

In our example, what would H_o and H_a be? The claim in the problem is that the mean exceeds \$3.80. Therefore we would write
$H_o : \mu = 3.80$ and $H_a : \mu > 3.80$.

Agreement of the data with H_o is measured by a **test statistic**, which combines the data and the value of the parameter specified in H_o. In our example, the test statistic we would use is

$$z_{obs} = \frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}}$$

When we would evaluate the test statistic in this example, we would use $\mu_o = 3.80$.

Once we have the test statistic, we must make a conclusion. There are 2 approaches used. The first uses the **p–value**, or observed significance level. Assuming that H_o is true, this is the probability of obtaining a value of the test statistic as or more extreme than the value actually observed. For our 3 H_a possibilities above, the p–values are calculated as:
(i) p–value $= P(z \geq z_{obs})$
(ii) p–value $= P(z \leq z_{obs})$
(iii) p–value $= 2P(z \geq |z_{obs}|)$

The strength of the evidence against H_o is determined by the size of the p–value:

RULE: The smaller the p–value, the greater the evidence against H_o.

The logic is that if H_o is true, extreme values for the test statistic are unlikely, and therefore a possible indication that H_o is not true. By convention we draw the following conclusions:

P–value	Strength of evidence against H_o
$> .10$	little or none
$.05 <$ p–value $\leq .10$	weak
$.01 <$ p–value $\leq .05$	strong
$< .01$	very strong

The second approach is sometimes termed the **rejection region** method. The rejection region consists of a range of values for the test statistic which will lead to rejection of H_o. Two types of error are possible with this approach. A **type I error** occurs if H_o is rejected when it is really true. A **type II error** occurs if H_o is not rejected when it is really false. Type I error is considered to be much more important than type II error. A common analogy is with court cases. The presumption of innocence (H_o) is rejected only when the evidence is very convincing against a defendant. The type I error would be to convict the client when they are innocent, while the type II error would be to release the client when they were really guilty. The type I error is seen as much more serious in this case. Recognizing the seriousness of the type I error, the rejection region is chosen so that the probability of making a type I error will not exceed a specified value. This P(type I error) is called α. We sometimes call α the **significance level** of the test.

In testing about μ, the rejection region criteria would be:

(i) If $z_{obs} > z_\alpha$, reject H_o.
(ii) If $z_{obs} < -z_\alpha$, reject H_o.
(iii) If $|z_{obs}| > z_{\alpha/2}$, reject H_o.

Keep in mind that we can only use the rejection region approach to draw a conclusion if a value of α has been specified or provided to us.

Is there a connection between using p–values and the rejection region approach? Yes, there is:

RULE: If p-value $\leq \alpha$, then reject H_o.