Chapter 6

Inferences Based on Two Samples

Frequently we want to use statistical techniques to compare two populations. For example, one might wish to compare the proportions of families with incomes below the poverty line in two regions of the country. Or we might want to determine whether electrical consumption in a community has increased during the past decade.

6.1 Comparison of Two Population Means

Take two populations with means μ_1 and μ_2. The central limit theorem tells us that sample means from these populations will be approximately normally distributed for large samples.

Suppose we select independent random samples of n_1 and n_2, both reasonably large, from the respective populations. We want to make inferences about the difference $\mu_2 - \mu_1$ on the basis of the two samples.

From the statistical theory developed in Chapter 3 (section 3.6) we know that

$$E\{\bar{Y} - \bar{X}\} = E\{\bar{Y}\} - E\{\bar{X}\} = \mu_2 - \mu_1$$

and, since the samples are independent,

$$\sigma^2\{\bar{Y} - \bar{X}\} = \sigma^2\{\bar{Y}\} + \sigma^2\{\bar{X}\}.$$

And it is natural to use

$$s^2\{\bar{Y} - \bar{X}\} = s^2\{\bar{Y}\} + s^2\{\bar{X}\}$$

as an unbiased point estimator of $\sigma^2\{\bar{Y} - \bar{X}\}$.

We can proceed in the usual fashion to construct confidence intervals and statistical tests. Suppose, for example, that a random sample of 200 households from a large community was selected to estimate the mean electricity use per household during February of last year and another simple random sample of 250 households was selected, independently of the first, to estimate mean electricity use during February of this year. The sample results, expressed in kilowatt hours, were

Last Year	This Year
$n_1 = 200$	$n_2 = 250$
$\bar{X} = 1252$	$\bar{Y} = 1320$
$s_1 = 157$	$s_2 = 151$

We want to construct a 99 percent confidence interval for $\mu_2 - \mu_1$.

An unbiased point estimate of $\mu_2 - \mu_1$ is

$$\bar{Y} - \bar{X} = 1320 - 1252 = 68.$$

The standard error of the difference between the sample means is

$$s\{\bar{Y} - \bar{X}\} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{157^2}{200} + \frac{151^2}{250}}$$

$$= \sqrt{123.45 + 91.20} = 14.64.$$

The 99 percent confidence interval will thus be

$$68 \pm z(1 - .01/2)(14.64) = 68 \pm (2.576)(14.64)$$

or

$$30.29 \leq \mu_2 - \mu_1 \leq 105.71.$$

The fact that the above confidence interval does not include zero makes it evident that a statistical test of the null hypothesis that $\mu_2 - \mu_1 \leq 0$ is likely to result in rejection of that null. To test whether the mean household use of electricity increased from February of last year to February of this year, controlling the α-risk at .005 when $\mu_2 = \mu_1$, we set

$$H_0\colon\ \mu_2 - \mu_1 \leq 0$$

and

$$H_1\colon\ \mu_2 - \mu_1 > 0.$$

The critical value of z is $z(.995) = 2.576$. From the sample,

$$z^* = \frac{68}{14.64} = 4.645,$$

which is substantially above the critical value. The P-value is

$$P(z > 4.645) = 0000017004.$$

We conclude that per-household electricity consumption has increased over the year.

6.2 Small Samples: Normal Populations With the Same Variance

The above approach is appropriate only for large samples. In some cases where samples are small (and also where they are large) it is reasonable to assume that the two populations are normally distributed with the same variance. In this case

$$E\{\bar{Y} - \bar{X}\} = E\{\bar{Y}\} - E\{\bar{X}\} = \mu_2 - \mu_1$$

as before but now

$$\sigma^2\{\bar{Y} - \bar{X}\} = \sigma^2\{\bar{Y}\} + \sigma^2\{\bar{X}\}.$$

$$= \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right].$$

To calculate confidence intervals we need an estimator for σ_2. It turns out that the *pooled* or *combined estimator*

$$\begin{aligned} s_c^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)} \\ &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \end{aligned} \qquad (6.1)$$

is an unbiased estimator of σ^2. We can thus use

$$s^2\{\bar{Y} - \bar{X}\} = s_c^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$$

as an unbiased estimator of $\sigma^2\{\bar{Y} - \bar{X}\}$.

We proceed as usual in setting the confidence intervals except that, given the small samples, the test statistic

$$\frac{(\bar{Y} - \bar{X}) - (\mu_2 - \mu_1)}{s\{\bar{Y} - \bar{X}\}}$$

is distributed as $t(n_1+n_2-2)$—that is, as a t-distribution with $v = n_1+n_2-2$ degrees of freedom.

By making the assumptions of normality and equal variance we can use small samples whereas in the general case of the previous section the sample sizes had to be large enough to justify approximate normality according to the Central Limit Theorem.

Now consider an example. Suppose we wish to estimate the difference in mean tread life for a certain make of automobile tire when it is inflated to the standard pressure as compared to a higher-than-standard pressure to improve gas mileage. Two independent random samples of 15 tires were selected from the production line. The tires in sample 1 were inflated to the standard pressure and the tires in sample 2 were inflated to the higher pressure. Tread-life tests were conducted for all tires with the following results, expressed in thousands of miles of tread life.

Standard Pressure	Higher Pressure
$n_1 = 14$	$n_2 = 15$
$\bar{X} = 43$	$\bar{Y} = 40.7$
$s_1 = 1.1$	$s_2 = 1.3$

Because one tire in sample 1 turned out to be defective it was dropped from that sample, reducing the sample size to 14.

Note that the respective populations here are the infinite populations of tread lives of non-defective tires of the make tested when inflated to the standard and higher pressures respectively. We suppose that on the basis of other evidence it is reasonable to assume that both populations are normal with the same variance.

So we have

$$\bar{Y} - \bar{X} = 40.7 - 43.0 = -2.3$$

as an unbiased point estimate of $\mu_2 - \mu_1$. In addition, we have

$$s_c^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(13)(1.1)^2 + (14)(1.3)^2}{14 + 15 - 2} = 1.45899,$$

so that

$$s^2\{\bar{Y} - \bar{X}\} = 1.45899 \left[\frac{1}{14} + \frac{1}{15}\right] = (1.45899)(.0714 + .0667) = .2015,$$

which implies that $s\{\bar{Y} - \bar{X}\} = .4488$. The 95 percent confidence interval is thus

$$-2.3 \pm t(1 - .05/2; 14 + 15 - 2)(.4488) = -2.3 \pm t(.975; 27)(.4488)$$

$$= -2.3 \pm (2.052)(.4488) = -2.3 \pm .9245.$$

Hence,

$$-3.2245 \leq \mu_2 - \mu_1 \leq -1.3755.$$

The mean life of non-defective tires inflated to the higher pressure is between 1.38 and 3.22 thousand miles less than that of non-defective tires inflated to the standard pressure, with 95 percent confidence.

The result of a test of the null hypothesis that $\mu_2 - \mu_1 > 0$ is obvious from the confidence interval above if we control the α-risk at .025 when $\mu_2 = \mu_1$. The critical value for t is -2.060 while

$$t^* = \frac{-2.3}{.4488} = -5.125.$$

The P-value is

$$P(t(27) < -5.125) = 0.00000050119.$$

We conclude that the tread life of tires inflated to the higher pressure is less than that for tires inflated to the standard pressure.

6.3 Paired Difference Experiments

Suppose that we want to find the weight loss in a shipment of bananas during transit. The procedures used above would suggest that we select and weigh a random sample of banana bunches before loading and then select and weigh another independent random sample of banana bunches after shipment and unloading. The differences in the mean weights before and after could then be used to make inferences about the weight loss during shipment. But there is a better way of handling this problem.

The better way would be to select and weigh a random sample of banana bunches before loading and then weigh the same bunch again after shipment and unloading. We could use the mean difference between the 'before' and 'after' weights of the sample of banana bunches to make inferences about the weight loss during shipping. It is important here that the sample of banana bunches be treated in the same way during transit as the rest of the shipment. To ensure that this is the case we would have to mark the selected bunches of bananas in a way that would identify them to us after shipment but not to the people handling the shipping process. The shipping company would therefore not be able to cover up weaknesses in its handling of the shipment by giving the sample of banana bunches special treatment. In this case we are using *matched samples* and making the inference on the basis of *paired differences*.

By using paired differences we can take advantage of the fact that the 'before' and 'after' means are positively correlated—banana bunches which were heavier than average before shipment will also tend to be heavier than average after shipment. The covariance between the 'before' and 'after' weights is therefore positive so that the variance of the difference between the 'before' and 'after' mean weights will be less than the variance of the difference between the mean weights of independently selected random samples before and after shipment. That is,

$$\sigma^2\{\bar{Y} - \bar{X}\} = \sigma^2\{\bar{Y}\} + \sigma^2\{\bar{X}\} - 2\sigma\{\bar{Y}\bar{X}\} < \sigma^2\{\bar{Y}\} + \sigma^2\{\bar{X}\}.$$

It is thus more efficient to work directly with the paired differences in weights than with differences of mean weights. Indeed, if we select matched samples it is inappropriate to use the procedures of the previous sections because the matched samples are not independent of each other as required by those procedures.

So we can set

$$D_i = Y_i - X_i$$

where Y_i is the weight of the ith bunch before shipment and X_i is the weight of that same bunch after shipment. We can then calculate

$$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}$$

and

$$s_D^2 = \sum_{i=1}^{n} \frac{(D_i - \bar{D})^2}{n-1}$$

from whence

$$s_{\bar{D}} = \sqrt{\frac{s_D^2}{n}}.$$

Consider another example. Suppose that a municipality requires that each residential property seized for non-payment of taxes be appraised independently by two licensed appraisers before it is sold. In the past 24 months, appraisers Smith and Jones independently appraised 50 such properties. The difference in appraised values $D_i = Y_i - X_i$ was calculated for each sample property, where X_i and Y_i denote Smith's and Jones' respective appraised values. The mean and standard deviation of the 50 differences were (in thousands of dollars)

$$\bar{D} = 1.21$$

and

$$s_D = 2.61$$

respectively. It thus follows that

$$s_{\bar{D}} = \frac{2.61}{\sqrt{50}} = \frac{2.61}{7.07} = .3692.$$

The 95 percent confidence interval for the mean difference in appraised values for these two appraisers is

$$1.21 \pm z(.975)(.3692) = 1.21 \pm (1.96)(.3692) = 1.21 \pm .724$$

which implies

$$.486 \leq \mu_D \leq 1.934.$$

The confidence interval applies to the hypothetical population of differences in appraised values given independently by Jones and Smith to properties of a type represented by those in the sample, namely, properties seized for non-payment of taxes.

Suppose that an observer who has not seen the sample suspects that Jones' appraised values tend to be higher on average than Smith's. To test whether this suspicion is true, setting the α-risk at .025 when $\mu_D = 0$, we set the null hypothesis

$$H_0\colon\ \mu_D \leq 0$$

and the alternative hypothesis

$$H_1\colon\ \mu_D > 0.$$

The critical value of z is 1.96. The value of z given by the sample is

$$z^* = \frac{1.21}{.3692} = 3.277.$$

We conclude that Jones' appraised values are on average higher than Smith's. The result of this hypothesis test is obvious from the fact that the confidence interval calculated above did not embrace zero.

Note that in the above example we used the normal approximation because the sample size of 50 was quite large. Had the sample size been small, say 8, we would have used the t-distribution, setting the critical value and confidence limits according to $t(.975; 7)$.

6.4 Comparison of Two Population Proportions

Inferences about two population proportions based on large samples can be made in straight-forward fashion using the relationships

$$E\{\bar{p}_2 - \bar{p}_1\} = p_2 - p_1$$

and

$$\sigma^2\{\bar{p}_2 - \bar{p}_1\} = \sigma^2\{\bar{p}_2\} + \sigma^2\{\bar{p}_1\}$$

and approximating the latter using

$$s^2\{\bar{p}_2 - \bar{p}_1\} = s^2\{\bar{p}_2\} + s^2\{\bar{p}_1\},$$

where

$$s^2\{\bar{p}_i\} = \frac{\bar{p}_i(1 - \bar{p}_i)}{n_i - 1}.$$

We use $(n_i - 1)$ in the denominator of the above expression for the same reason that $(n - 1)$ appears in the denominator of

$$s^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n - 1}.$$

Now consider an example. A manufacturer of consumer products obtains data on breakdowns of two makes of microwave ovens. In a sample of $n_1 = 197$ ovens of make 1 it is found that 53 broke down within 5 years of manufacture, whereas in a sample of $n_2 = 290$ ovens of make 2, only 38 ovens broke down within 5 years of manufacture. Assume that the samples are independent random samples from their respective populations. We want a 99 percent confidence interval for $p_2 - p_1$. We have

$$\bar{p}_2 - \bar{p}_1 = \frac{38}{290} - \frac{53}{197} = .13103 - .26904 = -.1380$$

$$s^2\{\bar{p}_1\} = \frac{(.26904)(.73096)}{196} = .00100335$$

$$s^2\{\bar{p}_2\} = \frac{(.13103)(.86897)}{289} = .00039439$$

$$s\{\bar{p}_2 - \bar{p}_1\} = \sqrt{.00100335 + .00039439} = .0374.$$

The 99 percent confidence interval is

$$-.1380 \pm z(.995)(.0374) = -.1380 \pm (2.576)(.0374) = -.1380 \pm .096$$

or

$$-.234 \leq p_2 - p_1 \leq -.042.$$

The percentage of units of make 1 that break down within 5 years of manufacture is between 4.2 and 23.4 percentage points more than that of make 2, with 99 percent confidence.

Now we want to test whether the proportion breaking down within one year for make 1 is larger than the proportion for make 2, controlling the α-risk at .005 when $p_2 = p_1$. We set

$$H_0\colon\ p_2 - p_1 \geq 0$$

and

$$H_1\colon\ p_2 - p_1 < 0.$$

The critical value of z is -2.576. To calculate z^* we need an estimate of $\sigma\{\bar{p}_2 - \bar{p}_1\}$ when $p_2 = p_1 = p$. The appropriate procedure is to use a *pooled estimator* of p to calculate an estimate of $\bar{p}$. We simply take a weighted average of $\bar{p}_1$ and $\bar{p}_2$ using the sample sizes as weights:

$$\bar{p}\,' = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2}.$$

We thus have

$$\bar{p}\,' = \frac{(197)(.26904) + (290)(.13103)}{197 + 290} = .185.$$

An appropriate estimator of $\sigma^2\{\bar{p}_2 - \bar{p}_1\}$ is thus

$$s^2\{\bar{p}_2 - \bar{p}_1\} = \bar{p}\,'(1 - \bar{p}\,')\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$$

which yields

$$s\{\bar{p}_2 - \bar{p}_1\} = \sqrt{(.185)(.815)\left[\frac{1}{197} + \frac{1}{290}\right]} = .0378.$$

The resulting value of z^* is thus

$$z^* = \frac{-.1380 - 0}{.0378} = -3.65.$$

We conclude that the proportion of microwave ovens of make 1 breaking down within 5 years of manufacture is greater than the proportion of microwave ovens of make 2 breaking down within 5 years of manufacture. The P-value is

$$P(z < -3.65) = .00013112.$$

6.5 Exercises

1. Two random samples are independently drawn from two populations. A two-tailed test is used to evaluate H_0: $\mu_x = \mu_y$.

	X	Y
Sample size (n)	3	5
Mean	7.0	3.0
Variance	1.0	2.5

Find the lowest value of α at which the researcher will reject the null hypothesis. (.015) What assumptions did the researcher have to make about the populations to do this test?

2. The following describe the results of independent samples drawn from different populations.

Sample 1	Sample 2
$n_1 = 159$	$n_2 = 138$
$\bar{X}_1 = 7.4$	$\bar{X}_2 = 9.3$
$s_1 = 6.3$	$s_2 = 7.1$

a) Conduct a test of the hypothesis H_0: $\mu_1 - \mu_2 \geq 0$ against the alternative H_1: $\mu_1 - \mu_2 < 0$ with a significance level $\alpha = 0.10$.

b) Determine the P-value for the test statistic of a) above.

3. A pharmaceutical company wishes to test whether a new drug that it is developing is an effective treatment for acne (a facial skin disorder that is particularly prevalent among teenagers). The company randomly selects 100 teenagers who are suffering from acne and randomly divides them into two groups of 50 each. Members of Group 1 receive the drug each day while members of Group 2 receive no medication. At the end of three months,

members of both groups are examined and it is found that 27 of the teenagers in Group 1 no longer have acne as compared with 19 of the teenagers in Group 2 who no longer have acne. Using a significance level of $\alpha = 0.01$, set up and conduct a test of whether the drug is effective or not. Determine the P-value for your test statistic. (.40675)

4. A public opinion research institute took independent samples of 500 males and 500 females in a particular U.S. state, asking whether the respondents favoured a particular constitutional amendment. It was found that 335 of the males and 384 of the females were in favour of the amendment. Construct a 90% confidence interval for difference between the proportions of males and females favouring the amendment and test the hypothesis that the proportions are the same.

5. A manufacturer of automobile shock absorbers was interested in comparing the durability of his shocks with that of the shocks of his biggest competitor. To make the comparison, one of the manufacturer's and one of the competitor's shocks were randomly selected and installed on the rear wheels of each of six cars. After the cars had been driven 20,000 miles, the strength of each test shock was measured, coded and recorded. The results were as follows

Car	Manufacturer's Shock	Competitor's Shock
1	8.8	8.4
2	10.5	10.1
3	12.5	12.0
4	9.7	9.3
5	9.6	9.0
6	13.2	13.0

a) Do the data present sufficient evidence to conclude that there is a difference in the mean strength of the two types of shocks after 20,000 miles of use?

b) Find the approximate observed significance level for the test and interpret its value?

c) What assumptions did you make in reaching these conclusions.

6. A sociologist is researching male attitudes toward women. For her study, random samples of male students from the City of Toronto are interviewed and their results are tabulated. Sample One was conducted in 1988 and consisted of $n_1 = 100$ boys aged 6 to 8. From this group, $x_1 = 90$ subjects indicated in their responses that "girls are ugly, girls have cooties, girls eat worms and that all girls should just go away." The researcher concluded that a large proportion of young boys just don't like girls. A second sample conducted in 1998 consisting of $n_2 = 225$ boys also aged 6 to 8. From this group $x_2 = 180$ subjects exhibited beliefs similar to those 90 boys in the first sample. Using both samples, develop an hypothesis test to evaluate whether the proportion of boys who don't like girls has changed significantly over the 10 year period. When required, manage the α-risk at 5%. Provide a P-value for the test. What does it say regarding attitudes?

7. You know from earlier studies that about 7% of all persons are left-handed. You suspect that left-handedness is more prevalent among men than among women and wish to use independent random samples to measure the difference between the proportions of men and women who are left-handed. You would like an 80% confidence interval for this difference to be accurate within ± 0.01.

a) How many persons should be included in your sample? (91)

b) Will the sample size determined in a) above be large enough to permit a 95% confidence interval for the proportion of men who are left-handed to be accurate to within ± 0.01?

8. In an economics class of 100 students the term mark, T, for each student is compared with his/her marks on two term texts, X_1 and X_2, with $T = X_1 + X_2$. The summary statistics for the entire class were:

	Mean Mark	Standard Deviation
First Term Test	32.0	8.0
Second Term Test	36.0	6.0
Term Mark	68.0	12.0

a) Determine the values for the covariance and the correlation between X_1 and X_2.

b) Calculate the mean and standard deviation of the paired difference in the marks between the first and second term tests.

c) Conduct a test as to whether students performed better on the first term test than the second.