STA 580 — Spring 2011 — Dr. Charnigo

Lecture 7

Two-sample problems

Introduction. In Lectures 4 through 6 we discussed how to make inferences about (parameters describing) a single population using a single sample. However, these approaches are not always adequate for addressing a question of scientific interest.

For example, consider the scenario in "Hypertension" on page 287. We can try to find whether there is a positive mean reduction in blood pressure for those who take the new medication, but that does not tell us whether people should take the new medication. First, finding that there is a positive mean reduction in blood pressure with the new medication does not give us any clue about how much of the reduction is attributable to the pharmacological effect of the new medication as opposed to the placebo effect. Second, finding that there is a positive mean reduction in blood pressure with the new medication does not reveal whether the new medication performs better than, similarly to, or worse than an existing medication.

Independent samples. Suppose that we intend to take a sample of size n_1 from one population and a sample of size n_2 from a second population. Let $X_1, \ldots, X_{n_1}$ denote the random conceptualizations of sample values for the first sample and $Y_1, \ldots, Y_{n_2}$ the random conceptualizations of sample values for the second sample. Suppose that $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ are mutually independent random variables. Then we say that the two samples are independent.

Intuitively, this means that how we choose the second sample is unrelated to how we choose the first sample; in particular, there is no specific subject in the second sample corresponding to a given subject in the first sample (Definition 8.5). If we let $\mu_1 := E[X_1]$ denote the mean for the first population and $\mu_2 := E[Y_1]$ the mean for the second population, then we may be interested in testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$ (or, equivalently, $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 \neq 0$).

For example, reconsidering the scenario in "Hypertension" on page 287, we can let μ_1 denote the mean reduction in blood pressure for people on the new medication and μ_2 the mean reduction in blood pressure for people on an existing medication. We can recruit n_1 subjects to receive the new medication; these subjects constitute an "experimental group". We can then recruit n_2 unrelated subjects to receive the existing medication; these subjects constitute a "control group".

We can use one of the methods described later in this lecture to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$.

Paired samples. Now suppose that we intend to take a sample of size n from one population and another sample of size n from a second population. Let $X_1, \ldots, X_n$ denote the random conceptualizations of sample values for the first sample and $Y_1, \ldots, Y_n$ the random conceptualizations of sample values for the second sample. Suppose that $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ are not mutually independent random variables but that the n differences $X_1 - Y_1, \ldots, X_n - Y_n$ are mutually independent. Then we say that the two samples are paired.

Paired samples generally arise in one of two ways (Definition 8.4). The first possibility is that, once the i^{th} member of the first sample is chosen, we deliberately choose someone "similar" (e.g., of the same gender and approximately the same age) to be the i^{th} member of the second sample. This strategy is referred to as "matching". The second possibility is that there is really only one group of subjects, but we obtain measurements for each subject under two different experimental conditions. In either case, we may be interested in testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$, where $\mu_1 := E[X_1]$ is the mean for the first population and $\mu_2 := E[Y_1]$ is the mean for the second population.

For example, reconsidering the scenario in "Hypertension" on page 287, we can again let μ_1 denote the mean reduction in blood pressure for people on the new medication and μ_2 the mean reduction in blood pressure for people on the existing medication. But now imagine recruiting just one group of n subjects. These subjects will take the new medication and have their blood pressure reductions recorded. Then they will not take any medication for a while, to clear their systems. Next they will take the existing medication and have their blood pressure reductions recorded. In effect, these n subjects will serve as their own controls.[1]

To test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$, we can use one of the methods from Lecture 5. Specifically, let $D_1 := X_1 - Y_1$, $D_2 := X_2 - Y_2$, and so forth. Then

$$\mu_D := E[D_1] = E[X_1 - Y_1] = E[X_1] - E[Y_1] = \mu_1 - \mu_2,$$

the second last equality being due to Equation 5.8. Hence, testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$ is logically equivalent to testing $H_0 : \mu_D = 0$ against $H_1 : \mu_D \neq 0$. If n is large, we can test $H_0 : \mu_D = 0$ against $H_1 : \mu_D \neq 0$ by calculating a z statistic from the realized values $d_1, \ldots, d_n$. If n is small and $D_1, \ldots, D_n$ are normally distributed, we can test $H_0 : \mu_D = 0$ against $H_1 : \mu_D \neq 0$ by calculating a t statistic from the realized values $d_1, \ldots, d_n$. The latter technique is referred to as the paired t-test (Equation 8.4).

---

[1] This is not the best experimental design available if we wish to use subjects as their own controls. For a better experimental design see Section 13.10, which will be covered in CPH 630/STA 681.

Comparing means (independent samples)

Preliminary formulation. Let μ_1 and σ_1^2 denote the mean and variance of the first population. Let μ_2 and σ_2^2 denote the mean and variance of the second population. Consider testing

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{against} \quad H_1 : \mu_1 - \mu_2 \neq 0.$$

Intuitively, a reasonable estimate for $\mu_1 - \mu_2$ is $\bar{x} - \bar{y}$ because $\bar{x}$ estimates μ_1 and $\bar{y}$ estimates μ_2. So, we can base our test on whether $\bar{x} - \bar{y}$ is "close" to 0.

If the populations are normal, we already know that $\bar{X}$ and $\bar{Y}$ are normal. Equation 5.13 says that the difference $\bar{X} - \bar{Y}$ is also normal. Moreover, Equations 5.8 and 5.9 yield

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

and

$$Var[\bar{X} - \bar{Y}] = Var[\bar{X}] + Var[\bar{Y}] = \sigma_1^2/n_1 + \sigma_2^2/n_2.$$

Hence, if H_0 is true, then

$$Z := \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

is standard normal. So, defining

$$z := \frac{\bar{x} - \bar{y}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

and rejecting H_0 if $|z| > z_{1-\alpha/2}$ yields a testing procedure with significance level α such that H_0 is rejected if $\bar{x} - \bar{y}$ is not "close" to 0.

All of the above computations are approximately valid if the populations are not normal but n_1 and n_2 are both large.

Practical formulation for large samples. The population variances σ_1^2 and σ_2^2 are not known in practice. With large sample sizes, the sample variances s_x^2 and s_y^2 closely approximate the population variances σ_1^2 and σ_2^2. So, we may redefine

$$z := \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/n_1 + s_y^2/n_2}}$$

and reject H_0 if $|z| > z_{1-\alpha/2}$.[2] The p-value is $2P(Z > z)$ if $z > 0$ and $2P(Z \leq z)$ if $z \leq 0$, where Z has a standard normal distribution and z is the numerical test statistic. Also, the $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is

$$\bar{x} - \bar{y} \pm z_{1-\alpha/2} \sqrt{s_x^2/n_1 + s_y^2/n_2}.$$

As noted in Lectures 4 and 5, there is not universal agreement about when samples are large enough to replace unknown population variances with sample variances. My recommendation is to use the large-sample procedure when n_1 and n_2 are both greater than 200, use the large-sample procedure when n_1 and n_2 are both greater than 30 but the populations cannot be assumed normal, use one of the small-sample procedures from this lecture when $n_1 < 200$ or $n_2 < 200$ and the populations can be assumed normal, and use a different strategy altogether (some options will be given in Lecture 8) when $n_1 < 30$ or $n_2 < 30$ but the populations cannot be assumed normal.

Small samples and equal variances. Suppose that both populations are normal and that $\sigma_1^2 = \sigma_2^2$. Let σ^2 denote the common (but unknown) value of

---

[2] If we wanted to test $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 > 0$, we would change the rejection criterion to $z > z_{1-\alpha}$. Similarly, if we wanted to test $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 < 0$, we would change the rejection criterion to $z < -z_{1-\alpha}$. However, in much the same way that the research community seems to have "settled on" using $\alpha = 0.05$, the research community almost always uses two-sided tests for two-sample problems. Analogous footnotes could be made later in this lecture for t statistics and f statistics.

σ_1^2 and σ_2^2. We can estimate σ^2 by taking a weighted average of s_x^2 and s_y^2, the weights being related to the sample sizes. This weighted average is

$$s^2 := \frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2}$$

and is referred to as a “pooled estimate” (Equation 8.10) since we are using the data from both samples to estimate σ^2.

If H_0 is true, then

$$T := \frac{\bar{X} - \bar{Y}}{\sqrt{S^2(1/n_1 + 1/n_2)}}$$

can be shown to have a T distribution on $(n_1 + n_2 - 2)$ degrees of freedom. Hence, our hypothesis testing procedure is to calculate

$$t := \frac{\bar{x} - \bar{y}}{\sqrt{s^2(1/n_1 + 1/n_2)}}$$

and reject H_0 if $|t| > t_{n_1+n_2-2,1-\alpha/2}$ (Equation 8.11). The p-value is $2P(T > t)$ if $t > 0$ and $2P(T \leq t)$ if $t \leq 0$, where T has a T distribution on $(n_1 + n_2 - 2)$ degrees of freedom and t is the numerical test statistic (Equation 8.12). Also, the $100(1 - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is (Equation 8.13)

$$\bar{x} - \bar{y} \pm t_{n_1+n_2-2,1-\alpha/2} \sqrt{s^2(1/n_1 + 1/n_2)}.$$

Example (small samples and equal variances). Refer to “Pulmonary Disease” on page 157 and {FEV.pdf}. Let μ_1 and σ_1^2 denote the mean and variance of FEV among nonsmoking children and adolescents. Let μ_2 and σ_2^2 denote the mean and variance of FEV among smoking children and adolescents. Consider testing

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{against} \quad H_1 : \mu_1 - \mu_2 \neq 0$$

at level $\alpha = 0.05$ using the method for small samples and equal variances. From page 1 of {FEV.pdf} we have $\bar{x} = 2.5661$, $\bar{y} = 3.2769$, $s_x^2 = 0.8505^2 = 0.7234$, $s_y^2 = 0.7500^2 = 0.5625$, $n_1 = 589$, and $n_2 = 65$. Thus,

$$s^2 = \frac{(589-1)0.7234 + (65-1)0.5625}{589+65-2} = \frac{461.4}{652} = 0.7077$$

and

$$t = \frac{2.5661 - 3.2769}{\sqrt{0.7077(1/589 + 1/65)}} = \frac{-0.7108}{0.1100} = -6.46.$$

Since $t = -6.46 < -1.964 = -t_{652,0.975}$, we easily reject H_0.[3]

Small samples and unequal variances. Suppose that both populations are normal but that $\sigma_1^2 \neq \sigma_2^2$. Let

$$t := \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/n_1 + s_y^2/n_2}}$$

and

$$df := \frac{(s_x^2/n_1 + s_y^2/n_2)^2}{(s_x^2/n_1)^2/(n_1 - 1) + (s_y^2/n_2)^2/(n_2 - 1)}.$$

We may round *df* to the next lower integer if using Table 5 to obtain critical values but may leave *df* unrounded if using SAS to obtain critical values. We reject H_0 if $|t| > t_{df,1-\alpha/2}$ (Equation 8.21). The p-value is $2P(T > t)$ if $t > 0$ and is $2P(T \leq t)$ if $t \leq 0$, where T has a T distribution on *df* degrees of freedom and t is the numerical test statistic (Equation 8.22). Also, the $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is (Equation 8.23)

$$\bar{x} - \bar{y} \pm t_{df,1-\alpha/2} \sqrt{s_x^2/n_1 + s_y^2/n_2}.$$

---

[3] This result seems to imply that smoking children and adolescents have better pulmonary function than nonsmoking children and adolescents. The problem is that age is a confounder: older children are more likely to smoke, but older children also tend to have higher FEV scores. How to “adjust for” such a confounder is addressed in CPH 630/STA 681.

Example (small samples and unequal variances). Continuing from the previous example, now consider testing

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{against} \quad H_1 : \mu_1 - \mu_2 \neq 0$$

at level $\alpha = 0.05$ using the method for small samples and unequal variances. We have

$$df = \frac{(0.8505^2/589 + 0.7500^2/65)^2}{(0.8505^2/589)^2/588 + (0.7500^2/65)^2/64} = \frac{0.00009765}{0.000001173} = 83.25$$

and

$$t = \frac{2.5661 - 3.2769}{\sqrt{0.8505^2/589 + 0.7500^2/65}} = \frac{-0.7108}{0.0994} = -7.15.$$

Since $t = -7.15 < -t_{83.25,0.975} = -1.989$, we easily reject H_0.

Comparing variances (independent samples)

Motivation. Often we will want to test

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{against} \quad H_1 : \sigma_1^2 \neq \sigma_2^2.$$

Even if we are not interested in making inferences about the population variances for scientific reasons, we may want guidance about which of the two small-sample procedures to use for making inferences about the population means. Most people feel that acceptance of $H_0 : \sigma_1^2 = \sigma_2^2$ justifies using the equal-variances small-sample procedure, while rejection of $H_0 : \sigma_1^2 = \sigma_2^2$ necessitates using the unequal-variances small-sample procedure.[4]

---

[4] This scheme for deciding which small-sample procedure to use is interesting because it represents a departure from the way we ordinarily use hypothesis tests. Ordinarily it is rejection of the null hypothesis that allows us to do as we would prefer (i.e., to claim a scientific discovery), but here it is acceptance of the null hypothesis that allows us to do as we would prefer (i.e., to use the equal-variances small-sample procedure). Why would we prefer to use the equal-variances small-sample procedure?

The hypothesis test. Consider testing

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{against} \quad H_1 : \sigma_1^2 \neq \sigma_2^2.$$

Define the test statistic

$$f := \frac{s_x^2}{s_y^2}.$$

If this ratio of sample variances is either very large or very small, then we can reject H_0. Assuming normality, we obtain a level α testing procedure by rejecting H_0 if (Equation 8.15) $f > f_{n_1-1,n_2-1,1-\alpha/2}$ or $f < f_{n_1-1,n_2-1,\alpha/2}$. Moreover, by Equation 8.14, $f_{n_1-1,n_2-1,\alpha/2} = 1/f_{n_2-1,n_1-1,1-\alpha/2}$, a fact that is useful if we are relying on Table 9.

Alternatively, we can define

$$f_{\text{SAS}} := \frac{\text{the larger of } s_x^2 \text{ and } s_y^2}{\text{the smaller of } s_x^2 \text{ and } s_y^2}$$

and then reject H_0 at level α if

$$f_{\text{SAS}} > f_{num-1,den-1,1-\alpha/2},$$

where *num* is the size of the sample whose variance appears in the numerator of f_{SAS} and *den* is the size of the sample whose variance appears in the denominator of f_{SAS}. This is referred to as a "folded f-test" and yields the same conclusion (i.e., decision to accept or reject H_0) that would be obtained

---

- Often *df* for the unequal-variances small-sample procedure is much less than $n_1 + n_2 - 2$, elevating the critical value required for rejection of H_0.
- When we have not just two but three or more samples, generalizing the equal-variances small-sample procedure is easier than generalizing the unequal-variances small-sample procedure. In fact, the generalization of the equal-variances small-sample procedure is the one-way analysis of variance that you will learn about in Lecture 10.
- In statistical modeling we try to adhere to a principle of parsimony. A model with only a few parameters that fits the data well is preferable to a model with many parameters for which the fit is only slightly better. By assuming $\sigma_1^2 = \sigma_2^2$ we reduce the number of parameters describing the two populations from four $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ to three (μ_1, μ_2, σ^2).

from Equation 8.15.

A $100(1-\alpha)\%$ confidence interval for the ratio of population variances σ_1^2/σ_2^2 is

$$\frac{s_x^2}{s_y^2}\,\frac{1}{f_{n_1-1,n_2-1,1-\alpha/2}} \quad \text{to} \quad \frac{s_x^2}{s_y^2}\,f_{n_2-1,n_1-1,1-\alpha/2}.$$

Example (the hypothesis test). We continue from the previous example. Assuming normality, let us test

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{against} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

at level $\alpha = 0.05$. We have

$$f_{\mathrm{SAS}} = \frac{0.8505^2}{0.7500^2} = 1.29.$$

Since 1.29 is not greater than $f_{588,64,0.975} = 1.484$, we do not reject H_0.