Contents

Increasing Power

Paired samples can increase the power of the test because, under some circumstances, the $\bar{D}$ estimator is a more efficient estimator of $\mu_x - \mu_y$ than $\bar{X}_I - \bar{Y}_I$, where $\bar{X}_I$ and $\bar{Y}_I$ are sample means from independent samples. To conclude this, we must show $\text{Var}(\bar{D}) \leq \text{Var}(\bar{X}_I - \bar{Y}_I)$. We know that $\text{Var}(\bar{X}_I - \bar{Y}_I) = \frac{\sigma_x^2 + \sigma_y^2}{n}$, when both samples are of the same size. On the other hand,

$$\text{Var}(\bar{D}) = \text{Var}(\bar{X}_i - \bar{Y}_i) = \frac{\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}}{n}$$

Therefore, it is clear that $\bar{D}$ is more efficient when $\sigma_{xy} > 0$. Since the pairs often exist because they are measurements on the same or two different experimental units that share many characteristics in common, including characteristics that could make their random variables more similar than random individuals, it is very often the case that paired samples lead to more efficient estimation.

Example: smoking and blood clots

Several decades ago there was a suspicion that smoking affected platelet aggregation, a normal process important for blood clotting. To test the effect, blood samples were taken from smokers before and after they smoked a cigarette. The resulting differences in a clotting measurement after minus before were

$$2, 4, 10, 12, 16, 15, 4, 27, 9, -1, 15$$

The sample mean of these $n = 11$ experimental units (smokers measured twice) is $\bar{D} = 10.27$. The sample standard deviation yields $S_{\bar{D}} = 2.40$. The statistic

$$T = \frac{10.27}{2.40} = 4.28$$

which exceeds $t_{10}(0.975) = 2.23$. We reject the null hypothesis that $H_0 : D = 0$, where D is the difference in population means, where the populations are before and after smoking. The p-value is `R> 2*pt(-4.28, df=10)`, which yields 0.0016, highly significant also.

This data is so obvious, we can detect a difference with far fewer assumptions. How likely is it to observe 9 positive values? If the null hypothesis is correct, then we expect $1/2$ of the individuals to have an increase in clotting and the other half a decrease (i.e. the median, and mean for normal data, of differences is 0). Let X be the number of positive values, then we seek the probability of seeing outcomes as rare or rarer than that. Under the null, the $X \sim \text{Binomial}(10, 0.5)$ has a symmetric distribution, so the p-value is `R> 2*pbinom(1, size=10, prob=0.5)`, with value 0.02148438. We have lost some power (the p-value is larger) as compared to the t-test, not surprising since we have thrown out everything but the sign of the data.

3.5.5 Nonparametric Two Sample Tests

Nonparametric Rank Sum Test

[See handout.]

Nonparametric Signed Rank Test

[See handout.]

Data Analysis Lessons

We have finished tests on population means for one or two samples. We now spend some time learning some general lessons about data analysis.

- **Placebo effect.** Most often the placebo effect is associated with human trials of medications, but it is more general. It is the idea that the process of going through a study, without actually receiving a "treatment" can result in a change from baseline. Therefore, it is essential that the test is between a "treatment" and a "control," rather than a "treatment" and an assumed baseline of 0 effect. When a treatment and control are included, the study becomes a two-sample setup.
- **Selection bias.** It is essential for individuals to be selected randomly from populations, or the treatments assigned randomly to individuals. Otherwise, there is a high risk that some bias will enter in the selection process. Imagine doctors selecting patients to assign treatments. As much as the doctors try to be unbiased, it is difficult to avoid hidden bias.
- **Blinding.** Ideally, both the subject and operator should be blinded so they do not know the treatment (or population) of the experimental units. Again, it is hard to get objective, impartial data if there is a predisposition to believe, expect, or want a certain outcome.
- **Confounded variables.** Sometimes one or more features of the experimental units is coincident with the treatment (or population). This is particularly problematic for observational studies, where randomization cannot be applied. For example, suppose that all the women applying for admission to school, by their own choice (self-selection), apply for majors with low acceptance rates. Even with no preference to reject women, it will appear that they are rejected at a higher rate. The erroneous conclusion that there is bias against women would be made unless the data were carefully considered for *confounding variables*, in this case major confounds sex.
- **Fishing expedition.** Observational studies are also often subject to fishing expeditions. In these cases, a lot of data is collected and each variable measured is tested for significant population means. If each of n tests is carried out controlling type I error at α, then the probability that at least one null hypothesis is rejected when all null hypotheses are correct (i.e. nothing interesting in the data) is

  $$\alpha' = P(\text{reject at least one } H_0) = 1 - P(\text{no } H_0 \text{ rejected}) = 1 - (1-\alpha)^n$$

  In the last step, we assume each of the tests are independent. For $\alpha = 0.05$, $n = 100$, $\alpha' = 0.994$, so almost all studies will result in some apparently significant result, even when there are no differences. The possible solutions to this problem are:

  - Treat the study as a pilot study that identifies possibly interesting measurements and is followed up by more focused, in depth studies.
  - Perform the 100 tests on 1/2 of the data, and test the predictions in the second 1/2.
  - Reduce α by choosing a desired α' and solving for α. In general, assuming independent tests, α/n is the new type I error rate to use.

3.6 Categorical Data

We now turn away from continuous data and focus on counts of outcomes in categories: categorical data.

Fisher's Exact Test

Suppose observations can be classified in two ways in two dimensions, into class C1 or C2 and also into R1 or R2. For example, you observe n indivuals and classify each one as male or female and BCB major or not. A summary of data like this can be presented as:

	C1	C2	
R1	N_{11}	N_{12}	$n_{1\cdot}$
R2	N_{21}	N_{22}	$n_{2\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

Null Hypothesis. The null hypothesis of the Fisher's exact test is that the row and column sums are fixed and that row and column categories are not associated, i.e. they are independent. For our example, the null hypothesis is that BCB major and sex of student are independent, i.e. BCB students don't tend to be male (or female).

$$H_0 : n_{1\cdot}, n_{2\cdot}, n_{\cdot 1}, n_{\cdot 2}, \text{rows and columns independent}$$

Because the row and column sums are fixed, there is one degree of freedom among the four random variables $N_{11}, N_{12}, N_{21}, N_{22}$. Once one is given, the others are determined.

Now, the idea is to compute how extreme the particular configuration observed is if the row and column features truly are independent. We answered this kind of question when computing probabilities. Suppose we know there are $n_{1\cdot}$ units of type R1, and $n_{2\cdot}$ of type R2. Now, we select $n_{\cdot 1}$ C1 labels to assign to all $n_{\cdot\cdot}$ units. What is the probability that we assign N_{11} C1 labels to the $n_{1\cdot}$ R1 types and the remaining N_{21} C1 labels the the remaining $n_{2\cdot}$ R2 types? Since all possible assignments are equally likely under the null hypothesis (no tendency to assign C1 to R1 rather than R2), then we can compute this probability by counting. Namely,

$$P(N_{11} = n_{11} \mid n_{\cdot 1}, n_{\cdot 2}, n_{1\cdot}, n_{2\cdot}) = \frac{\binom{n_{1\cdot}}{n_{11}}\binom{n_{2\cdot}}{n_{21}}}{\binom{n_{\cdot\cdot}}{n_{\cdot 1}}}$$

Example. To give a concrete example, return to the BCB, male and female students. If we observe $n_{\cdot\cdot} = 20$ students, of which $n_{1\cdot} = 10$ are BCB majors and $n_{\cdot 1} = 8$ are female, what is the probability of assigning sex to majors such that only $n_{11} = 2$ BCB students are female, as shown in the following table.

	Female	Male	
BCB	2	8	10
Other	6	4	10
	8	12	20

The number of ways to assign 8 females to the 20 students without regard to major is $\binom{20}{8}$. The number of ways to assign 8 females such that 2 of 10 BCB students are female and consequently 6 of the 10 other majors are female is $\binom{10}{2}\binom{10}{6}$. Overall, the probability is

$$P(N_{11} = 2 \mid 10, 10, 8, 12) = \frac{\binom{10}{2}\binom{10}{6}}{\binom{20}{8}} = 0.07501786$$

This is already quite likely (above $\alpha = 0.05$ without considering more extreme outcomes), and we can stop and accept the null hypothesis of independence.

p-value. In general, we compute the p-value by finding all outcomes that are as extreme or more extreme (as measured by their smaller probability) and sum up all their probabilities. These choices can be compiled in a table, for example as shown for the BCB example below. Notice that because row and columns sums are fixed, it is possible that some outcomes of N_{11} are not possible, for example, there cannot be mre than 8 female BCB students.

N_{11}	Probability
0	$\frac{\binom{10}{0}\binom{10}{8}}{\binom{20}{8}} = 0.0003572279$*
1	$\frac{\binom{10}{1}\binom{10}{7}}{\binom{20}{8}} = 0.009526078$*
2	$\frac{\binom{10}{2}\binom{10}{6}}{\binom{20}{8}} = 0.07501786$*
3	$\frac{\binom{10}{3}\binom{10}{5}}{\binom{20}{8}} = 0.2400572$
4	$\frac{\binom{10}{4}\binom{10}{4}}{\binom{20}{8}} = 0.3500834$
5	$\frac{\binom{10}{5}\binom{10}{3}}{\binom{20}{8}} = 0.2400572$
6	$\frac{\binom{10}{6}\binom{10}{2}}{\binom{20}{8}} = 0.07501786$*
7	$\frac{\binom{10}{7}\binom{10}{1}}{\binom{20}{8}} = 0.009526078$*
8	$\frac{\binom{10}{8}\binom{10}{0}}{\binom{20}{8}} = 0.0003572279$*

The starred* probabilities indicate configurations that are as extreme or more extreme than the observed $N_{11} = 2$, and their combined probability is the p-value:

$$2 * 0.0003572279 + 2 * 0.009526078 + 2 * 0.07501786 = 0.1698023$$