4/~~20~~21/09 Math 152

Paired Samples

e.g. • subjects matched by age and weight or severity of condition then one member of each pair randomly assigned to treatment or control.

• littermates divided into treatment and control randomly.

• before, after measurements on same subject

The purpose of such an experimental design is to reduce variance.

e.g. In a paired design let the pairs be (X_i, Y_i) $i = 1, \ldots, n$

assume $\text{mean}(X_i) = \mu_X$
$\text{mean}(Y_i) = \mu_Y$ $\forall i.$

and $Var(X_i) = \sigma_X^2$
$Var(Y_i) = \sigma_Y^2 \quad \forall i.$

~~Ass~~ Different pairs are independent but.
assume $Cov(X_i, Y_i) = \sigma_{XY}$

Then for $D_i = X_i - Y_i$ we have.

$$E(D_i) = \mu_X - \mu_Y$$

$$Var(D_i) = Cov(X_i - Y_i, X_i - Y_i)$$

$$= \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$$

$$= \sigma_X^2 + \sigma_Y^2 - 2\rho\,\sigma_X\sigma_Y$$

where $\rho = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$ is the correlation coefficient of (X, Y).

Using $\bar{D} = \bar{X} - \bar{Y}$ to estimate $\mu_X - \mu_Y$, we have.

$$E(\bar{D}) = \mu_X - \mu_Y$$

$$Var(\bar{D}) = \frac{1}{n}(\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)$$

In an unpaired design, we have.
(with sample size $= n$ for both groups)

$$E(\bar{X} - \bar{Y}) = \mu_X - \mu_Y$$

$$Var(\bar{X} - \bar{Y}) = \frac{1}{n}(\sigma_X^2 + \sigma_Y^2).$$

If $\rho > 0$ then $Var(\bar{D}) < Var(\bar{X} - \bar{Y})$.

If $\sigma_X = \sigma_Y$ we have.

$$\frac{Var(\bar{D})}{Var(\bar{X} - \bar{Y})} = 1 - \rho.$$

e.g. with $\rho = .5$.

$$Var(\bar{D}) = \frac{1}{2} Var(\bar{X} - \bar{Y}).$$

$$\frac{1}{n}(\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y) = \frac{1}{2n}(\sigma_x^2 + \sigma_y^2).$$

So twice as many subjects are required in the unpaired experiment to get the same precision.

A non-parametric method for the paired samples comparison

The Signed Rank Test

e.g.

Before	After	Diff	|Diff|	Rank	Signed Rank
25	27	+2	2	2	2
29	25	−4	4	3	−3
60	59	−1	1	1	−1
27	37	+10	10	4	4

$$W_+ = 2 + 4 = 6.$$

1. find differences D_i and also $|D_i|$
rank the $|D_i|$.

2. restore signs of differences to the ranks.

3. find W_+ = sum of positive ranks.

If there is no difference in groups, we expect $\approx$ half the D_i to be positive

so W_+ **should** not be too large or too small.

H_0: the distribution of the D_i is symmetric about zero.

Under the null hypothesis, any possible assignment of signs to the ranks is equally likely.

There are 2^n possible assignments each with prob. $\frac{1}{2^n}$.

Corresponding to each is a sum W^+, and so we can determine the null distribution of W^+.

If $n \geq 20$, the null distribution is approximately normal and we have

<u>Thm:</u>

$$E(W^+) = \frac{n(n+1)}{4}$$

$$Var(W^+) = \frac{n(n+1)(2n+1)}{24}$$

<u>Pf:</u>

$$W_+ = \sum_{k=1}^{n} k I_k$$

$$I_k = \begin{cases} 1 & k^{th} \text{ largest } |D_i| \text{ has } D_i > 0 \\ 0 & \text{else.} \end{cases}$$

Under H_0.

$$E(I_k) = \frac{1}{2}$$

$$Var(I_k) = 1/4$$

(Bernoulli with $p = 1/2$.

So
$$E(W^+) = \frac{1}{2}\sum_{k=1}^{n} k = \frac{n(n+1)}{4}$$

$$Var(W^+) = \frac{1}{4}\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{24}.$$

Too many ties can throw off the test.
