University of Washington

STATISTICS

# Stat 425

# Introduction to Nonparametric Statistics

# Paired Comparisons in a Population Model and the One-Sample Problem

Fritz Scholz

Spring Quarter 2009*

*June 2, 2009

# Paired Comparisons (Randomization Model)

We have looked at paired comparison procedures in a randomization context.

Treatment and control are randomly assigned to $N$ paired subjects.

Advantage: Conceptual simplicity. We fully control all random aspects.

All inferences are based on solid principles.

Disadvantage: Any drawn conclusion drawn are only pertinent to the $N$ subjects.

We can't generalize any results to other subjects without a leap of faith.

To generalize results we need a population model and a sampling scheme.

This issue is analogous to the parallels examined in Chapters 1 and 2.

# Population Model

Consider a population of homogeneous pairs of subjects, each providing a measurement. Pairs are randomly drawn from a population.

Treatment/control should be randomized for each pair.

Pairs of ears/hands/feet in a large population of subjects. Draw a random sample of subjects from a population. Randomize treatment/control between left/right.

The same subject is used for treatment and control. Draw a random sample of subjects from a population. Randomize the order of treatment and control for each subject in the sample.

Pairs could be measurements in the morning and afternoon of a day.

The days are randomly drawn from a population of days (limited scope).

# Population Distribution Model

We have $N$ randomly chosen pairs, with randomly assigned treatment and control.

Denote responses for the $i^{\text{th}}$ pair by $(X_i, Y_i)$ under (control,treatment), respectively.

Since each subject was drawn randomly from one population all the pairs $(X_i, Y_i)$ have the same bivariate distribution, say

$$P(X_i \le x, Y_i \le y) = M(x, y)$$

In contrast to the independence for two independently drawn samples from two populations we allow here for dependence between $X_i$ and $Y_i$ within the same pair.

This is a natural property for matched pairs.

We focus on differences $Z_i = Y_i - X_i$, again identically distributed with cdf

$$P(Z_i \le z) = L(z)$$

# Effect of Null Hypothesis on $M(x,y)$

Null hypothesis $H_0$: there is no difference between treatment and control.

$$\begin{aligned} P(\text{control response} \le y, \text{treatment response} \le x) &= P(X_i \le y, Y_i \le x) = M(y,x) \\ = P(\text{control response} \le x, \text{treatment response} \le y) &= P(X_i \le x, Y_i \le y) = M(x,y) \end{aligned}$$

If two measurements $(U_1, U_2)$ on a pair of subjects have a joint distribution

$H(x,y) = P(U_1 \le x, U_2 \le y)$ which is not symmetric, i.e., $H(x,y) \neq H(y,x)$

(e.g., when taking two measurements on the same subjects or when left and right side of the human body act differently) we can ensure symmetry by assigning treatment and control at random.

A formal argument is given on the next slide.

# Formal Symmetry Argument under $H_0$

Treatment/control are simply labels assigned at random, with no differential effect.

Let $U_1$ and $U_2$ be the $1^{\text{st}}$ and $2^{\text{nd}}$ (left and right) measurement.

Their joint distribution is $H(x,y) = P(U_1 \le x, U_2 \le y) \overset{??}{=} H(y,x)$

We can view $U_i$ as the response under control or treatment (no difference).

$$A = \begin{cases} 1 & \text{when the treatment is assigned to the first measurement} \quad \text{probability } 1/2 \\ 2 & \text{when the treatment is assigned to the second measurement} \quad \text{probability } 1/2 \end{cases}$$

Here $A$ and $(U_1, U_2)$ are independent.

$$\begin{aligned} M(x,y) = P(X \le x, Y \le y) &= P(X \le x, Y \le y, A = 1) + P(X \le x, Y \le y, A = 2) \\ &= P(U_2 \le x, U_1 \le y, A = 1) + P(U_1 \le x, U_2 \le y, A = 2) \\ &= [P(U_2 \le x, U_1 \le y) + P(U_1 \le x, U_2 \le y)]/2 \\ &= [H(y,x) + H(x,y)]/2 = [H(x,y) + H(y,x)]/2 = M(y,x) \end{aligned}$$

# Distribution of $Z = Y - X$ under $H_0$

Under $H_0$ the distribution of $Z = Y - X$ is symmetric around zero,

i.e., $Z$ and $-Z$ have the same distribution function $L(z)$.

$$\begin{aligned}
P_{H_0}(Z \le z) &= P_{H_0}(Y - X \le z) = P_{H_0}(Y - X \le z,\, A = 1) + P_{H_0}(Y - X \le z,\, A = 2) \\
&= P_{H_0}(U_1 - U_2 \le z,\, A = 1) + P_{H_0}(U_2 - U_1 \le z,\, A = 2) \\
&= \left[P_{H_0}(U_1 - U_2 \le z) + P_{H_0}(U_2 - U_1 \le z)\right]/2
\end{aligned}$$

$$\begin{aligned}
P_{H_0}(-Z \le z) &= P_{H_0}(X - Y \le z) = P_{H_0}(X - Y \le z,\, A = 1) + P_{H_0}(Y - X \le z,\, A = 2) \\
&= P_{H_0}(U_2 - U_1 \le z,\, A = 1) + P_{H_0}(U_1 - U_2 \le z,\, A = 2) \\
&= \left[P_{H_0}(U_2 - U_1 \le z) + P_{H_0}(U_1 - U_2 \le z)\right]/2 = P_{H_0}(Z \le z) \quad \square
\end{aligned}$$

The text gives the further equivalence $L(z) = 1 - L(-z)$, but that only holds under the additional condition that $L$ is continuous, i.e., $P_{H_0}(Z = z) = 0$ for any $z$.

# Alternatives to $H_0$

As alternative to $H_0$ we will consider distributions for $Z$ that tend to favor positive over negative values for $Z$.

Problem 44 makes this more precise, but assume continuous distributions.

A special case is the shift model.

As in the two-sample problem we assume that under the alternative the treatment adds a shift to the response $Y$.

Let $L_0{}^*$ denote the null distribution of $Y - X$ and assume a treatment effect of $\Delta$.
Then $Z - \Delta = Y - \Delta - X$ has again the null distribution $L_0$. Thus

$$P(Z \leq z) = P(Z - \Delta \leq z - \Delta) = L_0(z - \Delta) = L_\Delta(z)^\dagger ,$$

i.e., the null $L_0$ distribution of $Z$ is shifted to the right by the amount $\Delta$.

$^*$The Text uses $E(z)$ instead of $L_0(z)$. This avoids the expectation connotation of $E(z)$.

$^\dagger$Note that $L_\Delta(z) = L_0(z)$ when $\Delta = 0$. This harmonizes with the choice of $L_0$.

# The Large Population Model

As in the two-sample population case we assume that the matched pairs of subjects are randomly drawn from a large population.

This allows us to view the response pairs $(X_i, Y_i),\ i = 1, \ldots, N$ as independent (or at least approximately independent) random variable pairs.

The dependence between $X_i$ and $Y_i$ within each pair remains.

Thus $Z_1, \ldots, Z_N$ are independent, identically distributed random variables $\sim L(z)$.

The hypothesis to test: $L$ is symmetric around zero ($P_{H_0}(Z \leq z) = P_{H_0}(Z \geq -z)$).

As alternative we consider that positive values for $Z$ are favored over negative ones.

For the more specific shift model we test $H_0 : \Delta = 0$ against $A : \Delta > 0$, with the symmetric null cdf $L_0(z)$ left unspecified.

# The One-Sample Problem

So far we have viewed the $Z_i$ as the differences between treatment and control responses on a sampled matched subject pair.

However, we can also view these differences simply a priori as $N$ independent measurements $Z'_1, \ldots, Z'_N$ of some unknown quantity $\Delta$.

Often it is reasonable to view the measurement errors $Z'_i - \Delta$ as being symmetrically distributed around zero, with cdf $L_0$.

If we want to test a hypothesized value $\Delta_0$ for $\Delta$ (based on some theory), we can reduce this testing problem to our previous problem of testing $H_0 : \Delta = 0$.

Simply subtract $\Delta_0$ from all measurements: obtain $Z_i = Z'_i - \Delta_0, \ i = 1, \ldots, N$ and test whether the $Z_i$ distribution is symmetric around zero.

This is called the one-sample problem, in contrast to our two-sample problem considered in Chapter 2.

# Advantage of the Population Model

The advantage of the population model is again two-fold.

1. We can make inference about the full population (based on a sample from it).

2. We can assess the power of the test and plan the sample size $N$ appropriately.

This depends however on obtaining a true random sample, or a reasonable judgment that we can view the $Z_i (= Y_i - X_i)$ as a random sample.

These issues are completely analogous to those discussed in Chapter 2.

# The Sign Test

Recall that the sign test is based on the number $S_N$ of positive differences

$Z_i = Y_i - X_i$.

Under the randomization model we saw that the null distribution of $S_N$

is *binomial* $(N, .5)$, provided none of the $Z_i = 0$.

In the population model assume that $L(z)$ is continuous to avoid zero differences.

Under $H_0$ the distribution of $Z_i$ is symmetric around zero

$$\Longrightarrow \qquad P_{H_0}(Z_i > 0) = P_{H_0}(Z_i < 0) = {}^1\!/_2$$

Since the $Z_i$ are independent it follows that $S_N \sim$ *binomial* $(N, .5)$ under $H_0$.

We have the same null distribution as before in the randomization model.

# Distribution of $S_N$ under the Alternative

$Z_1,\ldots,Z_N$ are still independent but now with "success" probability

$$p = P_A(Z_i > 0) = 1 - P_A(Z_i \le 0) = 1 - q = 1 - L(0).$$

The power function $\Pi(p)$ of the sign test, which rejects for $S_N \ge c$ (for integer $c$),

$$\Pi(p) = P_p(S_N \ge c) = \sum_{k=c}^{N} \binom{N}{k} p^k (1-p)^{N-k} = 1 - \mathtt{pbinom}(\mathtt{c}-\mathtt{1},\mathtt{N},\mathtt{p})$$

depends only on $p$. Aside from $p$ the shape of $L$ does not enter.

It is easy to see that $\Pi(p)$ is strictly increasing in $p$.

$\Rightarrow$ any level $\alpha$ test with $P_{H_0}(S_N \ge c) = \alpha_c = \alpha$ is unbiased ($\Pi(p) \ge \alpha_c$ for $p > {}^1\!/_2$).
It is also unbiased for testing the wider hypothesis $H_0' : p \le {}^1\!/_2$ against $A : p > {}^1\!/_2$, since $\max\{P_p(S_N \ge c) : p \le {}^1\!/_2\} = \alpha_c$, i.e., it is still level $\alpha_c$ for testing $H_0'$.

# Normal Approximation for $\Pi(p)$

Although `pbinom` is readily available, for sample size determination it is still useful to introduce the normal approximation for $\Pi(p)$.

For binomial random variables we have $E_p(S_N) = Np$ and $\mathrm{var}_p(S_N) = Npq$.

The ordinary central limit theorem shows that for $0 < p < 1$

$$\frac{S_N - Np}{\sqrt{Npq}} \longrightarrow \mathcal{N}(0,1) \quad \text{as} \quad N \longrightarrow \infty$$

Thus we can treat $(S_N - Np)/\sqrt{Npq} \approx \mathcal{N}(0,1)$, provided the practical range of the former contains the practical range of the latter. (See next slide)

# Practical Range of Normal Approximation

$$
\begin{array}{ccccccr}
0 & \le & S_N & \le & N & & \text{absolute range} \\
Np - 3\sqrt{Npq} & \le & S_N & \le & Np + 3\sqrt{Npq} & & \text{approximate range} \\
-3 & \le & \frac{S_N - Np}{\sqrt{Npq}} & \le & 3 & & \text{standardized range} \\
-3 & \le & \mathcal{N}(0,1) & \le & 3 & & \text{normal range}
\end{array}
$$

Thus we should require

$$0 \le Np - 3\sqrt{Npq} \quad \text{and} \quad Np + 3\sqrt{Npq} \le N \iff N \ge 9 \max\left(\frac{p}{q}, \frac{q}{p}\right)$$

$p$ close to $0$ or $1$ will drive up the required $N$ for a reasonable approximation.

E.g., $p = .1 \;\Rightarrow\; N \ge 9(.9/.1) = 81$ and $p = .01 \;\Rightarrow\; N \ge 9(.99/.01) = 891$.

# Normal Approximation

Without continuity correction

$$\Pi(p) = P_p(S_N \geq c) = P_p\left(\frac{S_N - Np}{\sqrt{Npq}} \geq \frac{c - Np}{\sqrt{Npq}}\right) \approx 1 - \Phi\left(\frac{c - Np}{\sqrt{Npq}}\right)$$

With continuity correction

$$\Pi(p) = P_p(S_N \geq c) = P_p(S_N \geq c - .5) \approx 1 - \Phi\left(\frac{c - .5 - Np}{\sqrt{Npq}}\right)$$

# Example 1: Gas Mileage

A consumer organization wants to examine a dealer's claim of 22 miles/gallon in city driving for a particular car model.

They plan to test $N = 20$ cars and will reject the claim if too many of the cars have mileage $< 22$ mpg.

How likely are they going to reject the claim when in fact the mileage is $21$ mpg?

That will depend on the chosen $\alpha$ level. A high $\alpha$ will lead to higher power, higher rejection rate even under $H_0$ and even more so under $A$.

Let us work with a significance level close to $.05$.

$$1 - \texttt{pbinom}(14, 20, .5) = 0.020695 \quad \text{and} \quad 1 - \texttt{pbinom}(13, 20, .5) = 0.057659$$

$$\Longrightarrow \quad c = 14 \quad \text{with achieved} \quad \alpha = \alpha_c = 0.05765915\ .$$

# Gas Mileage: Power Calculation

In order to calculate power for some $p > {}^1\!/_2$ we need to have a value for $p$.

It is reasonable to assume that the average mileage per car (under like conditions) is approximately distributed like $\mathcal{N}(\mu, \sigma^2)$.

If $Z$ denotes the average mileage of a car we have

$$p = p_\mu = P(Z < 22) = P((Z-\mu)/\sigma < (22-\mu)/\sigma) = \Phi((22-\mu)/\sigma)$$

To evaluate $p$ for $\mu = 21$, the situation of interest for our power calculation, we need an appropriate value for $\sigma$. From past experience we assume that $\sigma = 1.5$ mpg.

$$p_\mu = p_{21} = \Phi((22-21)/1.5) = \Phi(1/1.5) = \Phi(2/3) = \texttt{pnorm}(2/3) = 0.74751$$

with power $\quad \Pi = P_p(S_{20} \geq 14) \approx 1 - \Phi\left(\dfrac{13.5 - 20 \cdot p_{21}}{\sqrt{20 p_{21} q_{21}}}\right) = \Phi(0.74639) = 0.77228$

or exactly $\qquad 1 - \texttt{pbinom}(13, 20, 0.7475075) = 0.7778726$

# Gas Mileage: Increasing Power

The power achieved by $N = 20$ is judged not to be adequate. We desire $\Pi = .9$.

How much higher do we need to choose $N$ to get there?

$$P_{H_0}(S_N \ge c) \approx 1 - \Phi\left(\frac{c - \frac{1}{2} - \frac{N}{2}}{\sqrt{\frac{N}{4}}}\right) = \alpha \implies c = u_\alpha\sqrt{\frac{N}{4}} + \frac{N}{2} + \frac{1}{2}$$

$$P_p(S_N \ge c) \approx 1 - \Phi\left(\frac{c - \frac{1}{2} - Np}{\sqrt{Npq}}\right) = \Pi$$

$$\implies u_\alpha\sqrt{\frac{N}{4}} + \frac{N}{2} + \frac{1}{2} - \frac{1}{2} - Np = u_\Pi\sqrt{Npq} \qquad \text{or} \qquad \frac{u_\alpha}{2} - u_\Pi\sqrt{pq} = \left(p - \frac{1}{2}\right)\sqrt{N}$$

$$\sqrt{N} = \frac{\frac{1}{2}u_\alpha - u_\Pi\sqrt{pq}}{p - \frac{1}{2}} \qquad \text{or} \qquad N = \left(\frac{\frac{1}{2}u_\alpha - u_\Pi\sqrt{pq}}{p - \frac{1}{2}}\right)^2 \Rightarrow \quad N = 34.07263$$

`1-pbinom(22,35,.5)=0.0448` and `1-pbinom(22,35,0.74751)=0.9192` will do.

# Alternate Power Approximation

Under the shift model $L(z) = L_0(z - \Delta)$ we get an alternate power approximation for small $\Delta$ and large $N$. For small $\Delta$ we have

$$p_\Delta = 1 - L(0) = 1 - L_0(-\Delta) \approx \frac{1}{2} \quad \text{and} \quad p_\Delta - \frac{1}{2} \approx \Delta \ell_0(0) \quad \text{with} \quad \ell_0(z) = \frac{\partial L_0(z)}{\partial z}.$$

$$P_{p_\Delta}(S_N \geq c_N(\alpha)) \approx 1 - \Phi\left(\frac{c_N(\alpha) - \frac{1}{2} - Np_\Delta}{\sqrt{Np_\Delta q_\Delta}}\right) \quad \text{and} \quad c_N(\alpha) = u_\alpha \sqrt{\frac{N}{4}} + \frac{N}{2} + \frac{1}{2}$$

combine to

$$\begin{aligned} P_{p_\Delta}(S_N \geq c) &\approx 1 - \Phi\left(\frac{u_\alpha \sqrt{\frac{N}{4}} + \frac{N}{2} + \frac{1}{2} - \frac{1}{2} - Np_\Delta}{\sqrt{Np_\Delta q_\Delta}}\right) \\ &\approx \Phi\left(2\sqrt{N}\left(p_\Delta - \frac{1}{2}\right) - u_\alpha\right) \approx \Phi\left(2\sqrt{N}\Delta \ell_0(0) - u_\alpha\right) \end{aligned}$$

# Population Model: $1^{\mathrm{st}}$ Extension

We assumed that $Z_1, \ldots, Z_N$ are independent and identically distributed (i.i.d.)

Suppose we wish to perform paired comparisons of treatment and control on samples drawn from stratified subpopulations.

For subpopulation $i$ we get $X_i$ and $Y_i$ with joint distribution $M_i(x,y) = M_i(y,x)$ and obtain $Z_i = Y_i - X_i \sim L_i$, which is symmetric around zero under $H_0$, provided we assign treatment and control randomly.

These subpopulations may not be large, but we sample independently from each.

Thus we have $Z_1, \ldots, Z_N$ independent and with respective distributions $L_1, \ldots, L_N$, which are symmetric around zero under $H_0$ (no treatment effect).

Under the alternative we expect $Z$ values that favor positive values.

# Population Model: $1^{st}$ Extension (continued)

A special case of such alternatives is that of a shift alternative, i.e., the treatment adds a shift $\Delta$, so that $Z_i - \Delta \sim L_{i0}$ is distributed symmetrically around zero, or $Z_i \sim P(Z_i \leq z) = P(Z_i - \Delta \leq z - \Delta) = L_{i,0}(z-\Delta) = L_{i,\Delta}(z)$.

This model is also appropriate when the same unknown quantity $\Delta$ is being measured under different circumstances and the measurement errors $Z_i - \Delta$ for some reason have different distributions $L_{i,0}$.

$L_{i,0}$ continuous $\Longrightarrow$ $P_{H_0}(Z_i > 0) = {}^1\!/_2$ for $i = 1, \ldots, N$ $\Longrightarrow$ $S_N \sim$ *binomial* $(N, {}^1\!/_2)$.
We have the same null distribution distribution as before.
We would view large values of $S_N$ as significant evidence against $H_0$.

# Population Model: $2^{\text{nd}}$ Extension

We assumed: $Z_i$ is symmetrically distributed around some value $\Delta$ with $H_0 : \Delta = 0$.

We now drop the symmetry assumption.

We still want to use a meaningful concept for the “center” of a distribution, without specifying the distribution in further detail, i.e., stay nonparametric.

The mean of a distribution is not a good measure of central tendency without further restrictions on the distribution $L$.

Very small changes in $L$ can lead to arbitrarily different means.

A more suitable measure of central tendency is the median $\mu$.

For $Z - \mu \sim L_0$, which is continuous with median zero, we have

$$P(Z > \mu) = P(Z < \mu) = P(Z \leq \mu) = P(Z - \mu \leq 0) = L_0(0) = {}^1\!/_2$$

# Population Model: $2^{\text{nd}}$ Extension (continued)

For some specified value $\mu_0$ consider now the problem of testing the hypothesis

$H_0'' : \mu = \mu_0$ against the alternative $A : \mu > \mu_0$.

Assume that we have $N$ i.i.d. measurements $Z_1, \ldots, Z_n \sim L$ with median $\mu$.

Under $H_0$ the differences $Z_1 - \mu_0, \ldots, Z_N - \mu_0$ are i.i.d. $\sim L_0$ with median zero.

Let $S_N$ be the number of $Z_i - \mu_0 > 0$ (or $Z_i > \mu_0$).

Under $H_0$ we have $S_N \sim$ *binomial* $(N, {}^1\!/_2)$, i.e., the same null distribution as before.

We reject $H_0$ in favor of $A$ when $S_N \geq c$ for appropriate $c$ with $P_{H_0}(S_N \geq c) = \alpha_c$.

# Hearing Test Example

In a hearing test patients are exposed to a sound at increasing pitch levels.

For patient $i$ the level $Z_i$ is determined (by iterative up and down approximation) at which the patient no longer hears the sound.

There is standard loss of hearing pitch level $\mu_0$, which is known.
$\mu_0$ represents the median of a normal or healthy population.

We wish to test the hypothesis $\mu = \mu_0$ (or $\mu \geq \mu_0$) against the alternative $\mu < \mu_0$.

We reject the hypothesis when too many of the $Z_i$ are $< \mu_0$.

$\Longrightarrow$ The sign test can be applied to alternatives in either direction.

# The Signed-Rank Test in the Population Model

The null distribution of the signed-rank Wilcoxon test in the population model is the same as in the randomization model.

**Theorem 1:**

Let $Z_1, \ldots, Z_N$ i.i.d. $\sim L$, which is assumed to be continuous.

Let $S_1, \ldots, S_n$ denote the ranks of the positive $Z$'s among $|Z_1|, \ldots, |Z_N|$.

Let $N_+$ be the number of positive $Z$'s.

Under the hypothesis $H_0 : L$ is symmetric around zero ($L(z) = 1 - L(-z)$ for all $z$), each of the possible sets $(n, s_1, \ldots, s_n)$ has the same probability

$$P_{H_0}(N_+ = n; S_1 = s_1, \ldots, S_n = s_n) = \left(1/2\right)^N$$

# Proof of Theorem 1

For each $Z_i$ its sign and absolute value are independent since

$$P(Z_i > 0, |Z_i| \le z) = P(0 < Z_i \le z) = {}^1\!/_2\, P(|Z_i| \le z) = P(Z_i > 0)\, P(|Z_i| \le z)$$

Since the $Z_i$ are all independent it follows that their set of signs and their set of absolute values are independent. The signs are independent among each other.

$\Longrightarrow$ the set of $N$ signs is independent of the set of ranks $1, \ldots, N$ for $|Z_1|, \ldots, |Z_N|$.

Each set of $\{N_+ = n, S_1 = s_1, \ldots, S_n = s_n\}$ with $s_1 < \ldots < s_n$ is in one-to-one correspondence with exactly one of the $2^N$ possible assignment of $+$ and $-$ signs to the ranks $1, 2, \ldots, N$.

Each assignment of a $+$ or $-$ sign has probability ${}^1\!/_2$. The $N$ sign assignments are independent. Thus the probability of any set of sign assignments is $({}^1\!/_2)^N$. $\square$

# Corollary and Consequences for $V_s$

Corollary to Theorem 1:

Theorem 1 still holds under our first extended model, namely for independent $Z_i \sim L_i,\ i = 1, \ldots, N$, with $L_i$ continuous and symmetric around zero under $H_0$.

While the ranks of the absolute values $|Z_1|, \ldots, |Z_N|$ may be affected by $L_1, \ldots, L_N$, the assigments of signs to the full set of ranks still have probability $(^1\!/_2)^N$ each.

Theorem 1 and its Corollary imply:

Under the original population model and its first extension the null distribution of $V_s$, the Wilcoxon signed-rank test statistic, is the same as under the randomization model.

Thus we can use the same procedures developed under the randomization model to obtain appropriate critical values and $p$-values (significance probabilities, observed significance levels). Exact calculation, simulation, normal approximation.

# The Power of the Wilcoxon Test in the Shift Model

Under the shift model: $Z_i \sim L_\Delta(z) = L_0(z-\Delta)$, with $L_0$ symmetric around zero, the power of the Wilcoxon test is

$$\Pi_{L_0}(\Delta) = \Pi(\Delta) = P_\Delta(V_s \geq c)$$

where $c$ is a critical value to achieve an appropriate significance level $\alpha$.

The subscript $L_0$ indicates that the power depends not only on $\Delta$, the center of symmetry, but also on the distribution $L_0$ of $Z-\Delta$.

It can be shown that for fixed $L_0$ the power function $\Pi(\Delta)$ is nondecreasing in $\Delta$.

$\Rightarrow$ the Wilcoxon test, rejecting for large $V_s$, is unbiased against alternatives $\Delta > 0$.
A level $\alpha$ test for $H_0 : \Delta = 0$ is also a level $\alpha$ test for the larger hypothesis $H_0' : \Delta \leq 0$.

Opposite monotonicity holds when rejecting for small $V_s$. Corresponding results hold concerning unbiasedness and maintaining a level $\alpha$ over a wider hypothesis.

# Exact Power of the Wilcoxon Test in the Shift Model

As was the case for the two-sample rank-sum test, computing the exact power of the Wilcoxon signed rank test is analytically difficult.
For more on this see Section 6C under Further Developments.

However, we can always use simulation to estimate the power for any $(L_0, \Delta)$.

We only need to simulate independent observations $U_1, \ldots, U_N$ from $L_0$ and then take $Z_i = U_i + \Delta$, and compute $V_s$ for each such sample. Do this $N_{\text{sim}}$ times.

Estimate $\Pi(\Delta)$ by the proportion of simulated $V_s$ values $\geq c_\alpha$.

Here $c_\alpha = \texttt{qsignrank}(1-\alpha, \texttt{N}) + 1$

since $\texttt{k} = \texttt{qsignrank}(1-\alpha, \texttt{N})$ is the smallest $\texttt{k}$ with $\texttt{P}(\texttt{V}_\texttt{s} \leq \texttt{k}) \geq 1-\alpha$,
i.e., the smallest $\texttt{k}$ with $\texttt{P}(\texttt{V}_\texttt{s} > \texttt{k}) \leq \alpha$ or $\texttt{P}(\texttt{V}_\texttt{s} \geq \texttt{k}+1) \leq \alpha$.

# PowerSignRank(...)

The function PowerSignRank is provided on the class web site.

It implements the previous power simulation process for various choices of sampled distributions that are symmetric around zero when $\Delta = 0$.

These distributions include:

the normal distribution,

the Student-$t_f$ distribution with $f$ degrees of freedom (Cauchy distribution for $f = 1$)

the logistic distribution with density $f(x) = \frac{1}{b}\exp(-x/b)/[1+\exp(-x/b)]^2$,

the double exponential or Laplace distribution with density $f(x) = \frac{1}{2b}\exp(-|x|/b)$,

the uniform distribution over the interval $(-a, a)$.

Other distributions could be added. See the addition of the double exponential distribution in the function body of `PowerSignRank`.

# Timing Results for `PowerSignRank(...)`

```
> system.time(PowerSignRank(alpha = 0.05, N = 20, Delta = 1,
      dist = "norm", parm = c(0, 1), Nsim = 10000))
   user  system elapsed
  5.205   0.112   5.322
  > system.time(PowerSignRank(alpha = 0.05, N = 20, Delta = 1,
      dist = "norm", parm = c(0, 1), Nsim = 100000))
   user  system elapsed
 54.011   0.896  54.981
> system.time(PowerSignRank(alpha = 0.05, N = 20, Delta = 1,
      dist = "dexp", parm = c(0, 1), Nsim = 100000))
   user  system elapsed
 54.703   1.312  56.084
```

I tried $\mathtt{Nsim} = 1{,}000{,}000$ on this machine but it degenerated into seemingly endless disk swapping activity (killed it after 50 minutes).

Use this function to check on normal approximations.

# Normal Approximation for $V_s$

For $Z_1, \ldots, Z_N$ i.i.d. $\sim L$ with $0 < P_L(Z_i < 0) < 1$ we have

$$\frac{V_s - E_L(V_s)}{\sqrt{\mathrm{var}_L(V_s)}} \longrightarrow \mathcal{N}(0,1) \quad \text{as} \quad n \longrightarrow \infty.$$

To use this approximation we need the mean and variance of $V_s$.

Using the representation

$$V_s = \sum_{i \le j} I_{[Z_i + Z_j > 0]} = \sum_{i<j} I_{[Z_i + Z_j > 0]} + \sum_i I_{[Z_i > 0]}$$

we easily see

$$E_L(V_s) = \frac{N(N-1)}{2} P_L(Z_1 + Z_2 > 0) + N P_L(Z_1 > 0) = \frac{N(N-1)}{2} p_1' + Np$$

For $Z_i$ symmetric around zero we have $p = p_1' = {}^1\!/_2$ and this reduces to

our previous expression $E_{H_0}(V_s) = N(N+1)/4$.

# The Variance of $V_s$

For the variance one gets the following expression from the same representation

$$\mathrm{var}_L(V_s) = N(N-1)(N-2)(p_2' - p_1'^2)$$
$$+\frac{N(N-1)}{2}\left[2(p-p_1')^2 + 3p_1'(1-p_1')\right] + Np(1-p)$$

where

$$p = P_L(Z_1 > 0), \quad p_1' = P_L(Z_1 + Z_2 > 0), \quad p_2' = P_L(Z_1 + Z_2 > 0,\ Z_1 + Z_3 > 0)$$

For $Z_i$ symmetric around zero we get $p = p_1' = {}^1\!/_2$ and

$$\begin{aligned} p_2' &= P_{H_0}(Z_1 + Z_2 > 0,\ Z_1 + Z_3 > 0) \\ &= P_{H_0}(Z_1 > -Z_2,\ Z_1 > -Z_3) = P_{H_0}(Z_1 > Z_2,\ Z_1 > Z_3) = {}^1\!/_3 \end{aligned}$$

$\Longrightarrow \qquad \mathrm{var}_{H_0}(V_s) = \dfrac{N(N+1)(2N+1)}{24}$ our previous variance expression

# Using the Normal Approximation

If reject for large values of $V_s$, i.e., for $V_s \geq c$, the power of the test is

$$\Pi(L) = P_L(V_s \geq c) = P\left(\frac{V_s - E_L(V_s)}{\sqrt{\text{var}_L(V_s)}} \geq \frac{c - E_L(V_s)}{\sqrt{\text{var}_L(V_s)}}\right)$$

and using the continuity correction in the normal approximation this becomes

$$\Pi(L) \approx 1 - \Phi\left(\frac{c - {}^1\!/_2 - E_L(V_s)}{\sqrt{\text{var}_L(V_s)}}\right)$$

When rejecting for low values of $V_s$ the power approximation becomes

$$\Pi(L) = P_L(V_s \leq c) \approx \Phi\left(\frac{c + {}^1\!/_2 - E_L(V_s)}{\sqrt{\text{var}_L(V_s)}}\right)$$

# Example 2: Treatment for Anemia

Vitamin $B_{12}$ is to be tested for effectiveness against pernicious anemia.

Hemoglobin levels are measured for 10 patients with that disease,

before and after the treatment $\longrightarrow \quad X_i, Y_i \quad \longrightarrow \quad Z_i = Y_i - X_i$.

We use the Wilcoxon signed-rank test at a target level $\alpha = .05$.

If $B_{12}$ is effective, we expect higher levels after treatment,

i.e., the $Z$ values will be slanted more toward positive values.

Thus we should reject the hypothesis $H_0$ of no effect when $V_s \geq c$.

`1-psignrank(43,10) = 0.05273` and `1-psignrank(44,10) = 0.04199`

Thus we compromise and take $c = 44$ with $\alpha = 0.05273$.

# Anemia Power Calculation

To calculate the power of our test we need to specify an alternative.

Assume that $Z \sim L = \mathcal{N}(\Delta, \tau^2)$ with $\Delta = 2g/\ell$ and $\tau = 2g/\ell$.

We need to determine $p$, $p_1'$ and $p_2'$ for this distribution.

$$p = P(Z > 0) = P\left(\frac{Z-\Delta}{\tau} > \frac{0-\Delta}{\tau}\right) = 1 - \Phi\left(-\frac{\Delta}{\tau}\right) = 1 - \Phi(-1) = \texttt{pnorm}(1) = 0.84134$$

$$p_1 = P(Z_1 + Z_2 > 0) = P\left(\frac{Z_1 + Z_2 - 2\Delta}{\tau\sqrt{2}} > \frac{0 - 2\Delta}{\tau\sqrt{2}}\right) = 1 - \Phi\left(-\frac{2\Delta}{\tau\sqrt{2}}\right) = 0.92135$$

$$p_2' = P(Z_1 + Z_2 > 0, Z_1 + Z_3 > 0) = P\left(\frac{Z_1 + Z_2 - 2\Delta}{\tau\sqrt{2}} > \frac{0 - 2\Delta}{\tau\sqrt{2}}, \frac{Z_1 + Z_3 - 2\Delta}{\tau\sqrt{2}} > \frac{0 - 2\Delta}{\tau\sqrt{2}}\right)$$

The random variables $U_1 = (Z_1 + Z_2 - 2\Delta)/(\tau\sqrt{2})$ and $U_2 = (Z_1 + Z_3 - 2\Delta)/(\tau\sqrt{2})$ have a bivariate normal distribution with means zero and variances one and with correlation coefficient $\rho = {}^1\!/_2$, as encountered before.

# Anemia Power Calculation (continued)

With $u_1 = u_2 = -2\Delta/\tau\sqrt{2}$ we have

$$p_2' = P(U_1 > u_1, U_2 > u_2) = P(-U_1 < -u_1, -U_2 < -u_2) = P(U_1 < -u_1, U_2 < -u_2)$$

$$= \texttt{pmnorm}(\texttt{c}(2 * \texttt{Delta}/(\texttt{sqrt}(2) * \texttt{tau}), 2 * \texttt{Delta}/(\texttt{sqrt}(2) * \texttt{tau})),$$

$$\texttt{c}(0,0), \texttt{varcov} = \texttt{matrix}(\texttt{c}(1, .5, .5, 1), \texttt{ncol} = 2)) = 0.8657672$$

$$\Longrightarrow \quad E_L(V_s) = 49.87422 \quad \text{and} \quad \text{var}_L(V_s) = 23.84759$$

$$\Pi(L) = 1 - \Phi\left(\frac{44 - .5 - 49.87422}{\sqrt{23.84759}}\right) = 0.904102$$

Klotz (1963) gives an exact value of $.8914$ based on integration.

Using `PowerSignRank` with `Nsim=100000` I got $0.8917$ by simulation

in 52 seconds on this laptop. ($0.8905$ in 17 seconds on my other laptop)

# Comments on the Normal Approximation

The previous normal approximation was illustrated when $Z_i \sim L = \mathcal{N}(\Delta, \tau^2)$.

The calculation of $p = P(Z_1 > 0)$ and $p'_1 = P(Z_1 + Z_2 > 0)$ was relatively easy, since it involved simple normal tail probabilities and since $Z_1 + Z_2 \sim \mathcal{N}(2\Delta, 2\tau^2)$.

The calculation of $p'_2$ used the fact that it involved a bivariate normal quadrant probability, again a consequence of $L$ being normal.

This is typically no longer so easy (especially for $p'_2$) when $L$ is not normal.

Of course, one could estimate these probabilities by simulating a large number of independent triplets $Z_1, Z_2, Z_3$ i.i.d. $\sim L$ and use these estimated probabilities $\hat{p}_1 = \#\{Z_{1i} > 0\}/N_{\text{sim}}, \quad \hat{p}'_2 = \#\{(Z_{1i} + Z_{2i}) > 0\}/N_{\text{sim}}$ and $\hat{p}'_3 = \#\{(Z_{1i} + Z_{2i}) > 0 \;\&\; (Z_{1i} + Z_{3i}) > 0\}/N_{\text{sim}}$ in the normal approximation.

However, one could then also simulate the distribution of $V_s$. The simulation effort would then be more substantial, depending on the sample size $N$.

# Alternate Normal Approximation

In the case of shift alternatives $L_\Delta(z) = L_0(z - \Delta)$ a different normal approximation is possible. It parallels that obtained for the Wilcoxon rank-sum test.

Let $L_0^*(u) = P_{L_0}(Z_1 + Z_2 \le u)$ with $Z_1, Z_2$ i.i.d. $\sim L_0(z)$ with density $\ell_0(z)$.

The densities of $L_0$ and $L_0^*$ evaluated at zero are denoted by $\ell_0(0)$ and $\ell_0^*(0)$.

The following alternate normal approximation should then be reasonable

for large $N$ and small $\Delta$

$$\Pi_{L_0}(\Delta) \approx \Phi\left(\frac{N(N-1)\ell_0^*(0) + N\ell_0(0)}{\sqrt{N(N+1)(2N+1)/24}}\Delta - u_\alpha\right)$$

# General Expression for $\ell_0^*(0)$

$$P_{L_0}(Z_1+Z_2\le u)=\int_{-\infty}^{\infty}P_{L_0}(Z_1\le u-z_2)\ell_0(z_2)dz_2=\int_{-\infty}^{\infty}L_0(u-z_2)\ell_0(z_2)dz_2$$

$$\Longrightarrow \quad \ell_0^*(u) \quad = \quad \int_{-\infty}^{\infty}\ell_0(u-z_2)\ell_0(z_2)dz_2$$

$$\Longrightarrow \quad \ell_0^*(0) \quad = \quad \int_{-\infty}^{\infty}\ell_0(-z_2)\ell_0(z_2)dz_2=\int_{-\infty}^{\infty}\ell_0^2(z)dz$$

Here the first $\Longrightarrow$ invokes interchange of $\partial/\partial u$ and integration,

and the last $=$ invokes the symmetry of $\ell_0$ around zero, i.e., $\ell_0(-z)=\ell_0(z)$.

The next slide gives $\ell_0(0)$ and $\ell_0^*(0)$ for five types of distributions,

all of which are symmetric around zero.

The double exponential distribution is also known as the Laplace distribution.

The Student-$t_1$ distribution is also known as the Cauchy distribution.

# $\ell_0(0)$ and $\ell_0^*(0)$ for Several Distributions

| distribution | density | $\ell_0(0)$ | $\ell_0^*(0)$ |
|---|---|---|---|
| Normal $\mathcal{N}(0,\tau^2)$ | $\frac{1}{\tau\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\tau^2}\right)$ | $\frac{1}{\tau\sqrt{2\pi}}$ | $\frac{1}{2\tau\sqrt{\pi}}$ |
| Uniform $\mathcal{U}(-\tau,\tau)$ | $\frac{1}{2\tau}I_{[-\tau,\tau]}(x)$ | $\frac{1}{2\tau}$ | $\frac{1}{2\tau}$ |
| Logistic | $\frac{\exp(-x/\tau)}{\tau(1+\exp(-x/\tau))^2}$ | $\frac{1}{4\tau}$ | $\frac{1}{6\tau}$ |
| Student-$t_\nu$ | $\left(1+\frac{(t/\tau)^2}{\nu}\right)^{-\frac{\nu+1}{2}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\tau\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}$ | $\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\tau\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}$ | $\frac{\Gamma^2\left(\frac{\nu+1}{2}\right)\Gamma\left(\nu+\frac{1}{2})\right)}{\tau\sqrt{\nu\pi}\,\Gamma^2(\frac{\nu}{2})\Gamma(\nu+1)}$ |
| Double Exponential | $\frac{1}{2\tau}\exp(-\lvert x/\tau\rvert)$ | $\frac{1}{2\tau}$ | $\frac{1}{4\tau}$ |

In R we get $\Gamma(x)$ via `gamma(x)`.

# Anemia Example Revisited

$L_0 = \mathcal{N}(0, \tau^2) \Rightarrow L_0^* = \mathcal{N}(0, 2\tau^2) \Longrightarrow \ell_0(0) = 1/(\tau\sqrt{2\pi})$ and $\ell_0^*(0) = 1/(2\tau\sqrt{\pi})$.

$$\Pi_{L_0}(\Delta) \approx \Phi\left(\frac{N(N-1)/2 + N/\sqrt{2}}{\sqrt{N(N+1)(2N+1)/24}}\,\frac{\Delta}{\tau\sqrt{\pi}} - u_\alpha\right)$$

For $\Delta = 2$, $\tau = 2$, $\alpha = .05$ and $N = 10$ we get $\Pi_{L_0}(\Delta) \approx 0.9114$

For the more appropriate $\alpha = .05273$ we get $\Pi_{L_0}(\Delta) \approx 0.9155$

as compared to the exact value $.8914$ and our previous approximation $0.9041$

# `PowerSignRank`

The function `PowerSignRank` (on the class web site) simulates the power of the one-sided Wilcoxon signed-rank test under shift alternatives for the five types of distributions given in the previous table for $\ell_0(0)$ and $\ell_0^*(0)$.

The time to run it is proportional to $\mathtt{Nsim}$. Thus try it first with $\mathtt{Nsim} = 10000$ before running it for larger $\mathtt{Nsim}$.

The time to run `PowerSignRank` grows only very slowly with the sample size $N$. When $N = 10$ needed 4.9 seconds, it took only 8 seconds for $N = 100$.

Thus the function can be used for sample size planning to obtain a desired power for an anticipated shift alternative.

```
PowerSignRank(alpha = 0.05, N = 10, Delta = 1, scale = 1,
    dist = "norm", df = 1, Nsim = 10000)
```
See internal documentation.

# Rough Sample Size Planning

To get a desired power $\Pi$ define $u_\Pi$ (with $Z \sim \mathcal{N}(0,1)$)

$$\Pi = P(Z \geq u_\Pi) = 1 - \Phi(u_\Pi) = \Phi(-u_\Pi)$$

$$\Pi = \Phi(-u_\Pi) = \Pi_{L_0}(\Delta) \approx \Phi\left(\frac{N(N-1)\ell_0^*(0) + N\ell_0(0)}{\sqrt{N(N+1)(2N+1)/24}}\Delta - u_\alpha\right)$$

$$\Longrightarrow \qquad -u_\Pi \approx \frac{N(N-1)\ell_0^*(0) + N\ell_0(0)}{\sqrt{N(N+1)(2N+1)/24}}\Delta - u_\alpha$$

We can either solve this for $N$ by trial and error (or write an R function to do it) or in anticipation of a large $N$ we may replace the numerator by $N^2\ell_0^*(0)$ and the denominator by $\sqrt{N^3/12}$ and get

$$u_\alpha - u_\Pi \approx \frac{N^2\ell_0^*(0)}{N^{3/2}}\Delta\sqrt{12} \qquad \Longrightarrow \qquad N \approx \frac{(u_\alpha - u_\Pi)^2}{12\Delta^2\ell_0^{*2}(0)}$$

# Example of Sample Size Planning

Suppose we want a level $\alpha = .01$ test to achieve a power of $\Pi = .95$ against a normal shift alternative with $\Delta/\tau = .5$ when $L_0 = \mathcal{N}(0,\tau^2)$, i.e., $L_\Delta = \mathcal{N}(\Delta,\tau^2)$

We have $u_\Pi = -1.645$ and $u_\alpha = 2.326$. With $\ell_0^{*2}(0) = 1/(4\pi\tau^2)$ we get

$$N \approx \frac{(u_\alpha - u_\Pi)^2}{12\Delta^2\ell_0^{*2}(0)} = \frac{4\pi\tau^2(u_\alpha - u_\Pi)^2}{12\Delta^2} = \frac{(2.326+1.645)^2\pi}{3\times .5^2} = 66.05$$

Thus $N = 67$ may be good first try to be followed up with

$\texttt{PowerSignRank}(\texttt{alpha} = .01, \texttt{N} = 67, \texttt{Delta} = .5, \texttt{dist} = \text{"}\texttt{norm}\text{"}, \texttt{Nsim} = 100000)$

$\Longrightarrow$ simulated power of `.94523` and normal approximations `.9513` and `.9508`.

Trying $N = 70$ gave a simulated power of `0.9535`.

# Power and Ties

The signed-rank test is no longer exactly distribution-free when we have ties.

However, the signed-rank test can be carried out conditionally given the tie or midrank pattern. Its null distribution can be simulated.
Appropriate critical points for a conditional level $\alpha$ test can be estimated.

The overall level of the test (over all tie patterns) is then still $\leq \alpha$.
The amount of conservatism will depend on the sampled distribution.

Shift alternatives with discrete distributions make little sense.

If ties are due to rounding of inherently continuous random variables, one could again explore the power behavior under shift alternatives for specific continuous distributions via simulation. Just round the observations to the relevant rounding grid and use shifts that are similarly rounded.

# Comparing Sign, Wilcoxon, and $t$-Tests

In the following we will compare the sign test, the Wilcoxon signed-rank test and Student's one-sample $t$-test for large samples in the shift model.

This parallels to a great extent our previous comparison in the two-sample situation.

Here we include one additional test in the comparison, namely the sign test.

First we have to discuss the one-sample $t$ test and its large sample properties.

# Student's One-Sample $t$-Test

Student's (Gosset's) one-sample $t$-test addresses the following testing problem:

Let $Z_1, \ldots, Z_N \sim \mathcal{N}(\Delta, \sigma^2)$ and test $H_0 : \Delta = \Delta_0$ for specified value of $\Delta_0$.

Considering as anticipated alternative that $\Delta > \Delta_0$, the $t$-test rejects $H_0$ whenever

$$\frac{\bar{Z} - \Delta_0}{S/\sqrt{N}} \geq c \qquad \text{where} \qquad S^2 = S_Z^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Z_i - \bar{Z})^2$$

and $c$ is the $(1-\alpha)$-quantile of the $t$-distribution with $N-1$ degrees of freedom.

Since we can always subtract $\Delta_0$ from all $Z_i$, i.e $Z_i' = Z_i - \Delta_0 \sim \mathcal{N}(\Delta' = \Delta - \Delta_0, \sigma^2)$,

we get $\bar{Z}_0' = \bar{Z} - \Delta_0$ and $S_{Z'}^2 = S_Z^2$.

Our hypothesis $H_0$ becomes $H_0' : \Delta' = 0$ which is rejected whenever $\sqrt{N}\bar{Z}'/S_Z$.

Thus we may as well assume $\Delta_0 = 0$ and drop the $'$ on $Z_i'$.

# The $t$-Test under $H_0$ Is $\approx$ Distribution-free

The $t$-test has exact level $\alpha$ only when $Z_1, \ldots, Z_N$ i.i.d. $\sim \mathcal{N}(0, \sigma^2) = \mathcal{N}(\Delta_0, \sigma^2)$.

For large $N$ it has approximate level $\alpha$ for other distributions with mean zero and finite variance $\sigma^2$. This follows from the CLT and the LLN.

$$\text{CLT} \quad \Longrightarrow \frac{\bar{Z}}{\sigma/\sqrt{N}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{and by the LLN} \quad \Longrightarrow S^2 \xrightarrow{P} \sigma^2 \ \text{ or } \ \frac{S}{\sigma} \xrightarrow{P} 1$$

$$\Longrightarrow \quad \frac{\sqrt{N}\bar{Z}}{S} = \frac{\sqrt{N}\bar{Z}/\sigma}{S/\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

$$\alpha = P_{H_0}\left(\frac{\sqrt{N}\bar{Z}}{S} \geq c_N(\alpha)\right) \approx 1 - \Phi(c_N(\alpha)) \implies c_N(\alpha) \approx u_\alpha = z_{1-\alpha} = \Phi^{-1}(1-\alpha)$$

Thus we can use $u_\alpha$ in place of $c_N(\alpha)$ in large samples from any distribution with mean zero and finite variance.

The $t$-test is approximately distribution-free under $H_0$.

# Approximate Power of the $t$-Test

Let $Z_i \sim L$ under the alternative, with mean $E(Z)$ and variance $\sigma^2$.

By the LLN we still have $S/\sigma \longrightarrow 1$ as $N \longrightarrow \infty$ and by the CLT we still have

$$\frac{\sqrt{N}(\bar{Z} - E(Z))}{\sigma} \longrightarrow \mathcal{N}(0,1) \qquad \text{as } N \longrightarrow \infty.$$

Thus we get the following approximation for the power of the $t$-test

$$\begin{aligned}
\Pi_t(L) = P_L\left(\frac{\sqrt{N}\bar{Z}}{S} \geq c_N(\alpha)\right) &= P_L\left(\frac{\sqrt{N}(\bar{Z} - E_L(Z))}{S} \geq c_N(\alpha) - \frac{\sqrt{N}E_L(Z)}{S}\right) \\
&= P_L\left(\frac{\sqrt{N}(\bar{Z} - E_L(Z))}{\sigma} \geq c_N(\alpha)\frac{S}{\sigma} - \frac{\sqrt{N}E_L(Z)}{\sigma}\right) \\
&\approx 1 - \Phi\left(u_\alpha - \frac{\sqrt{N}E_L(Z)}{\sigma}\right) = \Phi\left(\frac{\sqrt{N}\Delta}{\sigma} - u_\alpha\right)
\end{aligned}$$

where the last equality pertains to shift alternatives $L(z) = L_0(z - \Delta)$.

Note that $L_0$ affects the power expression through its standard deviation $\sigma$.

# Approximate Power for Sign and Wilcoxon Tests

Under shift alternatives the sign test had approximate power (slide 19)

$$\Pi_S(L) \approx \Phi\left(2\sqrt{N}\Delta\ell_0(0) - u_\alpha\right)$$

For the Wilcoxon signed-rank test we had the following approximate power under shift alternatives

$$\Pi_V(L) \approx \Phi\left(\frac{N(N-1)\ell_0^*(0) + N\ell_0(0)}{\sqrt{N(N+1)(2N+1)/24}}\,\Delta - u_\alpha\right)$$

where $\ell_0(0)$ and $\ell_0^*(0)$ were given on slide 41 for five different distributions.

# Comparing Sign, Wilcoxon, and $t$-Tests

Let us compare the three tests for the specific case of $N = 67$ for a normal shift alternative with $\Delta/\tau = .5$ and $\alpha = .01$

$$\Pi_t \approx \Phi\left(\frac{\Delta}{\tau}\sqrt{N} - u_{.01}\right) = \Phi\left(.5\sqrt{67} - 2.326\right) = 0.961$$

$$\Pi_S \approx \Phi\left(2\sqrt{N}\frac{\Delta}{\tau\sqrt{2\pi}} - u_\alpha\right) = \Phi\left(2\sqrt{67}\cdot .5 \cdot .3989 - 2.326\right) = .8262$$

$$\begin{aligned}
\Pi_V &\approx \Phi\left(\frac{N(N-1)\ell_0^*(0) + N\ell_0(0)}{\sqrt{N(N+1)(2N+1)/24}}\Delta - u_\alpha\right) \\
&= \Phi\left(\frac{67\cdot 66\cdot 0.2821 + 67\cdot 0.3989}{\sqrt{67\cdot 68\cdot 135/24}}.5 - 2.326\right) = 0.951
\end{aligned}$$

where $\ell_0^*(0) = 1/(2\sqrt{\pi}\tau) = 0.2821/\tau$ and $\ell_0(0) = 1/(\tau\sqrt{2\pi}) = 0.3989/\tau$.

# Sample Size Comparison for Equal Power

Plan the sample sizes for each of these test, for power $\Pi = .95$ and $\alpha = .01$.

From the previous we have for the Wilcoxon signed-rank test $N_V = 67$.

For the sign test we get

$$\begin{aligned} N_S &= \left( \frac{u_\alpha + \Phi^{-1}(\Pi)}{2} \frac{\sqrt{2\pi}}{\Delta/\tau} \right)^2 \\ &= \left( (2.326 + 1.645)\sqrt{2\pi} \right)^2 = 99 \end{aligned}$$

and for the $t$-test

$$N_t = \left( \frac{u_\alpha + \Phi^{-1}(\Pi)}{\Delta/\tau} \right)^2 = \left( \frac{2.326 + 1.645}{.5} \right)^2 = 63$$

# Relative Efficiencies

As in Chapter 2 we take the sample size ratios as a measure of the efficiencies of the tests relative to each other.

$$e_{S,t} = \frac{63}{99} = .636\,, \qquad e_{V,t} = \frac{63}{67} = .940\,, \qquad e_{S,V} = \frac{67}{99} = .677$$

For example, the Wilcoxon test requires with $N_V = 67$ only $67.7\%$ of the sample size $N_S = 99$ that is needed by the sign test to achieve the same power $\Pi = .95$ at the same significance level $\alpha = .01$.

# Relative Efficiencies $e_{S,t}$

The following table (Dixon, 1953) gives the power $\Pi_S$ of the Sign test over a spectrum of normal shift alternatives $\Delta/\sigma$ for two sample sizes $N = 10, 20$. Above each power row $\Pi_S$ are given the relative efficiencies $e_{S,t} = N_t/N_S$, which are remarkably stable.

| | $\Delta/\sigma$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | .1257 | .2534 | .3853 | .5244 | .6745 | .8416 | 1.0364 | 1.2816 | 1.6449 |
| $N = 10$ | | | | | | | | | | |
| $e_{S,t}$ | | .768 | .767 | .765 | .761 | .756 | .749 | .741 | .729 | .712 |
| | $\alpha$ | | | | | | | | | |
| $\Pi_S$ | $.1094$ | .13 | .18 | .27 | .38 | .53 | .68 | .82 | .93 | .99 |
| $N = 20$ | | | | | | | | | | |
| $e_{S,t}$ | | .698 | .696 | .693 | .688 | .683 | .677 | .669 | | |
| | $\alpha$ | | | | | | | | | |
| $\Pi_S$ | $.1153$ | .15 | .26 | .42 | .61 | .79 | .91 | .98 | | |

# Relative Efficiencies $e_{W,t}$

The table below (Klotz,1963) shows the corresponding efficiency comparison for the Wilcoxon test relative to the $t$-test for $N = 10$ and $\alpha \approx .05, .10$.

Again note the stability of the efficiencies relative to the spectrum of normal shift alternatives.

| | $\Delta/\sigma$ | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | .25 | .50 | .75 | 1.00 | 1.25 | 1.50 |
| $e_{V,t}$ | | .968 | .967 | .966 | .965 | .965 | .964 |
| | $\alpha$ | | | | | | |
| $\Pi_V$ | $.05237$ | .1844 | .4274 | .7013 | .8914 | .9734 | .9957 |
| $e_{V,t}$ | | .960 | .959 | .957 | .956 | .955 | |
| | $\alpha$ | | | | | | |
| $\Pi_V$ | $.09668$ | .2862 | .5669 | .8153 | .9476 | .9904 | .9989 |

# Asymptotic Relative Efficiencies

As $N \to \infty$ the following expressions can be derived for the asymptotic relative efficiences (ARE's or Pitman efficiencies) of the sign and Wilcoxon tests relative to the $t$-test.

$$e_{S,t}(L_0) = \lim_{N\to\infty} e_{S,t,N}(L_0) = 4\sigma^2 \ell_0^2(0)$$

and

$$e_{V,t}(L_0) = \lim_{N\to\infty} e_{V,t,N}(L_0) = 12\sigma^2 \left[\int_{-\infty}^{\infty} \ell_0^2(x)dx\right]^2 = 12\sigma^2 \ell_0^{*2}(0)$$

These are obtained by finding $N_V(N)$ and $N_t(N)$ to yield the same power $\Pi$ at an appropriate shift alternative $\Delta_N$. Here $\sigma^2$ is the variance of $L_0$.

To maintain power at a fixed level $\Pi$ with $\alpha < \Pi < 1$ it is necessary that $\Delta_N \to 0$.

The efficiencies do not depend on $\alpha$ or the targeted power $\Pi$.

# Sketch of Proof for $e_{V,t}$

Recall our second normal approximation for the power of the Wilcoxon test

$$\Pi_V(L_\Delta) \approx \Phi\left(\frac{N(N-1)\ell_0^*(0)+N\ell_0(0)}{\sqrt{N(N+1)(2N+1)/24}}\,\Delta - u_\alpha\right)$$

and for the $t$-test with $\sigma^2$ denoting the variance of $L_0$ we had

$$\Pi_t(L_\Delta) \approx 1-\Phi\left(u_\alpha - \frac{\sqrt{N}}{\sigma}\Delta\right) = \Phi\left(\frac{\sqrt{N}}{\sigma}\Delta - u_\alpha\right)$$

Note that in either case $\sqrt{N}\Delta$ has to converge to a finite value $A>0$ to maintain constant power $\Pi$ with $\alpha < \Pi < 1$. Thus let $\Delta_{N_t} = A/\sqrt{N_t}$.

Matching power for respective sample sizes $N_V$ and $N_t$ we get

$$\frac{N_V(N_V-1)\ell_0^*(0)+N_V\ell_0(0)}{\sqrt{N_V(N_V+1)(2N_V+1)/24}}\frac{A}{\sqrt{N_t}} = \frac{N_V(N_V-1)\ell_0^*(0)+N_V\ell_0(0)}{\sqrt{N_V^2(N_V+1)(2N_V+1)/24}}\frac{A\sqrt{N_V}}{\sqrt{N_t}} = \frac{A}{\sqrt{N_t}}\frac{\sqrt{N_t}}{\sigma}$$

$$\Longrightarrow \quad \sqrt{12}\ell_0^*(0)\sigma = \lim_{N_t\to\infty}\frac{\sqrt{N_t}}{\sqrt{N_V}} \quad \Longrightarrow \quad \lim_{N_t\to\infty}\frac{N_t}{N_V} = 12\ell_0^{*2}(0)\sigma^2 \qquad \Box$$

We clearly see how dependence on $\alpha$ ($u_\alpha$) and power (through $A$) drop out.

# Sketch of Proof for $e_{S,t}$

Recall the alternate power approximation for the sign test

$$\Pi_S(L_\Delta) \approx \Phi\left(2\sqrt{N_S}\,\Delta\,\ell_0(0) - u_\alpha\right)$$

and match that for common $\Delta$ with the approximate power of the $t$ test

$$\Pi_t(L_\Delta) \approx \Phi\left(\frac{\sqrt{N_t}}{\sigma}\Delta - u_\alpha\right)$$

For small shifts $\Delta_{N_t} = A/\sqrt{N_t}$ we get

$$2\sqrt{N_S}\,\frac{A}{\sqrt{N_t}}\,\ell_0(0) = \frac{\sqrt{N_t}}{\sigma}\,\frac{A}{\sqrt{N_t}} \qquad \Longrightarrow \qquad \lim_{N_t\to\infty}\frac{N_t}{N_S} = 4\sigma^2\ell_0^2(0) \qquad \square$$

Again we clearly see how any dependence on $\alpha$ ($u_\alpha$) and the common targeted power $\Pi$ (through $A$) falls away.

# ARE's for Some Distributions

From our previously tabulated expression for $\ell_0(0)$ and $\ell_0^*(0)$ we get the following

| | normal | logistic | Laplace | uniform | $t_1$ | $t_5$ |
|---|---|---|---|---|---|---|
| $e_{S,t}$ | $\frac{2}{\pi} = 0.6366$ | $\frac{\pi^2}{12} = 0.8225$ | 2 | $\frac{1}{3}$ | $\infty$ | 0.9607 |
| $e_{V,t}$ | $\frac{3}{\pi} = 0.9549$ | $\frac{\pi^2}{9} = 1.0966$ | $\frac{3}{2}$ | 1 | $\infty$ | 1.241184 |

Note the generally high efficiencies of the Wilcoxon test.

The efficiency of the Wilcoxon signed-rank test relative to the one-sample $t$-test is the same as that of Wilcoxon rank-sum test relative to the two-sample $t$-test, as long as the sampled distribution is symmetric.

The same lower bound of $.864 = 108/125$ applies to the Wilcoxon signed-rank test relative to the one-sample $t$-test.

# ARE Conversions

If we have three tests, indicated by indices 1,2,3 on the sample sizes $N_i$ required to attain power $\Pi$ at level $\alpha$ then the AREs are given by

$$e_{12} = \lim \frac{N_2}{N_1} \qquad e_{13} = \lim \frac{N_3}{N_1} \qquad e_{23} = \lim \frac{N_3}{N_2}$$

It follows immediately that

$$e_{ij} = \lim \frac{N_j}{N_i} = \frac{1}{\lim \frac{N_i}{N_j}} = \frac{1}{e_{ji}}$$

and

$$e_{ij} = \lim \frac{N_j}{N_i} = \lim \frac{N_k}{N_i} \times \lim \frac{N_j}{N_k} = e_{ik} \times e_{kj}$$

As an illustration we have

$$e_{S,W} = e_{S,t} \times e_{t,W} = \frac{e_{S,t}}{e_{W,t}} = \frac{2/\pi}{3/\pi} = \frac{2}{3}$$

# Estimating Location or Treatment Effect

A quantity $\theta$ is measured with error by observations $\mathbf{Z} = (Z_1, \ldots, Z_N)$.

We assume that the distribution $L(z) = L_\theta(z) = L_0(z - \theta)$ of $Z_i$ is symmetric around $\theta$, or the distribution $L_0(z)$ of $Z_i - \theta$ is symmetric around zero.

The most common estimate of $\theta$ is the average

$$\bar{\theta} = \bar{\theta}(\mathbf{Z}) = \frac{Z_1 + \ldots + Z_N}{N}$$

This estimation problem arises also quite naturally in the context of paired comparisons when we want to estimate the treatment effect $\theta = \Delta$.

When $L(z) = \Phi((z - \theta)/\sigma))$ the average has several optimality claims, i.e., it is the best estimator of $\theta$ among many broad classes of estimators.

# Unfavorable Aspects of $\bar{\theta}$

It is very sensitive to outliers or gross errors in the observations.

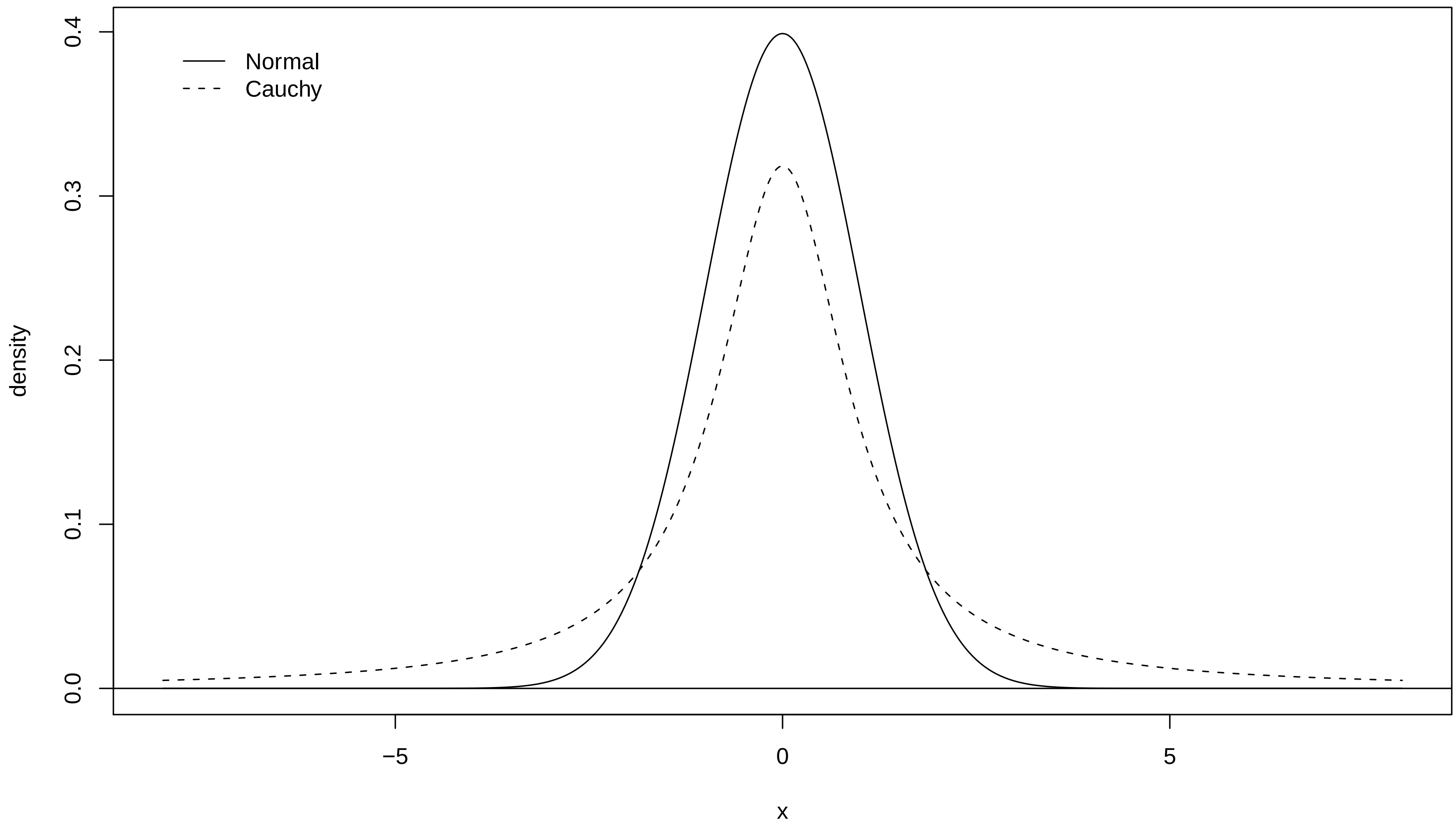
When sampling heavy tailed distributions it can be very inefficient in its use of $N$ independent measurements $Z_1, \ldots, Z_N$.

In fact, when sampling from the Cauchy distribution with center $\theta$ the average of $N$ observations is as accurate an estimate as using just one observation.

The Cauchy distribution with center 0 and scale 1 ($f(x) = [\pi(1+x^2)]^{-1}$) is the same as the $t$-distribution with 1 degree of freedom, see next slide.

For some distributions the average is even worse than using a single observation.

# Normal and Cauchy Distribution



# Relation of $\bar{\theta}$ to $t$-Test

The two-sided $t$-test for testing $H_{\theta_0} : \theta = \theta_0$ against $A_{\theta_0} : \theta \neq \theta_0$ rejects $H_{\theta_0}$ whenever $\sqrt{N}|\bar{\theta} - \theta_0|/S \geq c$, with $p$-value for observed $\bar{\theta}_{\text{obs}}$ and $S_{\text{obs}}$

$$p(\bar{\theta}_{\text{obs}}, S_{\text{obs}}) = P_{H_{\theta_0}}\left(\frac{|\bar{\theta} - \theta_0|}{S/\sqrt{N}} \geq \frac{|\bar{\theta}_{\text{obs}} - \theta_0|}{S_{\text{obs}}/\sqrt{N}}\right) = P\left(|t_{N-1}| \geq |t_{N-1,\text{obs}}|\right)$$

We will now find the $\theta_0$ for which the corresponding $p$-value gives the highest value, i.e., for which the data are best aligned with $H_{\theta_0}$.

That highest $p$-value is 1 when $\theta_0 = \bar{\theta}_{\text{obs}} = (z_1 + \ldots + z_N)/N$, i.e., $t_{N-1,\text{obs}} = 0$.

Thus the average coincides with the $\theta_0$ for which the observed data lend the strongest support to $H_{\theta_0}$ from the $t$-test perspective.

# Alignment Is Relative

On the previous slide we aligned $\theta_0$ with the data.

Recall that the distribution $L_0$ of $Z_i - \theta$ is symmetric around zero when $\theta$ is the correct location parameter.

We can turn the above alignment approach around and shift the data by $\theta_0$:
$Z_i \longrightarrow Z_i' = Z_i - \theta_0$ so that they best align around zero.

"Best alignment" depends on the discrepancy metric used,
here it is the $t$-test statistic for testing $H_0 : \theta = 0$

$$t(\mathbf{Z}') = \frac{|\sum_{i=1}^{N} Z_i'/N|}{S'/\sqrt{N}} = \frac{|\sum_{i=1}^{N}(Z_i - \theta_0)/N|}{S/\sqrt{N}} = \frac{|\sum_{i=1}^{N} Z_i/N - \theta_0|}{S/\sqrt{N}}$$

which becomes smallest (best alignment) when $\theta_0 = \bar{\theta} = (Z_1 + \ldots + Z_N)/N = \bar{Z}$.

Note that $S'^2 = \sum_{i=1}^{N}(Z_i' - \bar{Z}')^2/N = \sum_{i=1}^{N}(Z_i - \bar{Z})^2/N = S^2$.

# Alignment Using the Sign Test

Using again $Z_i' = Z_i - \theta_0$ we look for best alignment around zero when assessing best alignment by using the two-sided sign test for testing $H_0 : \theta = 0$.

Recall that $S_N = \sum_{i=1}^{N} I_{[Z_i' > 0]}$ has a distribution symmetric around $N/2$ under $H_0$.

We get greatest support for $H_0$ when $S_{N,\text{obs}} = N/2$ for an appropriate choice of $\theta_0$.

$$p\text{-value} = P_{H_0}(|S_N - N/2| \geq |S_{N,\text{obs}} - N/2|) = P_{H_0}(|S_N - N/2| \geq 0) = 1$$

$S_{N,\text{obs}} = N/2$ whenever half of the $Z_i'$ are $> 0$ and half are $< 0$, or

half of the $Z_i$ are $> \theta_0$ and half are $< \theta_0$, i.e., when

$$\theta_0 = \tilde{\theta} = \tilde{\theta}(\mathbf{Z}) = \text{median}(Z_1, \ldots, Z_N) = \text{med}(Z_i)$$

For even $N$ this is clear from the above but for odd $N$ it requires some special care of details invoking the modified definition of $S_N$ when we have zeros among the $Z_i'$.

# Alignment Using the Wilcoxon Test

Recall: the signed-rank statistic $V_s \equiv$ to the number of $(Z_i + Z_j)/2 > 0$ with $i \le j$.

Its null distribution is symmetric around $E_0(V_s) = N(N+1)/4$.

We get best alignment with $Z_i' = Z_i - \theta_0$ when half of the

$$(Z_i' + Z_j')/2 = ([Z_i - \theta_0] + [Z_j - \theta_0])/2 = (Z_i + Z_j)/2 - \theta_0, \quad i \le j,$$

are $> 0$ and half are $< 0$, i.e., when half of the $(Z_i + Z_j)/2$ with $i \le j$ are $> \theta_0$ and half are $< \theta_0$. This leads to

$$\theta_0 = \hat{\theta} = \hat{\theta}(\mathbf{Z}) = \text{median}\left(\frac{Z_i + Z_j}{2},\ i \le j\right) = \text{med}_{i \le j}\left(\frac{Z_i + Z_j}{2}\right)$$

as the best alignment estimator with respect to the signed-rank test.

This is also known as the Hodges-Lehmann estimator,

an interesting compromise between average and median.

# Example 3: Weight of One-year-old Boys

$\mathbf{Z} =$ `c(12.01, 8.99, 10.21, 12.15, 9.54, 9.85, 10.62, 9.52, 10.66, 9.87, 10.44, 10.51, 10.67, 11.16, 9.32, 9.62, 11.11, 9.14)`

`mat=outer(`$\mathbf{Z}$`,`$\mathbf{Z}$`,"+")/2` is the matrix of all pairwise averages $(Z_i + Z_j)/2$.

We would like to get the vector of all averages with $i \leq j$, and then take the median of this vector as $\hat{\theta}$. This is done as follows

$$\hat{\theta}(\mathbf{Z}) = \text{med}_{i \leq j}\left(\frac{Z_i + Z_j}{2}\right) = \mathtt{median}(\mathtt{mat}[\mathtt{row}(\mathtt{mat}) <= \mathtt{col}(\mathtt{mat})]) = 10.24$$

Here `row(mat)<=col(mat)` creates a matrix of `T` in any position $(i, j)$ with $i \leq j$ and `F` in all other positions. For the average and median we get

$$\bar{\theta}(\mathbf{Z}) = \mathtt{mean}(\mathbf{Z}) = 10.29944 \qquad \text{and} \qquad \tilde{\theta}(\mathbf{Z}) = \mathtt{median}(\mathbf{Z}) = 10.325$$

The estimates are fairly close to each other.

# General Estimator Properties

**Lemma 1:** With $Z_i \sim L_0(z-\theta),\ i=1,\ldots,N$, the distributions of $\bar{\theta}(\mathbf{Z})-\theta$, $\tilde{\theta}(\mathbf{Z})-\theta$ and $\hat{\theta}(\mathbf{Z})-\theta$ are independent of $\theta$.

**Proof:** The distribution of $Z_i - \theta \sim L_0$ is independent of $\theta$.

$$\Longrightarrow \quad \bar{\theta}(\mathbf{Z}-\theta) = \bar{\theta}(\mathbf{Z})-\theta, \quad \tilde{\theta}(\mathbf{Z}-\theta) = \tilde{\theta}(\mathbf{Z})-\theta \quad \text{and} \quad \hat{\theta}(\mathbf{Z}-\theta) = \hat{\theta}(\mathbf{Z})-\theta \ \square$$

**Theorem 2:** If the distribution $L$ of $Z_1,\ldots,Z_N$ is symmetric around $\theta$, the same is true of the distributions of $\hat{\theta}$, $\tilde{\theta}$ and $\bar{\theta}$.

**Proof:** Symmetry $\Longrightarrow Z_i - \theta \stackrel{\mathcal{D}}{=} \theta - Z_i = -(Z_i-\theta) \sim L_0$

$$\hat{\theta}(\mathbf{Z})-\theta = \hat{\theta}(\mathbf{Z}-\theta) \stackrel{\mathcal{D}}{=} \hat{\theta}(\theta-\mathbf{Z}) = \hat{\theta}(-(\mathbf{Z}-\theta)) = -\hat{\theta}(\mathbf{Z}-\theta) = -(\hat{\theta}(\mathbf{Z})-\theta) \qquad \square$$

The same argument works for $\bar{\theta}$ and $\tilde{\theta}$.

# Unbiasedness Properties

When $L = L_\theta(z) = L_0(z-\theta)$ is symmetric around $\theta$ and $Z_i \sim L_\theta,\ i = 1,\ldots,N$, then $\hat{\theta}$, $\tilde{\theta}$ and $\bar{\theta}$ are unbiased estimators of $\theta$, i.e.,

$$E_\theta\left(\hat{\theta}(\mathbf{Z})\right) = E_\theta\left(\tilde{\theta}(\mathbf{Z})\right) = E_\theta\left(\bar{\theta}(\mathbf{Z})\right) = \theta$$

provided these expectations exist.

Under the same conditions we also have

$$P_\theta(\hat{\theta}(\mathbf{Z}) < \theta) = P_\theta(\hat{\theta}(\mathbf{Z}) > \theta) \qquad \text{and} \qquad P_\theta(\hat{\theta}(\mathbf{Z}) \le \theta) = P_\theta(\hat{\theta}(\mathbf{Z}) \ge \theta)$$

and correspondingly with $\hat{\theta}(\mathbf{Z})$ replaced by $\tilde{\theta}(\mathbf{Z})$ or $\bar{\theta}(\mathbf{Z})$.

This means that these estimators are median unbiased, i.e., as likely to overestimate their target $\theta$ as to underestimate it.

# Continuity Assumption

Unless otherwise indicated we assume that $L$ is a continuous distribution function.

This implies $P_L(Z_i = c) = 0$ for any $c \in R$.

It also implies that the distributions of $\hat{\theta}(\mathbf{Z})$, $\tilde{\theta}(\mathbf{Z})$, and $\bar{\theta}(\mathbf{Z})$ are continuous,i.e.,

$$P_L(\hat{\theta}(\mathbf{Z}) = x) = P_L(\tilde{\theta}(\mathbf{Z}) = x) = P_L(\bar{\theta}(\mathbf{Z}) = x) = 0$$

for any $x \in R$.

The proof is technical but not hard (skipped here).

# Dispersion of $\hat{\theta}(\mathbf{Z})$, $\tilde{\theta}(\mathbf{Z})$, and $\bar{\theta}(\mathbf{Z})$

Although the variance of $\bar{\theta}(\mathbf{Z})$ is easily shown to be $\sigma_Z^2/N$ with $\sigma_Z^2 = \mathrm{var}(Z_i)$, it is analytically difficult to get the variances of the other two estimators.

As in the case of assessing dispersion of our treatment effect estimator $\hat{\Delta}$ in Chapter 2 we will thus measure the dispersion of our estimators in terms of the probability of our estimator being sufficiently close to their target $\theta$, i.e., in terms of $\hat{\theta}(\mathbf{Z})$ by the probability

$$P_\theta(|\hat{\theta}(\mathbf{Z}) - \theta| \le a)$$

We will again associate such probabilities with the power of the respective tests.

# Order Statistics Inequalities

**Theorem 3:** (i) If $Z_{(1)} < \ldots < Z_{(N)}$ denote the ordered observations (order statistics) of the sample $Z_1, \ldots, Z_N$, then for any $1 \leq i \leq N$ and $a \in R$

$$Z_{(i)} \leq a \qquad \Longleftrightarrow \qquad S_N(\mathbf{Z} - a) \leq N - i$$

or equivalently

$$Z_{(i)} > a \qquad \Longleftrightarrow \qquad S_N(\mathbf{Z} - a) \geq N - i + 1$$

(ii) Let $M = N(N+1)/2$ and denote by $A_{(1)} < \ldots < A_{(M)}$ the ordered averages $(Z_i + Z_j)/2$. Then for any $1 \leq i \leq M$ and $a \in R$

$$A_{(i)} \leq a \qquad \Longleftrightarrow \qquad V_s(\mathbf{Z} - a) \leq M - i$$

or equivalently

$$A_{(i)} > a \qquad \Longleftrightarrow \qquad V_s(\mathbf{Z} - a) \geq M - i + 1$$

**Proof:** $Z_{(i)} \leq a \Leftrightarrow$ at least $i$ of the $Z$'s are $\leq a \Leftrightarrow$ at most $N - i$ of the $Z$'s are $> a$, i.e., at most $N - i$ of $Z_i - a > 0$ or $S_N(\mathbf{Z} - a) \leq N - i$. Same argument for (ii).

# $P_\theta(|\tilde{\theta}(\mathbf{Z})-\theta|\le a)$

For $L_0$ continuous and symmetric around zero we have

$$
\begin{aligned}
P_\theta(|\tilde{\theta}(\mathbf{Z})-\theta|\le a) &= P_0(|\tilde{\theta}(\mathbf{Z})|\le a) = P_0(\tilde{\theta}(\mathbf{Z})\le a) - P_0(\tilde{\theta}(\mathbf{Z}) < -a) \\
&= P_0(\tilde{\theta}(\mathbf{Z})\le a) - \left[1 - P_0(\tilde{\theta}(\mathbf{Z}) \ge -a)\right] \\
&= P_0(\tilde{\theta}(\mathbf{Z})\le a) - \left[1 - P_0(\tilde{\theta}(\mathbf{Z}) \le a)\right] \\
&= 2P_0(\tilde{\theta}(\mathbf{Z})\le a) - 1 \stackrel{*}{=} 2P_0(S_N(\mathbf{Z}-a)\le N-k-1) - 1
\end{aligned}
$$

where by Theorem 3 $\stackrel{*}{=}$ holds for odd $N=2k+1$. For even $N=2k$ we have

$$2P_0(S_N(\mathbf{Z}-a)\le N-k-1)-1 \le P_\theta(|\tilde{\theta}(\mathbf{Z})-\theta|\le a) \le 2P_0(S_N(\mathbf{Z}-a)\le N-k)-1$$

$S_N(\mathbf{Z}-a) \sim$ Binomial$(N, p(a))$ with $p(a) = P_0(Z-a>0) = 1-L_0(a)$.

Thus the dispersion probability $P_\theta(|\tilde{\theta}(\mathbf{Z})-\theta|\le a)$ is easily calculated or bracketed using `pbinom` for any given $L_0$ and $a$.

# $P_\theta(|\hat\theta(\mathbf{Z}) - \theta| \le a)$

Again using Theorem 3 the same sequence of steps for odd $M = 2k+1$ leads to

$$P_\theta(|\hat\theta(\mathbf{Z}) - \theta| \le a) = 2P_0(V_s(\mathbf{Z} - a) \le M - k - 1) - 1$$

and for even $M = 2k$ we can bracket it as follows

$$2P_0(V_s(\mathbf{Z}-a) \le M-k-1)-1 \le P_\theta(|\hat\theta(\mathbf{Z})-\theta| \le a) \le 2P_0(V_s(\mathbf{Z}-a) \le M-k)-1$$

Evaluating these probabilities $P_0(V_s(\mathbf{Z} - a) \le \ell)$ when $Z_i \sim L_0$ involves the distribution of the Wilcoxon statistic under shift alternatives.

For given $a$, $L_0$ and $\ell$ these are easily evaluated via simulation.

Simply generate vectors $\mathbf{Z}$ from $L_0$, evaluate $V_s(\mathbf{Z} - a)$ for each shifted sample $\mathbf{Z} - a$ and observe the proportion of such $V_s$ values that are $\le \ell$.

# Large Sample Approximation for Odd $M = 2k+1$

$$
\begin{aligned}
P_\theta(\hat\theta(\mathbf{Z}) - \theta \le a) &= P_0(\hat\theta(\mathbf{Z}) \le a) = P_0(V_s(\mathbf{Z}-a) \le M-k-1) = P_0(\hat\theta(\mathbf{Z}) \le a) \\
&= P_0(V_s(\mathbf{Z}-a) \le k) = \Phi\left(\frac{k+\frac{1}{2} - E_0(V_s(\mathbf{Z}-a))}{\sqrt{\mathrm{var}_0(V_s(\mathbf{Z}-a))}}\right) \\
&= 1 - \Phi\left(\frac{\left(p_1' - \frac{1}{2}\right)\frac{N(N-1)}{2} + N\left(p - \frac{1}{2}\right)}{\sqrt{\mathrm{var}_0(V_s(\mathbf{Z}-a))}}\right) \\
P_\theta\left(\left|\hat\theta(\mathbf{Z}) - \theta\right| \le a\right) &= P_0(\hat\theta(\mathbf{Z}) \le a) - P_0(\hat\theta(\mathbf{Z}) < -a) \\
&= P_0(\hat\theta(\mathbf{Z}) \le a) - P_0(\hat\theta(\mathbf{Z}) > a) = 2P_0(\hat\theta(\mathbf{Z}) \le a) - 1 \\
&= 1 - 2\Phi\left(\frac{\left(p_1' - \frac{1}{2}\right)\frac{N(N-1)}{2} + N\left(p - \frac{1}{2}\right)}{\sqrt{\mathrm{var}_0(V_s(\mathbf{Z}-a))}}\right)
\end{aligned}
$$

$$p = P_0(Z - a > 0) = P_0(Z > a)\,,\ \ p_1' = P_0((Z_1 - a) + (Z_2 - a) > 0) = P_0((Z_1 + Z_2)/2 > a).$$

# $\mathrm{var}_0(V_s(\mathbf{Z}-a))$

$$\mathrm{var}_0(V_s(\mathbf{Z}-a)) = N(N-1)(N-2)(p_2' - p_1'^2) + \frac{N(N-1)}{2}\left[2(p-p_1')^2 + 3p_1'(1-p_1')\right] + Np(1-p)$$

where

$$p_2' = P_L((Z_1+Z_2)/2 > a,\ (Z_1+Z_3)/2 > a)$$

As in the case of the Wilcoxon rank-sum test statistic $W_s$ the above approximation suggests itself also in the case of $M$ even.

# A Further Approximation for Small $a$

As before one can take a further, simplifying approximation step for small $a$, namely

$$p - \frac{1}{2} \approx -a\ell_0(0) \qquad \text{and} \qquad p_1' - \frac{1}{2} \approx -2a\ell_0^*(0)$$

and

$$\mathrm{var}_0(V_s(\mathbf{Z}-a)) \approx \mathrm{var}_0(V_s(\mathbf{Z})) = \frac{N(N+1)(2N+1)}{24}$$

$$\Longrightarrow P_\theta\left(\left|\hat{\theta}(\mathbf{Z}) - \theta\right| \le a\right) \approx 2\Phi\left(\sqrt{\frac{24N}{(N+1)(2N+1)}}\left[(N-1)\ell_0^*(0) + \ell_0(0)\right]\right) - 1$$

When $\ell_0$ is normal with mean zero and variance $\tau^2$ we have

$$\ell_0(0) = \frac{1}{\tau\sqrt{2\pi}} \qquad \text{and} \qquad \ell_0^*(0) = \frac{1}{2\tau\sqrt{\pi}}$$

and the above approximation reduces to

$$P_\theta\left(\left|\hat{\theta}(\mathbf{Z}) - \theta\right| \le a\right) \approx 2\Phi\left(\sqrt{\frac{6N}{(N+1)(2N+1)\pi\tau^2}}\left[(N-1) + \sqrt{2}\right]\right) - 1$$

# Efficiency Carry Over

Because of the relationship between $P_\theta(|\tilde{\theta}-\theta| \le a)$ and $P_\theta(|\hat{\theta}-\theta| \le a)$ and the corresponding power of sign and Wilcoxon tests, and similarly between $P_\theta(|\bar{\theta}-\theta| \le a)$ and the asymptotic power of the $t$-test it will not surprise that all the efficiency results derived for the tests carry over to the estimators.

Here the efficiency of estimators is defined as the ratio of sample sizes required by the three estimators to match the respective dispersion probabilities

$$P_\theta(|\tilde{\theta}-\theta| \le a), \qquad P_\theta(|\hat{\theta}-\theta| \le a) \qquad \text{and} \qquad P_\theta(|\bar{\theta}-\theta| \le a).$$

For example, $\bar{\theta}$ needs about $N_t = N_{\bar{\theta}} = 3/\pi\, N_{\hat{\theta}} = .955 N_{\hat{\theta}} = .955\, N_V$ to match the dispersion probabilities

$$P_\theta(|\hat{\theta}_{N_t}-\theta| \le a) = P_\theta(|\bar{\theta}_{N_V}-\theta| \le a) \qquad \text{for any } a > 0.$$

# Dropping the Symmetry Assumption

When we drop the assumption that $L_0$ be symmetric around zero, the three estimators may no longer target the same value and it is less compelling to compare their performance in terms of their dispersion properties.

**Theorem 4:** (i) If $L$ has finite expectation $\lambda$ then $\bar{\theta}$ is a consistent estimator of $\lambda$.

(ii) If $L$ has a unique median $\mu$ then $\tilde{\theta}$ is a consistent estimator of $\mu$.

(iii) If $\nu$ is the only point with the property

$$P\left(\frac{Z_1+Z_2}{2} < \nu\right) = P\left(\frac{Z_1+Z_2}{2} > \nu\right) = \frac{1}{2},$$

where $Z_1, Z_2$ i.i.d. $\sim L$, then $\hat{\theta}$ is a consistent estimator of $\nu$.

A generic estimator $\hat{\theta}_N$ is a consistent estimator of $\theta$ if for any $\varepsilon > 0$

$$P(|\hat{\theta}_N - \theta| > \varepsilon) \longrightarrow 0 \qquad \text{as} \qquad N \longrightarrow \infty.$$

# Discussion of $\lambda$, $\mu$ and $\nu$

The parameters $\lambda$ and $\mu$: simple definitions and clear, intuitive meanings.

The parameter $\nu$ is more complex, not so intuitive.

$\lambda$ can change dramatically under minute changes in the distribution $L$.

Minute changes $=$ small shifts of probabilities.

With respect to $\mu$ we can move almost half the probability of the distribution to $\infty$ and the location of $\mu$ is still determined by the inertia of the $> 50\%$ majority.

The parameter $\nu$ is also quite resistant to large moves of probabilities to $\infty$ as long as the total moved probability stays $< 1 - 1/\sqrt{2} = .293$.

# Robustness Properties of $\bar{\theta}$, $\tilde{\theta}$ and $\hat{\theta}$

The distributional properties from the previous slide carry over to the estimators.

The sample mean or average will move to $\infty$ as soon as just a single observation moves to $\infty$ while the other observations stay put.

The sample median $\tilde{\theta}$ will remain finite as long as we move less than half of the observations off to $\infty$ while the other observations stay put.

The Hodges-Lehmann estimator $\hat{\theta}$ will remain finite as long as we move less than $29.3\%$ of the observations off to $\infty$ while the other observations stay put.

The same percentage was found for the Hodges-Lehmann shift estimator $\hat{\Delta}$.

# Explanation of $29.3\%$

We need to move less than $N(N+1)/4$ of the $M = N(N+1)/2$ averages $(Z_i + Z_j)/2$, $i \leq j$ to ∞ by moving $k$ of the $Z_i$ to ∞ while the others stay put.

If we move $Z_1$ to ∞ we move $N$ averages $(Z_1 + Z_j)/2$ with $1 \leq j$ to ∞.

If we move $Z_2$ to ∞ we move $N-1$ averages $(Z_2 + Z_j)/2$ with $2 \leq j$ to ∞.

…

If we move $Z_k$ to ∞ we move $N-k+1$ averages $(Z_k + Z_j)/2$ with $k \leq j$ to ∞.

$\Longrightarrow$ we need $\quad N + (N-1) + \ldots + (N-k+1) = kN - \dfrac{k(k-1)}{2} < \dfrac{N(N+1)}{4}$

This leads to a limiting quadratic inequality in $0 < x = k/N < 1$

$$4x - 2x^2 < 1 \Rightarrow 0 < (x-1)^2 - \frac{1}{2} = \left(x - 1 - \frac{1}{\sqrt{2}}\right)\left(x - 1 + \frac{1}{\sqrt{2}}\right) \Rightarrow x < 1 - \frac{1}{\sqrt{2}} = .293$$

# Confidence Interval Building Blocks

**Theorem 5:** (i) Let $Z_1,\ldots,Z_N$ i.i.d. $\sim L$ (a continuous cdf). If $\mu$ is any median of $L$

$$\Longrightarrow \quad P(Z_{(k)} < \mu \le Z_{(k+1)}) = P_0(S_N = k) = \frac{\binom{N}{k}}{2^N} \qquad \text{for all } k = 0,1,2,\ldots,N$$

with the convention $Z_{(0)} = -\infty$ and $Z_{(N+1)} = \infty$. Here $P_0 \sim \mathrm{Binomial}(N, p = 1/2)$, the null distribution of $S_N$. The median $\mu$ does not need to be unique.

(ii) Assume in addition that $L$ is symmetric around some $\theta$, which would then also serve as a median. Denote the $M = N(N+1)/2$ averages $(Z_i + Z_j)/2,\ i \le j$ in sorted order by $A_{(1)} < \ldots < A_{(M)}$.

$$\Longrightarrow \quad P_\theta(A_{(k)} < \theta \le A_{(k+1)}) = P_0(V_s = k) \qquad \text{for all } k = 0,1,2,\ldots,M$$

with the convention $A_{(0)} = -\infty$ and $A_{(M+1)} = \infty$.
Here $P_0(V_s = k)$ refers to the null distribution of the signed-rank statistic.

Because of the continuity of $L$ the $<$ inside the coverage probability can be replaced by $\le$ and vice versa, without changing the coverage probability.

# Proof of Theorem 5

(i) $Z_{(i)} < \mu \le Z_{(i+1)}$ means that exactly $i$ of the $Z_1, \ldots, Z_N$ are $< \mu$.

Each comparison $Z_j < \mu$ is one of $N$ independent Bernoulli trials,

each with probability $1/2 \;\Rightarrow\; P(Z_{(i)} < \mu \le Z_{(i+1)}) = P_0(S_N = i)$. $\square$

(ii) Since $Z_i - \theta \sim L_0$, symmetric around zero, we may as well assume $\theta = 0$.

$$V_s^* = \sum_{i \le j} I_{[(Z_i + Z_j)/2 < 0]} \overset{\mathcal{D}}{=} \sum_{i \le j} I_{[(Z_i + Z_j)/2 > 0]} = V_s$$

$$\Longrightarrow \qquad A_{(k)} < 0 \le A_{(k+1)} \quad \Longleftrightarrow \quad V_s^* = k$$

$$\Longrightarrow \qquad P_0(A_{(k)} < 0 \le A_{(k+1)}) \;=\; P_0(V_s^* = k) = P_0(V_s = k) \qquad \square$$

# Confidence Intervals Based on $Z_{(i)}$

By concatenating adjacent building block intervals $(Z_{(\ell)}, Z_{(\ell+1)}]$ with coverage probabilities $P_0(S_N = \ell)$, $\ell = 1, 2, \ldots$, we get

$$P(\mu \le Z_{(i)}) = \sum_{\ell=0}^{i-1} P(\mu \in (Z_{(\ell)}, Z_{(\ell+1)}]) = \sum_{\ell=0}^{i-1} P_0(S_N = \ell) = P_0(S_N \le i-1) = p_{i-1}$$

$$\begin{aligned} \Longrightarrow \quad P(Z_{(i)} < \mu \le Z_{(k)}) &= P(\mu \le Z_{(k)}) - P(\mu \le Z_{(i)}) \\ &= p_{k-1} - p_{i-1} = P_0(i \le S_N \le k-1) \\ &= 1 - P_0(S_N \le i-1) - P_0(S_N \ge k) \\ &=^* 1 - 2P_0(S_N \le i-1) = 1 - 2p_{i-1} \end{aligned}$$

where in $=^*$ we choose $k = N - (i-1)$ and exploit the symmetry of the $S_N$ null distribution around $N/2$, i.e., $P_0(S_N \le i-1) = P_0(S_N \ge N - (i-1))$.

$[Z_{(i)}, Z_{(N-i+1)}]$ is a confidence interval for the median $\mu$ with coverage probability $\gamma_i = 1 - 2p_{i-1}$ for any continuous $L$. Very distribution-free or nonparametric!!

# Discrete Confidence Levels $\gamma_i$

Given a desired confidence level $\gamma$ how do we find the smallest $\gamma_i \geq \gamma$ and the corresponding $i$, leading to the resulting interval $[Z_{(i)}, Z_{(N-i+1)}]$?

We want the largest $p_{i-1}$ or largest $i = i_0$ such that

$$1 - 2p_{i-1} = \gamma_i \geq \gamma \qquad \text{or} \qquad p_{i-1} \leq \frac{1-\gamma}{2}$$

$$i_0 = \mathtt{qbinom}((1-\gamma)/2, \mathtt{N}, .5) = \min\left(i:\ \mathtt{pbinom}(\mathtt{i}, \mathtt{N}, .5) = P_0(S_N \leq i) \geq \frac{1-\gamma}{2}\right)$$

$$\Longrightarrow \quad \mathtt{pbinom}(\mathtt{i}_0 - 1, \mathtt{N}, .5) < \frac{1-\gamma}{2}$$

If $\mathtt{pbinom}(\mathtt{i}_0, \mathtt{N}, .5) > (1-\gamma)/2$ then $\mathtt{i}_0 = \mathtt{qbinom}((1-\gamma)/2, \mathtt{N}, .5)$ is it.

If $\mathtt{pbinom}(\mathtt{i}_0, \mathtt{N}, .5) = (1-\gamma)/2$ then $\mathtt{i}_0 = \mathtt{qbinom}((1-\gamma)/2, \mathtt{N}, .5) + 1$ is it.

In either case $[Z_{(i_0)}, Z_{(N-i_0+1)}]$ is the interval $[Z_{(i)}, Z_{(N-i+1)}]$ with lowest confidence level $\gamma_i \geq \gamma$, namely with $\gamma_i = \gamma_{i_0} = 1 - 2 * \mathtt{pbinom}(\mathtt{i}_0 - 1, \mathtt{N}, .5) > \gamma$ in the first case and $= \gamma$ in the second case.

# Confidence Intervals Based on $A_{(i)}$

By concatenating adjacent building block intervals $(A_{(\ell)}, A_{(\ell+1)}]$

with coverage probabilities $P_0(V_s = \ell)$, $\ell = 1, 2, \ldots$, we get

$$P(\theta \le A_{(i)}) = \sum_{\ell=0}^{i-1} P(\theta \in (A_{(\ell)}, A_{(\ell+1)}]) = \sum_{\ell=0}^{i-1} P_0(V_s = \ell) = P_0(V_s \le i-1) = \tilde{p}_{i-1}$$

$$\begin{aligned}
\Longrightarrow \quad P(A_{(i)} < \theta \le A_{(k)}) &= P(\theta \le A_{(k)}) - P(\theta \le A_{(i)}) \\
&= \tilde{p}_{k-1} - \tilde{p}_{i-1} = P_0(i \le V_s \le k-1) \\
&=^* 1 - P_0(V_s \le i-1) - P_0(V_s \ge k) \\
&= 1 - 2P_0(V_s \le i-1) = 1 - 2\tilde{p}_{i-1}
\end{aligned}$$

where in $=^*$ we choose $k = M - i + 1$ and exploit the symmetry of the $V_s$

null distribution around $M/2$.

Thus $[A_{(i)}, A_{(M-i+1)}]$ is a confidence interval for the center of symmetry $\theta$ with

coverage probability $\tilde{\gamma}_i = 1 - 2\tilde{p}_{i-1}$ for any continuous $L$ symmetric around $\theta$.

Again, very distribution-free or nonparametric!!

# Discrete Confidence Levels $\tilde{\gamma}_i$

Given a desired confidence level $\gamma$ how do we find the smallest $\tilde{\gamma}_i \geq \gamma$ and the corresponding $i$, leading to the resulting interval $[A_{(i)}, A_{(M-i+1)}]$?

We want the largest $\tilde{p}_{i-1}$ or largest $i = i_0$ such that

$$1 - 2\tilde{p}_{i-1} = \gamma_i \geq \gamma \qquad \text{or} \qquad \tilde{p}_{i-1} \leq \frac{1-\gamma}{2}$$

$$i_0 = \mathtt{qsignrank}((1-\gamma)/2, \mathtt{N}) = \min\left(i : \mathtt{psignrank}(\mathtt{i}, \mathtt{N}) = P_0(S_N \leq i) \geq \frac{1-\gamma}{2}\right)$$

$$\Longrightarrow \quad \mathtt{psignrank}(\mathtt{i}_0 - 1, \mathtt{N}) < \frac{1-\gamma}{2}$$

If $\mathtt{psignrank}(\mathtt{i}_0, \mathtt{N}) > (1-\gamma)/2$ then $\mathtt{i}_0 = \mathtt{qsignrank}((1-\gamma)/2, \mathtt{N})$ is it.

If $\mathtt{psignrank}(\mathtt{i}_0, \mathtt{N}) = (1-\gamma)/2$ then $\mathtt{i}_0 = \mathtt{qsignrank}((1-\gamma)/2, \mathtt{N}) + 1$ is it.

In either case $[A_{(i_0)}, A_{(M-i_0+1)}]$ is the interval $[A_{(i)}, A_{(M-i+1)}]$ with lowest confidence level $\tilde{\gamma}_i \geq \gamma$, namely with $\tilde{\gamma}_i = \tilde{\gamma}_{i_0} = 1 - 2 * \mathtt{psignrank}(\mathtt{i}_0 - 1, \mathtt{N}) > \gamma$ in the first case and $= \gamma$ in the second case.

# Example 4: Effect of Muscle Training

12 first-graders were measured with respect to weight lifting capability before and after an 8-week muscle training program. The after $-$ before differences were recorded as

$$\mathtt{Z} = \mathtt{c}(6.0, 7.0, 5.0, 10.5, 8.5, 3.5, 6.1, 4.0, 4.6, 4.5, 5.9, 6.5)$$

To get the vector $\mathtt{A.vec}$ of ordered values $A_{(1)} \leq \ldots \leq A_{(M)}$ of the $M$ averages $(Z_i + Z_j)/2$, $i \leq j$, we proceed as follows

$$\mathtt{A} = \mathtt{outer}(\mathtt{Z}, \mathtt{Z}, \text{"}+\text{"})/2 \quad \Longrightarrow \quad \mathtt{A.vec} = \mathtt{sort}(\mathtt{A}[\mathtt{row}(\mathtt{A}) <= \mathtt{col}(\mathtt{A})])$$

For $\gamma = .95$: $\mathtt{qsignrank}(.025, 12) = 14$; $\mathtt{psignrank}(14, 12) = 0.0261 > .025$

$$\Longrightarrow \quad \mathtt{i}_0 = 14 \quad \text{and} \quad \mathtt{A.vec}[14] = 4.75\,, \;\; \mathtt{A.vec}[12 * 13/2 - 14 + 1] = 7.3$$

with achieved confidence level $= \tilde{\gamma}_{i_0} = 1 - 2 * \mathtt{psignrank}(\mathtt{i}_0 - 1, \mathtt{N}) = 0.9575195$

and with Hodges-Lehmann estimate $\quad \mathrm{med}_{i \leq j} \left( \dfrac{Z_i + Z_j}{2} \right) = \mathtt{median}(\mathtt{A.vec}) = 5.85$

# Example 4: R Function `MuscleGain`

```
MuscleGain=function(gamma=.9){
   weight.diff=c(6, 7, 5, 10.5, 8.5, 3.5, 6.1,
                              4, 4.6, 4.5, 5.9, 6.5)
   N=length(weight.diff); M=N*(N+1)/2
   A=outer(weight.diff,weight.diff,"+")/2
   A.vec=sort(A[row(A)<=col(A)])
   i0=qsignrank((1-gamma)/2,N)
   if(psignrank(i0,N)==(1-gamma)/2) i0=i0+1
   index=c(i0,M-i0+1); names(index)=c("i0","M-i0+1")
   Interval=c(A.vec[i0],A.vec[M-i0+1])
   names(Interval)=c("Lower","Upper")
   estimate=median(A.vec); names(estimate)="HL-estimate"
   list(estimate=estimate,index=index,Interval=Interval,
       achieved.confidence=1-2*psignrank(i0-1,N))
}
```

# MuscleGain(.9)

```
$estimate
HL-estimate
       5.85

$index
    i0 M-i0+1
    18     61

$Interval
Lower Upper
 4.95  7.00

$achieved.confidence
[1] 0.9077148
```

The Text erroneously has $[9.9, 14] = [2 \times 4.95, 2 \times 7]$ as $90\%$ interval, omitting the division by 2. Manual calculation was easier with $Z_i + Z_j$ than with $(Z_i + Z_j)/2$.

# wilcox.test

```
> wilcox.test(weight.diff,conf.int=T,exact=T,conf.level=.9)

Wilcoxon signed rank test

data:  weight.diff
V = 78, p-value = 0.0004883
alternative hypothesis: true location is not equal to 0
90 percent confidence interval:
 4.95 7.00
sample estimates:
(pseudo)median
          5.85
```

# Without Assuming Symmetry

Without assuming symmetry in the muscle gain example, we still can get confidence intervals for the median $\mu$ using $[Z_{(i_0)}, Z_{(N-i_0+1)}]$ for appropriate $i_0$.

This is implemented in the function `MuscleGain.median` on the next slide.

Achievable confidence levels are much coarser than in the previous case.

The distribution of $S_N$ has just $N+1$ possible values: $0, 1, \ldots, N$.

The distribution of $V_s$ has $N(N+1)/2+1$ possible values: $0, 1, \ldots, N(N+1)/2$,

$\Longrightarrow$ a much finer distributional graduation.

# MuscleGain.median

```
MuscleGain.median=function(gamma=.9){
   weight.diff=c(6, 7, 5, 10.5, 8.5, 3.5, 6.1,
                         4, 4.6, 4.5, 5.9, 6.5)
   N=length(weight.diff)
   Z.vec=sort(Z)
   i0=qbinom((1-gamma)/2,N,.5)
   if(pbinom(i0,N,.5)==(1-gamma)/2) i0=i0+1
   index=c(i0,N-i0+1); names(index)=c("i0","N-i0+1")
   Interval=c(Z.vec[i0],Z.vec[N-i0+1])
   names(Interval)=c("Lower","Upper")
   estimate=median(Z); names(estimate)="median"
   list(estimate=estimate,index=index,Interval=Interval,
       achieved.confidence=1-2*pbinom(i0-1,N,.5))
}
```

# MuscleGain.median(.9)

```
> MuscleGain.median(.9)
$estimate
median
  5.95  # not too different from 5.85 previously

$index
    i0 N-i0+1
     3      10

$Interval
Lower Upper
  4.5   7.0

$achieved.confidence
[1] 0.9614258
```

Note the more conservative achieved confidence of $.96$ over $.91$ previously.

Only the lower end point changed from $4.95$ to $4.5$.

# Normal Approximations

For large $N$ normal approximations with continuity correction may be useful

$$P(\mu \le Z_{(i)}) = P_0(S_N \le i-1) \approx \Phi\left(\frac{i-1+1/2-N/2}{\sqrt{N/4}}\right) = \Phi\left(\frac{2i-(N+1)}{\sqrt{N}}\right)$$

$$P(\theta \le A_{(i)}) = P_0(V_s \le i-1) \approx \Phi\left(\frac{i-1+1/2-\frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}\right) = \Phi\left(\frac{2i-1-\frac{N(N+1)}{2}}{\sqrt{\frac{N(N+1)(2N+1)}{6}}}\right)$$

These could then be used to approximate $i$ for prescribed coverage probability.

Given current state of software this is no longer so relevant.

# Confidence Interval Building Blocks for $z_p$

Our previous confidence intervals and bounds for the median $\mu$ are easily generalized to any quantile $z_p$ with $P(Z \le z_p) = L(z_p) = p$ for continuous $L$.

(The Text focusses on $\mu_p = z_{1-p}$. I don't quite understand why.)

$$P(Z_{(i)} < z_p \le Z_{(i+1)}) = P_p(S_N = i) = \binom{N}{i} p^i (1-p)^{N-i}$$

Here $P_p$ indicates that $S_N$ has a Binomial$(N, p)$-distribution.

$Z_{(i)} < z_p \le Z_{(i+1)}$ means that exactly $i$ of the independent $Z_1, \ldots, Z_N$ are $< z_p$, each with success probability $p = P(Z < z_p) = P(Z \le z_p)$ by continuity of $L$.

The binomial distribution gives the above formula.

# Confidence Intervals for $z_p$

As before, concatenation of such building block intervals easily leads to intervals

$$
\begin{aligned}
P(z_p \le Z_{(\ell)}) &= \sum_{i=0}^{\ell-1} P(Z_{(i)} < z_p \le Z_{(i+1)}) = P_p(S_N \le \ell - 1) \\
P(Z_{(k)} < z_p \le Z_{(\ell)}) &= P(z_p \le Z_{(\ell)}) - P(z_p \le Z_{(k)}) \\
&= P_p(S_N \le \ell - 1) - P_p(S_N \le k - 1) = P_p(k \le S_N \le \ell - 1)
\end{aligned}
$$

Since the distribution of $S_N$ is no longer symmetric when $p \ne 1/2$ there is no obvious natural relationship between $k$ and $\ell$ for a given desired confidence level.

However, one could still choose $k$ largest and $\ell$ smallest such that

$$
P_p(S_N \le k-1) \le \frac{1-\gamma}{2} \quad \text{and} \quad P_p(S_N \ge \ell) \le \frac{1-\gamma}{2} \quad \text{or} \quad P_p(S_N \le \ell - 1) \ge \frac{1+\gamma}{2}
$$

$$
\Longrightarrow \quad P(Z_{(k)} < z_p \le Z_{(\ell)}) = P_p(k \le S_N \le \ell - 1) \ge \frac{1+\gamma}{2} - \frac{1-\gamma}{2} = \gamma
$$

# Practical Determination of $k$ and $\ell$

Finding $\mathtt{k}$:

Let $\mathtt{k}_0 = \mathtt{qbinom}((1-\gamma)/2, \mathtt{N}, \mathtt{p})$ = smallest $\mathtt{i}$ such that $\mathtt{pbinom}(\mathtt{i}, \mathtt{N}, \mathtt{p}) \geq (1-\gamma)/2$,

then $\mathtt{pbinom}(\mathtt{k}_0 - 1, \mathtt{N}, \mathtt{p}) < (1-\gamma)/2$.

If $\mathtt{pbinom}(\mathtt{k}_0, \mathtt{N}, \mathtt{p}) > (1-\gamma)/2$ then $\mathtt{k} = \mathtt{qbinom}((1-\gamma)/2, \mathtt{N}, \mathtt{p})$ is it.

If $\mathtt{pbinom}(\mathtt{k}_0, \mathtt{N}, \mathtt{p}) = (1-\gamma)/2$ then $\mathtt{k} = \mathtt{qbinom}((1-\gamma)/2, \mathtt{N}, \mathtt{p}) + 1$ is it.

Finding $\ell$:

Let $\ell_0 = \mathtt{qbinom}((1+\gamma)/2, \mathtt{N}, \mathtt{p})$ = smallest $\mathtt{i}$ such that $\mathtt{pbinom}(\mathtt{i}, \mathtt{N}, \mathtt{p}) \geq (1+\gamma)/2$

Then $\ell = \ell_0 + 1$ is it.