C. The Need for Distribution-Free Tests

1. Table contains actual test levels for some tests of location parameters for four of the families described in §.

   a. True levels were determined via simulation.

   b. A large number of samples were drawn from each of the distributions under the $H_0$, the specified test statistic was calculated, the test of § was performed for each simulated data set, and the proportion of times the null hypothesis was rejected was tabulated.

   c. For now, restrict attention to the first line in each subtable, corresponding to the $t$-test.

   d. Null hypotheses in Table are in terms of the distribution median.

   e. The $t$-test, however, is appropriate for hypotheses involving the expectation.

   f. In the Gaussian, Laplace, and uniform cases, the median coincides with the expectation, and so standard asymptotic theory justifies the use of the $t$-test.

   g. In the Cauchy example, as noted before, even though the
distribution is symmetric, no expectation exists, and the $t$-test is inappropriate.

h. However, generally, data analysts do not have sufficient information to distinguish the Cauchy example from the set of distributions having enough moments to justify the $t$-test, and so it is important to study the implications of such an inappropriate use of methodology.

2. For both sample sizes, observations from a Gaussian distribution give the targeted level, as expected.
   a. Observations from the Laplace distribution give a level close to the targeted level.
   b. Observations from the Cauchy distribution give a level much smaller than the targeted level,
   c. Figure shows the density resulting from Studentizing the average of independent Cauchy variables.
   d. The resulting density is bimodal, with tails lighter than one would otherwise expect.
   e. This shows that larger values of the sample standard deviation
in the denominator of the Studentized statistic act more strongly than larger values of components of the average in the numerator.

3. In all cases above, the $t$-test succeeds in providing a test level not much larger than the target nominal level.

   a. On the other hand, in some cases the true level is significantly below that expected.

   b. This effect decreases as sample level increases.
D. One-Sample Median Methods

1. For moderate sample sizes, then, the standard one-sample $t$-test fails to control test level as the distribution of summands changes.
   a. Techniques that avoid this problem are developed in this section.
   b. These methods apply in broad generality, including in cases when the expectation of the individual observations does not exist.
   c. Because of this, inference about the population median rather than the expectation is pursued.
   d. Recall that the median $\theta$ of random variable $X_j$ is defined so that
   e. Below, the term median refers to the population version, unless otherwise specified.

2. Estimates of the Population Median
   a. An estimator $\text{smed} [X_1, \ldots, X_n]$ of the population median may be constructed by applying to the empirical distribution of $X_i$, formed by putting point mass on each of the $n$ values.
      i. In this case, with $n$ odd, the median is the middle value, and, with $n$ even, fails to uniquely define the estimator.
ii. In this case, the estimator is conventionally defined to be the average of the middle two values.

iii. By this convention, with $X_{(1)}, \ldots, X_{(n)}$ the ordered values in the sample,

$$\text{smed} [X_1, \ldots, X_n] = \begin{cases} X_{((n+1)/2)} & \text{for } n \text{ odd} \\ (X_{(n/2)} + X_{(n/2+1)})/2 & \text{for } n \text{ even.} \end{cases}$$

b. Alternatively, one might define the sample median to minimize the sum of distances from the median:

$$\text{smed} [X_1, \ldots, X_n] = \arg\min_{\eta} \sum_{i=1}^{n} |X_i - \eta|;$$

that is, the estimate minimizes the sum of distances from data points to the potential median value, with distance measured by the sum of absolute values.

i. This definition exactly coincides with the earlier definition for $n$ odd, shares in the earlier definition’s lack of uniqueness for even sample sizes, and typically shares the opposite resolution (averaging the middle two observations) of this non-uniqueness.

ii. In contrast, the sample mean $\bar{X}$ of satisfies

$$\bar{X} = \arg\min_{\eta} \sum_{i=1}^{n} |X_i - \eta|^2.$$
c. Under certain circumstances, the sample median is approximately Gaussian.

i. Central limit theorems for the sample median generally require only that the density of the raw observations be positive in a neighborhood of the population median.

d. In § it was claimed that the sample average is equivariant for affine transformation.

i. A stronger property holds for medians.

ii. If $Y_i = h(X_i)$, for $h$ monotonic, then for $n$ odd, 
$$\text{smed} [Y_1, \ldots, Y_n] = h(\text{smed} [X_1, \ldots, X_n]).$$

iii. For $n$ even, this is approximately true, except that the averaging of the middle two observations undermines exact equivariance for non-affine transformations.

e. Both and () are special cases of an estimator defined by

i. The sample mean uses $\varrho(z) = z^2/2$ and the sample median uses $\varrho(z) = |z|$.

ii. He suggests

iii. Denote this piecewise derivative as $\varrho'$.

3. Hypothesis Tests Concerning the Population Median
a. Consider iid random variables \( X_i \) for \( i = 1, \ldots, n \).

i. To test whether a putative median value \( \theta^0 \) is the true value, define new random variables
\[
Y_j = \begin{cases} 
1 & \text{if } X_j - \theta^0 \leq 0 \\
0 & \text{if } X_j - \theta^0 > 0 
\end{cases}.
\]

ii. Then under \( H_0 : \theta = \theta^0 \), \( Y_j \sim Bin(1/2, 1) \).

iii. This technique can be traced to , as described in the example below.

iv. This logic only works if
\[
P \left[ X_j = \theta^0 \right] = 0;
\]
assume this.

v. It is usually easier to assess this continuity assumption than it is for distributional assumptions.

b. Then the median inference problem reduces to one of binomial testing.

i. Let \( T(\theta^0) = \sum_{j=1}^{n} Y_j \) be the number of observations less than or equal to \( \theta^0 \).

ii. Pick \( t_l \) and \( t_u \) so that \( \sum_{j=t_l}^{t_u-1} (1/2)^n \binom{n}{j} \geq 1 - \alpha \).

iii. One might choose \( t_l \) and \( t_u \) symmetrically, so that \( t_l \) is the largest value such that
\[ \sum_{j=0}^{t_l-1} \binom{n}{j} \leq \alpha/2. \]

iv. That is, \( t_l \) is that potential value for \( T \) such that not more than \( \alpha/2 \) probability sits below it.

v. The largest such \( t_l \) has probability at least \( 1 - \alpha/2 \).

vi. Hence \( t_l \) is the \( \alpha/2 \) quantile of the \( \text{Bin}(n, 1/2) \) distribution.

vii. Generally, the inequality in is strict.

viii. That is, \( \leq \) is actually \( < \).

ix. For combinations of \( n \) and \( \alpha \) for which this inequality holds with equality, the quantile is not uniquely defined, and take the quantile to be the lowest candidate.

x. Symmetrically, one might choose the smallest \( t_u \) so that

c. Then, reject the \( H_0 \) if \( T \leq t_L^0 \) or \( T \geq t_U^0 \) for \( t_L^0 = t_l - 1 \) and \( t_U^0 = t_u \).

i. This test is called the (exact) sign test, or the binomial test.

ii. An approximate version of the sign test might be created by selecting critical values from the Gaussian approximation to the distribution of \( T(\theta^0) \).

d. Table presents levels of tests for various data distributions.
i. Both variants of the sign test succeed in keeping the test level no larger than the nominal value.

ii. However, the sign test variants, because of the discreteness of the binomial distribution, in some cases achieve levels much smaller than the nominal target.

iii. Subtable (a), for sample size 10, is the most extreme example of this.

iv. Subtable (b), for sample size 17, represents the smallest reduction in actual sample size, and subtable (c), for sample size 40, is intermediate.

v. Note further that, while the asymptotic sign test, based on the Gaussian approximation, is not identical to the exact version, for subtables (a) and (c) the levels coincide exactly, since for all simulated data sets, the $p$-values either exceed 0.05 or fail to exceed 0.05 for both tests.

vi. Subtable (b) exhibits a case in which for one data value, the exact and approximate sign tests disagree on whether $p$-values exceed 0.05.

e. Table presents characteristics of the exact two-sided binomial
test of the $H_0$ that the probability of success is half, with level $\alpha = 0.05$, applied to small samples.

i. In this case, the two-sided $p$-value is obtained by doubling the one-sided $p$ value.

f. For small samples ($n < 6$), the smallest one-sided $p$-value, $1/2^n$, is greater than .025, and the $H_0$ is never rejected.

i. Such small samples are omitted from Table .

ii. This table consists of two subtables side by side, for $n \leq 23$, and for $n > 23$.

iii. The first column of each subtable is sample size.

iv. The second is $t_l - 1$ from .

v. The third is the value taken from performing the same operation on the Gaussian approximation.

vi. That is, it is the largest $a$ such that

vii. The fourth is the observed test level.

viii. That is, it is double the right side of .

ix. Observations here agree with those from Table .

x. For sample size 10, the level of the binomial test is severely too
small, for sample size 17, the binomial test has close to the optimal level, and for sample size 40, the level for the binomial test is moderately too small.

g. A complication (or, to an optimist, an opportunity for improved approximation) arises when approximating a discrete distribution by a continuous distribution.

i. Consider the case with $n = 10$, exhibited in Figure.

---

**Fig. contcor: Approximate Probability Calculation for Sign Test**

Sample Size 10, Target Level 0.05, and from Table $a - 1 = 1$
ii. Bar areas represent the probability under the $H_0$ of observing the number of successes.

iii. Table indicates that the one-sided test of level 0.05 rejects the $H_0$ for $W \leq 1$.

iv. The actual test size is 0.0215, which is graphically represented as the sum of the areas in the bar centered at 1, and the very small area of the neighboring bar centered at 0.

v. Expression approximates the sum of these two bar areas by the area under the dotted curve, representing the Gaussian density with the appropriate expectation $n/2 = 5$ and standard deviation $\sqrt{n}/2 = 1.58$.

vi. In order to align the areas of the bars most closely with the area under the curve, the Gaussian area should be taken to extend to the upper end of the bar containing 1.

vii. That is, evaluate the Gaussian distribution function at 1.5, explaining the 0.5 in.

viii. More generally, for a discrete distribution with potential values $\Delta$ units apart, the ordinate is shifted by $\Delta/2$ before applying a Gaussian approximation.
ix. This adjustment is called a correction for continuity.

x. The power of the sign test is determined by

\[ P_{\theta^A} \left[ X_j \leq \theta^0 \right] \]

xi. Since \( \theta^A > \theta^0 \) if

h. If the population median is not unique, then there will be
   alternatives for which the sign test has power no larger than the
   test level.

i. That is, \( \theta^0 \) is the true population median of the \( X_j \), and if
   there exists a set of form \((\theta^0 - \epsilon, \theta^0 + \epsilon)\), with \( \epsilon > 0 \), such
   that \( P \left[ X_j \in (\theta^0 - \epsilon, \theta^0 + \epsilon) \right] = 0 \), then any other \( \theta \) in this
   set is also a population median for \( X_j \), and hence the test will
   have power against such alternatives no larger than the test
   level.

ii. Such occurrences are rare.

i. The alternative is chosen to make the \( t \)-test have power
   approximately .80 for the Gaussian and Laplace distributions,
   using.
ii. In this case both $\sigma_0$ and $\sigma_A$ for the Gaussian and Laplace distributions are $1/\sqrt{n}$.

iii. Formula is inappropriate for the Cauchy distribution, since in this case $\bar{X}$ does not have a distribution that is approximately Gaussian.

iv. For the Cauchy distribution, the same alternative as for the Gaussian and Laplace distributions is used.

j. Results in Table show that for a sample size for which the sign test level approximates the nominal level ($n = 17$), use of the sign test for Gaussian data results in a moderate loss in power relative to the $t$-test, while use of the sign test results in a moderate gain in power for Laplace observations, and in a substantial gain in power for Cauchy observations.

i. Its null value is 0.

ii. The one sided value of is trivially $P[T \geq 82] = (1/2)^{82}$, which is tiny.

4. Confidence Intervals for the Median

a. Apply the test inversion approach of § to the sign test that rejects
Lecture 2

$H_0 : \theta = \theta^0$ if fewer than $t_l$ or at least $t_u$ data points are less than or equal to $\theta^0$.

i. Let $X(\cdot)$ referring to the data values after ordering.

ii. When $\theta^0 \leq X(1)$, then $T(\theta^0) = 0$.

iii. For $\theta^0 \in (X(1), X(2)]$, $T(\theta^0) = 1$.

iv. For $\theta^0 \in (X(2), X(3)]$, $T(\theta^0) = 2$.

v. In each case, the ( at the beginning of the interval and the ) at the end of the interval arises from, because observations that are exactly equal to $\theta^0$ are coded as one.

vi. Hence the test rejects $H_0$ if $\theta^0 \leq X(t_l)$ or $\theta^0 > X(t_u)$, and, for any $\theta^0$.

vii. This relation leads to the confidence interval is $(X(t_l), X(t_u)]$.

viii. However, since the data have a continuous distribution, then $X(t_u)$ also has a continuous distribution, and

$$P_{\theta^0} \left[ \theta^0 = X(t_u) \right] = 0$$

for any $\theta^0$.

ix. Hence $P_{\theta^0} \left[ X(t_l) < \theta^0 < X(t_u) \right] \geq 1 - \alpha$, and one might exclude the upper end point, to obtain the interval $(X(t_l), X(t_u))$. 


b. Figure exhibits construction of the confidence interval in the previous example.

i. The confidence interval is the set of log medians that yield a test statistic for which the $H_0$ is not rejected.

ii. Values of the statistic for which the $H_0$ is not rejected are between the horizontal lines.

iii. Log medians in the confidence intervals are values of the test statistic within this region.

c. In this construction, order statistics (that is, the ordered values) are first plotted on the horizontal axis, with the place in the ordered data set on the vertical axis.

i. These points are represented by the points in Figure where the step function transitions from vertical to horizontal, as one moves from lower left to upper right.

ii. Next, draw horizontal lines at the values $t_l$ and $t_u$, given by
and () respectively.

iii. Finally, draw vertical lines through the data points that these horizontal lines hit.

iv. For this particular example, the exact one-sided binomial test of level 0.025 rejects the $H_0$ that the event probability is half if the sum of event indicators is 0, 1, 2, 3, 4, or 5; $t_l = 6$.

v. For $Y_j$ of , the sum is less than 6 for all $\theta$ to the left of the point marked $X(\theta)$.

vi. Similarly, the one-sided level 0.025 test in the other direction
rejects the $H_0$ if the sum of event indicators is at least $t_u = 16$.

vii. The sum of the $Y_j$ exceeds 15 for $\theta$ to the right of the point marked $X(t_u)$.

d. By symmetry, one might expect $t_l = n - t_u$, but this is not the case.

i. The asymmetry in definitions and () arises because construction of the confidence interval requires counting not the data points, but the $n - 1$ spaces between them, plus the regions below the minimum and above the maximum, for a total of $n + 1$ ranges.

ii. Then $t_l = n + 1 - t_u$.

e. This interval is not of the usual form $\hat{\theta} \pm 2\hat{\sigma}$, for $\hat{\sigma}$ with a factor of $1/\sqrt{n}$.

i. Chapter investigates estimation of this density.

ii. This estimate can be used to estimate the median variance, but density estimation is harder than the earlier confidence interval rule.

E. Inference for Other Quantiles
1. The quantile $\theta$ corresponding to probability $\gamma$ is defined by

$$P_{\theta} [X_j \leq \theta] = \gamma.$$ 

a. Suppose that $\theta$ is quantile $\gamma \in (0, 1)$ of distribution of iid continuous random variables $X_1, \ldots, X_n$.

b. Then one can produce a generalized sign test.

c. Define the null and alternative hypotheses $H_0 : \theta = \theta^0$ and $H_A : \theta \neq \theta^0$.

d. As before, $T(\theta)$ is the number of observations smaller than or equal to $\theta$.

e. For the true value $\theta$ of the quantile, $T \sim \text{Bin}(n, \gamma)$.

f. Choose $t_l$ and $t_u$ so that $
\sum_{j=t_l}^{t_u-1} \gamma^j (1-\gamma)^{n-j} \binom{n}{j} \geq 1 - \alpha.$

g. Often, one chooses the largest $t_l$ and smallest $t_u$ so that

$$\sum_{j=0}^{t_l-1} \gamma^j (1-\gamma)^{n-j} \binom{n}{j} < \alpha/2, \quad \sum_{j=t_u}^{n} \gamma^j (1-\gamma)^{n-j} \binom{n}{j} < \alpha/2$$

h. Hence

i. One then rejects $H_0$ if $T < t_l$ or $T \geq t_u$.

2. This test is then inverted to obtain $(X_{(t_l)}, X_{(t_u)})$ as the confidence interval for $\theta$.

a. Note that the confidence level is conservative:
b. For any given \( \theta \), the inequality is generally strict.

3. Dependence of the test statistic \( T(\theta) \) on \( \theta \) is relatively simple.
   a. Later inversions of more complicated statistics will make use of the simplifying device of first, shifting all or part of the data by subtracting \( \theta \), and then testing the \( H_0 \) that the location parameter for this shifted variable is zero.

F. Comparing Tests

1. For fixed level, alternative, and power, the test with a smaller sample size is better.
   a. Consider two families of one-sided tests, indexed by sample size, using statistics \( T_1 \) and \( T_2 \), both with test level \( \alpha \), and determine the sample sizes required to give power \( 1 - \beta \), for the same alternative.
   b. Compare the tests by taking ratio of these two sample sizes.
   c. The ratio is called relative efficiency.
   d. The notation dates back at least as far as , citing .
Lecture 2

2. Notation and Assumptions

a. Let $t_{j,n}^\circ$ represent the critical value for test $j$ based on $n$ observations.

i. That is, the test based on statistic $T_j$ and using $n$ observations, rejects the $H_0$ if $T_j \geq t_{j,n}^\circ$.

ii. Hence $t_{j,n}^\circ$ satisfies

$$P_{\theta_0} \left[ T_j \geq t_{j,n}^\circ \right] = \alpha.$$  

iii. Let $\omega_{j,n}(\theta_A)$ represent the power for test $T_j$ using $n$ observations, under the alternative $\theta_A$:

$$\omega_{j,n}(\theta_A) = P_{\theta_A} \left[ T_j \geq t_{j,n}^\circ \right].$$

iv. Assume that $\omega_{j,n}(\theta_A)$ is continuous and increasing in $\theta_A$ for all $j, n$,

$$\lim_{\theta_A \to \infty} \omega_{j,n}(\theta_A) = 1,$$

$$\lim_{n \to \infty} \omega_{j,n}(\theta_A) = 1 \text{ for all } \theta_A > \theta_0.$$ 

b. Two tests, tests 1 and 2, involving hypotheses about a parameter $\theta$, taking the value $\theta_0$ under the $H_0$, and with a simple alternative hypothesis of form $\{\theta_A\}$, for some $\theta_A > \theta_0$, with similar level and power, will be compared.
i. Pick a test level $\alpha$ and a power $1 - \beta$, and the sample size $n_1$ for test 1.

ii. The power and level conditions on $T_1$ imply a value for $\theta^A$ under the alternative hypothesis.

iii. That is, $\theta^A$ solves $P_{\theta^A} \left[ T_1 \geq t^\circ_{1,n_1} \right] = 1 - \beta$.

iv. Note that $\theta^A$ is a function of $n_1$, $\alpha$, and $\beta$.

v. Under conditions,

vi. Report $n_1/n_2$ as the relative efficiency of test 2 to test 1.

vii. This depends on $n_1$, $\alpha$, and $\beta$.

c. Define the asymptotic relative efficiency

i. Considering this quantity removes dependence on $n_1$.

d. This measure comparing efficiencies of two tests takes on a particularly easy form in a special, yet common, case, in which both statistics are asymptotically Gaussian.

i. In this case, the relative efficiency can be approximated in terms of standard deviations and derivatives of means under alternative hypotheses.

ii. General approximations for sample size, power, and effect sizes are investigated first.
iii. These are applied to relative efficiency later.

iv. Formulas for power, sample size, and effect size developed below may be used for efficiency comparisons, but are also useful on their own.

v. Gaussian approximations earlier in this chapter often applied a continuity correction.

vi. This correction will not be applied for large-sample power and sample size calculations, as the effect of this correction quickly becomes negligible as the sample size increases.

vii. WOLOG, take $\theta_0 = 0$.

3. Power

a. Consider test statistics satisfying

$$T_j \sim \mathcal{N}(\mu_j(\theta), \varsigma^2_j(\theta)), \quad \text{for } \varsigma_j(\theta) > 0, \mu_j(\theta) \text{ increasing in } \theta.$$  

i. The Gaussian distribution in does not need to hold exactly.

ii. Holding approximately is sufficient.

iii. In this case, one can find the critical values for the two tests, $t_{j,n_j}^\circ$, such that $P_0 \left[ T_j \geq t_{j,n_j}^\circ \right] = \alpha$.

iv. Since $(T_j - \mu_j(0))/\varsigma_j(0)$ is approximately standard Gaussian under the $H_0$, then
\[ \alpha = P_0 \left[ \left( T_j - \mu_j(0) \right) / \varsigma_j(0) \geq z_\alpha \right] = P_0 \left[ T_j \geq \mu_j(0) + \varsigma_j(0) z_\alpha \right]. \]

v. Hence

\[ t_{j,n_j}^0 = \mu_j(0) + \varsigma_j(0) z_\alpha. \]

vi. The power for test \( j \) is approximately

\[ \tau_{j,n_j}(\theta^A) \approx P_{\theta^A} \left[ T_j \geq \mu_j(0) + \varsigma_j(0) z_\alpha \right] = 1 - \Phi \left( \left[ \mu_j(0) + \varsigma_j(0) z_\alpha - \mu_j(\theta^A) \right] / \varsigma_j(\theta^A) \right). \]

vii. Often the variance of the test statistic changes slowly as one moves away from the \( H_0 \).

viii. In this case, the power for test \( j \) is approximately

\[ \tau_{j,n_j}(\theta^A) \approx 1 - \Phi \left( \left[ \mu_j(0) - \mu_j(\theta^A) \right] / \varsigma_j(0) + z_\alpha \right). \]