Some Remarks About Unit One Procedures

Inferences Pertaining to Skewed Distributions

Inferences about the mean

point estimates

When the skewness is large relative to the kurtosis, the sample mean should be used. But if the sample size is smallish and the kurtosis is large relative to the skewness, one should consider using a trimmed mean, perhaps trimming 10% (or 20% if the kurtosis is rather large). It can be noted that an M-estimator works nearly as well. However, if the sample size is large enough, then the bias associated with such alternative estimators more than offsets the reduction in variance that they can achieve, and in such cases one should just go with a sample mean. (See the handout pertaining to the overhead presentation for more details.)

interval estimates

The confidence interval based on Johnson's modified t statistic should be preferred over the one based on Student's t statistic. However, the difference can be very very slight (unlike the appreciable differences we often see in hypothesis testing).

hypothesis testing

Johnson's modified t test should be preferred over Student's t test. (Note: For Johnson's test, I sometimes just report one significant digit for the p-value. Also, if you're uncertain about the skewness issue, then one can try both Johnson's test and the regular t test to see if it makes a difference — sometimes it hardly matters when the skewness is slight.)

Inferences about the median

point estimates

When the sample size is smallish and the skewness is large relative to the kurtosis, an upper-teiled trimmed mean should be used (assuming positive skewness — otherwise use a lower-tailed trimmed mean), trimming only mildly (from 0 to 5 percent). If the sample size is smallish and the kurtosis is large relative to the skewness, one should consider using a (two-tailed) trimmed mean, trimming from 10% to 20% depending upon the tailweight. It can be noted that an M-estimator works nearly as well in these small sample situations except when the skewness is large relative to the kurtosis. However, if the sample size is large enough, then the bias associated with such alternative estimators more than offsets the reduction in variance that they can achieve, and in such cases one should just go with a Harrell-Davis estimate. (See the handout pertaining to the overhead presentation for more details.)

interval estimates

The confidence interval based on the sign test statistic is one possibility, and I think one can generally trust Minitab's interpolation version for cases when the desired coverage probability cannot be obtained with an exact interval. Another possibility is to seek a transformation to near-normality, use the t statistic to obtain a confidence interval (assuming median is same as mean when appropriate transformation is obtained), and then "back transform" to obtain interval estimate for median of the skewed distribution. But sometimes a suitable transformation cannot be obtained (and even if you find one it doesn't necessarily lead to a shorter interval), and I haven't seen adequate work done to investigate the accuracy of this scheme if one errs in choosing a transformation (you should be okay as long as you obtain symmetry, but I wonder what happens if the transformed distribution is slightly skewed (however, I suspect that in close cases the scheme will perform okay)).

hypothesis testing

The sign test is always valid for tests about the median of a continuous distribution. The other possibility would be to try the transformation ploy.

Inferences about quantiles other than the median

point estimates

As I discussed, for distributions having positive skewness, E1 and E9 (H-D estimator) are two good possibilities to choose between in a lot of cases, with preference going to E9 when sample size is suitably large, the skewness is not too severe, and the quantile is not too far out into the upper tail (assuming positive skewness). (Note: for negatively skewed distributions, one needs to modify E1.)

Here are some specific guidelines for estimating the qth quantile of a *postively* skewed distribution: for $n \leq 25$ and $q \geq 0.75$, use E1; for $n \geq 75$ and $q \in [0.15, 0.85]$, use E9 unless the skewness is great (e.g., for a lognormal distribution having a skewness greater than 6, one might use E1 for $q \geq 0.75$ even if n is as large as 300); for other cases it depends on the specifics (the values of q, n, and the skewness), but except possibly in some cases for which $n < 25$, I'd favor using E9 whenever $q \in [0.35, 0.65]$.

interval estimates

In STAT 657 I sometimes cover a method based on the binomial distribution, but I don't cover any methods for this in STAT 554.

hypothesis testing

One way that you can go is to recast the problem as a test about a Bernoulli parameter (the situation addressed the first week of class). For example, note that for continuous distributions the 70th percentile is greater than 25 only if the probability of an observation exceeding 25 is greater than 0.3.

Inferences Pertaining to Symmetric Distributions

Inferences about the median/mean

point estimates

For normal, near-normal, or light-tailed distributions, the sample mean is a good choice. But for heavy-tailed distributions, one can typically do better with another estimator. A nice choice that does pretty good in most cases with heavy-tailed distributions is an M-estimator. (The Huber variety does good, but one can wonder if some other type of M-estimator, say a Tukey or Andrews version, would do better for rather extreme heavy tails. For the Huber variety, one is pretty safe picking the bend parameter to be 1.345, but slightly improved performance may be obtained by going with a bend of 1.5 for distributions that appear to be only slightly heavier tailed than a normal distribution, and going with a bend of 1.2 for extreme heavy tails (say if the kurtosis is 3.5 or greater).) It can be noted that for a wide variety of moderately heavy-tailed distributions, a 10% trimmed mean seems to be as good as the Huber M-estimator with a bend of 1.345. Also, I'll point out that the one-step M-estimator seems to work nearly as good as the fully iterated version — the difference is so slight it's not worth being concerned about. In some heavier-tailed cases, trimming more than 10% is appropriate. (See the handout pertaining to the overhead presentation for more details.) An adaptive trimmed mean is another possibility, but when the sample size is small the estimated standard errors can be off and so one shouldn't blindly rely on the adaptive procedure.

interval estimates

For normal, near-normal, and light-tailed distributions, the confidence interval based on Student's t statistic is a good choice. But for heavy-tailed distributions one can sometimes do better with another method, with nonparametric choices (related to sign test and signed-rank test) among the possibilities, along with confidence intervals based on trimmed means. But such alternative choices can be inaccurate if distribution is skewed (although it may be that very mild skewness doesn't mess things up too much). The interval based on the sign test statistic will seldom be the best method. When using an interval based on a trimmed mean, it's important not to trim too much, since that can lead to the coverage probability dropping well below the nominal level. One is fairly safe in going with a 10% trimmed mean in most cases, as long as distribution has heavy tails — but always leave at least 10 observations untrimmed when using a trimmed mean for intervals and tests. I wouldn't trust the result if the trimmed mean interval is shorter than the t interval if the distribution doesn't appear to be heavy tailed.

hypothesis testing

For normal, near-normal, and light-tailed distributions, Student's t test is a good choice. But for heavy-tailed distributions one can sometimes do better with another method, with nonparametric choices (sign test and signed-rank test (and also normal scores test and permutation test)) among the possibilities, along with the approximate t test based on a trimmed mean. But such alternative choices can be inaccurate if distribution is skewed (although it may be that very mild skewness doesn't mess things up too much). The sign test statistic will seldom be the best method. When using a test based on a trimmed mean, it's important not to trim too much, since that can lead to an inflated type I error rate (and misleadingly small p-values). One is fairly safe in going with a 10% trimmed mean in most cases, as long as distribution has heavy tails — but always leave at least 10 observations untrimmed when using a trimmed mean for intervals and tests. I wouldn't trust the result if the trimmed mean test produces smaller p-value than the t test if the distribution doesn't appear to be heavy tailed.

Inferences about quantiles other than the median

point estimates

As I discussed, E1 and E9 (H-D estimator) are two good possibilities to choose between in a lot of cases, with preference going to E9 when sample size is suitably large, the tailweight isn't too extreme, and the quantile is not too far out into the upper or lower tail. (Note: for estimating quantiles below the median, one needs to modify E1.) For normal, near-normal, and most light-tailed distributions, E9 (the H-D estimator) is better than E1, and in fact unless the sample size is quite small (say less than 25) or the quantile is quite extreme (relative to the sample size), I prefer E9 to E1. Even for small sample sizes, E9 can do better than E1 for estimating quantiles related to the middle 50% (or more) of the distribution. If the distribution happens to be uniform, or very close to uniform, E4 can be a superior choice, but I don't encounter such cases too often. For near-normal distributions, if you want a simple estimator instead of the more complex H-D estimator, E8 is a good choice. Here are some specific guidelines for estimating the qth quantile of a symmetric distribution: for $n \leq 25$ and $q \leq 0.07$ or $q \geq 0.93$, use E1 (appropriately modified for $q \leq 0.07$) for heavy-tailed distributions, or E8 for near-normal distributions; for $n \geq 50$ and $q \in [0.1, 0.9]$, use E9; for other cases it depends on the specifics (the values of q, n, and the kurtosis).

interval estimates

In STAT 657 I cover a method based on the binomial distribution, but I don't cover any methods for this in STAT 554.

hypothesis testing

One way that you can go is to recast the problem as a test about a Bernoulli parameter (the situation addressed the first week of class). For example, note that for continuous distributions the 70th percentile is greater than 25 only if the probability of an observation exceeding 25 is greater than 0.3.

Testing for a Treatment Effect in a Matched-Pairs Experiment

All of the nonparametric tests from the first unit are valid (you ignore all differences of zero and work with a reduced sample size if zero values occur). The trimmed mean t test is okay provided that you don't overtrim (one is usually safe trimming 10%). The t test is okay in most cases — since it's consevstive for heavy-tailed distributions, one only has to worry about anticonservativeness when the sample size is quite small (less than 10) and the distribution is appreciably light-tailed. Only Johnson's test is eliminated from consideration. When using any of the other tests, you don't have to be concerned if the data suggests skewness, since if the treatment created a skewed distribution and that causes rejection on the null hypothesis, then fine — you want to reject if there was any sort of treatment effect. In fact, other than to screen for bad outliers (that may be bogus values due to some sort of mistake (and **on the HW you shouldn't assume any such mistakes and remove data values**)), you really don't have to look at the data. (Again, under H_0 the distribution will be symmetric (about 0), and if lack of symmetry causes a test to reject that's what we want.)

Note: If you are asked about the mean or the median of the difference distribution you just make use of the previous guidelines, taking the differences (including the zero values) as your sample — just because you have matched-pairs data, it doesn't necessarily mean that you are to do a test for a treatment effect.

One final comment: It should be noted that in many places guidelines refer to the true values of the skewness and the kurtosis, and not to estimates based on small samples. It should be kept in mind that in a small sample, only one or two somewhat extreme observations can have a great effect on the estimates. In particular, with regard to skewness, often a heavy-tailed symmetric distribution can yield a sample skewness value somewhat far from 0 if the sample size is rather small (since the sample skewness will only be close to 0 if the extreme observations "balance" properly in the two tails, which may be a bit unlikely).