Section 3.2: The Quantile Test

• Assume that the measurement scale of our data are at least ordinal. Then it is of interest to consider the <u>quantiles</u> of the distribution.

<u>Case I</u>: Suppose the data are continuous. Then the p^*th quantile is a number x^* such that

$$P(X \leq x^*) = p^*$$

• Consider testing the null hypothesis that the p^*th quantile is some specific number x^*, i.e.,

$$H_0: \underbrace{P(X \leq x^*)} = p^*$$

• If we denote $P(X \leq x)$ by p, then we see this is the same null as in the binomial test, and we can conduct the test in the same way.

• Assume the data are a random sample (i.i.d. random variables) measured on at least an ordinal scale.

• Which test statistic we use will depend on the alternative hypothesis. Consider

T_1 = number of sample observations $\leq x^*$

T_2 = number of sample observations $< x^*$

• **Note $T_1 \geq T_2$, and if none of the data values equal the number x^*, then:** $T_1 = T_2$

• **The null distribution of the test statistics T_1 and T_2 is again** binomial (n, p^*).

assuming continuous data

Three Possible Sets of Hypotheses

Two-tailed test:

H_0: p^*th population quantile $= x^*$

H_1: p^*th population quantile $\neq x^*$

Decision rule: Reject H_0 if $T_1 \leq t_1$ or $T_2 > t_2$ where t_1, t_2 are such that

$$P(Y \leq t_1) \approx \alpha_1 \quad \text{and} \quad P(Y \leq t_2) \approx 1 - \alpha_2$$

where $\alpha_1 + \alpha_2 \leq \alpha$.

• **The P-value of the test is:**

$$2\left[\min\left\{P(Y \leq t_1^{(obs)}),\ P(Y \geq t_2^{(obs)})\right\}\right]$$

where $Y \sim$ Binomial(n, p^*).

"Quantile greater than" alternative: book calls this the "Lower-Tailed Test"

H_0: p^*th population quantile $\leq x^*$

H_1: p^*th population quantile $> x^*$

Decision rule: Reject H_0 if $T_1 \le t_1$
where t_1 is such that

$$P(Y \le t_1) \le \alpha$$

- **The P-value of the test is:**

$$P(Y \le t_1^{(obs)})$$

where $Y \sim$ Binomial(n, p^*).

"Quantile less than" alternative: Book calls this "Upper-tailed" test

H_0: p^*th population quantile $\ge x^*$ **H_1:** p^*th population quantile $< x^*$

Decision rule: Reject H_0 if $T_2 > t_2$
where t_2 is such that

$$P(Y \le t_2) \ge 1 - \alpha$$

- **The P-value of the test is:**

$$P(Y \ge t_2^{(obs)})$$

where $Y \sim$ Binomial(n, p^*).

Example: Suppose the upper quartile (0.75 quantile) of a college entrance exam is known to be 193. A random sample of 15 students' scores from a particular high school are given on page 139. Does the population upper quartile for this high school's students differ from the national upper quartile of 193? Use $\alpha = 0.05$.

$\alpha/2 = .025$

$1-\alpha/2 = .975$

H₀: population 0.75 quantile = 193 **H₁:** population 0.75 quantile $\neq$ 193

Book: $P(X \le 193) = 0.75$ $P(X \le 193) \neq 0.75$

Decision Rule (using Table A3):

$n=15$, $p^* = 0.75$

$P(Y \le 7) = .0173$

$P(Y \le 14) = .9866$

Reject Ho if $T_1 \le 7$ or $T_2 > 14$

Observed test statistics:

$T_1 = 7 \leftarrow$ # obs ≤ 193

$T_2 = 6 \leftarrow$ # obs < 193

Since $T_1 \le 7$, we reject Ho.

P-value: $2[\min\{P(T_1 \le 7), P(T_2 \ge 6)\}]$

$= 2[\min\{.0173, 1-.0008\}] = 2(.0173) = \boxed{.0346}$

$\le \alpha$

Conclusion:

Reject Ho; conclude the population upper quartile for this high school differs from 193.

Example: Suppose the **median** (0.5 quantile) selling price (in $1000s) of houses in the U.S. from 1996-2005 was 179. Suppose a random sample of 18 house sale prices from 2011 is `120 500 64 104 172 275 336 55 535 251 214 1250 402 27 109 17 334 205`. Has the population median sale price decreased from 179? Use $\alpha = 0.05$.

H_0: 2011 population median ≥ 179

H_1: 2011 population median < 179

Decision Rule (using Table A3): Reject H_0 if $T_2 > 12$.

$n = 18$

$p^* = 0.5$

$P(Y \leq 12) = 0.9519$

(Table A3)

Observed test statistic: $T_2 = 8$, so fail to reject H_0.

obs < 179

P-value: $P(Y \geq 8) = 1 - P(Y \leq 7) = 1 - .2403$

$= \boxed{.7597}$

Conclusion: Fail to reject H_0. We cannot conclude the 2011 population median sale price is less than 179 thousand dollars.

- **See R code on course web page for examples using the `quantile.test` function.**

Confidence Interval for a Quantile

- Recall $X^{(1)} \le X^{(2)} \le \ldots \le X^{(n)}$ are called the ordered sample, or order statistics.

- The order statistics can be used to construct an exact CI for any population quantile.

- Suppose the desired confidence level is $1 - \alpha$ (e.g., 0.90, 0.95, 0.99, etc.).

- In Table A3, use the column for p^* (quantile desired).

- In Table A3, find a probability near $\alpha/2$ (call this α_1).

- The corresponding y in Table A3 is then called $r - 1$.

- Then find a probability near $1 - \alpha/2$ (call this $1 - \alpha_2$).

- The corresponding y in Table A3 is then called $s - 1$.

- Then the pair of order statistics $[X^{(r)}, X^{(s)}]$ yields a CI for the p^*th population quantile.

- This CI will have confidence level at least $1 - \alpha_1 - \alpha_2$ (exactly $1 - \alpha_1 - \alpha_2$ if the data are continuous).

$1-\alpha = .95 \Rightarrow \alpha = .05$

<u>Example</u> (House prices): Find an exact CI with confidence level at least 95% for the population median house price in 2011. $\alpha/2 = .025 \Rightarrow 1-\alpha/2 = .975$

Use $p^* = 0.5$ column

Use $n = 18$ (sample size)

Choose $\alpha_1 = .0154 \Rightarrow r-1 = 4 \Rightarrow r = 5$

Choose $1-\alpha_2 = .9846 \Rightarrow s-1 = 13 \Rightarrow s = 14$

$[X^{(5)}, X^{(14)}] = [104, 336]$ is a CI for the population median price with confidence level $.9846 - .0154 = \boxed{.9692}$

<u>Example</u> (House prices): Find an exact CI with confidence level at least 95% for the population 0.80 quantile of house prices in 2011.

$n=18$. Use $p^* = 0.80$ column

Choose $\alpha_1 = .0163 \Rightarrow r-1 = 10 \Rightarrow r = 11$

Choose $1-\alpha_2 = .9820 \Rightarrow s-1 = 17 \Rightarrow s = 18$

$[X^{(11)}, X^{(18)}] = [251, 1250]$ is a CI for the population 80th percentile with confidence level $.9820 - .0163 = \boxed{.9657}$

- **See R code on course web page for examples using the `quantile.interval` function.**

Comparison of the Quantile Test to Parametric Tests

• The quantile test is valid for data that are ordinal or interval/ratio, whereas the one-sample t-test about the mean requires that data be normal.

• So the quantile test is more applicable.

• Suppose our distribution is continuous and symmetric. Then: population median = population mean.

• So the quantile test about the median is testing the same thing as the t-test about the mean.

• Which is more efficient? Depends on true population distribution:

Population	A.R.E. of quantile test to t-test
Normal	0.637 (t-test better)
Uniform (light tails)	0.333 (t-test better)
Double exponential (heavy tails)	2.00 (quantile test better)

• See R code on course web page for power functions of quantile test and t-test for various population distributions.