STAT 333
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 **quantiles** of the distribution.

Case I:

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.,

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

- 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 that :

- $T_1 \geq T_2$, 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 Binomail(n,p^*).
  (assuming continues data)

Three Possible Sets of Hypotheses

Two-tailed test:

H_0: $p^{*\,th}$ population quartile = X^* or H_0: $P(X \le x_*) = p^*$

H_1: $p^{*\,th}$ population quartile $\neq X^*$ or H_0: $P(X \le x_*) \neq p^*$

Decision rule:

Reject H_0 if $T_1 \le t_1$ or $T_2 > t_2$ (Where t_1, t_2 are such that)

$P(y \le t_1) \approx \alpha_1$ and $P(y \le t_2) \approx 1 - \alpha_2$

where $\alpha_1 + \alpha_2 \le \alpha$

• The **P-value** of the test is:

P-value = $2x[\min\{ P(y \le t_1^{(obs)}) \ , P(y \le t_2^{(obs)}) \}]$

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

"**Quantile greater than" alternative** : (Book calls this" the lower-tailed test")

- **Hypothesis :**

  H_0:$p^{*\,th}$ population quartile $\le X^*$ or H_0: $P(X \le x_*) \ge p^*$

  H_1: $p^{*\,th}$ population quartile $> X^*$ or H_1 : $P(X \le x_*) < p^*$

- **Decision rule:**

  Reject H_0 if $T_1 < t_1$ (where t_2 is such that $P(y \le t_1) \le \alpha$

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

  P-value = $P(y \le t_1^{(obs)})$

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

"Quantile less than" alternative : (Book calls this" the upper-tailed test")

- **Hypothesis :**

  H_0: $p^{*\,th}$ population quartile $\geq X^*$ or H_0: $P(X < x*) \leq p^*$

  H_1: $p^{*\,th}$ population quartile $< X^*$ or $H1 : P(X < x*) > p^*$

- **Decision rule:**

  Reject H_0 if $T_2 > t_2$ (where t_2 is such that $P(y \leq t_2) \geq 1-\alpha$)

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

  P-value = $P(y \geq t_2^{(obs)})$

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

Example(1):

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 $\alpha = 0.05$.

where they take the exam and get the following scores

189 233 195 160 212 176 231 185 199 213 202 193 174 166 248

- **Hypothesis:**

H_0: Population 0.75 quartile =193 H_1: Population 0.75 quartile $\neq$ 193

(From book : $P(X \leq 193) = 0.75$ $P(X \leq 193) \neq 0.75$

- **Decision Rule** (using Table A3): Binomial (15 , 0.75)

$P(X \leq 7) = 0.0173 < \alpha_1$

$P(X \leq 14) = 0.9866 > 1-\alpha_2$ (where $\alpha_1+\alpha_2 = 0.0173+(1-0.9866) = 0.0307 \leq \alpha = 0.05$)

Reject H_0 if $T_1 \leq t_1$ or $T_2 > t_2$ ($t_1 = 7$, $t_2 = 14$)

$T_1 \leq 7$ or $T_2 > 14$

Observed test statistics:

T_1 : Scores ≤ 193 :

189 233 195 160 212 176 231 185 199 213 202 193 174 166 248

$T_1 = 7$ (# observation ≤ 193)

T_2 : Scores < 193 :
189 233 195 160 212 176 231 185 199 213 202 193 174 166 248

$T_2 = 6$ (# observation <193)

Since $T_1 \le 7$, we reject H_0

P-value:

P-value = 2x [min{$P(T_1 \le 7)$, $P(T_2 \ge 6)$}] = 2x [min{$P(T_1 \le 7)$, 1- $P(T_2 \le 5)$}]

= 2x [min{0.0173, (1-0.0008)}] = 2 x 0.0173 = 0.0346 $\le \alpha$

Conclusion:

Reject H_0 : Population 0.75 quartile = 193

Accept H_1 : Population 0.75 quartile $\neq$ 193

Conclude that the population upper quartile for this high school differs from 193

Example(2):

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 $\alpha = 0.05$.

Solution :

- **Hypothesis**: (upper -tailed)

  H_0: 2011 population median $\ge$ 179 **H_1:** 2011 population median <179

  Or H_0: $P(X < 179) \le 0.5$ H_1 : $P(X < 179) > 0.5$

- **Decision Rule (using Table A3):**

  Reject H_0 if $T_2 > 12$

(From Table (3) : n =18 , $p^* = 0.5$, $P(y \le 12) = 0.9519$)

[Take T_2 such that , $P(y \le t_2) \ge 1-\alpha$]

Observed test statistic:

$T_2 = 8$ (# of observation < 179)

Since 8 > 12 ,

We accept H_0 : 2011 population median $\ge$ 179

P-value:

P- value = P(y ≥ 8) = 1- P(y ≤ 7) = 1 – 0.2403 = 0.7597

Conclusion:

We accept H_0 : 2011 population median ≥ 179

We conclude that 2011 population median sale price is greater than or equal 179 thousand dollars .

• See R code on course web page for examples using the quantile.test function.

Confidence Interval for a Quantile

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

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

1. Suppose the desired confidence level is 1 – α (e.g., 0.90, 0.95, 0.99, etc.).
2. In Table A3, use the column for ***p** (quantile desired**).
3. In Table A3, **find a probability near α/2** (call this α_1).
4. The corresponding ***y*** in Table A3 is then **called (*r* – 1).**
5. Then **find a probability near 1 – α/2** (call this $1 - \alpha_2$).
6. The corresponding ***y*** in Table A3 is then **called *s* – 1.**
7. Then the **pair of order statistics [$X^{(r)}$, $X^{(s)}$]** yields a **CI for the *p**th** population quantile.
8. This **CI will have confidence level <u>at least</u> $1 - \alpha_1 - \alpha_2$**

   (exactly $1 - \alpha_1 - \alpha_2$ if the data are continuous).

<u>Example 3:</u> From Example (2):

(House prices): Find an exact CI with confidence level at least **95% for the population median** house price in 2011.

$1-\alpha = 0.95$
$\alpha = 0.05$
$\alpha/2 = 0.05/2 = 0.025$
$1-\alpha/2 = 1-0.025 = 0.975$

Now, use $p^*=0.5$, $n=18$

Choose $\alpha_1 = 0.0154$ → $r-1 = 4$ → $r = 5$

Choose $1-\alpha_2 = 0.9846$ → $s-1 = 13$ → $s = 14$

<u>Data:</u>120,500,64,104 ,172,275 ,336 , 55 , 535 , 251 , 214 , 1250 , 402 , 27 , 109, 17, 334, 205.

<u>Ordered data:</u>

17,27, 55, 64,**104**,109,120,172,205,214,**251**,275,334, **336**,402, 500,535,1250

$X^{(5)}$ $X^{(14)}$

[$X^{(5)}$, $X^{(14)}$] = [104, 336] is a C.I for the population median price with confidence level

$1-\alpha_1-\alpha_2 = 0.9846 – 0.0154 = 0.9692$

<u>Example 4:</u> From Example(2) :
(House prices): **Find an exact CI** with confidence level at least **95%** for the **population 0.80 quantile of house prices in 2011.**

$1-\alpha = 0.95$
$\alpha = 0.05$
$\alpha/2 = 0.05/2 = 0.025$
$1-\alpha/2 = 1-0.025 = 0.975$

Now, use $p^*=0.80$, $n=18$

Choose $\alpha_1 = 0.0163$ → $r-1 = 10$ → $r = 11$

Choose $1-\alpha_2 = 0.9820$ → $s-1 = 17$ → $s = 18$

<u>Data:</u>120,500,64,104 ,172,275 ,336 , 55 , 535 , 251 , 214 , 1250 , 402 , 27 , 109, 17, 334, 205.

<u>Ordered data:</u>

17,27, 55, 64,104,109,120,172,205,214,**251**,275,334, 336,402, 500,535,**1250**

$X^{(11)}$ $X^{(18)}$

[$X^{(11)}$, $X^{(18)}$] = [251, 1250] is a C.I for the population median price with confidence level

$1-\alpha_1-\alpha_2$ = 0.9820 – 0.0163 =0.9657

• 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 (quartile test better)

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

R Code:

Example(1): Test score

```
quantile.test<-function(x,xstar=0,quantile=.5,alternative="two.sided"){
  n<-length(x)
  p<-quantile
  T1<-sum(x<=xstar)
  T2<-sum(x< xstar)
  if (alternative=="quantile.less") {
    p.value<-1-pbinom(T2-1,n,p)}
  if (alternative=="quantile.greater"){
    p.value<-pbinom(T1,n,p)}
  if (alternative=="two.sided"){
    p.value<-2*min(1-pbinom(T2-1,n,p),pbinom(T1,n,p))}
  list(xstar=xstar,alternative=alternative,T1=T1,T2=T2,p.value=p.value)}
## Entrance example data:
testscores <- c(189,233,195,160,212,176,231,185,199,213,202,193,174,166,248)

quantile.test(testscores,xstar=193,quantile=0.75,alternative="two.sided")
```

output:

```
$xstar
[1] 193

$alternative
[1] "two.sided"

$T1
[1] 7

$T2
[1] 6

$p.value
[1] 0.03459968
```

Example 2:

```
quantile.test<-function(x,xstar=0,quantile=.5,alternative="two.sided"){
 n<-length(x)
 p<-quantile
 T1<-sum(x<=xstar)
 T2<-sum(x< xstar)
 if (alternative=="quantile.less") {
  p.value<-1-pbinom(T2-1,n,p)}
 if (alternative=="quantile.greater"){
  p.value<-pbinom(T1,n,p)}
 if (alternative=="two.sided"){
  p.value<-2*min(1-pbinom(T2-1,n,p),pbinom(T1,n,p))}
 list(xstar=xstar,alternative=alternative,T1=T1,T2=T2,p.value=p.value)}

## Entrance example data:
prices <- c(120, 500, 64, 104, 172, 275, 336, 55, 535, 251, 214, 1250, 402, 27, 109, 17, 334, 205)

quantile.test(prices,xstar=179, quantile=0.5, alternative="quantile.less")
```

output :

```
$xstar
[1] 179

$alternative
[1] "quantile.less"

$T1
[1] 8

$T2
[1] 8

$p.value
[1] 0.7596588
```

Example 3:

Confidence interval for house price example :

95% for the population median house price in 2011

```
quantile.interval<-function(x,quantile=.5,conf.level=.95){
  n<-length(x)
  p<-quantile
  alpha<-1-conf.level
  rmin1<-qbinom(alpha/2,n,p)-1
  r<-rmin1+1
  alpha1<-pbinom(r-1,n,p)
  smin1<-qbinom(1-alpha/2,n,p)
  s<-smin1+1
  alpha2<-1-pbinom(s-1,n,p)
  clo<-sort(x)[r]
  chi<-sort(x)[s]
  conf.level<-1-alpha1-alpha2
  list(quantile=quantile,conf.level=conf.level,r=r,s=s,interval=c(clo,chi))}

# Copy and paste this function into R, and then it can be implemented
# as in the following examples:


## House price example data:

prices <- c(120, 500, 64, 104, 172, 275, 336, 55, 535, 251, 214, 1250, 402, 27, 109, 17, 334, 205)

# 95% CI for median:

quantile.interval(prices, quantile=0.5, conf.level=0.95)
```

output:

```
$quantile
[1] 0.8

$conf.level
[1] 0.9657055

$r
[1] 11

$s
[1] 18

$interval
[1]  251 1250
```

Example 4:

95% for the population 0.8 quartile house price in 2011

```
quantile.interval<-function(x,quantile=.5,conf.level=.95){
  n<-length(x)
  p<-quantile
  alpha<-1-conf.level
  rmin1<-qbinom(alpha/2,n,p)-1
  r<-rmin1+1
  alpha1<-pbinom(r-1,n,p)
  smin1<-qbinom(1-alpha/2,n,p)
  s<-smin1+1
  alpha2<-1-pbinom(s-1,n,p)
  clo<-sort(x)[r]
  chi<-sort(x)[s]
  conf.level<-1-alpha1-alpha2
  list(quantile=quantile,conf.level=conf.level,r=r,s=s,interval=c(clo,chi))}

# Copy and paste this function into R, and then it can be implemented
# as in the following examples:


## House price example data:

prices <- c(120, 500, 64, 104, 172, 275, 336, 55, 535, 251, 214, 1250, 402, 27, 109, 17, 334, 205)
```

 # 95% CI for 0.80 quantile:

```
quantile.interval(prices, quantile=0.80, conf.level=0.95)
```

output :

```
$quantile
[1] 0.8

$conf.level
[1] 0.9657055

$r
[1] 11

$s
[1] 18

$interval
[1]  251 1250
```