Introduction to mathematical Statistics

1. Test on one population proportion: p

Here we will introduce two scenarios:

A. Large sample approximate Z test based on the CLT

B. Exact test based on the binomial distribution (*not required in exams)

Example 1. Suppose I wish to find the probability p that a given coin will turn up (head) when tossed. I tossed the coin for 100 times and obtained 64 heads and the rest tails.

1) Estimate the probability p

2) Find a 95% CI for p

3) Test if p=1/2

Solution:

1) $\hat{p}$ =64/100=0.64

2) Review: Large sample approximate C.I

Let $X_i = \begin{cases} 1, & \text{if the } i^{th} \text{ toss is head.} \\ 0, & \text{if the } i^{th} \text{ toss is tail.} \end{cases}$ A proportion of P is head in the population.

$$X_i \overset{iid}{\sim} Bernoulli(p) \Rightarrow P(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}, \quad x_i = 0, 1$$

Then $\mu = p$, $\sigma^2 = p(1-p)$ and the sample proportion $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$ is also the sample mean

By C.L.T, $Z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \overset{\cdot}{\sim} N(0,1)$

Alternatively $Z^* = \dfrac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}} \overset{\cdot}{\sim} N(0,1)$ if n is large

Therefore, 100(1-α)% CI for p is $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

3). A. Large sample test on one population proportion p

(Both # of successes and failures are ≥5)

By C.L.T:

Point estimator: $\hat{p} \dot{\sim} N(p, \frac{p(1-p)}{n})$

Pivotal quantity: $Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \dot{\sim} N(0,1)$ Alternatively $Z^* = \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \dot{\sim} N(0,1)$

$$\begin{cases} H_0 : p = p_0 \\ H_a : p < p_0 \end{cases}$$

T.S: $Z_0 = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{H_0}{\sim} N(0,1)$

At α, reject H_0 iff $Z_0 \le -Z_\alpha$

Homework: Finish the test problem given in part 3 of Example 1.

Example 2. Thanksgiving was coming up and Harvey's Turkey farm was doing a land-office business. Harvey sold 100 gobblers to Nedicks for their famous Turkey-dogs. Nedicks found that

90 of Harvey's turkeys were in reality peacocks. At α=0.05, Test if the proportion of turkey is

greater than 0.05

Solution:

$$\begin{cases} H_0 : p = 0.05 \\ H_a : p > 0.05 \end{cases}$$ n=100, X=10, n-X=90, $\Rightarrow$ large sample

T.S : $Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \overset{H_0}{\sim} N(0,1) \quad \Rightarrow \quad Z_0 = \dfrac{0.1 - 0.05}{\sqrt{\dfrac{0.05 * 0.95}{100}}} \approx 2.29$

p-value=0.011

At α=0.05, since Z_0=2.29≥ Z_α=1.625, reject H_0.

Also since p-value=0.011<0.05, reject H_0.

Power and sample size calculation

$$\begin{cases} H_0 : p = p_0 \\ H_a : p = p_1 > p_0 \end{cases}$$

At the significance level α, we reject H_0 if $Z_0 \ge z_\alpha$.

$$\beta = P(\,fail\ to\ reject\ H_0 \mid H_a)$$
$$1-\beta = P(reject\ H_0 \mid H_a) = P(Z_0 \ge z_\alpha \mid H_a : p = p_1 > p_0)$$
$$= P(\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \ge z_\alpha \mid p = p_1)$$

$$= P\left(\frac{\hat{p}}{\sqrt{\frac{p_0(1-p_0)}{n}}} - \frac{p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \ge z_\alpha \mid H_a: p = p_1 > p_0\right)$$

$$= P\left(\hat{p} - p_1 \ge z_\alpha\sqrt{\frac{p_0(1-p_0)}{n}} + p_0 - p_1 \mid p_1\right)$$

$$= P\left(\frac{\hat{p}-p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} \ge \frac{p_0-p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} + z_\alpha\sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}} \mid p_1\right)$$

$$\therefore -Z_\beta = \frac{p_0-p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} + z_\alpha\sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}}$$

$$\therefore\ n = \left[\frac{z_\alpha\sqrt{p_0(1-p_0)} + Z_\beta\sqrt{p_1(1-p_1)}}{p_1-p_0}\right]^2$$

Note, the above formula is true for any one-sided test. For a 2-sided test, replace z_α with $z_{\frac{\alpha}{2}}$

B. Exact test on one population proportion p (*optional).

Data: sample size n, # “S”: X, # “F”: n-X

X~Bin(n, p), P(X=x)= $\binom{n}{x} p^x(1-p)^{n-x}$, x = 0,…, n

(1) $\begin{cases} H_0: p = p_0 \\ H_a: p > p_0 \end{cases}$

$$p-value = P(X \ge x \mid H_0: p = p_0) = P(X = x \mid H_0) + P(X = x+1 \mid H_0) + \ldots + P(X = n \mid H_0)$$

$$= \sum_{i=x}^{n}\binom{n}{i} p_0^i(1-p_0)^{1-i}$$

(2) $\begin{cases} H_0 : p = p_0 \\ H_a : p < p_0 \end{cases}$

$$p-value = P(X \le x \mid H_0 : p = p_0) = \sum_{i=0}^{x} \binom{n}{i} p_0^i (1-p_0)^{1-i}$$

(3) $\begin{cases} H_0 : p = p_0 \\ H_a : p \ne p_0 \end{cases}$

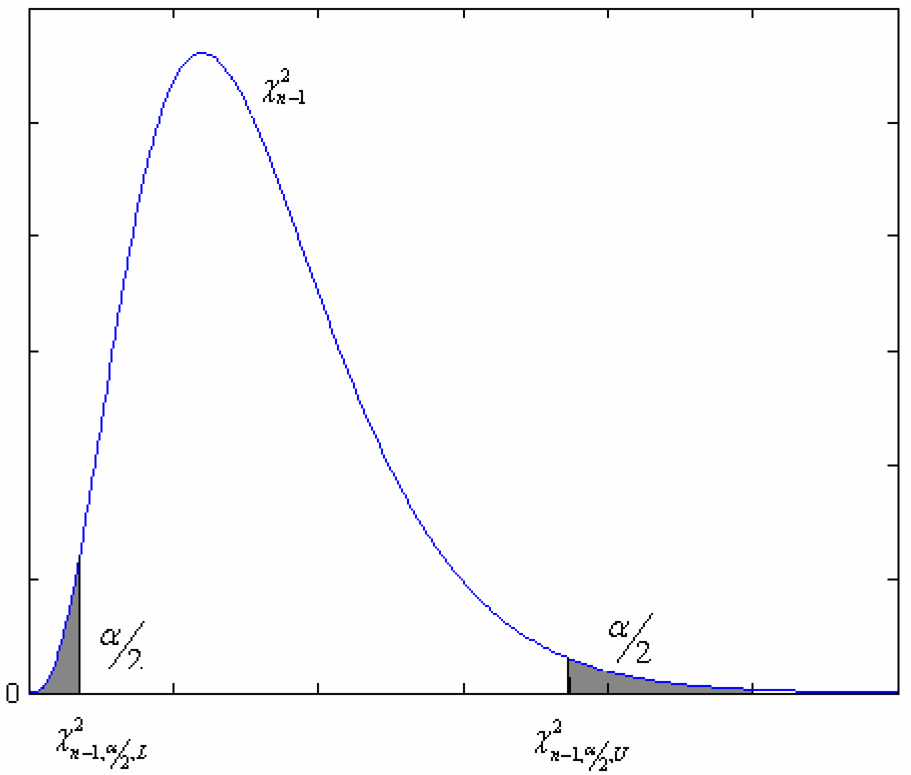
$$p-value = 2*\min\{P(X \le x \mid H_0), P(X \ge x \mid H_0)\}.$$

2. Inference on one population variance σ^2, population is normal

1) Review: Point estimator: $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ and s^2 is unbiased estimator of σ^2

Pivotal Quantity for the inference on σ^2 (P.Q.): $W = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$

2) Review: Confidence Interval for σ^2 :



$$P(\chi^2_{n-1,L,\alpha/2} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{n-1,U,\alpha/2}) = 1-\alpha$$

$$P(\frac{n-1}{\chi^2_{U,\alpha/2}} \le \frac{\sigma^2}{s^2} \le \frac{n-1}{\chi^2_{L,\alpha/2}}) = 1-\alpha$$

$$P(\frac{(n-1)s^2}{\chi^2_{U,\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{L,\alpha/2}}) = 1-\alpha$$

Hence, the $100(1-\alpha)\%$ confidence interval for σ^2 is $[\frac{(n-1)s^2}{\chi^2_{n-1,U,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,L,\alpha/2}}]$

3) Test for σ^2

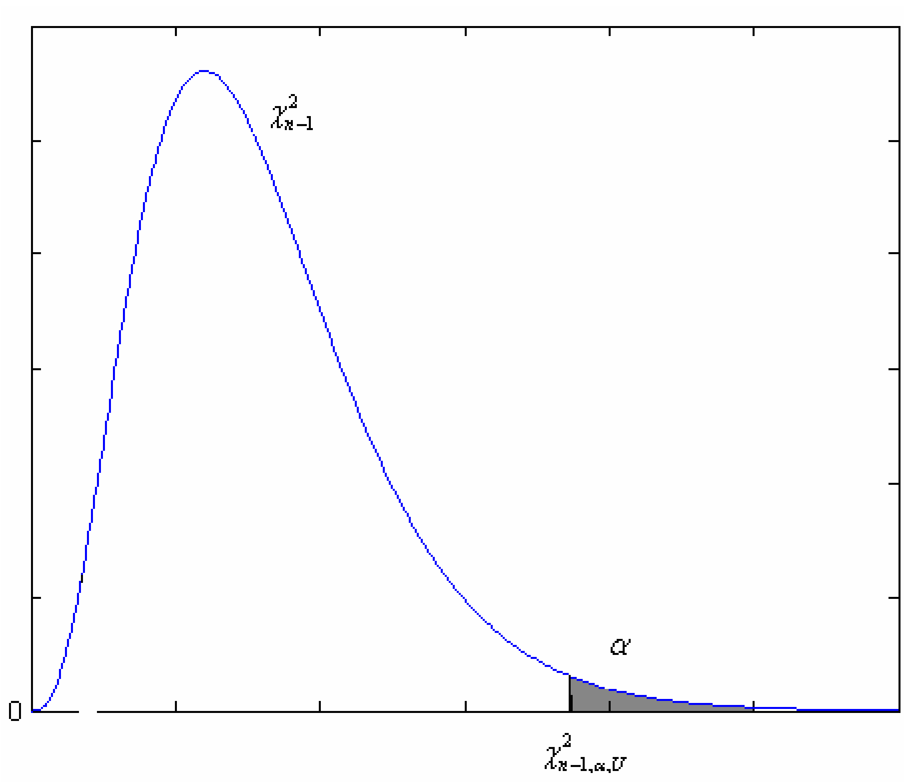
$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_a : \sigma^2 > \sigma_0^2$$

Test statistic: $W_0 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$ when Ho is true.

At the significance level α , we reject Ho if $W_0 \ge \chi^2_{n-1,\alpha,U}$

$\alpha = P(W_0 \ge c \mid \sigma^2 = \sigma_0^2)$

Then $c = \chi^2_{n-1,\alpha,U}$



Example 3.

Home buyers can choose a variety of ways to finance mortgages, ranging from fixed-rate thirty-year notes to one-year adjustable, where interest rates can move up or down from year to year. During the first quarter of 1994, Tennessee lenders were charging an average rate of 8.84% on a $100,000 loan amortized over thirty years; the standard deviation from bank to bank was 0.10%. n=9 lenders for one-year adjustable and the sample standard deviation for those nine interest rates is s=0.22%. Do these data lend credence to the speculation that rates for one-year adjustable are more variable than rates for conventional mortgages? Assume the population is normal.

Solution: This is inference on one population variance, normal population.

First, we set the hypotheses:

$$H_0 : \sigma^2 = (0.10)^2$$
$$H_a : \sigma^2 > (0.10)^2$$

Test statistic is: $W_0 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$ when the null hypothesis is true.

At significance level $\alpha = 0.05$, we reject H_0 if $W_0 \geq \chi^2_{n-1,\alpha,upper}$

In this problem, we now have W_0=38.72>15.507 we reject H_0 at $\alpha = 0.05$.

On the other hand, we can use P-value as well:

Since the P-value of W_0 is less than 0.01, which is less than $\alpha = 0.05$, we reject the null hypothesis at $\alpha = 0.05$

There are two other hypotheses:

1) $H_0 : \sigma^2 = \sigma_0^2$, $H_a : \sigma^2 < \sigma_0^2$ & 2) $H_0 : \sigma^2 = \sigma_0^2$, $H_a : \sigma^2 \neq \sigma_0^2$

The test statistic is the same as above: $W_0 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$

1) We reject Ho when $W_0 \le \chi^2_{n-1,\alpha,lower}$

2) We reject Ho when $W_0 \le \chi^2_{n-1,\alpha/2,lower}$ or $W_0 \ge \chi^2_{n-1,\alpha/2,upper}$

Note, the following notations are equivalent:

$$\chi^2_{n-1,\alpha/2,L} = \chi^2_{n-1,\alpha/2,lower} = \chi^2_{n-1,L,\alpha/2}$$

$$\chi^2_{n-1,\alpha/2,U} = \chi^2_{n-1,\alpha/2,upper} = \chi^2_{n-1,U,\alpha/2}$$

$$\chi^2_{n-1,\alpha,U} = \chi^2_{n-1,\alpha,upper} = \chi^2_{n-1,U,\alpha}$$

$$\chi^2_{n-1,\alpha,L} = \chi^2_{n-1,\alpha,lower} = \chi^2_{n-1,L,\alpha}$$