Ch 6 : Story of Sample Mean under Normality

Contents

1

Theoretical Consideration

[ToC]

1.1 Sum of Two Normals

[ToC]

- Suppose $X \sim N(2,3)$ and $Y \sim N(4,2)$. X and Y are independent.
- What is $P(X+Y<5) = ?$

2

Sample Mean and Confidence Interval

[ToC]

2.1 Sample Mean under Normality

[ToC]

$$\bar{X} \sim N\Big(\mu, \sigma^2/n\Big)$$

$100(1-\alpha)\%$ CI for μ is

$$\bar{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

2.2 Confidence Interval for μ

Suppose your data $X_1, \ldots, X_n$ are Random Sample from distributioin A with mean μ and variance σ^2. Now we know that $\bar{X} \sim N(\mu, \sigma^2/n)$.

Then we know that

$$P(\bar{X} \text{ is within } \mu \pm z_{\alpha/2}\sigma/n) = 1 - \alpha$$

Which is same thing as to say

$$P(\mu \text{ is within } \bar{X} \pm z_{\alpha/2}\sigma/n) = 1 - \alpha.$$

Thus, our $100(1 - \alpha)\%$ Confidence Interval for μ is

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{n}$$

2.3 Confidence Interval Formula

Let $X_1, \ldots, X_n$ be Random Sample from $\mathrm{N}(\mu, \sigma^2)$ distribution, and assume σ is known. then we have sample distribution,

$$\overline{X} \sim N(\mu, \sigma^2/n).$$

Confidence Interval

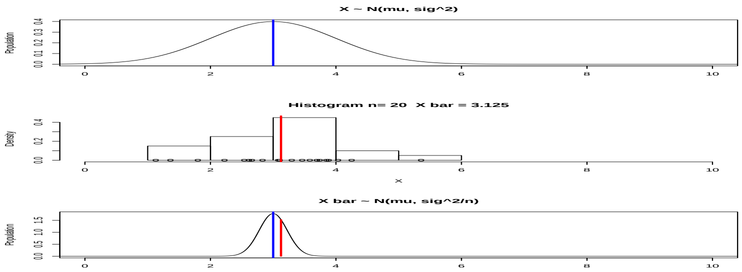
$100(1-\alpha)\%$ Confidence Interval for μ is

$$\left(\overline{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) \qquad \text{(Two-sided)}$$

$$\left(-\infty, \quad \overline{X} + z_{\alpha} \frac{\sigma}{\sqrt{n}}\right) \qquad \text{(one-sided upper-bound)}$$

$$\left(\overline{X} - z_{\alpha} \frac{\sigma}{\sqrt{n}}, \quad \infty\right) \qquad \text{(one-sided lower-bound)}$$

X bar Demo




```
n<-120; mean<-3; sd<-1;  layout( matrix(c(1,2,3), 3,1))

t <- seq(-1,11,.01); f0<-dnorm(t,mean,sd)
plot(t, f0, type='l', xlim=c(0,10), xlab="", ylab="Population", main="X ~ N(mu, sig^2)", axes=T)
abline(v=mean, col="blue", lwd=2)

X <- rnorm(n, mean, sd)
hist(X, freq=F, bins=10, xlim=c(0,10), main=paste("Histogram n=",n," X bar =", round(mean(X),3) ))
lines(X, rep(0,n), type="p")
abline(v=mean(X), col="red", lwd=2)

t <- seq(-1,11,.01); f0<-dnorm(t,mean,sd/sqrt(n))
plot(t, f0, type='l', xlim=c(0,10), xlab="", ylab="Population", main="X bar ~ N(mu, sig^2/n)", axes=T)
abline(v=mean, col="blue", lwd=2)
lines( rep(mean(X),2), c(0,dnorm(mean(X),mean,sd/sqrt(n))), col="red", lwd=2)
```

2.4 Examples:

Michelson 1879 Speed of Light Experiment

```
                                                  Certified Values
Sample Mean                                xbar:  299.852400000000
Sample Standard Deviation (denom. = n-1)      s:  0.0790105478190518

Number of Observations:                             100
```

Ex: Bus Routs

A metropolitan transit authority wants to determine whether there is any need for changes in the frequency of service over certain bus routes. Wants to know if average miles traveled per person by all residents in the area is 5miles or less. n=120, sample mean=4.66, assume population SD = 1.5

Ex: Daily Sales

Average daily sales at small food store are known to be \$452.8. The manager recently implemented some changes in store interior, and want to know if the sales have improved. For last 12 days, the sales averaged \$501.9, and sample SD=\$65. Is the change significant?

2.5 One Sided CI

- Take two-sided CI formula, change $\alpha/2$ to α.
- Pick one of the sign for upper-bound or lower-bound.

- $100(1-\alpha)\%$ One-sided upper-bound Confidence Interval for μ is

$$\left(-\infty, \quad \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

- $100(1-\alpha)\%$ One-sided lower-bound Confidence Interval for μ is

$$\left(\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}, \quad \infty\right)$$

Example: Hydro Turbines

- Old model of hydroelectric miniturbines averages 25.2 Kwatt output under the lab condition.
- Recently the model design was changed, and that supposed to improve the average output.
- Out of 10 units tested, sample mean 27.1, sigma is known to be 3.2.
- Is this result enough to claim the improvement?

3

When we can’t assume the normalilty

[ToC]

3.1 Central Limit Theorem

- If $X_1, \ldots, X_n$ are R.S. from a population with **any distribution** with mean μ and standard deviation σ, then

$$\bar{X} \text{ is approximately distributed as } N(\mu, \sigma^2/n)$$

  if n is large enough (> 40).

- Larger n, better the approximation.

- This means that when n is large, we can use z-test (and z-CI) regardless of the population distribution.

Example: Mercury in Bass

- Study to investigate the mercury contamination in largemouth bass.
- A sample of 53 fish in a lake was selected, and their mercury concentration in the muscle tissue was measured.
- Sample mean was 0.525(ppm) and Sample SD was 0.349.
- Is this the evidence that $\mu > .4ppm$?

```
D = c(1.230,  0.490,  0.490,  1.080,  0.590,  0.280, 0.940,  1.330,  0.190,  1.160,
      0.180,  0.100,  0.980,  0.340,  0.340,  0.190, 0.210,  0.400,  0.860,  0.430,
      0.040,  0.830,  0.050,  0.630,  0.340,  0.750, 0.044,  0.810,  0.150,  0.040,
      0.560,  0.840,  0.870,  0.490,  0.520,  0.250, 0.190,  0.410,  0.500,  0.560,
      1.200,  0.710,  1.100,  0.650,  0.270,  0.270, 0.500,  0.770,  0.730,  0.340,
      0.170,  0.160,  0.270)

mean(D); sd(D)
qqnorm(D)

c(mean(D) - 1.96*sd(D)/sqrt(53),  mean(D) + 1.96*sd(D)/sqrt(53))
```

Example: A/C unit lifetime

- Suppose lifetime of A/C unit can be modeled by Exponential distribution.
- Sample of 60 units have averaged in 4.75 years.
- Calculate 95% CI for true average of A/C units of same kind.

Example: A/C unit lifetime

- Suppose lifetime of A/C unit can be modeled by Exponential distribution.
- Sample of 60 units have averaged in 4.75 years.
- Calculate 95% CI for true average of A/C units of same kind.

Example: Casino Sales

- Customers pay \$1 to play, have probability p to win \$100.
- if $p = .05$, $\mathrm{E}(\text{profit}) = .5$, $\mathrm{Var}(\text{profit}) = 49.75$,

4

When answers are 1's and 0's

[ToC]

Example: Catching a Cheater

- Flip a coin. Let $X_i = 1$ if head, 0 otherwise.
- If p is .5, what is the confidence interval for p?

5

When σ is unknown

[ToC]

5.1 When we have to Estimate σ as well

Suppose $X_1, \ldots, X_n$ be Random Sample from $\mathrm{N}(\mu, \sigma^2)$ distribution, but now σ is unknown. Then we still have

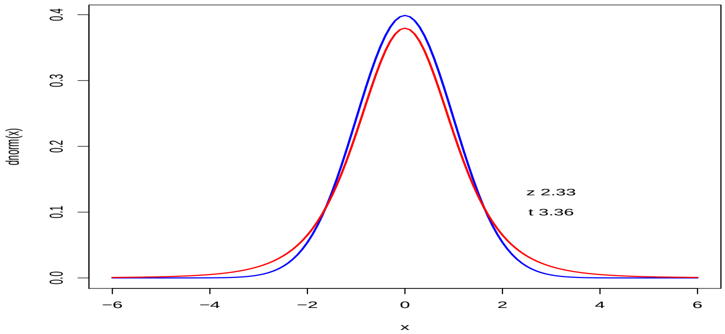
$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

But do not know σ. Then if we use S instead of σ, we have slightly different distribution,

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

Where $t(n-1)$ is t-distribution with degrees of freedom $n-1$.

5.1.1 Student t-distribution



```r
draw.t <- function(df=5, a=0){

  x <- seq(-6,6,.1)
  plot(x, dnorm(x), type='l', col="blue", lwd=2)
  lines(x, dt(x,df), col="red", lwd=2)

  if (a!=0) {
    c1<- qnorm(1-a);  c2 <- qt(1-a,df)
    text(3, .13, paste("z", round(c1,2)))
    text(3, .1, paste("t", round(c2,2)))
  }
}

draw.t(5)
draw.t(5, .05)
```

Confidence Interval with t-distribution

- $100(1-\alpha)\%$ (two-sided) Confidence Interval for μ is

$$\left(\overline{X} \pm t_{\frac{\alpha}{2},n-1} \frac{\sigma}{\sqrt{n}}\right)$$

- For one-sided upper- or lower- CI, pick one of $+$ or $-$ sign above, and change $\frac{\alpha}{2}$ to α.

Example: Tire Life

- The manufacturer of a new fiberglass tire claims that its average life will be at least 40,000 miles.
- To verify this claim a sample of 12 tires is tested, with their lifetimes (in 1,000s of miles) being as follows:
- Sample mean: 38.04. Sample SD: 2.49.

```
D = c(36.1,  40.2,  33.8,  40.9,  38.5,  42,    35.8,  37.2, 37,    41,    36.8,  37.2)

mean(D); sd(D)
qqnorm(D)
qt(.975, 12-1)

c(mean(D) - 2.200*sd(D)/sqrt(12),  mean(D) + 2.200*sd(D)/sqrt(12))
```

Example: Heat Transfer

- An article in the Journal of Heat Transfer (Trans. ASME, Sec. C, 96. 1974. p. 59) described a new method of measuring the thermal conductivity of Armco iron.
- Using a temperature of 100F and a power input of 550 wtts, 10 measurements of thermal conductivity (in Btu/hr-ft-F) were obtained:
- $\bar{X} = 41.92$ $S = .284$

```
D = c(41.60, 41.48, 42.34, 41.95, 41.81,  41.86, 42.18, 41.72, 42.26, 42.04)

mean(D); sd(D)
qqnorm(D)
qt(.975, 10-1)

c(mean(D) - 2.262*sd(D)/sqrt(10),  mean(D) + 2.262*sd(D)/sqrt(10))
```

Example: pH meter bias

- Suppose that an engineer is interested in testing the bias in a pH meter.
- Data are collected on a neutral substance (pH=7.0). A sample of the measurements were taken with the data as follows:
- Sample mean: 7.025, Sample SD 0.044.

```
D = c(7.07, 7.00, 7.10, 6.97, 7.00, 7.03, 7.01, 7.01, 6.98, 7.08)

mean(D); sd(D)
qqnorm(D)
qt(.975, 10-1)

c(mean(D) - 2.262*sd(D)/sqrt(10),  mean(D) + 2.262*sd(D)/sqrt(10))
```

6

Confidence Interval for σ^2

[ToC]

6.1 Confidence Interval for σ^2

- Sample SD S has sample distribution

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

- Two-sided $100(1-\alpha)\%$ Confidence Intervals for σ^2

$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \ \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right) \qquad \text{(two-sided)}$$

- For one-sided CI, take one of them, and change $\alpha/2$ to α.
- For CI for σ, take squareroot of above formulas.

6.2 Prediction Interval

If $X_1, \ldots, X_n$ are R.S from $N(\mu, \sigma^2)$, then

- $100(1-\alpha)\%$ prediction interval for the next ovservation X_{n+1} is

$$X_{n+1} \text{ is within } \left(\overline{X} \pm Z_{\frac{\alpha}{2}} \sigma \sqrt{\frac{1}{n} + 1}\right)$$

- If we need to replace σ with S, then replace $z_{\frac{\alpha}{2}}$ with $t_{\frac{\alpha}{2}, n-1}$.

7

Summary and Formulas

[ToC]

7.1 Summary and Formulas

[ToC]

Sampling distribution of sample mean

- If X_i are random sample from normal distribution with mean μ and SD σ,

$$\bar{X} \sim N\big(\mu, \sigma^2/n\big)$$

- If $n > 40$, then above is approximately true even if it is not from the normal distribution. (e.g. Exponential Distribution)

- If $np > 10$ and $n(1-p) > 10$ are both true, same goes for proportionn estimator

$$\bar{X} = \hat{p} \sim N\big(\mu, \sigma^2/n\big) = N\big(p, p(1-p)/n\big)$$

Confidence Interval

- Above result leads to $100(1-\alpha)\%$ two-sided Confidence Interval for μ,

$$\bar{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

- And $100(1-\alpha)\%$ two-sided Confidence Interval for p,

$$\hat{p} \pm z_{\frac{\alpha}{2}} \frac{p(1-p)}{\sqrt{n}}$$

- If σ is not known, replace with sample standard deviation S, and change $z_{\frac{\alpha}{2}}$ to $t_{\frac{\alpha}{2},n-1}$.

- For one-sided upper- or lower-bound CI, pick $+$ or $-$ sign, and change $\frac{\alpha}{2}$ to α.

Sample Variance

- Sampling distribution of the sample variance

$$(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$$

- $100(1-\alpha)\%$ two-sided Confidence Intervals for σ^2

$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \ \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)$$

- For one-sided CI, take one of them, and change $\alpha/2$ to α.

- For CI for σ, take squareroot of above formulas.