Parameters estimation

May 2, 2019

Parameter estimation

Estimation

- Parameters (unknown): Mean, variance, standard deviation ...
- Statistics (from data): Sample mean, sample variance ...
- Can use statistics to estimate parameter
- How acurate?

Distribution of statistics

- Suppose $X \sim \mathrm{N}(\mu, \sigma^2)$
- $X_1, \ldots, X_n$ sample of X
- $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$
- $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$
- $\sqrt{n}\dfrac{(\bar{X}-\mu)}{S} \sim t_{n-1}$

Estimate mean

- Let X be a random variable from a population
- Take samples $X_1, \ldots, X_n$ from the population,
- Calculate $\bar{X}$, then $E(\bar{X}) = \mu$
- $\bar{X}$ is *unbiased estimator* of μ
- Use $\bar{X}$ to estimate μ

Interval estimate

- We don't expect $\bar{X}$ to be exactly μ
- Want to find an interval around $\bar{X}$ so we can be sure that μ is in it.
- Ex: want to find $[a, b]$ so that 95% of the time $\mu \in [a, b]$
- $[a, b]$ is called *95% confidence interval estimate* of μ

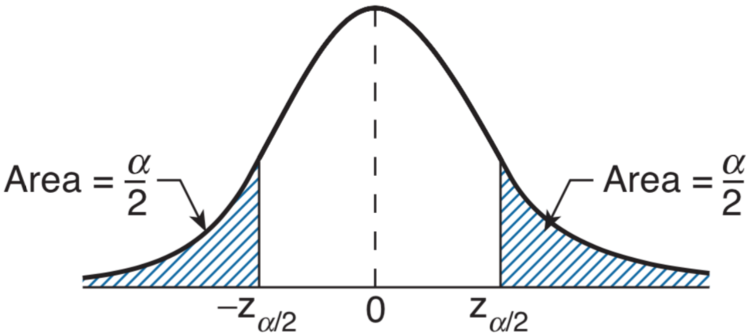
Confidence level

- $Z \sim \mathrm{N}(0, 1)$
- Let $\alpha \in (0, 1)$
- z_α is a positive number so that

$$\Phi(z_\alpha) = 1 - \alpha$$

The z-test

- $P(Z > z_\alpha) = \alpha$
- $P(Z < -z_\alpha) = \alpha$
- $P(-z_\alpha < Z < z_\alpha) = 1 - 2\alpha$
- $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$


Area = $\frac{\alpha}{2}$
Area = $\frac{\alpha}{2}$
$-z_{\alpha/2}$
0
$z_{\alpha/2}$

Estimate the mean when the variance is known

Estimate interval

- Suppose we know σ
- $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$
- $P(-z_{\alpha/2} < \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1-\alpha$
- $P(\bar{X} - z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}) = 1-\alpha$

Example

Suppose that when a signal having value μ is sent, then the value received is $\mu + N$ where the noise N is normal with mean 0 and variance 4. To reduce error, suppose the same value is sent 9 times. If the successive values received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, construct a 95 percent confidence interval for μ.

Solution

- $n = 9, \bar{X} = 9$
- $\alpha = 0.05, z_{\alpha/2} = 1.96$
- 95% confidence interval of μ is

$$(9 - 1.96\frac{2}{3}, 9 + 1.96\frac{2}{3}) = (7.69, 10.31)$$

One sided interval

- $P(\mu < \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$
- $(-\infty, \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}})$ is $100(1-\alpha)$ percent lower confidence interval of μ.
- $P(\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}} < \mu) = 1 - \alpha$
- $(\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty)$ is $100(1-\alpha)$ percent upper confidence interval of μ.

Example

From past experience it is known that the weights of salmon grown at a commercial hatchery are normal with a mean that varies from season to season but with a standard deviation that remains fixed at 0.3 pounds. If we want to be 95 percent certain that our estimate of the present season's mean weight of a salmon is correct to within ± 0.1 pounds, how large a sample is needed?

Solution

- 95% confidence interval of μ is

$$(\bar{X} - 1.96\frac{0.3}{\sqrt{n}}, \bar{X} + 1.96\frac{0.3}{\sqrt{n}})$$

- Want $1.96\frac{0.3}{\sqrt{n}} \leq 0.1$
- then $n \geq 35$

Estimate the mean when the variance is not known

The t-test

- Suppose $Z \sim \text{N}(0,1)$
- $C \sim \chi^2_n$
- then $T = \frac{Z}{\sqrt{C/n}}$ has $t-$distribution of n degree of freedom.
- $t_{\alpha,n}$ is a number so that

$$P(T > t_{\alpha,n}) = \alpha$$

Interval estimate

$$\frac{(\bar{X} - \mu)}{S/\sqrt{n}} \sim t_{n-1} \text{ then}$$

$$P(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}})$$
$$= 1 - \alpha$$

$(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}})$ is $100(1-\alpha)$ confidence interval of μ

Example

Suppose that when a signal having value μ is sent, then the value received is $\mu + N$ where N , representing noise, is normal with mean 0 and variance σ^2. To reduce error, suppose the same value is sent 9 times. If the successive values received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, construct a 95 percent confidence interval for μ.

Solution

- $\bar{X} = 9$, $S^2 = 9.5$, $S = 3.082$
- $\alpha = .05$, $t_{.025,8} = 2.306$
- 95% confidence interval for μ

$$(9 - 2.306\frac{3.082}{3}, 9 + 2.306\frac{3.082}{3}) = (6.63, 11.37)$$

Notice

- The $t-$test gives larger interval than the $z-$test
- We assumed X is normal
- If X is not normal but n is large we can still use these methods because $\bar{X}$ is approximately normal (central limit theorem).

Practice

Determine a 95 percent confidence interval for the average resting pulse of the members of a health club if a random selection of 15 members of the club yielded the data 54, 63, 58, 72, 49, 92, 70, 73, 69, 104, 48, 66, 80, 64, 77. Do 2 cases: know variance is 9 and not know variance.

Estimation of variance

Chi square test

- $Y \sim \chi^2_n$
- $\chi^2_{\alpha,n}$ is a number that

$$P(Y > \chi^2_{\alpha,n}) = \alpha$$

- $P(\chi^2_{1-\alpha/2,n} < Y < \chi^2_{\alpha/2,n}) = 1 - \alpha$

Estimate variance

Use S^2 to estimate σ^2

$$P(\chi^2_{1-\alpha/2,n} < (n-1)\frac{S^2}{\sigma^2} < \chi^2_{\alpha/2,n}) = 1-\alpha$$

or

$$P(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n}}) = 1-\alpha$$

Approximate confidence interval for mean of Bernoulli RV

- Sample n independent trials from a population, each success with unknown probability p
- X: number of successes in n sample trials
- $X \sim \text{Bino}(n, p) \approx \text{N}(np, np(1-p))$
- Want to find confidence interval for p

Estimator

- Let $\hat{p} = \dfrac{X}{n}$ then $E(\hat{p}) = p$
- Use $\hat{p}$ as unbiased estimator for p
- $np(1-p) \approx n\hat{p}(1-\hat{p})$
- $\dfrac{X - np}{\sqrt{n\hat{p}(1-\hat{p})}} \approx N(0,1)$

Interval estimate

$$P(-z_{\alpha/2} < \frac{X - np}{\sqrt{n\hat{p}(1-\hat{p})}} < z_{\alpha/2}) \approx 1 - \alpha$$

$$P(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$$
$$\approx 1 - \alpha$$

Example

On October 14, 2003, the New York Times reported that a recent poll indicated that 52 percent of the population was in favor of the job performance of President Bush, with a margin of error of ± 4 percent and 95% confidence level. Can we infer how many people were questioned?

Solution

- $\alpha = .05$, $z_{.025} = 1.96$
- $\hat{p} = .52$
- 95% confidence interval is given by

$$.52 \pm 1.96\sqrt{.52(.48)/n}$$

- $1.96\sqrt{.52(.48)/n} = .04$
- $n = 599$

Homework

Chapter 7: 9, 13, 14, 15, 18, 36, 48, 50