statistics

FOR

DATA SCIENCE

UNIT-3

feedback/corrections: vibha@pesu.pes.edu

Vibha Masti

point ESTIMATORS

- Sample statistic calculated from the data used to estimate an unknown parameter
- Single numeric value specified for the data
- Infers about population parameters

Point estimators $\longrightarrow$ unknown population parameters

$\bar{x} \longrightarrow \mu$

$s \longrightarrow \sigma$

$\hat{p} \longrightarrow p$

Properties of Point Estimators

1. Bias — difference in value
2. Consistency — large samples
3. Efficiency — small variance, unbiased, consistent

Mean Squared Error

- Combines bias (accuracy) and uncertainty (precision)

difference of mean of estimator and true value

standard deviation

Bias

$$\text{Bias}(\hat{\theta}) = \mu_{\hat{\theta}} - \theta = \mu_{\hat{\theta}} - \mu_{\theta} = \mu_{(\hat{\theta}-\theta)} = E(\hat{\theta}-\theta)$$

θ: unknown parameter

$\hat{\theta}$: estimator of $\hat{\theta}$

Uncertainty

$\text{Uncertainty}(\hat{\theta}) = \sigma_{\hat{\theta}}$ standard error

$$\boxed{MSE_{\hat{\theta}} = Var(\hat{\theta}) + (Bias(\hat{\theta}))^2}$$

$$MSE_{\hat{\theta}} = \mu_{\underbrace{(\hat{\theta}-\theta)^2}_{\text{error}^2}} = \sigma_{\hat{\theta}}^2 + (\mu_{\hat{\theta}} - \theta)^2$$

Derivation

mean of error²

$$= \mu_{(\hat{\theta}-\theta)^2}$$

$$= E[(\hat{\theta}-\theta)^2]$$ expectation value

$$= E[\hat{\theta}^2 + \theta^2 - 2\hat{\theta}\theta]$$

$$= E[\hat{\theta}^2] + E[\theta^2] - 2E[\hat{\theta}\theta]$$

$$= E[\hat{\theta}^2] - (E[\hat{\theta}])^2 + (E[\hat{\theta}])^2 + E[\theta^2] - 2E[\hat{\theta}\theta]$$

$$= \underbrace{\mu_{\hat{\theta}^2} - (\mu_{\hat{\theta}})^2}_{\text{variance}} + (E[\hat{\theta}])^2 + \theta^2 - 2\theta E[\hat{\theta}]$$

$$= var(\hat{\theta}) + (E[\hat{\theta}] - \theta)^2$$

$$= Var(\hat{\theta}) + \underbrace{(\mu_{\hat{\theta}} - \theta)^2}_{\text{bias}^2}$$

Q1. Let $X \sim Bin(n, p)$ where p is unknown. Find MSE of $\hat{p} = \frac{X}{n}$

$$MSE = (Bias)^2 + Variance$$

$$Bias = \mu_{\hat{p}} - p = \mu_{\frac{X}{n}} - p = \frac{\mu_X}{n} - p = \frac{np}{n} - p = p - p = 0$$

$$\text{Std. dev} = \sigma_{\hat{p}} = \sigma_{\frac{X}{n}} = \frac{\sigma_X}{n} = \frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{p(1-p)}{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$$Variance = \sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$$

$$\boxed{MSE = \frac{p(1-p)}{n}}$$

MAXIMUM LIKELIHOOD ESTIMATE

- Value of estimators that when substituted in for the parameters maximises the likelihood function

pdf / pmf or joint pdf / pmf

Steps

1. Write down likelihood function
2. Take natural log of likelihood function
3. Differentiate the log of likelihood function
4. Set derivative equal to 0 to get MLE.

Normal Distribution
LIKELIHOOD FUNCTION

$$f(x_i;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}$$

Joint pdf for $x_1, x_2 \dots x_n$

$$f(x_1,\dots,x_n;\mu,\sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}$$

$$L(\mu,\sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\sum_{i=1}^{n} \frac{-(x_i-\mu)^2}{2\sigma^2}}$$

- Taking ln on both sides

$$\ln(L(\mu,\sigma)) = -n\ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$f = \ln(L) = -n\ln\sigma - \frac{n}{2}\ln(2\pi) - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}$$

MLE of MEAN (μ)

$$\frac{\partial f}{\partial \mu} = \frac{-\partial}{\partial \mu}\left(\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right) = 0$$

$$-\frac{1}{2\sigma^2} \times \frac{\partial}{\partial \mu}\left(\sum_{i=1}^{n} (x_i-\mu)^2\right) = 0$$

$$\sum_{i=1}^{n} 2(x_i-\mu)(-1) = 0$$

$$\sum_{i=1}^{n} (x_i-\mu) = 0$$

$$\sum_{i=1}^{n} x_i - n\mu = 0$$

$$\boxed{\mu = \frac{\sum_{i=1}^{n} x_i}{n}} \qquad \boxed{\mu = \overline{x}}$$

MLE of STANDARD DEVIATION (σ)

$$\frac{\partial F}{\partial \sigma} = \frac{\partial}{\partial \sigma}(-n \ln \sigma) - \frac{\partial}{\partial \sigma}\left(\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right) = 0$$

$$\frac{-n}{\sigma} - \sum_{i=1}^{n}\left(\frac{(x_i-\mu)^2}{2} \frac{(-2)}{\sigma^3}\right) = 0$$

Multiply both sides by σ^3

$$-n\sigma^2 + \sum_{i=1}^{n} (x_i-\mu)^2 = 0$$

$$\boxed{\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i-\overline{x})^2}{n}}}$$

Bernoulli Distribution
LIKELIHOOD FUNCTION

- $X_1, X_2 \ldots X_n$ (random samples/variables)
- The probability mass function (discrete) or probability density function (continuous) of each X_i is given as

$$f(x_i ; \theta)$$

- The joint pmf / pdf (and – intersection)

$$L(\theta) = P(X_1 = x_1, X_2 = x_2 \ldots X_n = x_n)$$

if they are independent

$$= f(x_1; \theta) \ldots f(x_n; \theta)$$

$$= \prod_{i=1}^{n} f(x_i ; \theta) \quad \text{(product)}$$

parameter (population)

- For Bernoulli distribution, $\theta = p$

- pmf of each X_i is

$$f(x_i ; p) = p^{x_i} \; p^{1-x_i}$$

- $$f(x_1, x_2 \ldots x_n ; p) = p^{x_1}(1-p)^{1-x_1} \ldots p^{x_n}(1-p)^{1-x_n}$$

joint probability

$$p(X_1 = x_1, \ldots X_n = x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = L(p)$$

$$L(p) = p^{x_1}(1-p)^{1-x_1} \cdots p^{x_n}(1-p)^{1-x_n}$$

$$= p^{x_1} \cdots p^{x_n} (1-p)^{1-x_1} \cdots (1-p)^{1-x_n}$$

$$= p^{\sum_{i=1}^{n} x_i} (1-p)^{\sum_{i=1}^{n} 1-x_i}$$

- Taking ln on both sides

$$\ln L = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

- Taking derivative wrt p on both sides

$$\frac{d(\ln L)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{(1-p)} = 0$$

$$\frac{\sum_{i=1}^{n} x_i}{p} = \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

$$(1-\cancel{p}) \sum_{i=1}^{n} x_i = np - \cancel{p \sum_{i=1}^{n}} x_i$$

$$np = \sum_{i=1}^{n} x_i$$

$$\boxed{p = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}}$$

MLE of p is $\bar{x}$

Binomial Distribution

LIKELIHOOD FUNCTION

- pdf of $X \sim \text{Binomial}(n, p)$

$$f(x; n, p) = {}^nC_x \, p^x (1-p)^{n-x}$$

$$L(p; n, p) = {}^nC_x \, p^x (1-p)^{n-x}$$

- Taking ln on both sides

$$\ln(L) = \ln\left(\frac{n!}{(n-x)!\,x!}\right) + x \ln p + (n-x)\ln(1-p)$$

- Taking derivative wrt p

$$\frac{d}{dp}(\ln(L)) = \frac{x}{p} - \frac{(n-x)}{1-p} = 0$$

$$x(1-p) = p(n-x)$$

$$x - \cancel{xp} = np - \cancel{xp}$$

$$\boxed{p = \frac{x}{n}}$$

MLE of p is $\hat{p} = \frac{x}{n}$

Poisson Distribution

LIKELIHOOD FUNCTION

- pdf of $X \sim \text{Poisson}(\lambda)$

$$f(x;\lambda) = e^{-\lambda} \frac{\lambda^{x}}{x!}$$

- joint pmf / likelihood function

$$L(\lambda) = f(x_1, \dots x_n; \lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}$$

$$L(\lambda) = \frac{e^{-\lambda n} \lambda^{\sum_{i=1}^{n} x_i}}{x_1!\, x_2!\, x_3! \dots x_n!}$$

$$\ln(L) = -\lambda n + \left(\sum_{i=1}^{n} x_i\right) \ln \lambda - \sum_{i=1}^{n} \ln(x_i!)$$

- Differentiate wrt λ

$$\frac{d(\ln L)}{d\lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

$$\boxed{\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

MLE of λ is $\hat{\lambda} = \bar{x}$

Q2. The following data are the observed frequencies of occurrence of domestic accidents: we have n = 647 data as follows. Find MLE of λ

no. of accidents	frequency
0	447
1	132
2	42
3	21
4	3
5	2

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{0\times447 + 1\times132 + 2\times42 + 3\times21 + 4\times3 + 5\times2}{647}$$

$$\hat{\lambda} = \frac{301}{647} = 0.465$$

$$\hat{\lambda} = 0.465$$

PROBABILITY PLOTS

Steps

1. Sort the data
2. Assign ranks from $i=1$ to n (x_i for each data point)
3. Assign evenly-spaced values to the data (between 0 and 1)

for each x_i in the dataset, assign a value

$$r = \boxed{\frac{i-0.5}{n}} \leftarrow \text{Hazen method}$$

4. Find theoretical quantiles (Q_i) and plot (x_i, Q_i)

Method	Plotting Position Method
Blom	$\frac{i-0.375}{n+0.25}$
Benard	$\frac{i-0.3}{n+0.4}$
Hazen	$\frac{i-0.5}{n}$
Van der Waerden	$\frac{i}{n+1}$
Kaplan-Meier	$\frac{i}{n}$

Q3. Construct a normal probability plot for the following data. Do these data appear to come from an approximately normal distribution?
3.01, 3.35, 4.79, 5.96, 7.89

$i =$ 1 2 3 4 5

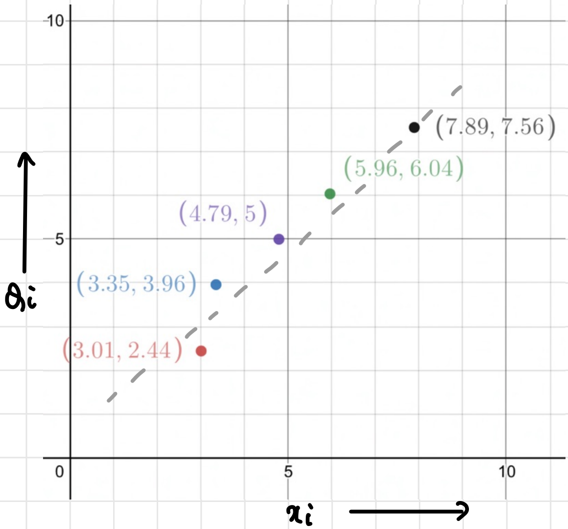
$n=5$ $\bar{x}=5$ $s=2\approx\sigma$

i	x_i	$(i-0.5)/5$	closest area	z-score	$Q_i = z*\sigma+\mu$
1	3.01	0.1	0.1003	-1.28	2.44
2	3.35	0.3	0.3015	-0.52	3.96
3	4.79	0.5	0.5000	0	5
4	5.96	0.7	0.6985	0.52	6.04
5	7.89	0.9	0.8997	1.28	7.56

(x_i, Q_i)

Q-Q Plot

(desmos)

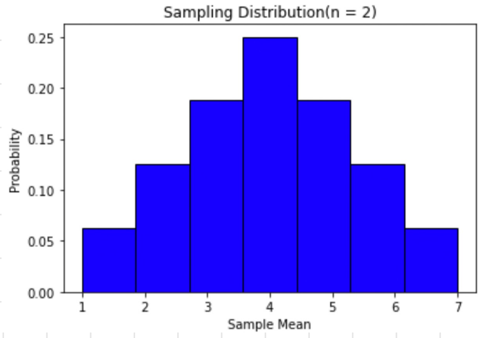


points are somewhat linear

- We can conclude that the data are approximately normally distributed
- Best to have at least 30 points for reliable conclusions

sampling DISTRIBUTION

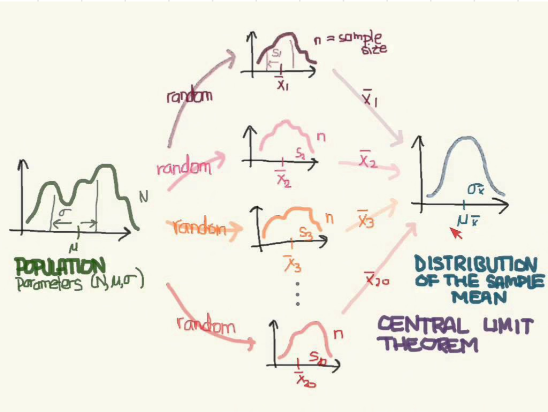
- Probability distribution of a statistic
- Eg: distribution of sample means for a sample size n where $n \leq$ size of population
- sample size = 2, population = $\{1, 3, 5, 7\}$ $\mu = 4$, $\sigma^2 = 5$
- mean of sample means $= \mu_{\bar{x}} = 4$



as $n \uparrow$, distribution approaches normal

Central Limit Theorem

- Distribution of sample means calculated from sampling a population will follow normal distribution as the size 'n' of the sample increases



https://medium.com/@garora039/what-exactly-is-central-limit-theorem-7c1531eb2987

- Let $x_1, x_2, \ldots x_n$ be a simple random sample from a population with mean μ and variance σ^2

- Let $\bar{x} = \dfrac{x_1 + \ldots + x_n}{n}$ be the sample mean

- Let $S_n = x_1 + \ldots + x_n$ be the sum of sample observations

- If n is sufficiently large,

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$S_n \sim N(n\mu, n\sigma^2)$$

Q4. A business client of FedEx wants to deliver urgently a large freight from Denver to Salt Lake City.
When asked about the weight of the cargo they could not supply the exact weight, however they have specified that there are total of 36 boxes.
You are working as a Business analyst for FedEx and you have been challenged to tell the executives quickly whether or not they can do certain delivery.

Since, we have worked with them for so many years and have seen so many freights from them we can confidently say that the type of cargo they follow is a distribution with a mean of μ= 32.66 kg standard deviation of σ = 1.36 kg. The plane you have can carry the max cargo weight up to 1193 kg. Based on this information what is the probability that all of the cargo can be safely loaded onto the planes and transported?

$\mu = 32.66 \quad \sigma = 1.36 \quad n = 36$

$S_n = 1193$

$S_n \sim N(36 \times 32.66,\ 36 \times 1.36^2)$

$S_n \sim N(\underset{\mu_{S_n}}{1175.76},\ \underset{\sigma^2_{S_n}}{66.5856})$

$P(S_n < \underset{x}{1193})$

$$z = \frac{1193 - 1175.76}{\sqrt{66.5856}} = 2.113$$

$P(z < 2.113) = 0.9826 = \boxed{98.26\%}$

Qs. Drums labeled 30 L are filled with a solution from a large vat. The amount of solution put into each drum is random with mean 30.01 L and standard deviation 0.1 L.

a) What is the probability that the total amount of solution contained in 50 drums is more than 1500 L?
b) If the total amount of solution in the vat is 2401 L, what is the probability that 80 drums can be filled without running out?
c) How much solution should the vat contain so that the probability is 0.9 that 80 drums can be filled without running out?

$\mu = 30.01 \quad \sigma = 0.1$

(a) $n = 50$, $P(S_n > 1500)$

$S_n \sim N(n\mu, n\sigma^2)$

$S_n \sim N(1500.5, 0.5)$

$$z_{1500} = \frac{1500 - 1500.5}{\sqrt{0.5}} = -0.707$$

$P(z > -0.707) = 1 - 0.2389 = 0.7611 = \boxed{76.11\%}$

(b) total of 80 < 2401, $n = 80$

$S_n \sim N(2400.8, 0.8)$

$$z_{2401} = \frac{2401 - 2400.8}{\sqrt{0.8}} = 0.2236 = 0.22$$

$P(z < 0.22) = 0.5871 = \boxed{58.71\%}$

(c) $S_n = ?$ area = 0.9 n = 80

for area = 0.9, $Z = 1.28$

for $Z \leq 1.28$

$\therefore S_n = Z \times \sigma + \mu = Z \times \sqrt{80 \times 0.01} + 80 \times 30.01$

$S_n = \boxed{2401.95 \text{ L}}$

continuity CORRECTION

- To approximate discrete distribution (binomial, Poisson) using continuous (normal) distribution
- Approximations for large no. of Binomial trials
- Correction: add or subtract 0.5 from discrete integer value for more accurate results

Probability	Discrete	Continuous
$P(X=n)$	$X=5$	$4.5 < x < 5.5$
$P(X>n)$	$X>5$	$x > 5.5$
$P(X \geq n)$	$X \geq 5$	$x > 4.5$
$P(X<n)$	$X<5$	$x < 4.5$
$P(X \leq n)$	$X \leq 5$	$x < 5.5$

for continuous, equality makes no difference

Normal Approximation to Binomial

- If $X \sim Bin(n,p)$
- $X = Y_1 + Y_2 + \dots + Y_n$ where $Y_1, \dots, Y_n$ are samples from a Bernoulli (p) population

$$\hat{p} = \frac{X}{n} = \frac{Y_1 + Y_2 + \dots + Y_n}{n}$$, which is also sample mean $\bar{Y}$

- Bernoulli(p) population has mean $\mu = p$ and variance $\sigma^2 = p(1-p)$
- Central Limit Theorem (n is large)

$$X \sim N(np, np(1-p))$$

$$\hat{p} \sim (p, \frac{p(1-p)}{n})$$

- Thumb rule: $np > 10$ and $n(1-p) > 10$
- If $X \sim Bin(n,p)$ and $np > 10$, $n(1-p) > 10$

$$\boxed{X \sim N(np, np(1-p))}$$

$$\boxed{\hat{p} \sim N(p, \frac{p(1-p)}{n})}$$

Q6. Imagine that a fair coin is tossed 100 times. Let X represent the number of heads. Then,

$$X \sim Bin(100, 0.5)$$

Imagine that we wish to compute the probability that X is between 45 and 55. This probability will differ depending on whether the endpoints, 45 and 55, are included or excluded.
Compute the following,

1. $P(45 \leq X \leq 55)$

2. $P(X \geq 60)$

$n = 100 \quad p = 0.5$

$np = 50 \quad n(1-p) = 50$

$$X \sim N(50, 25)$$

1. $P(44.5 < X < 55.5)$

using calculator, $P = 0.7287 = \boxed{72.87\%}$

2. $P(X > 59.5)$

using calculator, $P = 0.0287 = \boxed{2.87\%}$

Normal Approximation to Poisson

- If $X \sim \text{Poisson}(\lambda)$ with large n and $\lambda > 10$

$$\mu_X = \lambda \quad \sigma_X^2 = \lambda$$

$$X \sim N(\lambda, \lambda)$$

Q7. The number of hits on a website follows a Poisson distribution, with a mean of 27 hits per hour. Find the probability that there will be 90 or more hits in three hours.

$$\lambda = 27 > 10 \checkmark$$

$$X \sim N(27 \times 3, 27 \times 3)$$
$$X \sim N(81, 81)$$

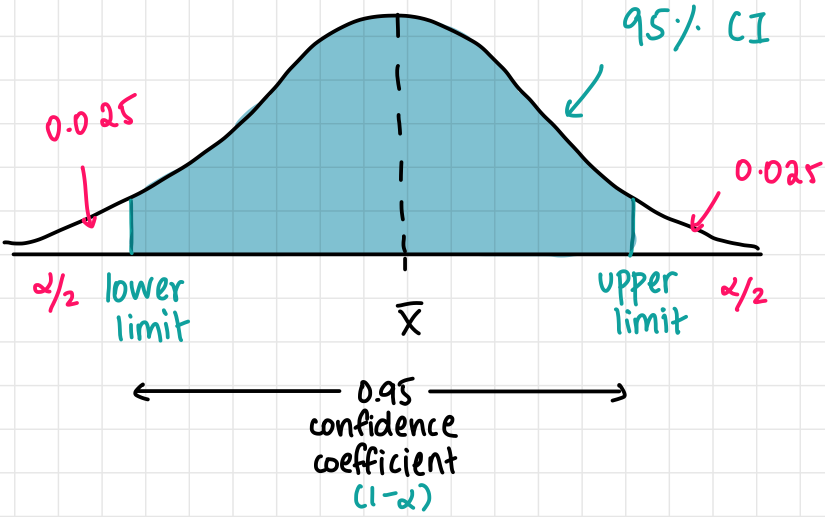
from calculator

without correction

$$P(X > 90) = 0.1587 = \boxed{15.87\%}$$

CONFIDENCE INTERVALS

- Range of plausible values for a parameter, based on sample data
- Differs from sample to sample
- Expressed as percentage



$$\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$\frac{\sigma}{\sqrt{n}}$ is std dev of sampling dist of sample mean $\bar{x}$

- 95% confidence that population parameter in that range

Q8. Find the value of $z_{a/2}$ to use to construct a confidence interval with level:
a) 95% b) 98% c) 99% d) 80%

(a) $\alpha/2 = 2.5\% = 0.025 \Rightarrow z = \pm 1.96$

(b) $\alpha/2 = 1\% = 0.01 \Rightarrow z = \pm 2.33$

(c) $\alpha/2 = 0.5\% = 0.005 \Rightarrow z = \pm 2.57$

(d) $\alpha/2 = 10\% = 0.1 \Rightarrow z = \pm 1.28$

Q9. Find the levels of confidence intervals that have the following values of $z_{a/2}$:
a) $z_{a/2} = 2.17$ b) $z_{a/2} = 3.28$

(a) $z_{\alpha/2} = 2.17$ $\quad$ $z_{\alpha/2} = -2.17$

area = 0.9850 $\quad$ area = 0.0150

$\therefore$ CI = 0.97 = 97%

(b) $z_{\alpha/2} = 3.28$ $\quad$ $z_{\alpha/2} = -3.28$

area = 0.9995 $\quad$ area = 0.0005

$\therefore$ CI = 0.999 = 99.9%

for CI of p

$\tilde{n} = n+4$

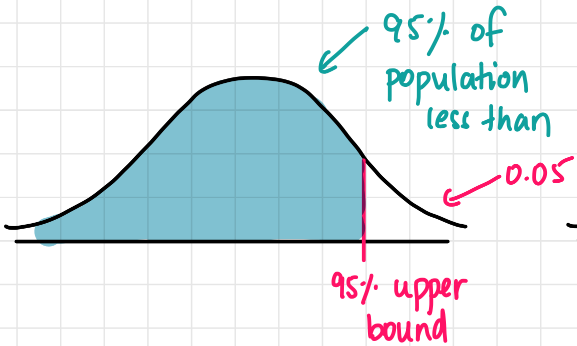
$\tilde{p} = \dfrac{X+2}{\tilde{n}}$

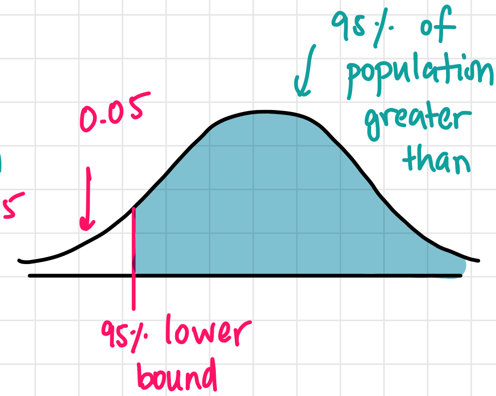
$\tilde{p} \pm z_{\alpha/2}\sqrt{\dfrac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$

Q10. A 90% confidence interval for the mean diameter (in cm) of steel rods manufactured on a certain extrusion machine is computed to be (14.73, 14.91). True or false: The probability that the mean diameter of rods manufactured by this process is between 14.73 and 14.91 is 90%.

False. 90% confidence that population mean in that range

ONE-SIDED CONFIDENCE BOUNDS





Q11. In a sample of 80 ten-penny nails, the average weight was 1.56g and the standard deviation was 0.1g.
a) Find a 90% upper confidence bound for the mean weight.
b) Find a 80% lower confidence bound for the mean weight.
c) Someone says that the mean weight is less than 1.585g. With what level of confidence can this statement be made?

(a) $z = 1.28 \Rightarrow x = \frac{1.28 \times 0.1}{\sqrt{80}} + 1.56 = 1.5743$

(b) area $= 0.2 \Rightarrow z = -0.84 \Rightarrow x = \frac{-0.84 \times 0.1}{\sqrt{80}} + 1.56 = 1.5506$

(c) upper $= 1.585 \Rightarrow z = \frac{1.585 - 1.56}{0.1/\sqrt{80}} = 2.236 \Rightarrow c = 98.75\%$

Q12. Of a random sample of n = 150 college students, 104 of the students said that they had played on a soccer team during their K-12 years. Estimate the proportion of college students who played soccer in their youth with a 90% confidence interval.

$\hat{p} = \frac{104}{150} = 0.693$ $n = 150$ $1-\hat{p} = 0.307$

90% CI $\Rightarrow \pm 1.65$

$\hat{p} \pm 1.65\sqrt{\frac{p(1-p)}{n}}$

$0.693 \pm 1.65\sqrt{\frac{0.693 \times 0.307}{150}}$

0.693 ± 0.0621

$\boxed{0.6309 \text{ to } 0.755}$

CONFIDENCE INTERVAL FOR DIFFERENCE BETWEEN MEANS

$$\boxed{(\bar{x}-\bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

Student's t- Distribution

- Samples of a full population
- Larger sample size $\rightarrow$ normal distribution
- Theoretical probability distribution — symmetrical, bell-shaped, similar to standard normal curve
- Degrees of freedom - another parameter

$$df = \text{sample size} - 1$$

- As df increases, approaches standard normal distribution (after df = 30, almost identical)
- t-score calculated like z-score
- The quantity $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ has a t-distribution with n-1 df

Q13. A random sample of size 10 is drawn from a normal distribution.
a) Find P(t >1.833)
b) Find P(t > 1.5)

$df = 9$

(a) $P(t > 1.833) = 0.05$

(b) $P(t > 1.5)$ = b/w 0.05 and 0.10

Q14. Find the value of $t_{n-1, a/2}$ needed to construct a two-sided confidence interval of the given level with the given sample size:

a) 90% with sample size 12

b) 95% with sample size 7

(a) df = 11 , area = 0.05

$t = \pm 1.796$

(b) df = 6 , area = 0.025

$t = \pm 2.447$

Note: if population standard deviation is known, and it is known to come from normal distribution, use z and not t