Statistics and Estimation for Computer Science

İstanbul Teknik Üniversitesi

Mustafa Kamasak, PhD

Version: 2022.3.15

Interval Estimation

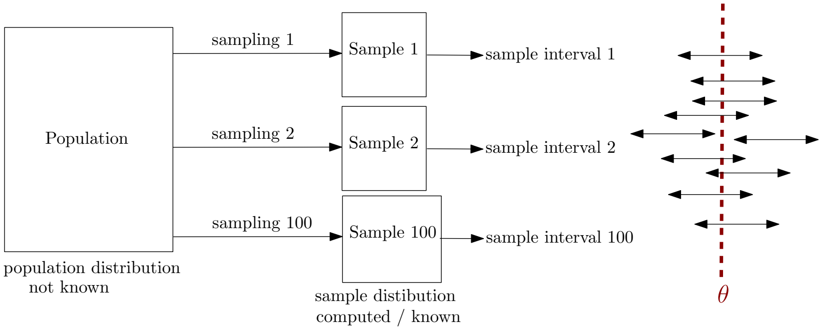
Interval Estimation

- Point estimates are not always useful
- Instead of a single estimate, an interval can be given
- For example, sample mean interval: $[L,\ U]$, where L is the lower limit and U is the upper limit
  - Short interval: High chance to miss population parameter
  - Long interval: Not very useful
- Define a confidence level to determine the lenth of the interval
- Commonly used confidence levels are
  - 90 % confidence level
  - 95 % confidence level
  - 99 % confidence level
- Given a confidence level, how to find the corresponding interval?

Confidence Interval (CI)

- Sample parameter estimation has certain sampling distr.
- For example, sample mean $\overline{x} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$
- Using sampling distr
  - High variance $\implies$ larger confidence interval
  - Low variance $\implies$ shorter confidence interval
- Determine confidence interval using area under sampling distribution
- For example, if 95 % CI for sample mean is [10, 20], then $P(10 \leq \overline{x} \leq 20) = 0.95$
- How do we interpret this interval in terms of population parameter? (note: population parameter is **not** rv)

Confidence Interval (CI)



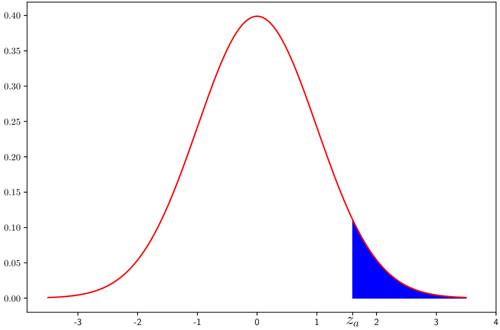
- If you take 100 different samples $\implies$ compute 100 different 95% confidence intervals,
- True population parameter θ will be included in 95 of these intervals.

Notation

- Define z_α as the value such that

$$\alpha = 1 - \Phi(z_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{z_\alpha}^{\infty} \exp\{-\frac{x^2}{2}\} dx$$

- Area of the blue shaded area is α



Confidence Interval for Sample Mean

- Sampling distribution of $\overline{x}$

$$\overline{x} \sim \mathcal{N}(\mu, \frac{\sigma^2}{\sqrt{N}})$$

$$Z = \frac{\overline{x} - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0, 1)$$

- We need z_α such that $P(|Z| < z_\alpha) = 0.95$
- $\alpha = 0.025$
- From z look-up table $z_\alpha = 1.96$
- Hence

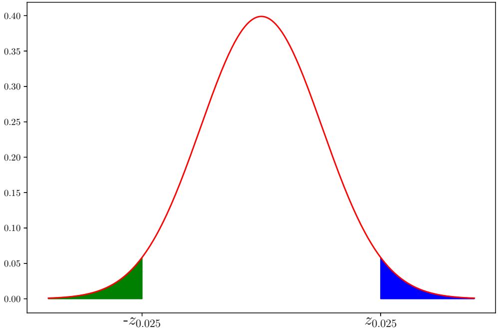
$$P\left(\frac{\sqrt{N}}{\sigma}|\overline{x} - \mu| \leq 1.96\right) = 0.95$$

- $\overline{x} \pm 1.96\sigma/\sqrt{N}$ will contain μ with 95% probability.

$$P(\mu \in \overline{x} \pm 1.96\sigma/\sqrt{N}) = 0.95$$

Confidence Interval for Sample Mean

- Why $\alpha = 0.025$



- Total area of red and green shades should be 0.05
- From symmetry of distr. green area = blue area = 0.025

Confidence Interval and Sample Size

- Confidence interval is

$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}}\ ,\ \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}\right]$$

- With a large sample size (N) $\implies$ shorter interval
- With a small sample size (N) $\implies$ larger interval

Confidence Interval (CI)

- Commonly used confidence levels and corresponding z_α values are

Confidence level	α	$z_{\alpha/2}$
90%	0.10	1.645
95%	0.05	1.96
99%	0.01	2.576

Confidence Interval (CI)

- sample size N given $\implies$ confidence interval is computed
- confidence interval is given $\implies$ sample size is computed

- For example, we want confidence interval $1-\alpha$ to be B, then

$$
\begin{aligned}
z_{\alpha/2}\frac{\sigma}{\sqrt{N}} &\leq \frac{B}{2} \\
N &\geq \left(\frac{2z_{\alpha/2}\sigma}{B}\right)^2 \\
N_{min} &= \left\lceil \left(\frac{2z_{\alpha/2}\sigma}{B}\right)^2 \right\rceil
\end{aligned}
$$

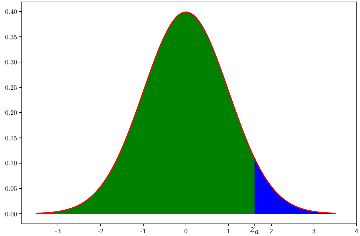
Confidence Level for Lower Limit

- Sometimes an upper or a lower limit is required with some confidence level $1 - \alpha$
- For example find the lower limit for population mean with confidence level $1 - \alpha$

$$P(\frac{\sqrt{N}}{\sigma}(\overline{x} - \mu) \leq z_\alpha) = 1 - \alpha$$

$$P(\overline{x} - z_\alpha \frac{\sigma}{\sqrt{N}} \leq \mu) = 1 - \alpha$$

- We need green area to be $1 - \alpha$, and blue are to be α



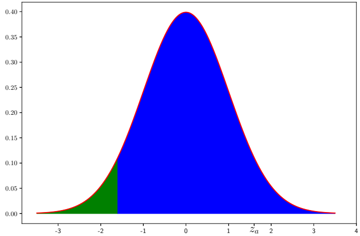
Confidence Level for Upper Limit

- Similarly, find the upper limit for population mean with confidence level α

$$P(\frac{\sqrt{N}}{\sigma}(\overline{x} - \mu) \geq z_\alpha) = 1 - \alpha$$

$$P(\mu \leq \overline{x} + z_\alpha \frac{\sigma}{\sqrt{N}}) = 1 - \alpha$$

- We need green area to be α, and blue are to be $1 - \alpha$



Confidence Level for Lower/Upper Limits

- Upper limit with confidence level of $1 - \alpha$ is

$$\overline{x} + z_{\alpha} \frac{\sigma}{\sqrt{N}}$$

- Lower limit with confidence level of $1 - \alpha$ is

$$\overline{x} - z_{\alpha} \frac{\sigma}{\sqrt{N}}$$

Confidence Interval

- Population std dev σ is used to compute confidence interval $\implies$ σ **not** known
- Use sample std dev s instead of population std dev σ
- Both $\overline{x}$ and s are rv $\implies$ their ratio is also rv
- Ratio of 2 rv does **not** have normal distr anymore

$$T \triangleq \frac{\sqrt{N}}{s}(\overline{x} - \mu)$$

- T has t distribution with $N-1$ dof as one dof is lost due to specification of sample std dev s

$$T \sim t_{N-1}$$

T-distribution History

- Discovered by William Gosset, a chemist working for the Guinness brewery
- Statistical methods were used to test quality of beer production
- Company policy at Guinness forbade its chemists from publishing their findings
- Gosset published his statistical work in the journal Biometrika 1908 under the pseudonym "Student"
- Known also as Student's t distribution

- For more info
  `https://en.wikipedia.org/wiki/Student's_t-test`

T Distribution

- Continuous distribution
- Distribution is symmetric around $x = 0$
- Single parameter degrees of freedom (υ)
- pdf

$$f(x) = a(\upsilon)\left(1 + \frac{1}{\upsilon}x^2\right)^{-\frac{1}{2}(\upsilon+1)}$$

where

$$a(\upsilon) = \frac{1}{\sqrt{\upsilon\pi}\frac{\Gamma(0.5(\upsilon+1)}{\Gamma(0.5\upsilon)}}$$

for integer values of υ and gamma function is

$$\Gamma(m) = (m-1)!$$

for positive integers.

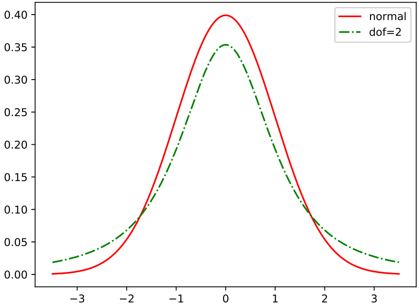
- No closed form expression for cdf of t-distr $\implies$ Need look-up table

T Distribution vs Normal Distribution

- t-distribution has fatter tails compared to normal distr
- This is due to the added uncertainty for using sample std dev instead of population std dev
- For a given α

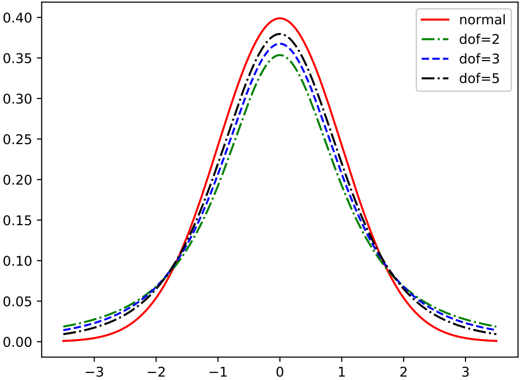
$$t_{N-1,\alpha} > z_\alpha$$

due to fat tails



T Distribution vs Normal Distribution

- As dof increases t distr converges to normal distr
- Difference between $t_{N-1,\alpha}$ z_α decreases with increasing dof (N)
- For dof$\geq$ 30, t-distr is just like normal distr



Confidence Interval of Sample Mean with Sample Std Dev

- $1-\alpha$ confidence interval is:

$$\left[\overline{x} - t_{N-1,\alpha/2}\frac{s}{\sqrt{N}}\ ,\ \overline{x} + t_{N-1,\alpha/2}\frac{s}{\sqrt{N}}\right]$$

- This interval is larger compared to case where population std dev is used
- This is due to added uncertainty
- How to find $t_{N-1,\alpha/2} \implies$ Use t-table

T-table

- Cdf of t-distr. is precomputed numerically
- Values of cdf at all t values are not given
- Only critical t values for some α are given
- T-lookup table

TABLE A.2

t Distribution: Critical Values of t

Degrees of freedom		*Significance level*					
	Two-tailed test:	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437

Confidence Interval for t Distr

- Commonly used confidence levels and corresponding $t_{10,\alpha/2}$ values are

Confidence level	α	$z_{\alpha/2}$	$t_{10,\alpha/2}$
90%	0.10	1.645	1.812
95%	0.05	1.960	2.228
99%	0.01	2.576	3.169

- Commonly used confidence levels and corresponding $t_{30,\alpha/2}$ values are

Confidence level	α	$z_\alpha/2$	$t_{30,\alpha/2}$
90%	0.10	1.645	1.697
95%	0.05	1.960	2.042
99%	0.01	2.576	2.750

- As sample size gets close to 30, there is small difference between critical values of normal distr and t distr

Confidence Level Lower/Upper Limits

- Upper limit with confidence level of $1 - \alpha$ is

$$\overline{x} + t_{N-1,\alpha} \frac{s}{\sqrt{N}}$$

- Lower limit with confidence level of $1 - \alpha$ is

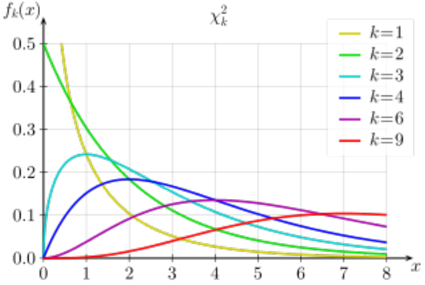
$$\overline{x} - t_{N-1,\alpha} \frac{s}{\sqrt{N}}$$

Confidence Interval for Population Variance

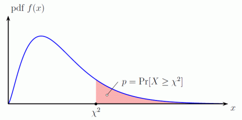
- Sampling distribution of the sample variance s^2 is

$$X^2 = \frac{(N-1)s^2}{\sigma^2} \sim \chi^2_{N-1}$$

chi-square distribution with $N-1$ dof



- Need χ^2 look-up table for intervals
- Critical values for χ^2 distribution:



Chi-square Distribution Table

d.f.	.995	.99	.975	.95	.9	.1	.05	.025	.01
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09

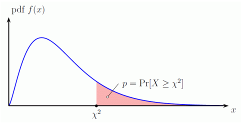
- For $1 - \alpha$ CI with dof=N-1
  - Lower limit for X^2 should be $\chi^2_{N-1,1-\alpha/2}$
  - Upper limit for X^2 should be $\chi^2_{N-1,\alpha/2}$

$$\chi^2_{N-1,1-\alpha/2} \leq X^2 \leq \chi^2_{N-1,\alpha/2}$$

$$\chi^2_{N-1,1-\alpha/2} \leq \frac{(N-1)s^2}{\sigma^2} \leq \chi^2_{N-1,\alpha/2}$$

$$\frac{(N-1)s^2}{\chi^2_{N-1,\alpha/2}} \leq \sigma^2 \leq \frac{(N-1)s^2}{\chi^2_{N-1,1-\alpha/2}}$$

Confidence Interval for Population Variance



Chi-square Distribution Table

d.f.	.995	.99	.975	.95	.9	.1	.05	.025	.01
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09

- 90% CI for population variance σ^2 with dof$=$5 is

$$\left[\frac{5s^2}{11.07}, \frac{5s^2}{1.15}\right]$$

Lower and Upper Limit for Population Variance

- Lower limit with $1 - \alpha$ confidence is

$$\frac{(N-1)s^2}{\chi^2_{N-1,\alpha}}$$

- Upper limit with $1 - \alpha$ confidence is

$$\frac{(N-1)s^2}{\chi^2_{N-1,1-\alpha}}$$

Confidence Interval for Population Proportion

- Sometimes a confidence interval is required for a proportion of some condition in a population
- For example, a confidence interval can be required for the accuracy of a binary classifier

- Let $\hat{p}$ be a zero bias point estimate of the proportion
- If p is not close to 0 or 1, and when the sample size is large ($Np > 5$ and $N(1-p) > 5$)

$$Z \triangleq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{N}}} \sim \mathcal{N}(0, 1)$$

- Why?

Confidence Interval for Population Proportion

- Let m be total number of cases in the sample that has this certain condition
- In binary classification example, let m be the number of correctly classified cases
- m has binomial distr

$$m \sim \text{Binomial}(p)$$

  with $E(m) = Np$ and $\sigma_m^2 = Np(1-p)$
- If p is not close to 0 or 1, and when the sample size is large ($Np > 5$ and $N(1-p) > 5$) binomial distr can be approximated with normal distr.
- Continuity correction - described in the next slide!!!
- Then, $\frac{m-Np}{\sqrt{Np(1-p)}}$ has standard normal distr.

$$Z \triangleq \frac{m - Np}{\sqrt{Np(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{N}}} \sim \mathcal{N}(0,1)$$

Confidence Interval for Population Proportion

$$Z \triangleq \frac{m - Np}{\sqrt{Np(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{N}}} \sim \mathcal{N}(0, 1)$$

- $1 - \alpha$ confidence interval is

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{N}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{N}}\right]$$

- This interval is useless as it uses population parameter p to get an interval about p!

Confidence Interval for Population Proportion

- Let $\hat{\Theta}$ be an estimator for population parameter θ
  - If $\hat{\Theta}$ has normal distr.
  - If $\hat{\Theta}$ is approximately unbiased
  - If estimator variance $\sigma^2_{\hat{\Theta}}$ can be estimated from sample

  then confidence interval for θ can be computed using $\sigma^2_{\hat{\Theta}}$ instead of population variance σ^2

$$[\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\Theta}},\ \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\Theta}}]$$

- For population parameter listed conditions hold
- $1 - \alpha$ confidence interval is

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{N}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}\right]$$

Lower and Upper Limit for Population Proportion

- Lower limit with $1 - \alpha$ confidence is

$$\hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

- Upper limit with $1 - \alpha$ confidence is

$$\hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

Continuity Correction

- Binomial is a discrete distr
- Normal distr. is continuous distr
- Binomial $\rightarrow$ Normal approximation will not work:
  - if p is close to 0 or 1,
  - if sample size is small

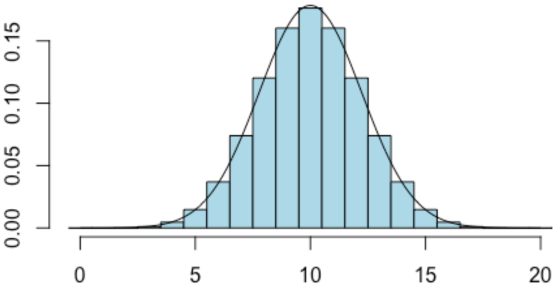


Image taken from http://mathcenter.oxford.emory.edu/site/math117/normalApproxToBinomial/

Continuity Correction

Binomial	Normal
$P(X = k)$	$P(k–0.5 < X < k + 0.5)$
$P(X \leq k)$	$P(X < k + 0.5)$
$P(X < k)$	$P(X < k–0.5)$
$P(X > k)$	$P(X > k + 0.5)$
$P(X \geq k)$	$P(X > k–0.5)$

One Sample Hypothesis Testing

Hypothesis Testing

- From estimation of a population parameter to test a hypothesis about the population parameter
- Concepts
  - H_0 : null hypothesis
  - H_1 : alternative hypothesis
- Typically, widely accepted assumption is used as null hypothesis
- Without a strong evidence (from sample) against null hypothesis, it will not be rejected

Hypothesis Types

- There are two types of hypothesis
- Simple hypothesis (two-sided)

$$\begin{aligned} \mathrm{H}_0 &: \theta = \theta_0 \\ \mathrm{H}_1 &: \theta \neq \theta_0 \end{aligned}$$

- Composite hypothesis (one-sided)

$$\begin{aligned} \mathrm{H}_0 &: \theta \leq \theta_0 \\ \mathrm{H}_1 &: \theta > \theta_0 \end{aligned}$$

or

$$\begin{aligned} \mathrm{H}_0 &: \theta \geq \theta_0 \\ \mathrm{H}_1 &: \theta < \theta_0 \end{aligned}$$

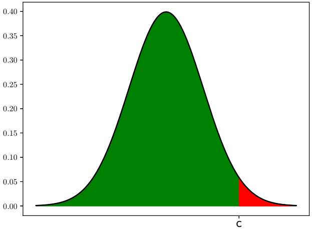
Hypothesis Testing

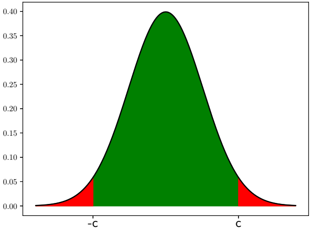
- A test statistics (T) obtained from the sample is used
- Depending on the value of this test statistics, we can reject or retain H_0
- Define a critical region (rejection region) using a critical value c

$$\mathcal{R}_{\text{one-sided}} = \{T(x) > c\}$$

or

$$\mathcal{R}_{\text{two-sided}} = \{|T(x)| > c\}$$





Hypothesis Testing

- If the value of test statistics is within critical region $\Longrightarrow$ reject $\mathrm{H_0}$
- If the value of test statistics is outside critical region (acceptance region) $\Longrightarrow$ do not reject $\mathrm{H_0}$ / retain $\mathrm{H_0}$

$$\begin{cases} T(x) \in \mathcal{R} & \text{reject } H_0 \\ T(x) \notin \mathcal{R} & \text{retain } H_0 \end{cases}$$

- Not rejecting $\mathrm{H_0}$ **does not** mean accepting $\mathrm{H_0}$
- Failing to reject (or retaining) $\mathrm{H_0}$ just means sample statistics does not supply a strong evidence against $\mathrm{H_0}$

Hypothesis Testing – Example

- Let H_0 to be average cholesterol level in the healthy population is 170
- Take a sample of 100 healthy individual, measure their cholesterol
- Let the test statistic to be $T = \overline{x} - 170$
- Let critical value to be 20
- Then critical region will be

$$\mathcal{R} = \{|\overline{x} - 170| \geq 20\}$$

- If mean cholesterol of the sample is
  - larger then or equal to 190
  - less than or equal to 150

  H_0 is rejected

Hypothesis Testing

- How to determine the test statistics T
- Use a test statistics where the sampling distribution is known

- How to determine the critical value or critical region?
- Determine a significance level α
- Using sampling distr of test statistics, find a critical value c such that

$$P(T(x) \in \mathrm{R} \mid \mathrm{H}_0\ ) = \alpha$$
$$P(\text{reject } \mathrm{H}_0 \mid \mathrm{H}_0\ ) = \alpha$$

Relation of Hypothesis Testing and Confidence Interval

- Consider the simple hypothesis:
  $\mathrm{H}_0 : \theta = \theta_0$
  $\mathrm{H}_1 : \theta \neq \theta_0$
- Let significance level to be α
- Let $[L, \ U]$ be the $1 - \alpha$ CI

$$\begin{cases} \theta_0 \in [L, \ U] & \text{retain } H_0 \\ \theta_0 \notin [L, \ U] & \text{reject } H_0 \end{cases}$$

Types of Errors in Hypothesis Testing

	H_0 retained	H_0 rejected
H_0 correct	$1-\alpha$	type-I error α
H_0 incorrect	type-II error β	$1-\beta$

$$P(\text{retain } H_0 \mid H_0\ ) = 1-\alpha$$

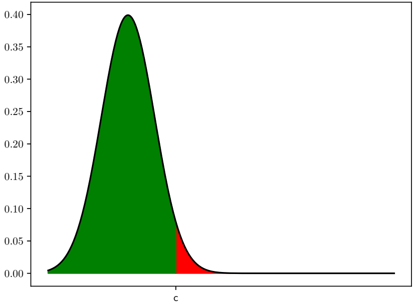
$$P(\text{reject } H_0 \mid H_0\ ) = P(\text{type-I error}) = \alpha$$

$$P(\text{reject } H_0 \mid H_1\ ) = 1-\beta$$

$$P(\text{retain } H_0 \mid H_1\ ) = P(\text{type-II error}) = \beta$$

Type-I Error

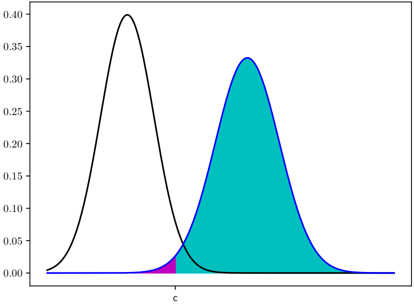
- Assume H_0 is correct for a composite hypothesis (one-sided)
- Consider the following sampling distribution (for a sample statistics) when H_0 is correct
- If a sample statistics fall in the critical (red) region, H_0 is rejected $\implies$ type-I error
- Red area is the probability of type-I error $= \alpha$
- Green are is the probability of correctly retaining $\mathrm{H}_0 = 1 - \alpha$



Factors of Type-I Error

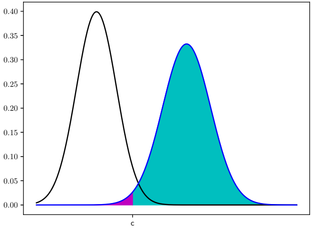
- Type-I error probability is the significance level α
- Depending on the cost of type-I error, its probability is selected during study design before data collection
- Typically used values for power of a test is 0.8, 0.85, or 0.9
- Hence there is no factor that affects type-I error

Type-II Error



- Type-II error: Retaining H_0 when H_1 is correct
- If the sample parameter falls within the acceptance region (outside critical region)
- In the figure, if sample parameter is less than c, H_0 is retained
- Probability of type-II error is the area shown in magenta $\implies \beta$

Type-II Error and Power



- Probability of type-II error is the area shown in magenta $\implies \beta$
- Probability of rejecting H_0 when H_1 is correct is called "power of the test" $= 1 - \beta$
- Shown in cyan area
- How to compute type-II error probability and power of the test
- Blue sampling distribution – $P(\text{sample parameter} \mid \mathrm{H}_1)$ – is not known

Type-II error and Power

- H_1 does not result in a sampling distribution
- For example:
  $\mathrm{H}_0 : \mu = 0$
  $\mathrm{H}_1 : \mu \neq 0$
  Alternative hypothesis do not result in a distribution for sample mean

- One possibility is to compute type-II error and power as a function of distance from H_0
- What about std. dev.? $\implies$ requires pilot study (small sample size) or technical prior information

- Another possibility is to compute type-II error and power of test at a meaningful (in terms of application) distance

Type-II Error and Power

- Power of a test is typically pre-selected at a given distance to find the sample size
- Typically used values for power of a test is 0.8, 0.85, or 0.9
- Distance is a meaningful difference
- Sample variance is computed from a pilot study (small sample size) or used from literature

Factors of Type-II Error and Power

- Sample size N and significance level α affects type-II error probability β and test power $1 - \beta$

- Sample size $\uparrow \implies \beta \downarrow$ and power $\uparrow$
- Significance level $\alpha \uparrow \implies \beta \downarrow$ and power $\uparrow$

Type I-II Errors in Binary Classification

- In binary classification, terminology is a bit different
  - Confusion matrix
  - True/False positive
  - True/False negative
- Think H_0 as negative and H_1 as positive
- Type-I error is like FP
- Type-II error is like FN

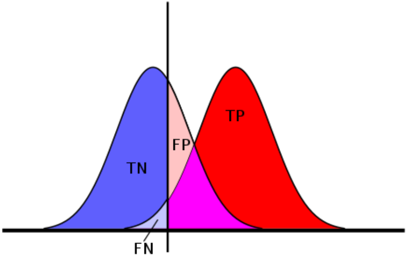
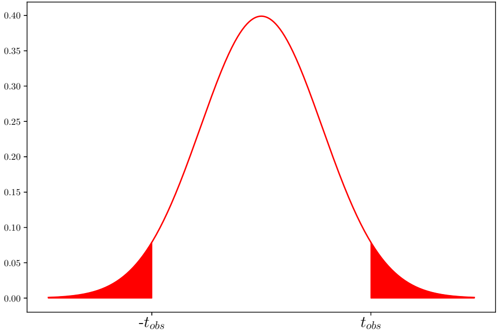


Figure taken from `https://en.wikipedia.org/wiki/Type_I_and_type_II_errors`

p-value

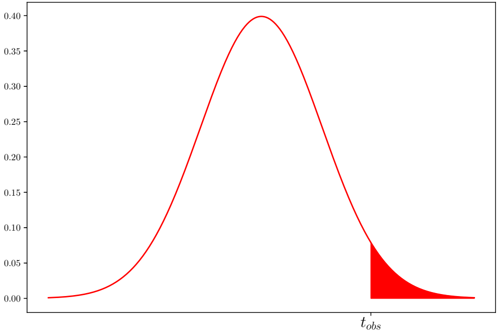
- From a sample, let the observed value of t statistic be t_{obs}
- Then p-value is the probability of observed sample statistics or more extreme observations when H_0 is correct

- For a two-tailed test (simple hypothesis), p-value is the red area shown below



p-value

- For a one-tailed test (composite hypothesis), p-value is the red area shown below



Interpretation of p-value

- p-value is a measure of evidence against H_0

p-value	Interpretation
<0.01	very strong evidence against H_0
0.01-0.05	strong evidence against H_0
0.05-0.10	weak evidence against H_0
>0.10	little or no evidence against H_0

- A large p-value is **NOT** a strong evidence in favor of H_0
- p-value is **NOT** the probability of H_0
- Why?

Interpretation of p-value

- t_{obs} is an rv $\implies$ p is also rv

$$p \triangleq P(T \geq t_{obs} | \mathrm{H}_0 \ ) = 1 - F_t(t_{obs})$$

  F_t is the cdf of sample statistics t when H_0 is correct
- Consider a new rv K whose pdf is $F_t(t)$
- Then $p = 1 - K$
- From probability course
  - $K \sim \mathrm{Uni}(0, 1)$
  - $p \sim \mathrm{Uni}(0, 1)$

- When H_0 is correct, p-value has uniform distribution in [0,1]
- Hence a large p-value cannot be interpreted as an evidence in favor of H_0

Significance Level vs p-value

- Significance level α is selected during study design (before data collection)
- Typical values for significance level is 0.1, 0.05 or 0.01
- If type-I error is very costly choose smaller significance level

- p-value is computed from sample after data collection

- If $p < \alpha$ $(t_{obs} > t_\alpha) \implies$ reject H_0
- t_{obs} is within critical region reject H_0

- If $p > \alpha$ $(t_{obs} < t_\alpha) \implies$ retain H_0
- t_{obs} is outside critical region retain H_0

Hypothesis Testing for Population Parameters

- Population parameters
  - Mean
  - Variance
  - Proportion

Hypothesis Testing for Population Mean

- Simple hypothesis: (two-sided test)
  $H_0 : \mu = \mu_0$
  $H_1 : \mu \neq \mu_0$

- Composite hypothesis: (one-sided test)
  $H_0 : \mu \leq \mu_0$
  $H_1 : \mu > \mu_0$
  or
  $H_0 : \mu \geq \mu_0$
  $H_1 : \mu < \mu_0$

Hypothesis Testing for Population Mean

- For population mean, normalized sample mean is used as test statistic
- If population variance σ^2 is known

$$T = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{N}}$$

- Sampling distribution for sample mean: $T \sim \mathcal{N}(0, 1)$
- If population variance σ^2 is unknown, sample variance s^2 is used

$$T = \frac{\overline{x} - \mu_0}{s/\sqrt{N}}$$

- Sampling distribution for sample mean: $T \sim t_{N-1}$

Hypothesis Testing for Population Mean

- Set significance level α
- Depending on the sampling distribution
  - Normal distr if population variance is known
  - T distr if population variance is unknown
- Critical values are found from tables
- For two-sided test (simple hypothesis), $t_{critical}$ is
  - $z_{\alpha/2}$ if population variance is known
  - $t_{N-1,\alpha/2}$ if population variance is unknown
- For one-sided test (composite hypothesis), $t_{critical}$ is
  - z_{α} if population variance is known
  - $t_{N-1,\alpha}$ if population variance is unknown
- Compute sample mean from sample $\overline{x}$ and observed test statistic t_{obs}
- If t_{obs} is within critical region $\implies$ reject H_0
- Otherwise, retain H_0

Hypothesis Testing for Population Mean

- p-value of the test is

- For known population variance
  - two-sided test

$$\text{p-value} = 2\Big(1 - F_z\big(\frac{\overline{x} - \mu_0}{\sigma/\sqrt{N}}\big)\Big) = 2(1 - F_z(t_{obs}))$$

  - one-sided test

$$\text{p-value} = 1 - F_z\big(\frac{\overline{x} - \mu_0}{\sigma/\sqrt{N}}\big) = 1 - F_z(t_{obs})$$

- For unknown population variance
  - two-sided test

$$\text{p-value} = 2\Big(1 - F_t\big(\frac{\overline{x} - \mu_0}{s/\sqrt{N}}\big)\Big) = 2(1 - F_t(t_{obs}))$$

  - one-sided test

$$\text{p-value} = 1 - F_t\big(\frac{\overline{x} - \mu_0}{s/\sqrt{N}}\big) = 1 - F_t(t_{obs})$$

where F_t is the cdf of t distr with $N-1$ dof.

Hypothesis Testing for Population Variance

- Simple hypothesis: (two-sided test)
  $\mathrm{H}_0 : \sigma^2 = \sigma_0^2$
  $\mathrm{H}_1 : \sigma^2 \neq \sigma_0^2$

- Composite hypothesis: (one-sided test)
  $\mathrm{H}_0 : \sigma^2 \leq \sigma_0^2$
  $\mathrm{H}_1 : \sigma^2 > \sigma_0^2$
  or
  $\mathrm{H}_0 : \sigma^2 \geq \sigma_0^2$
  $\mathrm{H}_1 : \sigma^2 < \sigma_0^2$

Hypothesis Testing for Population Variance

- Test statistics:

$$T = \frac{(N-1)s^2}{\sigma_0^2} \sim \chi^2_{N-1}$$

- Critical value for significance level of α
  - Two-sided test:
    - Lower critical value: $\chi^2_{N-1,1-\alpha/2}$
    - Upper critical value: $\chi^2_{N-1,\alpha/2}$
  - One-sided test:
    - $\chi^2_{N-1,1-\alpha}$ if $\mathrm{H}_0 : \sigma^2 \geq \sigma_0^2$
    - $\chi^2_{N-1,\alpha}$ if $\mathrm{H}_0 : \sigma^2 \leq \sigma_0^2$

Hypothesis Testing for Population Proportion

- Simple hypothesis: (two-sided test)
  $H_0 : p = p_0$
  $H_1 : p \neq p_0$

- Composite hypothesis: (one-sided test)
  $H_0 : p \leq p_0$
  $H_1 : p > p_0$
  or
  $H_0 : p \geq p_0$
  $H_1 : p < p_0$

Hypothesis Testing for Population Proportion

- Test statistic:

$$T = \frac{M/N - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}}$$

- $T \sim \mathcal{N}(0,1)$ if $Np_0 > 5$ and $N(1-p_0) > 5$
- Critical value
  - Two-sided test: $z_{\alpha/2}$
  - One-sided test: z_α
- p-value
  - Two-sided test:

$$\text{p-value} = 2(1 - F_z(t_{obs}))$$

  - One-sided test:

$$\text{p-value} = 1 - F_z(t_{obs})$$