stats formula sheet Cheat Sheet

by cicihex via cheatography.com/197422/cs/41607/

Basic Terms

parameter	fixed value describing population; usually unknown
statistic	value calculated from sample; used to estimate parameter
descriptive stats	- collecting, summarizing, describing data - graphical/numerical
inferential stats	- drawing conclusions/making predictions about pop based on sample
**	

data types

name	type	data
discrete	num	whole number
continuous	num	decimals
nominal	cat	no order
ordinal	cat	has order

sampling

graphical summary

numerical summary

percentile

quartile

standard deviation

IQR

outliers

symmetric	**skewed**

numerical summary (cont)

measure of center	mean	median
measure of spread	SD	IQR

histograms

association

probability

interpretation

properties

conditional probability

discrete RV

binomial RV

cont. RV

cont prob distribution properties

empirical rule

z stuff

normal distribution

sampling distribution - sample mean

CLT

standard error and bias of $\bar{X}$

estimation of μ

margin of error

confidence level & z-score

Confidence Interval - 3 cases

1. pop not normal; σ KNOWN ⇒ central limit theorem

the approx confidence interval for pop mean μ is

$\bar{x} \pm z^*(\sigma/\sqrt{n})$

$z^*=z\alpha/2$ is upper critical value

2. pop normal; σ UNknown ⇒ t-distribution

$T \equiv (\bar{X}-\mu)/(S/\sqrt{n})$

$S^2=1/n-1\sum(Xi-\bar{X})^2$

$S=\sqrt{S^2}$

t stuff

estimator and MOE from CI

sampling dist. - sample proportion

hyp test for one population proportion

hyp test - one population mean μ

normal pop, known σ	one sample z-test
normal pop, UNknown σ	one sample t-test

Cheatography

stats formula sheet Cheat Sheet

by cicihex via cheatography.com/197422/cs/41607/

decision errors

type 1	-reject a true Ho -false positive
type 2	-fail to reject false Ho -false negative

relationship

α	prob of type 1 error (same as sig level)
β	prob of type 2

hypothesis test steps

1. check validity of assumptions
a. randomness
b. sample size
c. population distribution
2. set up hypotheses
- identify parameter of interest
3. test statistic and its distribution
4. compute p-value
- confirm level of sig given in advance
5. conclusion interpretation

1. validity

2. hypotheses

3. test statistic

parameter	μ	p
Test Statistic	$\bar{X}$ (sample mean)	$\hat{p}$ (sample proportion)
Standardized Form	$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ or $\frac{\bar{X}-\mu_0}{S/\sqrt{n}}$	$\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

4. p-value

- When the parameter is the population mean μ: $H_0: \mu = \mu_0$

		population	
		Normal	Not normal or unknown (large enough n)
σ	Known	$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \to N(0,1)$
	unknown	$\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$	$\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \to N(0,1)$

- When the parameter is the population proportion p: $H_0: p = p_0$
$\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}} \to N(0,1)$, when $np \geq 10$ and $n(1-p) \geq 10$.

5. conclusion

hypothesis test

or significance testing

test an assumption regarding pop. parameter

method used depends on kind of data and reason

asses plausibility of hypothesis using sample data

hypothesis testing terms

hypothesis	a claim or statement about a characteristic of a population of interest
null hypothesis	statement about the value of a population parameter, such as the population mean (μ) or the population proportion (p)
alt hypothesis	claim to be tested, the opposite of the null hypothesis

hypothesis testing terms (cont)

test statistic	value computed from the sample data that is used in making a decision about the rejection of the null hypothesis; converts the sample mean (x̄) or sample proportion (p̂) to a Z- or t-score under the assumption that the null hypothesis is true;
p-value	area under the curve to the left or right of test statistic; compared to level of significance (α)

critical value

significance level

statistical significance

practical significance

effect size	degree of a relationship between two given variables

standardized effect size

one sided

two sided	tests whether the population parameter is equal to, versus not equal to, some specific value