UNIVERSITY OF HAWAIʻI
Office of Biostatistics & Quantitative Health Sciences
JOHN A. BURNS SCHOOL OF MEDICINE

Biostatistics for Med Students

Lecture 2

John J. Chen, Ph.D.
Professor & Director of Biostatistics Core
UH JABSOM

JABSOM MD7

February 22, 2017

Lecture Objectives

- To understand basic research design principles and data presentation approaches
- To build a foundation which will facilitate the active participation in clinical research
- To fully grasp descriptive statistics
- To introduce key concepts of inferential statistics
- To survey some commonly used statistical approaches
- To be prepared for the USMLE Step 1 biostat/epi questions

Outline

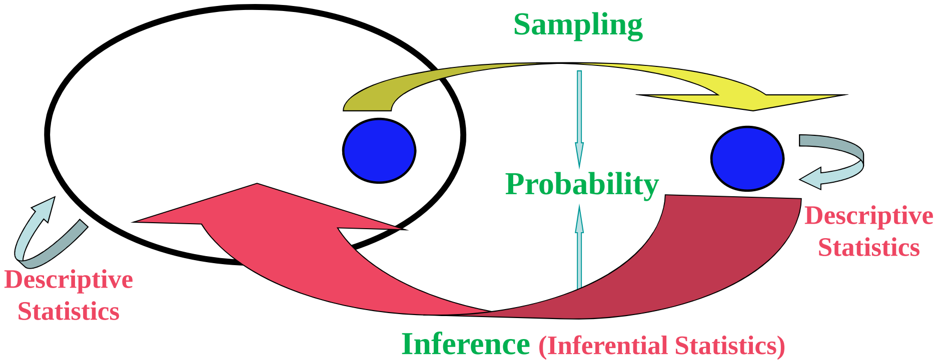
Lecture 1 (02/15/2017)

- The goal of statistics
- Introduction to descriptive biostatistics
- Basic research design principles and data presentation approaches

Lecture 2 (02/22/2017)

- Introduction to inferential statistics
- Commonly used statistical approaches

The Goal of Statistics



Sampling

Probability

Descriptive Statistics

Descriptive Statistics

Inference (Inferential Statistics)

Population

[Parameters]

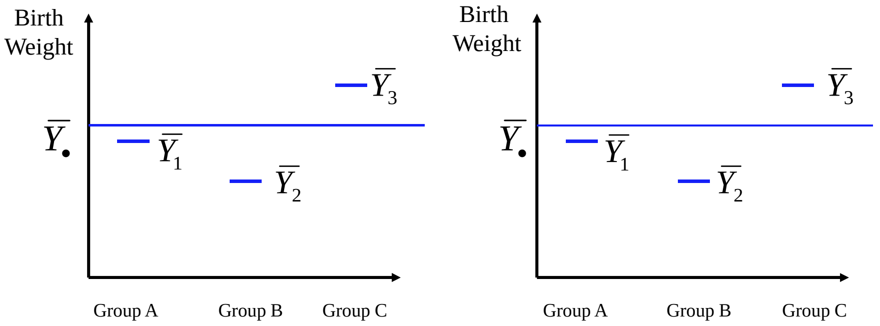
(μ, σ)

Sample

[Statistics]

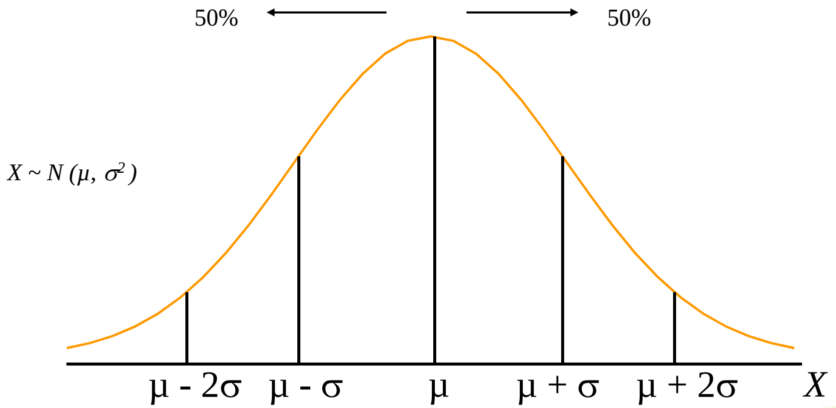
$(\bar{X}, S)$

Effects and Variability



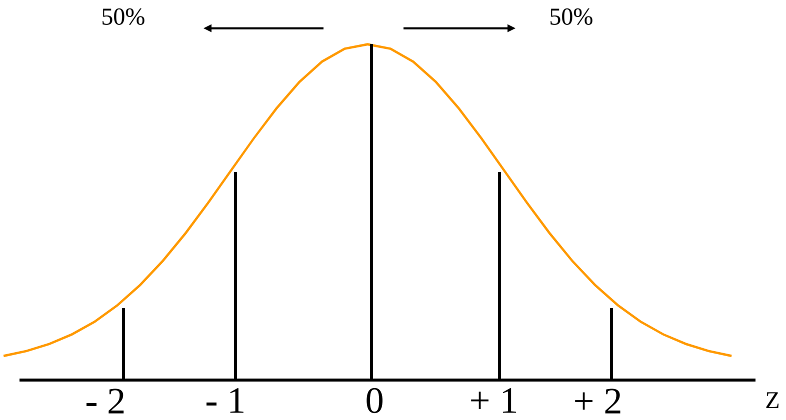
Note: Biological/clinical significance vs. statistical significance

The Normal Distribution



The Normal Distribution

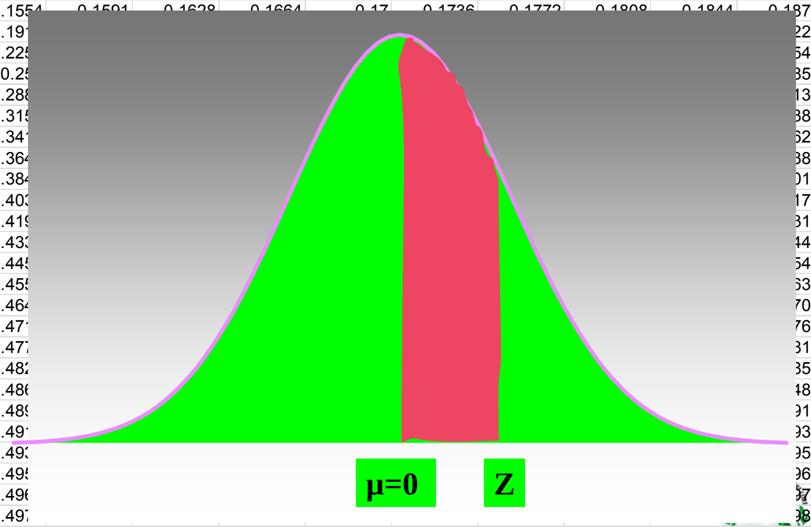
Standard normal distribution: $Z \sim N(\mu = 0, \sigma^2 = 1)$



Given $X \sim N(\mu, \sigma^2)$, we have $Z = (X - \mu)/\sigma$.

Standard Normal Table (From Zero to Z Under The Normal Curve)

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0	0.004	0.008	0.012	0.016	0.019	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.091	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.148	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.17	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.191									224
0.6	0.225									549
0.7	0.25									852
0.8	0.288									133
0.9	0.315									389
1.0	0.341									621
1.1	0.364									383
1.2	0.384									015
1.3	0.403									177
1.4	0.419									319
1.5	0.433									441
1.6	0.445									545
1.7	0.455									633
1.8	0.464									706
1.9	0.471									767
2.0	0.477									817
2.1	0.482									857
2.2	0.486									489
2.3	0.489									916
2.4	0.491									936
2.5	0.493									952
2.6	0.495									964
2.7	0.496									974
2.8	0.497									981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986



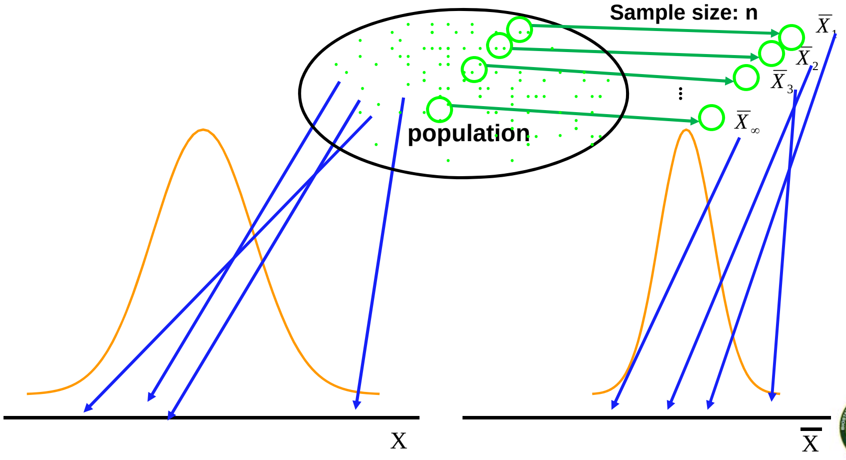
AUC For Normal Distribution

The Rule of Thumb:

Within one s.d.:	68.27% (2/3)
Within two s.d.:	95.45% (95%)
Within three s.d.:	99.74% (99%)

Sampling Distribution

The distribution of individual observations versus the distribution of sample means:



Central Limit Theorem

The distribution of sample means (sampling distribution) from a population is <u>approximately normal as long as the sample size is large</u>, i.e.,

$$\bar{X} \sim N(\mu_{\bar{X}}, \sigma^2_{\bar{X}}) \quad \rightarrow \quad Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

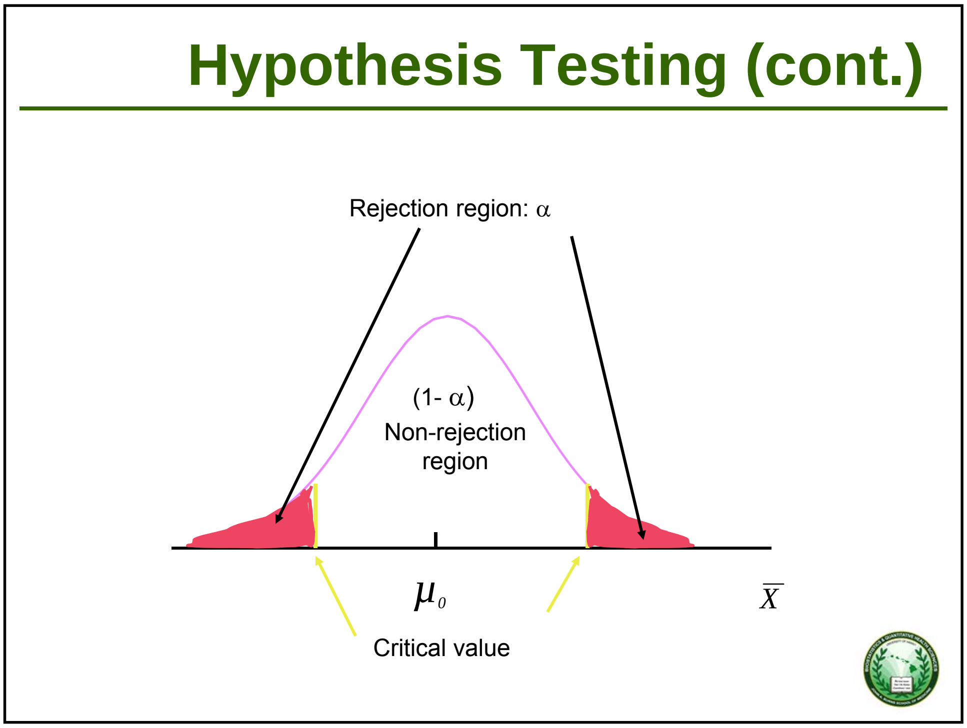
1. The population distribution can be non-normal.
2. Given the population has mean μ, then the mean of the sampling distribution, $\mu_{\bar{X}} = \mu$.
3. If the population has variance σ^2, the standard deviation of the sampling distribution, or the standard error (a measure of the amount of sampling error) is $\sigma_{\bar{X}} = s.e.(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.

Hypothesis Testing

<u>An Example:</u>

Normal serum creatinine level depends on the population studied. From the literature a 4th year JABSOM med student found that one well-established study showed an average sCr of 0.56 (with a standard deviation of 0.15 mg/dL) for 2nd trimester Caucasian pregnant women living on the east coast. But based on her knowledge and experience, she believed that the μ of sCr among Japanese pregnant women in Hawaii seemed different.

She decided to test this by measuring sCr of 49 local Japanese 2nd trimester pregnant women.



Hypothesis Testing (cont.)

Rejection region: α

$(1-\alpha)$
Non-rejection region

μ_0

$\overline{X}$

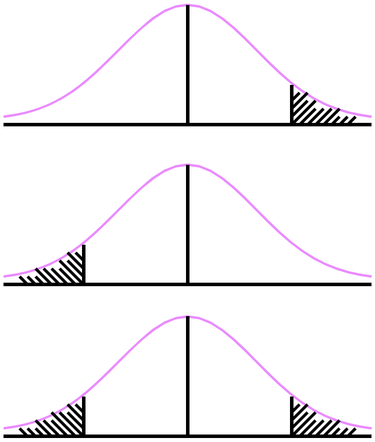
Critical value

Hypothesis Testing

Basic steps of hypothesis testing:

1. State null (H_0:) and alternative (H_1:) hypotheses
2. Choose a significance level, α (usually 0.05 or 0.01)
3. Determine the critical (or rejection) region and the non-rejection region, based on the sampling distribution and under the null hypothesis
4. Based on the sample, calculate the test statistic and compare it with the critical values
5. Make a decision, and state the conclusion

One-tailed vs. Two-tailed Test



Right – Sided Test

Left – Sided Test

Two – Sided Test

Statistical Decision: Errors & Power

		Truth	
		H_0 True	**H_0 False**
Decision	**Reject H_0**	α	**1- β**
	Not reject H_0	**1- α**	β

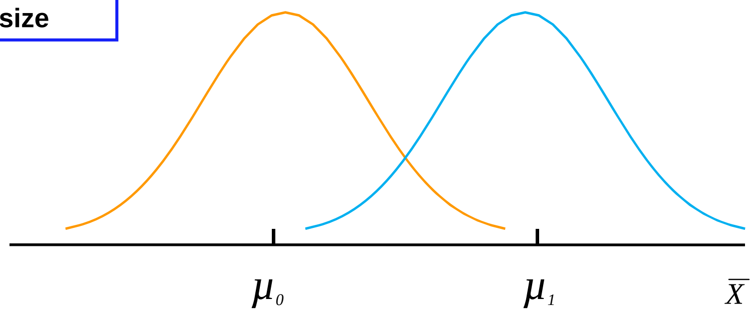
Type I Error (α) - False positives, errors due to chance; Reject H_0 when H_0 is true

Type II Error (β) - False negatives; Don't reject H_0 when H_1 is true

Power: (1- β) = 1 - P (Type II Error)

Statistical Decision

Design factors:
- **effect difference**
- **power**
- **alpha level**
- **std. dev.**
- **sample size**



Now, what is a p-value?

p-values

Interpretation:

The *p*-value is the probability of obtaining a result as extreme or more extreme than the one observed based on the current sample, given the null hypothesis is true.

Note: “Statistically significant” does not necessarily mean “biologically (or clinically) significant”!!!

Hypothesis Testing (cont.)

Example (cont.): Say, the average sCr of the sample of 49 locals is 0.60 mg/dL and the population standard deviation is 0.15 mg/dL (based on the literature).

Step 1. State H_0: and H_1:

$H_0 : \mu_{sCr} = 0.56$ vs. $H_1 : \mu_{sCr} \neq 0.56$

Step 2. Choose a significant level, say, α =0.05.

Step 3. Calculate the test statistic:

$$Z = \frac{\overline{X} - \mu_{sCr}}{\sigma / \sqrt{n}} = \frac{0.60 - 0.56}{0.15/\sqrt{49}} = 1.87.$$

Hypothesis Testing (cont.)

Step 4. Determine the critical region and the non-rejection region:

The critical value: $\pm$ *1.96.*

The rejection region: $| Z | \geq 1.96$.

The non-rejection region: $| Z | < 1.96$.

Step 5. Make a decision, based on the sample, and state the conclusion: As the test statistic $Z = 1.87 < 1.96$, it is within the non-rejection region. Therefore, we do not reject the null hypothesis. We conclude that there is no evidence that the average sCr among local Japanese 2nd trimester women is different from 0.56 mg/dL.

Confidence Intervals

Example (cont.):

A med student wanted to determine the average serum creatinine level among second trimester pregnant Japanese women in Honolulu. From the literature she found that σ for similar populations is about 0.15 mg/dL, but she could not find any information on μ of sCr among local pregnant women. She measured 49 Japanese 2nd trimester women and the sample mean sCr was 0.60 mg/dL. What should be the 95% CI for μ ?

Confidence Intervals

CIs for μ:

90% CI : $\bar{X} \pm 1.645 \frac{\sigma}{\sqrt{n}}$

95% CI : $\bar{X} \pm 1.960 \frac{\sigma}{\sqrt{n}}$

99% CI : $\bar{X} \pm 2.575 \frac{\sigma}{\sqrt{n}}$

The *t* - Distribution

- A small sample from normal distribution
- Unknown population standard deviation, σ

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$ with *n -1* degrees of freedom (d.f.).

The (Student's) t-distribution is very similar to normal distribution, with heavier tails.

t – Table (tail probabilities of the t-distributions)

Degrees of Freedom	2Q (Q)	0.10 (0.05)	0.05 (0.025)	0.01 (0.005)	0.005 (0.0025)	0.001 (0.0005)
1		6.3138	12.706	63.657	127.32	636.62
2		2.9200	4.3026	9.9251	14.0911	31.6075
3		2.3534	3.1825	5.8408	7.4533	12.9258
4		2.1318	2.7764	4.6040	5.5980	8.6087
5		2.0151	2.5706	4.0323	4.7734	6.8701
6		1.9432	2.4469	3.7075	4.3169	5.9590
7		1.8946	2.3646	3.4995	4.0293	5.4088
8		1.8595	2.3060	3.3555	3.8326	5.0421
9		1.8331	2.2621	3.2498	3.6895	4.7805
10		1.8125	2.2281	3.1693	3.5814	4.5871
11		1.7959	2.2010	3.1057	3.4967	4.4374
12		1.7823	2.1788	3.0545	3.4285	4.3184
13		1.7709	2.1604	3.0122	3.3726	4.2215
14		1.7613	2.1448	2.9768	3.3258	4.1412
15		1.7530	2.1314	2.9467	3.2862	4.0735
16		1.7459	2.1199	2.9207	3.2521	4.0157
17		1.7396	2.1098	2.8982	3.2226	3.9659
18		1.7341	2.1009	2.8784	3.1967	3.9224
19		1.7291	2.0930	2.8609	3.1738	3.8841
20		1.7247	2.0860	2.8453	3.1535	3.8502
21		1.7207	2.0796	2.8313	3.1353	3.8200
22		1.7171	2.0739	2.8187	3.1189	3.7928
23		1.7139	2.0687	2.8073	3.1041	3.7683
24		1.7109	2.0639	2.7969	3.0906	3.7461
25		1.7081	2.0595	2.7874	3.0783	3.7258
26		1.7056	2.0555	2.7787	3.0670	3.7073
27		1.7033	2.0518	2.7707	3.0566	3.6903
28		1.7011	2.0484	2.7632	3.0470	3.6746
29		1.6991	2.0452	2.7564	3.0382	3.6601
30		1.6973	2.0423	2.7500	3.0299	3.6466

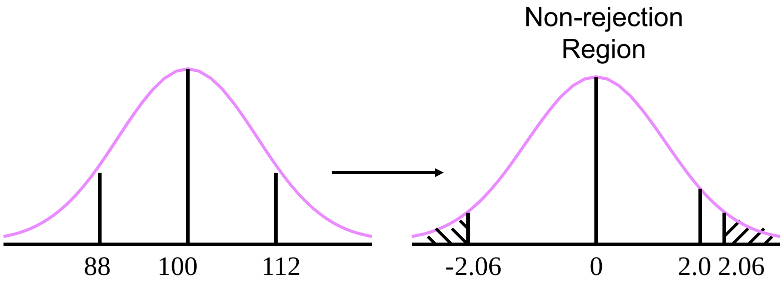
One Sample *t* – Test

Problem: Neonates gain (on average) 100 grams/wk in the first 4 weeks. A sample of 25 infants given a new nutrition formula gained 112 grams/wk (on average) with standard deviation = 30 grams. Is this statistically significant?

Test: H_0: $\mu = 100$ H_1: $\mu \neq 100$

One Sample *t* - Test

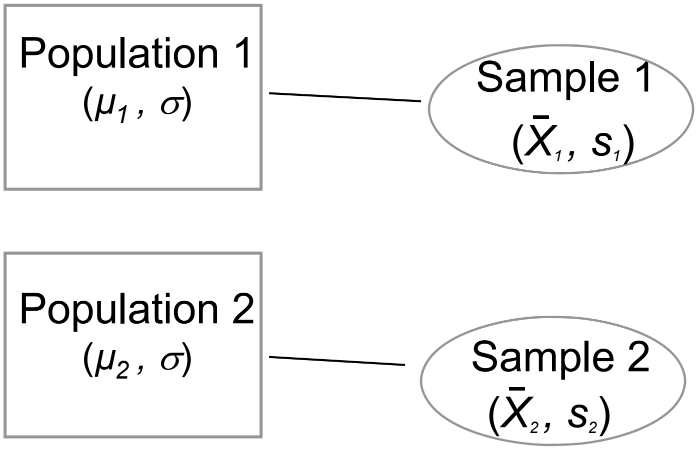
Solution:



$$t = \frac{112 - 100}{30 / \sqrt{25}} = 2.0 < t_{24,\, 0.025} = 2.06$$

Therefore, do not reject (p-value=0.057).

Two Independent Sample *t* - Test



Population 1
(μ_1, σ)

Sample 1
$(\bar{X}_1, s_1)$

Population 2
(μ_2, σ)

Sample 2
$(\bar{X}_2, s_2)$

Test: H_0: $\mu_1 = \mu_2$ versus H_1: $\mu_1 \neq \mu_2$,
assuming $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Two Independent Sample *t* - Test

Test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_{\bar{x}_1 - \bar{x}_2}} \quad \text{with } (n_1 + n_2 - 2)\ d.f.$$

$$S_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

$$S_{\bar{x}_1 - \bar{x}_2} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Two Independent Sample *t* - Test

Problem: Two headache remedies

Brand A: $\overline{X}_1 = 20.1,\ s_1 = 8.7,\ n_1 = 12$

Brand B: $\overline{X}_2 = 18.9,\ s_2 = 7.5,\ n_2 = 12$

Test: $H_0\!:\ \mu_1 = \mu_2 \Rightarrow \quad \mu_1 - \mu_2 = 0$

$H_1\!:\ \mu_1 \neq \mu_2$

Two Independent Sample *t* - Test

Solution:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}1 - \bar{x}2}} = \frac{20.1 - 18.9}{S_{\bar{x}1 - \bar{x}2}} = \frac{1.2}{S_{\bar{x}1 - \bar{x}2}}$$

$$S_p^2 = \frac{(12-1)\,8.7^2 + (12-1)\,7.5^2}{12 + 12 - 2} = \frac{832.59 + 618.75}{22} = 65.97$$

$$\longrightarrow \quad S_p = \sqrt{65.97} = 8.12$$

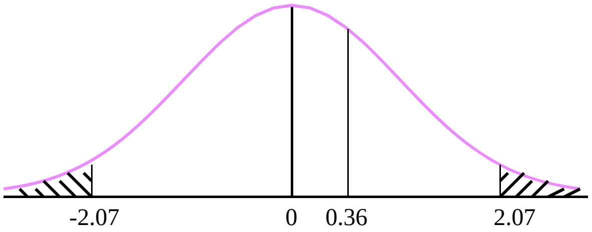
$$S_{\bar{x}1 - \bar{x}2} = 8.12 * \sqrt{\frac{1}{12} + \frac{1}{12}} = 3.3$$

$$t = \frac{1.2}{3.3} = 0.36$$

Two Independent Sample *t* - Test

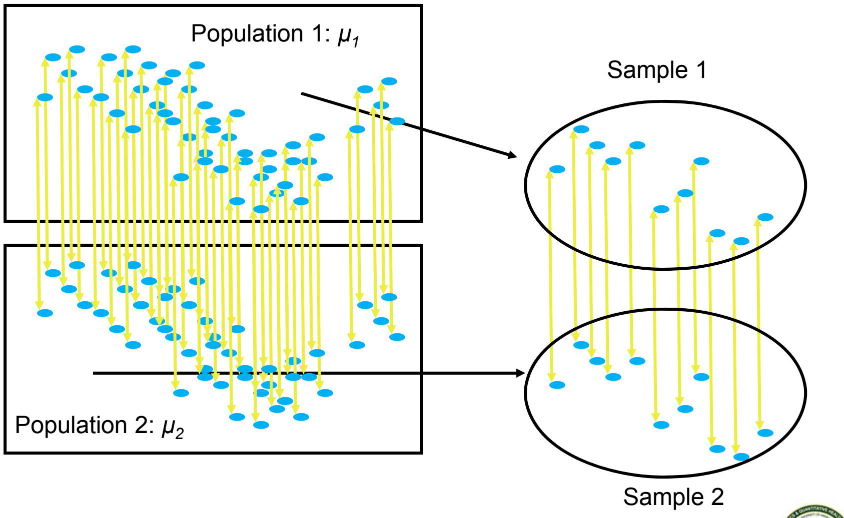
Solution (cont.):

Therefore, do not reject the null, i.e., no statistically significant difference was found between the two remedies.



$$t = \frac{1.2}{3.3} = 0.36, \qquad t_{0.05,\,22} = 2.07$$

Two Paired Sample *t* - Test



Test: H_0: $\mu_2 - \mu_1 = 0$ versus H_1: $\mu_2 - \mu_1 \neq 0$

Two Paired Sample *t* - Test

- Same subject for both treatments:
  -- placebo (X_1) versus active (X_2)
  -- before (X_1) versus after (X_2)
- Intra – individual comparison, e.g., left (X_1) versus right (X_2)

Approach: reduce data to one sample t-test problem. First, calculate the difference, $d = X_2 - X_1$, for each subject; then, perform one sample *t*-test on the *d* scores, with $d.f.=n-1$.

Two Paired Sample *t* - Test

Test: H_0: average difference is zero

Test statistic: $t = \frac{\bar{d} - 0}{S_{\bar{d}}}$

$$\bar{d} = \frac{\sum_i d_i}{n}$$

$$S_{\bar{d}} = S_d / \sqrt{n}$$

$$S_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}}$$

Two Paired Sample *t* - Test

Problem: Does the medication significantly lower blood pressure?

Subject	Reaction to Placebo	Reaction to Med.
1	150	130
2	180	148
3	148	126
4	172	150
5	160	136

Two Paired Sample *t* - Test

Subject	Reaction to Placebo	Reaction to Medication	d
1	150	130	20
2	180	148	32
3	148	126	22
4	172	150	22
5	160	136	24
Total			120

Two Paired Sample *t* - Test

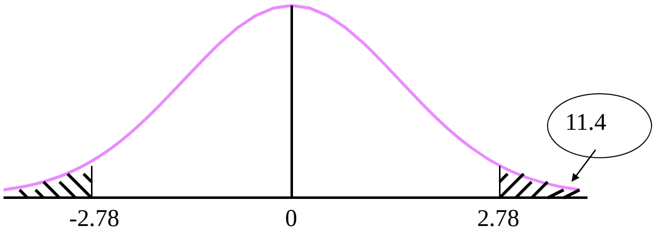
Solution: $\bar{d} = \frac{\sum_i d_i}{n} = \frac{120}{5} = 24$

$$S_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}} = \sqrt{22}$$

$$S_{\bar{d}} = S_d / \sqrt{n} = \sqrt{22} / \sqrt{5} = 2.1$$

$$t = \frac{\bar{d} - 0}{S_{\bar{d}}} = \frac{24}{2.1} = 11.4$$

Two Paired Sample *t* - Test



At 5% level, t - table = 2.78.

For a two-sided test: reject!
There is a statistically significant effect!

Contingency Tables

The most common two-way tables: 2-by-2 tables

For example, smoking and lung cancer status

Lung cancer ← Response variable

Explanatory variable ↓ Smoking

	Lung cancer: Yes	Lung cancer: No	
Smoking: Yes	$\mathbf{n_{11}}$	$\mathbf{n_{12}}$	$\mathbf{n_{1.}}$
Smoking: No	$\mathbf{n_{21}}$	$\mathbf{n_{22}}$	$\mathbf{n_{2.}}$
	$\mathbf{n_{.1}}$	$\mathbf{n_{.2}}$	$\mathbf{n_{..}}$

$$RR = \frac{\frac{n_{11}}{n_{1\bullet}}}{\frac{n_{21}}{n_{2\bullet}}}$$

$$RD = \frac{n_{11}}{n_{1\bullet}} - \frac{n_{21}}{n_{2\bullet}}$$

$$OR = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

Measures of association

- RD: risk difference (prospective studies)
- RR: relative risk (prospective studies)
- OR: odds ratio (prospective or retrospective)

Interpretations

Interpretation of estimates of RR/OR:

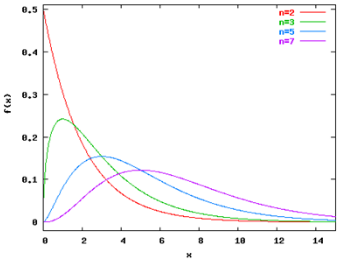
- If RR/OR is above 1, we say there is a positive relationship/association between risk factor and outcome.
- If RR/OR is below 1, we say there is a negative relationship/association between risk factor and outcome.

Interpretation of CI of population RR/OR:

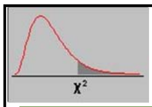
- If the entire interval is above 1, we conclude that the probability that having the outcome of interest is higher in those with risk factor.
- If the entire interval is below 1, we conclude that the probability is lower in those with risk factor.
- We conclude that there is no evidence that the probabilities of the outcome of interest are different between those with vs. without risk factor, if the interval contains 1.

χ^2 Tests

Expected and observed frequencies are compared



- Goodness of Fit of a single variable
- Test of Independence of two variables



Percentiles of the χ^2 - Distributions Table

Degrees of freedom	0.100	0.050	0.010	0.005	0.001
1	2.7055	3.8415	6.6349	7.8944	10.828
2	4.6052	5.9915	9.2102	10.5963	13.8173
3	6.2514	7.8147	11.3447	12.8383	16.2672
4	7.7795	9.4877	13.2768	14.8605	18.4667
5	9.2363	11.0705	15.0864	16.7495	20.5165
6	10.6447	12.5916	16.8118	18.5479	22.4599
7	12.0171	14.0671	18.4751	20.2776	24.3219
8	13.3616	15.5073	20.0900	21.9549	26.1237
9	14.6836	16.9190	21.6658	23.5891	27.8768
10	15.9872	18.3071	23.2095	25.1886	29.5871
11	17.2750	19.6751	24.7250	26.7569	31.2628
12	18.5493	21.0261	26.2170	28.2999	32.9099
13	19.8120	22.3621	27.6882	29.8195	34.5283
14	21.0641	23.6848	29.1409	31.3198	36.1258
15	22.3071	24.9958	30.5778	32.8014	37.6973
16	23.5418	26.2962	31.9998	34.2675	39.2520
17	24.7690	27.5871	33.4086	35.7186	40.7908
18	25.9894	28.8693	34.8052	37.1562	42.3131
19	27.2036	30.1435	36.1912	38.5823	43.8206
20	28.4120	31.4104	37.5660	39.9970	45.3141
21	29.6151	32.6706	38.9321	41.4017	46.7982
22	30.8133	33.9244	40.2893	42.7955	48.2678
23	32.0069	35.1725	41.6383	44.1808	49.7262
24	33.1962	36.4151	42.9797	45.2291	51.1831
25	34.3816	37.6525	44.3144	46.9280	52.6165

χ^2 Tests of Independence

Observed:

	C1	C2	
R1	A	C	A+C
R2	B	D	B+D
	A+B	C+D	A+B+C+D

Expected:

$$E = \frac{\text{(row total) * (column total)}}{\text{(grand total)}}$$

$d.f. = (r-1)*(c-1)$

χ^2 Tests of Independence

<u>Problem</u>: Is the NQO1 gene associated with endemic nephropathy (EN)?

<u>Observed:</u>

	EN cases	***EN controls***	
NQO1 Mutant	***50***	***20***	***70***
NQO1 WT	***10***	***20***	***30***
	60	***40***	***100***

χ^2 Tests of Independence

Solution:

O	E
50	?
10	?
20	?
20	?

50	20	70
10	20	30
60	40	100

$$E = \frac{60 * 70}{100}$$

$$= 42$$

χ^2 Tests of Independence

Solution:

O	E	(O-E)	$(O-E)^2$	$(O-E)^2 / E$
50	42	8	64	1.52
10	18	-8	64	3.56
20	28	-8	64	2.29
20	12	8	64	5.33

$$\chi^2 = 12.70$$

χ^2 with (R-1) * (C-1) d.f.

= (2-1) * (2-1) d.f. = 1 d.f.

χ^2 Tests of Independence

Solution: $\chi^2 = 12.70$

$\chi^2_{1,\, 0.001} = 10.83$

$\longrightarrow$ NQO1 is not independent of EN (an association exists), $p < 0.001$.

Fisher's Exact Tests

Fisher's Exact Test – for small expected frequencies
If any expected frequencies are < 2 or if half of the expected frequencies are < 5, you should use Fisher's Exact Test instead of χ^2.

Fisher's Exact Tests

Fisher's Tea Tasting Experiment

		Poured First (Guessed)		
		Milk	*Tea*	
Poured First (Actual)	*Milk*	*3*	*1*	*4*
	Tea	*1*	*3*	*4*
		4	*4*	*8*

Fisher's Exact Tests

Fisher's Tea Tasting Experiment (cont.):

Based on hypergeometric distribution, the p-value is the sum of all probabilities for tables that give even more evidence in favor of the lady's claim.

		Poured First (Guess)		
		Milk	*Tea*	
Poured First (Actual)	*Milk*	0	4	
	Tea	4	0	

… …

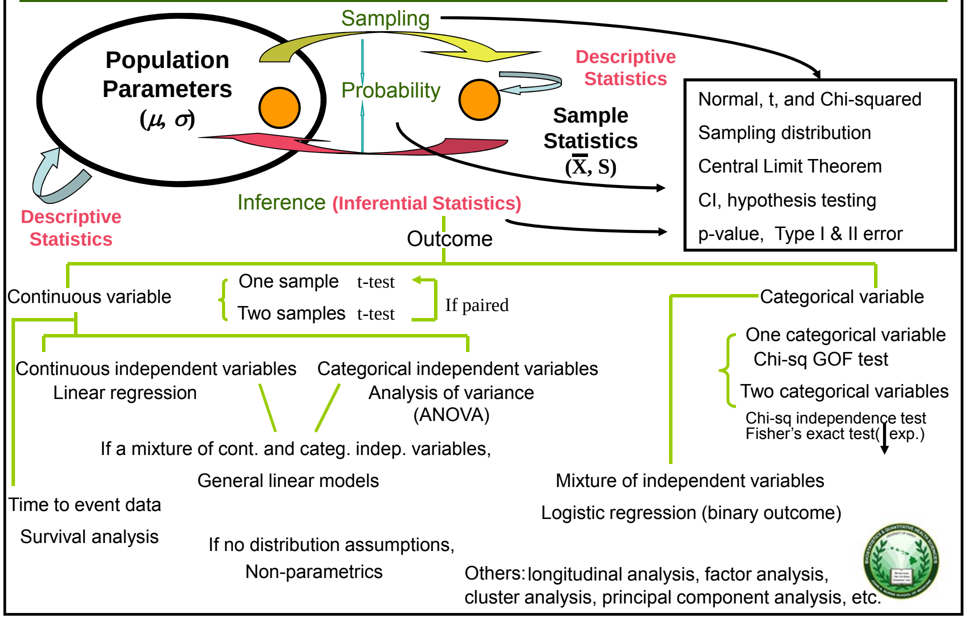
	Poured First (Guess)		
	Milk	*Tea*	
Milk	3	1	
Tea	1	3	

	Poured First (Guess)		
	Milk	*Tea*	
Milk	4	0	
Tea	0	4	

p-value = $P_{(1,1)}(3) + P_{(1,1)}(4) = 0.229 + 0.014 = 0.243$

Therefore, the experiment did not establish a significant association between the actual order of pouring and the woman's guess.

A Roadmap of Common Statistical Tools



Sampling

Population Parameters (μ, σ)

Probability

Descriptive Statistics

Sample Statistics ($\bar{X}$, S)

Normal, t, and Chi-squared
Sampling distribution
Central Limit Theorem
CI, hypothesis testing
p-value, Type I & II error

Descriptive Statistics

Inference (Inferential Statistics)

Outcome

Continuous variable

One sample t-test
Two samples t-test
If paired

Continuous independent variables
Linear regression

Categorical independent variables
Analysis of variance (ANOVA)

If a mixture of cont. and categ. indep. variables,
General linear models

Time to event data
Survival analysis

If no distribution assumptions,
Non-parametrics

Categorical variable

One categorical variable
Chi-sq GOF test
Two categorical variables
Chi-sq independence test
Fisher's exact test(exp.)

Mixture of independent variables
Logistic regression (binary outcome)

Others: longitudinal analysis, factor analysis, cluster analysis, principal component analysis, etc.

Collaboration with A Biostatistician

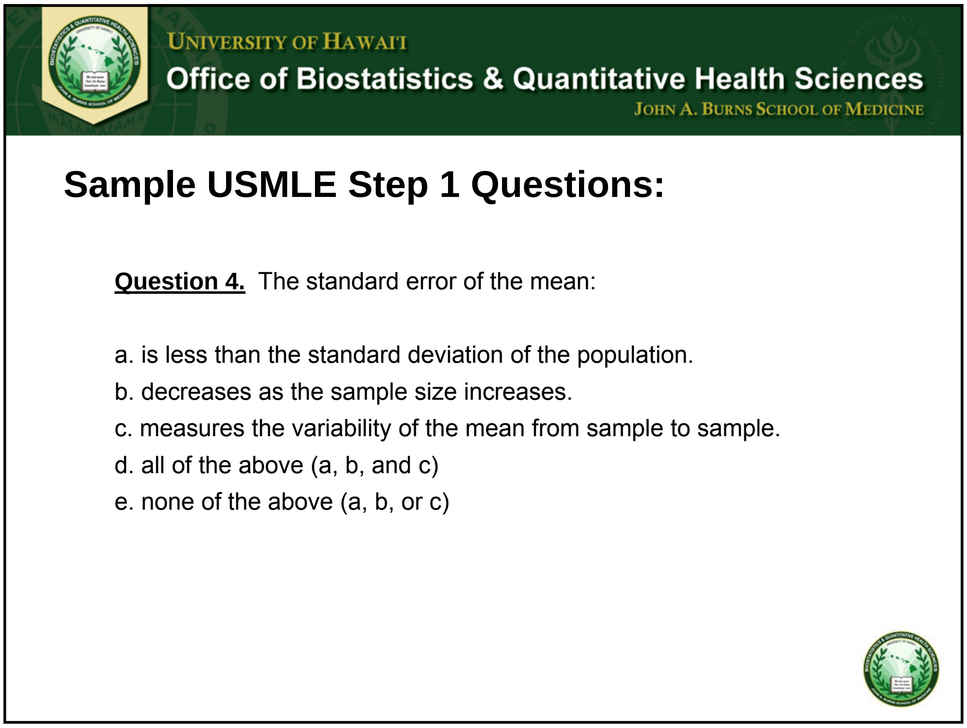
1. Early and often
2. Start the discussion when you have the initial idea
3. It is an iterative process
4. A collaborative effort: equal and fair
5. Ask questions so you can discuss about the general statistical approach without the statistician
6. Education and training in research design and biostatistics

http://biostat.jabsom.hawaii.edu

RMATRIX
NIH U54MD007584

BRIDGES
NIH G12MD007601

IDeA Network of Biomedical Research Excellence
INBRE III
Hawaii State Research and Education Partnership
NIH P20GM103466

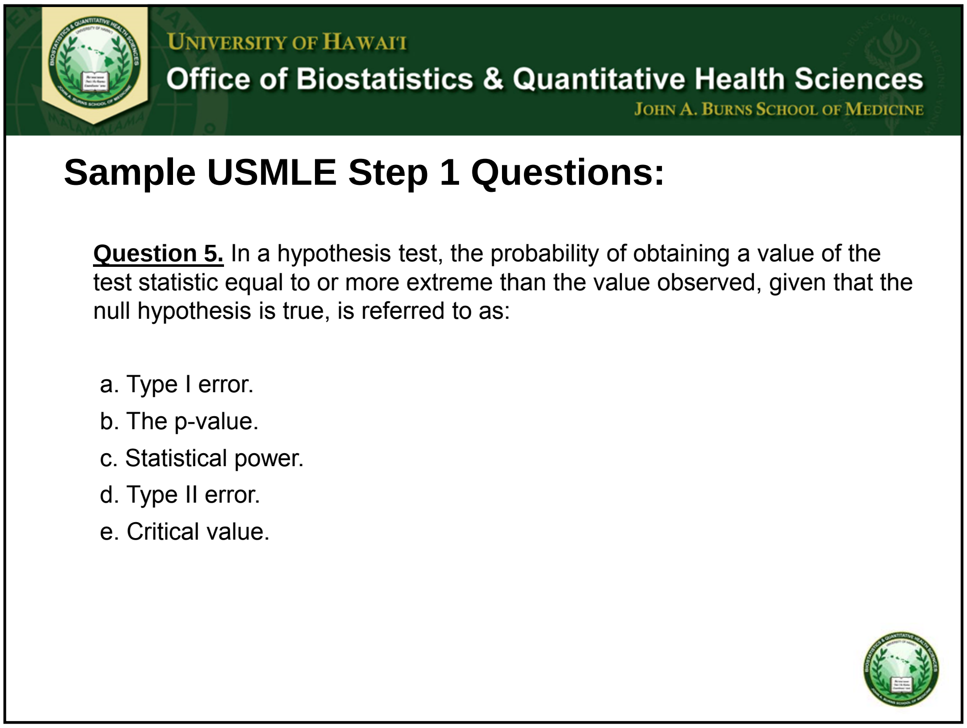


UNIVERSITY OF HAWAIʻI
Office of Biostatistics & Quantitative Health Sciences
JOHN A. BURNS SCHOOL OF MEDICINE

Sample USMLE Step 1 Questions:

Question 4. The standard error of the mean:

a. is less than the standard deviation of the population.
b. decreases as the sample size increases.
c. measures the variability of the mean from sample to sample.
d. all of the above (a, b, and c)
e. none of the above (a, b, or c)

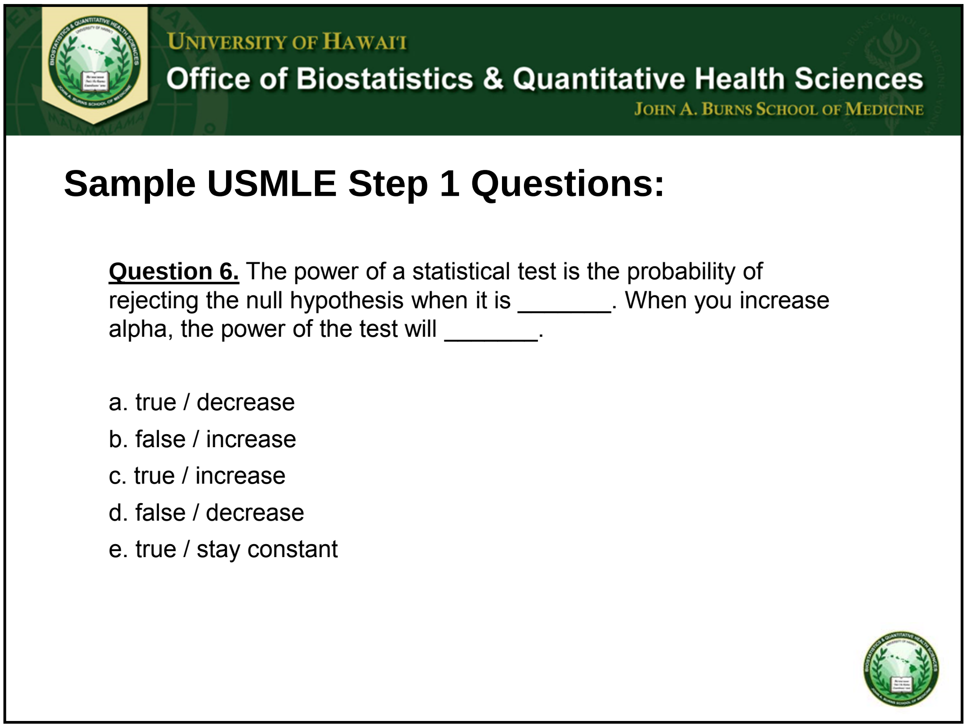


UNIVERSITY OF HAWAIʻI
Office of Biostatistics & Quantitative Health Sciences
JOHN A. BURNS SCHOOL OF MEDICINE

Sample USMLE Step 1 Questions:

Question 5. In a hypothesis test, the probability of obtaining a value of the test statistic equal to or more extreme than the value observed, given that the null hypothesis is true, is referred to as:

a. Type I error.
b. The p-value.
c. Statistical power.
d. Type II error.
e. Critical value.



UNIVERSITY OF HAWAI'I
Office of Biostatistics & Quantitative Health Sciences
JOHN A. BURNS SCHOOL OF MEDICINE

Sample USMLE Step 1 Questions:

Question 6. The power of a statistical test is the probability of rejecting the null hypothesis when it is _______. When you increase alpha, the power of the test will _______.

a. true / decrease
b. false / increase
c. true / increase
d. false / decrease
e. true / stay constant