STA 291
Lecture 19

- **Exam II   Next Tuesday 5-7pm**
- **Memorial Hall (Same place as exam I)**

- **Makeup Exam 7:15pm – 9:15pm**
- **Location CB 234**

Exam II Covers …

- Chapter 9
- 10.1; 10.2; 10.3; 10.4; 10.6
- 12.1; 12.2; 12.3; 12.4

- Formula sheet; normal table; t-table will be provided.

Confidence Interval

- A confidence interval for an unknown parameter is a range of numbers that is *likely* to cover (or capture) the true parameter. (for us, parameter is either p or mu)
- The probability that the confidence interval captures the true parameter is called the confidence level.
- The confidence level is a chosen number close to 1, usually 95%, 90% or 99%

- Why not chose confidence level 100%?

Confidence interval for mu

- For continuous type data, often the parameter is the population mean, mu.
- Chap. 12.1 – 12.4

Chap. 12.1 – 12.4: Confidence Interval for mu

- The ***random*** interval between

$$\bar{X} - 1.96\frac{s}{\sqrt{n}} \quad \text{and} \quad \bar{X} + 1.96\frac{s}{\sqrt{n}}$$

Will capture the population mean, mu, with 95% probability

- This is a confidence statement, and the interval is called a 95% confidence interval
- We need to know sigma. ☹

- confidence level 0.90, ←→ $Z_{a/2}$ =1.645
- confidence level 0.95 ←→ $Z_{a/2}$ =1.96
- confidence level 0.99 ←→ $Z_{a/2}$ =2.575
- Where do these numbers come from? (same number as the confidence interval for p).
- They are from normal table/web

"Student" t - adjustment

- If sigma is unknown, (often the case) we may replace it by s (the sample SD) but the value Z (for example z=1.96) needs adjustment to take into account of extra variability introduced by s
- There is another table to look up: t-table or another applet
- http://www.socr.ucla.edu/Applets.dir/Normal_T_Chi2_F_Tables.htm

William Gosset "student" for the t-table

works for Guinness Brewery 103 years ago

Degrees of freedom, n-1

- Student t - table is keyed by the df – degrees of freedom
- Entries with infinite degrees of freedom is same as Normal table
- When degrees of freedom is over 200, the difference to normal is very small

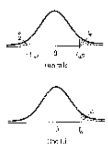
- With the t-adjustment, we ***do not*** require a large sample size n.
- Sample size n can be 25, 18 or 100 etc.

t **Table**

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416

A-78 APPENDIX C

Table T
Values of t_α



Two-tail probability / One-tail probability	0.20 / 0.10	0.10 / 0.05	0.05 / 0.025	0.02 / 0.01	0.01 / 0.005	
df						df
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26

Confidence Intervals

- Confidence Interval Applet
- http://bcs.whfreeman.com/scc/content/cat_040/spt/confidence/confidenceinterval.html

Example: Confidence Interval

- Example: Find and interpret the 95% confidence interval for the population mean, if the sample mean is 70 and the pop. standard deviation is 12, based on a sample of size

n = 100

First we compute $\frac{s}{\sqrt{n}}$ *=12/10= 1.2 ,*

1.96x 1.2=2.352

[70 – 2.352, 70 + 2.352] = [67.648, 72.352]

Example: Confidence Interval

- Now suppose the pop. standard deviation is unknown (often the case). Based on a sample of size *n = 100 , Suppose we also compute the s = 12.6 (in addition to sample mean = 70)*

First we compute $\frac{s}{\sqrt{n}}$ *=12.6/10= 1.26 ,*

From t-table 1.984 x 1.26 = 2.4998

[70 – 2.4998, 70 + 2.4998] = [67.5002, 72.4998]

Confidence Interval: Interpretation

- "Probability" means that "in the long run, 95% of these intervals would contain the parameter"

i.e. If we repeatedly took random samples using the same method, then, in the long run, in 95% of the cases, the confidence interval will cover the true unknown parameter

- For one given sample, we do not know whether the confidence interval covers the true parameter or not. (unless you know the parameter)
- The ***95% probability*** only refers to the ***method*** that we use, but not to the individual sample

Confidence Interval: Interpretation

- To avoid the misleading word "probability", we say:

  "We are 95% **confident** that the interval will contain the true population mean"
- Wrong statement:

  "With 95% probability, the population mean is in the interval from 3.5 to 5.2"

Wrong statement: "95% of all the future observations will fall within 3.5 to 5.2".

Confidence Interval

- If we change the confidence level from 0.95 to 0.99, the confidence interval changes

Increasing the probability that the interval contains the true parameter requires increasing the length of the interval

- In order to achieve 100% probability to cover the true parameter, we would have to increase the length of the interval to infinite -- that would not be informative, not useful.
- There is a tradeoff between length of confidence interval and coverage probability. Ideally, we want short length and high coverage probability (high confidence level).

Different Confidence Coefficients

- In general, a confidence interval for the mean, m has the form

$$\bar{X} \pm z \cdot \frac{s}{\sqrt{n}}$$

- Where z is chosen such that the probability under a normal curve within z standard deviations equals the confidence level

Different Confidence Coefficients

- We can use normal Table to construct confidence intervals for other confidence levels
- For example, there is 99% probability of a normal distribution within 2.575 standard deviations of the mean
- A 99% confidence interval for m is

$$\bar{X} \pm 2.575 \cdot \frac{s}{\sqrt{n}}$$

Error Probability

- The error probability (a) is the probability that a confidence interval does **not** contain the population parameter -- (missing the target)
- For a 95% confidence interval, the error probability a=0.05
- a = 1 - confidence level or

confidence level = 1 – a

Different Confidence Levels

Confidence level	*Error a*	*a/2*	*z*
90%	0.1		
95%	0.05	0.025	1.96
98%			
99%			2.575
99.74%			3
			1.5

- If a 95% confidence interval for the population mean, turns out to be [67.4, 73.6]

What will be the confidence level of the interval [67.8, 73.2]?

Interpretation of Confidence Interval

- If you calculated a 95% confidence interval, say from 10 to 14, The true parameter is either in the interval from 10 to 14, or not – we just don't know it (unless we knew the parameter).
- The 95% probability refers to probability before we do it: (before Joe shoot the free throw, I say he has 77% hitting the hoop. But after he did it, he either hit it or missed it).

Interpretation of Confidence Interval, II

- If you repeatedly calculate confidence intervals with the same method, then 95% of them will contain the true parameter, -- (using the long run average interpretation of the probability.)

Choice of sample size

- In order to achieve a margin of error smaller than B, (with confidence level 95%), how large the sample size n must we get?

Choice of Sample Size

$$\bar{X} \pm z \cdot \frac{s}{\sqrt{n}} = \bar{X} \pm B$$

- So far, we have calculated confidence intervals starting with *z*, *n* and s
- These three numbers determine the error bound *B* of the confidence interval
- Now we reverse the equation:
  - We specify a desired error bound *B*
  - Given *z* and s , we can find the minimal sample size n needed for achieve this.

Choice of Sample Size

- From last page, we have

$$z \cdot \frac{s}{\sqrt{n}} = B$$

- Mathematically, we need to solve the above equation for *n*
- The result is

$$n = s^2 \cdot \left(\frac{z}{B}\right)^2$$

Example

- About how large a sample would have been adequate if we merely needed to estimate the mean to within 0.5, with 95% confidence?
- (assume $s = 5$
- B=0.5, z=1.96
- Plug into the formula: $n = 5^2 \cdot \left(\frac{1.96}{0.5}\right)^2$ =384.16

Choice of sample size

- The most lazy way to do it is to guess a sample size n and

- Compute B, if B not small enough, then increase n;
- If B too small, then decrease n

- For the confidence interval for p

$$z \cdot \frac{\sqrt{p(1-p)}}{\sqrt{n}} = B$$

- Often, we need to put in a rough guess of p (called pilot value). Or, conservatively put p=0.5

- Suppose we want a 95%confidence error bound B=3% (margin of error + - 3%).

Suppose we do not have a pilot p value, so use p = 0.5

So, n = 0.5(1-0.5) [1.96/0.03]^2=1067.11

Attendance Survey Question

- On a 4”x6” index card
  - Please write down your name and section number
  - Today’s Question:

  Which t-table you like better?

Facts About Confidence Intervals I

- The width of a confidence interval
  - Increases as the confidence level increases
  - Increases as the error probability decreases
  - Increases as the standard error increases
  - Decreases as the sample size *n* increases

- www.webchem.sci.ru.nl/Stat/index.html

Try to teach us confidence interval but the interpretation is all wrong ☹

- For Bernoulli type data, the future observations NEVER fall into the confidence interval