Lecture 23

Last time:

- We want to find a confidence interval on the mean μ in the situation where we are sampling a normal distribution with unknown variance.

- In this situation, we consider

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

the student's t-distribution with $(n-1)$ degrees of freedom.

- Here, s is the sample standard deviation, n is the sample size.
- if $\bar{x}$ is a sample mean, then a $100(1-\alpha)\%$ CI for μ is given by

$$P\left(\bar{x} - t_{\alpha/2, n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2, n-1}\frac{s}{\sqrt{n}}\right) = 1-\alpha.$$

where $t_{\alpha/2, n-1}$ is the value of $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ such that $P(T \leq t_{\alpha/2, n-1}) = 1 - \alpha/2$.

Ex: Out of all the Stats 3Y/3J students 12 write their upcoming test at an alternate time. For these students, we found that $\bar{x} = 77\%$ with $s = 13.61\%$. Find a 95% CI for the average of the whole class.

Soln: First, we find α: $100(1-\alpha) = 95$

$$\Downarrow$$

$$1-\alpha = 0.95 \rightarrow \alpha = 0.05.$$

We have that the endpoints of our interval are $\bar{x} \pm \left(t_{\alpha/2,\ \underbrace{12-1}_{=11}}\right)\frac{s}{\sqrt{12}} = 77 \pm \left(t_{0.025,11}\right)\frac{13.6}{\sqrt{12}}$

So we need only to find $t_{0.025,11}$:

From the chart, $t_{0.025,11} = 2.201$.

Hence the CI is $77 \pm (2.201)\frac{13.6}{\sqrt{12}}$

$$= 77 \pm \underbrace{8.647}_{\text{Error.}}$$

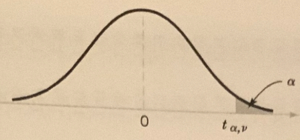

TABLE V **Percentage Points $t_{\alpha,\nu}$ of the t Distribution**

ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ν = degrees of freedom.

Confidence Intervals for Population Proportion.

- Suppose we take a population and we are interested in a certain subclass
  - ↳ eg the population could be 31/137 students and the subclass of interest could be the collection of students dressing up for Halloween.
- Let p is the proportion of the population in the class of interest.
- Suppose we take a large sample n. Let X $(\leq n)$ be the number of of samples which fall into this class.
- Then $\hat{P} = \frac{X}{n}$ is a point estimator of p
- Note that X is binomial $Bin(n, p)$ and so $E[\hat{P}] = E\left[\frac{X}{n}\right] = \frac{1}{n} E[X] = \frac{1}{n}(np) = p$.
  - ↳ i.e. $\hat{P}$ is an unbiased estimator of p.
- For large n, we have that

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

is approximately the standard normal.

- Suppose now that we want to find a $100(1-\alpha)\%$ CI for p.

- taking $z_{\alpha/2}$ s.t. $P(Z \leq z_{\alpha/2}) = 1 - \alpha/2$, we have that

$$P\left(-z_{\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}\right) \simeq 1-\alpha$$

and so

$$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) \simeq 1-\alpha.$$

- Here, $\sqrt{\frac{p(1-p)}{n}}$ is the standard error of $\hat{p}$, $\sigma_{\hat{p}}$ (from chapter 7).

- This is not ideal, since the upper and lower confidence bounds contain p, which is unknown.
  - ↳ however we can approximate p by $\hat{p}$, giving:

$$P\Bigg(\underbrace{\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}_{\text{approx. } 100(1-\alpha)\% \text{ CI on } p.}\Bigg) \simeq 1-\alpha.$$

- Note that this procedure depends on having a good approximation to the binomial distribution, and so we want $np \geq n(1-p) > 5$.

Ex: Out of a random sample of 50 engineers students; 18 are dressing up for Halloween. Find a 99% confidence interval for the proportion of eng students who are dressing up for Halloween.

<u>Soln</u>: $\alpha = 0.01$, so $1 - \alpha/2 = 0.995$.

<u>Z-scores</u>: $z_{\alpha/2} = 2.58$.

Now the point estimate $\hat{p}$ for P is $\hat{p} = 18/50 = 0.36$.

So the 99% CI is $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$$= 0.36 \pm (2.58)\sqrt{\frac{0.36(1-0.36)}{50}}$$

$$= 0.36 \pm 0.175.$$

Now, suppose we want a $100(1-\alpha)\%$ CI for the proportion with a certain amount of error, $E = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

- what should n be?

$$n = \left\lceil \frac{z_{\alpha/2}^2}{E^2} \underbrace{\left(\hat{p}(1-\hat{p})\right)}_{\text{but what if we don't know } \hat{p}\text{??}} \right\rceil$$

<u>2 methods</u>:

1. use an earlier study and compute $\hat{p}$ from that (eg: in the last example, found $\hat{p} = 0.36$; use that).

2. Observe that $p(1-p) \leq 0.25$ (when $p=0.5$) and so substitute $\hat{p} = 0.5$ as a worst case scenario: $n = \left\lceil \left(\frac{z_{\alpha/2}}{E}\right)^2 (0.25) \right\rceil$

Ex: We want a 95% CI for the proportion of eng students dressing up for halloween with error at most $E = 0.07$. What should the sample size be?

$\alpha = 0.05 \rightsquigarrow z_{\alpha/2} = 1.96.$

So $n = \left\lceil \left(\frac{1.96}{0.07}\right)^2 (\hat{p}(1-\hat{p})) \right\rceil$.

method 1: In the small test study (sample size 50) we had $\hat{p} = 0.36$.

So $n \geq \left\lceil \left(\frac{1.96}{0.07}\right)^2 (0.36)(1-0.36) \right\rceil = \lceil 180.36 \rceil = 181$

method 2: (worst case scenario) $\hat{p}(1-\hat{p}) = 0.25$

So $n \geq \left\lceil \left(\frac{1.96}{0.07}\right)^2 (0.25) \right\rceil = 196.$