Confidence intervals for unknown variances and proportions

Chrysafis Vogiatzis

Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Lecture 21

ILLINOIS
ISE | Industrial & Enterprise Systems Engineering
GRAINGER COLLEGE OF ENGINEERING

Last time..

We discussed **confidence intervals on means**.

- **Problem**: we do not know the mean of a population and we would like to find it out, please. **Strategy**: select a sample, take the average ($\overline{X}$).
- Point estimation: $\overline{X}$.
- Interval estimation: a confidence interval $[L, U]$ around the average.

1. Normally distributed population, with known variance σ^2.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
2. Normally distributed population, with unknown variance.
   - Do not have σ: we estimate it by the sample variance, s.
   - Find $t_{\alpha/2,n-1}$ (two-sided) or $t_{\alpha,n-1}$ (one-sided).
   - $\left[\overline{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \overline{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right]$
3. Not normally distributed population, but large enough sample ($n \geq 30$).
   - Do not have σ again: we use sample variance, s.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$

ILLINOIS

Last time..

We discussed **confidence intervals on means**.

- **Problem**: we do not know the mean of a population and we would like to find it out, please. **Strategy**: select a sample, take the average ($\overline{X}$).
- Point estimation: $\overline{X}$.
- Interval estimation: a confidence interval $[L, U]$ around the average.

1. Normally distributed population, with known variance σ^2.
2. Normally distributed population, with unknown variance.
3. Not normally distributed population, but large enough sample ($n \geq 30$).

ILLINOIS

Last time..

We discussed **confidence intervals on means**.

- **Problem**: we do not know the mean of a population and we would like to find it out, please. **Strategy**: select a sample, take the average ($\overline{X}$).
- Point estimation: $\overline{X}$.
- Interval estimation: a confidence interval $[L, U]$ around the average.

1. Normally distributed population, with known variance σ^2.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
2. Normally distributed population, with unknown variance.
   - Do not have σ: we estimate it by the sample variance, s.
   - Find $t_{\alpha/2,n-1}$ (two-sided) or $t_{\alpha,n-1}$ (one-sided).
   - $\left[\overline{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \overline{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right]$
3. Not normally distributed population, but large enough sample ($n \geq 30$).
   - Do not have σ again: we use sample variance, s.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$

Last time..

We discussed **confidence intervals on means**.

- **Problem**: we do not know the mean of a population and we would like to find it out, please. **Strategy**: select a sample, take the average ($\overline{X}$).
- Point estimation: $\overline{X}$.
- Interval estimation: a confidence interval $[L, U]$ around the average.

1. Normally distributed population, with known variance σ^2.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
2. Normally distributed population, with unknown variance.
   - Do not have σ: we estimate it by the sample variance, s.
   - Find $t_{\alpha/2,n-1}$ (two-sided) or $t_{\alpha,n-1}$ (one-sided).
   - $\left[\overline{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \overline{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right]$
3. Not normally distributed population, but large enough sample ($n \geq 30$).
   - Do not have σ again: we use sample variance, s.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$

ILLINOIS

Last time..

We discussed **confidence intervals on means**.

- **Problem**: we do not know the mean of a population and we would like to find it out, please. **Strategy**: select a sample, take the average ($\overline{X}$).
- Point estimation: $\overline{X}$.
- Interval estimation: a confidence interval $[L, U]$ around the average.

1. Normally distributed population, with known variance σ^2.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
2. Normally distributed population, with unknown variance.
   - Do not have σ: we estimate it by the sample variance, s.
   - Find $t_{\alpha/2,n-1}$ (two-sided) or $t_{\alpha,n-1}$ (one-sided).
   - $\left[\overline{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \overline{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right]$
3. Not normally distributed population, but large enough sample ($n \geq 30$).
   - Do not have σ again: we use sample variance, s.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$

ILLINOIS

Last time..

We discussed **confidence intervals on means**.

- **Problem**: we do not know the mean of a population and we would like to find it out, please. **Strategy**: select a sample, take the average ($\overline{X}$).
- Point estimation: $\overline{X}$.
- Interval estimation: a confidence interval $[L, U]$ around the average.

1. Normally distributed population, with known variance σ^2.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
2. Normally distributed population, with unknown variance.
   - Do not have σ: we estimate it by the sample variance, s.
   - Find $t_{\alpha/2,n-1}$ (two-sided) or $t_{\alpha,n-1}$ (one-sided).
   - $\left[\overline{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \overline{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right]$
3. Not normally distributed population, but large enough sample ($n \geq 30$).
   - Do not have σ again: we use sample variance, s.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$

ILLINOIS

Last time..

We discussed **confidence intervals on means**.

- **Problem**: we do not know the mean of a population and we would like to find it out, please. **Strategy**: select a sample, take the average ($\overline{X}$).
- Point estimation: $\overline{X}$.
- Interval estimation: a confidence interval $[L, U]$ around the average.

1. Normally distributed population, with known variance σ^2.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
2. Normally distributed population, with unknown variance.
   - Do not have σ: we estimate it by the sample variance, s.
   - Find $t_{\alpha/2,n-1}$ (two-sided) or $t_{\alpha,n-1}$ (one-sided).
   - $\left[\overline{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \overline{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right]$
3. Not normally distributed population, but large enough sample ($n \geq 30$).
   - Do not have σ again: we use sample variance, s.
   - Find $z_{\alpha/2}$ (two-sided) or z_α (one-sided).
   - $\left[\overline{X} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$

ILLINOIS

Quick check

1. For a 90% two-sided confidence interval for the mean of a population that is normally distributed, we need:

   a. a sample of size n, the variance of the population σ^2, and $z_{0.1}$.
   b. a sample of size n, the variance of the population σ^2, and $z_{0.05}$.
   c. a sample of size n, the variance of the population σ^2, and $t_{0.05,n-1}$.
   d. the variance of the population σ^2, and $z_{0.05}$.

2. For a 90% one-sided confidence interval for the mean of a population with unknown distribution, we need:

   a. a sample of size $n = 50$, the variance of the sample s^2, and $z_{0.1}$.
   b. a sample of size $n = 20$, the variance of the sample s^2, and $z_{0.05}$.
   c. a sample of size $n = 100$, the variance of the sample s^2, and $t_{0.05,n-1}$.
   d. we can't calculate anything when the population is not normally distributed.

ILLINOIS

Quick check

1. For a 90% two-sided confidence interval for the mean of a population that is normally distributed, we need:

   a. a sample of size n, the variance of the population σ^2, and $z_{0.1}$.
   b. a sample of size n, the variance of the population σ^2, and $z_{0.05}$.
   c. a sample of size n, the variance of the population σ^2, and $t_{0.05,n-1}$.
   d. the variance of the population σ^2, and $z_{0.05}$.

2. For a 90% one-sided confidence interval for the mean of a population with unknown distribution, we need:

   a. a sample of size $n = 50$, the variance of the sample s^2, and $z_{0.1}$.
   b. a sample of size $n = 20$, the variance of the sample s^2, and $z_{0.05}$.
   c. a sample of size $n = 100$, the variance of the sample s^2, and $t_{0.05,n-1}$.
   d. we can't calculate anything when the population is not normally distributed.

ILLINOIS

Quick check

1. For a 90% two-sided confidence interval for the mean of a population that is normally distributed, we need:
   a. a sample of size n, the variance of the population σ^2, and $z_{0.1}$.
   b. a sample of size n, the variance of the population σ^2, and $z_{0.05}$.
   c. a sample of size n, the variance of the population σ^2, and $t_{0.05,n-1}$.
   d. the variance of the population σ^2, and $z_{0.05}$.

2. For a 90% one-sided confidence interval for the mean of a population with unknown distribution, we need:
   a. a sample of size $n = 50$, the variance of the sample s^2, and $z_{0.1}$.
   b. a sample of size $n = 20$, the variance of the sample s^2, and $z_{0.05}$.
   c. a sample of size $n = 100$, the variance of the sample s^2, and $t_{0.05,n-1}$.
   d. we can't calculate anything when the population is not normally distributed.

ILLINOIS

Quick check

1. For a 90% two-sided confidence interval for the mean of a population that is normally distributed, we need:
   a. a sample of size n, the variance of the population σ^2, and $z_{0.1}$.
   b. a sample of size n, the variance of the population σ^2, and $z_{0.05}$.
   c. a sample of size n, the variance of the population σ^2, and $t_{0.05,n-1}$.
   d. the variance of the population σ^2, and $z_{0.05}$.

2. For a 90% one-sided confidence interval for the mean of a population with unknown distribution, we need:
   a. a sample of size $n = 50$, the variance of the sample s^2, and $z_{0.1}$.
   b. a sample of size $n = 20$, the variance of the sample s^2, and $z_{0.05}$.
   c. a sample of size $n = 100$, the variance of the sample s^2, and $t_{0.05,n-1}$.
   d. we can't calculate anything when the population is not normally distributed.

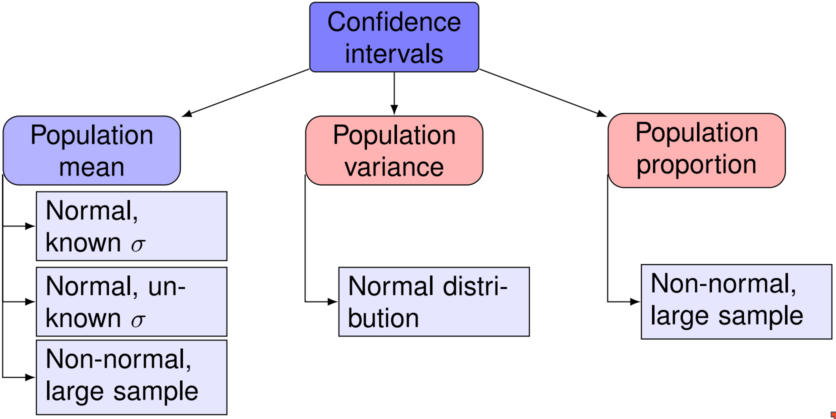
Quick check

1. For a 90% two-sided confidence interval for the mean of a population that is normally distributed, we need:
   - **a.** a sample of size n, the variance of the population σ^2, and $z_{0.1}$.
   - **b.** a sample of size n, the variance of the population σ^2, and $z_{0.05}$.
   - **c.** a sample of size n, the variance of the population σ^2, and $t_{0.05,n-1}$.
   - **d.** the variance of the population σ^2, and $z_{0.05}$.

2. For a 90% one-sided confidence interval for the mean of a population with unknown distribution, we need:
   - **a.** a sample of size $n = 50$, the variance of the sample s^2, and $z_{0.1}$.
   - **b.** a sample of size $n = 20$, the variance of the sample s^2, and $z_{0.05}$.
   - **c.** a sample of size $n = 100$, the variance of the sample s^2, and $t_{0.05,n-1}$.
   - **d.** we can't calculate anything when the population is not normally distributed.

Confidence intervals overview

In today's lecture, we will discuss population variances and proportions:



Confidence intervals for population variance

- We have a good estimator for the population variance σ^2 in the sample variance s^2 after collecting a sample of size n.
- We have shown earlier that $E\left[s^2\right] = \sigma^2$.
- The question now is: what is the sampling distribution of s^2?
  - It is distributed with a χ^2 distribution with $n-1$ degrees of freedom.

Much like our analysis for other confidence intervals, we focus on identifying critical values for the χ^2-distribution, that is:

$$P(X^2 \geq \chi^2_{\alpha,n-1}) = \alpha.$$

For example, here are some often used values:

- $\chi^2_{0.05,5} = 11.07$
- $\chi^2_{0.1,5} = 15.086$
- $\chi^2_{0.95,55} = 38.958$
- $\chi^2_{0.9,20} = 12.443$

Confidence intervals for population variance

- We have a good estimator for the population variance σ^2 in the sample variance s^2 after collecting a sample of size n.
- We have shown earlier that $E\left[s^2\right] = \sigma^2$.
- The question now is: what is the sampling distribution of s^2?
  - It is distributed with a χ^2 distribution with $n-1$ degrees of freedom.

Much like our analysis for other confidence intervals, we focus on identifying critical values for the χ^2-distribution, that is:

$$P(X^2 \geq \chi^2_{\alpha,n-1}) = \alpha.$$

For example, here are some often used values:

- $\chi^2_{0.05,5} = 11.07$
- $\chi^2_{0.1,5} = 15.086$
- $\chi^2_{0.95,55} = 38.958$
- $\chi^2_{0.9,20} = 12.443$

Confidence intervals for population variance

- We have a good estimator for the population variance σ^2 in the sample variance s^2 after collecting a sample of size n.
- We have shown earlier that $E\left[s^2\right] = \sigma^2$.
- The question now is: what is the sampling distribution of s^2?
  - It is distributed with a χ^2 distribution with $n-1$ degrees of freedom.

Much like our analysis for other confidence intervals, we focus on identifying critical values for the χ^2-distribution, that is:

$$P(X^2 \geq \chi^2_{\alpha,n-1}) = \alpha.$$

For example, here are some often used values:

- $\chi^2_{0.05,5} = 11.07$
- $\chi^2_{0.1,5} = 15.086$
- $\chi^2_{0.95,55} = 38.958$
- $\chi^2_{0.9,20} = 12.443$

ILLINOIS

Confidence intervals for population variance

- We have a good estimator for the population variance σ^2 in the sample variance s^2 after collecting a sample of size n.
- We have shown earlier that $E\left[s^2\right] = \sigma^2$.
- The question now is: what is the sampling distribution of s^2?
  - It is distributed with a χ^2 distribution with $n-1$ degrees of freedom.

Much like our analysis for other confidence intervals, we focus on identifying critical values for the χ^2-distribution, that is:

$$P(X^2 \geq \chi^2_{\alpha,n-1}) = \alpha.$$

For example, here are some often used values:

- $\chi^2_{0.05,5} = 11.07$
- $\chi^2_{0.1,5} = 15.086$
- $\chi^2_{0.95,55} = 38.958$
- $\chi^2_{0.9,20} = 12.443$

ILLINOIS

Confidence intervals for population variance

- We have a good estimator for the population variance σ^2 in the sample variance s^2 after collecting a sample of size n.
- We have shown earlier that $E\left[s^2\right] = \sigma^2$.
- The question now is: what is the sampling distribution of s^2?
  - It is distributed with a χ^2 distribution with $n - 1$ degrees of freedom.

Much like our analysis for other confidence intervals, we focus on identifying critical values for the χ^2-distribution, that is:

$$P(X^2 \geq \chi^2_{\alpha,n-1}) = \alpha.$$

For example, here are some often used values:

- $\chi^2_{0.05,5} = 11.07$
- $\chi^2_{0.1,5} = 15.086$
- $\chi^2_{0.95,55} = 38.958$
- $\chi^2_{0.9,20} = 12.443$

χ^2 table

ν	99.9%	99.5%	99.0%	97.5%	95.0%	90.0%	87.5%	80.0%	75.0%	66.7%	50.0%	40.0%	33.3%	25.0%	20.0%	12.5%	10.0%	5.0%	2.5%	1.0%	0.5%	0.1%
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	9.222	11.160	12.198	13.844	15.379	17.292	18.021	19.820	20.843	22.399	25.336	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	9.803	11.808	12.879	14.573	16.151	18.114	18.861	20.703	21.749	23.339	26.336	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	10.391	12.461	13.565	15.308	16.928	18.939	19.704	21.588	22.657	24.280	27.336	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	10.986	13.121	14.256	16.047	17.708	19.768	20.550	22.475	23.567	25.222	28.336	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	11.588	13.787	14.953	16.791	18.493	20.599	21.399	23.364	24.478	26.165	29.336	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	14.688	17.192	18.509	20.569	22.465	24.797	25.678	27.836	29.054	30.894	34.336	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	60.275	66.619
40	17.916	20.707	22.164	24.433	26.509	29.051	30.008	32.345	33.660	35.643	39.335	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	21.251	24.311	25.901	28.366	30.612	33.350	34.379	36.884	38.291	40.407	44.335	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	24.674	27.991	29.707	32.357	34.764	37.689	38.785	41.449	42.942	45.184	49.335	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661
55	28.173	31.735	33.570	36.398	38.958	42.060	43.220	46.036	47.610	49.972	54.335	57.016	58.945	61.665	63.577	67.211	68.796	73.311	77.380	82.292	85.749	93.168
60	31.738	35.534	37.485	40.482	43.188	46.459	47.680	50.641	52.294	54.770	59.335	62.135	64.147	66.981	68.972	72.751	74.397	79.082	83.298	88.379	91.952	99.607

ILLINOIS

Two-sided confidence interval on the variance

Once more, assume we have a sample $X_1, X_2, \ldots, X_n$. Then:

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

and hence:

$$P\left(\chi^2_{1-\alpha/2,n-1} \leq X^2 \leq \chi^2_{\alpha/2,n-1}\right) = 1 - \alpha.$$

By converting back to the σ^2 space, we get:

$$P\left(\frac{(n-1)\,s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)\,s^2}{\chi^2_{1-\alpha/2,n-1}}\right),$$

where the two bounds are (in $[L, U]$ form):

$$L = \frac{(n-1)\,s^2}{\chi^2_{\alpha/2,n-1}}$$

$$U = \frac{(n-1)\,s^2}{\chi^2_{1-\alpha/2,n-1}}$$

Two-sided confidence interval on the variance

Once more, assume we have a sample $X_1, X_2, \ldots, X_n$. Then:

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

and hence:

$$P\left(\chi^2_{1-\alpha/2,n-1} \leq X^2 \leq \chi^2_{\alpha/2,n-1}\right) = 1 - \alpha.$$

By converting back to the σ^2 space, we get:

$$P\left(\frac{(n-1)\, s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)\, s^2}{\chi^2_{1-\alpha/2,n-1}}\right),$$

where the two bounds are (in $[L, U]$ form):

$$L = \frac{(n-1)\, s^2}{\chi^2_{\alpha/2,n-1}}$$

$$U = \frac{(n-1)\, s^2}{\chi^2_{1-\alpha/2,n-1}}$$

ILLINOIS

Differences from the other confidence intervals

Assume we are interested in finding a $(1-\alpha)$-confidence interval. Then, be careful with the following.

1. There are no squares involved! You do not "square" the value: this is simply the name of the distribution!
2. Notice that the critical values are not symmetric.
   - For the lower bound, we use $\chi^2_{\alpha/2,n-1}$;
   - for the upper bound, $\chi^2_{1-\alpha/2,n-1}$.
3. There is no symmetry: on top of that, you are dividing the estimator (rather than adding/subtracting to the estimator).

ILLINOIS

Differences from the other confidence intervals

Assume we are interested in finding a $(1-\alpha)$-confidence interval. Then, be careful with the following.

1. There are no squares involved! You do not “square” the value: this is simply the name of the distribution!
2. Notice that the critical values are not symmetric.
   - For the lower bound, we use $\chi^2_{\alpha/2,n-1}$;
   - for the upper bound, $\chi^2_{1-\alpha/2,n-1}$.
3. There is no symmetry: on top of that, you are dividing the estimator (rather than adding/subtracting to the estimator).

Differences from the other confidence intervals

Assume we are interested in finding a $(1-\alpha)$-confidence interval. Then, be careful with the following.

1. There are no squares involved! You do not "square" the value: this is simply the name of the distribution!
2. Notice that the critical values are not symmetric.
   - For the lower bound, we use $\chi^2_{\alpha/2,n-1}$;
   - for the upper bound, $\chi^2_{1-\alpha/2,n-1}$.
3. There is no symmetry: on top of that, you are dividing the estimator (rather than adding/subtracting to the estimator).

Differences from the other confidence intervals

Assume we are interested in finding a $(1-\alpha)$-confidence interval. Then, be careful with the following.

1. There are no squares involved! You do not "square" the value: this is simply the name of the distribution!
2. Notice that the critical values are not symmetric.
   - For the lower bound, we use $\chi^2_{\alpha/2,n-1}$;
   - for the upper bound, $\chi^2_{1-\alpha/2,n-1}$.
3. There is no symmetry: on top of that, you are dividing the estimator (rather than adding/subtracting to the estimator).

ILLINOIS

Population proportions: motivation

Assume we are deciding for a new law, and want to make sure that the population of a city (estimated at 100,000) supports it. Moreover, assume that support means 50% or more people like the law.

What can we do?

- Ask a random set of n people whether they support the law.
- Count how many support the law. Let them be X.
- Estimate $\hat{p} = \frac{X}{n}$.

Suppose $\hat{p} = 0.6$ after asking $n = 30$ people.

Should we enact the law? *Are we 95% sure the majority supports it?*

Population proportions: motivation

Assume we are deciding for a new law, and want to make sure that the population of a city (estimated at 100,000) supports it. Moreover, assume that support means 50% or more people like the law.

What can we do?

- Ask a random set of n people whether they support the law.
- Count how many support the law. Let them be X.
- Estimate $\hat{p} = \frac{X}{n}$.

Suppose $\hat{p} = 0.6$ after asking $n = 30$ people.

Should we enact the law? *Are we 95% sure the majority supports it?*

ILLINOIS

Population proportions: motivation

Assume we are deciding for a new law, and want to make sure that the population of a city (estimated at 100,000) supports it. Moreover, assume that support means 50% or more people like the law.

What can we do?

- Ask a random set of n people whether they support the law.
- Count how many support the law. Let them be X.
- Estimate $\hat{p} = \frac{X}{n}$.

Suppose $\hat{p} = 0.6$ after asking $n = 30$ people.

Should we enact the law? *Are we 95% sure the majority supports it?*

Population proportions: motivation

Assume we are deciding for a new law, and want to make sure that the population of a city (estimated at 100,000) supports it. Moreover, assume that support means 50% or more people like the law.

What can we do?

- Ask a random set of n people whether they support the law.
- Count how many support the law. Let them be X.
- Estimate $\hat{p} = \frac{X}{n}$.

Suppose $\hat{p} = 0.6$ after asking $n = 30$ people.

Should we enact the law? *Are we 95% sure the majority supports it?*

Confidence intervals for population proportions

- $X \sim \text{binomial}(n, p)$.
- When n is big enough, then X is approximated by a normal distribution with mean np and variance $np(1-p)$.

Definition

Assume that X is binomially distributed with parameters n, p. Further assume that $np > 5$ and $n(1-p) > 5$. Then, X can be written as a normally distributed random variable $\mathcal{N}(np, np(1-p))$.

- Finally, $Z = \frac{X-np}{\sqrt{np(1-p)}}$ is follows the standard normal distribution.
- Note how we can rewrite Z as follows:

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1).$$

Confidence intervals for population proportions

- $X \sim$ binomial(n, p).
- When n is big enough, then X is approximated by a normal distribution with mean np and variance $np(1-p)$.

Definition

Assume that X is binomially distributed with parameters n, p. Further assume that $np > 5$ and $n(1-p) > 5$. Then, X can be written as a normally distributed random variable $\mathcal{N}(np, np(1-p))$.

- Finally, $Z = \frac{X-np}{\sqrt{np(1-p)}}$ is follows the standard normal distribution.
- Note how we can rewrite Z as follows:

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1).$$

ILLINOIS

Confidence intervals for population proportions

- $X \sim \text{binomial}(n, p)$.
- When n is big enough, then X is approximated by a normal distribution with mean np and variance $np(1-p)$.

Definition

Assume that X is binomially distributed with parameters n, p. Further assume that $np > 5$ and $n(1-p) > 5$. Then, X can be written as a normally distributed random variable $\mathcal{N}(np, np(1-p))$.

- Finally, $Z = \frac{X-np}{\sqrt{np(1-p)}}$ is follows the standard normal distribution.
- Note how we can rewrite Z as follows:

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1).$$

ILLINOIS

Confidence intervals for population proportions

- $X \sim \text{binomial}(n, p)$.
- When n is big enough, then X is approximated by a normal distribution with mean np and variance $np(1-p)$.

Definition

Assume that X is binomially distributed with parameters n, p. Further assume that $np > 5$ and $n(1-p) > 5$. Then, X can be written as a normally distributed random variable $\mathcal{N}(np, np(1-p))$.

- Finally, $Z = \frac{X - np}{\sqrt{np(1-p)}}$ is follows the standard normal distribution.
- Note how we can rewrite Z as follows:

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1).$$

ILLINOIS

Two-sided confidence intervals

- Let $\hat{p}$ be the proportion of observations that are of interest.
- Let n be the total sample selected.
- Then:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example

We asked 30 people and 18 said they support the law. What is the 95%-confidence interval for the true proportion supporting the law in the city?

Answer:

$$0.6 - 1.96 \cdot \sqrt{\frac{0.6 \cdot 0.4}{30}} \leq p \leq 0.6 + 1.96 \cdot \sqrt{\frac{0.6 \cdot 0.4}{30}} \Longrightarrow$$

$$0.4247 \leq p \leq 0.7753.$$

ILLINOIS

Two-sided confidence intervals

- Let $\hat{p}$ be the proportion of observations that are of interest.
- Let n be the total sample selected.
- Then:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example

We asked 30 people and 18 said they support the law. What is the 95%-confidence interval for the true proportion supporting the law in the city?

Answer:

$$0.6 - 1.96 \cdot \sqrt{\frac{0.6 \cdot 0.4}{30}} \leq p \leq 0.6 + 1.96 \cdot \sqrt{\frac{0.6 \cdot 0.4}{30}} \implies$$

$$0.4247 \leq p \leq 0.7753.$$

ILLINOIS

Two-sided confidence intervals

- Let $\hat{p}$ be the proportion of observations that are of interest.
- Let n be the total sample selected.
- Then:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example

We asked 30 people and 18 said they support the law. What is the 95%-confidence interval for the true proportion supporting the law in the city?

Answer:

$$0.6 - 1.96 \cdot \sqrt{\frac{0.6 \cdot 0.4}{30}} \leq p \leq 0.6 + 1.96 \cdot \sqrt{\frac{0.6 \cdot 0.4}{30}} \implies$$

$$0.4247 \leq p \leq 0.7753.$$

ILLINOIS

Two-sided confidence intervals

- Let $\hat{p}$ be the proportion of observations that are of interest.
- Let n be the total sample selected.
- Then:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example

We asked 30 people and 18 said they support the law. What is the 95%-confidence interval for the true proportion supporting the law in the city?

Answer:

$$0.6 - 1.96 \cdot \sqrt{\frac{0.6 \cdot 0.4}{30}} \leq p \leq 0.6 + 1.96 \cdot \sqrt{\frac{0.6 \cdot 0.4}{30}} \implies$$
$$0.4247 \leq p \leq 0.7753.$$

ILLINOIS

Bounding the error

The **estimation error** for our point estimate $\hat{p}$ is

$$E = |\hat{p} - p|\,.$$

- Assume we are given a $100 \cdot (1-\alpha)\%$ confidence interval.
- Then, the error is bounded above by:

$$E \leq z_{\alpha/2}\sqrt{p(1-p)/n}.$$

- Expectedly, as n increases, the error bound goes down.
- How big should n be for the error to be at a prespecified level?

$$n \geq \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p).$$

- p is unknown – but we can show that $p(1-p) \leq 0.25$. Hence, we use that:

$$n \geq 0.25\left(\frac{z_{\alpha/2}}{E}\right)^2.$$

ILLINOIS

Bounding the error

The **estimation error** for our point estimate $\hat{p}$ is

$$E = |\hat{p} - p|.$$

- Assume we are given a $100 \cdot (1 - \alpha)\%$ confidence interval.
- Then, the error is bounded above by:

$$E \leq z_{\alpha/2}\sqrt{p(1-p)/n}.$$

- Expectedly, as n increases, the error bound goes down.
- How big should n be for the error to be at a prespecified level?

$$n \geq \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p).$$

- p is unknown – but we can show that $p(1-p) \leq 0.25$. Hence, we use that:

$$n \geq 0.25\left(\frac{z_{\alpha/2}}{E}\right)^2.$$

Bounding the error

The **estimation error** for our point estimate $\hat{p}$ is

$$E = |\hat{p} - p|.$$

- Assume we are given a $100 \cdot (1 - \alpha)\%$ confidence interval.
- Then, the error is bounded above by:

$$E \leq z_{\alpha/2}\sqrt{p(1-p)/n}.$$

- Expectedly, as n increases, the error bound goes down.
- How big should n be for the error to be at a prespecified level?

$$n \geq \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p).$$

- p is unknown – but we can show that $p(1-p) \leq 0.25$. Hence, we use that:

$$n \geq 0.25\left(\frac{z_{\alpha/2}}{E}\right)^2.$$

Bounding the error

The **estimation error** for our point estimate $\hat{p}$ is

$$E = |\hat{p} - p|.$$

- Assume we are given a $100 \cdot (1-\alpha)\%$ confidence interval.
- Then, the error is bounded above by:

$$E \leq z_{\alpha/2}\sqrt{p(1-p)/n}.$$

- Expectedly, as n increases, the error bound goes down.
- How big should n be for the error to be at a prespecified level?

$$n \geq \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p).$$

- p is unknown – but we can show that $p(1-p) \leq 0.25$. Hence, we use that:

$$n \geq 0.25\left(\frac{z_{\alpha/2}}{E}\right)^2.$$

ILLINOIS

Bounding the error

The **estimation error** for our point estimate $\hat{p}$ is

$$E = |\hat{p} - p|.$$

- Assume we are given a $100 \cdot (1 - \alpha)\%$ confidence interval.
- Then, the error is bounded above by:

$$E \leq z_{\alpha/2}\sqrt{p(1-p)/n}.$$

- Expectedly, as n increases, the error bound goes down.
- How big should n be for the error to be at a prespecified level?

$$n \geq \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p).$$

- p is unknown – but we can show that $p(1-p) \leq 0.25$. Hence, we use that:

$$n \geq 0.25\left(\frac{z_{\alpha/2}}{E}\right)^2.$$

Example

Example

In the previous example, we want to have a 95%-confidence interval with an error of at most $E = 5\%$*. How many people should we ask?*

Answer: 95%-confidence level $\implies z_{0.025} = 1.96$

Overall: $n \geq 0.25 \cdot \left(\frac{1.96}{0.05}\right)^2 = 384.16 \implies n = 385.$

Observe that the number does **not** depend on the specific population, but only on the confidence level and the prespecified error.

ILLINOIS

Example

Example

In the previous example, we want to have a 95%-confidence interval with an error of at most $E = 5\%$. How many people should we ask?

Answer: 95%-confidence level $\implies z_{0.025} = 1.96$

Overall: $n \geq 0.25 \cdot \left(\frac{1.96}{0.05}\right)^2 = 384.16 \implies n = 385.$

Observe that the number does **not** depend on the specific population, but only on the confidence level and the prespecified error.

Example

Example

In the previous example, we want to have a 95%-confidence interval with an error of at most $E = 5\%$. How many people should we ask?

Answer: 95%-confidence level $\implies z_{0.025} = 1.96$

Overall: $n \geq 0.25 \cdot \left(\frac{1.96}{0.05}\right)^2 = 384.16 \implies n = 385.$

Observe that the number does **not** depend on the specific population, but only on the confidence level and the prespecified error.

ILLINOIS

Example

Example

In the previous example, we want to have a 95%-confidence interval with an error of at most $E = 5\%$. How many people should we ask?

Answer: 95%-confidence level $\implies z_{0.025} = 1.96$
Overall: $n \geq 0.25 \cdot \left(\frac{1.96}{0.05}\right)^2 = 384.16 \implies n = 385.$

Observe that the number does **not** depend on the specific population, but only on the confidence level and the prespecified error.