Confidence intervals for two populations

Chrysafis Vogiatzis

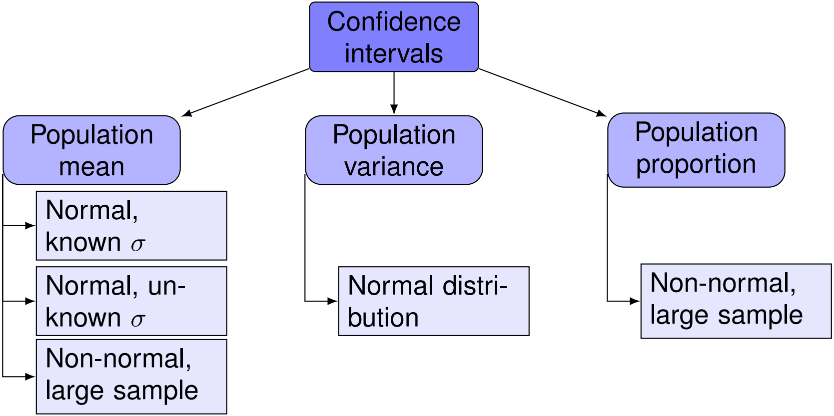
Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Lecture 23

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

Quick review

During the past lectures, we discussed a series of confidence intervals **for a single population**.



ILLINOIS

Confidence intervals for two populations

In many practical applications though, we want to be able to *compare* two populations:

- Comparing the performance of a drug in two groups of patients.
- Checking the differences in driving on ice between more and less experienced drivers.
- Seeing how people in two different areas of a city vote.
- Comparing the difference in variability in the performance of one student versus another.

In these cases, it is imperative to create confidence intervals for more than just one population.

ILLINOIS

Overview

Comparing	Want to estimate	Point estimate
Means of two populations	$\mu_1 - \mu_2$	$\overline{X}_1 - \overline{X}_2$
Variances of two populations	$\frac{\sigma_1^2}{\sigma_2^2}$	$\frac{s_1^2}{s_2^2}$
Proportions of two populations	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$

ILLINOIS

Difference in means with known variances

Let there be two *normally distributed* populations with known variances σ_1^2, σ_2^2. Also assume we have collected two samples of size n_1 from the first and size n_2 from the second population.

We know $\overline{X}_1 - \overline{X}_2$ is a good estimate for $\mu_1 - \mu_2$. But how about the interval around it?

First of all, we define a new kind of standard error, one that combines both populations. Its name is the **pooled standard deviation**:

$$\sigma_P = \sqrt{\frac{(n_1 - 1)\,\sigma_1^2 + (n_2 - 1)\,\sigma_2^2}{n_1 + n_2 - 2}}.$$

ILLINOIS

Difference in means with known variances

Let there be two *normally distributed* populations with known variances σ_1^2, σ_2^2. Also assume we have collected two samples of size n_1 from the first and size n_2 from the second population.

We know $\overline{X}_1 - \overline{X}_2$ is a good estimate for $\mu_1 - \mu_2$. But how about the interval around it?

First of all, we define a new kind of standard error, one that combines both populations. Its name is the **pooled standard deviation**:

$$\sigma_P = \sqrt{\frac{(n_1 - 1)\,\sigma_1^2 + (n_2 - 1)\,\sigma_2^2}{n_1 + n_2 - 2}}.$$

Difference in means with known variances

Let there be two *normally distributed* populations with known variances σ_1^2, σ_2^2. Also assume we have collected two samples of size n_1 from the first and size n_2 from the second population.

We know $\overline{X}_1 - \overline{X}_2$ is a good estimate for $\mu_1 - \mu_2$. But how about the interval around it?

First of all, we define a new kind of standard error, one that combines both populations. Its name is the **pooled standard deviation**:

$$\sigma_P = \sqrt{\frac{(n_1 - 1)\,\sigma_1^2 + (n_2 - 1)\,\sigma_2^2}{n_1 + n_2 - 2}}.$$

Difference in means with known variances

This leads to a confidence interval for $\mu_1 - \mu_2$ equal to:

$$\boxed{\overline{X}_1 - \overline{X}_2 - z_{\alpha/2}\sigma_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \overline{X}_1 - \overline{X}_2 + z_{\alpha/2}\sigma_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Example

Two populations have BMI with known variance equal to 9. We picked 1545 people from the first population and 1781 from the second population and found averages of 28.8 and 27.6, resp. Build a 99% confidence interval for the difference of the means between them.

Answer: $n_1 = 1545$, $n_2 = 1781$.

- We calculate: $\overline{X}_1 - \overline{X}_2 = 1.2$.
- We also find: $\sigma_P = \sqrt{\frac{1544 \cdot 9 + 1780 \cdot 9}{3324}} = 3$.
- Recall that $z_{\alpha/2} = z_{0.005} = 2.576$.
- CI: $1.2 \pm 2.576 \cdot 3 \cdot \sqrt{\frac{1}{1545} + \frac{1}{1781}} = [0.931, 1.469]$.

ILLINOIS

Difference in means with known variances

This leads to a confidence interval for $\mu_1 - \mu_2$ equal to:

$$\boxed{\overline{X}_1 - \overline{X}_2 - z_{\alpha/2}\sigma_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \overline{X}_1 - \overline{X}_2 + z_{\alpha/2}\sigma_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Example

Two populations have BMI with known variance equal to 9. We picked 1545 people from the first population and 1781 from the second population and found averages of 28.8 and 27.6, resp. Build a 99% confidence interval for the difference of the means between them.

Answer: $n_1 = 1545, n_2 = 1781.$

- We calculate: $\overline{X}_1 - \overline{X}_2 = 1.2.$
- We also find: $\sigma_P = \sqrt{\frac{1544 \cdot 9 + 1780 \cdot 9}{3324}} = 3.$
- Recall that $z_{\alpha/2} = z_{0.005} = 2.576.$
- CI: $1.2 \pm 2.576 \cdot 3 \cdot \sqrt{\frac{1}{1545}+\frac{1}{1781}} = [0.931, 1.469].$

Difference in means with known variances

This leads to a confidence interval for $\mu_1 - \mu_2$ equal to:

$$\boxed{\overline{X}_1 - \overline{X}_2 - z_{\alpha/2}\sigma_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \overline{X}_1 - \overline{X}_2 + z_{\alpha/2}\sigma_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Example

Two populations have BMI with known variance equal to 9. We picked 1545 people from the first population and 1781 from the second population and found averages of 28.8 and 27.6, resp. Build a 99% confidence interval for the difference of the means between them.

Answer: $n_1 = 1545, n_2 = 1781$.

- We calculate: $\overline{X}_1 - \overline{X}_2 = 1.2$.
- We also find: $\sigma_P = \sqrt{\frac{1544 \cdot 9 + 1780 \cdot 9}{3324}} = 3$.
- Recall that $z_{\alpha/2} = z_{0.005} = 2.576$.
- CI: $1.2 \pm 2.576 \cdot 3 \cdot \sqrt{\frac{1}{1545} + \frac{1}{1781}} = [0.931, 1.469]$.

ILLINOIS

Difference in means with known variances

This leads to a confidence interval for $\mu_1 - \mu_2$ equal to:

$$\boxed{\overline{X}_1 - \overline{X}_2 - z_{\alpha/2}\sigma_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \overline{X}_1 - \overline{X}_2 + z_{\alpha/2}\sigma_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Example

Two populations have BMI with known variance equal to 9. We picked 1545 people from the first population and 1781 from the second population and found averages of 28.8 and 27.6, resp. Build a 99% confidence interval for the difference of the means between them.

Answer: $n_1 = 1545, n_2 = 1781$.

- We calculate: $\overline{X}_1 - \overline{X}_2 = 1.2$.
- We also find: $\sigma_P = \sqrt{\frac{1544 \cdot 9 + 1780 \cdot 9}{3324}} = 3$.
- Recall that $z_{\alpha/2} = z_{0.005} = 2.576$.
- CI: $1.2 \pm 2.576 \cdot 3 \cdot \sqrt{\frac{1}{1545} + \frac{1}{1781}} = [0.931, 1.469]$.

Difference in means with unknown variances

If we do not know the population variances, then we can estimate them using the sample variances, s_1^2 and s_2^2.

- **Change 1**: we no longer have a z-value, but we have a t-value with $n_1 + n_2 - 2$ degrees of freedom.
- **Change 2**: as we do not know σ_1, σ_2, we estimate the pooled **sample** standard deviation as

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

All in all:

$$\overline{X}_1 - \overline{X}_2 - t_{\alpha/2, n_1+n_2-2}\, s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq$$

$$\leq \overline{X}_1 - \overline{X}_2 + t_{\alpha/2, n_1+n_2-2}\, s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

ILLINOIS

Difference in means with unknown variances

If we do not know the population variances, then we can estimate them using the sample variances, s_1^2 and s_2^2.

- **Change 1**: we no longer have a z-value, but we have a t-value with $n_1 + n_2 - 2$ degrees of freedom.
- **Change 2**: as we do not know σ_1, σ_2, we estimate the pooled **sample** standard deviation as

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

All in all:

$$\overline{X}_1 - \overline{X}_2 - t_{\alpha/2, n_1+n_2-2} s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq$$

$$\leq \overline{X}_1 - \overline{X}_2 + t_{\alpha/2, n_1+n_2-2} s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

ILLINOIS

Difference in means with unknown variances

If we do not know the population variances, then we can estimate them using the sample variances, s_1^2 and s_2^2.

- **Change 1**: we no longer have a z-value, but we have a t-value with $n_1 + n_2 - 2$ degrees of freedom.
- **Change 2**: as we do not know σ_1, σ_2, we estimate the pooled **sample** standard deviation as

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

All in all:

$$\overline{X}_1 - \overline{X}_2 - t_{\alpha/2, n_1+n_2-2} s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq$$
$$\leq \overline{X}_1 - \overline{X}_2 + t_{\alpha/2, n_1+n_2-2} s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Difference in means with unknown variances

Example

Assume that we run a smaller experiment on the BMI of two populations. We now have collected a sample of 4 from the first population with $\overline{X}_1 = 26.2$ and $s_1 = 2$, and a sample of 6 from the second population wtih $\overline{X}_2 = 28$ and $s_2 = 3.6$. Build a 99% confidence interval for $\overline{X}_1 - \overline{X}_2$.

Answer: We will need $t_{0.005,8} = 3.355$. We also have:

- $\overline{X}_1 - \overline{X}_2 = -1.8$.
- $s_P = \sqrt{\frac{3 \cdot 4 + 5 \cdot 12.96}{8}} = \sqrt{9.6} = 3.1$.
- $-1.8 \pm 3.355 \cdot 3.1 \cdot \sqrt{\frac{1}{4} + \frac{1}{6}} = [-8.513, 4.913]$.

ILLINOIS

Difference in means with unknown variances

Example

Assume that we run a smaller experiment on the BMI of two populations. We now have collected a sample of 4 from the first population with $\overline{X}_1 = 26.2$ and $s_1 = 2$, and a sample of 6 from the second population wtih $\overline{X}_2 = 28$ and $s_2 = 3.6$. Build a 99% confidence interval for $\overline{X}_1 - \overline{X}_2$.

Answer: We will need $t_{0.005,8} = 3.355$. We also have:

- $\overline{X}_1 - \overline{X}_2 = -1.8$.
- $s_P = \sqrt{\frac{3 \cdot 4 + 5 \cdot 12.96}{8}} = \sqrt{9.6} = 3.1$.
- $-1.8 \pm 3.355 \cdot 3.1 \cdot \sqrt{\frac{1}{4} + \frac{1}{6}} = [-8.513, 4.913]$.

Difference in means with unknown variances

Example

Assume that we run a smaller experiment on the BMI of two populations. We now have collected a sample of 4 from the first population with $\overline{X}_1 = 26.2$ and $s_1 = 2$, and a sample of 6 from the second population wtih $\overline{X}_2 = 28$ and $s_2 = 3.6$. Build a 99% confidence interval for $\overline{X}_1 - \overline{X}_2$.

Answer: We will need $t_{0.005,8} = 3.355$. We also have:

- $\overline{X}_1 - \overline{X}_2 = -1.8.$
- $s_P = \sqrt{\frac{3 \cdot 4 + 5 \cdot 12.96}{8}} = \sqrt{9.6} = 3.1.$
- $-1.8 \pm 3.355 \cdot 3.1 \cdot \sqrt{\frac{1}{4} + \frac{1}{6}} = [-8.513, 4.913]$.

Difference in means with unknown variances

Example

Assume that we run a smaller experiment on the BMI of two populations. We now have collected a sample of 4 from the first population with $\overline{X}_1 = 26.2$ and $s_1 = 2$, and a sample of 6 from the second population wtih $\overline{X}_2 = 28$ and $s_2 = 3.6$. Build a 99% confidence interval for $\overline{X}_1 - \overline{X}_2$.

Answer: We will need $t_{0.005,8} = 3.355$. We also have:

- $\overline{X}_1 - \overline{X}_2 = -1.8$.
- $s_P = \sqrt{\frac{3 \cdot 4 + 5 \cdot 12.96}{8}} = \sqrt{9.6} = 3.1$.
- $-1.8 \pm 3.355 \cdot 3.1 \cdot \sqrt{\frac{1}{4} + \frac{1}{6}} = [-8.513, 4.913]$.

ILLINOIS

Difference in variances

Let there be again two *normally distributed* populations with unknown variances σ_1^2, σ_2^2: we want to estimate their ratio σ_1^2/σ_2^2.

If we collect two samples (one from each population) and calculate the sample standard deviations, we may use s_1^2/s_2^2 to estimate their ratio! Again, though, we need more to estimate the interval around this ratio.

We will first need to introduce a new distribution, called the F distribution.

ILLINOIS

Difference in variances

Let there be again two *normally distributed* populations with unknown variances σ_1^2, σ_2^2: we want to estimate their ratio σ_1^2/σ_2^2.

If we collect two samples (one from each population) and calculate the sample standard deviations, we may use s_1^2/s_2^2 to estimate their ratio! Again, though, we need more to estimate the interval around this ratio.

We will first need to introduce a new distribution, called the F distribution.

Difference in variances

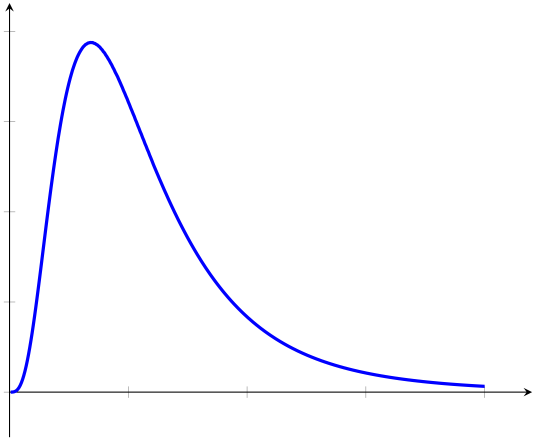
Let there be again two *normally distributed* populations with unknown variances σ_1^2, σ_2^2: we want to estimate their ratio σ_1^2/σ_2^2.

If we collect two samples (one from each population) and calculate the sample standard deviations, we may use s_1^2/s_2^2 to estimate their ratio! Again, though, we need more to estimate the interval around this ratio.

We will first need to introduce a new distribution, called the F distribution.

ILLINOIS

The F distribution



- the distribution of the ration of two χ^2 random variables.
- also **asymmetric**.
- has a table with two degrees of freedom: one for the numerator and one for the denominator!

ILLINOIS

F-table

PERCENTAGE POINTS OF THE F DISTRIBUTION

$\nu_2 \backslash \nu_1$	p	2	3	4	5	6	7	8	10	12	15	20	30	50	∞
1	0.900	49.5	53.6	55.8	57.2	58.2	59.1	59.7	60.5	61.0	61.5	62.0	62.6	63.0	63.3
	0.950	199	216	225	230	234	237	239	242	244	246	248	250	252	254
	0.975	800	864	900	922	937	948	957	969	977	985	993			
2	0.900	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.39	9.41	9.43	9.44	9.46	9.47	9.49
	0.950	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5
	0.975	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5
3	0.900	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.23	5.22	5.20	5.18	5.17	5.15	5.13
	0.950	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.74	8.70	8.66	8.62	8.58	8.53
	0.975	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.4	14.3	14.3	14.2	14.1	14.0	13.9
4	0.900	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.92	3.90	3.87	3.84	3.82	3.79	3.76
	0.950	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.91	5.86	5.80	5.75	5.70	5.63
	0.975	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.84	8.75	8.66	8.56	8.46	8.38	8.26
5	0.900	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.30	3.27	3.24	3.21	3.17	3.15	3.10
	0.950	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.68	4.62	4.56	4.50	4.44	4.36
	0.975	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.62	6.52	6.43	6.33	6.23	6.14	6.02
6	0.900	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.94	2.90	2.87	2.84	2.80	2.77	2.72
	0.950	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	4.00	3.94	3.87	3.81	3.75	3.67
	0.975	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.46	5.37	5.27	5.17	5.07	4.98	4.85
7	0.900	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.70	2.67	2.63	2.59	2.56	2.52	2.47
	0.950	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.64	3.57	3.51	3.44	3.38	3.32	3.23
	0.975	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.76	4.67	4.57	4.47	4.36	4.28	4.14
8	0.900	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.54	2.50	2.46	2.42	2.38	2.35	2.29
	0.950	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.35	3.28	3.22	3.15	3.08	3.02	2.93
	0.975	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.29	4.20	4.10	4.00	3.89	3.81	3.67

F-table (cont'd)

PERCENTAGE POINTS OF THE F DISTRIBUTION

$\nu_2 \backslash \nu_1$	q	2	3	4	5	6	7	8	10	12	15	20	30	50	∞
9	0.900	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.42	2.38	2.34	2.30	2.25	2.22	2.16
	0.950	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.14	3.07	3.01	2.94	2.86	2.80	2.71
	0.975	5.71	5.08	4.72	4.48	4.32	4.20	4.10	3.96	3.87	3.77	3.67	3.56	3.47	3.33
10	0.900	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.32	2.28	2.24	2.20	2.16	2.12	2.06
	0.950	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.98	2.91	2.84	2.77	2.70	2.64	2.54
	0.975	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.72	3.62	3.52	3.42	3.31	3.22	3.08
11	0.900	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.25	2.21	2.17	2.12	2.08	2.04	1.97
	0.950	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.85	2.79	2.72	2.65	2.57	2.51	2.40
	0.975	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.53	3.43	3.33	3.23	3.12	3.03	2.88
12	0.900	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.19	2.15	2.10	2.06	2.01	1.97	1.90
	0.950	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.75	2.69	2.62	2.54	2.47	2.40	2.30
	0.975	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.37	3.28	3.18	3.07	2.96	2.87	2.72
13	0.900	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.14	2.10	2.05	2.01	1.96	1.92	1.85
	0.950	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.67	2.60	2.53	2.46	2.38	2.31	2.21
	0.975	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.25	3.15	3.05	2.95	2.84	2.74	2.60
14	0.900	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.10	2.05	2.01	1.96	1.91	1.87	1.80
	0.950	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.60	2.53	2.46	2.39	2.31	2.24	2.13
	0.975	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.15	3.05	2.95	2.84	2.73	2.64	2.49
15	0.900	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.06	2.02	1.97	1.92	1.87	1.83	1.76
	0.950	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.54	2.48	2.40	2.33	2.25	2.18	2.07
	0.975	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.06	2.96	2.86	2.76	2.64	2.55	2.40
16	0.900	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.03	1.99	1.94	1.89	1.84	1.79	1.72
	0.950	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.49	2.42	2.35	2.28	2.19	2.12	2.01
	0.975	4.69	4.08	3.73	3.50	3.34	3.22	3.12	2.99	2.89	2.79	2.68	2.57	2.47	2.32
17	0.900	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.00	1.96	1.91	1.86	1.81	1.76	1.69
	0.950	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.45	2.38	2.31	2.23	2.15	2.08	1.96
	0.975	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.92	2.82	2.72	2.62	2.50	2.41	2.25

ILLINOIS

F-table (cont'd)

PERCENTAGE POINTS OF THE F DISTRIBUTION

$\nu_2 \backslash \nu_1$	q	2	3	4	5	6	7	8	10	12	15	20	30	50	∞
18	0.900	2.62	2.42	2.29	2.20	2.13	2.08	2.04	1.98	1.93	1.89	1.84	1.78	1.74	1.66
	0.950	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.41	2.34	2.27	2.19	2.11	2.04	1.92
	0.975	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.87	2.77	2.67	2.56	2.44	2.35	2.19
19	0.900	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.96	1.91	1.86	1.81	1.76	1.71	1.63
	0.950	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.38	2.31	2.23	2.16	2.07	2.00	1.88
	0.975	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.82	2.72	2.62	2.51	2.39	2.30	2.13
20	0.900	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.94	1.89	1.84	1.79	1.74	1.69	1.61
	0.950	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.28	2.20	2.12	2.04	1.97	1.84
	0.975	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.77	2.68	2.57	2.46	2.35	2.25	2.09
25	0.900	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.87	1.82	1.77	1.72	1.66	1.61	1.52
	0.950	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.24	2.16	2.09	2.01	1.92	1.84	1.71
	0.975	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.61	2.51	2.41	2.30	2.18	2.08	1.91
30	0.900	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.82	1.77	1.72	1.67	1.61	1.55	1.46
	0.950	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.16	2.09	2.01	1.93	1.84	1.76	1.62
	0.975	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.51	2.41	2.31	2.20	2.07	1.97	1.79
60	0.900	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.71	1.66	1.60	1.54	1.48	1.41	1.29
	0.950	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.99	1.92	1.84	1.75	1.65	1.56	1.39
	0.975	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.27	2.17	2.06	1.94	1.82	1.70	1.48
80	0.900	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.44	1.38	1.24
	0.950	3.11	2.72	2.49	2.33	2.21	2.13	2.06	1.95	1.88	1.79	1.70	1.60	1.51	1.32
	0.975	3.86	3.28	2.95	2.73	2.57	2.45	2.35	2.21	2.11	2.00	1.88	1.75	1.63	1.40
100	0.900	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.66	1.61	1.56	1.49	1.42	1.35	1.21
	0.950	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.93	1.85	1.77	1.68	1.57	1.48	1.28
	0.975	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.18	2.08	1.97	1.85	1.71	1.59	1.35
∞	0.900	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.60	1.55	1.49	1.42	1.34	1.26	1.00
	0.950	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.83	1.75	1.67	1.57	1.46	1.35	1.00
	0.975	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.05	1.94	1.83	1.71	1.57	1.43	1.00

ILLINOIS

Back to the variance confidence interval

We will need one more property for the *F* distribution:

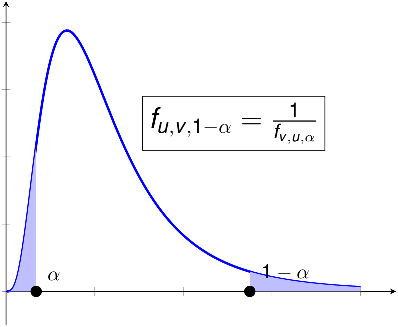
$$f_{u,v,1-\alpha} = \frac{1}{f_{v,u,\alpha}}$$

This leads to a confidence interval of:

$$f_{n_2-1,n_1-1,1-\alpha/2}\frac{s_1^2}{s_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq f_{n_2-1,n_1-1,\alpha/2}\frac{s_1^2}{s_2^2}$$

Back to the variance confidence interval

We will need one more property for the F distribution:



$$\boxed{f_{u,v,1-\alpha} = \frac{1}{f_{v,u,\alpha}}}$$

This leads to a confidence interval of:

$$\boxed{f_{n_2-1,n_1-1,1-\alpha/2}\frac{s_1^2}{s_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq f_{n_2-1,n_1-1,\alpha/2}\frac{s_1^2}{s_2^2}}$$

Difference in proportions

Finally, consider two populations with true proportions p_1, p_2: we want to bound their difference $p_1 - p_2$.

Ok... We've done something like this before:

- Collect n_1 elements from the first population and estimate $\hat{p}_1$.
- Collect n_2 elements from the second population and estimate $\hat{p}_2$.
- Report $\hat{p}_1 - \hat{p}_2$.

Finally, if we have that $n_1 p_1, n_1(1-p_1), n_2 p_2, n_2(1-p_2)$ are all ≥ 5, then $p_1 - p_2$ is normally distributed and:

$$\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \leq p_1 - p_2 \leq$$

$$\leq \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

Difference in proportions

Finally, consider two populations with true proportions p_1, p_2: we want to bound their difference $p_1 - p_2$.

Ok... We've done something like this before:

- Collect n_1 elements from the first population and estimate $\hat{p}_1$.
- Collect n_2 elements from the second population and estimate $\hat{p}_2$.
- Report $\hat{p}_1 - \hat{p}_2$.

Finally, if we have that $n_1 p_1, n_1 (1 - p_1), n_2 p_2, n_2 (1 - p_2)$ are all ≥ 5, then $p_1 - p_2$ is normally distributed and:

$$\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \leq p_1 - p_2 \leq$$

$$\leq \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

ILLINOIS

Difference in proportions

Finally, consider two populations with true proportions p_1, p_2: we want to bound their difference $p_1 - p_2$.

Ok... We've done something like this before:

- Collect n_1 elements from the first population and estimate $\hat{p}_1$.
- Collect n_2 elements from the second population and estimate $\hat{p}_2$.
- Report $\hat{p}_1 - \hat{p}_2$.

Finally, if we have that $n_1 p_1, n_1(1-p_1), n_2 p_2, n_2(1-p_2)$ are all ≥ 5, then $p_1 - p_2$ is normally distributed and:

$$\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \leq p_1 - p_2 \leq$$

$$\leq \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

ILLINOIS

Difference in proportions

Finally, consider two populations with true proportions p_1, p_2: we want to bound their difference $p_1 - p_2$.

Ok... We've done something like this before:

- Collect n_1 elements from the first population and estimate $\hat{p}_1$.
- Collect n_2 elements from the second population and estimate $\hat{p}_2$.
- Report $\hat{p}_1 - \hat{p}_2$.

Finally, if we have that $n_1 p_1, n_1 (1 - p_1), n_2 p_2, n_2 (1 - p_2)$ are all ≥ 5, then $p_1 - p_2$ is normally distributed and:

$$\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1 (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 (1 - \hat{p}_2)}{n_2}} \leq p_1 - p_2 \leq$$

$$\leq \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1 (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 (1 - \hat{p}_2)}{n_2}}.$$

ILLINOIS