University of California, Los Angeles
Department of Statistics

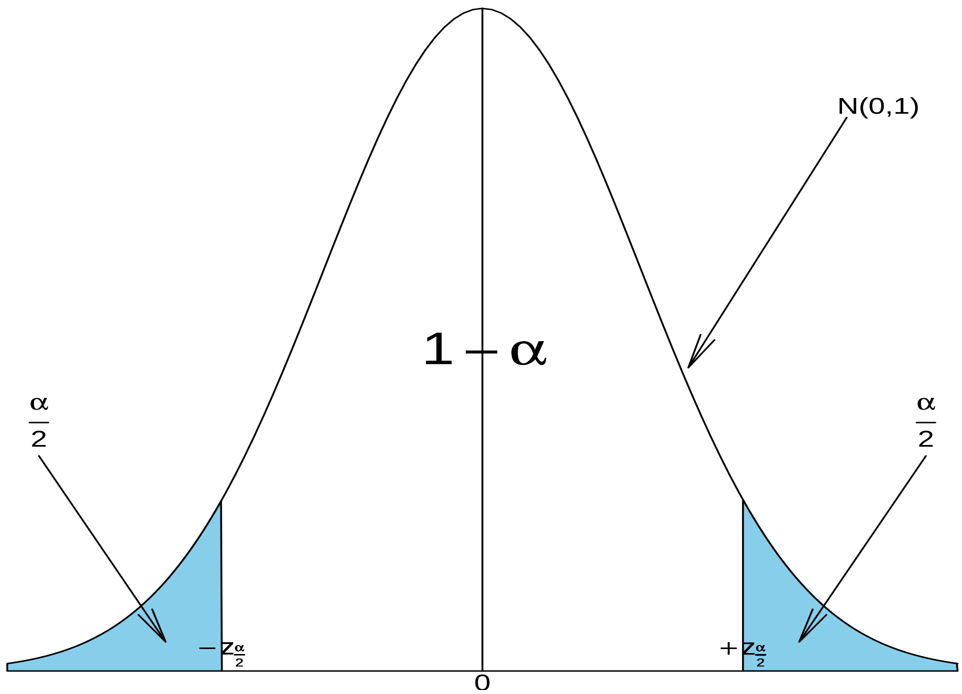
Statistics 100B **Instructor: Nicolas Christou**

Confidence intervals

A. Confidence intervals for the population mean μ of normal population with known standard deviation σ:

Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, \sigma)$. We know that $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. Therefore,

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}\right) = 1-\alpha, \quad \text{where } -z_{\frac{\alpha}{2}} \text{ and } z_{\frac{\alpha}{2}} \text{ are defined as follows:}$$



The area $1-\alpha$ is called *confidence level*. When we construct confidence intervals we usually use the following confidence levels:

$1-\alpha$	$z_{\frac{\alpha}{2}}$
0.90	1.645
0.95	1.960
0.98	2.325
0.99	2.575

The expression above can be written as:

$$P\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha. \tag{1}$$

It is tempting to read this statement as *"the probability ..."*. But we should not! Instead, we can say that we are $1-\alpha$ confident that μ falls in the interval $\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$. Why?

Example:
Suppose that the length of iron rods from a certain factory follows the normal distribution with known standard deviation $\sigma = 0.2\ m$ but unknown mean μ. Construct a 95% confidence interval for the population mean μ if a random sample of $n = 16$ of these iron rods has sample mean $\bar{x} = 6\ m$.

Sample size determination for a given length of the confidence interval:
Find the sample size n needed when we want the width of the confidence interval to be $\pm E$ with confidence level $1-\alpha$.

Answer:
In the expression $\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ the width of the confidence interval is given by $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ (also called margin of error). We want this width to be equal to E. Therefore,

$$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2.$$

Example:
For the example above, suppose that we want the entire width of the confidence interval to be equal to $0.05\ m$. Find the sample size n needed.

Question:
Is there a 100% confidence interval?

B. Confidence intervals for the population mean μ with known population standard deviation σ:

From the central limit theorem we know that when $n \geq 30$ the distribution of the sample mean $\bar{X}$ approximately follows:

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Therefore, the confidence interval for the populaton mean μ is given by the expression we found in part (A):

$$P\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) \approx 1 - \alpha.$$

The mean μ falls in the interval $\bar{x} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$.

Also the sample size determination is given by the same formula we found in part (A):

$$E = z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2.$$

Example:
A sample of size $n = 50$ is taken from the production of lightbulbs at a certain factory. The sample mean of the lifetime of these 50 lightbulbs is found to be $\bar{x} = 1570$ hours. Assume that the population standard deviation is $\sigma = 120$ hours.

a. Construct a 95% confidence interval for μ.

b. Construct a 99% confidence interval for μ.

c. What sample size is needed so that the length of the interval is 30 hours with 95% confidence?

Confidence intervals - An empirical investigation

Two dice are rolled and the sum X of the two numbers that occured is recorded. The probability distribution of X is as follows:

X	2	3	4	5	6	7	8	9	10	11	12
$P(X)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

This distribution has mean $\mu = 7$ and standard deviation $\sigma = 2.42$. We take 100 samples of size $n = 50$ each from this distribution and compute for each sample the sample mean $\bar{x}$. Pretend now that we only know that $\sigma = 2.42$, and that μ is unknown. We are going to use these 100 sample means to construct 100 confidence intervals each one with 95% confidence level for the true population mean μ. Here are the results:

Sample	$\bar{x}$	$95\% C.I. for \mu$: $\bar{x} - 1.96\frac{2.42}{\sqrt{50}} \le \mu \le \bar{x} + 1.96\frac{2.42}{\sqrt{50}}$	Is $\mu = 7$ included?
1	6.9	$6.23 \le \mu \le 7.57$	YES
2	6.3	$5.63 \le \mu \le 6.97$	NO
3	6.58	$5.91 \le \mu \le 7.25$	YES
4	6.54	$5.87 \le \mu \le 7.21$	YES
5	6.7	$6.03 \le \mu \le 7.37$	YES
6	6.58	$5.91 \le \mu \le 7.25$	YES
7	7.2	$6.53 \le \mu \le 7.87$	YES
8	7.62	$6.95 \le \mu \le 8.29$	YES
9	6.94	$6.27 \le \mu \le 7.61$	YES
10	7.36	$6.69 \le \mu \le 8.03$	YES
11	7.06	$6.39 \le \mu \le 7.73$	YES
12	7.08	$6.41 \le \mu \le 7.75$	YES
13	7.42	$6.75 \le \mu \le 8.09$	YES
14	7.42	$6.75 \le \mu \le 8.09$	YES
15	6.8	$6.13 \le \mu \le 7.47$	YES
16	6.94	$6.27 \le \mu \le 7.61$	YES
17	7.2	$6.53 \le \mu \le 7.87$	YES
18	6.7	$6.03 \le \mu \le 7.37$	YES
19	7.1	$6.43 \le \mu \le 7.77$	YES
20	7.04	$6.37 \le \mu \le 7.71$	YES
21	6.98	$6.31 \le \mu \le 7.65$	YES
22	7.18	$6.51 \le \mu \le 7.85$	YES
23	6.8	$6.13 \le \mu \le 7.47$	YES
24	6.94	$6.27 \le \mu \le 7.61$	YES
25	8.1	$7.43 \le \mu \le 8.77$	NO
26	7	$6.33 \le \mu \le 7.67$	YES
27	7.06	$6.39 \le \mu \le 7.73$	YES
28	6.82	$6.15 \le \mu \le 7.49$	YES
29	6.96	$6.29 \le \mu \le 7.63$	YES
30	7.46	$6.79 \le \mu \le 8.13$	YES
31	7.04	$6.37 \le \mu \le 7.71$	YES
32	7.06	$6.39 \le \mu \le 7.73$	YES
33	7.06	$6.39 \le \mu \le 7.73$	YES
34	6.8	$6.13 \le \mu \le 7.47$	YES
35	7.12	$6.45 \le \mu \le 7.79$	YES
36	7.18	$6.51 \le \mu \le 7.85$	YES
37	7.08	$6.41 \le \mu \le 7.75$	YES
38	7.24	$6.57 \le \mu \le 7.91$	YES
39	6.82	$6.15 \le \mu \le 7.49$	YES
40	7.26	$6.59 \le \mu \le 7.93$	YES
41	7.34	$6.67 \le \mu \le 8.01$	YES
42	6.62	$5.95 \le \mu \le 7.29$	YES
43	7.1	$6.43 \le \mu \le 7.77$	YES
44	6.98	$6.31 \le \mu \le 7.65$	YES
45	6.98	$6.31 \le \mu \le 7.65$	YES
46	7.06	$6.39 \le \mu \le 7.73$	YES
47	7.14	$6.47 \le \mu \le 7.81$	YES
48	7.5	$6.83 \le \mu \le 8.17$	YES
49	7.08	$6.41 \le \mu \le 7.75$	YES
50	7.32	$6.65 \le \mu \le 7.99$	YES

Sample	$\bar{x}$	$95\% C.I. for \mu$: $\bar{x} - 1.96\frac{2.42}{\sqrt{50}} \leq \mu \leq \bar{x} + 1.96\frac{2.42}{\sqrt{50}}$	Is $\mu = 7$ included?
51	6.54	$5.87 \leq \mu \leq 7.21$	YES
52	7.14	$6.47 \leq \mu \leq 7.81$	YES
53	6.64	$5.97 \leq \mu \leq 7.31$	YES
54	7.46	$6.79 \leq \mu \leq 8.13$	YES
55	7.34	$6.67 \leq \mu \leq 8.01$	YES
56	7.28	$6.61 \leq \mu \leq 7.95$	YES
57	6.56	$5.89 \leq \mu \leq 7.23$	YES
58	7.72	$7.05 \leq \mu \leq 8.39$	NO
59	6.66	$5.99 \leq \mu \leq 7.33$	YES
60	6.8	$6.13 \leq \mu \leq 7.47$	YES
61	7.08	$6.41 \leq \mu \leq 7.75$	YES
62	6.58	$5.91 \leq \mu \leq 7.25$	YES
63	7.3	$6.63 \leq \mu \leq 7.97$	YES
64	7.1	$6.43 \leq \mu \leq 7.77$	YES
65	6.68	$6.01 \leq \mu \leq 7.35$	YES
66	6.98	$6.31 \leq \mu \leq 7.65$	YES
67	6.94	$6.27 \leq \mu \leq 7.61$	YES
68	6.78	$6.11 \leq \mu \leq 7.45$	YES
69	7.2	$6.53 \leq \mu \leq 7.87$	YES
70	6.9	$6.23 \leq \mu \leq 7.57$	YES
71	6.42	$5.75 \leq \mu \leq 7.09$	YES
72	6.48	$5.81 \leq \mu \leq 7.15$	YES
73	7.12	$6.45 \leq \mu \leq 7.79$	YES
74	6.9	$6.23 \leq \mu \leq 7.57$	YES
75	7.24	$6.57 \leq \mu \leq 7.91$	YES
76	6.6	$5.93 \leq \mu \leq 7.27$	YES
77	7.28	$6.61 \leq \mu \leq 7.95$	YES
78	7.18	$6.51 \leq \mu \leq 7.85$	YES
79	6.76	$6.09 \leq \mu \leq 7.43$	YES
80	7.06	$6.39 \leq \mu \leq 7.73$	YES
81	7	$6.33 \leq \mu \leq 7.67$	YES
82	7.08	$6.41 \leq \mu \leq 7.75$	YES
83	7.18	$6.51 \leq \mu \leq 7.85$	YES
84	7.26	$6.59 \leq \mu \leq 7.93$	YES
85	6.88	$6.21 \leq \mu \leq 7.55$	YES
86	6.28	$5.61 \leq \mu \leq 6.95$	NO
87	7.06	$6.39 \leq \mu \leq 7.73$	YES
88	6.66	$5.99 \leq \mu \leq 7.33$	YES
89	7.18	$6.51 \leq \mu \leq 7.85$	YES
90	6.86	$6.19 \leq \mu \leq 7.53$	YES
91	6.96	$6.29 \leq \mu \leq 7.63$	YES
92	7.26	$6.59 \leq \mu \leq 7.93$	YES
93	6.68	$6.01 \leq \mu \leq 7.35$	YES
94	6.76	$6.09 \leq \mu \leq 7.43$	YES
95	7.3	$6.63 \leq \mu \leq 7.97$	YES
96	7.04	$6.37 \leq \mu \leq 7.71$	YES
97	7.34	$6.67 \leq \mu \leq 8.01$	YES
98	6.72	$6.05 \leq \mu \leq 7.39$	YES
99	6.64	$5.97 \leq \mu \leq 7.31$	YES
100	7.3	$6.63 \leq \mu \leq 7.97$	YES

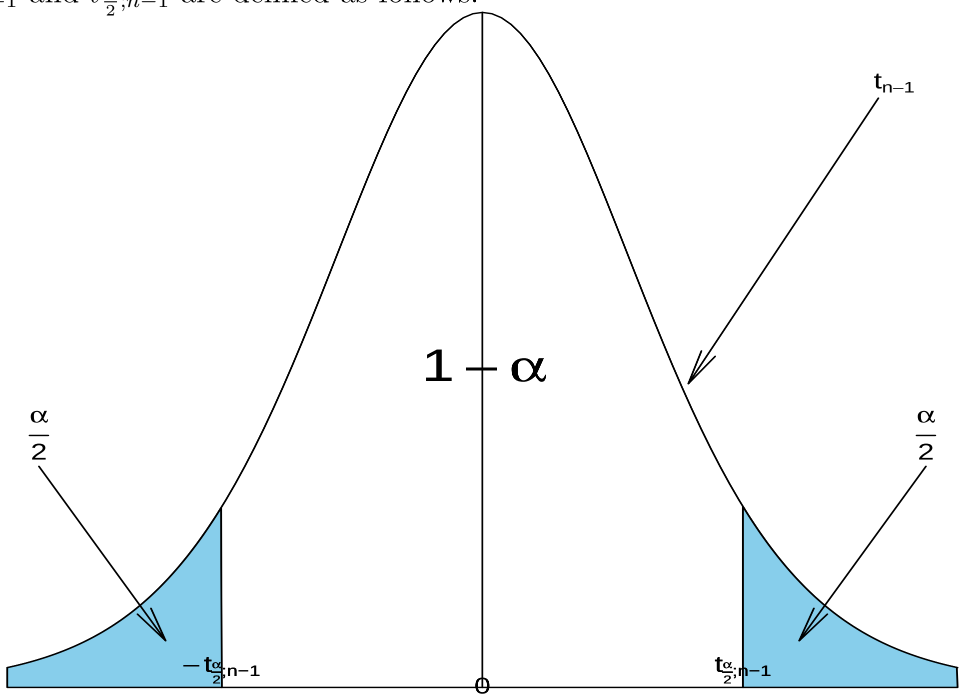
We observe that four confidence intervals among the 100 that we constructed fail to include the true population mean $\mu = 7$ (about 5%). It is also clear from this experiment why we should never use the word probability to interpret a confidence interval. Consider for example the first sample. Our confidence interval is $6.23 \leq \mu \leq 7.57$. Does it make sense to say "the probability is 95% that $\mu = 7$ falls between 6.23 and 7.57?" Of course the probability is 1 here. Look at sample 2. The resulting confidence interval is $5.63 \leq \mu \leq 6.97$. Here the probability that $\mu = 7$ included in this interval is 0. Therefore, the probability is either 0 or 1. The confidence interval either includes or not the population mean μ. We say: "we are 95% confident that μ falls in the interval we just constructed".

C. Confidence intervals for the population mean of normal distribution when the population standard deviation σ is unknown:

Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, \sigma)$. We know that $\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$. Therefore,

$$P\left(-t_{\frac{\alpha}{2};n-1} \leq \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \leq t_{\frac{\alpha}{2};n-1}\right) = 1-\alpha$$

where $-t_{\frac{\alpha}{2};n-1}$ and $t_{\frac{\alpha}{2};n-1}$ are defined as follows:



The area $1-\alpha$ is called *confidence level*. The values of $t_{\frac{\alpha}{2};n-1}$ can be found from the t table. Here are some examples:

$1-\alpha$	n	$t_{\frac{\alpha}{2};n-1}$
0.90	13	1.782
0.95	21	2.086
0.98	31	2.457
0.99	61	2.660

Note:
The sample standard deviation is computed as follows:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$$

or easier using the shortcut formula.

$$s = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right]}$$

After some rearranging the expression above can be written as:

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

We say that we are $1 - \alpha$ confident that μ falls in the interval:

$$\bar{x} \pm t_{\frac{\alpha}{2};n-1}\frac{s}{\sqrt{n}}.$$

Example:
The daily production of a chemical product last week in tons was: 785, 805, 790, 793, and 802.

a. Construct a 95% confidence interval for the population mean μ.

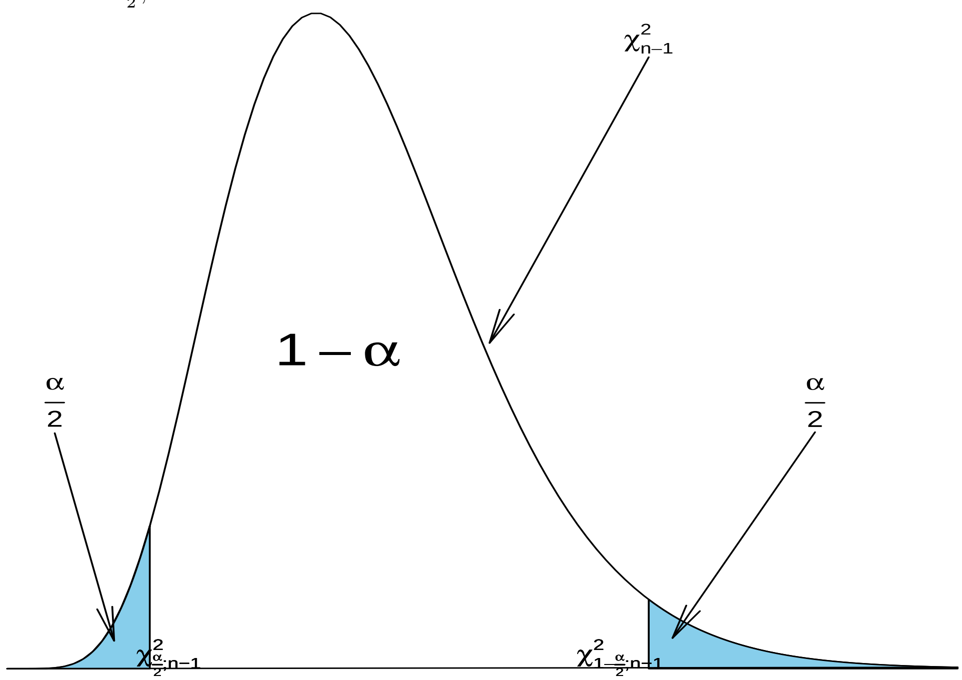
b. What assumptions are necessary?

D. Confidence interval for the population variance $\boldsymbol{\sigma^2}$ of normal distribution:

Let $X_1, X_2, \cdots, X_n$ random sample from $N(\mu, \sigma)$. We know that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. Therefore,

$$P\left(\chi^2_{\frac{\alpha}{2};n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2};n-1}\right) = 1 - \alpha$$

where $\chi^2_{\frac{\alpha}{2};n-1}$ and $\chi^2_{1-\frac{\alpha}{2};n-1}$ are defined as follows:



Some examples on how to find the values $\chi^2_{\frac{\alpha}{2};n-1}$ and $\chi^2_{1-\frac{\alpha}{2};n-1}$:

$1-\alpha$	n	$\chi^2_{\frac{\alpha}{2};n-1}$	$\chi^2_{1-\frac{\alpha}{2};n-1}$
0.90	4	0.352	7.81
0.95	16	6.26	27.49
0.98	25	10.86	42.98
0.99	41	20.71	66.77

After rearranging the inequality above we get:

$$P\left(\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2};n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2};n-1}}\right) = 1 - \alpha \tag{3}$$

We say that we are $1-\alpha$ confident that the population variance σ^2 falls in the interval:

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2};n-1}}, \quad \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2};n-1}}\right]$$

Comment: When the sample size n is large the χ^2_{n-1} distribution can be approximated by $N(n-1, \sqrt{2(n-1)})$. Therefore, in this situation, the confidence interval for the variance can be computed as follows:

$$\frac{s^2}{1 + z_{\frac{\alpha}{2}}\sqrt{\frac{2}{n-1}}} \leq \sigma^2 \leq \frac{s^2}{1 - z_{\frac{\alpha}{2}}\sqrt{\frac{2}{n-1}}}$$

Example:
A precision instrument is guaranteed to read accurately to within 2 units. A sample of 4 instrument readings on the same object yielded the measurements 353, 351, 351, and 355. Find a 90% confidence interval forn the population variance. What assumptions are necessary? Does the guarantee seem reasonable?

Note: if the population is not normal the coverage is poor (the χ^2 is not robust). In these situations (sampling from non-normal populations) an asymptotically distribution-free confidence interval for the variance can be obtained using the large sample theory result:

$$\sqrt{n}(s^2 - \sigma^2) \rightarrow N\left(0, \sqrt{\mu_4 - \sigma^4}\right)$$

or,

$$\frac{\sqrt{n}(s^2 - \sigma^2)}{\sqrt{\mu_4 - \sigma^4}} \rightarrow N(0, 1)$$

where, $\mu_4 = E(X - \mu)^4$ is the fourth moment of the distribution. Of course, μ_4 is unknown and will be estimated by the fourth sample moment $m_4 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4$. The confidence interval for the population variance is computed as follows:

$$s^2 - z_{\frac{\alpha}{2}}\frac{\sqrt{m_4 - s^4}}{\sqrt{n}} \leq \sigma^2 \leq s^2 + z_{\frac{\alpha}{2}}\frac{\sqrt{m_4 - s^4}}{\sqrt{n}}$$

E. Confidence interval for the population proportion p:

Let $Y_1, Y_2, \cdots, Y_n$ be a random sample from the Bernoulli distribution with probability of success p. Construct a confidence interval for p.

We know that when n is large:

$$\frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1), \text{ where } X = Y_1 + Y_2 + \ldots + Y_n.$$

Therefore,

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha, \quad \text{where } -z_{\frac{\alpha}{2}} \text{ and } z_{\frac{\alpha}{2}} \text{ as on page 1.}$$

After rearranging we get:

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

The ratio $\frac{x}{n}$ is the point estimate of the population p and it is denoted with $\hat{p} = \frac{x}{n}$. The problem with this interval is that the unknown p appears also at the end points of the interval. As an approximation we can simply replace p with its estimate $\hat{p} = \frac{x}{n}$. Finally the confidence interval is given below:

$$P\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha. \tag{4}$$

We say that we are $1 - \alpha$ confident that p falls in

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

Sample size determination:
Determine the sample size needed so that the resulting confidence interval will have margin of error E with confidence level $1 - \alpha$.

Answer:
In the expression $\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ the width of the confidence interval is given by the margin of error $z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. We simply solve for n:

$$E = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow n = \frac{z^2_{\frac{\alpha}{2}}\hat{p}(1-\hat{p})}{E^2}.$$

However the value of $\hat{p}$ is not known because we have not selected our sample yet. If we use $\hat{p} = 0.5$ we will obtain the largest possible sample size. Of course if we have an idea about its value (from another study, etc.) we can use it.

Example:
At a survey poll before the elections candidate A receives the support of 650 voters in a sample of 1200 voters.

a. Construct a 95% confidence interval for the population proportion p that supports candidate A.

b. Find the sample size needed so that the margin of error will be ± 0.01 with confidence level 95%.

Another formula for the confidence interval for the population proportion p:

A more accurate confidence interval can be obtained as follows:

$$
\begin{aligned}
P\left(-z_{\frac{\alpha}{2}} \le \frac{X-np}{\sqrt{np(1-p)}} \le z_{\frac{\alpha}{2}}\right) &= 1-\alpha \\
P\left(-z_{\frac{\alpha}{2}} \le \frac{\frac{X}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\frac{\alpha}{2}}\right) &= 1-\alpha \\
P\left(\frac{|\hat{p}-p|}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\frac{\alpha}{2}}\right) &= 1-\alpha \\
P\left(\frac{(\hat{p}-p)^2}{\frac{p(1-p)}{n}} \le z^2_{\frac{\alpha}{2}}\right) &= 1-\alpha
\end{aligned}
$$

We obtain a quadratic expression in p:

$$
\begin{aligned}
(\hat{p}-p)^2 - z^2_{\frac{\alpha}{2}}\frac{p(1-p)}{n} &\le 0 \\
\left(1+\frac{z^2_{\frac{\alpha}{2}}}{n}\right)p^2 - \left(2\hat{p}+\frac{z^2_{\frac{\alpha}{2}}}{n}\right)p + \hat{p}^2 &= 0
\end{aligned}
$$

Solving for p we get the following confidence interval:

$$
\frac{\hat{p}+\frac{z^2_{\frac{\alpha}{2}}}{2n} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2_{\frac{\alpha}{2}}}{4n^2}}}{1+\frac{z^2_{\frac{\alpha}{2}}}{n}}. \quad \text{When } n \text{ is large this is the same as (4).} \tag{5}
$$

Survey poll - an example:

Below we see part of a survey poll from Afghanistan. The entire survey can be accessed at:

`http://abcnews.go.com/PollingUnit/story?id=6787686&page=1`

```
Frustration With War, Problems in Daily Life Send Afghans' Support for U.S. Efforts Tumbling
ABC News/BBC/ARD National Survey of Afghanistan

ANALYSIS by GARY LANGER
Feb. 9, 2009

The United States, its NATO allies and the government of Hamid Karzai are losing not just ground in Afghanistan ´ but also the hearts and minds
of the Afghan people.

A new national public opinion poll in Afghanistan by ABC News, the BBC and ARD German TV finds that performance ratings and support levels for
the Kabul government and its Western allies have plummeted from their peaks, particularly in the past year. Widespread strife, a resurgent
Taliban, struggling development, soaring corruption and broad complaints about food, fuel, power and prices all play a role.

The effects are remarkable: With expectations for security and economic development unmet, the number of Afghans who say their country is headed
in the right direction has dived from 77 percent in 2005 to 40 percent now ´ fewer than half for the first time in these polls.

In 2005, moreover, 83 percent of Afghans expressed a favorable opinion of the United States ´ unheard of in a Muslim nation. Today just 47 percent
still hold that view, down 36 points, accelerating with an 18-point drop in U.S. favorability this year alone. For the first time slightly more
Afghans now see the United States unfavorably than favorably.

The number who say the United States has performed well in Afghanistan has been more than halved, from 68 percent in 2005 to 32 percent now.
Ratings of NATO/ISAF forces are no better. Just 37 percent of Afghans now say most people in their area support Western forces; it was 67 percent
in 2006. And 25 percent now say attacks on U.S. or NATO/ISAF forces can be justified, double the level, 13 percent, in 2006.

Nor does the election of Barack Obama hold much promise in the eyes of the Afghan public: While two in 10 think he'll make things better for their
country, nearly as many think he'll make things worse. The rest either expect no change, or are waiting to see.

This survey is ABC's fourth in Afghanistan since 2005, part of its ongoing "Where Things Stand" series there and in Iraq. It was conducted in late
December and early January via face-to-face interviews with a random national sample of 1,534 Afghan adults in all 34 of the country's provinces,
with field work by the Afghan Center for Socio-Economic and Opinion Research in Kabul.

The survey comes at a critical time for the conflict in Afghanistan, as the United States begins nearly to double its deployment of troops there,
adding as many as 30,000 to the 32,000 already present, and, under the new Obama administration, to rethink its troubled strategy. (Said Vice
President Joe Biden: "We've inherited a real mess.")

While Afghans likely will welcome a new strategy, they're far cooler on new troops: Contrary to Washington's plans, just 18 percent say the number
of U.S. and NATO/ISAF forces in Afghanistan should be increased. Far more, 44 percent, want the opposite ´ a decrease in the level of these forces.
(ISAF stands for International Security Assistance Force, the U.N.-mandated, NATO-led multinational force in Afghanistan.)

SECURITY ´ The failures to date to hold ground and provide effective security are powerful factors in Afghan public opinion. Far fewer than
in past years say Western forces have a strong presence in their area (34 percent, down from 57 percent in 2006), or ´ crucially ´ see them as
effective in providing security (42 percent, down from 67 percent).

Amid widespread experience of warfare ´ gun battles, bombings and air strikes among them ´ the number of Afghans who rate their own security
positively has dropped from 72 percent in 2005 to 55 percent today ´ and it goes far lower in high-conflict provinces. In the country's
beleaguered Southwest (Helmand, Kandahar, Nimroz, Uruzgan and Zabul provinces) only 26 percent feel secure from crime and violence; in Helmand
alone, just 14 percent feel safe.

Civilian casualties in U.S. or NATO/ISAF air strikes are a key complaint. Seventy-seven percent of Afghans call such strikes unacceptable, saying
the risk to civilians outweighs the value of these raids in fighting insurgents. And Western forces take more of the blame for such casualties, a
public relations advantage for anti-government forces: Forty-one percent of Afghans chiefly blame U.S. or NATO/ISAF forces for poor targeting, vs.
28 percent who mainly blame the insurgents for concealing themselves among civilians.

Given that view, more Afghans now blame the country's strife on the United States and its allies than on the Taliban. Thirty-six percent mostly
blame U.S., Afghan or NATO forces or the U.S. or Afghan governments for the violence that's occurring, up by 10 points from 2007. Fewer, 27
percent, now mainly blame the Taliban, down by 9 points.

Afghanistan's central and provincial governments have a stronger presence and greater public confidence than Western forces ´ but they, too, have
suffered. In 2005, still celebrating the Taliban's ouster in November 2001, 83 percent of Afghans approved of the work of President Karzai and 80
percent approved of the national government overall. Today those have slid to 52 and 49 percent respectively. (Karzai's expected to run for
re-election in August.) And fewer than half rate their provincial government positively.

IMPACT ´ Crucially, the Kabul government and its Western allies do better where they are seen as having a strong presence and as being effective
in providing security, as well as in areas where reported conflict is lower. Where security is weaker or these groups have less presence, their
ratings decline sharply.

For example, among people who say the central government, the provincial government or Western forces have a strong local presence, 58, 57 and
46 percent, respectively, approve of their performance. Where the presence of these entities is seen as weak, however, their respective approval
ratings drop to just 31, 22 and 25 percent...
```

Description of the methodology used for this survey poll (from ABC News Website):

```
METHODOLOGY  This ABC News/BBC/ARD poll is based on in-person interviews with a random national sample of 1,534 Afghan adults from Dec. 30, 2008
to Jan. 12, 2009. The results have a 2.5-point error margin. Field work by the Afghan Center for Socio-Economic and Opinion Research in Kabul, a
subsidiary of D3 Systems Inc. of Vienna, Va.
```

Other confidence intervals

Confidence interval for the difference between two population means $\boldsymbol{\mu_1 - \mu_2}$ when $\boldsymbol{\sigma_1^2, \sigma_2^2}$ are known:

$$\bar{x}_1 - \bar{x}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Where $\bar{x}_1, \bar{x}_2$ are the sample means of two samples independently selected from two populations with means μ_1, μ_2 and variances σ_1^2, σ_2^2 respectively.

Confidence interval for the difference between two normal population means $\boldsymbol{\mu_1 - \mu_2}$ when $\boldsymbol{\sigma_1^2 = \sigma_2^2}$ but unknown:

$$\bar{x}_1 - \bar{x}_2 - t_{\frac{\alpha}{2};n_1+n_2-2}\sqrt{s^2(\frac{1}{n_1} + \frac{1}{n_2})} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + t_{\frac{\alpha}{2};n_1+n_2-2}\sqrt{s^2(\frac{1}{n_1} + \frac{1}{n_2})}$$

Where $s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ is the pooled sample variance (the estimate of the true but unknown common population variance σ^2). This confidence interval is based on the fact that $\frac{(n_1+n_2-2)s^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$.

Confidence interval for the difference between two population proportions $\boldsymbol{p_1 - p_2}$:

$$\frac{x_1}{n_1} - \frac{x_2}{n_2} - z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{x_1}{n_1}(1-\frac{x_1}{n_1})}{n_1} + \frac{\frac{x_2}{n_2}(1-\frac{x_2}{n_2})}{n_2}} \leq p_1 - p_2 \leq \frac{x_1}{n_1} - \frac{x_2}{n_2} + z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{x_1}{n_1}(1-\frac{x_1}{n_1})}{n_1} + \frac{\frac{x_2}{n_2}(1-\frac{x_2}{n_2})}{n_2}}$$

Where x_1 is the number of successes among n_1 trials with probability of success p_1, and x_2 is the number of successes among n_2 trials with probability of success p_2.

Confidence interval for the ratio of two normal population variances $\boldsymbol{\frac{\sigma_1^2}{\sigma_2^2}}$:

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

Where s_1^2, s_2^2 are the sample variances based on two independent samples of size n_1, n_2 selected from two normal populations $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$.

Confidence intervals - Examples

Example 1
A sample of size $n = 50$ is taken from the production of lightbulbs at a certain factory. The sample mean is found to be $\bar{x} = 1570$ hours. Assume that the population standard deviation is $\sigma = 120$ hours.

a. Construct a 95% confidence interval for μ.

b. Construct a 99% confidence interval for μ.

c. What sample size is needed so that the length of the interval is 30 hours with 95% confidence?

Example 2
The *UCLA* housing office wants to estimate the mean monthly rent for studios around the campus. A random sample of size $n = 36$ studios is taken from the area around *UCLA*. The sample mean is found to be $\overline{x} = \$900$. Assume that the population standard deviation is $\sigma = \$150$.

a. Construct a 95% confidence interval for the mean monthly rent of studios in the area around *UCLA*.

b. Construct a 99% confidence interval for the mean monthly rent of studios in the area around *UCLA*.

c. What sample size is needed so that the length of the interval is \$60 with 95% confidence?

Example 3
We want to estimate the population proportion of students that are Democrats at UCLA.
A sample of size $n =$ is selected.
There are Democrats in the sample.

a. Construct a 95% confidencre interval for the population proportion p of students that are Democrats at UCLA. What do you observe?

b. What is the sample size needed in order to obtain a $\pm 2\%$ margin of error?

Example 4
A precision instrument is guaranteed to read accurately to within 2 units. A sample of 4 instrument readings on the same object yielded the measurements 353, 351, 351, and 355. Find a 90% confidence interval forn the population variance. What assumptions are necessary? Does the guarantee seem reasonable?

Example 5
A chemical process must produce, on the average, 800 tons of chemical per day. The daily yields for the past week are 785, 805, 790, 793, and 802 tons. Do these data provide evidence that the average production of this chemical is not 800 tons? Use 90% confidence level.

Example 6
A chemist has prepared a product designed to kill 60% of a particular type of insect. What sample size should be used if he desires to be 95% confident that he is within 0.02 of the true fraction of insects killed?

Example 7
Suppose that two independent random samples of n_1 and n_2 observations are selected from normal populations with means μ_1, μ_2 and variances σ_1^2, σ_2^2 respectively. Find a confidence interval for the variance ratio $\frac{\sigma_1^2}{\sigma_2^2}$ with confidence level $1 - \alpha$.

Example 8
The sample mean $\bar{X}$ is a good estimator of the population mean μ. It can also be used to predict a future value of X independently selected from the population. Assume that you have a sample mean $\bar{x}$ and a sample variance s^2, based on a random sample of n measurements from a normal population. Construct a prediction interval for a new observation x, say x_p. Use $1 - \alpha$ confidence level. Hint: Start with the quantity $X_p - \bar{X}$ and then use the definition of the t distribution.

Example 9
Let $10.5, 11.3, 12.8, 9.6, 5.3$ the times in seconds needed for downloading 5 files on your computer from a course website. If we assume that this sample we selected from a normal distribution, construct a 98% confidence interval for the population men μ. Also, construct a 99% confidence interval for the population variance σ^2.

Example 10
The sample mean lifetime of $n_1 = 100$ light bulbs was found to be equal $\bar{x}_1 = 1500$ hours. After new material was used in the production, another sample of size $n_2 = 100$ light bulbs was selected and gave $\bar{x}_2 = 1600$ hours. If assume that the standard deviation is $\sigma = 150$ hours in both case, construct a 95% confidence interval for the difference in the population means, $\mu_1 - \mu_2$.

Problem 11
Suppose the nicotine content of two kinds of cigarettes have standard deviations $\sigma_1 = 1.2$ and $\sigma_2 = 1.4$ milligrams. Fifty cigarettes of the first kind had a sample mean content $\bar{x}_1 = 26.1$ milligrams, while 40 cigarettes of the second kind had a sample mean content of $\bar{x}_2 = 23.8$ milligrams.

a. Construct a 95% confidence interval for the difference between the two population means $\mu_1 - \mu_2$ and make a conclusion.

b. What is the probability that the interval of part (a) contains the true difference between the two population means $\mu_1 - \mu_2$?

c. If you construct many intervals (as in part (a)), how many of them will contain the true difference $\mu_1 - \mu_2$ and how many will not?

Problem 12
One of the Nation's largest television networks wants to estimate the true proportion of American citizens that favor a certain political issue. In a random sample of 1200 American citizens 640 say that favor the issue.

a. Construct a 95% confidence interval for the true population proportion p of American citizens that favor the issue.

b. What is the probability that the confidence interval you constructed in part (a) includes the true population proportion p?

c. Suppose that another television network using the same data as above constructed the following interval for the true population proportion p: $0.4998 \leq p \leq 0.5668$. What confidence level did they use?

d. What is the sample size needed if the first television network wants to construct a confidence interval for which the margin of error is 1% with 95% confidence.

Problem 13
To test the effectiveness of a new pain-relieving drug, 120 patients at a clinic are given the drug while 120 others are given the placebo. In the first group 79 of the patients felt beneficial effect while 56 of those who received the placebo felt beneficial effect. Is there any difference in the effectiveness of the drug and the placebo? Use a 95% confidence level?

Problem 14
A study was conducted to study the effect of an oral anti-plaque rinse on plaque buildup on teeth. For this study, 14 subjects were divided into 2 groups of 7 subjects each. Both groups were assigned to use oral rinses for a 2-week period. Group 1 used rinse that contained an antiplaque agent, while group 2 received a similar rinse except that it contained no antilpaque agent. A plaque index that measures the plaque buildup was recorded after 2 weeks. The sample mean and sample standard deviation for the 2 groups are shown beklow:

	Group 1	Group 2
Sample size	7	7
Sample mean	0.78	1.26
Sample standard deviatuion	0.32	0.31

a. Assume that $\sigma_1^2 = \sigma_2^2 = \sigma^2$ but unknown, and that the 2 populations follow the normal distribution. Construct a 95% confidence interval for $\mu_1 - \mu_2$.

b. What is your conclusion based on your confidence interval of part (a)?

c. Why do we need the normality assumption?

Problem 15
Consider the simple regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, with $E(\epsilon_i) = 0$, $var(\epsilon_i) = \sigma^2$, and $\epsilon_i \sim N(0, \sigma)$. Once the parameters are estimated, the fitted line equation is written as $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Answer the following questions:

a. Find $E(\hat{y}_i)$.

b. Find $var(\hat{y}_i)$.

c. What is the distribution of $\hat{y}_i$?

d. We also discussed in class that $\frac{(n-2)s_e^2}{\sigma^2} \sim \chi^2_{n-2}$. Construct $1-\alpha$ confidence interval of $E(Y_i)$. Note: $E(Y_i) = \beta_0 + \beta_1 x_i$. *Hint:* Your answer to part (c) should be very helpful!

Confidence intervals - some numerical examples

Example 1:
Suppose that the length of iron rods from a certain factory follows the normal distribution with known standard deviation $\sigma = 0.2\ m$ but unknown mean μ. Construct a 95% confidence interval for the population mean μ if a random sample of $n = 16$ of these iron rods has sample mean $\bar{x} = 6\ m$.
Answer:

$$\mu \in \bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow \quad 6 \pm 1.96 \frac{0.2}{\sqrt{16}} \quad \text{or} \quad 6 \pm 0.098$$

Therefore we are 95% confident that $5.9 \le \mu \le 6.1$.

Example 2:
For the example above, suppose that we want the entire width of the confidence interval to be equal to $0.05\ m$. Find the sample size n needed.
Answer:

$$n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2 = \left(\frac{1.96 \times 0.2}{0.025}\right)^2 = 245.9 \approx 246.$$

Example 3:
A sample of size $n = 50$ is taken from the production of lightbulbs at a certain factory. The sample mean of the lifetime of these 50 lightbulbs is found to be $\bar{x} = 1570$ hours. Assume that the population standard deviation is $\sigma = 120$ hours.

a. Construct a 95% confidence interval for μ.
   Answer:

$$1570 \pm 1.96 \frac{120}{\sqrt{50}} \Rightarrow 1570 \pm 33.3 \Rightarrow 1536.7 \le \mu \le 1603.3.$$

b. Construct a 99% confidence interval for μ.
   Answer:

$$1570 \pm 2.575 \frac{120}{\sqrt{50}} \Rightarrow 1570 \pm 43.7 \Rightarrow 1526.3 \le \mu \le 1613.7.$$

c. What sample size is needed so that the length of the interval is 30 hours with 95% confidence?
   Answer:

$$n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2 = \left(\frac{1.96 \times 120}{15}\right)^2 = 245.9 \approx 246.$$

Example 4:
The daily production of a chemical product last week in tons was: 785, 805, 790, 793, and 802.

a. Construct a 95% confidence interval for the population mean μ.
   Answer: These data have $\bar{x} = 795$ and $sd(x) = 8.34$. Using the t distribution with 4 degrees of freedom we get:

$$\bar{x} \pm t_{\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}} \Rightarrow 795 \pm 2.776 \frac{8.34}{\sqrt{5}} \Rightarrow 795 \pm 10.35 \Rightarrow 784.65 \le \mu \le 805.35.$$

b. What assumptions are necessary?
   Answer: The sample was selected from normal distribution.

Example 5:
A precision instrument is guaranteed to read accurately to within 2 units. A sample of 4 instrument readings on the same object yielded the measurements 353, 351, 351, and 355. Find a 90% confidence interval for the population variance. What assumptions are necessary? Does the guarantee seem reasonable?
Answer: These data have $var(x) = 3.67$. Using the χ^2 distribution we get:

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2};n-1}}, \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2};n-1}}\right] \Rightarrow \left[\frac{(4-1)\times 3.67}{7.81}, \frac{(4-1)\times 3.67}{0.352}\right] \Rightarrow (1.41, 31.3)\,.$$

Example 6:
At a survey poll before the elections candidate A receives the support of 650 voters in a sample of 1200 voters.

a. Construct a 95% confidence interval for the population proportion p that supports candidate A.
   Answer: These data have $\hat{p} = \frac{650}{1200} = 0.54$.

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow 0.54 \pm 1.96\sqrt{\frac{0.54(1-0.54)}{1200}} \Rightarrow 0.54 \pm 0.028 \Rightarrow 0.512 \leq p \leq 0.568.$$

b. Find the sample size needed so that the margin of error will be ± 0.01 with confidence level 95%.
   Answer:

$$n = \frac{z^2_{\frac{\alpha}{2}}\hat{p}(1-\hat{p})}{E^2} = \frac{1.96^2(0.54)(1-0.54)}{0.01^2} = 9542.5 \approx 9543.$$

Example 7:
The *UCLA* housing office wants to estimate the mean monthly rent for studios around the campus. A random sample of size $n = 31$ studios is taken from the area around *UCLA*. The sample mean is found to be $\overline{x} = \$1300$ with sample standard deviation $s = \$150$. Assume that these data are selected from a normal distribution.

a. Construct a 90% confidence interval for the mean monthly rent of studios in the area around *UCLA*.
   Answer: Using the t distribution with 30 degrees of freedom

$$\bar{x} \pm t_{\frac{\alpha}{2};n-1}\frac{s}{\sqrt{n}} \Rightarrow 1300 \pm 1.697\frac{150}{\sqrt{31}} \Rightarrow 1300 \pm 45.7 \Rightarrow 1254.3 \leq \mu \leq 1345.7.$$

b. Construct a 99% confidence interval for the mean monthly rent of studios in the area around *UCLA*.
   Answer:

$$\bar{x} \pm t_{\frac{\alpha}{2};n-1}\frac{s}{\sqrt{n}} \Rightarrow 1300 \pm 2.750\frac{150}{\sqrt{31}} \Rightarrow 1300 \pm 74.1 \Rightarrow 1225.9 \leq \mu \leq 1374.1.$$

c. What sample size is needed so that the length of the interval is \$60 with 95% confidence? Assume $\sigma = \$150$.
   Answer:

$$n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2 = \left(\frac{1.96\times 150}{30}\right)^2 = 96.04 \approx 97.$$

Example 8:
A chemist has prepared a product designed to kill 60% of a particular type of insect. What sample size should be used if he desires to be 95% confident that he is within 0.02 of the true fraction of insects killed?
Answer:

$$n = \frac{z^2_{\frac{\alpha}{2}}\hat{p}(1-\hat{p})}{E^2} = \frac{1.96^2(0.60)(1-0.60)}{0.02^2} = 2304.96 \approx 2305.$$

Example 9
Let $10.5, 11.3, 12.8, 9.6, 5.3$ the times in seconds needed for downloading 5 files on your computer from a course website. If we assume that this sample we selected from a normal distribution, construct a 98% confidence interval for the population men μ. Also, construct a 98% confidence interval for the population variance σ^2.
Answer: We need $\bar{x}, s^2$:

```
x <- c(10.5, 11.3, 12.8, 9.6, 5.3)
> mean(x); var(x); sd(x)
[1] 9.9
[1] 7.995
[1] 2.827543
```

98% Confidence interval for the population mean:

$$\bar{x} \pm t_{\frac{\alpha}{2};n-1} \frac{s}{\sqrt{n}} \Rightarrow 9.9 \pm 3.747 \frac{2.83}{\sqrt{5}} \Rightarrow 9.9 \pm 4.74 \Rightarrow 5.16 \leq \mu \leq 14.64.$$

98% Confidence interval for the population variance:

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2};n-1}}, \quad \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2};n-1}}\right] \Rightarrow \left[\frac{(5-1)\times 7.995}{13.28}, \frac{(5-1)\times 7.995}{0.297}\right] \Rightarrow (2.41, 107.7)\,.$$

Example 10
The sample mean lifetime of $n_1 = 100$ light bulbs was found to be equal $\bar{x}_1 = 1500$ hours. After new material was used in the production, another sample of size $n_2 = 100$ light bulbs was selected and gave $\bar{x}_2 = 1600$ hours. If assume that the standard deviation is $\sigma = 150$ hours in both case, construct a 95% confidence interval for the difference in the population means, $\mu_1 - \mu_2$.
Answer:

$$\bar{x}_1 - \bar{x}_2 - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$1500 - 1600 \pm 1.96 \sqrt{\frac{150^2}{100} + \frac{150^2}{100}} \Rightarrow -100 \pm 41.6 \Rightarrow -141.6 \leq \mu_1 - \mu_2 \leq -58.4.$$

Confidence intervals: large sample theory for maximum likelihood estimates

Another method of constructing confidence intervals is based on the large sample theory of maximum likelihood estimates.

As the sample size n increases it can be shown that the maximum likelihood estimate $\hat{\theta}$ of a parameter θ follows approximately normal distribution with mean θ and variance equal to the lower bound of the Cramer-Rao inequality.

$$\hat{\theta} \sim N\left(\theta, \sqrt{\frac{1}{nI(\theta)}}\right), \quad \text{where} \quad \sqrt{\frac{1}{nI(\theta)}} \quad \text{is the lower bound of the Cramer-Rao inequality.}$$

Because $I(\theta)$ (Fisher's information - see previous handout) is a function of the unknown parameter θ we replace θ with its maximum likelihood estimate $\hat{\theta}$ to get $I(\hat{\theta})$.

Since,

$$Z = \frac{\hat{\theta} - \theta}{\sqrt{\frac{1}{nI(\hat{\theta})}}},$$

we can write

$$P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}) = 1 - \alpha$$

We replace Z with $Z = \frac{\hat{\theta}-\theta}{\sqrt{\frac{1}{nI(\hat{\theta})}}}$ to get

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{\hat{\theta} - \theta}{\sqrt{\frac{1}{nI(\hat{\theta})}}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

And finally,

$$P\left(\hat{\theta} - z_{\frac{\alpha}{2}}\sqrt{\frac{1}{nI(\hat{\theta})}} \leq \theta \leq \hat{\theta} + z_{\frac{\alpha}{2}}\sqrt{\frac{1}{nI(\hat{\theta})}}\right)$$

Therefore we are $1 - \alpha$ confident that θ falls in the interval

$$\hat{\theta} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{1}{nI(\hat{\theta})}}$$

Example:
Use the result above to construct a confidence interval for the Poisson parameter λ. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables from a Poisson distribution with parameter λ.

We know that the maximum likelihood estimate of λ is $\hat{\lambda} = \bar{x}$. We need to find the lower bound of the Cramer-Rao inequality:

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \Rightarrow lnf(x) = xln\lambda - \lambda - lnx!$$

Let's find the first and second derivatives w.r.t. λ.

$$\frac{\partial lnf(x)}{\partial \lambda} = \frac{x}{\lambda} - 1 \text{ and } \frac{\partial^2 lnf(x)}{\partial \lambda^2} = -\frac{x}{\lambda^2}.$$

Therefore,

$$\frac{1}{-nE\left(\frac{\partial^2 lnf(x)}{\partial \lambda^2}\right)} = \frac{1}{-nE(-\frac{X}{\lambda^2})} = \frac{\lambda^2}{n\lambda} = \frac{\lambda}{n}.$$

Therefore when n is large $\hat{\lambda}$ follows approximately

$$\hat{\lambda} \sim N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right)$$

Because λ is unknown we replace it with its mle estimate $\hat{\lambda}$:

$$\hat{\lambda} \sim N\left(\lambda, \sqrt{\frac{\hat{\lambda}}{n}}\right) \text{ or } \hat{\lambda} \sim N\left(\lambda, \sqrt{\frac{\bar{X}}{n}}\right)$$

Therefore, the confidence interval for λ is:

$$\bar{X} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{X}}{n}}$$

Application:
The number of pine trees at a certain forest follows the Poisson distribution with unknown parameter λ per acre. A random sample of size $n = 50$ acres is selected and the number of pine trees in each acre is counted. Here are the results:

```
7 4 5 3 1 5 7 6 4 3 2 6 6 9 2 3 3 7 2 5 5 4 4 8 8 7 2 6 3 5 0
5 8 9 3 4 5 4 6 1 0 5 4 6 3 6 9 5 7 6
```

The sample mean is $\bar{x} = 4.76$. Therefore a 95% confidence interval for the parameter λ is

$$4.76 \pm 1.96\sqrt{\frac{4.76}{50}} \text{ or } 4.76 \pm 0.31$$

Therefore $4.15 \leq \lambda \leq 5.34$.