Recap C.I

Constructing C.I:

① Determine parameter

* Mean
* Variance/St.Dev.
* proportion
* Difference

② If Mean

a) known σ

b) Unknown σ

③ Determine C.I level

④ Based on 1–3, find margin of error and setup C.I.

TABLE 8.1 **The Roadmap for Constructing Confidence Intervals and Performing Hypothesis Tests, One-Sample Case**

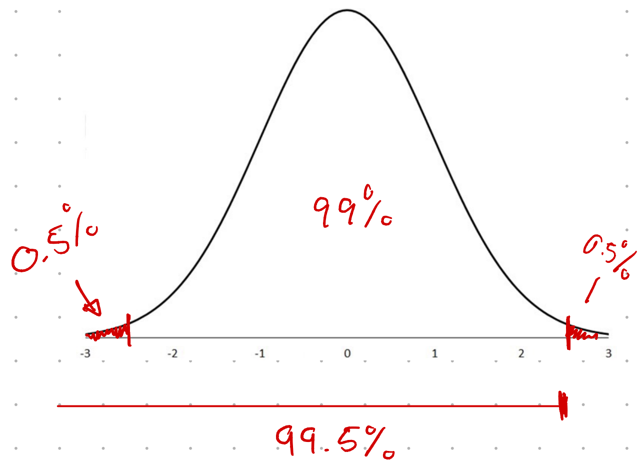
Parameter to Be Bounded by the Confidence Interval or Tested with a Hypothesis?	Symbol	Other Parameters?	Confidence Interval Section	Hypothesis Test Section	Comments
Mean of normal distribution	μ	Standard deviation σ known	8.1	9.2	Large sample size is often taken to be $n \geq 40$
Mean of arbitrary distribution with large sample size	μ	Sample size large enough that central limit theorem applies and σ is essentially known	8.1.5	9.2.3	
Mean of normal distribution	μ	Standard deviation σ unknown and estimated	8.2	9.3	
Variance (or standard deviation) of normal distribution	σ^2	Mean μ unknown and estimated	8.3	9.4	
Population proportion	p	None	8.4	9.5	

Exercise 8.1.7

8.1.7 **WP** A manufacturer produces piston rings for an automobile engine. It is known that ring diameter is normally distributed with $\sigma = 0.001$ millimeters. A random sample of 15 rings has a mean diameter of $\bar{x} = 74.036$ millimeters.

a. Construct a 99% two-sided confidence interval on the mean piston ring diameter.

b. Construct a 99% lower-confidence bound on the mean piston ring diameter. Compare the lower bound of this confidence interval with the one in part (a).



$n = 15$, $\sigma = 0.001$, $\bar{x} = 74.036$

a) $z_{1-\alpha/2} = 2.58$

$74.036 \pm 2.58 \cdot \frac{0.001}{\sqrt{15}}$

$[74.035 ; 74.037]$

b) $74.035 - 2.33 \cdot \frac{0.001}{\sqrt{15}}$

$74.0353 \leq \mu$

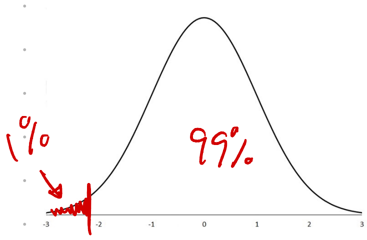
Text book example:

Confidence Interval on the Mean, Variance Known

If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval on μ is given by

$$\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{x} + z_{\alpha/2}\sigma/\sqrt{n} \quad (8.5)$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.



One-Sided Confidence Bounds on the Mean, Variance Known

A $100(1-\alpha)\%$ **upper-confidence bound** for μ is

$$\mu \leq \bar{x} + z_\alpha \sigma/\sqrt{n}$$

and a $100(1-\alpha)\%$ **lower-confidence bound** for μ is

$$\bar{x} - z_\alpha \sigma/\sqrt{n} \leq \mu$$

Sample Size for Specified Error on the Mean, Variance Known

If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error $|\bar{x}-\mu|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 \tag{8.6}$$

Large-Sample Confidence Interval on the Mean

When n is large, the quantity

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

has an approximate standard normal distribution. Consequently,

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}} \tag{8.11}$$

sample St.Dev.
unknown σ.

is a large-sample confidence interval for μ, with confidence level of approximately $100(1-\alpha)\%$.

***t* Distribution**

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. The random variable

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \tag{8.13}$$

has a t distribution with $n-1$ degrees of freedom.

I always use t-dist for all cases where σ is unknown; also if $n \geq 40$

as $n \to \infty$ $t \to z$

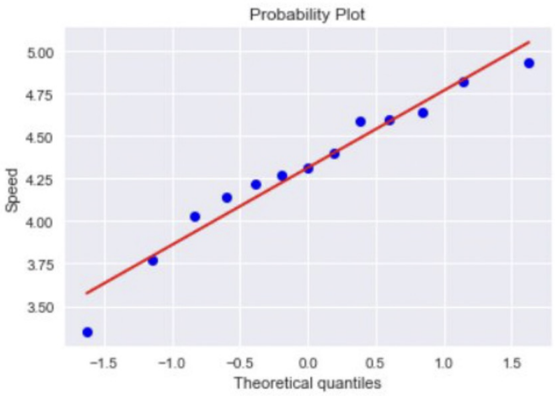
Exercise 8.2.10

When n is small, we must check whether data is normally distributed:

```
# The damn data is not saved as an excel file

a = [3.7753, 3.350679, 4.217981, 4.030324, 4.639692, 4.139665, 4.395575, 4.824257, 4.268119, 4.584193,
    4.930027, 4.315973, 4.600101]

# a) We check for normality using a normal probability plot and compute skewness and kurtosis and then create a plot of the pdf
stats.probplot(a, plot=plt)
plt.ylabel('Speed')
plt.show()
print('Skewness = ' + repr(round(stats.skew(a),4)))
print('Kurtosis = ' + repr(round(stats.kurtosis(a),4)))
df = pd.DataFrame(a)
fig, ax = plt.subplots()
df.plot.kde(ax=ax, legend=False, title='Distribution');
```




```
Skewness = -0.6732
Kurtosis = 0.0515
```

8.2.10 WP SS An article in *Computers & Electrical Engineering* ["Parallel Simulation of Cellular Neural Networks" (1996, Vol. 22, pp. 61–84)] considered the speedup of cellular neural networks (CNNs) for a parallel general-purpose computing architecture based on six transputers in different areas. The data follow:

3.775302	3.350679	4.217981	4.030324	4.639692
4.139665	4.395575	4.824257	4.268119	4.584193
4.930027	4.315973	4.600101		

a. Is there evidence to support the assumption that speedup of CNN is normally distributed? Include a graphical display in your answer.
b. Construct a 95% two-sided confidence interval on the mean speedup.
c. Construct a 95% lower confidence bound on the mean speedup.

b) $\bar{x} = 4.313$

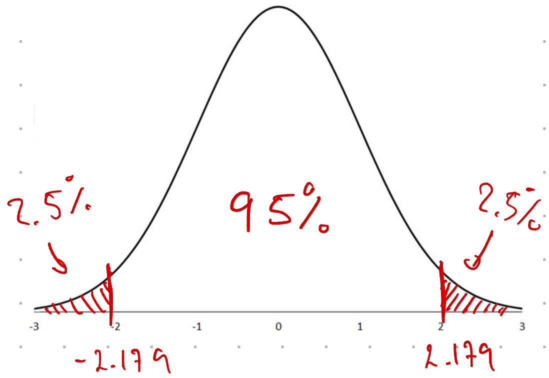
$s = 0.4328$ } Calculated from sample

$CI = 0.95$

found in table or Python

$$4.313 \pm 2.179 \cdot \frac{0.4328}{\sqrt{13}}$$

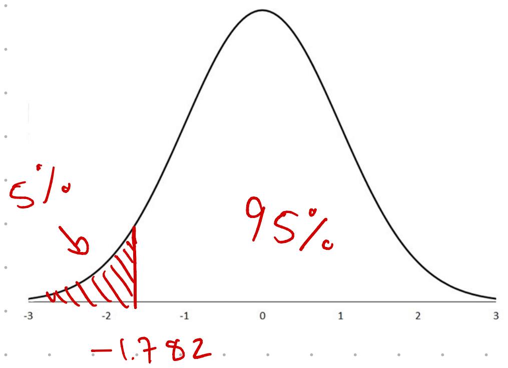
$[4.051 \; ; \; 4.574]$



c)

$$4.313 - 1.782 \cdot \frac{0.4328}{\sqrt{13}}$$

$4.099 \leq \mu$



Confidence Interval on the Mean, Variance Unknown

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, a $100(1 - \alpha)\%$ confidence interval on μ is given by

$$\bar{x} - t_{\alpha/2,n-1}s/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1}s/\sqrt{n} \tag{8.16}$$

where $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ percentage point of the t distribution with $n - 1$ degrees of freedom.

CI on Variance/St. Dev

χ^2 Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2, and let S^2 be the sample variance. Then the random variable

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \tag{8.17}$$

has a chi-square (χ^2) distribution with $n - 1$ degrees of freedom.

$$\rightarrow \sigma^2 = \frac{(n-1)S^2}{\chi^2}$$

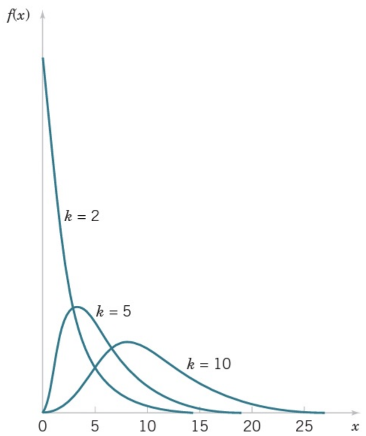


FIGURE 8.8

Probability density functions of several χ^2 distributions.

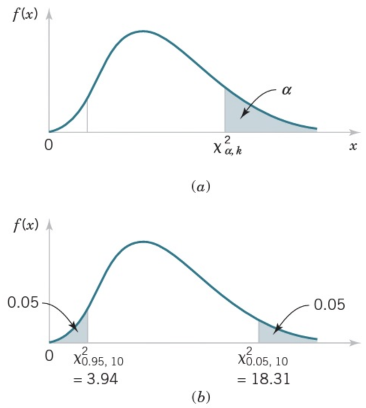


FIGURE 8.9

Percentage point of the χ^2 distribution. (a) The percentage point $\chi^2_{\alpha,k}$. (b) The upper percentage point $\chi^2_{0.05,10} = 18.31$ and the lower percentage point $\chi^2_{0.95,10} = 3.94$.

Left tail obtained from $\chi^2_{0.95,DF}$

Right tail obtained from $\chi^2_{0.05,DF}$

Example: $n = 11$, $CI = 95\%$

$\chi^2_{0.95,11} = 3.94$

$\chi^2_{0.05,11} = 18.31$

Confidence Interval on the Variance

If s^2 is the sample variance from a random sample of n observations from a normal distribution with unknown variance σ^2, then a $100(1-\alpha)\%$ confidence interval on σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \tag{8.19}$$

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the upper and lower $100\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom, respectively. A confidence interval for σ has lower and upper limits that are the square roots of the corresponding limits in Equation 8.19.

Example: $n=11$, $CI=95\%$

$\chi^2_{0.95,11} = 3.94$

$\chi^2_{0.05,11} = 18.31$

It is also possible to find a $100(1-\alpha)\%$ lower confidence bound or upper confidence bound on σ^2.

One-Sided Confidence Bounds on the Variance

The $100(1-\alpha)\%$ lower and upper confidence bounds on σ^2 are

$$\frac{(n-1)s^2}{\chi^2_{\alpha,n-1}} \le \sigma^2 \quad \text{and} \quad \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}} \tag{8.20}$$

respectively.

Exercise 8.3.5

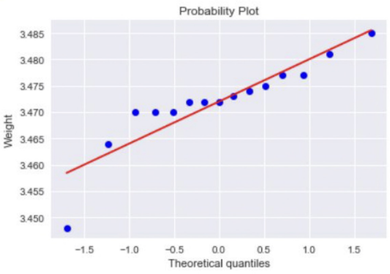
8.3.5 WP An article in *Technometrics* ["Two-Way Random Effects Analyses and Gauge R&R Studies" (1999, Vol. 41(3), pp. 202–211)] studied the capability of a gauge by measuring the weight of paper. The data for repeated measurements of one sheet of paper are in the following table. Construct a 95% one-sided upper confidence interval for the standard deviation of these measurements. Check the assumption of normality of the data and comment on the assumptions for the confidence interval.

Observations

3.481	3.448	3.485	3.475	3.472
3.477	3.472	3.464	3.472	3.470
3.470	3.470	3.477	3.473	3.474

```
a = [3.481, 3.448, 3.485, 3.475, 3.472, 3.477, 3.472, 3.464, 3.472, 3.470, 3.470, 3.470, 3.477, 3.473, 3.474]

# a) We check for normality using a normal probability plot and compute skewness and kurtosis and then create
stats.probplot(a, plot=plt)
plt.ylabel('Weight')
plt.show()
print('Skewness = ' + repr(round(stats.skew(a),4)))
print('Kurtosis = ' + repr(round(stats.kurtosis(a),4)))
df = pd.DataFrame(a)
fig, ax = plt.subplots()
df.plot.kde(ax=ax, legend=False, title='Distribution');
```




```
Skewness = -1.4381
Kurtosis = 2.986
```

$s = 0.0083$

$\chi_{95,14} = 6.57$

$$\sigma \le \sqrt{\frac{(15-1)\cdot 0.0082^2}{6.57}} = 0.012$$

CI on Proportions

Normal Approximation for a Binomial Proportion

If n is large, the distribution of

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

is approximately standard normal.

Approximate Confidence Interval on a Binomial Proportion

If $\hat{p}$ is the proportion of observations in a random sample of size n that belongs to a class of interest, an approximate $100(1-\alpha)\%$ confidence interval on the proportion p of the population that belongs to this class is

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (8.23)$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

Sample Size for a Specified Error on a Binomial Proportion

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) \quad (8.24)$$

Approximate One-Sided Confidence Bounds on a Binomial Proportion

The approximate $100(1-\alpha)\%$ lower and upper confidence bounds are

$$\hat{p} - z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \quad \text{and} \quad p \leq \hat{p} + z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (8.26)$$

respectively.

8.4.1 **WP** The 2004 presidential election exit polls from the critical state of Ohio provided the following results. The exit polls had 2020 respondents, 768 of whom were college graduates. Of the college graduates, 412 voted for George Bush.

a. Calculate a 95% confidence interval for the proportion of college graduates in Ohio who voted for George Bush.

b. Calculate a 95% lower confidence bound for the proportion of college graduates in Ohio who voted for George Bush.

a)

$$\hat{p} = \frac{412}{768} \approx 0.54$$

$$se = \sqrt{\frac{0.54(1-0.54)}{768}}$$

$$0.54 \pm 1.96 \cdot 0.018$$

$$[0.5012 ; 0.5717]$$

Sample Size for a Specified Error on a Binomial Proportion

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) \tag{8.24}$$

An estimate of p is required to use Equation 8.24. If an estimate $\hat{p}$ from a previous sample is available, it can be substituted for p in Equation 8.24, or perhaps a subjective estimate can be made. If these alternatives are unsatisfactory, a preliminary sample can be taken, $\hat{p}$ computed, and then Equation 8.24 used to determine how many additional observations are required to estimate p with the desired accuracy. Another approach to choosing n uses the fact that the sample size from Equation 8.24 will always be a maximum for $p = 0.5$ [that is, $p(1-p) \leq 0.25$ with equality for $p = 0.5$], and this can be used to obtain an upper bound on n. In other words, we are at least $100(1-\alpha)\%$ confident that the error in estimating p by $\hat{P}$ is less than E if the sample size is

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 (0.25) \tag{8.25}$$

EXAMPLE 8.8 | Crankshaft Bearings

Consider the situation in Example 8.7. How large a sample is required if we want to be 95% confident that the error in using $\hat{p}$ to estimate p is less than 0.05? Using $\hat{p} = 0.12$ as an initial estimate of p, we find from Equation 8.24 that the required sample size is

$$n = \left(\frac{z_{0.025}}{E}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.05}\right)^2 0.12(0.88) \cong 163$$

If we wanted to be *at least* 95% confident that our estimate $\hat{p}$ of the true proportion p was within 0.05 regardless of the value of p, we would use Equation 8.25 to find the sample size

$$n = \left(\frac{z_{0.025}}{E}\right)^2 (0.25) = \left(\frac{1.96}{0.05}\right)^2 (0.25) \cong 385$$

Practical Interpretation: Notice that if we have information concerning the value of p, either from a preliminary sample or from past experience, we could use a smaller sample while maintaining both the desired precision of estimation and the level of confidence.