Notes on Probability and Statistics

Contentsmkboth CONTENTSCONTENTS

I Statistics

6 Normal Random Samples 9

6.1.2 Estimators 23
6.2 Confidence intervals **27**
6.2.1 Confidence intervals for μ 30
6.2.2 Confidence intervals for σ^2 49
6.3 Hypothesis Testing **53**
6.3.1 Basic t-test about μ 55
6.3.2 Paired t-*test* 70
6.3.3 The two-sample t-*test* 74
6.3.4 Pearson's χ^2 test (goodness-of-fit test) 79

Statistics

6 Normal Random Samples 9

6.1 Theoretical building blocks

6.2 Confidence intervals

6.3 Hypothesis Testing

6. Normal Random Samples

The basic assumption in statistic is that the random variable of interest, X, is distributed across a **population** according to a known distribution, typically Normal(μ, σ^2).

Parameters μ and σ^2 are **unknown**. **Statistics** is all about estimating them using a sample, evaluating the potential errors when the sample is small, testing hypotheses and making predictions using the available data.

Random sample. A sample of size n is a realization $\mathbf{x} = (x_1, \ldots, x_n)$ of the random vector:

$$\mathbf{X} = (X_1, \ldots, X_n)$$

which we call a **random sample** in this chapter. We assume $X_1, \ldots, X_n$ to be independent and identically distributed **(iid)** Normal random variables having expected value μ and variance σ^2, denoted as:

$$\boxed{X_i \overset{iid}{\sim} \mathrm{N}(\mu, \sigma^2), \qquad i = 1, 2 \ldots n.} \tag{6.1}$$

Sample mean and sample variance

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The sample mean is the "best" estimator of μ

$$\sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

and is normally distributed.

and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

The sample variance, "best" estimator of σ^2

$$= \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} X_i^2\right) - n\bar{X}^2\right)$$

(shortcut formula)

Proof. (Shortcut formula for S^2)

$$
\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 = \sum_{i=1}^{n}\left(X_i^2+\bar{X}^2-2X_i\bar{X}\right) \\
&= \sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n}\bar{X}^2 - 2\bar{X}\sum_{i=1}^{n} X_i \\
&= \sum_{i=1}^{n} X_i^2 + n\bar{X}^2 - 2\bar{X}(n\bar{X}) \\
&= \sum_{i=1}^{n} X_i^2 + n\bar{X}^2 - 2n\bar{X}^2 \\
&= \sum_{i=1}^{n} X_i^2 - n\bar{X}^2
\end{aligned}
$$

Therefore,

$$
\boxed{S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)}
$$

STOP! **Law of large numbers:** Since $\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$ then

$$\bar{X} \to \mu \quad \text{as} \quad n \to \infty$$

The standard error is the standard deviation of the sample mean:
$\mathrm{SE} = \sigma/\sqrt{n}$ or $s/\sqrt{n}$ if σ is unknown.

Example 1. The times between successive of vehicles arrivals at a toll booth were observed as follows:

$$\{1.2,3.0,6.3,10.1,5.2,2.4,7.1\} \text{ in sec}$$

(a) Find the sample mean
(b) Find the sample variance and the standard error

Solution: (a)

$$\begin{aligned} \bar{x} &= \frac{1.2+3.0+6.3+10.1+5.2+2.4+7.1}{7} \\ &= 5.04 \end{aligned}$$

(b)

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^{7}(x_i-\bar{x})^2}{7-1} \\ &= 9.56 \\ SE &= \frac{s}{\sqrt{n}} \\ &= 1.17 \end{aligned}$$

□

6.1 Theoretical building blocks

6.1.1 The Z,T and C^2-statistics

The standardized version of $\bar{X}$ is the Z-statistic:

The Z-statistic

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

STOP! The problem with the Z-statistic is that in practice the variance σ^2 is not known.

Solution: use S^2 instead.

Fact 6.1 — The Chi-square distribution with r degrees of freedom, χ^2_r. If $Y_1, Y_2, \ldots, Y_r$ are independent standard normal random variables, $Y_i \sim N(0,1)$, then

$$\sum_{j=1}^{r} Y_j^2 \sim \chi^2_r.$$

Additive property: If $Y_1 \sim \chi^2_r$ and $Y_2 \sim \chi^2_s$, and are independent, then

$$Y_1 + Y_2 \sim \chi^2_{r+s}$$

Chi-Sqr probability tables.

The C^2 statistic

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

Proof.

$$
\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 = \sum_{i=1}^{n}\Big((X_i-\mu)-(\bar{X}-\mu)\Big)^2 \\
&= \sum_{i=1}^{n}(X_i-\mu)^2 - n\left(\bar{X}-\mu\right)^2
\end{aligned}
$$

Dividing by σ^2,

$$
\begin{aligned}
\frac{(n-1)\,S^2}{\sigma^2} &= \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 - n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 \\
&= \underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}_{\sim\chi^2_n} - \underbrace{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2}_{\sim\chi^2_1}
\end{aligned}
$$

By the additive property of the chi-square distribution in Fact (6.1) the result follows. ■

Fact 6.2 — Student's t-distribution. If $U \sim N(0,1)$ and $V \sim \chi_r^2$ are independent, then

$$T = \frac{U}{\sqrt{V/r}} \sim t_r$$

has a t-distribution with r degrees of freedom.

t-distribution tables.

Fact 6.3 — Independence of $\bar{X}$ and S^2 for normal samples.. If $X_1, X_2, \ldots, X_n$ are independent, identically distributed random variables with normal distribution $N\left(\mu, \sigma^2\right)$, then $\bar{X}$ and S^2 are independent, and:

(i) $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

(ii) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

The *T-statistic*

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Proof Recall that:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \ , \qquad C^2 = \frac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{n-1}$$

which are independent since $\bar{X}$ and S^2 are independent. It turns out that T can be expressed as:

$$T = \frac{Z}{\sqrt{C^2/(n-1)}}$$

which corresponds to the definition of the t-distribution with $n-1$ degrees of freedom!

6.1.2 Estimators

Parameters μ and σ^2 are **unknown**, so how can we estimate them?

Suppose we have a random sample $X = X_1, X_2, \ldots, X_n$ drawn from a distribution with some parameter θ.

Estimators An *estimator* $\hat{\theta}$ of θ is a function of the observed data which (we hope) forms a useful approximation of the parameter:

$$\hat{\theta} = g(X_1, X_2, \ldots, X_n).$$

Note that $\hat{\theta}$:

1. can depend only on the observed data, and not on any unknown parameters,
2. is itself a random variable, with a distribution, mean, variance, etc.

Desirable properties of a "good" estimators

Unbiased: $\hat{\theta}$ is said to be *unbiased* if

$$E\left(\hat{\theta}\right) = \theta,$$

Consistency: $\hat{\theta}$ is said to be *consistent* if $\hat{\theta} \to \theta$ as $n \to \infty$

Efficiency (Minimum variance): $\hat{\theta}_A$ is said to be more *efficient* than $\hat{\theta}_B$ if

$$V\left(\hat{\theta}_A\right) < V\left(\hat{\theta}_B\right)$$

Even if we assume unbiasedness and consistency to be desirable, it is possible to have more than one such estimator.

Example 2. — Show that $\bar{X}$ and S^2 are unbiased

$$\begin{aligned}\mathrm{E}\left(\bar{X}\right) &= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu \\ &= \mu\end{aligned}$$

Now for S^2,

$$\begin{aligned}(n-1)S^2 &= \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \sum_{i=1}^{n}\left((X_i-\mu)-(\bar{X}-\mu)\right)^2 \\ &= \sum_{i=1}^{n}(X_i-\mu)^2 \qquad\qquad - n\left(\bar{X}-\mu\right)^2\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{E}\left[(n-1)S^2\right] &= \mathrm{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2\right] - n\,\mathrm{E}\left[\left(\bar{X}-\mu\right)^2\right] \\
&= \sum_{i=1}^{n}\mathrm{V}(X_i) - n\,\mathrm{V}\left(\bar{X}\right) \\
&= n\sigma^2 - n\frac{\sigma^2}{n} \\
&= (n-1)\sigma^2
\end{aligned}$$

Therefore,

$$\mathrm{E}\left(S^2\right) = \sigma^2$$

which means that S^2 is unbiased.

6.2 Confidence intervals

Let $\mathbf{X} = (X_1, X_2, \ldots X_n)$ represent a random sample, and $\mathbf{x}$ a realization. If $(a(\mathbf{X}), b(\mathbf{X}))$ is a random interval such that

$$P\left(a(\mathbf{X}) < \theta < b(\mathbf{X})\right) = 1 - \alpha,$$

then a realization of that interval, $(a(\mathbf{x}), b(\mathbf{x}))$ is said to be a $100\,(1-\alpha)\,\%$ confidence interval for the parameter θ.

Interpretation of confidence intervals It is not easy to get to grips with what is meant by a confidence interval. One **cannot** say:

"the parameter μ has probability $(1-\alpha)$ of lying within the calculated interval $(a(\mathbf{x}), b(\mathbf{x}))$"

because that statement has no random variables.
The only **viable interpretation** is to say that we have used a procedure which, if repeated over and over again, would give intervals containing the parameter $100\,(1-\alpha)\,\%$ of the time.

Critical value z_α. The symbol z_α is the $(1-\alpha)-$percentile of Z. Note that since the $\mathrm{N}(0,1)$ is symmetric with respect to zero, so:

$$z_{1-\alpha} = -z_\alpha$$

The most common value of α in use is 0.05, in which case:

$$z_\alpha = z_{0.05} = 1.654, \qquad z_{\alpha/2} = z_{0.025} = \mathbf{1.96}$$

6.2.1 Confidence intervals for μ

Two-sided $100(1-\alpha)\%$ confidence intervals. Since $Z \sim N(0,1)$, and the definition of critical values $z_{\alpha/2}$, it follows that

$$\boxed{P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1-\alpha.}$$

Since $Z = \dfrac{\left(\bar{X} - \mu\right)}{\sigma/\sqrt{n}}$, then

$$\begin{aligned} P\left(-z_{\alpha/2} \leq \frac{\left(\bar{X} - \mu\right)}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) &= 1-\alpha \\ \Longrightarrow \quad P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) &= 1-\alpha. \end{aligned}$$

Hence the appropriate **random interval** is

$$\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\,,\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right), \qquad \text{whose realization gives:}$$

Confidence interval for μ when σ is known The $100\,(1-\alpha)\,\%$ confidence interval is

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\,,\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

The margin of error is the half-width, h, of the confidence interval:

$$h = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

If we solve for n in the above equation we obtain:

Sample size needed for a prescribed margin of error h:

$$n = \left(\frac{\sigma \, z_{\alpha/2}}{h} \right)^2$$

Example 3. The times between successive arrivals of vehicles at a toll booth were observed as follows:

$$\{1.2, 3.0, 6.3, 10.1, 5.2, 2.4, 7.1\} \text{ in sec}$$

a) Find the sample mean and the sample standard deviation.

b) Find the 95%-confidence interval for μ when σ **is assumed to be equal to the sample standard deviation**.

c) Find the margin of error.

d) Find the sample size needed to reduce the margin of error by a factor of two.

Solution:

a) Find the sample mean and the sample standard deviation.
$\bar{x} = 5.04, s = \sqrt{9.56} = 3.1$

b) Find the confidence interval for μ when σ **is assumed to be equal to the sample standard deviation**.

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\,,\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) =$$

c) Find the margin of error.

$$h = z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} =$$

d) Find the sample size needed to reduce the margin of error by a factor of two.

$$n = \left(\frac{\sigma\, z_{\alpha/2}}{h}\right)^2 =$$

$\square$

Recall that the only **viable interpretation** of confidence intervals is to say that we have used a procedure which, if repeated over and over again, would give intervals containing the parameter $100(1-\alpha)\%$ of the time.

This is shown in the figure, we were repeated the following procedure 100 times:

1. take a sample of size $n = 7$: in Excel we generate seven random numbers from the $\mathrm{N}(\mu = 50, \sigma^2 = 10)$ distribution.
2. compute the confidence interval with the above recipe
3. plot all these intervals next to each other

→ See and modify this chart in Excel

STOP! Unfortunately we do not know σ, so what should we do?

$\rightarrow$ We **replace σ by s,** which boils down to replacing the Z-statistic with the T-statistic. Writing $t_{\alpha/2}$ for the critical values from the distribution t_{n-1}, we have

$$P\left(-t_{\alpha/2} < \frac{\left(\bar{X} - \mu\right)}{S/\sqrt{n}} < t_{\alpha/2}\right) = 1 - \alpha.$$

Re-arranging gives the random interval

$$\left(\bar{X} - t_{\alpha/2}\frac{s}{\sqrt{n}},\ \bar{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right),$$

and the $100\,(1-\alpha)\,\%$ confidence interval is a **realization** of this interval.

Confidence interval for μ when σ is unknown The $100\,(1-\alpha)\,\%$ confidence interval is

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

In some circumstances, it can make more sense to express the confidence interval in only one direction – to either the lower or upper confidence limit.

One-sided CIs for μ when σ is unknown :

1. The $100\,(1-\alpha)\,\%$ **lower** confidence interval is

$$\left(-\infty,\, \bar{x} + t_\alpha \frac{s}{\sqrt{n}}\right), \qquad \text{Note: } \bar{x} + t_\alpha \tfrac{s}{\sqrt{n}} \text{ is an } \textbf{upper} \text{ bound.}$$

2. The $100\,(1-\alpha)\,\%$ **upper** confidence interval is

$$\left(\bar{x} - t_\alpha \frac{s}{\sqrt{n}}, \infty\right), \qquad \text{Note: } \bar{x} - t_\alpha \tfrac{s}{\sqrt{n}} \text{ is a } \textbf{lower} \text{ bound.}$$

Example 4. The times between successive arrivals of vehicles at a toll booth were observed as follows:

$$\{1.2, 3.0, 6.3, 10.1, 5.2, 2.4, 7.1\} \text{ in sec}$$

a) Find the sample mean and the sample standard deviation.

b) Find the 95%-confidence interval for μ.
c) Find a margin of error.
d) Find the sample size needed to reduce the margin of error by a factor of two.

Solution:

a) Find the sample mean and the sample standard deviation.
$\bar{x} = 5.04, s = \sqrt{9.56} = 3.1$

b) Find the confidence interval for μ.

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

c) Find a margin of error.

$$h = t_{\alpha/2} \frac{s}{\sqrt{n}} =$$

d) Find the sample size needed to reduce the margin of error by a factor of two.

$$n = \left(\frac{s\, t_{\alpha/2}}{h} \right)^2 =$$

$\square$

Example 5. — * **Radioactive-carbon dating** was undertaken on 8 samples from a single early site.

Sample number	Radiocarbon age determination
C-288	2419
M-26	2485
M-195	2575
M-911	2521
M-912	2451
Y-1279	2550
Y-1280	2540

Compute the two-sided and one-sided confidence intervals of the age of the site, and calculate the sample size required to estimate the age to within ± 10 years.

Solution:
In order to estimate the age of the site, we estimate the mean of the distribution by the sample mean and write

$$\bar{x} = 2505.86$$

a) Two-sided CI:
Use a T-statistic to find a 95% confidence interval which gives a range of plausible values for the mean age. This is

$$\left(\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right),$$

and, putting in $n = 7$ and $\bar{x} = 2505.86$, $t_{0.025} = 2.447$, from a t-distribution with 6 degrees of freedom, $s = 56.44$, results in a 95% confidence interval of $(2453.5, 2558.3)$, thereby giving a range of *plausible* values for μ.

b) The $100\,(1-\alpha)\,\%$ **lower** confidence interval is

$$\left(-\infty,\ \bar{x} + \frac{t_\alpha}{\sqrt{n}}\,s\right) = (-\infty\ ;\ 2,547.31)$$

here $t_{0.05} = 1.943$.

c) The $100\,(1-\alpha)\,\%$ **upper** confidence interval is

$$(2,464.4\ ;\ \infty)$$

d) To simplify the sample size calculation, we assume a normal approximation to avoid the dependency between $t_{\alpha/2}$ and the sample size, so that we can use:

$$n \approx \left(\frac{s\, z_{\alpha/2}}{h}\right)^2 = 122$$

in this case, since $h = 10$ and $z_{\alpha/2} = 1.96$.
Note: $t_{\alpha/2} = 1.98$ for $n = 122$.

$\square$

Example 6. — * Confidence intervals for Proportion In an opinion poll prior to the 2016 US Presidential election, out of 688 constituents chosen at random 368 said they would vote for Hillary Clinton (53.5%). The newspapers typically use these data to estimate p, the probability that a constituent selected at random would vote Clinton, but they rarely give any idea of the quality of the estimate.
Calculate a 95% confidence interval for p.

Solution:
First identify the random sample. Constituents questioned are labeled 1,. . . , 688. Let

$$X_i = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ constituent says "I will vote Clinton"}, \\ 0, & \text{otherwise.} \end{cases}$$

Then $X_i \sim \text{Ber}\,(p)$, and the sample size n is 688. Recall that for a $\text{Ber}\,(p)$,

$$\mu = E\,(X_i) = p$$
$$\sigma^2 = V\,(X_i) = p(1-p)$$

We know that μ, and therefore p, can be estimated by the sample mean $\bar{x} = \dfrac{368}{688} = 0.535$, and we can use the usual confidence interval for the mean

$$\left(\bar{x} - \frac{\sigma}{\sqrt{n}}\, z_{\alpha/2},\ \bar{x}\ + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right)$$

assuming that the variance is known, $\sigma^2 \approx \bar{x}(1-\bar{x})$. This gives **(0.498, 0.572)** as a 95% confidence interval for p, with point estimate 0.535. Margin of error = (0.572 -0.498)/2=**0.037**.
For 99% confidence we would use $z_{0.005} = 2.576$ to replace 1.960, and get a wider interval **(0.486, 0.584)** that is less accurate! Margin of error = (0.584 -0.486)/2=**0.049**. □

Example 7. At a weigh station, the weighs of trailer trucks were observed before crossing a highway bridge.
(a) Suppose observations on 30 trucks yielded a sample mean of 12.5 tons. Assume that the standard deviation of truck weights is known to be 2.5 tons. Determine the two-sided 99% intervals of the mean weight of trailer trucks on the particular highway.
(b) In part (a), how many additional trucks should be observed such that the mean truck weight can be estimated to within ± 1 ton with 99% confidence

Solution: (a)

$$\begin{aligned} CI &= (\bar{x} + z_{0.005}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.995}\frac{\sigma}{\sqrt{n}}) \\ &= (12.5 - 2.58 \times \frac{2.5}{\sqrt{30}}, 12.5 + 2.58 \times \frac{2.5}{\sqrt{30}}) \\ &= (11.3, 13.7) \end{aligned}$$

(b) Sample size calculation: $n' = \left(\frac{\sigma\, z_{\alpha/2}}{h}\right)^2 = \left(\frac{2.5 \times 2.58}{1}\right)^2 = 42$

So $42 - 30 = 12$ additional observation of truck weights would be required.

$\square$

Example 8. In a traffic study, the speed of vehicles are measured by laser guns for the purpose of determining the mean vehicle speed on a particular city street. It is known that the posted speed limit is 45 kph. The following results were obtained from ten test vehicles:

$$45,\ 39,\ 50,\ 41,\ 47,\ 42,\ 44,\ 48,\ 48,\ 44 \text{ kph}$$

(a) Determine the 95% two-sided confidence interval of the mean vehicle speed.

Solution: (a)

$$
\begin{aligned}
\bar{x} &= 44.8 \\
s^2 &= \frac{1}{10-1}[\sum x_i^2 - 10 \times \bar{x}^2] \\
&= 12.178 \\
t_{0.025}(r=9) &= 2.262 \\
CI &= \bar{x} \pm t_{0.025} \times \frac{s}{\sqrt{n}} \\
&= 44.8 \pm (2.262) \times \frac{3.49}{\sqrt{10}} \\
&= (42.3\ ; 47.3)
\end{aligned}
$$

□

6.2.2 Confidence intervals for σ^2

Recall: $C^2 = \dfrac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{n-1}$. Therefore,

$$P\left(\chi^2_{1-\alpha/2} < C^2 < \chi^2_{\alpha/2}\right) = 1-\alpha$$

where $\chi^2_{\alpha/2}$ is the $1-\frac{\alpha}{2}$ quantile from the distribution χ^2_{n-1}

Proceedings similarly as we did with the mean μ, we find:

The $100\,(1-\alpha)\,\%$ **Confidence intervals for** σ^2 :

a) The **two-sided** CI:

$$\left(\frac{(n-1)\,S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2}}\right)$$

b) The **lower** CI:

$$\left(-\infty, \frac{(n-1)\,S^2}{\chi^2_{1-\alpha}}\right)$$

c) The **upper** CI:

$$\left(\frac{(n-1)\,S^2}{\chi^2_{\alpha}}, \infty\right)$$

Example 9. Suppose that $n = 9$, $\bar{x} = 51.2$, and $s = 11.7$. Find the 95% Confidence intervals for σ^2.

Solution: Here α =0.05, $\chi^2_{1-\alpha/2}$ =2.18, $\chi^2_{\alpha/2}$ =17.535, $\chi^2_{1-\alpha}$ =2.733, χ^2_{α} =15.507 , which gives

a) The **two-sided** CI:

$$\left(\frac{(n-1)\,S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2}}\right) = (62.6\ ; 503.6)$$

b) The **lower** CI:

$$\left(-\infty, \frac{(n-1)\,S^2}{\chi^2_{1-\alpha}}\right) = (-\infty\ ; 401.7)$$

c) The **upper** CI:

$$\left(\frac{(n-1)\,S^2}{\chi^2_{\alpha}}, \infty\right) = (70.8\ ; \infty)$$

□

6.3 Hypothesis Testing

Basic example Suppose the throughput of a laptop production system, in number of laptops per hour, was $\mathrm{N}(\mu_0 = 20, \sigma^2 = 100)$ before an *improvement* performed to the system. We want to see if the improvement was effective by testing the mean after improvement, μ.

The **null hypothesis** is

$$H_0 : \mu = \mu_0 \qquad \text{no difference after the improvement}$$

The **alternative hypothesis** H_1 can be either:

a) $H_1 : \mu \neq \mu_0$ before and after are different (*2-sided*), or
b) $H_1 : \mu > \mu_0$ new system is better (*1-sided*), or
c) $H_1 : \mu < \mu_0$ old system is better (*1-sided*).

Example 10. A coin with probability μ of coming up tails is tossed and we hypothesize that it is fair. Therefore:

$$H_0 : \mu = \frac{1}{2} \quad \text{vs} \quad H_1 : \mu \neq \frac{1}{2}$$

Now, if there is reason to believe that the coin is biased towards tails (we suspect that $\mu > \frac{1}{2}$) then:

$$H_0 : \mu = \frac{1}{2} \quad \text{vs} \quad H_1 : \mu > \frac{1}{2}$$

6.3.1 Basic t-test about μ

Recall that:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Suppose: $H_0 : \mu = \mu_0$. Then, **under the null hypotheses we have that**

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \quad \text{if } H_0 \text{ is true.}$$

Key idea: We take a sample to observe a realization of the random variable T:

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

If it falls "far from" its mean $\mathrm{E}(T) = 0$ then we reject H_0, else we **fail to reject** it.

The meaning of "far from" depends on the alternative hypothesis H_1, and on the **significance level,** α. Suppose we wanted to test

$$H_0 : \mu = \mu_0 \quad \text{against} \quad H_1 : \mu \neq \mu_0$$

In this case, we fail to reject H_0 if: $\quad -t_{\alpha/2} < t_0 < t_{\alpha/2} \quad$ Why? because

$$P\left(-t_{\alpha/2} < T < t_{\alpha/2}\right) = 1 - \alpha$$

means that if H_0 is true then $100(1-\alpha)\%$ of the realizations of T should fall in that range.

Summary of rejection regions:

a) if $H_1 : \mu \neq \mu_0 \Rightarrow$ **reject** H_0 if: $|t_0| > t_{\alpha/2}$

b) if $H_1 : \mu > \mu_0 \Rightarrow$ **reject** H_0 if: $t_0 > t_\alpha$

c) if $H_1 : \mu < \mu_0 \Rightarrow$ **reject** H_0 if: $t_0 < -t_\alpha$

STOP! Not rejecting the hypothesis does not mean that there is strong evidence that H_0 is true. It is recommendable to use the terminology "reject hypothesis H_0" or "not reject hypothesis H_0" but not to say "accept H_0".

Example 11. — * In a traffic study, the speed of vehicles are measured by laser guns for the purpose of determining the mean vehicle speed on a particular city street. It is known that the posted speed limit is 45 kph. The following results were obtained from ten test vehicles:

$$45\ 39\ 55\ 50\ 47\ 45\ 44\ 48\ 51\ 44$$

(a) Determine the 95% two-sided confidence interval of the mean vehicle speed.
(b) Test the hypothesis that the vehicles are speeding at a 5% level of significance.
(c) If we wish to determine the mean vehicle speed to within ± 1 kph with a 99% confidence, what should be the sample size of our observations?

Solution: (a)

$$\begin{aligned}
\bar{x} &= 46.8 \\
\sum x_i^2 &= 22,082 \\
s^2 &= \frac{1}{10-1}[\sum x_i^2 - 10 \times \bar{x}^2] = 20 \\
t_{0.025}(r=9) &= 2.262 \\
CI &= \bar{x} \pm t_{0.025} \times \frac{s}{\sqrt{n}} \\
&= 46.8 \pm (2.262) \times \frac{4.47}{\sqrt{10}} \\
&= (43.6\ ; 50)
\end{aligned}$$

(b) $H_0 : \mu = 45,\ H_1 : \mu > 45$

$$\begin{aligned}
t_{0.05}(r=9) &= 1.833 \\
t_0 &= \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{46.8-45}{4.47/\sqrt{10}} = 1.27
\end{aligned}$$

Since $1.27 < 1.833$ we fail to reject H_0: "There is no evidence to suggest that vehicles are speeding". Note, we can't say vehicles are not speeding.

(c) Sample size calculation: Here

$$\begin{aligned} h &= 1 \\ z_{0.995} &= 2.58 \end{aligned}$$

so, $n \approx \left(\frac{s\, z_{\alpha/2}}{h}\right)^2 = \left(\frac{4.47 \times 2.58}{1}\right)^2 = 133$

The sample size should be 133. $\square$

Possible error in hypothesis testing. There are two types of possible error in hypothesis testing:
Type I error: rejecting the null hypothesis when it is, in fact, true.
Type II error: not rejecting the null hypothesis when it is, in fact, false.

	H_0 not rejected	H_0 rejected
H_0 true	no error	Type I error
H_0 false	Type II error	no error

Thus,

$$\begin{aligned} P(\text{Type I error}) &= P(\text{reject } H_0 \mid H_0) = \alpha \\ P(\text{Type II error}) &= P(\text{accept } H_0 \mid H_1) = \beta. \end{aligned}$$

The p-*value*, denoted p:

1. probability of a test statistic, say T, taking a value **at least as extreme as** its observed value t_0, assuming H_0 is true,
2. or equivalently: the smallest level α such that we would reject the null-hypothesis with the observed data.

It depends on the alternative hypothesis:

- if $H_1 : \mu > \mu_0 \Rightarrow p = P(T > t_0 \mid H_0)$
- if $H_1 : \mu < \mu_0 \Rightarrow p = P(T < t_0 \mid H_0)$
- if $H_1 : \mu \neq \mu_0 \Rightarrow p = 2\min\{P(T < t_0 \mid H_0), P(T > t_0 \mid H_0)\}$

Three ways of testing hypotheses. For a null hypothesis $\mu = \mu_0$, the following are a equivalent procedures to **reject** H_0:

1. the T-test t_0 falls in the rejection region
2. the p-value $< \alpha$
3. the $(1-\alpha)100\%$ confidence interval does not contain t_0:

$$|t_0| > t \quad \Leftrightarrow \quad t_0 \notin (-t_{\alpha/2}, t_{\alpha/2}) \quad \Leftrightarrow \quad \mu_0 = \bar{x} + t_0 \frac{s}{\sqrt{n}} \notin \left(\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right).$$

Hence we can see that there is an equivalence between the test and the interval.

Example 12. — * Concrete placed on a structure was subsequently cored after 28 days, and the following results were obtained of the compressive strengths from five test specimens:

$$4042,\ 3505,\ 3402,\ 3939,\ 3472 \text{ psi}$$

(a) Determine the 90% two-sided confidence interval of the mean concrete strength.
(b) Suppose the confidence interval established in part (a) is too wide, and the engineer would like to have a confidence interval to be ±300 psi of the computed sample mean concrete strength. Generally, more specimens of concrete would be needed to keep the same confidence level. However, without additional samples, what is the confidence level associated with the specified interval based on the five measurements given above?
(c) If the required minimum compressive strength is 3500 psi, test whether the concrete satisfies these requirements by performing a one-sided hypothesis test at the 2% significance level.

Solution: (a) Sample mean:

$$\begin{aligned}\bar{x} &= \frac{4042+3505+3402+3939+3472}{5} \\ &= 3672\end{aligned}$$

Sample standard deviation:

$$\begin{aligned} s &= \sqrt{\frac{\sum_1^5 (x_i - \bar{x})}{5-1}} \\ &= 295.37 \end{aligned}$$

For a 90% two-sided confidence interval:

$$\begin{aligned} t_{0.05,4} &= -2.1318 \\ t_{0.95,4} &= 2.1318 \\ CI &= (3672 - 2.1318\frac{295.37}{\sqrt{5}}, 3672 + 2.1318\frac{295.37}{\sqrt{5}}) \\ &= (3672 - 281.6, 3672 + 281.6) \\ &= (3390.4, 3953.6) \end{aligned}$$

(b) If the half-width of the confidence interval is 300, it means:

$$\begin{aligned} t_{\alpha/2}\frac{295.37}{\sqrt{5}} &= 300 \\ t_{\alpha/2} &= 2.2711 \end{aligned}$$

Refer to T-table with 4 deg. of freedom; we have:

$$\begin{aligned} t_{0.05} &= 2.1318 \\ t_{0.025} &= 2.7764 \end{aligned}$$

We may use linear interpolation to get an approximate answer:

$$\begin{aligned} 1-\alpha/2 &= 0.95+(0.975-0.95)*\frac{2.2711-2.1318}{2.7764-2.1318} \\ &= 0.9554 \\ \Rightarrow \alpha &= 0.911 \end{aligned}$$

The required confidence level is 91.1%.
(c) If the required minimum compressive strength is 3500 psi, test whether the concrete satisfies these requirements by performing a one-sided hypothesis test at the 2% significance level.

$$\begin{aligned} H_0 &: \mu = 3500 \\ H_1 &: \mu < 3500 \end{aligned}$$

In this case, the test statistic will now be:

$$t_0 = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{3672-3500}{295.37/\sqrt{5}} = 1.302$$

With $f = 5 - 1 = 4$ d.o.f, we obtain the critical value of t at the 2% significance level to be $t_\alpha \approx -3$. Therefore the value of the test statistic is outside of the region of rejection, hence the null hypothesis cannot be rejected, and therefore we conclude that the concretes meet the minimum requirement. $\square$

Example 13. — Shoshoni bead rectangles The table below gives width-to-length ratios for 20 rectangles, analyzed as part of a study in experimental aesthetics.

Table 6.1: Shoshoni bead rectangles

Width-to-length ratios			
0.693	0.670	0.654	0.749
0.606	0.553	0.601	0.609
0.672	0.662	0.606	0.615
0.844	0.570	0.933	0.576
0.668	0.628	0.690	0.611

We want to test whether the Shoshoni instinctively made their rectangles conform to the golden ratio. That is we want to test

$$H_0 : \mu = 0.618 \quad \text{against} \quad H_1 : \mu \neq 0.618.$$

We have 20 measurements so, under the null hypothesis $\mu = 0.618$ gives

$$T = \frac{\sqrt{20}(\bar{X} - 0.618)}{S} \sim t\,(19)\,,$$

where $S^2 = \frac{1}{20}\sum_{i=1}^{20}(X_i - \bar{X})^2$. For these data, $\bar{x} = 0.660$, $s = 0.093$, and the observed value of T is

$$t_0 = \frac{\sqrt{20}(\bar{x} - 0.618)}{s} = \frac{\sqrt{20}(0.660 - 0.618)}{0.093} = 2.019$$

With $r = 20 - 1 = 19$ d.o.f, we obtain the critical value of t at the 5% significance level to be $t_{\alpha/2} \approx 2.086$. Therefore the value of the test statistic is too close to call. In fact, the p-value is very close to α :

$$\text{p}-value = 2\min\{P(T < 2.019 \mid H_0), P(T > 2.019 \mid H_0)\} = 0.058$$

This says that the case for it is very weak.

6.3.2 Paired t-*test*

Suppose that we have pairs of random variables (X_i, Y_i) and that $D_i = X_i - Y_i$, $i = 1, \ldots, n$, is a random sample from a normal distribution, i.e. $D \sim N(\mu, \sigma^2)$ with unknown parameters. We use the test statistic

$$T = \frac{\bar{D} - \mu_0}{S_D/\sqrt{n}} \sim t_{n-1}$$

under the null hypothesis $H_0 : \mu = \mu_0$. Here S_D^2 is the sample variance of the differences D_i.

Example 14. — Patients with glaucoma in one eye Here is we ask "Is there a difference in corneal thickness between the eyes?"

Table 4.1 Glaucoma in one eye

Corneal thickness		
Glaucoma	Normal	Difference
488	484	4
478	478	0
480	492	−12
426	444	−18
440	436	4
410	398	12
458	464	−6
460	476	−16

Formally we are testing the difference μ between the corneal thicknesses.

$$H_0 : \mu = 0 \quad \text{against} \quad H_1 : \mu \neq 0.$$

The mean difference is $\bar{d} = -4$ and the estimated standard deviation is $s_D = 10.744$. Under H_0 we obtain a t-statistic of

$$t_0 = \frac{\bar{d} - \mu_0}{s_D/\sqrt{n}} = \frac{-4\sqrt{8}}{10.744} = -1.053.$$

With $r = 8 - 1 = 7$ d.o.f, we obtain the critical value of t at the 5% significance level to be $t_{\alpha/2} \approx 2.365$. Therefore the value of the test statistic is outside of the region of rejection, hence the null hypothesis cannot be rejected, and therefore we cannot reject the null hypothesis of no difference in corneal thickness.

0.05
−2.365
−1.053
2.365

6.3.3 The two-sample t-*test*

Consider two random samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ which are independent, normally distributed **with the same variance**. The null hypothesis is

$$H_0 : \mu_X = \mu_Y$$

Fact 6.8 Under H_0 the test statistic T is such that

$$\boxed{T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)}} \sim t_{m+n-2},} \tag{6.2}$$

where $S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$ is known as the "pooled" variance.

Proof.
Step 1: Under H_0,

$$\bar{X} - \bar{Y} \sim N\left(0, \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right), \text{ where } \sigma^2 \text{ is the common variance.}$$

Step 2: recall that

$$\frac{(m-1)S_X^2}{\sigma^2} \sim \chi^2_{m-1}, \quad \frac{(n-1)S_Y^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\Longrightarrow \quad \frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} \sim \chi^2_{m+n-2}$$

Step 3: Thus, writing

$$S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2},$$

we obtain the result:

$$T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)}} \sim t_{m+n-2}$$

■

Example 15. — Etruscan and Italian skull widths

Table 4.3 Ancient Etruscan and modern Italian skull widths

Ancient Etruscan skulls							Modern Italian skulls					
141	147	126	140	141	150	142	133	124	129	139	144	140
148	148	140	146	149	132	137	138	132	125	132	137	130
132	144	144	142	148	142	134	130	132	136	130	140	137
138	150	142	137	135	142	144	138	125	131	132	136	134
154	149	141	148	148	143	146	134	139	132	128	135	130
142	145	140	154	152	153	147	127	127	127	139	126	148
150	149	145	137	143	149	140	128	133	129	135	139	135
146	158	135	139	144	146	142	138	136	132	133	131	138
155	143	147	143	141	149	140	136	121	116	128	133	135
158	141	146	140	143	138	137	131	131	134	130	138	138
150	144	141	131	147	142	152	126	125	125	130	133	
140	144	136	143	146	149	145	120	130	128	143	137	

The width measurements are taken with the aim of comparing modern day Italians with ancient

Etruscans. The null hypothesis is therefore that the mean skull width is the same. In what follows X refers to Ancient Etruscan measurements and Y refers to Modern Italian.

$$\bar{x} - \bar{y} = 11.33,\ m = 84,\ n = 70.$$

Using the formulas above, the value of the test statistic turns out to be $T = 11.92 >> t_{\alpha/2} \approx 1.98$. (As we are just asking "*is there a difference?*", we need a 2-sided alternative hypothesis.) The test provides overwhelming evidence to suggest that the two populations are ancestrally of different origin.

6.3.4 Pearson's χ^2 test (goodness-of-fit test)

This is a method for testing how well a particular distribution F_X fits the histogram of a single random variable X from a sample of size n. **Hypothesis H_0:** We claim that

$$\mathrm{P}(\text{“Experiment falls in bin } i\text{”}) = \mathrm{P}(X \in \text{bin } i), \quad i = 1, \ldots, k.$$

where $\mathrm{P}(X \in \text{bin } i)$ is calculated using F_X. Let O_i be the observed number in bin i, $i = 1, 2, \ldots, k$ and E_i be the expected number in bin i:

$$E_i = n \cdot \mathrm{P}(X \in \text{bin } i) \tag{6.3}$$

Pearson's statistic is

$$\boxed{Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-1-n_p}.} \tag{6.4}$$

where n_p is the number of parameters to be estimated, if any.

How large Q should be to reject the hypothesis? Reject H_0 if $Q > \chi^2_\alpha(k-1-n_p)$. Further, in order to use the test, as a rule of thumb one should check that $nE_i > 5$ for all i.

Example 16. — Flying bomb hits on London The number of flying bomb hits recorded in each of 576 small areas of $\frac{1}{4}km^2$ in the south of London during World War II.

Flying bomb hits on London							
Number of hits in an area	0	1	2	3	4	5	≥ 6
Frequency	229	211	93	35	7	1	0

Propaganda broadcasts claimed that the weapon could be aimed accurately. If, however, this was not the case, the hits should be randomly distributed over the area and should therefore be fitted by a Poisson distribution. Is this the case?

Solution:
The first thing to do is estimate the the Poisson parameter. Since we know that $\mathrm{E}(X) = \theta$ a good candidate is:

$$\hat{\theta} = \bar{x} = \frac{535}{576} = 0.927$$

Second, we pool small cells and we obtain

Number of hits in an area	0	1	2	3	≥ 4
Frequency O_i	229	211	93	35	8
Expected frequency E_i	211.7	196.3	91.	28.1	7.9

and hence

$$Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 4.25$$

This is tested against $\chi^2(r = 5 - 1 - 1) = 7.814$. Clearly there is not a shred of evidence in favor of rejection. We have therefore found no evidence to reject the hypothesis that the Poisson distribution is a good model. Therefore, there is no evidence that V1 flying bomb could be aimed with any degree of precision.

$\square$

Example 17. — Fair dice? In 1882, R. Wolf rolled a dice $n = 20\,000$ times and recorded the number of eyes shown

Number of eyes i	1	2	3	4	5	6
Frequency O_i	3407	3631	3176	2916	3448	3422

Was his dice fair?

Solution: Sample space $\mathcal{S} = \{1, \ldots, 6\}$ and let random variable X be the number shown. Since the dice is assumed fair, all results are equally probable hence $\mathrm{P}(X \in \text{bin } i) = \mathrm{P}(X = i) = 1/6$.
In our example $k = 6$.
For Wolf's data Q is

$$Q = 1.6280 + 26.5816 + 7.4261 + 52.2501 + 3.9445 + 2.3585 = 94.2$$

Since $r = k - 1 = 5$ and the quantile $\chi^2_{0.05}(r) = 11.1$, we have $Q > \chi^2_{0.05}(5)$ which leads to **rejection of the hypothesis of a fair dice**. □