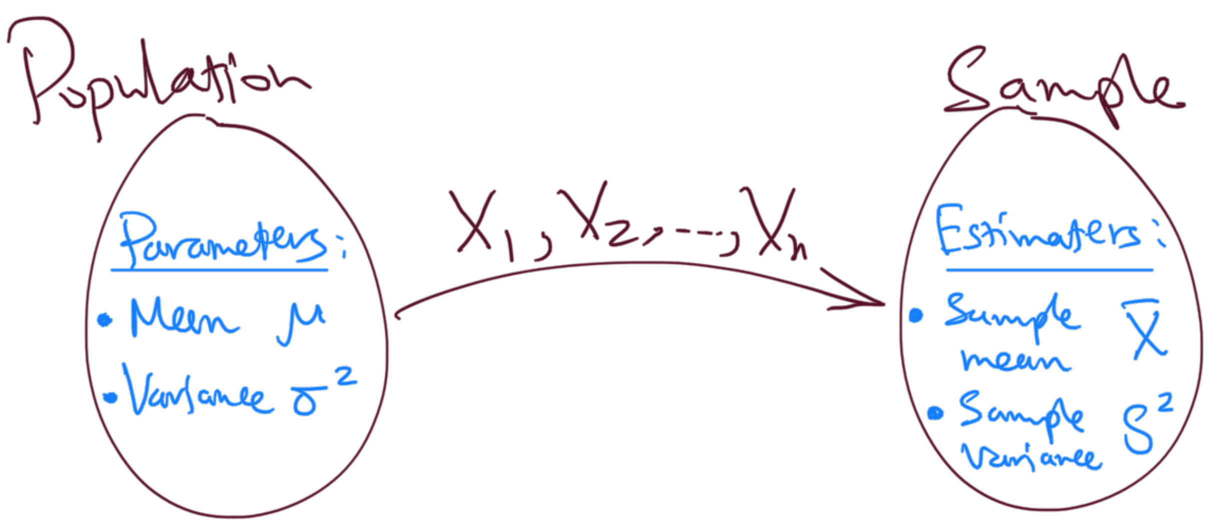
Sampling distributions

The distribution of the sample mean.

Let $X_1, X_2, \ldots, X_n$ represent n independent observations from a population with mean μ and standard deviation σ^2.
(We are drawing a random sample of n observations from this distribution).
Let $\overline{X}$ represent the mean of these n observations: $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.



Then:

1) The mean of the sampling distribution of $\overline{X}$ is $\mu_{\overline{X}} = \mu$. (i.e., $E(\overline{X}) = \mu$).

The mean of the sampling distribution of the sample mean.

The mean of the population from which we are sampling.

2) The standard deviation of the sampling distribution of $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \quad \left(\text{i.e., } Var(\overline{X}) = \frac{\sigma^2}{\sqrt{n}}\right).$$

The standard deviation of the sampling distribution of the sample mean.

The standard deviation of the population from which we are sampling.

3) If the population is normally distributed, then $\overline{X}$ is also normally distributed.

i.e., if $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

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

For probability calculations, we will standardize in the usual way. (If we are sampling from a normally distributed population).

For a single value:

$Z = \dfrac{X - \mu}{\sigma}$ has the standard normal distribution.

For the mean of n observations:

$Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ has the standard normal distribution.

Ex. Suppose that the weights of a certain population are normally distributed with mean $\mu = 70$ Kgs. and standard deviation $\sigma = 10$ Kgs. If a sample of size $n = 25$ persons is to be drawn, what is the probability that:

a) their average weight will be less than 73 Kgs.?

b) their total weight exceeds 1800 Kgs?

<u>Soln.</u> $X_1, X_2, \dots, X_{25} \sim N(70, 10^2)$.

Then $\bar{X} \sim N(70, 10^2/25)$, i.e. $\bar{X} \sim N(70, 2^2)$.

a) $P(\bar{X} < 75) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{73-70}{2}\right)$

$= P(Z < 1.5)$

$= 0.9332$.

b) $P\left(\sum_{i=1}^{25} X_i > 1800\right) = P\left(\bar{X} > \frac{1800}{25}\right)$

$= P(\bar{X} > 72)$

$= P\left(Z > \frac{72-70}{2}\right)$

$= P(Z > 1)$

$= 1 - P(Z \leq 1)$

$= 1 - 0.8413$

$= 0.1587$.

<u>Ex:</u> Suppose that the grades in a general eximination are normally distributed with mean 68 and standard deviation of 12 points. If a sample of four grades are to be drawn, what is the probability that the average of the grades drawn will be:

a) more than 71?

b) less than 65?

c) between 66 and 74?

<u>Solu.</u> $X_1, X_2, X_3, X_4 \sim N(68, 12^2)$.

Then $\bar{X} \sim N(68, 12^2/4)$ or $\bar{X} \sim N(68, 36)$.

a) $P(\bar{X} > 71) = P\left(Z > \frac{71-68}{2}\right)$

$= P(Z > 0.5)$

$= 1 - P(Z \leq 0.5)$

$= 1 - 0.6915$

$= 0.3085.$

b) $P(\bar{X} < 65) = P\left(Z < \frac{65-68}{6}\right)$

$= P(Z < -0.5)$

$= 0.3085$

c) $P(66 < \bar{X} < 74) = P\left(\frac{66-68}{6} < Z < \frac{74-68}{2}\right)$

$= P(-0.33 < Z < 1)$

$= P(Z < 1) - P(Z < -0.33)$

$= 0.8413 - 0.3707$

$= 0.4706.$

Ex- Suppose that the weights of orange boxes are normally distributed with mean 10 Kgs. and standard deviation of 1.5 Kgs. If a number of boxes will be loaded in a car with threshold 1000 kgs. Find the number of boxes that will be loaded so that their total weight does not exceed the threshold of the car with probability about 0.95.

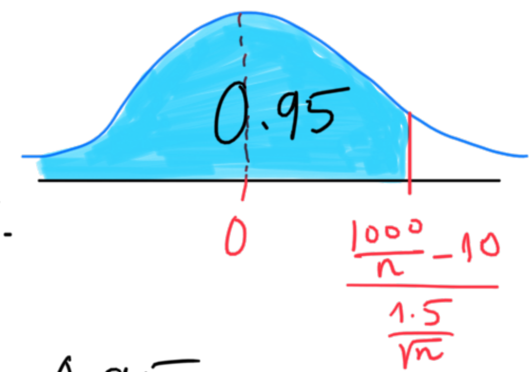
Soln. $X_1, X_2, \ldots, X_n \sim N(10, 1.5^2)$.

Then $\bar{X} \sim N(10, 1.5^2/n)$.

$$P\left(\sum_{i=1}^{n} X_i \le 1000\right) = 0.95.$$

$$P\left(\bar{X} \le \frac{1000}{n}\right) = 0.95.$$

$$P\left(Z \le \frac{\frac{1000}{n} - 10}{\frac{1.5}{\sqrt{n}}}\right) = 0.95.$$



From tables, we have $\dfrac{\frac{1000}{n} - 10}{\frac{1.5}{\sqrt{n}}} = 1.64.$

Then $\left(\frac{1000}{n} - 10\right) \frac{\sqrt{n}}{1.5} = 1.64.$

Let $x=\sqrt{n}$, then $x^2=n$, and hence

$$\left(\frac{1000}{x^2}-10\right)\frac{x}{1.5}=1.64$$

$$\vdots$$

$x^2=97.57$. Thus, $n\approx 98$.

Distribution of the sample mean

When σ is known

In this case, we transform $\bar{X}$ to an approximately standard normal variable

$$Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

as we discussed above.

When σ is unknown

In this case, we cannot transform $\bar{X}$ to standard normal, but we can estimate σ using the sample standard deviation S and transform $\bar{X}$ to to a variable with a similar distribution, called the t-distribution.

If the population variance σ^2 is unknown, then σ^2 can be estimated by the sample variance

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

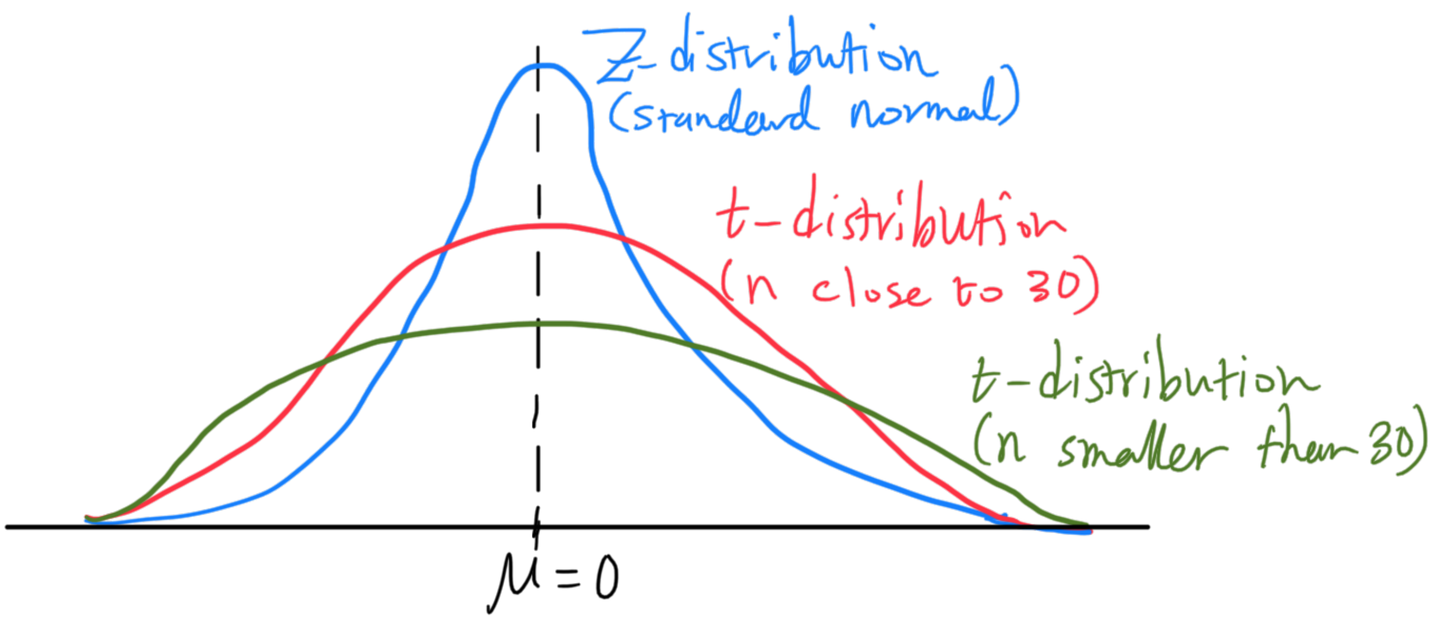
It can be shown that $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}$ has a sampling distribution called the t-distribution with (n-1) degrees of freedom (d.f.). ← To be explained later.

For this purpose, we use the following notation

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim T(n-1).$$

However, for $n \geq 30$, we estimate with Z-dist.:

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



The t-distribution has the following characteristics:-

1) It is symmetric about zero.

2) It has heavier tails (i.e., the curve does not converge to the horizontal line as quickly as the normal line).

3) The shape of t-distribution depends on the sample size n.

Q2: What are degrees of freedom (d.f.)?

Ans: The number of independent values or quantities which can be assigned to a statistical distribution.

e.g., the d.f. associated with 3 numbers is 2 when their mean is known.

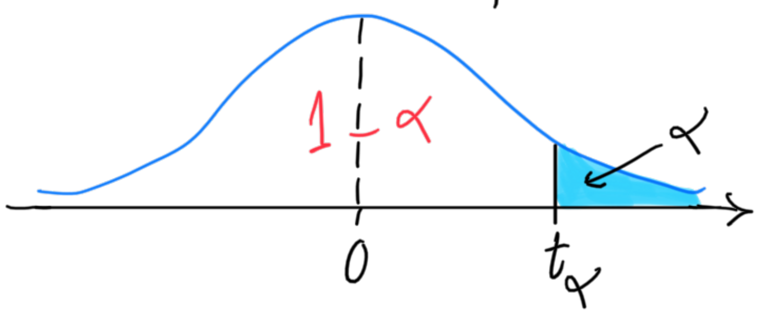
There numbers: $\boxed{?}$ $\boxed{10}$ $\boxed{15}$ with $\bar{X} = 10$

To ensure $\bar{X} = 10$, this number must be 5.

We were able to decide 2 numbers only.

So d.f. = 2 here.

A portion of T-table is presented below.



d.f.	$t_{0.1}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
1	*	*	*	*	*
2	*	*	*	*	*
3	*	*	*	*	*
4	*	*	*	*	*
5	*	*	*	*	*
6	*	*	*	*	*
7	*	*	*	*	*
8	*	*	*	*	*
9	1.383	1.833	2.262	2.821	3.250
10	*	*	*	*	*

What does this number mean?

It means that, when d.f. = 9, we have

$P(T \geq 1.383) = 0.1$ and $P(T < 1.383) = 0.9$.

Ex. Suppose that the weights of newborn babies

It means that, when d.f. = 9, we have

$P(T \geq 1.383) = 0.1$ and $P(T < 1.383) = 0.9$.

Ex. Suppose that the weights of newborn babies

b) What is the 90^{th} - percentile of the distribution of $\bar{X}$?

unknown.

Soln. $X_1, X_2, \ldots, X_{10} \sim N(3, \sigma^2)$

$n = 10$, $S = 2$. So we have t-dist. (as $n < 30$)

(if $n > 30$ use normal!)

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim T(9).$$

a) $$P(\bar{X} < 4.16) = P\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} < \frac{4.16 - 3}{2/\sqrt{10}}\right)$$

$$= P(T < 1.834)$$

$$= 0.95.$$

b) $P(\bar{X} < P_{90}) = 0.90$. Then

$$P\left(T < \frac{P_{90} - 3}{3/\sqrt{10}}\right) = 0.9.$$

From the T-table, $\frac{P_{90} - 3}{2/\sqrt{10}} = 1.383$.

Thus, $P_{90} = 3.87$.

<u>Summary:</u> $X_1, X_2, \dots, X_n \overset{r.s.}{\sim} N(\mu, \sigma^2)$ random sample

σ is known

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

or $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

σ is unknown

$n < 30$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim T(n-1).$$

$n \geq 30$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0,1).$$

<u>Ex.</u> If $T \sim T(10)$, find

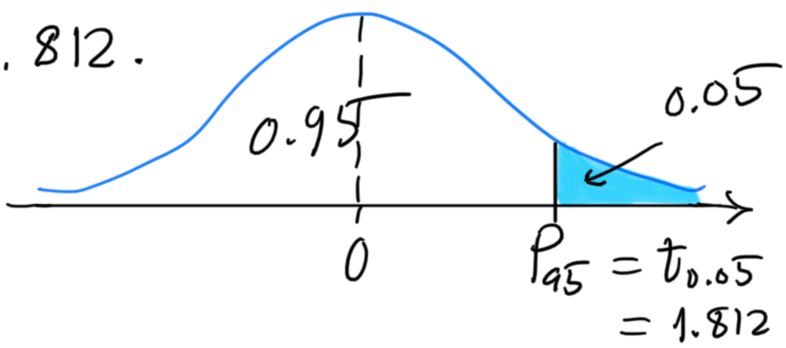
a) the 95^{th} - percentile of T.

b) the 10^{th} - percentile of T.

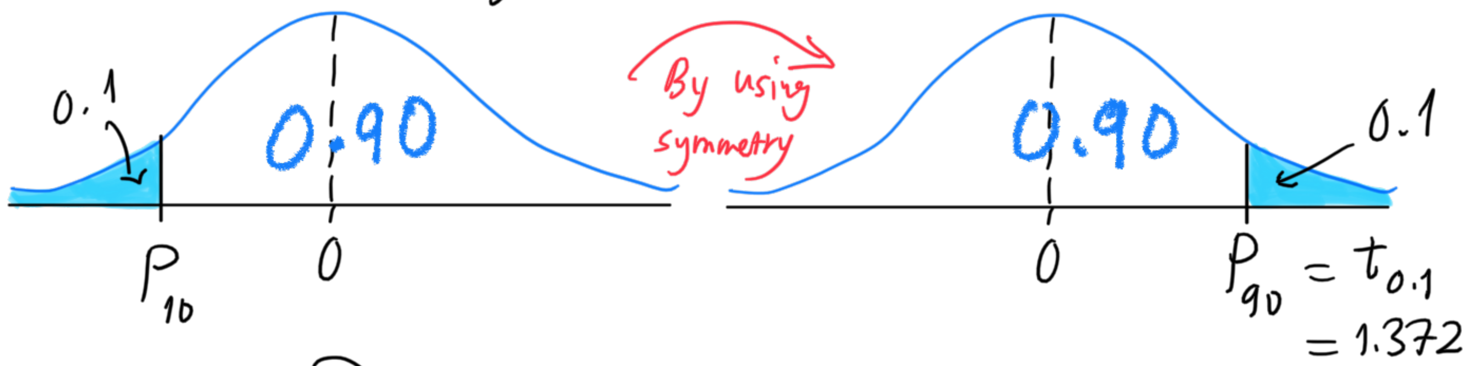
<u>Soln</u> d.f. = 10.

a) $P(T \leq P_{95}) = 0.95$

From tables $P_{95} = 1.812$.



b) $P(T \le P_{10}) = 0.10.$



Thus, $P_{10} = -1.372.$

Searching keywords:

- Sampling distributions, distribution of the sample mean
- T-distribution and T-tables
- Find the probability of
- The University of Jordan الجامعة الأردنية
- Principles of Statistics مبادئ الإحصاء
- Baha Alzalg بهاء الزالق

References: See the course website

http://sites.ju.edu.jo/sites/Alzalg/Pages/131.aspx

For any comments or concerns, please use my email to contact me.

د. بهاء محمود الزالق
The University of Jordan
Dr. Baha Alzalg
baha2math@gmail.com