STAT22000 Autumn 2013 Lecture 22

Yibi Huang

November 18, 2013

7.1 Inference for the Mean of a Population

Outline

- The t distributions
- The one-sample t confidence interval
- The one-sample t test
- Matched pairs t procedures
- Robustness
- Power of the t-test (p.419-420) . . . . Skip
- Inference for non-normal distributions (p.420-425) . . . . Skip

What if σ is Unknown?

We have $X_1, X_2, \ldots, X_n$ i.i.d. (or SRS) from a population with **unknown mean** μ and standard deviation σ.

Based on the CLT, we can construct **confidence intervals** for μ

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

and use the z-statistic for hypothesis testing

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

In all the above, we assume that the population SD σ is KNOWN. But in reality, σ is usually UNKNOWN. We usually estimated it with the **sample standard deviation**

$$s = \sqrt{\frac{1}{n-1}\sum\nolimits_{i=1}^{n}(X_i - \overline{X})^2}.$$

The t-Distributions

Suppose that i.i.d. sample of size n: $X_1, X_2, \ldots, X_n$, is drawn from an $N(\mu, \sigma)$ population.

- When σ is known, then

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

- When σ unknown and is estimated from the **sample standard deviation** s, then the z-statistic becomes the t-statistic defined as follows

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}}, \quad \text{in which } s = \sqrt{\frac{1}{n-1}\sum\nolimits_{i=1}^{n}(X_i - \overline{X})^2}.$$

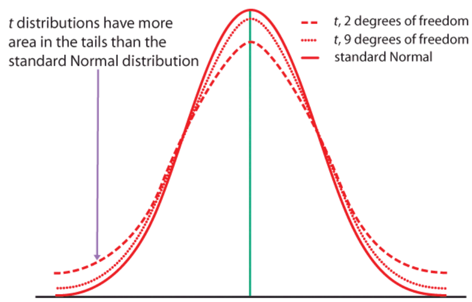
The t-statistic has a ***t*-distribution with degrees of freedom** $n-1$, denoted as

$$t \sim t_{n-1}$$

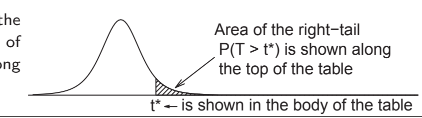
What is a ***t*-distribution with degrees of freedom** $n-1$?

The density curves of a t-distribution

- are symmetric about 0,
- are bell-shaped,
- more spread out than normal — **heavier tails**
- Exact shape of the curves depend on the degrees of freedom
- As the number of degrees of freedom increases, the t-curve approaches the standard normal curve.



t-Table (**Table D** in the Text), with degrees of freedom shown along the left of the table.



	Upper-tail probability p											
df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.90	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.22	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.842	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.090	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Exercise

t-table

	Upper-tail probability p											
df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.90	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.22	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869

Let T_d be a random variable with t-distribution with d degrees of freedom. Find

(a) $P(T_3 > 1.25) = 0.15$

(b) $P(T_5 > 2.015) = 0.05$

(c) $P(|T_5| > 2.015) = 2 \times P(T_5 > 2.015) = 2 \times 0.05 = 0.1$

(d) $P(T_5 > 5) =$ between 0.0025 and 0.001

(e) $P(|T_5| > 5) = 2 \times P(T_5 > 5) =$ between 0.005 and 0.002

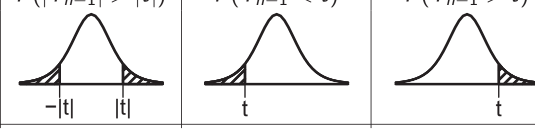
One-Sample t-test

Suppose a simple random sample (or i.i.d. sample) of size n, $X_1, \dots, X_n$, is drawn from a $N(\mu, \sigma)$ population with both μ and σ unknown. The t-statistic,

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}}, \quad \text{in which} s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

has the t distribution with $n-1$ d.f. To test $H_0 : \mu = \mu_0$, first calculate the t-statistic above and then find p-value as follows.

	Two-Sided	Lower One-Sided	Upper One-Sided
H_1	$\mu \neq \mu_0$	$\mu < \mu_0$	$\mu > \mu_0$
P-value	$P(\lvert T_{n-1}\rvert > \lvert t\rvert)$	$P(T_{n-1} < t)$	$P(T_{n-1} > t)$



The bell curve above is the t-curve with $n-1$ degrees of freedom, not normal curve

	Upper-tail probability p											
df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	.0005
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725

Example 1. The one-sample t statistic for testing

$$H_0 : \mu = 10 \quad \text{v.s.} \quad H_a : \mu > 10$$

from a sample of $n = 21$ observations has the value $t = 2.10$. Between what two values does the P-value of the test fall?

- The P-value $= P(T_{20} > 2.1)$ is between .025 and 0.02.

Example 2. The one-sample t statistic for testing

$$H_0 : \mu = 60 \quad \text{v.s.} \quad H_a : \mu \neq 60$$

from a sample of $n = 24$ observations has the value $t = 2.6$. Between what two values does the P-value of the test fall?

- Ans: $P(T_{23} > 2.6)$ is between 0.01 and 0.005. The P-value $= 2P(T_{23} > 2.6)$ is between 0.02 and 0.01.

	Upper-tail probability p											
df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	.0005
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300

Example 3. The one-sample t statistic for testing

$$H_0 : \mu = 20 \quad \text{v.s.} \quad H_a : \mu < 20$$

from a sample of $n = 115$ observations has the value $t = -1.55$. Between what two values does the P-value of the test fall?

- Ans: The df $115 - 1 = 114$ is not on the table. Look at the available dfs above and below 114, which are 1000 and 100. $P(T_{100} < -1.55) = P(T_{100} > 1.55)$ is between 0.1 and 0.05. $P(T_{1000} < -1.55)$ is also between 0.1 and 0.05. So the P-value $P(T_{114} < -1.55)$ is also between 0.1 and 0.05.

Example: Growth of Tumor (1)

- Let X (in millimeter, or mm) be the growth in 15 days of a tumor induced in a mouse. It is known from a previous experiment that the average tumor growth is 4mm.
- A sample of 20 genetically variant mice used in the tumor growth study yielded $\overline{x} = 3.8$mm, $s = 0.3$mm.
- We want to test $\mu = 4$ or not (assuming growths are normally distributed).

Example: Growth of Tumor (2)

1. State the hypotheses

$$H_0 : \mu = 4 \qquad H_a : \mu \neq 4$$

2. Calculate the t-statistic

$$t = \frac{3.8 - 4.0}{0.3/\sqrt{20}} = -2.98$$

3. Determine the P-value
   From the t-table we know $P(T_{19} > 2.98)$ is between 0.005 and 0.0025. So the P-value $= 2P(T_{19} > 2.98)$ is between 0.01 and 0.005.

Since p is less than 0.01, we reject H_0 at significance level $\alpha = 0.01$. There is evidence that the population mean growth is not $4mm$.

Confidence Intervals with Unknown σ

Suppose that i.i.d. sample of size n: $X_1, X_2, \ldots, X_n$, is drawn from an $N(\mu, \sigma)$ population.
Recall when σ is known, the $(1-\alpha)$ confidence interval for μ is

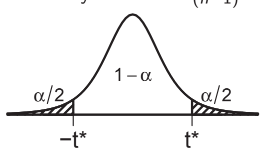
$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

When σ is unknown, and is estimated using the **sample standard deviation** $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$, the $(1-\alpha)$ confidence interval for μ becomes

$$\overline{X} \pm t^*\frac{s}{\sqrt{n}}.$$

The **critical value** $t^* = t_{n-1,\alpha/2}$ is chosen such that $(1-\alpha)$ of the area under the $t_{(n-1)}$ density lies between $-t^*$ and t^*.

density curve of $t_{(n-1)}$



	Upper-tail probability p											
df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	.0005
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.842	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.090	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Find the critical value t^* from Table D for calculating a confidence interval in each of the following situations.

(a) A 95% confidence interval based on $n = 22$ observations.
$df = 22 - 1 = 21$. $t^* = t_{21,0.025} = 2.080$

(b) A 95% confidence interval from an SRS of 25 observations.
$df = 25 - 1 = 24$. $t^* = t_{24,0.025} = 2.064$

(c) A 90% confidence interval from a sample of size 115.
$df = 115 - 1 = 114$ is not in the Table. Use the largest df available below 114, which is 100. $t^* = t_{100,0.05} = 1.660$

Example: Sitcom

Is your favorite TV program often interrupted by advertising? CNBC presented statistics on the average number of programming minutes in a half-hour sitcom[1]. The following data (in minutes) are representative of their findings.

21.06, 22.24, 20.62, 21.66, 21.23,
23.86, 23.82, 20.30, 21.52, 21.52,
21.91, 23.14, 20.02, 22.20, 21.20,
22.37, 22.19, 22.34, 23.36, 23.44

Assume the population is approximately normal.

$$\overline{X} = 22.00, \quad s \approx 1.12, \quad n = 20, \quad t_{19,0.025} = 2.093$$

A 95% confidence interval for the population mean (the mean number of programming minutes during a half-hour TV sitcom) is:

$$22.00 \pm 2.093 \times 1.12/\sqrt{20} \approx 22.00 \pm 0.52 = (21.48, 22.52)$$

[1] CNBC, February 23, 2006

Does Lack of Caffeine Increase Depression?

— a Matched-Pair Study

Individuals diagnosed as caffeine-dependent are deprived of caffeine-rich foods and assigned to receive daily pills. Sometimes, the pills contain caffeine and other times they contain a placebo. Depression was assessed.

	depression with		diff-
subject	caffeine	placebo	erence
1	5	16	11
2	5	23	18
3	4	5	1
4	3	7	4
5	8	14	6
6	5	24	19
7	0	6	6
8	0	3	3
9	2	15	13
10	11	12	1
11	1	0	−1

- In matched pairs designs, there are 2 measurements taken on the same subject or on 2 similar subjects.
- To conduct statistical inference on such a sample, we analyze the *difference* using the *one-sample* procedures.

Example – Matched Pairs t-test

For each individual in the sample, we have calculated a difference in depression score (placebo minus caffeine).

There were 11 "differences" observations, thus $df = 11 - 1 = 10$ (not $22 - 1$). We calculate that $\overline{X} = 7.36$; $s = 6.92$.
To test whether lack of caffeine *increase* depression, let

$$H_0 : \mu = 0 \qquad H_a : \mu > 0$$

where μ is the mean difference (placebo minus caffeine).

The t-statistic is $t = \dfrac{\overline{X} - 0}{s/\sqrt{n}} = \dfrac{7.36 - 0}{6.92/\sqrt{11}} = 3.53.$
For $df = 10$,

$$t_{10,0.005} = 3.169 < t = 3.53 < t_{10,0.0025} = 3.581,$$

thus the P-value $= P(t_{10} \geq 3.53)$ is between 0.005 and 0.0025.

Conclusion: Caffeine deprivation causes a significant increase in depression.

Robustness

The t procedures are exactly correct when the population is distributed exactly normally. However, most real data are not exactly normal.
The t procedures are robust to small deviations from normality — the results will not be affected too much. Factors that strongly matter are:

- The sample must be an **SRS** or **i.i.d.** from the population.
- **Outliers and skewness**. They strongly influence the mean and therefore the t procedures. However, their impact diminishes as the sample size gets larger because of the Central Limit Theorem.

Robustness (2)

Specifically, to use the t procedures

- When $n < 15$, the data must be close to normal and without outliers.
- When $15 > n > 40$, mild skewness is acceptable but not outliers.
- When $n > 40$, the t-statistic will be valid even with strong skewness. (Outlier is still a problem.)

Comparison of the z-Procedures and t-Procedures

For the same data set and at the same confidence level, if we pretend that the population SD σ is identical to the sample SD s, then

- a t-interval is **wider** than a z-interval, since $t_{n-1,\frac{\alpha}{2}} > z_{\frac{\alpha}{2}}$.
  - That is the price for the extra uncertainty in the estimation of σ.
- the P-value of a one-sample z-test calculated using the normal curve is smaller than that of a one-sample t-test calculated using a t-curve.
  - A z-test will be more significant and more likely to reject H_0 than a t-test