## Class 15: When $\sigma$ is Not Known: The T-Distribution (Text: Section 7.1)

**Statistical Inference**

We take a sample to learn about a population.

- We have made estimates using **confidence intervals.**
- We have decided whether to believe statements using **Hypothesis Testing**:

In both cases, we assumed we knew the population standard deviation, $\sigma$. What if we don't know $\sigma$?

**Standard Error: What if we don't know $\sigma$ for population?**

CLT says Standard error $= \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. If we don't know $\sigma$, we use $s$, the standard deviation of the sample (instead of the population), so

$$\text{Standard error} \approx \frac{s}{\sqrt{n}}$$

This expression for the standard error is an *estimate* of standard deviation of the sampling distribution, $\sigma_{\bar{x}}$.

**Central Limit Theorem:** We know $Z$ is distributed normally with mean 0, standard deviation 1, that is

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

**T-distribution**

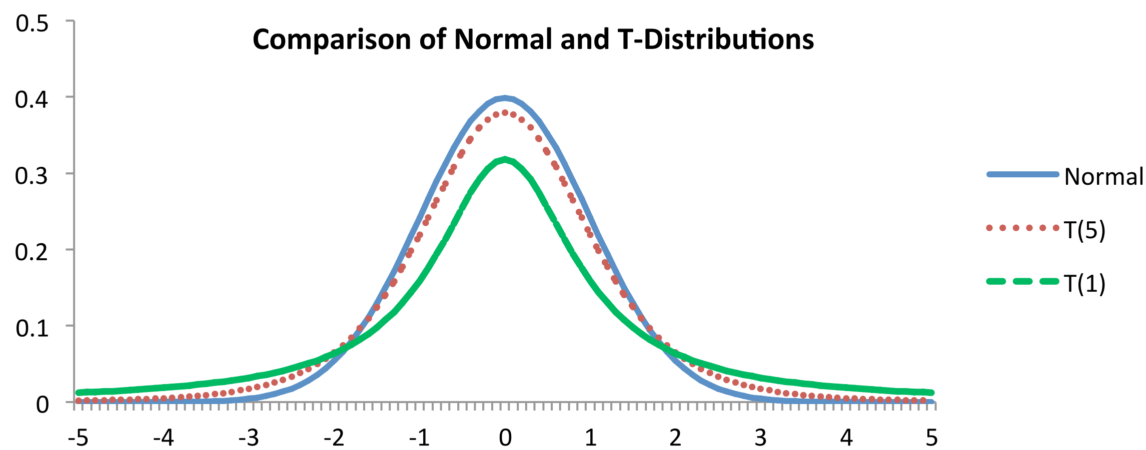
If we replace $\sigma$ by the sample standard deviation, how is the test statistic distributed?
If the original distribution is *norma*l, the test statistic has the T-distribution.

If population is $N(\mu, \sigma)$, for a SRS of size $n$, we calculate a statistic that is very like $z$:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim \text{T}(\text{n}-1)$$

This variable has the T-distribution with $(n-1)$ degrees freedom.

Ex:. Graph $N(0,1), T(5), T(1)$ on $5 \le x \le 5, 0 \le y \le 0.5$.



**What is the T-distribution? What is a degree of freedom?**

- The $T$-distributions is similar to normal, but has thicker tails, since there is more variability in $T$ than in $Z$ because $s$ is variable, whereas $\sigma$ was not.
- For large sample sizes, the $T$-distribution is approximately equal to the standard normal. For $n > 100$, you can use $Z$ instead of $T$.
- A *degree of freedom* (df) is a number that tells us how variable the T-distribution is. In the case of Z, there was only one standard normal distribution (with $\mu = 0, \sigma = 1$). In the case of $T$, there is a whole family of $T$-distributions, one for each sample size. The larger the sample size, the less $s$ is likely to vary (because bigger samples give better estimates of $s$), so the less T varies.
  --No matter what the size of the sample:

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

--The distribution of the $t$ statistic depends on $n$. The df $= n - 1$:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim \mathrm{T}\,(\mathrm{n} - 1)$$

**Reading the T-table**

In $Z$-table, probabilities measured from the left are in body of table.
In $T$-table, probabilities are across the top, and they are measured from the *right*. The $T$-values are in the body of the table. Down the left are the degrees of freedom—take the closest one if the one you want is not there. The bottom row is $Z$ scores, corresponding to an infinite df.

Ex. For 1 df, what does the 15.89 in top row mean?
$P(T > 15.89) = 0.02 = 2\%$.

Ex. For 1 df, what does the 96% at the bottom of this column mean?
Same 15.89 gives 96% confidence interval. In other words $P(-15.89 < T < 15.89) = 96\%$.

Ex. For 3 df, find $P(T > 1.638) = 0.10 = 10\%$
For 3df, find $P(T > 6) \approx 0.004 = 0.4\%$.
For 3df, find $P(T < -1.638) = 10\%$,
For 3 df, find $P(-1.638 < T < 1.638) = 80\%$.

Ex. For 10 df, find $t$ for an upper tail probability of 2.5%. What confidence interval does this give you?
$P(T > 2.228) = 0.025$, so $t = 2.228$
This gives a 95% confidence interval, since $P(-2.228 < T < 2.228) = 95\%$.

Ex. For 10 df, find the $T$-values for a 99% confidence interval
$t = \pm 3.169$
Note that the $P(-3.169 < T < 3.169) = 99\%$.

**With TI-83/84**

Ex: For 1 df: find $P(|T|) < 15.89) = \text{tcdf}(-15.89, 15.89, 1) = 0.96$
For 3 df: find $P(T > 1.638) = \text{tcdf}(1.638, 100, 3) = 0.10$
For 10 df: find$P(T > 2.228) = \text{tcdf}(2.228, 100, 10) = 0.025$

tcdf $(a, b, n) = P(a < X < b)$ if $X$ is distributed according to $t$-distribution with $n$ degrees of freedom

**Confidence intervals for mean μ using the T-Distribution**

If $\sigma$ is not known, and $s$ calculated from data, replace $z$-values by $t$-values. Confidence interval is

$$\left(\bar{x} - t\frac{s}{\sqrt{n}}, \bar{x} + t\frac{s}{\sqrt{n}}\right)$$

The margin of error is $ME = t\frac{s}{\sqrt{n}}$. The value of $t$ depends on the confidence level.

Ex. **A coaching service claims to raise SAT scores.** Without coaching, scores are normally distributed with mean of 475. If a random sample of 40 students who are coached has mean score 500 with standard deviation 90:
(a) Find a 95% confidence interval for the mean of all coached students.
(b) Decide if the mean of the coached group is significantly different than the mean, 475.

(a) For 95% confidence, $df = 39 \approx 40$, from table find $t = 2.021$ and interval is

$$(500 - 2.021\frac{90}{\sqrt{40}}, 500 + 2.021\frac{90}{\sqrt{40}}) = (471.2, 528.8).$$

(b) Population mean is in this interval, so the mean of the coached groups is *not* significantly higher.

A confidence interval can always be used to do a hypothesis test:
—If the mean of the population is *outside* the interval, result is *significant*
—If the mean of the population is *inside* the interval, result is *not significant*

Ex: If we take 100 samples and the find 100 corresponding 95% confidence intervals, how often do we expect the true mean to be in the confidence interval?
If we take 100 such samples, and find 95% confidence intervals for each one, we expect the true mean to lie in about 95 of them. It may not be exactly 95 because of random variation, but we expect it to be close to 95% of the number of intervals.

**Hypothesis Testing**

Ex: Use the same SAT coaching data. Use a hypothesis test to decide: **Does coaching service *change* scores?** Give null and alternative hypothesis; $P$-value. Use 5% and 1% significance levels.

**Step 1**:

$H_0$: $\mu = 475$

$H_a$: $\mu \neq 475$

This is a **two**-sided test because the alternate hypothesis is ≠. So the alternate hypothesis is different than in the one-sided case.

**Step 2**:

Calculate the test statistic, which is the same as in the one-sided case:

$$t = \frac{500 - 475}{90/\sqrt{40}} = 1.757 \text{ which has the } T \text{ distribution with 39 df.}$$

**Step 3:**

The $p$-value changes for a two-sided test.

With the table, we find the probability for the one-sided case. Here we have to double it:

$$1.757 \to \left.\begin{matrix} P(T > 1.684) = 0.05 \\ P(T > 2.021) = 0.025 \end{matrix}\right\} \text{ so } P\text{-value is between } 2(0.025) \text{ and } 2(0.05)\text{; that is between 5\% and 10\%.}$$

With a calculator, we can find the $p$-value more precisely:

$$P(|T| > 1.757) = 2 \cdot \mathrm{tcdf}(1.757, 100, 39) = 2(0.043) = 8.6\%$$

**Step 4**:

Since $8.6\% > 5\%$, the increase in scores is *not* significant at 5% level or at the 1% level, so we do not reject the null hypothesis. We do *not* have evidence that the mean is changed.

**Interpretation: If the mean is unchanged by coaching**,

- There is **not** a less than 5% chance that would observe this change by chance—so we can't say there's been a change—coached students could have had the original mean.
- This is the same result as from the confidence interval.

Ex. Use the same SAT coaching data. Use a hypothesis test to decide: **Does coaching service raise scores?** Give null and alternative hypothesis; $p$-value. Use 5% and 1% significance levels.

**Step 1**:
Null and alternate hypothesis are
$H_0$: $\mu = 475$
$H_a$: $\mu > 475$
This is a **one**-sided test because the alternate hypothesis is >

**Step 2**:
Calculate the test statistic:

$$t = \frac{500-475}{90/\sqrt{40}} = 1.757 \text{ which has the } T \text{ distribution with 39 df.}$$

**Step 3:**
Computing P-values
With the table, we can only find a range of values of P-values (which is perfectly adequate):

$$1.757 \rightarrow \left.\begin{matrix} P(T > 1.684) = 0.05 \\ P(T > 2.021) = 0.025 \end{matrix}\right\}$$ so $P$-value is between 0.025 and 0.05; that is between 2.5% and 5%.

With a calculator, we can find the P-value more precisely:

$$P(T > 1.757) = \text{tcdf}(1.757, 100, 39) = 0.043 = 4.3\%$$

**Step 4**:
Since 4.3% < 5% so the increase in scores *is* significant at 5% level so, at 5%, reject hypothesis that mean is unchanged. We conclude the mean has been raised.
But 4.3% ≮ 1%, so the increase *not* significant at 1% level. So, at 1% level, we cannot say mean has been raised.

Ex: Compare the results of a one sided two-sided test for the previous example.
The $p$-value for a one sided test is 4.3%; for a two-sided test it is 8.6%. Thus the one-sided test shows a significant increase; the two-sided test does not.

Notice that it is possible to draw one conclusion from a two sided test and a different one from a one-sided test.

Ex. **Alternate method: Instead of P-values, use table to compare critical t-values**
The T-values in the table are called *critical values*. Instead of using $P$-values, we can compare the $t$-value we get with the critical values.
For 39 ≈ 40 df,
For $p = 0.05$, critical $t = 1.684$. Since $1.684 < 1.757$, increases *is* significant at 5% level.
For $p = 0.01$, critical $t = 2.423$. Since $1.757 < 2.423$, increases is *not* significant at 1% level.

If our $t$-value is *larger* in magnitude than the critical value corresponding to the significance level, the result is significant.

Ex. A school wonders whether its SAT scores are comparable to the national average, 475. A SRS of 70 students in this school has mean $\mu = 460$ and standard deviation $s = 80$.
**Does the school have a significantly different SAT average?** Test at 5% level, 1% level, 10% level.
Notice that the standard deviation is of the sample, not the population.
**Step1**:
Null hypothesis $H_0$: $\mu = 475$,
Alternate hypothesis $H_a$: $\mu \neq 475$
This is a two sided test because of the $\neq$
**Step2**: We find the standard error

$$SE_{\bar{x}} = \frac{80}{\sqrt{70}} = 9.56,$$

and the test statistic

$$t = \frac{460 - 475}{80/\sqrt{70}} = 1.57.$$

The test statistic has the $T$-distribution with $\text{df} = 69$
**Step3**: Find the $P$-value:
Table: Since $\text{df} = 69$ is not in the table, we look at $\text{df} = 60$ and $\text{df} = 80$: We have
$P\,(T > 1.671) = 0.05$ for $\text{df} = 60$ and
$P(T > 1.664) = 0.05$ for $\text{df} = 80$
Since this is a two-sided test, we have $P = 2(0.05) = 0.10$.
Calculator:
$P = 2 \cdot \text{tcdf}\,(1.57, 100, 69) = 0.12$.
**Step 4:** Since $P = 0.1$ is *not* smaller than $\alpha = 0.05$, we cannot reject Ho at 5% level (or at the 1%, 10% level). The scores are *not* significantly different than the national average.

Ex. **The same school wonders if its scores are lower than national average.**
**Step1**:
Null hypothesis $H_0$: $\mu = 475$,
Alternate hypothesis $H_a$: $\mu < 475$
This is a now a one-sided test because of the $<$ in $H_a$. The test statistic will be the same, but the $p$-value will be different because we will not multiply by 2.
**Step2:** As before $SE_{\bar{x}} = 9.56$ and $t = -1.57$.
**Step3:**
From table and symmetry
$P(T < -1.671) = 0.05$ and our $p$-value is larger than 0.05. (Because $-1.57$ is less far out on the tail than $-1.671$.
From calculator
$P(T < -1.57) = \text{tcdf}\,(-100, -1.57, 69) = 0.06$.
**Step4:** Since 0.05 or 0.06 is *not* smaller than 5% significance level, we cannot reject $H_0$ at 5% level or the 1% level
Since 0.05 or 0.06 is smaller than 10% significance level, we can reject $H_0$ at 10% level. Thus there is some evidence, but not strong, that the school has lower scores.

Notice again that we have one conclusion from a two sided test and a different one from a one sided test.