

All solutions and teacher notes in blue

AP Statistics Handout Key: Lesson 8.1

Topics: review sampling distribution for \bar{x} , 95% confidence interval for \bar{x} , four step process

Lesson 8.1 Guided Notes

YouTube Creator



Mr. Beast
2020: **\$24 Million**

Insta Influencer



Kendall Jenner
2019: **\$16 Million**

“TikToker”



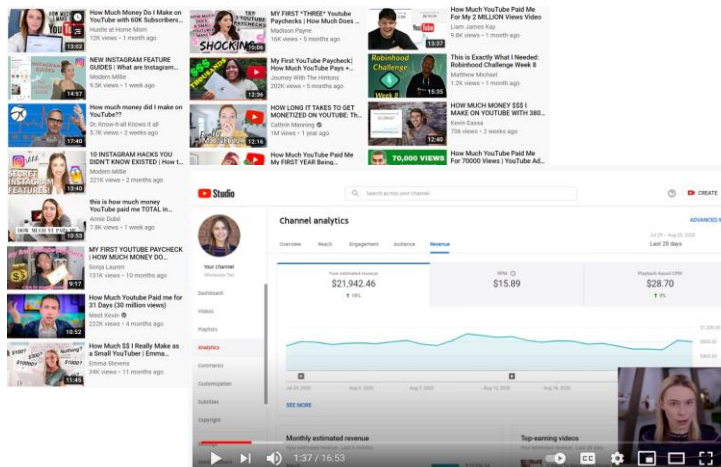
Josh Richards
2020: **\$1.5 Million**

Source: Forbes

Today’s Key Analysis: Do social media creators, on average, make a livable wage?

The Data: YouTubers

- Searched “How much I make on YouTube” and found hundreds of videos in which YouTubers show off how much they make on the platform. We randomly sampled 35 of these videos.
- From the figures shown in each video, we estimated their yearly salary.
- Since these videos show their private channel revenue pages, we know the data is reliable.



Review: Sampling Distribution for \bar{x}

The Population vs. The Sample Mean

μ = **population** mean

- Parameter
- Ex: mean salary among all YouTubers

\bar{x} = **sample** mean

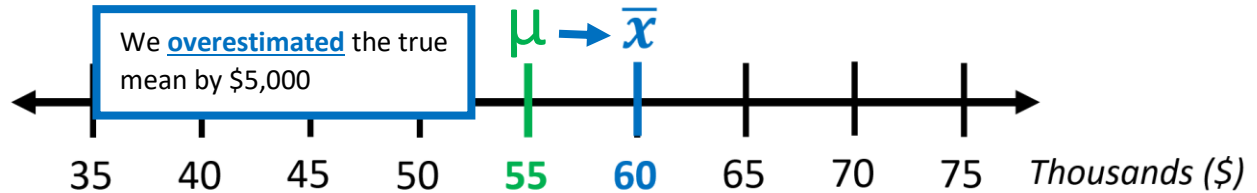
- Statistic used to estimate μ
- Ex: mean salary in our sample



In a world where...

1. The true mean yearly salary among YouTubers is \$55,000
2. The true standard deviation of salaries is \$29,500

In most scenarios, **we don't know the true mean (μ)!** So, we collect a sample and estimate using the sample mean (\bar{x}). Imagine we randomly selected 35 YouTubers. Among them, the average yearly earnings were $\bar{x} = \$60,000$.



Is a \$5,000 overestimate a typical estimation error? To explore that, we need the **sampling distribution for a mean**:

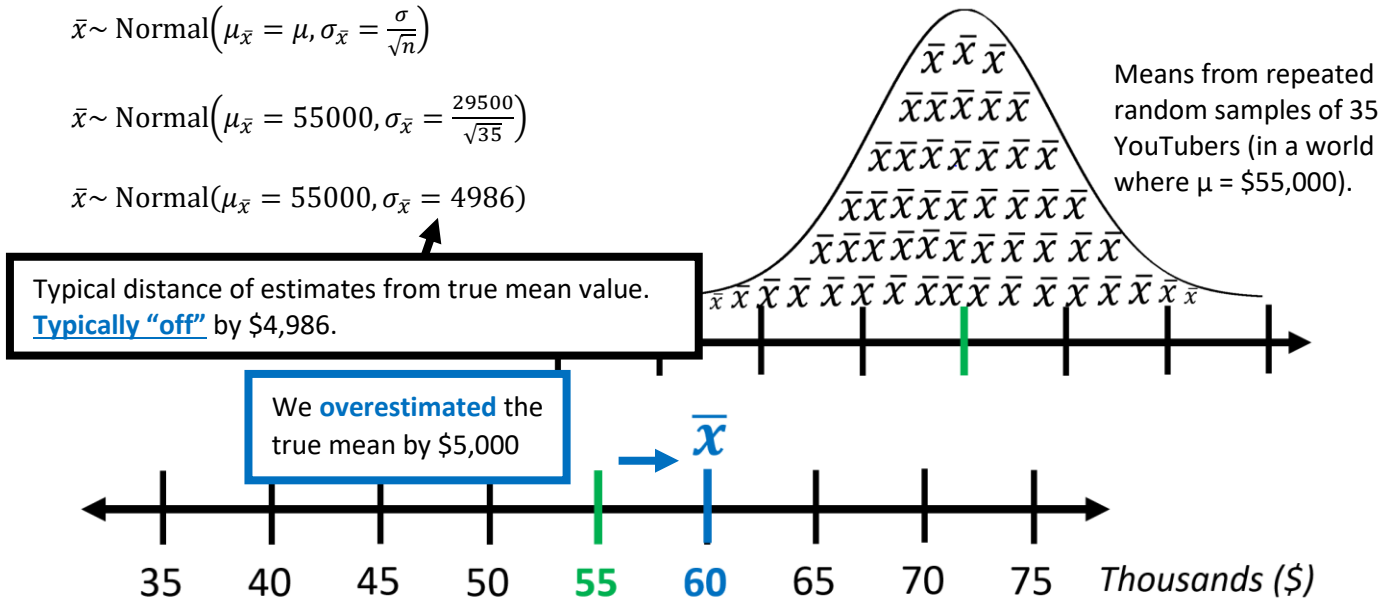
Under certain conditions: $\bar{x} \sim \text{Normal}(\mu_{\bar{x}} = \mu, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}})$

Calculations:

$$\bar{x} \sim \text{Normal}(\mu_{\bar{x}} = \mu, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}})$$

$$\bar{x} \sim \text{Normal}(\mu_{\bar{x}} = 55000, \sigma_{\bar{x}} = \frac{29500}{\sqrt{35}})$$

$$\bar{x} \sim \text{Normal}(\mu_{\bar{x}} = 55000, \sigma_{\bar{x}} = 4986)$$

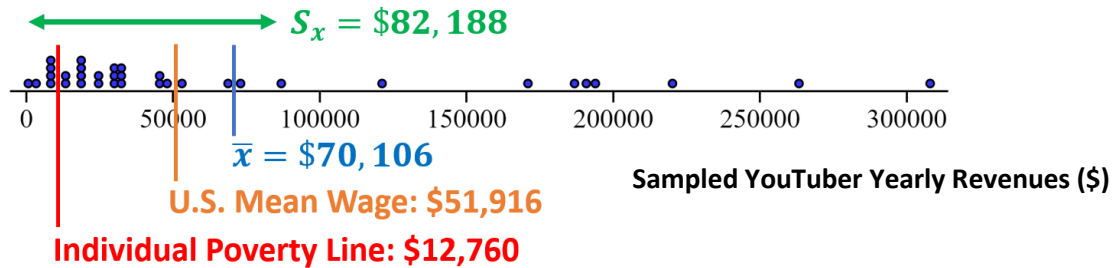


1) Was our estimate's error amount (\$5,000) pretty typical? Justify using the sampling distribution above.

This is a pretty typical estimation error in this situation. According to the sampling distribution, the typical estimate error is $\sigma_{\bar{x}} = \$4,986$. Our actual error of \$5,000 is pretty close to this.



The t-distribution and interval for \bar{x}



2) Label the U.S. individual poverty line, the U.S. mean wage, the sample mean, and the sample standard deviation on the dotplot above.

3) Do you believe it's likely that the true mean YouTuber salary is exactly the same as the mean salary among our random sample? Why or why not?

There's a chance that, due to random variation, we happened to select some unusually high-earning or unusually low-earning YouTubers, which would lead to over or under-estimating the true mean. Note: students may also point out selection bias in the sampling method. This will be covered in the discussion question.

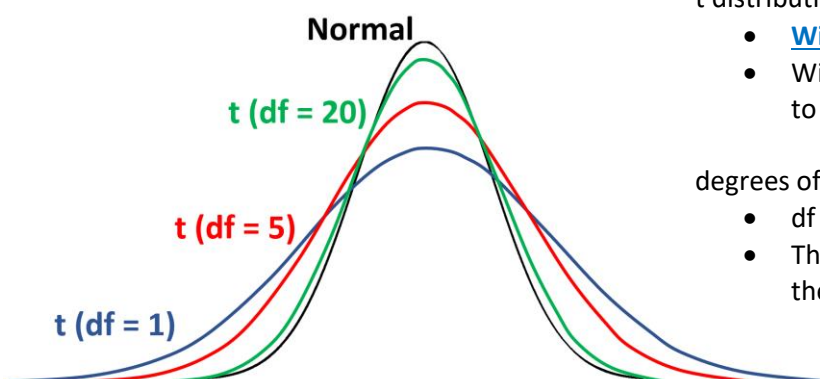
First Attempt to Create the Confidence Interval

$$\bar{x} \sim \text{Normal}\left(\mu_{\bar{x}} = \mu, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\right)$$

4) Why can't we simply use the normal distribution to calculate our confidence interval?

We don't know the true standard deviation of YouTuber salaries (σ). We plug in our sample standard deviation ($S_x = \$82,188$) in its place to provide a reasonable estimate. However, because we used an S_x in place of σ when calculating the standard error, there is some extra uncertainty involved. So, we need a distribution (besides the normal curve) that reflects this greater uncertainty.

The t-distribution to the Rescue!



t distribution

- **Wider** than normal curve
- Width captures **greater uncertainty** when using S_x to estimate σ

degrees of freedom (df) = $n - 1$

- df measures how **precisely** S_x estimates σ
- The higher n , the more precisely we estimate σ , and the more our t-curve **approaches normal!**

t critical values

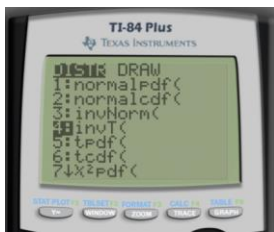
t^* : the **critical value** of the t-interval

- Tells you how many **standard errors** you're including in your interval.
- Determines the **confidence level**.

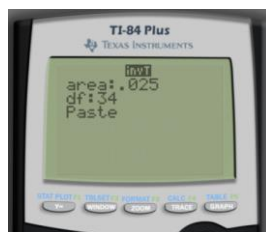
5) For a z-interval (based on a normal curve), we use a critical value (z^*) of 1.96 to capture 95% confidence. For a t-interval (based on a t-distribution) to capture 95% confidence, will the critical value (t^*) need to be higher or lower than 1.96? Explain.

It will need to be higher than 1.96. The t-distribution has extra uncertainty (i.e. wider tails) than the normal distribution, so we need to travel more standard errors away from the center to capture 95% of the confidence area.

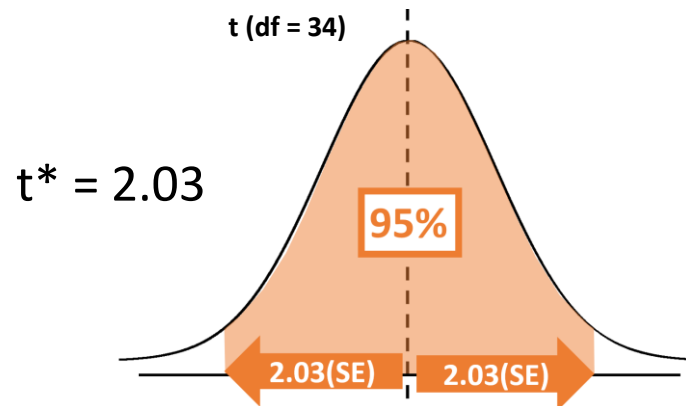
Calculator steps to find t^*



2nd → VARS → 4: InvT



-Area: percent below interval (2.5% = 0.025)
-df: $n - 1$ ($35 - 1 = 34$)



Confidence Interval for a Mean

6) Using the formula, calculate and interpret the confidence interval for the mean YouTube salary:

Formula:

- point estimate \pm margin of error
- $\bar{x} \pm t^*(SE_{\bar{x}})$

$$\bar{x} \pm t^*(SE_{\bar{x}})$$

$$70106 \pm 2.03(13892) \rightarrow (\$41905, \$98307)$$

We are 95% confident the interval from \$41,905 to \$98,307 captures the true mean yearly income of YouTubers.



The Four Steps Process

a) **Construct and interpret** a 95% confidence interval for the true mean salary of all YouTubers.

State:

We are estimating the true mean yearly YouTuber salary (μ), at 95% confidence.

Plan:

We will calculate a one-sample t-interval for μ , if all conditions are met.

Conditions

Random: The sample of 35 YouTubers was collected randomly ✓

10%: $n \leq 0.10N$

$35 \leq 0.10$ (all YouTubers)

It's reasonable to assume 35 is less than 10% of all YouTubers ✓

Normal/Large Sample:

$n \geq 30$

$35 \geq 30$ ✓

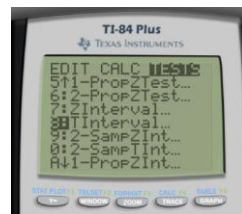
The sample size is large

Do:

T-interval ($\bar{x} = 70106$,
 $s_x = 82188$, $n = 35$,
confidence = 0.95):

(41877, 98339)

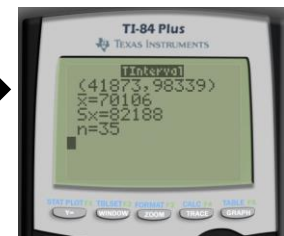
For "DO" phase: Calculator steps



STAT → TESTS →
8: Interval



→Stats
 \bar{x} : sample mean
 S_x : sample stdev
 n : sample size
Calculate



Output

Conclude:

We are 95% confident the interval from \$41,877 to \$98,339 captures the true mean yearly income of YouTubers.

Lesson 8.1 Discussion

We can directly estimate¹ the true mean income of YouTubers from Ad Revenue:

$$\begin{array}{r} \text{Creator Share of 2019 YouTube Ad Revenue} \\ \$8,331,950,000 \end{array} \bigg/ \begin{array}{r} \text{\# of Creators} \\ 18,000,000 \end{array} = \begin{array}{r} \text{Average Yearly Pay Per Creator} \\ \mathbf{\$463} \end{array}$$

Confidence interval we estimated: **\$41,877 to \$98,339**

Discussion Question: Why was our confidence interval so far off?

Recommended discussion norms:
skewthescript.org/discussion-norms

We only sampled among people who made videos about their YouTube incomes. These people likely have more to brag about than those who don't make videos about their YouTube incomes.

Even though we sampled randomly *among* that group, the selection of that group itself was biased. So, we overestimated.

Takeaway: When checking the random condition, check that you're sampling from whole population.

Lesson 8.1 Practice

Teachers: We recommend providing additional practice exercises from your AP Stats textbook or from prior AP Stats exams. The following textbook sections and AP exam questions are aligned to this lesson.

- [The Practice of Statistics \(AP Edition\)](#), 4th-6th editions section 8.3
 - 6th edition update (CED-aligned): section 10.1
- [Stats: Modeling the World \(AP Edition\)](#), 4th/5th editions: ch 22, 3rd edition: ch 23
- [Statistics: Learning from Data \(AP Edition\)](#), 2nd edition: section 12.2
- [Advanced High School Statistics](#), section 7.1
- [AP Exam Free Response Questions \(FRQs\)](#): 2013 Q1 (part b)

¹ Sources: Alphabet 4th quarter earnings release: https://abc.xyz/investor/static/pdf/2019Q4_alphabet_earnings_release.pdf?cache=05bd9fe, "How YouTube Ad Revenue Works," *Investopedia*, June 4, 2020: <https://www.investopedia.com/articles/personal-finance/032615/how-youtube-ad-revenue-works.asp#:~:text=Enabling%20ads%20on%20your%20YouTube,get%20the%20remaining%2055%20percent,> "How Many YouTube Channels Are There?", *Tubics*, <https://www.tubics.com/blog/number-of-youtube-channels>, *SocialBlade*, data accessed 2019

