STAT 311: Confidence Intervals for Means

Y. Samuel Wang

Summer 2016

Logistics

- Homework has been posted

Sampling Distribution of the Mean

From the Central Limit Theorem, we know that as n, the size of my sample, increases

$$\sqrt{n}\frac{\bar{x}_n - \mu}{\sigma} \to \mathcal{N}(0, 1)$$

Sampling Distribution of the Mean

From the Central Limit Theorem, we know that as n, the size of my sample, increases

$$\sqrt{n}\frac{\bar{x}_n - \mu}{\sigma} \rightarrow \mathcal{N}(0, 1)$$

...But in almost every case, we don't know σ

Sampling Distribution of the Mean

So typically we consider using s_x, the standard deviation of the data, to estimate σ, the population standard deviation

$$\sqrt{n}\frac{\bar{x}_n - \mu}{s_x}$$

The sample standard deviation, s_x, is also a random variable and changes from sample to sample. Yesterday in lab we talked about a distribution which arises from the ratio of certain random variables.

Sampling Distribution of the Mean

So typically we consider using s_x, the standard deviation of the data, to estimate σ, the population standard deviation

$$\sqrt{n}\frac{\bar{x}_n - \mu}{s_x}$$

The sample standard deviation, s_x, is also a random variable and changes from sample to sample. Yesterday in lab we talked about a distribution which arises from the ratio of certain random variables.

$$\sqrt{n}\frac{\bar{x}_n - \mu}{s_x} \rightarrow t_{n-1}$$

Turns out, when we plug in s_x for σ, we end up with a T distribution with $n-1$ degrees of freedom

T Distribution

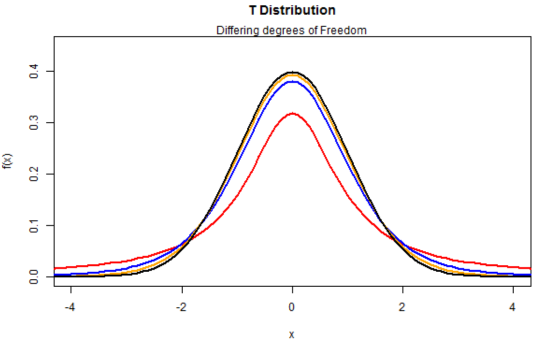
The T Distribution was developed by William Gosset while working at the Guiness Brewery in Dublin, Ireland.

Figure : William Gosset

In particular, Gosset was interested in estimating the "degree of saccharine" in malt extract

T Distribution

The T distribution looks very similar to the Normal, but gives more weight to "extreme" observations



As the degrees of freedom increase, the distribution looks more and more like a normal distribution

Normal Probabilities

In particular-

$$
\begin{aligned}
.95 &= P\left(-t^\star_{n-1} \leq \sqrt{n}\frac{\bar{x}_n - \mu}{s_x} \leq t^\star_{n-1}\right) \\
&= P\left(-t^\star_{n-1}\frac{s_x}{\sqrt{n}} \leq \bar{x}_n - \mu \leq t^\star_{n-1}\frac{s_x}{\sqrt{n}}\right) \\
&= P\left(\bar{x} - t^\star_{n-1}\frac{s_x}{\sqrt{n}} \leq \mu \leq \bar{x} + t^\star_{n-1}\frac{s_x}{\sqrt{n}}\right)
\end{aligned} \tag{1}
$$

Normal Probabilities

In particular-

$$
\begin{aligned}
.95 &= P\left(-t^\star_{n-1} \leq \sqrt{n}\frac{\bar{x}_n - \mu}{s_x} \leq t^\star_{n-1}\right) \\
&= P\left(-t^\star_{n-1}\frac{s_x}{\sqrt{n}} \leq \bar{x}_n - \mu \leq t^\star_{n-1}\frac{s_x}{\sqrt{n}}\right) \\
&= P\left(\bar{x} - t^\star_{n-1}\frac{s_x}{\sqrt{n}} \leq \mu \leq \bar{x} + t^\star_{n-1}\frac{s_x}{\sqrt{n}}\right)
\end{aligned} \tag{1}
$$

So there is a 95% chance that when I form sample a $\bar{x}$ and form the interval $(\bar{x} - t^\star_{n-1}\frac{s_x}{\sqrt{n}}, \bar{x} + t^\star_{n-1}\frac{s_x}{\sqrt{n}})$, it will contain the true parameter μ

Example

Confidence Intervals in general

Hopefully you've picked up on a general pattern for Confidence intervals

$$\text{point estimate} \pm \text{multiplier} \times \text{SE}$$

Type	Point Estimate	Multiplier	Standard Error
Single Proportion	$\hat{p}$	standard normal	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Diff in Proportions	$\hat{p}_1 - \hat{p}_2$	standard normal	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
Mean	$\bar{x}$	T distribution	$\frac{s_x}{n}$

Means of paired differences

Now suppose that I want to estimate the average weight loss after using a specific work out regimen for 12 weeks. I record the weight of each individual before and after the program. For each individual, I denote the difference in weights as d_i.

Means of paired differences

Now suppose that I want to estimate the average weight loss after using a specific work out regimen for 12 weeks. I record the weight of each individual before and after the program. For each individual, I denote the difference in weights as d_i.

I can still form confidence intervals for the mean of the differences in the same exact way, except now the differences d_i are my observations.

$$\bar{d} \pm t_{n-1} \frac{s_d}{\sqrt{n}}$$

Difference in Means (independent samples)

Suppose I am interested in comparing the average IQ of students at UW agains the average IQ of students at WSU. My parameter of interest would be

$$\mu_{UW} - \mu_{WSU}$$

Can I create a confidence interval for the difference given a sample of UW students and WSU students? In this case, the IQ of a randomly selected UW student is independent of the IQ of a randomly selected WSU student and it doesn't quite make sense to pair two randomly selected students together and take the difference.

Difference in Means (independent samples)

For two populations, if I am interested in estimating

$$\mu_1 - \mu_2$$

I can use the point estimate of

$$\bar{x}_1 - \bar{x}_2$$

which has a mean of $\mu_1 - \mu_2$. When they are independent, the variance of the point estimate is $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$, so we use a standard error of

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Difference in Means (independent samples)

So I can form a confidence interval

$$(\bar{x}_1 - \bar{x}_2) \pm t_{df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Difference in Means (independent samples)

So I can form a confidence interval

$$(\bar{x}_1 - \bar{x}_2) \pm t_{df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The degrees of freedom we should be using can get complicated. For this class, we will simply use $df = \min(n_1 - 1, n_2 - 1)$.

Difference in Means (independent samples)

So I can form a confidence interval

$$(\bar{x}_1 - \bar{x}_2) \pm t_{df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Difference in Means (independent samples)

So I can form a confidence interval

$$(\bar{x}_1 - \bar{x}_2) \pm t_{df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The degrees of freedom we should be using can get complicated. For this class, we will simply use $df = \min(n_1 - 1, n_2 - 1)$.

Difference in Means (independent samples)

So I can form a confidence interval

$$(\bar{x}_1 - \bar{x}_2) \pm t_{df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Difference in Means (independent samples)

So I can form a confidence interval

$$(\bar{x}_1 - \bar{x}_2) \pm t_{df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The degrees of freedom we should be using can get complicated. For this class, we will simply use $df = \min(n_1 - 1, n_2 - 1)$.

Difference in Means (independent samples)

So I can form a confidence interval

$$(\bar{x}_1 - \bar{x}_2) \pm t_{df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Difference in Means (independent samples)

So I can form a confidence interval

$$(\bar{x}_1 - \bar{x}_2) \pm t_{df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The degrees of freedom we should be using can get complicated. For this class, we will simply use $df = \min(n_1 - 1, n_2 - 1)$.

Example