**Colorado Technical University**

 **Course:** MATH366 – Probability and Statistics

#### Unit 10 Part 19 Readings: Confidence Intervals for t and p

**What if you have a sample size smaller than 20???**

You must use a different (bigger) critical value

You will have a smaller interval if you have a larger value for n

**Law of Large Numbers** -if you take a larger and larger sample, your sample

statistic will become closer and closer to the real population parameter

We don't generally take repeated samples ourselves - we just take one sample and

assume it is from a population of samples that have these characteristics

We use this principal to test hypotheses about population parameters being within a

certain range

**Confidence intervals when you aren’t sure you have a normal distribution**

If you have samples from a normally-distributed population, no matter how small your

sample is, it’s values will be normally-distributed

If you have a “large-enough” sample size, irrespective of the population’s distribution, the

samples will be normally-distributed

BUT…

What if you have neither of these?

A very clever man W.S. Gossett (who worked for the Guinness brewery) was working with

sensory testing (flavor) of the vats of stout

Only a limited number of trained testers were available (about 5-7)

It seemed unlikely that flavor characteristics would follow a normal distribution (but might be

close)

Mr. Gossett invented a new continuous probability distribution for small samples from close-

to-normal populations

He called this the “t” (always small letter) distribution because he was using it to compare

two groups (“t” for two): the standard flavor and the flavor for a particular vat

Because he worked for a company (and companies don’t like their employees publishing

things) he published his ideas under the pseudonym “A Student”

So this distribution is called “**Student’s t-Distribution**”

The t distributions is wide (has thicker tails) than a standard normal distribution

The thick tails ensure that the confidence intervals are wider than those using a standard

normal distribution (and are better at including the population mean)

The formula for t:

t = $\frac{\overbar{x}-μ}{s/\sqrt{n}}$

depends on the sample size n

The sample distribution of t = $\frac{\overbar{x}-μ}{s/\sqrt{n}}$ is a t-distribution with n − 1 “degrees of freedom”

Written: tn-1

The **degrees of freedom** (df) is a measure of how well s estimates σ

The larger the degrees of freedom, the better σ is estimated

We use n-1 rather than n degrees of freedom as a small penalty for reusing the same

dataset twice: once to calculate the mean $\overbar{x}$ (estimating the population mean μ) and a second time to calculate the sample standard deviation s (estimating the population standard deviation 𝛔)

95% confidence interval for the mean: x̄ – tn-1 s/$\sqrt{n}$ ≤ μ ≤ $\overbar{x}$x̄ + tn-1 s/$\sqrt{n}$

The “tn-1” in the equations is called the “**critical value**”

It comes from the t-distribution

Traditionally, people had to use tables of these values

You looked up the “n – 1” value and the % confidence you needed to find the t-value

Nowadays, we use Excel or a variety of other on-line software systems:

There are several “flavors” of “t” in

 Excel

For confidence intervals, use:

 **t.inv.2t**

 the two-tail (we’re using “minus” the

value, and “plus” the value, so it’s two

 tails)

Excel will want to know α (usually 5%)

 and the degrees of freedom (n-1):

 t.inv.2t(α,n-1)

**Confidence Intervals for “p”**

So far we have been doing confidence intervals for measurement data for population means

μ based on sample means x-bar

What about confidence intervals for population “p” values based on the “$\hat{p}$” of a sample?

$\hat{p}$ values are continuous – they can be ANY value from 0 to 1 on the real number line

These are easy, once you know the confidence intervals for the means

We’ll use the normal curve for proportions:



Notice the se for this type of data is: $\sqrt{\frac{pq}{n}}$

For normally distributed data or large sample sizes the margin of error will be 2se

The confidence interval will be:

$\hat{p}$ – 2$\sqrt{\frac{pq}{n}}$ ≤ p ≤ $\hat{p}$ + 2$\sqrt{\frac{pq}{n}}$

For small sample sizes from near-normal data, the confidence interval will be:

$\hat{p}$ – tn-1 $\sqrt{\frac{pq}{n}}$ ≤ p ≤ $\hat{p}$ + tn-1 $\sqrt{\frac{pq}{n}}$