Math 10 Handout Chapter 8 **Confidence Intervals**

Suppose we are interested in the following question:

- What is the mean number μ of hours of sleep De Anza students got last night?

We take our class as a sample. Suppose the sample average is 7.5 hours per night. Then, we say that $\bar{x} = 7.5$. The value 7.5 is called a **point estimate** of μ.

Using the CLT, we know $\bar{X} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$.

Another way to give the mean estimate is in the form of a **confidence interval**. A confidence interval has the from ($\bar{x} -$ margin of error, $\bar{x} +$ margin or error).

- Example: Suppose we want to estimate the average number of Giants home games a fan attends per season. A sample of 30 local fans is taken and the average number of games attended is calculated from the sample. We calculate that $\bar{x} = 10$. Suppose we know that the population standard deviation of games attended is $\sigma = 2$. Construct a 95% confidence interval for the true mean μ number of games attended.

Confidence level = CL = _____________

Associated with the CL is something called the α **-value**. The α -value is the probability that our confidence interval will **<u>not</u>** contain the true population mean. Thus:

$$\alpha = 1 - CL$$

So: α = _________________ Thus, $\frac{\alpha}{2}$ = ___________________

Graphically, we have:

Recall z-scores $z = \frac{\bar{x} - \mu_X}{\sigma_X}$. Since $\bar{X} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$, a bit of algebra (that we won't do) tells us z-scores for $\bar{X}$ have the normal distribution Z~N(0,1).

The confidence interval has the from ($\bar{x} -$ EBM, $\bar{x} +$ EBM) where EBM = $z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

The value $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ -critical value for the standard normal distribution. (Huh?) In other words, the area to the **<u>right</u>** of $z_{\frac{\alpha}{2}}$ is $\frac{\alpha}{2}$. We find this by using invNorm $\left(CL + \frac{\alpha}{2}, 0, 1\right)$ on our calculator.

So $z_{\frac{\alpha}{2}}$ = $Z_{\square}$ = invNorm(_______, _____, _____) = _____________

So EB = __________ and the confidence interval is (________, _________)

Let's draw a picture:

We are 95% confident that the true population mean is between ______________ and _____________.

In other words, we are 95% confident that the average number of Giants home games a fan attends per season is between ______________ and _____________.

Here's another (slightly easier) way to calculate of this:

Even though $\overline{X} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$, to create the confidence interval, we will use the distribution $\overline{X} \sim N\left(\bar{x}, \frac{\sigma_x}{\sqrt{n}}\right)$ (with $\bar{x}$ instead on μ_x since we don't know μ_x). To find the 95% confidence interval, we need to find the values that correspond to the percentiles for $\frac{\alpha}{2}$ and CL+$\frac{\alpha}{2}$.

In this example, $\overline{X} \sim$N(____, ______), CL = _______ and $\frac{\alpha}{2}$ = ________.

So we use the calculator commands invNorm(______, _____, ____) and invNorm(______, ____, ____)

to get the same confidence interval (_______, ________) and the picture:

- Example: Suppose we are interested in finding the mean price of an HDTV that is currently on the market. A random sample of 34 HDTV prices is taken that gives a sample mean of \$1585.00. If it is known that the population standard deviation of HDTV prices is \$390.00, find the 95% confidence interval for the true mean price of an HDTV.

Math 10 Handout Chapter 8 **Student-t Distribution & Confidence Intervals for Proportions**

Consider the previous problem:

- Test scores for statistics classes are normally distributed with unknown population mean but <u>population</u> standard deviation 3. A sample of 36 scores is taken that gives a sample mean of 68. Find a 90% confidence interval for the true population mean of test scores.

To solve this problem, we used: $\text{EB} = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

If the population standard deviation σ is **<u>not</u>** given, we will use the sample standard deviation s_x but will not use the $z_{\frac{\alpha}{2}}$ -score. Instead, we will use a ***t*-score** from a different distribution called the **Student-t distribution**. The Student-t distribution is denoted by $T \sim t_{df}$ where $df = n - 1$.

<u>5 Facts about the Student-t distribution:</u>

1. The graph of a Student-t distribution is similar to a normal curve.
2. Student-t tails have more probability under them because their spread is a little greater.
3. The Student-t distribution can only be used if the underlying population distribution is close to normal (or at least large and bell shaped). Also, we'll only use the Student-t distribution if the population mean and standard deviation are unknown.
4. $\text{EB} = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$
5. The Student-t distribution was discovered by a beer maker named William Gosset who worked for the Guinness brewery in Dublin, Ireland.

EXAMPLE

- Suppose we do a study on acupuncture to determine how effective it is in relieving pain. We measure sensory rates for 15 subjects. The results are given below. Use the sample data to construct a 95% confidence interval for the population mean sensory rate. Assume that the distribution of sensory rates is normal.

  8.6, 9.4, 7.9, 6.8, 8.3, 7.3, 9.2, 9.6, 8.7, 11.4, 10.3, 5.4, 8.1, 5.5, 6.9