Probability Plots

Scientists and engineers often work with data that can be thought of as a random sample from some population. In many cases, it is important to determine the probability distribution that approximately describes the population.

More often than not, the only way to determine an appropriate distribution is to examine the sample to find a sample distribution that fits.

Probability plots are a good way to determine an appropriate distribution.

Here is the idea (the computer does this for us):
• Suppose we have a random sample $X_1, \ldots, X_n$.
• We first arrange the data in ascending order. Then assign a evenly spaced values between 0 and 1 to each $X_i$. There are several acceptable ways to this; the simplest is to assign the value $(i - 0.5)/n$ to $X_i$.
• It then calculates the quantile ($Q_i$) corresponding to that number from the distribution of interest. Then it plots each $(X_i, Q_i)$. If this plot is a reasonably straight line then you may conclude that the sample came from the distribution that we used to find quantiles.

Most of the time, the probability plot that we are interested in is a Normal Probability Plot—Normal Quantile Plot in Minitab.

How do we read a Normal Quantile Plot?
If the points lie close to a straight line, the data are reasonably Normal.

• Curves indicate skewness in the data.
• Outliers appear as points far away from the overall pattern.
• Granularity, or a stepladder pattern, indicates limited precision of the measurements.
Give an interpretation for each of the following Normal Probability Plots.

1. Newcomb’s Data with two outliers
2. Survival Times (days) of guinea pigs in a medical experiment.

![Normal Probability Plot for C1](image1)

- ML Estimates
- Mean: 141.847
- StDev: 108.448
- Goodness of Fit
- AD*: 7.603

3. The final exam scores in STA 2122 in summer 2002.

![Normal Probability Plot for Exam](image2)

- ML Estimates
- Mean: 70.2573
- StDev: 13.7255
- Goodness of Fit
- AD*: 1.141
Simulation in Statistics

Sampling Variability -
Each random sample will give different values of the statistic. However, if we look at the statistics resulting from many random samples, there should be a pattern to them.

Random phenomenon-
We cannot predict the next outcome, but a regular and predictable pattern emerges in the long run. This pattern follows the laws of probability.

Simulations are useful for "pretending" we are performing an experiment a very large number of times.

Ex. You are a reliability engineer who has been ask to approximate the lifetime of several different systems with multiple components and to determine which is best. You know the individual lifetimes of the components and know that the components act independently of each other. For the following systems, find the average lifetime, create a histogram of the lifetimes, find the percentage of systems that last longer than 5 months.

Individual Components Lifetimes:

P ~ Exponential with mean equal to 1 month
R ~ Exponential with mean equal to 5 months

System 1

You have two components P and R. These components are in series with each other. Both have to work for the system to work. The system is drawn below.

```
P    R
```
Computing Instructions for System 1

These are the computing instructions for simulating the lifetimes for system 1. All other systems lifetimes can be approximate using similar instructions.

1. Enable the Commands in Minitab. Click on the session window. Go to Editor and Click on Enable Commands. (This will allow you to type in commands in the session window and will cause a history of the commands that you use to be recorded in the session window.)
2. Generate N random numbers that are Exponential with mean 1 and put them into C1. These are your simulated P values, P*.
3. Generate N random numbers that are Exponential with mean 5 and put them into C2. These are your simulated R values, R*.
4. Determine how long the system lasts by finding L* the minimum of P* and R*. Use Calc → Calculator and find the RMIN(C1, C2) and store the value in C3.
5. Find the mean by using the descriptive statistics option for variable L*.
6. Find the histogram by going to Graph → Histogram.
7. Determine how many systems last longer than 5 months by using the following two commands.
   a. Go to the session window. Enter
   b. MTB> let c4 = c3>5
   c. MTB> tally c4
Chapter 4 Section 10 and 11

Populations have ________________.

Samples have ________________.

Probability

• We know the parameter

• Take repeated samples.

• Study the behavior of the statistic

Inferential Statistics (the rest of the semester)

• We don’t know the parameter

• We want to estimate it

• Take one sample, compute the statistic

• Use Probability results to determine how close our statistic is likely to be to the unknown parameter
**Sampling Distribution of a Statistic:** The distribution of values taken by the statistic in all possible samples of the same size, from the same population. The idea of the sampling distribution is to see the pattern that emerges when we take repeated samples, and compute a statistic from each one of them.

Ex. How much does the sample mean change from sample to sample? What is the sampling distribution of the sample mean number of dots on the top of dice when \( n = 1, n = 2, n = 5, n = 20 \)?

1. If \( X \) is the number on the up-face of ONE rolled die, what is:
   a) the shape of the distribution of \( X \)? ________________
   b) the mean of the distribution of \( X \)? ________________

2. Simulations with Dice: The table below directs you to roll a die a certain number of times, \( n \). Each student will roll the die \( n \) times and record the number they got on each roll. Then you will compute the average of the numbers in your sample: divide the sum by the number of times the die was rolled (\( n \)). This is only one possible value of \( \bar{x} \) - to see the distribution, each student in the class will report their average and they will all be plotted on the board. You will sketch that plot on the space provided in the table.

<table>
<thead>
<tr>
<th>sample size (number of rolls)</th>
<th>Results of your die</th>
<th>Average for your sample</th>
<th>Sketch all the averages for the whole class, including the scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n=2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n=5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Now, use the computer to sketch the distribution of the sample means.

<table>
<thead>
<tr>
<th>sample size (number of rolls)</th>
<th>Sketch all the averages for the whole class, including the scale if the class had a 1,000 people</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=1</td>
<td></td>
</tr>
<tr>
<td>n=2</td>
<td></td>
</tr>
<tr>
<td>n=20</td>
<td></td>
</tr>
</tbody>
</table>

**Sampling Distribution**

- Suppose $\mu$ is known (Probability)
- Take a sample, compute $\overline{X}$.
- $X$-bar is (most likely) not exactly the same as $\mu$
- But if the sample was random and representative, $x$-bar should be close to $\mu$
- Any other sample would give a slightly different value of $x$-bar
- If each student got their sample and computed $x$-bar, a plot of all those $x$-bars should show most values close to the true $\mu$, and less values the farther away we move from $\mu$ -- Bell Shape
Example: Consider the distribution of weights of 20 year old women. Suppose this distribution is known to be Normal with mean = 130 and standard deviation equal to 10.

a.) Sketch this distribution. $x =$ the weight of each woman.

b.) Now consider the distribution of the average weight of 5 women at a time.
$\bullet =$ average weight of each sample.

c.) Now consider the distribution of the average weight of 30 women at a time.
$\bullet =$ average weight of each sample
What can we tell about the Sampling Distribution of x-bar?

- Mean of the Sampling Distribution of x-bar

- Standard Deviation of the Sampling Distribution of x-bar (Note: This is called the standard error.)

- Shape of the Sampling Distribution of x-bar

We need to formalize these results – Central Limit Theorem

**Central Limit Theorem**

Take a SRS of size \( n \) from a population having mean \( \mu \) and standard deviation \( \sigma \). Then, for large \( n \), the distribution of x-bar, the sample mean, will be approximately Normal, with the mean \( \mu \), the same as the original population and the standard deviation \( \frac{\sigma}{\sqrt{n}} \).
How large does the sample size, n, have to be?

- It depends on the shape of the original population.

- If the population is Normal, the Sampling Distribution of x-bar will be Normal for any n.

- If the population is far from Normal, n=30 is large enough in most cases for the Sampling Distribution of x-bar to be considered Normal.

- In general, the closer to Normal (bell, shaped, symmetric and continuous) the original distribution is, the smaller n needs to be.

- And for any shape distribution, as n increases, the Sampling Distribution of x-bar will get closer to Normal.

<table>
<thead>
<tr>
<th>Original Population</th>
<th>Sampling Distribution of x-bar for n=2</th>
<th>Sampling Distribution of x-bar for n=5</th>
<th>Sampling Distribution of x-bar for n=25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bimodal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skewed Right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sampling Distribution of the Sample Proportion

• In this chapter, we are also interested in the **proportion of the population** with a particular characteristic (not means or averages).

• Examples:

• The response variable for each experimental unit can be thought of as a **Yes/No** answer.

• We have used the **sample proportion** to estimate the true but unknown **population proportion** of successes:

\[ \hat{p} = \frac{x}{n} = \frac{\text{number of successes}}{\text{number of samples}} \]

Questions

• What is the distribution of X? __________________________

• What is the mean of X? _______________________________

• What is the standard deviation of X?

• What is the mean of the sampling distribution of the sample proportion? ______

• What is the standard deviation of the sampling distribution of the sample proportion?

• If you have at least 10 failures and 10 successes and you can assume that you have independent/identical trials (the process needs to be a stable), the sampling distribution of the sample proportion is roughly Normal \( \left( p, \sqrt{\frac{p(1-p)}{n}} \right) \).
Ex. Explore these concepts by looking at the following applet and answering the following questions.


a. How good is the normal distribution at approximating the binomial distribution at n=12 and p=0.5?

b. How good is the normal distribution at approximating the binomial distribution at n=12 and p = 0.9?

c. How good is the normal distribution at approximating the binomial distribution at n=12 at p=0.1?

d. How good is the normal distribution at approximating the binomial distribution at n=100 p = 0.1?

e. When is the normal a good approximation of the normal distribution?
Problems with the Sampling Distribution of the Sample Mean and the Sample Proportion

\( \bar{X} \sim \text{________________} \) if n is large or \( X \sim \text{Normal} \)

\( \bar{X} \sim \text{________________} \) if n is small and X is not Normally distributed.

\( \hat{p} \sim \text{________________} \) if \( np > 10 \) and \( n(1-p) > 10 \)

**For the following examples:** Identify the type of problem (sample mean or sample proportion) and determine if the sampling distribution of the statistic is approximately Normal before computing the probabilities.

1. The distribution of lawyer’s salaries at a firm has mean of $90,000 and standard deviation of $50,000. What is the probability that the average salary of lawyers of a random sample of five lawyers from this firm is less than $100,000?
   a. Is this problem about the sample proportion or about the sample mean?
   b. What is the sampling distribution of the statistic?
   c. Can we answer the question posed? If so, answer the question, if not, explain why not.

2. In 2002, the Global Social Survey asked 18-22 year olds, how many times they frequented a bar in the past month. 19.3% said that they had been to a bar only once in the past month. Suppose that this sample is representative of all 18-22 year olds. If 500 students were sampled at the University of Florida, what is the probability that at least 25% of the sample has been to a bar only once in the past month.
   a. Is this problem about the sample proportion or about the sample mean?
   b. What is the sampling distribution of the statistic?
   c. Can we answer the question posed? If so, answer the question, if not, explain why not.
3. X= number of accidents per week at a certain intersection. The distribution of X has a mean of 2.2 accidents, with a standard deviation of 1.4 accidents.

   a. Can the random variable X have a Binomial distribution?

   b. Can the random variable X have a Normal Distribution?

   c. Let $\bar{X}$ = the average number of accidents per week in one year. Find the sampling distribution.

   d. Find the probability that $x$-bar is less than 2.

   e. Find the probability that there are less than 100 accidents in one year.

4. In 1998, the Global Social Survey found that 75% of 18-22 year olds ate out last week. Let’s suppose that this is representative of all 18-22 year olds. Suppose that a random sample of 10 18-22 year old was taken in Gainesville, Florida. What are the chances that over half of the sample ate out last week?

   a. Is this problem about the sample proportion or about the sample mean?

   b. What is the sampling distribution of the statistic?

   c. Can we answer the question posed? If so, answer the question, if not, explain why not?
Chapter 5 – Confidence Intervals  
Section 5.1 Large Sample Confidence Intervals for a Population Mean

Idea Behind a Confidence Interval
- We don’t know \( \mu \), but we want to estimate it
- We could use the point estimate for \( \mu \), \( \bar{x} \).
- We know that \( \mu \) does not equal \( \bar{x} \).
- It would be better if we had an interval in which \( \mu \) would exist 95% of the time.

A confidence interval for a parameter is a data based interval of numbers likely to contain the parameter with a stated probability-based confidence.

Ex.: Suppose that we are trying to determine the true mean amount of daily sulfur dioxide emissions from a Nuclear Power Plant. We have taken a sample and \( n = 50 \), \( \bar{X} = 18.86 \).
For some reason we know that \( \sigma^2 = 30.77 \).

a. What is the sampling distribution of the sample mean?

b. If we assume \( \sigma^2 = 30.77 \)g, where will we find the central 95% of the all sample means?

c. So 95% of all sample means will be lucky enough to be within _____________ grams of the true mean.

d. We can flip this reasoning around to say that the true mean \( \mu \) will be within _____ grams of x-bar for 95% of all samples.

e. We can also determine this using mathematics.
f. Determine the 95% Confidence Interval for the true mean, $\mu$ with a known $\sigma$.

Confidence Interval for $\mu$:
Take a random sample of size $n$ from a population that has UNKNOWN mean $\mu$ and known $\sigma$. A (1-$\alpha$)*100% CI for $\mu$ is $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

where $\alpha/2$ is the area in the right hand tail.

How do you find $z$?

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>$\alpha$</th>
<th>$\alpha/2$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A Generally Applicable Large-n Confidence Interval for $\mu$

What happens if you don’t know the population standard deviation?

For large $n$, you can replace $\sigma$ with $s$.

CI Equation: \[ \bar{x} \pm z \frac{s}{\sqrt{n}}. \]

Example: A random sample of 110 lightning flashes resulted in a sample average radar echo duration equal to 0.81 seconds and sample standard deviation equal to 0.34 seconds. Find the 90% confidence interval for the true mean echo duration.

Characteristics of the Confidence Interval

a.) What would happen if you increased $n$?

b.) What would happen if you increased the confidence level?

c.) Suppose that we conducted a survey of students and asked them how many hours they spent studying for one course per week. We then found the 95% confidence interval for the true mean number of hours studied per one week for one course to be (2.3, 5.4).

Could $\mu$ be 2.4?
Could $\mu$ be 5.3?
Could $\mu$ be 2.21?

We don’t know if our interval is one of the lucky intervals to contain the true mean.
IMPORTANT NOTES:

- Confidence intervals are statements about the POPULATION MEAN, not about the sample mean or about individuals.

- We can only talk about “Probability” BEFORE we take the sample. After that, we talk about “Confidence”.

Determine what is wrong (if anything) with each interpretation of the confidence interval for the mean number of hours students study.

1. We are 95% confident that the true average number of hours students spend studying for one course in one week is between 2.3 and 5.4 hours.

2. We are 95% confident that all students study between 2.3 and 5.4 hours for one course in one week.

3. The probability that a student studies between 2.3 and 5.4 hours for one course in one week is 0.95.

4. The probability that the true mean, mu, is in the 95% confidence interval is 0.95.

5. We are 95% confident that the sample mean number of hours that students study is between 2.3 and 5.4 hours for one course in one week.
Determining Sample Size

We can find out what the sample size needs to be to have a desired margin of error at a certain level of confidence.

Example
a.) Suppose that we want to determine the average breaking strengths of fibers produced by Machine A. We want to find a 95% confidence interval. Because we have worked with this type of fiber previously, we know that $s = 5.1$ and we want to have a width of 2 mm. How many times should the sample be run through the machine?
Some Additional Comments Concerning Confidence Intervals

One Sided Confidence Bounds
Upper Bound: $\mu < \bar{x} + z_{\alpha} \frac{s}{\sqrt{n}}$

Lower Bound: $\mu > \bar{x} - z_{\alpha} \frac{s}{\sqrt{n}}$

How do you get $z^*$ for a one sided confidence interval?

To find the value of $z$, find the $z$ for the two sided confidence interval with $(1-2\alpha)\%$ confidence interval.

For example to find the $z$ for a one sided 90% ($\alpha = 0.10$) confidence interval, find the $z$ for the two sided 80% confidence interval.

Fill in the table for a one-sided confidence interval.

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td></td>
</tr>
</tbody>
</table>
Ex. The aluminum content of 36 hourly samples of recycled PET plastic from a recycling facility has a mean of 140 ppm and a standard deviation of 90 ppm. Find the 95% lower bound for the true mean average content of aluminum in samples of recycled PET plastic.

Interpretation:
6.1 Large Sample Test for a Population Mean

Two different types of Statistical Inference:

1. Confidence Intervals
   • Give a region that is likely to contain the parameter.
   • We have no preconceived notion of what \( \mu \) should be, simply want to estimate it.

2. Hypothesis Test
   • Assess the evidence for a claim about the population.
   • Someone proposes a value of \( \mu \). We take a sample to try to disprove it.
   • They have a very elaborate vocabulary, but the basic idea behind them is quite simple.

The reasoning of Tests of Significance

Example: The melting point of each of 16 samples of a certain brand of hydrogenated vegetable oil was determined, resulting in \( \bar{X} = 94.32 \). Assume that the distribution of melting points is normal with \( \sigma = 1.20 \). The melting point should be less than 95 degrees on average to achieve the “best” oil. Is the melting point significantly less than 95 degrees?

Solution: Pretend that the melting temperature is 95 degrees and determine how likely (or unlikely) it would be to get results as low as what you got in your sample.

• If we pretend that the melting temperature is 95 degrees, what is the distribution of the sample means from all possible samples?

• According to that distribution, what is the probability that a random sample would give a result as low as the one you observed?

• Considering that this is a very unlikely result, what can you conclude about the claim that the average melting temperature is 95 degrees?
<table>
<thead>
<tr>
<th>Elements of a Significance Test</th>
<th>In our example:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Null Hypothesis: Ho</strong></td>
<td></td>
</tr>
<tr>
<td>states the value we want to disprove</td>
<td></td>
</tr>
<tr>
<td><strong>Alternative Hypothesis: H₁</strong></td>
<td></td>
</tr>
<tr>
<td>states what we suspect is true</td>
<td></td>
</tr>
<tr>
<td>(This is denote as H₁ in the book.)</td>
<td></td>
</tr>
<tr>
<td><strong>Reference (null) distribution:</strong></td>
<td></td>
</tr>
<tr>
<td>the probability distribution describing the statistic; provided that the null hypothesis is in fact true</td>
<td></td>
</tr>
<tr>
<td><strong>Test Statistic: z-score</strong></td>
<td></td>
</tr>
<tr>
<td>summarizes the info from the sample</td>
<td></td>
</tr>
<tr>
<td><strong>p-value: &quot;corner&quot; area</strong></td>
<td></td>
</tr>
<tr>
<td>Probability that the test statistic will take on values at least as extreme as the one observed if Ho is true.</td>
<td></td>
</tr>
<tr>
<td>Small p-values support H₁.</td>
<td></td>
</tr>
</tbody>
</table>
More detail: Stating the Hypotheses

- The Null and Alternative Hypotheses are always statements about the unknown parameters, not the sample statistics.

- The Null Hypothesis always sets the parameter equal to the value we want to disprove.
  \[ H_0: \mu \leq # \quad \text{or} \quad H_0: \mu \geq # \quad \text{or} \quad H_0: \mu = # \]

- The Alternative Hypothesis can be either
  - one-sided - we are only interested in deviations from Ho in only one direction
    \[ H_1: \mu < # \quad \text{or} \quad H_1: \mu > # \]
  - two-sided - we simply want to prove the parameter is different from the number claimed in Ho
    \[ H_1: \mu \neq # \]

- We always determine the sign of the Alternative Hypothesis from the story - not from the data. Looking at the data for this is cheating.
More details:  P-value and Statistical Significance

- The p-value is a **probability**, and thus a number between 0 and 1.

- **P-value is the probability of getting an outcome as extreme or more extreme than the actually observed outcome, if in fact, Ho were true.**

- Extreme here means too far from the value specified in Ho - too far in the direction specified by $H_1$.

- **Test Statistic close to 0 ⇒ big p-value**

- **Test Statistic far from 0 ⇒ small p-value**

- The p-value represents the strength of the evidence for the Null Hypothesis.

- **The smaller the p-value, the more evidence against the Null and for the Alternative Hypothesis (the one we are trying to prove).**
6.2 Drawing Conclusions from the Results of Hypothesis Tests

There are two ways to interpret the p-value.

1.) The p-value can be compared to a pre-specified significance level, $\alpha$.

$\alpha$ = the level of significance = the level of willingness to make an error.

$\alpha = P(\text{Rejecting the null hypothesis when the null hypothesis is true})$

Usual $\alpha$ levels are 0.10, 0.05, 0.01

If a result is significant at the $100\alpha\%$ level, we can say that the null hypothesis is “rejected at the level $100\alpha\%$”

2.) Look at the p-value and determine if it is small.

How small should the p-value be to prove $H_1$? The smaller the better

- **How small a p-value is convincing?** It depends on the situation: Does $H_0$ represents firmly established beliefs? Does going with $H_a$ cost a lot of money? Then you need a more convincing (smaller) p-value. You can simply report the p-value, and let someone else decide if it is convincing enough for him or her.

<table>
<thead>
<tr>
<th></th>
<th>p-value range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>very small p-values</td>
<td>p-value &lt; 0.001</td>
<td>This constitutes very strong evidence against $H_0$.</td>
</tr>
<tr>
<td>small p-values</td>
<td>p-value between 0.001 and 0.05</td>
<td>This constitutes strong evidence against $H_0$.</td>
</tr>
<tr>
<td>marginal p-values</td>
<td>p-value between 0.05 and 0.10</td>
<td>This constitutes marginal evidence against $H_0$.</td>
</tr>
<tr>
<td>non-small p-values</td>
<td>P-values greater than 0.10</td>
<td>No evidence against $H_0$.</td>
</tr>
</tbody>
</table>

There is never a time when you should say that the p-value gives you evidence for $H_0$.

Note: Although many people and books treat $\alpha = .05$ as a magical cutoff between significant and insignificant, there really is no practical difference between a p-value of .049 and one of .051.
In general:

- Small p-value ⇒ Ho must be wrong (or we got an unlucky sample)
  ⇒ Results are significant
  ⇒ Reject Ho

- Large p-value ⇒ Cannot prove Ho wrong
  ⇒ Results are NOT significant
  ⇒ Fail to Reject Ho

Examples:
- p-value = 0.03
- p-value = 0.0004
- p-value = 0.12
- p-value = 0.17
Five Steps to Conducting a Significance Test

1. State the null hypothesis.
2. State the alternative hypothesis.
3. State the test criteria. (Formula)
4. Calculate the test statistic.
5. Report the p-value.

Examples:

1. A large university uses thousands of light bulbs every year. The brand they use now has a mean life of 900 hours. Another manufacturer claims that their new light bulbs (that sell for the same price) lasts longer. The university tests 64 of these new light bulbs, and their average lifetime was 920 hours, with a known standard deviation of 80 hours. Do the new light bulbs have a significantly longer average lifetime?
6.2.3 General Applicable Large-n Significance Tests for $\mu$

If you don’t know $\sigma$ and you have a large sample size, you can replace $\sigma$ with $s$, to get the following test statistic.

$$ Z = \frac{\bar{x} - \mu}{s / \sqrt{n}} $$

2. Suppose that you work at a water treatment plant. The level of radioactivity allowed in the drinking water is 5 pCi/L. For the past 32 bi-hourly measurements, the level of radioactivity has a sample mean of 4.85 and a sample standard deviation of 0.75 pCi/L. Is the level of radioactivity significantly less than 5 pCi/L?
3. A machine fills boxes with 12 oz. of cereal. Thirty-one boxes receive a sample mean of 11.68 oz. of cereal with a sample standard deviation of 1.3 oz.

a.) Are the boxes average content significantly different from 12 oz.?

b.) Make a 95% CI for μ:

c.) Compare the results of the confidence intervals to those of the significance test.

Note: The results of a two-sided confidence interval and two sided hypothesis test will always agree at the same level of α.
Use and Abuse of Tests

Use of Tests

1.) Sometimes, there are legal guidelines that require significance testing at a prescribed level of alpha.
2.) There are some areas in Engineering where hypothesis test just make more sense, i.e. Acceptance Sampling.
3.) Really large p-values—tell engineers how inconclusive their data is. They need more data!!

Abuse of Tests

1. Statistical Significance is NOT the same as Practical Significance.

Statistical Significance says that we believe the differences seen in the sample can be extended to the population. But these differences can be quite small, and of no practical significance whatsoever. Very large samples can have this effect. What should we do? Plot the data to get a feel for it. Make a CI. It will tell you how big the difference is.

Example: A device called a “Pass Master” was given EPA approval for making a “statistically significant” decrease in the amount of gasoline used by a car. The manufacturer claimed that this new device decreased the amount of gasoline needed by as much as 4%. Suppose that X=the amount of gasoline used is normally distributed. Are the results practically significant? Assume $\alpha = 0.05$.

Suppose that you were testing the following hypothesis.

$$H_0: \mu = 33.5 \quad H_1: \mu > 33.5$$

Suppose that the sample mean is 34.84 and the standard deviation is 5.1.

a. If n equals 25, the test statistic equals 1.31 and the p-value equals 0.0951. What decision do you make? ______________

b. If n equals 1000, the test statistic equals 8.3 and the p-value equals 0.000. What decision do you make? ______________

c. The 95% confidence interval for $\mu$ is (32.84, 36.84). What decision do you make?

d. The 95% confidence interval for $\mu$ is (34.52, 35.156). What decision do you make?

e. Conclusion.
2. Beware of Searching for Significance

When gathering information about a lot of different groups, there will usually be one group with the highest average, and another one with the lowest average. However, unless you had planned in advance to compare those particular two groups, it would be misleading to apply the formal statistical inference procedures learned here. Those conclusions only work on random data - and picking the largest and the smallest to compare to each other does not qualify as random. If you notice something from a study and want to investigate it formally, gather new data with the specific purpose of testing for that effect.

WARNING LABELS for Confidence Intervals and Significance Tests

- The data must be a SRS from the population of interest. Bad data will give you useless results.
- Outliers can have a large effect on the sample mean, and that would also affect our conclusions. Outliers should be investigated, and if they cannot be fixed or removed, other statistical procedures that are resistant to outliers should be used. *
- These procedures also require either a large n or a Normal population. If your sample is quite small and outliers or great skew are present, they should not be used. Instead, statistical procedures that are resistant to outliers should be used. *

* These procedures will not be covered in this class. If you ever need to use them, consult a statistician.
5.3 and 6.3 Small Sample Confidence Intervals and Hypothesis Test for a Population Mean

- In chapter 6, we learned how to do CI’s and Sig Tests for $\mu$ using the Z table, based on:
  \[ \overline{X} \sim N \left( \mu, \frac{\sigma}{\sqrt{n}} \right) \]
  for large enough $n$, so $\frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$ has a $Z \sim N(0, 1)$ distribution.

- We made the unrealistic assumption that $\sigma$ was known. In practice, we substitute $s$ for $\sigma$
  \[ \frac{s}{\sqrt{n}} \]
  is the standard deviation of the distribution of $\overline{X}$

- $\frac{s}{\sqrt{n}}$ is called the standard error – it’s an estimator of $\frac{\sigma}{\sqrt{n}}$

- $\frac{\overline{x} - \mu}{s/\sqrt{n}}$ does NOT have a Normal distribution.

- Instead, if the X’s are Normal, it will have a t distribution with n-1 degrees of freedom.

**t-Distribution**

\[
f(t) = \frac{\Gamma \left( \frac{v+1}{2} \right)}{\Gamma \left( \frac{v}{2} \right) \sqrt{\pi v} \left( 1 + \frac{t^2}{v} \right)}^{-(v+1)/2}
\]
Examples: Find the critical values for $t$ for following situations:

1.) Find the $t$ that has probability of 0.95 to the left of it in a $t(5)$ distribution.

2.) Find the $t$ that has probability of 0.01 to the right of it in a $t(12)$ distribution.

3.) Find the $t$ that would be used in a 95% Confidence Interval for a sample of $n=10$.

4.) Find the $t$ that would be used in a 99% Confidence Interval for a sample of $n=25$.

5.) Find the probability that you are to the right of 2.7 for a sample of $n=25$.

6.) Find the probability that you are to the left of $-2.7$ for a sample of $n=25$. 
Small Sample Confidence Interval for the mean:

\[ \bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \]

where df = n-1

Small Sample Hypothesis Test for the mean:

\[ T = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \]

where df = n-1

Assumptions – These procedures give exact results if the original distribution was Normal and approximately correct for large n in other cases. The data must have been from a SRS.

Examples

1.) Engineers were interested in estimating the stability viscosity of asphalt specimens. They measured the asphalt at five random locations.

Data: 2,781  2,900  3,013  2,856  2,888

a. Discuss the normality or non-normality of the data.

b. Are the assumptions for a t confidence interval for the mean stability viscosity of asphalt specimens met?
c. What t critical point do you use for 90% confidence interval?

d. Determine a 90% confidence interval for the mean of the viscosity of the specimens.

e. Use Minitab to calculate a 90% confidence interval for the mean.

```
One-Sample T: C1
Var N Mean   StDev SEMean  90.0% CI
C1  5 2887.6 84.0  37.6 (2807.5,2967.7)
```

f. Give an interpretation.
2. A new type of fire prevention system was being tested. The system was designed so that it would take less than 25 seconds to activate. Use the data below to test this hypothesis. Make sure that you test the assumptions.

25  24  27  24  23  25  26  26  25  23  24  23  23  23  23  22
Some Guidelines for using the t procedures:

- **Always:** data must be a simple random sample from the population of interest to extend the conclusions.

- **When n > 30:**
  - t procedures are quite robust to departures from Normality in the original population
  - s is a good estimator of σ
  - t distribution gets close to Normal
  - makes very little difference to use Z or t
  - if df are not on table -
    - conservative approach: use next lowest
    - liberal approach: use Z table instead
    - easiest approach: use Minitab

- **When n < 30:**
  - t procedures are very sensitive to skewness or outliers in the original population
  - s might be far from σ
  - t distribution still far from Normal
  - need to use t table
  - need original population to be Normal
  - impossible to check population, we only have a small sample, so plot data and make sure it could have come from a Normal distribution perfect symmetry of the sample not important, but there should be no major outliers

- **If σ is given** with any sample size and sample is from a Normal Population
  - Use Z
5.7 and 6.8 Inference for the Mean of Paired Differences

- Two treatments are given to the same experimental unit or two very similar ones
- Examples:
  - 
  - 
- To do this, first compute the differences between the two treatments.
- Find $\bar{d}$ and the standard deviation is $s_d$.
- Statistics ($\bar{d}$) estimates $\mu_d$
- Standard error of the statistic $= \frac{s_d}{\sqrt{n}}$ where $n$ is the number of pairs.
- Sampling Distribution
  - $t$ with $n-1$ df if $n<30$
  - $z$ if $n>30$
- What are the assumptions of the test?

CI Formula:

Significance Test
Ex. Three scientists measured the resistances for seven different copper wires at 0 degrees C and at 21.8 degrees C. Is there a difference between the resistances of the wires at the two temperatures?

<table>
<thead>
<tr>
<th>0 degrees C</th>
<th>21.8 degrees C</th>
<th>Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.52</td>
<td>1.72</td>
<td></td>
</tr>
<tr>
<td>1.44</td>
<td>1.56</td>
<td></td>
</tr>
<tr>
<td>1.52</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>1.52</td>
<td>1.64</td>
<td></td>
</tr>
<tr>
<td>1.56</td>
<td>1.69</td>
<td></td>
</tr>
<tr>
<td>1.49</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td>1.56</td>
<td>1.72</td>
<td></td>
</tr>
</tbody>
</table>

a.) Make a quick plot of the differences. Is it reasonable to assume that they are possibly normally distributed?

b.) b. Conduct a significance test to if there is a difference between the resistances at the two temperatures.

c.) Compute a 90% CI for the true mean difference between the resistances at the two temperatures.
## Interpreting Confidence Intervals with Two Groups (Group 1 - Group 2)

<table>
<thead>
<tr>
<th>Type of Interval</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-a, +b)</td>
<td>We have no evidence that there is a difference between group 1 and group 2. If group 1’s mean is higher, it is by $b$ units. If group 2’s mean is higher it is by $a$ units.</td>
</tr>
<tr>
<td>(-a, -b)</td>
<td>We are ____ % confident that the true mean of group 2 is higher than the group 1 by as little as $b$ units to as much as $a$ units.</td>
</tr>
<tr>
<td>(a, b)</td>
<td>We are _____ % confident that the true mean of group 1 is higher than the true mean of group 2 by as little as $a$ units to as much as $b$ units.</td>
</tr>
</tbody>
</table>

d. Interpret the Interval above.
5.4, 5.6, 6.5 and 6.7 Large and Small Sample Comparisons of Two Means
(Based on Independent Samples)

Two Independent Mean Test

- Comparative studies.
- Experimental units are randomly allocated to receive one of the two treatments, or selected at random from two populations.
- We want to know if there is a difference in the population means.
- First, find the average of responses for each group, $\bar{x}$ and $\bar{y}$, and their standard deviations $s_x$ and $s_y$.
- Examples

- Parameter of interest: $\mu_x - \mu_y$
- Statistic Used: $\bar{x} - \bar{y}$
- Standard Deviation of Statistic: $\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$
- Theoretical Test Statistic:
However, we don’t have $\sigma_1$ or $\sigma_2$. Depending on the sample size and the distribution of the population, there are several possible ways of creating a test statistic to test for the difference of two means.

<table>
<thead>
<tr>
<th>If both samples have an $n$ greater than or equal to 30, you can estimate $\sigma_1$ or $\sigma_2$, with $s_1$ and $s_2$.</th>
<th>If the one of the samples has an $n$ smaller than 30 and you can assume that $\sigma_1$ equals $\sigma_2$, you can pool the variances to estimate $\sigma$ and the test statistic is a t with $n_1 + n_2 - 2$ degrees of freedom. (homoscedasticity)</th>
<th>If one the samples has an $n$ smaller than 30 and you can not assume that $\sigma_1$ equals $\sigma_2$, use $s_1$ and $s_2$ in the test statistic. The test statistic is t with $\nu$ degrees of freedom. <strong>(heteroscedasticity)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{s \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$</td>
<td>$\hat{T} = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{s_x^2 + s_y^2}}$</td>
<td>$\hat{\nu} = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{s_x^4}{(n_x - 1)n_x^2} + \frac{s_y^4}{(n_y - 1)n_y^2}}$</td>
</tr>
<tr>
<td>Where $s^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* If the ratio of standard deviations (larger over smaller) is at least 2, the assumption of unequal variances is violated.

**Always round $\nu$ down to the nearest whole integer.
### Confidence Interval for the Difference of Two Independent Means:

<table>
<thead>
<tr>
<th>Situation</th>
<th>Formula</th>
<th>Assumptions</th>
</tr>
</thead>
</table>
| If both samples have an n greater than or equal to 30, you can estimate \( \sigma_1 \) or \( \sigma_2 \), with \( s_1 \) and \( s_2 \). | \((\bar{x} - \bar{y}) \pm t_{n_1+n_2-2, \nu} \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \) | - SRS  
  - That the populations could have come from Normal Distributions.  
  - Variances are equal. |
| If the one of the samples has an n smaller than 30 and you can assume that \( \sigma_1 \) equals \( \sigma_2 \), you can pool the variances to estimate \( \sigma \) and the test statistic is a t with \( n_1 + n_2 - 2 \) degrees of freedom. | \((\bar{x} - \bar{y}) \pm t_{\nu, \alpha/2} \sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}} \) | (heteroscedasticity) |
| If one the samples has an n smaller than 30 and you cannot assume that \( \sigma_1 \) equals \( \sigma_2 \), use \( s_1 \) and \( s_2 \) in the test statistic. The test statistic is t with \( \nu \) degrees of freedom. | \( \nu = \frac{s_ar{x}^4 + s_ar{y}^4}{(n_x - 1)n_x^2 + (n_y - 1)n_y^2} \) |  
  - SRS  
  - That the populations could have come from Normal Distributions. |
For the following problems, determine what would be the null and alternative hypothesis, test statistic, and sampling distribution of the test statistic.

a.) Two supplier’s manufacturer a plastic gear used in a laser printer. The impact strength of these gears measured in foot-pounds is an important characteristic. The strength of these gears from both suppliers was tested. Are supplier 2’s gears significantly stronger than supplier 1’s?


Descriptive Statistics: Supplier 1, Supplier 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>SE Mean</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier 1</td>
<td>40</td>
<td>0</td>
<td>308.12</td>
<td>4.43</td>
<td>28.01</td>
<td>239.03</td>
<td>287.38</td>
<td>307.29</td>
<td>329.37</td>
</tr>
<tr>
<td>Supplier 2</td>
<td>31</td>
<td>0</td>
<td>315.93</td>
<td>4.24</td>
<td>23.63</td>
<td>264.34</td>
<td>298.19</td>
<td>317.04</td>
<td>333.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier 1</td>
<td>362.88</td>
</tr>
<tr>
<td>Supplier 2</td>
<td>355.81</td>
</tr>
</tbody>
</table>
Interpret the hypothesis test and the confidence intervals. Remember to also check the assumptions. Use a 0.05 level of significance.

**Descriptive Statistics: Supplier 1, Supplier 2**

![Boxplot of Supplier 1, Supplier 2](image)

**Two-Sample T-Test and CI: Supplier 1, Supplier 2**

Two-sample T for Supplier 1 vs Supplier 2

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier 1</td>
<td>40</td>
<td>308.1</td>
<td>28.0</td>
<td>4.4</td>
</tr>
<tr>
<td>Supplier 2</td>
<td>31</td>
<td>315.9</td>
<td>23.6</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Difference = mu (Supplier 1) - mu (Supplier 2)

<table>
<thead>
<tr>
<th>Estimate for difference:</th>
<th>-7.80916</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% upper bound for difference:</td>
<td>2.41986</td>
</tr>
</tbody>
</table>

T-Test of difference = 0 (vs <): T-Value = -1.27  P-Value = 0.104  DF = 68

95% CI for difference: (-20.04953, 4.43120)
b.) A study is designed to compare gas mileage with a fuel additive to gas mileage without the additive. We wish to prove that using the additive increases gas mileage. A group of 10 Ford Mustangs are randomly divided into two groups and the gas mileage is recorded for one tank of gas.

Treatment 1 (with Additive): 26.3, 27.4, 25.1, 26.8, 27.1

Treatment 2 (without Additive): 24.5, 25.4, 23.7, 25.9, 25.7

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additive</td>
<td>5</td>
<td>26.540</td>
<td>26.800</td>
<td>26.540</td>
<td>0.902</td>
<td>0.403</td>
</tr>
<tr>
<td>No_Addit</td>
<td>5</td>
<td>25.040</td>
<td>25.400</td>
<td>25.040</td>
<td>0.921</td>
<td>0.412</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additive</td>
<td>25.100</td>
<td>27.400</td>
<td>25.700</td>
<td>27.250</td>
</tr>
<tr>
<td>No_Addit</td>
<td>23.700</td>
<td>25.900</td>
<td>24.100</td>
<td>25.800</td>
</tr>
</tbody>
</table>
Interpret the hypothesis test and the confidence intervals. Remember to also check the assumptions.

**Descriptive Statistics: Additive, No_Additive**

Dotplot for Additive-No_Additive

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additive</td>
<td>5</td>
<td>26.540</td>
<td>0.902</td>
<td>0.40</td>
</tr>
<tr>
<td>w/o addi</td>
<td>5</td>
<td>25.040</td>
<td>0.921</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Difference = mu Additive - mu w/o additive
Estimate for difference: 1.500
95% CI for difference: (0.171, 2.829)
T-Test of difference = 0 (vs not =): T-Value = 2.60 P-Value = 0.031 DF = 8
Both use Pooled StDev = 0.911

T-Test of difference = 0 (vs >): T-Value = 2.60 P-Value = 0.018 DF = 7
c.) Two different methods were used to make 1045 steel. The amount of manganese was measured for ten samples made from the two methods. Is there a difference in the amount of manganese for the two methods? (the units are points, or .01%).

Method 1: 87, 74, 78, 81, 78, 77, 84, 80, 85, 78

Method 2: 86, 86, 82, 87, 85, 84, 84, 82, 82, 85

Descriptive Statistics: Method 1, Method 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>10</td>
<td>80.20</td>
<td>79.00</td>
<td>80.13</td>
<td>4.05</td>
<td>1.28</td>
</tr>
<tr>
<td>Method 2</td>
<td>10</td>
<td>84.30</td>
<td>84.50</td>
<td>84.25</td>
<td>1.83</td>
<td>0.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>74.00</td>
<td>87.00</td>
<td>77.75</td>
<td>84.50</td>
</tr>
<tr>
<td>Method 2</td>
<td>82.00</td>
<td>87.00</td>
<td>82.00</td>
<td>86.00</td>
</tr>
</tbody>
</table>
Interpret the hypothesis test and the confidence intervals. Remember to also check the assumptions.

**Descriptive Statistics: Method 1, Method 2**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>10</td>
<td>80.20</td>
<td>79.00</td>
<td>80.13</td>
<td>4.05</td>
<td>1.28</td>
</tr>
<tr>
<td>Method 2</td>
<td>10</td>
<td>84.30</td>
<td>84.50</td>
<td>84.25</td>
<td>1.829</td>
<td>0.578</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>74.00</td>
<td>87.00</td>
<td>77.75</td>
<td>84.25</td>
</tr>
<tr>
<td>Method 2</td>
<td>82.000</td>
<td>87.000</td>
<td>82.000</td>
<td>86.000</td>
</tr>
</tbody>
</table>

**Dotplot for Method 1-Method 2**

![Dotplot](image)

**Two-Sample T-Test and CI: Method 1, Method 2**

Two-sample T for Method 1 vs Method 2

<table>
<thead>
<tr>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>10</td>
<td>80.20</td>
<td>4.05</td>
</tr>
<tr>
<td>Method 2</td>
<td>10</td>
<td>84.30</td>
<td>1.83</td>
</tr>
</tbody>
</table>

Difference = mu Method 1 - mu Method 2
Estimate for difference: -4.10
95% CI for difference: (-7.16, -1.04)
T-Test of difference = 0 (vs not =): T-Value = -2.92  P-Value = 0.013  DF = 12
5.2 and 6.3 Confidence Intervals and Hypothesis Test for the one sample proportion

Inference for a Single Proportion
- In this section, we are interested in the **proportion of the population** with a particular characteristic (not means or averages).
- Examples:
  - proportion of females in Law school
  - proportion of people who favor abortion
  - proportion of students older than 25
  - proportion of students in live sections getting an A in the class

- The response variable for each experimental unit can be thought of as a *Yes/No* answer.
- We have used the **sample proportion** to estimate the true but unknown **population proportion** of successes:
  \[ \hat{p} = \frac{x}{n} = \frac{\text{number of successes}}{\text{number of samples}} \]

- We have learned that, for large enough \( n \), \( \hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) \)

- Since \( p \) is unknown, we need to use an estimate of it to estimate the standard deviation.

Confidence Interval for a Population Proportion
We could use the Normal distribution to make CI for the true proportion of successes \( p \), based on \( \hat{p} \). Unfortunately, modern computer studies show that they can be quite inaccurate, even for large samples.

To get better results, we will modify our estimator by moving it slightly away from 0 and 1. We do this by pretending we had an extra two successes and two failures in our sample. This is called a Wilson estimate.

- **Wilson estimate:** \( \tilde{p} = \frac{X + 2}{n + 4} \)
- **Standard error of \( \tilde{p} \):** \( \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + 4}} \)
• **Confidence Interval for a Population Proportion:**

\[
\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n+4}}
\]

• We can use this formula when the sample size is 5 or more, and the confidence level is 90%, 95% or 99%.

**Significance Test for a Population Proportion**

Since the null hypothesis proposes a value of p (let’s call it \( p_0 \)), we can use that value as our estimator of \( p \) in the standard deviation.

• **Test Statistic**

\[
z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}
\]

• We can use this test when the population is at least 10 times larger than the sample (for stdev formula to work) and the expected number of successes (\( np_0 \)) and the expected number of failures (\( n(1- p_0) \)) under the null hypothesis are 10 or more each.
**Example:** The Harvard School of Public Health College Alcohol Study Survey surveys college students in about 200 colleges in 1993, 1997, and 1999. They asked the students demographic questions as well as questions about their drinking habits. They were especially interested in the binging habits of college students. The survey defines an drink as “12 oz bottle or can of beer, a 4 oz (120 mL) glass of wine, a 12 oz. (360mL) bottle or can of wine cooler, or a shot (1.25 oz or 37 mL) of liquor straight or in a mixed drink.” Binge Drinking is considered drinking 5 drinks in a row for males and 4 drinks in a row for females. This information is from the 2001 study.

Do the majority of male college students binge drink?

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binge Drinkers</td>
<td>1908</td>
<td>2854</td>
<td>4762</td>
</tr>
<tr>
<td>Non-Binge Drinkers</td>
<td>2017</td>
<td>4125</td>
<td>6142</td>
</tr>
<tr>
<td></td>
<td>3925</td>
<td>6979</td>
<td>N=10,904</td>
</tr>
</tbody>
</table>

Make a 95% CI for p and interpret.

Are the assumptions met for this problem?

Compare these results with the Minitab output.

**Test and CI for One Proportion**

Test of p = 0.5 vs p not = 0.5

<table>
<thead>
<tr>
<th>Sample</th>
<th>X</th>
<th>N</th>
<th>Sample p</th>
<th>95.0% CI</th>
<th>Z-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1908</td>
<td>3925</td>
<td>0.486115</td>
<td>(0.470478, 0.501751)</td>
<td>-1.74</td>
<td>0.082</td>
</tr>
</tbody>
</table>
Example: Suppose that you have some system that outputs a marble at 4 different locations.

There are four positions on the apparatus 10, 20, 30 and 40. Tests the hypothesis that the marble will fall into the fourth position less often than the other positions.

a.) Write a null and alternative hypothesis statement.
b.) Collect the data. (Notice: We stated the hypothesis statement before we collected the data.)

<table>
<thead>
<tr>
<th>Number of Marbles</th>
<th>Number of Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>None</td>
</tr>
<tr>
<td>25</td>
<td>10 points</td>
</tr>
<tr>
<td>22</td>
<td>20 points</td>
</tr>
<tr>
<td>26</td>
<td>30 points</td>
</tr>
<tr>
<td>12</td>
<td>40 points</td>
</tr>
</tbody>
</table>

c.) Conduct a significance test to determine if the marble falls into the “40 point” position less often than the other positions.

d.) Are the assumptions met for this problem?

c.) Minitab: Compare the by-hand calculations with Minitab output.

**Test and CI for One Proportion**
Test of $p = 0.25$ vs $p < 0.25$

<table>
<thead>
<tr>
<th>Sample</th>
<th>X</th>
<th>N</th>
<th>Sample p</th>
<th>95.0% Upper Bound</th>
<th>Z-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>100</td>
<td>0.120000</td>
<td>0.173451</td>
<td>-3.00</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Sample Size Determination

- When planning a study, we can determine how large the sample needs to be to estimate the parameter within a given margin of error, with the desired confidence.
- Here, we will do the same for estimating a population proportion.
- Margin of error for a CI for \( p \) is:

\[
\Delta = \frac{z_{\alpha/2} \sqrt{\hat{p}(1 - \hat{p})}}{\sqrt{n}} = \frac{z_{\alpha/2} \sqrt{\hat{p}(1 - \hat{p})}}{n + 4}
\]

- This margin of error depends on \( \hat{p} \), which we won't know until we conduct the actual study (remember - we are doing this in the planning stage of the study).
- **If you are no clue what the value of \( p \) is**, you can use a formula that will ensure that you get the confidence and margin of error that you want. This formula is the result of plugging \( \hat{p} = 0.5 \) into the above equation and solving for \( n \). A \( \hat{p} = 0.5 \) will give you the largest sample size over any other size \( \hat{p} \).

\[
n = \left(\frac{z_{\alpha/2}}{2\Delta}\right)^2 - 4
\]

- **If we have a guess about \( \hat{p} \)**, we can plug that value into the below equation. This equation was found by solving for \( n \) of the equation at the top of the page.

\[
n = \frac{z_{\alpha/2}^2 \hat{p}(1 - \hat{p})}{\Delta^2} - 4
\]

- Remember to **round up to the next largest integer**.
Example: You want to know what proportion of UF students drink when they are underage. Find the number of students that should be sampled to estimate the true population proportion to within 0.02, with 95% confidence if you have no idea what the true proportion is.

Example: Compute the sample size again with a guess for the population proportion.
5.5 and 6.6 Comparing Two Proportions

We want to compare the true proportion of successes $p_1$ and $p_2$, in two independent populations. We take SRS from each population, and base our inference on the sample proportions $\hat{p}_x$ and $\hat{p}_y$.

We want to compare the true proportion of successes $p_1$ and $p_2$, in two independent populations. We take SRS from each population, and base our inference on the sample proportions $\hat{p}_1$ and $\hat{p}_2$. As before, it is better to adjust our sample proportions with a Wilson estimate when making confidence intervals.

Confidence Interval for $p_1 - p_2$:

$$\hat{p}_x - \hat{p}_y \pm z \sqrt{\frac{\hat{p}_x (1 - \hat{p}_x)}{n_x} + \frac{\hat{p}_y (1 - \hat{p}_y)}{n_y}}$$

- Wilson’s Estimate for Two Sample Proportions:

$$\hat{p}_x = \frac{X + 1}{n_x + 2} \quad \hat{p}_y = \frac{Y + 1}{n_y + 2}$$

- Use this formula when each sample is 10 or more, and the confidence level is 90%, 95% or 99%.

Significance Test for $H_0: p_1 - p_2 = 0$

TS: $z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_x} + \frac{1}{n_y} \right)}}$

- $\hat{p}$ in the standard error is a pooled proportion of successes for both samples taken together: pooled proportion $= \hat{p} = \frac{all \ successes}{all \ trials} = \frac{x + y}{n_x + n_y}$

- The distribution of $\hat{p}_x - \hat{p}_y$ is approximately Normal, so we can use the Z table as long as we have at least 10 successes and 10 failures in each sample.
Ex: A few engineers were interested in reducing the number of defective fuel pellets. They were interested in determining if the amount of reground powder in the pellets affected the proportion of defectives.

<table>
<thead>
<tr>
<th></th>
<th>20% reground</th>
<th>50% reground</th>
</tr>
</thead>
<tbody>
<tr>
<td># of defectives</td>
<td>38</td>
<td>29</td>
</tr>
<tr>
<td>n</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

What is $\hat{p}_1$?  What is $\hat{p}_2$?  What is $\hat{p}$?

Conduct a Significance Test to determine if there is a difference in the proportion of defective pellets for pellets made out of 20% reground and 50% reground powder.

Create a 95% confidence interval to estimate the true difference between the proportion of defective pellets made from 20% and 50% reground powder.

Compare your results with the output from Minitab.

**Test and CI for Two Proportions**

<table>
<thead>
<tr>
<th>Sample</th>
<th>X</th>
<th>N</th>
<th>Sample p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38</td>
<td>100</td>
<td>0.380000</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>100</td>
<td>0.290000</td>
</tr>
</tbody>
</table>

Estimate for $p(1) - p(2)$: 0.09
95% CI for $p(1) - p(2)$: (-0.0402307, 0.220231)
Test for $p(1) - p(2) = 0$ (vs not = 0): $Z = 1.35$  P-Value = 0.178
# Review

**Sampling Distribution** – distribution of values a sample statistic takes in repeated sampling.

**Sample Means**

**Sampling Distribution of X-bar:**
- has the same mean as original population
- has smaller standard deviation than original population
- (CLT) for any shape population, shape of distribution of X-bar goes towards Normal as \( n \) increases

\[
\bar{X} \sim N \left( \mu, \left( \frac{\sigma}{\sqrt{n}} \right)^2 \right)
\]

When can you use the Normal Table to find probabilities? Which z-score do you use?

<table>
<thead>
<tr>
<th></th>
<th>( X ) is Normal</th>
<th>( X ) is not Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>One individual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample mean when ( n&lt;30 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample mean when ( n&gt;30 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Introduction to Formal Statistical Inference

- **Statistical Inference (CI and Sig Tests)** – use random and representative sample to draw conclusions about population AND attach measure of reliability to it.

Confidence Intervals

Find Formulas for CI by using the table:

**Interpretation** –
- CI is a statement about a PARAMETER, not about statistic or individuals.
- “Probability” applies BEFORE we take data. After we use the word “confidence”.
- Interpret Results (CI include zero or \(H_0\))

**Behavior** –
- as confidence level increases – CI bigger
- as n increases – CI smaller

Finding n for one sample mean–
- margin of error = \(z^* \frac{\sigma}{\sqrt{n}}\)
- solve for n (always round up)

Finding n for one sample proportion
- \(n = \frac{z^2 \tilde{p}(1 - \tilde{p})}{\Delta^2} - 4\)
- \(\tilde{p}\) equals 0.5 if you have no idea what the true proportion is
- \(\tilde{p}\) can also be last years p-hat or if you are knowledgeable about the subject your guess
Significance Tests for $\mu$:

Ho                                                                 Ha

TS =

p-value="corner" area

Conclusions – small p-value supports Ha
Assumptions the t for one sample mean, paired means and two independent means

Why do we use the t table?

• What do we need (Assumptions)?
  1. Random Samples
  2. Population is Normal

• How do we check?
  1. Read story and THINK.
  2. Plot Data as long as there are NO OUTLIERS proceed.

Assumptions for the CI and Sig. Tests for p and p1-p2 using the z

• Why do we use the z table?

• What do we need (Assumptions)?
  1. Random Samples-to extend conclusions
     *Data is independent and comes from a stable system.
  2. Check specific assumption for that test
     *Have to have at least 5 successes and 5 failures.

How do we check these?

  1. Think about the story.
  2. Look at the data.
Additional Items

- READ MINITAB OUTPUT.

- What is the correct sampling distribution for the two independent means case?