Bring Back the Binomial . . .

- *Number of successes in n independent trials*
- *Each trial has probability p of a success*
- *Notation:* $X \sim B(n, p)$
- *Mean and Variance of a Binomial :*

  Mean = np

  Variance = $np(1-p)$

- *the **sample proportion** is calculated as* $\hat{p} = \frac{X}{n}$

Recall the CLT for binomials :

Central Limit Theorem (For Binomials)

If X_n is the number of successes observed from a binomial distribution, and if np and $n(1-p)$ are both larger than 10, then

$$\hat{p} = \frac{X_n}{n} \sim N(p, \sqrt{\frac{p(1-p)}{n}})$$

Add Confidence Intervals . . .

Suppose we have a random binomial sample of size n. A A level C confidence interval for the parameter p, the true probability of success in each trial is given by

$$\hat{p} \pm z_C \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

sample proportion ± z $_{\text{critical value}}$ * sqrt($\frac{\text{Sample prop. (1-sample prop.)}}{\text{Sample Size}}$)

Example *: Deer Ticks in CT. Deer ticks are the vector which transmits Lyme Disease to humans. In 2001, Michael Benjamin (et all) investigated the use of a naturally occurring fungus <u>M.anisopliae</u> to control the tick population. Fungus spores were sprayed on some ticks, while water was sprayed on other ticks. Here were the results of one study :*

	Water	*Fungus*
Dead Ticks	*3*	*40*
Total Ticks	*92*	*76*

For the ticks sprayed with Fungus, $\hat{p} = \frac{40}{76} = 0.52$

A 95% Confidence Interval for the true probability a tick will die when sprayed with the fungus is

$$\hat{p} \pm z_C \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.52 \pm 1.96\sqrt{\frac{0.52(1-0.52)}{76}}$$

$$= 0.52 \pm 0.11$$

***Example** : **Evolution vs. Creation.** Recall that a sample of 1000 Americans taken last June found that 64% believe that people are the result of creation rather than evolution. What is a 99% Confidence Interval for the true proportion of Americans who believe we were created?*

$$\hat{p} \pm z_C \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.64 \pm 2.58\sqrt{\frac{0.64(1-0.64)}{1000}}$$

$$= 0.64 \pm 0.04$$

*0.04 is called the "**Margin of Error**" in a poll, i.e. 4% Margin of Error (usually based on a 95% Confidence Interval).*

NOTE : when making confidence intervals for the probability of success in a binomial distribution, **use z (standard normal) critical values.** Reason : there is only one unknown parameter, p, since the variance is a function of p!! Don't need to correct for not knowing the standard deviation

Confidence intervals for proportions in MINITAB : use Stat → Basic Statistics → One proportion. Enter variable with data or summarize data. Choose options for different CIs.

Test and CI for One Proportion

```
Sample    X     N  Sample p         95% CI
1       640  1000  0.640000  (0.609372, 0.669801)
```

Confidence intervals for proportions in SPSS : use Analyze → Nonparametric tests → Legacy Dialogues → Binomial. Enter variable with data of zeros and ones. Haven't found a way to use summarized data.

Hypothesis Testing

- Like proof by contradiction : (think back to geometry)

__Example__ : there are infinitely prime numbers (attributed to Euclid)

<u>Proof</u> (by contradiction). Assume we have finitely many primes $p_1, p_2, ..., p_n$. Consider the number $q = p_1 p_2 ... p_n + 1$ (the product of all primes + 1). q is not divisible by any of the primes $p_1, p_2, ..., p_n$ since it always leaves remainder 1. So, q must be a prime number. BUT q is not in our list of primes so we have a

Contradiction

Therefore, there must be infinitely many primes.

Math, Logic	Statistics, Real Life
• Want to prove a statement.	• Want to use data to provide evidence for a hypothesis.
• Assume the statement is not true.	• Assume the opposite hypothesis is true.
• Reason to a contradiction.	• Show the observed data is <u>very unlikely</u>
(i.e. prove something is impossible)	*(not impossible, just very unlikely)*

Example : IQ Tests.
The general population has a mean IQ score of 100 and the population standard deviation of scores is $\sigma = 16$.

Suppose a sample of 9 CIA operatives yields a sample mean IQ score of $\bar{X}_n$=126.

We're wondering if on average, CIA operatives are smarter than the general population.

Read as 'H naught'

"Null Hypothesis" $H_0 : \mu_{CIA\ IQ} = 100$

"Alternative Hypothesis" $H_a : \mu_{CIA\ IQ} > 100$

We want to assess evidence for H_a, the alternative hypothesis.

Assume the opposite; that is, assume the null hypothesis $\mathbf{H_0}$ is true!!

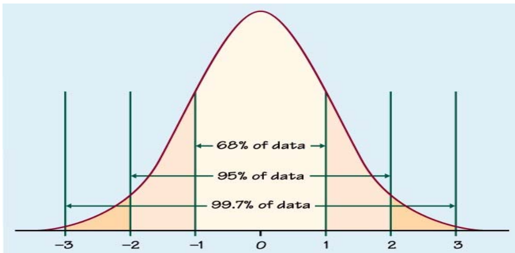
- *If the null hypothesis is true, then (according to the Central Limit Theorem), our sample mean $\bar{X}_n$ should look like an observation from a normal distribution with mean 100 and standard deviation 16/sqrt(9) :*

$$\bar{X}_n \sim N(100, \frac{16}{3})$$

- *How likely is it to observe a value of 126 or larger in a normal distribution with mean 100 and standard deviation 16/3? That is, what is* $P(\overline{X}_n \geq 126)$

- *This is a problem we know how to solve!!*

Calculate z-score : $\dfrac{126-100}{16/3} = 4.875$



$P(\overline{X}_n \geq 126)$
$= P(Z_n \geq 4.875)$

100 110 .6 126

(or just look up probability in MINITAB)

In General : Calculate z-score as $\dfrac{\overline{x}_n - \mu_o}{\sigma/\sqrt{n}}$

z-score = $\dfrac{\text{Sample mean – true mean under null hypothesis}}{\text{True standard deviation / sqrt(sample size)}}$

What is the probability of seeing a value of 4.875 of greater in a standard normal distribution? Basically ZERO!!!!!

$P(Z_n \geq 4.875) = 0.000$ ⟵ ***p-value***

The p-value is the probability of observing our sample statistic (the sample mean) or something more extreme IF the null hypothesis is true.

Memorize This!

For the CIA IQ data, we REJECT THE NULL HYPOTHESIS because the p-value is quite SMALL.

How small is small?

Reject the null hypothesis when

- **p-value is less than a pre-specified threshold**
- **This threshold is called** α *(alpha)*
- **The generally accepted minimum threshold is** $\alpha = .05$
- **If the p-value is not less than** α**, we <u>FAIL TO REJECT</u> the null hypothesis**

Memorize This too!

Rejection, Acceptance, and the Law

- Say p-value is 0.07. Do we 'accept the null hypothesis'?
- **NO. We 'fail to reject the null hypotheses'**
- Legal analogy "Innocent until proven guilty beyond a reasonable doubt". If someone is 'not guilty', it means the state failed to prove guilt, not that the defendant was proven innocent.

***Example** : **OJ Simpson.** In October 1995, OJ Simpson was found not-guilty of the criminal murder of his ex-wife Nicole Brown Simpson and her friend Ronald Goldman.*

14 months later, in a civil case, a jury found that Simpson 'willfully and wrongfully caused the death of Ronald Goldman' and was ordered to pay $8.5 million in damages.

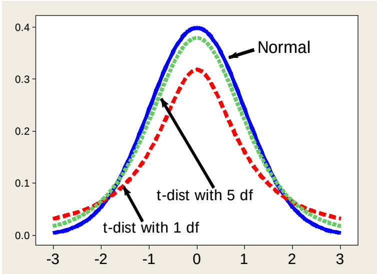
- *In a criminal trial, a defendant must be 'guilty beyond a reasonable doubt'.*
- *In a civil trial, a jury decides based on the 'preponderance of the evidence'.*

Thus, the criminal trial did not prove OJ was innocent – it failed to provide sufficient evidence of guilt.

The threshold for guilt is more stringent in a criminal trial – ***i.e. alpha is smaller!***

What to do when the true standard deviation σ is unknown *(almost always!!)*

Use the *t*-distribution !!



- If σ is unknown, estimate with <u>sample standard deviation</u> s.
- Correct for fact that we don't know σ by using a *t* distribution with (*n-1*) degrees of freedom.

In General : Calculate t-score as $\dfrac{\bar{x}_n - \mu_o}{s/\sqrt{n}}$

t-score = $\dfrac{\text{Sample mean – true mean under null hypothesis}}{\text{Sample standard deviation / sqrt(sample size)}}$

***Example** (Cartoon Guide). Comparison of two gas treatments in 10 cabs, calculate differences. Evaluate null hypothesis of no difference between gas types.*

$H_0 : \mu = 0$ *(no difference in gas types)*

$H_a : \mu \neq 0$ *(difference in gas types)*

$\bar{x}_{10}$ = -0.61, s_x=0.61, n=10, *What is probability of a sample mean of –0.61* ***<u>or something more extreme</u>*** *if the null hypothesis is true?*

Get standardized t-score : $\frac{-0.61-0}{0.61/3} = -3$

Having no prior opinion about which gas would be better, the p-value is

$2 * P(T_{9df} \leq -3)$ *= 2*0.0075 = 0.0150. This is less than 0.05 so we* ***reject the null hypothesis*** *and conclude that the gas brands are <u>not the same.</u>*

What's up with the 2 ???

One Sided vs. Two Sided Hypothesis tests

Only difference is in the alternative hypothesis :

$H_0 : \mu = 0$

$H_a : \mu \neq 0$ (***Two Sided Alternative Hypothesis*** *– mean could be either larger* ***OR*** *smaller than zero)*

--or --

$H_a : \mu > 0$ (***One Sided Alternative Hypothesis*** *– mean could be larger than zero)*

$H_a : \mu < 0$ (***One Sided Alternative Hypothesis*** *– mean could be larger than zero)*

- Use a one-sided test if you have a **prior opinion** about the how the true mean will differ. Otherwise, use a **two-sided test.**
- **P-values for a two sided test = 2 times p-value for a one-sided test.**

Hypothesis Tests of the mean in MINITAB : use `Stat` → `Basic Statistics` → `One sample t.` Click on Test Mean. Enter variable with data. For hypothesis test, enter a value for the mean under the null hypothesis (the 'Test Value'). To choose alpha and one/two sided test, click on `Options.`

SPSS: use `Analyze` → `Compare Means` → `One Sample t-test.` This only gives two-sided p-value, so cut in half for one-sided p-value.

Example *:* ***Gas in Cabs.*** *Results from MINITAB :*

Two-Sided Test

One-Sample T: Gas Difference

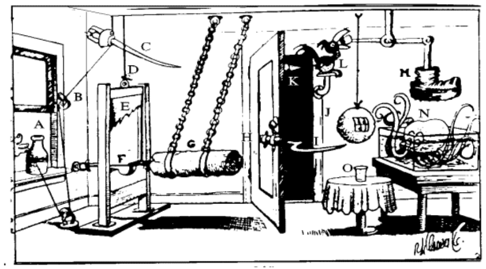
```
Test of mu = 0 vs not = 0

Variable         N       Mean     StDev            95% CI
Gas Difference  10  -0.606000  0.614278  (-1.045428, -0.166572)

Variable            T      P
Gas Difference  -3.12  0.012
```

One-Sided Test

One-Sample T: Gas Difference

```
Test of mu = 0 vs < 0
                                          95% Upper
Variable         N     Mean   StDev         Bound      T      P
Gas Difference  10  -0.6060  0.6142      -0.24991  -3.12  0.006
```

Hypothesis Testing
A General Structure



Hypothesis testing is a four step process :

Step 1 : Formulate two hypotheses

a) H_o : <u>The null hypothesis</u> - usually this says that the observations we observe are by chance and that the usual state of nature is in good operation.

b) H_a: <u>The alternative hypothesis</u> - the state of nature has been altered. Choose one or two sided.

Step 2 : Identify a <u>test statistic</u> which will provide evidence against the null hypothesis. *(i.e. a z-statistic or a t-statistic).*

Step 3 : Calculate a **p-value = the probability that if the null hypothesis is true, we observe a test statistic at least as extreme as the one we observed.**

Step 4 : Compare the p-value to a fixed **significance level,** α. If p-value $\le \alpha$, reject the null hypothesis. If p-value$>\alpha$, we fail to reject the null hypothesis **: we <u>do not</u> accept the null hypothesis – we simply failed to show it was false.**

Let's apply this to the Binomial Distribution :

Hypothesis Tests for Proportions

Step 1 : Formulate hypotheses

1. The null hypothesis is that the true population proportion p is equal to some number p_o.

2. The alternative hypothesis takes one of three forms
   (1) H_a: $p > p_o$.
   (2) H_a: $p < p_o$.
   (3) H_a: $p \neq p_o$.

Example : Ticks and fungus. *Recall that researchers are trying to see if spraying the fungus <u>M.anisopliae</u> on ticks causes them to die. Past experience suggests that in general, about 3% of ticks die each month in captivity. After one month, 40 of 76 ticks sprayed with the fungus die. Did this happen just by chance, or is there evidence that more ticks are dying?*

One-sided test : p_o

H_0: p_{fungus} = 0.03 *(the usual death rate for ticks over this time span)*

H_a: $p_{fungus} > 0.03$ *(i.e. fungus kills ticks) : a one-sided test is reasonable since we are pretty sure the fungus won't help the ticks!*

Step 2 : After taking a binomial sample of size *n,* calculate the statistic $z_{obs} = \frac{\hat{p} - p_o}{\sqrt{p_o(1-p_o)/n}}$.

z-observed = $\frac{\text{Sample proportion – true proportion under null hypothesis}}{\text{Standard deviation for a sample of size } n \text{ under the null hypothesis}}$

z_{obs} should have an approximately **standard normal distribution**

This is because . . . . .

- The sample proportion is $\hat{p} = \frac{\#\,successes}{\#\,trials}$
- The CLT states that **under the null hypothesis**, if sample size large enough, p_0 not too large or small, the sample proportion has an approximately normal distribution with mean p_0

$$\hat{p} \sim N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$$

- This implies that (subtract the mean) $(\hat{p} - p_0) \sim N\left(0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$

  (divide by standard deviation) $z_{obs} = \frac{(\hat{p} - p_0)}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$

i.e. z_{obs} should have an approximately standard normal distribution **IF** the null hypothesis is true.

Example : Ticks and fungus. $\hat{p} = \frac{\#\,successes}{\#\,trials} = \frac{40}{76} = 0.52$

$$= \frac{.52 - 0.03}{\sqrt{0.03(1-0.03)/76}} = 25.04$$

Step 3 : Calculate p-value = Prob of observing z_{obs} or something more extreme in a standard normal distribution. Multiply by 2 for a two-sided test.

Example : Ticks and fungus.

$\Pr(Z > 25.04) = 0.00000$ *: this is the p-value*

Step 4 : Compare the p-value to the pre-determined value of α (alpha). **Reject null hypothesis if p-value < α (alpha)**

Example : Ticks and fungus.

Pretty much any α **we pick is greater than 0.00000, so we reject the null hypothesis and conclude that the fungus death rate is significantly higher than the null death rate of ticks!**

Hypothesis tests for proportions in MINITAB : use Stat → Basic Statistics → One proportion. Enter variable with data **or** summarize data. Choose options to set null proportion (p_0) and to choose one sided or two-sided hypothesis tests.

SPSS : use Analyze → Nonparametric tests → Legacy Dialogues → Binomial. Enter variable with data of zeros and ones. Haven't found a way to use summarized data. Only does two-sided

Example : *Ticks and fungus.*

Test and Confidence Interval for One Proportion

```
Test of p = 0.03 vs p not = 0.03

                                                  Exact
Sample       X      N  Sample p        95.0 % CI       P-Value
1           40     76  0.526316  (0.408437, 0.642082)    0.000
```


Types of Errors

When bad things happen to good researchers

- or -

Crime and Punishment

There are two types of errors that can be made.

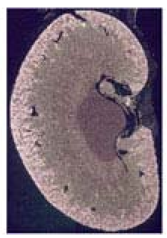
	We choose H_o	We choose H_a
H_o is true	No error P(correct decision)= 1-α ***[acquit the innocent]***	**Type I error** **P(Type I error) = α** **= <u>significance level</u>** ***[convict the innocent]***
H_a is true	**Type II error** **P(Type II error)= β** ***[acquit the guilty]***	No error P(correct decision)= 1-β = <u>Power of the test</u> ***[convict the guilty]***

We want :

- **Small** α - usually, this is fixed at 0.05 or 0.01.
- **Small** β - often, specify a **power** $(1-\beta)$ of 80% or 90%.

- For a fixed sample size, α and β have an **inverse** (but not linear) **relationship** : if β goes up, then α goes down.
- In general, the Type I error rate α is fixed.
- Often in experiments it is desirable to have a fixed Type I error rate and still have a specified level of power.
- **This requires a larger sample size or a better experiment!**
- Many U.S. government agencies require that studies they fund have a sample size large enough that for a significance level of α=0.05, the probability the experiment will find a significant result under a particular alternative hypothesis is 80% (i.e. power = 80%)

***Example** : An experiment on mice exposed to a pesticide anticipates a 1% increase in the size of the kidneys relative to baseline values. 1% is a meaningful change. The standard deviation of changes is estimated to be about 2%, and a sample size of 25 is used. A significance of 0.05 is specified.*



$$H_o : \mu_{change} = 0\%$$

$H_o : \mu_{change} > 0\%$. *In particular, we test the alternative* $\mu_{change} = 1\%$

Find the value such that the probability we reject H_o when it's true is 0.05. From standard normal tables, we find

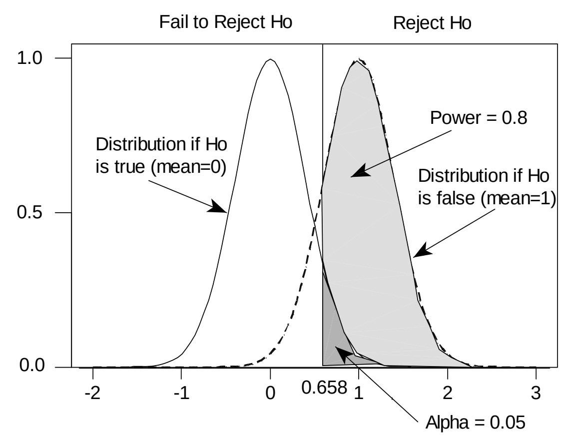
$$\frac{\bar{x} - 0}{2/\sqrt{25}} > 1.645 \quad \textit{i.e. } \bar{x} > 0.658$$

Now, assume the alternative hypothesis is true : $\mu_{change} = 1\%$*. What is probability of seeing a value greater than 0.658 under the alternative hypothesis ?*

$$P(\bar{x} > 0.658 \text{ when } H_a \text{ is true})$$

$$= P(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{0.658 - 1}{2/\sqrt{25}}) = P(Z > -0.855) = 0.80,$$

*that is, the experiment will have **80% power under the given conditions.***



*If we want α =0.01, this moves the rejection region to the left, which decreases the power. **To reduce alpha and maintain power, we must increase the sample size** (more on this later . . .)*