Simple example

- A box contains 1000 balls some white some black
- It is believed (hypothesized) that there are 990 W and 10 B
- 5 balls are drawn at random (with replacement). They were all found to be B
- Would you believe in the hypotheses ?
- No
- Because,if the hypotheses is true prob of getting all 5 B is small?
- Reasonable conclusion wrong reason !!

Simple example

- Suppose I have 10 cards numbered 1,2,..10
- They are shuffled
- I find 1 on top, 2 next, 3 in the third positions and so on
- Did the shuffle result in random ordering ?
- No?
- All orders have same probability $1/10!$.
- Small prob of observation not equivalent to doubt on hypotheses!!

Simple example

- A box contains 1000 balls some white some black
- It is believed (hypothesized) that there are 990 W (H_0)and 10 B or 995 W and 5 B (H_1)
- 5 balls are drawn at random (with replacement). They were all found to be B
- which of the two hypotheses would you chose?
- What if H_1 is 980 W and 20 B?
- The decision depends on the alternative

Example

A medicine is effective only if the concentration of a certain chemical is at least 200 ppm. At the same time the medicine would produce side effects if the concentration exceeds 200 ppm. A sample of 50 observations had an average of 201.5 ppm with a standard deviation of .75 ppm. Do you think the medicine would be effective?

- Hypotheses: concentration is more than 220 ppm
- Hypotheses: concentration is less than 220 ppm
- choose between the above two hypotheses

Testing Statistical Hypotheses

- Null and Alternative Hypothesis
- p-value
- significance level
- Type I and II errors, Power

Null and Alternative hypotheses

- An assertion about a parameter that we hold as true unless we have sufficient (statistical) evidence against it
- Alternative Hypotheses H1 Specification of values of a population parameter deemed as possible alternative values

Example

- A medicine is effective only if the concentration of a certain chemical is at least 200 ppm. At the same time the medicine would produce side effects if the concentration exceeds 200 ppm. A sample of 50 observations had an average of 201.5 ppm with a standard deviation of .75 ppm. Do you think the medicine would be effective?
  - Hypotheses: concentration is more than 220 ppm
  - Hypotheses: concentration is less than 220 ppm
  - Which of the two hypotheses is Null and which is alternative?
  - Let the concentration be μ
  - Null: $\mu \leq 220$ Alternative $\mu > 220$
  - $H_0 : \mu \leq 220$ $H_1 : \mu > 220$

Example

- An automobile manufacturer substitutes a different engine in cars that were known to have an average mpg of 31.5 miles. The manufacturer wants to test if the new engine changes the mpg. A random sample of 100 trial runs gives $\bar{X} = 29.8$, $s.d = 6.6$. Test the appropriate hypotheses at level of significance is $\alpha = .05$.

- μ: Average mpg of all cars with the current engine.
- $H_0 : \mu = 31.5 \qquad H_1 : \mu \neq 31.5$

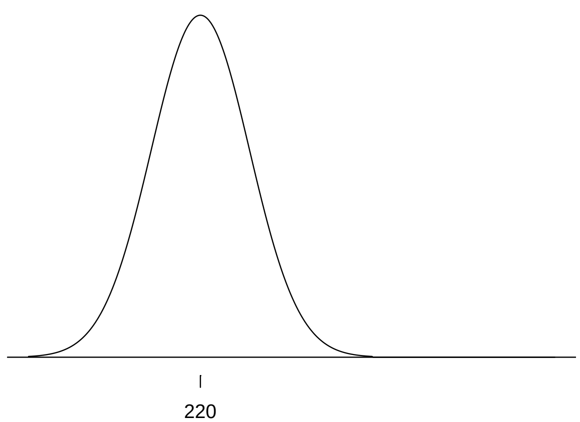
Type I and II errors

Suppose we have to choose between hypotheses H_0 and H_1

Truth	H_0 True	H_0 False
Decision		
Accept H_0	correct	Type II error
Reject H_0	Type I error	correct

We would ideally like a procedure than makes the probability both the error zero or small. This is not possible

- A medicine is effective only if the concentration of a certain chemical is at least 200 ppm. At the same time the medicine would produce side effects if the concentration exceeds 200 ppm. A sample of 50 observations had an average of 201.5 ppm with a standard deviation of .75 ppm. Do you think the medicine would be effective?
  - Let the average concentration be μ
  - Assume the concentration is distributed normally with $\sigma = .75$
  - $H_0 : \mu \leq 220 \qquad H_1 : \mu > 220$
  - $\bar{X}$ from 50 samples is normal with mean μ and s.d $\frac{.75}{\sqrt{50}} = 0.106$.
  - Large values of $\bar{X}$ favors H_1



Neyman Pearson approach

- Type I error is considered more serious than Type II error .
- So, choose Significance Level α (Usually 5%, 1%,). This is the probability of Type I error that we are willing to tolerate
- subject to this level of Type I error, choose a method that minimizes Type II error

Mechanics of Testing

- Suppose μ: Mean of a population
- $H_0 : \mu = \mu_0$(a specified value) $\quad H_1 = \mu > \mu_0$
- *barX* average of a sample from the population
- Under H_0: $\bar{X}$ will be close to μ_0
  Under H_1: $\bar{X}$ will be large
- So large values of $\bar{X}$ will evidence against H_0.

p-value

- Suppose that H_0 (that is if $\mu = \mu_0$) is true
  Let $\bar{x}$ be the observed sample average
- $P(\bar{X} > \bar{x})$ is called the p-value.

p-value

- If the NULL HYPOTHESES were true, this is the probability of observing as extreme a value of $\bar{X}$ as has been observed
- Smaller the p-value, lesser the evidence in favor of H_0
- If p-value is less than the specified significance level α, then Reject the NULL

Steps in Hypotheses testing

- Formulate the NULL and the Alternative
- Choose a test statistic that will be close to the value specified in the NULL
- Use the alternative to decide what is the ‘extreme’ value of the statistic (Large? small ?, two sided ?)

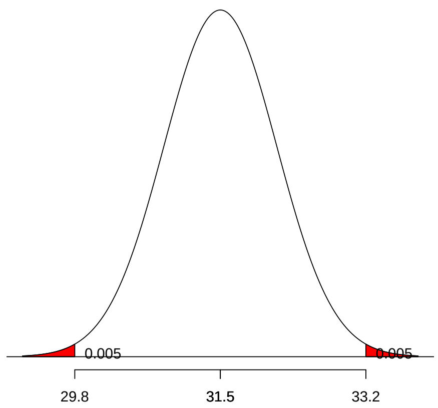
Steps in Hypotheses testing

- compute p-value (use software)
- smaller the p-value less evidence for Null hyp
- make a decision
- If p-value $< \alpha$ - Reject NULL.

Example

- An automobile manufacturer substitutes a different engine in cars that were known to have an average mpg of 31.5 miles. The manufacturer wants to test if the new engine changes the mpg. A random sample of 100 trial runs gives $\bar{X} = 29.8$, $s.d = 6.6$. Test the appropriate hypotheses at level of significance $\alpha = .05$

- μ: Average mpg of all cars with the current engine.
- $H_0 : \mu = 31.5 \qquad H_1 : \mu \neq 31.5$
- A value far from 5 indicates that the null hypotheses is untenable



some tests

- for testing for the mean of a normal population
  - If s.d is known: Z-test using normal distribution
  - If s.d is not known: t-test using t-distribution
  - test for s.d.: chi-square distribution
- for testing equality of mean of two populations, use t-test
- for testing equality of variance : F-test

NONPARAMETRIC TESTS

- Parametic models: concerned parameters of distributions. In order to apply these tests, certain conditions about the distributions must be verified. In practice, these tests are applied when the sampling distributions of the data variables reasonably satisfy the normal model.
- Non-parametric tests make no assumptions regarding the distributions of the data variables
- they are adequate to small samples
- there are nonparametric tests that can be applied to ordinal and/or nominal data
- The non-parametric tests are, in general, less powerful than their parametric counterparts

Mann-Whitney test

- also known as Wilcoxon-Mann-Whitney or rank-sum test,
- used to assess whether two independent samples were drawn from the same population

Mann-Whitney test

- start by assigning ranks to the samples
- Let the samples be denoted $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$
  The ranking of the x_i and y_i assigns ranks in 1, 2, . . ., n $+$ m.
-

x_i	12	21	15	8
y_i	9	13	19	

- The null hypothesis tested is $P(X > Y) = \frac{1}{2}$
- under the null hypothesis, one expects the ranks to be randomly distributed between the x_i and y_i, therefore resulting in approximately equal average ranks in each of the two samples.
- For small samples, there are tables with the exact probabilities of W_X.
- For large samples (say m or n above 10), the sampling distribution of W_X rapidly approaches the normal distribution

- The ranking of x_i and y_i would then yield the result

Variable	X	Y	X	Y	X	Y	X
Data	8	9	12	13	15	19	21
Rank	1	2	3	4	5	6	7

- The test statistic is the sum of the ranks for one of the variables, say X
  $W_X = 16$