Programming Tidbit
DO Loops

- The advantage of a DO loop is a counter that changes from a set value, by a set increment, while it does not pass the another set value.

  DO counter = starting_value, limiting_value, increment

  - Example: DO counter = 1, 10
  - The counter cycles through the values 1, 2, 3, …, 10
  - Example: DO counter = 1, 11, 3
  - The counter cycles through the values 1, 4, 7, 10.
  - Example: DO counter = 10, 1, -2
  - The counter cycles through the values 10, 8, 6, 4, 2.
- The lines of code between the DO and the ENDDO are repeated until the value of the counter passes the limiting value.
  - DO loops are often used to sum variables or to set values of arrays.
  - The index of the array should be the counter, or some mathematical function of the counter.

WHAT NOT TO DO!!!
How Not to Sum

```
DO j = 1, 12
    READ(7,*) monthly_rent
ENDDO
annual_rent = annual_rent + monthly rent
PRINT*, annual rent
```

- What is wrong with this approach?
  - Annual_rent is not initialized, so the computer gets to chose.
  - The summation is outside the DO loop
- So let's do it right….

```
annual_rent = 0.0
DO j = 1, 12
    READ(7,*) monthly_rent
   annual_rent = annual_rent + monthly rent
ENDDO
annual_rent = annual_rent + monthly rent
PRINT*, annual rent
```

WHAT NOT TO DO!!!
How Not to Store Data in an Array

```
REAL (dim 366) :: temperature_max
DO j = 1, 366
    READ(7,*) temperature_max(index)
ENDDO
```

- What is wrong?
  - The value of index is not defined. The computer gets to choose, which is likely to lead to a bad value.
  - Even if the computer picks an OK value (between 1 and 366), it will never change. The variable index does not depend on j!

```
REAL (dim 366) :: temperature_max
DO j_day = 1, 366
   READ(7,*) temperature_max(j_day)
ENDDO
```

MET3220C & MET6480 Computational Statistics

Hypothesis Testing

(Chapter 5 of Wilk's book)

Background: key concepts

Key Points:

1) Parametric vs. Nonparametric tests

2) The elements of any hypothesis test

3) Confidence intervals

Tests to Deal with Uncertainty

- If we know the exact and singular values for everything we are considering, then testing questions about these values would be trivial.
  - The answers would be obvious.
- In the real world, there is always some uncertainty, and variables seldom have only one value (they have a distribution of values).
  - If the uncertainty is small compared to what we are considering, then the answer is more likely to be obvious.
  - However, in many weather and climate related studies, the uncertainty can be large enough (relative to the test statistic that we are considering), that there can be some doubt in an answer to the test.
- Consequently, a very important part of hypothesis testing is being able to make a good estimate of the uncertainty.
- Another key consideration is the type of test.

Parametric vs. Nonparametric Tests

- Parametric tests assume that the data is well described by a theoretical distribution (one that can be parameterized).
  - Parametric tests are often designed to test under what conditions data fits the theoretical distribution.
    - Tests could examine fitting parameters,
    - Tests could examine a fit to the distribution, or
    - The likeliness of a value being different from another value.
- Nonparametric tests do not rely on a presumed distribution.
  - Nonparametric tests are designed to avoid the assuming an underlying distribution of the data.
  - There are two broad categories of nonparametric tests/applications.
    - Classical nonparametric tests are not dependent on the distribution – any distribution would be OK.
    - Resampling techniques attempt to use the available observations to construct a distribution.

Uncertainty in Fitting Parameters

- Fitting parameters are usually determined from a sample of the population. Since samples differ, it is reasonable to expect the fitting parameters to differ. This situation results in uncertainty in the fitting parameters.
  - This type of error is called sampling error.
  - Observational error also contributes to this uncertainty.
- The set of fitting parameters determined from different samples will also have a distribution.
- If this distribution can be characterized, then it can be used in a parametric test.

The Elements of A Hypothesis Test

- In general, hypothesis tests will work through 5 steps.
- 1) Identify a *test statistic*
  - Chose a statistic that is appropriate to the data and the question.
  - The test statistics must be computed from the data.
    - For parametric tests, the statistic will often be a fitting parameter for the assumed distribution.
    - Possibilities for non-parametric tests are enormous.
- 2) Define a *null hypothesis*.
  - The null hypothesis is usually denoted as H_0.
  - The null hypothesis defines a logical structure which will be used to examine the test statistic.
  - The null hypothesis is often designed as the compliment to what we would like to test for.
    - Example: student A is not taller than student B.
    - Example: global warming is not occurring.

The Elements of A Hypothesis Test

- 3) Define the *alternative hypothesis* (H_A).
  - This hypothesis is the compliment of the null hypothesis.
    - Example: the null hypothesis in not true.
  - A more complicated hypothesis is possible.
  - Hint: think about whether it is easier to clearly state null hypothesis or an alternative hypothesis, then define the other hypothesis as the compliment of the one that is more easily defined.
- 4) Determine the *null distribution*
  - The distribution associated with a true null hypothesis.
  - Think of this as a distribution based on uncertainty in H_0.
    - This distribution could be a parametric distribution, or
    - A distribution based on resampling.
  - Knowing this distribution helps determine a good test.
- 5) *Compare* the test statistic to the null distribution.

Interpretation of the Comparison

- The interpretation hinges on two factors:
  - The test statistic, and
  - Uncertainty in the statistic.
    - Which is then uncertainty in the null hypothesis.

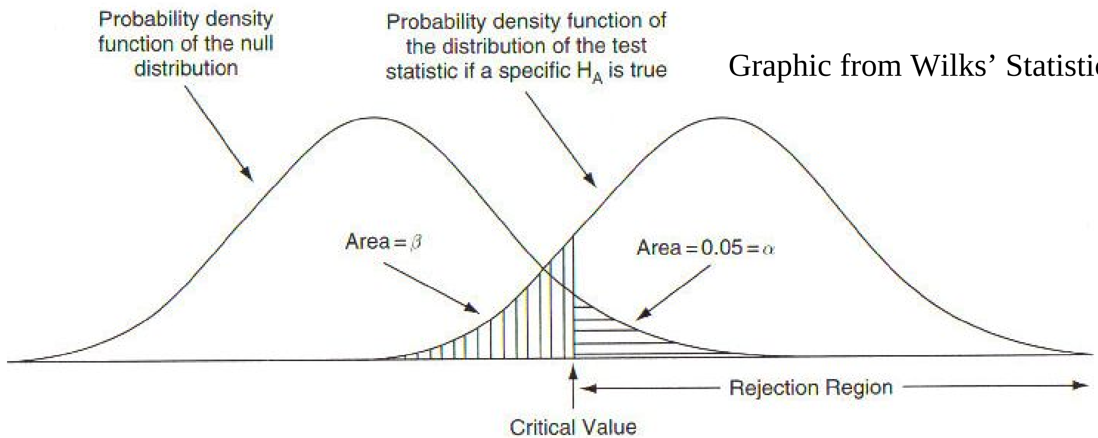
- If the test statistic is outside of reasonable uncertainty in the null hypothesis, then we can reject the null hypothesis.
  - If our error statistic is based on absolute error, then we can be certain that the null hypothesis is rejected.
  - If our error statistic is a measure of spread, then we cannot be certain that the null hypothesis is wrong.
- If the test statistic not outside reasonable uncertainty in H_0, then
  - We cannot distinguish the result from the null hypothesis.
  - We cannot say if either H_0 or H_A is correct.

Outside of Reasonable Uncertainty: Test Levels and *p* Values

- How do we define *sufficiently improbable* in the context of the *null hypothesis*?
- We must define a *test probability* (or test level) below which the null hypothesis is considered *sufficiently improbable*.
- The *null hypothesis* is rejected if the probability, as defined by the *test statistic* and the *null distribution*, is less than or equal to the *test probability*.
- The test level is defined in advance of the comparison.
  - It depends on the investigator's judgment and preferences.
  - Consequently it is somewhat arbitrary.
  - Common values are 1%, 5%, and 10%, with 5% being most common.
- In situations where there are consequences associated with correct and incorrect results, the test level can be optimized for the application.
- The *p* value is the specific test probability (the chance of the null hypothesis being rejected based on the null distribution).

Error Types

- There are two types of errors in the context of binary (true or false) outcomes.
- *False Rejection* (or false negative):
  - Incorrect rejection of the null hypothesis (it is actually true).
  - Also known as a Type I error, often denoted as α.
- False Positive:
  - Incorrect acceptance of the alternative hypothesis (when the null hypothesis is actually correct).
  - Also known as a Type II error, often denoted as β.



Graphic from Wilks' Statistical Methods

Error Types

- The probabilities of false rejections and false positives can be somewhat controlled by the choice of the *test probability*.
- Ideally both types of errors would be eliminated; however, there are many applications where this ideal cannot be achieved.
- The probability of a false rejection can be estimated if the null distribution has been estimated (this is your p value).
  - Keep in mind that the null distribution is estimated – not known.
  - Sometimes the estimation is good, other times it will not be.
  - It is useful to know the sensitivity of the value associated with the test probability to errors in the null distribution.
    - One of the reasons to avoid any extremely small test probability is the relatively large sensitivity in the tails of distributions.

One Sided vs. Two Sided Tests

- Statistical tests can be either one sided or two sided.
- One sided tests have an alternative hypothesis that is true only on one side of the null distribution.
  - Example: H_A = Student A is taller than student B.
  - Example: H_A = Global warming is occurring.
- Two sided tests have an alternative hypothesis that is true on both sides of the null distribution.
  - Example: H_A = Student A has a different height than student B.
  - Example: H_A = The global temperature is changing.
  - In these examples, a substantial difference in either direction would violate the null hypothesis.
  - In these cases the chance of a false rejection (α) is split on *both sides* of the *null distribution*.
  - If the null distribution is symmetric, then α is split equally on both sides of the null distribution.

Confidence Intervals

- A confidence interval indicates the region where the alternative hypothesis is found to be true for a set chance of a false rejection.
- Example: the regions in North America were the rate of change of temperature is positive (and different from no change), with a 5% chance of a false rejection.
  - Confidence levels of 10% and 1% could also be plotted
  - The result would be a contour map indicating differing levels of confidence in the alternative hypothesis.

Example

- Consider the word of an advertiser that in the summer after an El Niño, it is rain free for 6 out of 7 days in the Pacific Northwest.
  - Most people would not take this seriously, knowing that the Pacific Northwest is best known for fog, light rain, depression related to insufficient sunlight, and a rather large rain forest.
  - However, during an El Niño event the weather is relatively sunny.
  - But 6/7 days sounds a bit fanciful. So let's test if the number is within reasonable bounds of observations.
- Assume that we acquire observations for an appropriate location.
  - We chose days that are well separated, so that we can treat the data as independent.
  - We chose from within the summer season to be consistent with the advertisement, and because that will reduce the change in statistics with time.
  - We find that for 15 of 25 days the weather is rain free.
  - Is 15/25 close enough to 6/7 that we can't tell the difference?

Example Continued

- Consider our 5 steps:
- 1) Identify a test statistic.
- The difference in the means.
- 2) Define a null hypothesis
- The observed fraction of rain free days is similar (or greater) than 6/7.
    - *Fraction of rain free days* ≥ 0.857 minus statistical uncertainty
- 3) Define the alternative hypothesis.
- The observed fraction of rain free days is substantially less than 6/7.
- 4) Obtain the null distribution
- Rain vs. no rain is binary, so a binary distribution is appropriate.
- 5) Compare the observed statistic to the null distribution.

$$\Pr\{X \le 15\} = \sum_{x=0}^{15}\left[\binom{25}{x} 0.857^{x}\left(1-0.857\right)^{25-x}\right] = 0.0015$$

- This is the probability of finding 15/25 (or fewer) rain free days if the advertised statistics are true. *What is missing from this example*?