0 Preliminaries

First introductory question: When do we use nonparametric statistics?

Example: Construct a 95% confidence interval on the population mean household income of your community. Test if the population mean household income equals $72,000 versus the alternative that the population mean household income does not equal $72,000.

(1) How is this done?

(2) Which test (or table) is often used?

(3) What assumptions are needed?

"Essentially, all models are wrong, but some are useful." – George Box (a great statistician, 1919 – 2013).

(4) Suppose the sample size is small, and normality is not reasonable, and a transformation to normality is not reasonable, and methods based on other non-normal distributions are not available or are not appropriate (or assumptions are not made on the distribution).

Second introductory question: What are nonparametric statistics?

Details on *nonparametric* methods begin in section 0.5.

0.1 Cumulative Distributions and Probability Density Functions

A **population** consists of all possible values for some variable.

A **sample** consists of a subset of the population, and is typically drawn randomly.

A **random variable** denotes some numerical value of an observation selected randomly from the population.

A **cumulative distribution function** (cdf) of a random variable X is $P(X \le x)$ for all real x.

Example: Suppose X denotes the income of a randomly selected household. What does the cumulative distribution of X evaluated at $100,000 represent?

□

If the random variable is *continuous*, then probabilities may be illustrated by the area under a curve known as the **probability density function** (pdf).

What are some continuous distributions (which have names)?

Homework Learn *R*. Read chapter 0, if you have the textbook.

0.2 Common Continuous Probability Distributions

Using *R* or *RStudio* {see https://www.r-project.org or https://rstudio.com}

To download *R*: Click "CRAN", "UCLA" (or some other mirror site), "Download R for Windows", "base", "Download R 4.0.4 for Windows" (or the latest version). Click "Save" and then "Save" again. The saving should take less than five minutes using high-speed internet. Click "Run", "Run", and "OK", and continue to click "Next" until the process is complete.

Alternatively, use *Rweb* by going to https://rweb.webapps.cla.umn.edu/Rweb/

Normal distribution, $N(\mu, \sigma)$

The probability density function is

$$f(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}, \ -\infty < x < \infty \quad \text{(need not memorize)}$$

What is $f(x)$ for a *standard normal* distribution?

Example: Use *R* regarding the *normal* distribution for the following.

(a) Graph the *probability density function* (pdf) of a standard normal distribution, say, from -3 to 3.

```
> x <- 0:100 # Generate numbers from 0 to 100.
```

```
> x = 0:100 # Generate numbers from 0 to 100.
```

Alternatively, use:

```
> x = c(0:100) # 'c' means combine.

> x = x / 100 - 0.5 # Generates numbers between -0.5 and 0.5.

> x = x * 6 # Generates numbers between -3 and 3.
```

Alternatively, type:

```
> x = 6 * ( c(0:100) / 100 - 0.5 )
```

```
> y = dnorm(x) # f(x); i.e., probability density function for a standard
  normal r.v.
```

```
> # Note that different means and standard deviations may be used.
> plot( x, y )
> plot( x, dnorm(x) )
> plot( x, dnorm(x), type="s" ) # Use stair step for a nicer looking
  graph.


> plot( x, dnorm(x), type="s" )
```

(b) Graph the *probability density function* (pdf) of a standard normal distribution, say, from -3 to 3, using only one command.

(c) Graph the *probability density function* (pdf) of a $N(\mu = 1.5,\ \sigma = 0.4)$ random variable from -3 to 3.

```
> x1 = x
> plot( x=x1, y=dnorm( x=x1, mean=1.5, sd=0.4 ), type="s" ) # different
  mean and sd.
> plot( x1, dnorm( x1, 1.5, 0.4 ), type="s" )
```

(d) Graph the *probability density function* (pdf) of a $N(\mu = 1.5,\ \sigma = 0.4)$ random variable from -3 to 3, using only one command.

(e) Graph the *cumulative distribution function* (cdf) of a $N(\mu = 1.5,\ \sigma = 0.4)$ random variable from -3 to 3, using only one command.

(f) Use the function `plotDist` to graph the *probability density function* (pdf) of a $N(\mu = 50,\ \sigma = 5)$ random variable from 35 to 65.

```
> source( "http://educ.jmu.edu/~garrenst/math324.dir/Rfunctions" ) #
  Read in the functions.
> ls( ) # List the functions and variables.

> ?plotDist # Help menu.
> plotDist( "dnorm", 50, 5, xmin=35, xmax=65 )
```

(g) Use the function `plotDist` to graph the *cumulative distribution function* (cdf) of a $N(\mu = 50,\ \sigma = 5)$ random variable from 35 to 65.

(h) Use the function `plotDist` to graph the *probability density function* (pdf) and the *cumulative distribution function* (cdf) of a standard normal random variable.

(i) What will the following *R* commands generate?

```
> pnorm( 1.96 )

> pnorm( -1.96 )

> pnorm( 1.96 ) - pnorm( -1.96 )

> qnorm( 0.975 ) # 'q' in qnorm means 'quantile.'

> pnorm( 319.6, 300, 10 )

> qnorm( 0.975, 300, 10 )
```

(j) Generate 1000 independent observations from a $N(\mu = 80,\ \sigma = 10)$ distribution.

Determine the sample mean, $\bar{X}$, and the sample standard deviation, s, and graph these observations in a histogram.

```
> x1 = rnorm( 1000, 80, 10 ) # The 'r' in 'rnorm' means 'random.'
> mean( x1 )
> sd( x1 )
> hist( x1 )
> history( ) # To view the history.
> q( ) # To quit!
```

Central Limit Theorem: If n independent observations are sampled from a population with mean μ and positive finite standard deviation σ, then the sample mean $\bar{X}$ has approximately a $N(\mu,\ \sigma/\sqrt{n})$ distribution, for **large** n.

Furthermore, if n independent observations are sampled from an approximately $N(\mu,\ \sigma)$ population, then the sample mean $\bar{X}$ has approximately a $N(\mu,\ \sigma/\sqrt{n})$ distribution.

A **parameter** is an (often unknown) quantity which is a characteristic of a population.

More on standardizing

In general, if X is any random variable and a and b are constants such that the distribution of $(X - a)/b$ does not depend on a or b, then a is a **location** parameter and b is a **scale** parameter.

Then, the probability density function (pdf) of X can be written:

$$f_{a,b}(x) = \frac{1}{b}\, h\left(\frac{x-a}{b}\right).$$

Example: Let $f_{\mu,\sigma}(x)$ be the pdf of a $N(\mu, \sigma)$ distribution.

$$f_{\mu,\sigma}(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}, \ -\infty < x < \infty$$

Let $h(z)$ be the pdf of a $N(0, 1)$ distribution. Then,

$$h(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}, \ -\infty < z < \infty$$

Note:

$$\frac{1}{\sigma}\, h\left(\frac{x-\mu}{\sigma}\right) = \left(\frac{1}{\sigma}\right) \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}} = f_{\mu,\sigma}(x)$$

Uniform distribution

The pdf for a **standard** uniform distribution is

$$h_1(z) = 1, \ 0 < z < 1.$$

Example: Use R regarding the **uniform** distribution for the following.

(a) Graph the **pdf** of a *standard* **uniform** distribution.

```
> ?runif
```

(b) Graph the **cdf** of a *standard* **uniform** distribution.

(c) Graph the **pdf** and **cdf** of a **uniform** distribution with endpoints 30 and 40.

(d) What is the **mean** of a **uniform** random variable?

(e) Sample 1000 observations from a Uniform(30, 40) distribution, calculate the sample mean and plot the histogram.

```
> x = runif( 1000, 30, 40 )
> x[ 1 :  50 ] # List the first 50 observations.
```

□

Example: Applying the Central Limit Theorem to the uniform distribution using *R*.

Let $U \sim \text{Uniform}(40,\ 60)$. Let $\bar{U}_n$ be the sample mean of n independent realizations of U.

(a) Plot the pdf and cdf of U.

(b) Sample 30 independent observations of $\bar{U}_2$, using the function 'replicate'.

```
> u.bar = replicate( 30, mean( runif( 2, 40, 60 ) ) )
```

(c) Construct a histogram of 10,000 independent observations of $\bar{U}_2$.

Interpret your graph.

(d) Repeat part (c) for $\bar{U}_5$, $\bar{U}_{20}$, and $\bar{U}_{100}$.

```
> hist( replicate( 1e4, mean( runif( 5, 40, 60 ) ) ) )
```

□

Exponential distribution

An **exponential** random variable is *continuous* and **lacks memory**.

Example: Suppose that a particular type of computer chip, running continuously, has lifetime distributed according to an *exponential* distribution.

Suppose that an **old** computer chip, which has been running for two years, is compared to a **new** computer chip, which has been running for only one day.

Which computer chip is better?

Can the *lifetime* of a computer chip be negative?

□

The pdf for a **standard** exponential distribution is

$$h_2(z) = e^{-z}, \; z > 0,$$

and has mean and standard deviation equal to 1.

```
> # Graph the pdf of a standard exponential distribution.
> plotDist
> ?dexp
```

```
> # Graph the pdf of an exponential distribution using 'rate' equal to
   0.1.

> # What are the new mean and standard deviation?

> # Generate 10,000 observations from an exponential distribution using
   'rate' equal to 0.1.

> # Construct a histogram of the data.

> # Determine the sample mean of x.

> # Determine the sample standard deviation of x.
```

□

Laplace or Double Exponential distribution

A **Laplace** distribution is symmetric about its mean, such that each half of the *pdf* is *exponential*.

The pdf for a *standard* **Laplace** distribution is

$$h_3(z) = \frac{e^{-|z\sqrt{2}|}}{\sqrt{2}}, \; -\infty < z < \infty.$$

The *standard* Laplace distribution has mean zero and standard deviation one.

```
> # Graph the pdf of the standard Laplace distribution.

> ?dlaplace
```

```
> # Graph the pdf of the Laplace distribution with mean equal to 2 and
   standard deviation equal to 3.


> # Generate 10,000 observations from a Laplace distribution with mean
   equal to 2 and standard deviation equal to 3.


> # Construct a histogram of the data.


> # Determine the sample mean of x.


> # Determine the sample standard deviation of x.
```

□

Cauchy distribution

A **Cauchy** distribution is symmetric and has **heavy** tails.

The pdf for a *standard* **Cauchy** distribution is

$$h_4(z) = \frac{1}{\pi(1+z^2)}, \ -\infty < z < \infty.$$

(Note the typo in the textbook on p. 3.)

The population mean (μ) and the population standard deviation (σ) of a Cauchy distribution **DO NOT EXIST**.

```
> # Graph the pdf of the standard Cauchy distribution.
> plotDist
> ?dcauchy
```

```
> # Graph the pdf of a Cauchy distribution with location parameter 40
  and scale parameter 5.
```

```
> # If ‘X’ has the above pdf, determine P(X<40), using R.
```

```
> # Determine P(X<35).
```

```
> ?shadeDist
> shadeDist( 35, "dcauchy", 40, 5 )
> # Determine P(X<45).
```

□

Comparing Normal and Cauchy distributions

Ignoring the constants, the probability density functions of the normal and Cauchy distributions are:

$$e^{-z^2} \text{ and } (1+z^2)^{-1}.$$

Which of these converges to zero faster as $z \to \infty$ (or $z \to -\infty$)?

If Z is standard normal, determine $P(-1 < Z < 1)$.

```
> shadeDist( c(-1, 1), "dnorm", lower.tail=FALSE )
> install.packages( "mosaic", depends=FALSE )
> library( mosaic )
> xpnorm( c( -1, 1 ) )
```

If X is standard Cauchy, determine $P(-1 < X < 1)$.

Generate 100 observations from a standard normal distribution, and 100 observations from a standard Cauchy distribution.

```
> z = rnorm(100) ;   x = rcauchy(100)
```

Look at the observations. Are there any outliers?

Determine the minimum and maximum, for both z and x.

```
> min(z) ;   max(z)
> min(x) ;   max(x)
```

Plot the histograms.

May want to try:

```
> truncHist( x )
> truncHist( x, -3, 3 ) # Construct a truncated histogram, excluding
  values outside -3 and 3.
```

Consider the **Shapiro-Wilk** normality test.

H_0 : The distribution is normal, vs. H_a : The distribution is nonnormal.

P-values ≤ 0.05 imply evidence of nonnormality. However, p-values > 0.05 do NOT imply or prove normality.

Light-tailed distributions, such as *uniform* and *normal*, rarely produce outliers.

Heavy-tailed distributions, such as *exponential*, *Laplace,* and *Cauchy*, tend to produce outliers.

The *Cauchy* distribution has very heavy tails.

What is the distribution of the *sample mean* of n independent $N(\mu,\ \sigma)$ random variables?

What is the distribution of the *sample mean* of n independent Cauchy(θ, 1) random variables?

Example: Compute the sample **mean** of **10** observations from a standard **normal** distribution; repeat for a standard **Cauchy** distribution. Then, increase the sample size to **1000** and **100,000**.

Example: Compute the sample **median** of **10** observations from a standard **normal** distribution; repeat for a standard **Cauchy** distribution. Then, increase the sample size to **1000** and **100,000**.

□

Homework **C0.2.1*:** Let $X \sim N(\mu = 500,\ \sigma = 200)$. Let U be a uniform random variable with endpoints 50 and 60. Let $Y \sim \text{Cauchy}(\text{location} = 500,\ \ \text{scale} = 200)$. Let $\bar{X}$ be the sample mean of 100 independent realizations of X. Let $\bar{U}$ be the sample mean of 100 independent realizations of U. Let $\bar{Y}$ be the sample mean of 100 independent realizations of Y. Let $\widetilde{Y}$ be the sample median of 100 independent observations of Y. Use R for all graphs below, and show both your source code and output. Introduce the question number and letter as a comment using "#" or in red using .html code; e.g., `<span style="color: red"> Exercise C0.2.1(a) </span>`.

(a) Graph the pdf of X.

(b) Graph the pdf of U.

(c) Graph the pdf of Y.

(d) What distribution (i.e., the shape, location, and scale) does $\bar{X}$ have? (There is no need to use R for this part.)

(e) Graph the pdf of $\bar{X}$.

After completing part (e),

sample 10,000 independent realizations of $\bar{X}$.

(f) Compute the sample mean and sample standard deviation of your 10,000 values of $\bar{X}$.

(g) Graph your 10,000 values of $\bar{X}$ in a histogram.

(h) Are your results from parts (f) and (g) consistent with your answer from part (d)? Explain.

Next, sample 10,000 independent realizations of $\bar{U}$.

(i) Graph your 10,000 values of $\bar{U}$ in a histogram.

(j) What is the approximate *shape* of the distribution of $\bar{U}$?

Next, sample 10,000 independent realizations of $\bar{Y}$.

(k) Graph your 10,000 values of $\bar{Y}$ in a histogram.

(l) Graph your 10,000 values of $\bar{Y}$ in a *truncated* histogram.

(m) What distribution (i.e., the shape, location, and scale) does $\bar{Y}$ have? (There is no need to use R for this part.)

(n) When trying to estimate the population median (i.e., location parameter) of a Cauchy distribution, which is more informative: one observation from the Cauchy distribution OR the sample mean based on 100 independent observations from the Cauchy distribution? Explain.

Next, sample 10,000 independent realizations of $\widetilde{Y}$.

(o) Graph your 10,000 values of $\widetilde{Y}$ in a histogram.

(p) When trying to estimate the population median (i.e., location parameter) of a Cauchy distribution, which is more informative: the sample mean or the sample median? Explain.

End of Homework C0.2.1*. □

0.3 The Binomial Distribution

Example: Toss a coin ten times where the probability of heads is 40%. Let X be the number of heads. Then X is a binomial random variable with parameters $n = 10$ and $p = 0.4$.

Example: Sample ten people (independently) from a population consisting of 40% Democrats. Let X be the number of Democrats. Then X is a binomial random variable with parameters $n = 10$ and $p = 0.4$.

For both above examples, what are the possible values of X?

In general: A *Bernoulli* trial may result in a *success* or a *failure*. Suppose X represents the number of *successes* from n independent Bernoulli trials, where n is fixed (not random) and $p = P(\text{success})$. Then, $X \sim \text{Binomial}(n, p)$.

The probability density function (pdf) of X can be shown to be

$$P(X = x) = \binom{n}{x} p^x \ (1-p)^{n-x}, \ x = 0, 1, 2, \ldots, n.$$

Note well: The definition of the **pdf** of a discrete random variable differs from that of a continuous random variable.

In the above example regarding political affiliation, what is the probability that exactly 30% of the sample consist of Democrats?

Alternatively, in the above example regarding coin tosses, what is the probability that exactly 30% of the ten coin tosses result in heads?

```
> dbinom(3, 10, 0.4)
```

Determine $P(X = x)$, for all $x = 0, 1, 2, \ldots, 10$.

On average, how many heads do we expect from the 10 coin tosses?

On average, how many Democrats do we expect from the 10 people sampled?

In general, what is the mean of X?

The variance of X can be shown to be $np(1-p)$.

A special case of the Central Limit Theorem: If $X \sim \text{Binomial}(n, p)$ and $\hat{p} = X/n$, where $np \geq 10$ and $n(1-p) \geq 10$, then X is approximately $N\left(\mu_x = np,\ \sigma_x = \sqrt{np(1-p)}\right)$ and $\hat{p}$ is approximately $N\left(\mu_{\hat{p}} = p,\ \sigma_{\hat{p}} = \sqrt{p(1-p)/n}\right)$.

Hence, for sufficiently large sample sizes, a binomial random variable (or a sample proportion) may be reasonably approximated by a normal random variable.

Example: Graph the pdf of a Binomial$(n, p = 0.3)$ random variable for various values of n, say $n = 1$, 2, 5, 10, 30, 50 and 100.

```
> plotDist( "dbinom", 1, 0.3 )
```

□

Homework **C0.3.1*:** Let $X \sim \text{Binomial}(n,\ p = 0.01)$. Introduce the question number and letter as a comment using "#" or in red using .html code; e.g., `<span style="color: red"> Exercise C0.3.1(a) </span>`.

(a) Graph the pdf of X for $n = 10$, 100, 500, and 1000. (Make sure that you select appropriate domains for X so that your four graphs are neat, especially for $n = 10$.)

(b) State for which of these graphs in part (a) a normal approximation seems reasonable.

End of Homework C0.3.1*. □

Homework C0.3.2: Let $X \sim \text{Binomial}(n = 500,\ p = 0.8)$.

(a) Determine (exactly) $P(X \leq 390)$.

(b) Determine the mean and standard deviation of X.

(c) Determine the z-score (i.e., standard normal score) corresponding to $x = 390$ (or you may use $x = 390.5$).

(d) Using a normal approximation to the binomial, compute $P(X \leq 390)$. (Use R for this computation, not a standard normal table.)

(e) Compare your answers from parts (a) and (d).

(f) Sample 1000 independent realizations of X. Determine the sample mean, $\bar{X}$, and the sample standard deviation, s, and graph these observations in a histogram.

(g) Compare your answers from part (b) with $\bar{X}$ and s from part (f).

End of Homework C0.3.2. □

0.4 Confidence Intervals and Tests of Hypotheses

Suppose **(a)** the observations from a population are independent, and **(b)** the original population is approximately normal or the sample size n is large (and $0 < \sigma < \infty$). Let $\bar{X}$ be the sample mean, and let s be the sample standard deviation. Then, an approximate confidence interval on the population mean, μ, is

$$\bar{X} \pm t_{n-1} s/\sqrt{n}.$$

Hypothesis testing on μ involves converting the standardized test statistic $T = (\bar{X} - \mu)/(s/\sqrt{n})$ to a p-value.

Example: Consider the following data set regarding birthweight of humans in grams: $\{3615, 3105, 4217, 3234, 3551, 4023, 3098\}$.

(a) Graph the data in a Q-Q plot.

```
> ?qqnorm

> x = c( 3615, 3105, 4217, 3234, 3551, 4023, 3098 )
```

(b) Construct a 90% confidence interval on μ, the population mean birthweight.

```
> ?t.test

> t.test( x, conf.level = 0.9 )
```

(c) Test at level $\alpha = 0.05$, $H_0: \ \mu = 3400$ g versus $H_a: \ \mu > 3400$ g.

```
> t.test( x, alternative = "greater", mu = 3400 )
```

0.5 Parametric versus Nonparametric Methods

Parametric methods often require that the population be approximately **normal** or the sample size be **large**.

When n is large (and the population standard deviation is finite), **normal** theory often may be used.

Much theory has been developed for the **normal** distribution.

Random variables from other distributions (whose theory has also been developed quite thoroughly) arise from **normal** populations.

What is an example?

◇

Some populations are *non-normal* but may be transformed to **normality** by taking a **transformation**.

For example, a data set generated by a **log-normal** random variable may be transformed to a **normal** random variable. How?

◇

Other distributions (in addition to the **normal** distribution) also have theory that is fairly well developed.

For example, data from an **exponential** population may be analyzed using techniques based on the **exponential** distribution.

◇

How can we determine whether or not we believe a data set is from an approximately **normal** population?

How well can we distinguish **normal** data from **exponential** data for small sample sizes?

◇

With **parametric** methods (such as the t-test), typically the sample size needs to be large or the shape of the population is assumed (such as **normality**), and some or all of the numeric parameters (such as μ and σ) are assumed unknown.

◇

Nonparametric methods generally do not require that the shape of the population be known or that the sample size be large.

With **nonparametric** methods, **normality** is not assumed. The assumptions required on the distributions for **nonparametric** methods tend to be quite weak (e.g., the population is continuous).

Nonparametric methods are sometimes called **distribution-free** methods, since use of such methods is not restricted to particular (e.g., **normal**) distributions.

◇

Typically, when the distribution is heavy-tailed and the sample size is small, the **nonparametric** methods are more valid than the **parametric** methods.

Suppose the population really is **normal**. Should **parametric** or **nonparametric** methods be used?

◇

Even when the population is **normal**, **nonparametric** methods often perform well, even in comparison to **parametric** methods.

0.6 Classes of Nonparametric Methods

Nonparametric methods are based on four statistical ideas.

Methods Based on the Binomial Distribution

(chapter 1)

Suppose we perform hypothesis testing or construct confidence intervals on the **median** of a *continuous* population.

We sample n independent observations from the distribution.

An observation *above* the median might be labeled as a *success*, whereas an observation *below* the median might be labeled as a *failure*.

Let Y be the number of successes. Then Y has what distribution?

Permutation (Shuffling / Scrambling) Methods

(chapters 2 through 7)

Example: Is the *new* drug better than the *old* drug for lowering cholesterol?

Suppose there are 20 volunteers for the study.

Obtain some **statistic** to account for the difference between the two groups regarding cholesterol change.

How do we convert our **statistic** to a p-value?

Permute (shuffle, at random) the 20 observations, and compute a *permuted* **statistic**.

Repeat the permuting many times to produce many (perhaps 10,000) *permuted* **statistics**.

Compare your original **statistic** to the 10,000 *permuted* **statistics**.

Interpret: Suppose the original **statistic** falls far into the (appropriate) tail of the histogram of the *permuted* **statistics**.

Bootstrap (Resampling) Methods (chapter 8 and section 9.1)

Sample 100 independent observations from an unknown population with unknown finite mean and unknown median.

$\bar{X}$ estimates μ, and the **sample median** estimates the **population median**.

What is the standard deviation of $\bar{X}$?

From Math 220, what is a reasonable estimator of the standard deviation of $\bar{X}$?

What is the standard deviation of the **sample median**?

What is a reasonable estimator of the standard deviation of the **sample median**?

One method for estimating is to use the **bootstrap**.

Bootstrap:

(1) Sample (i.e., **RESAMPLE**) 100 observations WITH replacement from the original 100 observations, and obtain a *bootstrapped* **sample median**.

(2) Repeat step (1) many times to produce many (perhaps 10,000) *bootstrapped* **sample medians**.

(3) Compute the sample standard deviation of the 10,000 *bootstrapped* **sample medians**. That is your estimate!

Smoothing and Non-Least Squares Methods

Smoothing (sections 10.1 and 10.2)

Example using one variable:

```
> x = c( rnorm(1000), rnorm(1000, 6, 1.5) )

> # Think of 'x' as just being some data set of 2000 numbers.


> hist( x )
```

□

Example using two variables: Start with a scatterplot (i.e., some 2-dimensional graph based on a sample).

Draw a smooth curve through the data. Why?

□

Non-Least Squares Methods (section 10.3)

Least squares methods (e.g., linear regression) often require **normality** and residuals (error terms) get **squared**.

Suppose a *heavy-tailed* distribution is of interest.

What do *heavy-tailed* distributions tend to produce?

Why is this a problem?

When **normality** is not a valid assumption, **non-least squares** methods may be more appropriate than **least squares** methods.