Chapter 5, Evans & Rosenthal

section 5.1

The source of uncertainty in a probability model -
- an uncertainty associated with an outcome or response as described by a probability model. The probability model itself (the probability measure) is fixed. An additional uncertainty comes from the fact that we don't exactly know what this probability measure is. E.g. $X \sim N(0, \sigma^2)$ - probabilistic uncertainty; in practice we may know (roughly) that X is Gaussian but σ^2 is unknown - thus, only a general model class is known. This is a statistical model.

Statistical inference must be based on data - observed outcomes/responses from a statistical model.

Ex. Stanford Heart Transplant study (March 1974, JASA)

Can only hope to determine whether a patient lived longer if receiving a new heart transplant - causality is very difficult to establish. Comparisons are made between e.g. the pdf of lifelength with transplant f_T and pdf of lifelength without the transplant f_C, T, C - Treatment and Control. The time is measured from the date of determination that the patient is a heart transplant candidate until the end of study. Typically, some characteristics of f_T and f_C will be measured e.g. their expectations. Note that clinical vs. statistical significance is still important here.

In practice, only a small number of obs. from f_T & f_C are available e.g. 30 patients in a control group and 52 patients in the treatment group. In table 5.1 (control) we know X and S - the indicator of whether the patient died by the end of the study.

In the table 5.2, there are additional covariates:
Y - # of days they waited for the transplant
Z - # of days they were alive after the date they received the transplant until the end of the study. Note that $X = Y + Z$. Typically, also age, sex etc. may be available as covariates.

Section 5.2 (Inference Using a Probability Model).

Assume that the probability measure P is known on a collection of subsets of a sample space S for a response s. Typically, we would know P but uncertain about the future response $s \in S$. Inference about s may be of interest -- prediction/estimate of s; e.g. $\mathbb{E}S$ may be taken as a prediction. Alternatively, we may be asked to construct a subset that has a high probability of containing s and is in some sense small.

Ex. Lifelength of a machine in years $X \sim Exp(1)$ [assumed]
$\Rightarrow \mathbb{E}X = 1$ - a prediction. The smallest interval containing 95% of all possible values of X is $(0, c)$ where

$$.95 = \int_0^c e^{-x}dx = 1 - e^{-c} \Rightarrow c = -\log(0.05) = 2.9957$$

Is $x_0 = 5$ a plausible value for some machine? Well,

$$P(X > 5) = \int_5^\infty e^{-x}dx = e^{-5} = 0.0067\ldots$$

Sometimes, additional prior information is available e.g. that $s \in C \subset S$. $\Rightarrow$ Condition $P(\cdot/C)$ when deriving our inferences. This is a basic axiom $\rightarrow$ it cannot be proven and may be a source of disagreement sometimes. Principle of conditional probability.

Suppose that $X \sim Exp(1)$ and a specific machine has been running for a year. — condition on $X>1$. The density of the conditional distribution is $e^{-(x-1)}$ for $x>1$. The predicted lifelength is, then

$$E(X/X>1) = \int_1^{\infty} x e^{-(x-1)} dx = 2 \quad \text{— memoryless property}$$

of exponential distribution. The tail probability for $x_0 = 5$ is $P(X>5/X>1) = \int_5^{+\infty} e^{-(x-1)} dx = e^{-4} =$

$= 0.0183\ldots$ now $x_0 = 5$ a little more plausible than before. Finally, $0.95 = \int_1^c e^{-(x-1)} dx = -e^{-(x-1)}\Big|_1^c =$

$= 1 - e^{-(c-1)} \Rightarrow e^{-(c-1)} = 0.05 \Rightarrow -(c-1) = \log 0.05$, or

$1 - c = -\log 20$, or $c = 1 + \log 20$

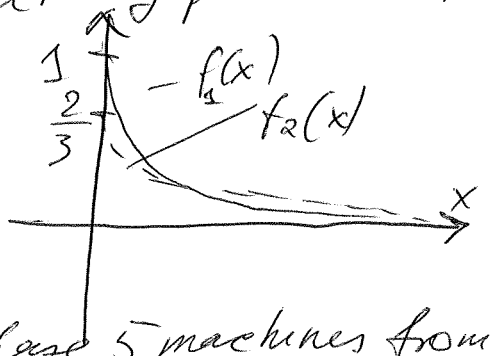
5.3 Statistical models.

How to conduct statistical inference for a statistical model $\{P_\theta : \theta \in \Omega\}$?! θ is the parameter of the model that has to be estimated, Ω is the parameter space. Hopefully, the model is identifiable, e.g. $P_{\theta_1} = P_{\theta_2} \Leftrightarrow \theta_1 = \theta_2$. If there is a density function, we write $\{f_\theta : \theta \in \Omega\}$ instead. Problems we face: 1) to estimate the true parameter value θ, 2) construct regions in Ω that contain the true value 3) assess whether the data are in agreement with a suggested value θ_0

For a sample $X_1, \ldots X_n$ the statistical model for a sample of $X_1, \ldots, X_n$ are independent is $f_\theta(x_1) f_\theta(x_2) \cdots f_\theta(x_n)$ for a sample from the statistical model $\{f_\theta : \theta \in \Omega\}$

Distribution: $X_1, \dots X_n$ - RV's. $x_1, \dots, x_n$ - their observed values

Two different groups of machines produced by two different manufacturing plants: lifelength $X \sim Exp(1)$, $Y \sim Exp(1.5)$



Here $Exp(\theta)$ mean $\frac{1}{\theta} e^{-x/\theta}$ so $EX = \theta$!

Purchase 5 machines from the same plant - don't know from which one. Observations - $(X_1, \dots, X_5)$. Statistical Model is $\{P_\theta : \theta \in \Omega\}$, $\Omega = \{1, 2\}$. Longer observed lifelengths favor $\theta = 2$; i.e. if $(X_1, \dots, X_5) = (5.0, 3.5, 3.3, 4.1, 2.8)$ we are more certain that $\theta = 2$ than if $(X_1, \dots, X_5) = (2.0, 2.5, 3.0, 3.1, 3.8)$

Parameter θ is a label but may be an important characteristic of a distribution. Reparametrization sometimes useful ($1 \leftrightarrow 1$ transformation). Two important examples:

Ex. 5.3.3. Bernoulli model: $\Omega = [0, 1]$ if $X_1, \dots X_n \sim Ber(\theta)$

$\Rightarrow f_\theta(x_i) = \theta^{x_i}(1-\theta)^{1-x_i}$. for a sample

$\prod_{i=1}^{n} f_\theta(x_i) = \theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})}$. Can parametrize using $\psi = \log \frac{\theta}{1-\theta}$!!

Ex. 5.3.4. Location-scale model (Normal)

$(X_1, \dots X_n) \sim N(\mu, \sigma^2)$ e.g. distribution of heights in a population

$\Rightarrow \prod_{i=1}^{n} f_{(\mu,\sigma^2)}(x_i) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\} =$

$= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2 - \frac{n-1}{2\sigma^2}s^2\right\}$. derive it here?!

Alternative parametrization: (μ, σ)

5.4 Data Collection – draw the sharp distinction between observational studies and experiments

5.4.1. Finite Populations – a finite set Π of objects (population), a real-valued measurement function X defined on Π; $X: \Pi \to \mathbb{R}$. E.g. $X(\pi)$ – height of a student π Height – quantitative variable – or $X(\pi) = \begin{cases} 1 - \text{female} \\ 2 - \text{male} \end{cases}$ – categorical variable. Population and the measurement produce a population distribution ... defined by the population CDF $F_X: \mathbb{R}^1 \to [0,1]$, $F_X(x) = \dfrac{|\{\pi: X(\pi) \le x\}|}{N}$,

$N = |\Pi|$. Consider example 5.4.1 – Π a population of $N = 20$ plots of land for the same use; $X(\pi)$ – a measure of fertility of plot π on a 10-point scale.

Measurements: 4 8 6 7 8 3 7 5 4 6

9 5 7 5 8 3 4 7 8 3.

$$\Rightarrow F_X(x) = \begin{cases} 0, & x < 3 \\ 3/20, & 3 \le x < 4 \\ 6/20, & 4 \le x < 5 \\ 9/20, & 5 \le x < 6 \\ 11/20, & 6 \le x < 7 \\ 15/20, & 7 \le x < 8 \\ 19/20, & 8 \le x < 9 \\ 1, & 9 \le x. \end{cases}$$

To know $F_X(x)$ exactly, one must conduct a census – hard!

In practice, we select a subset $\{\pi_1, \dots \pi_n\} \subset \Pi$ where $n < N$.

Approximate (estimate) $F_X(x)$ by the empirical distribution function: $\hat{F}_X(x) = \dfrac{|\{\pi_i: X(\pi_i) \le x,\ i = 1, \dots n\}|}{n}$

$$= \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(-\infty, x]}\left(X(\pi_i)\right).$$

Note: sometimes multivariate measurements $X: \Pi \to \mathbb{R}^k$ are needed

e.g. $X(\pi) = (X_1(\pi), X_2(\pi))$ – height, weight

How should we select the subset $(\pi_1, \dots \pi_n)$ and how large should n be?!

5.4.2. Simple random sampling

An important issue is how to select $\{\pi_1, \dots, \pi_n\}$ from an entire population in such a way that it does not turn out to have been sampled from a distinct subpopulation? 1) E.g. sampling river creatures from only the topmost layer of water. 2) Selecting only students with low ID numbers in order to estimate the population F_X may result in sampling senior students only. - Called a selection bias (or a selection effect).

This is the main problem with observational studies - the way the data has been generated is unknown to statisticians. Practical solution — select the set $\{\pi_1, \dots, \pi_n\}$ randomely. SRS (simple random sampling) - each subset of size n has the same probability $\frac{1}{\binom{N}{n}}$ of being selected. In this case $(X(\pi_1), \dots, X(\pi_n))$ is random. Thus, for $n=1$, we have $P(X(\pi_1) \le x) = F_X(x)$. Go back to the example of $N=20$ agricultural plots, each one has the prob. of $\frac{1}{20}$ to be selected. Thus, $P(X(\pi_1) \le x) = \frac{\#\{\pi : X(\pi) \le x\}}{20}$ $= F_X(x)$ for every $x \in \mathbb{R}^1$.

Prior to observing the sample, we also have $P(X(\pi_2) \le x) = F_X(x)$. However, given that $X(\pi_1) = x_1$, we removed one population member with measurement value x_1 so then $NF_X(x) - 1$ is the # of individuals left in Π with $X(\pi) \le x_1$. Therefore,

$$P(X(\pi_2) \le x \mid X(\pi_1) = x_1) = \begin{cases} \frac{NF_X(x) - 1}{N-1}, & x \ge x_1 \\ \frac{NF_X(x)}{N-1}, & x < x_1 \end{cases}$$

— neither result is equal to $F_X(x)$. $\Rightarrow X(\pi_1)$ and $X(\pi_2)$ are not independent in a random sample. For large N, however, $P(X(\pi_2) \le x \mid X(\pi_1) = x_1) \approx F_X(x)$ so $X(\pi_1)$ and $X(\pi_2)$ are independent and identically distributed.

Section 5.4.2.

Similarly, when N is large and n is small relative to N, $X(\vec{u}_1), \dots, X(\vec{u}_n)$ are approximately iid with the cdf $F_X(x)$. Thus, we will always treat values $(x_1, \dots, x_n)$ of $(X(\vec{u}_1), \dots, X(\vec{u}_n))$ as a sample. By the WLLN, $\hat{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty, x]}(X(\vec{u}_i)) \xrightarrow{P} F_X(x)$ as $n \to \infty$.

When the data were collected using SRS, the investigation is called a sampling study. Whenever available, they are always preferred to observational studies. Nevertheless, observational studies often have to be used for lack of better opportunities; selection bias always has to be acknowledged. The highest level of statistical evidence is an experiment which is appropriate when cause-effect relationships between variables have to be investigated.

Question #2: how to select the sample size n? As large as possible, seemingly... but beware of 1) High costs 2) Increasing possibility of corruption by various errors arising during the collection process. Thus, the best answer – as much as needed for the given accuracy but no more. Therefore, the needed accuracy has to be specified first. These are called sample-size calculations.

If we define $f_X(x) = \frac{|\{\vec{u}: X(\vec{u}) = x\}|}{N} = \frac{1}{N}\sum_{\vec{u} \in I} I_{\{x\}}(X(\vec{u}))$, this $f_X(x)$ is the proportion of population members that satisfy $X(\vec{u}) = x$; it plays the role of probability mass function because $F_X(x) = \sum_{z \le x} f_X(z)$. This f_X is the population relative frequency function.

Based on the sample $\{\vec{u}_1, \dots, \vec{u}_n\}$, $f_X(x)$ may be estimated as $\hat{f}_X(x) = \frac{|\{\vec{u}_i: X(\vec{u}_i) = x, i = 1, \dots, n\}|}{n} = \frac{1}{n}\sum_{i=1}^{n} I_{\{x\}}(X(\vec{u}_i))$ – the proportion of sample members satisfying $X(\vec{u}) = x$. It is appropriate to estimate for categorical variables where $f_X(x)$ is the proportion in the category specified by x. With quantitative variables, in general, $f_X(x)$ is not an appropriate estimation target.

Section 5.4.3. Histograms.

Now suppose that X is a continuous quantitative variable. Group values into intervals: $(h_1, h_2]$, $(h_2, h_3]$, $\ldots$, $(h_{m-1}, h_m]$ where $h_1 < h_2 < \ldots < h_m$, (h_1, h_m) should cover the range of possible values for X. Then,

$$h_X(x) = \begin{cases} \dfrac{\#\{\vec{a} : X(\vec{a}) \in (h_i, h_{i+1}]\}}{N(h_{i+1} - h_i)}, & x \in (h_i, h_{i+1}] \\ 0 \text{ otherwise.} \end{cases}$$

Then, h_X is a density histogram function. This implies that $h_X(x)(h_{i+1} - h_i)$ is the proportion of individuals for whom $X(\vec{a})$ is in $(h_i, h_{i+1}]$. Next, $F_X(h_j) = \int_{-\infty}^{h_j} h_X(x)\,dx$ for each interval endpoint, and $F_X(h_j) - F_X(h_i) = \int_{h_i}^{h_j} h_X(x)\,dx$, where $h_i \le h_j$. If the intervals $(h_i, h_{i+1}]$ are small, we expect $F_X(b) - F_X(a) \approx \int_a^b h_X(x)\,dx$ for any choice of $a < b$. If the lengths $h_{i+1} - h_i$ are small and N is very large, h_X would be approximated by a smooth continuous function $f_X(x)$ in the sense that $\int_a^b f_X(x)\,dx \approx \int_a^b h_X(x)\,dx$ for any choice of $a < b$. Then, we will also have $\int_a^b f_X(x)\,dx \approx F_X(b) - F_X(a)$; f_X - density function for the population distribution. This is typically how continuous distributions arise in practice!!

Section 5.4.4. Survey sampling.

Survey sampling (polling) is entirely based on finite population sampling. A survey is a set of questions asked of a sample $\{\vec{a}_1, \ldots, \vec{a}_n\}$ from a population Π. If there are m questions $\Rightarrow m$ measurements so we have a vector $(X_1(\vec{a}), X_2(\vec{a}), \ldots, X_m(\vec{a}))$.

A good example is: 1) pre-election polling
2) market surveys done by consumer product companies

Typically, the analysis of results is concerned not just with population distribution of the individual X_i, but also the joint population distributions. E.g. the joint cdf of (X_1, X_2) is given by

$$F_{(X_1,X_2)}(x_1, x_2) = \frac{\#\{\omega : X_1(\omega) \le x_1,\ X_2(\omega) \le x_2\}}{N}$$

It is also possible to define $f_{(X_1,X_2)}(x_1, x_2)$ and joint density histograms if (X_1, X_2) are both continuous quantitative variables.

Ex 4 mayoral candidates exactly, 1000 voters randomly selected and asked if they will vote, for whom they vote and their age. The 1st question is $X_1(\omega) = 1$ - will vote & $X_1(\omega) =$ no, won't vote. Next, preference is $X_2(\omega) = i$, $i = 1, 2, 3, 4$. X_3 - the age of the respondent in years. Typically, we would be interested in joint distributions of X_1 and X_2 but also in joint distributions of (X_1, X_3) and (X_2, X_3)

Serious problem in survey sampling - how to handle the nonresponse error

Chapter

~~Section~~ 5.5. Some Basic Inferences

Section 5.5.1. Descriptive Statistics

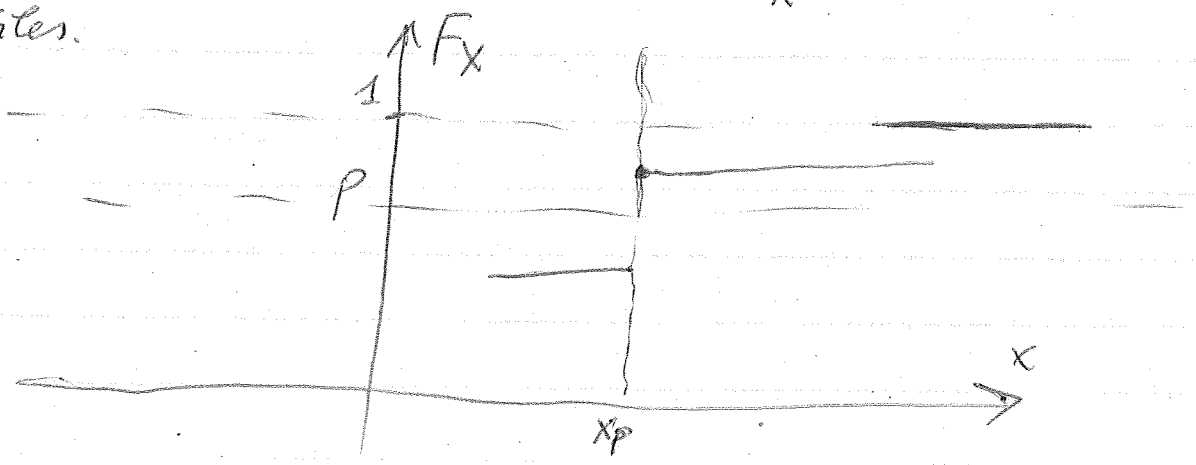
Ex. 5.5.1. $f_X(x)$ - proportion of population members equal to x; $F_X(x)$ - proportion of those that don't exceed x. From a sample $(x_1, \dots, x_n)$ we can get $\hat{f}_X(x)$ - proportion of sample values equal to x. Then, $\hat{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty, x]}(x_i)$ - an empirical cdf.

If $n=10$: 1, 2, 2.1, 0.4, 3.3, -2.1, 4.0, -0.3, 2.2, 1.5, 5.0

Here, $\hat{f}_X(x) = 0.1$ for any x that is a data value and 0 ~~else~~ elsewhere. $\hat{F}_X(x) = ?$ $\hat{F}_X(-3) = 0/10 = 0$, $\hat{F}_X(0) = 2/10 = 0.2$, $\hat{F}_X(4) = 0.9 (9/10)$

For $p \in [0,1]$, the p^{th} quantile (or $100p^{th}$ percentile) x_p is the smallest number x_p s.t. $p \le F_X(x_p)$; in other words, $x_p = F_X^{-1}(p) = \min\{x : p \le F_X(x)\}$;

If your test result is at the 90^{th} percentile, your grade is $x_{0.9}$ and 90% are at or below your level. If F_X is strictly increasing and continuous $F_X^{-1}(p)$ is the unique x_p s.t. $F_X(x_p) = p$. $x_{0.5} = F_X^{-1}(0.5)$ is a median, $x_{0.25} = F_X^{-1}(0.25)$ and $x_{0.75} = F_X^{-1}(0.75)$ - the 1^{st} and 3^{rd} quartiles.



Section 5.5.1 (continuation).

A natural estimate of a population quantile $x_p = F_X^{-1}(p)$ is $\hat{x}_p = \hat{F}_X^{-1}(p)$ (*) $\hat{F}_X$ is not continuous $\Rightarrow$ there may not be a solution to (*). To solve this issue: 1) order $X_1, \dots, X_n$ to obtain the order statistics $X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$.

2) $X_{(i)}$ is the $\left(\frac{i}{n}\right)$th quantile of cdf because $\hat{F}_X(X_{(i)}) \ge \frac{i}{n}$ and $\hat{F}_X(x) < \frac{i}{n}$ when $x < X_{(i)}$.

3) In general, define the sample $p^{(th)}$ quantile as $\hat{x}_p = X_{(i)}$ whenever $\frac{i-1}{n} < p \le \frac{i}{n}$ (**) A number of modifications to this are also used; for example, if for some (i) (**) is satisfied, we can define

$$\hat{x}_p = X_{(i-1)} + n\left(X_{(i)} - X_{(i-1)}\right)\left[p - \frac{i-1}{n}\right] \quad (5.5.3)$$

so that $\hat{x}_p$ is the linear interpolation between $X_{(i-1)}$ and $X_{(i)}$. When n is even, this definition gives the sample median $\hat{x}_{0.5} = X_{\left(\frac{n}{2}\right)}$. An alternative way of defining the sample median is as $\hat{x}_{0.5} = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & n \text{ is odd} \\ \frac{1}{2}\left[X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right], & n \text{ is even.} \end{cases}$ For n large enough, the exact choice doesn't matter much.

Here, use the data from Example (5.5.1) using the definition (5.5.3) for $p = 0.5$, $p = 0.25$, $p = 0.75$.

Ex. 5.5.3. Measuring location and scale of a Population Distribution.

Often, we need inferences about the population mean $\mu_X = \frac{1}{|\Pi|}\sum_{\pi \in \Pi} X(\pi)$ and the population variance $\sigma_X^2 = \frac{1}{|\Pi|}\sum_{\pi \in \Pi}\left(X(\pi) - \mu_X\right)^2$; here Π is the finite population and X is a real-valued measurement on it.

For discrete X, we can also write $\mu_X = \sum_x x f_X(x)$ because $|\Pi| f_X(x)$ is the

Section 5.5.1 (continuation)

is the number of elements $\in \Omega$ with $X(\omega) = x$. In the continuous case, $\mu_X \approx \int x f_X(x)\,dx$ where $f_X(x)$ is the approximating density. A natural estimate of the population mean is $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and the one for σ_X^2 is $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. An alternative choice would be the population median $x_{0.5}$ as a measure of location and the population IQR $x_{0.75} - x_{0.25}$ as a variability measure. This 2nd choice is preferred when the distribution is skewed. - an example: house prices in a specific area. The mean & median are very different - example of χ^2_4 distn where the median is 3.3567. In the 2nd case, natural estimators are the sample IQR defined as $\widehat{IQR} = \hat{x}_{0.75} - \hat{x}_{0.25}$. Again, in the example 5.5.1, check that $\hat{x}_{0.5} = 1.5$, $\widehat{IQR} = 2.75 - 2.05 = 2.70$

Section 5.5.2 Plotting data.

Using the example 5.5.1, using the same data, show the boxplot using R. Whiskers run from the quartiles to the adjacent values. The adjacent values are given by the greatest value $\leq$ upper limit $\left(3^{rd}\text{ quartile} + 1.5\,\widehat{IQR}\right)$ ... the lower limit $\left(1^{st}\text{ quartile} - 1.5 * \widehat{IQR}\right)$. Beyond adjacent values are outliers -- with * ngn; Change $x_{(10)} = 5.0$ to $x_{(10)} = 15.0$ and you will get an outlier. Outliers may occur simply because the data has a long-tailed distribution or they may signify a mistake in collecting or recording data.

For categorical variables:

Flavor	Count	Proportion
Chocolate	42	0.42
Vanilla	28	0.28
Butterscotch	22	0.22
Strawberry	8	0.08

- bars don't need to touch each other.

Chapter 5.5.3, Types of Inference.

Descriptive statistics are not that straightforward — we may not know which one we should use. If there's an additional information about a distribution family for the data, e.g. that $f_X \in \{f_\theta : \theta \in \Omega\}$ then descriptive statistics don't use this information at all.

⇒ using it will help us develop the theory of statistical inference. Recall the 3 types of inference for an unobserved response value s. (i) Predict an unknown response value s.

(ii) Construct a subset C of the sample space S that has a high probability of containing an unknown response s

(iii) Assess whether or not $s_0 \in S$ is a plausible value from the probability distribution specified by f.

In applications, we typically begin with determining some characteristics of the unknown distribution f_θ, e.g. its mean or the median. Denote such a characteristic $\psi(\theta)$. E.g. if it is a mean, $\psi(\theta) = \int x f_\theta(x)\,dx$; or $\psi(\theta) = F_\theta^{-1}(0.5)$

⇒ We consider three types of inference for $\psi(\theta)$:

(i) Choose an estimate $T(s)$ of $\psi(\theta)$ — estimation problem

(ii) Construct a subset $C(s)$ of the set of possible values for $\psi(\theta)$ that we believe contains the true value — the problem of credible region (confidence region) construction

(iii) Assess whether or not ψ_0 is a plausible value for $\psi(\theta)$ after having observed s — hypothesis assessment problem.

Ex. Location-Scale Normal Model Inference

SRS of heights (in inches) of 30 students:

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, $\underset{\sim}{\theta} = (\mu, \sigma^2) \in \Omega = \mathbb{R}^1 \times \mathbb{R}^+$

Step 1. Use the histogram to check the normality assumption.

Let's say, we are interested in $\psi(\mu, \sigma^2) = \mu \ldots$ or $\psi(\mu, \sigma^2) = x_{0.90} = \mu + \sigma z_{0.90} \rightarrow$ 90th percentile of $N(0,1)$.

By intuition, $T(x_1, \ldots, x_n) = \bar{x}$ is sensible for μ;

$T(x_1, \ldots, x_n) = \bar{x} + s z_{0.90}$ is sensible for $\mu + \sigma z_{0.90}$ later, we'll learn to justify them. If $\bar{x} = 64.577$, $s = 2.379$, $z_{0.90} = 1.2816$,

$\Rightarrow \bar{x} + s z_{0.90} = 64.577 + 2.379(1.2816) = 67.586$.

To describe the accuracy of $\bar{x}$ as an est. of μ, we will learn to construct a CI of the form $[\bar{x} - c, \bar{x} + c]$ for some constant c ... specifically, $c = 0.3734$ will give a 95% CI for μ. The half-length of this interval is $c = 0.888$ - the measure of the accuracy of $\bar{x}$ as an est. of μ.

Finally, suppose we hypothesized μ_0 for μ ... says $\mu_0 = 65$. Does it make sense? If $\bar{x}$ is far from μ, would seem to be evidence against μ_0 ... later, we will base our assessment on

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{64.577 - 65}{2.379/\sqrt{30}} = -1.112 \ldots$$ will turn out

to be a plausible value for t