University of Washington

STATISTICS

STAT 498 B

Industrial Statistics

Applications of the Noncentral t-Distribution©

Fritz Scholz

Spring Quarter 2008

May 7, 2008

Uses of the Noncentral t-Distribution

The noncentral t–distribution is intimately tied to statistical inference procedures for samples from normal populations.

For simple random samples from a normal population the applications of the noncentral t–distribution include (extendable to regression situations):

- basic power calculations,
- variables acceptance sampling plans (MIL–STD–414)
- confidence bounds for
  - percentiles,
  - tail probabilities,
  - statistical process control parameters C_L, C_U and C_{pk} and for
  - coefficients of variation.

Checking Normality of a Sample

$X_1, \ldots, X_n$ is a random sample from $\mathcal{N}(\mu, \sigma^2)$.

CDF $F(x) = P(X_i \le x) = \Phi((x-\mu)/\sigma)$ & density $f(x) = F'(x) = \varphi((x-\mu)/\sigma)/\sigma$.

The p-quantile of $\mathcal{N}(\mu, \sigma^2)$ is $x_p = \mu + \sigma z_p$, z_p is the standard normal p-quantile.

Sort the sample $X_1, \ldots, X_n$ in increasing order $X_{(1)} \le \ldots \le X_{(n)}$ assigning fractional ranks $p_i \in (0,1)$ to these order statistics in one of several ways for $i = 1, \ldots, n$:

$$p_i = \frac{i-.5}{n} \quad \text{or} \quad p_i = \frac{i}{n+1} \quad \text{or} \quad p_i = \frac{i-.375}{n+.25}\,.$$

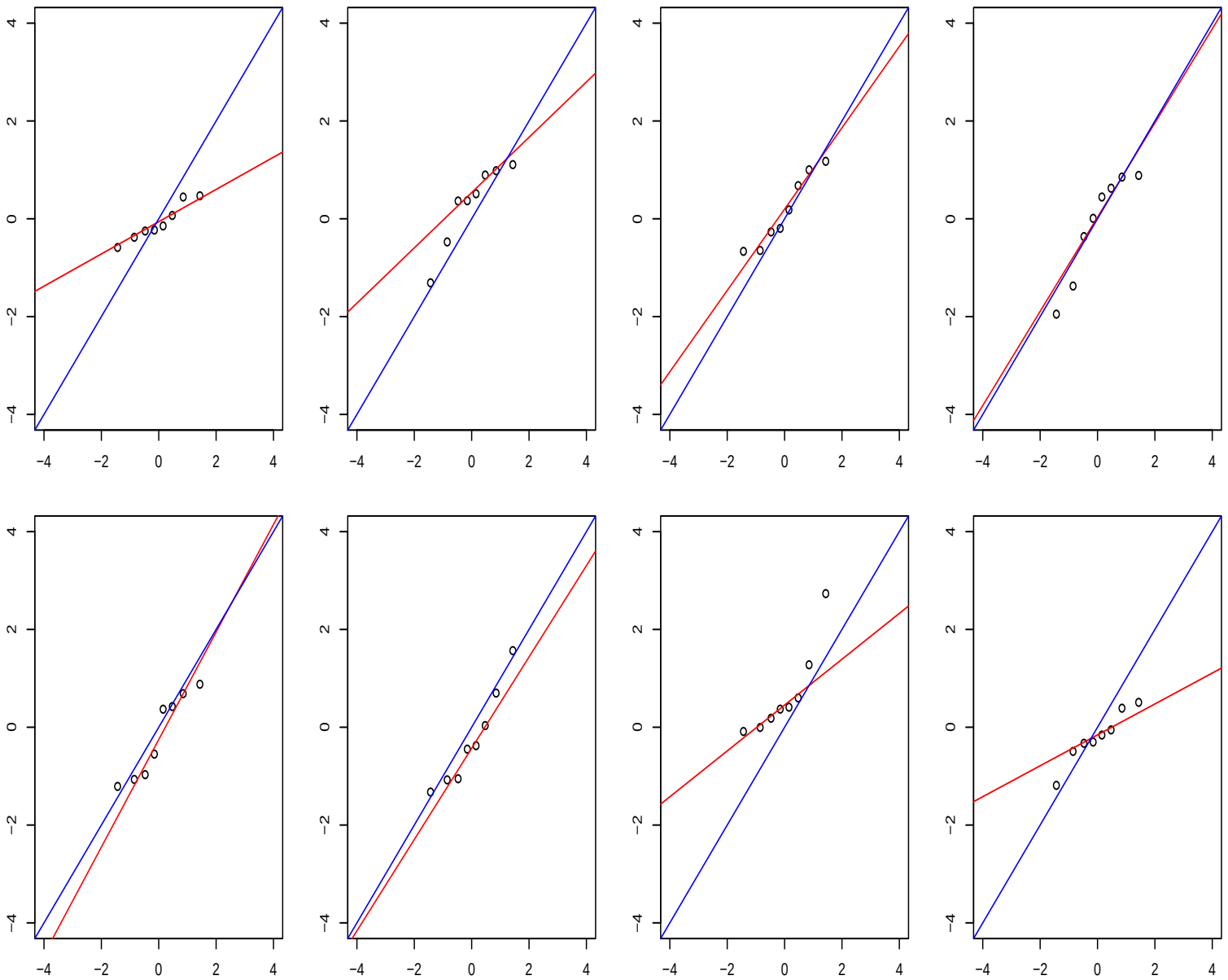
Plot $X_{(i)}$ against the standard normal p_i-quantile $z_{p_i} = \texttt{qnorm}(\texttt{p}_\texttt{i})$ for $i = 1, \ldots, n$.

We would expect $X_{(i)} \approx x_{p_i} = \mu + \sigma z_{p_i}$, i.e., $X_{(i)}$ should look $\approx$ linear against z_{p_i} with intercept $\approx \mu$ and slope $\approx \sigma$. Judging approximate linearity takes practice.

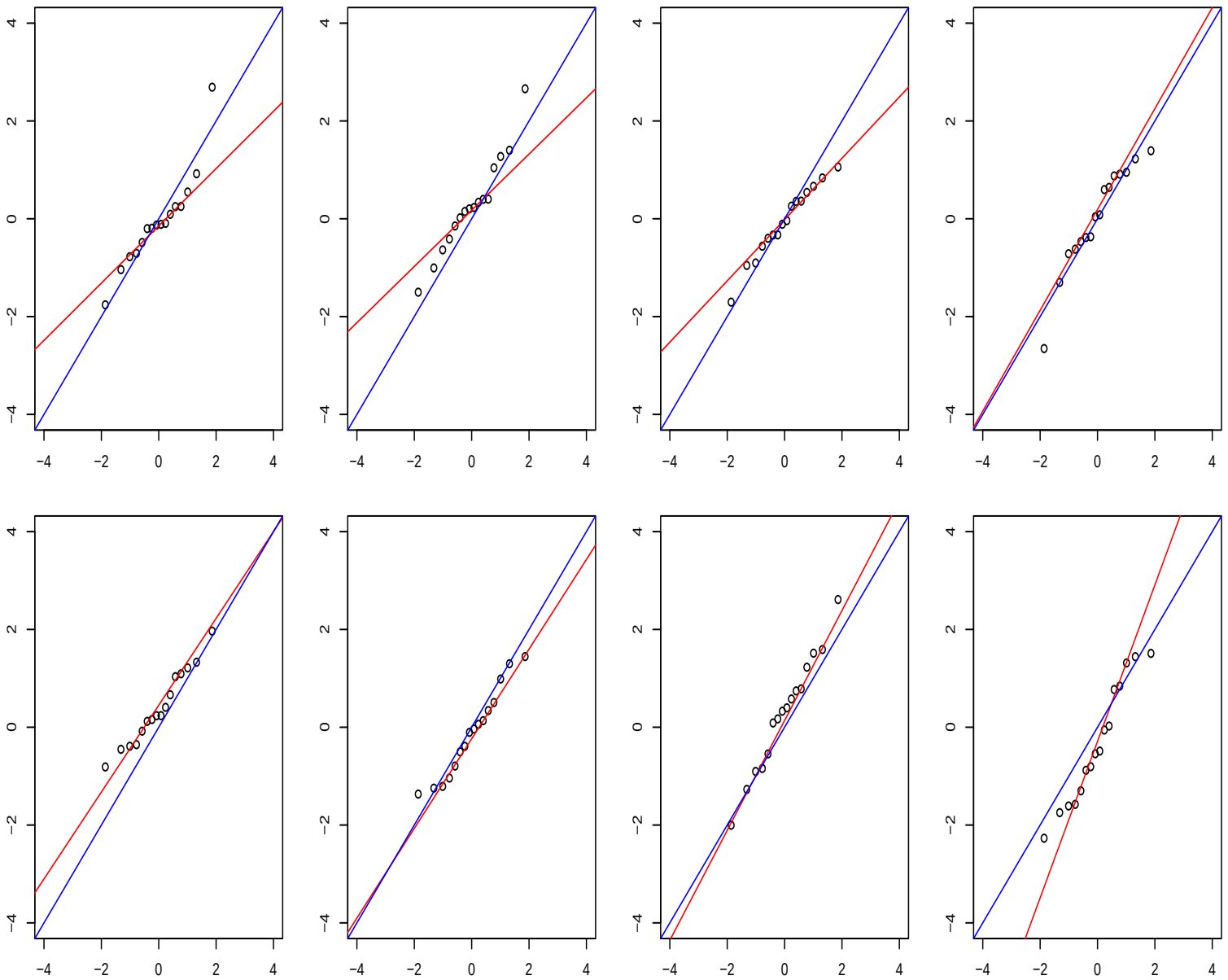
The third choice for p_i is used by R in `qqnorm(x)` for a given sample vector `x`.

`qqline(x)` (invoked after `qqnorm(x)`) fits a line to the middle half of the data.

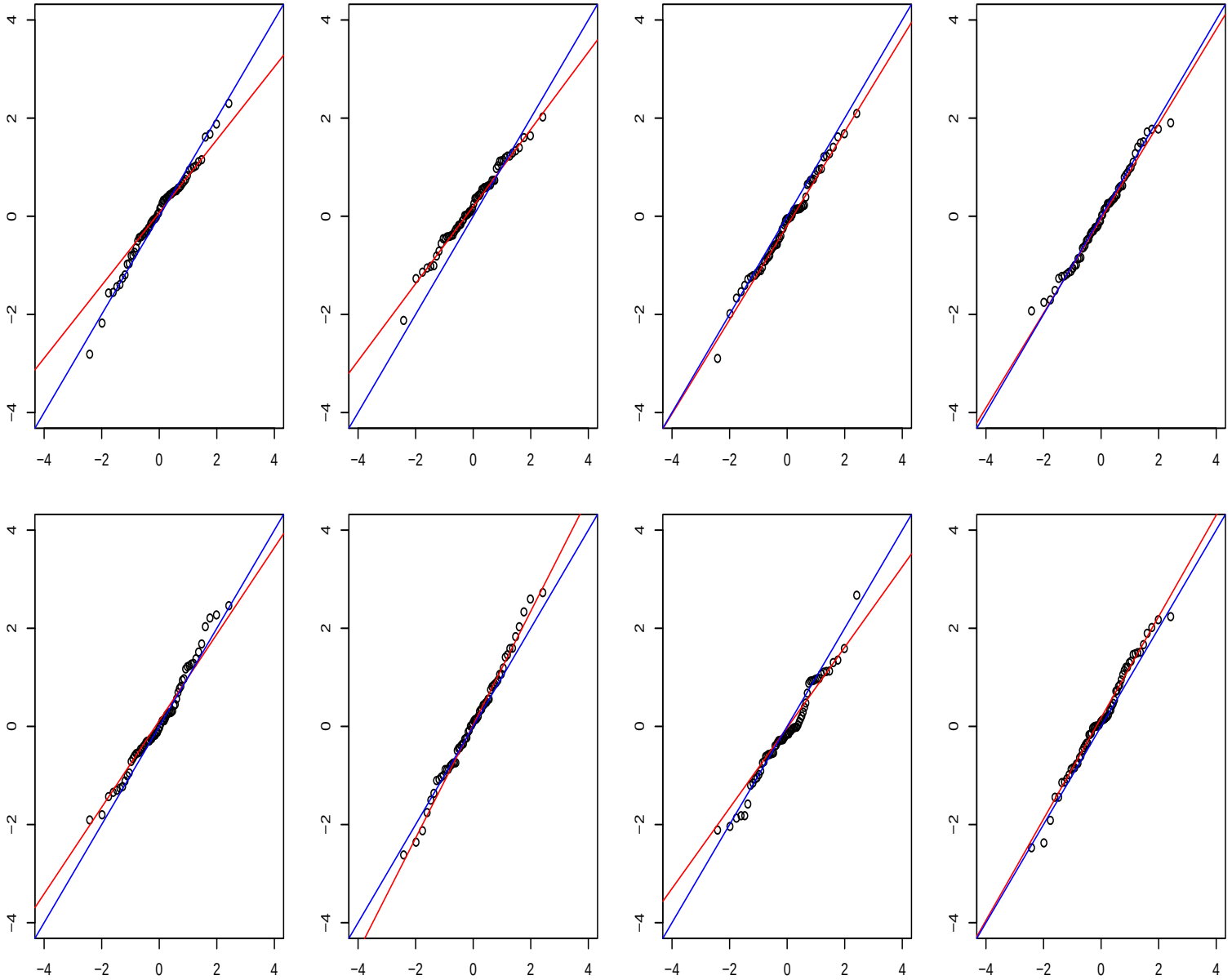
Normal QQ-Plot: $n = 8$



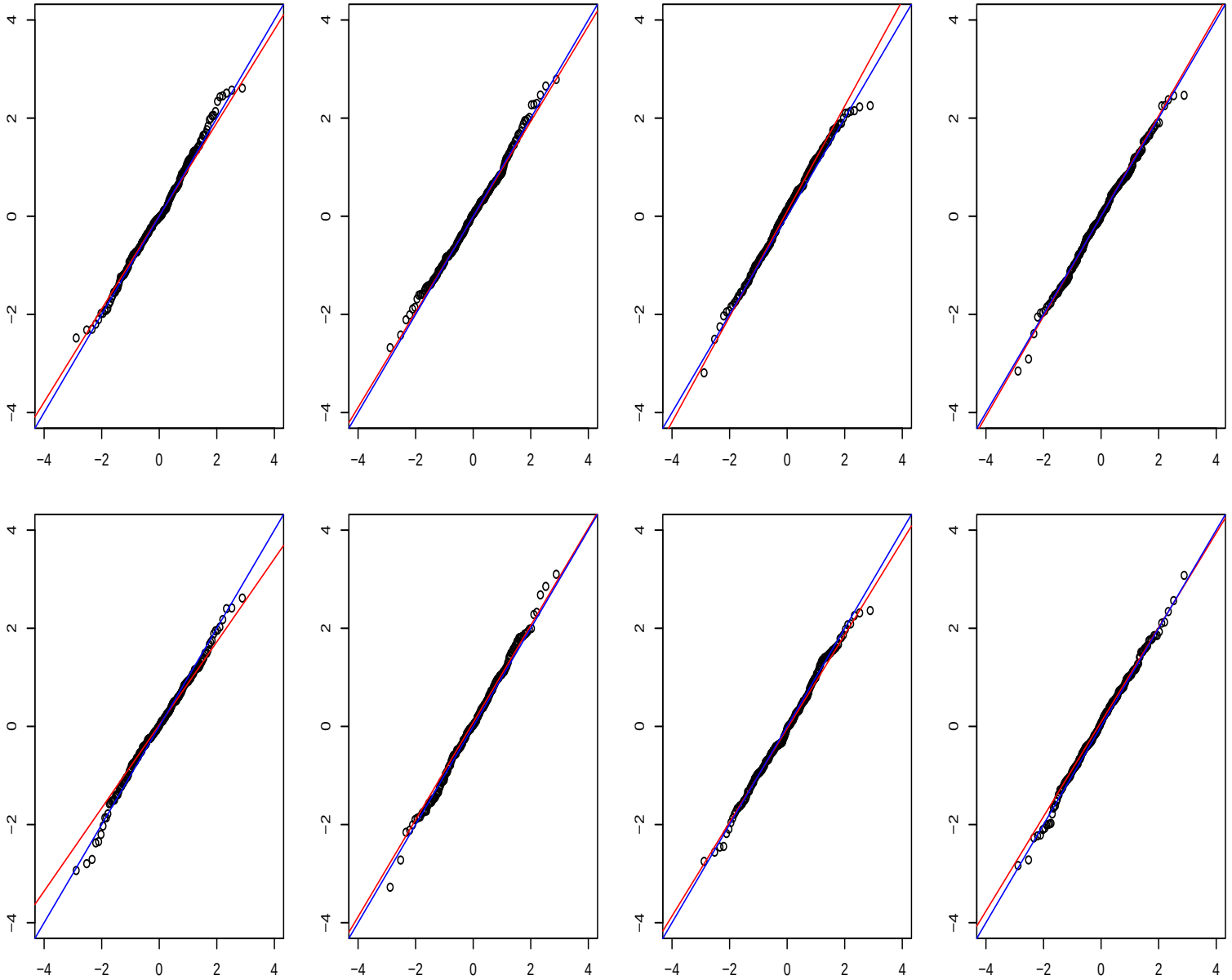
Normal QQ-Plot: $n = 16$



Normal QQ-Plot: $n = 64$



Normal QQ-Plot: $n = 256$



EDF-Based Tests of Fit

Judgment?? We can also carry out formal EDF-based tests of fit for normality.

Assume $X_1, \ldots, X_n \sim F$. Test $H_0 : F(x) = \Phi((x-\mu)/\sigma)$ for some μ and σ.

The empirical distribution function (EDF) is defined as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,\, x]}(X_i) \quad \text{with} \quad I_{(-\infty,\, x]}(X_i) = 1 \text{ or } 0 \quad \text{as} \quad X_i \le x \text{ or } X_i > x\,.$$

$F_n(x)$ is the proportion of sample values $\le x$.

Law of Large Numbers (LLN) $\implies$ $F_n(x) \overset{n\to\infty}{\longrightarrow} F(x)$ for all x.

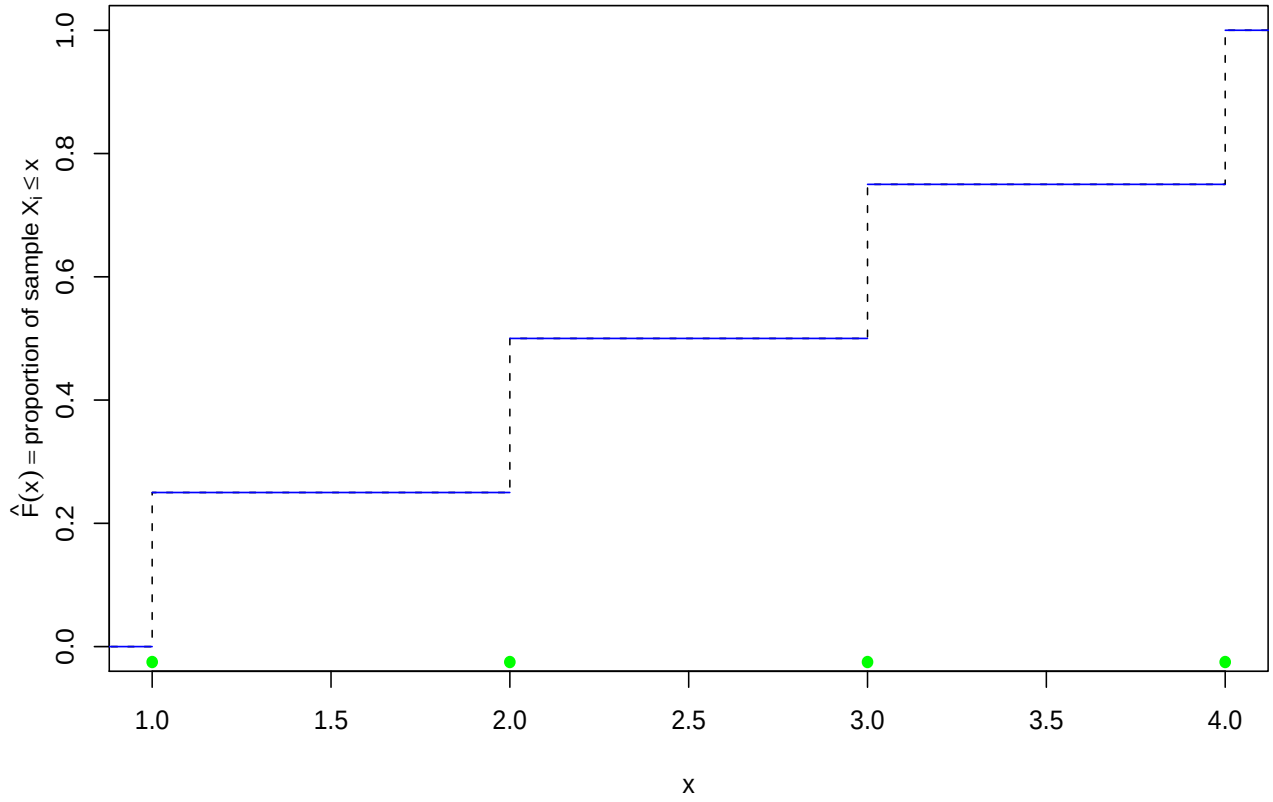
Compare $F_n(x)$ with $\hat{F}(x) = \Phi((x-\bar{X})/S)$ via some discrepancy metric $D(F_n, \hat{F})$.

Using the null distribution for $D(F_n, \hat{F})$ we reject H_0 whenever $D(F_n, \hat{F})$ is too large.

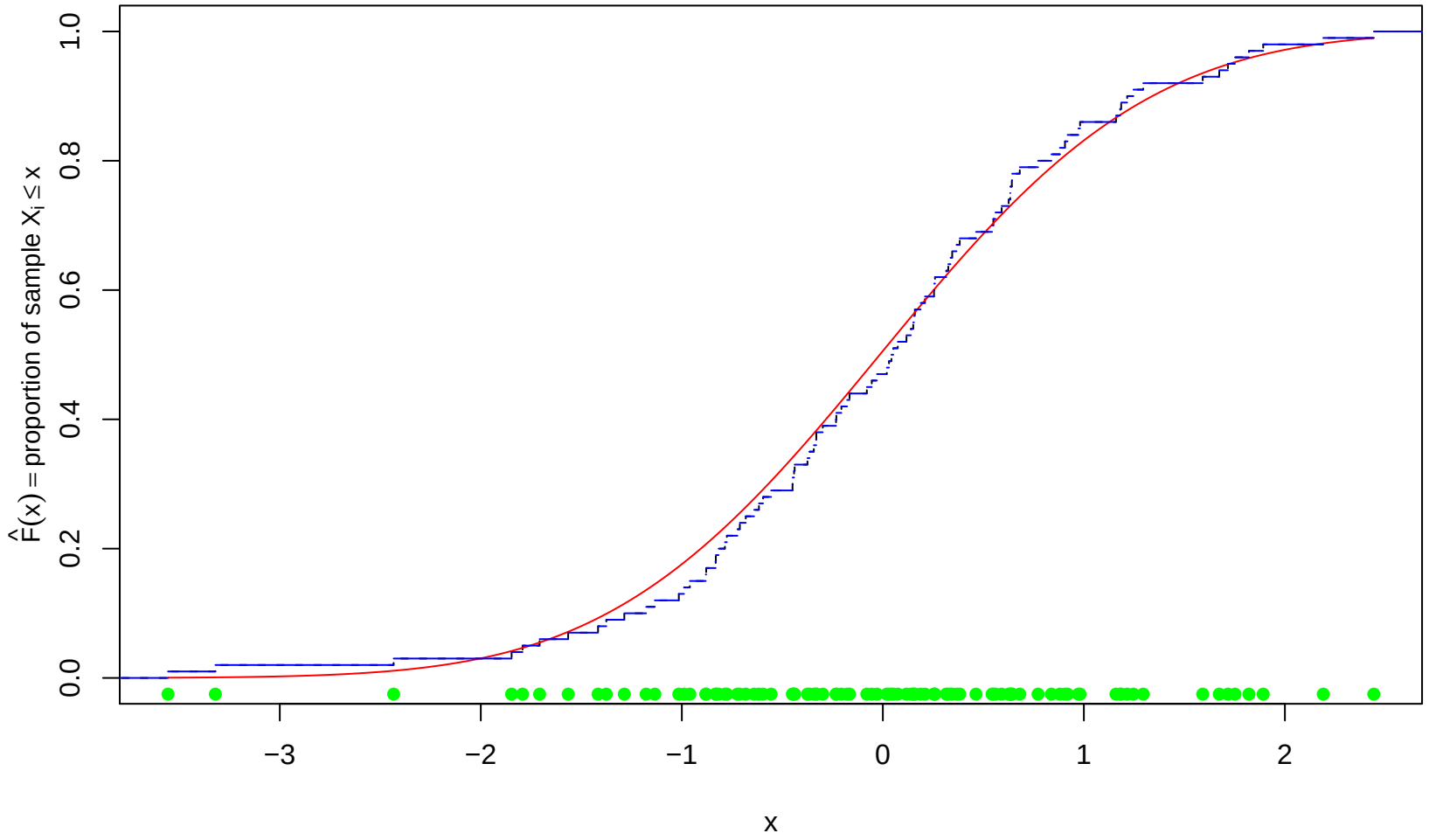
Empirical Distribution Function (EDF)

$$\hat{F}_n(x) = \frac{\text{number of } X_1, \ldots, X_n \leq x}{n}$$

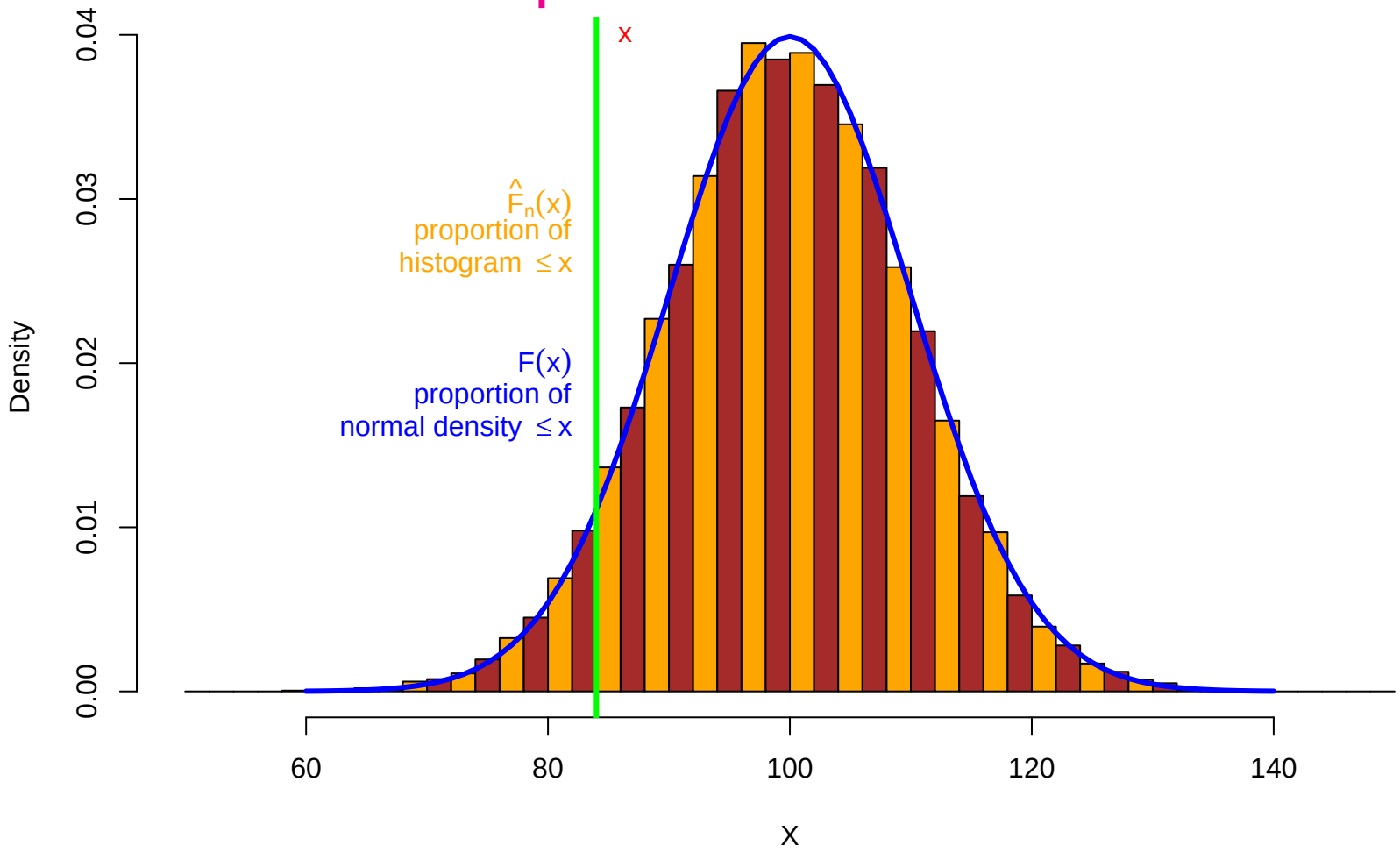
here $n = 4$, i.e., step size $1/4$



EDF of Normal Sample $n = 100$



Comparison of EDF & CDF



Comparing CDFs is not as fickle as comparing histograms with densities.

There is smoothing due to averaging.

Discrepancy Metrics

- The Kolmogorov-Smirnov metric (local discrepancy)

$$D = \max_x \left\{ \left| \hat{F}_n(x) - \Phi\left(\frac{x-\bar{X}}{S}\right) \right| \right\}$$

- The Cramer-von-Mises metric (cumulative discrepancies)

$$W^2 = \int_{-\infty}^{\infty} \left[\hat{F}_n(x) - \Phi\left(\frac{x-\bar{X}}{S}\right) \right]^2 \frac{1}{S} \varphi\left(\frac{x-\bar{X}}{S}\right) dx \quad \text{with} \quad \varphi(x) = \Phi'(x)$$

- The Anderson-Darling metric (cumulative but sensitive to tail behavior)

$$A^2 = \int_{-\infty}^{\infty} \frac{\left[\hat{F}_n(x) - \Phi\left(\frac{x-\bar{X}}{S}\right) \right]^2}{\Phi\left(\frac{x-\bar{X}}{S}\right)\left[1 - \Phi\left(\frac{x-\bar{X}}{S}\right)\right]} \frac{1}{S} \varphi\left(\frac{x-\bar{X}}{S}\right) dx$$

Computing these metrics seems challenging, but ...

Computational Formulas for Discrepancy Metrics

- The Kolmogorov-Smirnov metric

$$D = \max\left[\max\left\{\frac{i}{n} - \Phi\left(\frac{X_{(i)} - \bar{X}}{S}\right)\right\}, \max\left\{\Phi\left(\frac{X_{(i)} - \bar{X}}{S}\right) - \frac{i-1}{n}\right\}\right]$$

- The Cramer-von-Mises metric

$$W^2 = \sum_{i=1}^{n}\left\{\Phi\left(\frac{X_{(i)} - \bar{X}}{S}\right) - \frac{2i-1}{2n}\right\}^2 + \frac{1}{12n}$$

- The Anderson-Darling metric

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}\left[(2i-1)\log\left(\Phi\left(\frac{X_{(i)} - \bar{X}}{S}\right)\right)\right.$$
$$\left. + (2n+1-2i)\log\left(1 - \Phi\left(\frac{X_{(i)} - \bar{X}}{S}\right)\right)\right]$$

Installing the Package `nortest`

Approximate null distributions have been developed for all three metrics.

Goodness-of-Fit Techniques, (1986) ed. by R.B. D'Agostino and M.A. Stephens

Download `nortest_1.0.zip` from the class web site to the directory that houses your R work space.

Under the R `Packages` menu item install this package.

This installation is done only once on your computer for the installed version of R.

After this installation you need to invoke `library(nortest)` in any R session that wants to use the functions in the package `nortest`.

These functions are `lillie.test`, `cvm.test` and `ad.test` and you get documentation on them by placing a `?` in front of the respective function names, e.g., `?lillie.test`.

Kolmogorov-Smirnov Test for Normality

```
> lillie.test(rnorm(7))

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  rnorm(7)
D = 0.287, p-value = 0.08424

> lillie.test(runif(137))

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  runif(137)
D = 0.0877, p-value = 0.01169
```

Anderson-Darling Test for Normality

```
> ad.test(rnorm(10))

        Anderson-Darling normality test

data:  rnorm(10)
A = 0.4216, p-value = 0.2572

> ad.test(runif(30))

        Anderson-Darling normality test

data:  runif(30)
A = 0.8551, p-value = 0.02452
```

Some Comments

For $n = 8$ QQ-plots can exhibit strong nonlinear patterns. It improves as $n \nearrow$.

For large n one can still expect some fluctuating behavior in the tails. That is not unusual and should not necessarily be construed as evidence of nonlinearity and thus nonnormality.

Intuitively such sample tail fluctuations can be understood by the fact that near the sample extremes the data are not hemmed in quite as strongly as they are in the main part of the sample.

When QQ-plots are not clearly linear use formal EDF goodness-of-fit tests to clarify the issue. However, even such tests may then give an ambiguous verdict.

Do both routinely, the QQ-plot for visual impression of the data and the EDF goodness-of-fit test (sample selection bias would invalidate the p-values).

Definition of the Noncentral t-Distribution

If $Z \sim \mathcal{N}(0,1)$ and $V \sim \chi^2_f$ are (statistically) independent then the ratio

$$T_{f,\,\delta} = \frac{Z+\delta}{\sqrt{V/f}}$$

is said to have a noncentral t-distribution with f degrees of freedom and noncentrality parameter δ.

Although $f \geq 1$ originally was intended to be an integer closely linked to sample size, it is occasionally useful to extend its definition to any real $f > 0$.

The noncentrality parameter δ may be any real number.

The cdf of $T_{f,\,\delta}$ is denoted by $G_{f,\,\delta}(t) = P(T_{f,\,\delta} \leq t)$.

$\delta = 0 \implies G_{f,\,0}(t)$ is the usual central or Student t cdf.

Properties of the Noncentral t-Distribution

$G_{f,\,\delta}(t)$ increases strictly from 0 to 1 as t increases from $-\infty$ to $+\infty$.

(standard property of any cdf with positive density)

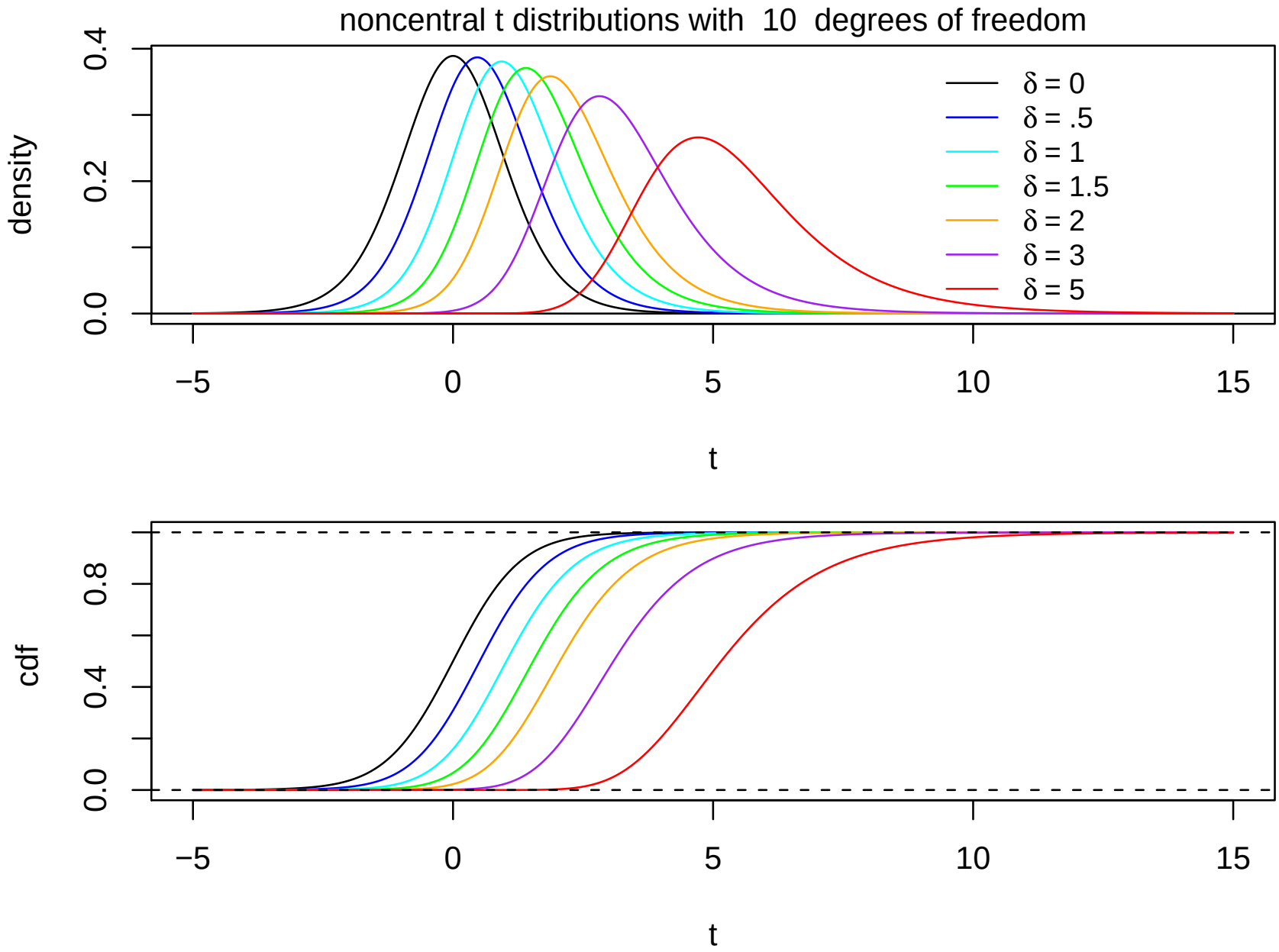
$G_{f,\,\delta}(t)$ decreases strictly from 1 to 0 as δ increases from $-\infty$ to $+\infty$.

$$G_{f,\,\delta}(t) = P\left(\frac{Z+\delta}{\sqrt{V/f}} \le t\right) = P\left(Z - t\sqrt{V/f} \le -\delta\,.\right)$$

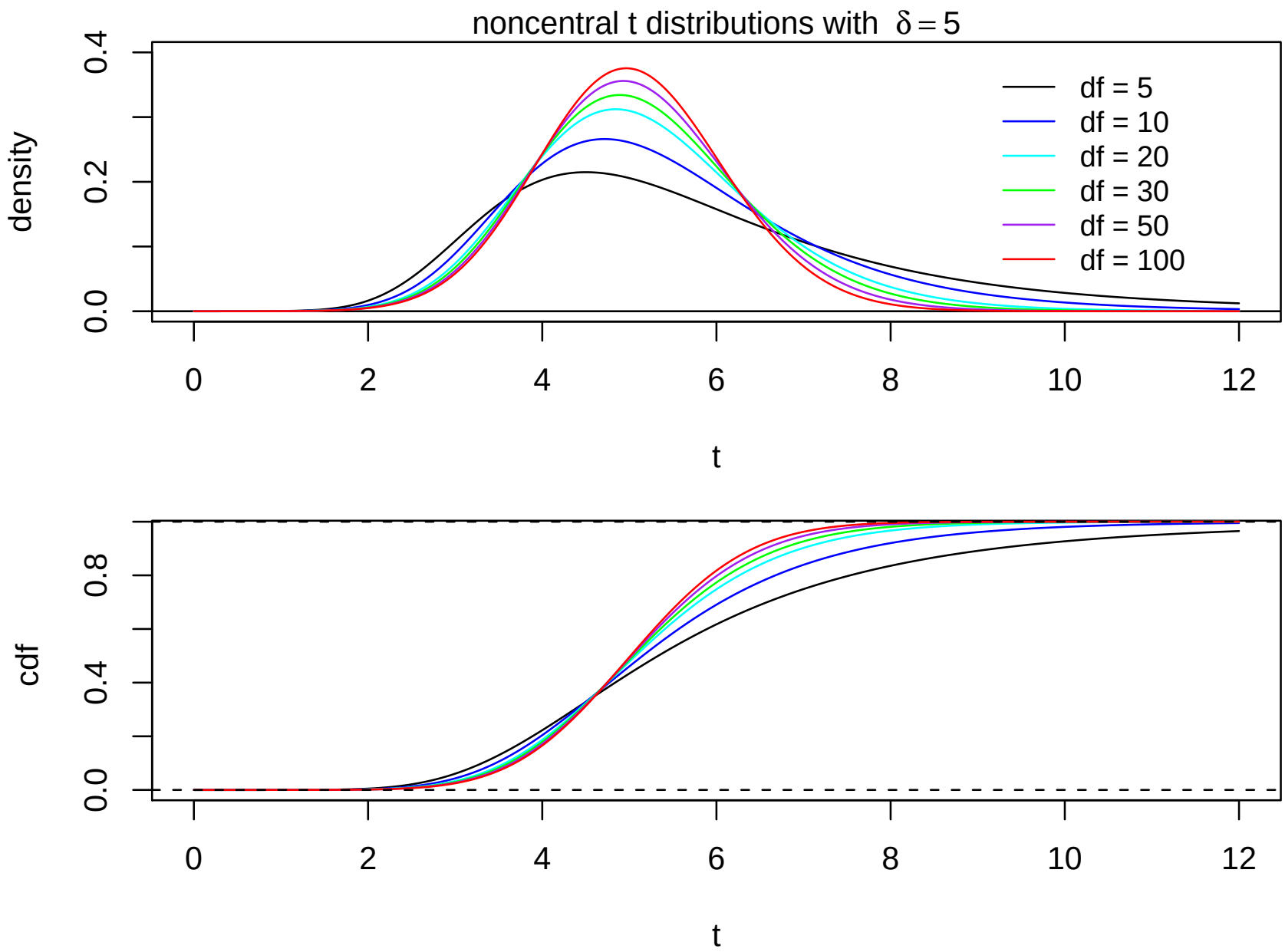
We have the following identity relating $G_{f,-\delta}$ to $G_{f,\delta}$

$$\begin{aligned} G_{f,-\delta}(-t) &= P\left(\frac{Z-\delta}{\sqrt{V/f}} \le -t\right) = P\left(\frac{-Z+\delta}{\sqrt{V/f}} \ge t\right) \\ &= P\left(\frac{Z+\delta}{\sqrt{V/f}} \ge t\right) = 1 - P\left(\frac{Z+\delta}{\sqrt{V/f}} \le t\right) = 1 - G_{f,\delta}(t) \end{aligned}$$

The Noncentral t-Distribution and δ



The Noncentral t-Distribution and f



Basic Normal Sample Situation

Assume $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \text{and} \qquad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}\,.$$

$\bar{X}$ and S are statistically independent

$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ or equivalently $Z = \sqrt{n}(\bar{X} - \mu)/\sigma \sim \mathcal{N}(0,1)$

$V = (n-1)S^2/\sigma^2 \sim \chi^2_f$ with $f = n-1$ degrees of freedom

V and Z are statistically independent.

Canonical Use of the Noncentral t-Distribution

All one–sample applications involving the noncentral t–distribution can be reduced to calculating the following probability

$$\gamma = P(\bar{X} - aS \le b) \; .$$

$$\bar{X} - aS \le b \iff \frac{\sqrt{n}(\bar{X}-\mu)/\sigma - \sqrt{n}(b-\mu)/\sigma}{S/\sigma} \le a\sqrt{n} \iff T_{f,\,\delta} \stackrel{\text{def}}{=} \frac{Z+\delta}{\sqrt{V/f}} \le a\sqrt{n}$$

with $f = n-1$, $\delta = -\sqrt{n}(b-\mu)/\sigma$, and with Z and V as defined previously in terms of $\bar{X}$ and S. Thus

$$\gamma = P(T_{f,\,\delta} \le a\sqrt{n}) = G_{f,\,\delta}(a\sqrt{n}) \; .$$

Three of the four parameters n, a, δ and γ are usually given and the fourth needs to be determined either by direct computation of $G_{f,\,\delta}(t)$ or by root solving techniques, using `qnct` or `del.nct`, or by iterative trial and error with n.

All referenced R functions are part of the work space provided on the web.

The One-Sample t-Test

Assuming $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ consider the following testing problem.

A hypothesis $H : \mu \leq \mu_0$ is tested against the alternative $A : \mu > \mu_0$.

The intuitive and in many ways optimal procedure rejects H in favor of A whenever

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \geq t_{n-1}(1-\alpha) \qquad \text{or equivalently when} \qquad \bar{X} - \frac{t_{n-1}(1-\alpha)\, S}{\sqrt{n}} \geq \mu_0 \,.$$

$t_{n-1}(1-\alpha)$ is the $(1-\alpha)$-percentile of the central t-distribution with $n-1$ df.

The test has chance $\leq \alpha$ of rejecting H when $\mu \leq \mu_0$, i.e., when H is true.

As will become clear below, the chance of rejection is $< \alpha$ when $\mu < \mu_0$.

It is $= \alpha$ when $\mu = \mu_0$.

Thus α is the maximum chance of rejecting H falsely,

i.e., the maximum type I error probability.

The Power Function of the One-Sample t-Test

An important characteristic of a test is its power function, which is defined as the probability of rejecting H as a function of (μ, σ), i.e.,

$$\beta(\mu, \sigma) = P_{\mu, \sigma}\left(\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \geq t_{n-1}(1-\alpha)\right) .$$

For $\mu > \mu_0$ the value of $1 - \beta(\mu, \sigma)$ represents the probability of falsely accepting H, i.e., the probability of type II error.

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma + \sqrt{n}(\mu - \mu_0)/\sigma}{S/\sigma} = \frac{Z + \delta}{\sqrt{V/(n-1)}} ,$$

$$\Longrightarrow \beta(\mu, \sigma) = \beta(\Delta) = P_{\mu, \sigma}\left(\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \geq t_{n-1}(1-\alpha)\right) = 1 - G_{n-1, \delta}(t_{n-1}(1-\alpha)) ,$$

strictly increasing in $\delta = \sqrt{n}(\mu - \mu_0)/\sigma = \sqrt{n}\Delta$ with $\Delta = (\mu - \mu_0)/\sigma$.

Some Comments

With increasing n the noncentrality parameter δ can become arbitrarily large.

Thus we will reject H for any alternative $\mu > \mu_0$ with probability increasing to 1, no matter how close μ is to μ_0 and no matter how large σ is.

Of course one should address the practical significance issue of any difference $\mu - \mu_0$ and weigh that against the cost of a large sample size.

In doing so, the magnitude of $\mu - \mu_0$ would typically be judged in relation to the inherent population variability σ.

$\beta(\mu, \sigma) = \beta(\Delta)$ depends on μ and σ only through $\Delta = (\mu - \mu_0)/\sigma$.

Minimum Sample Size Determination

What is the minimum sample size n to achieve power β for a specific $\Delta = \Delta_1$?
This also controls the type II error probability $1 - \beta$.

Problem: The power function depends on n not only through $\delta = \sqrt{n}\Delta_1$ but also through the degrees of freedom in $t_{n-1}(1-\alpha)$ and in the cdf $G_{n-1,\,\delta}$.

The smallest n for which $\beta(\Delta_1) = \beta$ can be found through iteration, starting with a crude initial guess $\tilde{n} = \Big((z_\beta - z_\alpha)/\Delta_1\Big)^2$ rounded up to the next integer.

Here z_p denotes the p-quantile of the standard normal distribution.

This crude initial guess is based on treating the noncentral t-distribution as a $\mathcal{N}(\delta, 1)$ distribution, which it approaches as n gets large.

Crude Initial Guess

$$\beta(\Delta) = P_\Delta\left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{S} \geq k\right) \approx P_\Delta\left(Z+\delta \geq k\right)$$

$$\alpha = P_{\Delta=0}\left(Z+\delta \geq k\right) = P(Z \geq k)$$

$$\Longrightarrow\ k = z_{1-\alpha} = -z_\alpha \qquad \text{with} \quad z_p = \Phi^{-1}(p).$$

$$\beta = P_{\Delta_1}\left(Z+\delta \geq k\right) = 1-\Phi(k-\sqrt{n}\Delta_1) \ \Longrightarrow\ z_{1-\beta} = -z_\beta = k-\sqrt{n}\Delta_1$$

$$\Longrightarrow\ \sqrt{n}\Delta_1 = z_\beta + k = z_\beta - z_\alpha \ \Longrightarrow\ n = \left(\frac{z_\beta - z_\alpha}{\Delta_1}\right)^2$$

The R Function `min.sample.size`

The R function `min.sample.size` (available in the R work space) carries out this iterative process and reports the initial $\tilde{n}$ and resulting initial power, in addition to the final n and its achieved power $\geq \beta$.

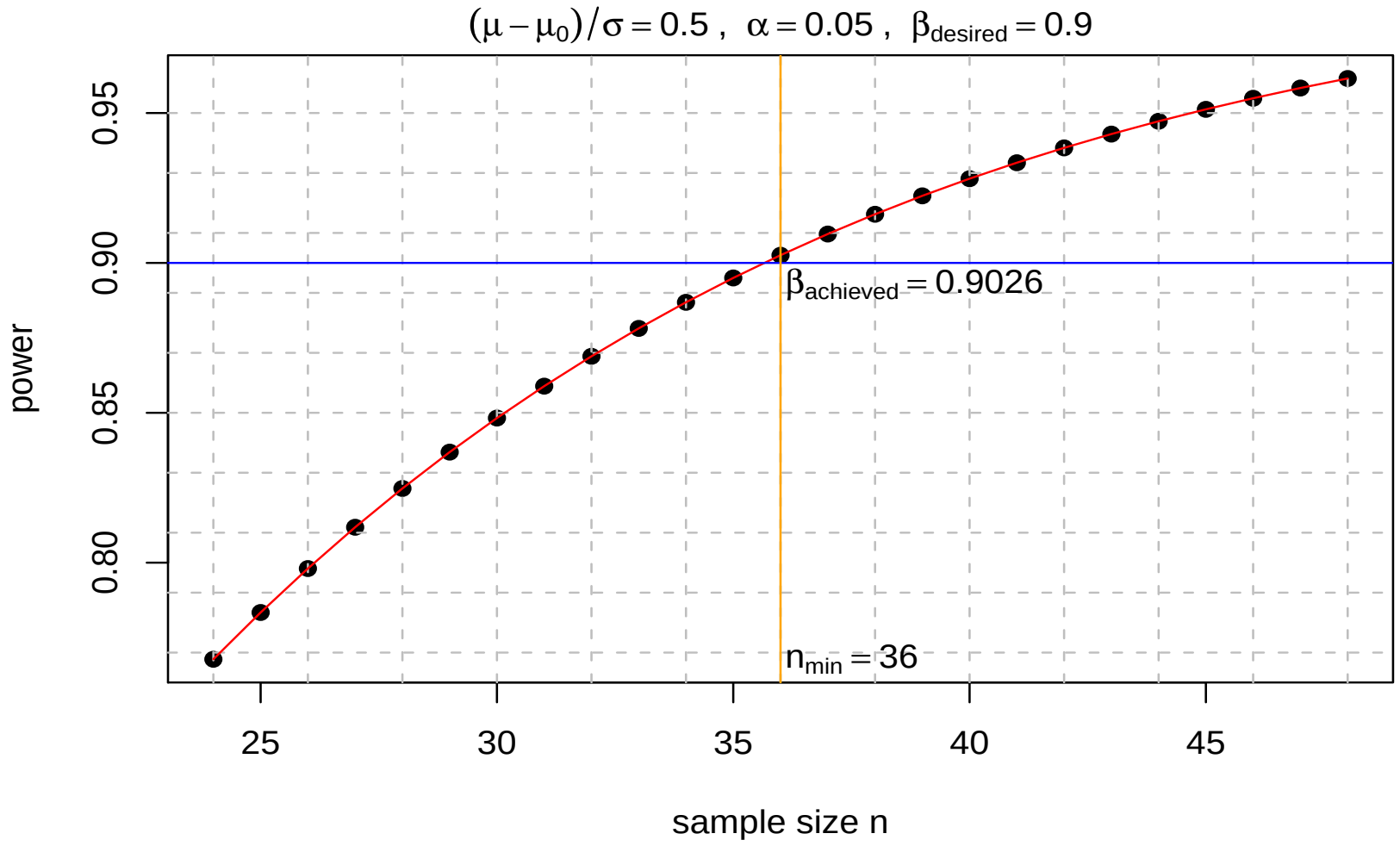
Please study the implementation of `min.sample.size`.

This function also produces the following plots.

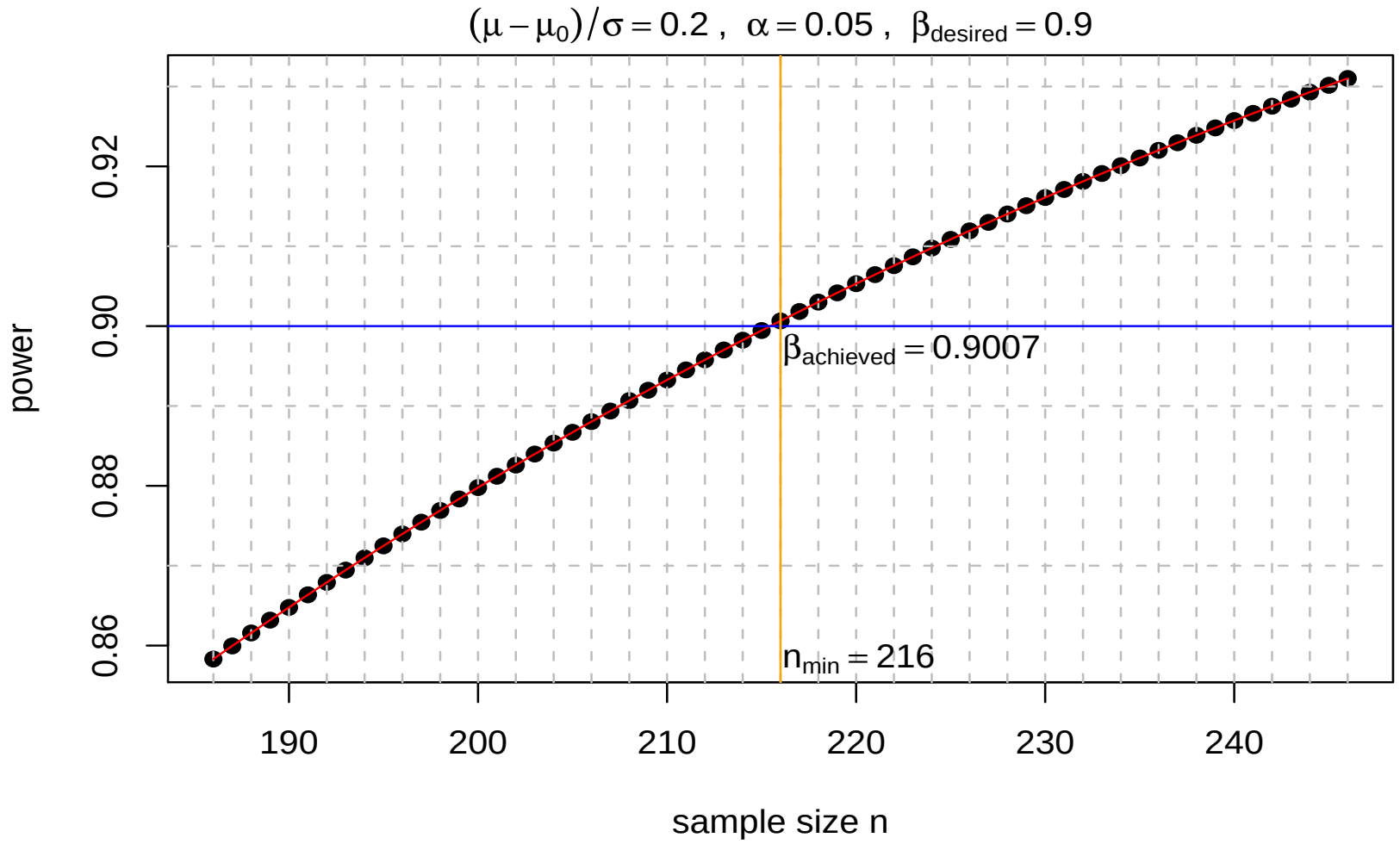
Similarly deal with the dual problem of testing the hypothesis $H' : \mu \geq \mu_0$ against the alternative $A' : \mu < \mu_0$.

The modifications, which consist of reversing certain inequalities, e.g., rejecting H' when $\sqrt{n}(\bar{X} - \mu_0)/S \leq t_{n-1}(\alpha)$, are straightforward and are omitted.

Sample Size Determination Plots



Sample Size Determination Plots



Two-Sided Alternatives

Testing $H^\star : \mu = \mu_0$ against the alternative $A^\star : \mu \neq \mu_0$ the relevant test rejects $H^\star$ in favor of $A^\star$ whenever

$$\frac{\sqrt{n}|\bar{X} - \mu_0|}{S} \geq t_{n-1}(1 - \alpha/2) \; .$$

The power function $\beta(\mu, \sigma)$ of this test is

$$P_{\mu,\,\sigma}\left(\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \leq -t_{n-1}(1-\alpha/2) \;\text{ or }\; \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \geq t_{n-1}(1-\alpha/2) \right)$$

$$= G_{n-1,\,\delta}(-t_{n-1}(1-\alpha/2)) + 1 - G_{n-1,\,\delta}(t_{n-1}(1-\alpha/2)) = \beta^\star(\mu,\, \sigma) \; ,$$

where $\delta = \sqrt{n}(\mu - \mu_0)/\sigma$.

The power function $\beta^\star(\mu,\, \sigma) = \beta^\star(|\delta|)$ is strictly increasing in $|\delta|$.

`min.sample.size` $\Longrightarrow$ minimum n for H' vs A' and for $H^\star$ vs $A^\star$.

Variables Acceptance Sampling Plans

Quality control applications governed by MIL–STD–414 deal with

variables acceptance sampling plans (VASP).

In a VASP the quality of sampled items is measured on a quantitative scale.

An item is judged defective when its measured quality exceeds a certain threshold.

The samples are drawn randomly from a population of items.

Objective: Make inferences about the proportion of defectives in the population.

$\Longrightarrow$ acceptance or a rejection of the population quality as a whole.

Meaning of Population

In various applications the term “population” can have different meanings.

It represents that collective of items from which the sample is drawn.

It could be a shipment, a lot or a batch or any other collective entity.

For the purpose of this discussion the term “population” will be used throughout.

Any batch, lot or shipment consists of items that come from a certain process.

If that process were to run indefinitely it would produce an infinite population of such items. Thus the sampled items from the batch, lot or shipment could be considered as a sample from that larger conceptual population.

If the sample indicates that something is wrong the producer would presumably adjust the process appropriately.

Assumptions

A VASP assumes that measurements (variables) $X_1, \ldots, X_n$ for a random sample of n items from a population are available.

Item i is defective $\Longleftrightarrow X_i < L$, where $L =$ given lower specification limit.

Or, item i is defective $\Longleftrightarrow X_i > U$, where $U =$ given upper specification limit.

The methodology of any VASP depends on the assumed underlying distribution for the measured variables $X_1, \ldots, X_n$.

Here we assume that we deal with a random sample from a normal population with mean μ and standard deviation σ.

The following discussion will be in terms of a lower specification limit L.

The corresponding procedure for an upper specification limit U is only summarized without derivation.

Consumer/Producer Interests

If L is a lower specification limit, then

$$p = p(\mu, \sigma, L) = P_{\mu,\,\sigma}(X < L) = P_{\mu,\,\sigma}\left(\frac{X-\mu}{\sigma} < \frac{L-\mu}{\sigma}\right) = \Phi\left(\frac{L-\mu}{\sigma}\right)$$

represents the probability that a random item in the population will be defective.

p can be interpreted as the proportion of defective items in the population.

It is in the consumer's interest to keep the proportion p of defective items in the population below a tolerable value p_1.

Keeping the proportion p low is typically costly for the producer.

Hence the producer will try too keep p only so low as to remain cost effective but sufficiently low as not to trigger too many costly rejections.

Hence the producer will aim for keeping $p \leq p_0$ ($< p_1$, in order to provide a sufficient margin between producer and consumer interest).

Producer Positioning

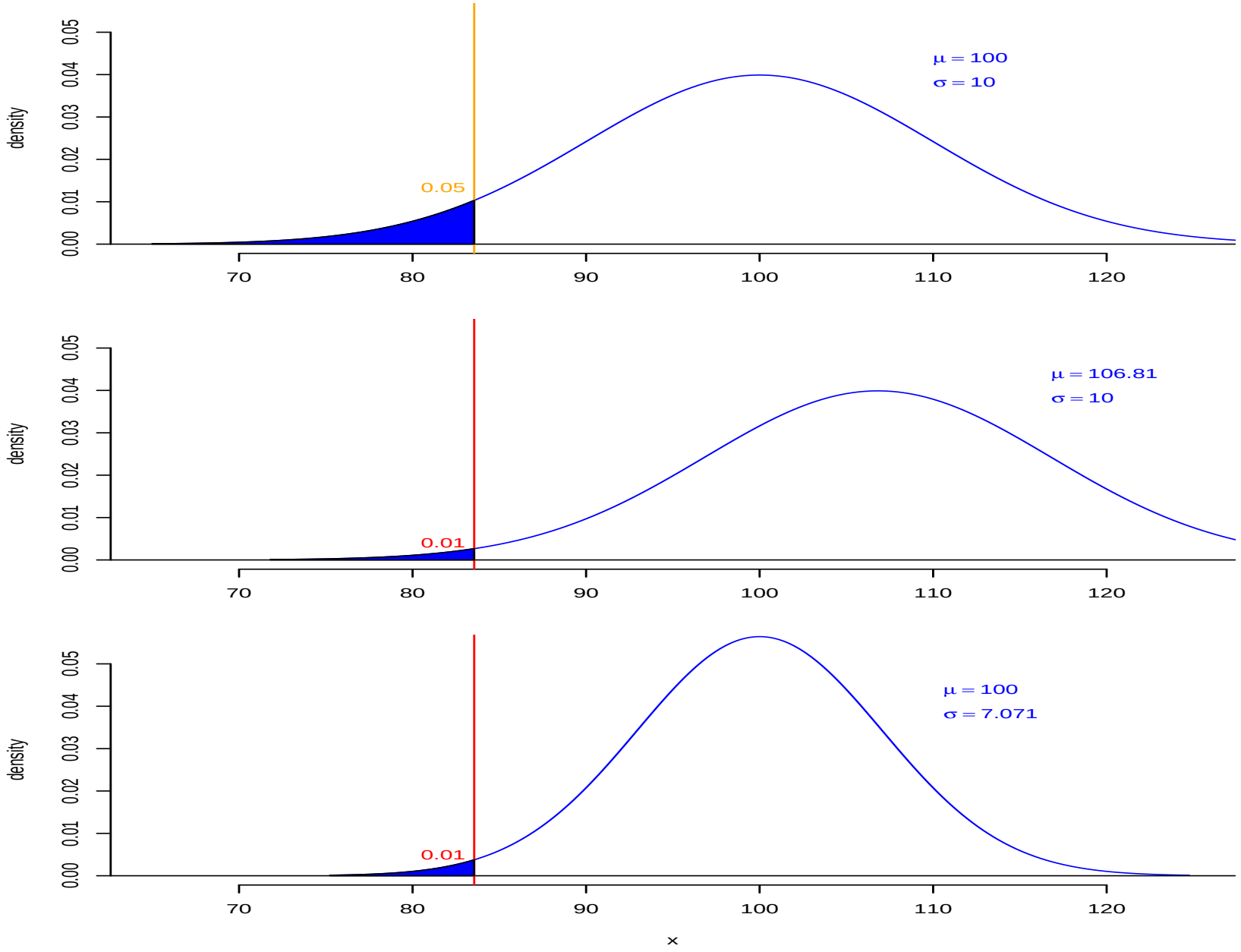
The consumer's demand $p \leq p_1$ does not specify how to accomplish this in terms of μ and σ.

The producer can control $p \leq p_0$ by either increasing μ sufficiently or by reducing σ, provided $\mu > L$.

Reducing σ is usually more difficult since sources of variation have to be controlled more tightly.

Increasing μ is mainly a matter of biasing the process in some way and is usually easier to accomplish.

Corrective Options



The Basic VASP Process

The standard VASP consists in computing $\bar{X}$ and S from the obtained sample of n items and in comparing $\bar{X} - kS$ with L for an appropriately chosen constant k.

If $\bar{X} - kS \geq L$, the consumer accepts the population from which the sample was drawn and otherwise it is rejected.

Rejection/acceptance is not based on the sample proportion of items with $X_i < L$.

Such classification ignores how far above or below L each measurement X_i is.

Basing decisions on just such attributes $X_i < L$ or $X_i \geq L$ is much less effective than using the values X_i in their entirety to estimate the underlying normal population and from that get a better idea about p for much smaller sample size.

Attribute data should only be used when the direct measurements are not available or not feasible. In that case one needs to employ attribute sampling plans based on the binomial distribution, requiring typically much higher sample sizes.

Consumer and Producer Risks

Due to the random nature of the sample there is some chance that the sample misrepresents the population at least to some extent and thus may induce us to take incorrect action.

The consumer's risk is the probability of accepting the population when in fact the proportion p of defectives in the population is greater than the acceptable limit p_1.

The producer's risk is the probability of rejecting the population when in fact the proportion p of defectives in the population is $\leq p_0$.

The probability of acceptance for a given VASP(k) depends on μ, σ, L only through $p = \Phi((L-\mu)/\sigma)$, the proportion of defectives in the population.

This function will thus be denoted by $\gamma(p)$. It is also known as operating characteristic or OC–curve of the VASP.

The OC-Curve

$\gamma(p)$ can be expressed in terms of $G_{n-1,\,\delta}(t)$ as follows:

$$\begin{aligned}\gamma(p) &= P_{\mu,\,\sigma}(\bar{X} - kS \geq L) = P_{\mu,\,\sigma}\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} + \frac{\sqrt{n}(\mu - L)}{\sigma} \geq k\sqrt{n}\,\frac{S}{\sigma}\right) \\ &= P_{\mu,\,\sigma}\left(\frac{Z+\delta}{\sqrt{V/(n-1)}} \geq k\sqrt{n}\right) = P(T_{n-1,\,\delta} \geq k\sqrt{n})\end{aligned}$$

where the noncentrality parameter

$$\delta = \delta(p) = \frac{\sqrt{n}\,(\mu - L)}{\sigma} = -\sqrt{n}\,\frac{L-\mu}{\sigma} = -\sqrt{n}\,\Phi^{-1}(p) = -\sqrt{n}\,z_p$$

depends on μ, σ and L only through p. This greatly streamlines such VASPs.

The Choice of k for Consumer Risk

$$\delta(p) = -\sqrt{n}\, z_p \;\searrow \text{ in } p \quad \Longrightarrow \quad \gamma(p) = 1 - G_{n-1,\,\delta(p)}(k\sqrt{n}) \;\searrow \text{ in } p$$

To control the consumer's risk, $\gamma(p)$ has to be kept $\leq \beta$ for $p \geq p_1$.

Since $\gamma(p)$ is decreasing in p, we need to insure $\gamma(p_1) = \beta$ by proper choice of k.

The factor k is then found by solving the equation

$$\beta = 1 - G_{n-1,\,\delta(p_1)}(k\sqrt{n}) \qquad \text{for } k, \text{ i.e.,} \qquad k = G^{-1}_{n-1,\,\delta(p_1)}(1-\beta)/\sqrt{n}\,.$$

This is accomplished in R by the command

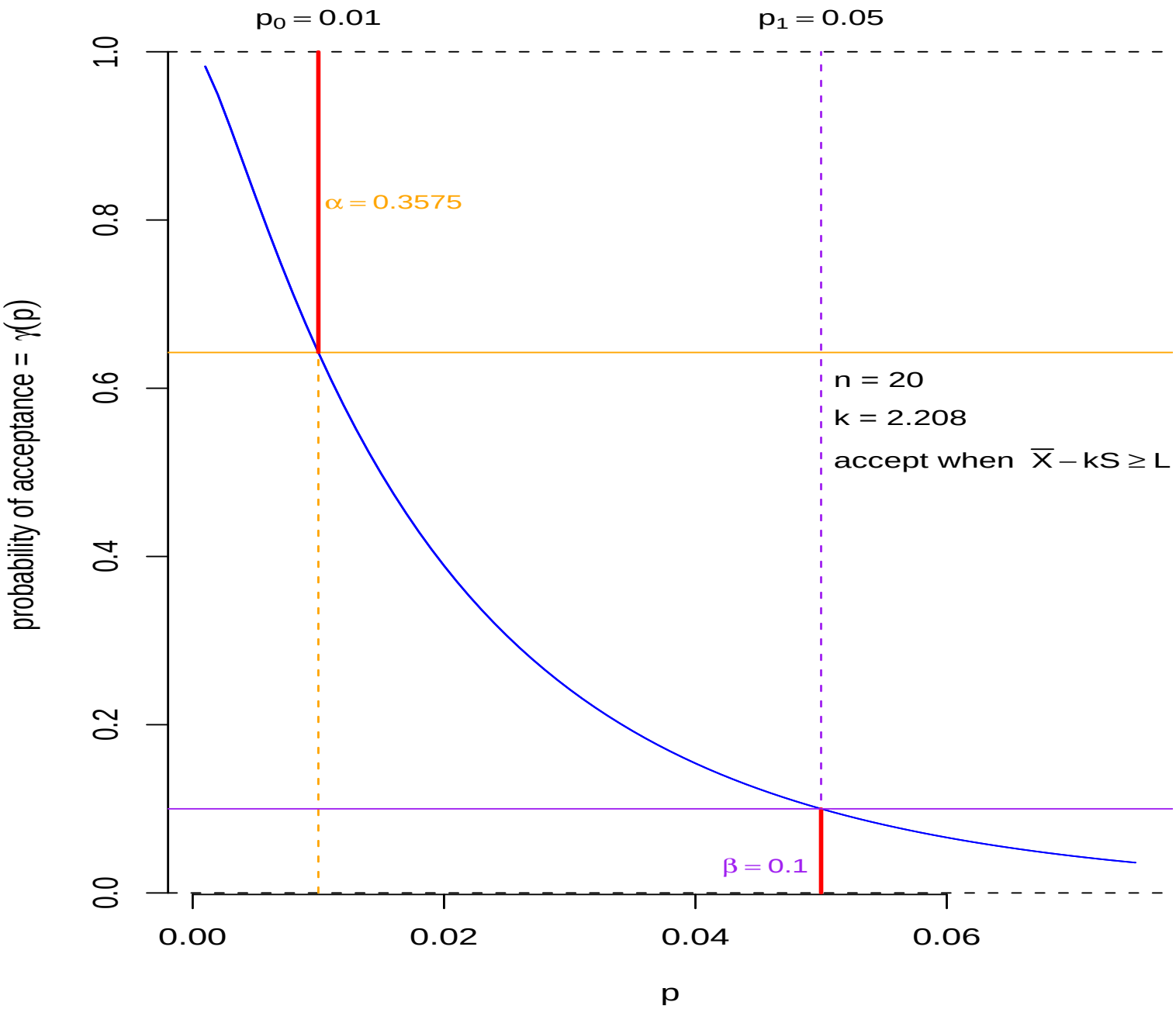
```
k=qnct(1-beta,n-1,-sqrt(n)*qnorm(p1))/sqrt(n) ,
```

where `beta` $= \beta$ and `p1` $= p_1$. It is customary to choose $\beta = .10$.

`qnct` is not intrinsic to R, it was added by me to the supplied R work space.

`OC.curve.n1` shows the resulting k and the OC-curve when $n = 20$.

The Choice of k: Controlling Consumer Risk β



The Choice of k for Producer Risk

This solves the problem as far as the consumer is concerned.

It does not address the producer's risk requirements.

The producer's risk is $1-\gamma(p)$, maximal for $p \le p_0$ at $p = p_0$.

In the previous plot that risk is as high as $.3575$ when $p_0 \le .01$.

The producer wants to limit $1-\gamma(p_0)$ by some value α, customarily $\alpha = .05$.

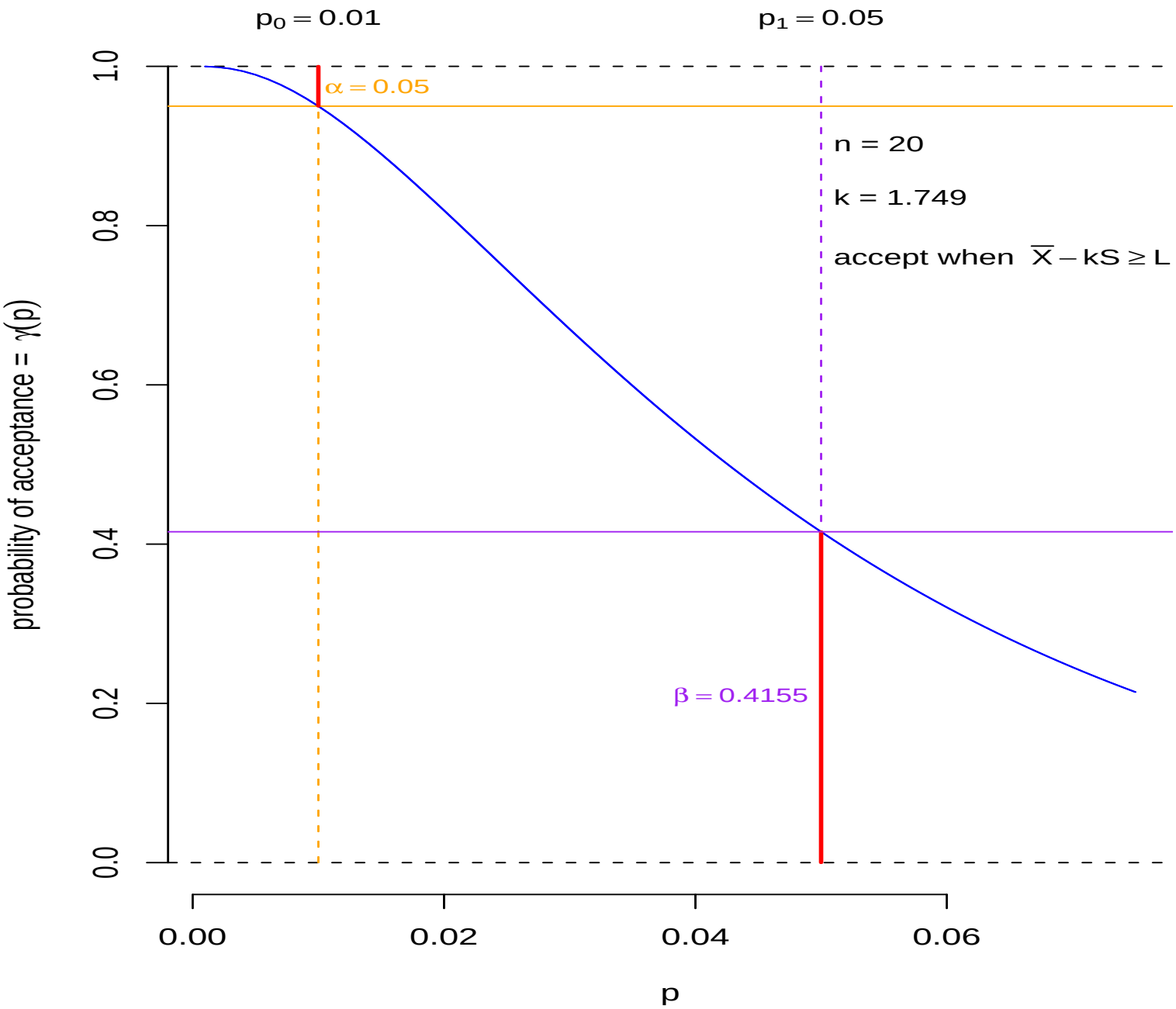
Solving

$$\alpha = 1-\gamma(p_0) = G_{n-1,\,\delta(p_0)}(k\sqrt{n}) \qquad \text{for } k, \text{ i.e., } \quad k = G^{-1}_{n-1,\,\delta(p_0)}(\alpha)/\sqrt{n}\,.$$

or $\qquad \texttt{k} = \texttt{qnct(alpha, n-1, -sqrt(n)*qnorm(p0))/sqrt(n)}\,,$

$\Longrightarrow$ different choice of k $\Longrightarrow$ a conflict. This is illustrated on the next slide.

The Choice of k: Controlling Producer Risk α



Conflict Resolution

This conflict can be resolved by leaving n flexible.

We then have two variables k and n to satisfy two inequalities

$$\gamma(p_1) \leq \beta \qquad \text{and} \qquad \gamma(p_0) \geq 1 - \alpha\,.$$

Find the smallest n. One slight problem: n is an integer.

Thus it may not be possible to satisfy both equations (in $\leq$ and $\geq$) exactly.

For a given value n find $k = k(n)$ to solve $\gamma(p_1) = \beta$.

If that $k(n)$ also yields $\alpha \geq G_{n-1,\,\delta(p_0)}(k(n)\sqrt{n})$,

then n was possibly chosen too high and a lower value of n should be tried.

If we have $\alpha < G_{n-1,\,\delta(p_0)}(k(n)\sqrt{n})$,

then n was definitely chosen too small and a larger value of n should be tried next.

Iteration

Through iteration find the smallest sample size n such that $k(n)$ and n satisfy both

$$\gamma(p_1) \leq \beta \qquad \text{and} \qquad \gamma(p_0) \geq 1 - \alpha\,.$$

This iteration process will lead to a solution provided $p_0 < p_1$.

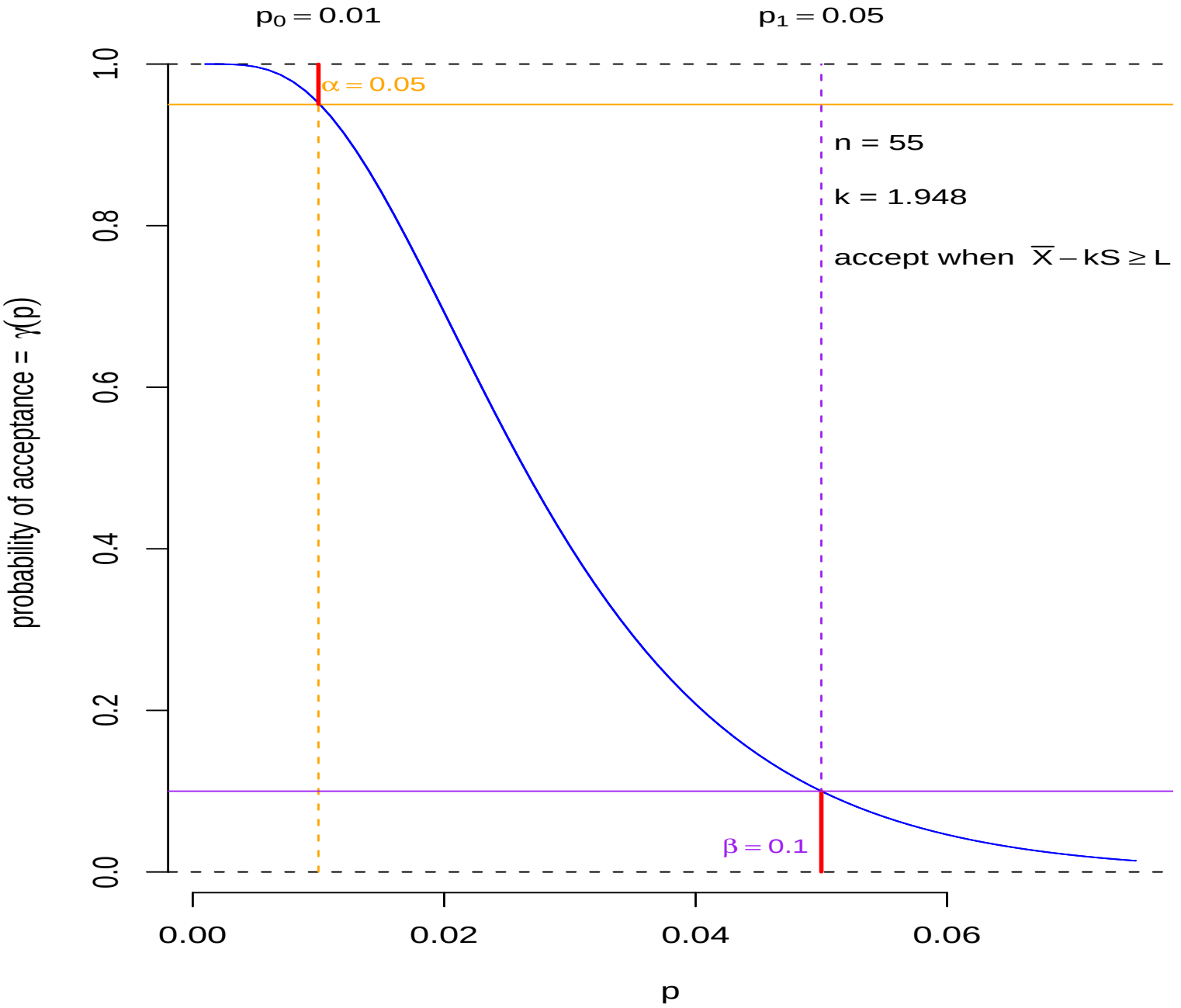
If p_0 and p_1 are too close to each other, very large sample sizes will be required.

Note that the search for the minimal sample size n does not involve L, μ and σ.

Only p_0, p_1, α and β are required.

Such a process is carried out by the R function `OC.curve` which also produces the next plot, indicating the appropriate choice for n and k.

Optimal Choice of n



Some Comments

The VASP does not say how the producer accomplishes the value $p \leq p_0$.

This is usually based on extensive testing or the producer's broad experience.

$\Longrightarrow$ upper confidence bounds for $P(X < L)$ based on sufficient data.

This is addressed in a later section.

Also, the consumer cannot set p_1 arbitrarily low since there may not be a producer that will deliver that quality or will deliver it only at exorbitant costs.

Comparison with Attributes Acceptance Sampling Plan

We compare the VASP with the Attributes Acceptance Sampling Plan (AASP).

To understand the effect on the needed sample size n when all requirements are kept at the same levels.

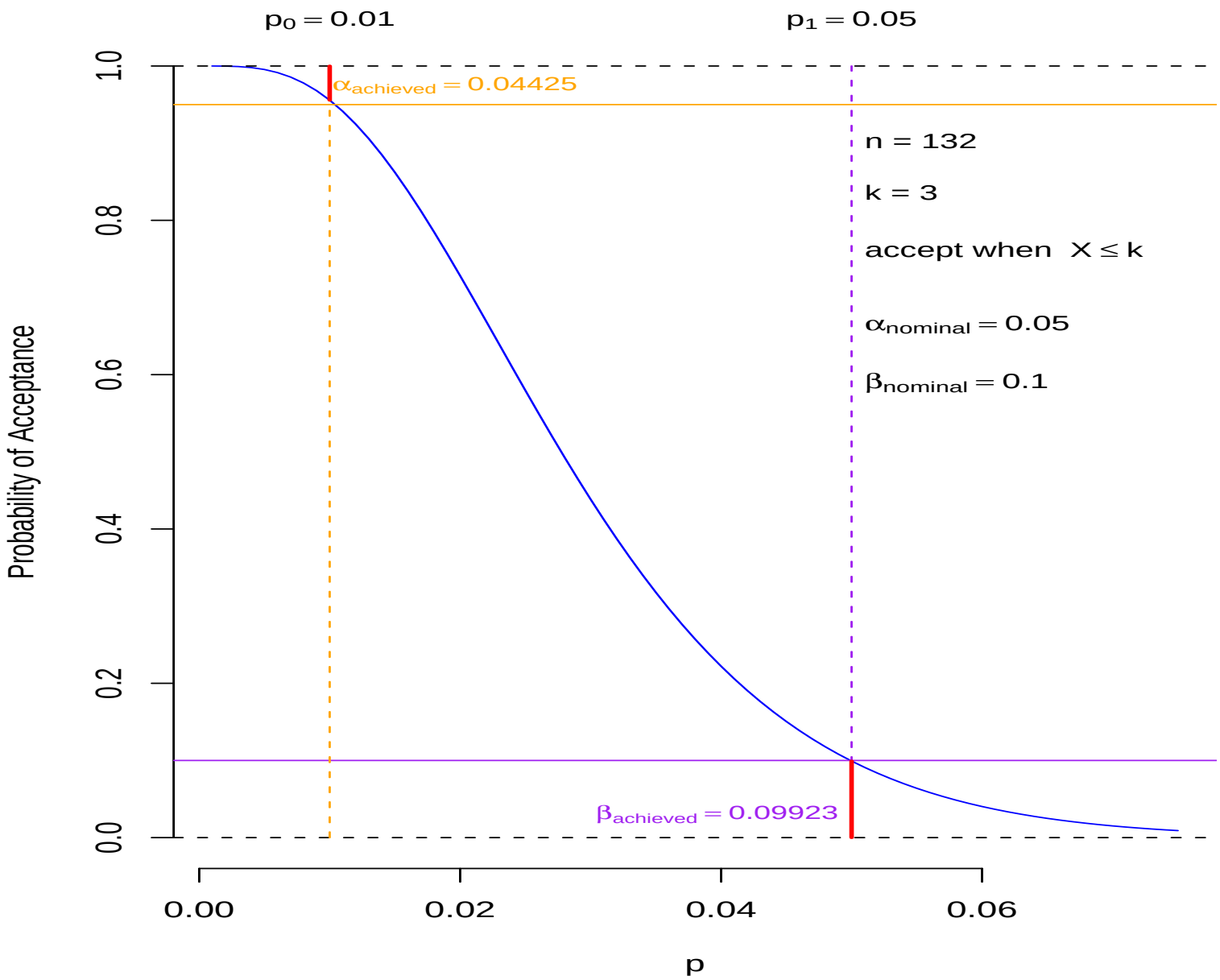
In an AASP the number X of defective items is counted and the population is accepted when $X \leq k$.

Here k and the smallest sample size n are determined such that for given $p_0 < p_1$ and $\alpha > 0$, $\beta > 0$ with $\alpha + \beta < 1$ we have

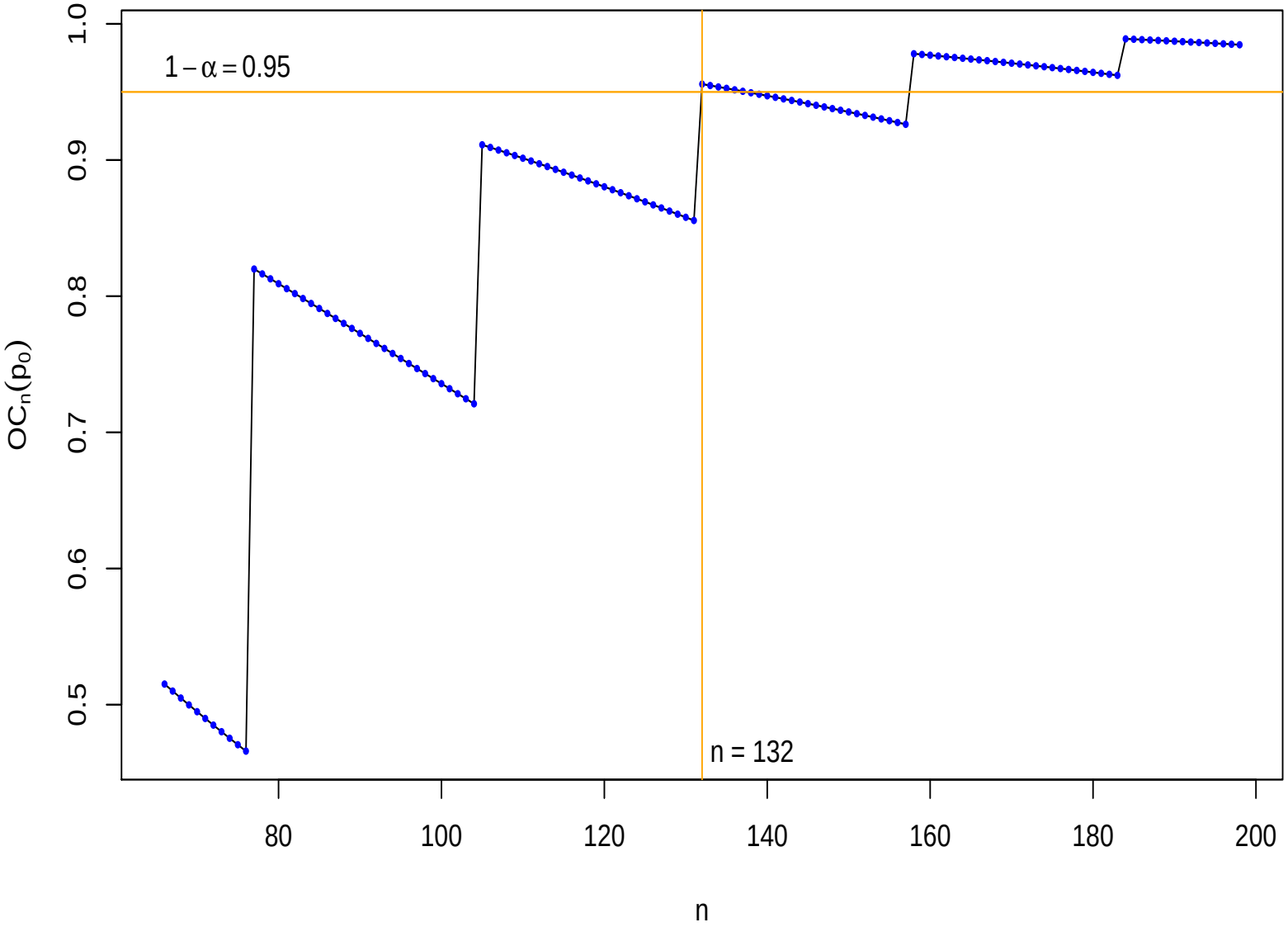
$$P_{p_1}(X \leq k) \leq \beta \qquad \text{and} \qquad P_{p_0}(X \leq k) \geq 1 - \alpha\,.$$

The AASP OC-curve on the next slide was produced by the R function `OC.binom`.

AASP OC-Curve



Zigzag Behavior of $\mathbf{OC}_n(p_0)$



Tolerance Bounds or Tolerance Limits

Tolerance bounds or tolerance limits are lower or upper confidence bounds on population percentiles or quantiles.

We assume a normal population.

The p-percentile or p-quantile x_p of $\mathcal{N}(\mu, \sigma^2)$ can be expressed as

$$x_p = \mu + z_p \, \sigma \, ,$$

where $z_p = \Phi^{-1}(p)$ is the p-quantile of the standard normal distribution.

The discussion will mainly focus on lower confidence bounds.

A $100\gamma\%$ lower bound $\hat{x}_{p,L}(\gamma)$ for x_p is also a $100(1-\gamma)\%$ upper bound for x_p.

The lower confidence bound for x_p is then computed as $\hat{x}_{p,L}(\gamma) = \bar{X} - kS$

where k satisfies $P_{\mu,\,\sigma}(\bar{X} - kS \leq x_p) = \gamma$ for all $(\mu,\, \sigma)$.

Finding the k-Factor

We have

$$\begin{aligned} P_{\mu,\,\sigma}(\bar{X} - kS \le x_p) &= P_{\mu,\,\sigma}\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} + \frac{\sqrt{n}(\mu - x_p)}{\sigma} \le k\sqrt{n}\,\frac{S}{\sigma}\right) \\ &= P_{\mu,\,\sigma}\left(\frac{Z+\delta}{\sqrt{V/(n-1)}} \le k\sqrt{n}\right) \\ &= P(T_{n-1,\,\delta} \le k\sqrt{n}) = G_{n-1,\delta}(k\sqrt{n}) \end{aligned}$$

where $\delta = -\sqrt{n}(x_p - \mu)/\sigma = -\sqrt{n}z_p$.

Thus solve $\quad G_{n-1,\,\delta}(\sqrt{n}k) = \gamma \quad$ for $k \quad \Longrightarrow \quad k = G^{-1}_{n-1,\,\delta}(\gamma)/\sqrt{n}$.

In R this is done by invoking the command

$$\texttt{k} = \texttt{qnct}(\texttt{gam}, \texttt{n}-1, -\texttt{sqrt}(\texttt{n}) * \texttt{qnorm}(\texttt{p}))/\texttt{sqrt}(\texttt{n})\ ,$$

where `gam`=γ. Avoid the variable name `gamma` in R since it is the intrinsic Γ-function.

Usage Background

In structural engineering the 95% lower bounds for $x_{.01}$ and $x_{.10}$ are called A- and B-Allowables, respectively, and are mainly used to limit material strength properties from below.

In the lumber industry the interest is in 75% lower bounds for $x_{.05}$, see page 4 of

`https://www.aitc-glulam.org/shopcart/Pdf/aitc_402-2005.pdf`

`402.4.8. Beam Performance. The beam strength 5% tolerance limit with 75% confidence determined in accordance with ASTM D2915 shall be a minimum of 2.1 times the design value for the beam. ....`

Data Illustration

As an illustration we will use some data from MIL-HDBK-5J*, see

`http://www.weibull.com/mil_std/mil_hdbk_5j.pdf`.

In particular, we will use the TUS (tensile ultimate strength) data set, designated as Group 5 on page 9-165. It consists of $n = 100$ values, measured in KSI (1000 pounds per square inch).

See `m5dat5` in the referenced R work space.

*Note that this file is about 68.5MB and consists of 1733 pages

Test for Normality

The normal QQ-plot of this data set is shown on the next slide.

Produced by `m5dat5.qqnorm`, it shows no significant deviation from normality.

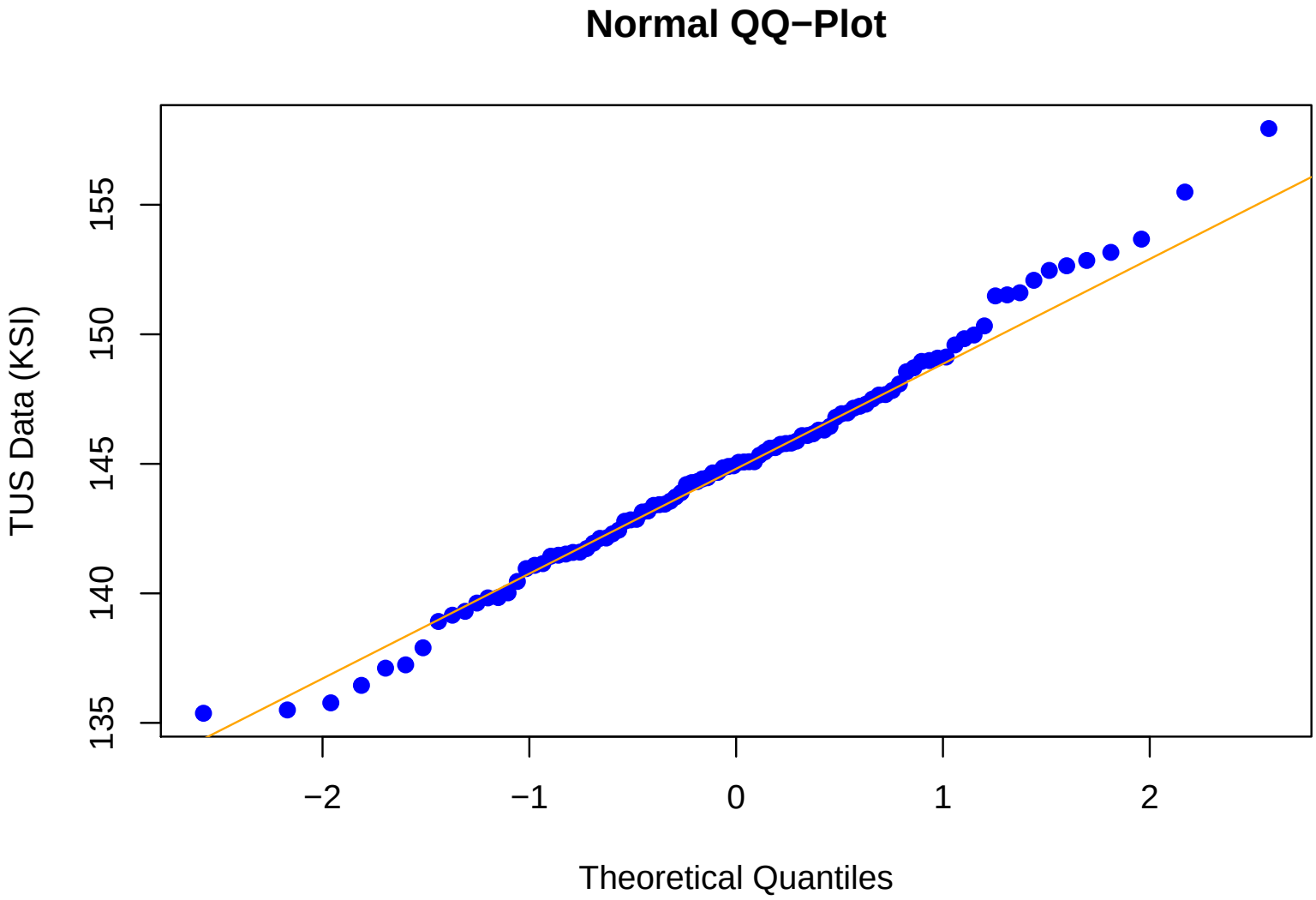
Formal tests for normality,

Lilliefors (Kolmogorov-Smirnov),

Cramér-von Mises, and

Anderson-Darling,

confirm this with p-values above $.63$ for all three discrepancy metrics.


Normal QQ−Plot
TUS Data (KSI)
135
140
145
150
155
−2
−1
0
1
2
Theoretical Quantiles

The Calculation of Allowables

The sample mean and standard deviation are $\bar{X} = 145$ and $S = 4.469965$, respectively. The k-factors for A- and B-allowables are respectively

$$k_A = \texttt{qnct}(.95, 99, -\texttt{sqrt}(100) * \texttt{qnorm}(.01))/\texttt{sqrt}(100) = 2.683957$$

and

$$k_B = \texttt{qnct}(.95, 99, -\texttt{sqrt}(100) * \texttt{qnorm}(.1))/\texttt{sqrt}(100) = 1.526749$$

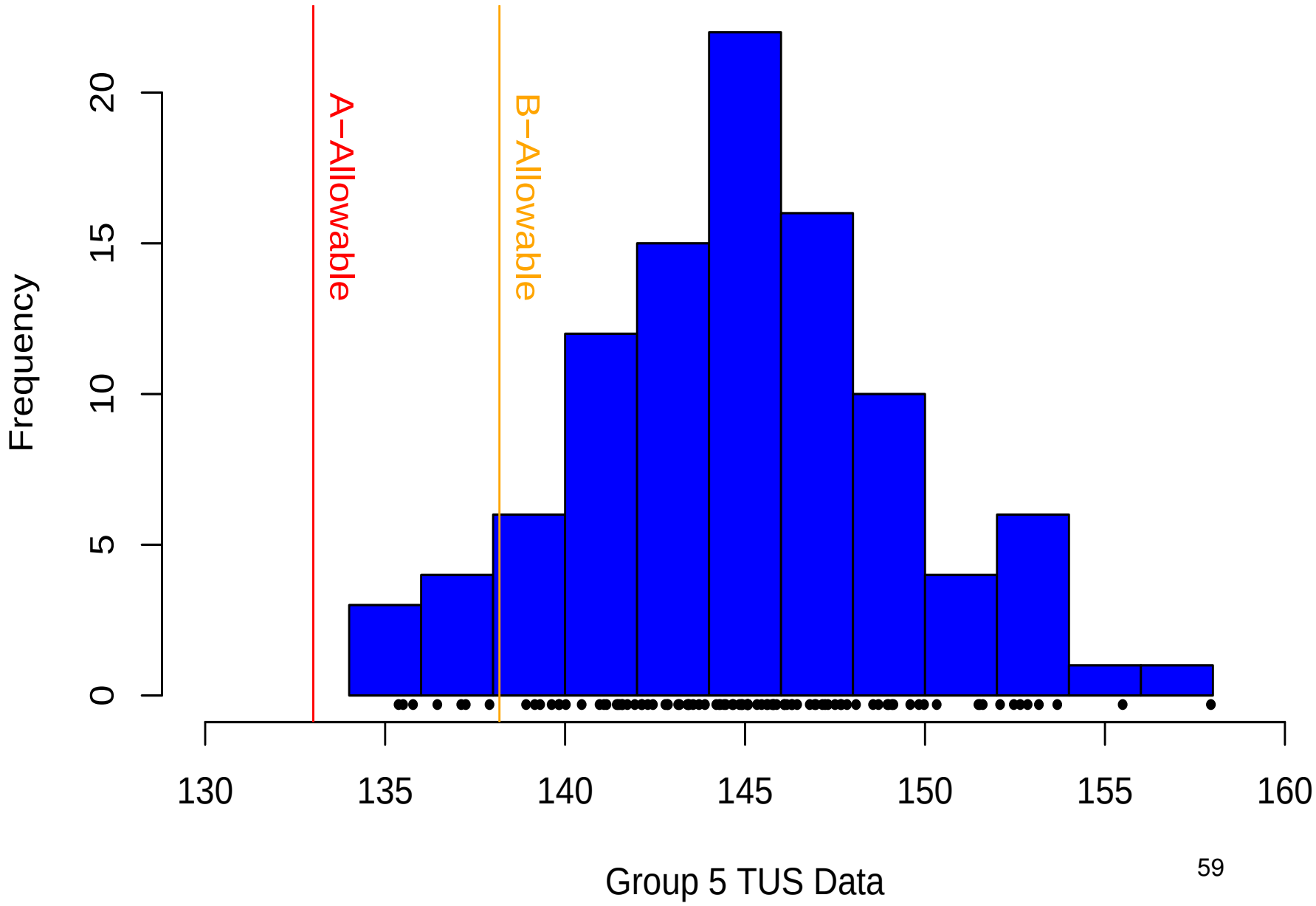
so that the A-and B-allowables are

$$A = \hat{x}_{.01,L}(.95) = \bar{X} - k_A \times S = 145 - 2.683957 \times 4.469965 = 133.0028$$

and

$$B = \hat{x}_{.10,L}(.95) = \bar{X} - k_B \times S = 145 - 1.526749 \times 4.469965 = 138.1755\ .$$

The next slide shows these allowables in relation to the data and their histogram.


A−Allowable
B−Allowable
Frequency
0
5
10
15
20
130
135
140
145
150
155
160
Group 5 TUS Data

Tail Probabilities

For a given threshold value x_0 we are interested in the normal tail probability

$$p = p(x_0) = p(x_0, \mu, \sigma) = P_{\mu,\,\sigma}(X \leq x_0) = \Phi\left(\frac{x_0 - \mu}{\sigma}\right) .$$

For VASPs this came up as the probability $p = P(X < L)$ of an item being defective.

Upper bounds for such probabilities p could give a producer the needed assurance of having a proportion of defectives $\leq p_0$, the value used in setting up the VASP.

$\hat{p} = \Phi\left((x_0 - \bar{X})/S\right)$ is a natural estimate of p but it is not unbiased.

Constructing confidence bounds seems not so obvious.

Somehow one feels/suspects a connection with bounds on the p-quantile x_p

Left/Right Tail and Upper/Lower Confidence Bounds

If $\hat{p}_U(\gamma)$ denotes an upper bound for p with confidence level γ, i.e., for all $(\mu,\ \sigma)$

$$P_{\mu,\,\sigma}(\hat{p}_U(\gamma) \geq p) = P_{\mu,\,\sigma}(\hat{p}_U(x_0,\gamma) \geq p(x_0,\mu,\sigma)) = \gamma\,,$$

then we also have for all $(\mu,\ \sigma)$ $\qquad P_{\mu,\,\sigma}(\hat{p}_U(\gamma) \leq p) = 1-\gamma\,.$

$\Longrightarrow$ $\hat{p}_U(\gamma) = \hat{p}_L(1-\gamma) = \hat{p}_L(\gamma')$ also is a $\gamma' = (1-\gamma)$-level lower bound for p.

If the upper tail probability $q = 1-p$ of the normal distribution is of interest,

then $\quad \hat{q}_U(1-\gamma) = 1-\hat{p}_U(\gamma)$ $\quad$ is a γ-level lower bound for q

and thus a $(1-\gamma)$-level upper bound for q.

Thus it suffices to limit the discussion to upper confidence bounds for p.

Monotone Quantile Bounds

Recall: $\bar{X} - k_p(\gamma)S = \bar{X} + h_\gamma(p)S$ is a $100\gamma\%$ lower bounds for x_p.

$h(p) = h_\gamma(p) = -k_p(\gamma)$ is strictly increasing in p

thus has a well defined strictly increasing inverse $h^{-1}(\cdot)$.

Proof: For $p_1 < p_2$ we have $x_{p_1} < x_{p_2}$. Suppose that $h(p_1) \geq h(p_2)$, then

$$\begin{aligned}
\gamma &= P(\bar{X} + h(p_2)S \leq x_{p_2}) && \text{by definition of } h(p) = h_\gamma(p) = -k_p(\gamma) \\
&= P(\bar{X} + h(p_1)S \leq x_{p_2} + (h(p_1) - h(p_2))S) \\
&\geq P(\bar{X} + h(p_1)S \leq x_{p_2}) && \text{since } (h(p_1) - h(p_2))S \geq 0 \\
&= P(\bar{X} + h(p_1)S \leq x_{p_1}) + P(x_{p_1} < \bar{X} + h(p_1)S \leq x_{p_2}) = \gamma + \delta > \gamma
\end{aligned}$$

since $(\bar{X}, S)$ has positive density over the half-plane $R \times (0, \infty)$ and thus $\delta > 0$.

$\Longrightarrow$ contradiction $\Longrightarrow$ our supposition must be wrong, i.e., $\Longrightarrow h(p_1) < h(p_2)$.

Confidence Bounds by Inversion

Conceptually simple step $\Longrightarrow$ $100\gamma\%$ upper confidence bounds for $p(x)$, i.e.,

$$\gamma = P(\bar{X} + h(p)S \leq x_p) = P(h(p) \leq (x_p - \bar{X})/S) = P\left(p \leq h^{-1}\left((x_p - \bar{X})/S\right)\right)$$

for all $p \in (0,1)$ and thus also for $p = p(x)$ for all $x \in R$, i.e.,

$$\gamma = P\left(p(x) \leq h^{-1}\left((x - \bar{X})/S\right)\right) \qquad \text{since } x_{p(x)} = x.$$

Thus $\hat{p}_{U\gamma}(x) = h^{-1}\left((x - \bar{X})/S\right)$ is a $100\gamma\%$ upper confidence bound for $p(x)$.

The only remaining practical problem is the calculation of $h^{-1}(y)$ for any y,

in particular for $y = (x - \bar{x})/s$, where $(\bar{x}, s)$ is the observed value of $(\bar{X}, S)$.

A Characterization of $h^{-1}(y)$

$$\begin{aligned}
\gamma &= P(\bar{X} + h(p)S \leq x_p) \\
&= P(\bar{X} + h(p)S \leq \mu + z_p\sigma) \\
&= P((\bar{X} - \mu)/\sigma + h(p)S/\sigma \leq z_p) \\
&= P(\bar{Z} + h(p)\sqrt{V/(n-1)} \leq \Phi^{-1}(p)) \\
&= P\left(\Phi\left(\bar{Z} + h(p)\sqrt{V/(n-1)}\right) \leq p\right)
\end{aligned}$$

for all p, hence also for $p = h^{-1}(y)$

$$= P\left(\Phi\left(\bar{Z} + y\sqrt{V/(n-1)}\right) \leq h^{-1}(y)\right)$$

Hence $a = h^{-1}(y)$ is the γ-quantile of the $\Phi\left(\bar{Z} + y\sqrt{V/(n-1)}\right)$ random variable.

How to Find $h^{-1}(y)$

To find $a = h^{-1}(y)$ we note that

$$\begin{aligned}
\gamma &= P\left(\Phi\left(\bar{Z} + y\sqrt{V/(n-1)}\right) \le a\right) \\
&= P\left(\bar{Z} + y\sqrt{V/(n-1)} \le \Phi^{-1}(a)\right) \\
&= P\left(\left(\sqrt{n}\bar{Z} - \sqrt{n}\Phi^{-1}(a)\right) \Big/ \sqrt{V/(n-1)} \le -\sqrt{n}\, y\right) \\
&= G_{n-1,\delta}\left(-\sqrt{n}\, y\right)
\end{aligned}$$

This equation needs to be solved for $\delta = -\sqrt{n}\Phi^{-1}(a)$ using `del.nct`.

Denote that solution by $\hat{\delta}$ then $a = \Phi(-\hat{\delta}/\sqrt{n}) = h^{-1}(y)$ is our desired value.

Using `del.nct`

This upper confidence bound is found by invoking the following R command

$$\hat{p}_{U\boldsymbol{\gamma}}(x) = \texttt{pnorm}(-\texttt{del.nct}(-\texttt{sqrt(n)} * (\texttt{x} - \texttt{Xbar})/\texttt{S}, \texttt{gam}, \texttt{n} - 1)/\texttt{sqrt(n)})$$

where `gam`$=\boldsymbol{\gamma}$, `Xbar`$=\bar{X}$, `S`$=S$.

Again avoid `gamma` as a variable name.

Relating $\hat{p}_U(\gamma, x_0)$ to $\hat{x}_L(\gamma, p)$

The upper bounds for left tail probabilities $p(x) = P(X \leq x)$ are just the inverse to the lower bounds for the $x_{p(x)}$-quantile and vice versa, see next slide.

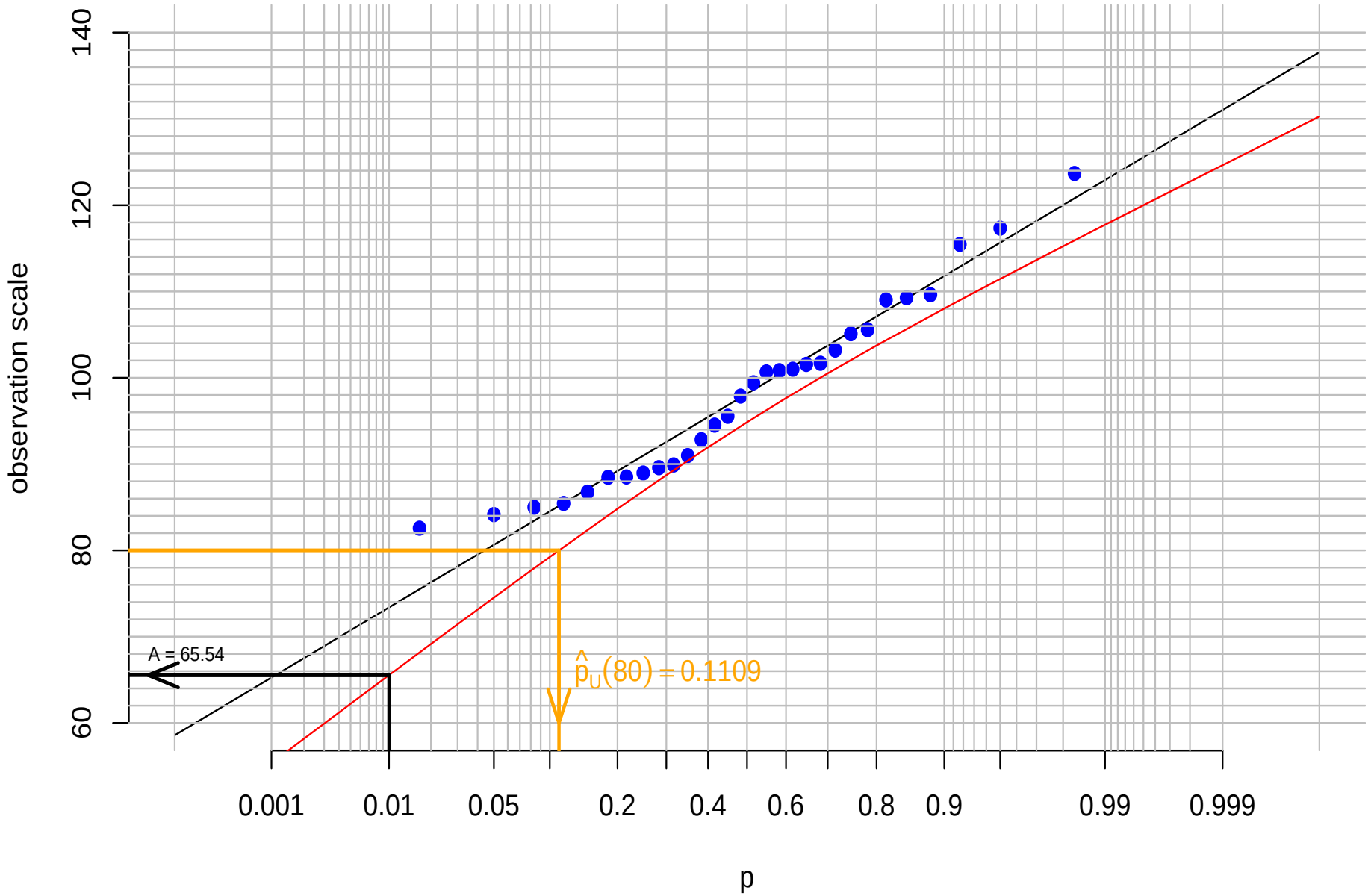
Using a random sample of size $n = 30$ from $\mathcal{N}(\mu, \sigma^2)$ with $\mu = 100$ and $\sigma = 10$, it shows a QQ-plot of the sample, i.e., the i^{th} smallest sample value $X_{(i)}$ is plotted against the standard normal p_i-quantile z_{p_i}, with $p_i = (i - .5)/n$.

However, the markings on the abscissa are given in terms of p which makes it a normal probability plot.

Expect $X_{(i)} \approx x_{p_i} = \mu + \sigma z_{p_i}$, expect an $\approx$ linear pattern when plotting $X_{(i)}$ vs z_{p_i}.

The line through the data is $\bar{X} + z_p S$. The curve below that line represents either the 95% lower bound for x_p when read sideways from the curve at the p intercept, or it represents the 95% upper bound $\hat{p}_U(x)$ for the left tail probability $p(x)$ when read vertically down from the curve at the horizontal x intercept.

Normal Probability Plot with Confidence Curve for $\hat{x}_L(p)$ and $\hat{p}_U(x)$



Some Comments

The binomial upper bound for $P(X \leq 80) = \Phi((80-100)/2) = 0.02275$ is based on $\#\{X_i \leq 80\} = 0 \implies$ `qbeta(.95,1,30)=0.09503385`.

This is lower than $\hat{p}_U(80) = 0.1109$ as obtained from $\bar{X}$ and S,

The lowest sample value is somewhat high compared to the line $\bar{X} + z_p S$. If it had been ≤ 80 we would get an upper bound $\geq$ `qbeta`$(.95, 2, 29) =$ `0.1485961`.

1) Confidence bounds based on the same data but different methods are different.

2) Even if method A (based on $\bar{X}$ and S) is generally superior to method B (binomial method), it can happen (as in this instance) that the bound produced by B is "better" than the bound produced by A. Both upper bounds are above 0.02275.

Interpret the 95% confidence curve point-wise, i.e., the probability for several such upper bounds simultaneously covering their respective targets is $< .95$.

Process Control Capability Indices

The process control capability indices C_L, C_U and C_{pk} are relatively new in quality control applications. They are defined as

$$C_L = \frac{\mu - x_L}{3\sigma}\,, \quad C_U = \frac{x_U - \mu}{3\sigma} \quad \text{and} \quad C_{pk} = \min(C_L, C_U)\,,$$

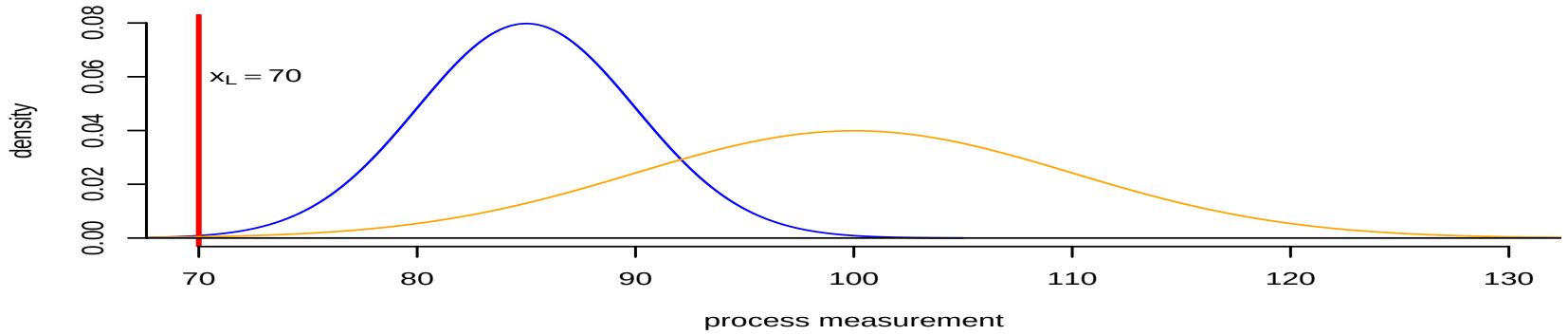
where x_L and x_U are given lower and upper specification limits.

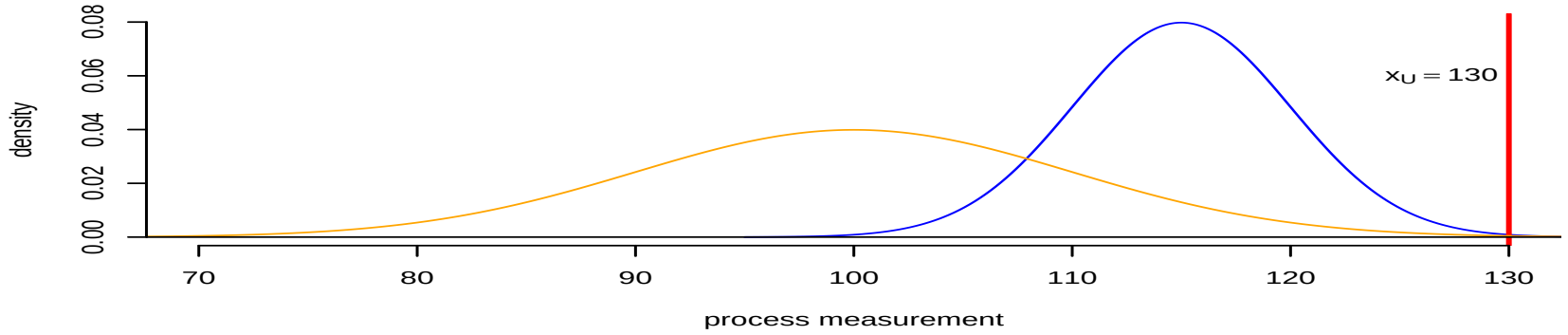
Assume that process output $X \sim \mathcal{N}(\mu, \sigma^2)$.

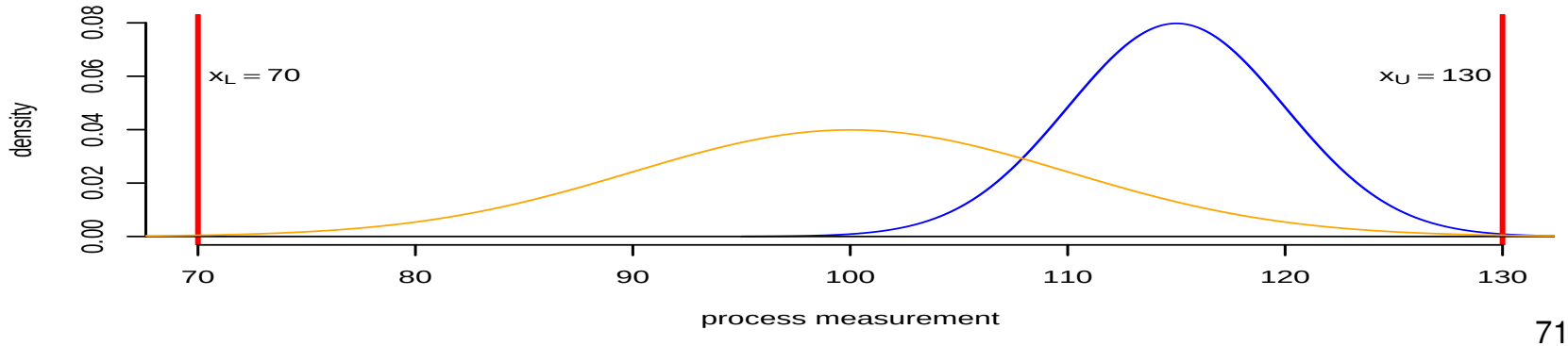
Values $C_L \geq 1$, $C_U \geq 1$ and $C_{pk} \geq 1$ indicate that the process output is at least 3σ units on the safe side from any specification limit, since

$$\begin{aligned} C_L \geq 1 &\iff \mu - 3\sigma \geq x_L \\ C_U \geq 1 &\iff \mu + 3\sigma \leq x_U \\ C_{pk} \geq 1 &\iff x_L \leq \mu - 3\sigma \quad \& \quad \mu + 3\sigma \leq x_U\,. \end{aligned}$$

$C_L = 1, \quad C_U = 1, \quad \text{and} \quad C_{pk} = 1$







Some Comments

Previous slide shows: there are many (μ, σ) for which these indices are are 1.

As $\sigma \nearrow$ we need $\mu - L \nearrow$.

This does not work when we have a specification interval.

In order to have $C_{pk} \geq 1$ we must have $6\sigma \leq x_U - x_L$.

Typically the parameters μ and σ are unknown and only limited sample data, say $X_1, \ldots, X_n$, are available from this population (assumed normal).

We now address how to obtain lower confidence bounds for these indices.

Lower bounds are of primary interest here since it is typically desired to show that the process capability index meets at least a certain threshold, say 1 or $4/3$.

Lower Confidence Bounds for C_L

A natural estimate for C_L is $\widehat{C}_L = (\bar{X} - x_L)/3S$

It will be the basis for constructing $100\gamma\%$ lower confidence limits for C_L. We have

$$\begin{aligned} P\left(\widehat{C}_L \le k\right) &= P\left(\frac{\bar{X} - x_L}{3S} \le k\right) \\ &= P\left(\frac{\sqrt{n}(\bar{X} - \mu)/\sigma + \sqrt{n}(\mu - x_L)/\sigma}{S/\sigma} \le 3\sqrt{n}k\right) \\ &= P\left(T_{n-1,3\sqrt{n}C_L} \le 3\sqrt{n}k\right). \end{aligned}$$

We define $k = k(C_L)$ as that unique number which for given C_L solves

$$P\left(T_{n-1,3\sqrt{n}C_L} \le 3\sqrt{n}k(C_L)\right) = \gamma\,.$$

From the previously cited monotonicity properties of the noncentral t-distribution we know that $k(C_L)$ is a strictly increasing function of C_L. Thus we have

$$\gamma = P\left(\widehat{C}_L \le k(C_L)\right) = P\left(k^{-1}(\widehat{C}_L) \le C_L\right) \qquad \text{for all } \mu \text{ and } \sigma.$$

Lower Confidence Bounds for C_L (cont.)

We can treat $\widehat{B}_L = k^{-1}\left(\widehat{C}_L\right)$ as a $100\gamma\%$ lower confidence bound for C_L.

How is $\widehat{B}_L$ actually computed for each such observed value $\widehat{c}_L$ of $\widehat{C}_L$?

Rewrite the defining equation for $k(C_L)$ by taking $C_L = k^{-1}(\widehat{c}_L)$:

$$\gamma = P\left(T_{n-1,3\sqrt{n}k^{-1}(\widehat{c}_L)} \leq 3\sqrt{n}k\left(k^{-1}(\widehat{c}_L)\right)\right) = P\left(T_{n-1,3\sqrt{n}k^{-1}(\widehat{c}_L)} \leq 3\sqrt{n}\widehat{c}_L\right) = \gamma.$$

If, for fixed $\widehat{c}_L$, we solve the equation:

$$P\left(T_{n-1,\widehat{\delta}} \leq 3\sqrt{n}\widehat{c}_L\right) = \gamma$$

for $\widehat{\delta}$, then we get the following expression for the observed value $\hat{b}_L$ of $\widehat{B}_L$:

$$\widehat{b}_L = k^{-1}\left(\widehat{c}_L\right) = \frac{\widehat{\delta}}{3\sqrt{n}} = \texttt{del.nct}(3*\texttt{sqrt}(\texttt{n})*\texttt{cL.hat},\texttt{gam},\texttt{n}-1)/(3*\texttt{sqrt}(\texttt{n}))\,,$$

where `gam`$=\gamma$ and `cL.hat`$=\hat{c}_L$.

Lower Confidence Bounds for C_U

In a similar fashion we develop lower confidence bounds for

$$C_U = \frac{x_U - \mu}{3\sigma}, \qquad \text{using its natural estimate} \qquad \widehat{C}_U = \frac{x_U - \bar{X}}{3S}.$$

$$P\left(\widehat{C}_U \le k\right) = P\left(\frac{x_U - \bar{X}}{3S} \le k\right) = P\left(T_{n-1,3\sqrt{n}C_U} \le 3\sqrt{n}k\right).$$

We define $k = k(C_U)$ as that unique number which for given C_U solves

$$P\left(\widehat{C}_U \le k(C_U)\right) = P\left(T_{n-1,3\sqrt{n}C_U} \le 3\sqrt{n}k(C_U)\right) = \gamma \qquad k(C_U) \nearrow \text{ as } C_U \nearrow$$

$\Longrightarrow \widehat{B}_U = k^{-1}(\widehat{C}_U) = 100\gamma\%$ lower confidence bound for C_U.

For an observed value $\widehat{c}_U$ of $\widehat{C}_U$ get the observed value $\widehat{b}_U$ of $\widehat{B}_U$ as $\widehat{\delta}/(3\sqrt{n})$,

where $\widehat{\delta}$ solves $$P\left(T_{n-1,\widehat{\delta}} \le 3\sqrt{n}\,\widehat{c}_U\right) = \gamma.$$

or $$\widehat{b}_U = k^{-1}(\widehat{c}_U) = \frac{\widehat{\delta}}{3\sqrt{n}} = \texttt{del.nct}(3*\texttt{sqrt}(\texttt{n})*\texttt{cU.hat}, \texttt{gam}, \texttt{n}-1)/(3*\texttt{sqrt}(\texttt{n})),$$

where `gam`$= \gamma$ and `cU.hat`$= \hat{c}_U$.

Lower Confidence Bounds for C_{pk}

Putting the bounds on C_U and C_L together, we can obtain (slightly conservative) confidence bounds for the two-sided statistical process control parameter

$$C_{pk} = \min(C_L, C_U) \qquad \text{simply by taking} \qquad \widehat{B} = \min\left(\widehat{B}_L, \widehat{B}_U\right) .$$

If $C_L \leq C_U$, i.e., $C_{pk} = C_L$, then

$$\begin{aligned} P\left(\min\left(\widehat{B}_L, \widehat{B}_U\right) \leq \min(C_L, C_U)\right) &= P\left(\min\left(\widehat{B}_L, \widehat{B}_U\right) \leq C_L\right) \\ &\geq P\left(\widehat{B}_L \leq C_L\right) = \gamma \end{aligned}$$

and if $C_U \leq C_L$, i.e., $C_{pk} = C_U$, then

$$\begin{aligned} P\left(\min\left(\widehat{B}_L, \widehat{B}_U\right) \leq \min(C_L, C_U)\right) &= P\left(\min\left(\widehat{B}_L, \widehat{B}_U\right) \leq C_U\right) \\ &\geq P\left(\widehat{B}_U \leq C_U\right) = \gamma . \end{aligned}$$

Comments on Coverage Probability for C_{pk} Bounds

$\widehat{B}$ can be taken as lower bound for C_{pk} with confidence level at least γ.

The exact confidence level of $\widehat{B}$ is somewhat higher than γ for $C_L = C_U$

$C_L = C_U$ when μ is the midpoint of the specification interval: $\mu = (C_L + C_U)/2$.

As $|\mu - (C_L + C_U)/2| \nearrow$ and as $\sigma \searrow$ in order to maintain a constant C_{pk} then the actual confidence level of $\widehat{B}$ gets arbitrarily close to γ

Hence the confidence coefficient of $\hat{B}$ is indeed γ.

Getting the Message to Suppliers

The supplier may understand the meaning of C_{pk}
but not the impact of sampling uncertainty in $\hat{C}_{pk}$, double whammy!

The following tables show the $\hat{C}_{pk}$ required to get a C_{pk} lower bound $\hat{B}$
to come out at the desired value, given in the top row of that table.

For example, when $n = 20$ we need $\hat{C}_{pk} \geq 1.298$ in order to get $\hat{B} \geq 1$,
i.e., to be 90% confident that the actual $C_{pk} \geq 1$.

For $n = 60$ this margin can be pushed down to $.150$, about half of $.298$.

This should easily bring home the message that it pays to have a larger sample.

Of course, larger sample sizes do not guarantee better quality.

If the quality is poor we are likely to see small values of $\hat{B}$ or even $\hat{C}_{pk}$, i.e., < 1.

This becomes clearer as $n \nearrow$. For small n it may hide that.

Required $\hat{C}_{pk}$: 90% Confidence

n	desired C_{pk}										
	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
10	1.499	1.644	1.789	1.934	2.079	2.225	2.370	2.516	2.662	2.808	2.954
12	1.432	1.570	1.708	1.847	1.986	2.125	2.265	2.404	2.544	2.683	2.823
14	1.384	1.518	1.652	1.786	1.921	2.056	2.190	2.325	2.460	2.596	2.731
16	1.349	1.479	1.610	1.741	1.872	2.004	2.135	2.267	2.398	2.530	2.662
18	1.321	1.449	1.577	1.706	1.834	1.963	2.092	2.221	2.350	2.479	2.608
20	1.298	1.424	1.551	1.677	1.804	1.930	2.057	2.184	2.311	2.438	2.565
40	1.191	1.308	1.424	1.541	1.658	1.775	1.892	2.009	2.126	2.243	2.365
60	1.150	1.263	1.376	1.489	1.602	1.715	1.828	1.944	2.058	2.171	2.285
80	1.127	1.238	1.349	1.460	1.571	1.683	1.795	1.906	2.018	2.129	2.240
100	1.112	1.222	1.331	1.442	1.552	1.661	1.771	1.881	1.991	2.101	2.211
120	1.101	1.210	1.319	1.428	1.537	1.646	1.755	1.864	1.973	2.082	2.191
140	1.093	1.201	1.309	1.417	1.525	1.634	1.742	1.850	1.958	2.067	2.175
160	1.087	1.194	1.301	1.409	1.516	1.624	1.732	1.839	1.947	2.055	2.162
180	1.081	1.188	1.295	1.402	1.509	1.616	1.723	1.831	1.938	2.045	2.152
200	1.077	1.183	1.290	1.396	1.503	1.610	1.716	1.823	1.930	2.037	2.144
400	1.053	1.157	1.262	1.366	1.471	1.576	1.680	1.785	1.890	1.994	2.099

See `http://www.boeing.com/companyoffices/doingbiz/supplier/d1-9000-1.pdf` on page 196 in Boeing's AQS D1-9000-1 Advanced Quality Systems Tools document for suppliers.

Required $\hat{C}_{pk}$: 95% Confidence

n	desired C_{pk}										
	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
10	1.686	1.847	2.009	2.171	2.333	2.496	2.659	2.822	2.985	3.148	3.312
12	1.588	1.740	1.892	2.045	2.198	2.351	2.505	2.659	2.812	2.966	3.121
14	1.520	1.665	1.811	1.958	2.105	2.252	2.399	2.546	2.693	2.841	2.989
16	1.469	1.610	1.752	1.894	2.036	2.178	2.320	2.463	2.606	2.748	2.891
18	1.430	1.568	1.706	1.844	1.982	2.121	2.260	2.399	2.538	2.677	2.816
20	1.399	1.534	1.669	1.804	1.939	2.075	2.211	2.347	2.483	2.619	2.756
40	1.252	1.373	1.495	1.617	1.739	1.862	1.984	2.107	2.229	2.352	2.487
60	1.197	1.313	1.430	1.547	1.665	1.782	1.899	2.023	2.141	2.259	2.377
80	1.166	1.280	1.394	1.509	1.623	1.741	1.856	1.971	2.086	2.201	2.317
100	1.146	1.258	1.371	1.486	1.599	1.712	1.825	1.938	2.051	2.164	2.278
120	1.132	1.243	1.356	1.467	1.579	1.691	1.802	1.914	2.026	2.138	2.250
140	1.121	1.232	1.343	1.453	1.564	1.675	1.785	1.896	2.007	2.118	2.229
160	1.113	1.223	1.332	1.442	1.552	1.662	1.772	1.882	1.992	2.102	2.212
180	1.106	1.215	1.324	1.433	1.542	1.652	1.761	1.870	1.980	2.089	2.199
200	1.100	1.208	1.317	1.426	1.534	1.643	1.752	1.861	1.969	2.078	2.187
400	1.069	1.174	1.280	1.386	1.492	1.598	1.704	1.810	1.916	2.022	2.128

Coefficient of Variation Confidence Bounds

Coefficient of variation $=$ ratio of standard deviation to mean, i.e., as $\nu = \sigma/\mu$.

It expresses measurement variability relative to what is being measured.

We will instead give confidence bounds for its reciprocal $\rho = 1/\nu = \mu/\sigma$.

Reason: $\bar{X} \approx 0$ in the natural estimate $S/\bar{X}$ for ν could cause certain problems.

If the coefficient of variation is sufficiently small, usually the desired situation, then the distinction between it and its reciprocal is somewhat immaterial since typical bounds for ν can be inverted to bounds for ρ and vice versa.

This is easily recognized by the sign of the upper or lower bound, respectively.

When Do We Have Problems?

If $\hat{\rho}_L$ as lower bound for ρ is positive, then $\hat{\nu}_U = 1/\hat{\rho}_L$ is an upper bound for $\nu > 0$.

If $\hat{\rho}_U$ as upper bound for ρ is negative, then $\hat{\nu}_L = 1/\hat{\rho}_U$ is a lower bound for $\nu < 0$.

In either case ρ is bounded away from zero which implies that the reciprocal $\nu = 1/\rho$ is bounded.

If $\hat{\rho}_L$ as lower bound for ρ is negative, then ρ is not bounded away from zero and the reciprocal values could be arbitrarily large.

In that case $\hat{\nu}_U = 1/\hat{\rho}_L$ is useless as an upper bound for ν since no finite upper bound on the values of ν can be derived from $\hat{\rho}_L$.

Lower and Upper Bounds for ρ

To construct a lower confidence bound for $\rho = \mu/\sigma$ consider

$$\sqrt{n}\,\frac{\bar{X}}{S} = \frac{\sqrt{n}(\bar{X}-\mu)/\sigma + \sqrt{n}\mu/\sigma}{S/\sigma} = T_{n-1,\,\delta}$$

with $\delta = \sqrt{n}\mu/\sigma$. $\quad\Longrightarrow\quad U = G_{n-1,\,\delta}(\sqrt{n}\,\bar{X}/S) = G_{n-1,\,\delta}(T_{n-1,\,\delta}) \sim U(0,1)$

$$\gamma = P(U \le \gamma) = P(G_{n-1,\,\delta}(\sqrt{n}\,\bar{X}/S) \le \gamma) = P(\hat{\delta}_L \le \delta)$$

where $\hat{\delta}_L$ is the solution of $\quad G_{n-1,\,\hat{\delta}_L}(\sqrt{n}\,\bar{X}/S) = \gamma$

$\hat{\rho}_L \overset{\text{def}}{=} \hat{\delta}_L/\sqrt{n} = \texttt{del.nct(sqrt(n)*Xbar/S, gam, n-1)/sqrt(n)}$

$= 100\gamma\%$ lower confidence bound for $\rho = \mu/\sigma$. Here `Xbar`$= \bar{X}$ and `gam`$= \gamma$.

To obtain an upper bound for ρ with confidence level γ one finds $\hat{\delta}_U$ as solution of

$$G_{n-1,\,\hat{\delta}_U}(\sqrt{n}\,\bar{X}/S) = 1-\gamma$$

$\hat{\rho}_U \overset{\text{def}}{=} \hat{\delta}_U/\sqrt{n} = \texttt{del.nct(sqrt(n)*Xbar/S, 1-gam, n-1)/sqrt(n)}$

$= 100\gamma\%$ upper bound for $\rho = \mu/\sigma$. Here `Xbar`$= \bar{X}$ and `gam`$= \gamma$.

Batch Effects

Eventually these methods were also used in the context of composite materials where batch effects can be quite significant.

Chemical compositions change for each batch of material.

A good portion of the strength variation of tested specimens from that material is due to the variation from batch to batch.

It does not help to have lots of observations from few batches!

At least in the early production stage the tendency is to make do with few batches.

Make up for this deficiency by sampling each batch many times (questionable!).

Two Extreme Situations

1. The variation from batch to batch is insignificant.

   Treat all specimen strengths as one big sample of size $N = n_1 + \ldots + n_k$,

   k is the number of batches involved

   n_i is the number of strength measurements from the i^{th} batch.

2. Batch to batch variation $\gg$ within batch variation

   It is a wasted effort to have more than one observation per batch.

   It is like writing down same observation n_i times.

   Having $n_i > 1$ can only confirm variability mismatch.

   Treating all $N = n_1 + \ldots + n_k$ as one large random sample greatly inflates the "effective" sample size.

Effective Sample Size Solution

This problem was addressed by Scholz and Vangel (1998) by interpolating between these two extreme situations. See class web site for preprint.

The problem was reduced to that of a simple random sample with some "effective" sample size $N^\star$ somewhere between k and N.

$N^\star$ reflects the ratio of within to between batch variability.

This reduced a rather messy situation in a simple and intuitive fashion to the previous process for a pure random sample.

The effective sample size concept was intuitively very appealing to the customer.

Solutions were developed for tolerance bounds and capability index bounds. We deal only with tolerance bounds here (see paper for the other bounds).

Measurement Variation Model

$$X_{ij} = \mu + b_i + e_{ij}, \quad j = 1, \ldots, n_i, \quad \text{and} \quad i = 1, \ldots, k\,,$$

where $b_i \sim \mathcal{N}(0, \sigma_b^2)$ (between batch variation effect)

and $e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ (within batch variation effects) .

b_i and $\{e_{ij}\}$ are assumed to be mutually independent $\implies X_{ij} \sim \mathcal{N}(\mu, \sigma_b^2 + \sigma_e^2)$

for $j \neq j'$ $\quad \operatorname{cov}(X_{ij}, X_{ij'}) = \operatorname{cov}(b_i + e_{ij}, b_i + e_{ij'}) = \operatorname{cov}(b_i, b_i) = \sigma_b^2$.

The correlation of two different observations within the same batch is

$$\rho = \frac{\operatorname{cov}(X_{ij}, X_{ij'})}{\sqrt{\operatorname{var}(X_{ij})}\ \sqrt{\operatorname{var}(X_{ij'})}} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}$$

which can range anywhere within $[0, 1]$.

Typical Industrial Statistics Example

The individual sample sizes n_i from each batch can vary.

However, in developing the ultimate solution we were guided strongly by the special case $n_1 = \ldots = n_k$. Even in that case we invoked an interpolation approximation.

This was augmented with a further approximation (Satterthwaite) when allowing the n_i to be different.

Simulations validated these approximations as reasonable.

This problem arose when a supplier was trying to build his case based on one large sample $N = n_1 + \ldots + n_k$ without accounting for the possible batch effects.

After confirming the significance of that effect it was essential to find a middle ground, which was easily captured by the “effective sample size” concept.

It reduced the calculations in a simple manner to a previously accepted method.

Equivalent/Effective Sample Size

Conceptualize a pure random sample $X_1^\star, \ldots, X_{N^\star}^\star$ from $\mathcal{N}(\mu, \sigma_b^2 + \sigma_e^2)$ that carries the "same kind of information" as the original data.

$N^\star$ then represents the "equivalent sample size."

$$\bar{X} = \sum_{i=1}^{B} \sum_{j=1}^{n_i} X_{ij}/N \sim \mathcal{N}(\mu, \sigma_{\bar{X}}^2) \quad \text{and} \quad \bar{X}^\star = \sum_{i=1}^{N^\star} X_i^\star/N^\star \sim \mathcal{N}(\mu, \sigma_{\bar{X}^\star}^2)$$

we choose $N^\star$ to match the variances of $\bar{X}$ and $\bar{X}^\star$, i.e., find $N^\star$ such that

$$\begin{aligned} \operatorname{var}(\bar{X}) &= \operatorname{var}\left(\mu + \frac{\sum_{i=1}^{k} n_i b_i + \sum_{i=1}^{k} \sum_{j=1}^{n_i} e_{ij}}{N}\right) = \sigma_b^2 \sum_{i=1}^{k} \left(\frac{n_i}{N}\right)^2 + \sigma_e^2 \frac{1}{N} \\ &= \operatorname{var}(\bar{X}^\star) = \frac{\sigma_b^2 + \sigma_e^2}{N^\star} . \end{aligned}$$

$$\Longrightarrow N^\star = \left[\frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2} \sum_{i=1}^{k} \left(\frac{n_i}{N}\right)^2 + \frac{1}{N} \frac{\sigma_e^2}{\sigma_b^2 + \sigma_e^2}\right]^{-1} = \left[\rho \frac{1}{f+1} + (1-\rho) \frac{1}{N}\right]^{-1} ,$$

where we write $1/(f+1) = \sum_{i=1}^{k} (n_i/N)^2$ for reasons to become clear later.

Note that $N^\star$ is the weighted harmonic mean of $f+1$ and N.

Some Comments

$\rho = 0 \Rightarrow N^\star = N$ and $\rho = 1 \Rightarrow N^\star = f + 1 \quad (= k \quad$ when $n_1 = \ldots = n_k)$.

When $n_1 = \ldots = n_k$ the effective sample size formula for $N^\star$ agrees with our previous notion of what the effective sample size should be in these two extreme situations, namely N and k.

$N^\star$ may not be an integer, but an actual conceptual sample $X_1^\star, \ldots, X_{N^\star}^\star$ is never used in our procedure.

All calculations are based on the original batch data $\{X_{ij}\}$.

Estimated Equivalent/Effective Sample Size

The within batch correlation ρ is unknown. Find reasonable estimates from the data as follows. Compute the between batch and error sums of squares

$$SS_b = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\bar{X}_{i\cdot} - \bar{X})^2 = \sum_{i=1}^{k} n_i (\bar{X}_{i\cdot} - \bar{X})^2 \qquad \text{and} \qquad SS_e = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i\cdot})^2 \, .$$

Take $\hat{\sigma}_e^2 = SS_e/(N-k)$ as unbiased estimate of σ_e^2 and $\hat{\tau}^2 = SS_b/(k-1)$ as unbiased estimate of

$$\tau^2 = \sigma_e^2 + \sigma_b^2 \frac{N}{k-1} \left(1 - \sum_{i=1}^{k} \left(\frac{n_i}{N} \right)^2 \right) = \sigma_e^2 + \sigma_b^2 \frac{N}{k-1} \frac{f}{f+1} \, .$$

$$\Rightarrow \ \hat{\sigma}_b^2 = \left(\hat{\tau}^2 - \hat{\sigma}_e^2 \right) (k-1)(f+1)/(N f) \text{ as unbiased estimate for } \sigma_b^2.$$

Redefine $\hat{\sigma}_b^2 = \max(0, \hat{\sigma}_b^2)$, it will no longer be unbiased.

Estimate ρ by $\hat{\rho} = \hat{\sigma}_b^2/(\hat{\sigma}_b^2 + \hat{\sigma}_e^2)$.

This estimate is used in place of ρ in estimating $N^\star$ by $\hat{N}^\star = N^\star(\hat{\rho})$.

Tolerance Bounds with No Between Batch Variation

Here assume $\sigma_b = 0$ and $\sigma_e > 0$, i.e, $\rho = 0$, $\Rightarrow$ X_{ij} are mutually independent.

$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/N)$ and $SS_T = SS_b + SS_e = (N-1)S^2 \sim \sigma^2 \cdot \chi^2_{N-1}$

and both are independent of each other.

It was shown that $100\gamma\%$ lower tolerance bounds are of the form $\bar{X} - k\,S$, where k

$$k = k_0(N) = \frac{1}{\sqrt{N}}\, t_{N-1,-z_p\sqrt{N},\gamma} = \sqrt{\frac{N-1}{N}}\,\frac{1}{\sqrt{N-1}}\, t_{N-1,-z_p\sqrt{N},\gamma}\,, \qquad (1)$$

where $t_{N-1,-z_p\sqrt{N},\gamma}$ is the γ-quantile of $T_{N-1,-z_p\sqrt{N}}$.

No Within Batch Variation: Satterthwaite Approximation

Here assume $\sigma_b > 0$ and $\sigma_e = 0$, i.e, $\rho = 1$,

$\sigma^2 = \sigma_b^2$ and all observations within each batch are identical.

$\Rightarrow\ SS_e = 0$, and thus $S^2 = SS_b/(N-1)$.

Using Satterthwaite's method approximate the distribution of $SS_T = SS_b$ by $a \cdot \chi^2_g$, where a and g are determined to match mean and variance on either side.

$$\Rightarrow\ g = \frac{(1-\sum w_i^2)^2}{\sum w_i^2 - 2\sum w_i^3 + (\sum w_i^2)^2} \quad \text{and} \quad a = \frac{N}{g}\sigma_b^2\left(1 - \sum_{i=1}^{k} w_i^2\right) = \frac{N}{g}\sigma_b^2 \frac{f}{f+1},$$

where $w_i = n_i/N$ and summations are over $i = 1, \ldots, k$. (Notes Appendix A)

The Satterthwaite approximation is exact when the n_i are all the same.

g can be approximated very well by $f = \left(\sum w_i^2\right)^{-1} - 1$. (Notes Appendix B)

Tolerance Bounds: No Within Batch Variation

With f replacing g we have $a \approx N\sigma_b^2/(f+1)$ and we can treat

$$V^2 = \frac{SS_T}{a\,f} = \frac{SS_b}{a\,f} = S^2\,\frac{(N-1)(f+1)}{f\,N\,\sigma_b^2}$$

as an approximate χ_f^2/f random variable.

Further, $\bar{X} \sim \mathcal{N}(\mu, \tau^2)$ with $\tau^2 = \sigma_b^2 \cdot \sum_{i=1}^k w_i^2 = \sigma_b^2/(f+1)$,
$\Longrightarrow\ Z = \sqrt{f+1}\,(\bar{X}-\mu)/\sigma_b \sim \mathcal{N}(0,1)$.

When all samples sizes n_i are the same ($= n$), then $f = k-1$ and $a = n\sigma_b^2$.
In that case SS_b actually is exactly distributed like $n\sigma_b^2 \cdot \chi_{k-1}^2$ and then $SS_T = SS_b$ is independent of $\bar{X}$. When the samples sizes are not the same, then SS_T is approximately distributed like the above chi-square multiple $a\chi_f^2$ and the strict independence property no longer holds. We will ignore this latter flaw in our derivation below. The simulations show that this is of no serious consequence.

Tolerance Bounds: No Within Batch Variation

Again we have

$$\gamma = P\left(\bar{X} - k\,S \le x_p\right) = P\left(\frac{\sqrt{f+1}\,(\bar{X}-\mu)}{\sigma_b} - \frac{\sqrt{f+1}\,(x_p-\mu)}{\sigma_b} \le \frac{k\sqrt{f+1}\,S}{\sigma_b}\right)$$

$$= P\left(\frac{Z - z_p\sqrt{f+1}}{V} \le k\sqrt{\frac{f\,N}{N-1}}\right)$$

$$= P\left(T_{f,-z_p\sqrt{f+1}} \le k\sqrt{\frac{f\,N}{N-1}}\right)$$

leading to

$$k = k_1(N) = \sqrt{\frac{N-1}{N}}\,\frac{1}{\sqrt{f}}\,t_{f,-z_p\sqrt{f+1},\gamma}\,. \tag{2}$$

Tolerance Bounds: The Interpolation Step

We note the strong parallelism between equations (1) and (2) for the k-factor.

Common factor $\sqrt{(N-1)/N}$ in both.

The rest of the expressions match via $N \leftrightarrow f+1$.

Note that the actual tolerance bound $= \bar{X} - kS$ in both these extreme cases.

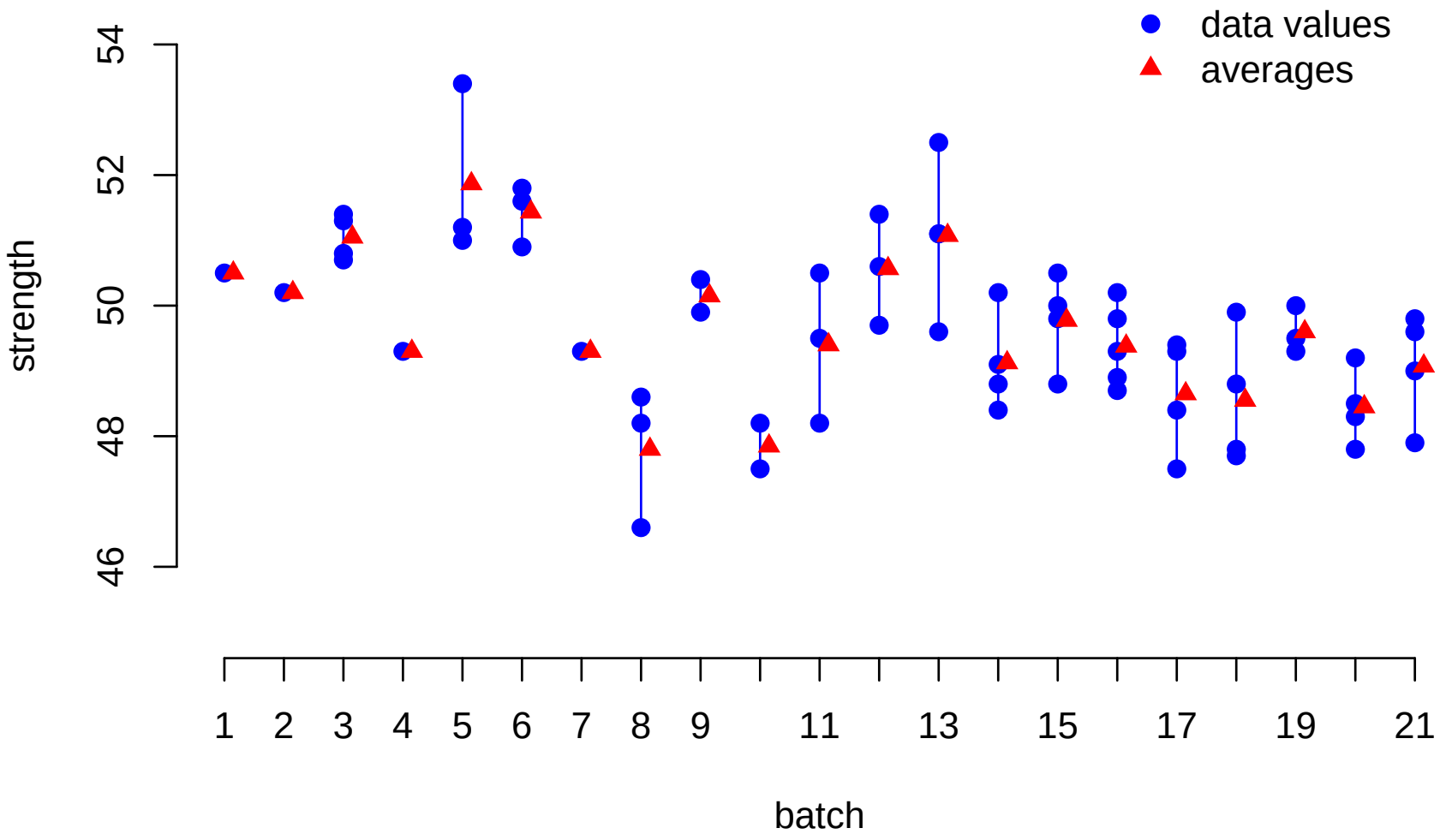
For batch effect situations that are positioned between these two extreme cases we propose to use the previously developed estimated effective sample size $\hat{N}^\star$ as a simple interpolation between $f+1$ and N and use as k-factor in the general case

$$k^\star(N) = \sqrt{\frac{N-1}{N}} \frac{1}{\sqrt{\hat{N}^\star - 1}} t_{\hat{N}^\star - 1, -z_p\sqrt{\hat{N}^\star}, \gamma}\,.$$

Batch Data

batch	n_i	sample data	$\bar{X}_{i\cdot}$
1	1	50.5	50.5
2	1	50.2	50.2
3	4	50.7, 50.8, 51.4, 51.3	51.05
4	1	49.3	49.3
5	3	51.0, 51.2, 53.4	51.867
6	3	50.9, 51.6, 51.8	51.433
7	1	49.3	49.3
8	3	48.6, 48.2, 46.6	47.8
9	2	50.4, 49.9	50.15
10	2	48.2, 47.5	47.85
11	3	50.5, 48.2, 49.5	49.4
12	3	49.7, 51.4, 50.6	50.567
13	4	49.6, 51.1, 51.1, 52.5	51.075
14	4	48.4, 50.2, 48.8, 49.1	49.125
15	4	48.8, 49.8, 50.0, 50.5	49.775
16	5	49.3, 50.2, 49.8, 48.9, 48.7	49.38
17	4	49.3, 47.5, 49.4, 48.4	48.65
18	4	47.8, 47.7, 48.8, 49.9	48.55
19	3	50.0, 49.5, 49.3	49.6
20	4	48.5, 49.2, 48.3, 47.8	48.45
21	4	47.9, 49.6, 49.8, 49.0	49.075

Batch Data


data values
averages
strength
46
48
50
52
54
1
2
3
4
5
6
7
8
9
11
13
15
17
19
21
batch

Calculations of A-Allowables (Without Batch Effect)

The Table gives composite material property data for 21 batches.

$\Longrightarrow \quad \bar{X} = 49.638$ and $S = 1.320$.

Ignoring the batch effects and assuming that we deal with $N = 63$ independent observations we obtain as k-factor for the A-allowable

$$k_A = \texttt{qnct}(.95, 63-1, -\texttt{qnorm}(.01) * \texttt{sqrt}(63))/\texttt{sqrt}(63) = 2.793392$$

and thus $A = \bar{X} - k_A S = 49.638 - 2.793392 * 1.320 = 45.95072$ as A-allowable.

The data plot shows strong batch effects.

The above A-allowable may not be appropriate.

Calculations of A-Allowables (Using $\hat{N}^{\star}$)

When adjusting by the "effective" sample size we obtain

$$SS_b = 78.921\ , \quad SS_e = 29.148\ , \quad f = 17.123\ , \quad \hat{\sigma}_e^2 = .6939\ , \quad \hat{\sigma}_b^2 = 1.093$$

and thus $\hat{\rho} = .6116$ and $N^{\star} = 25.056$. As k-factor for the A allowable we now get

$$k_A = \texttt{sqrt}((63-1)/63) * \texttt{qnct}(.95, 25.056-1, -\texttt{qnorm}(.01) * \texttt{sqrt}(25.056))/\texttt{sqrt}(25.056-1) = 3.195986$$

and thus $A = \bar{X} - k_A S = 49.638 - 3.195986 * 1.320 = 45.4193$ as A-allowable.

If the threshold, against which these allowables are compared, had been 45 then the allowables by either analysis fall on the same side of 45, namely above.

If the threshold had been 45.5 then the allowables fall on opposite sides of 45.5. The one accounting for the batch effect falls a little bit short. This may be mainly because of the "effective" sample size being too small.

Closer Examination

The data plot suggests that the measured values stabilize from batch 14 onward. Prior to that point the batch to batch variation seems quite strong.

Also, there may have been selective decisions on how many data points to gather, depending on the first and/or second measurement in each batch.
Such a selection bias would put in doubt any of the calculations made so far.

If we disregard these first 13 batches and obtain an A-allowable from the remaining 8 batches with a total of 32 observations we find $\bar{X} = 49.06875$ (not much changed) and $S = 0.8133711$ (quite a bit smaller) and the k-factor becomes

$$\texttt{qnct}(.95, 32-1, -\texttt{qnorm}(.01)*\texttt{sqrt}(32))/\texttt{sqrt}(32) = 3.033847$$

with resulting A-allowable $A =$`49.06875-3.033847*0.8133711=46.60111`.

Using the above interpolation method we find $N^\star = 22.44343$, $k_A = 3.243241$ and $A = 46.43079$, which is not that much different from 46.60111 and both values are significantly higher than the previous ones based on the full data set.

Validation Simulation

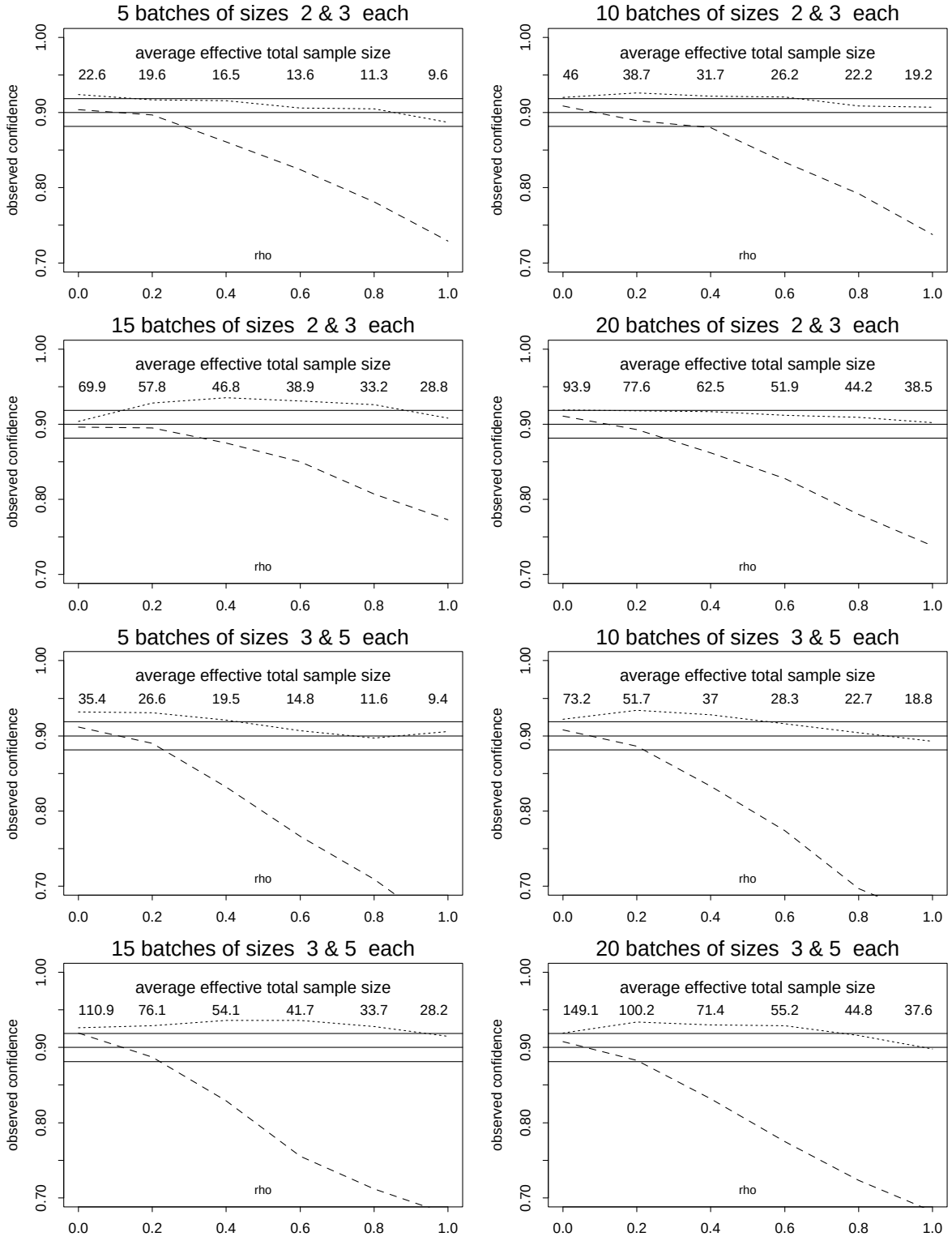
Simulations of the above process were run for the corresponding bounds on C_{pk}.

Since these were not discussed here we will only show one plot.

For various magnitudes of batch effects we observed the coverage rate of these bounds and found that the actual coverage came close to the nominal one, if not a bit higher.

The coverage probability of the method that ignored the batch effect fell off strongly as the batch variation became more and more dominant.

For details see the reference or the posted preprint.


5 batches of sizes 2 & 3 each
average effective total sample size
22.6 19.6 16.5 13.6 11.3 9.6
10 batches of sizes 2 & 3 each
average effective total sample size
46 38.7 31.7 26.2 22.2 19.2
15 batches of sizes 2 & 3 each
average effective total sample size
69.9 57.8 46.8 38.9 33.2 28.8
20 batches of sizes 2 & 3 each
average effective total sample size
93.9 77.6 62.5 51.9 44.2 38.5
5 batches of sizes 3 & 5 each
average effective total sample size
35.4 26.6 19.5 14.8 11.6 9.4
10 batches of sizes 3 & 5 each
average effective total sample size
73.2 51.7 37 28.3 22.7 18.8
15 batches of sizes 3 & 5 each
average effective total sample size
110.9 76.1 54.1 41.7 33.7 28.2
20 batches of sizes 3 & 5 each
average effective total sample size
149.1 100.2 71.4 55.2 44.8 37.6
observed confidence
rho
0.70 0.80 0.90 1.00
0.0 0.2 0.4 0.6 0.8 1.0

Tolerance Bounds in Regression

Applications of the noncentral t-distribution can easily be extended to more complex data situations, such as to regression.

Standard linear regression model: Responses $Y_1, \ldots, Y_n$ are observed under respectively varying but known conditions $x_1' = (x_{11}, \ldots, x_{1p}), \ldots, x_n' = (x_{n1}, \ldots, x_{np})$

$$Y_i = x_{i1}\beta_1 + \ldots + x_{ip}\beta_p + e_i = x_i'\beta + e_i\,, \qquad i = 1, \ldots, n\,.$$

$\beta_1, \ldots, \beta_p$ are unknown parameters, to be estimated from the data $(Y_1, x_1'), \ldots, (Y_n, x_n')$.

The terms $e_1, \ldots, e_n$ are the error terms that capture to what extent the observed values Y_i differ from the model values $x_i'\beta$.

It is typically assumed that these error terms are statistically independent with common $\mathcal{N}(0, \sigma^2)$ distribution, where the variance σ^2 is also unknown, to be estimated from the data as well.

A Composite Material Example

Consider the tensile strength of coupons of composite materials.

These consist of laminates, i.e., are built up from layers of lamina, typically using lamina with varying fiber ply orientations, such as 90°, 45° and 0°.

Such laminates are usually characterized by the percent of lamina in each orientation. Since these percentages have to add up to 100% it is only necessary to specify $k-1=2$ percentages when $k=3$ orientations are involved.

The response $Y=$ the coupon tensile strength.

$x=(x_1,x_2)$ gives the two percentages for lamina at 45° and 0° orientation.

In addition to the simple linear model in the covariates (x_1,x_2) one may also want to explore any quadratic effects, i.e., $x_3=x_1^2, x_4=x_2^2, x_5=x_1x_2$.

Testing Is Costly

Testing such coupons is costly.

Many possible lay-up orientation combinations make it prohibitive to test all these combinations extensively.

Test coupons in moderate numbers for several such combinations, carefully chosen to cover the space of lay-up percentages reasonably well.

Upfront it is not known which lay-up combination will give the best strength results.

It is entirely possible that coupons at such an optimal combination have not been tested at all for the initial experiment. However, such test runs can be added later in confirmatory testing or in order to tighten up the tolerance bounds.

Sources of Variation

The full data set would then consist of $(Y_1, x_{11}, x_{21}), \ldots, (Y_n, x_{1n}, x_{2n})$.

For the quadratic model this expands to $(Y_1, x_{11}, \ldots, x_{51}), \ldots, (Y_n, x_{1n}, \ldots, x_{5n})$.

Much of the variation in the strength measurement Y comes from testing itself.

Both the orientation at which the stress is applied and the orientation of the coupon as it is cut from the manufactured laminate can vary $\Rightarrow$ significant strength impact.

Other factors can cause response variation, e.g., chemical batch effects.

We confine ourselves to the pure regression model.

It should be possible to blend the batch effect methods with the solution for this pure regression model.

Matrix Notation

The pure regression model can be written more concisely in matrix notation

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{11}\beta_1 + \ldots + x_{1p}\beta_p \\ x_{21}\beta_1 + \ldots + x_{2p}\beta_p \\ \vdots \\ x_{n1}\beta_1 + \ldots + x_{np}\beta_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

$$= \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = X\beta + e\,.$$

It is usually assumed that $n > p$ and that the matrix X is of full rank p, i.e.,

its p columns $x_1, \ldots, x_p \in R^n$ are linearly independent.

This means that the equation $a_1x_1 + \ldots + a_px_p = 0$ only admits the solution

$a' = (a_1, \ldots, a_p) = (0, \ldots, 0) \Longrightarrow p \times p$ matrix $X'X$ has full rank p as well,

since $\quad X'Xu = 0 \implies u'X'Xu = 0 \implies Xu = 0 \implies u = 0.$

The Solution Process

$\Longrightarrow$ the equation $X'Xa = b$ has a unique solution $a = (X'X)^{-1}b$ for each b.

Here $(X'X)^{-1}$ is the inverse matrix to $X'X$.

A $p \times p$ matrix A is the inverse to a $p \times p$ matrix B if $AB = I = I_p$, where I_p is a $p \times p$ matrix with 1's on the diagonal and 0's off the diagonal.

Multiplying the above matrix data model by X' and then by $(X'X)^{-1}$ we get

$$X'Y = X'X\beta + X'e \implies$$

$$\begin{aligned}(X'X)^{-1}X'Y &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'e \\ &= \beta + (X'X)^{-1}X'e\end{aligned}$$

$$\hat{\beta} = \beta + (X'X)^{-1}X'e \quad \text{where} \quad \hat{\beta} = (X'X)^{-1}X'Y$$

$\hat{\beta}$ is also known as the least squares estimate of β.

The Least Squares Solution

$\hat{\beta}$ is the vector β that minimizes the following sum of squares

$$\sum_{i=1}^{n}(Y_i - x_{i1}\beta_1 - \ldots - x_{ip}\beta_p)^2 = \sum_{i=1}^{n}(Y_i - x_i'\beta)^2$$

$$\begin{aligned}
&= (Y - X\beta)'(Y - X\beta) = (Y - X\hat{\beta} + X\hat{\beta} - X\beta)'(Y - X\hat{\beta} + X\hat{\beta} - X\beta) \\
&= (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (Y - X\hat{\beta})'(X\hat{\beta} - X\beta) \\
&\qquad + (X\hat{\beta} - X\beta)'(Y - X\hat{\beta}) + (X\hat{\beta} - X\beta)'(X\hat{\beta} - X\beta) \\
&= (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (X\hat{\beta} - X\beta)'(X\hat{\beta} - X\beta) \qquad (3)
\end{aligned}$$

since $$(Y - X\hat{\beta})'(X\hat{\beta} - X\beta) = (Y - X(X'X)^{-1}X'Y)'X(\hat{\beta} - \beta)$$

$$= Y'(X - X(X'X)^{-1}X'X)(\hat{\beta} - \beta) = Y'(X - X)(\hat{\beta} - \beta) = 0\,.$$

The second term in (3) is minimized by taking $\beta = \hat{\beta}$ as is seen from

$$(X\hat{\beta} - X\beta)'(X\hat{\beta} - X\beta) = (X(\hat{\beta} - \beta))'X(\hat{\beta} - \beta) = (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \geq 0\,,$$

with equality if and only if $X(\hat{\beta} - \beta) = 0$, i.e., if $\hat{\beta} - \beta = 0$.

The other term in (3) does not depend on β.

The Response $Y(x_0)$

$Y(x_0) = (x_{01}\ldots,x_{0p})\beta + e = x_0'\beta + e \sim \mathcal{N}(x_0'\beta, \sigma^2)$

The natural estimate of $x_0'\beta$ is

$\hat{Y}(x_0) = x_0'\hat{\beta} = x_0'\beta + x_0'(X'X)^{-1}X'e \sim \mathcal{N}(x_0'\beta, \sigma^2 x_0'(X'X)^{-1}x_0) = \mathcal{N}(\mu(x_0), \tau^2(x_0))\,,$

where the mean $\mu(x_0) = x_0'\beta$ derives from the fact that $E(e_i) = 0$ for $i = 1,\ldots,n$ and the variance expression $\tau^2(x_0) = \sigma^2 x_0'(X'X)^{-1}x_0$ comes from

$$\begin{aligned} \mathrm{var}(x_0'(X'X)^{-1}X'e) &= \mathrm{var}(u'e) = \sigma^2 \sum_{i=1}^{n} u_i^2 = \sigma^2 u'u \\ &= \sigma^2 x_0'(X'X)^{-1}X'X(X'X)^{-1}x_0 = \sigma^2 x_0'(X'X)^{-1}x_0\,. \end{aligned}$$

The unknown parameter σ^2 can be estimated by the unbiased estimator

$$S^2 = \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - x_i'\hat{\beta})^2 = \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \frac{1}{n-p}\sum_{i=1}^{n}\hat{e}_i^2\,,$$

$\hat{Y}_i = x_i'\hat{\beta} =$ fitted values and $\hat{e}_i = Y_i - \hat{Y}_i =$ residuals, $i = 1,\ldots,n$.

$(n-p)S^2/\sigma^2 \sim \chi^2_{n-p}$ is independent of $\hat{\beta}$, i.e., also independent of $\hat{Y}(x_0) = x_0'\hat{\beta}$.

Tolerance Bounds at x_0

The p-quantile of the response $Y(x_0)$ is $y_p(x_0) = \mu(x_0) + \sigma z_p$

Its natural estimate is $\hat{Y}(x_0) + S z_p$. $\quad \bar{X} - kS \Longrightarrow \hat{Y}(x_0) - k(x_0)S.$

Note that the k-factor here depends on x_0.

$$Z = \frac{\hat{Y}(x_0) - \mu(x_0)}{\tau(x_0)} = \frac{\hat{Y}(x_0) - \mu(x_0)}{\sigma\sqrt{x_0'(X'X)^{-1}x_0}} \sim \mathcal{N}(0,1) \qquad \text{and} \qquad V = \frac{S^2(n-p)}{\sigma^2} \sim \chi^2_{n-p}$$

are independent. Abbreviating $\kappa(x_0) = \sqrt{x_0'(X'X)^{-1}x_0}$ we continue with

$$\begin{aligned}
\gamma = P(\hat{Y}(x_0) - kS \le y_p(x_0)) &= P(\hat{Y}(x_0) - kS \le \mu(x_0) + \sigma z_p) \\
&= P\left(\frac{\hat{Y}(x_0) - \mu(x_0) - \sigma z_p}{\sigma\kappa(x_0)} \le \frac{kS}{\sigma\kappa(x_0)}\right) = P\left(\frac{Z - z_p/\kappa(x_0)}{\sqrt{V/(n-p)}} \le \frac{k}{\kappa(x_0)}\right) \\
&= P\left(T_{n-p,\delta(x_0)} \le k/\kappa(x_0)\right) = G_{n-p,\delta(x_0)}(k/\kappa(x_0)), \quad \delta(x_0) = -z_p/\kappa(x_0)
\end{aligned}$$

Thus $k = \kappa(x_0) G^{-1}_{n-p,\delta(x_0)}(\gamma) =$ `kappa * qnct(gam, n - p, delta)`,

reg.tolbd and poly.tolbd

The R workspace contains a function `reg.tolbd` that calculates such $100\gamma\%$ lower confidence bounds for $y_p(x_0)$ for any specified (γ, p, x_0).

The intercept covariate is not input into this function, it is created internally.

A $100(1-\gamma)\%$ lower bound is a $100\gamma\%$ upper bound for $y_p(x_0)$

The documentation to `reg.tolbd` is given in the function body.

`poly.tolbd` is tailored to polynomial fits using a univariate explanatory variable.

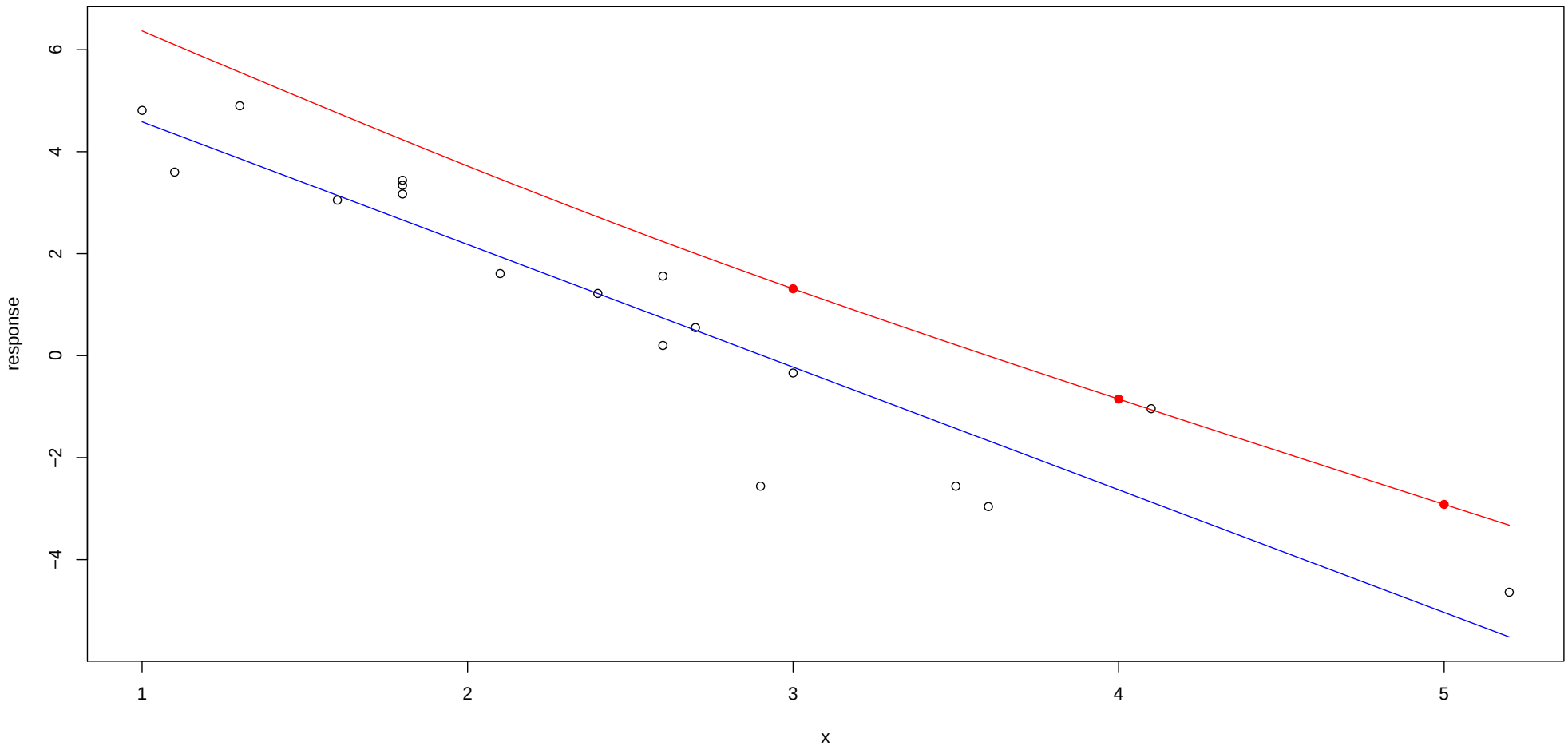
$$Y_i = x_{i0}\beta_0 + x_{i1}\beta_1 + \ldots + x_{ik}\beta_k + e_i\,, \quad \text{with} \quad x_{ij} = x_i^j,\ j = 0, 1, \ldots, k$$

The data are $(x_1, Y_1), \ldots, (x_n, Y_n)$

`poly.tolbd` has graphical output, see next slides.

Example data taken from Graybill (1976), pp. 274-276.

Linear Fit



Quadratic Fit