Lynch & Wash Chapter 2 - Properties of distributions

Jhonathan P. R. dos Santos

Ph.D Student at Universidade de São Paulo / ESALQ, Brazil
Advisor: Dr. Antonio Augusto Franco Garcia

Msc Genetics and Plant Breeding
Universidade Federal de Lavras, Brazil

Research Interest: Genomic selection, statistical and machine learning models

Outline

1. Parameters of Univariate Distributions
   - Characters
   - Distributions
   - The Normal distribution
   - Truncated Normal distribution
   - Confidence intervals

Types of Characters

Types of characters studied by biologists:

- **Meristic characters** (range of discrete classes)
  - Example: Leaf numbers
- **Metric characters** (range of discrete classes)
  - Example: Plant height
- **Binary characters** (range of discrete classes)
  - Example: Survival at a fixed age

Types of Characters

Types of characters studied by biologists:

- **Meristic characters** (range of discrete classes)
  - Example: Leaf numbers
- **Metric characters** (range of discrete classes)
  - Example: Plant height
- **Binary characters** (range of discrete classes)
  - Example: Survival at a fixed age

Types of Characters

Types of characters studied by biologists:

- **Meristic characters** (range of discrete classes)
  - Example: Leaf numbers
- **Metric characters** (range of discrete classes)
  - Example: Plant height
- **Binary characters** (range of discrete classes)
  - Example: Survival at a fixed age

Types of Distributions

We usually can describe data using three different types of distributions:

- Univariate distribution (UD):
  Definition: **"UD describes the relative frequency of phenotypes for a single trait"**
- Bivariate distribution (BD):
  Definition: **"BD describes the mutual distribution of two characters"**
- Multivariate distribution (MD):
  Definition: **"MD corresponds to the joint distribution of more than two traits"**

Types of Distributions

We usually can describe data using three different types of distributions:

- Univariate distribution (UD):
  Definition: **"UD describes the relative frequency of phenotypes for a single trait"**
- Bivariate distribution (BD):
  Definition: **"BD describes the mutual distribution of two characters"**
- Multivariate distribution (MD):
  Definition: **"MD corresponds to the joint distribution of more than two traits"**

Types of Distributions

We usually can describe data using three different types of distributions:

- Univariate distribution (UD):
  Definition: **"UD describes the relative frequency of phenotypes for a single trait"**
- Bivariate distribution (BD):
  Definition: **"BD describes the mutual distribution of two characters"**
- Multivariate distribution (MD):
  Definition: **"MD corresponds to the joint distribution of more than two traits"**

Types of Distributions

- **Goal of statistics:** Fit simple mathematical functions to data

- Types of probability distributions:
  - Discrete: **some variable z (e. g. offspring number) is completely described by $P(z = z_i)$**

    Book's example: $P(z = z_1) = \frac{15}{1005}$, given z_1=1, that is, proportion of mothers that produce a single offspring

  - Continuous: **Point probability is infinitesimal small (we should compute interval prob.)**

    Book's example: $P(a \leq z \leq b) = \int_a^b p(z)\, dz$

Types of Distributions

- **Goal of statistics:** Fit simple mathematical functions to data

- Types of probability distributions:
  - Discrete: **some variable z (e. g. offspring number) is completely described by $P(z = z_i)$**

    Book's example: $P(z = z_1) = \frac{15}{1005}$, given z_1=1, that is, proportion of mothers that produce a single offspring

  - Continuous: **Point probability is infinitesimal small (we should compute interval prob.)**

    Book's example: $P(a \leq z \leq b) = \int_a^b p(z)\, dz$

Distributions properties

- **Key property:**
  - $\int_{z_{max}}^{z_{min}} p(z)\, dz = 1$ (Continuous case)
  - $\sum_{i=1} p(z_i) = 1$ (Discrete case)
- What is a easy way to integrate probability distributions?
  - **Monte Carlo Integration!**

Distributions properties

- **Key property:**
  - $\int_{z_{max}}^{z_{min}} p(z)\, dz = 1$ (Continuous case)
  - $\sum_{i=1} p(z_i) = 1$ (Discrete case)
- What is a easy way to integrate probability distributions?
  - **Monte Carlo Integration!**

Example of computing probabilities

Example 1: Suppose that z is a continuously distributed in the range of 0 to ∞ with pdf negative exponential distribution:

$$p(z) = \frac{1}{\lambda} \exp\left[\frac{-z}{\lambda}\right]$$

What is the probability that a random sample will have z in the range of $\frac{1}{4}$ to $\frac{1}{2}$?

$$P(\tfrac{1}{4} \leq z \leq \tfrac{1}{2}|\lambda) = \int_{\frac{1}{4}}^{\frac{1}{2}} p(z)\,dz = \int_{\frac{1}{4}}^{\frac{1}{2}} \frac{1}{\lambda} \exp\left[\frac{-z}{\lambda}\right] dz$$

Rule: $\int e^{ax} dx = \frac{1}{a} e^{ax}$, in our case:

$$\int_{\frac{1}{4}}^{\frac{1}{2}} p(z)\,dz = \frac{1}{\lambda} - \lambda \exp\left[\frac{-z}{\lambda}\right]\Big|_{\frac{1}{4}}^{\frac{1}{2}} = -\exp\left[\frac{-z}{\lambda}\right]\Big|_{\frac{1}{4}}^{\frac{1}{2}} =$$
$$\exp\left[\frac{-1}{4\lambda}\right] - \exp\left[\frac{-1}{2\lambda}\right]$$

for $\lambda = \frac{1}{2}$, then $P(\frac{1}{4} \leq z \leq \frac{1}{2}|\lambda) = 0.239$

Example of computing probabilities

Monte Carlo integration: **R code!**

Parameters *vs* Estimates

- True parameters are only obtained if all population units are measured with high accuracy
- Estimates are statistics obtained from analysing sampled data
- Accuracy of estimates rely on **experimental setting**, **measurement apparatus** and the **sample size**

Parameters *vs* Estimates

- True parameters are only obtained if all population units are measured with high accuracy
- Estimates are statistics obtained from analysing sampled data
- Accuracy of estimates rely on **experimental setting**, **measurement apparatus** and the **sample size**

Parameters *vs* Estimates

- True parameters are only obtained if all population units are measured with high accuracy
- Estimates are statistics obtained from analysing sampled data
- Accuracy of estimates rely on **experimental setting**, **measurement apparatus** and the **sample size**

Parameters *vs* Estimates

- Most useful probability density function are defined by:
- **Central location:**
  - $\mu = \int_{-\infty}^{+\infty} z\, p(z)\, dz = E(z)$
  - $\mu = \sum_{i=1} p(z = z_i)$
- **Variance parameter:**

$$\sigma^2 = \int_{-\infty}^{+\infty} (z-\mu)^2 p(z)\, dz = E\left[(z-\mu)^2\right]$$

- Some useful expectation properties:
  - $E(x+y) = E(x) + E(y)$
  - $E(c\,x) = c\,E(x)$

Parameters *vs* Estimates

- Most useful probability density function are defined by:
- **Central location**:
  - $\mu = \int_{-\infty}^{+\infty} z\, p(z)\, dz = E(z)$
  - $\mu = \sum_{i=1} p(z = z_i)$
- **Variance parameter:**

  $\sigma^2 = \int_{-\infty}^{+\infty} (z-\mu)^2 p(z)\, dz = E\left[(z-\mu)^2\right]$
- Some useful expectation properties:
  - $E(x+y) = E(x) + E(y)$
  - $E(c\,x) = c\,E(x)$

Parameters *vs* Estimates

- Most useful probability density function are defined by:
- **Central location**:
  - $\mu = \int_{-\infty}^{+\infty} z\, p(z)\, dz = E(z)$
  - $\mu = \sum_{i=1} p(z = z_i)$
- **Variance parameter**:

$$\sigma^2 = \int_{-\infty}^{+\infty} (z-\mu)^2 p(z)\, dz = E\left[(z-\mu)^2\right]$$

- Some useful expectation properties:
  - $E(x+y) = E(x) + E(y)$
  - $E(c\,x) = c\,E(x)$

Parameters *vs* Estimates

- Most useful probability density function are defined by:
- **Central location**:
  - $\mu = \int_{-\infty}^{+\infty} z\, p(z)\, dz = E(z)$
  - $\mu = \sum_{i=1} p(z = z_i)$
- **Variance parameter**:

$$\sigma^2 = \int_{-\infty}^{+\infty} (z - \mu)^2 p(z)\, dz = E\left[(z - \mu)^2\right]$$

- Some useful expectation properties:
  - $E(x + y) = E(x) + E(y)$
  - $E(c\, x) = c\, E(x)$

Parameters *vs* Estimates

- Variance has the problem of being harder to estimate with small samples

  Recommendation: $E\left[E(z-\mu)^2\right] = \frac{1}{n-1}\sum_{i=1}(y_i - \bar{y})^2$

- Square root of the variance: standard deviation or scale parameter
- Relative measure of dispersion: $CV(z) = \frac{SD(z)}{\bar{z}}$
- Measure of a asymmetry of a distribution (skewness):

  $\mu_3 = E\left[(z-\mu)^3\right] = \int_{-\infty}^{+\infty}(z-\mu)^3 p(z)\,dz$

  $\mu_3 = E\left[z^3\right] - 3\mu E\left[z^2\right] + 2\mu^3$

Parameters *vs* Estimates

- Variance has the problem of being harder to estimate with small samples

  Recommendation: $E\left[E(z-\mu)^2\right] = \frac{1}{n-1}\sum_{i=1}(y_i-\bar{y})^2$

- Square root of the variance: standard deviation or scale parameter
- Relative measure of dispersion: $CV(z) = \frac{SD(z)}{\bar{z}}$
- Measure of a asymmetry of a distribution (skewness):

  $\mu_3 = E\left[(z-\mu)^3\right] = \int_{-\infty}^{+\infty}(z-\mu)^3 p(z)\,dz$

  $\mu_3 = E\left[z^3\right] - 3\mu E\left[z^2\right] + 2\mu^3$

Parameters *vs* Estimates

- Variance has the problem of being harder to estimate with small samples

  Recommendation: $E\left[E(z-\mu)^2\right] = \frac{1}{n-1}\sum_{i=1}(y_i - \bar{y})^2$

- Square root of the variance: standard deviation or scale parameter
- Relative measure of dispersion: $CV(z) = \frac{SD(z)}{\bar{z}}$
- Measure of a asymmetry of a distribution (skewness):

  $\mu_3 = E\left[(z-\mu)^3\right] = \int_{-\infty}^{+\infty}(z-\mu)^3 p(z)\,dz$

  $\mu_3 = E\left[z^3\right] - 3\mu E\left[z^2\right] + 2\mu^3$

Parameters *vs* Estimates

- Variance has the problem of being harder to estimate with small samples

  Recommendation: $E\left[E(z-\mu)^2\right] = \frac{1}{n-1}\sum_{i=1}(y_i-\bar{y})^2$

- Square root of the variance: standard deviation or scale parameter

- Relative measure of dispersion: $CV(z) = \frac{SD(z)}{\bar{z}}$

- Measure of a asymmetry of a distribution (skewness):

  $\mu_3 = E\left[(z-\mu)^3\right] = \int_{-\infty}^{+\infty}(z-\mu)^3 p(z)\,dz$

  $\mu_3 = E\left[z^3\right] - 3\mu E\left[z^2\right] + 2\mu^3$

Parameters *vs* Estimates

- The degree of asymmetry can also be obtained by the coefficient of skewness:
  - $k_3 = \frac{Skw(z)}{Var(z)^{\frac{3}{2}}}$
  - $k_3 > 0$ longer tail to the right
  - $k_3 < 0$ longer tail to the left
  - $k_3 = 0$ perfectly symmetrical distribution
- General expression for the rth moment about the mean:

$\mu_r = \int_{-\infty}^{+\infty} (z-\mu)^r p(z)\, dz$

Parameters *vs* Estimates

- The degree of asymmetry can also be obtained by the coefficient of skewness:
  - $k_3 = \frac{Skw(z)}{Var(z)^{\frac{3}{2}}}$
  - $k_3 > 0$ longer tail to the right
  - $k_3 < 0$ longer tail to the left
  - $k_3 = 0$ perfectly symmetrical distribution
- General expression for the rth moment about the mean:

$\mu_r = \int_{-\infty}^{+\infty} (z - \mu)^r p(z)\, dz$

Normal p.d.f. features

- In most cases, large datasets when plotted in form of histograms often approximate bell-shaped distribution
- Normal distribution developed by DeMoivre(1738), LaPlace(1778) and **Gauss(1801)**
- Probability density function ($z(\mu, \sigma^2)$)

$$p(z) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right]$$

- **Some properties:** mean = median, symmetrical and has two parameters, the mean (location) and variance (scale)
- **Key properties:** Nice mathematical properties, random variables that deviates normality can be rescaled by transformation

Normal p.d.f. features

- In most cases, large datasets when plotted in form of histograms often approximate bell-shaped distribution
- Normal distribution developed by DeMoivre(1738), LaPlace(1778) and **Gauss(1801)**
- Probability density function ($z(\mu, \sigma^2)$)

$$p(z) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right]$$

- **Some properties:** mean = median, symmetrical and has two parameters, the mean (location) and variance (scale)
- **Key properties:** Nice mathematical properties, random variables that deviates normality can be rescaled by transformation

Normal p.d.f. features

- In most cases, large datasets when plotted in form of histograms often approximate bell-shaped distribution
- Normal distribution developed by DeMoivre(1738), LaPlace(1778) and **Gauss(1801)**
- Probability density function ($z(\mu, \sigma^2)$)

$$p(z) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right]$$

- **Some properties:** mean = median, symmetrical and has two parameters, the mean (location) and variance (scale)
- **Key properties:** Nice mathematical properties, random variables that deviates normality can be rescaled by transformation

Normal p.d.f. features

- In most cases, large datasets when plotted in form of histograms often approximate bell-shaped distribution
- Normal distribution developed by DeMoivre(1738), LaPlace(1778) and **Gauss(1801)**
- Probability density function ($z(\mu, \sigma^2)$)

  $p(z) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right]$
- **Some properties:** mean = median, symmetrical and has two parameters, the mean (location) and variance (scale)
- **Key properties:** Nice mathematical properties, random variables that deviates normality can be rescaled by transformation

Normal p.d.f. features

- In most cases, large datasets when plotted in form of histograms often approximate bell-shaped distribution
- Normal distribution developed by DeMoivre(1738), LaPlace(1778) and **Gauss(1801)**
- Probability density function ($z(\mu, \sigma^2)$)

$$p(z) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right]$$

- **Some properties:** mean = median, symmetrical and has two parameters, the mean (location) and variance (scale)
- **Key properties:** Nice mathematical properties, random variables that deviates normality can be rescaled by transformation

Normal p.d.f. features

- Why most of the quantitative traits follows a normal p.d.f. ?
  - **Reason:** Central limit theorem + infinitesimal genetic architecture hypothesis
- Central limit theorem states that "the sum of a number of independent random variables approaches normality as the number of variables increases"
- **Tiny limitation:** possibility of negative values realizations (sampling)

Normal p.d.f. features

- Why most of the quantitative traits follows a normal p.d.f. ?
  - **Reason:** Central limit theorem + infinitesimal genetic architecture hypothesis
- Central limit theorem states that "the sum of a number of independent random variables approaches normality as the number of variables increases"
- **Tiny limitation:** possibility of negative values realizations (sampling)

Normal p.d.f. features

- Why most of the quantitative traits follows a normal p.d.f. ?
  - **Reason:** Central limit theorem + infinitesimal genetic architecture hypothesis
- Central limit theorem states that "the sum of a number of independent random variables approaches normality as the number of variables increases"
- **Tiny limitation:** possibility of negative values realizations (sampling)

Normal p.d.f. features

- Discrete data with large number of categories approximate normality as these numbers increases (result of central limit theorem)
- It is frequently useful to work with the standardized form of the normal density:
  - $p(z') = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{(z')^2}{2}\right]$
- **Other features:**
  - $\mu_3 = 0$ (symmetrical distribution)
  - $k_4 = \frac{Kur(z)-3[Var(z)]^2}{[Var(z)]^2} = 0$ (index to measure kurtosis)
  - **Deviation from normality:** $k_4 > 0$ (leptokurtic, e. g. Cauchy density), $k_4 < 0$ (platykurtic, e. g. Uniform density)

Normal p.d.f. features

- Discrete data with large number of categories approximate normality as these numbers increases (result of central limit theorem)
- It is frequently useful to work with the standardized form of the normal density:
  - $p(z') = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{(z')^2}{2}\right]$
- **Other features:**
  - $\mu_3 = 0$ (symmetrical distribution)
  - $k_4 = \frac{Kur(z)-3[Var(z)]^2}{[Var(z)]^2} = 0$ (index to measure kurtosis)
  - **Deviation from normality:** $k_4 > 0$ (leptokurtic, e. g. Cauchy density), $k_4 < 0$ (platykurtic, e. g. Uniform density)

Normal p.d.f. features

- Discrete data with large number of categories approximate normality as these numbers increases (result of central limit theorem)
- It is frequently useful to work with the standardized form of the normal density:
  - $p(z') = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{(z')^2}{2}\right]$
- **Other features:**
  - $\mu_3 = 0$ (symmetrical distribution)
  - $k_4 = \frac{Kur(z) - 3[Var(z)]^2}{[Var(z)]^2} = 0$ (index to measure kurtosis)
  - **Deviation from normality:** $k_4 > 0$ (leptokurtic, e. g. Cauchy density), $k_4 < 0$ (platykurtic, e. g. Uniform density)

Truncated Normal distribution (TND) features

- The TND corresponds to some subset values of the normal distribution
- Extremely import for theory of **truncation selection**
- Truncation selection:
  - Critical phenotype T is called truncation point
  - Mean of the subset of the distribution (mean of the selected fraction of genotypes) can be computed by:
    $\mu_S = \frac{\int_T^{+\infty} zp(z)\,dz}{\int_T^{+\infty} p(z)\,dz}$
  - After reparametrizations and the integration we got:
    $\mu_S = \mu + \frac{\sigma\, p_T}{\Phi_T}$, where $p_T = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{(T-\mu)^2}{2\sigma^2}\right]$
- p_T can be obtained by the *dnorm(T,0,1)* R function

Truncated Normal distribution (TND) features

- The TND corresponds to some subset values of the normal distribution
- Extremely import for theory of **truncation selection**
- Truncation selection:
  - Critical phenotype T is called truncation point
  - Mean of the subset of the distribution (mean of the selected fraction of genotypes) can be computed by:
    $\mu_S = \frac{\int_T^{+\infty} zp(z)\,dz}{\int_T^{+\infty} p(z)\,dz}$
  - After reparametrizations and the integration we got:
    $\mu_S = \mu + \frac{\sigma\, p_T}{\Phi_T}$, where $p_T = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{(T-\mu)^2}{2\sigma^2}\right]$
- p_T can be obtained by the *dnorm(T,0,1)* R function

Truncated Normal distribution (TND) features

- The TND corresponds to some subset values of the normal distribution
- Extremely import for theory of **truncation selection**
- Truncation selection:
  - Critical phenotype T is called truncation point
  - Mean of the subset of the distribution (mean of the selected fraction of genotypes) can be computed by:
    $\mu_s = \frac{\int_T^{+\infty} zp(z)\, dz}{\int_T^{+\infty} p(z)\, dz}$
  - After reparametrizations and the integration we got:
    $\mu_s = \mu + \frac{\sigma\, p_T}{\Phi_T}$, where $p_T = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{(T-\mu)^2}{2\sigma^2}\right]$
- p_T can be obtained by the *dnorm(T,0,1)* R function

Truncated Normal distribution (TND) features

- The TND corresponds to some subset values of the normal distribution
- Extremely import for theory of **truncation selection**
- Truncation selection:
  - Critical phenotype T is called truncation point
  - Mean of the subset of the distribution (mean of the selected fraction of genotypes) can be computed by:
    $\mu_s = \frac{\int_T^{+\infty} zp(z)\,dz}{\int_T^{+\infty} p(z)\,dz}$
  - After reparametrizations and the integration we got:
    $\mu_s = \mu + \frac{\sigma\, p_T}{\Phi_T}$, where $p_T = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{(T-\mu)^2}{2\sigma^2}\right]$
- p_T can be obtained by the *dnorm(T,0,1)* R function

Truncated Normal distribution (TND) features

- *Directional selection differential*: $\mu_s - \mu = \frac{\sigma\, p_T}{\Phi}$
- Units of phenotypic standard deviations: $\frac{(\mu_s - \mu)}{\sigma} = \frac{p_T}{\Phi_T}$
- Variance after selection: $\sigma_s^2 = \left[1 + \frac{p_T\, z'}{\Phi_T} - \left(\frac{p_T}{\Phi_T}\right)^2\right] \sigma^2$
- which gives the fraction of phenotypic variance after selection $\left(\frac{\sigma_s^2}{\sigma^2}\right)$

Truncated Normal distribution (TND) features

- *Directional selection differential*: $\mu_s - \mu = \frac{\sigma\, p_T}{\Phi}$
- Units of phenotypic standard deviations: $\frac{(\mu_s - \mu)}{\sigma} = \frac{p_T}{\Phi_T}$
- Variance after selection: $\sigma_s^2 = \left[1 + \frac{p_T\, z'}{\Phi_T} - \left(\frac{p_T}{\Phi_T}\right)^2\right] \sigma^2$
- which gives the fraction of phenotypic variance after selection $\left(\frac{\sigma_s^2}{\sigma^2}\right)$

Truncated Normal distribution (TND) features

- *Directional selection differential*: $\mu_s - \mu = \frac{\sigma\, p_T}{\Phi}$
- Units of phenotypic standard deviations: $\frac{(\mu_s - \mu)}{\sigma} = \frac{p_T}{\Phi_T}$
- Variance after selection: $\sigma_s^2 = \left[1 + \frac{p_T\, z'}{\Phi_T} - \left(\frac{p_T}{\Phi_T}\right)^2\right] \sigma^2$
- which gives the fraction of phenotypic variance after selection $\left(\frac{\sigma_s^2}{\sigma^2}\right)$

Truncated Normal distribution (TND) features

- *Directional selection differential*: $\mu_s - \mu = \frac{\sigma \, p_T}{\Phi}$
- Units of phenotypic standard deviations: $\frac{(\mu_s - \mu)}{\sigma} = \frac{p_T}{\Phi_T}$
- Variance after selection: $\sigma_s^2 = \left[1 + \frac{p_T \, z'}{\Phi_T} - \left(\frac{p_T}{\Phi_T}\right)^2\right] \sigma^2$
- which gives the fraction of phenotypic variance after selection $\left(\frac{\sigma_s^2}{\sigma^2}\right)$

Truncated Normal distribution (TND) features

- All estimates have sampling error
- Interval estimates is a range of values estimated from the data that is likely given a α level of significance probability that the true parameter value is present.
- Interval estimates are useful to mitigate wrong conclusions about parameter values realization.

Truncated Normal distribution (TND) features

- All estimates have sampling error
- Interval estimates is a range of values estimated from the data that is likely given a α level of significance probability that the true parameter value is present.
- Interval estimates are useful to mitigate wrong conclusions about parameter values realization.

Truncated Normal distribution (TND) features

- All estimates have sampling error
- Interval estimates is a range of values estimated from the data that is likely given a α level of significance probability that the true parameter value is present.
- Interval estimates are useful to mitigate wrong conclusions about parameter values realization.

Truncated Normal distribution (TND) features

- Confidence intervals are obtained by theoretical mathematical expressions based upon frequentist theory:
  $$\alpha = P\left[(\bar{z} - \Delta) \leq \mu \leq (\bar{z} - \Delta)\right] = \int_{\frac{-\Delta}{\sigma(\bar{z})}}^{\frac{+\Delta}{\sigma(\bar{z})}} p(z')\, dz'$$
  in which $z' = \frac{(\bar{z}-\mu)}{\sigma(\bar{z})}$ and $\frac{\Delta}{\sigma(\bar{z})}$ distance in standard errors that the observed statistic and parametric value will lie with probability α
- Mean case: $\bar{z} \pm 1.96 \left[\frac{var(z)}{n}\right]^{\frac{1}{2}}$
- Variance case (approximation): $2\sqrt{var(z)}$ (large sample) - 95% confidence interval

Truncated Normal distribution (TND) features

- Confidence intervals are obtained by theoretical mathematical expressions based upon frequentist theory:

  $$\alpha = P\left[(\bar{z} - \Delta) \leq \mu \leq (\bar{z} - \Delta)\right] = \int_{\frac{-\Delta}{\sigma(\bar{z})}}^{\frac{+\Delta}{\sigma(\bar{z})}} p(z')\, dz'$$

  in which $z' = \frac{(\bar{z}-\mu)}{\sigma(\bar{z})}$ and $\frac{\Delta}{\sigma(\bar{z})}$ distance in standard errors that the observed statistic and parametric value will lie with probability α

- Mean case: $\bar{z} \pm 1.96 \left[\frac{var(z)}{n}\right]^{\frac{1}{2}}$

- Variance case (approximation): $2\sqrt{var(z)}$ (large sample) - 95% confidence interval

Truncated Normal distribution (TND) features

- Confidence intervals are obtained by theoretical mathematical expressions based upon frequentist theory:
  $$\alpha = P\left[(\bar{z} - \Delta) \leq \mu \leq (\bar{z} - \Delta)\right] = \int_{\frac{-\Delta}{\sigma(\bar{z})}}^{\frac{+\Delta}{\sigma(\bar{z})}} p(z')\, dz'$$
  in which $z' = \frac{(\bar{z}-\mu)}{\sigma(\bar{z})}$ and $\frac{\Delta}{\sigma(\bar{z})}$ distance in standard errors that the observed statistic and parametric value will lie with probability α
- Mean case: $\bar{z} \pm 1.96 \left[\frac{var(z)}{n}\right]^{\frac{1}{2}}$
- Variance case (approximation): $2\sqrt{var(z)}$ (large sample) - 95% confidence interval

Truncated Normal distribution (TND) features

Thanks!

"les charmes enchanteurs de cette sublime science ne se décèlent dans toute leur beauté qu'à ceux qui ont le courage de l'approfondir" - Letter from Gauss to Sophie Germain (30 April 1807).