Review of mathematical statistics

STATISTICAL INFERENCE

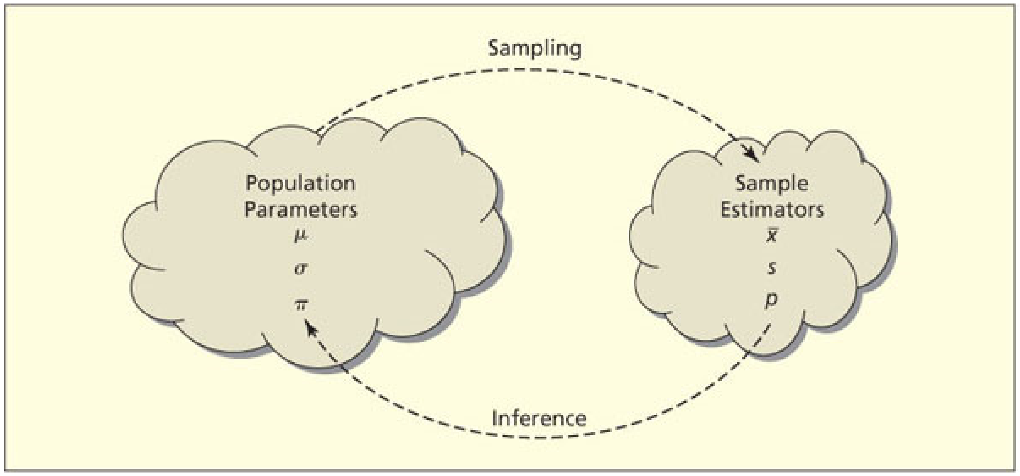


FIGURE 8.3

Sample Estimators of Population Parameters

Main point: Remember that our main concern is to use the sample to conclude about unknown aspects of the population

- **Population** – Model specification
  - Parametric and non-parametric models – What is our previous knowledge about the population?
  - How to define a model?
- **Sampling process –** How to collect the sample?
  - **Random sample**: Independent and identically distributed observations
  - Other sampling processes: stratified sampling, cluster sampling or non-random processes like convenience sampling or snowball (you ask the participants to provide you with names of those that will be able to provide you with important information) ….
  - Understanding variability

- **Statistical inference** – The role of uncertainty
  - Parametric and non-parametric inference: Population $X \sim f(x|\theta)$

    If the density (probability) function $f(.)$ is known (and θ is unknown) we face a parametric inference problem.

    If $f(.)$ (and possibly θ) is unknown we face a non-parametric problem.
- **Random sample** $(X_1, X_2, \cdots, X_n)$
  - Sample space
  - Sample distribution (this is a central concept in “classical” statistics)
  - Example: $X \sim Ber(\theta)$ sample size $n = 3$ $2^3 = 8$ possible sample

$$
\begin{aligned}
f(x_1, x_2, x_3) &= \Pr(X_1 = x_1) \times \Pr(X_2 = x_2) \times \Pr(X_3 = x_3) && \text{independent observations} \\
&= \Pr(X = x_1) \times \Pr(X = x_2) \times \Pr(X = x_3) && \text{identically distributed} \\
&= \theta^{x_1} (1-\theta)^{1-x_1}\ \theta^{x_2} (1-\theta)^{1-x_2}\ \theta^{x_3} (1-\theta)^{1-x_3} && \text{Bernoulli distribution} \\
&= \theta^{x_1+x_2+x_3} (1-\theta)^{3-(x_1+x_2+x_3)} \qquad x_i \in \{0,1\} \quad i = 1,2,3
\end{aligned}
$$

- **Statistic**
    - Definition: Real valued or vector-valued function of the random sample. The domain of the function is the sample space
    - Sampling distribution of a statistic
    - Example: Return to previous example and obtain the sampling distribution of $T = X_1 + X_2$
- How to get the sampling distribution of a statistic?
    - General approach: $F_{\mathbf{X}}(t) = \Pr\left(T\left(X_1, X_2, \cdots, X_n\right) \le t\right)$
    - Theoretical results – most of them proved using the moment generating function of X (the characteristic function)
    - Approximate procedures
        - Central limit theorem
        - Monte-Carlo simulation (to be developed latter)

- Examples – Sampling distribution of a statistic
  - Sample average from a normal population with known mean and variance;
  - Sampling distribution of $T = \sum_{i=1}^{n} X_i$ when we are sampling from a Bernoulli population.

- **Sample moments**
  - k-th sample moment about 0: $M'_k = (1/n)\sum_{i=1}^{n} X_i^k$ $\tilde{\mu}'_k$
    - Sample mean $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$
  - k-th sample moment about $\bar{X}$: $M_k = (1/n)\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^k$ $\tilde{\mu}_k$
    - Sample variance

$$M_2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

- Sample moments *versus* population moments

- Some results (we assume that the corresponding population moments exist)
  - Sample mean

$$E(\bar{X}) = E(X) = \mu \;; \qquad \operatorname{var}(\bar{X}) = \frac{\operatorname{var}(X)}{n} = \frac{\sigma^2}{n}.$$

  - Central limit theorem

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} \overset{\circ}{\sim} n(0,1)$$

  - Sample variance

$$E(M_2) = \frac{n-1}{n}\sigma^2 \qquad \operatorname{var}(M_2) = \frac{\mu_4-\mu_2^2}{n} - 2\frac{\mu_4-2\mu_2^2}{n^2} + \frac{\mu_4-3\mu_2^2}{n^3} \text{ with } \mu_k = E(X-\mu)^k$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \qquad n>1$$

$$E(S^2) = \sigma^2 \qquad \operatorname{var}(S^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right)$$

Order statistics

- **Definition:** The **order statistics** of a random sample $(X_1, X_2, \cdots, X_n)$ are the sample values placed in ascending order. They are denoted by $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ or by $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ or $Y_1 \le Y_2 \le \cdots \le Y_n$

- **Comments:**
  - Unlike the random sample itself, the order statistics are **not independent**. If $Y_j > y$ then $Y_s > y$ for $s > j$.
  - The sample minimum and the sample maximum are examples of order statistics.
  - Remember that the sample median is defined to be the middle order statistic if n is odd ($Y_{(n+1)/2}$) or the average of the middle two order statistics if n is even ($0.5 \times Y_{n/2} + 0.5 \times Y_{1+n/2}$).

- Marginal cumulative distribution of the *r*-th order statistic: Let $(X_1, X_2, \cdots, X_n)$ denote a random sample of size n from a population with cumulative distribution function $F_X(x)$.

  The marginal cumulative distribution will be $F_{Y_r}(y) = \sum_{j=r}^{n} \binom{n}{j} (F_X(y))^j (1 - F_X(y))^{n-j}$

Proof:

$$
\begin{aligned}
F_{Y_r}(y) &= \Pr(Y_r \le y) \\
&= \Pr(Y_r \le y \wedge Y_{r+1} > y) + \Pr(Y_{r+1} \le y \wedge Y_{r+2} > y) + \cdots + \Pr(Y_{n-1} \le y \wedge Y_n > y) + \Pr(Y_n \le y) \\
&= \binom{n}{r} (F_X(y))^r (1 - F_X(y))^{n-r} + \binom{n}{r+1} (F_X(y))^{r+1} (1 - F_X(y))^{n-(r+1)} + \\
&\quad + \cdots + \binom{n}{n-1} (F_X(y))^{n-1} (1 - F_X(y))^{n-(n-1)} + \binom{n}{n} (F_X(y))^n (1 - F_X(y))^{n-n} \\
&= \sum_{j=r}^{n} \binom{n}{j} (F_X(y))^j (1 - F_X(y))^{n-j}
\end{aligned}
$$

- If X is a continuous random variable the **density function of the *r*-th order statistic** will be

$$f_{Y_r}(y) = \frac{n!}{(r-1)!1!(n-r)!}\left(F_X(y)\right)^{r-1}\left(1-F_X(y)\right)^{n-r} f_X(y)$$

Proof: see Casella and Berger, 2nd edition, p 229.

- **Examples**: Let us consider a continuous random variable following an exponential distribution with mean θ and a sample of size 5.

1. The density function of the sample median will be

$$f_{Y_3}(y) = \frac{5!}{2!1!2!}\left(1-e^{-y/\theta}\right)^2\left(e^{-y/\theta}\right)^2\theta^{-1}e^{-y/\theta} = 30\,\theta^{-1}\left(1-e^{-y/\theta}\right)^2 e^{-3y/\theta} \qquad y>0$$

2. Compute the density and the distribution function of the sample maximum.
3. Identify the distribution of the sample minimum.

- Let X be a continuous random variable with distribution function $F(x)$ and density $f(x)$. $F(x)$ is strictly monotone for $0 < F(x) < 1$, and let m be the population median (m is the unique solution of $F(m) = 1/2$). Let M be the sample median. Then, it can be proved that M is asymptotically distributed as a normal variable with mean m and variance $(4n\,f(m)^2)^{-1}$, i.e. $(M-m)\left(2\,f(m)\sqrt{n}\right)\overset{\circ}{\sim} n(0;1)$

- **Example**: What is the asymptotic distribution of the median of a sample from a normal population?

  As we know, for the normal distribution, the population mean is equal to the population median, $m = \mu$. However, sample mean and sample median are different (but expected not to be very different). We get $(M-\mu)\left(\dfrac{2\sqrt{n}}{\sigma\sqrt{2\pi}}\right) = \dfrac{M-\mu}{\left(\sigma\sqrt{\pi/2}\right)/\sqrt{n}} \overset{\circ}{\sim} n(0;1)$ and we know that $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim n(0;1)$. Similar distribution, same mean but the variability of M is greater than the variability of $\bar{X}$.

POINT ESTIMATION

- We are in the core of parametric inference i.e. we have a model and we want to estimate the unknown parameter(s), i.e. $X \sim f(x|\theta)$, $\theta \in \Theta$ where $f(.)$ is a known density (probability) function and θ is an unknown parameter.
- In real world we could also consider that our knowledge of $f(.)$ is questionable but, at this stage, we will not proceed in such direction.
- They are 2 main problems in point estimation:
  - How to find estimators? To be discussed later.
  - How to evaluate the "quality" of an estimator?

At this point we only look for an answer to the second question.

How to evaluate the "quality" of an estimator?

- The important thing to notice is that we **will evaluate the procedure that generates the estimate** and not the estimate itself.

  We must distinguish between **estimator** and **estimate**.

- Keep in mind that a good procedure can lead to a poor estimate and conversely a poor procedure can originate a good estimate. However good procedures are more likely to produce good estimates than poor procedures.

- This evaluation is performed considering the set of results that could have been generated by the procedure and not a particular one.

- **Example**: If we want to estimate the mean, θ, of a normal population with known variance σ^2, the intuitive procedure is to use the sample average, i.e. $\overline{X} = (1/n)\sum_{i=1}^{n} X_i$ as the estimator or $\overline{x} = (1/n)\sum_{i=1}^{n} x_i$ as an estimate. The quality of the procedure (the estimator) is evaluated using the sampling distribution of $\overline{X}$ (and not the value given by $\overline{x}$).

Unbiasedness

- **Definition 10.1 (12.1)**: An estimator $\hat{\theta}$ is **unbiased** if $E\left(\hat{\theta} \mid \theta\right) = \theta, \forall \theta \in \Theta$. The **bias** is $bias_{\hat{\theta}} = E\left(\hat{\theta} \mid \theta\right) - \theta$.
- **Comments**:
  - The point is to verify the equality $\forall \theta \in \Theta$ (see example 2)
  - The bias depends on the estimator being used but also on the particular value of θ.
  - An estimator with a positive bias tends to overestimate the parameter.
- **Example 1:** Prove that the sample mean is an unbiased estimator for the population mean, μ (assume that the population mean exists).

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu$$

- **Example 2:** Consider a Bernoulli population with mean θ and $T_2 = 0.3$ as an estimator for θ. As it is obvious T_2 is a bad estimator since it does not take into account the sample values. For $\theta = 0.3$, $E(T_2) = \theta$ but T_2 is a biased estimator since the equality $E(T_2) = \theta$ is not true $\forall \theta \in \Theta$.

- **Example 10.4 (12.4)**: A population has an exponential distribution with mean θ. We want to estimate the population mean using a sample of size 3. Determine the bias of the sample mean and the sample median as estimators of the population mean.

Sample mean: $E(\bar{X} \mid \theta) = \theta$ No bias

Sample median: Let T be the sample median. $E\left(T \mid \theta\right) = 5\theta/6$ $bias = -\theta/6$

How to compute the bias?

What is the meaning of the bias?

Sol:

Computation: (next slide)

Meaning:

On average, the estimator (sample median) underestimates the population mean θ which is not a surprise. Remember that the median of the population is $\theta \ln 2 < \theta$ – the sample median is also a biased estimator for the population median (ln2 < 5/6), but now the bias is positive ($= \left((5/6) - \ln 2\right)\theta$ □

$$f_T(t) = \frac{3!}{1!1!1!}\left(1-e^{-t/\theta}\right)\left(e^{-t/\theta}\right)\theta^{-1}\, e^{-t/\theta} \qquad t>0$$

$$= 6\theta^{-1}\left(1-e^{-t/\theta}\right)e^{-2t/\theta} = 3(2/\theta)e^{-2t/\theta} - 2(3/\theta)e^{-3t/\theta} \qquad t>0$$

$$E(T\mid\theta) = \int_0^\infty t\, f_T(t)\, dt$$

$$= \int_0^\infty t\left(3(2/\theta)e^{-2t/\theta} - 2(3/\theta)e^{-3t/\theta}\right)dt = 3\int_0^\infty t\,(2/\theta)e^{-2t/\theta}\,dt - 2\int_0^\infty t\,(3/\theta)e^{-3t/\theta}\,dt$$

$$= 3\times\frac{\theta}{2} - 2\times\frac{\theta}{3} = \frac{9\theta-4\theta}{6} = \frac{5\theta}{6}$$

Then $bias_{\hat{\theta}} = -\theta/6$.

- **Definition 10.2 (12.2)**: An estimator $\hat{\theta}$ is **asymptotically unbiased** if $\lim_{n\to\infty} E\left(\hat{\theta} \mid \theta\right) = \theta, \forall \theta \in \Theta$.

- **Example 10.5 (12.5)** $X \sim U(0;\theta)$, sample $(X_1, X_2, \cdots, X_n)$ and $\hat{\theta} = \max X_i$.

$$f_X(x) = 1/\theta \qquad 0 < x < \theta$$

$$f_{\hat{\theta}}(y) = \frac{n!}{(n-1)!1!0!}(F_X(y))^{n-1}(1 - F_X(y))^0 f_X(y)$$

$$= n\left(y/\theta\right)^{n-1}(1/\theta) = n\theta^{-n}y^{n-1} \qquad 0 < y < \theta$$

$$E\left(\hat{\theta} \mid \theta\right) = \int_0^\theta y\, f_{\hat{\theta}}(y)\, dy = \int_0^\theta y\, n\, \theta^{-n} y^{n-1}\, dy = \theta^{-n}\int_0^\theta n\, y^n\, dy$$

$$= \frac{n}{n+1}\theta \qquad \text{The estimator is biased}$$

However, it is asymptotically unbiased as

$$\lim_{n\to\infty} E\left(\hat{\theta} \mid \theta\right) = \lim_{n\to\infty} \frac{n}{n+1}\theta = \theta$$

- **How to compare 2 unbiased estimators**?
  - Let T and T' be 2 unbiased estimators for the parameter θ. We will say that T is better than T' if $\operatorname{var}(T \mid \theta) \le \operatorname{var}(T' \mid \theta)$, $\forall \theta \in \Theta$ (the inequality has to be strict for, at least, one value of θ).
  - Example: $X \sim Po(\theta)$ and let us consider $T = \bar{X}$ and $T' = S^2$ as estimators of θ.

$$E(T \mid \theta) = E(\bar{X} \mid \theta) = \theta \qquad E(T' \mid \theta) = E(S^2 \mid \theta) = \operatorname{var}(X \mid \theta) = \theta$$

$$\operatorname{var}(T \mid \theta) = \operatorname{var}(\bar{X} \mid \theta) = \sigma^2 / n = \theta / n$$

$$\operatorname{var}(T' \mid \theta) = \operatorname{var}(S^2 \mid \theta) = \frac{1}{n}\left(\theta + 3\theta^2 - \frac{n-3}{n-1}\theta^2\right) = \frac{\theta}{n} + \frac{1}{n}\left(\frac{3n-3-n+3}{n-1}\right)\theta^2 = \frac{\theta}{n} + \frac{2}{n-1}\theta^2 > \operatorname{var}(T \mid \theta)$$

- **Definition** (CB): An estimator T for $\tau(\theta)$ is **a best unbiased estimator** of $\tau(\theta)$ if it satisfies $E(T \mid \theta) = \tau(\theta)$ for all θ and, for any other estimator W with $E(W \mid \theta) = \tau(\theta)$, we have $\operatorname{var}(T \mid \theta) \le \operatorname{var}(W \mid \theta)$ for all θ. T is also called a uniform minimum variance unbiased estimator (**UMVUE**) of $\tau(\theta)$.

- **Cramér-Rao Inequality applied to unbiased estimators**

Let $(X_1, X_2, \cdots, X_n)$ be a random sample from a population with probability density function $f_X(x \mid \theta)$ and let $T = T(X_1, X_2, \cdots, X_n)$ be an unbiased estimator of $\tau(\theta)$ satisfying

$\frac{d}{d\theta} E(T \mid \theta) = \int_{D_{\mathbf{x}}} \frac{\partial}{\partial \theta}\big(T(\mathbf{x}) \times f(\mathbf{x} \mid \theta)\big) d\,\mathbf{x}$ and $\operatorname{var}(T \mid \theta) < \infty$.

Then $\operatorname{var}(T \mid \theta) \geq \dfrac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2}{n\Im(\theta)}$ where $\Im(\theta) = E\left(\frac{\partial}{\partial \theta} \ln f_X(X \mid \theta)\right)^2 = -E\left(\frac{\partial^2}{\partial \theta^2} \ln f_X(X \mid \theta)\right)$

- Comments
  - The original Cramér-Rao inequality is proved for any estimator and for non-independent sampling – see Casella and Berger, 2nd edition, page 335.
  - $\frac{d}{d\theta}E(T\,|\,\theta)=\int_{D_{\mathrm{x}}}\frac{\partial}{\partial\theta}\big(T(\mathbf{x})\times f(\mathbf{x}\,|\,\theta)\big)d\,\mathbf{x}$. We can swap the derivation (in order to θ) with the integration (in order to x). The set of support of X cannot depend on θ (the uniform density function doesn't fulfill this condition).
  - $\mathrm{var}(T\,|\,\theta)<\infty$: The variance of T should exist.
  - When we have an unbiased estimator of θ we can compare its variance with the lower bound given by the Cramér-Rao inequality. If they are equal we have an UMVUE. If not, nothing can be concluded (nothing is said about the possibility that an unbiased estimator with a variance equal to the lower bound exists).
  - $\Im(\theta)$ is called Fisher information for each observation (as the observations are *iid* the information contained in each of them is the same). $n\Im(\theta)$ is Fisher iformation for the sample.

- **Example** – Consider a Poisson population with mean θ and show that $\bar{X}$ in an UMVUE estimator for θ.

We have already shown that $\bar{X}$ is an unbiased estimator for θ and that $\operatorname{var}(\bar{X})=\theta/n$.

Let us now calculate the lower bound of the Cramér-Rao inequality.

$$f_X(x\,|\,\theta)=\frac{e^{-\theta}\,\theta^x}{x!} \qquad \ln f_X(x\,|\,\theta)=-\theta+x\ln\theta-\ln(x!)$$

$$\frac{\partial}{\partial\theta}\ln f_X(x\,|\,\theta)=-1+\frac{x}{\theta} \qquad \frac{\partial^2}{\partial\theta^2}\ln f_X(x\,|\,\theta)=-\frac{x}{\theta^2}$$

$$\Im(\theta)=-E\left(\frac{\partial^2}{\partial\theta^2}\ln f_X(X\,|\,\theta)\right)=-E\left(-\frac{X}{\theta^2}\right)=\frac{\theta}{\theta^2}=\frac{1}{\theta} \qquad \tau(\theta)=\theta \qquad \left(\frac{d}{d\theta}\tau(\theta)\right)^2=1$$

The lower bound is then $\dfrac{\left(\dfrac{d}{d\theta}\tau(\theta)\right)^2}{n\Im(\theta)}=\dfrac{1}{n/\theta}=\dfrac{\theta}{n}$

As $\bar{X}$ is an unbiased estimator of θ with a variance equal to the lower bound, we can conclude that $\bar{X}$ in an UMVUE estimator for θ.

Mean-squared error

- How to compare estimators that are not unbiased?
- **Definition 10.4** **(12.4)**: The mean-squared error of an estimator is $MSE_{\hat{\theta}}(\theta) = E\left((\hat{\theta}-\theta)^2 \mid \theta\right)$
- The mean-squared error can be rewrite as

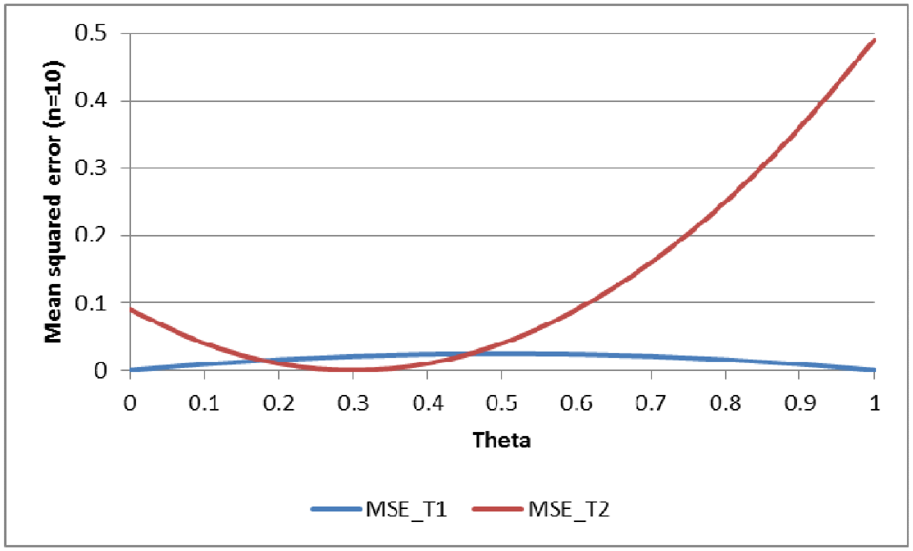
$$MSE_{\hat{\theta}}(\theta) = E\left((\hat{\theta}-\theta)^2 \mid \theta\right) = \mathrm{var}(\hat{\theta} \mid \theta) + \left(bias_{\hat{\theta}}(\theta)\right)^2$$

- Comments
  - ○ The mean-squared error is a function of the true value of the unknown parameter, θ: some estimator can perform well for some values of θ and poorly for other values of θ.
  - ○ Using the MSE with an unbiased estimator of θ is the same as using its variance.

- **Example**: Let us consider a Bernoulli population with parameter θ and two estimators for θ obtained using a sample of size n: $T_1 = \overline{X}$ and $T_2 = 0.3$. Compare these estimators using their MSE.

$$MSE_{T_1}(\theta) = E\left((T_1-\theta)^2 \mid \theta\right) = E\left((\overline{X}-\theta)^2 \mid \theta\right) = \mathrm{var}(\overline{X} \mid \theta) + \left(E(\overline{X} \mid \theta)-\theta\right)^2 = \frac{\theta(1-\theta)}{n} + 0 = \frac{\theta(1-\theta)}{n}$$

$$MSE_{T_2}(\theta) = E\left((T_2-\theta)^2 \mid \theta\right) = E\left((0.3-\theta)^2 \mid \theta\right) = (0.3-\theta)^2$$



Although T_2 is an inadequate estimator of θ (the estimator does not take into account the collected sample) we see that $MSE_{T_1}(\theta)$ is less than $MSE_{T_2}(\theta)$ for some values of θ

- It is convenient to use a qualification criterion before using the MSE and only compare estimator that fulfill such criterion.

Consistency

- **Definition 10.3 (12.3) –** An estimator is consistent (often called, in this context, weakly consistent) if, for all $\delta > 0$ and any θ, $\lim_{n\to\infty} \Pr(\left|\hat{\theta}_n - \theta\right| > \delta) = 0$.
- **Comments:**
    - A sufficient although not necessary condition for weak consistency is that $\lim_{n\to\infty} E(\hat{\theta}_n \mid \theta) = \theta$ and $\lim_{n\to\infty} \mathrm{var}(\hat{\theta}_n \mid \theta) = 0$. Such statement can be proved using Markov inequality ($\Pr(\left|X\right| \geq a) \leq E(\left|X\right|)/a$)[1].
    - Consistency is a property of the sequence of estimators, $\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_n, \cdots$, and not of the estimator itself.
    - The idea behind consistency is that the estimator will work well for large samples.

[1] $\Pr\left(\left|\hat{\theta}_n - \theta\right| > \delta\right) = \Pr\left(\left(\hat{\theta}_n - \theta\right)^2 > \delta^2\right) \leq \Pr\left(\left(\hat{\theta}_n - \theta\right)^2 \geq \delta^2\right) \leq \dfrac{E\left(\hat{\theta}_n - \theta\right)^2}{\delta^2} = \dfrac{\mathrm{var}(\hat{\theta})}{\delta^2} + \dfrac{\left(E(\hat{\theta}) - \theta\right)^2}{\delta^2}$

- **Example 10.6 (12.6) –** Prove that, if the variance of a random variable is finite, the sample mean is a consistent estimator of the population mean.

$$E(\overline{X}) = \mu$$

$$\operatorname{var}(\overline{X}) = \sigma^2 / n$$

Then

$$\lim_{n\to\infty} E(\overline{X}) = \lim_{n\to\infty} \mu = \mu$$

$$\lim_{n\to\infty} \operatorname{var}(\overline{X}) = \lim_{n\to\infty} \sigma^2 / n = 0$$

INTERVAL ESTIMATION

- Unlike **point estimation, interval estimation** leads to a set of values.
- The idea is to associate a level of confidence to such intervals.
- **Definition 10.6 (12.6)** – A $100(1-\alpha)\%$ confidence interval for a parameter θ is a pair of random values, *L* and *U*, computed from a random sample such that $\Pr(L \le \theta \le U) \ge 1-\alpha$ for all θ.
- **Comments:**
  - ○ The definition does not uniquely define an interval;
  - ○ When we replace the random variables by their observed values, nothing is said about whether or not the interval encloses θ;
  - ○ The level of confidence is a property of the process and not a property of the particular values obtained;
  - ○ Note that the inequality concerns discrete random populations (more theoretical).

- How to construct a confidence interval?
  - Not an easy question when considering a general case
  - Usually we follow the pivotal method
- **Pivotal quantity –** A random variable $Q(X_1, X_2, \cdots, X_n, \theta)$ is a pivotal quantity if the distribution of $Q(X_1, X_2, \cdots, X_n, \theta)$ does not depend on θ.
- **Comments**: The function $Q(X_1, X_2, \cdots, X_n, \theta)$
  - depends only on the sample $(X_1, X_2, \cdots, X_n)$, on θ and, possibly, on some known values;
  - is completely known;
  - usually, is monotonic in θ.

- **Pivotal method** (we will assume that $Q(X_1, X_2, \cdots, X_n, \theta)$ follows a continuous distribution)
  - Step 1 – Find q_1 and q_2 such that $\Pr\left(q_1 \leq Q(X_1, X_2, \cdots, X_n, \theta) \leq q_2\right) = 1 - \alpha$.
  - Step 2 – From $q_1 \leq Q(X_1, X_2, \cdots, X_n, \theta) \leq q_2$ define L and U such that $q_1 \leq Q(X_1, X_2, \cdots, X_n, \theta) \leq q_2 \Leftrightarrow L \leq \theta \leq U$.

L and U define a confidence interval for θ. How to choose the pair q_1 and q_2?

Optimally q_1 and q_2 are chosen to minimize the length (or its expected value if such length is random) of the confidence interval. As this task is difficult to fulfill in most situations we can follow a practical approximation and choose q_1 and q_2 such that

$$\Pr\left(Q(X_1, X_2, \cdots, X_n, \theta) < q_1\right) = \Pr\left(Q(X_1, X_2, \cdots, X_n, \theta) > q_2\right) = \alpha / 2$$

- **Using R to calculate confidence intervals will be discussed later**

- Some well-known pivotal quantities:
  - For normal populations or when we have a large sample some pivotal quantities are well-known for usual situations;
  - For other situations we try to find and estimator $\hat{\theta}$ for θ with a known distribution (independent of θ). If the sample is large enough and the estimator well behaved we can use $\frac{\hat{\theta}-E(\hat{\theta})}{\sqrt{\mathrm{var}(\hat{\theta})}} \overset{\circ}{\sim} n(0;1)$. Note that, as this result is asymptotic, we can use an adequate approximation for $E(\hat{\theta})$ and $\mathrm{var}(\hat{\theta})$

A) Gaussian (normal) populations:

	Pivotal Quantity	**Confidence Interval**
Mean (known variance)	$Q(X_1,X_2,...,X_n,\mu)=Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$	$\left(\bar{X}-z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\,,\,\bar{X}+z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right)$
Mean (unknown variance)	$Q(X_1,X_2,...,X_n,\mu)=T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}$	$\left(\bar{X}-t_{\alpha/2}\dfrac{S}{\sqrt{n}}\,,\,\bar{X}+t_{\alpha/2}\dfrac{S}{\sqrt{n}}\right)$
Variance	$Q(X_1,X_2,...,X_n,\sigma^2)=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$	$\left(\dfrac{(n-1)\,S^2}{q_2}\,,\,\dfrac{(n-1)\,S^2}{q_1}\right)$
$z_{\alpha/2}:\ \Phi(z_{\alpha/2})=1-\alpha/2;\quad t_{\alpha/2}:\ \ P(T_{(n-1)}>t_{\alpha/2})=\alpha/2;\ q_1,q_2:\ P(Q_{(n-1)}<q_1)=P(Q_{(n-1)}>q_2)=\alpha/2$		

B) Large samples (Confidence interval for the mean):

	Pivotal Quantity	Confidence Interval (aprox)
Case 1	$Q(X_1, X_2, ..., X_n, \mu) = Z = \dfrac{\bar{X} - \mu}{\sqrt{\text{var}(\bar{X})}} \overset{\circ}{\sim} N(0,1)$	
Case 2	$Q(X_1, X_2, ..., X_n, \mu) = Z = \dfrac{\bar{X} - \mu}{\sqrt{\hat{\text{var}}(\bar{X})}} \overset{\circ}{\sim} N(0,1)$	$\left(\bar{X} - z_{\alpha/2}\sqrt{\hat{\text{var}}(\bar{X})}\ ,\ \bar{X} + z_{\alpha/2}\sqrt{\hat{\text{var}}(\bar{X})}\right)$
Case 3	$Q(X_1, X_2, ..., X_n, \mu) = Z = \dfrac{\bar{X} - \mu}{S/\sqrt{n}} \overset{\circ}{\sim} N(0,1)$	$\left(\bar{X} - z_{\alpha/2}\dfrac{S}{\sqrt{n}}\ ,\ \bar{X} + z_{\alpha/2}\dfrac{S}{\sqrt{n}}\right)$
$z_{\alpha/2} : \Phi(z_{\alpha/2}) = 1 - \alpha/2\,;$		

Bernoulli populations: $X \sim Ber(\theta)$

Usual approach – Use $\overset{\wedge}{\text{var}}(\bar{X}) = \frac{\bar{X}\,(1-\bar{X})}{n} \rightarrow \left(\bar{X} - z_{\alpha/2}\sqrt{\frac{\bar{X}\,(1-\bar{X})}{n}}\,;\bar{X} + z_{\alpha/2}\sqrt{\frac{\bar{X}\,(1-\bar{X})}{n}} \right)$

More accurate solution – Use $\text{var}(\bar{X}) = \frac{\text{var}(X)}{n} = \frac{\theta\,(1-\theta)}{n}$

$$\left(\frac{\left(2n\bar{X} + z_{\alpha/2}^{2}\right) - z_{\alpha/2}\sqrt{z_{\alpha/2}^{2} + 4n\,\bar{X}\,(1-\bar{X})}}{2\left(n + z_{\alpha/2}^{2}\right)}\,;\frac{\left(2n\bar{X} + z_{\alpha/2}^{2}\right) + z_{\alpha/2}\sqrt{z_{\alpha/2}^{2} + 4n\,\bar{X}\,(1-\bar{X})}}{2\left(n + z_{\alpha/2}^{2}\right)} \right)$$

Poisson Populations: $X \sim Po(\theta)$

Usual approach – Use $\overset{\wedge}{\text{var}}(\bar{X}) = \frac{\bar{X}}{n} \rightarrow \left(\bar{X} - z_{\alpha/2}\sqrt{\frac{\bar{X}}{n}}\,;\bar{X} + z_{\alpha/2}\sqrt{\frac{\bar{X}}{n}} \right)$

More accurate solution – Use $\text{var}(\bar{X}) = \frac{\text{var}(X)}{n} = \frac{\theta}{n} \rightarrow$

$$\left(\frac{\left(2n\bar{X} + z_{\alpha/2}^{2}\right) - z_{\alpha/2}\sqrt{z_{\alpha/2}^{2} + 4n\,\bar{X}}}{2n}\,;\frac{\left(2n\bar{X} + z_{\alpha/2}^{2}\right) + z_{\alpha/2}\sqrt{z_{\alpha/2}^{2} + 4n\,\bar{X}}}{2n} \right)$$

TEST OF HYPOTHESES

- Null, H0, and alternative, H1, hypotheses
- The two hypotheses are not treated symmetrically (Neyman-Pearson approach). We do not reject H0 unless there is strong statistical evidence against it.
- The result of a test is the rejection (or not) of the null hypothesis. What so ever the decision is, an error is always possible:
  - ○ Type I error: Rejection of the null when the null is true;
  - ○ Type II error: Not rejecting the null when the null is false.

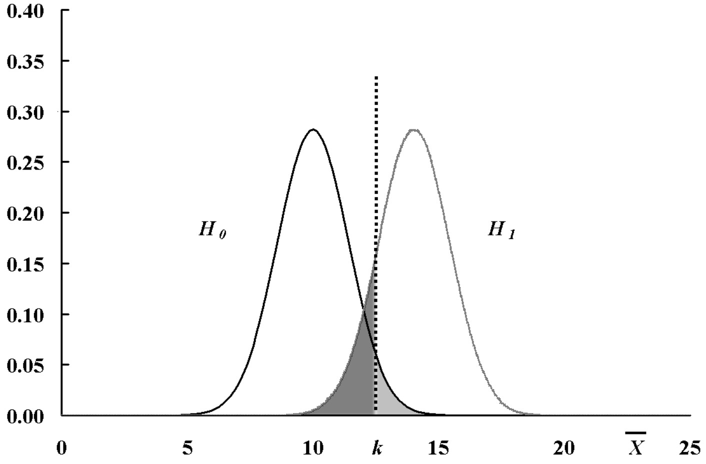
	H_0 **true**	H_0 **false**
Reject H_0	Type I error	Correct
Do not reject H_0	Correct	Type II error

- Using a simple example it can be shown that it is not possible to minimize both errors (unless we increase the sample size)

  $X \sim N(\mu, \sigma^2)$ with $\sigma^2 = 4$. The test is $H_0 : \mu = 10$ against $H_1 : \mu = 14$.

  a) Let us assume that our sample has only one observation and that the rejection region is given by $W = \{x : x > 12.5\}$. Determine the probabilities associated with type 1 and type 2 errors. ($\alpha \approx 0.1056$, $1 - \beta \approx 0.2266$).

  b) Show that decreasing the probability of a type 1 error implies increasing the probability of a type 2 error and vice-versa



- **Definition 10.7 (12.7)** – The significance level of a hypothesis test is the probability of making a Type I error given that the null is true. If it can be in more than one way, the level of significance is the maximum of such probabilities. The significance level is usually denoted by α.
- **Comments**:
  - This definition is conservative since we are considering the worst case;
  - Typically, the worst case is on the boundaries between H0 and H1;
  - Usual values for the level of significance are 1%, **5**% or 10%.
- Using the Neyman-Pearson approach one should control the probability associated with the Type I error, i.e. one must control the significance level of the test, and choose the test with a smaller probability of a Type II error, given the significance level.
- **Comments**:
  - The approach give more importance to the type I error;
  - Such a test is called a most powerful (uniformly most powerful test);

- **Definition 10.8 (12.8)** – A hypotheses test is **uniformly most powerful** (UMP) if no other test exists that has the same or lower significance level and, for a particular value within the alternative hypothesis, has a smaller probability of making a Type II error.
- **Test statistic** –The test statistic is a function of the sample observations with a known distribution under the null. The design of a test procedure looks at all the samples that might have been observed and not at the particular sample that was observed.
- **Rejection region** – The test specification is completed by defining a rejection region. If the observed value of the test statistic falls in the rejection region we will reject the null, otherwise we will not reject the null.
- **How develop a test of hypotheses?**
    - Define the hypotheses H0 and H1 and
    - Choose an adequate significance level
    - Obtain a test statistic and determine the rejection region
    - Calculate the observed value of test statistic and conclude

Open questions: How to obtain the test statistic and, given the test statistic, how to determine the rejection region?

- **Theoretical results**: Neyman-Pearson's lemma and Karlin-Rubin theorem
- **Empirical rule of thumb**: When testing a mean, a variance or a proportion (Bernoulli populations) using the "natural" test statistic the rejection region is on the side of the alternative.
- In most situations a UMP test does not exist, namely when the null hypothesis is an equality and the alternative is both sides ("=" against "≠").

- Some useful results - **Normal populations (1 sample)**:

Test about the mean, variance known $$Z=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\sim n(0;1)$$

Test about the mean, variance unknown $$T=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\sim t_{(n-1)} \qquad S^2=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n-1}$$

Test about the variance $$Q=\frac{(n-1)\,S^2}{\sigma^2}\sim\chi^2_{(n-1)} \qquad S^2=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n-1}$$

- Some useful results – **Larges samples** (populations with finite variance)- **1 sample**:

Test about the mean, variance unknown $T=\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}\overset{\circ}{\sim}n(0;1)$ $S^2=\dfrac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1}$

Bernoulli population* $Z=\dfrac{\overline{X}-p_0}{\sqrt{p_0\left(1-p_0\right)/n}}\overset{\circ}{\sim}n(0;1)$

Poisson population $Z=\dfrac{\overline{X}-\mu_0}{\sqrt{\mu_0/n}}\overset{\circ}{\sim}n(0;1)$

* As discussed for confidence intervals the way to approximate a Bernoulli population to a normal is not unique. This formula is the most common solution.

- **Examples 10.13 to 10.15 (12.13 to 12.15)** – Your company has been basing its premium on an assumption that the average claim is 1200. You want to raise the premiums, and a regulator has insisted that you provide evidence that the average now exceeds 1200. To provide such evidence, the following have been obtained:

  27 82 115 126 155 161 243 294 340 384

  457 680 855 877 974 1193 1340 1884 2558 15743

  a) What are the hypotheses for this problem (example 10.13)?

  b) Complete the test using the test statistic and rejection region that is promoted in most statistics books ($\alpha = 0.05$). Assume that the population has a normal distribution with standard deviation 3435 (example 10.14).

  c) Determine the probability of making a Type II error when the alternative hypothesis is true with $\mu = 2000$ (example 10.15).

Answers:

a) $H_0 : ? \quad H_1 : ?$ We assume a normal distribution.

b)

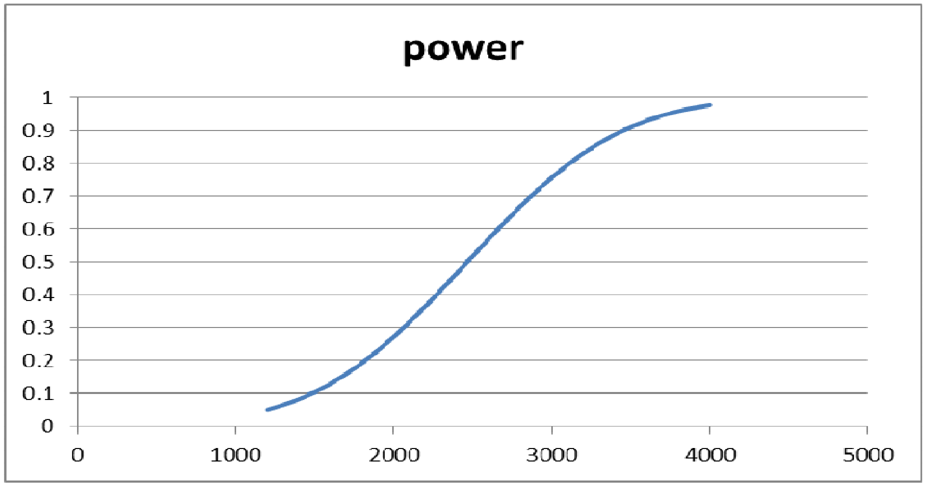
$\bar{x} = 1424.4 \qquad z = (1424.4 - 1200) \times \sqrt{20} / 3435 = 0.292154$

Test (N-P procedure): $\alpha = 0.05 \quad z_\alpha = 1.645$ (one side test)

Rejection region: $W = \{(x_1, x_2, \cdots, x_{20}) : z > 1.645\}$ or
$W = \{(x_1, x_2, \cdots, x_{20}) : \bar{x} > 1200 + 1.645 \times 3435 / \sqrt{20}\}$

conclusion: do not reject H0

c) $\Pr(\text{Accept } H_0 \mid \mu = 2000) = \Pr(\bar{X} \leq 2463.507 \mid \mu = 2000) = \Pr(Z \leq 0.603455) = 0.7269$


power
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1000
2000
3000
4000
5000

p-values

- Under the "classical" (Neyman-Pearson) approach a test will produce a decision on whether or not to reject H_0 for a predetermined value of α.
- Sometimes this procedure does not provide the recipient of the result with clear information on the strength of the evidence against H_0.
- A more informative approach is to calculate and quote the ***p-value*** of the observed test statistic. This is the significance level of the test statistic, i.e.
  - The probability, assuming H_0 is true, of observing a test statistic at least as "extreme" (inconsistent with H_0) as the value observed;
  - The significance level that originates a critical value equal to the observed value of the test statistic.

If α is greater than the p-value we reject H_0 and if α is smaller than the significance level we do not reject H_0

- **Definition 10.9 (12.9)** – For a hypothesis test, the p-value is the probability that the test statistic takes on a value that is less in agreement with the null hypothesis than the value obtained from the sample. Tests conducted at a significance level that is greater than the p-value will lead to a rejection of the null hypothesis, while tests conducted at a significance level that is smaller than the p-value will lead to a failure to reject the null hypothesis.
- **Comment** – The definition should refer less than or equal to. This point has no practical influence when the test statistic follows a continuous distribution as it is generally the case.
- **Example:** Resume previous example using p-value.

Test (p-value): p-value=$\Pr(Z \geq z) = \Pr(\overline{X} \geq \overline{x} \mid \mu = 1200) = 0.3851$ do not reject H0 for $\alpha = 0.05$

Using R

- As a calculator namely to compute probabilities and quantiles
- In R, confidence intervals are obtained “simulating” test of hypothesis
- Function t.test(….) for one sample or two samples tests for the means, ideally for normal populations but can be used with large samples

```
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

- Function var.test for the comparison of the variances of 2 independent normal population using 2 independent samples

```
var.test(x, y, ratio = 1,
         alternative = c("two.sided", "less", "greater"),
         conf.level = 0.95, ...)
```

- Function prop.test() for one sample or two samples tests for the proportions (means) of 2 Bernoulli populations – Larges samples only – should be used carefully as the options go beyond the usual formula presented above

```
prop.test(x, n, p = NULL,
          alternative = c("two.sided", "less", "greater"),
          conf.level = 0.95, correct = TRUE)
```

And, as usual, you can search for many other options using specific libraries

- **Example**: In a survey with 1000 answers, we got 510 YES (and 490 NO) to the question: “are you satisfied with your motor insurance company?”. Can we conclude that more 50% of the insured people in the population are satisfied?

  $H_0 : p \leq 0.5 \quad H_1 : p > 0.5$

  Usual procedure: $Z_{obs} = \dfrac{(510/1000) - 0.5}{\sqrt{0.5 \times (1-0.5)/1000}} = 0.6325 \qquad p - value = 0.2635$

  Using R (next slide)

```
> n=1000; s.x=510; xb=s.x/n
>
> z.obs=(xb-0.5)/sqrt(0.5*0.5/1000); z.obs
[1] 0.6324555
> p.value=pnorm(z.obs,0,1,lower.tail=F); p.value
[1] 0.2635446
>
> prop.test(510,1000,p=0.5,alternative="greater",correct=F)

        1-sample proportions test without continuity
correction

data:  510 out of 1000, null probability 0.5
X-squared = 0.4, df = 1, p-value = 0.2635
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval: Only when alternative is 2 sided
 0.4840059 1.0000000
sample estimates:
   p
0.51
```

More results for 2 independent samples $\left(X_1, X_2, \cdots, X_{n_1}\right)$ and $\left(Y_1, Y_2, \cdots, Y_{n_2}\right)$

- **Normal populations:** $X \sim n(\mu_X, \sigma_X^2)$**;** $Y \sim n(\mu_Y, \sigma_Y^2)$**;**
  - ○ σ_X^2 and σ_Y^2 known → $Z = \dfrac{\left(\bar{X} - \bar{Y}\right) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n_1} + \dfrac{\sigma_Y^2}{n_2}}} \sim N(0,1)$
  - ○ σ_X^2 and σ_Y^2 unknown but $\sigma_X^2 = \sigma_Y^2$ → $T = \dfrac{\left(\bar{X} - \bar{Y}\right) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{(n_1 - 1)\, S_X^2 + (n_2 - 1)\, S_Y^2}{n_1 + n_2 - 2}} \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$

- σ_X^2 and σ_Y^2 unknown (only an approximate distribution) → $T=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_X^2}{n_1}+\dfrac{S_Y^2}{n_2}}}\sim t(r)$

  r being the largest integer contained in $r^*=\dfrac{\left(\dfrac{s_X^2}{n_1}+\dfrac{s_Y^2}{n_2}\right)^2}{\dfrac{1}{n_1-1}\left(\dfrac{s_X^2}{n_1}\right)^2+\dfrac{1}{n_2-1}\left(\dfrac{s_Y^2}{n_2}\right)^2}$

- Ratio between 2 variances: $F=\dfrac{S_X^2}{S_Y^2}\dfrac{\sigma_Y^2}{\sigma_X^2}\sim F(n_1-1,n_2-1)$

Example 1 – To compare 2 risks we collect 2 independent samples of claims and get

Risk *A*: 8.0, 8.4, 8.0, 6.4, 8.6, 7.7, 7.7, 5.6, 5.6, 6.2;

Risk *B*: 5.6, 7.4, 7.3, 6.4, 7.5, 6.1, 6.6, 6.0, 5.5, 5.5;

```
> x=c(8.0, 8.4, 8.0, 6.4, 8.6, 7.7, 7.7, 5.6, 5.6, 6.2)
> y=c(5.6, 7.4, 7.3, 6.4, 7.5, 6.1, 6.6, 6.0, 5.5, 5.5)
>
> var.test(x,y,1,alternative="two.sided")

        F test to compare two variances
data:  x and y
F = 2.1433, num df = 9, denom df = 9, p-value = 0.2715
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5323637 8.6288860
sample estimates:
ratio of variances
          2.143293
```

```
> t.test(x,y,alternative="two.sided",0,var.equal=T)
        Two Sample t-test
data:  x and y
t = 1.882, df = 18, p-value = 0.07611
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.09655188  1.75655188
sample estimates:
mean of x mean of y
     7.22      6.39
> t.test(x,y,alternative="two.sided",0,var.equal=T)
        Two Sample t-test
data:  x and y
t = 1.882, df = 18, p-value = 0.07611
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.09655188  1.75655188
sample estimates:
mean of x mean of y
     7.22      6.39
```

- Correlation coefficient (Fisher transformation) →

$$Z=\sqrt{n-3}\left(\frac{1}{2}\ln\left(\frac{1+R}{1-R}\right)-\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)\right)\overset{a}{\sim}n(0;1)$$

- Correlation coefficient: When $\rho=0$, $T=\dfrac{R\sqrt{n-2}}{\sqrt{1-R^2}}\sim t(n-2)$

```
cor.test(x, y,
         alternative = c("two.sided", "less", "greater"),
         method = c("pearson", "kendall", "spearman"),
         exact = NULL, conf.level = 0.95, continuity = FALSE,...)
```

The null is always $\rho=0$

Example 2 – The vice president of marketing for a large firm is concerned about the effectiveness o advertising in generating sales of the firm’s product. To investigate the relationship between advertising and sales, data were gathered from a random sample of 20 sales districts on the two variables (file advertising.csv) in an adequate monetary unit. What can you conclude?

```
> direct="G:/Risk Models 2018 global/datasets/"; file="advertising.csv"
> dta=read.csv(paste(direct,file,sep=""),header=T,sep=";")
> attach(dta)
> cor.test(Sales,AdvExpenditures)
        Pearson's product-moment correlation
data:  Sales and AdvExpenditures
t = 10.7023, df = 18, p-value = 3.111e-09
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8275196 0.9722002
sample estimates:
      cor
0.9296189
> cor.test(Sales,AdvExpenditures,alternative="greater")
        Pearson's product-moment correlation
data:  Sales and AdvExpenditures
t = 10.7023, df = 18, p-value = 1.556e-09
alternative hypothesis: true correlation is greater than 0
95 percent confidence interval:
 0.8501354 1.0000000
sample estimates:
      cor
0.9296189
```

Paired samples

What is a paired sample?

The main idea of a paired sample is to observe (measure), each subject or entity twice, resulting in *pairs* of observations, ideally one before and one after a given event. The point is to relate the event with a possible difference between the populations mean. Common applications of the paired sample *t*-test include case-control studies or repeated-measures designs. Suppose you are interested in evaluating the effectiveness of a company training program. One approach you might consider would be to measure the performance of a sample of employees before and after completing the program, and analyze the differences using a paired sample *t*-test.

Formally, the sample is now composed by independent pairs of observations. It should be noted that although the pairs of observations are independent of each other nothing is said about the independence between the elements of the same pair, X and Y since there is generally no independence.

Usually the point is to test $H_0 : \mu_X = \mu_Y$ against a one-side or a two-tails alternative.

Define $Z_i = X_i - Y_i \sim N(\mu_X - \mu_Y ; \sigma^2_{Z_i}),\ i = 1,2,\ldots,n$, with $\sigma^2_{Z_i} = \sigma^2_X + \sigma^2_Y - 2\sigma_{XY},\ i = 1,2,\ldots,n$

As σ^2_Z is unknown we use $T = \dfrac{\bar{Z} - \mu_0}{S_Z / \sqrt{n}} \sim t_{(n-1)}$ where $\mu_0 = \mu_X - \mu_Y$ under the null, $\bar{Z} = \bar{X} - \bar{Y}$

and $S^2_z = \dfrac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$.

Example – Time needed to complete a task was measured before and after a course intended to improve the performance of the employees for a sample of 20 employees randomly chosen among those who attended the course (file time.csv). Assuming that the variables are normally distributed can we conclude that the course has been a success?

```
> rm(list=ls(all=TRUE))
> direct="G:/Risk Models 2018 global/datasets/"
> file="time.csv"
> dta=read.csv(paste(direct,file,sep=""),header=T,sep=";")
> attach(dta)
> diff=After-Before
> t.test(diff,alternative="less")


        One Sample t-test

data:  diff
t = -1.3087, df = 19, p-value = 0.1031
alternative hypothesis: true mean is less than 0
95 percent confidence interval:
    -Inf 44.3717
sample estimates:
mean of x
   -138.1
```