# Oncological Propedeutics – 4th year General Medicine – Oral Questions

## 1) Natural history of cancer, cancer as an ancient disease

The earliest written record regarding cancer is from 3000 BC in the Egyptian Edwin Smith Papyrus and describes cancer of the breast. Cancer however has existed for all of human history. Hippocrates (ca. 460 BC – ca. 370 BC) described several kinds of cancer, referring to them with the Greek word carcinos (crab or crayfish). This name comes from the appearance of the cut surface of a solid malignant tumor, with "the veins stretched on all sides as the crab has its feet, whence it derives its name". Celsus (ca. 25 BC – 50 AD) translated carcinos into the Latin cancer, also meaning crab and recommended surgery as treatment. Galen (2nd century AD) disagreed with the use of surgery and recommended purgatives instead. These recommendations largely stood for 1000 years.

In the 15th, 16th and 17th centuries, it became more acceptable for doctors to dissect bodies to discover the cause of death. The German professor Wilhelm Fabry believed that breast cancer was caused by a milk clot in a mammary duct. The Dutch professor Francois de la Boe Sylvius, a follower of Descartes, believed that all disease was the outcome of chemical processes, and that acidic lymph fluid was the cause of cancer. His contemporary Nicolaes Tulp believed that cancer was a poison that slowly spreads, and concluded that it was contagious.

The physician John Hill described tobacco snuff as the cause of nose cancer in 1761. This was followed by the report in 1775 by British surgeon Percivall Pott that cancer of the scrotum was a common disease among chimney sweeps. With the widespread use of the microscope in the 18th century, it was discovered that the 'cancer poison' spread from the primary tumor through the lymph nodes to other sites ("metastasis"). This view of the disease was first formulated by the English surgeon Campbell De Morgan between 1871 and 1874.

## 2) Cancer occurrence in developed countries

In 2008 approximately 12.7 million cancers were diagnosed (excluding non-melanoma skin cancers and other non-invasive cancers) and 7.6 million people died of cancer worldwide. Cancers as a group account for approximately 13% of all deaths each year with the most common being: lung cancer (1.4 million deaths), stomach cancer (740,000 deaths), liver cancer (700,000 deaths), colorectal cancer (610,000 deaths), and breast cancer (460,000 deaths). This makes invasive cancer the leading cause of death in the developed world and the second leading cause of death in the developing world. Over half of cases occur in the developing world.

Global cancer rates have been increasing primarily due to an aging population and lifestyle changes in the developing world. The most significant risk factor for developing cancer is old age. Although it is possible for cancer to strike at any age, most people who are diagnosed with invasive cancer are over the age of 65. According to cancer researcher Robert A. Weinberg, "If we lived long enough, sooner or later we all would get cancer." Some of the association between aging and cancer is attributed to immunosenescence, errors accumulated in DNA over a lifetime, and age-related changes in the endocrine system.

Some slow-growing cancers are particularly common. Autopsy studies in Europe and Asia have shown that up to 36% of people have undiagnosed and apparently harmless thyroid cancer at the time of their deaths, and that 80% of men develop prostate cancer by age 80. As these cancers did not cause the person's death, identifying them would have represented overdiagnosis rather than useful medical care.

The three most common childhood cancers are leukemia (34%), brain tumors (23%), and lymphomas (12%). Rates of childhood cancer have increased by 0.6% per year between 1975 to 2002 in the United States and by 1.1% per year between 1978 and 1997 in Europe.

### 3) Environment and cancer occurrence

Cancer related to one's occupation is believed to represent between 2–20% of all cases. Every year, at least 200,000 people die worldwide from cancer related to their workplace. Most cancer deaths caused by occupational risk factors occur in the developed world. It is estimated that approximately 20,000 cancer deaths and 40,000 new cases of cancer each year in the U.S. are attributable to occupation. Millions of workers run the risk of developing cancers such as lung cancer and mesothelioma from inhaling asbestos fibers and tobacco smoke, or leukemia from exposure to benzene at their workplaces.
Up to 10% of invasive cancers are related to radiation exposure, including both ionizing radiation and non-ionizing UV radiation. Additionally, the vast majority of non-invasive cancers are non-melanoma skin cancers caused by non-ionizing ultraviolet radiation.

Sources of ionizing radiation include medical imaging, and radon gas. Radiation can cause cancer in most parts of the body, at any age, although radiation-induced solid tumors usually take 10–15 years, and can take up to 40 years, to become clinically manifest, and radiation-induced leukemias typically require 2–10 years to appear. Some people, such as those with nevoid basal cell carcinoma or retinoblastoma, are more susceptible than average to developing cancer from radiation exposure. Children and adolescents are twice as likely to develop radiation-induced leukemia as adults; radiation exposure before birth has ten times the effect. Ionizing radiation is not a particularly strong mutagen. Residential exposure to radon gas, for example, has similar cancer risks as passive smoking. Low-dose exposures, such as living near a nuclear power plant, are generally believed to have no or very little effect on cancer development. Radiation is a more potent source of cancer when it is combined with other carcinogens, such as radon gas exposure plus smoking tobacco.

### 4) Basic tumor characteristics: Malignant tumor (infiltrative growth)

Cancer, also called malignant neoplasm, is a broad group of various diseases, all involving unregulated cell growth. In cancer, cells divide and grow uncontrollably, forming malignant tumors, and invade nearby parts of the body. The cancer may also spread to more distant parts of the body through the lymphatic system or bloodstream. Not all tumors are cancerous. Benign tumors do not grow uncontrollably, do not invade neighboring tissues, and do not spread throughout the body. There are over 200 different known cancers that afflict humans.
Malignant cells have a phenotype that is fixed. Malignancy is evidently related to a dysfunction of mechanisms regulating normal cell function.
Malignant cells are morphologically altered, they have a vascular stroma and many times outgrow their blood supply, causing necrosis in their central mass. Microscopically, cells show nuclear atypia, hyperchromatism and increased mitotic activity. They spread locally and subsequently metastasise to distant locations, first to locoregional lymphnodes and then through the bloodstream to other organs.

### 5) Basic tumor characteristics: Benign tumor (expansive growth)

By definition, benign tumors do not penetrate (invade) adjacent tissue borders, nor do they spread (metastasise) to distant sites. They remain as localized overgrowths in the area in which they arise. As a rule, benign tumors are more differentiated than malignant ones – that is, they more closely resemble their tissue of origin.
In common usage, the characteristic "benign" refers to the overall biological behaviour of a tumor rather than to its morphologic characteristics. However, benign tumors in critical locations can be deadly. For example, a benign intracranial tumor (meningioma) can kill by excreting pressure on the brain. A minute benign ependymoma can block the circulation of CSF, resulting in lethal hydrocephalus.

**6) Molecular control of cell replication**

Regulation of the cell cycle involves processes crucial to the survival of a cell, including the detection and repair of genetic damage as well as the prevention of uncontrolled cell division. The molecular events that control the cell cycle are ordered and directional; that is, each process occurs in a sequential fashion and it is impossible to "reverse" the cycle.
Two key classes of regulatory molecules, cyclins and cyclin-dependent kinases (CDKs), determine a cell's progress through the cell cycle. Many of the relevant genes were first identified by studying yeast, especially Saccharomyces cerevisiae; genetic nomenclature in yeast dubs many of these genes cdc (for "cell division cycle") followed by an identifying number, e.g., cdc25 or cdc20.
Cyclins form the regulatory subunits and CDKs the catalytic subunits of an activated heterodimer; cyclins have no catalytic activity and CDKs are inactive in the absence of a partner cyclin. When activated by a bound cyclin, CDKs perform a phosphorylation that activates or inactivates target proteins to orchestrate coordinated entry into the next phase of the cell cycle. Different cyclin-CDK combinations determine the downstream proteins targeted. CDKs are constitutively expressed in cells whereas cyclins are synthesised at specific stages of the cell cycle, in response to various molecular signals.
Two families of genes, the cip/kip family (CDK interacting protein/Kinase inhibitory protein) and the INK4a/ARF (Inhibitor of Kinase 4/Alternative Reading Frame) prevent the progression of the cell cycle. Because these genes are instrumental in prevention of tumor formation, they are known as tumor suppressors.

**7) Cell replication and differentiation, the role of cytokines and receptors**

Mitosis is the process by which a eukaryotic cell separates the chromosomes in its cell nucleus into two identical sets in two nuclei. It is generally followed immediately by cytokinesis, which divides the nuclei, cytoplasm, organelles and cell membrane into two cells containing roughly equal shares of these cellular components. Mitosis and cytokinesis together define the M phase of the cell cycle - the division of the mother cell into two daughter cells, genetically identical to each other and to their parent cell. This accounts for approximately 10% of the cell cycle.
The process of mitosis is complex and highly regulated. The sequence of events is divided into phases, corresponding to the completion of one set of activities and the start of the next. These stages are prophase, prometaphase, metaphase, anaphase and telophase. During the process of mitosis the pairs of chromosomes condense and attach to fibers that pull the sister chromatids to opposite sides of the cell.
Regulation of the cell cycle involves processes crucial to the survival of a cell, including the detection and repair of genetic damage as well as the prevention of uncontrolled cell division. The molecular events that control the cell cycle are ordered and directional; that is, each process occurs in a sequential fashion and it is impossible to "reverse" the cycle.
Two key classes of regulatory molecules, cyclins and cyclin-dependent kinases (CDKs), determine a cell's progress through the cell cycle. Many of the relevant genes were first identified by studying yeast, especially Saccharomyces cerevisiae; genetic nomenclature in yeast dubs many of these genes cdc (for "cell division cycle") followed by an identifying number, e.g., cdc25 or cdc20.

**8) Exponential growth of primary malignant tumor**

The growth rate is defined as the rate of increase in volume (or the number of cells) in relation to time. Exponential growth occurs when the growth rate of the tumor is proportional to the tumor's current size. For example, one cell divides into two cells, two cells divide into four cells, four cells divide into eight cells, eight cells divide into sixteen cells etc.
Testicular carcinomas, pediatric tumors, and some mesenchymal tumor are examples of rapidly proliferating cell populations, for which the tumor volume doubling time (TVDT) can be counted in days. Cancers from the breast, prostate, and colon are frequently slow-growing, displaying a TVDT of months or years. Irrespective of their growth rates, most human tumors have been found: to start

from one single cell, to have a long subclinical period, to grow at constant rates for long periods of time, to start to metastasize often even before the primary is detected, and to have metastases that often grow at approximately the same rate as the primary tumor. The recognition of basic facts in tumor cell kinetics is essential in the evaluation of important present-day strategies in oncology.

**9) Biological factors influencing exponential tumor growth**

Some hormones play a role in the development of cancer by promoting cell proliferation. Insulin-like growth factors and their binding proteins play a key role in cancer cell proliferation, differentiation and apoptosis, suggesting possible involvement in carcinogenesis.
Hormones are important agents in sex-related cancers such as cancer of the breast, endometrium, prostate, ovary, and testis, and also of thyroid cancer and bone cancer. For example, the daughters of women who have breast cancer have significantly higher levels of estrogen and progesterone than the daughters of women without breast cancer. These higher hormone levels may explain why these women have higher risk of breast cancer, even in the absence of a BRCA1 gene. Similarly, men of African ancestry have significantly higher levels of testosterone than men of European ancestry, and have a correspondingly much higher level of prostate cancer.
Other factors are also relevant: obese people have higher levels of some hormones associated with cancer and a higher rate of those cancers. Women who take hormone replacement therapy have a higher risk of developing cancers associated with those hormones. On the other hand, people who exercise far more than average have lower levels of these hormones, and lower risk of cancer. Osteosarcoma may be promoted by growth hormones. Some treatments and prevention approaches leverage this cause by artificially reducing hormone levels, and thus discouraging hormone-sensitive cancers.

**10) "Immortality" of tumor cells**

Cancer is fundamentally a disease of failure of regulation of tissue growth. In order for a normal cell to transform into a cancer cell, genes, which regulate cell growth and differentiation must be altered.

The affected genes are divided into two categories. Oncogenes are genes which promote cell growth and reproduction. Tumor suppressor genes are genes which inhibit cell division and survival. Malignant transformation can occur through the formation of novel oncogenes, the inappropriate over-expression of normal oncogenes, or by the under-expression or disabling of tumor suppressor genes. Typically, changes in many genes are required to transform a normal cell into a cancer cell.

Genetic changes can occur at different levels and by different mechanisms. The gain or loss of an entire chromosome can occur through errors in mitosis. More common are mutations, which are changes in the nucleotide sequence of genomic DNA.

Large-scale mutations involve the deletion or gain of a portion of a chromosome. Genomic amplification occurs when a cell gains many copies (often 20 or more) of a small chromosomal locus, usually containing one or more oncogenes and adjacent genetic material. Translocation occurs when two separate chromosomal regions become abnormally fused, often at a characteristic location. A well-known example of this is the Philadelphia chromosome, or translocation of chromosomes 9 and 22, which occurs in chronic myelogenous leukemia, and results in production of the BCR-abl fusion protein, an oncogenic tyrosine kinase.

Small-scale mutations include point mutations, deletions, and insertions, which may occur in the promoter region of a gene and affect its expression, or may occur in the gene's coding sequence and alter the function or stability of its protein product. Disruption of a single gene may also result from integration of genomic material from a DNA virus or retrovirus, and resulting in the expression of viral oncogenes in the affected cell and its descendants.

**11) Tumor growth and immune response, tumor-associated antigens**

It used to be believed that malignant tumors elicit a chronic inflammatory response that is unrelated to necrosis or infection of the tumor. The inflammatory reaction is correlated with a better prognosis in some tumors, such as medullary carcinoma of the breast and seminoma, but in general no clear correlation exists. Although the infiltrate is composed principally of T cells and macrophages, suggesting a cell-mediated immune response, the antigen to which the cells respond has not yet been identified.
Studies in mice have shown than cells of a chemically or virally-induced tumor can cause cancer when transplanted into other mice. If those cells are then removed before the cancer metastasizes, the mouse is cured. However, when those same cells are reinjected into that mouse, the body "rejects" them and cancer does not develop, because of immunity aquired as a result of the first tumor transplant. This whole process proves an antigenic nature of cancer.
Early studies on melanoma showed that certain HLA-associated peptide antigens correspond to proteins that are present in small amounts in the adult but are abundant during development. Such tumor-associated oncodevelopmental antigens are shared in cancers in different patients and sometimes of varying histologic type. Examples include CEA (carcinoembryonic antigen) and AFP (alpha fetoprotein)

**12) Humoral regulation of cell division, endocrine, paracrine, autocrine regulation factors**

Cell replication is regulated by growth factors produced by genes. Once these genes are pathologically activated, the cell is abnormally forced to replicate, leading to malignant transformation.
Examples include: PDGF (platele-derived growth factor) from thrombocytes and TGFb (transforming growth factor beta) from macrophages.
Regulating factors can be:
→ endocrine: targets of hormones are in distant locations and are reached through the blood (steroids act in this manner)
→ paracrine: targets of hormones are in the vicinity of the cell that produces them (lymphocytes secrete cytokines this way)
→ autocrine: the cell that produces the hormone is also the target of the substance (growth factors and kinases act like this)

**13) The role of autocrine stimulation in the process of oncogenesis**

**14) Genome of tumor cells, chromosomal basis of malignancy**

Genetic changes can occur at different levels and by different mechanisms. The gain or loss of an entire chromosome can occur through errors in mitosis. More common are mutations, which are changes in the nucleotide sequence of genomic DNA.

Large-scale mutations involve the deletion or gain of a portion of a chromosome. Genomic amplification occurs when a cell gains many copies (often 20 or more) of a small chromosomal locus, usually containing one or more oncogenes and adjacent genetic material. Translocation occurs when two separate chromosomal regions become abnormally fused, often at a characteristic location. A well-known example of this is the Philadelphia chromosome, or translocation of chromosomes 9 and 22, which occurs in chronic myelogenous leukemia, and results in production of the BCR-abl fusion protein, an oncogenic tyrosine kinase.

Small-scale mutations include point mutations, deletions, and insertions, which may occur in the promoter region of a gene and affect its expression, or may occur in the gene's coding sequence and alter the function or stability of its protein product. Disruption of a single gene may also result from integration of genomic material from a DNA virus or retrovirus, and resulting in the expression of viral oncogenes in the affected cell and its descendants.

Replication of the enormous amount of data contained within the DNA of living cells will probabilistically result in some errors (mutations). Complex error correction and prevention is built into the process, and safeguards the cell against cancer. If significant error occurs, the damaged cell can "self-destruct" through programmed cell death, termed apoptosis. If the error control processes fail, then the mutations will survive and be passed along to daughter cells.

**15) Humoral regulation of cell replication, wound healing**

Two key classes of regulatory molecules, cyclins and cyclin-dependent kinases (CDKs), determine a cell's progress through the cell cycle. Cyclins form the regulatory subunits and CDKs the catalytic subunits of an activated heterodimer; cyclins have no catalytic activity and CDKs are inactive in the absence of a partner cyclin.
Upon receiving a pro-mitotic extracellular signal, $G_1$ cyclin-CDK complexes become active to prepare the cell for S phase, promoting the expression of transcription factors that in turn promote the expression of S cyclins and of enzymes required for DNA replication. The $G_1$ cyclin-CDK complexes also promote the degradation of molecules that function as S phase inhibitors by targeting them for ubiquitination. Once a protein has been ubiquitinated, it is targeted for proteolytic degradation by the proteasome.

Active S cyclin-CDK complexes phosphorylate proteins that make up the pre-replication complexes assembled during $G_1$ phase on DNA replication origins. The phosphorylation serves two purposes: to activate each already-assembled pre-replication complex, and to prevent new complexes from forming. This ensures that every portion of the cell's genome will be replicated once and only once. The reason for prevention of gaps in replication is fairly clear, because daughter cells that are missing all or part of crucial genes will die. However, for reasons related to gene copy number effects, possession of extra copies of certain genes is also deleterious to the daughter cells.

Mitotic cyclin-CDK complexes, which are synthesized but inactivated during S and $G_2$ phases, promote the initiation of mitosis by stimulating downstream proteins involved in chromosome condensation and mitotic spindle assembly. A critical complex activated during this process is a ubiquitin ligase known as the anaphase-promoting complex (APC), which promotes degradation of structural proteins associated with the chromosomal kinetochore. APC also targets the mitotic cyclins for degradation, ensuring that telophase and cytokinesis can proceed.

**16) Wound healing and cancer: similarities and diversities**

**17) Cancer as a genetic disease**

A cancer syndrome is a genetic disorder in which genetic mutations in one or more genes predispose the affected individuals to the development of cancers and may also cause the early onset of these cancers. Cancer syndromes often show not only a high lifetime risk of developing cancer, but also the development of multiple independent primary tumors. Many of these syndromes are caused by mutations in tumor suppressor genes. Other genes that may be affected are oncogenes and genes involved in angiogenesis. Common examples of inherited cancer syndromes are hereditary breast-ovarian cancer syndrome and hereditary non-polyposis colon cancer (Lynch syndrome).
Hereditary breast-ovarian cancer syndrome is an autosomal dominant genetic disorder caused by genetic mutations of the BRCA1 and BRCA2 genes. In women this disorder primarily increases the risk of breast and ovarian cancer, but also increases the risk of fallopian tube carcinoma and papillary serous carcinoma of the peritoneum. In men the risk of prostate cancer is increased. Other cancers that are inconsistently linked to this syndrome are pancreatic cancer, male breast cancer, colorectal cancer and cancers of the uterus and cervix. Genetic mutations account for approximately 7% and 14% of breast and ovarian cancer, respectively, and BRCA1 and BRCA2 account for 80% of these cases. BRCA1 and BRCA2 are both tumor suppressor genes.

Hereditary non-polyposis colon cancer, also known as Lynch syndrome, is an autosomal dominant cancer syndrome that increases the risk of colorectal cancer. It is caused by genetic mutations in DNA mismatch repair (MMR) genes, notably MLH1, MSH2, MSH6 and PMS2. In addition to colorectal cancer many other cancers are increased in frequency. These include: endometrial cancer, stomach cancer, ovarian cancer, cancers of the small bowel and pancreatic cancer. MMR genes are involved in repairing DNA when the bases on each strand of DNA do not match. Defective MMR genes allow continuous insertion and deletion mutations in regions of DNA known as microsatellites. These short repetitive sequences of DNA become unstable, leading to a state of microsatellite instability (MSI). Mutated microsatellites are often found in genes involved in tumor initiation and progression, and MSI can enhance the survival of cells, leading to cancer.

**18) Genes involved in the process of oncogenesis**

Cancer is fundamentally a disease of failure of regulation of tissue growth. In order for a normal cell to transform into a cancer cell, the genes which regulate cell growth and differentiation must be altered.

The affected genes are divided into two broad categories. Oncogenes are genes which promote cell growth and reproduction. Tumor suppressor genes are genes which inhibit cell division and survival. Malignant transformation can occur through the formation of novel oncogenes, the inappropriate over-expression of normal oncogenes, or by the under-expression or disabling of tumor suppressor genes. Typically, changes in many genes are required to transform a normal cell into a cancer cell.

Genetic changes can occur at different levels and by different mechanisms. The gain or loss of an entire chromosome can occur through errors in mitosis. More common are mutations, which are changes in the nucleotide sequence of genomic DNA.

Large-scale mutations involve the deletion or gain of a portion of a chromosome. Genomic amplification occurs when a cell gains many copies (often 20 or more) of a small chromosomal locus, usually containing one or more oncogenes and adjacent genetic material. Translocation occurs when two separate chromosomal regions become abnormally fused, often at a characteristic location. A well-known example of this is the Philadelphia chromosome, or translocation of chromosomes 9 and 22, which occurs in chronic myelogenous leukemia, and results in production of the BCR-abl fusion protein, an oncogenic tyrosine kinase.

Small-scale mutations include point mutations, deletions, and insertions, which may occur in the promoter region of a gene and affect its expression, or may occur in the gene's coding sequence and alter the function or stability of its protein product. Disruption of a single gene may also result from integration of genomic material from a DNA virus or retrovirus, and resulting in the expression of viral oncogenes in the affected cell and its descendants.

**19) Regulation mechanisms of cell replication: Cell transduction signaling pathway**

Regulation of the cell cycle involves processes crucial to the survival of a cell, including the detection and repair of genetic damage as well as the prevention of uncontrolled cell division. The molecular events that control the cell cycle are ordered and directional; that is, each process occurs in a sequential fashion and it is impossible to "reverse" the cycle.

Two key classes of regulatory molecules, cyclins and cyclin-dependent kinases (CDKs), determine a cell's progress through the cell cycle. Many of the genes encoding cyclins and CDKs are conserved among all eukaryotes, but in general more complex organisms have more elaborate cell cycle control systems that incorporate more individual components. Many of the relevant genes were first identified by studying yeast, especially Saccharomyces cerevisiae; genetic nomenclature in yeast dubs many of these genes cdc (for "cell division cycle") followed by an identifying number, e.g., cdc25 or cdc20.

Hereditary non-polyposis colon cancer, also known as Lynch syndrome, is an autosomal dominant cancer syndrome that increases the risk of colorectal cancer. It is caused by genetic mutations in DNA mismatch repair (MMR) genes, notably MLH1, MSH2, MSH6 and PMS2. In addition to colorectal cancer many other cancers are increased in frequency. These include: endometrial cancer, stomach cancer, ovarian cancer, cancers of the small bowel and pancreatic cancer. MMR genes are involved in repairing DNA when the bases on each strand of DNA do not match. Defective MMR genes allow continuous insertion and deletion mutations in regions of DNA known as microsatellites. These short repetitive sequences of DNA become unstable, leading to a state of microsatellite instability (MSI). Mutated microsatellites are often found in genes involved in tumor initiation and progression, and MSI can enhance the survival of cells, leading to cancer.

**18) Genes involved in the process of oncogenesis**

Cancer is fundamentally a disease of failure of regulation of tissue growth. In order for a normal cell to transform into a cancer cell, the genes which regulate cell growth and differentiation must be altered.

The affected genes are divided into two broad categories. Oncogenes are genes which promote cell growth and reproduction. Tumor suppressor genes are genes which inhibit cell division and survival. Malignant transformation can occur through the formation of novel oncogenes, the inappropriate over-expression of normal oncogenes, or by the under-expression or disabling of tumor suppressor genes. Typically, changes in many genes are required to transform a normal cell into a cancer cell.

Genetic changes can occur at different levels and by different mechanisms. The gain or loss of an entire chromosome can occur through errors in mitosis. More common are mutations, which are changes in the nucleotide sequence of genomic DNA.

Large-scale mutations involve the deletion or gain of a portion of a chromosome. Genomic amplification occurs when a cell gains many copies (often 20 or more) of a small chromosomal locus, usually containing one or more oncogenes and adjacent genetic material. Translocation occurs when two separate chromosomal regions become abnormally fused, often at a characteristic location. A well-known example of this is the Philadelphia chromosome, or translocation of chromosomes 9 and 22, which occurs in chronic myelogenous leukemia, and results in production of the BCR-abl fusion protein, an oncogenic tyrosine kinase.

Small-scale mutations include point mutations, deletions, and insertions, which may occur in the promoter region of a gene and affect its expression, or may occur in the gene's coding sequence and alter the function or stability of its protein product. Disruption of a single gene may also result from integration of genomic material from a DNA virus or retrovirus, and resulting in the expression of viral oncogenes in the affected cell and its descendants.

**19) Regulation mechanisms of cell replication: Cell transduction signaling pathway**

Regulation of the cell cycle involves processes crucial to the survival of a cell, including the detection and repair of genetic damage as well as the prevention of uncontrolled cell division. The molecular events that control the cell cycle are ordered and directional; that is, each process occurs in a sequential fashion and it is impossible to "reverse" the cycle.

Two key classes of regulatory molecules, cyclins and cyclin-dependent kinases (CDKs), determine a cell's progress through the cell cycle. Many of the genes encoding cyclins and CDKs are conserved among all eukaryotes, but in general more complex organisms have more elaborate cell cycle control systems that incorporate more individual components. Many of the relevant genes were first identified by studying yeast, especially Saccharomyces cerevisiae; genetic nomenclature in yeast dubs many of these genes cdc (for "cell division cycle") followed by an identifying number, e.g., cdc25 or cdc20.

Cyclins form the regulatory subunits and CDKs the catalytic subunits of an activated heterodimer; cyclins have no catalytic activity and CDKs are inactive in the absence of a partner cyclin. When activated by a bound cyclin, CDKs perform a common biochemical reaction called phosphorylation that activates or inactivates target proteins to orchestrate coordinated entry into the next phase of the cell cycle. Different cyclin-CDK combinations determine the downstream proteins targeted. CDKs are constitutively expressed in cells whereas cyclins are synthesised at specific stages of the cell cycle, in response to various molecular signals.

**20) Growth factors, receptors and protein kinases in the regulation of cell replication**

**21) Transfection of normal cells by tumor cell DNA, explanation**

A disregulation of the cell cycle components may lead to tumor formation. Some genes like the cell cycle inhibitors, RB, p53 etc., when they mutate, may cause the cell to multiply uncontrollably, forming a tumor. Although the duration of cell cycle in tumor cells is equal to or longer than that of normal cell cycle, the proportion of cells that are in active cell division (versus quiescent cells in $G_0$ phase) in tumors is much higher than that in normal tissue. Thus there is a net increase in cell number as the number of cells that die by apoptosis or senescence remains the same.
The cells which are actively undergoing cell cycle are targeted in cancer therapy as the DNA is relatively exposed during cell division and hence susceptible to damage by drugs or radiation. This fact is made use of in cancer treatment; by a process known as debulking, a significant mass of the tumor is removed which pushes a significant number of the remaining tumor cells from $G_0$ to $G_1$ phase (due to increased availability of nutrients, oxygen, growth factors etc.). Radiation or chemotherapy following the debulking procedure kills these cells which have newly entered the cell cycle.
The fastest cycling mammalian cells in culture, crypt cells in the intestinal epithelium, have a cycle time as short as 9 to 10 hours. Stem cells in resting mouse skin may have a cycle time of more than 200 hours. Most of this difference is due to the varying length of $G_1$, the most variable phase of the cycle. M and S do not vary much.
In general, cells are most radiosensitive in late M and $G_2$ phases and most resistant in late S.
For cells with a longer cell cycle time and a significantly long $G_1$ phase, there is a second peak of resistance late in $G_1$.

**22) Oncogenes, tumor suppressor genes, DNA repair genes, role in the carcinogenesis**

An oncogene is a gene that has the potential to cause cancer. In tumor cells, they are often mutated or expressed at high levels. Most normal cells undergo apoptosis. Activated oncogenes can cause those cells that ought to die to survive and proliferate instead. Most oncogenes require an additional step, such as mutations in another gene, or environmental factors, such as viral infection, to cause cancer. Many cancer drugs target the proteins encoded by oncogenes.
A tumor suppressor gene, or anti-oncogene, is a gene that protects a cell from one step on the path to cancer. When this gene is mutated to cause a loss or reduction in its function, the cell can progress to cancer, usually in combination with other genetic changes. Tumor-suppressor genes, or more precisely, the proteins for which they code, either have a dampening or repressive effect on the regulation of the cell cycle or promote apoptosis, and sometimes do both.
DNA repair is a collection of processes by which a cell identifies and corrects damage to the DNA molecules that encode its genome. Both normal metabolic activities and environmental factors such as UV light and radiation can cause DNA damage, resulting in as many as 1 million individual molecular lesions per cell per day. Many of these lesions cause structural damage to the DNA molecule and can alter or eliminate the cell's ability to transcribe the gene that the affected DNA encodes. Other lesions induce potentially harmful mutations in the cell's genome, which affect the survival of its daughter cells after it undergoes mitosis. As a consequence, the DNA repair process is constantly active as it responds to damage in the DNA structure. When normal repair processes fail, and when cellular apoptosis does not occur, irreparable DNA damage may occur, including double-strand breaks and DNA crosslinkages (interstrand crosslinks or ICLs). Because of inherent limitations in the DNA repair mechanisms, if humans lived long enough, they would all eventually

develop cancer. There are at least 34 inherited human DNA repair gene mutations that increase cancer risk. Many of these mutations cause DNA repair to be less effective than normal. In particular, Hereditary nonpolyposis colorectal cancer is strongly associated with specific mutations in the DNA mismatch repair pathway. *BRCA1* and *BRCA2*, two famous genes whose mutations confer a hugely increased risk of breast cancer on carriers, are both associated with a large number of DNA repair pathways, especially NHEJ and homologous recombination.

**23) Carcinogenesis and mutagenesis, carcinogenic agents (biological, physical, chemical)**

A carcinogen is any substance, radionuclide, or radiation that is an agent directly involved in causing cancer. This may be due to the ability to damage the genome or to the disruption of cellular metabolic processes.
A mutagen is a physical or chemical agent that changes the genetic material, usually DNA, of an organism and thus increases the frequency of mutations above the natural background level. As many mutations cause cancer, mutagens are therefore also likely to be carcinogens.
Biological agents are usually viruses. A virus that can cause cancer is called an oncovirus. These include HPV (cervical carcinoma), EBV (B-cell lymphoproliferative disease and nasopharyngeal carcinoma), HHV–8 (Kaposi's sarcoma and primary effusion lymphomas), HBV/HCV (hepatocellular carcinoma), and HTLV-1 (T-cell leukemias). Bacterial infection may also increase the risk of cancer, as seen in H.pylori-induced gastric carcinoma. Parasitic infections strongly associated with cancer include Schistosoma haematobium (squamous cell carcinoma of the bladder) and the liver flukes, Opisthorchis viverrini and Clonorchis sinensis (cholangiocarcinoma).
A prominent example of physical agents causing cancer is prolonged exposure to asbestos, naturally occurring mineral fibers which are a major cause of mesothelioma. Other substances in this category, including both naturally occurring and synthetic asbestos-like fibers such as wollastonite, attapulgite, glass wool, and rock wool, are believed to have similar effects.
Decades of research have demonstrated the link between tobacco use and cancer in the lung, larynx, head, neck, stomach, bladder, kidney, esophagus and pancreas. Tobacco smoke contains over fifty known carcinogens, including nitrosamines and polycyclic aromatic hydrocarbons. Tobacco is responsible for about 1/3 of all cancer deaths in the developed world, and about 1/5 worldwide. Lung cancer death rates in the United States have mirrored smoking patterns, with increases in smoking followed by dramatic increases in lung cancer death rates and, more recently, decreases in smoking rates since the 1950s followed by decreases in lung cancer death rates in men since 1990.

**24) Environmental factors related to cancer occurrence**

Up to 10% of invasive cancers are related to radiation exposure, including both ionizing radiation and non-ionizing ultraviolet radiation. Additionally, the vast majority of non-invasive cancers are non-melanoma skin cancers caused by non-ionizing ultraviolet radiation.
Sources of ionizing radiation include medical imaging, and radon gas. Radiation can cause cancer in most parts of the body, in all animals, and at any age, although radiation-induced solid tumors usually take 10–15 years, and can take up to 40 years, to become clinically manifest, and radiation-induced leukemias typically require 2–10 years to appear. Some people, such as those with nevoid basal cell carcinoma syndrome or retinoblastoma, are more susceptible than average to developing cancer from radiation exposure. Children and adolescents are twice as likely to develop radiation-induced leukemia as adults; radiation exposure before birth has ten times the effect. Ionizing radiation is not a particularly strong mutagen. Residential exposure to radon gas, for example, has similar cancer risks as passive smoking. Low-dose exposures, such as living near a nuclear power plant, are generally believed to have no or very little effect on cancer development. Radiation is a more potent source of cancer when it is combined with other cancer-causing agents, such as radon gas exposure plus smoking tobacco.

**25) Chemical carcinogenic compounds, characteristics, division**

Decades of research have demonstrated the link between tobacco use and cancer in the lung, larynx, head, neck, stomach, bladder, kidney, esophagus and pancreas. Tobacco smoke contains over fifty known carcinogens, including nitrosamines and polycyclic aromatic hydrocarbons. Tobacco is responsible for about one in three of all cancer deaths in the developed world, and about one in five worldwide. Lung cancer death rates in the United States have mirrored smoking patterns, with increases in smoking followed by dramatic increases in lung cancer death rates and, more recently, decreases in smoking rates since the 1950s followed by decreases in lung cancer death rates in men since 1990.
Cancer related to one's occupation is believed to represent between 2–20% of all cases. Every year, at least 200,000 people die worldwide from cancer related to their workplace. Most cancer deaths caused by occupational risk factors occur in the developed world. Millions of workers run the risk of developing cancers such as lung cancer and mesothelioma from inhaling asbestos fibers and tobacco smoke, or leukemia from exposure to benzene at their workplaces.

**26) Radiation-induced oncogenesis, mutagenic radiation**

Up to 10% of invasive cancers are related to radiation exposure, including both ionizing radiation and non-ionizing ultraviolet radiation. Additionally, the vast majority of non-invasive cancers are non-melanoma skin cancers caused by non-ionizing ultraviolet radiation.
Sources of ionizing radiation include medical imaging, and radon gas. Radiation can cause cancer in most parts of the body, in all animals, and at any age, although radiation-induced solid tumors usually take 10–15 years, and can take up to 40 years, to become clinically manifest, and radiation-induced leukemias typically require 2–10 years to appear. Some people, such as those with nevoid basal cell carcinoma syndrome or retinoblastoma, are more susceptible than average to developing cancer from radiation exposure. Children and adolescents are twice as likely to develop radiation-induced leukemia as adults; radiation exposure before birth has ten times the effect. Ionizing radiation is not a particularly strong mutagen. Residential exposure to radon gas, for example, has similar cancer risks as passive smoking. Low-dose exposures, such as living near a nuclear power plant, are generally believed to have no or very little effect on cancer development. Radiation is a more potent source of cancer when it is combined with other cancer-causing agents, such as radon gas exposure plus smoking tobacco.

**27) Biological oncogenic factors, viruses, mycotoxins**

Biological agents are usually viruses. A virus that can cause cancer is called an oncovirus. These include HPV (cervical carcinoma), EBV (B-cell lymphoproliferative disease and nasopharyngeal carcinoma), HHV–8 (Kaposi's sarcoma and primary effusion lymphomas), HBV/HCV (hepatocellular carcinoma), and HTLV-1 (T-cell leukemias). Bacterial infection may also increase the risk of cancer, as seen in H.pylori-induced gastric carcinoma. Parasitic infections strongly associated with cancer include Schistosoma haematobium (squamous cell carcinoma of the bladder) and the liver flukes, Opisthorchis viverrini and Clonorchis sinensis (cholangiocarcinoma).
A mycotoxin is a toxic secondary metabolite produced by fungi, commonly known as molds. The term mycotoxin is usually reserved for the toxic chemical products produced by fungi that readily colonize crops. One mold species may produce many different mycotoxins, and the same mycotoxin may be produced by several species. Typical mycotoxins include alfatoxins (by Aspergillus), ochratoxin (by Penicillium and Aspergillus), citrinin (by Penicillium and Aspergillus) and ergot (by Claviceps).

**28) Oncogenesis versus apoptosis**

**29) Gene products involved in the control of tumor cell replication**

# 3. Evaluation of Measurements

## 3.1. Probability and Theoretical Distributions

### 3.1.1. Introduction

Without variability in attributes (parameters, signs, characteristics, quantities) or responses, events would be entirely predictable, and there would be no need for statistical methods. Because there is variability, we need statistical analysis to unravel what is going on. Ability to detect this effect, whether in observational or experimental studies, depends upon both the magnitude of the effect on average, and the variability of the effect.

Another essential concept in the application of statistical methods is that of **probability**. What is the probability that something happens? Given appropriate data, statistical methods help to answer this type of questions. It must be remembered, though, that statistical analysis rarely leads to a definite answer, so that we should indicate (or at least be aware of) a degree of uncertainty in our answer.

The **probability** of some specific outcome can be defined as the proportion of how many times the outcome would occur if we repeated the experiment or observation a large number of times. By definition, a probability lies between 0 and 1; something that cannot happen has the probability of 0, while something that is certain to happen has the probability of 1. In practice, we have to estimate most probabilities, as there is no way of knowing the true value.

There are two simple rules regarding probabilities that we need to consider at this stage:

- For a given event, for any two outcomes that might happen, the probability of either occurring is the sum of the individual probabilities. If the probability of event A is $p(A)$ and the probability of event B is $p(B)$, the probability that either A or B occurs is $p(A \text{ or } B) = p(A) + p(B)$. It follows that the probabilities of all possible outcomes must add up to 1, since one of these possibilities must occur, assuming that all outcomes are mutually exclusive.
- If we consider two or more different events, which are independent of each other, then to get the probability of a combination of specific outcomes for each of the events we must multiply the probabilities of these outcomes $p(A \text{ and } B) = p(A) \cdot p(B)$.

By **independent** we mean that if we know the outcome of one event this tells us nothing about the outcome of the over event. More formally, the probability of each possible outcome for the second event is the same regardless of the outcome for the first event. If the two events are not independent, the multiplicative property does not apply.

Nearly all statistical analysis is based on the principle that one acquires data on a **sample** of individuals, and uses the information to make inferences about all such

individuals. The set of all subjects (or whatever is being investigated) is called the **population** of interest. We take samples to study because it is rarely, if ever, possible to study the whole population. The relation between sample and population is subject to uncertainty, and we use ideas of probability to indicate this uncertainty. The idea of a theoretical probability distribution is important in this context.

Many statistical methods use the related idea of a **probability distribution**, which is specified mathematically. A probability distribution is used to calculate the theoretical probability of different value occurring, and is thus a theoretical equivalent of an empirical relative frequency distribution.

For example, if we know the mean and the standard deviation of some quantity $X$ of a population, we can calculate the probability of getting a measured value being greater than some chosen value $x_0$, if we assume that the distribution of the quantity $X$ in the population is the same as a particular probability distribution. The value $x_0$ is a **parameter** of the distribution as the population mean $\mu$ and the population standard deviation $\sigma$ (or more exactly the population variance (dispersion) $\sigma^2$) are. All probability distributions are described by one or more parameters.

Many statistical methods are based on the assumption that the observed data are a sample from a population with a distribution that has a known theoretical form. If this assumption is reasonable (we cannot establish if it is true), then the statistical methods of analysis are simple to use and wide-ranging. If the distributional assumption is not reasonable and we proceed as it were, then we may end up with misleading (and invalid) answers. When analysing data, we have a choice between methods that make distributional assumptions, called **parametric methods**, and those which make no assumption about distributions, called **distribution-free** or **non-parametric methods**. The importance of probability distributions in statistical analysis reflects the dominance of parametric methods.

The following text is focused on normal distribution and on $t$ distribution as they are needed for statistical calculations in practical tasks in this textbook. Next used theoretical distributions for statistical evaluation of data in medicine are for example, the $F$ distribution (Fisher distribution), the $\chi^2$ distribution (chi square distribution), the lognormal distribution, the binomial distribution, the Poisson distribution, the uniform distribution (for more detailed information see textbooks on statistics listed in References – Chapter 11).

### 3.1.2. The Gaussian (Normal) Distribution

The **Gaussian** or the **normal distribution** is by far the most important probability distribution in statistics. Figure 3.11 in Chapter 3.4.2 shows a histogram of a continuous variable – age – based on 85 observations. If there had been thousands of observations and age had been recorded more precisely, the age values could be divided into many tiny intervals, and a **frequency polygon** of the

data would appear more like a smooth curve. It is not difficult to imagine that a frequency polygon of some observed data is an approximation to some "underlying" smooth frequency distribution (Figure 3.1).

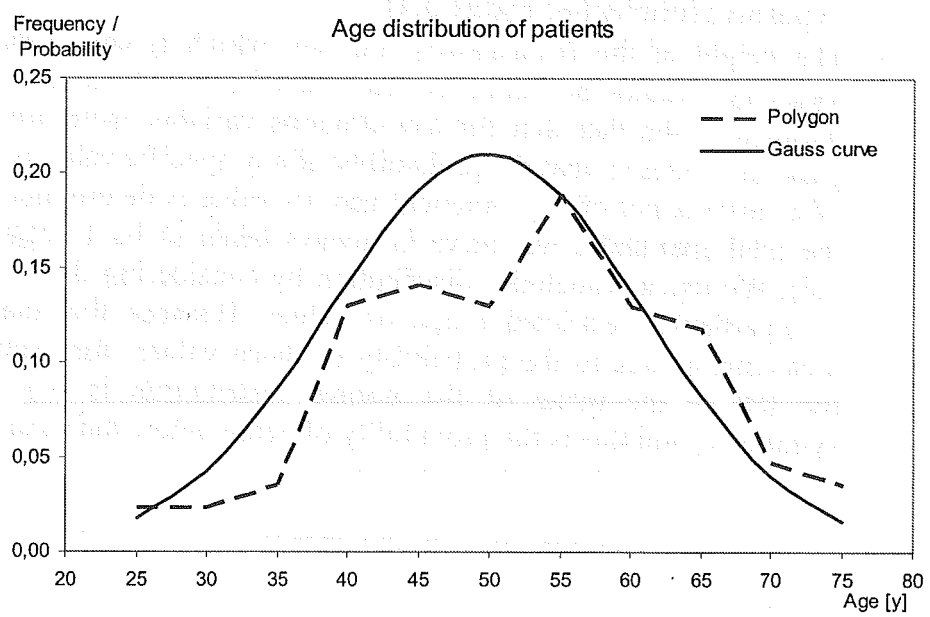


Figure 3.1: A comparison of a frequency polygon and the corresponding Gauss curve. Data as in Figures 3.11 and 3.12.

Frequency distributions for continuous data tend have a single peak - they are called **unimodal**. They may be fairly **symmetric** or **asymmetric** (Figure 3.2). The Gaussian distribution is a probability distribution which is unimodal and symmetric. Frequency distributions with two peaks are occasionally seen (Figure 3.3). These are called **bimodal** and are usually the result of mixing subgroups with different means, what should be avoided.

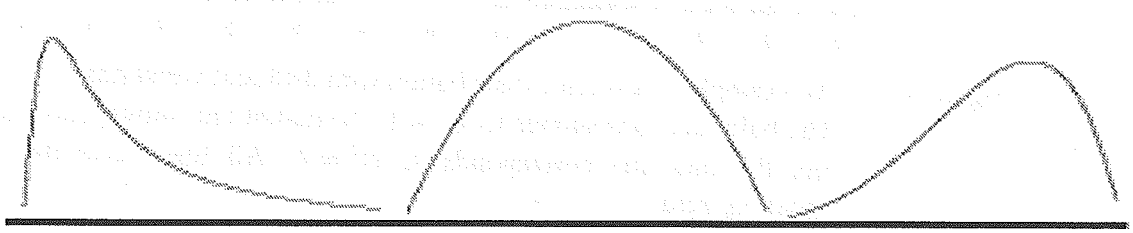

Figure 3.2: Examples of a left asymmetric (to the left), a symmetric (in the middle), and a right asymmetric (to the right) distribution.

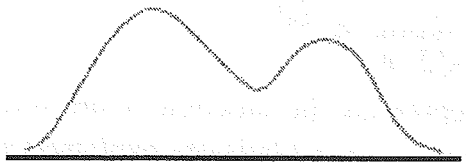

Figure 3.3: An example of a bimodal distribution.

Continuous probability distributions have some common features:

- They usually do not have upper limit and some have no lower limit either. In theory, the normal (Gaussian) distribution extends from negative infinity ($-\infty$) to positive infinity ($\infty$; Figure 3.4).
- The height of the **frequency curve**, which is called the **probability density**, cannot be taken as the probability of a particular value. This is because of the fact that for a continuous variable there are infinitely many possible values so that the probability of any specific value is zero. The height of a curve is not of any practical use; its value is determined by the fact that the total area under the curve is always taken to be 1 (Figures 3.4 through 3.6). We use a probability distribution by considering the area corresponding to a particular restricted range of values. Because the total area is 1, this area corresponds to the probability of those values, for example, the area to the left of the mean of the normal distribution is 0.5 (because of the symmetry) and this is the probability of being below the mean.

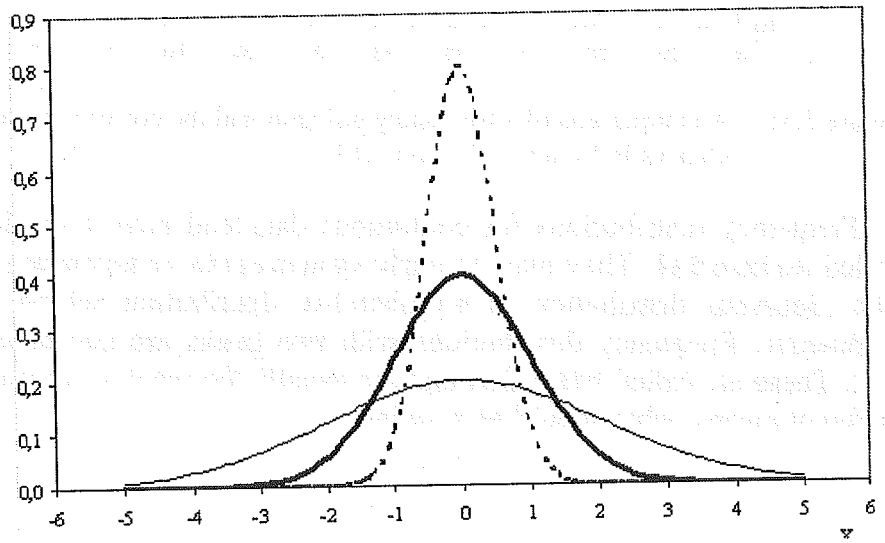


Figure 3.4: An example of a normal distribution with different variations. The bold line corresponds to $\sigma^2 = 1$, the dotted line corresponds to $\sigma^2 = 0.25$, and the fine line corresponds to $\sigma^2 = 4$. All lines have the same mean equalling zero.

The **probability density** for normal distribution is given by the equation:

$$\phi(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{\frac{(x-\mu)^2}{2 \cdot \sigma^2}}$$

where $\phi(x)$ represents the function of the variable $x$, $\mu$ is the mean value of the distribution, next used symbols are explained above. The probability distribution is symmetrical around its mean value, in this point it has the maximal value, in infinity

(both negative -∞ and positive ∞) it approaches zero (the $x$ axis). It is always a bell-shaped curve the width of which is influenced by the value of variance (the smaller the variance, the narrower the curve).

When using so-called **normalised variable** $z$ defined as the difference of the variable $x$ and the mean $\mu$ divided by the standard deviation:

$$z = \frac{x - \mu}{\sigma}$$

the probability density becomes more simple

$$\phi(z) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{\frac{z^2}{2}}$$

where $\mu = 0$ and $\sigma^2 = 1$. This distribution is called **standard normal (Gaussian) distribution** (Figure 3.5). Any normal distribution can be converted (transformed) into a standard normal distribution.

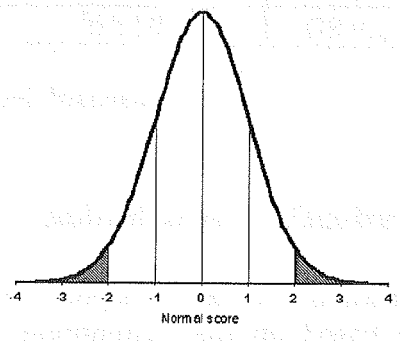


Figure 3.5: An example of a standard normal distribution with normal scores (standard normal deviates).
The lower tail (left to -2) and the upper tail (right to 2) are shaded.

Whatever values have the mean or the standard deviation, the normal distribution is related to them in the manner shown in Figure 3.5. Any position along the horizontal axis can be expressed as a distance of a multiple of standard deviations from the mean. This distance is known as **a standard normal deviate** or **normal score**.

When a set of measurements has a distribution that is similar to a normal distribution, we assume that in the population the distribution of the variable is actually normal, and we carry out calculations on this basis. Table 10.11 gives the lower tail areas of the standard normal distribution. The lower tail means the area under the curve from -∞ up to the value of interest (Figure 3.5). This area is equivalent to the probability of a value lower than the specified value. This idea can be also expressed as the **cumulative relative frequency distribution**. For example, the area below -2 is 0.023 and below 2 is 0.977, so that the area corresponding to the range from -2 to +2 is $0.977 - 0.023 = 0.954$. In other words, for data with an exactly normal distribution there is a probability of 0.954 of

being within two standard deviations of the mean. Repeating these calculations for other numbers of standard deviations, we get the values in the Table 3.1 (values that are more detailed are given in Table 10.11). In each case the probability of not being within the standard range equals one minus the probability of being within the standard range (Chapters 3.3.4 and 3.3.5).

Table 3.1: Often used standard deviation ranges and corresponding probabilities

| Range | Probability of being | |
|---|---|---|
| | within range | outside range |
| Mean ± 1 SD | 0.6827 | 0.3173 |
| Mean ± 1.96 SD | 0.9500 | 0.0500 |
| Mean ± 2 SD | 0.9545 | 0.0455 |
| Mean ± 2.58 SD | 0.9901 | 0.0099 |
| Mean ± 3 SD | 0.9973 | 0.0027 |
| Mean ± 3.9 SD | 0.9999 | 0.0001 |

SD – standard deviation

### 3.1.3. The Student's $t$ Distribution

Calculating confidence intervals (Chapter 3.3.4) and performing hypothesis tests (Chapter 3.3.5) are based on the assumption that the estimates of interest (for example, means) have a normal distribution. This is a reasonable assumption for large samples ($n > 100$; n is sample size), but not all samples are large. In the analysis of continuous data the calculation of means plays a prominent part. Consideration of distributions of the mean for smaller samples is needed.

The mean of a sample from a normal distribution with unknown variance has a distribution that is similar to, but not quite the same as, a normal distribution. This is called the **Student's $t$ distribution**. As the sample size increases, the sampling distribution of the mean becomes closer to the normal distribution. The $t$ distribution is used for estimation and hypothesis testing relating to the means of one or two samples. Although the normal distribution can be used for large samples, there is a little point of doing so, since for large samples the methods give virtually identical answers. It is simpler to use the same method regardless the sample size.

The $t$ distribution has one parameter, a quantity called the **degrees of freedom** $f$. In general, they are calculated as the sample size minus the number of estimated parameters. The degrees of freedom for the $t$ distribution relate to the estimated standard deviation, which is calculated as variation around the estimated mean. Hence, for a single sample of $n$ observations we have $f = n - 1$ degrees of freedom. (Notice: For the appropriate degrees of freedom look always at the used test. They differ for different applications.)

If comparing the $t$ distribution with different degrees of freedom with the normal distribution we can see that the higher are the degrees of freedom, the closer it is to the normal distribution (Figure 3.6). The difference is most marked in the tails of the distribution, which is usually the part we are interested in.

The **probability density** for the $t$ distribution is given by the equation:

$$\psi(x) = \frac{\left[\frac{f-1}{2}\right]!}{\left[\frac{f-2}{2}\right]! \cdot \sqrt{\pi \cdot f}} \cdot \left(1 + \frac{x^2}{f}\right)^{-\frac{f+1}{2}},$$

where $\psi(x)$ represents the function of the variable $x$, $f$ are the degrees of freedom. The probability distribution is symmetrical around its mean value (for $f > 0$ this mean equals zero), the probability distribution has the maximal value in the point of mean. In infinity (both negative $-\infty$ and positive $\infty$) it approaches zero (the $x$ axis). The geometrical form of the corresponding curve is influenced only by the degrees of freedom.

The critical values for one-tailed probability of the $t$ distribution are given in Table 10.12 and for two-tailed probability in Table 10.13.

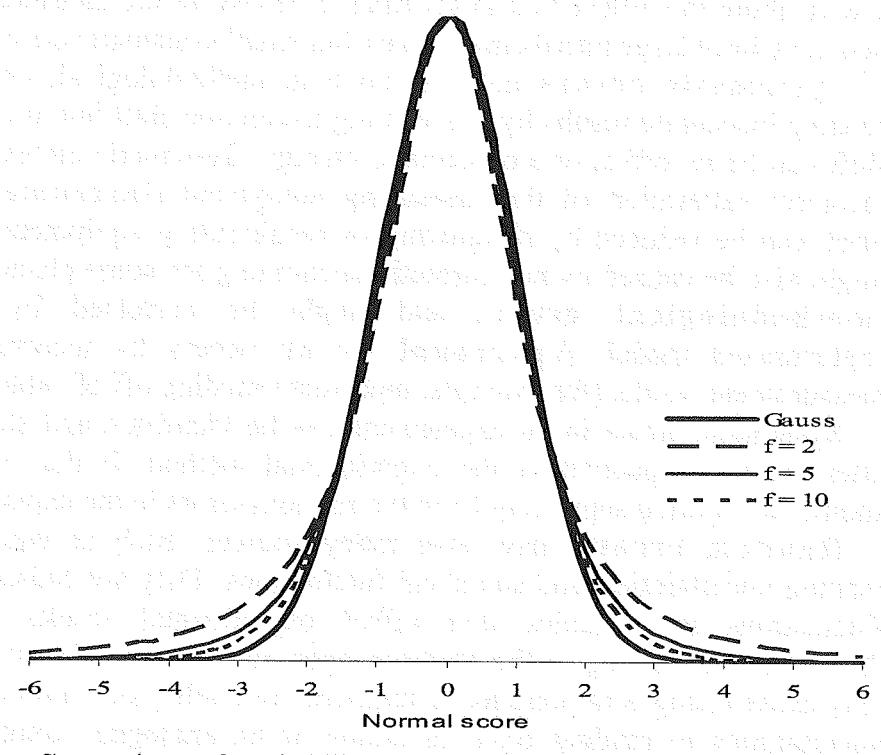


Figure 3.6: Comparison of probability densities for a Gauss curve (normal distribution) and for Student's t distributions with different degrees of freedom ($f = 2, f = 5, f = 10$).
All probability densities are adjusted to their maximal values to stress the different behaviour on the tails.

## 3.2. Numerical Evaluation of Measurements

### 3.2.1. Uncertainties or Errors of Measurement

In science, the terms **uncertainties** or **errors** refer to those uncertainties that are inherent in all measurements and can never be completely eliminated. A large part of a scientist's effort is devoted to understanding these uncertainties (**error analysis**) so that appropriate conclusions can be drawn from variable observations. A common complaint of students is that the error analysis is more tedious than the calculation of the numbers they are trying to measure. This is generally true. However, measurements can be quite meaningless without knowledge of their associated errors. In science and engineering, numbers without accompanying error estimates are suspect and possibly useless. **For every measurement, you must record the uncertainty in the measured quantity.**

Experimental errors may be divided into three classes: systematic errors, random errors, and gross errors. **Precision** refers to reproducibility of a measurement while **accuracy** is a measure of the closeness to the true value. **Precise** data exhibit small **random errors** and therefore have small fluctuations. However, there is a systematic shift of the data points away from the expected line. We attribute this effect to a **systematic error** in the measurements. **Accurate** data may have large **random errors** but small systematic errors.

**Systematic errors** are instrumental, methodological, or personal mistakes causing inaccurate results by introducing a common shift into measured values. This shift can be an offset or a percentage change. Systematic errors may be caused by incorrect calibration of the measuring equipment (**instrumental error**) and often can be reduced by readjusting or recalibrating equipment. Systematic errors might also be caused by not correctly accounting for some phenomena in the model (**methodological error**) and might be corrected by adopting a more sophisticated model. A **personal error** occurs by incorrect reading of the measurement results (for example, incorrect rounding off of values).

Systematic errors in an experiment can be identified and should be eliminated after careful inspection of the experimental method. If they are important, they should be reported separately from the random errors in the experimental results.

**Random errors** may have many sources, such as equipment limitations, reading uncertainties, and statistical fluctuations. They are caused by uncontrollable fluctuations in variables that affect experimental results (for example, air fluctuations occurring as the students open and close laboratory windows or doors may cause changes in pressure or temperature readings). Common examples are the uncertainties in reading scale divisions of an analogue voltmeter or a ruler and statistical fluctuations in counting rates from random processes (for example, dosimetry measurements). We can often reduce these uncertainties by repeated measurements (Chapter 3.3).

However, while it may be possible to reduce random errors by increasing the number of repeated measurement, they can never be completely eliminated.

**Gross errors** are undetected mistakes that cause a measurement to be very much farther from the mean measurement than other measurements ("outliers"). The gross error in the measurement may propagate through the calculation procedure and may indicate a bias in the resulting value. These errors are caused by experimenter carelessness or equipment failure. They have always to be avoided or, if possible, the measurement has to be repeated. It is possible to discard the outliers in some cases.

### 3.2.2. Error Estimates

When making only a few observations, the laws of probability are not applicable to the determination of uncertainties. The number of observations may be too small to justify using the standard deviation to estimate the uncertainty in a measurement (Chapter 3). However, it is usually possible to set limits on the range in which the true value is most likely to lie from an inspection of the measuring instruments. In reading most scales we should attempt to estimate some fraction of the smallest division, usually one half of the smallest scale division.

This type of uncertainty may be estimated as an **absolute uncertainty** $\Delta x$ and is given in the same units as is the measured quantity. This uncertainty is comparable to a standard deviation or a standard error of mean (Chapter 3.3).

If comparing the absolute uncertainty to the measured quantity, we will get a **relative error (fractional uncertainty)** $\Delta_{rel}x$ that equals

$$\Delta_{rel}x = \frac{\Delta x}{|x_0|}.$$

This error is a non-dimensional quantity (physical unit equals 1) and often is given as a device characteristic by its producer. $x_0$ represents the measured value (single value, mean, median, mode).

We may often meet a **percentage uncertainty** $\Delta_{\%}x$ defined as

$$\Delta_{\%}x = \frac{\Delta x}{|x_0|}\cdot 100\% \qquad \text{or} \qquad \Delta_{\%}x = \Delta_{rel}x \cdot 100\%.$$

The percentage uncertainty is comparable to a variation coefficient.

Uncertainties estimated in this way are referred to as **external errors**, i.e. estimating the uncertainties requires additional steps beyond making the measurements. For a complete uncertainty analysis, both internal errors (Chapter 3.3) and external errors should be calculated and checks should be made that the results are consistent.

The final uncertainty of a measurement task can be obtained in principle in three different ways:

a) The final quantity is an average of a set of repeated direct measurements. In this case, statistical methods can be applied (for example, quantities in the anthropometry or in the dosimetry task; Chapter 3.3).

b) The final quantity is a result of a mathematically formulated law, in which the incomes are directly measured with accompanying uncertainties (for example, surface tension in the stalagmometric method task). The final uncertainty is established according to error propagation rules (Chapter 3.2.3).
c) The final quantity is a functional relation between more variables (for example, correlation in the spirometry task). In this case, the uncertainty of the established functional relation is of interest rather than the error of each measured quantity itself.

### 3.2.3. Error Propagation

In many cases, the quantity $Y$ that we wish to determine is derived from several measured quantities $y = f(x_1; x_2; ...; x_n)$. Uncertainty contribution $\Delta x_i$ of each variable $x_i$ can be considered separately as long as the contributing variables are independent of each other. The total (absolute) error $\Delta y = \Delta f(x_1; x_2; ...; x_n)$ in $Y$ is obtained by combining the squared individual contributions:

$$\Delta f(x_1; x_2; ...; x_n) = \sqrt{\left(\frac{\partial f}{\partial x_1}\right)^2 \cdot (\Delta x_1)^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \cdot (\Delta x_2)^2 + ... + \left(\frac{\partial f}{\partial x_n}\right)^2 \cdot (\Delta x_n)^2}$$

or

$$\Delta f(x_1; x_2; ...; x_n) = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \cdot (\Delta x_i)^2},$$

where $\sum_{i=1}^{n}$ stands for the sum of variables starting with $i = 1$ and ending with $i = n$ ($i$ is so called the **running number**). The notation

$$\frac{\partial f}{\partial x_i} = \frac{\partial f(x_1; x_2; ...; x_n)}{\partial x_i}$$

represents the **partial derivative** of $f(x_1; x_2; ...; x_n)$ after the variable $x_i$. When taking a partial derivative with respect to one variable, treat any other variable as a constant (Appendix F). The individual contribution to the uncertainty in $f(x_1; x_2; ...; x_n)$ from a measured uncertainty in $x_i$ is the product of the uncertainty in $x_i$ with the partial derivative of $f(x_1; x_2; ...; x_n)$ with respect to $x_i$.

The basis of the addition in quadrature is an assumption that the measured quantities have a Gaussian distribution about their mean values (Chapter 3.1.1). When two or more independent Gaussians are added, the width of the new, combined distribution of values is given by this same quadrature rule. Each $\Delta$ describes the width of corresponding distribution.

When using simple mathematical operations, it is helpful to use simple formulas for error propagation derived from the general one instead of tiring (and sometimes difficult) calculating the derivatives according to the general formula above. These simple formulas can also be applied step-by-step to more complicated situations always starting with the highest mathematical operation (i.e. first raising into power, then multiplication, then addition). In any case, remember that there are no negative uncertainties, uncertainties are always positive numbers and always add (i.e. each uncertain variable will increase the final uncertainty).

If the result is a **sum** (or a **difference**) of an **independent variable** $x$ and a **constant** $k$, the final absolute uncertainty $\Delta f$ will equal the absolute uncertainty of the variable $\Delta x$ :

$$\Delta f(x+k)=\Delta x \qquad \text{or} \qquad \Delta f(x-k)=\Delta x .$$

If the result is a **product** (or a **ratio**) of an **independent variable** $x$ and a **constant** $k$, the final absolute uncertainty $\Delta f$ will equal the product (or the ratio) of the absolute uncertainty of the variable $\Delta x$ and the constant $k$ :

$$\Delta f(k\cdot x)=k\cdot\Delta x \qquad \text{or} \qquad \Delta f\left(\frac{x}{k}\right)=\frac{\Delta x}{k} .$$

If the result is a **sum** (or a **difference**) of two **independent variables** $x_1$ and $x_2$ the final absolute uncertainty $\Delta f$ will equal the square root of the sum of the squares of the absolute uncertainties of both variables $\Delta x_1$ and $\Delta x_2$ :

$$\Delta f(x_1+x_2)=\sqrt{(\Delta x_1)^2+(\Delta x_2)^2} \qquad \text{or} \qquad \Delta f(x_1-x_2)=\sqrt{(\Delta x_1)^2+(\Delta x_2)^2} .$$

If the result is a **product** (or a **ratio**) of two **independent variables** $x_1$ and $x_2$ the final relative uncertainty $\Delta_{rel} f$ will equal the square root of the sum of the squares of the relative uncertainties of both variables $\Delta_{rel} x_1$ and $\Delta_{rel} x_2$ :

$$\Delta_{rel} f(x_1\cdot x_2)=\sqrt{(\Delta_{rel} x_1)^2+(\Delta_{rel} x_2)^2} \qquad \text{or} \quad \Delta_{rel} f\left(\frac{x_1}{x_2}\right)=\sqrt{(\Delta_{rel} x_1)^2+(\Delta_{rel} x_2)^2}$$

It is also worth mentioning that relative uncertainties in multiplication and division behave like absolute uncertainties in addition and subtraction.

If the result is based on **raising to a power** (or **extracting of roots**), the final relative uncertainty equals the relative of the variable $\Delta_{rel} x$ multiplied by that power (or divided by that root):

$$\Delta_{rel} f(x^k)=k\cdot\Delta_{rel} x \qquad \text{or} \qquad \Delta_{rel} f\left(\sqrt[k]{x}\right)=\Delta_{rel} f\left(x^{\frac{1}{k}}\right)=\frac{\Delta_{rel} x}{k}$$

If the result is a **natural logarithm** of an **independent variable** $x$, the final absolute uncertainty will numerically equal the relative uncertainty of the variable $\Delta_{rel} x$ :

$$\Delta f(\ln x) = \Delta_{rel} x \quad .$$

### 3.2.4. Significant Figures

**Significant figures (significant digits)** are those figures about which there exists no or very little uncertainty. Care should be taken to distinguish between significant figures and decimal places. The scale reading could have been expressed as 0.0751 kg, but it would still have 3 significant figures. The zeros preceding the "7" are "place markers" and are not significant figures. On the other hand, quoting the result as 75.10 g would imply that the "1" is well known while the last "0" is also known, but with some degree of uncertainty. The zero after the decimal point is thus considered to be significant. The greater the number of significant figures, the more accurate the experiment is presumed to be.

It is important to report the results with the correct number of significant figures to tell the reader useful information. Suppose you have obtained from your calculations some dimension $l = 9.9238$ m with uncertainty $\Delta l = 0.086695$ m.

Begin by rounding the uncertainty in your result to 1 (or 2) significant figure, i.e. $\Delta l = 0.09$ m (or $\Delta l = 0.087$ m). Then give $l$ to the same number of decimal places, i.e. $l = (9.92 \pm 0.09)$ m (or $l = (9.924 \pm 0.087)$ m).

The **final uncertainty** should have only as many digits as the quantity with the least number of significant digits used in the calculation if presenting a result of a multiplication or a division. In a result of an addition or a subtraction the last significant figure is in the position of the least accurate result of any direct measurement.

In any case, since the uncertainty only identifies how well the result has been measured, it doesn't make sense to give any uncertainty to more than 1 or 2 significant figures.

The example above has a form sometimes referred to as a **measurement interval**. If using scientific notation, **always use the same power of 10** for both the quantity and its uncertainty. For example, quote $h = (6.4 \pm 0.3) \cdot 10^{-34}$ J·s (Table 10.9). Never forget to include the units!

To summarise: No physical quantity is ever known with perfect exactness. All physical measurement are approximations, but improvements in equipment and technique make these approximations better and better as time goes on, and they finally converge toward the actual value of the quantity being measured. The accuracy of any measurement depends on the apparatus used, the technique of the experimenter, and her/his skills in analysing the possible errors inherent in any physics experiment.

Probable experimental error comprises a range of values specified for the result of a physics experiment, within which the experimenter is confident the actual (true) value is likely to fall. Probable experimental error includes statistical (random) and systematic errors.

# 3.3. Statistical Evaluation of Measurements

## 3.3.1. Introduction

If we make repeated measurements of the same quantity, we can apply statistical analysis to study the uncertainties in our measurements. This type of analysis yields **internal errors**. The uncertainties are determined from the data themselves without requiring further estimates.

Statistical data collected from individual subjects or patients are either **numerical** (**quantitative**, for example, age, mass, pressure, ...) or **non-numerical** (**categorical** or **qualitative**, for example, gender, smoker or non-smoker, ...). The type of data collected for individual subjects determines to some extent how we summarise data for a group of subjects. Chapters 3.3.2. and 3.3.3 deal with **summary statistics** in more detail.

**Hypothesis testing** and **confidence intervals** are the names of two aspects of drawing conclusions from sample data about the populations from which the samples were taken (**inductive statistics**). In a medical study the selected patients (or subjects) should ideally be a **random sample** from a defined **population**. If this is not the case, it may not be possible to draw any useful conclusions about the population in question.

In the above, **population** means all the measurements (or counts) of interest in a study; a **sample** means a subset of the population; a **random sample** is such that each measurement in the population has the same probability of being included in the sample. Confidence intervals are described in Chapter 3.3.4; hypothesis testing in Chapter 3.3.5.

A high proportion of statistical analyses are carried out to study the relation between two (or more) variables within a group of subjects (Chapter 3.3.6.). Three main purposes of such analyses might be:

- To asses, whether the two variables are associated (correlation).
- To enable the value of one variable to be predicted from any known value of the other variable (linear regression).
- To asses the amount of agreement between the values of the two variables (not discussed in this text).

## 3.3.2. Measures of Location

To obtain **the best estimate** of a measured quantity $X$ from $n$ measurements of the quantity, calculate the **mean** (**average**) $\bar{x}$ of the measurements as the sum of all measured values in a sample divided by the number of measurements:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

where $n$ is the sample size, $x_i$ is the $i^{th}$ measurement of the quantity $X$, $i = 1, 2, ..., n$.

Other possibilities to obtain an estimate of a measured quantity $X$ from $n$ measurements of the quantity, is given by **median** $\tilde{x}$, that divides an ordered sample into two equal parts. This is used in the case when average cannot be used (the distribution of data is not normal, for example, very skewed data as in asymmetric distributions, Figure 3.2). The value of median is calculated according to one of the following formulas:

- if the sample size $n$ is even, the median equals the average of two "middle" measurements in an ordered sample (ordered according to the values, not to the measurement order):

$$\tilde{x} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

- if the sample size $n$ is odd, the median equals the average of the "middle" measurement in an ordered sample:

$$\tilde{x} = x_{\frac{n+1}{2}}$$

The most frequent value in the sample is the **mode** $\hat{x}$. This measure of location can be also used for categorical data and for data not distributed normally. (If data are distributed evenly, there is no mode in the sample.)

### 3.3.3. Measures of Variability

Probably the simplest measure of variability is the **range** $R$ given as the difference between the highest (maximal, $x_{max}$) and the lowest (minimal, $x_{min}$) measured value:

$$R = x_{max} - x_{min}.$$

The range is the length of the smallest interval, which contains all the data. It is usually used together with the median.

The **standard deviation** $s_{\bar{x}}$ describes the scatter of measurements about the average, and is given by the square root of the sum of squared differences between each measurement and the average divided by degrees of freedom

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

For population studies, instead of degrees of freedom, the population size $N$, the population standard deviation $\sigma$ and the population mean $\mu$ is used:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}.$$

The expression in denominator of sample standard deviation $f = n - 1$ represents the **degrees of freedom.**

The standard deviation is used if we want to know where a measurement is likely to fall compared to the mean value. The **variance** $s_{\bar{x}}^2$ is the square of the standard deviation:

$$s_{\bar{x}}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}.$$

The analogue formula holds for the population variance:

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}.$$

The **standard error of the mean** $se_{\bar{x}}$ is given by the square root of the sum of squared differences between each measurement and the average divided by degrees of freedom and by the sample size:

$$se_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)\cdot n}}$$

or (based on known standard deviation)

$$se_{\bar{x}} = \frac{s_{\bar{x}}}{\sqrt{n}}.$$

The standard error of the mean is also referred to as the **standard deviation of the mean**, because it is an estimate of the standard deviation of the distribution of means that would be obtained if the mean were measured many times. If we take a large sample from a population and calculate the sample mean $\bar{x}$ and the standard error $se_{\bar{x}}$, then there is approximately a 95% chance that the population mean will lie between $(\bar{x} - 2 \cdot se_{\bar{x}})$ and $(\bar{x} + 2 \cdot se_{\bar{x}})$. (To be exact, the value 1.96 should be used instead of 2 (Table 3.1). Remember that this is valid only for large samples.)

Taking more measurements of a given quantity might not improve the standard deviation, but it should make the standard error of the mean smaller (scaling it as $\sqrt{n}$). The standard error of the mean is used if we want to know how well we have determined the average value itself. This measure of variability is in the same units as the standard deviation, i.e. the same as the units of the variable in question.

The **coefficient of variation** $V_{\bar{x}}$ indicates the quality of a population estimate. It is defined as the ratio of the standard deviation to the mean:

$$V_{\bar{x}} = \frac{s_{\bar{x}}}{\bar{x}} \cdot 100\%$$

The coefficient of variation is a dimensionless number (reported on a scale of 0 % to 100 %) that allows comparison of the variation of populations that have significantly different mean values.

A coefficient of variation of 16.7 % - 25.5 % indicates moderate sampling variability. A coefficient of variation of 25.6 % - 33.3 % indicates high sampling variability. Estimates with either a moderate or high coefficient of variation should be interpreted with caution. In some situations, the sample may be too small to produce a reliable estimate.

### 3.3.4. Confidence Interval

The statistical procedures are designed to help to dedicate, whether or not a set of observations (measurements) is compatible with some hypothesis. These procedures yield $p$ values to estimate the chance reporting that, for example, a treatment has an effect when it really does not (see Chapter 3.3.5.). It can be more informative to think not only in terms of the accept/reject approach of statistical hypothesis testing but also to estimate the size of the treatment affect together with some measure of the uncertainty in that estimate.

The resulting estimates, called **confidence intervals**, can also be used to test hypotheses. Both approaches yield the same conclusions. Confidence intervals are also used to estimate the range of values that include a specified proportion of all members of a population, such as the "normal range" of values for a laboratory test.

#### 3.3.4.1. Confidence Interval for a Mean

Let $\mu$ be the **population mean**, which is not known, to distinguish it from the **sample mean** $\bar{x}$ with **sample standard deviation** $s_{\bar{x}}$, which is known, if a random sample has been taken. If the two values

$$\bar{x} - \frac{t_{f;\alpha} \cdot s_{\bar{x}}}{\sqrt{n}} \quad \text{and} \quad \bar{x} + \frac{t_{f;\alpha} \cdot s_{\bar{x}}}{\sqrt{n}}$$

are calculated, the unknown value $\mu$ lies between these limits with the **probability** $p = 1 - \alpha$. These limits are called **confidence limits, lower** and **upper**, respectively. The interval between them is called **confidence interval** and the probability $p$ corresponds to the **confidence level**, while $\alpha$ corresponds to the **significance level**. These formulas can be used as long as the variable in question is normally distributed (Chapter 3.1.1). Values $t_{f;\alpha}$ arise from Student's

distribution of probability, depend on degrees of freedom $f = n-1$ ($n$ is the sample size), and are given in tables (Tables 10.12 and 10.13).

Notice that the confidence interval is the wider the higher is the confidence level (the probability $p$), i.e. the smaller is the significance level $\alpha$, for example, a 99% confidence interval will be wider than a 95% confidence interval but narrower than a 99.9% confidence interval. Therefore, the confidence interval for one mean corresponds to the expression

$$\bar{x} - \frac{t_{f;\alpha} \cdot s_{\bar{x}}}{\sqrt{n}} < \mu < \bar{x} + \frac{t_{f;\alpha} \cdot s_{\bar{x}}}{\sqrt{n}}.$$

### 3.3.4.2. Confidence Interval for Paired Observations

If two populations have to be compared, the data may be paired or unpaired. Paired data may arise in the case that the same subjects are measured before and after an assumption changed (for example, before and after some exercise, before and after some medical treatment, in two independent states each receiving different treatments, early in the morning and late in the afternoon, ...).

In this case, we are interested in the differences between the pairs

$$d_i = x_i^{(2)} - x_i^{(1)}$$

and not in the individual sets of measurements of the same quantity. Here $x_i^{(1)}$ is the $i^{th}$ subject in the first set of the measurement of quantity $X$; $x_i^{(2)}$ is the $i^{th}$ subject in the second set. The confidence interval for the mean difference $\mu_{\bar{d}}$ is calculated in analogy with $\mu$:

$$\bar{d} - \frac{t_{f;\alpha} \cdot s_{\bar{d}}}{\sqrt{n}} < \mu_{\bar{d}} < \bar{d} + \frac{t_{f;\alpha} \cdot s_{\bar{d}}}{\sqrt{n}}$$

where $\bar{d}$ is the mean (average) value of differences calculated in analogy with $\bar{x}$:

$$\bar{d} = \frac{1}{n} \cdot \sum_{i=1}^{n} d_i$$

and $s_{\bar{d}}$ is the standard deviation of differences calculated in analogy with $s_{\bar{x}}$:

$$s_{\bar{d}} = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}.$$

Corresponding degrees of freedom are $f = n-1$, $n$ is the number of pairs (not the number of measurements). If the $100 \cdot (1-\alpha)$ % calculated confidence interval includes zero, there is no statistical difference between the paired measurements at the significance level $100 \cdot \alpha$ %.

### 3.3.4.3. Confidence Interval for Comparing Two Means (Unpaired Observations)

We often have to compare two populations by taking independent samples from each population. There is usually no question of pairing an observation from one sample with an observation from the other sample, for example, treated group versus control group.

If the one population has the true mean value $\mu_1$ (sample mean $\bar{x}_1$) and the second one has the true mean value $\mu_2$ (sample mean $\bar{x}_2$), the confidence interval for the difference of both means (assuming normal distributions of the variable in both populations) will be:

$$|\bar{x}_1 - \bar{x}_2| - t_{f;\alpha} \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < |\mu_1 - \mu_2| < |\bar{x}_1 - \bar{x}_2| + t_{f;\alpha} \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where the subscripts 1 and 2 refer to corresponding population measures (see Chapters 3.2.2 and 3.2.3), $s^2$ is the **pooled variance**

$$s^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2} .$$

Under the assumption of equal population variances $s_1^2$ and $s_2^2$, the pooled variance represents the best estimate of this unknown population variance. It is a weighted average of the variance within each group. Corresponding degrees of freedom are $f = n_1 + n_2 - 2$, $n_1$ and $n_2$ are the sample sizes of both samples.

If the $100 \cdot (1 - \alpha)$ % calculated confidence interval includes zero, there is no statistical difference between the measurements in the two populations at the significance level $100 \cdot \alpha$ %.

## 3.3.5. Hypothesis Testing

The main idea of a **hypothesis test** (sometimes called "a test of significance") is to set up two hypotheses about a population parameter. After the data are collected, decision is made which of the two hypotheses is better supported by them. The first hypothesis is called the **null hypothesis** ($H_0$) and indicates the idea that there is no difference between the true value and the corresponding sample characteristic. The second hypothesis is called the **alternative hypothesis** ($H_A$) and indicates the idea that there is some difference between the true value and the corresponding sample characteristic.

Following seven steps should be used in any hypothesis testing assuming that the variable is normally distributed:

1. Set a null hypothesis $H_0$ (a two-tailed or a one-tailed).

2. Set an alternative hypothesis $H_A$ (a two-tailed or a one-tailed).
3. State the **significance level** of the test, which is the risk of rejecting the hypothesis $H_0$ when $H_0$ is the correct hypothesis. $\alpha = 5\%$ is a conventional significance level used in hypothesis testing, next often used levels are 1 %, 0.5 %, 0.1 %.
4. Calculate **a test statistic** *t* using an appropriate formula.
5. Look up tabulated test statistics for a **two-tailed hypothesis** or for a **one-tailed hypothesis** (Tables 10.12 and 10.13).
6. Compare calculated and tabulated test statistics. If the calculated statistics is numerically greater than the tabulated one, reject $H_0$ (i.e. the alternative hypothesis $H_A$ is valid automatically).
7. Draw a conclusion.

Table 3.2: Examples of rejection or acceptation of null hypothesis and the corresponding *p* values.

| Value *p*: | Null hypothesis $H_0$ |
|---|---|
| $p \geq 0.05$ | $H_0$ is not rejected at the 5 % significance level |
| $p < 0.05$ | $H_0$ is rejected at the 5 % significance level (but is not rejected at the 1 % significance level) |
| $p < 0.01$ | $H_0$ is rejected at the 1 % significance level (but is not rejected at the 0.5 % significance level) |
| $p < 0.005$ | $H_0$ is rejected at the 0.5 % significance level (but is not rejected at the 0.1 % significance level) |
| $p < 0.001$ | $H_0$ is rejected at the 0.1 % significance level (but is not rejected at any lower significance level) |

Statistical computer programs often give $p$ values rather than stating, whether a null hypothesis is or is not rejected at a particular level of significance. The statements given in Table 3.1 are equivalent in pairs. The $p$ value is the probability of being wrong when asserting that a true difference exists.

Hypotheses tests can be used under the assumption that the variable has a normal distribution or also when having a large sample ($n > 30$), but the assumption is usually no longer needed in that case.

### 3.3.5.1. Hypothesis Test for a Mean

This method is used if we want to know, whether a population mean equals some value of interest.

Start with setting the null hypothesis that the population mean μ equals some value $M$ (here the $M$ is a number):

$H_0: \ \mu = M$

Then set the alternative hypothesis that the population mean μ does not equal some value $M$ (two-tailed test):

$H_A: \ \mu \neq M$

After choosing the appropriate significance level, calculate the test statistics $t$ using the formula (all used symbols are explained above):

$$t = \frac{\bar{x} - M}{s_{\bar{x}}} \cdot \sqrt{n}$$

Then continue with the steps 5 - 7 described at the beginning of Chapter 3.3.5.

When a one-tailed test is needed, the null hypothesis remains the same, but the alternative hypothesis changes to one of the following possibilities:

$H_A: \ \mu > M$ or $H_A: \ \mu < M$

Next process is the same as described above.

### 3.3.5.2. Hypothesis Test for Paired Observations

This method is used if we want to know whether two population means are significantly different when using paired data.

Start with setting the null hypothesis that the difference between population means $\mu_{\bar{d}}$ equals zero (that the differences come from a population with a mean of zero; there is no effect between measurement conditions of the same subjects):

$H_0: \ \mu_{\bar{d}} = 0$

Then set the alternative hypothesis that the differences does not come from a population with a mean $\mu_{\bar{d}}$ of zero (there is some effect; two-tailed test):

$H_A: \ \mu_{\bar{d}} \neq 0$

After choosing the appropriate significance level, calculate the test statistics $t$ using the formula (all used symbols are explained above):

$$t = \frac{\bar{d} - 0}{s_{\bar{d}}} \cdot \sqrt{n}$$

Then continue with item 5 - 7 described at the beginning of Chapter 3.3.5.

When a one-tailed test is needed, the null hypothesis remains the same, but the alternative hypothesis changes to one of the following possibilities:

$H_A: \ \mu_{\bar{d}} > 0$ or $H_A: \ \mu_{\bar{d}} < 0$

Next process is the same as described above.

### 3.3.5.3. Hypothesis Test for Comparing Two Means (Unpaired Observations)

This method is used if we want to know, whether means of two populations are significantly different.

Start with setting the null hypothesis that the difference between the two different population means $\mu_1 - \mu_2$ equals zero:

$H_0: \ \mu_1 = \mu_2$

Then set the alternative hypothesis that the difference between the two different population means $\mu_1 - \mu_2$ does not equal zero (two-tailed test):

$H_A: \ \mu_1 \neq \mu_2$

After choosing the appropriate significance level, calculate the test statistics $t$ using the formula (all used symbols are explained above, the pooled variance is the same as in Chapter 3.3.4.3):

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Then continue with items 5 - 7 described at the beginning of Chapter 3.3.5.

When a one-tailed test is needed, the null hypothesis remains the same, but the alternative hypothesis changes to one of the following possibilities:

$H_A: \ \mu_1 > \mu_2$ or $H_A: \ \mu_1 < \mu_2$

Next process is the same as described above.

## 3.3.6. Regression and Correlation

So far, the problem of observing two (or more) numerical variables on a number of subjects has not been considered yet, but this kind of problem is quite common. We might be interested in two variables of a sample of subjects. We could draw a scatter diagram, one variable $x$ (the independent variable, measured quantity $X$) on the horizontal axis, the next one $y$ (the dependent variable, measured quantity $Y$ depending on the quantity $X$) on the vertical axis (Chapter 3.4.2, Figure 3.8). If there appears to be a linear relation (or any other mathematically reasonable relation), we can do two types of analysis:

- Obtain the equation of **the best-fit** for the data of the form

  $y = a + b \cdot x$

  The "best fit" here means that the sum of squares of distances from the points to the line in vertical dimension is minimised. The equation is called the **simple linear regression equation** of $y$ on $x$ and would allow predicting $y$ from the $x$ of the subject (Chapter 3.3.6.1).

- Calculate the correlation coefficient $r$ between the variables $x$ and $y$. The aim in this type of analysis is to measure the degree to which the two variables are linearly related. The value of correlation coefficient can also be tested whether the null hypothesis that there is no correlation in the bivariate population of variable $x$ and variable $y$ is valid.

### 3.3.6.1. Simple Linear Regression

Assume we have $n$ subjects. On each subject, the two quantities were measured and we obtained $n$ pairs of measured values: $(x_1; y_1), (x_2; y_2), \ldots, (x_n; y_n)$.

The **slope** $b$ and the **intercept** $a$ of the linear regression line are given by the following formulas:

$$b = \frac{\sum_{i=1}^{n} x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2}$$

and

$a = \bar{y} - b \cdot \bar{x}$ .

To draw the linear regression line, proceed as follows:

1. Choose the smallest value of $x = x_{min}$ and calculate the predicted value of $y_{min}$ (use the formula on the previous page).
2. Plot the corresponding point $[x_{min}; y_{min}]$ in the graph.
3. Choose the largest value of $x = x_{max}$ and calculate the predicted value of $y_{max}$.
4. Plot the corresponding point $[x_{max}; y_{max}]$ in the graph.
5. Join the two plotted points. This is the regression line. About the half of the measured values should be above the line and about the half below the line.

### 3.3.6.2. Correlation Coefficient

**Pearson's product moment correlation coefficient** $r$ can be calculated for samples of bivariate data, and measures the degree to which two variables are linearly related using the following formula:

$$r = \frac{\sum_{i=1}^{n} x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\left[\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2\right] \cdot \left[\sum_{i=1}^{n} y_i^2 - n \cdot \bar{y}^2\right]}}$$

It can be shown that $r$ lies within the range –1 to +1. If the variable $y$ increases with increasing variable $x$, $r$ will be positive. If the variable $y$ decreases with increasing variable $x$, $r$ will be negative. If there is no trend one way or other, the value of $r$ will be close to zero. If all the points lie exactly on a straight line, $r$ will be either +1 or -1, depending on the trend.

If we know that the sample taken is a random one from a population, we can test the hypothesis that the population correlation coefficient $\rho$ is zero. Next seven steps are required (under the assumption that both variables $x$ and $y$ are normally distributed; Chapter 3.1.2):

1. Set the null hypothesis that there is no correlation between the variables $x$ and $y$ in the population:
   $H_0: \ \rho = 0$
2. Set the alternative hypothesis that there is some correlation between $x$ and $y$ in the population:
   $H_A: \ \rho \neq 0$
3. State the significance level of the test, which is the risk of rejecting the hypothesis $H_0$ when $H_A$ is the correct hypothesis (see Chapter 3.5.).
4. Calculate the test statistics $t$ using the formula:
   $$t = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$
5. Look up tabulated test statistics for a two-tailed hypothesis (Table 10.13).
6. Compare calculated and tabulated test statistics. If the calculated statistic is numerically greater than the tabulated one, reject $H_0$ (i.e. the alternative hypothesis $H_A$ is valid).
7. Draw a conclusion (look whether $r > 0$ or $r < 0$).

Correlation coefficient can be used only if linear dependence between the variables $x$ and $y$ is assumed. If this is not the case, appropriate transformations of both variables may be used to get a linear dependence of the new pair of variables (for next study see, for example, textbooks on statistics, data processing and medical research listed in References – Chapter 11).

## 3.4. Graphical Presentation of Measurements

### 3.4.1. Tables

A **table** is both a mode of visual communication and a means of arranging **data**. The use of tables is pervasive throughout all communication, research and data analysis. The precise conventions and terminology for describing tables varies depending on the context. Tables differ in variety, structure, flexibility, notation, representation, and use. As a communication tool, a table allows a form of generalisation of information. It provides a familiar way to convey information that might otherwise not be obvious or readily understood.

A table consists of an ordered arrangement of **rows** and **columns**. A column is usually identified by a name that can consist of a word, a phrase or a numerical index. The first row is not counted, because it is only used to display the column names. This is traditionally called a **header row (heading row)**. The intersection of a column and a row on a table that can contain data is called a **cell** (Figure 4.1).

Above each table has to be a **title** that briefly explains the content of the table. Below the table, a **source** has to be given if the author of the text (paper, protocol) has not measured the data by him/her. Also **comments** and **explanations** needed to understand the table content should be here (explanatory footnotes).

Table number: Title of the table

| | | Heading row | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| | 1 | **Cell** | **Row** | | | |
| **Legend** | 2 | **Column** | | | | |
| | 3 | | | | | |
| | 4 | | | | | |

*Source: .....*
*Footnotes: .....*

Figure 3.7: A scheme of a 4 × 5 table (4 rows × 5 columns)

Whether or not to put descriptive data in tables may depend on the number of variables and groups of subjects. It is preferable, although not always possible, to put a like-kind data in columns rather than in rows. Different statistical characteristic

listed in one cell should be clearly distinguished (for example, average and standard deviation; median and range).

Tables can be used to show **raw data** (from each measured or observed subject), although this is only reasonable when there are not too many observations. It is always helpful to order the data by one of the variables (starting from the column on the left side, if reasonable, or emphasising the column with the data in sequence).

For summarising data in a table, keep in mind the following **guidelines**:

- Include a title at the top of the table that describes the content, including outcome measure, comparison groups or population, and time period, if relevant. Omit unnecessary words.
- Clearly label columns and rows of the table.
- If you report means, also report standard errors of the means or standard deviations (for example, in parentheses, as a separate column). Label the columns appropriately. Confidence intervals for means are useful to indicate precision.
- If you report rates or percentages, also report the numerators and the denominators (total sample size) for the rates, along with confidence intervals to indicate precision.
- When a table represents comparisons between groups (columns) on multiple variables (rows), indicate statistically significant comparisons in the appropriate row with footnotes denoting the level of significance.
- Avoid giving too much information in one table. As a rule of thumb, tables should not be bigger than 8 columns and 10 rows.

### 3.4.2. Graphs

A **graph** is a drawing representing the relationship between two sets of data in pictorial form, one set represented on a horizontal scale or axis, the other on a vertical scale or axis. The relationship is plotted where the two scales intersect, the line between meeting points generally being called the graph. Graphs are especially useful in showing the trends in the data. They give a visual display of information or data organized to help people interpret, understand, or quickly find information.

In order to construct an accurate graph, consider the following **guidelines**:

- Always include a **title**. The title should be clear and concise and describe what the graph represents (used outcome measure and the form, in which it is reported as individual measurements or group means; comparison groups if two or more groups are being compared or the population under study if one group is being described; time period under study, if relevant).
- Clearly label the **axes** (vertical and horizontal) and include the scale of measurement of the variables depicted. The **independent variable** (for example, time) is plotted along the horizontal (x) axis. The **dependent variable** is plotted along the vertical (y) axis.

- Mark **scale divisions** with tick marks and number them from 0 (zero) to the maximum value included in the data (or just above it). The tick marks should face outwards not to be confused with any data in the graph.
- Clearly label **comparison groups**. Whether groups are depicted by lines or bars, each should be represented by a separate symbol (different shading, for illustration different shading is used in the examples of graphs) that is clearly indicated in the legend. Perhaps the most standard symbols are open and closed circles, triangles and squares or the multiplication sign (○, △, □, ●, ▲, ■, ×). Different types of connecting lines can also be used (solid ——, dashed – – –, dash dot — · –, dotted (dot) ·····).
- Keep the graph simple and uncluttered. Do not try to include too much information (such as too many subgroups, more than one set of axes, multiple graphs on a page).

**Scatter plots** usually display raw data (Figure 3.8). They can be accompanied by a **regression line** to emphasize the mutual dependence of variables. Single points must not to be connected by a zigzag line!

**Bar graphs** (horizontal) usually compare different variables at one time (Figure 3.9).

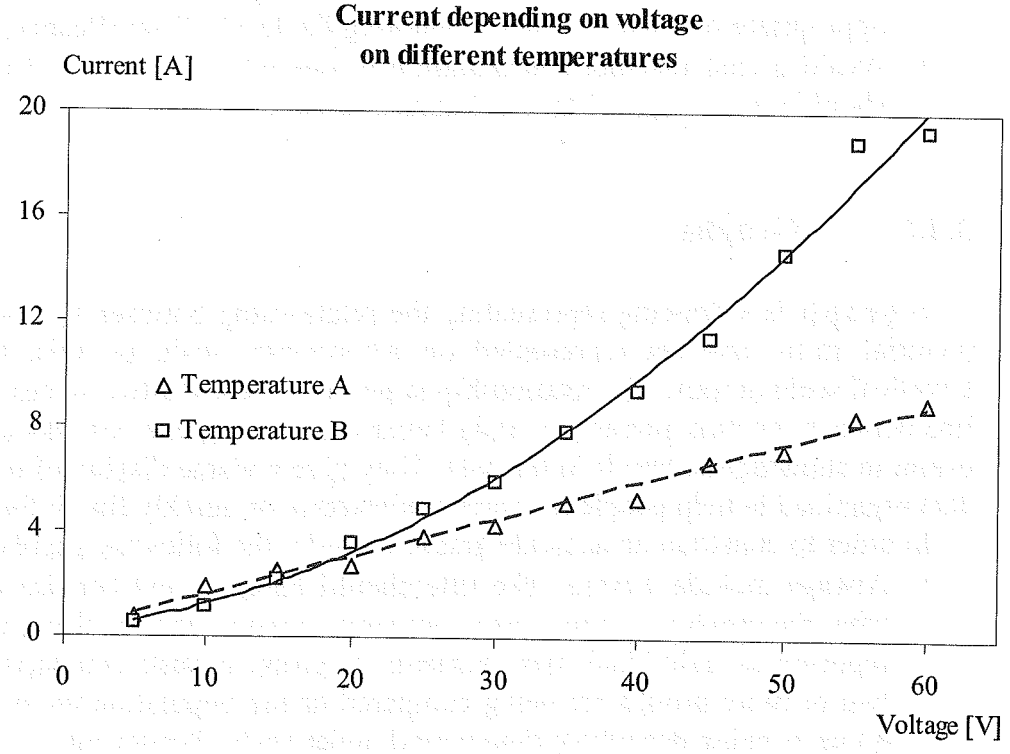


Figure 3.8: An example of a scatter plot displaying 2 sets of measured data with corresponding regression lines (linear for Temperature A, quadratic for Temperature B).

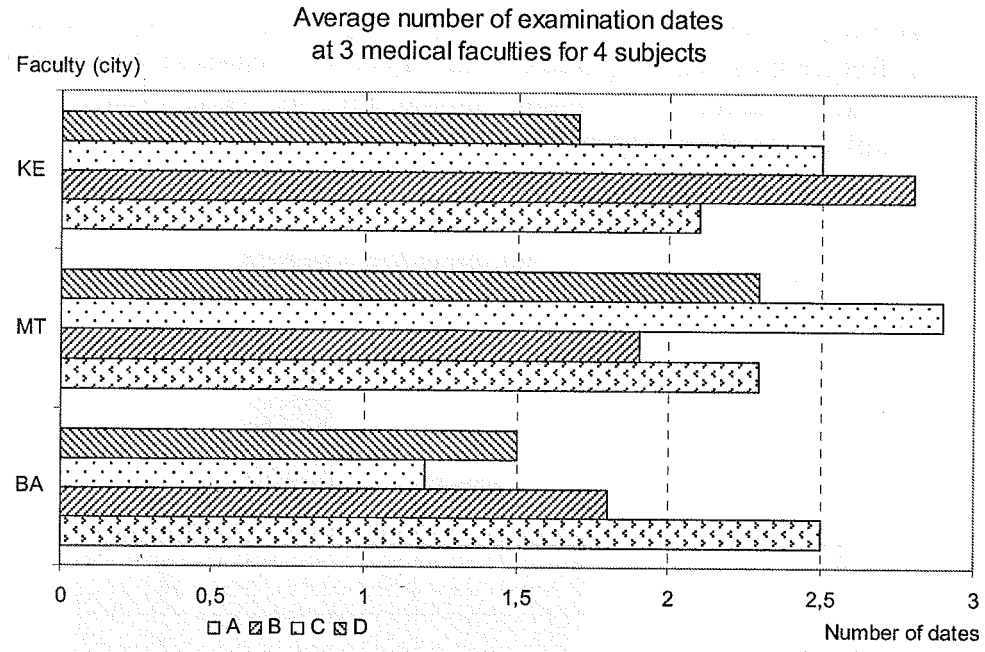


Figure 3.9: An example of a bar graph.

**Column graphs** (vertical) usually compare the same variable at different times or in different groups (Figure 3.10). Column graphs can be used to show the means and standard deviations or standard errors of the means, i.e. the manipulated data, but it has to be clear that the horizontal top of the column represents the mean value and what type of error represents the bar.

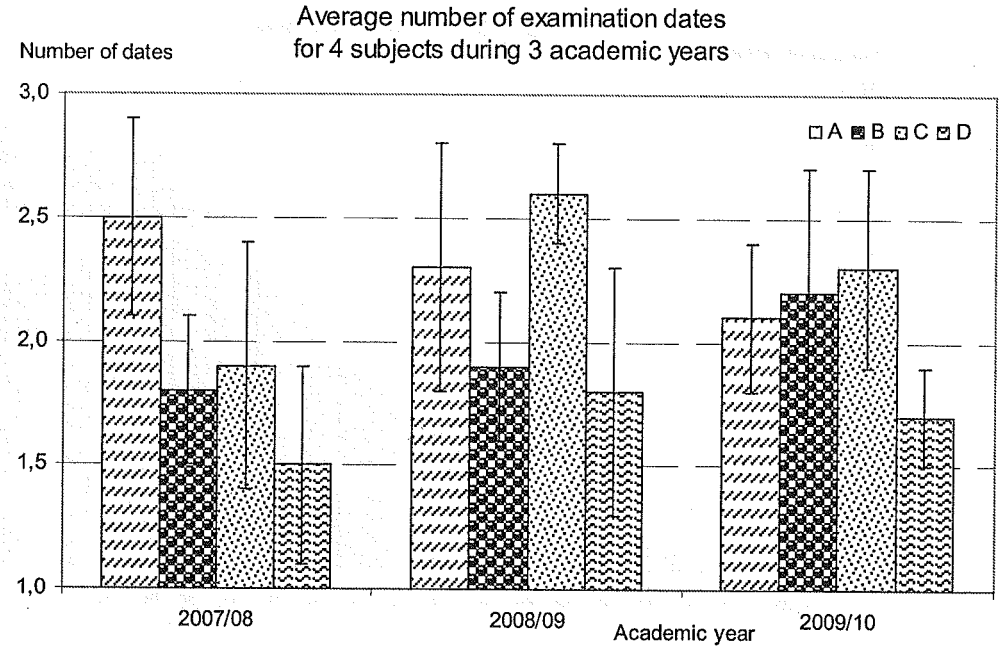


Figure 3.10: An example of a column graph.
Error bars represent the standard deviations.

**Histograms** also look like column graphs, but they illustrate continuous data (frequencies, data separated into intervals - classes) and compare the areas under each column. Each column should have the same width corresponding with the width of the class (Figures 3.11. and 3.12.).

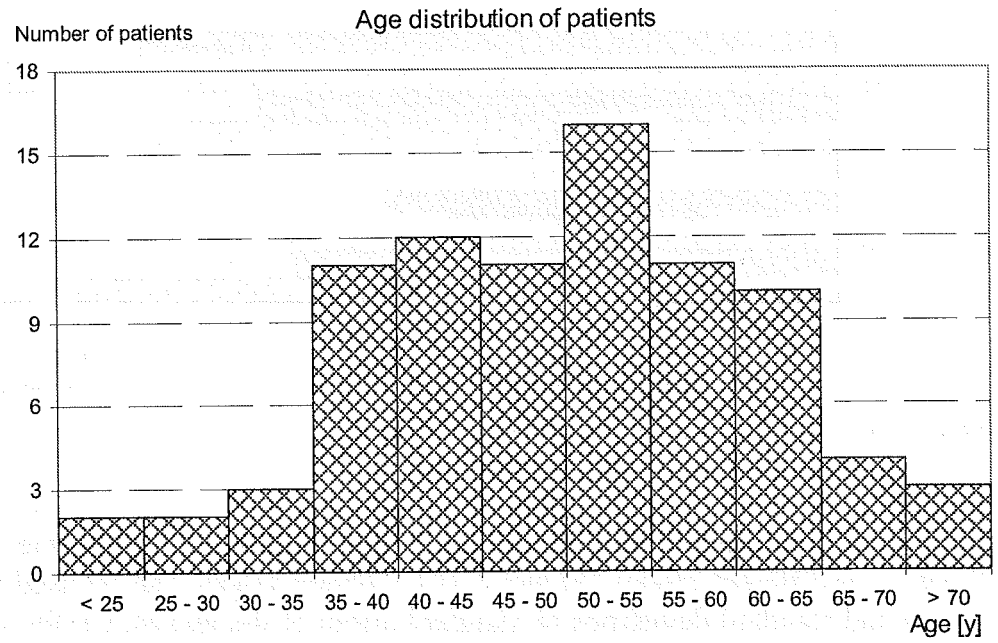


Figure 3.11: An example of an absolute frequency histogram.
Same data as in Figures 3.1 and 3.12.

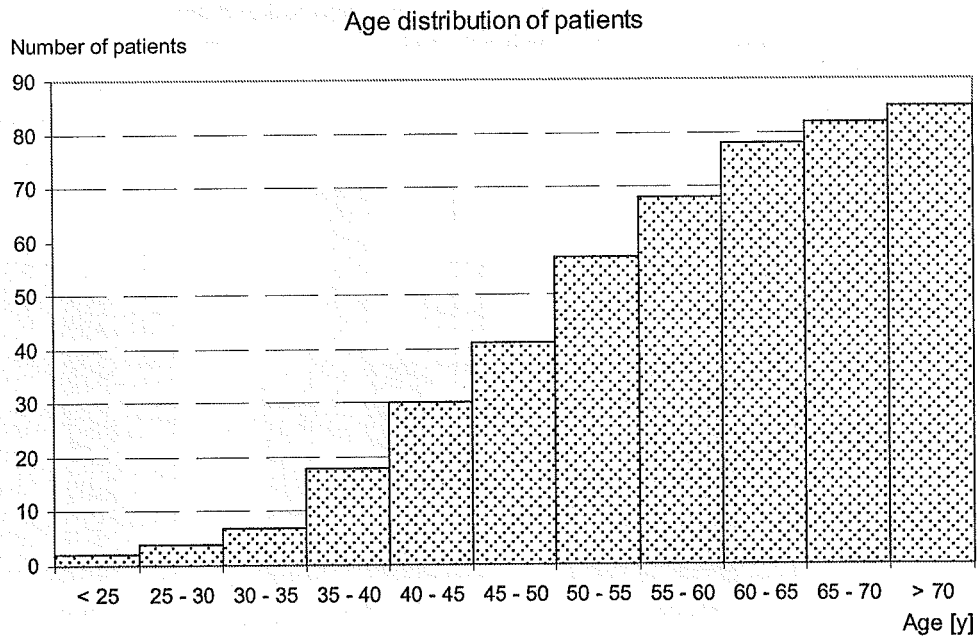


Figure 3.12: An example of a cumulative absolute frequency histogram.
Same data as in Figures 3.1 and 3.11.

**Pie charts** are used for showing the distribution of items within a total (Figure 3.13). The included **slices** should not be too small.

Three-dimensional charting, although popular, is usually unnecessary when two-dimensional variables are under consideration, and may even lead to less clear reading of the chart. In opposite, it can be very useful for three-dimensional data.

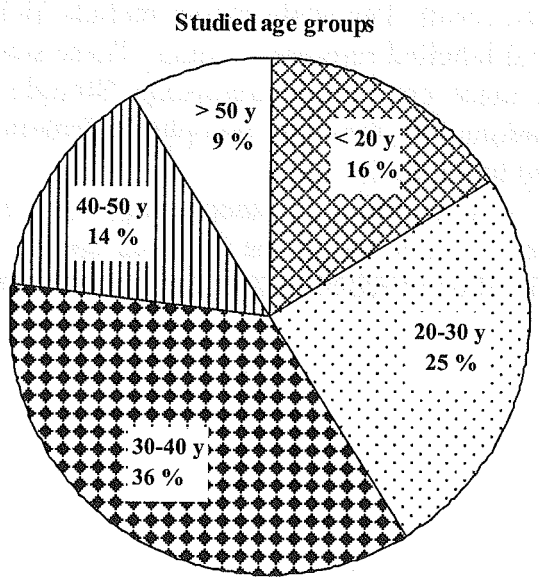


Figure 3.13: An example of a pie chart

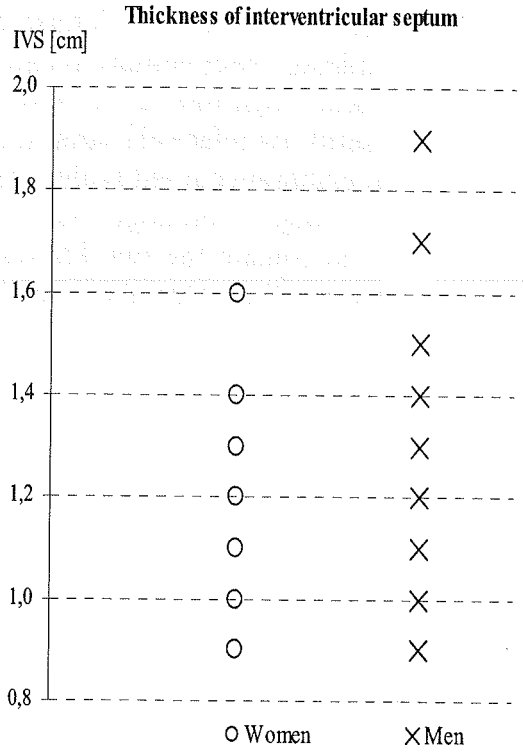


Figure 3.14: An example of a dot plot

The **dot plots** (vertical scatter plots, Figure 3.14) are a variant of the column graph, but give much more information including distributions, ranges and outliers, to which the mean, median, mode, and different types of quantiles are easily added.

**A nomogram** or **a nomograph** is a graphical calculating device, a two-dimensional diagram designed to allow the approximate graphical computation of a function. Like a slide rule, it is a graphical analogue computation device; and, like the slide rule, its accuracy is limited by the precision with which physical markings can be drawn, reproduced, viewed, and aligned. Most nomograms are used in applications where an approximate answer is appropriate and useful. Otherwise, the nomogram may be used to check an answer obtained from an exact calculation method.

The slide rule is intended to be a general-purpose device. Nomograms are usually designed to perform a specific calculation, with tables of values effectively built in to the construction of the scales, for example, the psychrometric chart (Figure 4.6).

A nomogram typically has three scales: two scales represent known values and one scale is the scale where the result is read off. The known scales are placed on the outside; i.e. the result scale is in the centre. Each known value of the calculation is marked on the outer scales and a line is drawn between each mark. Where the line and the inside scale intersects is the result. The scale marks include 'tick marks' to indicate exact number locations, and labelled reference values. These scales may be linear, logarithmic or have some more complex relationship. Straight scales are useful for relatively simple calculations, but for more complex calculations, simple or elaborate curved scales may need to be used.

Usage of nomograms is simple — a taut string or other straight edge is placed so as to contact the two known values on their lines. The required answer is read off another line. This allows calculation of one variable when the other two are known.

# 4. Evaluation of Physical Factors of the Working Place

## 4.1. Measurement of Chosen Microclimatic Factors

### 4.1.1. Temperature

**Temperature** is a physical property of a system that underlies the common notions of hot and cold; something that is hotter has the greater temperature. Temperature is one of the principal parameters of thermodynamics. The temperature of a system is related to the average energy of microscopic motions in the system. For a solid, these microscopic motions are principally the vibrations of the constituent atoms about their sites in the solid. For an ideal monatomic gas, the microscopic motions are the translational motions of the constiutent gas particles.

Temperature is a measure of the average energy of the particles (atoms or molecules) of a substance. This energy occurs as the translational motion of an article or as internal energy of a particle, such as a molecular vibration or the excitation of an electron energy level. Although highly specialised laboratory equipment is required to directly detect the translational thermal motions, thermal collisions by atoms or molecules with small particles suspended in a fluid can be seen with an ordinary microscope (the Brownian motion). The thermal motions of atoms are very fast and temperatures close to absolute zero are required to observe them directly.

Molecules, such as $O_2$, have more degrees of freedom than single atoms: they can have rotational and vibrational motions as well as translational motion. An **increase in temperature** will increase the average translational energy. It will also increase the energy associated with vibrational and rotational modes. Thus a diatomic gas, with extra degrees of freedom like rotation and vibration, will require a higher energy input to change the temperature by a certain amount, i.e. it will have a higher heat capacity than a monatomic gas.

The process of **cooling** involves removing energy from a system. When there is no more energy able to be removed, the system is said to be at absolute zero, which is the point on the thermodynamic (absolute) temperature scale where all kinetic motion in the particles comprising matter ceases and they are at complete rest in the "classic" (non-quantum mechanical) sense. By definition, **absolute zero** is a temperature of exactly 0 K = –273.15°C.

Temperature plays an important role in almost all fields of science, including physics, chemistry, and biology. Many physical properties of materials including the phase (solid, liquid, gaseous or plasma), density, solubility, vapor pressure, and electrical conductivity depend on the temperature. Temperature also plays an important role in determining the rate and extent to which chemical reactions occur. This is one reason why the human body has several elaborate mechanisms for maintaining the temperature at 37°C, since temperatures only a few degrees higher

can result in harmful reactions with serious consequences. Temperature also controls the type and quantity of thermal radiation emitted from a surface. One application of this effect is the incandescent light bulb, in which a tungsten filament is heated electrically to a temperature at which significant quantities of visible light are emitted.

The temperature of 0°C corresponds to **water freezing**; 100°C corresponds to the **boiling point of water** at the sea level. In this scale, a temperature difference of 1 degree is the same as a 1 K temperature difference, so the scale is essentially the same as the Kelvin scale, but offset. By the everyday applications, it is often convenient to use the Celsius scale, in which 0°C corresponds to 273.15 K. Thus, the following equation can be used to convert from degree Celsius to Kelvin (just Kelvin, not degree!):

$$\mathrm{K} = [°\mathrm{C}]\left(\frac{1\mathrm{K}}{1°\mathrm{C}}\right) + 273.15\,\mathrm{K}$$

The world's average surface air temperature is 15°C. Room temperature, in common usage, is taken to be 21°C – 23°C. In chemistry and other sciences, standard temperature and pressure is a standard set of conditions for experimental measurements, to enable comparisons to be made between sets of data. Internationally, the current standard temperature and pressure defined by the International Union of Pure and Applied Chemistry is an absolute pressure of 100 kPa and a temperature of 273.15 K. Other organisations have established a variety of alternative definitions for the standard reference conditions of temperature and pressure.

Many methods have been developed for measuring temperature. Most of them rely on measuring some physical property of a working material that varies with temperature.

**A glass thermometer** is one of the most common devices for measuring temperature. It is based on volume expansion of liquids

$$V = V_0 \cdot (1 + \beta \cdot \Delta T),$$

where $V_0$ is the original volume at some temperature $T_0$ [K], $V$ is the changed volume at the temperature $T$ [K],

$$\Delta T = T - T_0$$

is the temperature change, $\beta$ [$K^{-1}$] is the coefficient of volume expansion.

A glass thermometer consists of a glass tube filled with a suitable liquid (for example, coloured spirit; mercury should not be used), which acts as the working fluid (Figure 4.1). Temperature increases cause the fluid to expand, so the temperature can be determined by measuring the volume of the fluid. Such thermometers are usually calibrated, so that one can read the temperature simply by observing the level of the fluid in the thermometer.

**A bi-metal mechanical thermometer** is based on linear expansion of metals:

$$L = L_0 \cdot (1 + \alpha \cdot \Delta T),$$

where $L_0$ [m] is the original lenght at some temperature $T_0$ [K], $L$ [m] is the changed lenght at the temperature $T$ [K], $\Delta T$ is the temperature change as above, $\alpha$ [$K^{-1}$] is the coefficient of linear expansion.

A bi-metallic strip is used to convert the temperature change into mechanical displacement. The strip consists of two thin strips of different metals, which expand at different rates as they are heated (usually steel and copper). The strips are joined together throughout their length. The different expansions force the flat strip to bend in one direction if heated, and in the opposite direction if cooled below its normal temperature. The metal with the higher expansion is on the outer side of the curve when the strip is heated and on the inner side when cooled. Such strips can be wound onto a coil and attached to a pointer, which indicates the temperature on a circular scale (Figure 4.2). This type of thermometers is also used in thermostats to control the electric power to keep the temperature of a region of space (for example, room).

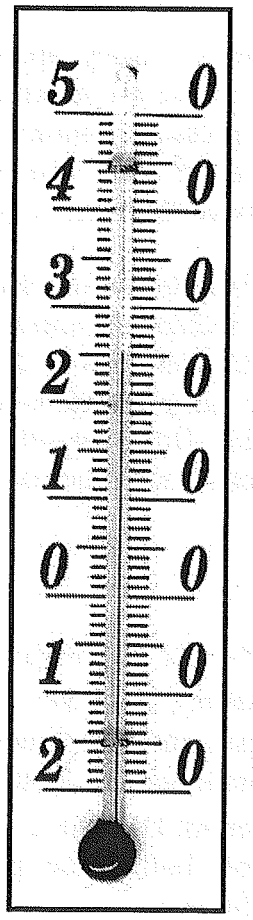


Figure 4.1: A glass thermometer.

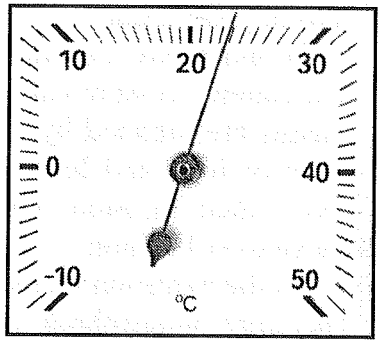


Figure 4.2: A bimetal thermometer.

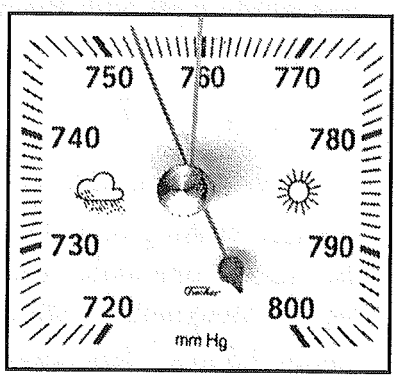


Figure 4.3: An aneroid barometer.

### 4.1.2 Atmospheric pressure

Air is a tangible material substance and as a result has mass. Any object with mass is influenced by the universal force known as gravity. Gravity shapes and influences all atmospheric processes. It causes the density and pressure of air to decrease exponentially as one moves away from the surface of the Earth.

The weight of the earth's atmosphere pressing on each unit of surface constitutes atmospheric pressure, which is 101 325 Pa at the sea level. In other commonly used units (although not accepted by the International System of Units), it equals 760 mm Hg (or 760 Torr). Pressure 101.325 kPa is the amount of pressure that can lift water approximately 10.3 m. Thus, a diver at a depth 10.3 m under water in a fresh-water lake experiences a pressure of about 200 kPa (100 kPa for the air and 100 kPa for the water).

Atmospheric pressure varies widely on Earth, and these variations are important in studying weather and climate. Atmospheric pressure shows a diurnal (twice-daily) cycle caused by the tides. This effect is very strong in tropical zones, and almost zero in polar areas.

In most circumstances atmospheric pressure is closely approximated by the hydrostatic pressure caused by the weight of air above the measurement point. Air masses are affected by the general atmospheric pressure within the mass, creating areas of high and low pressure. Low pressure areas have less atmospheric mass above their location, whereas high pressure areas have more atmospheric mass above their location.

As the elevation increases, there are exponentially fewer and fewer air molecules. Therefore, atmospheric pressure decreases with increasing altitude at a decreasing rate. The barometric formula (below) is used to model how the pressure of air changes with the altitude. It is based on the simplifying assumption that the temperature does not depend on the altitude. However, this formula agrees reasonably well with the actual pressure variations above the earth's surface up to a height of 140 km:

$$p = p_o \cdot e^{\frac{M \cdot g \cdot z}{R \cdot T}},$$

where $p$ [Pa; kPa] is the atmospheric pressure, $p_0$ [Pa; kPa] is the atmospheric pressure at the ground level (mean sea level pressure), $M$ ($M = 0.029\ \mathrm{kg \cdot mol^{-1}}$) is the mass of one mole of air, $R$ is the universal gas constant (Appendix D), $T$ [K] is the thermodynamic temperature, $g$ is the gravitational acceleration due to gravity (about $9.8\ \mathrm{m \cdot s^{-2}}$ depending on the location; Appendix D), and $z$ [m] is the vertical height above the earth's surface. As a rule of thumb, the pressure decreases approximately 1 % for every 80 m increase in altitude.

Atmospheric pressure serves as a reference level for other types of pressure measurements. One of these is **gauge pressure**. Gauge pressure is either positive or negative, depending on its level above or below the atmospheric pressure reference. Atmospheric pressure represents zero gauge pressure. For example, an ordinary tire gauge is showing the excess pressure above atmospheric. In other

words, what the gauge shows is the difference between atmospheric pressure and the pressure of the air pumped into the tire.

Mean sea level pressure is the pressure at the sea level or (when measured at a given elevation on land) the station pressure reduced to the sea level assuming an isothermal layer at the station temperature. This is the pressure normally given in weather reports on radio, television, and newspapers. When barometers at home are set to match the local weather reports, they measure pressure reduced to the sea level, not the actual local atmospheric pressure.

**A barometer** is the name of any instrument that measures the air pressure. The most common type of a barometer used at home is an **aneroid barometer** (Figure 4.3). Inside this instrument is a small, flexible metal box called an **aneroid cell**. In the construction of the device, a vacuum is created inside the box so that small changes in outside air pressure cause the box to expand or contract. The size of the aneroid cell is then calibrated and any change in its volume is transmitted by springs and levers to an indicating arm that points to the corresponding atmospheric pressure. It is often necessary to tap this type of barometer before reading it. This loosens the needle and causes the reading to jump in the direction of the most recent changes in pressure.

### 4.1.3 Air humidity

There is always water vapour in the bottom layers of atmosphere. If we have some air of volume $V$ [$m^3$] containing the water vapour of mass $m$ [g], the quantity $H_{abs}$ defined by formula:

$$H_{abs} = \frac{m}{V}$$

is called the **absolute air humidity**, [$g \cdot m^{-3}$]. It equals the water vapour density in the air.

The amount of water vapour in the air is usually overheated vapour. Its amount cannot be increased arbitrarily. It may become saturated when reaching other temperature (for example, by additional evaporation). The absolute air humidity may reach the maximal value $H_{max}$ (**maximal air humidity**, [$g \cdot m^{-3}$]) at given temperature when it equals the density $\rho$ of saturated water vapour at the same temperature.

The water evaporation from surfaces of living subjects does not depend only on absolute air humidity, but also on the fact how the state of water vapour in the air differs from the state of saturated water vapour. This is described by the relative air humidity $H_{rel}$ [%] defined by the formula:

$$H_{rel} = \frac{H_{abs}}{H_{max}} \cdot 100 \ .$$

The relative air humidity of dry air equals 0 %, of air saturated with water vapour it equals 100 %. The most suitable relative air humidity for human life and the best working capability is between 50 % and 70 %.

People may complain that humid air feels "heavy", but in fact, the more moisture in the air, the lower the air density. That is because the molecular weight of water vapours is lower than the average molecular weight of the constituents of dry air. The discomfort associated with high humidity is somewhat analogous to the wind chill effect, where high winds make people feel colder. As relative humidity increases, so does human discomfort. For example, at an air temperature of 32°C and 50 % relative humidity, the air "feels" as if it were 36°C. The reason is that the moister the air, the larger the resistance to moisture loss (and therefore to heat loss via evaporation) from the human body to the air, because the air is closer to saturation. The humidity effect on comfort operates at low temperatures as well: people are more comfortable in cold air when humidity is high than low.

When decreasing the air temperature with overheated water vapour (without changing the absolute air humidity) it becomes saturated at the temperature of the **dew point** $t_r$. If the temperature continues decreasing, the vapour becomes liquid and dew, fog or clouds arise. If $t_r < 0°C$, hoarfrost or snow arises. At very high temperatures, air is rarely if ever close to saturation because the saturation vapour pressure is very large.

**A microclimate** is the distinctive climate of a small-scale area, such as a garden, park, valley or part of a city. The weather variables in a microclimate, such as temperature, rainfall, wind or humidity, may be subtly different from the conditions prevailing over the area as a whole and from those that might be reasonably expected under certain types of pressure or cloud cover.

**A hair hygrometer** (Figure 4.4) allows measuring relative air humidity directly. It is based on the phenomenon that the human hair cleared from fats changes in length in response to changes of air humidity, with the hair becoming longer when the relative humidity increases. To make the change more marked, the hair bunch is fixed to a string on one end and goes over a small padlock connected with a handle on the next end. The handle shows the relative air humidity on a scale.

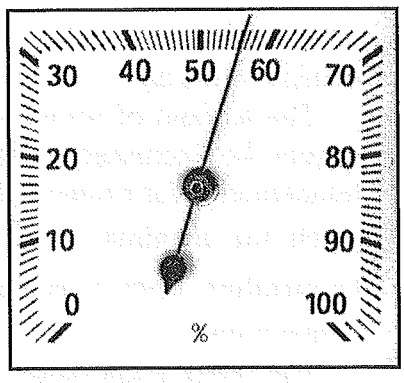


Figure 4.4: A hair hygrometer.

**An Assman's ventilated psychrometer** consists of two identical thermometers (Figure 4.5). Thermometers are placed in tubes open at the bottom. At the top, both tubes are connected to a common one directed to a fan. A spring-driven motor, wound up by a key, operates the fan that draws air across the bulbs of two thermometers.

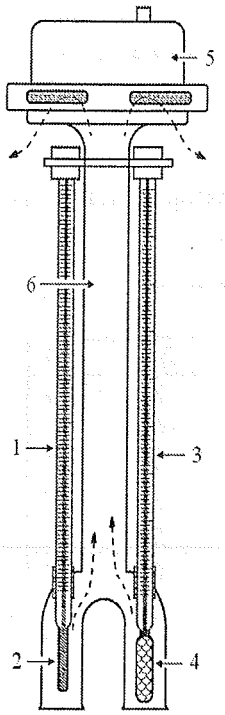


Figure 4.5: Assman's psychrometer:
1 – dry-bulb thermometer
2 – dry-bulb
3 – wet-bulb thermometer
4 – wet-bulb with muslin wick
5 – fan
6 – connecting tube
(modified after: http://www.kar.elf.stuba.sk/predmety/sap/cvika/vlhkost.pdf)

One thermometer bulb remains dry during measurement (dry-bulb thermometer), while the other bulb is covered with a muslin wick, which is moistened with distilled water (wet-bulb thermometer). This wet-bulb thermometer is cooled by evaporation (due to the air stream passing over it, which is generated by the fan) to a value below the temperature shown by the dry-bulb thermometer. The computation of the humidity is carried out by comparing the two readings of the thermometers, since the difference between them depends on humidity and pressure (the pressure is measured independently using a barometer). The evaporation rate depends on relative air humidity (assuming that next conditions remain unchanged). The lower is the relative air humidity in the vicinity, the earlier the water evaporates, and the greater is the temperature difference between both thermometers. The temperature difference

$$\Delta t = t_d - t_w$$

is given in connected psychrometric tables (Table 4.1) and the relative air humidity is established. It can be also established from the connected psychrometric chart (Figure 4.6) using measured "wet" and "dry" temperature ($t_w$ and $t_d$, respectively).

On the psychrometric chart, the horizontal axis displays the dry bulb thermometer temperature (0°C - 50°C). Constant temperatures are given on vertical lines. The wet bulb thermometer scale is on the left side and is curved (values 0°C - 30°C are marked). Constant temperatures are given on oblique lines moderately descending from left to right. At the same time, the most left upper curve corresponds to 100% relative air humidity ($t_d = t_w$). The scale for relative air humidity starts at the upper left corner and continues to the right and then

downward. Values from 100% to 10% are marked and are connected to the curves that start on the left edge (between the two zeroes) and continue to the right and upward.

Table 4.1: Psychrometric table - values of relative humidity using Assman's psychrometer due to temperature difference $\Delta t$

| $t_d$ | $\Delta t = t_d - t_w$ [°C] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| [°C] | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4 |
| 0 | 91 | 82 | 73 | 65 | 56 | 48 | 39 | 31 |
| 2 | 92 | 84 | 76 | 68 | 60 | 52 | 45 | 37 |
| 4 | 92 | 85 | 78 | 70 | 63 | 56 | 49 | 42 |
| 6 | 93 | 86 | 79 | 73 | 66 | 60 | 53 | 47 |
| 8 | 94 | 87 | 87 | 75 | 69 | 63 | 57 | 51 |
| 10 | 94 | 88 | 82 | 76 | 71 | 65 | 60 | 54 |
| 12 | 94 | 89 | 83 | 78 | 73 | 68 | 62 | 57 |
| 14 | 95 | 90 | 85 | 79 | 75 | 70 | 65 | 60 |
| 16 | 95 | 90 | 86 | 81 | 76 | 71 | 67 | 62 |
| 18 | 95 | 91 | 86 | 82 | 77 | 73 | 68 | 64 |
| 20 | 96 | 91 | 87 | 83 | 79 | 74 | 70 | 66 |
| 25 | 96 | 92 | 88 | 84 | 80 | 77 | 73 | 70 |
| 30 | 96 | 93 | 89 | 86 | 83 | 79 | 76 | 73 |
| 35 | 97 | 93 | 90 | 87 | 84 | 81 | 78 | 75 |
| 40 | 97 | 94 | 90 | 87 | 84 | 81 | 79 | 76 |

| $t_d$ | $\Delta t = t_d - t_w$ [°C] | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [°C] | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 14 | 16 | 18 | 20 |
| 2 | 22 | | | | | | | | | | |
| 4 | 29 | | | | | | | | | | |
| 6 | 35 | 23 | | | | | | | | | |
| 8 | 40 | 29 | 18 | | | | | | | | |
| 10 | 44 | 34 | 24 | | | | | | | | |
| 12 | 48 | 38 | 29 | 20 | 11 | | | | | | |
| 14 | 51 | 42 | 33 | 25 | 17 | 9 | | | | | |
| 16 | 54 | 45 | 37 | 30 | 22 | 15 | | | | | |
| 18 | 56 | 48 | 41 | 34 | 26 | 20 | | | | | |
| 20 | 59 | 51 | 44 | 37 | 30 | 24 | 12 | | | | |
| 25 | 63 | 57 | 51 | 44 | 38 | 33 | 22 | 12 | | | |
| 30 | 67 | 61 | 55 | 50 | 44 | 39 | 30 | 21 | 13 | 5 | |
| 35 | 69 | 64 | 58 | 53 | 48 | 44 | 35 | 27 | 20 | 13 | 7 |
| 40 | 71 | 66 | 61 | 56 | 51 | 47 | 39 | 32 | 26 | 20 | 17 |

$t_d$: temperature of the dry-bulb thermometer (in the first column)

$t_w$: temperature of the wet-bulb thermometer

(modified after: Brož et al., 1980)

Table 4.2: Temperature dependence of saturated vapour density and pressure

| Temperature $t$ [°C] | Saturated vapour density $\rho$ [g·m$^{-3}$] | Saturated vapour pressure $p$ [Pa] | Temperature $t$ [°C] | Saturated vapour density $\rho$ [g·m$^{-3}$] | Saturated vapour pressure $p$ [Pa] |
|---|---|---|---|---|---|
| 1 | 4.20 | 656 | 16 | 13.63 | 1817 |
| 2 | 4.57 | 705 | 17 | 14.48 | 1937 |
| 3 | 4.95 | 757 | 18 | 14.36 | 2062 |
| 4 | 6.37 | 813 | 19 | 16.29 | 2196 |
| 5 | 6.80 | 872 | 20 | 17.29 | 2337 |
| 6 | 7.27 | 935 | 21 | 18.32 | 2486 |
| 7 | 7.79 | 1005 | 22 | 19.41 | 2642 |
| 8 | 8.28 | 1072 | 23 | 20.57 | 2809 |
| 9 | 8.83 | 1148 | 24 | 21.78 | 2984 |
| 10 | 9.40 | 1227 | 25 | 23.04 | 3168 |
| 11 | 10.01 | 1312 | 26 | 24.37 | 3361 |
| 12 | 10.66 | 1401 | 27 | 24.76 | 3565 |
| 13 | 11.35 | 1497 | 28 | 27.23 | 3780 |
| 14 | 12.06 | 1597 | 29 | 28.75 | 4005 |
| 15 | 12.82 | 1704 | 30 | 30.35 | 4242 |

(modified after: Brož et al., 1980)

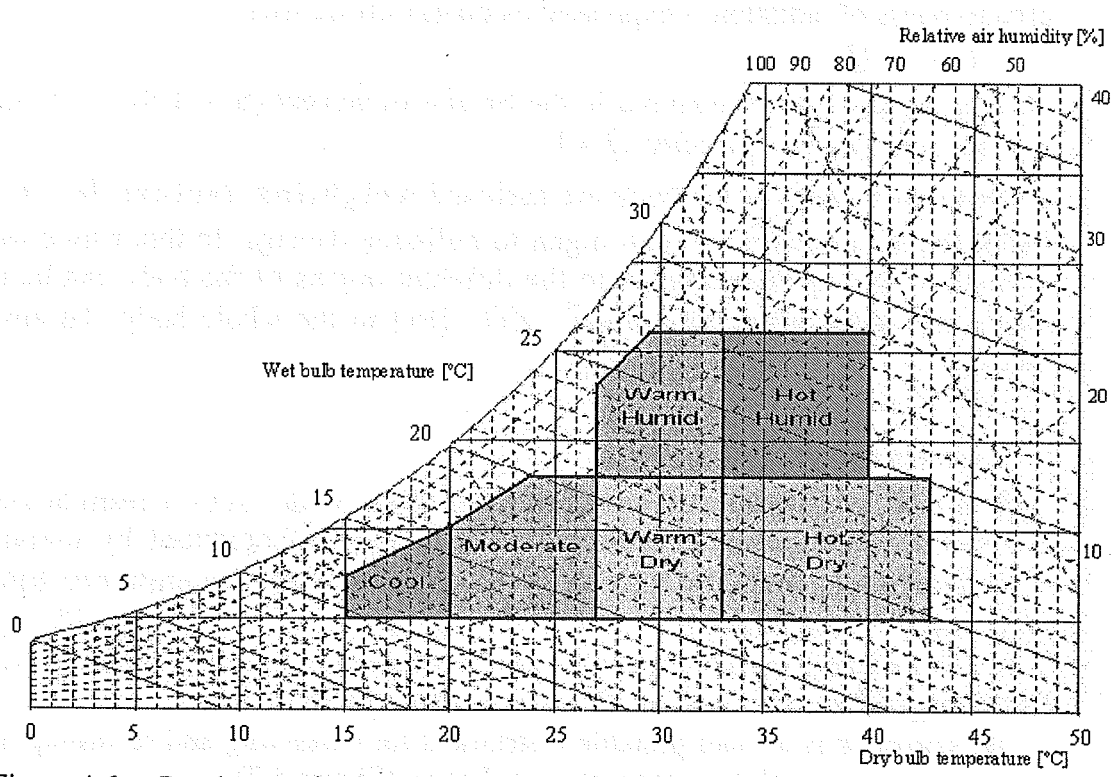


Figure 4.6: Psychrometric chart.
(modified after: http://www.arch.hku.hk/~cmhui/teach/fig6.jpg)

## 4.2. Measurement of Background Radiation – Dosimetry

As any ionising radiation passes through body tissue or any other matter, it creates ionised atoms (and molecules) and liberates fast electrons. The fast electrons in turn create more ionisation until all the energy of the incident radiation is spent. Many of the ionised atoms and molecules recombine with electrons before any permanent harm is done or else damage that is done to body cells is repaired without lasting effect. Some radiation damage however causes permanent chemical change, which can lead to harmful **biological effects** (carcinogenic and mutagenic).

Biological effects are found to be directly related to the amount of energy that is deposited by the radiation per unit mass of tissue - **the absorbed radiation dose** $D$. Assessments of the adverse biological effects of ionising radiation must take account of the type of radiation involved and the nature of the organ of the body that is exposed. However, it is accepted that the biological effects in any tissue are proportional to the amount of energy absorbed per unit mass of tissue. Absorbed dose is measured in the SI unit of the **gray** (Gy): 1 Gy = 1 J · kg$^{-1}$.

Different radiation has different biological effectiveness for the same amount of energy absorbed. To consider this, each radiation type has been given a radiation-weighting factor, $W_R$ (called a **quality factor,** $Q$). Different biological effectiveness of radiation is expressed as **equivalent dose** $H$:

$$H = D \cdot Q.$$

Equivalent dose is measured in the SI unit of **sievert** (Sv): 1 Sv = 1 J · kg$^{-1}$. For gamma rays and beta particles $Q = 1$.

Different organs of the body are assigned **weighting factors** $W_T$, to reflect the different sensitivity of each organ to radiation damage. In this way a weighted average of the equivalent doses to the different organs of the body can be used as a measure of the **effective dose** $ED$ [Sv] to the whole body. To any organ holds:

$$ED = H \cdot W_T.$$

For the whole body $W_T = 1$.

**Exposure** $X$ is a measure of ionisation of air produced by a beam of radiation. It expresses the electric charge (positive or negative) produced by ionisation in volume $dV$ of mass $dm$. Exposure is given in units **coulomb per kilogram.** (Older unit: 1 R = 1 röntgen ≅ 2.6 ·10$^{-4}$ C · kg$^{-1}$.)

**Background radiation** is naturally occurring ionising radiation, including cosmic rays and radiation from naturally occurring radioactive materials.

**Dosimeter** is a small portable instrument for measuring and recording the total accumulated personal dose of ionising radiation (Figure 4.7).

The dosimeter is based on gas ionisation. It is turned on and off at the bottom. Sound signal of registration can be chosen at the top.

The regime switch ("REŽIM") is used to choose between registration of beta particles (β) and gamma rays (γ). When in position "γ", equivalent dose rate $\dot{H}$ or exposure rate $\dot{X}$ can be measured. When in position "β", number of beta particles $\varphi$ per unit area and unit time or specific activity of a radioactive source ("Am") can be measured. With the metal cover on the back, only gamma rays are measured; when removed, beta particles and gamma rays are measured at once.

The time switch ("ČAS") aloowing the choice between 20 s or 200 s stands for measurement of $\dot{H}$, $\dot{X}$, and $\varphi$. Specific activity has to be measured for at least 10 min.

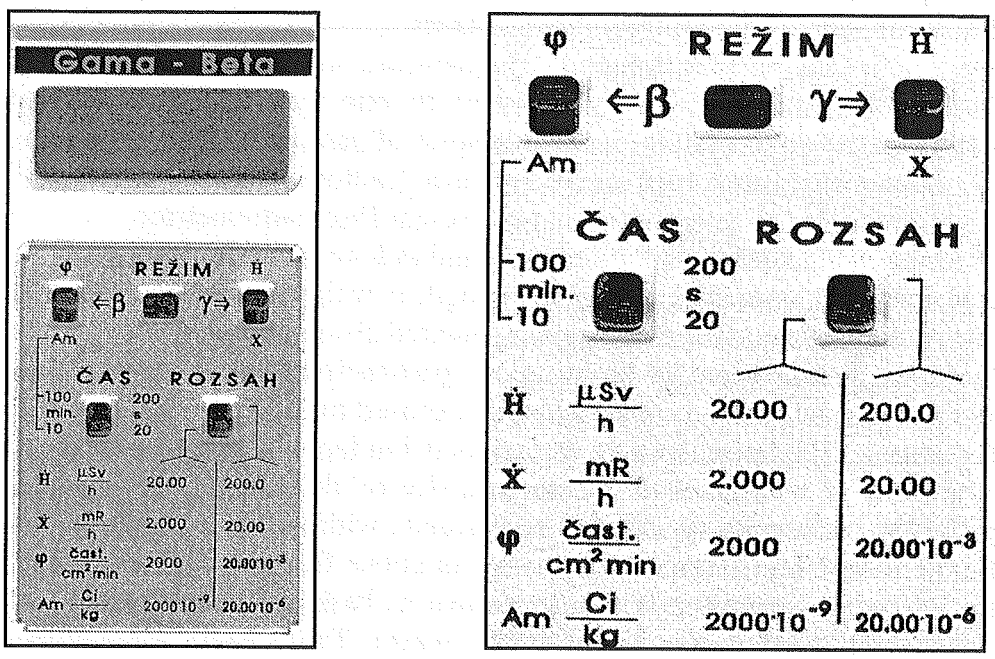


Figure 4.7: The digital Gamma-Beta dosimeter (frontal view and detail).

## 4.3. Measurement of Illuminance

**Photometry** is the science of measurement of **light**, in terms **of** its **perceived brightness** to the human eye. It is distinct from radiometry, which is the science of measurement of light in terms of absolute **power**.

The human eye is not equally sensitive to all wavelengths of light. Photometry attempts to account for this by weighting the measured power at each wavelength with a factor that represents how sensitive the eye is at that wavelength. The standardised model of the eye's response to light as a function of wavelength is given by the luminosity function. The eye has different responses as a function of wavelength when it is adapted to light conditions (**photopic vision**) and dark conditions (**scotopic vision**). Photometry is based on the eye's photopic response, and so photometric measurements will not accurately indicate the perceived brightness of sources in dim lighting conditions.

There are two parallel systems of quantities known as **photometric** and **radiometric** quantities. Every quantity in one system has **an analogous quantity** in the other system. Some examples of parallel quantities include:

- Luminance (photometric) and radiance (radiometric)
- Luminous flux (photometric) and radiant flux (radiometric)
- Luminous intensity (photometric) and radiant intensity (radiometric)

In photometric quantities every wavelength is weighted according to how visible it is, while radiometric quantities use unweighted absolute power.

Photometric measurement is based on **photodetectors**, devices (of several types) that produce an electric signal when exposed to light.

In photometry, **illuminance** is the total luminous flux incident on a surface, per unit area. It is a measure of the intensity of the incident light, wavelength-weighted by the luminosity function to correlate with human brightness perception. Similarly, **luminous emittance** is the luminous flux per unit area emitted from a surface. Luminous emittance is also known as **luminous exitance**. These are both measured in lux or lumen per square metre (Table 10.6). Recommended and typical values of illuminance for different places are given in Tables 4.3 and 4.4.

Table 4.3: Recommended values of illuminance

| Room type or work type | Illuminance [lx] |
|---|---|
| Corridors in buildings | 50 |
| Common household work | 100 |
| Classrooms | 200 |
| Laboratories - fine work | 500 |
| Very fine work - drawing | 1 000 |
| Extremely fine work – jewellery | 2 000 |

Table 4.4: Typical values of illuminance

| Illuminance [lx] | Example |
|---|---|
| 0.000 05 | Starlight |
| <1 | Moonlight |
| 10 | A candle at a distance of 30 cm |
| 400 | A brightly lit office |
| 400 | Sunrise or sunset on a clear day |
| 1 000 | Typical TV studio lighting |
| 32 000 | Sunlight on an average day (minimal) |
| 100 000 | Sunlight on an average day (maximal) |

Illuminance was formerly often called **brightness**, but this lead to confusion with other uses of the word. The term "brightness" should never be used for quantitative description, but only for nonquantitative references to physiological sensations and perceptions of light.

**Luxmeter** is used for illuminance measuring. The measured light (luminous flux) falls on a selenium barrier of a photocell or a semiconductor photoelectric resistor. A microampere-meter is attached to this detector and is calibrated directly in luxes.

The luxmeter PU 150 (Figure 4.8) allows switching to different measuring ranges (10 lx - used only with a photo-resistor; then 200 lx, 1000 lx, and 5000 lx). The measuring probe is provided with a photocell. The spectral sensitivity of the photocell is corrected with a colour filter for spectral sensitivity of the eye, by which the necessity of using conversion factors for light sources of different spectral emission is avoided.

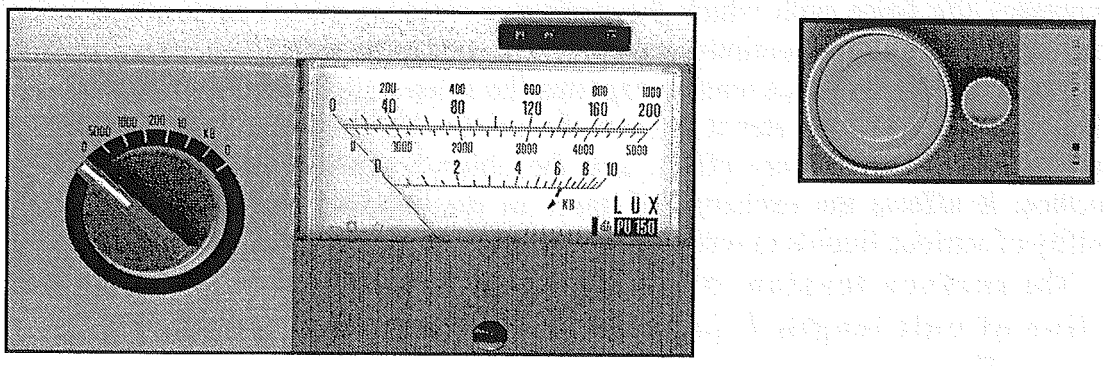


Figure 4.8: Front view of the luxmeter PU 150 (left) and its measuring probe (right).

# 5. Evaluation of Physical Properties of Liquids

## 5.1. Measurement of Surface Tension – Stalagmometric Method

The molecules of a liquid have widely varying speeds and kinetic energies. However, the intermolecular forces between liquid molecules, unlike those between gas molecules, are quite large and bind the liquid molecules together in a compact mass. Hence liquids do not expand to fill a container as gases do, but settle down, under the influence of gravity, to the bottom of any container in which they are placed. The intermolecular forces are not as strong in liquids as in solids and do not restrict the molecules to fixed positions in space. As a result, if external forces are applied to a liquid, it can be easily made to change shape. By applying sufficiently large forces, it is possible to split up the liquid into pieces, as when water drips from a faucet. Here the force of gravity acting on the unsupported liquid leaving the faucet pulls the water apart into small droplets that gradually change their shape as they fall, owing to the surface tension of the liquid.

A molecule in the interior of a liquid is subjected to attractive forces from the other liquid molecules. Since these forces are in all directions, they cancel out, and there is no net unbalanced force on the molecule. Contrary to inner molecules, molecules at the surface of the liquid experience a net force pulling them back toward the interior of the liquid. As a consequence, a molecule in the surface layer has a higher potential energy than a molecule in the interior of the liquid. Therefore, increasing of the surface layer will increase the potential energy of the liquid. Since the liquid tends to minimise its potential energy so it tends to minimise its surface area. As a result, the surface of a liquid behaves like a stretched membrane and produces what is called the **surface tension** of the liquid. The surface tension expresses the force with which the surface molecules attract each other; it is the tendency of a liquid to minimise the area of its surface by contracting.

The concept of surface tension explains the spherical shape of liquid droplets, the "hanging" of drops of water at the end of a faucet, the behaviour of liquids in narrow capillary tubes – capillary effect, and the ability of insect walking on the water surface. It affects the exchange of gases in the pulmonary alveoli and alters the ability of various liquids to wet another surface.

The **surface tension** $\sigma$ [$N \cdot m^{-1}$] is defined as a **force** $F$ [N] acting **along a line of unit length** $l$ [m] **perpendicular to the surface**:

$$\sigma = \frac{F}{l}$$

For practical reasons, the unit [$mN \cdot m^{-1}$] is used instead of the large unit [$N \cdot m^{-1}$]. The surface tension of liquids is a material constant expressing the amount of energy (or work, [J]) needed to increase the surface of a liquid by 1 $m^2$ (1 $N \cdot m^{-1}$ = 1 $J \cdot m^{-2}$).

The surface tension does not depend on the area of the surface. It decreases with increasing temperature.

A **stalagmometer** is a device used to measure the surface tension of a solution (Figure 5.1). It uses the drop weight method, in which the number and weight of drops are compared to those from a reference liquid.

Non-wettable liquid may not flow out through a small opening. In some cases it forms a hanging drop. The drop will drop down if the gravity force $\vec{F}_g$ [N] is greater than the force of surface tension $\vec{F}_\sigma$ [N] acting along the perimeter of the opening $F_g > F_\sigma$, where

$$F_\sigma = 2 \cdot \pi \cdot r \cdot \sigma$$

and

$$F_g = m \cdot g ,$$

$m$ [kg] is the mass of the drop, $g$ [m·s$^{-2}$] is the gravitational acceleration, $r$ [m] is the radius of the capillary.

The mass of one drop of a liquid can be calculated from the known volume $V$ [m$^3$] of the liquid and the number $N$ of drops involved in this volume:

$$m = \frac{V \cdot \rho}{N} ,$$

where $\rho$ is the liquid density [kg·m$^{-3}$]. After substitution the formulas for forces and one drop mass into the relation $F_g = F_\sigma$ corresponding to the moment just before the drop falls down, we will get:

$$\frac{\sigma \cdot N}{\rho} = \frac{V \cdot g}{2 \cdot \pi \cdot r} .$$

The unknown surface tension (for example, surface tension of a citric acid solution, CA) can be established from a known one - the reference (for example, surface tension of water, W) as the liquid volume in the stalagmometer is the same in both cases ($V_{CA}$ is the volume of citric acid solution, $V_W$ is the volume of water):

$$V_{CA} = V_W : \qquad \frac{\sigma_{CA} \cdot N_{CA}}{\rho_{CA}} = \frac{\sigma_W \cdot N_W}{\rho_W} .$$

The unknown surface tension is then calculated from the relation

$$\sigma_{CA} = \sigma_W \cdot \frac{\rho_{CA}}{\rho_W} \cdot \frac{N_W}{N_{CA}} .$$

Table 5.1: The temperature dependence of the surface tension and distilled water density.

| Temperature $t$ [°C] | Surface tension $\sigma$ [mN·m] | Density $\rho$ [kg·m$^{-3}$]. |
|---|---|---|
| 15 | 73.49 | 999.96 |
| 16 | 73.34 | 999.94 |
| 17 | 73.19 | 999.90 |
| 18 | 73.05 | 999.85 |
| 19 | 72.90 | 999.78 |
| 20 | 72.75 | 998.20 |
| 21 | 72.59 | 997.99 |
| 22 | 72.44 | 997.77 |
| 23 | 72.28 | 997.54 |
| 24 | 72.13 | 997.30 |
| 25 | 71.97 | 997.05 |

(modified after:
http://www.fpharm.uniba.sk/fileadmin/user_upload/english/Fyzika/The_surface_tension_of_liquids_measured_with_the_stalagmometer.pdf)

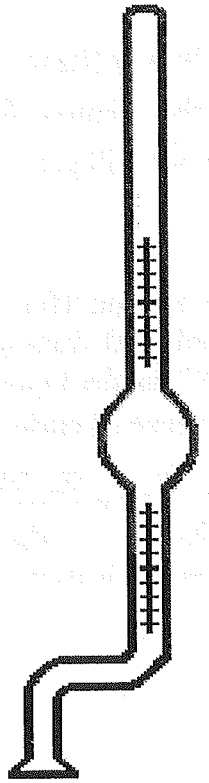

Figure 5.1: A schematic drawing of Traube's stalagmometer.
Above and below the widened part ("the reservoir") of the capillary is a scale that is used to establish the exact volume.

## 5.2. Measurement of Viscosity – Viscometry

**Viscosity** describes a fluid's internal resistance to flow and may be thought of as a measure of fluid friction.

Viscosity depends on other physical parameters such as pressure and temperature. Viscosity of low molecular weight fluids, including water and many aqueous solutions, liquid metals, organic compounds, and silicones, is independent on velocity of flowing. These fluids are called **Newtonian fluids**. Other fluids, with viscosity dependent on velocity of flowing (i. e. blood), are called **non-Newtonian fluids**.

In liquids, viscosity is independent of pressure (except for very high pressure) and depends on temperature (falls with increasing temperature). For diagnostic purposes, measurement of viscosity takes place mainly in haematology.

Two quantities are used to describe internal friction of fluids: **dynamic viscosity** $\eta$ [Pa·s] and **kinematic viscosity** $\nu$ [$m^2 \cdot s^{-1}$], their mutual relation is:

$$\nu = \frac{\eta}{\rho}$$

where $\rho$ [$kg \cdot m^{-3}$] is the density of the measured liquid .

Dynamic viscosities of liquids are typically several orders higher than dynamic viscosities of gases. The viscosity of water is $8.90 \cdot 10^{-4}$ Pa·s at about 25°C.

Viscosity is measured with various types of viscometer, typically at 25 °C. Different types of viscometers are used, the two main types are the tube and the rotational instruments. We describe the first type, based on measuring of the rate of flow through tubes due to a known pressure difference.

For tube viscometers (Figure 5.2) the **Hagen-Poisseuille law** applies, which describes the flow of liquids of the Newtonian type through a tube. The amount of liquid (volume $V$ [$m^3$]) flowing per unit time $t$ [s] through a narrow tube is proportional to the 4th power of the tube radius $r$ [m] and to the pressure difference $\Delta p$ [Pa] appearing at the tube ends, and indirectly proportional to the tube length $l$ [m] and to the dynamic viscosity of the liquid $\eta$:

$$V = \frac{\pi \cdot r^4 \cdot \Delta p}{8 \cdot \eta \cdot l} \cdot t$$

If the tube is in vertical position, the pressure difference is equal to the hydrostatic pressure:

$$\Delta p = h \cdot \rho \cdot g$$

where the height $h$ is equal to the length of the tube ($h = l$). Hagen-Posseuille low enables calculation of time $t$, in which the given volume $V$ of the measured liquid passes through the tube. All parameters of the tube can be measured and are represented by the **constant of the viscometer** $K$, the only unknown is the viscosity. Thus, if we measure the time, in which the given volume of liquid passes through the tube, we can calculate the kinematic viscosity $\nu$:

$$\nu = K \cdot t$$

where $K$ is a constant value for a given viscometer depending on its geometry.

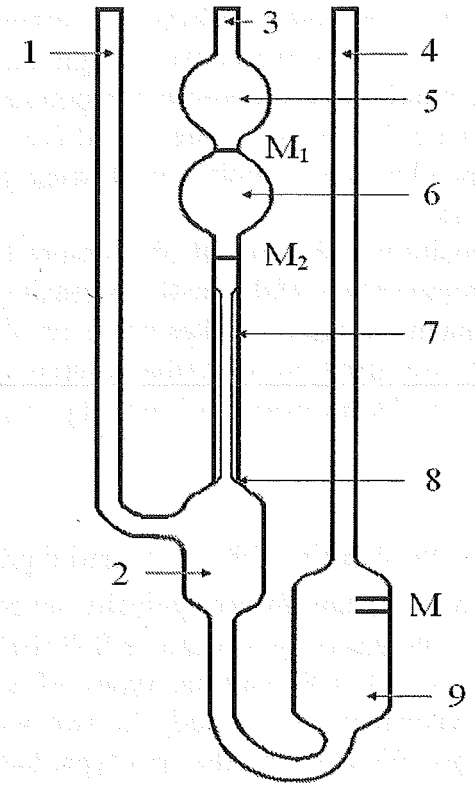


Figure 5.2: A schematic drawing of an Ubbelohde´s viscometer
1 - deaerating tube
2 - exhaust bulb
3 - deaerating aperture
4 - filling tube
5 - storage bulb
6 - measurement bulb
7 - capillary
8 - exhaust aperture of the capillary
9 - collecting bulb
(modified after:
http://www.kchsz.sjf.stuba.sk/pedagogika/latkove_vlastnosti/Meranie%20reologickej%20vlastnosti%20kvapaliny.doc)

# 6. Evaluation of Concentration of Solutions Using Optical Methods

## 6.1. Refractometry

The **absolute index of refraction** $n$ of a medium is defined as the ratio

$$n = \frac{c}{v},$$

where $c$ is the **speed of light in vacuum** (Table 10.9) and $v$ is the **speed of light in a given medium**. In general, the ray passes from the medium A to the medium B. The behaviour of light rays on the boundary of two optically isotropic media, which have refraction indexes $n_A$ and $n_B$ is described by the Snell's law

$$\frac{\sin\alpha}{\sin\beta} = \frac{n_B}{n_A}.$$

The **angle of incidence** $\alpha$ is the angle between the incident ray and the normal of incidence in medium A. The **angle of refraction** $\beta$ is the angle between the refracted ray and the normal in medium B. The ratio

$$n_{AB} = \frac{n_B}{n_A}$$

is called the **relative index of refraction.**

The medium of higher refraction index is denoted as **optically denser** medium, in comparison to the **optically rarer** medium. A ray passing from the optically denser medium into the optically rarer medium refracts away the normal (Figure 6.1). The angle $\beta$ is larger than the angle $\alpha$. The angle of incidence for which $\beta = 90^o$ is called **the critical angle** ($\varepsilon$).

Most refractometry methods are based on the measurement of this angle. If we know the refraction index of the optically denser medium, the whole measurement is reduced to the measurement of the critical angle as $\sin\beta = 1$. According to the Snell's law

$$n_B = n_A \cdot \sin\alpha.$$

If the angle of incidence is larger than the critical angle, **total reflection** of light occurs. This phenomenon is utilised in fibre-optic instruments (endoscopes). Conversely, rays from the optically rarer medium, which strike the interface under the angle of 90° (parallel to the interface), are refracted into the optically denser medium under the critical angle.

**Refractometers** are instruments for the measurement of refraction index, mainly in liquids. They are also often used for an indirect determination of concentration of organic or inorganic substances.

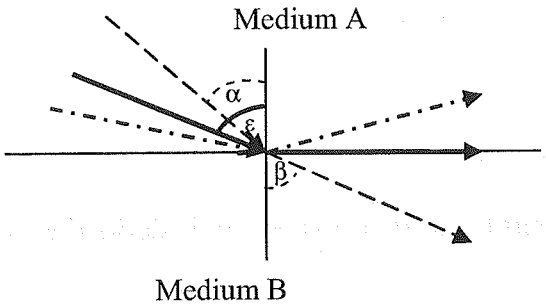


Figure 6.1: Light refraction and total light reflection.
The dashed arrow corresponds to the light refraction from the optically denser to the optically rarer medium; the solid line arrow corresponds to the critical angle; the dash-dot arrow corresponds to the total reflection. For next explanation see text.

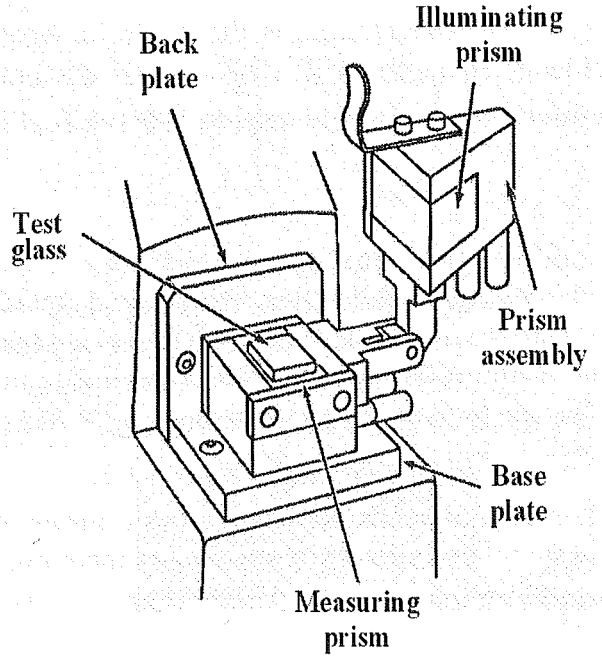


Figure 6.2: A detailed drawing of the inside view of the double-prism Abbe refractometer. (modified after: http://www.chemistry.oregonstate.edu/)

The **double-prism Abbe refractometer** is the most commonly used instrument based on the principle of critical angle measurement (Figures 6.1 and 6.2). It comprises illumination and observation systems (Figure 6.3: Mi – mirror, T -

telescope) and between them, two prisms (I – illuminating, M - measuring). Between the hypotenuse walls of these prisms is a narrow gap filled by the liquid to be measured. The hypotenuse wall of the illuminating prism is dull; the hypotenuse wall of the measuring prism is polished. Both prisms are made of flint glass, which has a high refraction index (approximately $n \cong 1.6$).

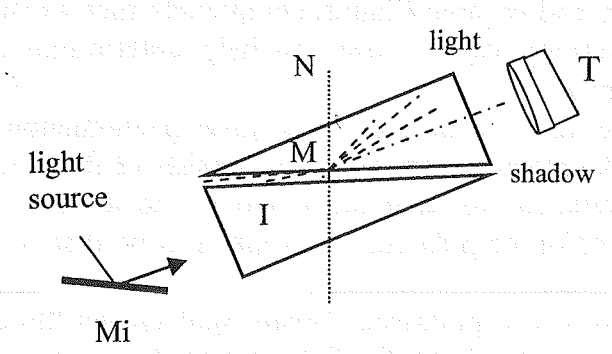


Figure 6.3: A schematic drawing how the light passes through both prismas.

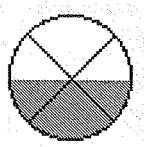

Figure 6.4: Reflection border-line adjusted between light and dark regions.

**Measurement in transmitted light:** Light passes through the short wall of the illuminating prism to a dull surface where it is scattered. Then it passes into the measured solution and is refracted into the measuring prism. The maximum refraction angle equals the critical angle ε (Figure 6.1). A boundary between light and shadow can be observed using a telescope focused in infinity (right eyepiece). It is found by changing the orientation of the prisms and the telescope axis (by rotating the prism pair, in practice). The rotation mechanism (left adjustment hand wheel) is coupled with the instrument scale, which is observed in another optical system. When the boundary is aligned with cross hairs (Figure 6.4), the value of refraction index can be read from the scale through the left eyepiece.

Refractive index values depend directly on the concentration $c$ of solutions. If very high accuracy is required, the metallic housing of the prism pair can be maintained at constant temperature by water circulated from a thermostatically controlled water bath. Universal Abbe refractometers use a source of white (polychromatic) light, which causes that the boundary between light and shadow is coloured and blurry due to optical dispersion. Compensation prisms (compensator dial – right adjustment hand wheel) can remove this aberration.

## 6.2. Polarimetry

**Optical activity** is the ability of a chiral molecule to rotate the plane of plane-polarised light. It can be measured using a **polarimeter**. **Optically active** materials can change the plane of polarisation of a light beam. This process comes about because of the molecular structure of these materials, and has been observed in crystalline materials such as quartz and organic (liquid) compounds such as sugar solutions. The degree of optical activity can be used to help determining the molecular structure of these compounds.

**Polarisation**, like interference and diffraction, is a wave phenomenon. It occurs with transverse waves only. An electromagnetic wave consists of fluctuating electric and magnetic fields, perpendicular to each other and to the direction of propagation. By convention, the direction of polarisation is taken to be that of the electric field vector.

Light from ordinary light sources is not polarised. When light passes through a sample that can rotate the plane of polarised light, the light appears to dim because it no longer passes straight through the polarising filters (Figure 6.5). The amount of rotation is quantified as the number of degrees that the analysing lens must be rotated so that it appears as if no dimming of the light has occurred.

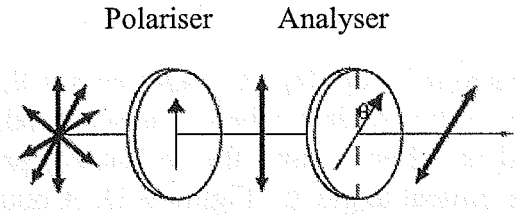


Figure 6.5: Rotation of the plane of polarised light.
The rotated plane is represented by a circle with an arrow.

When rotation is quantified using a polarimeter, it is known as an **observed rotation,** because rotation is affected by path length $l$ [dm], expressing the path the light travels through a sample, and concentration $c$ [%], expressing how much of the sample is present that will rotate the light. When these effects are eliminated, a standard for comparison of all molecules is obtained, which is called the **specific rotation** $[\alpha]$

$$[\alpha]=\frac{100\cdot\alpha}{c\cdot l},$$

where $\alpha$ is the angle of rotation, [°, deg]. **Specific rotation** $[\alpha]$ is defined as the arc of rotation through which the plane of polarized light moves when it is in a light path one decimetre in length passing through a solution containing one gram of a compound per one millilitre of water (1 ml = 1 $cm^3$). It is expressed in angular

degrees (the correct unit is [deg·cm$^2$·g$^{-1}$], but is usually just given as degrees) and is used for monochromatic light of wavelength of the $Na_D$ line (sodium line with wavelength $\lambda = 589$ nm).

Enantiomers (the nonsuperimposable mirror image of a chiral compound) will rotate the plane of polarisation in exactly equal amounts (same magnitude) but in opposite directions: **dextrorotary** designated as *d* or (+), clockwise rotation (to the right), **levorotary** designated as *l* or (-), anti-clockwise rotation (to the left). If only one enantiomer is present, a sample is considered to be **optically pure**. When a sample consists of a mixture of enantiomers, the effect of each enantiomer cancels out, molecule for molecule. For example, a 50:50 mixture of two enantiomers or a racemic mixture will not rotate plane polarised light and is **optically inactive**. A mixture that contains excess of one enantiomer, however, will display a net plane of polarisation in the direction characteristic of the enantiomer that is in excess.

Polarised light has many applications, chief among them being liquid crystal displays. Saccharimetry and photoelasticity are other applications, in which polarised light is used to provide information about test materials. Optical rotation dispersion and circular dichroism are research methods derived from polarimetry and are used to study the structure of biopolymers.

**A polarimeter** is an instrument used to measure the rotation of the plane of polarisation of light as it passes through a substance, especially a liquid or solution. It is an important tool in the analysis of sugar solutions. A simple polarimeter consists of at least a light source, a polarising lens, a sample tube, and an analysing lens (Figure 6.6). The used visual circle polarimeter consists of more parts (Figures 6.7 and 6.8).

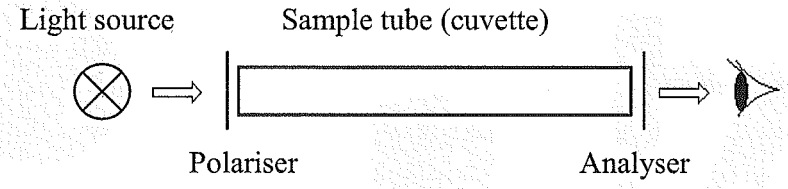


Figure 6.6: A principal scheme of a simple polarimeter.

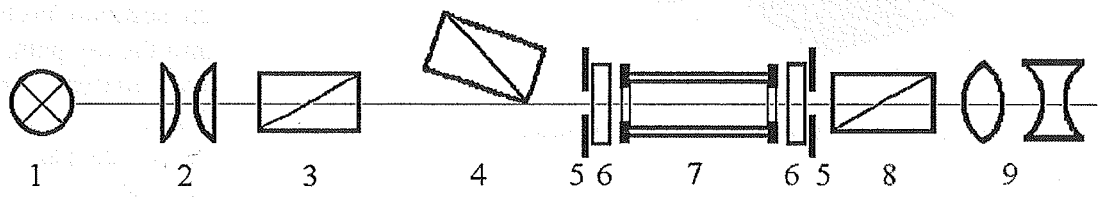


Figure 6.7: A scheme of a laboratory polarimeter:
1 - light source, 2 - condenser, 3 - polariser, 4 - half-shadow device,
5 - diaphragms, 6 - protection plates, 7 - cuvette, 8 - analyser, 9 - telescope
(modified after: Hrazdira and Mornstein, 2001)

The light beam comes from the source to the polariser (Iceland crystal, Nicol prism, Polaroid) that transmits only the plane polarised component of the beam (Figure 6.5). The polarised light then passes further through a cylindrical cuvette containing a solution of optically active substance into the analyser (very similar to the polariser). The analyser can be rotated (Figures 6.5 and 6.7). Its position is indicated on a scale. Rays that pass through the analyser are observed by means of a telescope eyepiece.

The following is valid for a cuvette without any optically active substance. If the polarisation planes of the polariser and the analyser are **parallel**, all rays pass through to the telescope, so the optical field is **bright**; if the polarisation planes of the polariser and the analyser are mutually **perpendicular**, all rays are absorbed, so the optical field is **dark**.

To measure the optical activity of the substance in the cuvette we find an angle of rotation of the analyser against the polariser at which the optical field of the telescope is illuminated either maximally or minimally. The rotation angle is read from the instrument scale.

The accuracy of measurement can be improved by using a **half-shadow method** (Figure 6.9). The human eye resolves differences in brightness between neighbouring areas better than the minimal or maximal brightness of some area. In the half-shadow method a device able to cover one half of the optical field is placed behind the polariser. The polarisation planes of the light rays in the two halves of the field of view make the so-called **half-shadow angle**. The user rotates the analyser until both halves of the optical field are illuminated equally (are in the same half-shadow), indicating that the polarisation plane of the analyser is parallel with the plane of symmetry of light polarisation plates of both halves of the field.

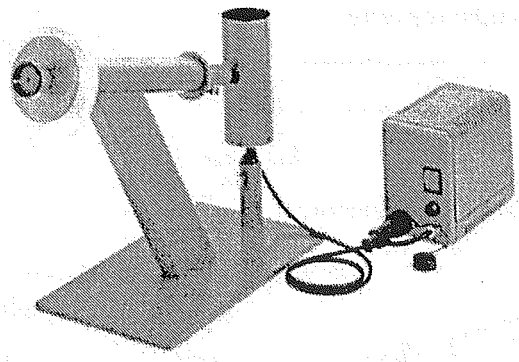

Figure 6.8: A visual circle polarimeter.

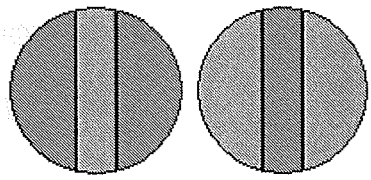

Figure 6.9: Field of view for monochromatic light after the solution has been put into the ray path.
Left: analyser position 0°.
Right: maximum darkness in the middle part of the field of view.

The unknown concentration is then calculated according to the formula:

$$c = \frac{\alpha \cdot 100}{[\alpha] \cdot 1} \ [\%]$$

## 6.3. Spectrophotometry

**Spectroscopy** is the use of the absorption, emission, or scattering of electromagnetic radiation passing through a medium, to study qualitatively or quantitatively the matter or physical processes. The matter (material, medium) can be composed of atoms, molecules, atomic or molecular ions, or solids. The interaction of radiation with matter can cause redirection of the radiation and/or transitions between the energy levels of the atoms or molecules.

**Absorption** is a transition from a lower level to a higher level with transfer of energy from the radiation field to an absorber, atom, molecule, or solid.

**Emission** is a transition from a higher level to a lower level with transfer of energy from the emitter to the radiation field. If no radiation is emitted, the transition from higher to lower energy levels is called **non-radiative decay**.

**Scattering** is a redirection of light due to its interaction with matter. Scattering might or might not occur with a transfer of energy, i.e., the scattered radiation might or might not have a slightly different wavelength compared to the light incident on the sample.

**Absorption spectroscopy** is a widely used physical-chemical analytical method for detecting and determining the concentration of organic and inorganic substances. It measures the absorbed wavelength and the intensity of the absorption.

**Absorption** is attenuation of the radiant flux when light passes through a clear medium. If light of suitable wavelength passes through a sample, part of the energy is transmitted to the molecules. As a result, the emergent beam $I$ has less energy than the incident beam $I_0$ (Figure 6.1). The amount of light absorbed generally follows the Lambert-Beer Law and is therefore proportional to the number of absorbing molecules and the path length traversed ($x$).

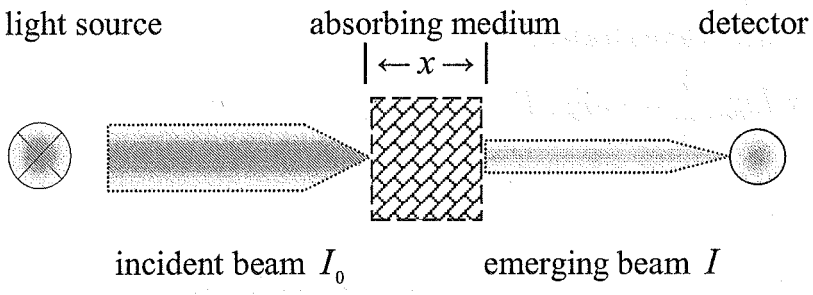


Figure 6.10: A scheme of radiation absorption.

**Absorbance** is the degree of absorption of light or other radiant energy by a medium through which the radiant energy passes. It is expressed as the logarithm of the ratio of energy transmitted through a pure solvent to the intensity of energy

transmitted through the medium. Absorbance varies with wavelength, solution, concentration, and path length.

The **Lambert-Beer law** (also called the Beer-Lambert-Bouguer law or simply Beer's law) is the linear relationship between absorbance and concentration of an absorber of electromagnetic radiation.

When the light passes through a very thin layer of absorbing solution of concentration $c$ [mol·m$^{-3}$] and thickness $dx$ [m], the decrease of intensity $dI$ [W·sr$^{-1}$] is proportional to the intensity of light $I$ [W·sr$^{-1}$] entering this layer, the concentration $c$, and a constant $k$:

$$-dI = k \cdot c \cdot I \cdot dx .$$

If introducing a new constant

$$\varepsilon = k \cdot \log e ,$$

the Lambert-Beer law can be expressed in integral form:

$$I = I_0 \cdot 10^{-\varepsilon \cdot c \cdot x} .$$

The constant $\varepsilon$ [m$^2$·mol$^{-1}$] is the **absorption coefficient**, $x$ [m] is the total thickness of the absorbing layer in the direction of light propagation, $I_0$ [W·sr$^{-1}$] is the entering light intensity, $I$ [W·sr$^{-1}$] is the intensity of light when leaving the layer.

The absorption coefficient is an important substance constant. It is always given for a specified wavelength and can be influenced by many chemical factors. When expressing the concentration in moles per litre, we speak about **molar absorption coefficient.**

The ration of transmitted and incident light intensities is called **transmittance** $T$. It can be also expressed as a percentage. The logarithm to the base 10 of reciprocal value of the transmittance:

$$T = \frac{I}{I_0}$$

is called **absorbance** $A$:

$$A = \log_{10} \frac{1}{T} = -\log_{10} T .$$

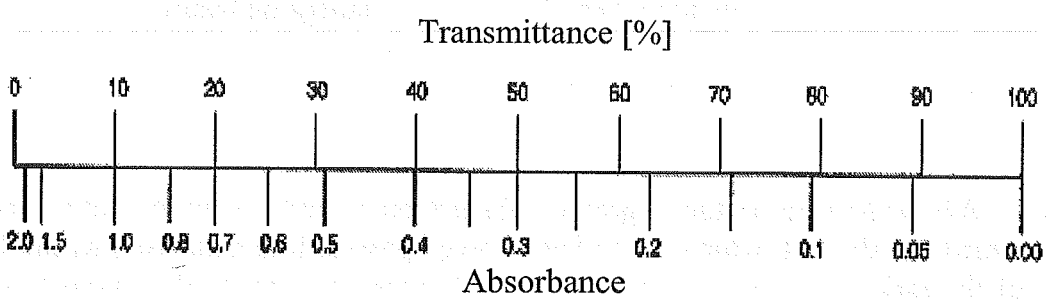


Figure 6.11: Comparison of scales for transmittance and absorbance

After substitution from Lambert-Beer's law, the above formula holds:

$A = \varepsilon \cdot c \cdot x$,

i.e. the absorbance is directly proportional to the concentration of the solution (of dispersed particles) and thickness of the absorbing layer. Transmittance and absorbance are dimensionless quantities (Figure 6.11). If the transmittance $T$ is expressed in per cents as:

$T_{\%} = 100 \cdot T$

the absorbance is expressed as:

$A = 2 - log_{10} T_{\%}$

The linearity of the Beer-Lambert law is limited by chemical and instrumental factors. Non-linearity can be caused, for example, by deviations in absorption coefficients at high concentrations due to electrostatic interactions between molecules in close proximity; by scattering of light due to particles in the sample; fluorescence or phosphorescence of the sample; by changes in refractive index at high analyte concentration; by shifts in chemical equilibrium as a function of concentration; by non-monochromatic radiation - deviations can be minimised by using a relatively flat part of the absorption spectrum such as the maximum of an absorption band; stray light.

**Spectrophotometers** are laboratory instruments used to measure concentration of substances absorbing or emitting infrared, visible, and ultraviolet light, and indirectly for the study of their chemical structure. The spectral dependence of **light absorption**, which reveals the electron shell structure of the substance under study, is measured by **absorption spectrophotometers.** The material under study absorbs some wavelengths of light selectively.

The **Spekol series spectrophotometer** (Figure 6.12) is a typical single-beam spectrophotometer where one beam of light passes through both the measured and reference samples (the cuvettes containing solutions have to be movable). Polychromatic light from the light source passes through a condenser and reflects from a mirror into input slit of the monochromatic (Figure 6.12, parts 4 to 8 and part 12). A lens directs the light onto a reflection optical grating, which decomposes the light into a colour spectrum. A part of this spectrum is projected by an objective into the exit slit of the monochromator. The slit delimits a narrow beam of almost monochromatic light (spectral width 11 nm). The grating can be rotated by means of a screw and determines the wavelength directed into the monochromator exit slit. The light then passes through a cuvette containing the studied solution. A portion of light is absorbed when passing through the cuvette. The intensity of the transmitted light is measured by a photocell. The photocurrent originating in the detector is amplified and processed. The value of absorbance is displayed at a scale with the needle. The intensity of the light transmitted through the reference solution (often the pure solvent) is always compared with the intensity of the same beam passed through the measured sample.

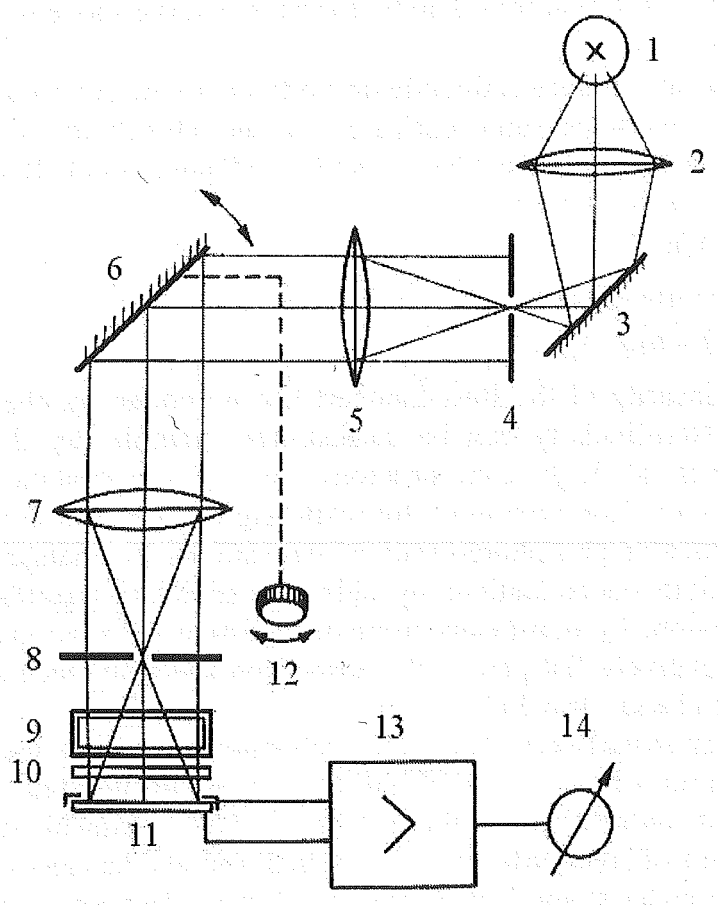


Figure 6.12: A scheme of a Spekol spectrophotometer:
1 – light source;
2 – condenser;
3 – mirror;
4 – input slit;
5 – objective;
6 – optical grating;
7 – objective;
8 – exit slit;
9 – cuvette with solution;
10 – colour filter;
11 – photoelectric detector;
12 – wavelength selector;
13 – amplifier;
14 – display.
(modified after: Keller et al., 1980)