Hoofdstuk 1

Inleidende begrippen over foutentheorie

Doelstellingen

1. leren omgaan met fouten op een meting
2. kennis van statistische basisbegrippen
3. meetgegevens verwerken en interpreteren (in Excell)

1.1 Meetfout

Een fysische grootheid kan gemeten worden maar op elke meting is er meetfout, hoe nauwkeurig uw meettoestel ook is. Er is een zekere onbetrouwbaarheid op elke meting. De exacte waarde van uw meting ligt in wat men noemt een betrouwbaarheidsinterval.We noteren een meting x met zijn betrouwbaarheids-interval B als volgt

$$x \pm B_x$$

Als men dan met deze meetwaarden gaat rekenen gaat de berekende waarde ook een afwijking vertonen op de werkelijke waarde omdat die fout inherent is aan de meting. De fout propageert zich in de berekeningen.

Men definieert ook een

1. absolute fout: de grootste afwijking die kan bestaan tussen de gemeten waarde en de werkelijke waarde. Men neemt voor 1 meting de verdeling van het meettoestel.
2. relatieve fout: de absolute fout gedeeld door de meting

Voorbeeld: We meten de breedte van een A1 blad met behulp van een klassieke lat of regel. We meten 21.1 cm en de verdeling van de lat is in mm. De fout is dus 1mm en we noteren

1. meetresultaat: 21.1 ± 0.1 cm
2. absolute fout: AF=1mm
3. relatieve fout: RF=0.0473933 of 4.73933%

Hierbij moeten we nog vermelden dat we best afronden maar dat het afronden van een getal opnieuw een bron van fouten is die het eindresultaat zal beinvloeden. Het afronden gebeurt gewoonlijk tot op twee beduidende cijfers afhankelijk van de meetprecisie.

voorbeeld:

We meten 15.329m met betrouwbaarheisinterval B=0.26m dan wordt dit genoteerd als 15.33 ± 0.26 m

Er bestaan twee soorten fouten

1. **systematische fout**: Een fout die inherent aan het experiment verbonden zit zoals verkeerde aflezing door parrallax of een verkeerde ijking van een instrument. Deze fout kan enkel aan het licht komen door ze te vergelijken met resultaten van andere onderzoekers of de proef te herhalen met andere instrumenten en/of andere onderzoekers.

2. **toevallige fout**: Deze is uiteraard onvoorspelbaar en te wijten aan wat men noemt ruis op de meting. Deze fout kan men voor een deel wegcompenseren door de meting meermaals te herhalen en dan uit te middelen. De metingen zullen verdeeld of gespreid zijn rond dit gemiddelde met wat men een spreiding noemt.

We moeten nog twee termen definieren die heel wat verwarring kunnen veroorzaken

1. **accuraatheid**(eng. accuracy): is een term die het verschil weergeeft tussen de werkelijke waarde en de gemeten waarde. Wij benaderen de werkelijke waarde door het gemiddelde te nemen van een reeks gemeten waarden. Dit is dan eerder gerelateerd aan de systematische fout omdat we hier rekening houden met een afwijking ten gevolge van de nauwkeurigheid van de meting.

2. **precisie**(eng.precision):Is een term voor de toevallige fout en wordt gekenmerkt door de standaarddeviatie.

1.2 Berekenen van het betrouwbaarheidsinterval

1.2.1 Gekende functie

We hebben een functie die de grootheid die we willen meten beschrijft. (bv een volume) We moeten aan de hand van onze meetresultaten , dus meting en betrouwbaarheidsinterval, een resultaat berekenen met behulp van die functie.

Opgelet, in dit geval veronderstellen we dat de variabelen onafhankelijk zijn van mekaar anders moeten we ook nog rekening houden met hun correlatie.

$$y = f(x_i)$$

dan wordt het betrouwbaarheidsinterval

$$B_y = \sqrt{\sum_{i=1}^{n} (\frac{\partial f}{\partial x_i} B_{xi})^2}$$

Hieruit volgen een aantal rekenregels

1. som en verschil:
$$y = ax_1 \pm bx_2$$
wordt
$$B_y^2 = a^2.B_{x1}^2 + b^2.B_{x2}^2$$
2. produkt en deling:
$$RF_y^2 = RF_{x1}^2 + RF_{x2}^2$$
en het betrouwbaarheidinterval wordt dan
$$B_y = y.RF_y$$
3. machtsverheffing:
$$y = a.x^n$$
dan is
$$RF_y = |n|.RF_x$$
4. sinus:y=a. sinx dan is
$$B_y = |a.cosx|.B_x$$
(opm.:als y=a.cosx draaien de functies gewoon om)
5. logaritme:
$$y = lnx$$
dan is
$$B_y = \frac{B_x}{x}$$
6. exponentieel:
$$y = e^x$$
dan is
$$B_y = e^x.B_x$$

1.2.2 Oefening

We schakelen twee weerstanden in parallel. Bereken de vervangingsweerstand met zijn betrouwbaarheidsinterval. $R_1 = (4.1 \pm 0.2)k\Omega$ *en* $R_2 = (1.01 \pm 0.01)k\Omega$

1.3 Statistische begrippen

1.3.1 Verdelingen

Elke meting is een greep uit een verzameling van mogelijke waarden. Deze verzameling waarden noemt een waarschijnlijkheidsverdeling of kansverdeling f(x).

Deze verdelingen kunnen discreet zijn, dus er is een soort 'aanwijsbaarheid' van een element.

Deze verdelingen kunnen ook continu zijn dus is er eigenlijk geen echte 'aanwijsbaarheid' meer.

Naar gelang het gedrag van de elementen, dus metingen, krijgt men een andere verdeling.

Een belangrijke eigenschap is dat kansverdelingen genormeerd zijn, met andere woorden, de som van alle kansen is gelijk aan 1.

- discrete verdeling:
$$\sum_{k=1}^{n} f(k) = 1$$
- continue verdeling:
$$\int_{-\infty}^{+\infty} f(x)\, dx = 1$$

Zoals reeds vermeld zijn er verschillende verdelingen die elk hun toepassingsgebied hebben. We gaan hieronder enkele belangrijke verdelingen bespreken, zowel discreet als continu.

1. **De binomiaalverdeling**

   Als een meting slechts twee waarden kan aannemen (waar/onwaar, 0/1,..) dan wordt de kans dat van n metingen er k waarde 1 hebben beschreven door
$$f(k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}$$

2. **De Poissonverdeling**

   Deze verdeling wordt gebruikt als aantallen geteld worden zoals een aantal fotonen door in een tijdsinterval of een aantal bacteriekolonien in een Petrischaal. Als μ het gemiddeld aantal is dat verwacht wordt bij telling is de kans dat k gebeurtenissen zich voordoen
$$f(k) = \frac{\mu^k}{k!} e^{-\mu}$$

Deze verdeling is eigenlijk een limietgeval van de binomiaalverdeling en voor grote waarden voor k benadert deze verdeling de normaalverdeling.

Een belangrijke eigenschap van deze verdeling is dat de spreiding gelijk is aan $\sqrt{\mu}$. Ze wordt veel gebruikt voor tijdsafhankelijke processen.(bv.slijtage) In Excell is dit de EXPONDIST functie.

3. **De Normaalverdeling**

   Deze verdeling komt zeer vaak voor in de natuur en ten gevolge van de centrale limietstelling kan deze verdeling in limiet van grote aantallen metingen altijd gebruikt worden.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

4. **De Chi-kwadraat verdeling**

   Deze verdeling wordt gebruikt om betrouwbaarheidsintervallen te verkrijgen voor de afwijkingen tussen meetwaarden en voorspelde waarden in het geval dat de standaarddeviatie van de meetwaarden gekend is.

5. **De Student t verdeling**

   Deze verdeling wordt gebruikt om betrouwbaarheidsintervallen te verkrijgen voor het gemiddelde van een reeks waarnemingen waarbij de standaarddeviatie van de metingen niet gekend is. In Excell is dit de TDIST functie.

6. **De Fisher verdeling**

   Deze verdeling wordt gebruikt om twee theorieen te bekijken, gegeven een stel data.

1.3.2 Momentgenererende functie

Dit is een algemene methode om gemiddelde en variantie (en andere eigenschappen) van een verdeling te berekenen.

Per definitie is het n^e moment van een waarschijnlijkheidsverdeling

$$\mu_n = E[x^n]$$

en het n^e centrale moment

$$\mu_n^c = E[(x-\mu)^n]$$

waarin

$$E[g(x)] = \sum_{k=1}^{n} g(k)f(k)$$

voor een discrete verdeling en

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x) f(x)\, dx$$

voor een continue verdeling

Zo is is het derde centrale moment gerelateerd aan de scheefheid (skewness) van de verdeling en het vierde centrale moment aan de spitsheid (kurtosis) van de verdeling.

1.3.3 Gemiddelde

Het gemiddelde $\bar{x}$ is het eerste centrale moment van een verdelingsfunctie.

Voor een meetreeks kan in Excell dit gemiddelde snel berekend worden met de functie'gemiddelde'. We moeten echter nog vermelden dat we ook met een gewogen gemiddelde kunnen werken als bepaalde metingen in de meetreeks belangrijker zijn dan andere. Men kent een gewicht w toe aan elke meting en men berekent de gewenste waarde als volgt

$$\langle g \rangle = \frac{1}{w} \sum_{i=1}^{n} w_i x_i$$

en

$$w = \sum_{i=1}^{n} w_i$$

1.3.4 Mediaan

De mediaan van een verzameling metingen is de waarde waar evenveel metingen onder als boven liggen in een gesorteerde rij meetresultaten. Voor een verdeling F is dit het punt m waar F(m)=0.5. Als de verdeling een oneven aantal metingen bevat is dit de middelste meting bij een even aantal neemt men dan de twee middelste getallen of het gemiddelde van die twee.

1.3.5 Spreiding

De spreiding is de vierkantswortel van de variantie van een verdelingsfunctie en de variantie is het tweede centrale moment van deze verdelingsfunctie.

In Excell wordt voor een meetreeks de functie 'var' gebruikt. Van deze waarde neemt men dan de vierkantswortel.

1.3.6 Uitschieters

Uitschieters (eng. outliers) zijn eigenlijk een stoorzender in het verwerken van meetgegevens want zij zullen de berekende waarden vervormen ondanks het feit dat deze waarden echt gemeten zijn. De oorzaken van deze uitschieters zijn legio maar het verwerken ervan kan aanleiding geven tot verkeerde resultaten en dus verkeerde besluiten. Er zijn rond deze problematiek al heel wat onderzoeken gebeurt maar wij gaan het simpel houden. Voorzichtigheid is geboden als men overgaat tot het weglaten van gemeten waarden op bais van een te grote afwijking. Men stelt als vuistregel dat een waarde mag verworpen worden als men de afwijking ten op zichte van het gemiddelde bekijkt met andere woorden, wordt de spreiding als maatstaf gebruikt. Het criterium moet zo gekozen worden dat het toevallig voorkomen van zo'n afwijking onwaarschijnlijk is. De grens die men dan vooropstelt hangt dus duidelijk af van het aantal metingen. Als vuistregel stelt men:

aantal meetpunten	**verwerpen bij**
minder dan 10	2.5σ
10 tot 50	3σ
meer dan 50	3.5σ

1.4 Regressie

Als men als resultaat een verband wil zoeken tussen twee parameters gaat men een wiskundige vergelijking (een functie) willen opstellen om dit verband te beschrijven. Men krijgt aan de hand van gegevens echter bijna nooit een éénduidig klaar en duidelijk verband maar wel een puntenwolk waar men een verband wil uit berekenen. Hiertoe heeft men in de statistiek een methode gevonden om dit verband uit een puntenwolk te berekenen en deze methode noemt men regressie.

Deze methode noemt men de 'kleinste kwadratenmethode'. Als men deze techniek gebruikt komt het hierop neer dat men een kromme trekt doorheen de puntenwolk en daarbij de som van de afstand van elk punt tot die kromme gaat minimaliseren. Eigenlijk gaat men de totale fout op de benadering zo klein mogelijk houden. De afwijking kan men dan berekenen via de correlatiecoefficient.

Een speciaal geval hiervan is de lineaire regressie die een lineair verband distilleert uit die puntenwolk. (Deze berekening trekt dus een rechte door de puntenwolk die de totale som van de afstanden van elk punt tot de rechte minimaal maakt.)

De vergelijking van beste rechte is in het algemene geval

$$y = a.x + b$$

dus moeten we a en b zoeken en hiervoor gebruikt men volgende formules

$$A = \sum x_i.y_i - n.\bar{x}.\bar{y}$$

$$D = \sum x_i^2 - n.(\bar{x})^2$$

$$E = \sum y_i^2 - n.(\bar{y})^2$$

waaruit de waarden voor a en b volgen

$$a = \frac{A}{D}$$

en

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

De correlatie r berekend men met

$$r = \frac{A}{\sqrt{D.E}}$$

. Deze factor geeft een idee hoe goed het verband opgaat.

In Excell gebruikt men 'scatterplot' of 'spreiding met alleen markeringen'. Dan kiest men onder 'indeling' het vak 'trendlijn' en gaat men naar 'meer opties voor trendlijn' en men vinkt onderaan het venster 'vergelijking in grafiek weergeven' en 'R-kwadraat in grafiek weergeven' aan.

Als men meetgegevens verwerkt zal men in de meeste gevallen gemiddeldes van de meetpunten gebruiken om weer te geven in een grafiek. Dan wordt ook de standaarddeviatie per meetpunt berekend en dan kan men in de grafiek de standaarddeviatie weergeven met wat men een foutenvlag noemt. Deze foutenvlag kan men in excell aanbrengen via 'trendlijn' dan 'foutenbalken' dan naar onderaan 'aangepast' gaan en de standaardvariatie aanduiden in positieve en negatieve waarden.

1.5 Toetsen van hypothesen

1.5.1 Algemeen overzicht

Men gaat van een meetreeks een gemiddelde en een variantie berekenen maar men moet zich dan wel ervan bewust zijn dat dit slechts een schatting is aan de hand van die meetuitkomsten. Men moet dus kijken naar de betrouwbaarheid van de metingen en men gaat een betrouwbaarheidsinterval berekenen voor die 'schatting'. Dit noemt men toetsen van hypothesen. Men stelt een hypothese voorop, de zogenaamde nulhypothese H_0 en men gaat berekenen of deze hypothese opgaat voor een zeker percentage (99% of 95%). Men gaat dan vergelijken met een zogenaamde alternatieve hypothese de zogenaamde H_1 hypothese.Dat percentage noemt men ook wel het betrouwbaarheidsniveau. (eng. confidence level)

Elke toets kan berekend worden boven OF onder een bepaalde waarde en dan noemt men dit een ***éénzijdige*** *toets. Als men echter zoekt naar een interval van twee bepaalde waarden, dus boven EN onder een bepaalde waarde of anders VERSCHILLEND van een bepaalde waarde berekend men een* ***tweezijdige*** *toets.*

Het betrouwbaarheidsniveau wordt $(1 - \alpha)$ genoemd. De waarschijnlijkheid waarmee H_0 onterecht verworpen wordt is dan α.

Hieruit verkrijgt men volgende tabel

	H_0 **is waar**	H_1 **is waar**
H_0 **wordt geaccepteerd**	juiste beslissing kans= $1-\alpha$	onjuiste beslissing kans= β
H_0 **wordt verworpen**	onjuiste beslissing kans= α	juiste beslissing kans= $1-\beta$

Voor het toetsen van hypothesen wordt er een standaardprocedure gebruikt

1. Formuleer een nulhypothese
2. Zoek de toetsvariabele en de geschikte verdelingsfunctie
3. Bepaal het onbetrouwbaarheidsniveau (α dus)
4. Bepaal het acceptatiegebied
5. Bepaal de actuele waarde voor de toetsvariabele en vergelijk met het acceptatiegebied
6. Beslis of H_0 verworpen wordt

1.5.2 Enkele veel gebruikte toetsen

1. **u toets**

   Deze toets wordt gebruikt om te onderzoeken of het verschil tussen een steekproefgemiddelde $\bar{x}$ en een referentiewaarde μ_0 van dat gemiddelde significant is. Hierbij moet de variantie van de populatie (dus van de referentie) σ en het gemiddelde van de steekproef gekend zijn. Het aantal elementen in de steekproef is n.

   In Excell gebruikt men hiervoor de functie NORMINV.

   De statistische variabele is in dit geval

   $$u = \left(\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

   en deze is standaard-normaal verdeeld. Het acceptatiegebied wordt gevonden met Excell $NORMINV(\alpha; 0; 1)$

2. **t toets**

   (a) **voor één populatiegemiddelde**

   Deze toets wordt gebruikt om te onderzoeken of het verschil tussen een steekproefgemiddelde $\bar{x}$ en een referentiewaarde van dat gemiddelde μ significant is. Hierbij moet het gemiddelde en de variantie van de steekproef s^2 gekend zijn. Dit is de de normale situatie bij meetcampagnes!

   In Excell gebruikt men hiervoor de functie TINV.

De statistische variabele is in dit geval

$$t = (\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}})$$

en deze is student verdeeld met het aantal vrijheidsgraden $\nu = n - 1$. Het acceptatiegebied wordt gevonden met Excell $TINV(\alpha; \nu)$

(b) **voor twee populatiegemiddelden**

In dit geval zijn er weer twee versies, ofwel zijn de populatiestandaardeviaties gelijk (homoscedastisch) ofwel zijn ze ongelijk (heteroscedastisch).

Ook in dit geval gebruikt men in Excell de functie TINV

3. **chi kwadraat toets**

Deze toets wordt gebruikt om te onderzoeken of het verschil tussen een populatievariantie en zijn referentiewaarde significant is. De steekproefvariantie is gekend.

In Excell gebruikt men de functie CHINV

De statistische variabele is in dit geval

$$\chi^2 = \nu \frac{s^2}{\sigma^2}$$

met $\nu = n - 1$. In Excell berekend men dan $CHINV(\alpha, \nu)$

4. **F toets**

Deze toets wordt gebruikt om na te gaan of het verschil tussen twee populatievarianties significant is, ook hier zijn de steekproefvarianties gekend. De populaties moeten *niet* hetzelfde gemiddelde hebben.

In Excell gebruikt men voor deze berekening de functie FINV.

Α α	alpha	Ν ν	nu
Β β	beta	Ξ ξ	ksi
Γ γ	gamma	Ο ο	omicron
Δ δ	delta	Π π	pi
Ε ε	epsilon	Ρ ρ	rho
Ζ ζ	zeta	Σ σς	sigma
Η η	eta	Τ τ	tau
Θ θ	theta	Υ υ	upsilon
Ι ι	iota	Φ φ	phi
Κ κ	kappa	Χ χ	chi
Λ λ	lambda	Ψ ψ	psi
Μ μ	mu	Ω ω	omega

USEFUL CONVERSION FACTORS AND RELATIONSHIPS

Length

SI unit: meter (m)

1 km = 0.62137 mi
1 mi = 5280 ft
= 1.6093 km
1 m = 1.0936 yd
1 in. = 2.54 cm (exactly)
1 cm = 0.39370 in.
1 Å = 10^{-10} m

Mass

SI unit: kilogram (kg)

1 kg = 2.2046 lb
1 lb = 453.59 g
= 16 oz
1 amu = $1.66053873 \times 10^{-24}$ g

Temperature

SI unit: Kelvin (K)

0 K = −273.15°C
= −459.67°F
K = °C + 273.15
$°C = \frac{5}{9}(°F - 32°)$
$°F = \frac{9}{5}°C + 32°$

Energy (derived)

SI unit: Joule (J)

1 J = 1 kg-m^2/s^2
1 J = 0.2390 cal
= 1 C × 1 V
1 cal = 4.184 J
1 eV = 1.602×10^{-19} J

Pressure (derived)

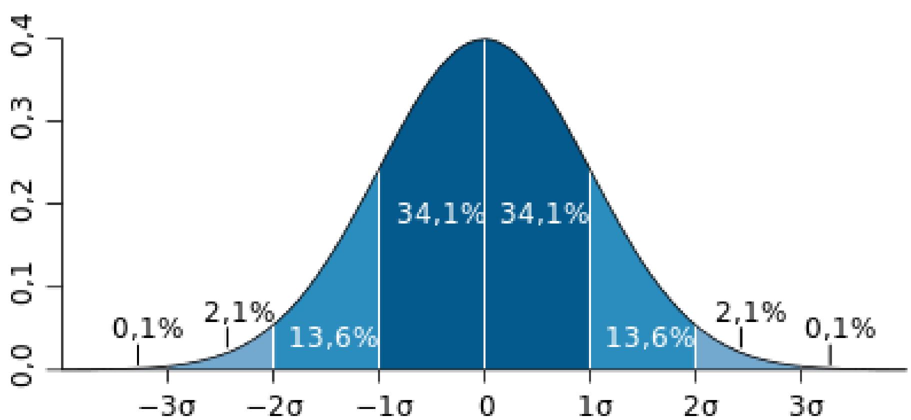
SI unit: Pascal (Pa)

1 Pa = 1 N/m^2
= 1 kg/m-s^2
1 atm = 101,325 Pa
= 760 torr
= 14.70 lb/in^2
1 bar = 10^5 Pa
1torr = 1 mm Hg

Volume (derived)

SI unit: cubic meter (m^3)

1 L = 10^{-3} m^3
= 1 dm^3
= 10^3 cm^3
= 1.0567 qt
1 gal = 4 qt
= 3.7854 L
1 cm^3 = 1 mL
1 in^3 = 16.4 cm^3


0,4
0,3
0,2
0,1
0,0
34,1%
34,1%
13,6%
13,6%
2,1%
2,1%
0,1%
0,1%
−3σ
−2σ
−1σ
0
1σ
2σ
3σ

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)