ΣΤΑΤΙΣΤΙΚΗ :

Παράσταση – Περιγραφή δεδομένων

Σύγκριση δεδομένων – Εξαγωγή συμπερασμάτων

Σχέση αιτίου - αιτιατού

Με τις στατιστικές μεθόδους επιδιώκεται αφενός η συνοπτική αλλά εμπεριστατωμένη παρουσίαση των ευρημάτων μιας μελέτης (**περιγραφική στατιστική**) και αφετέρου η συναγωγή συμπερασμάτων που βασίζονται στα ευρήματα αυτά (**συμπερασματολογική στατιστική / επαγωγική στατιστική**)

Πιθανότητα (P, Probability) είναι μέτρο του πόσο αναμενόμενο να συμβεί ένα γεγονός ή μια θέση (ισχυρισμός) να είναι αληθής. Οι πιθανότητες παίρνουν τιμές μεταξύ 0 (δεν θα συμβεί) και 1 (θα συμβεί). Όσο μεγαλύτερη η πιθανότητα ενός γεγονότος, τόσο βέβαιοι είμαστε ότι θα συμβεί.

Ως **μεταβλητή** θεωρούμε κάθε χαρακτηριστικό το οποίο μπορεί να μεταβληθεί ή να διαφοροποιηθεί κατά μήκος του χρόνου, από τόπο σε τόπο, από άτομο σε άτομο ή από ομάδα σε ομάδα (π.χ. ηλικία, ύψος, εισόδημα, συγκέντρωση χοληστερόλης, αρτηριακή πίεση, ρυθμός γεννητικότητας κτλ)

- **Ποιοτική** ονομάζεται η μεταβλητή που περιγράφει κάποιο ποιοτικό χαρακτηριστικό ενός ατόμου ή μιας ομάδας.
- **Ποσοτική** ονομάζεται η μεταβλητή που μπορεί να μετρηθεί με τη συνήθη έννοια του όρου
  - Συνεχής
  - Ασυνεχής

- Ως **ανεξάρτητη** χαρακτηρίζεται μια μεταβλητή όταν επηρεάζει μια άλλη μεταβλητή.
- Ως **εξαρτημένη** χαρακτηρίζεται μια μεταβλητή όταν επηρεάζεται από μια άλλη μεταβλητή.

ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ

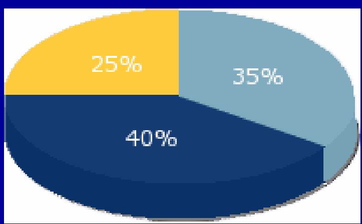
Γραφικές μέθοδοι

- **Ραβδογράμματα - Ιστογράμματα (Συχνότητα)**
- **Κυκλικά διαγράμματα**
- **Διαγράμματα πλαισίου**

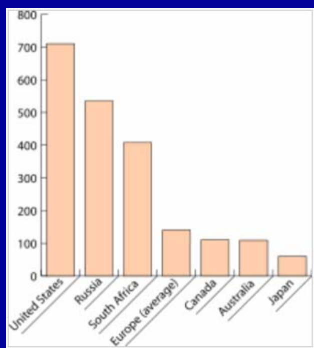
Αριθμητικοί στατιστικοί δείκτες ή μέτρα *(statistics)*

- **Κεντρικής τάσης**
  - **Μέση τιμή *(mean)***
  - **Διάμεσος *(median)***
  - **Επικρατούσα τιμή *(mode)***
- **Διασποράς**
  - **Εκατοστημόρια ή ποσοστιαία σημεία *(percentiles)***
  - **Διακύμανση ή Διασπορά *(Variance)***
  - **Τυπική απόκλιση *(standard deviation)***

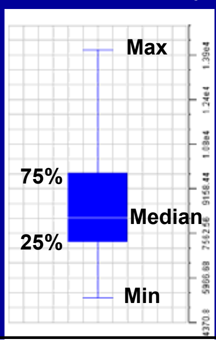
Διάγραμμα πίτας (Pie chart)



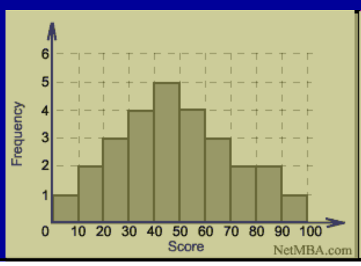
Ραβδόγραμμα (bar chart)

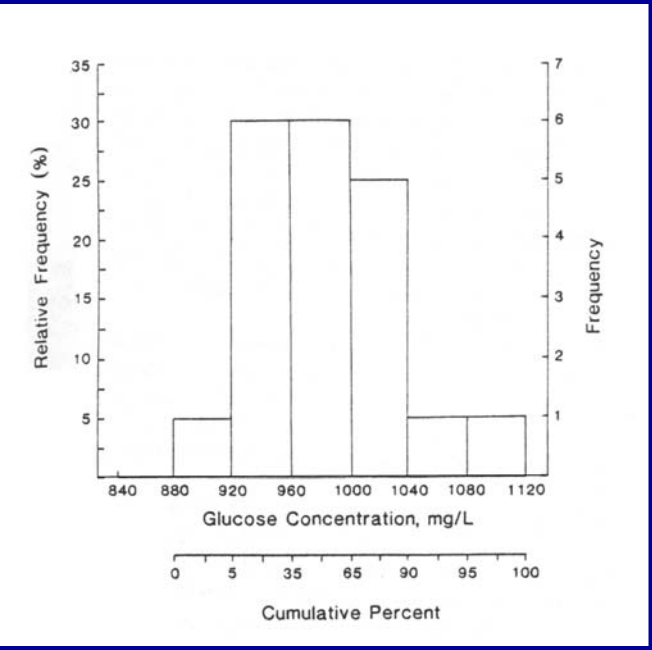


Διάγραμμα πλαισίου (Box plot)



Ιστόγραμμα (Histogram)





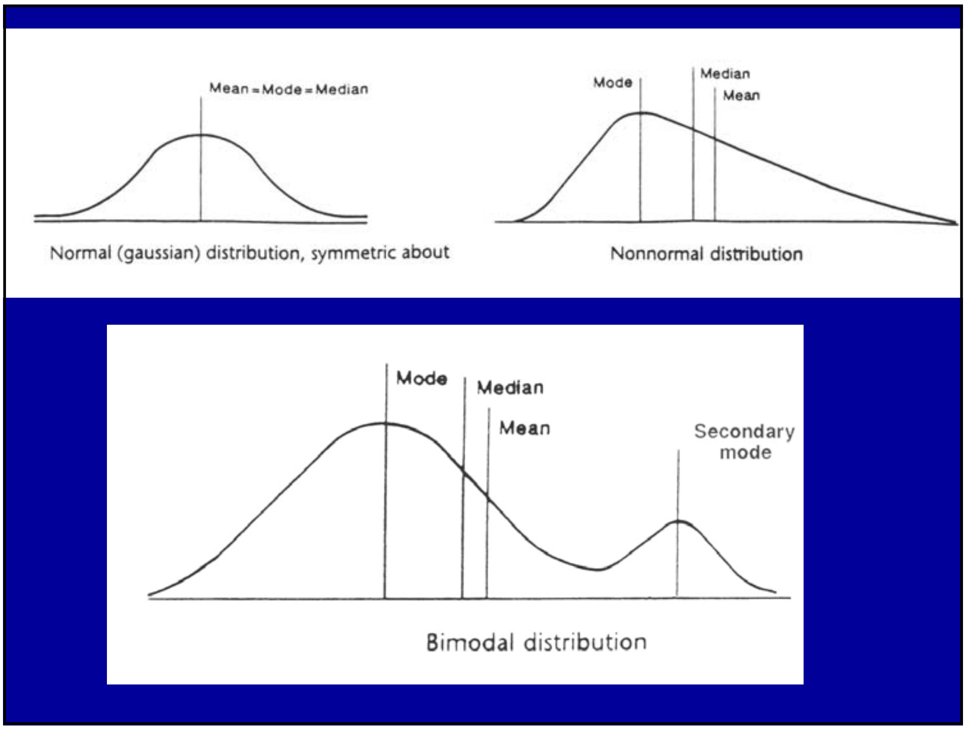
Μέση τιμή :
(mean ή average)

$$\overline{x} = \frac{\sum_{1}^{N} x_i}{N}$$

Διάμεσος :
(median)

- **Ταξινόμηση των μετρήσεων κατά μέγεθος**
- **Επιλογή της τιμής στο μέσον των μετρήσεων**

Επικρατούσα τιμή : Η συχνότερη τιμή
(mode)



Εκατοστημόρια ή ποσοστιαία σημεία *(percentiles)*

- **Διατάσσουμε τα δεδομένα κατά τάξη μεγέθους**
- **Το p-εκατοστημόριο είναι η τιμή που έχει p% των μετρήσεων μικρότερες από αυτήν**

Τεταρτημόρια *(quartiles)*

Ειδικά εκατοστημόρια

- $Q_1 \rightarrow 25\%$
- $Q_2 \rightarrow 50\%$
- $Q_3 \rightarrow 75\%$

ΜΕΤΡΑ ΤΗΣ ΔΙΑΣΠΟΡΑΣ

Έκταση ή εύρος *(range)* : $x_{max} - x_{min}$

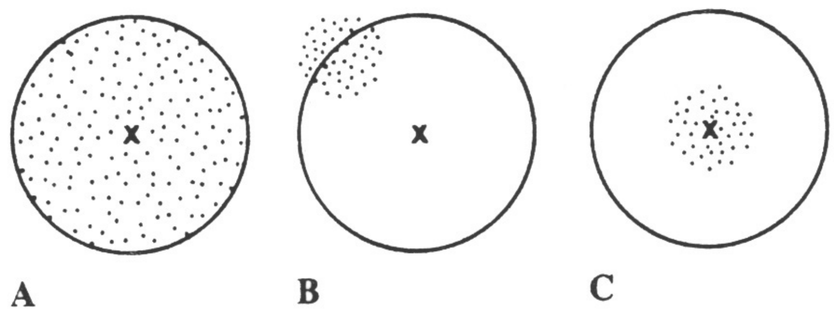
Διακύμανση – Διασπορά *(variance)* :
$$s^2 = \frac{\sum_{1}^{N}(x_i - \bar{x})^2}{N-1}$$

Πληθυσμού «σ²» Δείγματος «s²»

Τυπική απόκλιση *(standard deviation)* :
$$s = \sqrt{\frac{\sum_{1}^{N}(x_i - \bar{x})^2}{N-1}}$$

Συντελεστής διακύμανσης *(coefficient of variance)* :
$$\%CV = \frac{100\% s}{\bar{x}}$$

Accurate but imprecise **inaccurate but precise** **Accurate and precise**



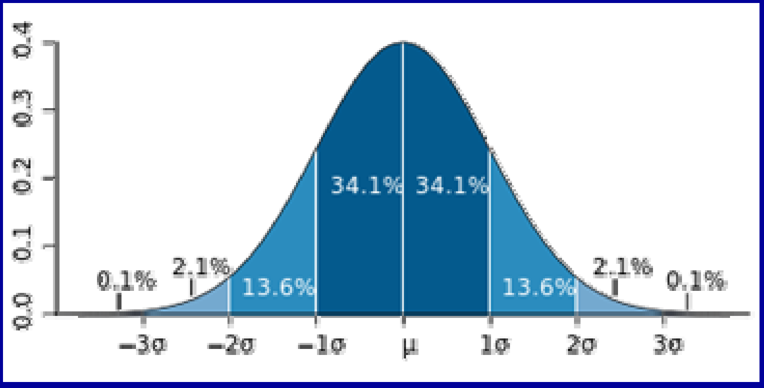
Precision: ορθότητα ή επαναληψιμότητα

Accuracy: ακρίβεια

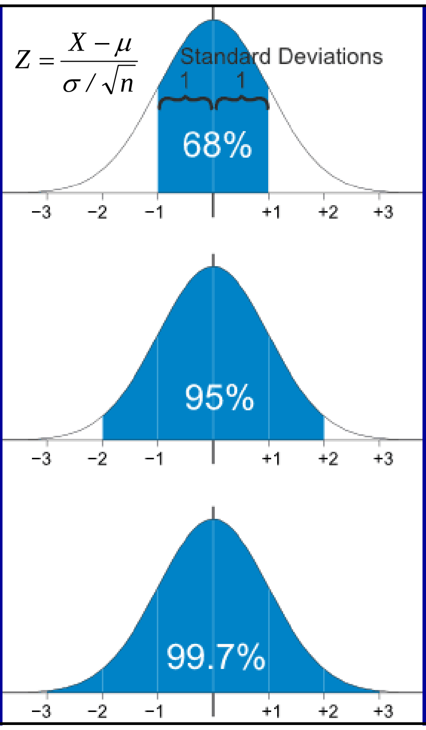
Κανονική κατανομή

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

N(μ,σ²)



$$Z = \frac{X-\mu}{\sigma/\sqrt{n}}$$

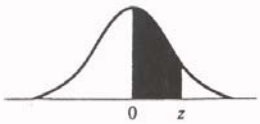


N(0,1)

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

ΠΙΝΑΚΑΣ Ι

Των Πιθανοτήτων P $(0 < Z < z)$ για την N(0,1) κατανομή

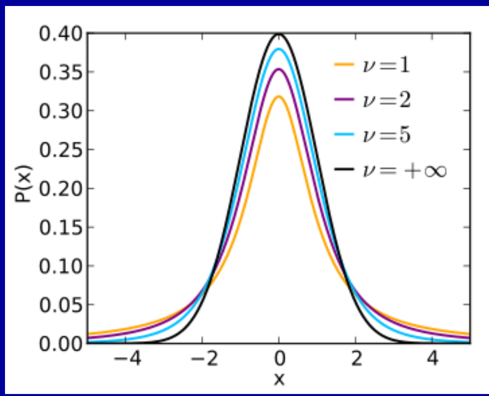


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817

Κεντρικό οριακό θεώρημα (central limit theorem)

Για αρκούντως μεγάλα δείγματα από έναν πληθυσμό, οι μέσες τιμές ακολουθούν περίπου την κανονική κατανομή, ανεξάρτητα από το είδος της κατανομής του πληθυσμού. Όσο μεγαλύτερα τα δείγματα τόσο καλύτερα προσεγγίζεται η κανονική κατανομή.

Κατανομή t του Student



$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

ΕΚΤΙΜΗΤΙΚΗ

- **Εκτίμηση σε σημείο**

Δίνουμε τους αριθμητικούς δείκτες του δείγματος ως προσεγγιστική (αβέβαιη) εκτίμηση αυτών του πληθυσμού

εκτιμώμενο τυπικό σφάλμα : $s_{\hat{x}} = \frac{s}{\sqrt{N}}$

τυπικό σφάλμα : $\sigma_{\hat{x}} = \frac{\sigma}{\sqrt{N}}$

Εκτίμηση σε διαστήματα εμπιστοσύνης (*confidence intervals*)

Δίνουμε ένα διάστημα μέσα στο οποίο αναμένεται με συγκεκριμένη πιθανότητα να εμπίπτει μια παράμετρος του πληθυσμού

Στάθμη σημαντικότητας (σ.σ): 1-a→100(1-a)% δ.ε.

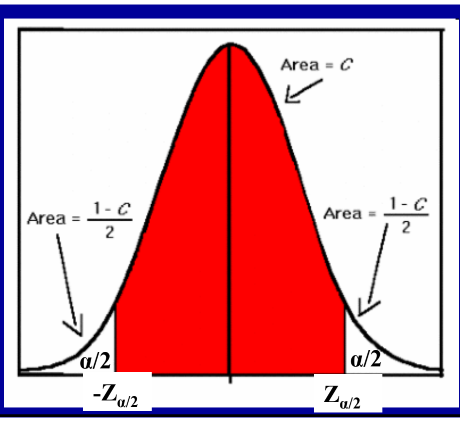
$$\mu = \bar{x} \pm t \cdot s_{\hat{x}}$$

- **Διάστημα εμπιστοσύνης για τη μέση τιμή (μεγάλα δείγματα)**

Η μεταβλητή "Z" ακολουθεί την κανονική κατανομή $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$

Αναζητούμε την πιθανότητα για κάποιο "z" από πίνακες της τυποποιημένης κανονικής κατανομής

$$P\left[\bar{X}_n - Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{X}_n + Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right] = 1 - \alpha$$



Ο δείκτης σωματικής μάζας X(kgr/m²) ενός ατόμου υπολογίζεται αν διαιρέσουμε το βάρος του με το τετράγωνο του ύψους του. Για άνδρες ηλικίας 30-40 ετών είναι γνωστό ότι $X \sim N(\mu,\sigma^2)$. Να προσδιορισθεί το 95% δ.ε. για την σωματική μάζα μ των ανδρών εάν από τυχαίο δείγμα 49 ανδρών από αυτόν τον πληθυσμό προέκυψε $\overline{X} = 25$ και $s^2=9$.

Δ.ε. 95% → $Z_{\alpha/2}=Z_{0.025}$ –(πίνακες)-> 1.96

$$\overline{X}_n - 1.96\left(\frac{3}{\sqrt{49}}\right) \le \mu \le \overline{X}_n + 1.96\left(\frac{3}{\sqrt{49}}\right)$$

- **Διάστημα εμπιστοσύνης για τη μέση τιμή (μικρά δείγματα)**

Η μεταβλητή "t" ακολουθεί την κατανομή "t"

$$t = \frac{\overline{Y} - \mu}{s/\sqrt{n}}$$

Αναζητούμε την τιμή "t" από πίνακες για "v" β.ε και a

$$\overline{X}_n - t_{n-1}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \overline{X}_n + t_{n-1}\left(\frac{s}{\sqrt{n}}\right)$$

Από δείγμα 15 υγιών γυναικών ηλικίας 25-40 ετών, υπολογίσθηκε για την αμυλάση του ορού ότι $\overline{X} = 96$ μονάδες/100ml και $s=35$ μονάδες/ml. Να υπολογισθεί το 90% διάστημα εμπιστοσύνης για την αληθή τιμή της μέσης τιμής μ της αμυλάσης στον πληθυσμό των υγιών γυναικών στις ίδιες ηλικίες.

Δ.ε. 90% → $t_{14,0.05}$ –(πίνακες)-> 1.76

$$\overline{X}_n - 1.76\left(\frac{35}{\sqrt{15}}\right) \le \mu \le \overline{X}_n + 1.76\left(\frac{35}{\sqrt{15}}\right)$$

• **Διάστημα εμπιστοσύνης για τη διασπορά (μεγάλα και μικρά δείγματα)**

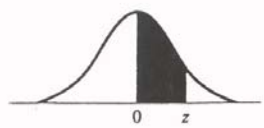
Ακολουθεί την κατανομή "X^2"

Αναζητούμε τις τιμές "X^2" από πίνακες για $P(X^2 > X^2_{v;a}) = a$

$$\frac{(n-1)\cdot s_n^2}{X^2_{(n-1),a/2}} < \sigma^2 < \frac{(n-1)\cdot s_n^2}{X^2_{(n-1),(1-a/2)}} \qquad X^2 = \frac{(n-1)\cdot s^2}{\sigma^2}$$

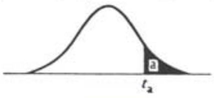
ΠΙΝΑΚΑΣ Ι

Των Πιθανοτήτων P (0 < Z < z) για την N(0,1) κατανομή

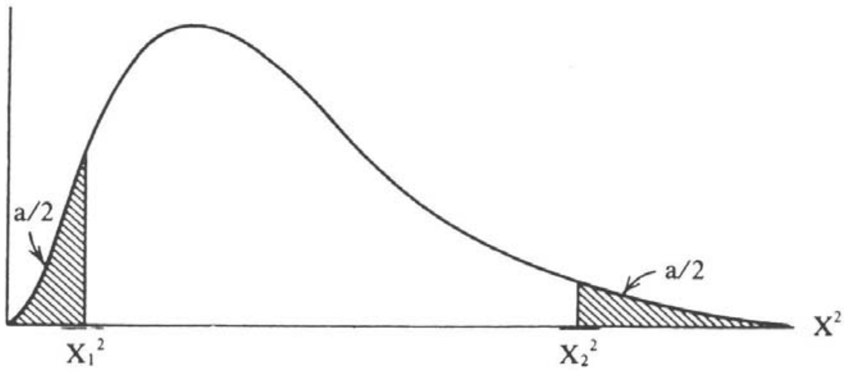


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817

Τιμών $t_{v,a}$ της t_v-κατανομής ώστε $P(t_v > t_{v,a}) = a$



Β.ε	a = .10	a = .05	a = .025	a = .010	a = .005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	Γ.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
inf.	1.282	1.645	1.960	2.326	2.576



• Διάστημα εμπιστοσύνης για τη διαφορά μέσων δύο πληθυσμών (μεγάλα ανεξάρτητα δείγματα)

- Ακολουθεί την κανονική κατανομή
- Αναζητούμε τις τιμές *“Z”* από πίνακες της τυποποιημένης κανονικής κατανομής

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\mu_1 - \mu_2 = \bar{\mathbf{X}} - \bar{Y} \pm Z_{a/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

• Διάστημα εμπιστοσύνης για τη διαφορά μέσων δύο κανονικών πληθυσμών με κοινό “σ²” (μικρά ανεξάρτητα δείγματα)

- Ακολουθεί την κατανομή *“t”* με $v_1 + v_2$ β.ε.
- Αναζητούμε την τιμή *“t”* από πίνακες για $v_1 + v_2$ και a/2

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\mu_1 - \mu_2 = \bar{x} - \bar{y} \pm t \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

• Διάστημα εμπιστοσύνης για τη διαφορά “μέσων” ζευγαρωτών δειγμάτων

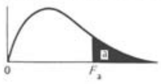
- Διαφορά *“δ”* μεταξύ των μέσων τιμών των δύο δειγμάτων
- Ακολουθεί την κατανομή *“t”* με $v = n-1$ β.ε.
- Αναζητούμε την τιμή *“t”* από πίνακες για v και a/2

• Διάστημα εμπιστοσύνης για τον λόγο των “διασπορών” δύο κανονικών πληθυσμών

- Ακολουθεί την κατανομή *“F”* με v_1, v_2 β.ε.
- Αναζητούμε την τιμή *“F”* από πίνακες για v_1, v_2 και a/2

ΠΙΝΑΚΑΣ IV

Των τιμών $F_{v_1, v_2, a}$ της F-κατανομής για τις οποίες $P(F > F_{v_1, v_2, a}) = a$



Βαθμοί Ελευθερίας (a = .05) **αριθμητής**

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

ΔΟΚΙΜΑΣΙΑ ΥΠΟΘΕΣΕΩΝ (ΕΛΕΓΧΟΣ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ)
Significance testing

Συγκρίνοντας δύο πληθυσμούς διαπιστώνουμε πάντοτε διαφορές μεταξύ των μέσων $\hat{\mathbf{x}}$ αλλά και μεταξύ των διασπορών τους $\mathbf{s}^2$.

Απηχούν πραγματικές διαφορές μεταξύ των πληθυσμών;

Μηδενική υπόθεση (null) $H_0 : \mu_1 = \mu_2$ ή $\sigma_1^2 = \sigma_2^2$

Εναλλακτική υπόθεση (alternative) $H_1 : \mu_1 \neq \mu_2$ ή $\sigma_1^2 \neq \sigma_2^2$

Σφάλμα 1ου είδους (α) : Πιθανότητα εσφαλμένης απόρριψης της H_0

Σφάλμα 2ου είδους (β) : Πιθανότητα εσφαλμένης απόρριψης της H_1

- Έλεγχος για τη μέση τιμή "μ" (μεγάλα δείγματα), σ.σ. "α"

$$z = \frac{\bar{y} - \mu_0}{s / \sqrt{n}} = \frac{\textit{δειγματικός μέσος - υποτιθέμενος μέσος}}{\textit{τυπικό σφάλμα του δειγματικού μέσου}}$$

Κ.Ο.Θ→Το «z» ακολουθεί την κανονική κατανομή

Για κάθε "α" αντιστοιχεί ένα κρίσιμο "z_α"

Η H_0 απορρίπτεται όταν $z > z_\alpha$

- Έλεγχος για τη μέση τιμή "μ" (μικρά δείγματα), σ.σ. "α"

$$t = \frac{\bar{y} - \mu_0}{s / \sqrt{n}}$$

ακολουθεί την κατανομή "t"

- **Έλεγχος μέσων από δύο πληθυσμούς**

Όταν $n_1 \neq n_2$ → unpaired t-test

Δείγματα από ανεξάρτητους πληθυσμούς
Διακυμάνσεις όχι σημαντικά διαφορετικές
Unpaired

Π.χ. Σύγκριση τιμών γλυκόζης από ασθενείς δύο διαφορετικών νοσοκομείων

Όταν $n_1 = n_2$ → κατά ζεύγη (paired) t-test

Διακυμάνσεις όχι σημαντικά διαφορετικές

ΑΝΕΞΑΡΤΗΤΑ

- Έλεγχος για διαφορά δύο μέσων τιμών (μεγάλα δείγματα), σ.σ. "*α*"

$$z = \frac{(\bar{y}_1 - \bar{y}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

ακολουθεί την κανονική κατανομή

Για κάθε "*α*" αντιστοιχεί ένα κρίσιμο "z_α"

Η H_0 απορρίπτεται όταν $z > z_\alpha$

- Έλεγχος για τη διαφορά δύο μέσων τιμών (μικρά δείγματα), σ.σ. "*α*"

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

ακολουθεί την κατανομή "*t*"

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

β.ε.$=n_1+n_2-2$

t-test_immunoglobulin.xls

ΚΑΤΆ ΖΕΥΓΗ

- Έλεγχος σημαντικότητας για τη σύγκριση μέσων τιμών κατά ζεύγη (paired ή split sample)

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - \delta_0}{s_d / \sqrt{n}}$$

ακολουθεί την κατανομή *"t"*

s_d: για τις διαφορές

- Έλεγχος σημαντικότητας για το «***p***» της διωνυμικής κατανομής (μεγάλα δείγματα)

Έστω «***x***» επιτυχίες σε «***n***» δοκιμές. Συγκρίνουμε την αναλογία ***x/n*** των επιτυχιών με μια δοθείσα πιθανότητα p_0 με την βοήθεια της τυποποιημένης μεταβλητής ***Z***.

$$Z = \frac{\frac{x}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

6_17.xls

- Έλεγχος σημαντικότητας για τη διασπορά, σ.σ. *"α"*

$$X^2 = \frac{(n-1)\cdot s^2}{\sigma^2}$$

ακολουθεί την κατανομή *"X^2"*

Για κάθε *"α"* και *"ν"* αντιστοιχεί ένα κρίσιμο *"$X^2_{\alpha;\nu}$"*

Η H_0 απορρίπτεται όταν $X^2 > X^2_{\alpha;\nu}$

- Σύγκριση των διασπορών δύο πληθυσμών, σ.σ. *"α"*

$$F = \frac{s_1^2}{s_2^2}$$

ακολουθεί την κατανομή *"F"*

Για κάθε *"α"* και *"ν_1,ν_2"* αντιστοιχεί ένα κρίσιμο *"$F_{\nu 1,\nu 2;\alpha}$"*

Η H_0 απορρίπτεται όταν $F \geq F_{\nu 1,\nu 2;\alpha}$ για $s_1^2 > s_2^2$

Η H_0 απορρίπτεται όταν $F \geq F_{\nu 2,\nu 1;\alpha}$ για $s_1^2 < s_2^2$

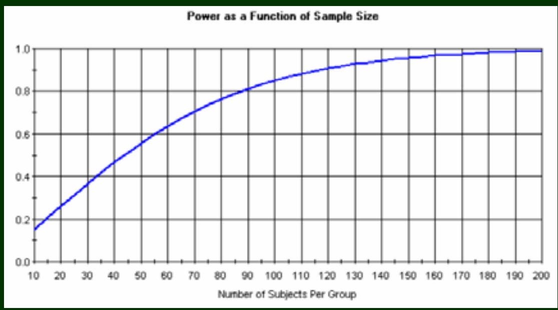
Folate.xls

Power analysis

Στατιστική ισχύς (Statistical Power) : P = 1 – β, P τουλάχιστον .80

Σφάλμα 2ου είδους (β) : Πιθανότητα εσφαλμένης απόρριψης της H_1

Ποιο μέγεθος δείγματος θα μου δώσει την απαιτούμενη ακρίβεια;



Η στατιστική ισχύς εξαρτάται από το μέγεθος της διαφοράς που μελετάμε και από το μέγεθος του δείγματος.

Εάν η στατιστική ισχύς είναι υψηλή, η πιθανότητα να κάνουμε σφάλμα 2ου είδους (να συμπεράνουμε ότι δεν υπάρχει διαφορά ενώ υπάρχει) είναι μικρή.

	Effect size Index	Small	Medium	Large
t-test on Means	d	0.20	0.50	0.80
t-test on Correlations	r	0.10	0.30	0.50
F-test ANOVA	f	0.10	0.25	0.40
F-test regression	f^2	0.02	0.15	0.35
Chi-Square Test	w	0.10	0.30	0.50

$$-1.64 = Z_{cr} = \frac{\bar{x}_{cr} - \mu_0}{s/\sqrt{n}} \Rightarrow -1.64 = \frac{\bar{x}_{cr} - 30}{10} \Rightarrow \bar{x}_{cr} = 13.6$$

$$P\left(z > \frac{13.6 - 26}{10}\right) = P(z > -1.24) = 0.8925$$

Η πιθανότητα ότι ορθώς απορρίψαμε την H_0 είναι 0.8925.

$$-1.64 = Z_{cr} = \frac{\bar{x}_{cr} - \mu_0}{s/\sqrt{n}} \Rightarrow -1.64 = \frac{\bar{x}_{cr} - 30}{10} \Rightarrow \bar{x}_{cr} = 13.6$$

$$P\left(z > \frac{13.6 - 26}{10}\right) = P(z > -1.24) = 0.8925$$

ΑΝΑΛΥΣΗ ΤΗΣ ΔΙΑΚΥΜΑΝΣΗΣ (ANOVA)

Σύγκριση περισσοτέρων των δύο ανεξαρτήτων δειγμάτων από κανονικούς πληθυσμούς με ίδια διακύμανση

Συνολική εκτίμηση για το αν οι μέσοι όλων των δειγμάτων είναι μεταξύ τους ίσοι ή αν τουλάχιστον ένας από αυτούς διαφέρει.

Ελέγχουμε αν υπάρχει σημαντική διαφορά μεταξύ της διασποράς των μέσων τιμών και της συνολικής διασποράς των δειγμάτων.

$$s_\alpha^2 = \sum_1^\kappa n_i (\bar{x}_i - \bar{x})^2 \qquad s_\upsilon^2 = \sum_1^\kappa (n_i - 1) s_i^2$$

$$F = \frac{\overline{\text{τετράγωνα μεταξύ δειγμάτων}}}{\overline{\text{τετράγωνα εντός δειγμάτων}}} = \frac{s_\alpha^2/(\kappa-1)}{s_\upsilon^2/(n-\kappa)}$$

$n = n_1 + n_2 + \ldots + n_\kappa$

κ: πλήθος δειγμάτων

Hb_ANOVA.xls

H$_0$: $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = \cdots = \bar{x}_\kappa$

H$_1$: Τουλάχιστον ένας από τους δειγματικούς μέσους δεν είναι ίσος με κάποιον άλλον

ΠΙΝΑΚΑΣ ΑΝΑΛΥΣΗΣ ΔΙΑΣΠΟΡΑΣ

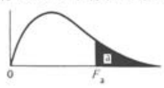
Μεταβολή	Αθρ. Τετραγ.	Β.ε.	Μέση μεταβλητ.	F
Μεταξύ δειγμάτων	s_α^2	κ-1	$s_\alpha^2/(\kappa-1)$	$\dfrac{s_\alpha^2/(\kappa-1)}{s_\upsilon^2/(n-\kappa)}$
Εντός δειγμάτων	s_υ^2	n-κ	$s_\upsilon^2/(n-\kappa)$	

Προϋπόθεση : διασπορές των δειγμάτων ίσες.

09__3_Example.xls

ΠΙΝΑΚΑΣ IV

Των τιμών $F_{v_1, v_2; a}$ της F-κατανομής για τις οποίες $P(F > F_{v_1, v_2; a}) = a$



Βαθμοί Ελευθερίας (a = .05) αριθμητής κ-1

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

Προϋπόθεση όλων των «παραμετρικών δοκιμασιών» που αναφέρθηκαν είναι η «κανονικότητα» των δεδομένων. Πώς ελέγχουμε την κανονικότητα;

- Γράφημα συσχέτισης των ποσοστιαίων σημείων (quantiles) των δεδομένων με αυτά της κανονικής κατανομής (**QQ-plot**).

  Ευθεία γραμμή → κανονικότητα

- Έλεγχος **λοξότητας (skewness)** και **κύρτωσης (kurtosis)**

  συντελεστής λοξότητας: μετρά την ασυμμετρία της κατανομής

  >0→ η δεξιά ουρά μεγαλύτερη από την αριστερή

  συντελεστής κύρτωσης: μετρά το πόσο πιο απότομη ή πιο επίπεδη είναι η κατανομή σε σχέση με την κανονική.

  >0→ απότομη στο κέντρο με μακρές ουρές

  <0→ επίπεδη στο κέντρο με μικρές ουρές

Αποδεχόμαστε ότι τα δεδομένα ακολουθούν την κανονική κατανομή, αν οι συντελεστές ασυμμετρίας και κύρτωσης είναι στο διάστημα (-1,1).

ΠΟΙΟΤΙΚΑ ΔΕΔΟΜΕΝΑ

Κατηγορίας (nominal) - Διάταξης (ordinal)

Ποσοστά - Σχετικές Συχνότητες - Αναλογίες

- $$p = \frac{\text{Πλήθος εμφανίσεων κάποιου φαινομένου}}{\text{Σύνολο παρατηρήσεων}}$$

- **Σε μεγάλα δείγματα ακολουθεί κανονική κατανομή**

  Δηλ. όταν *np(1-p)* $\geq$ ***10***

ΣΥΓΚΡΙΣΗ ΑΝΑΛΟΓΙΑΣ ΕΝΟΣ ΔΕΙΓΜΑΤΟΣ ΜΕ ΚΑΠΟΙΑ ΑΛΛΗ

(Η τ.μ. παίρνει δύο τιμές)

$$z = \frac{p - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Διάστημα εμπιστοσύνης : $= p \pm z_{1-a/2}\sqrt{\dfrac{p(1-p)}{n}}$

ΣΥΓΚΡΙΣΗ ΑΝΑΛΟΓΙΩΝ ΔΥΟ ΑΝΕΞΑΡΤΗΤΩΝ ΔΕΙΓΜΑΤΩΝ

$$z = \frac{p_1 - p_2}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$$

ΣΥΓΚΡΙΣΗ ΑΝΑΛΟΓΙΩΝ ΔΥΟ ΔΕΙΓΜΑΤΩΝ ΚΑΤΑ ΖΕΥΓΗ
Ανάλυση "χ^2" (Μη παραμετρικός έλεγχος)

- Η τ.μ. παίρνει δύο τιμές (Κατηγορίες 1, 2)
  - Πίνακες 2x2

Ομάδα	Κατηγορία 1	Κατηγορία 2	Σύνολο
1 2	π_{11} π_{21}	π_{12} π_{22}	$\pi_{11}+\pi_{12}=N_1$ $\pi_{21}+\pi_{22}=N_2$
Σύνολο	$\pi_{11}+\pi_{21}$	$\pi_{12}+\pi_{22}$	$N=N_1+N_2$

- Αναμενόμενες τιμές κατ. 1 ομάδας 1, 2

$\theta_{11} = (\pi_{11}+\pi_{12})*(\pi_{11}+\pi_{21})/N$, $\theta_{21} = (\pi_{21}+\pi_{22})*(\pi_{11}+\pi_{21})/N$

- Αναμενόμενες τιμές κατ. 2 ομάδας 1, 2

$\theta_{12} = (\pi_{11}+\pi_{12})*(\pi_{12}+\pi_{22})/N$, $\theta_{22} = (\pi_{21}+\pi_{22})*(\pi_{12}+\pi_{22})/N$

Εξετάζουμε την τιμή της μεταβλητής :

$$X^2 = \frac{(\pi_{11}-\theta_{11})^2}{\theta_{11}} + \frac{(\pi_{12}-\theta_{12})^2}{\theta_{12}} + \frac{(\pi_{21}-\theta_{21})^2}{\theta_{21}} + \frac{(\pi_{22}-\theta_{22})^2}{\theta_{22}}$$

ή $$X^2 = \frac{\pi_{11}^2}{\theta_{11}} + \frac{\pi_{12}^2}{\theta_{12}} + \frac{\pi_{21}^2}{\theta_{21}} + \frac{\pi_{22}^2}{\theta_{22}} - N$$

η οποία ακολουθεί κατανομή "χ^2" με β.ε.=1,

όταν οι αναμενόμενες αναλογίες δεν είναι < 5

Η χ^2 εξετάζει αμφίπλευρα, αλλά στον πίνακα αναζητούμε για σ.σ=0.05 και όχι 0.05/2.

x_square_schmerzlind.xls x_square_operation.xls

ΣΥΚΡΙΣΗ ΑΝΑΛΟΓΙΩΝ S ΔΕΙΓΜΑΤΩΝ ΜΕ K ΚΑΤΗΓΟΡΙΕΣ
Ανάλυση "χ^2"

Πίνακες Συνάφειας (contingency tables)

	Κατηγορία "j"				
Ομάδα "i"	1	2	.	k	Σύνολο γραμμών
1	π_{11}	π_{12}	.	π_{1k}	$n_{1k}=\Sigma n_{1j}$
2	π_{21}	π_{22}	.	π_{2k}	$n_{2k}=\Sigma n_{2j}$
.	.	.	.	.	.
s	π_{s1}	π_{s2}		π_{sk}	$n_{sk}=\Sigma n_{sj}$
Σύνολο στηλών	$n_{s1}=\Sigma n_{i1}$	$n_{s2}=\Sigma n_{i2}$	.	$n_{sk}=\Sigma n_{ik}$	$N=\Sigma\Sigma n_{ij}$

$$\theta_{ij} = \frac{(\sum_{i=1}^{s} n_{ij}) \cdot (\sum_{j=1}^{k} n_{ij})}{N}$$

$$X^2 = \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{(\pi_{ij} - \theta_{ij})^2}{\theta_{ij}}$$

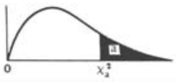
ή απλούστερα

$$X^2 = \sum_{i=1}^{s}\sum_{j=1}^{k} \left[\frac{\pi_{ij}^2}{\theta_{ij}} - N \right]$$

β.ε. = (*s*-1)×(*k*-1)

ΠΙΝΑΚΑΣ III

Των τιμών $X^2_{v,a}$ της X^2 κατανομής για τις οποίες $P(X^2 > X^2_{v,a}) = a$



B.ε	a = .995	a = .990	a = .975	a = .950	a = .900	a = .10	a = .05	a = .025	a = .010	a = .005
1	0.0000393	0.0001571	0.0009821	0.0039321	0.0157908	2.70554	3.84146	5.02389	6.63490	7.87944
2	0.0100251	0.0201007	0.0506356	0.102587	0.210720	4.60517	5.99147	7.37776	9.21034	10.5966
3	0.0717212	0.114832	0.215795	0.351846	0.584375	6.25139	7.81473	9.34840	11.3449	12.8381
4	0.206990	0.297110	0.484419	0.710721	1.063623	7.77944	9.48773	11.1433	13.2767	14.8602
5	0.411740	0.554300	0.831211	1.145476	1.61031	9.23635	11.0705	12.8325	15.0863	16.7496
6	0.675727	0.872085	1.237347	1.63539	2.20413	10.6446	12.5916	14.4494	16.8119	18.5476
7	0.989265	1.239043	1.68987	2.16735	2.83311	12.0170	14.0671	16.0128	18.4753	20.2777
8	1.344419	1.646482	2.17973	2.73264	3.48954	13.3616	15.5073	17.5346	20.0902	21.9550
9	1.734926	2.087912	2.70039	3.32511	4.16816	14.6837	16.9190	19.0228	21.6660	23.5893
10	2.15585	2.55821	3.24697	3.94030	4.86518	15.9871	18.3070	20.4831	23.2093	25.1882
11	2.60321	3.05347	3.81575	4.57481	5.57779	17.2750	19.6751	21.9200	24.7250	26.7569
12	3.07382	3.57056	4.40379	5.22603	6.30380	18.5494	21.0261	23.3367	26.2170	28.2995
13	3.56503	4.10691	5.00874	5.89186	7.04150	19.8119	22.3621	24.7356	27.6883	29.8194
14	4.07468	4.66043	5.62872	6.57063	7.78953	21.0642	23.6848	26.1190	29.1413	31.3193
15	4.60094	5.22935	6.26214	7.26094	8.54675	22.3072	24.9958	27.4884	30.5779	32.8013
16	5.14224	5.81221	6.90766	7.96164	9.31223	23.5418	26.2962	28.8454	31.9999	34.2672
17	5.69724	6.40776	7.56418	8.67176	10.0852	24.7690	27.5871	30.1910	33.4087	35.7185
18	6.26481	7.01491	8.23075	9.39046	10.8649	25.9894	28.8693	31.5264	34.8053	37.1564
19	6.84398	7.63273	8.90655	10.1170	11.6509	27.2036	30.1435	32.8523	36.1908	38.5822
20	7.43386	8.26040	9.59083	10.8508	12.4426	28.4120	31.4104	34.1696	37.5662	39.9968
21	8.03366	8.89720	10.28293	11.5913	13.2396	29.6151	32.6705	35.4789	38.9321	41.4010
22	8.64272	9.54249	10.9823	12.3380	14.0415	30.8133	33.9244	36.7807	40.2894	42.7956
23	9.26042	10.19567	11.6885	13.0905	14.8479	32.0069	35.1725	38.0757	41.6384	44.1813
24	9.88623	10.8564	12.4011	13.8484	15.6587	33.1963	36.4151	39.3641	42.9798	45.5585
25	10.5197	11.5240	13.1197	14.6114	16.4734	34.3816	37.6525	40.6465	44.3141	46.9278
26	11.1603	12.1981	13.8439	15.3791	17.2919	35.5631	38.8852	41.9232	45.6417	48.2899
27	11.8076	12.8786	14.5733	16.1513	18.1138	36.7412	40.1133	43.1944	46.9630	49.6449
28	12.4613	13.5648	15.3079	16.9279	18.9392	37.9159	41.3372	44.4607	48.2782	50.9933
29	13.1211	14.2565	16.0471	17.7083	19.7677	39.0875	42.5569	45.7222	49.5879	52.3356
30	13.7867	14.9535	16.7908	18.4926	20.5992	40.2560	43.7729	46.9792	50.8922	53.6720
40	20.7065	22.1643	24.4331	26.5093	29.0505	51.8050	55.7585	59.3417	63.6907	66.7659
50	27.9907	29.7067	32.3574	34.7642	37.6886	63.1671	67.5048	71.4202	76.1539	79.4900
60	35.5346	37.4848	40.4817	43.1879	46.4589	74.3970	79.0819	83.2976	88.3794	91.9517
70	43.2752	45.4418	48.7576	51.7393	55.3290	85.5271	90.5312	95.0231	100.425	104.215
80	51.1720	53.5400	57.1532	60.3915	64.2778	96.5782	101.879	106.629	112.329	116.321
90	59.1963	61.7541	65.6466	69.1260	73.2912	107.565	113.145	118.136	124.116	128.299
100	67.3276	70.0648	74.2219	77.9295	82.3581	118.498	124.342	129.561	135.807	140.169

ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ - ΣΥΣΧΕΤΙΣΗ
(Simple Linear Regression - Correlation)

Εύρεση μιας μαθηματικής ευθείας που εξηγεί τα δεδομένα

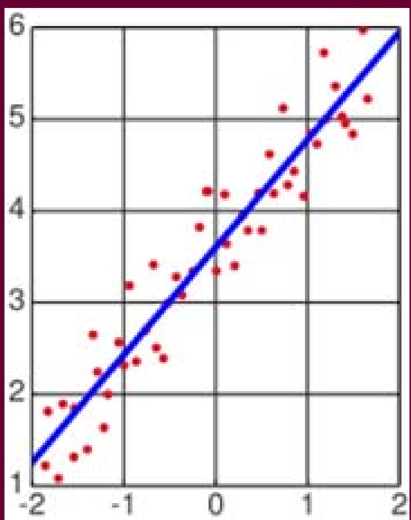
$$y = \alpha + \beta x + e$$

α : τεταγμένη επί την αρχή (intercept)   β : κλίση (slope)

e : τυχαίο σφάλμα
x : ελεγχόμενη (predictor)
y : απόκριση (response)

Μέθοδος ελαχίστων τετραγώνων

$$y_i = a + bx_i + e_i \qquad \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

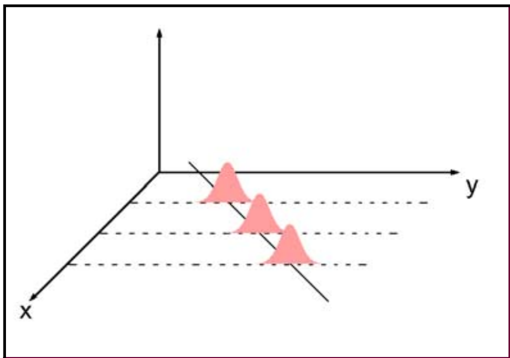


$$b = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} \qquad a = \overline{y} - b\overline{x}$$

Προϋποθέσεις για τη χρήση της απλής γραμμικής παλινδρόμησης

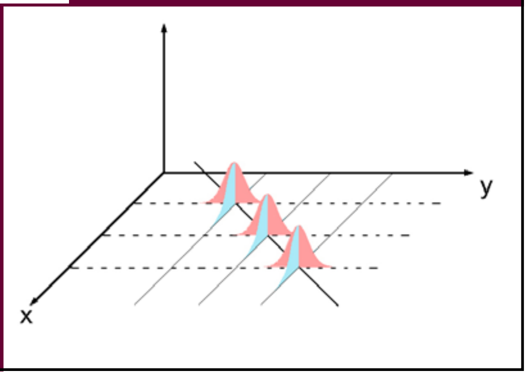
- **Τα τυχαία σφάλματα στην “x” είναι αμελητέα**
- **Για κάθε τιμή της “x” υπάρχει μια κανονική κατανομή τιμών της “y”.**
- **Η κατανομή του “y” για κάθε τιμή του “x” έχει την ίδια διακύμανση**

Linear_Regr.xls



Απλή γραμμική παλινδρόμηση

Παλινδρόμηση Deming



Τυπικό σφάλμα της εκτίμησης
(standard error of the estimation)

$$s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}}$$

Τυπικό σφάλμα για το “b”

δ.ε. β = b ± t·s_b, β. ε. n-2, a

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

Τυπικό σφάλμα για το “a”

δ.ε. α = a ± t·s_a, β. ε. n-2, a

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

ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ
(Multiple Linear Regression)

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e$$

Έστω δείγμα μεγέθους n

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_k x_{1k} + e_1$$
$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_k x_{2k} + e_2$$
.
$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_k x_{nk} + e_n$$

$$\boxed{Y = X \cdot \beta + e}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots x_{1k} \\ 1 & x_{21} & x_{22} & \dots x_{2k} \\ \dots \\ 1 & x_{n1} & x_{n2} & \dots x_{nk} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$

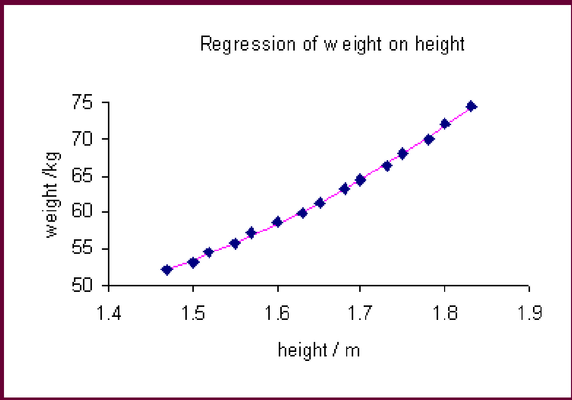
$$\widehat{\beta} = \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \dots \\ \widehat{\beta}_n \end{bmatrix}$$

$$\widehat{\beta} = \left(X'X\right)^{-1} X'Y$$

Απώλεια βάρους Y[gr]	Χρόνος έκθεσης X_1[h]	Σχετική υγρασία X_2
4.3	4	0.2
5.5	5	0.2
6.8	6	0.2
8.0	7	0.2
4.0	4	0.3
5.2	5	0.3
6.6	6	0.3
7.5	7	0.3
2.0	4	0.4
4.0	5	0.4
5.7	6	0.4
6.5	7	0.4

$$\widehat{y} = 5.49 + 1.32x_1 - 8x_2$$

ΜΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ (Non Linear Regression)



$$y = \alpha + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + e$$

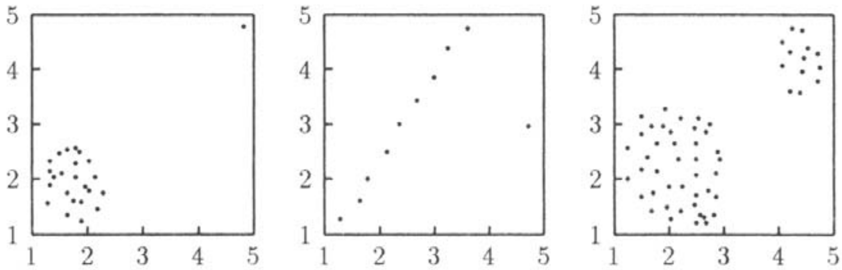
ΣΥΣΧΕΤΙΣΗ (Correlation)

Μέθοδος για την μέτρηση του βαθμού συμμεταβλητότητας των μεταβλητών.

Συντελεστής συσχέτισης Pearson

(Correlation coefficient)

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



$r = 0.70$

Διάστημα εμπιστοσύνης

(Confidence interval)

$$y \pm s \cdot t_{n-2;a/2}\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

Διάστημα πρόβλεψης

(Prediction interval)

$$y \pm s \cdot t_{n-2;a/2}\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

y **: τιμή που προβλέπεται από την γρ. παλ. για το “x”**

$\hat{x}$ **: μέση τιμή του “x”**

x **: τιμή της ανεξ. μετ. για την οποία αναζητούμε την “y”**

x_i **: τιμή της ανεξ. μετ. από τις μετρήσεις**

Δοκιμασία ανεξαρτησίας

H_0 : b=0,

απορρίπτεται όταν $t > t_{n-2;a/2}$

$$t = \frac{b}{s_b}$$

Δοκιμασία μη συσχέτισης

H_0 : r=0,

απορρίπτεται όταν $t > t_{n-2;a/2}$

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$$

Γραμμική παλινδρόμηση Deming

$$Deming\ b = U + \sqrt{U^2 + \frac{1}{\lambda}} \qquad \lambda = \frac{S_x^2}{S_y^2}$$

$$U = \frac{\sum_{i=1}^{N}(y_i - \overline{y})^2 - \lambda^{-1} \cdot \sum_{i=1}^{N}(x_i - \overline{x})^2}{2\sum_{i=1}^{N}(y_i - \overline{y})(x_i - \overline{x})}$$

Μη παραμετρική συσχέτιση
Συντελεστής συσχέτισης του Spearman

- Συσχετίζει τάξεις τ.μ.
- Έστω «n» ζεύγη τ.μ. x_i, y_i.
- Δημιουργούμε τις διαφορές των τάξεων των ζευγών των τ.μ.

$$d_i = Tx_i - Ty_i$$

$$r_T = 1 - \frac{6 \cdot (d_1^2 + d_2^2 + \cdots + d_n^2)}{n^3 - n}$$

np_Spearman_10_21_Exer.xls

Δοκιμασία ανεξαρτησίας

H_0: $r_T=0$

- $n \le 5$ Δεν μπορούμε να αποφανθούμε
- $5 < n \le 10$ Απορρίπτουμε την H_0 όταν $|r_T| \ge S_{n;a}$
- $n > 10$ Το $t_T = \frac{r_T \cdot \sqrt{n-2}}{\sqrt{1-r_T^2}}$ ακολουθεί την κατανομή "t".

Απορρίπτουμε την H_0 όταν $|t_T| \ge t_{n-2;a}$

ΜΗ ΠΑΡΑΜΕΤΡΙΚΕΣ ΔΟΚΙΜΑΣΙΕΣ

Δεν προαπαιτούν να ακολουθούν οι πληθυσμοί κάποια κατανομή.

Ενδιαφέρει μόνο η τάξη και όχι η τιμή της τυχαίας μεταβλητής.

- **Δοκιμασία Kolmogorov –Smirnov**
  (Καλής προσαρμογής, Ομογένειας)
- **Κριτήριο των Ροών ή Wald-Wolfowitz**
  (Δοκιμασία τυχαιότητας)
- **Δοκιμασία προσήμου (sign test)**
- **Αθροίσματα τάξεων (rank sum test)**
  - **Δοκιμασία Wilcoxon**
  - **Δοκιμασία Mann-Whithey**
  - **Δοκιμασία Kruskal-Wallis**

<u>ΔΟΚΙΜΑΣΙΑ ΟΜΟΓΕΝΕΙΑΣ (Kolmogorov – Smirnov)</u>

- **Ελέγχει αν δύο δείγματα προέρχονται από την ίδια κατανομή.**

H_0= Τα δύο δείγματα προέρχονται από την ίδια κατανομή

H_1= Τα δύο δείγματα προέρχονται διαφορικές κατανομές

- **Χρησιμοποιεί τον δείκτη $D = \max |F_1(x) - F_2(x)|$**

  $F_1(x)$ και $F_2(x)$ είναι οι αθροιστικές συχνότητες των δύο δειγμάτων

- **Η μηδενική υπόθεση απορρίπτεται, όταν το $D \geq D_{a,m,n}$ από τον πίνακα VII.**

Για μεγάλα m,n (>15):

α=0.05 $$D_{a,m,n} = 1.36\sqrt{\frac{m+n}{m\cdot n}}$$

α=0.01 $$D_{a,m,n} = 1.63\sqrt{\frac{m+n}{m\cdot n}}$$

10__2_Example.xls

ΔΟΚΙΜΑΣΙΑ ΤΩΝ ΡΟΩΝ ή WALD-WOLFOWITZ

(Δοκιμασία τυχαιότητας)

Εξετάζει αν μια ακολουθία παρατηρήσεων είναι τυχαία ή όχι. Π.χ.

ΕΕ ΑΑ ΕΕ ΑΑΑ Ε ΑΑΑ ΕΕΕΕ ΑΑΑ

Κάθε διαδοχή ομοίων συμβόλων λέγεται «ροή».

Το πλήθος των συμβόλων μιας ροής λέγεται «μήκος ροής».

Στο παράδειγμα έχουμε 8 ροές

n_1=9 Ε, n_2=11 Α

Εάν n_1, n_2 $\geq$ 10 τότε το πλήθος των ροών U ακολουθεί την

$N(\mu_u,\sigma^2)$ με $\mu_u = \frac{2n_1n_2}{n_1+n_2}+1$ και $\sigma_u^2 = \frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$

$$Z = \frac{U-\mu_u}{\sigma_u}$$

Alignment.xls

ΔΟΚΙΜΑΣΙΑ ΠΡΟΣΗΜΟΥ (Sign test)

- Ανάλογη προς τη δοκιμασία *"t"*
- Χρησιμοποιεί τον διάμεσο (median)
- Ζευγαρωτές παρατηρήσεις

H_0: $M_1 = M_2$

Η H_0 γίνεται αποδεκτή, όταν το πλήθος των αρνητικών (N_-) και θετικών (N_+) διαφορών είναι "ίσο". Αυτό συμβαίνει όταν το N_- εμπίπτει στο διάστημα εμπιστοσύνης που προβλέπεται από πίνακες της διωνυμικής κατανομής για το πλήθος των μη μηδενικών διαφορών και για συγκεκριμένο "a".

Sign_test.xls

Όρια εμπιστοσύνης για το N_p (διωνυμική κατανομή), $p=0.05$; $N=0$ έως 99

N	0	1	2	3	4	5	6	7	8	9
0	-	-	-	-	-	-	0-6	0-7	0-8	1-8
10	1-9	1-10	2-10	2-11	2-12	3-12	3-13	4-13	4-14	4-15
20	5-15	5-16	5-17	6-17	6-18	7-18	7-19	7-20	8-20	8-21
30	9-21	9-22	9-23	10-23	10-24	11-24	11-25	12-25	12-26	12-27
40	13-27	13-28	14-28	14-29	15-29	15-30	15-31	16-31	16-32	17-32
50	17-33	18-33	18-34	18-35	19-35	19-36	20-36	20-37	21-37	21-38
60	21-39	22-39	22-40	23-40	23-41	24-41	21-42	25-42	25-43	25-44
70	26-44	26-45	27-45	27-46	28-46	28-47	28-48	29-48	29-49	30-49
80	30-50	31-50	31-51	32-51	32-52	32-53	33-53	33-54	34-54	34-55
90	35-55	35-56	36-56	36-57	37-57	37-58	37-59	38-59	38-60	39-60

N: πλήθος μη μηδενικών διαφορών

ΔΟΚΙΜΑΣΙΑ ΑΘΡΟΙΣΜΑΤΩΝ ΤΑΞΕΩΝ ΚΑΤΑ WILCOXON

- **Ζευγαρωτές παρατηρήσεις ($n_1=n_2$)**
- **H_0: $M_1 = M_2$**
- **Κατατάσσουμε τις απόλυτες τιμές των διαφορών κατά αύξουσα σειρά.**
- **Η σειρά εμφάνισης των διαφορών αποτελεί την τάξη τους.**
- **Γράφουμε δίπλα σε κάθε διαφορά την τάξη της και το πρόσημό της.**
- **Αθροίζουμε τις τάξεις των θετικών διαφορών (T_+)**
- **Αθροίζουμε τις τάξεις των αρνητικών διαφορών (T_-)**

- **Η μηδενική υπόθεση απορρίπτεται όταν:**

  $T = \min\{T_+, T_-\} \leq T_c$. ($T_c$ από τον πίνακα XI, n=πλήθος μη μηδενικών παρατηρήσεων)

- **Για n > 25 η T ακολουθεί κανονική κατανομή.**

$$Z = \frac{T - \mu_T}{\sigma_T}$$

$$\mu_T = \frac{n(n+1)}{4} \qquad \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Εναλλακτικά, όταν $T = T_+ - T_- > T_p$

(n: πλήθος μη μηδενικών διαφορών)

Wilc_Sign_Rank_Sum.xls

10_12_Example.xls

ΠΙΝΑΚΑΣ XI

Πίνακας κρίσιμων τιμών του Τ στο test **Wilcoxon** για ζευγαρωτές παρατηρήσεις

$n = 5(1)50$

Μονόπλευρο test a	Δίπλευρο test a	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$	$n = 11$	$n = 12$	$n = 13$	$n = 14$	$n = 15$	$n = 16$
.05	.10	1	2	4	6	8	11	14	17	21	26	30	36
.025	.05		1	2	4	6	8	11	14	17	21	25	30
.01	.02			0	2	3	5	7	10	13	16	20	24
.005	.01				0	2	3	5	7	10	13	16	19
		$n = 17$	$n = 18$	$n = 19$	$n = 20$	$n = 21$	$n = 22$	$n = 23$	$n = 24$	$n = 25$	$n = 26$	$n = 27$	$n = 28$
.05	.10	41	47	54	60	68	75	83	92	101	110	120	130
.025	.05	35	40	46	52	59	66	73	81	90	98	107	117
.01	.02	28	33	38	43	49	56	62	69	77	85	93	102
.005	.01	23	28	32	37	43	49	55	61	68	76	84	92
		$n = 29$	$n = 30$	$n = 31$	$n = 32$	$n = 33$	$n = 34$	$n = 35$	$n = 36$	$n = 37$	$n = 38$	$n = 39$	
.05	.10	141	152	163	175	188	201	214	228	242	256	271	
.025	.05	127	137	148	159	171	183	195	208	222	235	250	
.01	.02	111	120	130	141	151	162	174	186	198	211	224	
.005	.01	100	109	118	128	138	149	160	171	183	195	208	
		$n = 40$	$n = 41$	$n = 42$	$n = 43$	$n = 44$	$n = 45$	$n = 46$	$n = 47$	$n = 48$	$n = 49$	$n = 50$	
.05	.10	287	303	319	336	353	371	389	408	427	446	466	
.025	.05	264	279	295	311	327	344	361	379	397	415	434	
.01	.02	238	252	267	281	297	313	329	345	362	380	398	
.005	.01	221	234	248	262	277	292	307	323	339	356	373	

Όρια ασφαλείας του δείκτη για τη δοκιμασία προσήμου αθροιστικών τάξεων κατά Wilcoxon

	a=0.05	a=0.025
Πλήθος διαφορών≠0	$T_{0.95}$	$T_{0.975}$
5	15	-
6	17	21
7	22	24
8	26	30
9	29	35
10	35	39
11	40	46
12	44	52
13	49	57
14	55	63
15	60	70
16	66	76
17	71	83
18	77	91
19	84	98
20	90	106

Για n > 20

$$T_p = z_p \sqrt{\frac{n(n+1)(2n+1)}{6}}$$

Z_p : το p εκατοστιαίο σημείο της κανονικής κατανομής

ΔΟΚΙΜΑΣΙΑ ΑΘΡΟΙΣΜΑΤΩΝ ΤΑΞΕΩΝ ΚΑΤΑ MANN-WHITNEY (U-test)

- Δείγματα διαφορετικού μεγέθους n_1,n_2
- Διατάσσουμε τις παρατηρήσεις των δύο δειγμάτων σε αύξουσα σειρά, σαν να ανήκαν στον ίδιο πληθυσμό.
- Σημειώνουμε την τάξη κάθε παρατήρησης.
- Υπολογίζουμε τα αθροίσματα T_1 και T_2 των τάξεων για κάθε δείγμα.
- $T = \min \{T_1,T_2\}$
- Εξετάζουμε, αν το T εμπίπτει στα όρια αποδοχής που βρίσκουμε σε πίνακες για n_1, n_2 και συγκεκριμένη σ.σ.

Στον πίνακα το N_1 είναι το n του δείγματος με το μικρότερο T.

Acceptance region for the rank sum T (Mann-Whitney-Wilcoxon 2 sample test), $p = 0.05$

N_1 / N_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2	—	—	—	—	—	—	—	36-52	45-63	55-75	66-88	79-101	92-116	106-132	121-149
3	—	—	—	—	15-30	22-38	29-48	38-58	47-70	58-82	69-96	82-110	95-126	110-142	125-160
4	—	—	—	10-26	16-34	23-43	31-53	40-64	49-77	60-90	72-104	85-119	99-135	114-152	130-170
5	—	—	6-21	11-29	17-38	24-48	33-58	42-70	52-83	63-97	75-112	89-127	103-144	118-162	134-181
6	—	—	7-23	12-32	18-42	26-52	34-64	44-76	55-89	64-104	79-119	92-136	107-153	122-172	139-191
7	—	—	7-26	13-35	20-45	27-57	36-69	46-82	57-96	69-111	82-127	96-144	111-162	127-181	144-201
8	—	3-19	8-28	14-38	21-49	29-61	38-74	49-87	60-102	72-118	85-135	100-152	115-171	131-191	149-211
9	—	3-21	8-31	14-42	22-53	31-65	40-79	51-93	62-109	75-125	89-142	104-160	119-180	136-200	154-221
10	—	3-23	9-33	15-45	23-57	32-70	42-84	53-99	65-115	78-132	92-150	107-169	124-188	141-209	159-231
11	—	3-25	9-36	16-48	24-61	34-74	44-89	55-105	68-121	81-139	96-157	111-177	128-197	145-219	164-241
12	—	4-26	10-38	17-51	26-64	35-79	46-94	58-110	71-127	84-146	99-165	115-185	132-206	150-228	169-251
13	—	4-28	10-41	18-54	27-68	37-83	48-99	60-116	73-134	88-152	103-172	119-193	136-215	155-237	174-261
14	—	4-30	11-43	19-57	28-72	38-88	50-104	62-122	76-140	91-159	106-180	123-201	141-223	160-246	179-271
15	—	4-32	11-46	20-60	29-76	40-92	52-109	65-127	79-146	94-166	110-187	127-209	145-232	164-256	184-281

ΔΟΚΙΜΑΣΙΑ KRUSKAL - WALLIS

- Μη παραμετρικό ανάλογο της ANOVA
- Η τ.μ. πρέπει να έχει συνεχή κατανομή και να είναι διατάξιμη.
- Χρησιμοποιούμε τις τάξεις των παρατηρήσεων και τις υποβάλουμε σε “ανάλυση των διακυμάνσεων”.
- Υπολογίζουμε τον δείκτη *“H”* ο οποίος ακολουθεί την κατανομή “χ^2”

$$H = \frac{12}{n(n+1)} \cdot \sum_{j=1}^{\kappa} \frac{R_j^2}{n_j} - 3(n+1)$$

κ: πλήθος δειγμάτων, n_j: πλήθος του δείγματος “j”, $n=\Sigma n_j$

R_j: άθροισμα των τάξεων στο δείγμα “j”

Για κ=3 και n_j μέχρι 5 ανατρέχουμε στον πίνακα XII.

Στην περίπτωση πολλαπλότητας «μ» των τιμών διορθώνουμε την τιμή του H διαιρώντας με τον αριθμό C, που λαμβάνει υπόψη το άθροισμα των κύβων των βαθμών πολλαπλότητας καθώς και το άθροισμά των:

$$C = 1 - \frac{\left(\mu_1^3 + \mu_2^3 + \ldots + \mu_{\rho 1}^3\right) - \left(\mu_1 + \mu_{21} + \ldots + \mu_\rho\right)}{n^3 - n}$$

Παραμετρικές δοκιμασίες	Προϋποθέσεις παραμετρικών δοκιμασιών	Μη παραμετρικές εναλλακτικές
Δύο ανεξάρτητα δείγματα Student's t test	1) Τα δεδομένα και των δύο δειγμάτων έχουν συλλεγεί τυχαία 2) Τα δεδομένα και των δύο δειγμάτων προέρχονται από πληθυσμούς με κανονική κατανομή 3) Οι διακυμάνσεις τους είναι ίσες	Mann-Whitney U test
Κατά ζεύγη Student's t test	1) Η διαφορές (d_i) πρέπει να προέρχονται από πληθυσμό διαφορών με κανονική κατανομή	Wilcoxon signed rank
ANOVA	1) Τα δεδομένα όλων των δειγμάτων έχουν συλλεγεί τυχαία 2) Τα δεδομένα όλων των δειγμάτων προέρχονται από πληθυσμούς με κανονική κατανομή 3) Οι διακυμάνσεις τους είναι ίσες	Kruskal-Wallis H test
Συντελεστής συσχέτισης του Pearson	1) Τα δεδομένα Y για κάθε X πρέπει να έχουν συλλεγεί τυχαία από κανονική κατανομή των τιμών Y. 2) Τα δεδομένα X για κάθε Y πρέπει να έχουν συλλεγεί τυχαία από κανονική κατανομή των τιμών X.	Συντελεστής συσχέτισης τάξεως του Spearman