統計学の基本

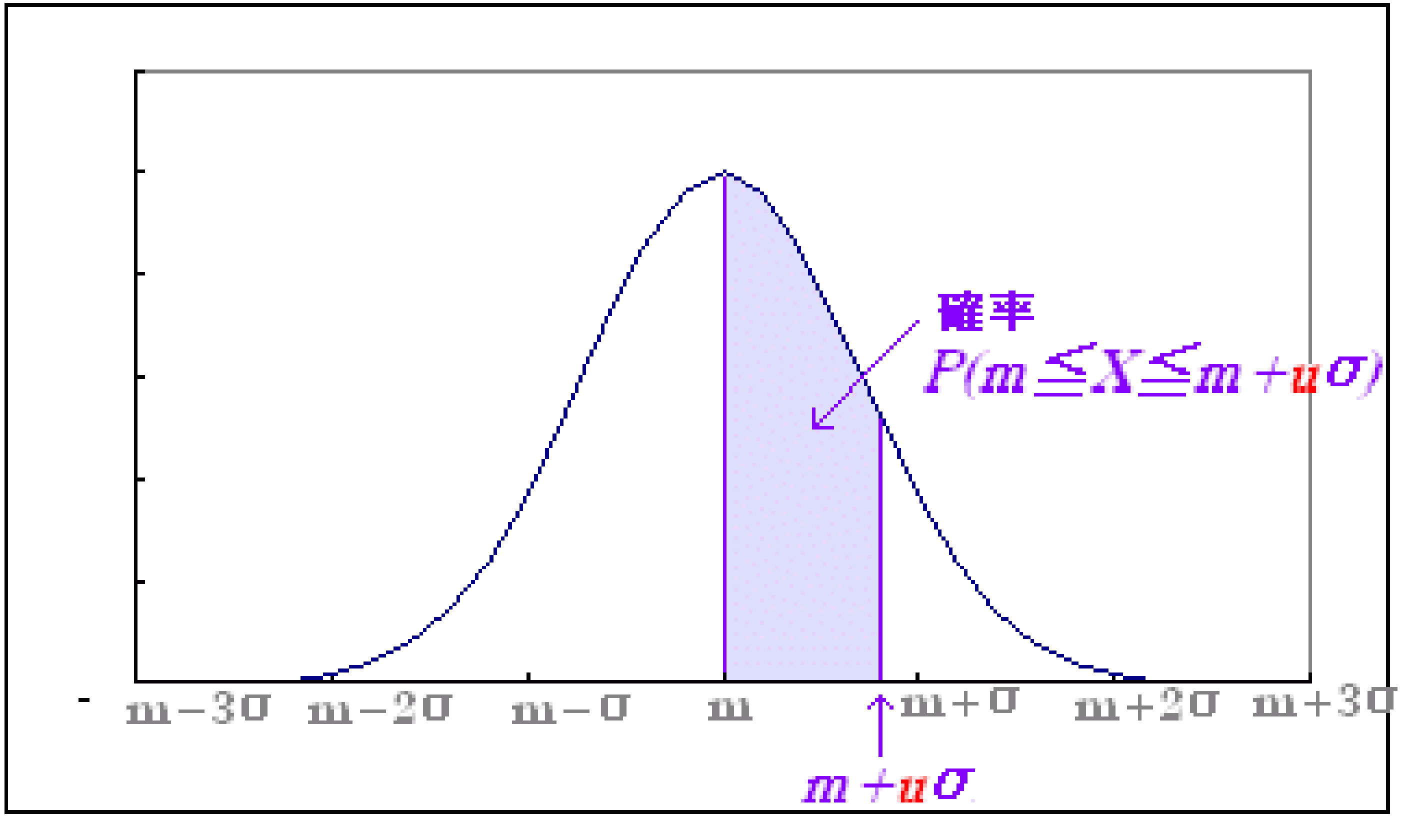
国際総合学類
内藤久裕

平均と標準偏差

- 今模擬試験の点数を考える。
- 平均は、点数の総和を受験生の数で割ったもの
- 標準偏差は、各個人の点数から模擬試験の平均点を差し引き、それを２乗し、全員に関して足し合わせ、平方根をとったもの。
- 直感的には、平均した「平均からのかい離」をあらわす標準偏差は、分布の散らばり具合をあらわす。

正規分布と標準正規分布

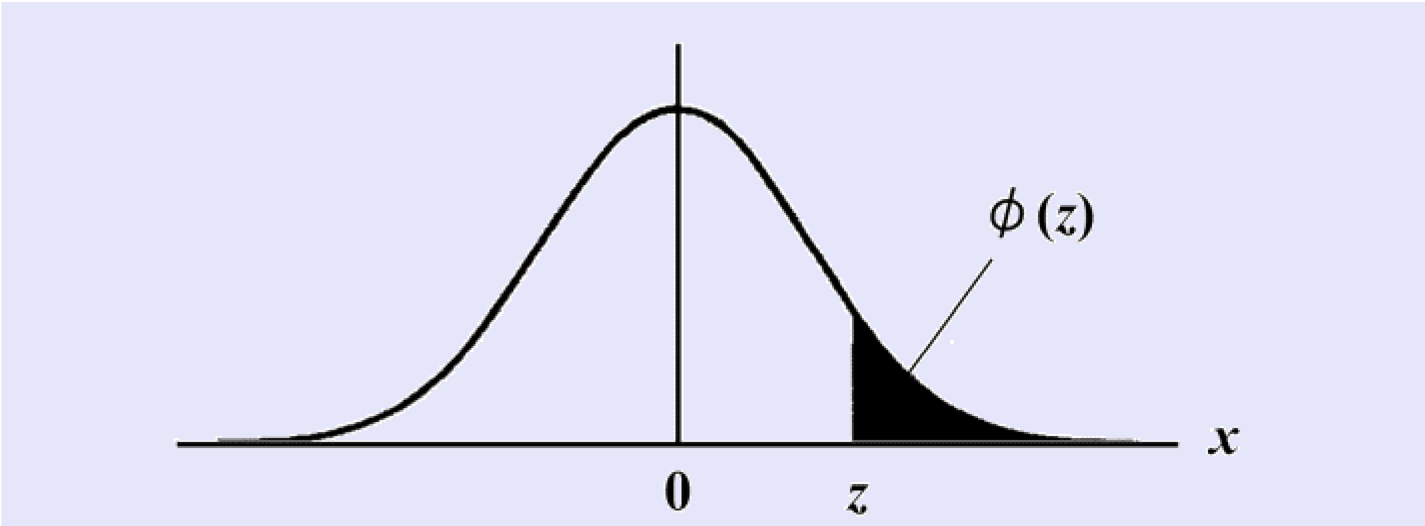
- 平均がm,標準偏差がmuの正規分布は次の通り


確率
$P(m \leqq X \leqq m+u\sigma)$
$m-3\sigma$ $m-2\sigma$ $m-\sigma$ m $m+\sigma$ $m+2\sigma$ $m+3\sigma$
$m+u\sigma$

統計学の基本

- 特に平均が０、標準偏差（分散の平方根）が１の場合、標準正規分布と呼ぶ
- 標準正規分布は、その特性が詳しく調べられているので、標準正規分布に変換し直して、統計的な検定を行う。
- 具体的には、平均　mu, 標準偏差sigmaで正規分布に従う変数xがあるとする。この時
- (X-mu)/sigmaは標準正規分布にしたがう。

標準正規分布



標準正規分布表

正規分布	z	0	1	2	3	4	5	6	7	8	9
	0	0.5	0.496011	0.492022	0.488033	0.484047	0.480061	0.476078	0.472097	0.468119	0.46414 4
	0.1	0.460172	0.456205	0.452242	0.448283	0.44433	0.440382	0.436441	0.432505	0.428576	0.424655
	0.2	0.42074	0.416834	0.412936	0.409046	0.405165	0.401294	0.397432	0.39358	0.389739	0.385908
	0.3	0.382089	0.378281	0.374484	0.3707	0.366928	0.363169	0.359424	0.355691	0.351973	0.348268
	0.4	0.344578	0.340903	0.337243	0.333598	0.329969	0.326355	0.322758	0.319178	0.315614	0.312067
	0.5	0.308538	0.305026	0.301532	0.298056	0.294598	0.29116	0.28774	0.284339	0.280957	0.277595
	0.6	0.274253	0.270931	0.267629	0.264347	0.261086	0.257846	0.254627	0.251429	0.248252	0.245097
	0.7	0.241964	0.238852	0.235762	0.232695	0.22965	0.226627	0.223627	0.22065	0.217695	0.214764
	0.8	0.211855	0.20897	0.206108	0.203269	0.200454	0.197662	0.194894	0.19215	0.18943	0.186733
	0.9	0.18406	0.181411	0.178786	0.176186	0.173609	0.171056	0.168528	0.166023	0.163543	0.161087
σ	1	0.158655	0.156248	0.153864	0.151505	0.14917	0.146859	0.144572	0.14231	0.140071	0.137857
	1.1	0.135666	0.1335	0.131357	0.129238	0.127143	0.125072	0.123024	0.121001	0.119	0.117023
	1.2	0.11507	0.11314	0.111233	0.109349	0.107488	0.10565	0.103835	0.102042	0.100273	0.098525
	1.3	0.096801	0.095098	0.093418	0.091759	0.090123	0.088508	0.086915	0.085344	0.083793	0.082264
	1.4	0.080757	0.07927	0.077804	0.076359	0.074934	0.073529	0.072145	0.070781	0.069437	0.068112
	1.5	0.066807	0.065522	0.064256	0.063008	0.06178	0.060571	0.05938	0.058208	0.057053	0.055917
	1.6	0.054799	0.053699	0.052616	0.051551	0.050503	0.049471	0.048457	0.04746	0.046479	0.045514
	1.7	0.044565	0.043633	0.042716	0.041815	0.040929	0.040059	0.039204	0.038364	0.037538	0.036727
	1.8	0.03593	0.035148	0.034379	0.033625	0.032884	0.032157	0.031443	0.030742	0.030054	0.029379
	1.9	0.028716	0.028067	0.027429	0.026803	0.02619	0.025588	0.024998	0.024419	0.023852	0.023295
2σ	2	0.02275	0.022216	0.021692	0.021178	0.020675	0.020182	0.019699	0.019226	0.018763	0.018309

標準正規分布表（続き）

[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	1.1	0.135666	0.1335	0.131357	0.129238	0.127143	0.125072	0.123024	0.121001	0.119	0.117023
	1.2	0.11507	0.11314	0.111233	0.109349	0.107488	0.10565	0.103835	0.102042	0.100273	0.098525
	1.3	0.096801	0.095098	0.093418	0.091759	0.090123	0.088508	0.086915	0.085344	0.083793	0.082264
	1.4	0.080757	0.07927	0.077804	0.076359	0.074934	0.073529	0.072145	0.070781	0.069437	0.068112
	1.5	0.066807	0.065522	0.064256	0.063008	0.06178	0.060571	0.05938	0.058208	0.057053	0.055917
	1.6	0.054799	0.053699	0.052616	0.051551	0.050503	0.049471	0.048457	0.04746	0.046479	0.045514
	1.7	0.044565	0.043633	0.042716	0.041815	0.040929	0.040059	0.039204	0.038364	0.037538	0.036727
	1.8	0.03593	0.035148	0.034379	0.033625	0.032884	0.032157	0.031443	0.030742	0.030054	0.029379
	1.9	0.028716	0.028067	0.027429	0.026803	0.02619	0.025588	0.024998	0.024419	0.023852	0.023295
2σ	2	0.02275	0.022216	0.021692	0.021178	0.020675	0.020182	0.019699	0.019226	0.018763	0.018309
	2.1	0.017864	0.017429	0.017003	0.016586	0.016177	0.015778	0.015386	0.015003	0.014629	0.014262
	2.2	0.013903	0.013553	0.013209	0.012874	0.012545	0.012224	0.011911	0.011604	0.011304	0.011011
	2.3	0.010724	0.010444	0.01017	0.009903	0.009642	0.009387	0.009137	0.008894	0.008656	0.008424
	2.4	0.008198	0.007976	0.00776	0.007549	0.007344	0.007143	0.006947	0.006756	0.006569	0.006387
	2.5	0.00621	0.006037	0.005868	0.005703	0.005543	0.005386	0.005234	0.005085	0.00494	0.004799
	2.6	0.004661	0.004527	0.004397	0.004269	0.004145	0.004025	0.003907	0.003793	0.003681	0.003573
	2.7	0.003467	0.003364	0.003264	0.003167	0.003072	0.00298	0.00289	0.002803	0.002718	0.002635
	2.8	0.002555	0.002477	0.002401	0.002327	0.002256	0.002186	0.002118	0.002052	0.001988	0.001926
	2.9	0.001866	0.001807	0.00175	0.001695	0.001641	0.001589	0.001538	0.001489	0.001441	0.001395
3σ	3	0.00135	0.001306	0.001264	0.001223	0.001183	0.001144	0.001107	0.00107	0.001035	0.001001
	3.1	0.000968	0.000936	0.000904	0.000874	0.000845	0.000816	0.000789	0.000762	0.000736	0.000711

統計学の基本

- 標準正規分布に従う変数 z が1.96以上をとる確率は、0.02498ほぼ、2.5%
- つまり標準正規分布に従う変数 z が1.96以上をあるいは、-1.96以下をとる確率は、5%
- 正規分布から標準正規分布への変換を利用したのが模擬試験でつかう偏差値である。

模擬試験の偏差値

- 大規模なテストの点数は正規分布に従うことが知られている。
- そこで、i君のテストの点をX_iとする、この模擬試験のテストの平均点をmu、標準偏差をsigmaとすると
- (X_i-mu)/sigmaは標準正規分布になる。
- 標準正規分布表から、i君は上位何パーセントにいるかわかる。

模擬試験の偏差値

- たとえば、i君が英語の模擬試験をうけて、点数が８０点。このテストでは、平均点が６０点、標準偏差が１０点だったとしよう。
- この場合、標準化された点数は
- (80-60)/10=2
- ここで標準正規分布表をみると、２より大きくなる確率は、0.02275
- つまりi君は上位2.275%にいることがわかる。

統計学の基本

- ただし、標準化されたテストの点は、平均０でだいたい[-2.5,2.5]ぐらいのところに散らばっている。
- 標準化されたテストの点が、0.3　点とか,

１点とかは直観的に分かりにくい。

- そこで標準化されたテストの点を１０倍して、そしてそれに５０を足した値を偏差値としている。つまり５０を中心にばらつくように変換する

統計学の基本

- つまり、偏差値が７０の場合は、標準化された点数の値は２であるということが分かる。
- この場合、この人は、上位２．３パーセントにいることが分かる。
- 質問、あるひとが偏差値で７５を取ったとしよう。この人は上位何パーセントに入っているだろうか。
- 標準化された点数は、(75-50)/10=2.5
- 標準正規分布表より、2.5より多くなる確率は0.006.つまり、上位0.6パーセントに属する

統計学の基本

- 回帰分析とは、データをプロットして一番フィットするように直線を当てはめる方法
- 今、推定された直線の傾き、beta1^hatを考える
- beta1_hatは平均beta1,標準偏差SEの正規分布に従うことが知られている。
- つまり(beta1^hat-beta1)/SEは、標準正規分布に従う
- ただし、ここでbeta1は分析者にとっては、未知の変数
- beta1_hatが回帰分析から(stataで）得られた値
- SEは、standard errorの略で、beta1_hatの標準偏差の事。これもstataが計算してくれる。

仮説検定の考え方

- データの分析者にとって、beta1^hatは回帰分析から計算できる。
- ただし、真の値 beta1は観察できない。
- 今,beta1=0であることが、本当らしいか、ありそうにないかをチェックしたい。
- つまり説明変数　ｘ１のｙに与える影響が０であることは、本当か、ありそうにないか、検討したい。

仮説検定の考え方

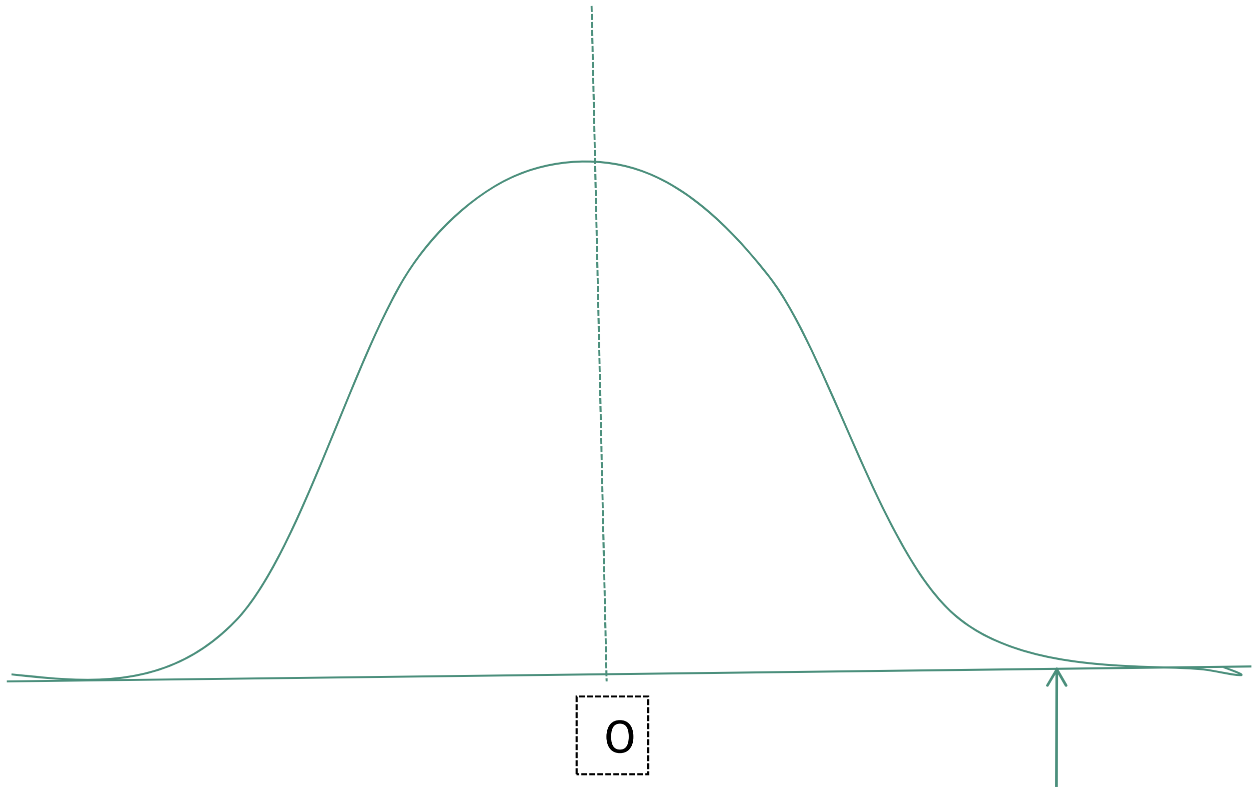
- 直観的に考えて、今beta1=0が本当かどうかを検定したい場合、beta1^hatが0から遠く離れていれば、真の値beta1が０ということはあり得ないのではないかと推論してよい。
- なぜなら、beta1^hatは、beta1を平均として正規分布に従うから。
- 「beta1^hatが0から遠く離れていれば」と書いたが、遠く離れているという、その遠さはどのように定義するのか。

仮説検定の考え方

- ここでbeta1^hatに関する正規分布の性質を使う。
- (beta1^hat-beta1)/SEは標準正規分布に従う。
- つまり (beta1^hat-beta1)/SEは平均０、標準偏差１の正規分布にしたがう。
- 仮に、ここでbeta1=0であるとすると、
- 　beta1^hat/SEが標準正規分布に従うことになる

仮説検定の考え方

- いまここで、 beta1^hat/se を計算して、この値が３であったとしよう。



３はこのあたり
これより右にくる確率は
１％以下。

仮説検定の考え方

- つまり、beta1=0とした場合、beta1^hat/se＝３となる場合、そのような事が起こる確率はほぼ１％以下と言える。
- つまりそのようなケースは、ほぼおこらないと言える。
- つまりbeta1=0というケースはあり得そうでないといえる。
- この場合、beta1^hatを統計的に有意で、0ではないという。

具体的な方法

- 具体的には、通常５％の有意水準を定める。
- 標準正規分布でこの値は、1.96か-1.96
- (beta1^hat-beta1)/seが標準正規分布に従うことは、分かっている。
- beta1=0と仮定してみる。
- するとbeta1^hat/seが標準正規分布に従うことになる。

具体的な方法

- beta1^hat/seを計算し、これが1.96以上かあるいは、-1.96以下かチェックする
- これが1.96以上あるいは、-1.96以下であると、５%の以下の事が起こったということになる。
- 起こり得難いことが起こってしまった。
- 仮定がおかしい。
- 　beta1=0は、ありえない、と結論する。