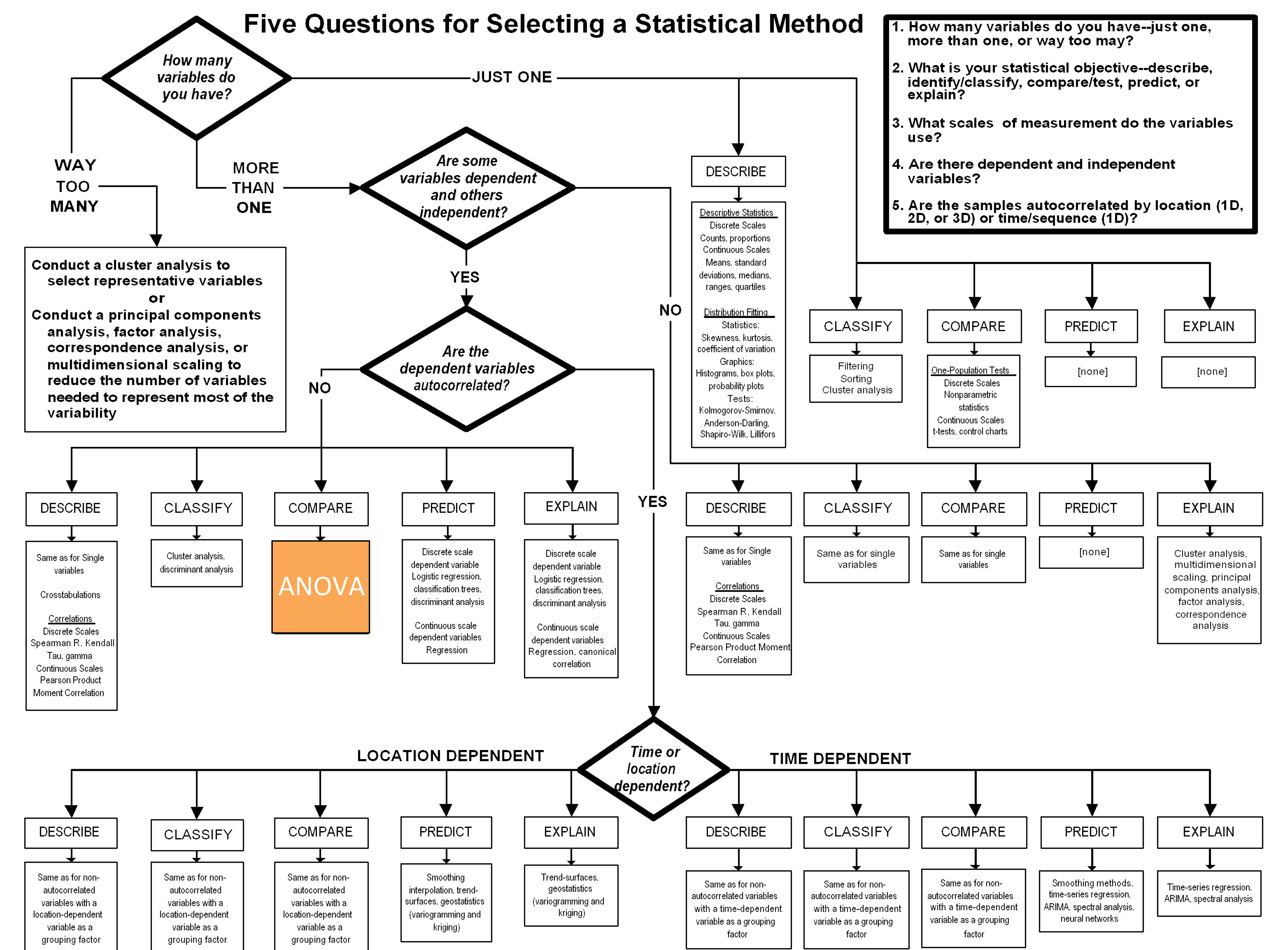
BIOL933 Design, Analysis, and Interpretation of Experiments

Class 1
Aug 29, 2017

$$
\begin{aligned}
\mathrm{P}(\mu \in T_n^u) &= \mathrm{P}\left(\bar{X}_n - \sqrt{\frac{S_n^2}{n}}\, z \le \mu \le \bar{X}_n + \sqrt{\frac{S_n^2}{n}}\, z \right) \\
&= \mathrm{P}\left(\left\{ \bar{X}_n - \sqrt{\frac{S_n^2}{n}}\, z \le \mu \right\} \cap \left\{ \mu \le \bar{X}_n + \sqrt{\frac{S_n^2}{n}}\, z \right\} \right) \\
&= \mathrm{P}\left(\left\{ \bar{X}_n - \mu \le \sqrt{\frac{S_n^2}{n}}\, z \right\} \cap \left\{ \bar{X}_n - \mu \ge -\sqrt{\frac{S_n^2}{n}}\, z \right\} \right) \\
&= \mathrm{P}\left(\left\{ \frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \le z \right\} \cap \left\{ \frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \ge -z \right\} \right) \\
&= \mathrm{P}\left(-z \le \frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \le z \right) \\
&= \mathrm{P}\left(-\sqrt{\frac{n-1}{n}}\, z \le \sqrt{\frac{n-1}{n}}\, \frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \le \sqrt{\frac{n-1}{n}}\, z \right) \\
&= \mathrm{P}\left(-\sqrt{\frac{n-1}{n}}\, z \le Z_{n-1} \le \sqrt{\frac{n-1}{n}}\, z \right)
\end{aligned}
$$

BIOL933 Design, Analysis, and Interpretation of Experiments



https://statswithcats.wordpress.com/2010/08/27/the-right-tool-for-the-job/

http://www.unh.edu/halelab/BIOL933

Design, Analysis, and Interpretation of Experiments

BIOL 933 (4 credits)

Fall 2017
(CRN 16386)

Lectures: Spaulding 230 (Tuesdays & Thursdays 9:40 – 11:00 am)
Lab: Kingsbury N134 (Thursdays 4:10 - 5:30 pm)

Group office hours
Location TBD (Mondays & Wednesdays, hours TBD)

Instructor:

Iago Hale
Assistant Professor of Specialty Crop Improvement
Department of Agriculture, Nutrition, and Food Systems; UNH
iago.hale@unh.edu

Click here for the 2017 welcome letter.
Click here for some great readings/links.

Syllabus | **Detailed schedule and course materials** | Calendar

Last updated: August 2017

http://www.unh.edu/halelab/BIOL933/schedule.htm

BIOL 933: Design, Analysis, and Interpretation of Experiments

Detailed schedule and course materials

Home | Syllabus | Calendar

***** Hit your browser's REFRESH button *****
***** every time you visit this page! *****

(Get Adobe Acrobat Reader)

Date (Fall 2017)		Lectures	Labs	Assignments
Aug 29	Tu	**Introduction to the course and review of fundamental statistical concepts** Reading	.	**Homework tips**
Aug 31	Th	**Topic 1: Introduction to the principles of experimental design** Lecture Reading	Lab 1 Scripts 1 2 Datasets 2 3 All files [ZIP]	HW 1 (Topics 1&2) [Word] [PDF] KEY
Sept 5	Tu	**Topic 2: Distributions, hypothesis testing, and sample size determination** Lecture Reading Supp Materials: Z table, t table, χ^2 table Importing Data into RStudio	Lab 2 Scripts 1 2 3 Datasets 3 All files [ZIP]	.
		NO CLASS – **Video lecture HERE** **Topic 3: General linear models and the fundamentals of analysis of**	Lab 3 Scripts 1 2 3 Datasets 1 2 3	.

Student Evaluation

Homeworks (10)	45%	
In-class quiz	5%	(Sept 14)
First exam	25%	(Oct 9)
Second exam	25%	(Dec 1)

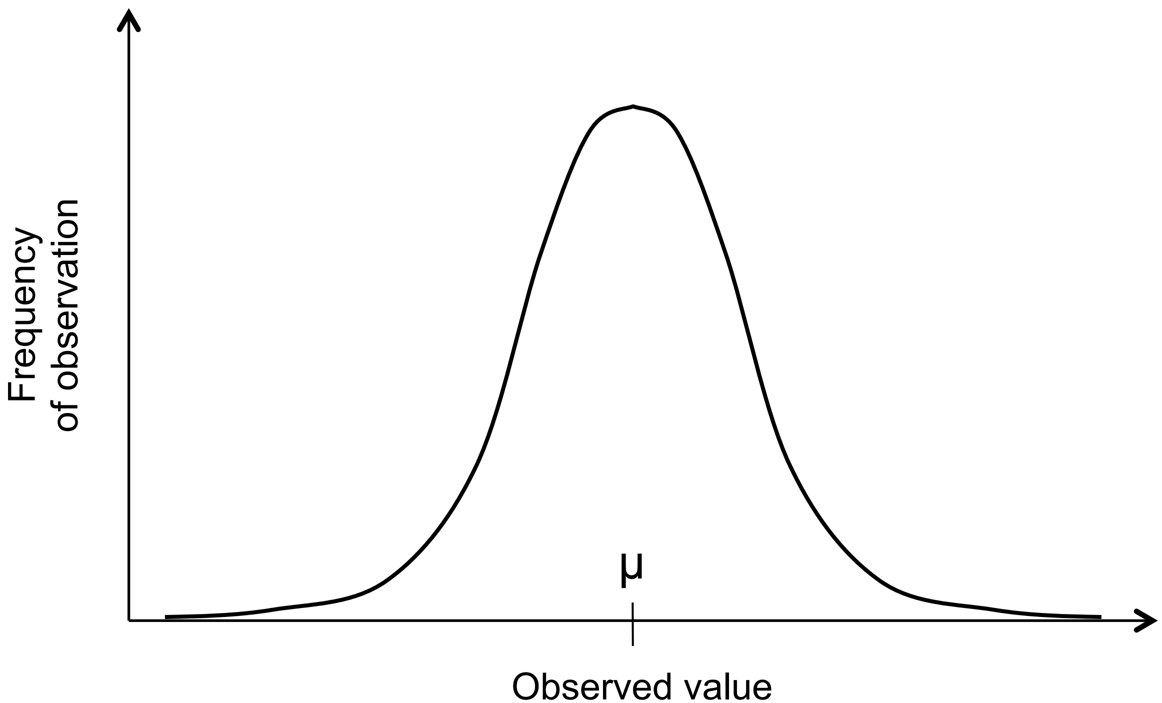
More details can be found in the online syllabus.

RStudio™

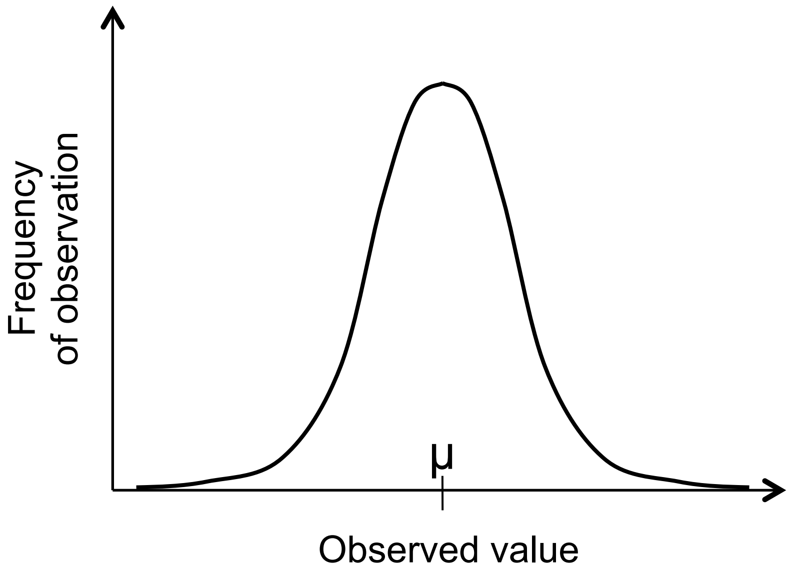
Free & Open-Source IDE for R

questions?

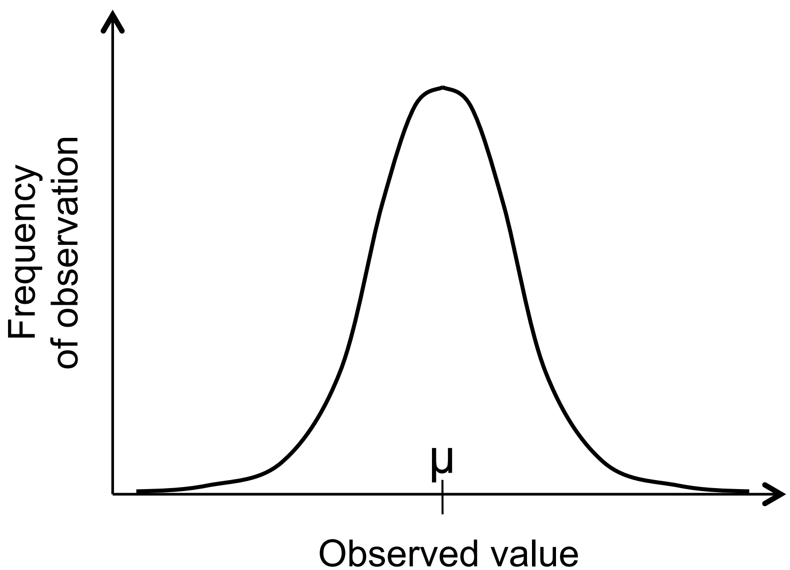
The Normal Distribution



The Normal Distribution

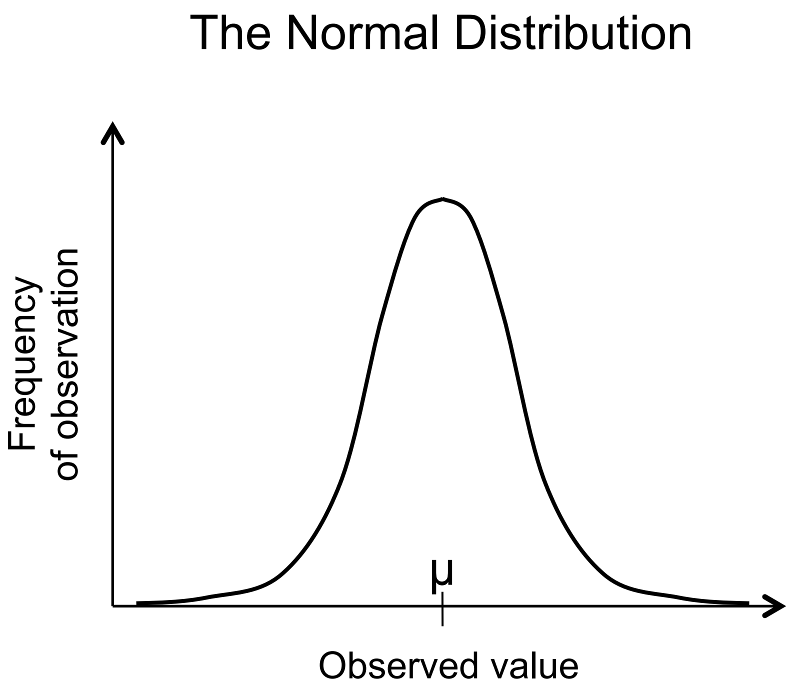


The Normal Distribution



1. A purely mathematical entity.

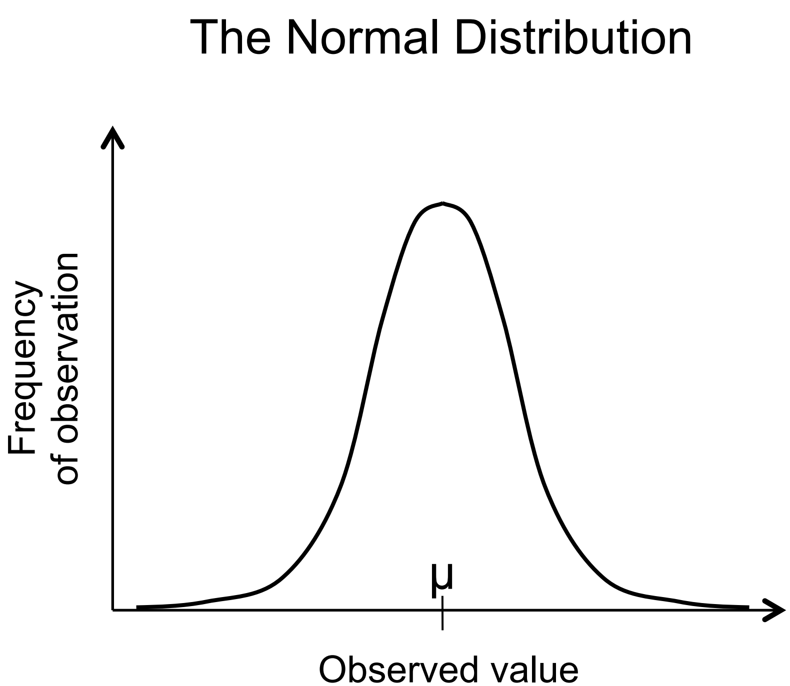
$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



1. A purely mathematical entity.

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

2. A distribution of related individuals.
(what we see)



1. A purely mathematical entity.

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

2. A distribution of related individuals.
(what we see)

3. A distribution of errors.
(what we think underlies what we see)

$$N = 150$$

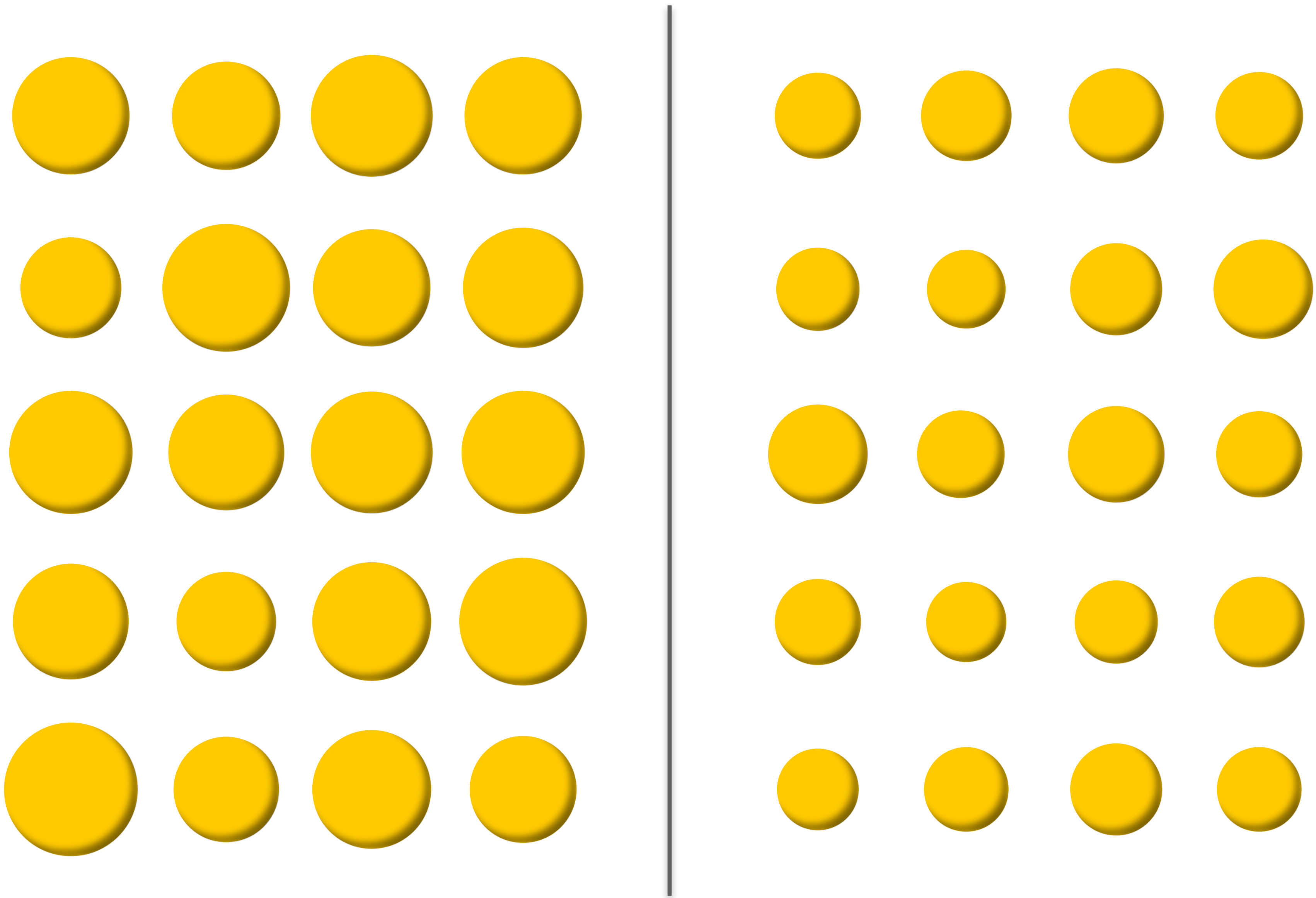
$$\mu = \frac{\sum_{i=1}^{N} Y_i}{N} = 200g$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (Y_i - \mu)^2}{N} = 100g^2$$

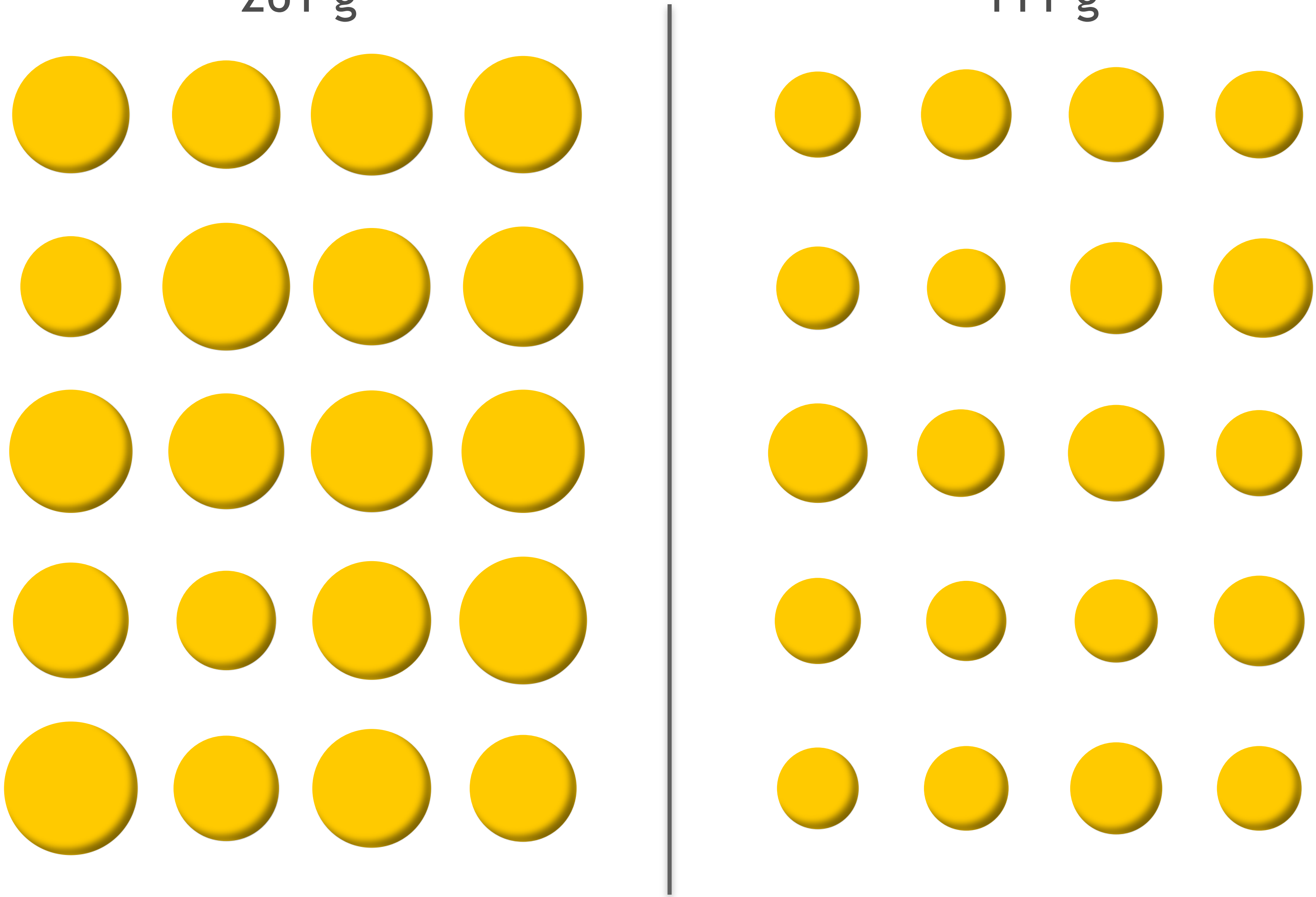
$$n < N$$

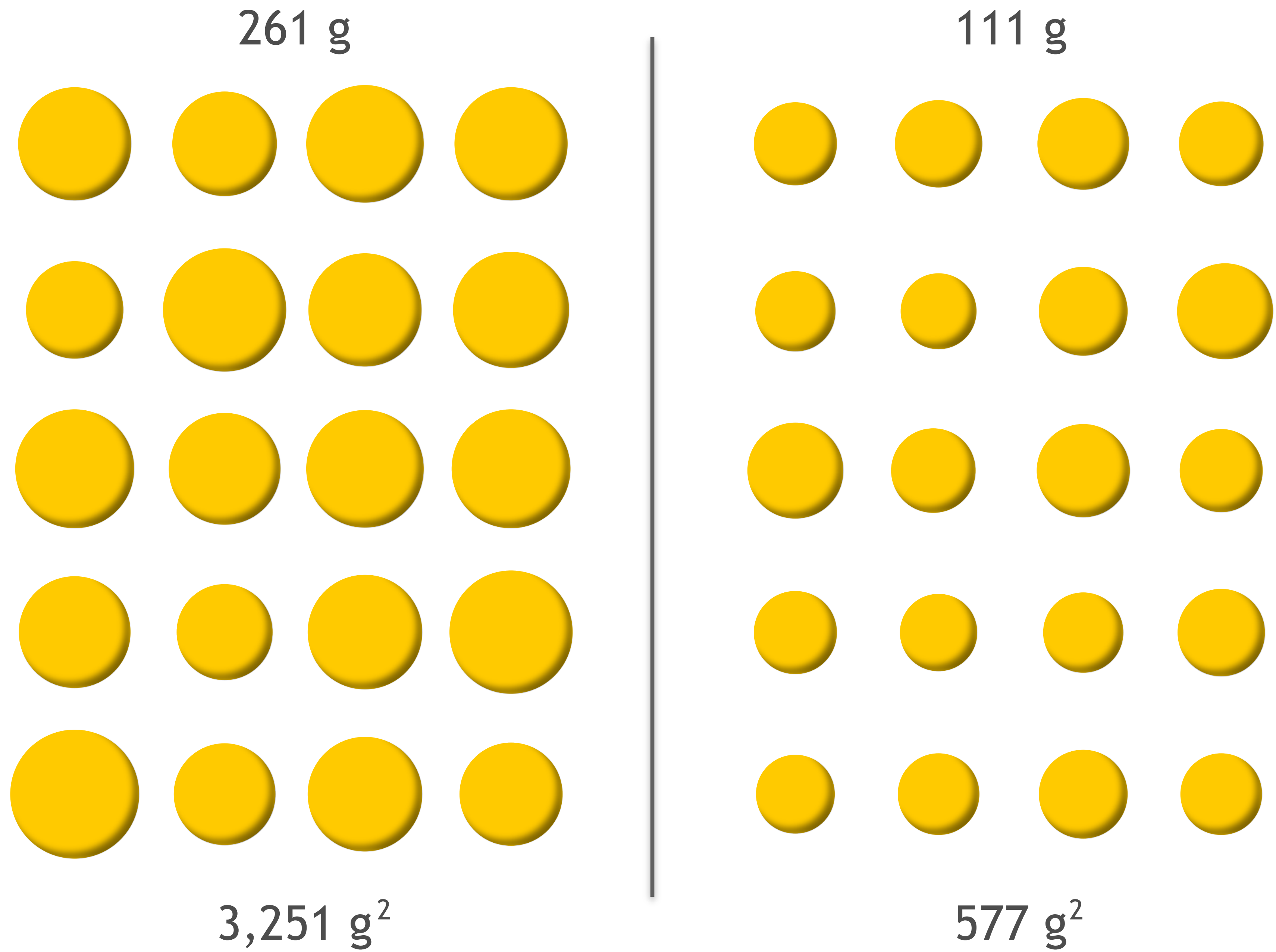
$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \cong 200g$$

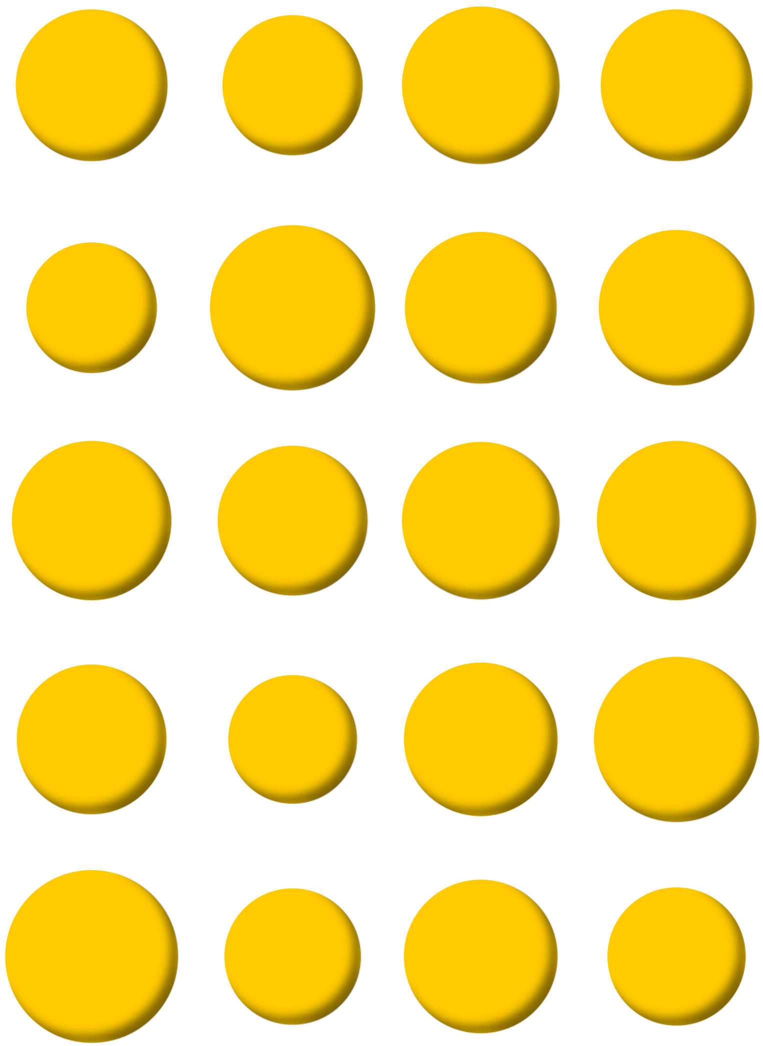
$$s^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1} \cong 100g^2$$



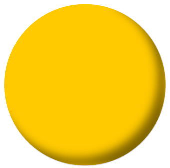
261 g | 111 g



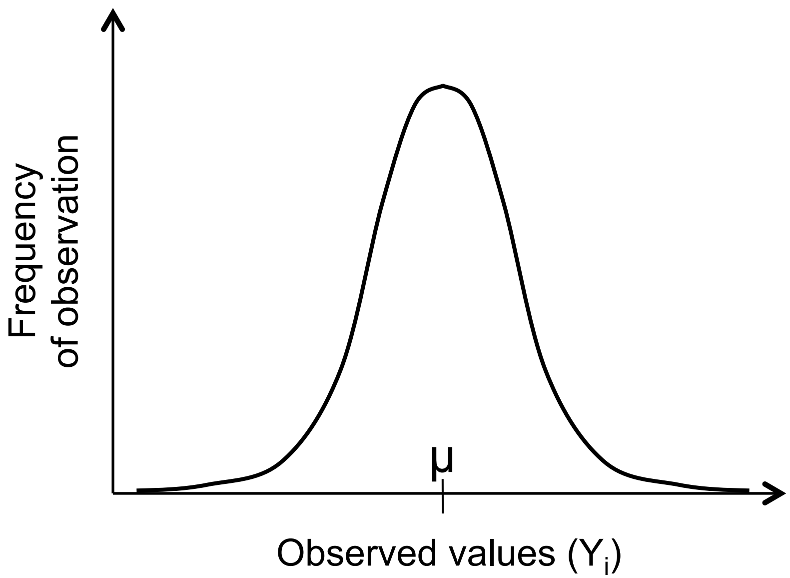

261 g
111 g
3,251 g²
577 g²



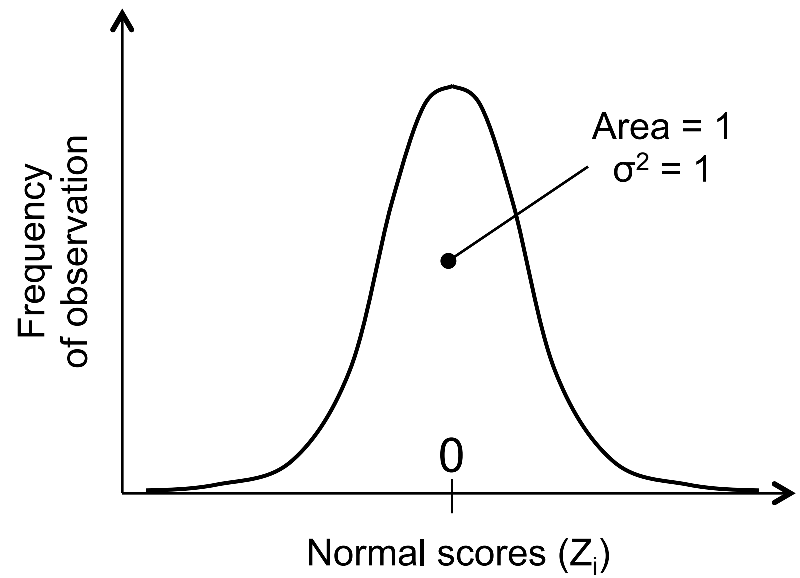
An unusual individual?

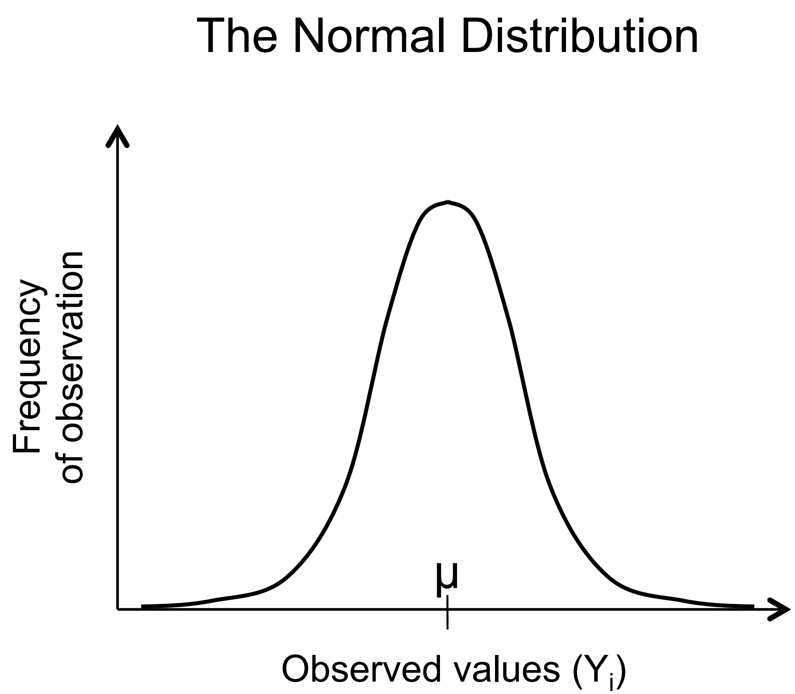


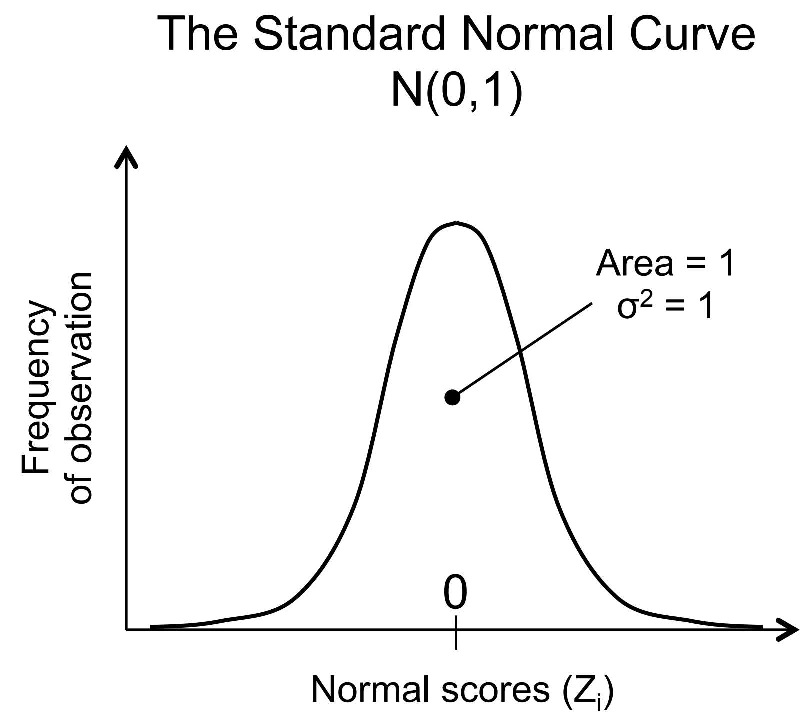
The Normal Distribution



The Standard Normal Curve N(0,1)







$$Z_i = \frac{Y_i - \mu}{\sigma}$$

In a normal frequency distribution,

$\mu \pm 1\sigma$ contains 68.27% of the items
$\mu \pm 2\sigma$ contains 95.45% of the items
$\mu \pm 3\sigma$ contains 99.73% of the items

Thought of in another way,

50% of the items fall between $\mu \pm 0.674\sigma$
95% of the items fall between $\mu \pm 1.960\sigma$
99% of the items fall between $\mu \pm 2.576\sigma$

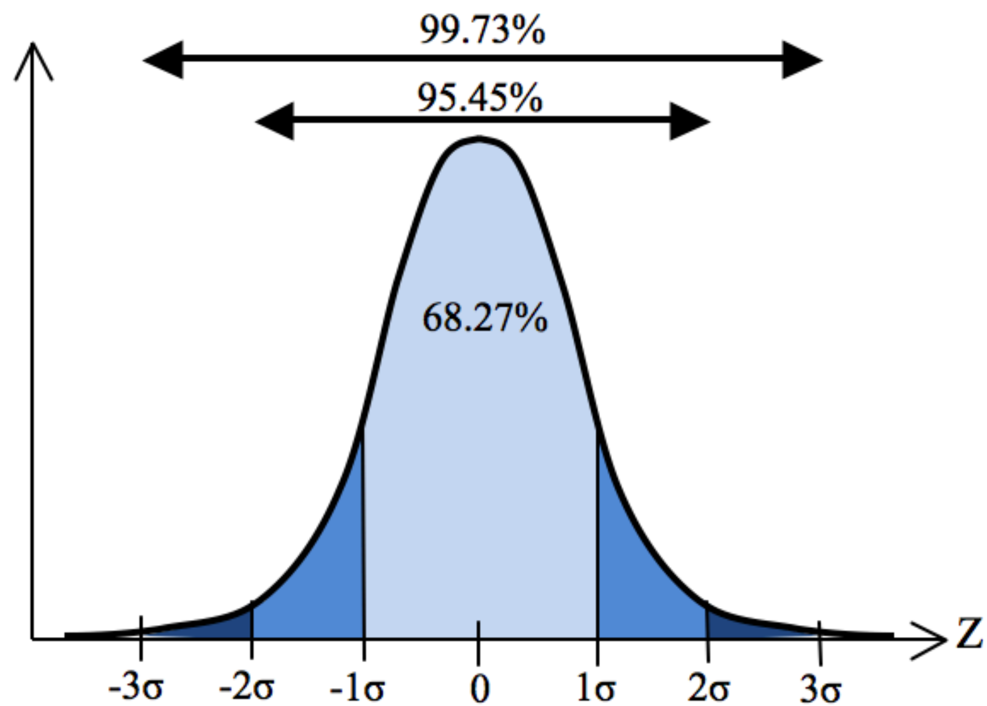
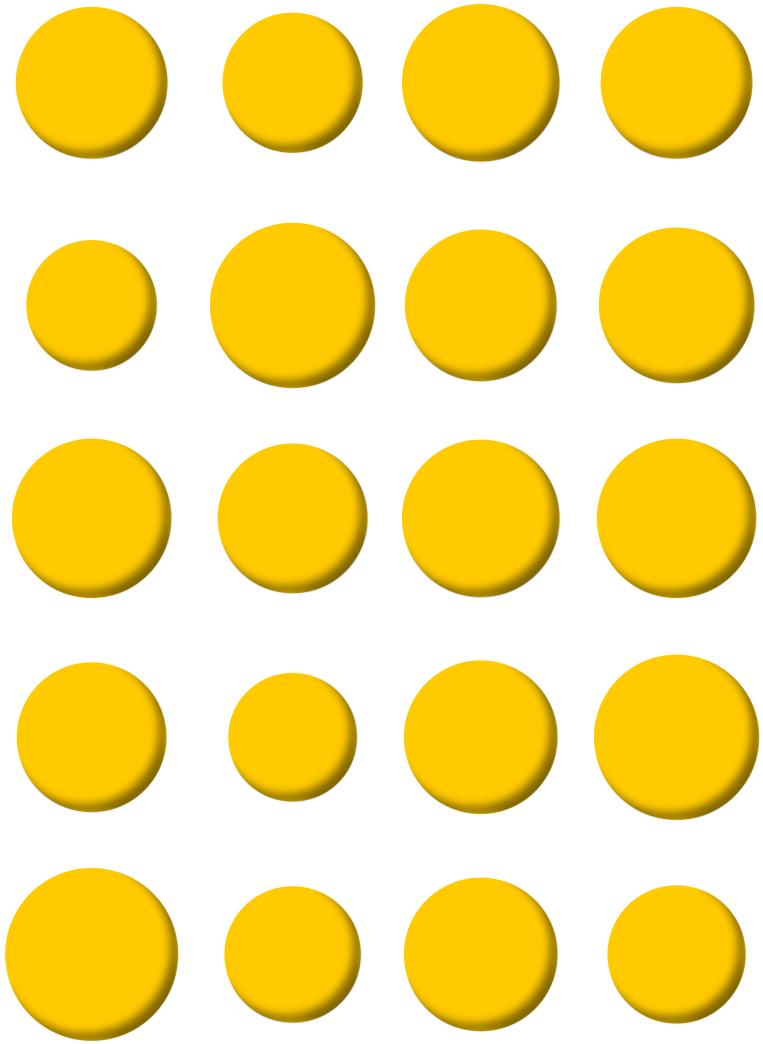
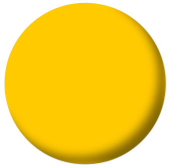


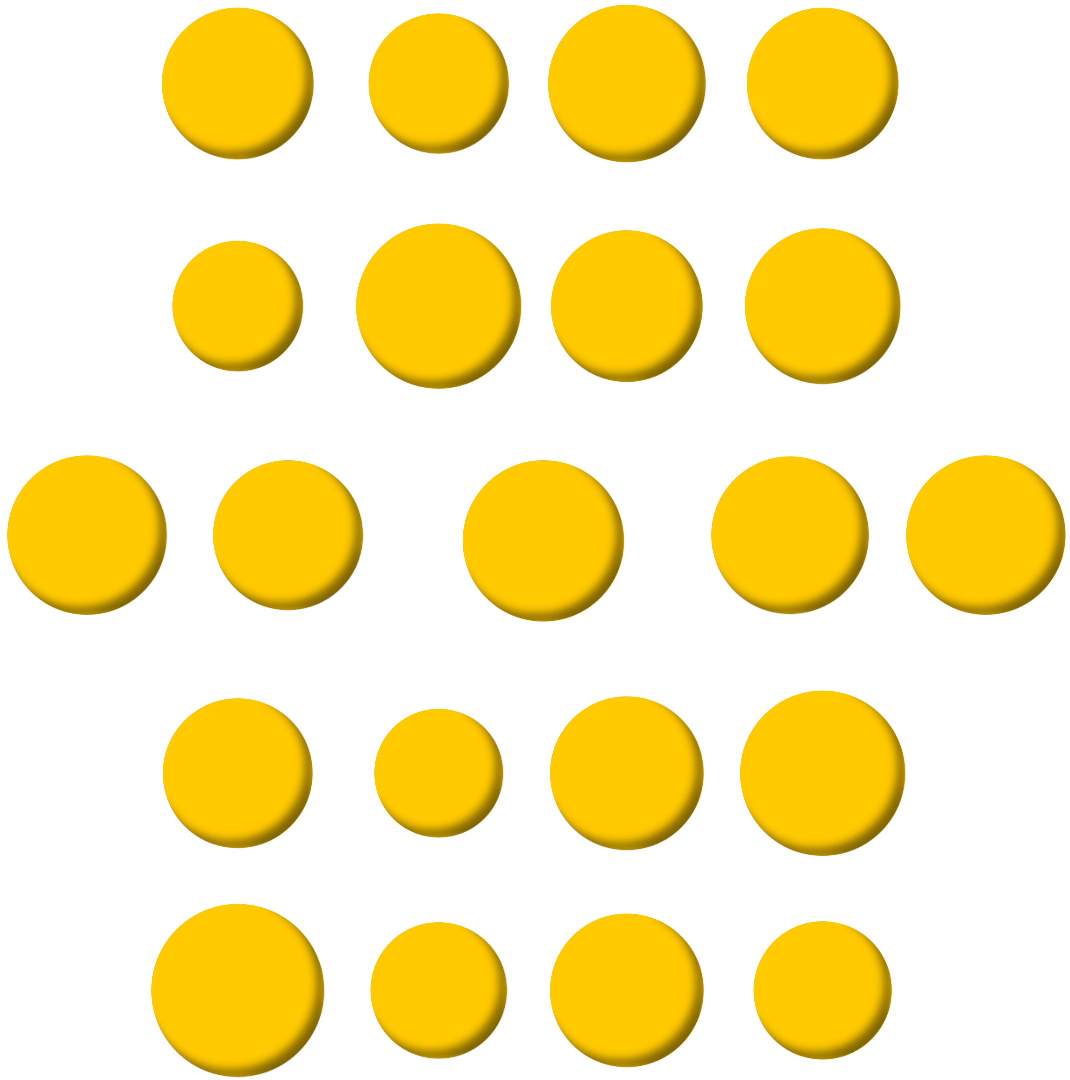
TABLE A.4
Probability of a random value of $Z = (Y - \mu)/\sigma$ being greater than the values tabulated in the margins

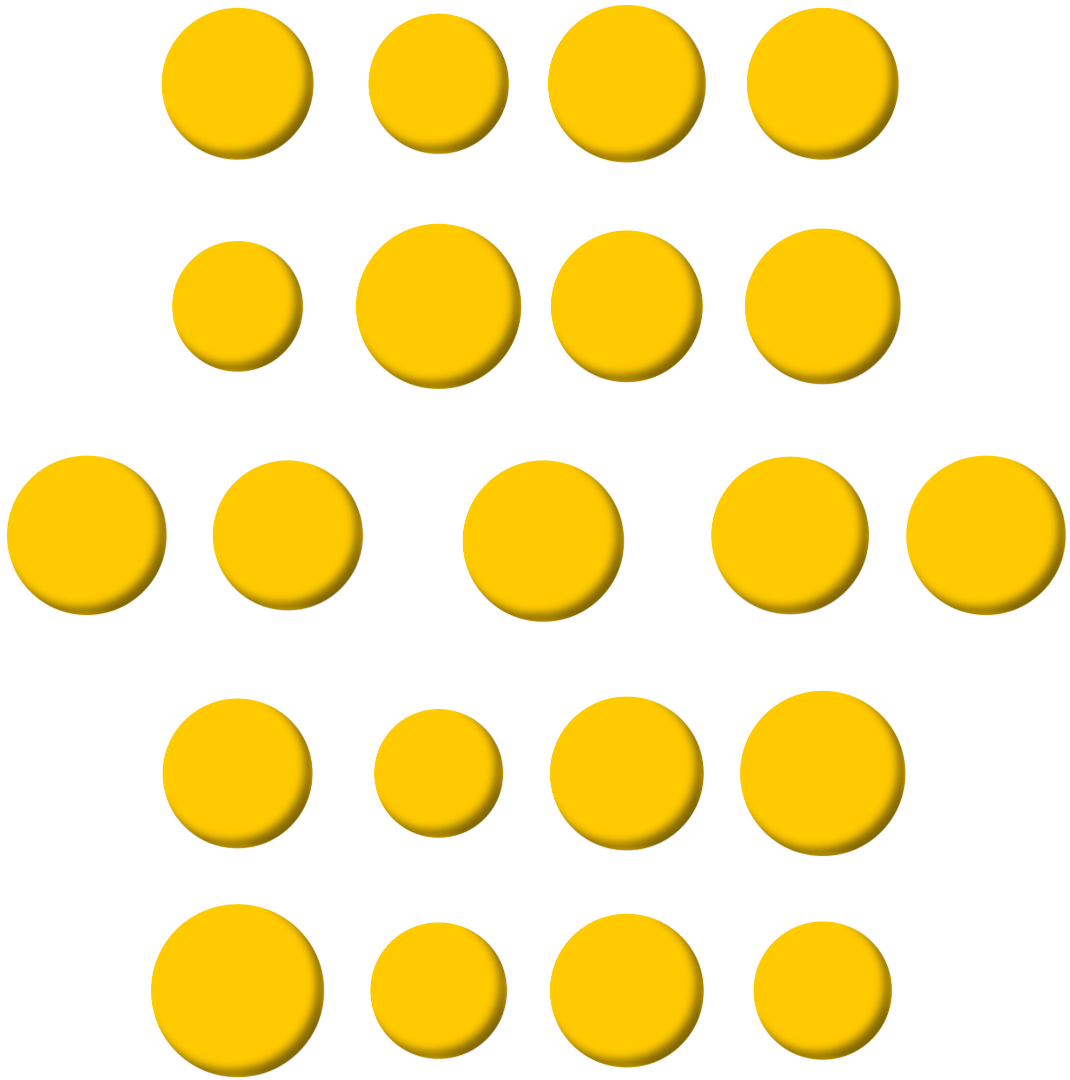
Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
3.9	.0000									



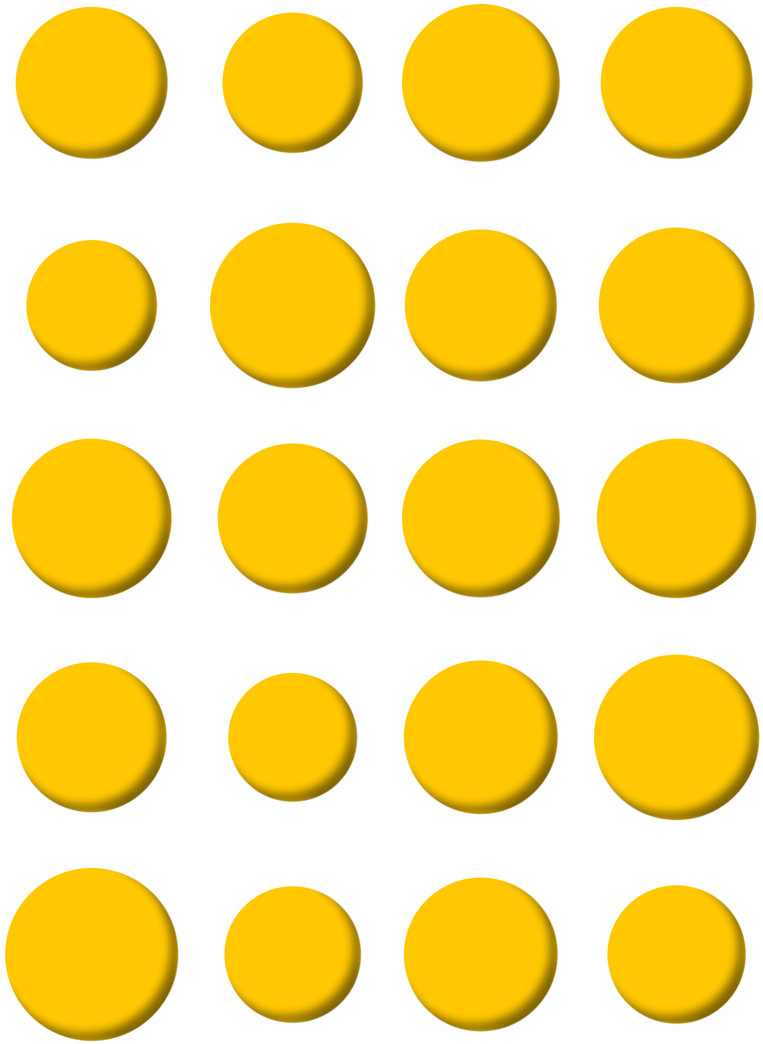
An unusual individual?



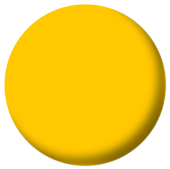


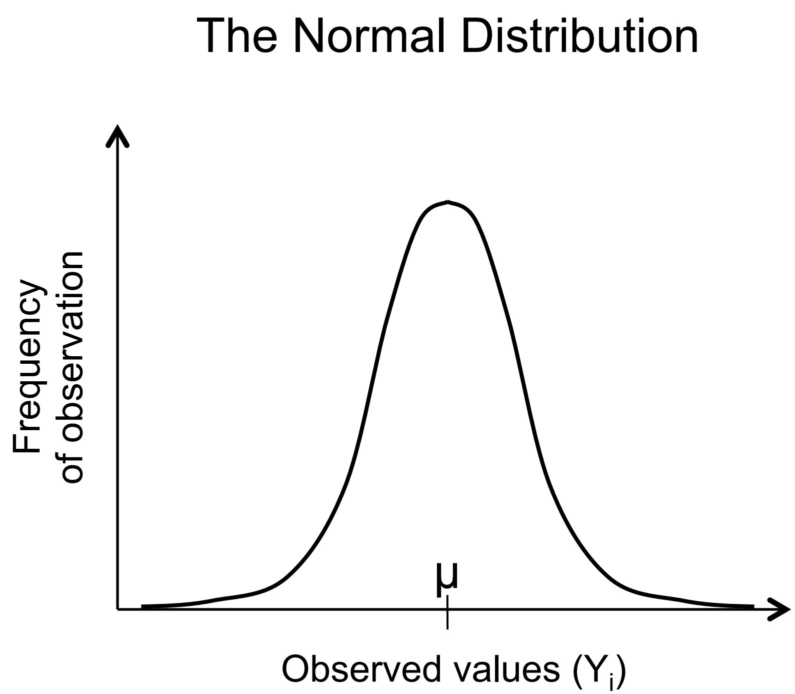


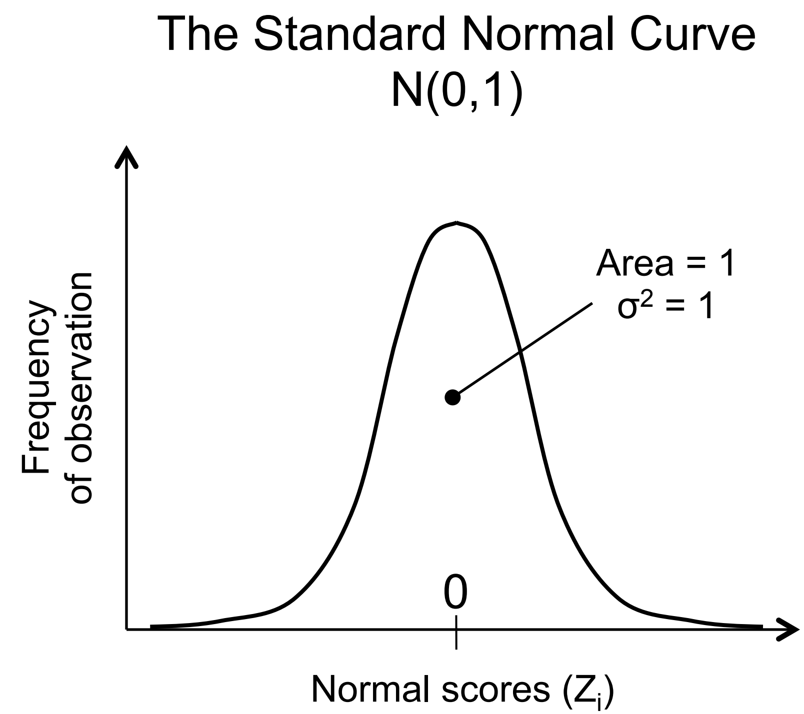
p = 0.20



An unusual *sample mean* (n=20)?

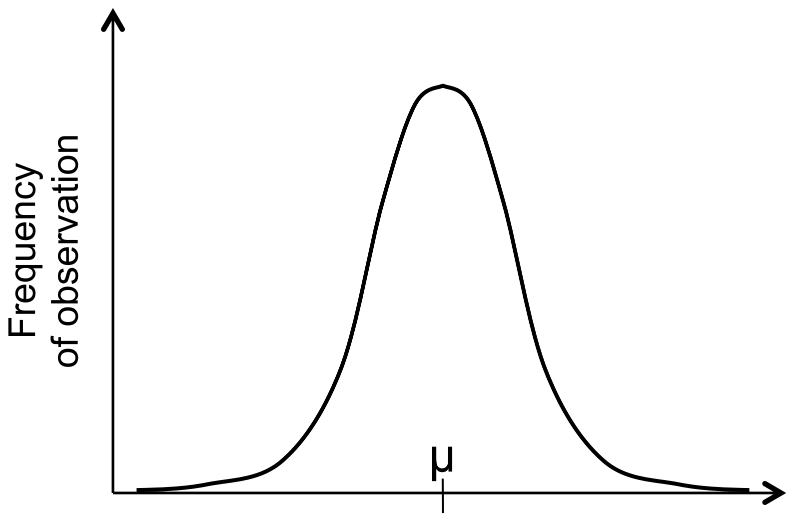




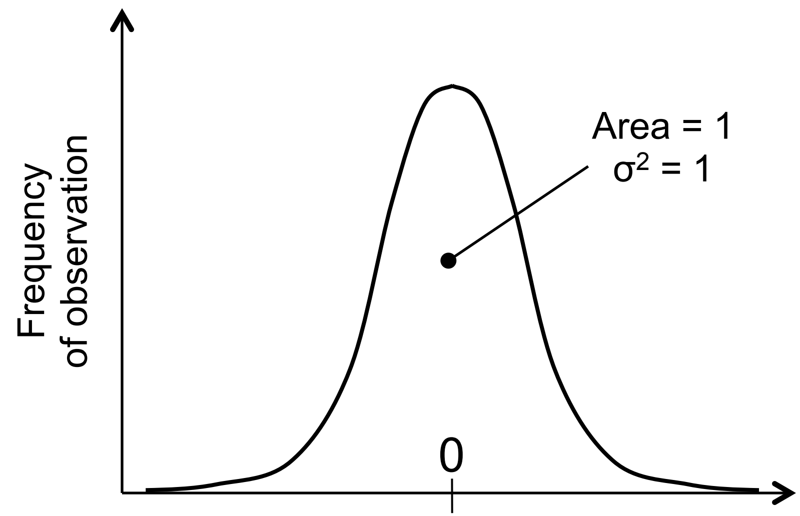


$$Z_i = \frac{Y_i - \mu}{\sigma}$$

The Normal Distribution



The Standard Normal Curve N(0,1)



$$\bar{Z}_{i(n)} = \frac{\bar{Y}_{i(n)} - \mu}{\sigma_{\bar{Y}(n)}}$$

$$\sigma_{\bar{Y}(n)} = \frac{\sigma}{\sqrt{n}}$$

The following six samples of 10 observations each (n = 10) were randomly drawn from a normally distributed population with $\mu = 40$ and $\sigma^2 = 100$:

Sample 1	38	20	22	50	46	25	45	40	39	43
Sample 2	37	24	27	25	38	58	50	32	57	38
Sample 3	42	46	45	40	50	43	45	29	38	51
Sample 4	39	45	31	42	21	60	48	26	51	38
Sample 5	49	36	33	50	42	42	46	33	44	25
Sample 6	39	40	50	30	47	72	60	34	47	34

The following six samples of 10 observations each (n = 10) were randomly drawn from a normally distributed population with $\mu = 40$ and $\sigma^2 = 100$:

											Mean	StDev
Sample 1	38	20	22	50	46	25	45	40	39	43	36.8	10.7
Sample 2	37	24	27	25	38	58	50	32	57	38	38.6	12.6
Sample 3	42	46	45	40	50	43	45	29	38	51	42.9	6.3
Sample 4	39	45	31	42	21	60	48	26	51	38	40.1	11.8
Sample 5	49	36	33	50	42	42	46	33	44	25	40.0	8.0
Sample 6	39	40	50	30	47	72	60	34	47	34	45.3	13.0
											40.6	10.4

The following six samples of 10 observations each (n = 10) were randomly drawn from a normally distributed population with $\mu = 40$ and $\sigma^2 = 100$:

											Mean	StDev
Sample 1	38	20	22	50	46	25	45	40	39	43	36.8	10.7
Sample 2	37	24	27	25	38	58	50	32	57	38	38.6	12.6
Sample 3	42	46	45	40	50	43	45	29	38	51	42.9	6.3
Sample 4	39	45	31	42	21	60	48	26	51	38	40.1	11.8
Sample 5	49	36	33	50	42	42	46	33	44	25	40.0	8.0
Sample 6	39	40	50	30	47	72	60	34	47	34	45.3	13.0
											40.6	10.4

The following six samples of 10 observations each ($n = 10$) were randomly drawn from a normally distributed population with $\mu = 40$ and $\sigma^2 = 100$:

											Mean	StDev
Sample 1	38	20	22	50	46	25	45	40	39	43	36.8	10.7
Sample 2	37	24	27	25	38	58	50	32	57	38	38.6	12.6
Sample 3	42	46	45	40	50	43	45	29	38	51	42.9	6.3
Sample 4	39	45	31	42	21	60	48	26	51	38	40.1	11.8
Sample 5	49	36	33	50	42	42	46	33	44	25	40.0	8.0
Sample 6	39	40	50	30	47	72	60	34	47	34	45.3	13.0
											40.6	10.4

Sample Means	36.8	38.6	42.9	40.1	40.0	45.3

The following six samples of 10 observations each ($n = 10$) were randomly drawn from a normally distributed population with $\mu = 40$ and $\sigma^2 = 100$:

											Mean	StDev
Sample 1	38	20	22	50	46	25	45	40	39	43	36.8	10.7
Sample 2	37	24	27	25	38	58	50	32	57	38	38.6	12.6
Sample 3	42	46	45	40	50	43	45	29	38	51	42.9	6.3
Sample 4	39	45	31	42	21	60	48	26	51	38	40.1	11.8
Sample 5	49	36	33	50	42	42	46	33	44	25	40.0	8.0
Sample 6	39	40	50	30	47	72	60	34	47	34	45.3	13.0
											40.6	10.4

							Mean	StDev
Sample Means	36.8	38.6	42.9	40.1	40.0	45.3	40.6	3.0

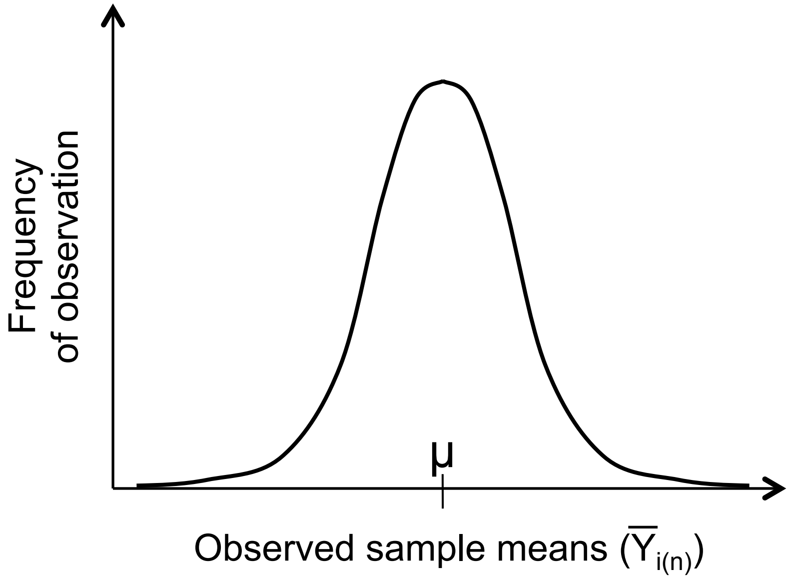
The following six samples of 10 observations each (n = 10) were randomly drawn from a normally distributed population with μ = 40 and σ^2 = 100:

											Mean	StDev
Sample 1	38	20	22	50	46	25	45	40	39	43	36.8	10.7
Sample 2	37	24	27	25	38	58	50	32	57	38	38.6	12.6
Sample 3	42	46	45	40	50	43	45	29	38	51	42.9	6.3
Sample 4	39	45	31	42	21	60	48	26	51	38	40.1	11.8
Sample 5	49	36	33	50	42	42	46	33	44	25	40.0	8.0
Sample 6	39	40	50	30	47	72	60	34	47	34	45.3	13.0
											40.6	10.4

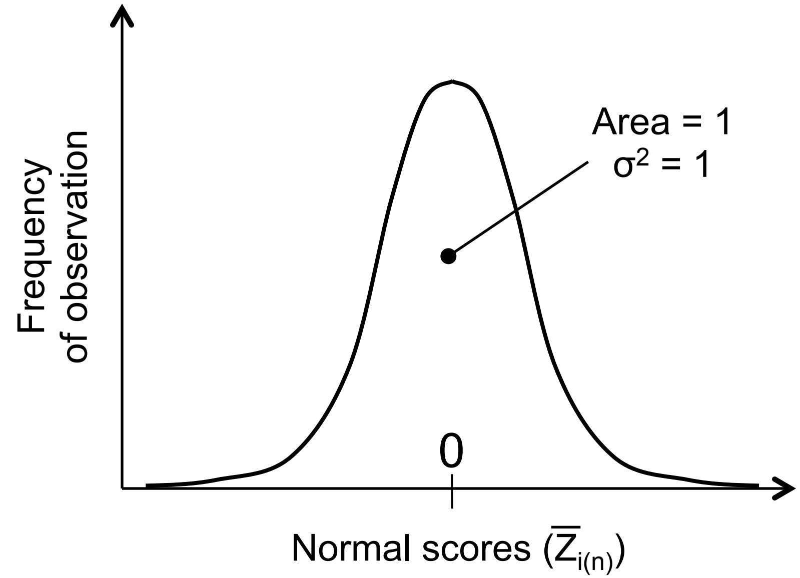
							Mean	StDev
Sample Means	36.8	38.6	42.9	40.1	40.0	45.3	40.6	3.0

$$\sigma_{\bar{Y}(n)} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{100}}{\sqrt{10}} = \sqrt{10} = 3.16$$

The Normal Distribution

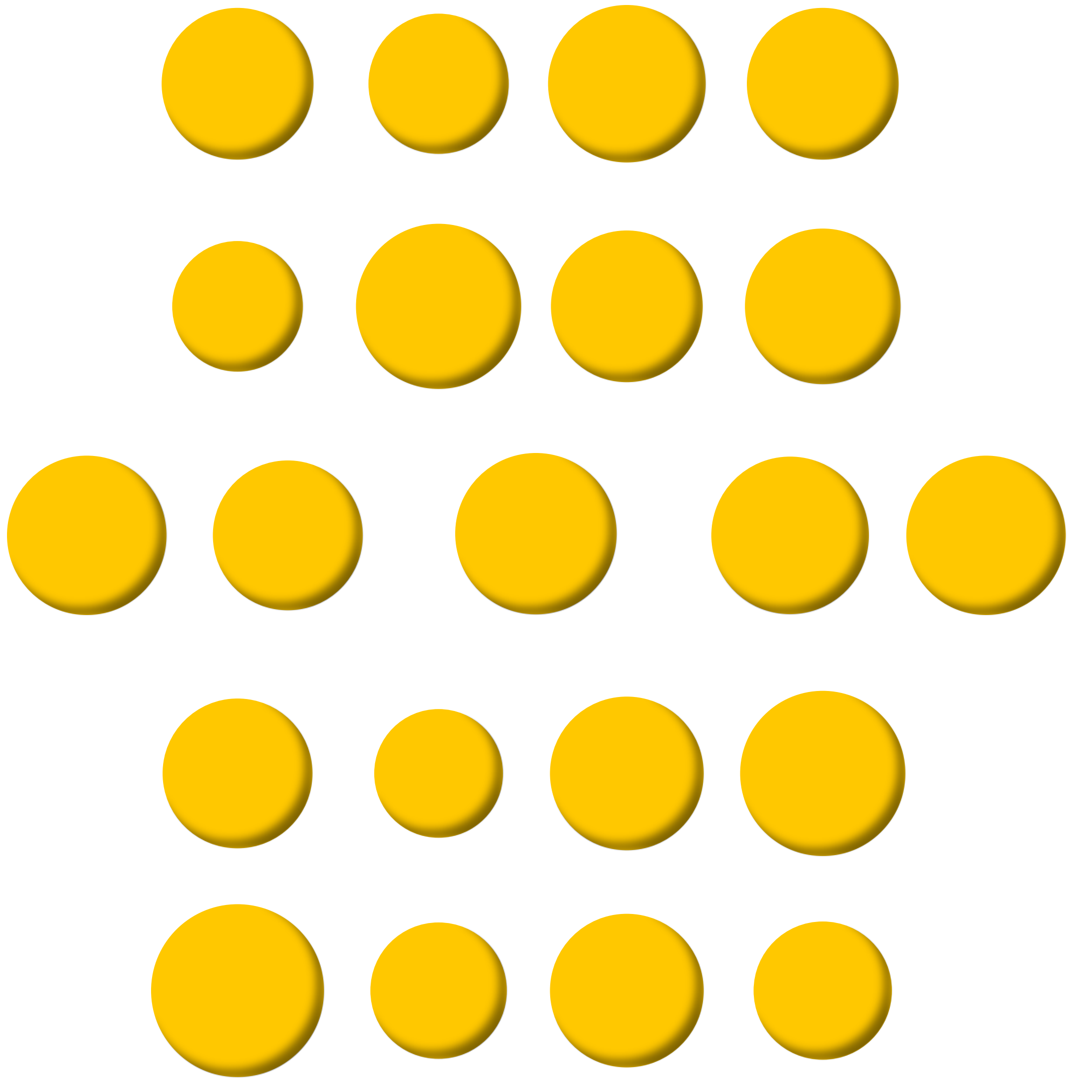


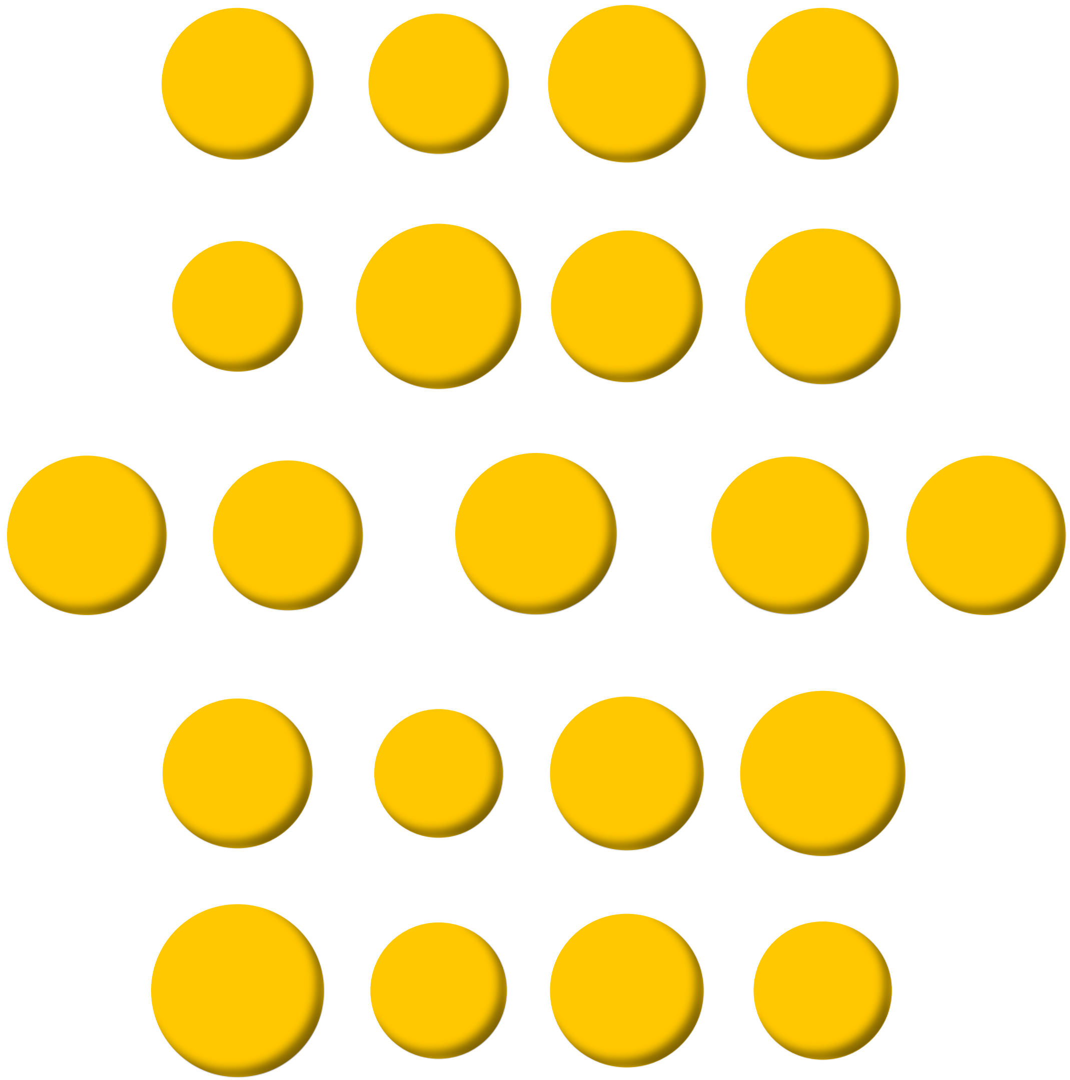
The Standard Normal Curve N(0,1)



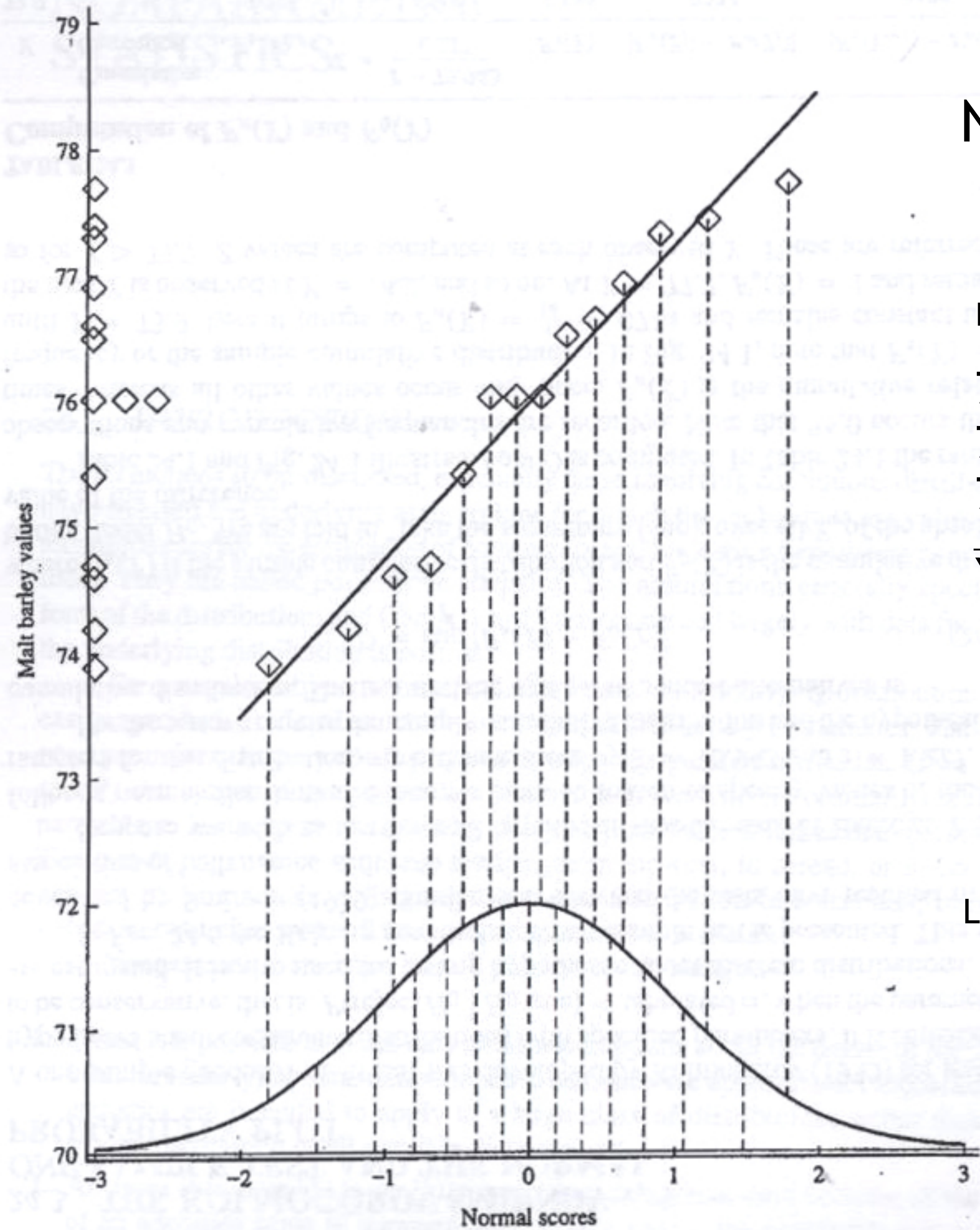
$$\bar{Z}_{i(n)} = \frac{\bar{Y}_{i(n)} - \mu}{\sigma_{\bar{Y}(n)}}$$

$$\sigma_{\bar{Y}(n)} = \frac{\sigma}{\sqrt{n}}$$





p = 0.000065



Normal Probability Plot

Figure 24.2 (page 566)

n=14

77.7, 76.0, 76.9, 74.6, 74.7, 76.5, 74.2, 75.4, 76.0, 76.0, 73.9, 77.4, 76.6, 77.3

$\overline{Y}$ = 75.943, s = 1.227

1/14 = 0.0714 square units

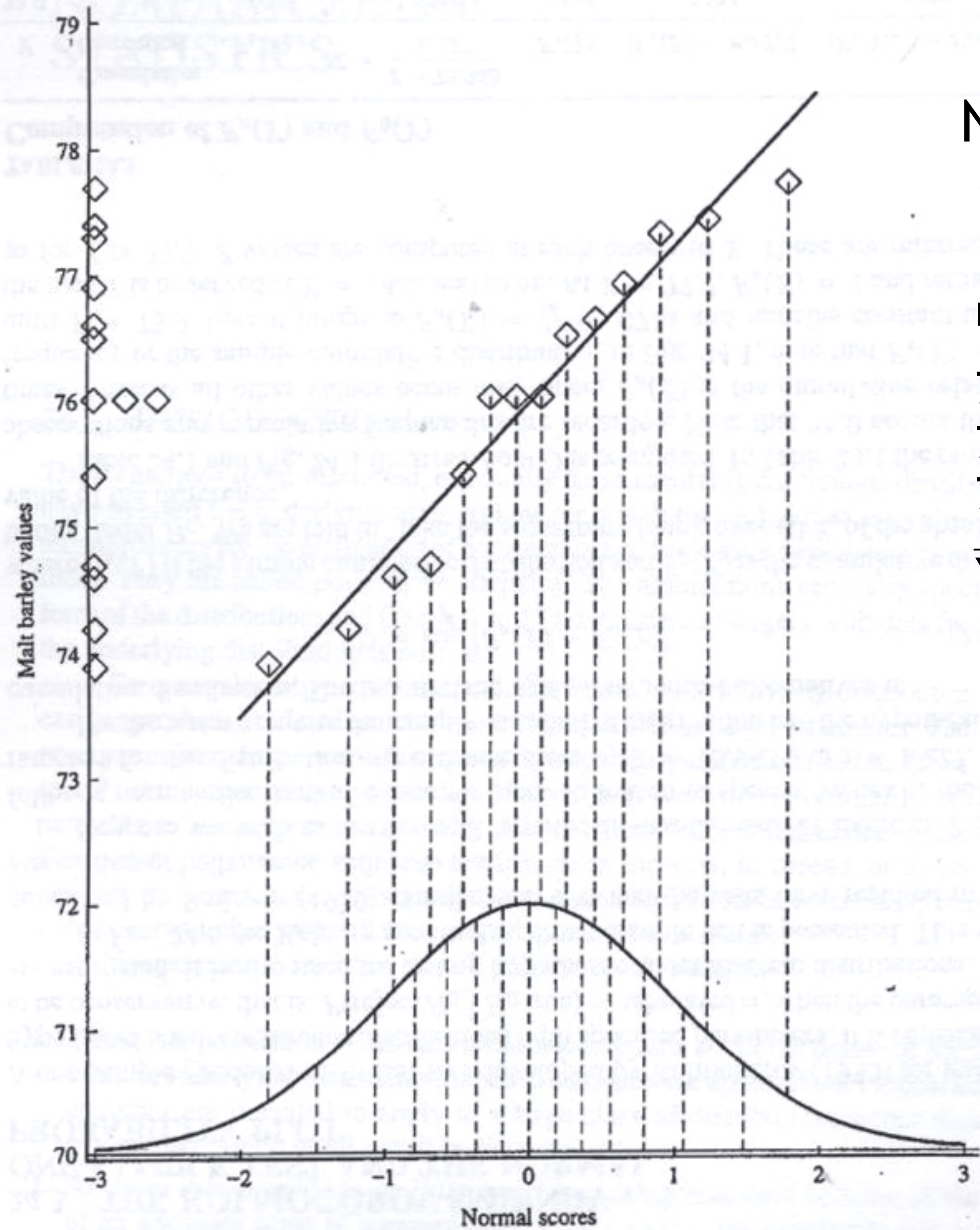
1/28 = 0.0357 square units

Lower bound of 14th quantile ≈ 1.465

“Expected” Z_{14} ≈ 1.80

“Expected” $Y_{14} = sZ_{14} + \overline{Y}$ = 78.1516

Observed Y_{14} = 77.7



Normal Probability Plot

Figure 24.2 (page 566)

n=14

77.7, 76.0, 76.9, 74.6, 74.7, 76.5, 74.2, 75.4, 76.0, 76.0, 73.9, 77.4, 76.6, 77.3

$\overline{Y} = 75.943$, $s = 1.227$

W = 0.9484

(Pr < W) = 0.5092