XXXV. The Central Limit Theorem

Setting

- We have a population of stuff, e.g. students of different heights.
- There is some mean μ and variance σ^2, but we don't know them.
- We will take samples: $Y_1 :=$ height of one student. $Y_2 :=$ height of another, and so on. $Y_n :=$ height of last sample student.

Assumptions and Notation

- **Assumption forever**: Our samples are independent.
- We say the Y_i are independent identically distributed (i.i.d.) random variables.
- **We no longer assume that the population is normally distributed.**
- The population could have any distribution at all.

The Central Limit Theorem

- Suppose $Y_1, \dots, Y_n$ are independent samples from any population with mean μ, variance σ^2.

- Then $\boxed{\overline{Y} \to N\left(\mu, \frac{\sigma^2}{n}\right)}$.
- That is, $\overline{Y}$ approaches a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.
- In practice, when $n \geq 30$, we may assume $\overline{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Standard Normal Distribution

	Second decimal place of z									
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	.0722	.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	.0352	.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

Converting to Standard Normal

- **Recall**: If $\overline{Y}$ is normal, then

$$Z := \frac{\overline{Y} - \text{mean}}{\text{standard deviation}}$$

is standard normal, that is, $Z \sim N(0,1)$.

- **Corollary to Theorem**:

$$Z := \frac{\sqrt{n}\left(\overline{Y} - \mu\right)}{\sigma} \sim N(0,1)$$

- So we can use charts to find probabilities for Z and then work back to find them for Y.

Example I

Homework problems take you an average of 12 minutes, with standard deviation 10 minutes. Your assignment is 36 problems. What is the probability that it will take you more than 9 hours?

(Graph. Most are quick, but if you get stumped, they take forever.) Note that this is definitely not a normal distribution (there are no negative values). It's probably exponentialish, but it can't be exactly exponential because $\mu \neq \sigma$.

Taking $>$ 12 hours (720 minutes) means you would average > 20 minutes per problem. Ferrari says that $\overline{Y}$ is normal with mean $\mu = 15$, variance $\frac{\sigma^2}{n} = \frac{100}{36} = \frac{25}{9}$. So $Z := \frac{\overline{Y} - 15}{\sqrt{\frac{25}{9}}} = \frac{3}{5}(\overline{Y} - 15)$ is standard normal. $\overline{Y} > 20$ iff $Z > 3$, which by the chart has probability $\boxed{0.135\%}$ (one tenth of one percent, not 13%). [See how nice I am to worry about such things and make sure that it's very unlikely you'll spend too much time on them?]

9 hours is 540 minutes, so you would have to average $\frac{540}{36} = 15$ minutes per problem.

$$\begin{aligned}
\overline{Y} &> 15 \\
\overline{Y} - \mu &> 15 - 12 = 3 \\
\frac{\sqrt{n}\left(\overline{Y} - \mu\right)}{\sigma} &> \frac{6 \cdot 3}{10} = 1.8 \\
Z &> 1.8 \Longrightarrow P \approx \boxed{0.0359 \approx 3.6\%}
\end{aligned}$$

	Second decimal place of z									
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	.0722	.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	.0352	.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

Example II

A bakery sells an average of 30 muffins per day, with a standard deviation of 8 muffins. What is the chance that they will sell more than 1,000 muffins in the next 36 days?

She would need to average $\overline{Y} > \frac{1000}{36} = 27\frac{7}{9}$ muffins per batch. CLT tells us that $\overline{Y} \sim N\left(30, \frac{64}{36}\right) = N\left(30, \frac{16}{9}\right)$. Recall that $Z := \frac{\overline{Y} - \mu}{\sigma} \sim N(0, 1)$.

$$\begin{aligned} \overline{Y} &> 27\frac{7}{9} \\ \text{iff} \quad \overline{Y} - \mu &> 27\frac{7}{9} - 30 = -2\frac{2}{9} = -\frac{20}{9} \\ \text{iff} \quad \frac{\overline{Y} - \mu}{\sigma} &> \frac{-\frac{20}{9}}{\frac{4}{3}} = -\frac{5}{3} \approx -1.67 \\ \text{iff} \quad Z &> -1.67 \end{aligned}$$

From the standard normal table, $P(Z > 1.67) \approx 0.0475$. [Graph.] So $P(Z < -1.67) = 0.0475$, so $P(Z > -1.67) = 1 - 0.0475 = \boxed{95\%}$.

	Second decimal place of z									
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	.0722	.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	.0352	.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

Example III

A technician checks whether a soda dispenser gives the correct amount of soda by taking 100 samples. Assuming the standard deviation is 2.5ml, what the chance that her sample mean is within 0.5ml of the true average amount of soda dispensed?

We want $P\left(\left|\overline{Y} - \mu\right| \leq 0.5\right)$. We know

$$Z := \frac{\sqrt{n}\left(\overline{Y} - \mu\right)}{\sigma} \sim N(0,1).$$

$\sqrt{n} = 10, \sigma = 2.5$

$$\begin{aligned} \left|\overline{Y} - \mu\right| &\leq 0.5 \\ \frac{10\left|\overline{Y} - \mu\right|}{2.5} &\leq \frac{10 \cdot 0.5}{2.5} = 2 \\ P(Z > 2.0) &= 0.0228 \\ P(-2.0 < Z < 2.0) &= 1 - 2(0.0228) = 1 - 0.0456 = \boxed{0.9544 \approx 95\%} \end{aligned}$$

	Second decimal place of z									
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	.0722	.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	.0352	.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

Example IV

The technician from Example III wants to guarantee (with probability 95%) that her sample mean will be within 0.4ml of the true average. How many samples should she take?

We want $P\left(\left|\overline{Y}-\mu\right| \leq 0.5\right) = 0.95$. We know

$$Z := \frac{\sqrt{n}\left(\overline{Y}-\mu\right)}{\sigma} \sim N(0,1).$$

Draw the normal curve. We want 95% of the area in the middle, so we want the outer area to be $\frac{1-0.95}{2} = 0.025$. This corresponds by the chart to $Z = 1.96$.

$$\begin{aligned}
\left|\overline{Y}-\mu\right| &\leq 0.4 \\
\left|\frac{\sqrt{n}\left(\overline{Y}-\mu\right)}{\sigma}\right| &\leq \frac{0.4\sqrt{n}}{\sigma} \\
|Z| &\leq \frac{0.4\sqrt{n}}{\sigma} \\
1.96 &\leq \frac{0.4\sqrt{n}}{\sigma} \\
\frac{1.96\sigma}{0.4} &\leq \sqrt{n} \\
\frac{(1.96)(2.5)}{0.4} &\leq \sqrt{n} \\
n &\geq \left[\frac{(1.96)(2.5)}{0.4}\right]^2 \approx 150.063 \\
n &= \boxed{151}
\end{aligned}$$

	Second decimal place of z									
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	.0722	.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	.0352	.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

Example V

A restaurant finds that its customers spend an average of \$30, with a standard deviation of \$10. If they have 25 reservations tonight, what is the chance that their total revenue will be between \$725 and \$800?

We know

$$Z := \frac{\sqrt{n}\left(\overline{Y} - \mu\right)}{\sigma} \sim N(0,1).$$

$\mu = 30$, $\sigma = 10$

$$\begin{array}{rcl} 725 \leq & Y_1 + \cdots + Y_{25} & \leq 800 \\ \frac{725}{25} \leq & \frac{Y_1 + \cdots + Y_{25}}{25} & \leq \frac{800}{25} \\ 29 \leq & \overline{Y} & \leq 32 \\ \frac{5(29-30)}{10} \leq & \frac{\sqrt{n}\left(\overline{Y} - \mu\right)}{\sigma} & \leq \frac{5(32-30)}{10} \\ -0.5 \leq & Z & \leq 1 \end{array}$$

$$\begin{array}{rcl} P(Z \leq -0.5) & = & 0.3085 \\ P(Z \geq 1) & = & 0.1587 \\ P(0.5 \leq Z \leq 1) & = & 1 - 0.3085 - 0.1587 = \boxed{0.5328 \approx 53\%} \end{array}$$

	Second decimal place of z									
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	.0722	.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	.0352	.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