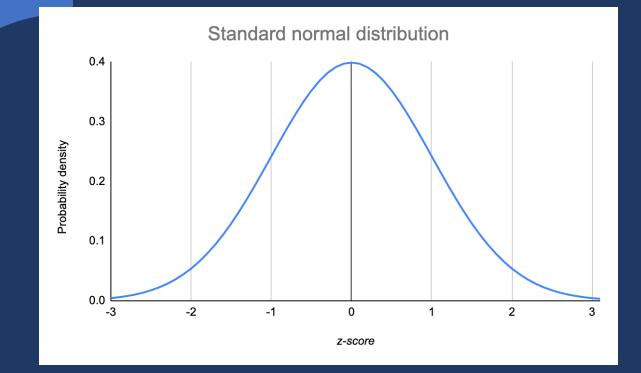
Unit 3: Lecture 2:

Standard Normal Distribution and calculating probabilities



Standard normal distribution

- The standard normal distribution, also called the *z*distribution, is a special normal distribution where the mean is 0 and the standard deviation is 1.
- Any normal distribution can be standardized by converting its values into *z* scores.
- Z scores tell you how many standard deviations from the mean each value lies.
- Converting a normal distribution into a z-distribution allows you to calculate the probability of certain values occurring and to compare different data sets.

Standardizing Data: What is Z-Score Normalization?

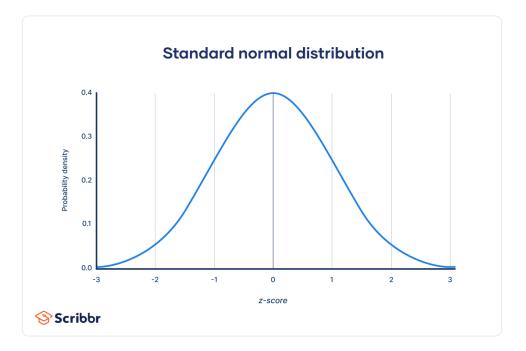
- Z-score normalization, also called standardization, transforms data so that it has a mean of 0 and a standard deviation of 1.
- This process adjusts data values based on how far they deviate from the mean, measured in units of standard deviation.

The Standard Normal Distribution

The *standard normal distribution*

has

$$\mu$$
=0 and σ =1.

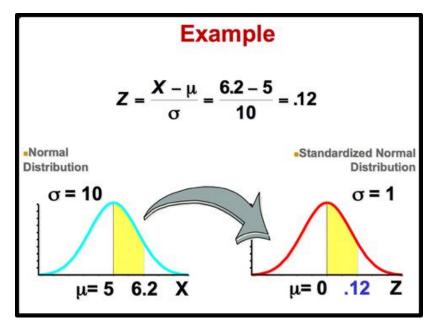


Why we use Z score

- Life is full of instances where we want to compare an observed value against the norm. Here are some examples:
- You get offered a job out of college at a salary of 42,000 USD a year. Is this salary high? Low? How does it compare to the incomes of other recent college graduates?
- Your niece was just born weighing 6.9 pounds. Is that normal?
- Your score on the SAT was 1479. How does that compare to other students?
- We can answer all of these questions in using Z-scores.
- A Z-score is a standardized number that tells you how far away a given data point is from the mean.

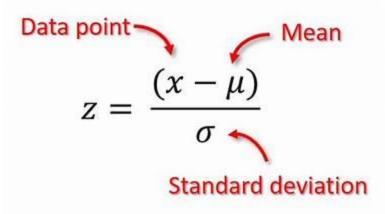
Normal Distribution

Z has a μ = 0 and a σ = 1. This is the Z distribution, also called the standard *normal* distribution. It is one of trillions of normal distributions we could have selected.



Standard Z Score

•The *standard z score* is obtained by creating a variable z whose value is



•Given the values of the sample mean and the sample standard deviation, we can convert a value of x to a value of z and find its probability using the table of normal curve areas.

Z-score Formula?

- How to find a Z-score is a simple process where you need to know three things:
- x. The value for which you want to calculate the Zscore. We sometimes call this the raw score.
- μ . The population mean.
- σ . The population standard deviation.
- If you know these three things, calculating a Z-score is easy.
- In the z-score formula, you simply subtract the population mean from your raw score and divide by the population standard deviation.

sample _ sample value _ mean Zscore standard deviation 6.5-7.96 $\frac{5}{-1.46}$ =-3.4762 0.42 *negative answer \rightarrow below the mean *positive answer \rightarrow above the mean

Interpretation of Z scores

- Z-scores are measured in standard deviation units.
- For example, a Z-score of 1.2 shows that your observed value is 1.2 standard deviations from the mean.
- A Z-score of 2.5 means your observed value is 2.5 standard deviations from the mean and so on.
- Typically, you will not see Z-scores that are more than 3 standard deviations from the mean. This is because most data points lie within 3 standard deviations of the mean.

Interpretation of Z scores

- Values of x that are larger than the mean have positive z -scores, and values of x that are smaller than the mean have negative z -scores.
- The closer your Z-score is to zero, the closer your value is to the mean. The further away your Z-score is from zero, the further away your value is from the mean.

Interpretation of Z scores

- If I tell you your income has a Z-score of -0.8, you immediately know that your income is below average. How far below average? 0.8 standard deviations.
- If I tell you that an SAT score has a Z-score of 2, you know the score is above average. How far above the average? 2 standard deviations. Note that a Zscore of zero shows that your value is equal to the mean

Example

- A college admissions officer reviewing an application.
- The applicant has a 1500 on their SAT and a 3.2 GPA.
- It is not immediately obvious how to compare these two figures, but if you calculate a Z-score relative to the average test scores and high school GPAs of students enrolled in your college, the comparison becomes much easier.
- Say the applicant's SAT Z-score is equal to 2.8 and their GPA Z-score is equal to -1.2. Immediately, you can infer that the applicant is well above average on their test scores but below average with their GPA.
- Similarly, you can compare Z-scores across metrics like height and weight, household income, resting heart rates for men versus women, and more.

What is the Usefulness of a Standard Normal Score?

- It tells you how many SDs (s) an observation is from the mean
- Thus, it is a way of quickly assessing how "unusual" an observation is
- Suppose the mean *BP* is 125 mmHg, and standard deviation = 14 mmHg
 - Is 167 mmHg an unusually high measure? If we know Z = 3.0, does that help us?

Standard Normal Scores

- Example: Male Blood Pressure, mean = 125, s = 14 mmHg
 - BP = 167 mmHg

$$Z = \frac{167 - 125}{14} = 3.0$$

• BP = 97 mmHg

$$Z = \frac{97 - 125}{14} = -2.0$$

Standard Normal Scores

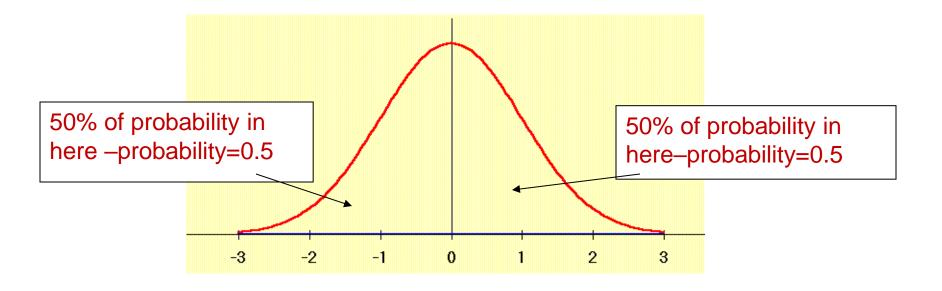
A standard score of:

- Z = 1: The observation lies one SD above the mean
- Z = 2: The observation is two SD above the mean
- Z = -1: The observation lies 1 SD below the mean
- Z = -2: The observation lies 2 SD below the mean

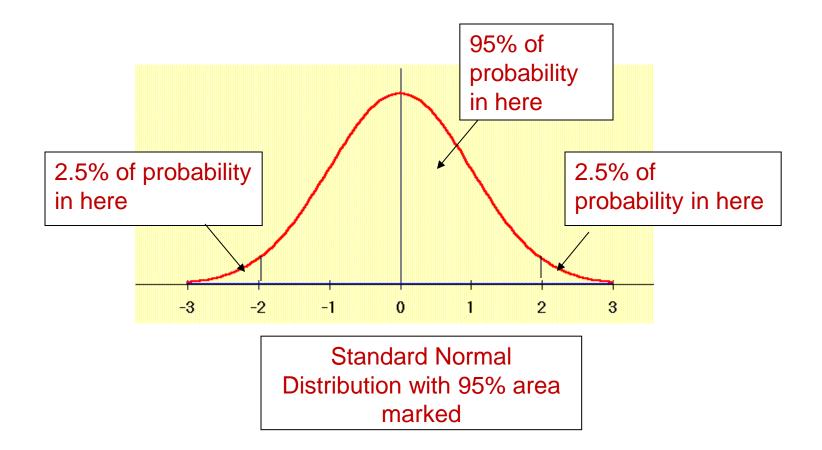
Standardizing Data: Z-Scores

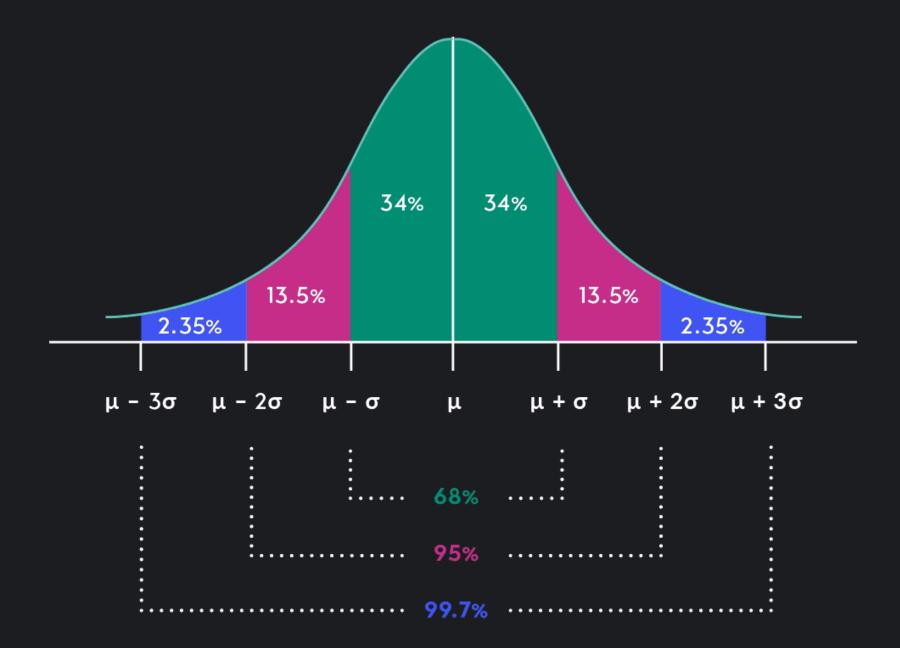
- No matter what you are measuring, a Z-score of more than +5 or less than – 5 would indicate a very, very unusual score.
- For standardized data, if it is normally distributed, 95% of the data will be between ±2 standard deviations about the mean.
- If the data follows a normal distribution,
 - 95% of the data will be between -1.96 and +1.96.
 - 99.7% of the data will fall between -3 and +3.
 - 99.99% of the data will fall between -4 and +4.

Standard Normal Distribution



Standard Normal Distribution





Calculating Probabilities

- •Probability calculations are always concerned with finding the probability that the variable assumes any value in an interval between two specific points *a* and *b*.
- •The probability that a continuous variable assumes the a value between *a* and *b* is the area under the graph of the density between *a* and *b*.

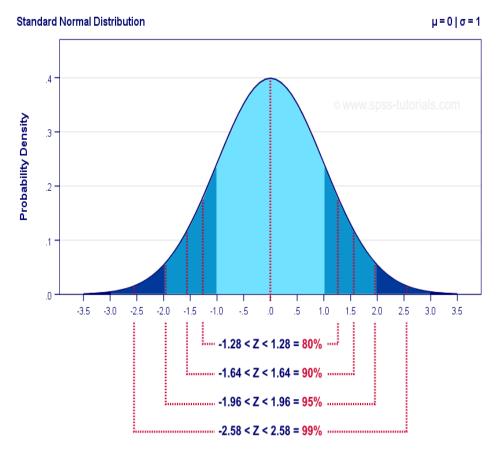


Table of Normal Curve Areas

3.80

.9999

.99999

.9999

TABLE D Normal Curve Areas $P(z \le z_i)$. Entries in the Body of the Table Are Areas Between -∞ and :

> .9750 0 1.96

z	- 0.09	- 9.06	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	0.00	1
- 3.80	.0001	.0001	.0001	.0001	1000.	1000.	.0001	.0001	.0001	.0001	- 3.80
-3.70	.0001	.0001	.0001	.0001	10001	.0001	1000.	.0001	.0001		-3.70
-3.60	1000.	.0061	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	- 3.60
-3.50	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	- 3.50
-3.40	.0000	,0003	.0003	.0003	.0003	.0003	.0003	.9003	.0003	.0003	-3.40
- 3.30	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005	-3.30
-5.20	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007	-3.20
-3.10	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010	-3.10
-3.00	.0000	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013	-3.00
-2.90	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019	-2.90
-2.80	.0009	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026	-2.80
-2.70	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035	-2.70
-2.60	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047	-2.60
-2.50	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062	-2.50
-2.40	.0054	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.00800.	.0082	-2.40
-2.30	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107	-2.30
-2.20	0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139	-2.20
-2.10	.0145	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179	- 2.10
-2.00	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228	- 2.00
-1.90	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287	-1.90
-1.80	.0294	.0304	.0507	.0314	.0322	.0329	.0336	.0344	.0351	.0359	-1.80
-1.70	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427			-1.70
-1.60	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526			-1.60
-1.50	.0559	.0571	:0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668	-1.50
-1.40	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793		-1.40
-1.30	.0823	.0838	.0853	.0869	.0685	.0901	.0918	.0934	.0951		-1.30
-1.20	.0985	.1003	.1620	.1038	.1056	.1075	.1098	.1112	.1131	.1151	
-1.10	.1170	.1190	.1210	.1250	.1251	,1271	.1292	.1314	.1335		-1.10
- 1.00	.1379	.1401	.1423	.1445	.1469	.1492	.1515	.1539	.1562	.1587	-1.00
-0.90	.1611	.1635	.1660	.1685	.1711	.1736	1762	.1788	.1814	.1841	-0.90
-0.80	.1867	.1894	.1922	.1949	.1977	.2005	.2063	.2061	.2090	.2119	-0.80
-0.70	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	2389	.2420	-0.70
-0.60	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	2709		-0.60
-0.50	.2776	.2810	2843	.2877	.2912	.2346	.2981	.3015	.3050	.3085	-0.50
-0.40	.3121	3156	.3192	,3228	.3254	.3300	.3336	.3372	.3409	.3446	-0.40
-0.30	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821	-0.30
-0.20	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207	
-0.10	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602	-0.10
0.00	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000	0.00

0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 . 0.09 . 0.00 .5040 .5080 .5120 .5160 .5199 5239 .5000 5279 .5319 .5359 0.00 5478 5438 5517 .5557 0.10 5398 .5596 .5636 .5675 .5714 .5753 0.10 .5832 5871 .5910 .5948 0.205793 .5987 .6026 .6064 .6103 .6141 0.20 6179 .6217 .6255 6293 6331 .6368 0.30.6406 6443 .6480 .6517 0.30 .6700 6664 0.40 .6554 6591 .6628.6736 .6772 6808 .6844 .6879 0.40 6950 .6985 .7019 .7054 0.50 .6915 .7088 .7123 .7157 .7190 .7224 0.50 .7324 .7357 .7389 .7454 7486 0.60 .7257 7291 .7422 .7517 .7549 0.60 0.70 .7580 .7611 .7642 .7673 .7704 .7734 .7764 .7794 .7823 .7852 0.70 .7910 .7967 .7995 0.80 .7881 .7939 .8023 .805t .8078 .8106 .8133 0.80 8212 .8238 .8264 .8289 .8315 .8340 .8365 .8389 0.90 .8159 8186 0.90 1.00 .8413 .8438 8461 .8485 .8508 .8531 .8554 8577 8599 .8621 1.00 1.10 .8643 .8665 8686 .8708 .8729 .8749 .8770 8790 .8810 .8830 1.10 8888. 1.208849 .8869 .8907 8925 .8944 8962 .8980 .8997 .9015 1.20 1.30 9032 .9049 9066 .9082 .9099 .9115 .9131 .9147 .9162 .9177 1.30 1.40 .9192 9207 .9222 .9236 .9251 .9265 .9279 .9292 .9306 .9319 1.40 1.50 9332 .9345 9357 9370 9382 .9394 9406 .9418 .9429 .9441 1.50 1.60 9452 9463 .9474 .9484 .9495 .9505 9515 .9525 .9535 .9545 1.601.70 9554 9564 .9573 .9582 .9591 .9599 9608 .9616 9625 .9633 1.70 1.80 .9641 .9649 .9656 .9664 .9671 .9678 9686 .9693 .9699 .9706 1.80 1.90 .9713 .9719 .9726 .9732 .9738 .9744 .9750 .9756 .9761 .9767 1.90 2.00 .9772 .9778 .9783 .9788 .9793 .9798 9803 9808 .9812 .9817 2.00 .9826 9830 .9834 .9838 .9842 .9846 .9850 .9854 .9857 2.10 .9821 2.102.20 9861 9864 9868 .9871 .9875 .9878 .9881 .9884 .9887 .9890 2.20 .9896 .9898 .9901 .9904 .9906 .9909 .9913 2.30 .9893 .9911 .9916 2.30.9927 .9934 .9918 .9920 .9922 .9925 .9929 .9931 .9932 .9936 2.40 2.40 .9940 .9941 9943 .9945 .9946 9948 9949 .9951 .9952 2.50 2.50 9938 2.60 .9953 .9955 .9956 .9957 .9959 .9960 .9961 .9962 .9963 .9964 2.609965 .9966 .9967 .9968 .9969 .9970 .9971 .9973 .9974 2.70 2.70 .9972 2.809974 .9975 .9976 .9977 .9977 .9978 .9979 .9979 .9980 .9981 2.80 2.90.9981 .9982 .9982 .9983 .9984 .9984 .9985 .9985 .9986 .9986 2.90 9987 9987 .9987 9988 .9988 .9989 9989 .9989 9990 .9990 3.00 3.00 9992 3.10 .9990 .9991 .9991 .9991 .9992 .9992 9992 .9993 .9993 3.10 3.20 .9993 .9993 .9994 .9994 9994 .9994 9994 9995 .9995 .9995 3.20.9995 .9996 .9996 .9996 9996 .9996 .9996 .9997 3.30 3.30 .9995 .9995 .9997 3.40 .9997 .9997 .9997 .9997 .9997 .9997 9997 .9997 .9998 3.40 9998 9998 9998 9998 .9998 9998 9998 9998 9998 3.50 9998 3.50 9998 9999 9999 9999 .99999 .99999 .9999 .99999 9999 3.60 3.60 .9998 .99999 3.70 .99999 .99999 .99999 .9999 .99999 .99999 .9999 .99999 9999 3,70 .99999 .99999 .9999 .99999 .9999

TABLE D (continued)

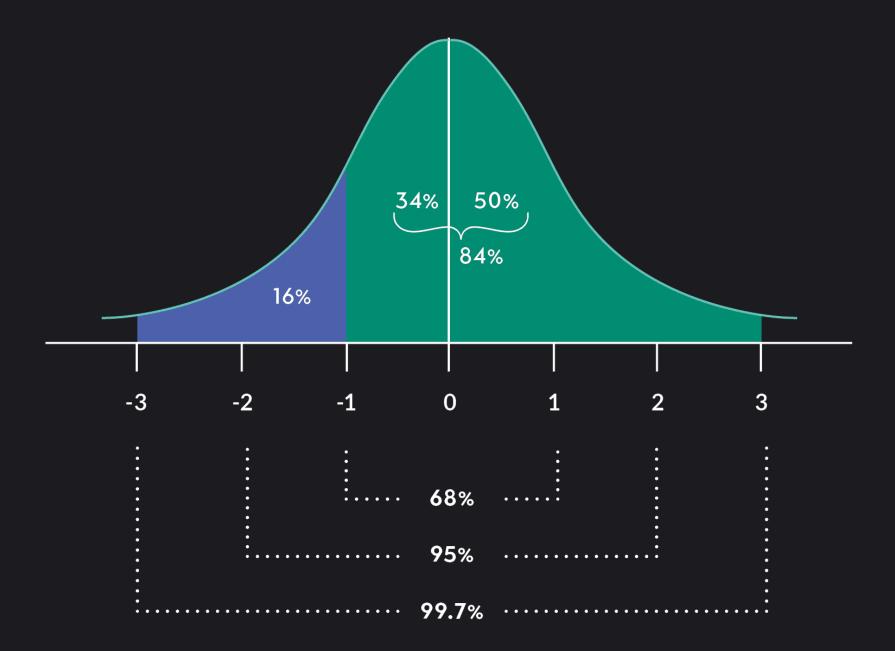
3.80

.9999

.99999

Example

- Say that a recent college graduate named Ben has an annual income, which when compared to the incomes of other recent college graduates, has a Z-score of -1.
- Assuming incomes are normally distributed, you can use the empirical rule to find the percentage of recent college graduates whose incomes are above and below Ben's.
- Roughly 16% of recent college graduates will have an income below Ben's, and roughly 84% of recent graduates will have an income above Ben's.
- In other words, a Z-score of -1 puts Ben at roughly the 16th percentile of the distribution.



(a) What is the probability that z < -1.96?

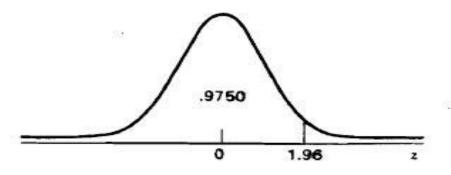
(1) Sketch a normal curve

(2) Draw a line for z = -1.96

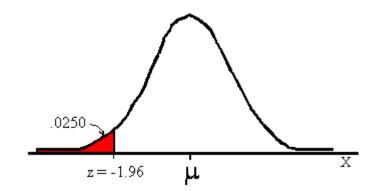
(3) Find the area in the table

(4) The answer is the area to the left of the line P(z < -

1.96) = .0250



z	- 0.09	- 0.08	- 0.07	-0.06	- 0.05	-0.04	-0.03	-0.02	-0.01	0.00	z
- 3.80	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	- 3.80
-3.70	10001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	-3.70
-3.60	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	-3.60
-3.50	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	-3.50
-3.40	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	-3.40
-3.30	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005	-3.30
-3.20	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007	-3.20
-3.10	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010	-3.10
-3.00	.0040	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013	-3.00
-2.90	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019	-2.90
-2.80	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026	-2.80
-2.70	.002%	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035	-2.70
-2.60	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047	-2.60
-2.50	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062	-2.50
-2.40	:0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082	-2.40
-2.30	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107	-2.30
-2.20	.01 10	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139	-2.20
-2.10	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179	-2.10
-2.00	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228	-2.00
(-1.90)	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287	-1.90
-1.80	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359	-1.80
-1.70	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446	-1.70
-1.60	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548	261.60



(b) What is the probability that -1.96 < z < 1.96?

(1) Sketch a normal curve

(2) Draw lines for lower z = -1.96, and

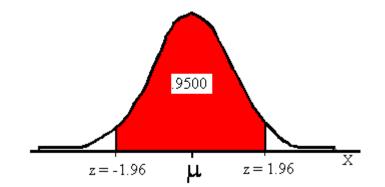
upper z = 1.96 (3) Find the area in the table corresponding to each value (4) The answer is the area between the values.

Subtract lower from upper:

P(-1.96 < z < 1.96) = .9750 - .0250 = .9500

TABLE D (continued)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
0.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359	0.00
0.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753	0.10
0.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141	0.20
0.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517	0.30
0.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879	0.40
0.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224	0.50
0.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549	0.60
0.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852	0.70
0.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133	0.80
0.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389	0.90
1.00	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621	1.00
1.10	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830	1.10
1.20	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015	1.20
1.30	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177	1.30
1.40	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319	1.40
1.50	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441	1.50
1.60	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545	1.60
1.70	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633	1.70
1.80	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706	1.80
1.90	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767	1.90
2.00	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817	2.00
2.10	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857	2.10
2.20	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890	2.20
2.30	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916	2.30
2.40	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936	2.40

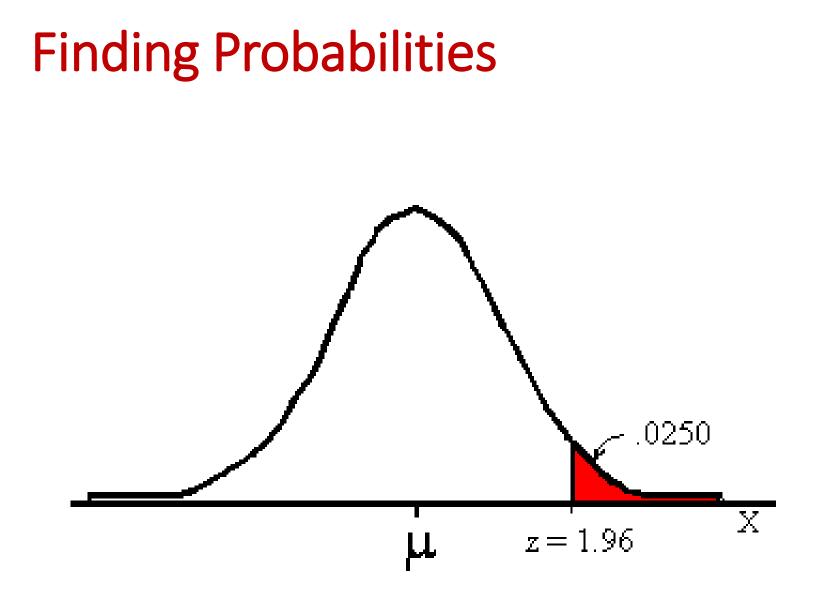


(c) What is the probability that z > 1.96?

- (1) Sketch a normal curve
- (2) Draw a line for z = 1.96
- (3) Find the area in the table

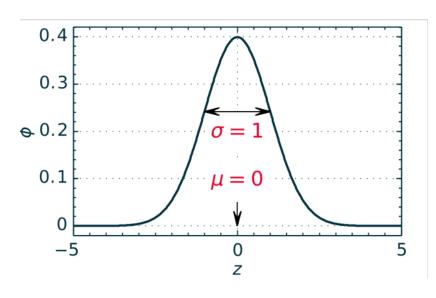
(4) The answer is the area to the right of the line. It is found by subtracting the table value from 1.0000:

P(z > 1.96) = 1.0000 - .9750 = .0250

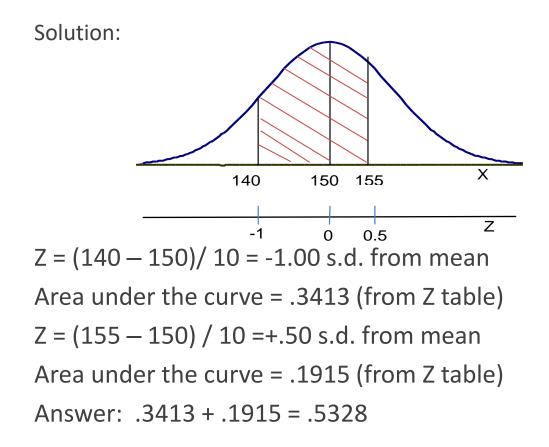


Example: Weight

If the weight of males is N.D. with μ =150 and σ =10, what is the probability that a randomly selected male will weigh between 140 Ibs and 155 lbs?



Example: Weight



Example: IQ

If IQ is ND with a mean of 100 and a S.D. of 10, what percentage of the population will have

(a) IQs ranging from 90 to 110?

(b)IQs ranging from 80 to 120?

Solution:

Z = (90 - 100)/10 = -1.00

Z = (110 - 100) / 10 = +1.00

Area between 0 and 1.00 in the Z-table is .3413; Area between 0 and -1.00 is also .3413 (Z-distribution is symmetric).

Answer to part (a) is .3413 + .3413 = .6826.

Example: IQ

• (b) IQs ranging from 80 to 120? Solution:

$$Z = (80 - 100)/10 = -2.00$$

Z = (120 - 100) / 10 = +2.00

Area between =0 and 2.00 in the Z-table is .4772; Area between 0 and -2.00 is also .4772 (Zdistribution is symmetric).

Answer is .4772 + .4772 = .9544.

Example

- Assume that SAT scores are normally distributed and that an SAT score of 1150 has a Z-score of 0.44. What percent of students score below 1150 and what percent of students score above 1150?
- To answer questions like these, you can look up the probabilities associated with the given Z-score (in this case 0.44) in a standard normal table. If you look at a standard normal table that shows probabilities to the left of the Z-score, you will find that roughly 0.67 or 67% of the scores fall below 1150. Knowing this, we can also say that 1-0.67 or 33% of scores are above 1150.

Community medicine: Biostatistics Attendance Barcode

- Register your attendance with your university number
- Make sure that the settings of your phone allow tracking location. The location should be turned on to complete this process. Go to settings > privacy> location> services> make sure that location services is ON
- If the student's name doesn't pop up, he should stop by the examination office to add it manually.

