STAT7: Discussion Section

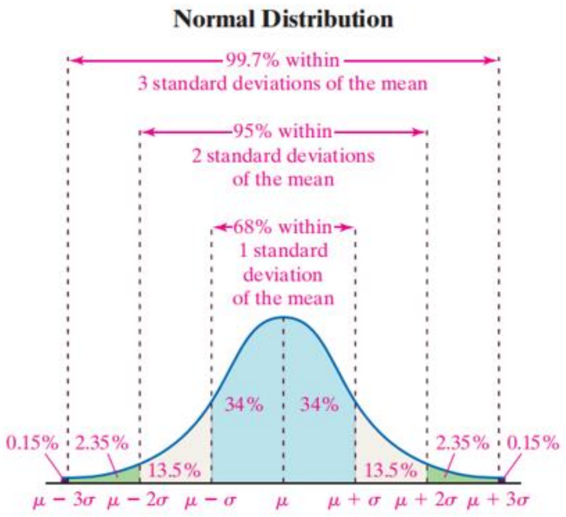
UC SANTA CRUZ

Ch7.1 Normal Distribution

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

- Normal Distribution:

- A continuous random variable is normally distributed, or has a normal probability distribution, if its relative frequency histogram has the shape of a normal curve. (Approximately a bell-shaped curve)



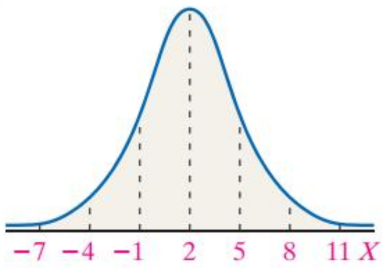
Properties of the Normal Density Curve

1. The normal curve is symmetric about its mean, μ.
2. Because mean = median = mode, the normal curve has a single peak and the highest point occurs at $x = \mu$.
3. The normal curve has inflection points at $\mu - \sigma$ and $\mu + \sigma$.
4. The area under the normal curve is 1.
5. The area under the normal curve to the right of μ equals the area under the curve to the left of μ, which equals $\frac{1}{2}$.
6. As x increases without bound (gets larger and larger), the graph approaches, but never reaches, the horizontal axis. As x decreases without bound (gets more and more negative), the graph approaches, but never reaches, the horizontal axis.

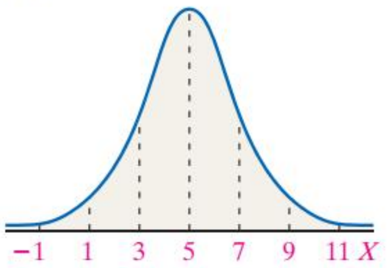
Ch7.1 Normal Distribution

In Problems 25–28, the graph of a normal curve is given. Use the graph to identify the values of μ and σ.

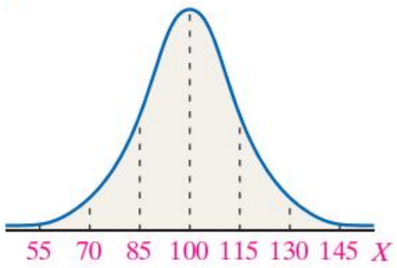
NW **25.**



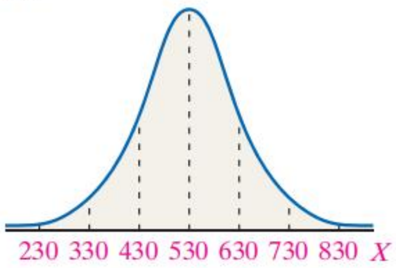
26.



27.



28.

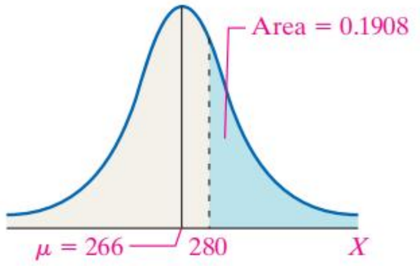


- How to specify the mean from the picture?
- The peak
- How to specify the standard deviation?
- The range of X.

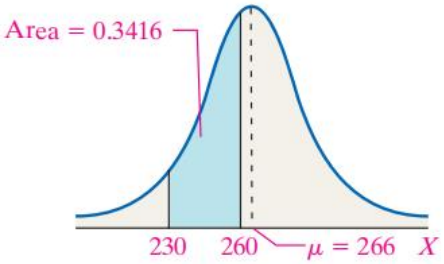
Ch7.1 Normal Distribution

NW **35. You Explain It! Gestation Period** The lengths of human pregnancies are normally distributed with $\mu = 266$ days and $\sigma = 16$ days.

(a) The figure represents the normal curve with $\mu = 266$ days and $\sigma = 16$ days. The area to the right of $x = 280$ is 0.1908. Provide two interpretations of this area.



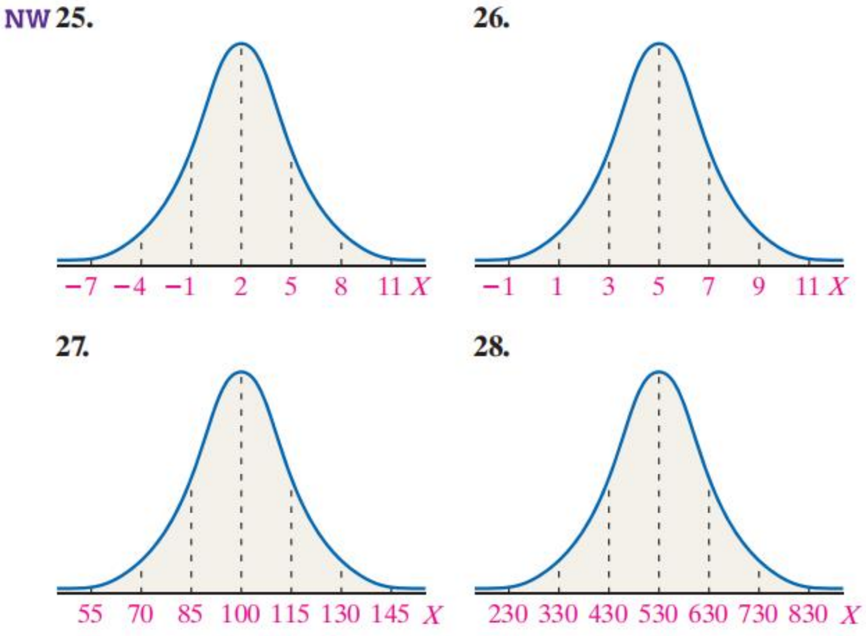
(b) The figure represents the normal curve with $\mu = 266$ days and $\sigma = 16$ days. The area between $x = 230$ and $x = 260$ is 0.3416. Provide two interpretations of this area.



- What is the sum of the area under the normal curve?
- One
- Why does the sum of the area sum to one?
- Axiom of probability.
- What does certain partition of area under the curve mean?
- Probability of x belongs to certain set

Ch7.1 Normal Distribution

In Problems 25–28, the graph of a normal curve is given. Use the graph to identify the values of μ and σ.

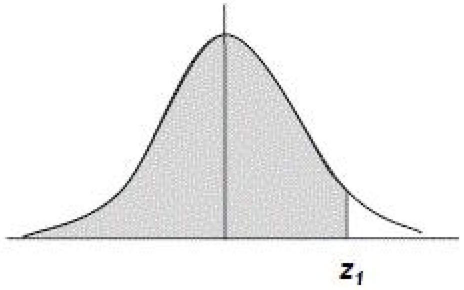


- How to specify the mean from the picture?
- The peak
- How to specify the standard deviation?
- The range of X.

UC SANTA CRUZ

Ch7.2 Application of Normal Distribution

z_1	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000



$$p(z \le z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{1}{2}z^2}\, dz$$

11. Determine the total area under the standard normal curve

(a) to the left of $z = -2$ or to the right of $z = 2$

(b) to the left of $z = -1.56$ or to the right of $z = 2.56$

(c) to the left of $z = -0.24$ or to the right of $z = 1.20$

Remember:

- All the area under the curve sums to 1.
- The distribution is symmetric about the mean.

UC SANTA CRUZ

Ch7.2 Application of Normal Distribution

41. Gestation Period The lengths of human pregnancies are approximately normally distributed, with mean $\mu = 266$ days and standard deviation $\sigma = 16$ days.

(a) What proportion of pregnancies lasts more than 270 days?
(b) What proportion of pregnancies lasts less than 250 days?
(c) What proportion of pregnancies lasts between 240 and 280 days?
(d) What is the probability that a randomly selected pregnancy lasts more than 280 days?
(e) What is the probability that a randomly selected pregnancy lasts no more than 245 days?
(f) A "very preterm" baby is one whose gestation period is less than 224 days. Are very preterm babies unusual?

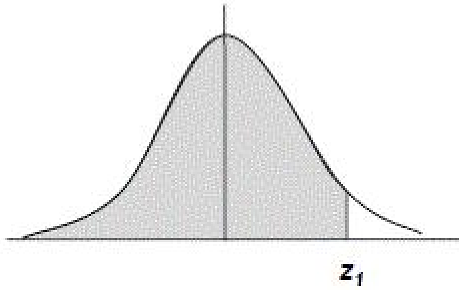
- How to calculate when you are only provided standard normal table, but not a table that you want?
- Transfer to z-score!

Population z-Score	Sample z-Score
$z = \frac{x-\mu}{\sigma}$	$z = \frac{x-\bar{x}}{s}$

Remember:

z-score follows an approximately standard normal distribution.

Ch7.2 Application of Normal Distribution



$$p(z \le z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{1}{2}z^2} dz$$

(a) What proportion of pregnancies lasts more than 270 days?

μ = 266, σ = 16

$$P(x > 270) = P(x - 266) > (270 - 266) = P(\frac{\boldsymbol{x - 266}}{\boldsymbol{16}} > \frac{270 - 266}{16} = 0.25)$$

Check the table again to get the number, or by technical methods.
Other questions similar, leave for your exercise.

UC SANTA CRUZ

Ch7.3 Assessing Normality

Why do we need to assess normality?

- Traditional judgement from common sense is not rigorous enough. Need clear and more scientific method.

Drawing a Normal Probability Plot

Step 1 Arrange the data in ascending order.

Step 2 Compute $f_i = \frac{i - 0.375}{n + 0.25}$,* where i is the index (the position of the data value in the ordered list) and n is the number of observations. The expected proportion of observations less than or equal to the ith data value is f_i.

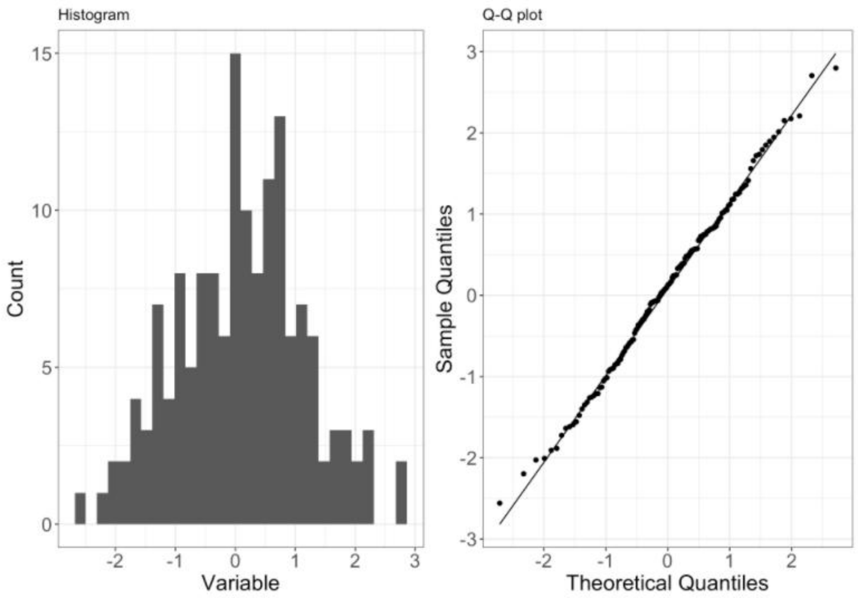
Step 3 Find the z-score corresponding to f_i from Table V.

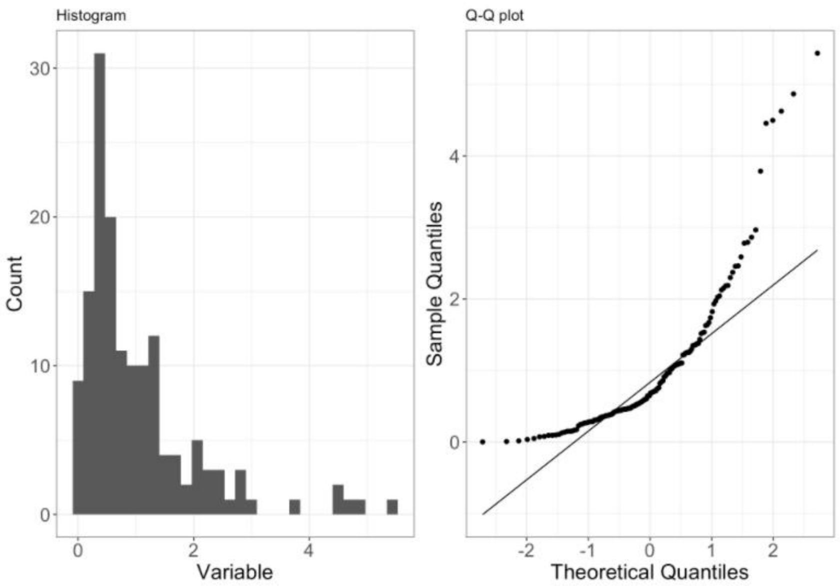
Step 4 Plot the observed values on the horizontal axis and the corresponding expected z-scores on the vertical axis.

Ch7.3 Assessing Normality

How to judge from the plot?

- If the points approximately are on a straight line, the slope is approximately 1 for quantile-quantile plot, then it satisfies perfect normality.



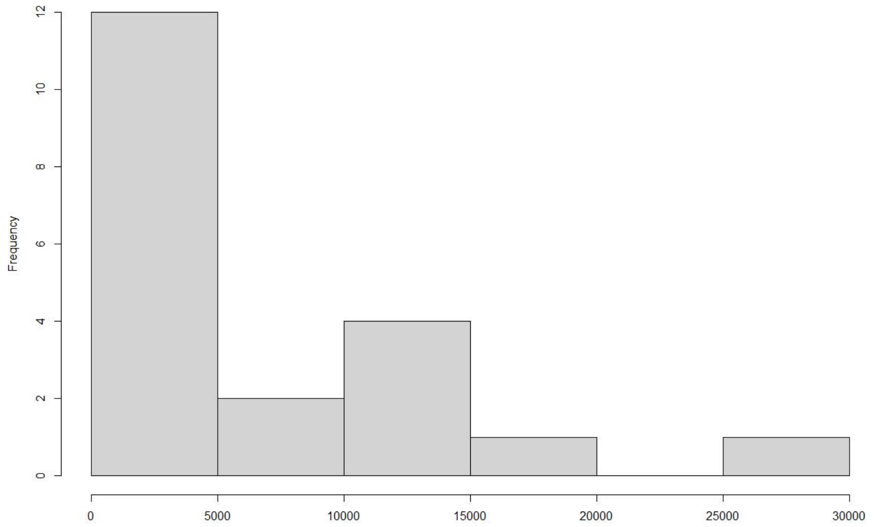


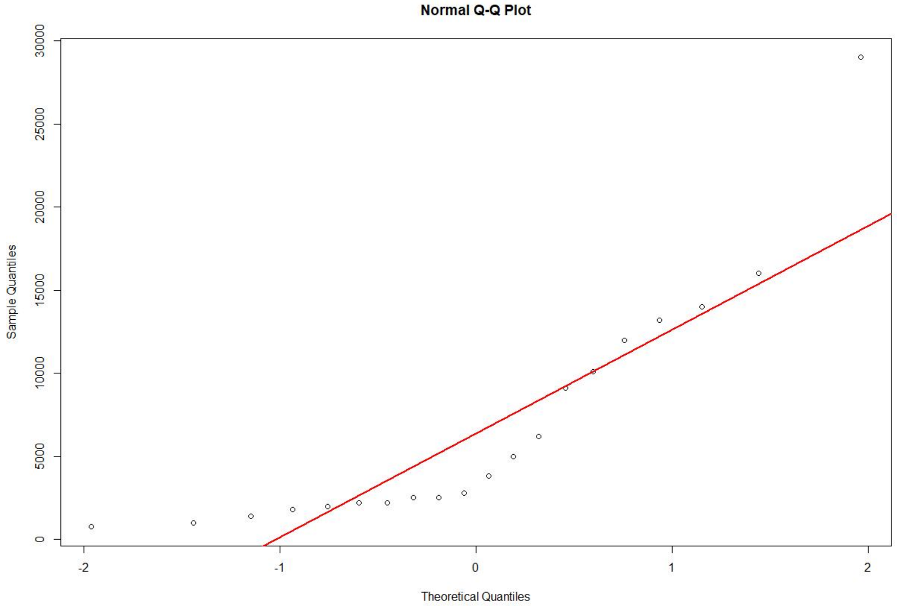
UC SANTA CRUZ

Ch7.3 Assessing Normality

9. School Loans A random sample of 20 undergraduate students receiving student loans was obtained, and the amount of their loans for the 2014–2015 school year was recorded.

2,500	1,000	2,000	14,000	1,800
3,800	10,100	2,200	29,000	16,000
5,000	2,200	6,200	9,100	2,800
2,500	1,400	13,200	750	12,000





Not normal. Also, we can judge the correlation coefficient (you will learn later) between theoretical and true z-score. If the correlation is greater than 0.951, then we can say it is normally distributed. But the correlation in this case is: 0.88 < 0.951.

UC SANTA CRUZ

Law of Large Number

Weak law [edit]

The **weak law of large numbers** (also called Khinchin's law) states that the sample average converges in probability towards the expected value[17]

$$\overline{X}_n \xrightarrow{P} \mu \qquad \text{when } n \to \infty.$$

UC SANTA CRUZ

Central Limit Theorem

Central limit theorem formula

Fortunately, you don't need to actually repeatedly sample a population to know the shape of the sampling distribution. The parameters of the sampling distribution of the mean are determined by the parameters of the population:

- The mean of the sampling distribution is the mean of the population.

  $$\mu_{\bar{x}} = \mu$$

- The standard deviation of the sampling distribution is the standard deviation of the population divided by the square root of the sample size.

  $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

We can describe the sampling distribution of the mean using this notation:

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$