Normal distribution

- The normal distribution is the most important distribution. It describes well the distribution of random variables that arise in practice, such as the heights or weights of people, the total annual sales of a firm, exam scores etc. Also, it is important for the central limit theorem, the approximation of other distributions such as the binomial, etc.

- We say that a random variable X follows the normal distribution if the probability density function of X is given by

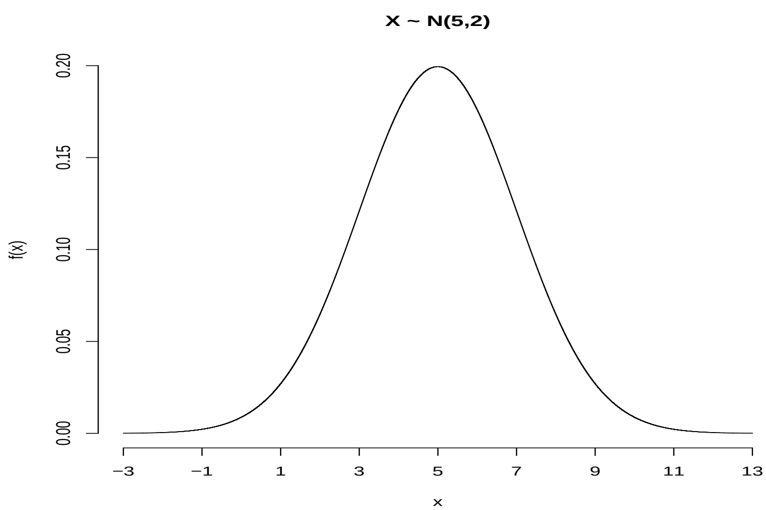
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \quad -\infty < x < \infty$$

  This is a bell-shaped curve.

- We write $X \sim N(\mu, \sigma)$. We read: X follows the normal distribution (or X is normally distributed) with mean μ, and standard deviation σ.

- The normal distribution can be described completely by the two parameters μ and σ. As always, the mean is the center of the distribution and the standard deviation is the measure of the variation around the mean.

- Shape of the normal distribution. Suppose $X \sim N(5, 2)$.



- The area under the normal curve is 1 (100%).

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

- The normal distribution is symmetric about μ. Therefore, the area to the left of μ is equal to the area to the right of μ (50% each).

- Useful rule (see figure above):
  The interval $\mu \pm 1\sigma$ covers the middle $\sim 68\%$ of the distribution.
  The interval $\mu \pm 2\sigma$ covers the middle $\sim 95\%$ of the distribution.
  The interval $\mu \pm 3\sigma$ covers the middle $\sim 100\%$ of the distribution.

- Because the normal distribution is symmetric it follows that
  $P(X > \mu + \alpha) = P(X < \mu - \alpha)$

- The normal distribution is a continuous distribution. Therefore,

  $$P(X \geq a) = P(X > a), \text{ because } P(X = a) = 0. \text{ Why?}$$

- How do we compute probabilities? Because the following integral has no closed form solution

  $$P(X > \alpha) = \int_{\alpha}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = \ldots$$

  the computation of normal distribution probabilities can be done through the standard normal distribution Z:

  $$Z = \frac{X - \mu}{\sigma}$$

  Theorem:
  Let $X \sim N(\mu, \sigma)$. Then $Y = \alpha X + \beta$ follows also the normal distribution as follows:

  $$Y \sim N(\alpha\mu + \beta, \alpha\sigma)$$

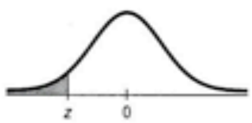
  Therefore, using this theorem we find that

  $$Z \sim N(0, 1)$$

  It is said that the random variable Z follows the standard normal distribution and we can find probabilities for the Z distribution from tables (see next pages).

The standard normal distribution table:

Table Z
Areas under the standard normal curve



Second decimal place in z										
0.09	*0.08*	*0.07*	*0.06*	*0.05*	*0.04*	*0.03*	*0.02*	*0.01*	*0.00*	*z*
									0.0000†	−3.9
0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	−3.8
0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	−3.7
0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	−3.6
0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	−3.5
0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	−3.4
0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005	−3.3
0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007	−3.2
0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010	−3.1
0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013	−3.0
0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019	−2.9
0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026	−2.8
0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035	−2.7
0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047	−2.6
0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062	−2.5
0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082	−2.4
0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107	−2.3
0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139	−2.2
0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179	−2.1
0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228	−2.0
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	−1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	−1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	−1.7
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	−1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	−1.5
0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808	−1.4
0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968	−1.3
0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151	−1.2
0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357	−1.1
0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587	−1.0
0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841	−0.9
0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119	−0.8
0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420	−0.7
0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743	−0.6
0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085	−0.5
0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446	−0.4
0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821	−0.3
0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207	−0.2
0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602	−0.1
0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000	−0.0

†For $z \leq 3.90$, the areas are 0.0000 to four decimal places.

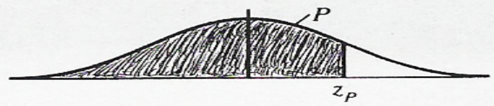
Table Z (cont.)
Areas under the
standard normal curve

z	Second decimal place in z 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†									

†For $z \geq 3.90$, the areas are 1.0000 to four decimal places.

TABLE 2

Cumulative Normal Distribution—Values of P Corresponding to z_p for the Normal Curve



z is the standard normal variable. The value of P for $-z_p$ equals 1 minus the value of P for $+z_p$; for example, the P for -1.62 equals $1 - .9474 = .0526$.

z_p	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Example:
Suppose the diameter of a certain car component follows the normal distribution with $X \sim N(10, 3)$. Find the proportion of these components that have diameter larger than 13.4 mm. Or, if we randomly select one of these components, find the probability that its diameter will be larger than 13.4 mm.
Answer:

$$P(X > 13.4) = P(X - 10 > 13.4 - 10) =$$

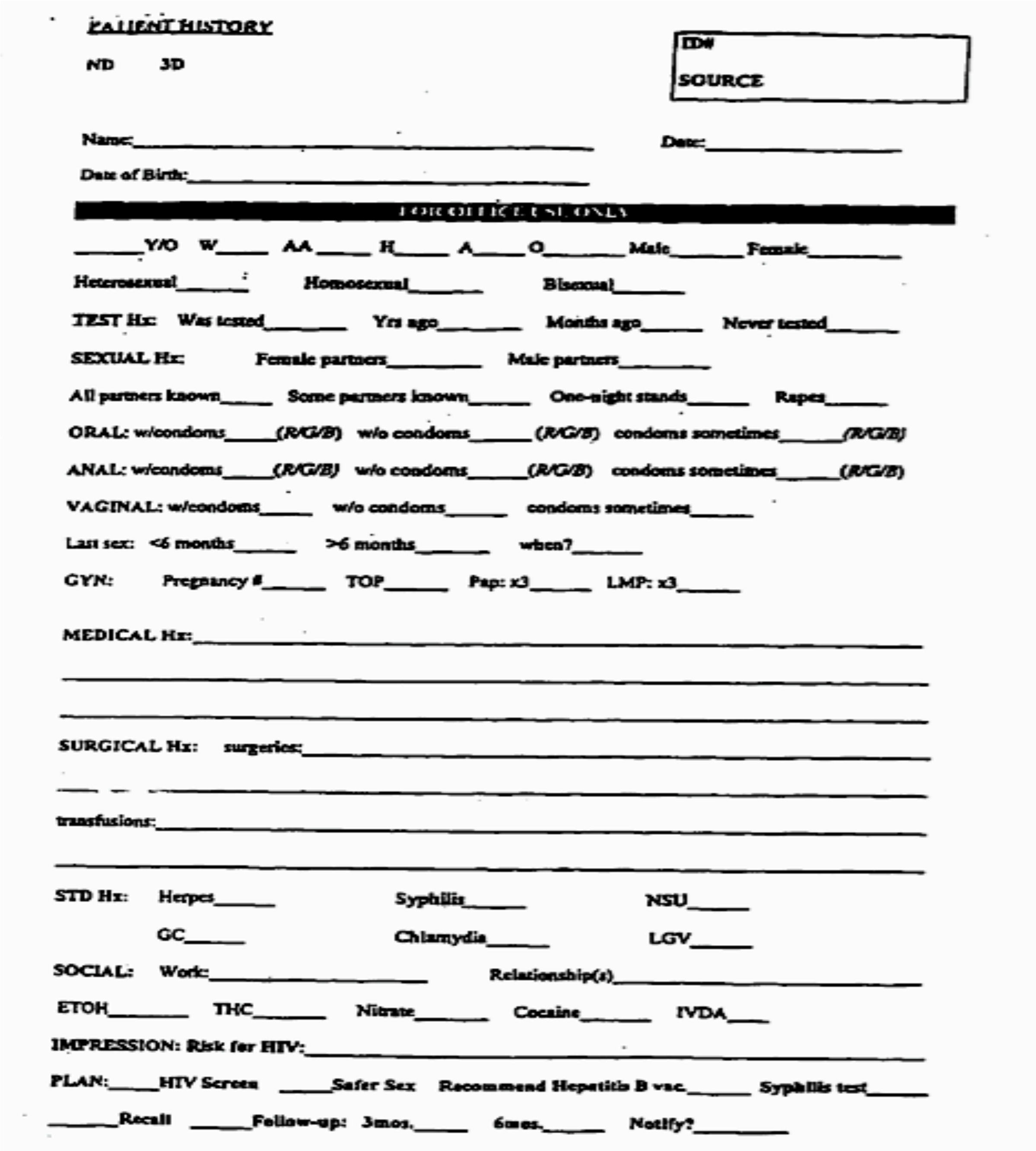
$$P\left(\frac{X-10}{3} > \frac{13.4-10}{3}\right) = P(Z > 1.13) = 1 - 0.8708 = 0.1292.$$

We read the number **0.8708** from the table. First we find the value of $z = 1.13$ (first column and first row of the table - the first row gives the second decimal of the value of z). Therefore the probability that the diameter is larger than 13.4 mm is 12.92%.

Question:
What is z? The value of z gives the number of standard deviations the particular value of X lies above or below the mean μ. In other words, $X = \mu \pm z\sigma$, and in our example the value $x = 13.4$ lies 1.13 standard deviations away from the mean. Of course z will be negative when the value of x is below the mean.
Example: Find the proportion of these components with diameter less than 5.1 mm.
Answer: $P(X < 5.1) = P(Z < \frac{5.1-10}{3}) = P(Z < -1.63) = 0.0516$. Here, the value of $x = 5.1$ lies 1.63 standard deviations below the mean μ.

Finding percentiles of the normal distribution:
Find the 25_{th} percentile (or first quartile) of the distribution of X. In other words, find c such that $P(X \leq c) = 0.25$.
Answer:
First we find (approximately) the probability 0.25 from the table and we read the corresponding value of z. Here it is equal to $z = -0.675$. It is negative because the first percentile is below the mean. Therefore, $x_{25} = 10 - 0.675(3) = 7.975$.

Normal distribution - example:



PATIENT HISTORY

ND 3D

ID#

SOURCE

Name:________________ Date:________

Date of Birth:________________

FOR OFFICE USE ONLY

_____Y/O W_____ AA_____ H_____ A_____ O_____ Male_____ Female_____

Heterosexual_____ Homosexual_____ Bisexual_____

TEST Hx: Was tested_____ Yrs ago_____ Months ago_____ Never tested_____

SEXUAL Hx: Female partners_____ Male partners_____

All partners known_____ Some partners known_____ One-night stands_____ Rapes_____

ORAL: w/condoms_____(R/G/B) w/o condoms_____(R/G/B) condoms sometimes_____(R/G/B)

ANAL: w/condoms_____(R/G/B) w/o condoms_____(R/G/B) condoms sometimes_____(R/G/B)

VAGINAL: w/condoms_____ w/o condoms_____ condoms sometimes_____

Last sex: <6 months_____ >6 months_____ when?_____

GYN: Pregnancy #_____ TOP_____ Pap: x3_____ LMP: x3_____

MEDICAL Hx:________________

SURGICAL Hx: surgeries:________________

transfusions:________________

STD Hx: Herpes_____ Syphilis_____ NSU_____

GC_____ Chlamydia_____ LGV_____

SOCIAL: Work:________ Relationship(s)________

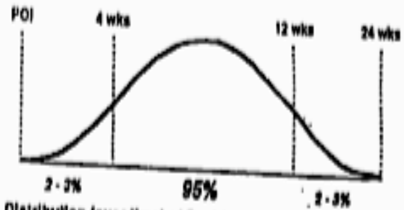
ETOH_____ THC_____ Nitrate_____ Cocaine_____ IVDA_____

IMPRESSION: Risk for HIV:________________

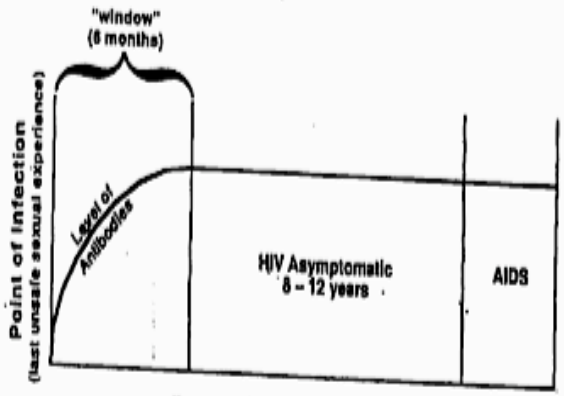
PLAN:_____HIV Screen _____Safer Sex Recommend Hepatitis B vac._____ Syphilis test_____

_____Recall _____Follow-up: 3mos._____ 6mos._____ Notify?_____

Risk	Protection
exchange of body fluids	abstain
– semen } orally	condom
– vaginal fluid } vaginally	dental dam
– blood } anally	N-9
– menstruation	K-Y



Distribution (over time) of Positive Results After Infection



Progression of HIV Infection to AIDS

ELISA	+	+	–
Western Blot	+	unreadable	☒
patient results	+	inconclusive	–

PCR

Confidentiality

Partner Notification

Feelings

Life Changes (+ or –)

Normal distribution - finding probabilities and percentiles

Suppose that the weight of navel oranges is normally distributed with mean $\mu = 8$ ounces, and standard deviation $\sigma = 1.5$ ounces. We can write $X \sim N(8, 1.5)$. Answer the following questions:

a. What proportion of oranges weigh more than 11.5 ounces? (or if you randomly select a navel orange, what is the probability that it weighs more than 11.5 ounces?).

$$P(X > 11.5) = P(Z > \frac{11.5-8}{1.5}) = P(Z > 2.33) = 1 - 0.9901 = 0.0099.$$

b. What proportion of oranges weigh less than 8.7 ounces?

$$P(X < 8.7) = P(Z < \frac{8.7-8}{1.5}) = P(Z < 0.47) = 0.6808.$$

c. What proportion of oranges weigh less than 5 ounces?

$$P(X < 5) = P(Z < \frac{5-8}{1.5}) = P(Z < -2.00) = 0.0228.$$

d. What proportion of oranges weigh more than 4.9 ounces?

$$P(X > 4.9) = P(Z > \frac{4.9-8}{1.5}) = P(Z > -2.07) = 1 - 0.0192 = 0.9808.$$

e. What proportion of oranges weigh between 6.2 and 7 ounces?

$$P(6.2 < X < 7) = P(\frac{6.2-8}{1.5} < Z < \frac{7-8}{1.5}) = P(-1.2 < Z < -0.67) =$$
$$0.2514 - 0.1151 = 0.1363.$$

f. What proportion of oranges weigh between 10.3 and 14 ounces?

$$P(10.3 < X < 14) = P(\frac{10.3 - 8}{1.5} < Z < \frac{14 - 8}{1.5}) = P(1.53 < Z < 4) \approx$$
$$1 - 0.9370 = 0.0630.$$

g. What proportion of oranges weigh between 6.8 and 8.9 ounces?

$$P(6.8 < X < 8.9) = P(\frac{6.8 - 8}{1.5} < Z < \frac{8.9 - 8}{1.5}) = P(-0.8 < Z < 0.6) =$$
$$0.7257 - 0.2119 = 0.5138.$$

h. Find the 80_{th} percentile of the distribution of X. This question can also be asked as follows: Find the value of X below which you find the lightest 80% of all the oranges.

$$z = \frac{x - \mu}{\sigma} \Rightarrow 0.845 = \frac{x - 8}{1.5} \Rightarrow x = 9.27.$$

i. Find the 5_{th} percentile of the distribution of X.

$$z = \frac{x - \mu}{\sigma} \Rightarrow -1.645 = \frac{x - 8}{1.5} \Rightarrow x = 5.53.$$

j. Find the interquartile range of the distribution of X.

Normal distribution - Examples

Example 1
The lengths of the sardines received by a certain cannery is normally distributed with mean 4.62 inches and a standard deviation 0.23 inch. What percentage of all these sardines is between 4.35 and 4.85 inches long?

Example 2
A baker knows that the daily demand for apple pies is a random variable which follows the normal distribution with mean 43.3 pies and standard deviation 4.6 pies. Find the demand which has probability 5% of being exceeded.

Example 3
Suppose that the height of $UCLA$ female students has normal distribution with mean 62 inches and standard deviation 8 inches.

a. Find the height below which is the shortest 30% of the female students.

b. Find the height above which is the tallest 5% of the female students.

Example 4
A firm's marketing manager believes that total sales for next year will follow the normal distribution, with mean of \$2.5 million and a standard deviation of $\$300,000$.

a. What is the probability that the firm's sales will fall within \$150000 of the mean?

b. Determine the sales level that has only a 9% chance of being exceeded next year.

Example 5
To avoid accusations of sexism in a college class equally populated by male and female students, the professor flips a fair coin to decide whether to call upon a male or female student to answer a question directed to the class. The professor will call upon a female student if a tails occurs. Suppose the professor does this 1000 times during the semester.

a. What is the probability that he calls upon a female student at least 530 times?

b. What is the probability that he calls upon a female student at most 480 times?

c. What is the probability that he calls upon a female student exactly 510 times?

Example 6
$MENSA$ is an organization whose members possess IQs in the top 2% of the population.

a. If IQs are normally distributed, with mean 100 and a standard deviation of 16, what is the minimum IQ required for admission to $MENSA$?

b. If three individuals are chosen at random from the general population what is the probability that all three satisfy the minimum requirement for $MENSA$?

Example 7
A manufacturing process produces semiconductor chips with a known failure rate 6.3%. Assume that chip failures are independent of one another. You will be producing 2000 chips tomorrow.

a. Find the expected number of defective chips produced.

b. Find the standard deviation of the number of defective chips.

c. Find the probability (approximate) that you will produce less than 135 defects.

EXERCISE 8
Suppose that the height (X) in inches, of a 25-year-old man is a normal random variable with mean $\mu = 70$ inches. If $P(X > 79) = 0.025$ what is the standard deviation of this random normal variable?

EXERCISE 9
Suppose that the weight (X) in pounds, of a 40-year-old man is a normal random variable with standard deviation $\sigma = 20$ pounds. If 5% of this population is heavier than 214 pounds what is the mean μ of this distribution?

Problem 10
At *Heinz* ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.1 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of the bottle goes below 35.8 oz. or above 36.2 oz., then the bottle will be declared out of control.

a. If the process is in control, meaning $\mu = 36$ oz. and $\sigma = 0.1$ oz., find the probability that a bottle will be declared out of control.

b. In the situation of (a), find the probability that the number of bottles found out of control in an eight-hour day (16 inspections) will be zero.

c. In the situation of (a), find the probability that the number of bottles found out of control in an eight-hour day (16 inspections) will be exactly one.

d. If the process shifts so that $\mu = 37$ oz and $\sigma = 0.4$ oz, find the probability that a bottle will be declared out of control.

Problem 11
Suppose that a binary message -either 0 or 1- must be trasmitted by wire from location A to location B. However, the data sent over the wire are subject to a channel noise disturbance, so to reduce the possibilty of error, the value 2 is sent over the wire when the message is 1 and the value -2 is sent when the message is 0. If x, $x = \pm 2$, is the value sent from location A, then R, the value received at location B, is given by $R = x + N$, where N is the channel noise disturbance. When the message is received at location B the receiver decodes it according to the following rule:

If $R \geq 0.5$, then 1 is concluded

If $R < 0.5$, then 0 is concluded

If the channel noise follows the standard normal distribution compute the probability that the message will be wrong when decoded.

Normal distribution - Examples
Solutions

Example 1
We are given $X \sim N(4.62, 0.23)$. We want to compute

$$P(4.35 < X < 4.85) = P(\frac{4.35 - 4.62}{0.23} < Z < \frac{4.85 - 4.62}{0.23}) \quad = \\ P(-1.17 < z < 1) = 0.8413 - 0.1210 = 0.7203.$$

Example 2
We are given $X \sim N(43.3, 4.6)$. We want to find the demand d such that $P(X > d) = 0.05$. From the standard normal table this corresponds to $z = 1.645$. Therefore $1.645 = \frac{d-43.3}{4.6} \Rightarrow d = 50.9$ pies.

Example 3
We are given $X \sim N(62, 8)$.

a. We want to find the height h such that $P(X < h) = 0.30$. From the standard normal table this corresponds to $z = -0.525$. Therefore $-0.525 = \frac{h-62}{8} \Rightarrow h = 57.8$ inches.

b. We want to find the height h such that $P(X > h) = 0.05$. From the standard normal table this corresponds to $z = 1.645$. Therefore $1.645 = \frac{h-62}{8} \Rightarrow h = 75.16$ inches.

Example 4
We are given $X \sim N(2500000, 300000)$.

a. $P(2350000 < X < 2650000) = P(\frac{2350000-250000}{300000} < z < \frac{2650000-2500000}{300000}) =$
$P(-0.5 < z < 0.5) = 0.6915 - 0.3085) = 0.3830$.

b. We want to find the sales level s such that $P(X > s) = 0.09$. This corresponds to $z = 1.345$. Therefore $1.345 = \frac{s-2500000}{300000} \Rightarrow s = 2903500$.

Example 5
This is a binomial problem but we are going to use the normal distribution as an approximation. We need μ and σ. These are: $\mu = np = 1000\frac{1}{2} = 500$. And $\sigma^2 = np(1-p) = 1000\frac{1}{2}(1-\frac{1}{2}) = 250 \Rightarrow \sigma = 15.81$.

a. $P(X \geq 530) = P(z > \frac{529.5-500}{15.81}) = P(z > 1.87) = 1 - 0.9693 = 0.0307$.

b. $P(X \leq 480) = P(z < \frac{480.5-500}{15.81}) = P(z < -1.23) = 0.1093$.

c. $P(X = 510) = P(\frac{509.5-500}{15.81} < z < \frac{510.5-500}{15.81}) = P(0.60 < z < 0.66) = 0.7454 - 0.7257 = 0.0197$.

Example 6
We are given $X \sim N(100, 16)$.

a. We want to find the IQ q such that $P(X > q) = 0.02$. This corresponds to $z = 2.055$. Therefore $2.055 = \frac{q-100}{16} \Rightarrow q = 132.88$.

b. This is binomial with $X \sim b(3, 0.02)$. We want $P(X = 3) = \binom{3}{3}0.2^3(1-0.02)^0 = 0.000008$.

Example 7

This is binomial with $X \sim b(2000, 0.063)$.

a. $E(X) = np = 2000(0.063) = 126$.

b. $\sigma^2 = np(1-p) = 2000(0.063)(1-0.063) = 118.06 \Rightarrow \sigma = 10.87$.

c. $P(X < 135) = P(z < \frac{134.5-126}{10.87}) = P(z < 0.78) = 0.7823$.

EXERCISE 8

We are given $X \sim N(70, \sigma)$. From $P(X > 79) = 0.025$ we find the corresponding z-value: $z = 1.96$. Therefore $1.96 = \frac{79-70}{\sigma} \Rightarrow \sigma = 4.59$ inches.

EXERCISE 9

We are given $X \sim N(\mu, 20)$. From $P(X > 214) = 0.05$ we find the corresponding z-value: $z = 1.645$. Therefore $1.645 = \frac{214-\mu}{20} \Rightarrow \mu = 181.1$ pounds.

Problem 10

The process is out of control if $P(X < 35.8)$ or $P(X > 36.2)$.

a. We are given $X \sim N(36, 0.1)$. We compute the probability:
$P(X < 35.8) + P(X > 36.2) = P(z < \frac{35.8-36}{0.1}) + P(z > \frac{36.2-36}{0.1}) = P(z < -2) + P(z > 2) = 0.0228 + (1 - 0.9772) = 0.0456$.

b. This is binomial with $n = 16, p = 0.0456$.
$P(X = 0) = \binom{16}{0}(0.0456)^0(1-0.0456)^{16} = 0.4739$.

c. This is binomial with $n = 16, p = 0.0456$.
$P(X = 1) = \binom{16}{1}(0.0456)^1(1-0.0456)^{15} = 0.3623$.

d. Now $X \sim N(37, 0.4)$. We compute the probability:
$P(X < 35.8) + P(X > 36.2) = P(z < \frac{35.8-37}{0.4}) + P(z > \frac{36.2-37}{0.4}) = P(z < -3) + P(z > -2) = 0.0013 + (1 - 0.0028) = 0.9785$.

Problem 11

The channel noise N follows the standard normal distribution, $N(0,1)$.
If the message was 1: It will be wrong when decoded if $R < 0.5$. Or $x + N < 0.5 \Rightarrow 2 + N < 0.5 \Rightarrow N < -1.5$. This probability is equal to $P(z < -1.5) = 0.0668$. If the message was 0: It will be wrong when decoded if $R \geq 0.5$. Or $x + N \geq 0.5 \Rightarrow -2 + N \geq 0.5 \Rightarrow N \geq 2.5$. This probability is equal to $P(z \geq 2.5) = 1 - 0.9938 = 0.0062$.

Normal distribution - Practice problems

Problem 1

The chickens of the Ornithes farm are processed when they are 20 weeks old. The distribution of their weights is normal with mean 3.8 lb, and standard deviation 0.6 lb. The farm has created three categories for these chickens according to their weight: petite (weight less than 3.5 lb), standard (weight between 3.5 lb and 4.9 lb), and big (weight above 4.9 lb).

a. What proportion of these chickens will be in each category? Show these proportions on the normal distribution graph.

b. Find the 60_{th} percentile of the distribution of the weight. In other words find c such that $P(X < c) = 0.60$.

c. Suppose that 5 chickens are selected at random. What is the probability that 3 of them will be petite?

Problem 2

A television cable company receives numerous phone calls throughout the day from customers reporting service troubles and from would-be subscribers to the cable network. Most of these callers are put "on hold" until a company operator is free to help them. The company has determined that the length of time a caller is on hold is normally distributed with a mean of 3.1 minutes and a standard deviation 0.9 minutes. Company experts have decided that if as many as 5% of the callers are put on hold for 4.8 minutes or longer, more operators should be hired.

a. What proportion of the company's callers are put on hold for more than 4.8 minutes? Should the company hire more operators? Show these probabilities on a sketch of the normal curve.

b. At another cable company (length of time a caller is on hold follows the same distribution as before), 2.5% of the callers are put on hold for longer than x minutes. Find the value of x, and show this on a sketch of the normal curve.

Problem 3

Answer the following questions:

a. Suppose that the height (X) in inches, of a 25-year-old man is a normal random variable with mean $\mu = 70$ inches. If $P(X > 79) = 0.025$ what is the standard deviation of this random normal variable?

b. Suppose that the weight (X) in pounds, of a 40-year-old man is a normal random variable with standard deviation $\sigma = 20$ pounds. If 5% of this population weigh less than 160 pounds what is the mean μ of this distribution?

c. Find an interval that covers the middle 95% of $X \sim N(64, 8)$.

Problem 4

A bag of cookies is underweight if it weighs less than 500 grams. The filling process dispenses cookies with weight that follows the normal distribution with mean 510 grams and standard deviation 4 grams.

a. What is the probability that a randomly selected bag is underweight?

b. If you randomly select 5 bags, what is the probability that exactly 2 of them will be underweight?

Problem 5

Answer the following questions:

a. Suppose that X follows the normal distribution with mean $\mu = 5$. If $P(X > 9) = 0.2$ find the variance of X.

b. Let X be a normal random variable with mean $\mu = 12$ and standard deviation $\sigma = 2$. Find the $10th$ percentile of this distribution.

c. The weight X of water melons is normally distributed with mean $\mu = 10$ pounds and standard deviation $\sigma = 2$ pounds. Find c such that $P(X > c) = 0.60$.

d. The montly return of a particular stock follows the normal distribution with mean 0.02 and standard deviation 0.1. Find the 85_{th} percentile of this distribution.

e. Find the probability that the monthly return of the stock in question (b) will be larger that 0.2.

f. Find the probability that in one year (12 months), the return of the stock in question (e) will be larger than 0.2 on exactly 4 months. Assume that the returns are independent from month to month.

g. The annual rainfall X (in inches) at a certain region is normally distributed with mean $\mu = 40$ pounds and standard deviation $\sigma = 4$. What is the probability that starting with this year, it will take more than 10 years before a year occurs having a rainfall of over 50 inches?

h. Let $X \sim N(100, 20)$. Find $P(X > 70 | X < 90)$.

Problem 6

The diameters of apples from the Milo Farm follow the normal distribution with mean 3 inches and standard deviation 0.3 inch. Apples can be size-sorted by being made to roll over a mesh screens. First the apples are rolled over a screen with mesh size 2.5 inches. This separates out all the apples with diameters < 2.5 inches. Second, the remaining apples are rolled over a screen with mash size 3.2 inches. Find the proportion of apples with diameters < 2.5 inches, between 2.5 and 3.2 inches, and greater than 3.2 inches. Use only *SOCR* to find the answers and print the appropriate snapshots.

Normal distribution - Practice problems
Solutions

Problem 1

The chickens of the Ortnithes farm are processed when they are 20 weeks old. The distribution of their weights is normal with mean 3.8 lb, and standard deviation 0.6 lb. The farm has created three categories for these chickens according to their weight: petite (weight less than 3.5 lb), standard (weight between 3.5 lb and 4.9 lb), and big (weight above 4.9 lb).

a. What proportion of these chickens will be in each category? Show these proportions on the normal distribution graph.
Answer:
Petite: $P(X < 3.5) = P(Z < \frac{3.5-3.8}{0.6}) = P(Z < -0.50) = 0.3085$.
Standard: $P(3.5 < X < 4.9) = P(\frac{3.5-3.8}{0.6} < Z < \frac{4.9-3.8}{0.6}) = P(-0.5 < Z < 1.83) = 0.9664 - 0.3085 = 0.6579$.
Big: $P(X > 4.9) = P(Z > \frac{4.9-3.8}{0.6}) = P(Z > 1.83) = 1 - 0.9664 = 0.0336$.

b. Find the 60_{th} percentile of the distribution of the weight. In other words find c such that $P(X < c) = 0.60$.
Answer:
From the z table approximately $z = 0.2055$. Therefore, $0.2055 = \frac{x-3.8}{0.6} \Rightarrow x = 3.8 + 0.2055(0.6) = 3.92$.

c. Suppose that 5 chickens are selected at random. What is the probability that 3 of them will be petite?
Answer:
This is binomial with $n = 5, p = 0.3085$. Therefore, $P(Y = 3) = \binom{5}{3}0.3085^3(1 - 0.3085)^2 = 0.1404$.

Problem 2

A television cable company receives numerous phone calls throughout the day from customers reporting service troubles and from would-be subscribers to the cable network. Most of these callers are put "on hold" until a company operator is free to help them. The company has determined that the length of time a caller is on hold is normally distributed with a mean of 3.1 minutes and a standard deviation 0.9 minutes. Company experts have decided that if as many as 5% of the callers are put on hold for 4.8 minutes or longer, more operators should be hired.

a. What proportion of the company's callers are put on hold for more than 4.8 minutes? Should the company hire more operators? Show these probabilities on a sketch of the normal curve.
Answer:
$P(X > 4.8) = P(Z > \frac{4.8-3.1}{0.9}) = P(Z > 1.89) = 1 - 0.9706 = 0.0294$.

b. At another cable company (length of time a caller is on hold follows the same distribution as before), 2.5% of the callers are put on hold for longer than x minutes. Find the value of x, and show this on a sketch of the normal curve.
Answer:
From the z table we find that $z = 1.96$. Therefore, $1.96 = \frac{x-3.1}{0.9} \Rightarrow x = 3.1+1.96(0.9) = 4.86$.

Problem 3
Answer the following questions:

a. Suppose that the height (X) in inches, of a 25-year-old man is a normal random variable with mean $\mu = 70$ inches. If $P(X > 79) = 0.025$ what is the standard deviation of this random normal variable?
Answer:
$1.96 = \frac{79-70}{\sigma} \Rightarrow \sigma = \frac{9}{1.96} = 4.59.$

b. Suppose that the weight (X) in pounds, of a 40-year-old man is a normal random variable with standard deviation $\sigma = 20$ pounds. If 5% of this population weigh less than 160 pounds what is the mean μ of this distribution?
Answer:
$-1.645 = \frac{160-\mu}{20} \Rightarrow \mu = 160 + 20(1.645) = 192.9.$

c. Find an interval that covers the middle 95% of $X \sim N(64, 8)$.
Answer:
We have 2.5% probability at each one of the two tails. Therefore
$-1.96 = \frac{x-64}{8} \Rightarrow x = 64 - 1.96(8) = 48.32.$
$1.96 = \frac{x-64}{8} \Rightarrow x = 64 + 1.96(8) = 79.68.$

Problem 4
A bag of cookies is underweight if it weighs less than 500 grams. The filling process dispenses cookies with weight that follows the normal distribution with mean 510 grams and standard deviation 4 grams.

a. What is the probability that a randomly selected bag is underweight?
Answer:
$P(X < 500) = P(Z < \frac{500-510}{4}) = P(Z < -2.5) = 0.0062.$

b. If you randomly select 5 bags, what is the probability that exactly 2 of them will be underweight?
Answer:
$P(Y = 2) = \binom{5}{2} 0.0062^2 (1 - 0.0062)^3 = 0.0004.$

Normal approximation to binomial

Suppose that X follows the binomial distribution with parameters n and p. We can approximate binomial probabilities using the normal distribution as follows:

- Calculate np and $n(1-p)$. If both are ≥ 5 continue.
- Compute $\mu = np$ and $\sigma = \sqrt{np(1-p)}$.
- Here is how you can approximate binomial probabilities:

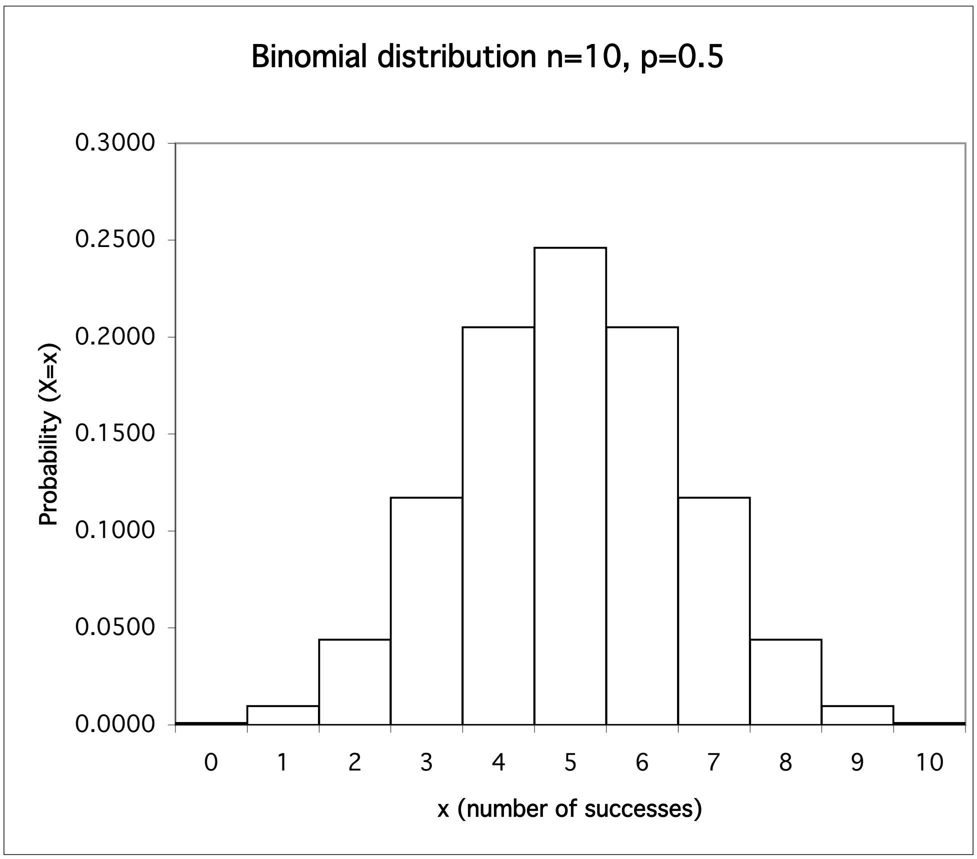
At least k successes	$P(X \geq k) = P(Z > \frac{k-0.5-\mu}{\sigma}) = \cdots$
More than k successes	$P(X > k) = P(Z > \frac{k+0.5-\mu}{\sigma}) = \cdots$
At most k successes	$P(X \leq k) = P(Z < \frac{k+0.5-\mu}{\sigma}) = \cdots$
Less than k successes	$P(X < k) = P(Z < \frac{k-0.5-\mu}{\sigma}) = \cdots$
Exactly k successes	$P(X = k) = P(\frac{k-0.5-\mu}{\sigma} < Z < \frac{k+0.5-\mu}{\sigma}) = \cdots$

- Some comments: The approximation is good if both np and $n(1-p)$ are ≥ 5. These 2 requirements hold if n is large, or if n is not very large but $p \approx 0.5$. The ± 0.5 in the formulas above is the so called continuity correction and we should use it for better approximation.
- Example: A coin is flipped 1000 times. Find the probability that in these 1000 tosses we obtain at least 530 heads?

- See figures below for the shape of the binomial distribution for different values of n and p.

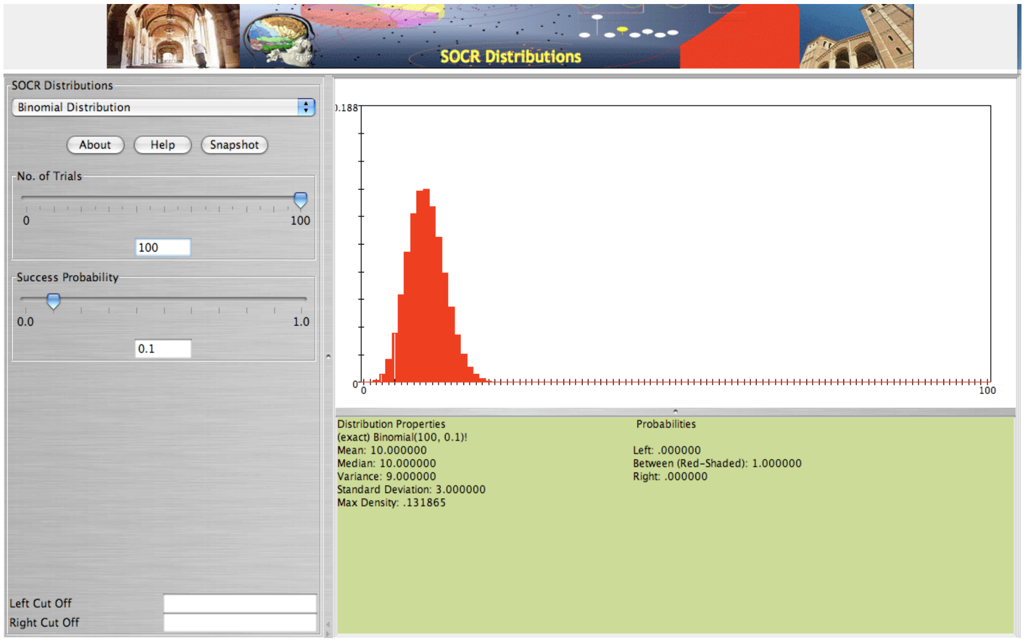
Normal approximation to binomial

x	P(x)
0	0.0010
1	0.0098
2	0.0439
3	0.1172
4	0.2051
5	0.2461
6	0.2051
7	0.1172
8	0.0439
9	0.0098
10	0.0010

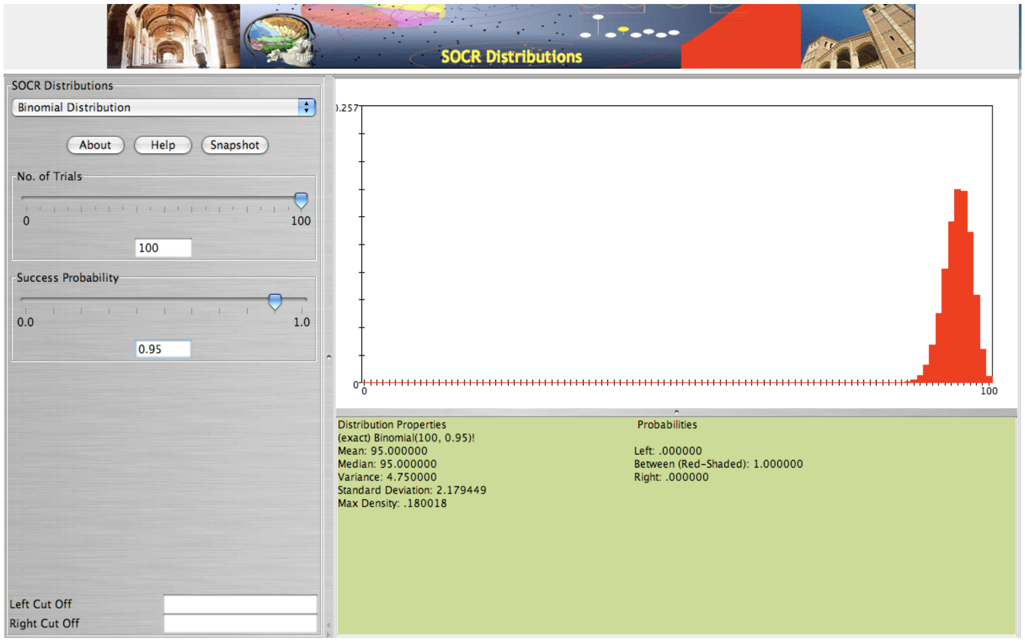


Examples:

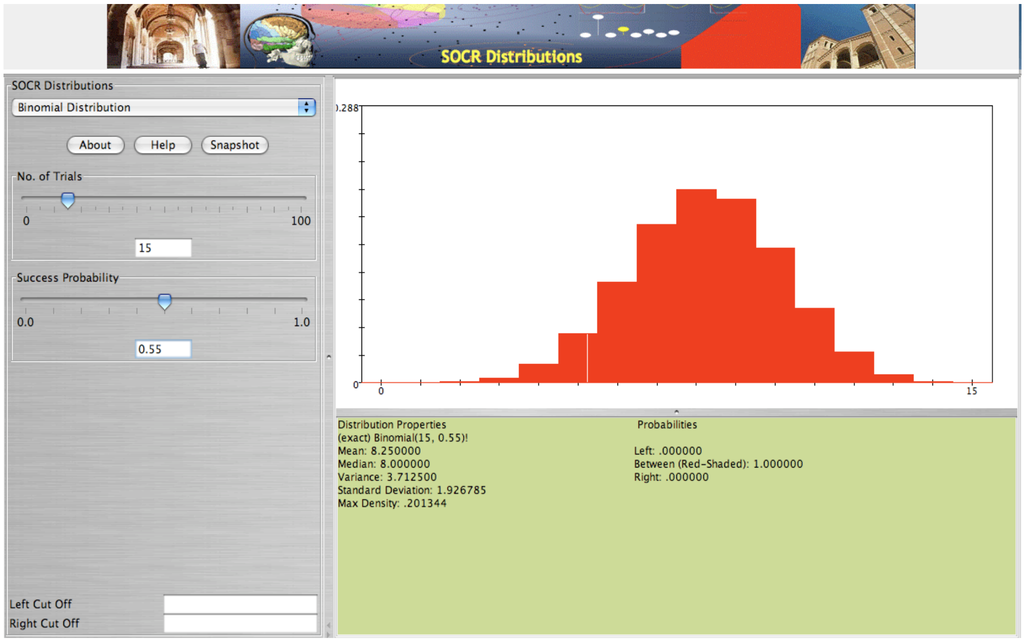
$X \sim b(100, 0.1)$



$X \sim b(100, 0.95)$



$X \sim b(15, 0.55)$



Example:
A manufacturing process produces semiconductor chips with a known failure rate 6.3%. Assume that chip failures are independent of one another. You will be producing 2000 chips tomorrow.

a. Find the expected number of defective chips produced.

b. Find the standard deviation of the number of defective chips.

c. Find the probability (approximate) that you will produce less than 135 defects.