UNIVERSITY OF MASSACHUSETTS
Department of Biostatistics and Epidemiology
BE540w - Introduction to Biostatistics
Fall 2004

Exercises with Solutions– Topic 5
Normal Distribution

Due: Monday November 15, 2004

READINGS

1. Text (Rosner B. Fundamentals of Biostatistics, 5th Edition) Chapter 5, pp 117-133.
2. Study Guide (Rosner B. Study Guide for Fundamentals of Biostatistics, 5th Edition) Chapter 5, pp 45-49.

EXERCISES:

1. Find the proportion of observations from a standard normal distribution that satisfies each of the following statements.

a. $Z < 2.85$
b. $Z > 2.85$
c. $Z > -1.66$
d. $-1.66 < Z < 2.85$
e. $Z < -2.25$
f. $Z > -2.25$
g. $Z > 1.77$
h. $-2.25 < Z < 1.77$

2. The height, X, of young American women is distributed normal with mean μ=65.5 and standard deviation σ=2.5 inches. Find the probability of each of the following events.

a. $X < 67$
b. $64 < X < 67$

3. Suppose the distribution of GRE scores satisfies the assumptions of normality with a mean score of μ=600 and a standard deviation of σ=80.

a. What is the probability of a score less than 450 or greater than 750?
b. What proportion of students have scores between 450 and 750?
c. What score is equal to the 95th percentile?

4. The Chapin Social Insight Test evaluates how accurately the subject appraises other people. In the reference population used to develop the test, scores is normally distributed with mean μ=25 and standard deviation σ=5. The range of possible scores is 0 to 41.

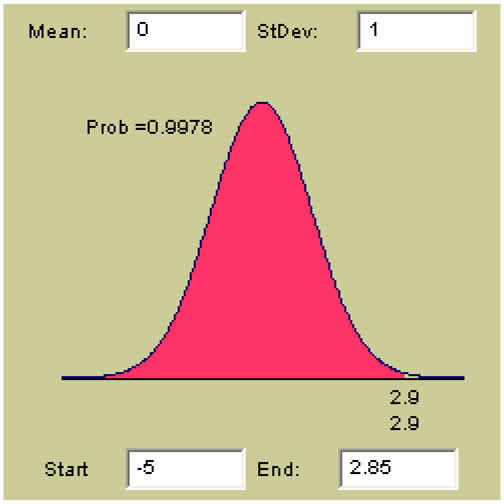
a. What proportion of the population has scores below 20 on the Chapin test?
b. What proportion has scores below 10?
c. How high a score must you have in order to be in the top quarter of the population in social insight?

5. A normal distribution has mean μ=100 and standard deviation σ=15 (for example, IQ). Give limits, symmetric about the mean, within which 95% of the population would lie:

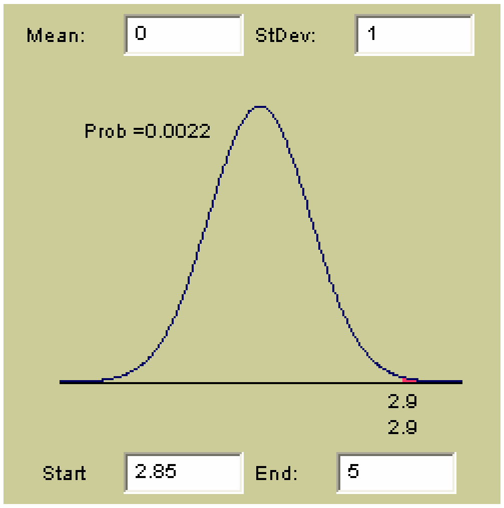
a. Individual observations.
b. Means of 4 observations.
c. Means of 16 observations.
d. Means of 100 observations.
e. Write down an expression for the width of the limits symmetric about the mean, within which 95% of the population of means of samples of size n would lie.

Solutions

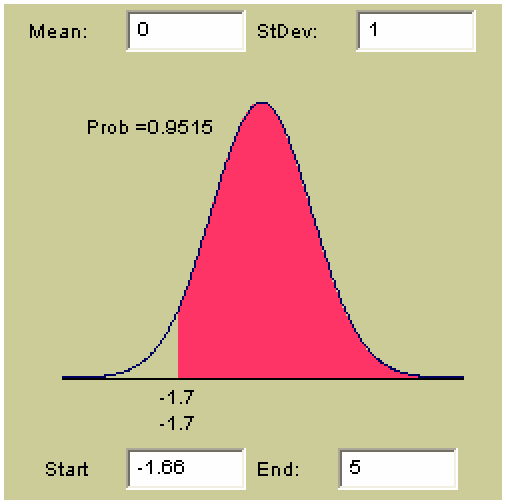
Notes – (1) To obtain the pictures that you see below, I used the link
http://psych.colorado.edu/%7Emcclella/java/normal/accurateNormal.html
(2) Since I couldn't enter $-\infty$ or $+\infty$ I replaced these entries with -5 or +5 as extremes

#1a. 

Pr (Z < 2.85) = .9978

#1b. 

Pr (Z > 2.85) = .0022



#1c.

Pr (Z > -1.66) = Pr (Z < +1.66)

= .9515

#1d.

Pr (-1.66 < Z < 2.85)

= Pr (Z < 2.85) - Pr (Z < -1.66)

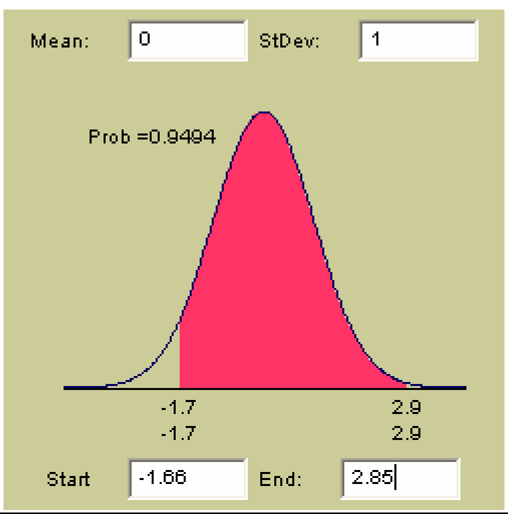
= Pr (Z < 2.85) - Pr (Z > +1.66)

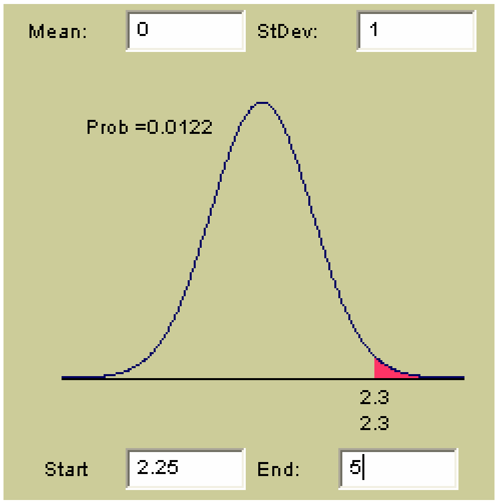
= Pr (Z < 2.85) - { 1 – Pr (Z < 1.66) }

= Pr (Z < 2.85) – 1 + Pr (Z < 1.66)

= .9978 - 1 + .9515

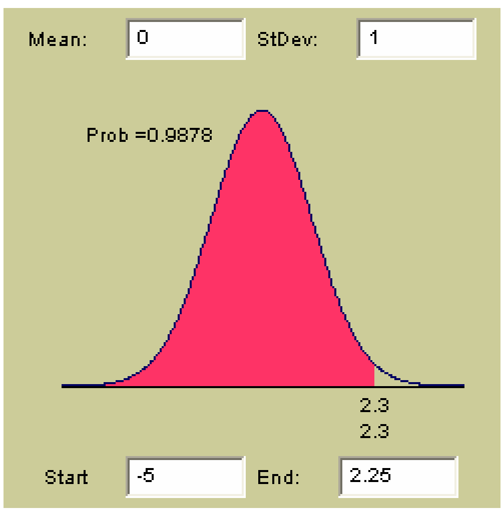
= .9493 which is pretty close to the applet solution on the web





#1e.

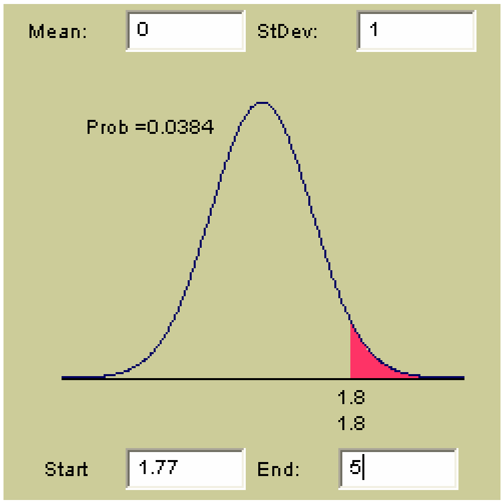
Pr (Z < -2.25)
= Pr (Z > + 2.25)
= .0122



#1f.

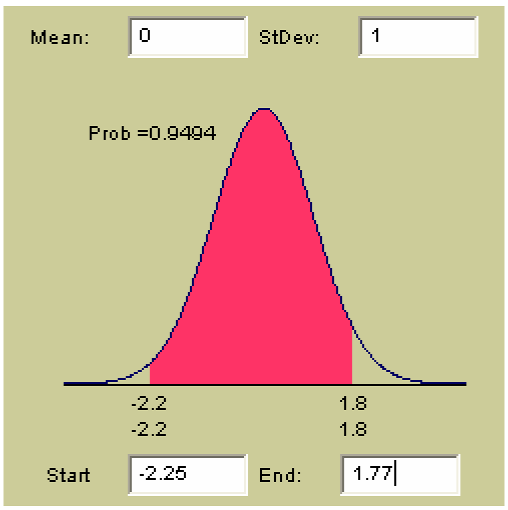
Pr (Z > -2.25)
= Pr (Z < +2.25)
= .9878

#1g.



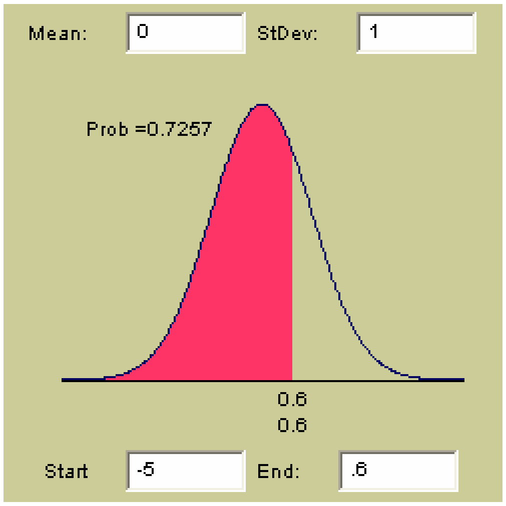
Pr (Z > 1.77)
= .0384

#1h.



Pr (-2.25 < Z < 1.77)
= Pr (Z < 1.77) - Pr (Z < -2.25)
= Pr (Z < 1.77) – Pr (Z > +2.25)
= .9616 - .0122
= .9494

#2a.



$$pr(X < 67) = pr\left[\left(\frac{X-\mu}{\sigma}\right) < \left(\frac{67-\mu}{\sigma}\right)\right]$$

$$= pr\left[Z < \left(\frac{67-65.5}{2.5}\right)\right]$$

$$= pr[Z < .6]$$

$$= .7257$$

#2b.

$$pr(64 < X < 67) = pr\left[\left(\frac{64-65.5}{2.5}\right) < Z < \left(\frac{67-65.5}{2.5}\right)\right]$$

$$= pr[-.6 < Z < +.6]$$

$$= .4515$$

#3a. Probability { score < 450 OR score > 750 }

$$= pr[X < 450] + pr[X > 750]$$

$$= pr\left[Z < \left(\frac{450-600}{80}\right)\right] + pr\left[Z > \left(\frac{750-600}{80}\right)\right]$$

$$= pr[Z < -1.875] + pr[Z > +1.875]$$

$$= 2\,pr[Z > +1.875]$$

$$= 2(.0304)$$

$$= .0608$$

#3b. Proportion of students with scores between 450 and 750

$$= pr[450 < X < 750]$$
$$= 1 - pr[x < 450 or X > 750]$$
$$= 1 - .0608$$
$$= .9392$$

#3c.

For Z~Normal(0,1)

$$pr[Z_{.95} < 1.645] = .95$$

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

$$1.645 = \frac{X_{.95} - 600}{80}$$

Thus, $X_{.95} = \sigma Z_{.95} + \mu$

$$= (80)[1.645] + 600$$
$$= 731.2$$

#4a.

$$pr(X < 20) = pr\left[Z < \frac{20 - 25}{5}\right] = pr[Z < -1] = .1587$$

#4b.

$$pr(X < 10) = pr[Z < \frac{10 - 25}{5}] = pr[Z < -3] = .0014$$

#4c.

$pr(Z > 0.67) = .25$ (I got this out of STATISTIX)

Thus, $X = \sigma Z + \mu$

$$= (5)[0.67] + 25$$
$$= 28.35$$

#5. First obtain an interval for Z~Normal(0,1)

$$pr\left[-1.96 < Z < +1.96\right] = .95$$

Next, recall that the standard error, SE, of $\overline{X}$, is related to σ via $se[\overline{X}] = \sigma/\sqrt{n}$

And the mean of $\overline{X}$ is $E[\overline{X}] = \mu$

Thus, the standardization formula can be manipulated to yield a formula for $\overline{X}$ in terms of Z.

From $Z = \dfrac{\overline{X} - E[\overline{X}]}{SE[\overline{X}]}$, solve for $\overline{X}$.

$$\overline{X} = \left\{se\left(\overline{X}\right)\right\}Z + \mu$$

$$= \left(\sigma/\sqrt{n}\right)Z + \mu$$

Now we can make a little table

	Mean	**SE**	**Lower limit**	**Upper limit**
Z	0	1	-1.96	+1.96
X	100	15	$(15)(-1.96)+100 = 70.6$	$(15)(+1.96)+100 = 129.4$
$\overline{X}_{n=4}$	100	15/2	$(15/2)(-1.96)+100 = 85.3$	$(15/2)(+1.96)+100 = 114.7$
$\overline{X}_{n=16}$	100	15/4	$(15/4)(-1.96)+100 = 92.65$	$(15/4)(+1.96)+100 = 107.35$
$\overline{X}_{n=100}$	100	15/10	$(15/10)(-1.96)+100 = 97.06$	$(15/10)(+1.96)+100 = 102.94$

#5e.

Width of confidence interval

= (Upper endpoint) - (Lower endpoint)

$$= \left[\frac{\sigma}{\sqrt{n}}(Z) + \mu\right] - \left[\frac{\sigma}{\sqrt{n}}(-Z) + \mu\right]$$

$$= \frac{2\sigma Z}{\sqrt{n}}$$

Thus, the width gets smaller as the sample size n gets larger.