MATH 3339
Statistics for the Sciences
Sec 5.5;6.5

Wendy Wang
wwang60@central.uh.edu

Lecture 11 - 3339

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Not 1, 2 or 3 Standard Deviations

$X \sim \exp(\lambda = 3)$

$P(X < 6.5) = \text{pexp}(6.5, 3)$

An orange juice producer buys all his oranges from a large orange grove. The amount of juice squeezed from each of these oranges is approximately normally distributed, with a mean of 4.70 ounces and a standard deviation of 0.40 ounce. The random variable is X = the amount of juice squeezed from one orange.

- What is the probability that an orange will have less than 4 ounces of juice? $P(X < 4)$.
  $= \text{pnorm}(\ )$ $X \sim N(\mu = 4.7, \sigma = 0.40)$
- There are a couple of ways to answer this question.
  - R: pnorm(x,mean,sd), pnorm(4,4.7,0.4) = 0.04005916
  - Z-Table: In your textbook. This table is for Standard Normal Distribution.

$P(X < 4) = P(X \leq 4) = \text{pnorm}(4, 4.7, 0.40)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Determine the probabilities?

X = amount of juice squeezed with $N(4.7, 0.4)$.

1. What is the probability that more than four ounces of juice will be squeezed?

$$P(X > 4) = 1 - P(X < 4) = 1 - \text{pnorm}(4, 4.7, 0.4)$$

2. What is the probability that between 3.5 and 4.5 ounces of juice will be squeezed?

$$P(3.5 < X < 4.5) = P(X < 4.5) - P(X < 3.5)$$
$$= \text{pnorm}(4.5, 4.7, 0.4) - \text{pnorm}(3.5, 4.7, 0.4)$$

3. What is the probability that beyond 3 or beyond 5 ounces of juice will be squeezed?

3 5

$$P(X < 3 \text{ or } X > 5) = P(X < 3) + P(X > 5)$$
$$= \text{pnorm}(3, 4.7, 0.4) + 1 - P(X < 5)$$

Standard Normal Distribution

$N(\mu = 0, \sigma = 1)$

To compute $P(a \leq X \leq b)$ when X is a Normal random variable with parameters μ and σ, we must evaluate:

$X \sim N(\mu, \sigma)$

$$\int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx$$

None of the standard integration techniques can be used to evaluate this, thus we "standardize" the values by:

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

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

where $\mu_Z = 0$ and $\sigma_Z = 1$ to get the pdf:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

The cdf of Z is $P(Z \leq z) = \int_{-\infty}^{z} \phi(y)dy$ which we denote by $\Phi(z)$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Key concepts for z-scores

$Z \sim N(0, 1)$

- The z-score is the number of standard deviations a value is from the mean.

$$z = \frac{x - \mu}{\sigma}, \quad \sigma > 0$$

- Z-scores have no units
- They measure the distance an observation is from the mean in standard deviations.
- Positive z-scores indicate that the observation is above the mean.
- Negative z-scores indicate that the observation is below the mean.
- Z-scores usually are between -3 and 3. Anything beyond these two values indicates that the observation is extreme.

$P(Z < -4) \approx 0$ $\qquad 99.7\% = P(-3 < Z < 3)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example of Z-scores

A certain town has a mean monthly high temperature in January of 35° F and a standard deviation of 8° F. This town also has a mean monthly high temperature in July of 75° F with a standard deviation of 10° F. In which month is it more unusual to have a day with a high temperature of 55 degrees?

Jan: $\mu = 35$ $\sigma = 8$ $z = \frac{55-35}{8} = 2.5$

July: $\mu = 75$ $\sigma = 10$ $z = \frac{55-75}{10} = -2$

$X_0 = 55$ is more unusual in Jan, because $|2.5| > |-2|$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Another Example

The score on a test has a mean of 75 with standard deviation 15. If I said your standard score (z-score) is 2.25, what is your actual test score?

$\mu = 75$ $\sigma = 15$

$z = 2.25$

$z = \frac{x - \mu}{\sigma}$

$x = ?$

$x = \mu + z \cdot \sigma$

$= 75 + 2.25 * 15$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Normal Distribution Calculations

- Area under a Normal curve represent proportions (probability) of observations within a range of values. There is no easy way to find the area under a Normal curve.
- We use a table or software that calculates the desired areas. The table we use is Z-table. It uses a **cumulative proportion**. A cumulative proportion is the proportion (probability) of observations in a distribution that lie at or below a given value. This is $\Phi(z)$.
- When the distribution is given by a density curve, the cumulative proportion is the area under the curve to the left of a given value.

pnorm( )

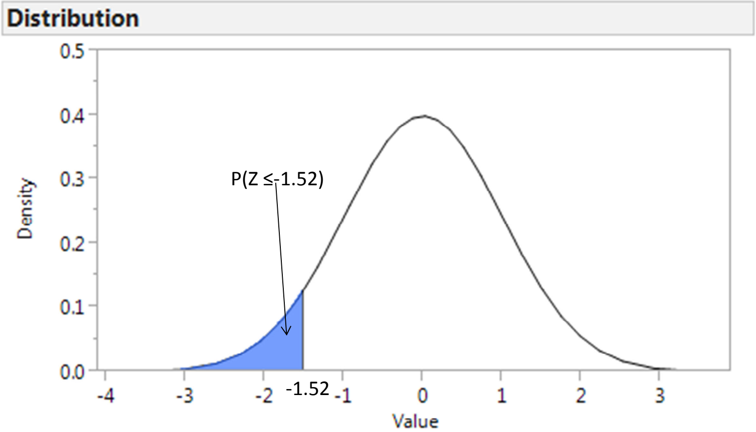
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Using The Z-table

https://www.math.uh.edu/~wwang/z-table.pdf

- The vertical margin are the left most digits of a z-score.
- The top margin is the hundredths place of a z-score.
- The numbers inside the table represents the area from $-\infty$ to that z-score.
- Remember that the standard Normal density curve is symmetric and the total area is equal to 1.
- *Note*: R can calculate these probabilities and also some calculators. Without having to convert to z-scores.

$P(Z \leq -1.52)$



Distribution
0.5
0.4
0.3
0.2
0.1
0.0
Density
P(Z ≤-1.52)
-4
-3
-2
-1.52
-1
0
1
2
3
Value

$P(Z \leq -1.52) = 0.0643$

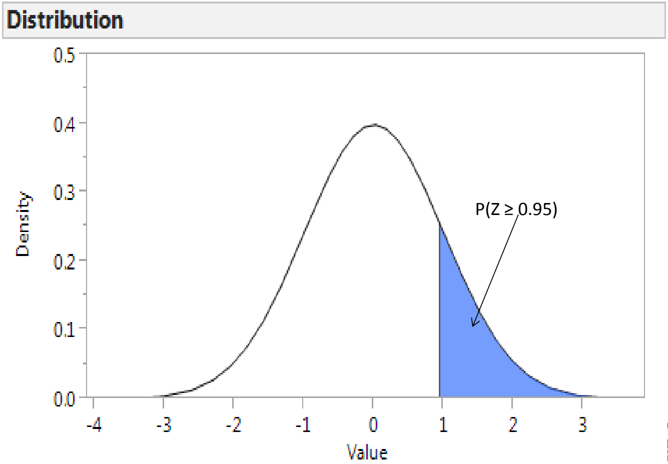
R: pnorm(-1.52,0,1) = 0.06425549
or, pnorm(-1.52)=0.06425549

Table A: P(Z < z)

z	0.00	0.01	0.02	0.03
-3.4	0.0003	0.0003	0.0003	0.0003
-3.3	0.0005	0.0005	0.0005	0.0004
-3.2	0.0007	0.0007	0.0006	0.0006
⋮	⋮	⋮	⋮	⋮
-1.5	0.0668	0.0655	0.0643	0.0630

P(Z ≤ -1.52)

$P(Z \geq 0.95)$



$P(Z \geq 0.95) = 0.1711$

R: 1 - pnorm(0.95,0,1) = 0.1710561

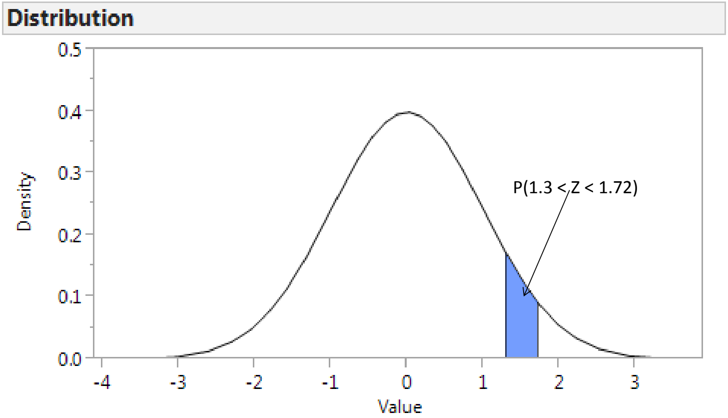
Table A: P(Z < z)

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289

$P(Z \geq 0.95) = 1 - P(Z < 0.95) = 1 - 0.8289 = 0.1711$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

$P(1.3 < Z < 1.72)$



Distribution
0.5
0.4
0.3
0.2
0.1
0.0
Density
P(1.3 < Z < 1.72)
-4
-3
-2
-1
0
1
2
3
Value

$P(1.3 < Z < 1.72) = 0.0541$

R: pnorm(1.72,0,1) - pnorm(1.3,0,1) = 0.05408426

z	0.00	0.01	0.02	0.03
0.0	0.5000	0.5040	0.5080	0.5120
0.1	⋮	⋮	⋮	⋮
1.3	0.9032	0.9049	0.9066	0.9082
1.4	⋮	⋮	⋮	⋮
1.7	0.9554	0.9564	0.9573	0.9582

P(1.3 < Z < 1.72) = 0.9573 – 0.9032 = 0.0541

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Inverse Normal

Inverse Normal: how to find a value, when we are given a proportion

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Finding a value when given a proportion

- Called inverse Normal.
- This is working “Backwards” using Z-Table.
- Finding the observed values when given a percent.
- In R: qnorm(proportion,mean,sd).

given a range of value, to find prob, pnorm( )

given a prob/proportion, to find obs. , qnorm( )

"Backward" Normal calculations Using Z-Table

1. State the problem. Since, Z-Table, qnorm and invNorm gives the areas to the left of z-scores, always state the problem in terms of the area to the left of x. Keep in mind that the total area under the standard Normal curve is 1.

2. Use Table A to find c. This is the value from the table not a value that we calculate.

3. Unstandardized to transform the solution from the z-score back to the original x scale. Solving for x using the equation

$$c = \frac{x - \mu}{\sigma}$$

gives the equation $x = \sigma(c) + \mu$.

Examples to Work "Backwards" with the Normal Distribution

Find the value of *c* so that:

$Z \sim N(0, 1)$

1. $P(Z < c) = 0.7704$

$c = \text{qnorm}(\underset{0.7704}{?}, 0, 1)$

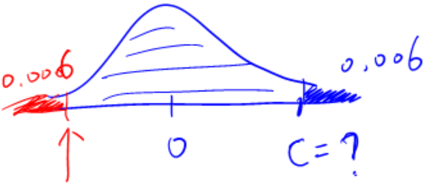
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Examples to Work "Backwards" with the Normal Distribution

-qnorm(0.006, 0, 1)

Find the value of c so that:

2. $P(Z > c) = 0.006$



qnorm(0.006, 0, 1)

qnorm(?, 0, 1)

1 - 0.006

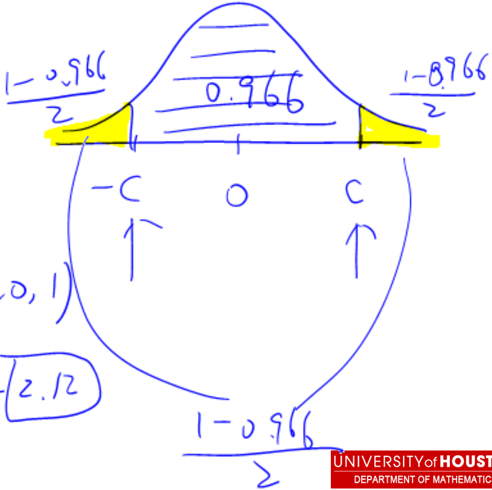
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Examples to Work "Backwards" with the Normal Distribution

$Z \sim N(0, 1)$

Find the value of c so that:

3. $P(-c < Z < c) = \underline{0.966}$



$P(X < -c) = \frac{1-0.966}{2}$

$-c = \text{qnorm}\left(\frac{1-0.966}{2}, 0, 1\right)$

$= -2.12 \Rightarrow c = \boxed{2.12}$

```
> qnorm((1-0.966)/2,0,1)
[1] -2.120072
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

- given $\underline{X < a}$, to find prob.

$P(X<a) = \text{pnorm}(a, \mu, \sigma)$

- given prob. or proportion, to find X.

$P(X<b) = \underline{0.3}$, to find $\underline{b}$.

$b = \text{qnorm}(\underset{0.3}{\boxed{?}}, \mu, \sigma)$

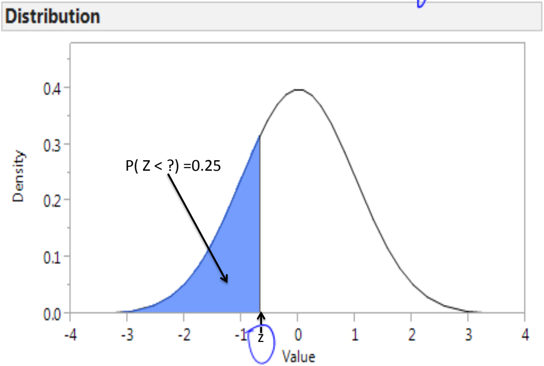
MPG for Prius

$X \sim N(\mu = 49, \sigma = 3.5)$

The miles per gallon for a Toyota Prius has a Normal distribution with mean $\mu = 49$ mpg and standard deviation $\sigma = 3.5$ mpg. 25% of the Prius have a MPG of what value and lower? $P(X < ?) = 0.25$

1. We want c, such that $P(Z < c) = 0.25$. That is we want to know what z-score cuts off the lowest 25%.

$z = \text{qnorm}(0.25, 0, 1)$



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Find c such that $P(Z < c) = 0.25$

3. From Table A, find something close to 0.25 **inside** the table.

z	0.00	0.01	0.02	0.07	0.08	0.09
-3.4	0.0003	0.0003	...	0.0003	0.0003	0.0002
⋮	⋮	⋮	⋮	⋮	⋮	⋮
-0.7	0.2420	0.2389	...	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	...	0.2514	0.2483	0.2451

P(Z < ?) = 0.25 (closes value is 0.2514)

z = -0.67 (-0.6 “row” + 0.07 “column”)

$P(Z < -0.67) = 0.2514 \approx 0.25$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Find c such that $P(Z < c) = 0.25$

$z = \frac{x-\mu}{\sigma}$

$X = \mu + z \cdot \sigma$

$= 49 + (-0.67) * 3.5$

an approx.

4. Unstandardized. $x = \sigma(c) + \mu = 3.5(-0.67) + 49 = 46.655$
5. This means that 25% of the Prius has a mpg of less than 46.655 mpg.

Using R: qnorm(0.25,49,3.5) = 46.63929

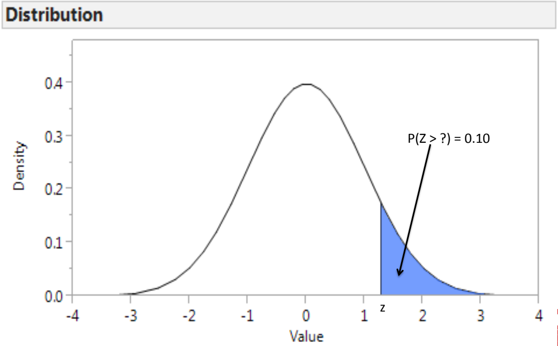
qnorm (0.25 , 49 , 3.5)

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Top 10%

Suppose you rank in the 10% of your class. If the mean GPA is 2.7 and the standard deviation is 0.59, what is your GPA? (Assume a Normal distribution)

1. We want c, such that $P(Z > c) = 0.10$. That is we want to know what z-score cuts off the highest 10%.



Find c such that $P(Z > c) = 0.1$

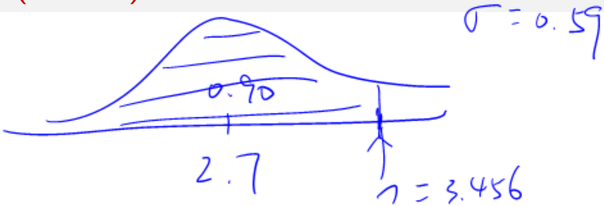
3. From Table A, the areas are **below** or to the left of a z-score thus we want to find something close to 0.90 **inside** the table.

z	0.00	0.01	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5675	0.5714	0.5753
		⋮	⋮	⋮	⋮
0.2	0.5793				
1.2	0.8849	0.8869	0.8980	0.8997	0.9015

P(Z < ?) = 0.90 (close value is 0.8997)

z = 1.28 (1.2 “row” + 0.08 “column”)

Find c such that $P(Z > c) = 0.1$



4. Unstandardized: $x = \sigma(c) + \mu = 0.59(1.28) + 2.7 = 3.4552$

5. This means that your gpa is 3.4375 if you rank at the 10% of your class.

In R: qnorm(0.9,2.7,0.59) = 3.456115

qnorm (0.90 , 2.7 , 0.59) = 3.456

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

Let X = amount of juice in ounces in a orange, $X \sim N(4.7, 0.4)$.

4.7

1. Determine the third quartile.

Q_3

$$x_0 = Q_3 = \text{qnorm}(0.75, 4.7, 0.4)$$

2. Determine the 95th percentile.

$$x_{95^{th}} = \text{qnorm}(0.95, 4.7, 0.4)$$

Recall $E(X + Y)$

- If X and Y are two different random variables, then the expected value (mean) of the sums of the pairs of the random variable is the same as the sum of their means:

  $$\mu_{X+Y} = E(X + Y) = E(X) + E(Y) = \mu_X + \mu_Y.$$

  This is called the addition rule for means.

- The expected value (mean) of the difference of the pairs of the random variable is the same as the difference of their means:

  $$\mu_{X-Y} = E(X - Y) = E(X) - E(Y) = \mu_X - \mu_y.$$

Recall $VAR(X+Y)$

If X and Y are independent random variables

$$\sigma^2_{X+Y} = \underline{Var(X+Y)} = Var(X) + Var(Y) = \underline{\sigma^2_X + \sigma^2_Y}$$

and

$$\sigma^2_{X-Y} = Var(X-Y) = Var(X) + Var(Y) = \sigma^2_X + \sigma^2_Y$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

If X & Y are dependent

If X and Y are dependent random variables then

$$\sigma^2_{X+Y} = Var(X+Y) = Var(X) + Var(Y) + 2cov(X, Y) = \sigma^2_X + \sigma^2_Y + 2cov(X, Y)$$

and

$$\sigma^2_{X-Y} = Var(X-Y) = Var(X) + Var(Y) - 2cov(X, Y) = \sigma^2_X + \sigma^2_Y - 2cov(X, Y)$$

if X and Y are independent,

cov(X, Y) = 0

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

Suppose we have two independent random variables, X and Y where $\mu_X = 10$, $\sigma_X = 2$, $\mu_Y = 10$ and $\sigma_Y = 2$.

a. Determine: μ_{X+Y} and σ_{X+Y}

$$\mu_{X+Y} = E(X+Y) = E(X) + E(Y)$$
$$= \mu_X + \mu_Y$$
$$= 10 + 10 = 20$$

b. Suppose we want the mean of X and Y, what would be the expected value of the mean?

$$\sigma_{(X+Y)} = \sqrt{Var(X+Y)} = \sqrt{Var(X) + Var(Y)}$$
$$= \sqrt{2^2 + 2^2} = \sqrt{8}$$

given: X and Y are independent.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS