

## One sided tests:

So what is a one sided test?

In our original set-up, we formulated a hypothesis such as:

$$H_0: \mu = 72$$

(your text uses  $H_0: \mu \leq 72$  or  $H_0: \mu \geq 72$ . There's nothing wrong with that, but most people don't do it that way - just make sure  $H_0$  is opposite of  $H_1$  if you do want to do it this way)

our alternative was:

$$H_1: \mu \neq 72$$

Now we want to look at the following alternative hypotheses:

$$H_1: \mu > 72 \quad \text{or} \quad H_1: \mu < 72$$

2) The last two alternatives give rise to what is known as a one-sided test.

a) not all tests that you will see have one-sided alternatives (an example is ANOVA).

b) So what happens in a one sided test?

i) we REJECT if we conclude  $\mu > 72$  or, of course  $\mu < 72$ .

ii) we do *not* test both alternatives at once!

3) How do you decide if you need to do a one sided test?

- **based on your subject matter knowledge!** Do you know something about the situation that would lead you to believe that a one sided test would do better?

- for example - you test a new medication to lower blood pressure.

- what outcome are you interested in?

- that the medicine LOWERS blood pressure. If  $\mu_1 =$  "average blood pressure before medication", then you're interested in  $H_1: \mu_1 > \mu_2$ .

- note that this example is for a two sample test (more on that soon).

- you want to show that the medication lowers blood pressure; you're not interested if it somehow raises blood pressure.

- you make your decision based on what you know about the problem.

- NEVER, NEVER decide to do a one sided test by looking at the results and then making your decision (e.g., comparing  $\bar{y}_1$  with  $\bar{y}_2$ ). That's cheating!

4) Here's an outline:

a) Set up our hypotheses:

i)  $H_0: \mu = 72$

ii)  $H_1: \mu > 72$  or  $H_1: \mu < 72$  (not both)

b) decide on  $\alpha$ .

c) verify that your results agree with your alternative hypothesis

- if your alternative hypothesis is  $H_1: \mu_1 < \mu_2$ , make sure  $\bar{y}_1 \leq \bar{y}_2$  (or vice-versa)

- if this is not true, STOP - don't do anything else. Don't calculate anything else.

*Your results contradict your alternative hypothesis!*

d) calculate  $t^*$  in the usual way

(don't don't forget your assumptions, etc.)

make your comparison in the usual way:

if  $|t^*| \geq t_{\text{table}}$  then reject (otherwise “fail to reject”).

*make sure that you now use the 2<sup>nd</sup> row of your t-tables to select  $\alpha$*

or (as usual):

if  $p\text{-value} \leq \alpha$  then reject (otherwise “fail to reject”)

(note that R will give you one sided values by default; you need to either explicitly tell R to use both sides or simply multiply the result by 2)

e) note that this only works if you do part (c) first; otherwise this can lead you to make a mistake.

5) An example:

a) Example 7.3, p. 104

i) We're investigating the weight loss in 12 people after giving them a drug for weight loss. Data are the change in weight after taking the drug:

ii) what are you interested in?

- that they gained weight? No, that's silly

- that they lost weight - that makes sense.

b) So form your hypotheses:

i)  $H_0: \mu = 0$  (i.e., no change in weight)

ii)  $H_1: \mu < 0$  (i.e., the average weight change was negative (they lost weight))

c) pick  $\alpha = .05$  (book picks  $\alpha$  for us).

d) verify that our results agree with the alternative hypothesis;

- yes,  $\bar{y} < 0.00$ , so we can continue

d) calculate  $t^*$ :

$$t^* = \frac{\bar{y} - \mu}{\frac{s_y}{\sqrt{n}}} = \frac{-0.61}{\frac{.6331}{\sqrt{12}}} = -3.389$$

f) now we look up  $t$  with 11 *d.f.* and  $\alpha = .05$  and get 1.796.

g) Our comparison:

$|t^*| = 3.389$  is greater than or equal to 1.796,

so we reject  $H_0$ .

We could also use R to get our p-value:

`pt(-3.389, 11)`

and we get 0.003022778 (which is less than  $\alpha$ )

If you're using R, make sure you specify “lower.tail = FALSE” if you need it (say, if our  $t^*$  had been positive):

`pt(3.389, 11, lower.tail = FALSE)`

h) We conclude that we do have evidence to show the weight loss drug is working.

6) Some comments:

a) note that your  $p$ -values are now half of what they were before.

Just examine the t-tables to see this and note the top two rows.

Why? because you don't have to put half of your probability in the other tail!! This is good, because you have a much better chance of rejecting  $H_0$  (you get more power).

(to be illustrated on board)

b) remember: if you look at your data after performing the test and find that  $\bar{y}_1 \gg \bar{y}_2$ , you just can't go back and test for  $\mu_1 > \mu_2$ . You'd be cheating!!

(this is a two sample test example)

**c) if you decide to do a one-tailed test, you must decide to do so before looking (snooping) through your data. *You should have a reason for choosing a one-tailed test before you even perform the experiment.***

**Assumptions:**

All hypothesis tests rely on assumptions.

If these assumptions are not met, then the test may be invalid.

(sometimes it depends on how badly you violate the assumptions).

An assumption that all statistical tests share:

**The data are random!**

If the data are not random, then everything you're doing is potentially totally meaningless.

As an extreme example, suppose you want to prove that the average height of men is 7 feet.

You go out and collect the heights of 25 basketball players and merrily conclude that you're right.

Obviously this is wrong!

Note that there isn't an easy way to fix non-random data; often this is impossible.

You should, nevertheless, verify that the data are in somehow random (examine the methods used, etc.)

Most tests can be divided into two groups:

Parametric and non-parametric (some people prefer “distribution free”)

Parametric tests assume a particular distribution (usually normal) for the data.

Non-parametric tests don't assume any kind of distribution for the data.

There are other tests as well (e.g. permutation tests), but we won't discuss these.

We will see examples of both soon. But we'll start with parametric tests.

A very broad generalization (not always true):

In general, if the data follow the distribution you're assuming for the test, then the parametric test will be more powerful.

If not, then the non-parametric test will do better.

A common assumption of parametric tests:

**Data (in each sample) are distributed normally.**

There are many other assumptions, but many of those vary from test to test, so we'll cover those with the specific test.

The normal distribution assumption and how to check for normal distributions:

There are several ways of trying to figure out if your data have a normal distribution.

Hypothesis tests:

Graphical methods:

Some plots like histograms and boxplots will very quickly tell you if your data are seriously not normal. But they don't do well for data that may superficially appear to be normal.

Much preferred is the q-q plot or normal probability plot (not quite the same thing, but close enough).

A simplified explanation: a normal probability plot is a special kind of q-q plot.

Your text doesn't do a good job explaining this, so we'll use the method from my introductory class (and illustrate this with R).

In general, we'll let the computer do this. What the computer does is roughly as follows:

It takes your data, figures out your sample size, and then calculates what you should “expect” from a sample of that size.

It then plots these “expected” values against what you actually got. If this turns out to be a straight line, then things are probably normal.

For example, if you took many, many samples of size 12, where would you expect the smallest value to be? The second smallest? etc.

This is easy to see for the median, but you can do all of your values in a similar way.

Let's do an example using the data from example 6.3, p. 74.

These represent the systolic pressures of 12 chimpanzees.

If systolic blood pressure has a normal distribution, where do we expect the lowest blood pressure to be?

There are several ways to do this. We'll do things the same way that R does (**caution:** this is a bit different from what some of you learned in my introductory class):

$$Z_{area} \text{ where } area = \frac{i - \frac{1}{2}}{n}$$

What we calculate is the area to the left for the  $i^{\text{th}}$  value. So for the 1<sup>st</sup> value (the lowest of our 12 blood pressures), we would expect the area to the left of the first or lowest value to be:

$$area = \frac{1 - \frac{1}{2}}{12} = 0.04167$$

Now we convert this into a normal score (i.e., we get the  $z$  value) by looking up this area (using reverse lookup) in our table:

if  $area = 0.04167$ , our  $z$ -score would be:

( we take advantage of symmetry and look up 0.04167 in our table)

$$z = 1.73$$

However, since we want the  $z$ -score for the area to the left of 0.04167 (in the lower tail), we need the negative of that and so we get:

$$z = -1.73$$

Very quickly, let's try the 2<sup>nd</sup> value:

$$area = \frac{2 - \frac{1}{2}}{12} = 0.125$$

And getting the  $z$ -score for this we get:

$$z = -1.15$$

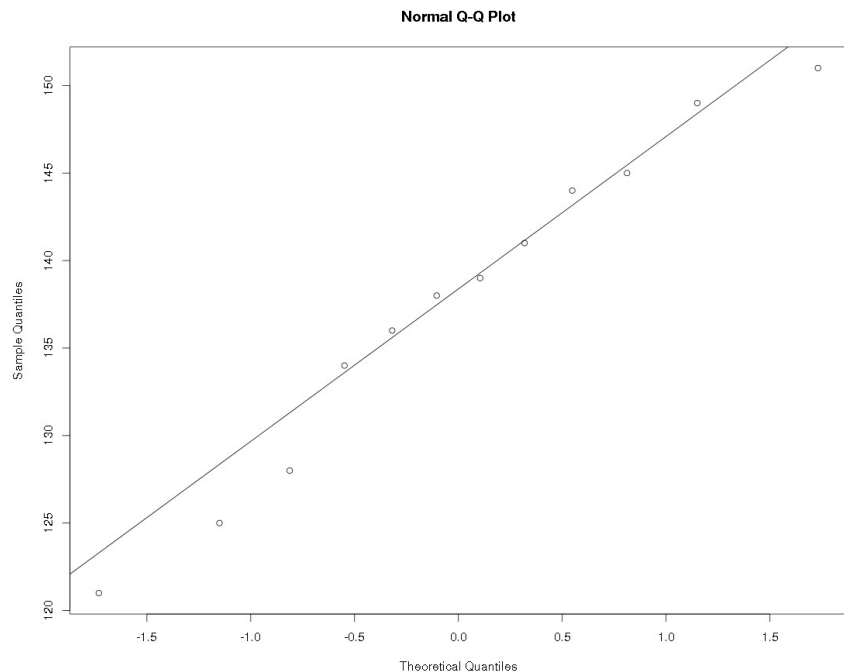
You would keep going until you've calculated the  $z$ -scores for all 12 values (here they are, directly from R):

-1.73 -1.15 -0.81 -0.55 -0.32 -0.10 0.10 0.32 0.55  
0.81 1.15 1.73

(Incidentally, when your area goes over 0.5, remember to subtract your area from 1 before using our table).

When you're all done, you would plot these  $z$ -scores vs. the actual data values (see the graph).

Anyway, you hardly ever do this by hand. You have the computer graph your samples, and then look at them:



Some notes on R:

Here's how to do the above example in R:

```
bp <- c(121, 125, 128, 134, 136, 138, 139, 141, 144, 145, 149, 151)
qqnorm(bp)
qqline(bp)
```

(if you want the actual normal scores, do):

```
bb <- qqnorm(bp)
bb$x
```

Incidentally, we never explained what we're looking for in this plot.

We're looking for our data to line up closely with the line:

- 1) Our data should be in a straight line.
- 2) Our data shouldn't be too far off the drawn line.

What is sufficiently straight? That takes some practice. Some basics:

Don't worry about every little bump. There'll be lots of bumps.

Worry if:

You see a strong overall curve

If you see a backwards "S" (if the ends curve towards the vertical) - this indicates long tails which is BAD. If you see a regular "S" that indicates short tails, that's usually okay, even if it isn't quite normal.

Some people (& some software (e.g., Minitab) like to reverse the axes, so always make sure you look at the labels of the axes before you interpret a probability plot.

If the axes are reversed, then the above comments need to be "reversed" as well (a regular "S" becomes long tailed).

### **Fixing problems with our normal distribution assumptions:**

It is often possible to change our data and "make" it normal.

What we do is we "transform" our data so it becomes normal and then use the "transformed" data in our analysis.

There are disadvantages to this, since interpretation becomes difficult:

If you show that blood pressure is not equal to 140mm using a transform, that's not actually what you're doing.

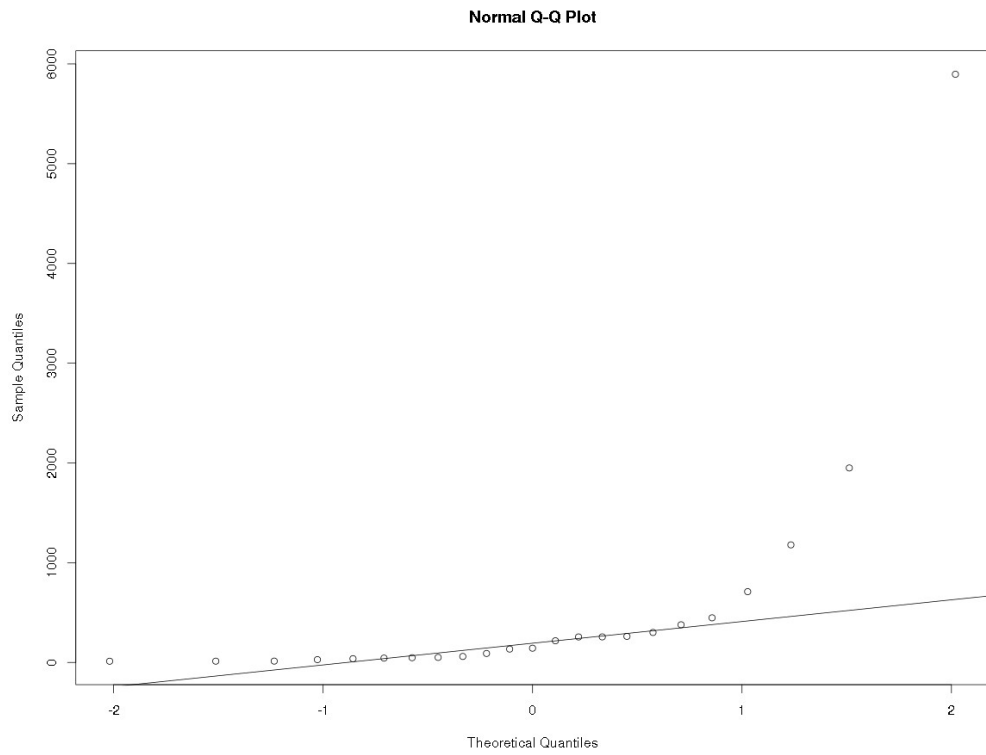
You're showing that the transformed data are not equal to the transformed value (the "140mm" above also needs to be transformed).

Before we get confused, let's do a simple example using a very common transformation (see chapter 13 in your text):

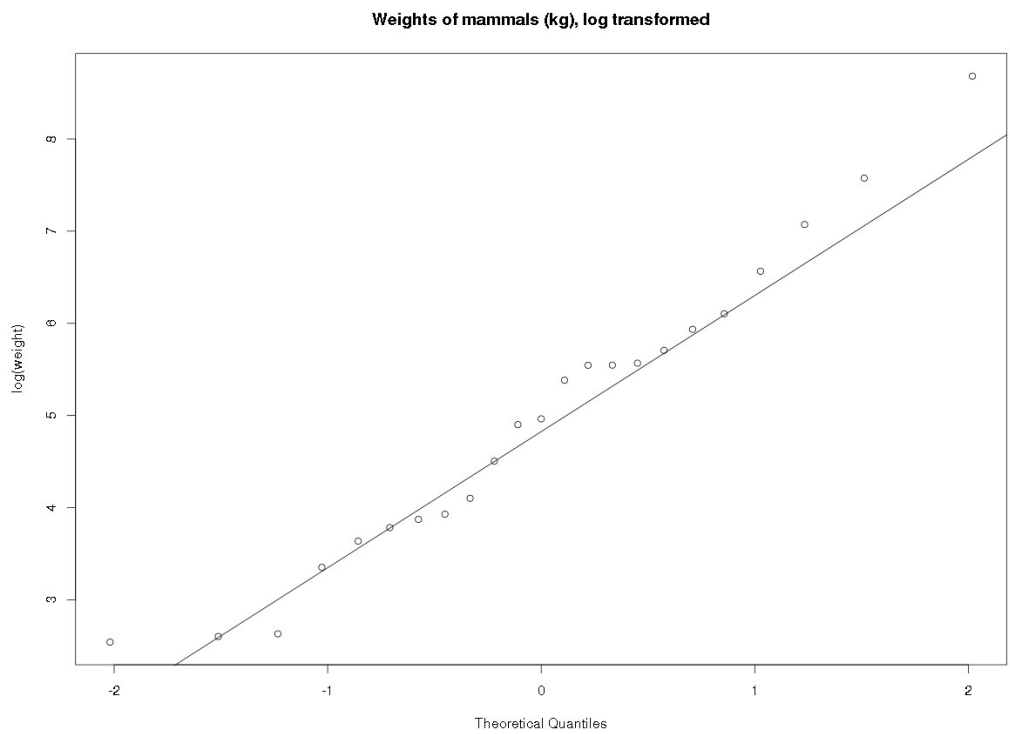
Let's use the weights of some mammals in kg (also used in a lab in 312):

Coyote	12.7
Clouded leopard	13.5
Duiker	13.9
Yellow baboon	28.6
Cheetah	38.0
Cougar	44.0
Wolf	48.1
Bushbuck	50.9
Impala	60.5
Warthog	90.5
Nyala	134.5
Lion	143.2
Black bear	217.7
Grizzly bear	256.1
Blue wildebeest	256.5
Cape Mountain Zebra	262
Kudu	301
Burchell's Zebra	378
Polar bear	447.7
Giraffe	710
Bison	1179
Hippo	1950
Elephant	5896.7

If we plot these data using a qq plot, we get the following graph:



The data are obviously not normal, so let's try to transform the data using a log transform and then plot the data again:



This is much better!

There are many (many) transformations. Some of the most popular ones include:

- |                     |  |
|---------------------|--|
| log                 | - useful when there is a large disparity in values (like above).<br>- also useful when variance is non-constant (you'll have to wait for the explanation of this one until later). |
| square root         | - also useful when the variance is non-constant.   |
| arcsine-square root | - useful if your data are proportions or percentages   |
| Box-Cox             | - the most versatile, because it can “automatically” select a good distribution for you. If you need it, look it up; it's beyond what we can do in here.                           |

Often we can transform data that are not normal.

Just remember - we are now working with transformed data, not the original data, and interpreting your results **MUST** take this into considerations:

Another example: you transform your data to logs. You show that the log(blood pressure) of people on medication is lower than the log(blood pressure) of people not on medication.

What, exactly, does this mean?

It does NOT mean blood pressure for people on medication is lower than for people not on medication.

A better alternative (frequently) is to use non-parametric tests (if available or possible).