

Data Analysis and Statistical Methods

Statistics 651

<http://www.stat.tamu.edu/~suhasini/teaching.html>

Lecture 15 (MWF) The principles of statistical testing

Suhasini Subba Rao

The underlying idea of statistical testing

- So far we have considered confidence intervals, which is one type of statistical inference.
- The second type of inference is statistical testing. In this case we have a research hypothesis which is proposed by the person conducting the study. We have come across this concept before. Specifically:
 - In Lecture 7, Example 2, where we assessed the validity of Jack's claim that he guessed the answers in this exams.
 - In Lecture 12, Example 5, where the aim was to determine if the mean weight of new born calves was over 91 pounds.
- We revisit these examples.

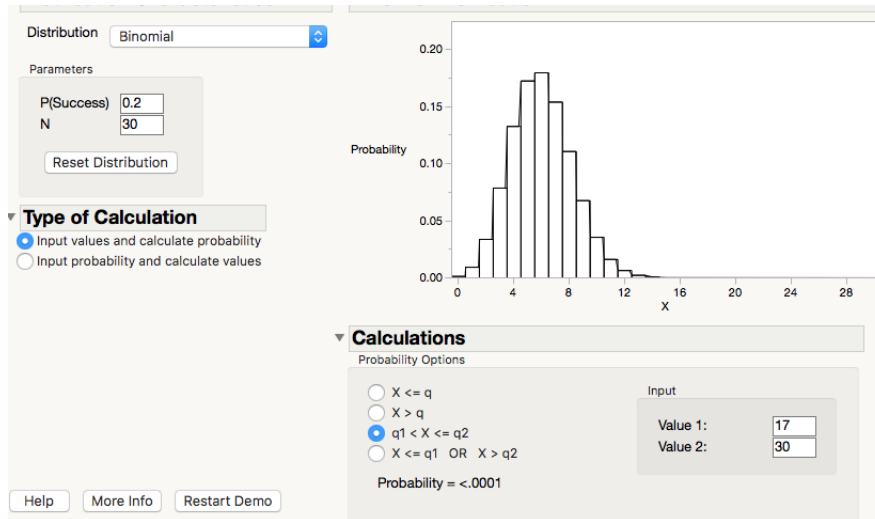
Example 1: Jack's exams

We recall that Jack did some multiple choice exams (where in each exam there were 5 choices) and Jack claimed that he guessed. We want to assess the validity of his claim. Let p denote the probability he gets a question correct. Then

$$\underbrace{H_0 : p \leq 0.2}_{\text{claim}} \quad \underbrace{H_A : p > 0.2}_{\text{knew something}} .$$

As evidence, we observe this scores in a multiple choice exam, where Jack scored 18 points out of 30. We calculate the chance of scoring 18 points or more out of 30 under the null hypothesis of his guess ($p = 0.2$).

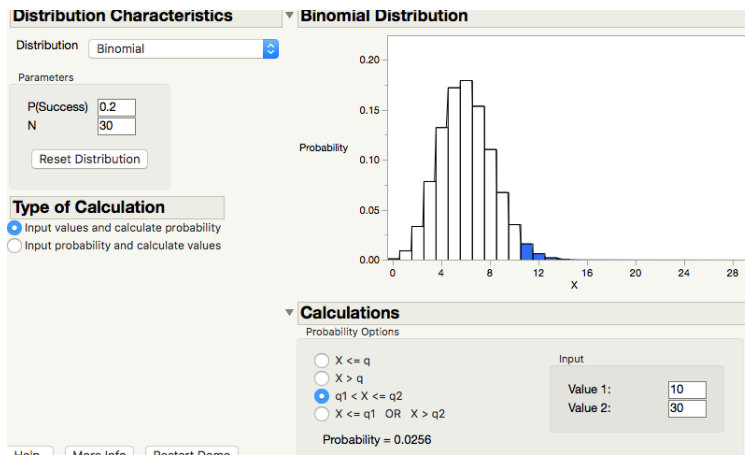
Lecture 15 (MWF) Introduction to statistical testing (testing on the mean)



From the graphic above we observe that this probability is less than 0.0001 (0.01%). This is so small that we believe the null is implausible. In other words, **we reject the null and accept the alternative**. We paraphrase this as saying there is evidence in the data to suggest that the $p > 0.2$ (he did not guess).

How small is small?

- Traditionally, we use 5% as the threshold. And we say the null becomes implausible if the probability (p-value) is less than 5%.



The blue area is called the critical region (about 5%). The white area is the non-rejection region. A score in the white region means the null is plausible (at the 5% level). A score in the blue area means the null is not plausible under the null.

Look carefully at the plot. If the null were true the mean score would be $0.2 \times 30 = 6$.

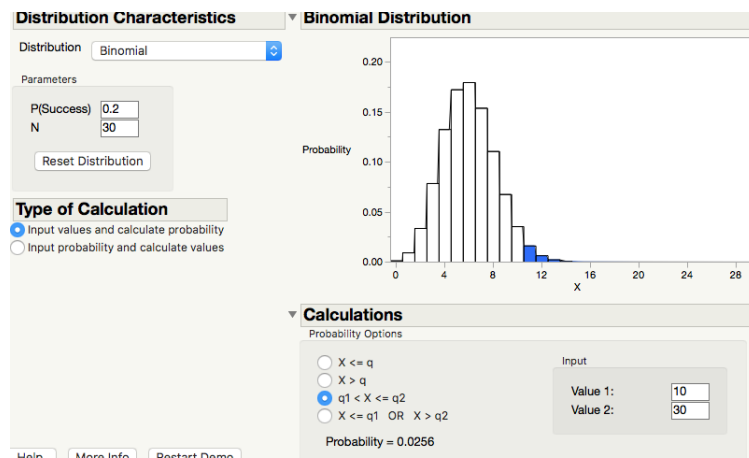
Summary at the 5% level

If we test $H_0 : p \leq 0.2$ against $H_A : p > 0.2$.

- The white region is the **non-rejection region**. We **do not reject** the null if the score is in the non-rejection region.
- The blue region is the **critical region**. We **reject the null** if the score is in the critical region.
- For this hypothesis, the p-value is the area equal to and to the **right** of the score. If the score is 11 or more, than the p-value will be less than 5% (less 3%). The larger the score, the smaller the p-value. The more implausible the null.

An alternative scenario

- If Jack were to score 9 out of 30 in this exam. His score is in the white, non-rejection region. With p-value 12.87%. This tells us that his score is plausible under the null. However, we cannot say that the null is true, only that it is plausible.



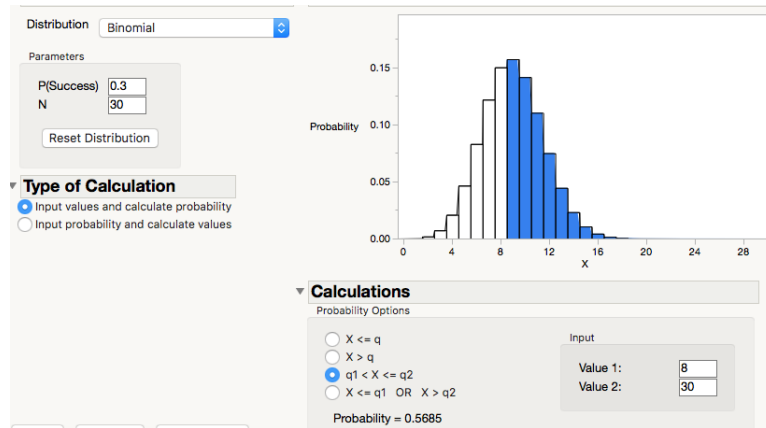
- If we want to place a p-value to this test, the p-value is 12.87%.

Tests and confidence intervals

- Tests and confidence intervals are related but they convey different bits of information.
- A tests says whether a parameter value is plausible or not.
Bear in mind plausible does not mean it **it is that value**.
- Whereas a confidence interval gives an interval of plausible values.
- Consider the previous example where Jack scores 9 out of 30.

We determined it was plausible he was guessing and $p = 0.2$. But it is also plausible $p = 0.3$. Look at the Binomial plot below which is the Binomial distribution with $p = 0.3$ and the probability of scoring 9 or more out of 30. The probability is also quite large.

Lecture 15 (MWF) Introduction to statistical testing (testing on the mean)



This $p = 0.3$ is also a plausible outcome.

- In general, we can accept the alternative but we cannot accept the null.
- Based on this, how we should pose the hypothesis should depend on what we want to prove (which is the alternative).

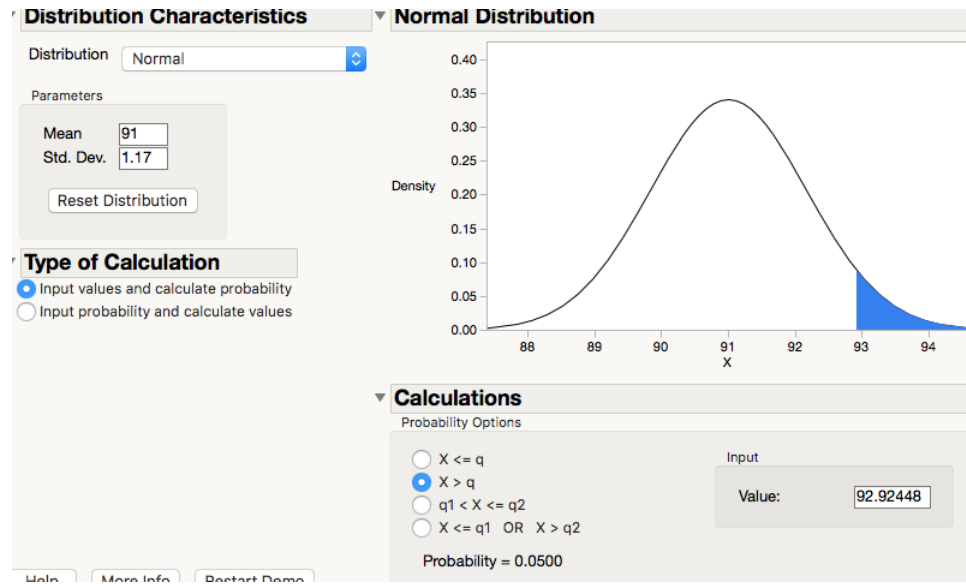
Example 2: One sided-test pointing RIGHT

In Lecture 12, Example 5, the aim was to determine if the mean weight of new born calves was over 91 pounds. We test the hypothesis

$$H_0 : \mu \leq 91 \quad H_A : \mu > 91.$$

- The alternative hypothesis is sometimes called the research hypothesis as this is in the only one we can prove.
- We assume that the population standard deviation is known and $\sigma = 7.7$ pounds (this means using a normal distribution in all calculation).
- A sample of 44 calves is “drawn” and the sample mean evaluated. The standard error is $7.7/\sqrt{44} = 1.17$ pounds.
- If the null were true, the distribution of the sample mean is given below.

Lecture 15 (MWF) Introduction to statistical testing (testing on the mean)



- The white area is the non-rejection region. If the sample mean is in this region the null hypothesis is plausible and we cannot reject the null.
Observe that this can include sample means which are far below 91, which is plausible if $\mu \leq 91$.
- The blue area is the critical region. If the sample mean is in the blue

region the p-value (the area to the **right** of that sample mean) will be less than 5%. We can only reject the null if the sample mean is in this region (at the 5% level).

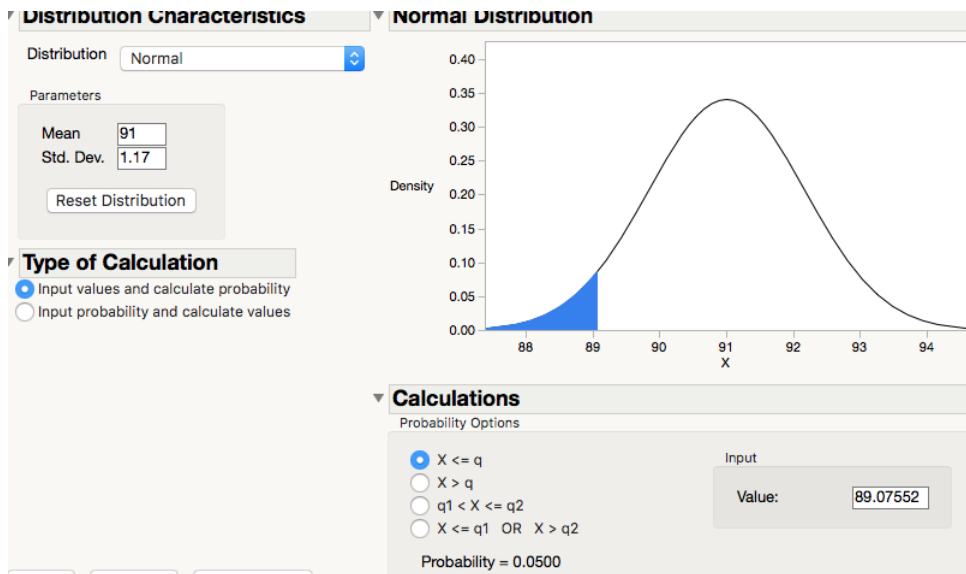
- In the data set we considered in Lecture 12, the sample mean is $\bar{X} = 93.21$. As this is in the critical region, we reject the null at the 5% level.
- The p-value is the area to the RIGHT of \bar{X} which is 2%. Since $2\% < 5\%$ we reject the null, and determine that there is evidence to suggest the mean is greater than 91 pounds.

Example 2: One sided-test pointing LEFT

If we test the hypothesis

$$H_0 : \mu \geq 91 \quad H_A : \mu < 91.$$

Under the same specifications as before. If the null were true the distribution of the sample mean looks like this



- The white area is the non-rejection region. If the sample mean lies inside this region the null hypothesis is plausible and we cannot reject the null.

Any sample mean in this region tell us a mean of $\mu \geq 91$ is plausible.

- The blue area is the critical region. If the sample mean is in the blue region the p-value (the area to the **left** of that sample mean) will be less than 5%. We can only reject the null if the sample mean is in this region (at the 5% level).
- Suppose the sample mean is $\bar{X} = 93.21$. As this is in the non-rejection region, we cannot reject the null at the 5% level. The area to the left of 93.21 is the p-value = 98%.

- For the same data set where we have $H_0 : \mu \leq 91$ and $H_A : \mu > 91$ and $H_0 : \mu \geq 91$ and $H_A : \mu < 91$:

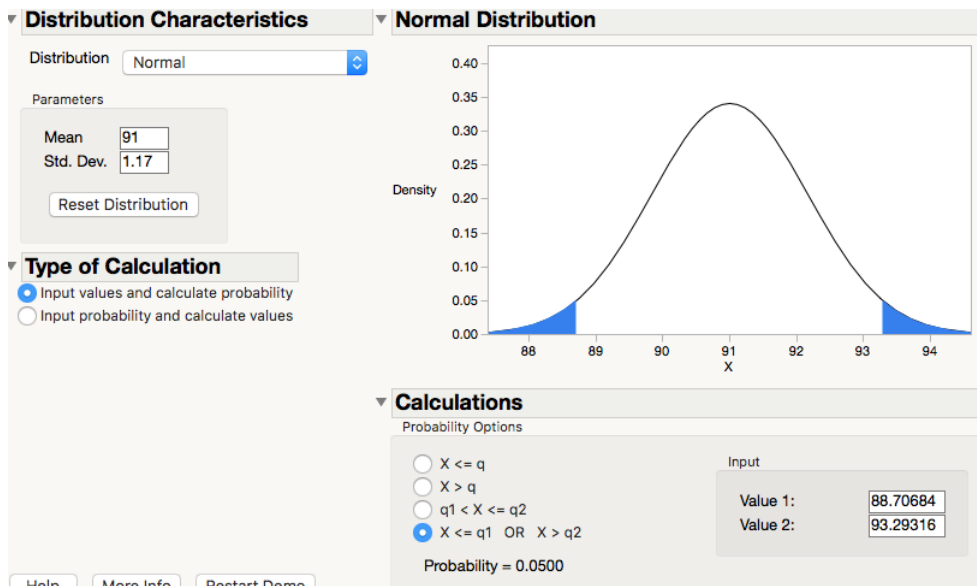
p-value for Right sided test + p-value of the Left sided test = 100%.

Example 2: Two sided-test

If the test that mean **is** 91 against the alternative that it is not:

$$H_0 : \mu = 91 \quad H_A : \mu \neq 91.$$

The null does not contain a range of values, just one value. This is a **two-sided** test. If the null were true the distribution of the sample mean is



- The white area is the non-rejection region. If the sample mean is in this region the null hypothesis is plausible and we cannot reject the null.
- The blue area is the critical region (total area for both left and right is 5%). If the sample mean is in the blue region the p-value (the area to the **left** of that sample mean) will be less than 5%. We can only reject the null if the sample mean is in this region (at the 5% level).
- In the data set we considered in Lecture 12, the sample mean is $\bar{X} = 93.21$. As this is in the non-rejection region, we cannot reject the null at the 5% level.
- The area to the right of 93.21 is 2%. The p-value is $2 \times 2\% = 4\%$.

We multiple the smallest area by **two** (to calculate the p-value), because it is a **two-sided** test.

Discussion

- We observe from the examples above, that we have a *hypothesis* of interest, something we want to investigate. The hypothesis of interest is called the research/alternative hypothesis.
- The only way prove the research/alternative hypothesis is to show that it is implausible to “generate” the observed data under the null hypothesis. The null hypothesis is the complement of the research hypothesis.
- The basic philosophy in statistical testing is if we observe something, the probability of what we observe is unlikely to be too small (else we probably could not observe it).
- **Small probabilities and how to interpret them** You observe some data (an event) and calculate the probability of that event under some conditions.

If the probability is very small, then this suggests that the probability has been calculated under the wrong set of conditions.

- On the other hand, if the probability is not very small we cannot say anything about these conditions; whether they are right or wrong.
- Below we do exactly what we did in Example 2, but do the calculations by hand.

Running example: The mean price of wheat

- Suppose you want to test whether the price of wheat has changed since last year. It is known that last year the mean price of wheat was 4 dollars a bushel (bushel is approximately 35 liters). Let μ be the true mean price (this is unknown)

We test $H_0 : \mu = 4$ (price has stayed the same) against $H_A : \mu \neq 4$ (price has changed).

- We assume throughout this discussion that the sample mean is normally distributed, so as not to distract from the main ideas.
- Later we discuss the implications of this assumption when the sample mean is not normally distributed.

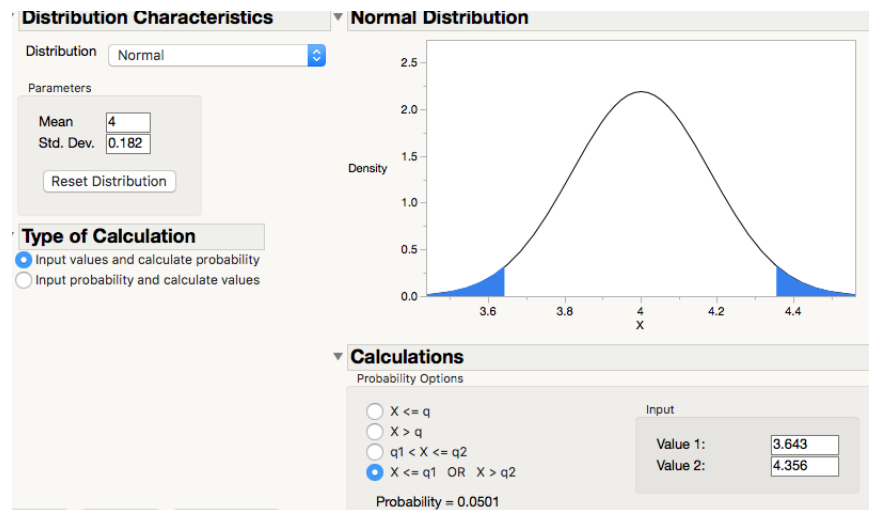
- We do not know the population mean this year, but we do have the various sample means (based on $n = 30$). We apply the test under different conditions:
 - Ex 1: Known $\sigma = 1$ and $\bar{X} = 5$.
 - Ex 2: Known $\sigma = 1$ and $\bar{X} = 4.2$
 - Ex 3: Known $\sigma = 1$, $\bar{X} = 4.2$ but now $n = 100$.
 - Ex 4: Larger known $\sigma = \sqrt{2}$, $\bar{X} = 4.2$, $n = 100$.
 - Ex 5: σ unknown and using sample standard deviation, $\bar{X} = 5$, $n = 30$.
 - Ex 6: σ unknown and using sample standard deviation, $\bar{X} = 4.2$, $n = 30$.
- We do all the above calculations by hand. However, I explain how Example 5 (and 6) can be done in JMP. You should learnt both.

Ex 1: Sample size $n = 30$, $\sigma = 1$ and sample mean = 5

- Suppose you sample 30 farmers, and from these 30 farmers, $\bar{X} = 5$.
<http://www.stat.tamu.edu/~suhasini/teaching651/wheat5.txt>
- What do you think? Has the price changed?
- This means we have to *decide* whether 5 is *significantly* far from 4. Or determine if the difference between $\bar{X} = 5$ and $\mu = 4$ can be explained by sample variation.
- We need to know how variable the sample mean (is it highly variable or not). We measure this with the standard error $1/\sqrt{30} = 0.182$.
- Heuristically, several standard errors (0.182) can fit between $\bar{X} = 5$ and $\mu = 4$, this does not seem plausible.

- We do a formal test at the 5% level. We work under the assumption that the sample mean is normally distributed (evoking the CLT).

Suppose the price of wheat has remained the same since last year at 4 dollars. Then $\bar{X} \sim \mathcal{N}(4, 1/\sqrt{30})$



The white area is 1.96 standard errors from the conjectured mean $\mu = 4$. We would expect the sample mean price to lie here if the null is true.

The white interval is the **non-rejection region**. If the sample means lies here we cannot reject it.

This interval is not a confidence interval (it is centered around the mean under the null, not the sample mean).

Ex 1 (cont): Comparing the plausible region with the observed sample mean

- From the data $\bar{X} = 5$. 5 lies in the blue critical region. Hence there is only a very small chance that the sample average can be 5, when the true mean is 4.
- We can calculate this 'small chance' by making the z-transform

$$z = \frac{5 - 4}{0.182} = 5.5$$

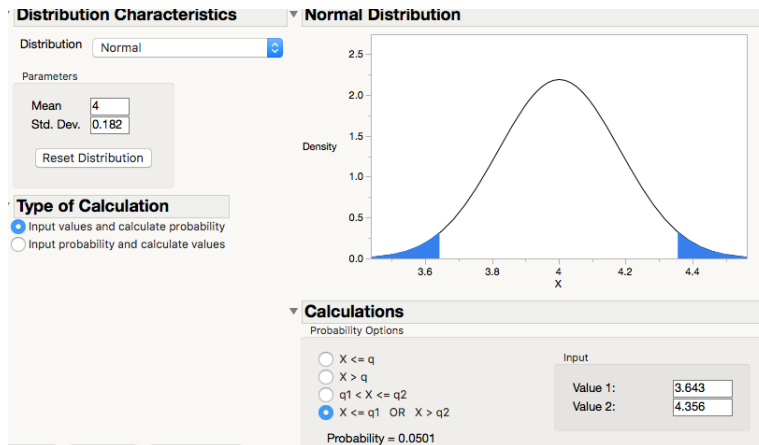
and looking up smallest area (the area to the RIGHT of 5) in the normal tables:

$$P(\bar{X} \geq 5 | \mu = 4) = P\left(Z \geq \frac{5 - 4}{0.182}\right) = P(Z \geq 5.5) = 0.00000001959.$$

- Since this is a two-sided test (we take the absolute of the z-transform) the p-value is $P(|Z| \geq 5.5) = 2 \times 0.000001959\%$. This probability is very small.
- The conclusion of this test: Since $\bar{X} = 5$ lies in the critical region (or equivalently the p-value is tiny) the null hypothesis seems implausible and **there is evidence to reject the null** at the 5% level.

Ex 2: example size $n = 30$, $\sigma = 1$ and sample mean = 4.2

- Suppose that the sample mean is 4.2 and it is calculated using a sample of 30 farmers. Here is the data http://www.stat.tamu.edu/~suhasini/teaching651/wheat4_2.txt.
- Again we want to see whether 4.2 is close to 4. In other words, can this difference $\bar{X} = 4.2$ from $\mu = 4$ be explained by sampling variation. The standard error for the sample mean is $\sqrt{1/30} = 0.182$.



The white area is 1.96 standard errors from the conjectured mean $\mu = 4$. We would expect the sample mean price to lie here if the null is true.

- Since 4.2 lies in the white non-rejection region, this means that $\mu = 4$ is plausible. We **cannot reject the null** at the 5% level.
- In fact the chance that we observe $\bar{X} = 4.2$ or greater when the true mean is $\mu = 4$ (smallest area) is

$$P(\bar{X} \geq 4.2 | \mu = 4) = P\left(Z \geq \frac{4.2 - 4}{0.182}\right) = P(Z \geq 1.1) = 0.1359.$$

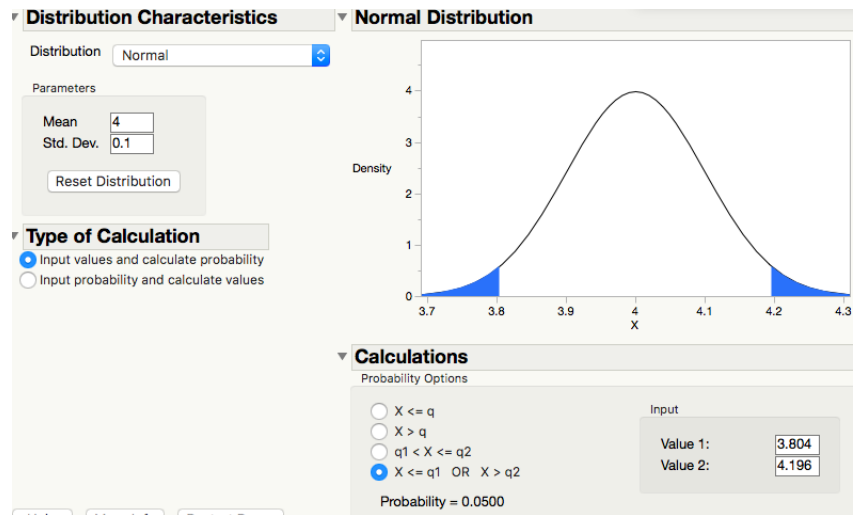
- The p-value is $P(|Z| \geq 1.1) = 2 \times 13.59\% = 27\%$ which is greater than 5%, we cannot reject the null, but of course we cannot accept it either. The mean $\mu = 4$ is plausible.

Ex 3: Sample size $n = 100$, $\sigma = 1$ and sample mean = 4.2

- Suppose that the sample mean is $\bar{X} = 4.2$ and it is calculated using a sample of 100 farmers.
- Again we want to see whether 4.2 is close to 4. In other words, can this difference of 4.2 from the last year's mean of 4, be explained by random variation is there really a change in price?
- The standard error for the sample mean is $1/\sqrt{100} = 0.1$ (observe it is now much smaller compared to the standard error with sample size 30). If the mean price has not changed we expect that about 95% of the time the sample mean to lie in the interval

$$[4 - 1.96 \times 0.1, 4 + 1.96 \times 0.1] = [3.804, 4.196].$$

Lecture 15 (MWF) Introduction to statistical testing (testing on the mean)



- The blue region is the critical region. If the sample mean lies here, then the null is considered implausible (at the 5% level) and we reject the null. The p-value will be less than 5%.
- However, observe that there is a 5% chance the sample mean will lie in this interval, even if the null were true.

- We observe that 4.2 lies in the critical region, but only just!
- In fact the chance that we observe $\bar{X} = 4.2$ when the true mean is $\mu = 4$ is

$$P(\bar{X} \geq 4.2 | \mu = 4) = P\left(Z \geq \frac{4.2 - 4}{0.1}\right) = P(Z > 2) = 0.0227.$$

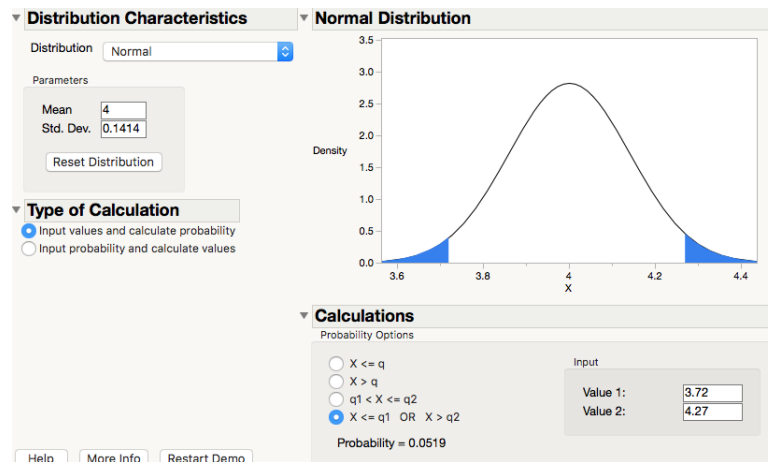
- This means the p-value is $P(|Z| > 2) = 2 \times 2.27\%$ which is relatively small. Thus we are **able to reject the null** at the 5% level (since p-value = 4.54% < 5%). There is evidence in the data to suggest the alternative is true.
- However, keep in mind there is a 4.52% chance that the z-value is $|Z| > 2$ when the $\mu = 4$.

Ex 4: Sam. size $n = 100$, $\sigma = 1.4$ and sample mean = 4.2

- Let us consider the price of Rye. Last year its mean price was 4 dollars. However, the price of rye is more variable, the standard deviation is $\sigma = 1.414$ (in contrast with the price of wheat where $\sigma = 1$).
- Suppose that the sample mean is 4.2 and it is calculated using a sample of 100 farmers.
- Again we want to see whether 4.2 is close to 4. In other words, can this difference of 4.2 from the last year's mean of 4 be due to random variation.
- The standard error for the sample mean is $1.414/\sqrt{100} = 0.1414$ (because there is more variation in the price of wheat there is more

variation in the sample mean). If the mean price has not changed we expect that about 95% of the sample means lie in the interval

$$[4 - 1.96 \times 0.1414, 4 + 1.96 \times 0.1414] = [3.72, 4.27].$$



- $\bar{X} = 4.2$ lies the non-rejection region. This means that it is not unlikely that the mean has stayed the same. In fact the chance that we observe $\bar{X} = 4.2$ when the true mean is $\mu = 4$ is

$$P(\bar{X} \geq 4.2 | \mu = 4) = P\left(Z \geq \frac{4.2 - 4}{0.141}\right) = 0.078.$$

- This means the p-value is $P(|Z| \geq 1.41) = 2 \times 7.8\% = 15.6\%$. We are **unable to reject the null** at the 5% level (since $2 \times 7.8 > 5$).
- The difference between this example and the previous example where $\sigma = 1$ is the increase in variability of the data. This results in the sample mean being more variable (larger standard error) and a wider non-rejection region. A wider white/non-rejection region means the sample mean has to “far” from the null in order to deem the implausible and to reject it.

Ex 5: Using the estimated standard deviation: $\bar{X} = 5$

Let us return to the wheat data set at

<http://www.stat.tamu.edu/~suhasini/teaching651/wheat5.txt>.

- I previously claimed that the population standard deviation, σ , is **known**. In reality this piece of information is usually unknown and needs to be **estimated** from the data.
- Loading this data set into JMP we see that the estimated standard deviation and sample mean are $s = 1.02$ and $\bar{X} = 5$ respectively.
- The procedure is exactly the same as before, but because we have estimated the standard deviation from the data we use the t-distribution instead of a normal distribution (but we will assume the sample mean \bar{X} is normally distributed).

- The sample standard error is $1.02/\sqrt{30} = 0.182$.
- To construct the non-rejection region (at the 5% level) we use exactly what was done for the normal distribution but replace 1.96 with $t_{29}(0.025) = 2.045$

$$[4 - 2.045 \times 0.182, 4 + 2.045 \times 0.182] = [3.63, 4.37]$$

- If the the sample mean lies outside this interval we reject the null at the 5% level and the p-value is less than 5%.
- We observe that $\bar{X} = 5$ does not lie in the non-rejection region. Thus we reject the null.

- To calculate the exact p-value we make the t-transform

$$t = \frac{5 - 4}{0.182} = 5.5, \quad (1)$$

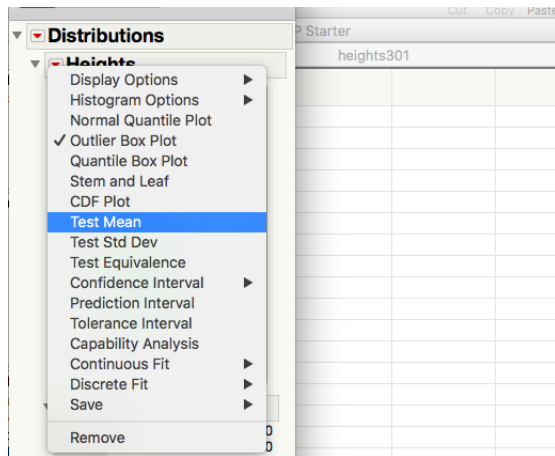
and look up the area to the right of 5.5 using the t-distribution with 29df (30-1 = 29) to deduce the p-value (or bounds for the p-value). Observe it is larger than any number on the t-table, thus the area is very small. Using software we see that the area to the right of 5.5 is 0.0000031, hence the p-value is $P(|t_{29}| \geq 5.5) = 2 \times 0.0000031 = 0.0000062$.

From the tables, since 5.5 is larger than 3.659 (which corresponds to the area 0.05%). The p-value must be less than 0.1%.

- As this probability is so small (far smaller than 5%) there is strong evidence to suggest the mean is **not** 4 and **we reject the null**.

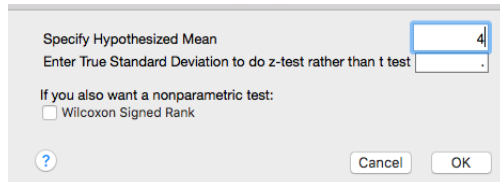
Ex 5: One sample test in JMP

- Analyze > Distribution > Highlight variable of interest (up to now the same set of instructions for plotting a histogram).
- You should see the pop-up window



Click on the red arrow next to the variable of interest and select Test Mean.

- A new window pops up



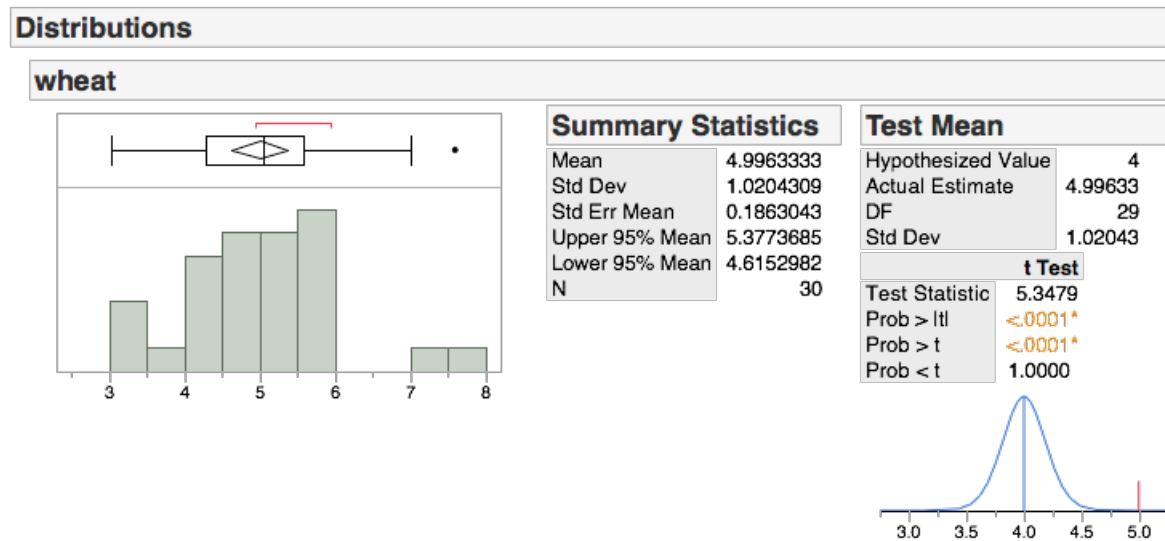
A screenshot of a dialog box with a light gray background. At the top, it says "Specify Hypothesized Mean" followed by a text input field containing the number "4". Below that, it says "Enter True Standard Deviation to do z-test rather than t test" followed by an empty text input field. Underneath, there is a section titled "If you also want a nonparametric test:" with a checkbox labeled "Wilcoxon Signed Rank" which is currently unchecked. At the bottom left is a question mark icon, and at the bottom right are "Cancel" and "OK" buttons.

In the part where it says Specify Hypothesized Mean place the mean under the null. For example, in the tests above we test $H_0 : \mu = 4$ (or $H_0 : \mu \leq 1$ or $H_0 : \mu \geq 4$), in which case we place 4 in the box.

- Observe the box where it says Enter Standard Deviation to do z-test rather than t-test. If the population standard deviation is known, then place it here. If not, leave it blank (in which case a t-test will be done).
- Ignore the Wilcoxon Signed Rank test for now. We do this later.

Ex 5 (cont): Interpreting the JMP output

This is the JMP output.



JMP gives the results of both the one-sided and the two-sided tests. In this example we are testing $H_0 : \mu = 4$ against $H_A : \mu \neq 4$. The p-value for this two-sided test corresponds to $\text{Prob} > |t|$. We see the p-value is less than 0.01%, and we reject the null. This matches our calculation.

Ex 6: Estimated standard deviation: $\bar{X} = 4.2$

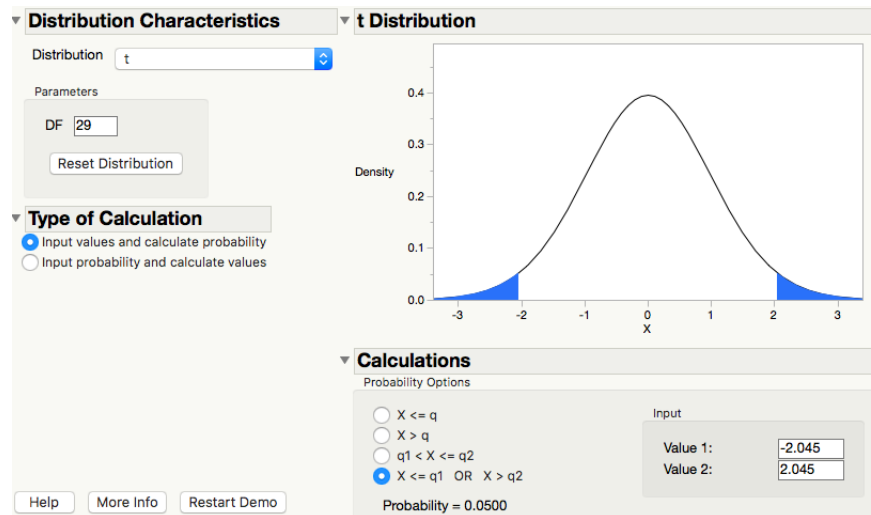
http://www.stat.tamu.edu/~suhasini/teaching651/wheat4_2.txt.

- The sample standard deviation and sample mean are $s = 0.936$ $\bar{X} = 4.2$ (respectively).
- The sample standard error is $0.936/\sqrt{30} = 0.17$.
- The non-rejection region is

$$[4 - 2.045 \times 0.17, 4 + 2.045 \times 0.17] = [3.65, 4.35]$$

- Since $\bar{X} = 4.2$ lies in this region, the null is a plausible mean, the p-value is **greater than 5%** and we **cannot reject the null**.

- The equivalent non-rejection region in the t-transform world is that the t-transform $t = (\bar{X} - 4)/0.17$ is $[-2.045, 2.045]$. If the t-transform is the blue region we reject the null (at the 5% level)



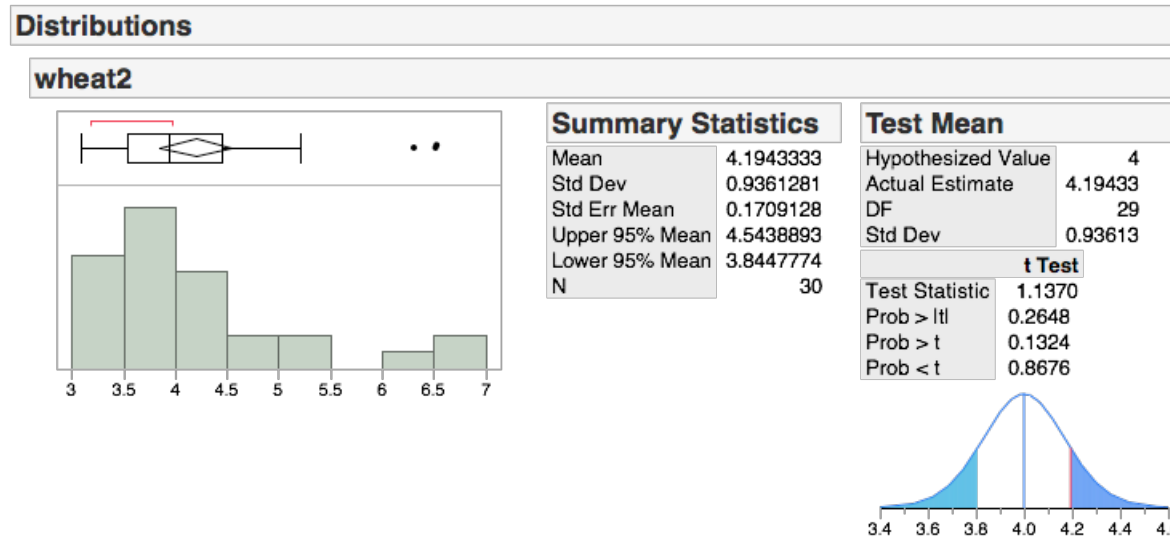
- The t-transform

$$t = \frac{4.2 - 4}{0.17} = 1.17. \quad (2)$$

This does not lie in the blue region, thus we cannot reject the null.

- To deduce the p-value we evaluate the area to the right of 1.17. From the t-tables, 1.17 is between 1.055 and 1.311 (which means the area is between 10% to 15%). The p-value is two times this, which is between 20 – 30%.
- The exact p-value (using software) is $2 \times 12.57 = 26\%$.
- As this p-value is larger than 5%, the data cannot disprove the null. We cannot know whether the mean is really 4 or not, but we certainly cannot discount this possibility (we cannot reject the null).
- Using JMP we observe

Lecture 15 (MWF) Introduction to statistical testing (testing on the mean)



From JMP output we see that the p-value is 26%, which matches our calculation. This means this data could easily be generated with a mean $\mu = 4$ and there is no evidence to prove otherwise.

P-values vs critical regions

- In the examples, above we used both p-values and rejections regions to do the test.
- They are the same. Suppose the test is done at the 5% level:
 - If the sample mean is in the critical region, then the p-value will be less than 5%. If the p-value is less than 5%, then the sample mean is in the critical region.
 - Conversely, If the sample mean is not in the critical region then the sample mean is greater than 5%. If the sample mean is greater than 5%, it will be in the non-rejection region.
- The critical region helps us to understand which sample means are sufficient far from the null for us to determine the alternative to be true (the null is implausible).

- On the other hand, p-values helps us to assess how strongly significant the alternative is. The smaller the p-value the “stronger” the evidence against the null.
- In many papers, the p-value is simply quoted.

Statistics in article

Let us return to the article on the influence of CO₂ on diet by Eweis et al. (2017). Below are the average daily food consumption and ghrelin levels in rats after drinking only regular water, a sugared soda, diet soda and decarbonated sugared soda (for 6 months).

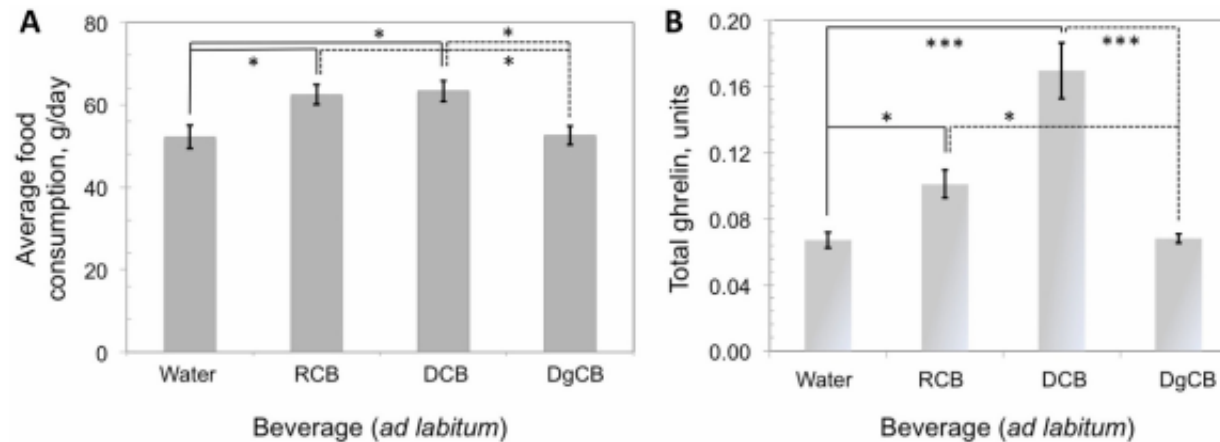


Figure 2 Rats on CBs consume more food daily due to a higher secretion of ghrelin. (A) The average food consumed per cage was monitored for all drinking groups. (B) Tail vein blood was collected from each rat to measure the levels of ghrelin hormone in the morning after the rats have been allowed to feed ad libitum (* $p < 0.05$; *** $p < 0.005$). Solid lines show statistical significance between CBs and tap water, whereas dashed lines indicate the differences between CBs and DgCB. RCB: regular carbonated beverage; DCB: diet carbonated beverage; DgCB: degassed regular carbonated beverage.

- The stars between the groups denote tests which are statistically significant. For example, to understand if there are differences in gherlin between those on water and those on DCB we test $H_0 : \mu_{water} - \mu_{DCB} = 0$ against $H_A : \mu_{water} - \mu_{DCB} \neq 0$. The *** between these two groups means the results of the test are significant, with p-value less than 0.5%. This implies the differences between sample means are so large to suggest they cannot share the same mean.

Type I errors (this is the significance level)

- The decision process: If the p-value is below $\alpha\%$ (say 5%), then it is deemed 'small' and we rejected the null

p-value = $P(\text{evidence} | \text{when } H_0 \text{ is true})$

$$\begin{cases} \leq \alpha\% & \text{Evidence to suggest } H_A \text{ is true} \\ > \alpha\% & \text{not enough evidence to suggest } H_A \text{ is true} \end{cases}$$

Often we set $\alpha = 5\%$.

- If the p-value is below α (say 5%) it does not mean that the null is not true, it is just unlikely to be true. This is why we reject the null. But we could be making a mistake and rejecting the null when it is in fact true. The significance level (often called the type I error) gives us the

probability of this happening:

$$P(\text{Conclude } H_A | H_0 \text{ is true}) = 5\%.$$

This is known as a Type I error.

- In plain terms, the Type I error, is finding a difference when there isn't any
- In a criminal trial, it is the probability of finding someone guilty when he is innocent.
- We always control the Type I error, it is something that we pre-set.
- But there is a trade off. The smaller we make the type I error, the less likely we are to reject the null.

- For example, in the wheat example, recall the case that $\bar{X} = 5$, we rejected the null at the 5% level. Suppose we set $\alpha = 0.0000000001$. Now comparing $2 \times P(\bar{X} \geq 5 | \mu = 4) = 2 \times 0.00000001959 = 0.00000004$ with 0.0000000001 , since $0.00000004 > 0.0000000001$, we are unable to reject the null. But this basically means that we cannot reject - even when the alternative is true!

If we use this rejection rate ie. $\alpha = 0.0000000001$, the mean price of wheat has to leap from 4 dollars to over 5.3 dollars (when the standard error is 0.182) for us to be able to reject the null (often we want to be able to detect subtle changes)!

- If we dropped the threshold of 5% to zero, what would happen? We would not be able to reject anything - no type I errors, but no rejections either.
- Analogy: Innocent people will not be convicted, but guilty people will

not be convicted either.

How to choose the significance level α ?

- The defacto level is 5%, and the level we will usually we use in this class.
- But the level really depends on how sure you want to be rejection of the null. If you want to extremely certain that when you reject the null, that the alternative is true, then you should choose a very small level.
- For example, in the case of detection of Higgs Boson in experiments, the physicists used an extremely low level $\alpha = 10^{-6}$ (in order to be sure that they did not make an embarrassing mistake). Thus the probability of observing the strange traces seen in an experiment given that it was down to random chance had to be less than 10^{-6} .