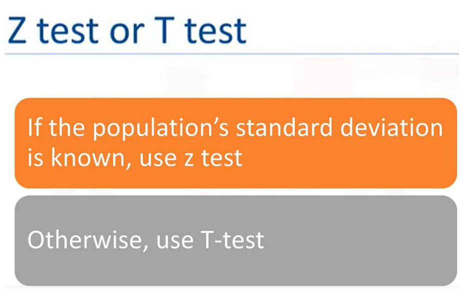
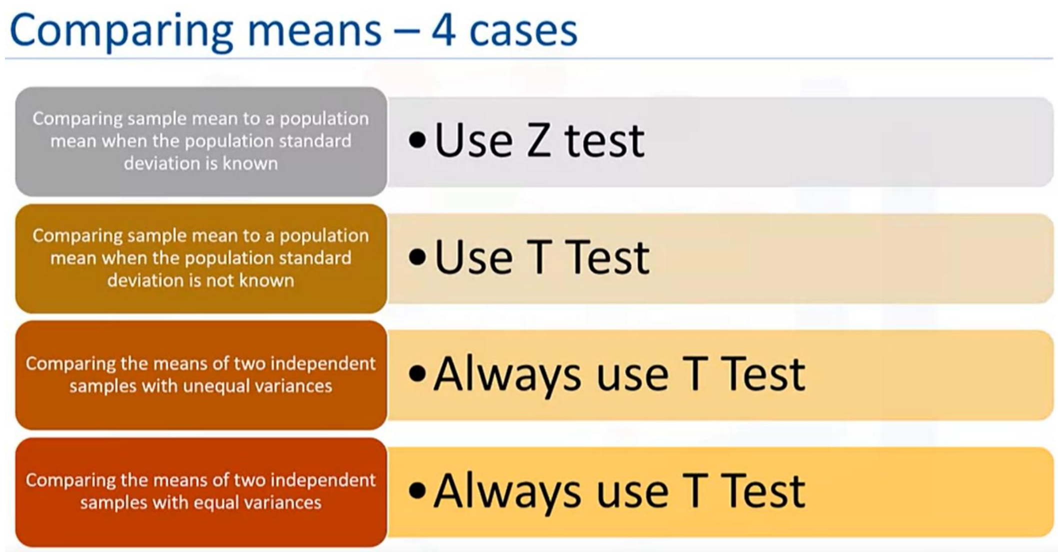
Week 4 - Hypothesis Testing

1. z-test or t-test





In traditional hypothesis testing, one has the option to perform z-test or a t-test, and the question is, under what circumstances should one perform a z-test or a t-test?

- Well, the answer is rather simple, if one is aware of the populations standard deviation or variance, we use the z-test. And that is when we are comparing the sample mean to a hypothetical or a population mean.
- And if the population standard deviation is not known and we're comparing the sample mean against the population mean within unknown standard deviation, then we use the t-test.

Now there are four scenarios in which we perform these tests.

- First scenario is where we are comparing a sample mean to a population mean and the population standard deviation is known, in that particular case, we use a z-test.
- And in cases where we are comparing a sample mean to a population mean with an unknown standard deviation, we use the t-test. Now this I covered earlier in the last slide.
- The new thing here is that when we compare the means of two independent samples, that is comparing the means of two independent samples with unequal variances. If we are faced with this kind of a question, we use a t-test.
- Again, if we are comparing the means of two independent samples with equal variances, we still use a t-test.

The underlying theory is that when you're using a z-test, you're basing your results on normal distribution, and when you are deploying t-test, you're basing your results on t-distribution.

Rules of Thumb

Type of Test	z or t Statistics*	Expected p-value	Decision
Two-tailed test	The absolute value of the calculated z or t statistics is greater than 1.96	Less than 0.05	Reject the null hypothesis
One-tailed test	The ~~absolute~~ value of the calculated z or t statistics is greater than 1.64 or less than -1.64	Less than 0.05	Reject the null hypothesis

* In large samples this rule of thumb holds true for the t-test because in large sample sizes, the t-distribution is approximate to a normal distribution

And the process of hypothesis testing could be made rather simple by looking at these thresholds. If you are comparing the means and in the particular case you are looking at the null being that the two averages are the same against two averages not being the same. Then you're using a **two-tailed test**, and in that particular case you're looking for a t-statistic or z-statistics of 1.96, the absolute value of 1.96. If that were to be the case, you reject the null hypothesis. That is, you're conducting a two-tailed test. You can be using normal distribution or a t-distribution, and you get the calculated z or t-statistics of greater than absolute value of 1.96. And the expected p-value, the probability of that happening would be less than 0.05 and you reject the null. The null being that the two means are equal.
In the case of **one-tailed test**, where you're testing whether the mean or average of one entity is greater or less than the other, here the absolute value for z or t-statistic is 1.64, and the probability would still be less than 0.05. If that were to be the case, you reject the null.

总结:

2. Dealing with tails and rejections

Rules of Thumb

Type of Test	z or t Statistics*	Expected p-value	Decision
Two-tailed test	The absolute value of the calculated z or t statistics is greater than 1.96	Less than 0.05	Reject the null hypothesis
One-tailed test	The ~~absolute~~ value of the calculated z or t statistics is greater than 1.64 or less than -1.64	Less than 0.05	Reject the null hypothesis

* In large samples this rule of thumb holds true for the t-test because in large sample sizes, the t-distribution is approximate to a normal distribution

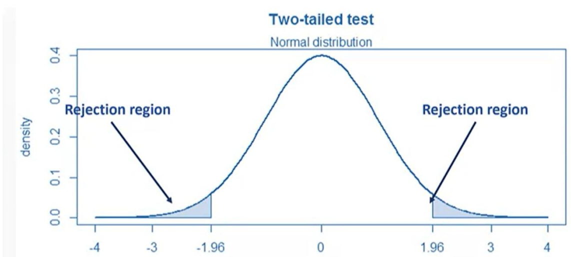
The Difference in means equals 0

- If the mean values of two variables is the same, the difference between the two means is essentially 0.
- Mathematically:
  - If $\mu_a = \mu_b$,
  - then $\mu_a - \mu_b = 0$
- Therefore, the alternative hypothesis comes in three flavors:
  - The difference in means is not equal to 0
  - The difference is greater than 0
  - The difference is less than 0.

One needs to understand the theory behind the hypothesis testing and how do you reject a null hypothesis or otherwise. There are rules of thumb. That is in case of a two-tail test, one can use 1.96 as the calculated threshold for either Z or T statistics to reject a null hypothesis or for a one-tail test, absolute value of 1.64 to reject a null hypothesis. What does it mean, how do you get to these 1.64 or 1.96? There is some theory to it. It involves statistical distributions and now perhaps is a good time to, to learn about those.

Imagine if the mean values of two variables is the same. That is, we are assuming that the difference between the two means is essentially zero. Let's say the mean of variable A and the mean of Variable B, we assume that they are equal, that is mu_A equals mu_B. And if that were to be the case, the difference between the two should be equal to zero. So the alternative hypothesis could be that the differences mean is not equal to zero. So, and you would say that mu_A is not equal to mu_B, or the difference is greater than zero. That is, mu_A is greater than mu_B, or the difference is less than zero, where mu_A is less than mu_B. And in these three circumstances, the rejection region or the how do you reject the null hypothesis means three different things.

Normal distribution and rejection regions



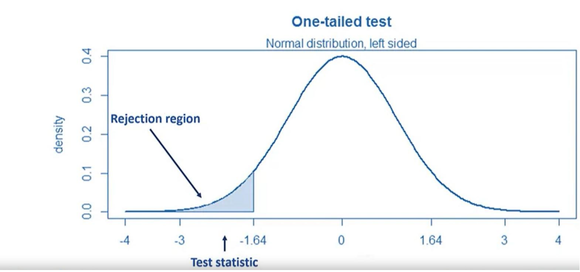
Rejection Regions and Rules of Thumb

- Alternative hypothesis: mean difference is not equal to 0
  - Difference could be greater or less than zero
  - We call this the two-tailed test
- We will define a rejection region in both tails (left and right) of the Normal distribution
- Remember, we only consider 5% of the area under the Normal curve to define the rejection region
- For a two-tailed test, we divide 5% into two halves and define rejection regions covering 2.5% under the curve in each tail

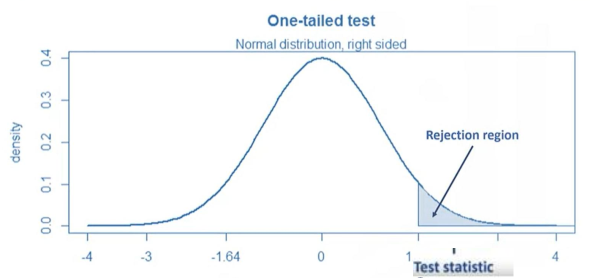
And for this, we were to revert to the normal distribution curve. Imagine that you're conducting a Z-test using a normal distribution and the shape of the curve would be very similar. This image will be very similar if we were to do a two-tailed test for T-distribution.

But let's assume that we are working with normal distribution and you have a rejection region that is to the left and to the right of the curve, as you saw the normal distribution curve. Let's say our alternative hypothesis is that the mean difference is not equal to zero. It could be greater than or less than zero, but it's not equal to zero. So we will call this a two-tail test because we are not making an assumption of the difference being greater or less than zero. Then we have to define the rejection region in both tails, that is the left tail and the right tail of the normal distribution. Remember we only consider 5 percent of the area under the normal curve to define the rejection region and for the two-tail tests that 5 percent gets divided into half of it goes into the left tail and the other half goes into the right tail, so two and half percent under the curve in each tail.
Graphically you can see this again as the same as we saw earlier, that this is a normal distribution curve and two-and-a-half percent is in the left tailed into the other two-and-a-half percent is in the right tail. If the test statistic is 1.96, if the absolute value of the test statistic is greater than 1.96 or less than 1.96, it falls in the rejection region and you can safely reject the null. The null would be that the difference of mean equals to zero or in common parlance, which is that the two means are not the same.

Left tailed test



Right tailed test



Now let us work with the assumption of the situation where we're testing if the difference of mean is less than zero, we are only interested in the left tail. Our alternative hypothesis is that the difference of mean is less than zero. In this case, the entire rejection region, that is 5 percent of the rejection region is to the left and in any situation in, for a one-tailed test, if we were to get the T-stat of 1.64 or less, we would reject the null that the mean is greater than 0 in favor of the alternative that the difference is less than 0.

The exact opposite to this would be the right tailed test. Where the alternative hypothesis is that the mean is greater than zero. And if you get the T-test statistics of greater than 1.64, for a right-tailed test, you reject the null in favor of the alternative that the mean difference is greater than zero.

总结:

3. Equal vs unequal variances

Equal versus Unequal Variances

Variances – equally unequal!

Case Summaries

teaching evaluation

female instructor	N	Mean	Std. Deviation
female	195	3.9010	.53880
male	268	4.0690	.55665
Total	463	3.9983	.55487

A t-test is the comparison of average values between two groups. For example, you could be comparing whether teaching evaluations of male instructors is the same for female instructors. You can either assume that the variance or standard deviation is equal or unequal. How do we determine this?

We have the teaching evaluation data. We calculated the average teaching evaluation for female instructors to be 3.9 with a standard deviation of 0.53. The average teaching evaluations for male instructors on the other hand, was calculated as 4.06 with a standard deviation of 0.55.

Levene's Test

Levene's Test is an inferential statistic to assess the equality of variances.

Null hypothesis:

- Population variances are equal
- If the p-value < 0.05, reject the Null Hypothesis of Equal Variances

When we conduct a t-test, we are faced with whether to assume equal or unequal variances. We have a t-test called **Levene's test** to determine the equality of variances. The null hypothesis of the Levene's test is that population variances are equal, if the p-value of the test is less 0.05, reject the null hypothesis of equal variances and assume that the variances are unequal.

Conclusion: Variance does not differ

```
scipy.stats.levene(ratings_df[ratings_df['gender'] == 'female']['eval'],
                   ratings_df[ratings_df['gender'] == 'male']['eval'], center='mean')
```

```
LeveneResult(statistic=0.1903292243529225, pvalue=0.6628469836244741)
```

Variance is equal

```
scipy.stats.ttest_ind(ratings_df[ratings_df['gender'] == 'female']['eval'],
                      ratings_df[ratings_df['gender'] == 'male']['eval'], equal_var = True)
```

Let us look at the example for that of teaching evaluation scores for male and female instructors.

- We will use the Levene's function in the ***scipy.stats*** package. We will run it against both samples and we will specify the center argument as **mean**. Since our t-tests, we test for mean differences, we will get a p-value of 0.66, which is greater than 0.05. That means we will fail to reject the null hypothesis and we will assume equal variances when conducting our t-test.
- When you run your t-test, you set the equal underscore var option to true and if you got a p-value less than 0.05, you set that option to false.

总结:

Equal and Unequal Variances

Equal Variances

$$sdev = \sqrt{\frac{vpool*(n_1+n_2)}{n_1*n_2}}$$

$$vpool = \frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}$$

$$t = \frac{x_1-x_2}{sdev}$$

Unequal Variances

$$t = \frac{x_1-x_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

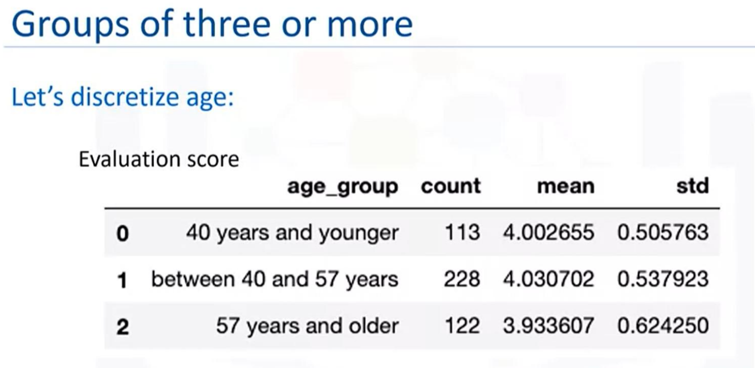
$$\boxed{dof = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}}$$

If you were to do a t-test by hand, the formula would be different for calculating equal versus unequal variances.

- You must calculate the pooled variance as shown here. Then calculate the standard deviation that you will use in the t-test.
- If the variances were unequal, calculate the t-test with each of the individual standard deviations and sample size. You will need to find the degree of freedom to check the t-test table. With this formula the rule of thumb for assuming equal variance when calculating by hand, is defined by the ratio of the larger groups variance to the smaller groups variance to be less than 1.5.

总结：

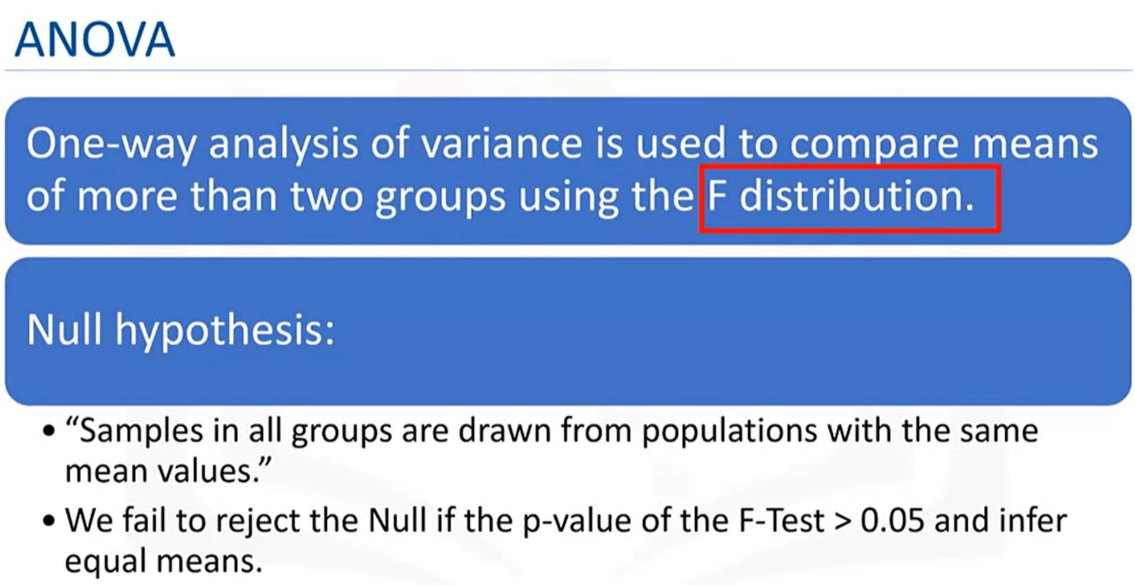
4. ANOVA



	age_group	count	mean	std
0	40 years and younger	113	4.002655	0.505763
1	between 40 and 57 years	228	4.030702	0.537923
2	57 years and older	122	3.933607	0.624250

Most of us are familiar with comparing the average values between two groups. For example, the average teaching evaluation square group four male instructors compared with that of female instructors, and the groups are two, and we know that such comparisons are made using the t-test. But what if you're dealing with more than two groups? What if there are three, four or more groups? In that particular case, we would use **ANOVA** or **analysis of variance**, where our intent or goal is to compare the means of more than two groups.

So in order to accomplish this, we return back to our teaching evaluation data. In that particular case, we have a variable called age, where the age of the instructor is recorded in a number of years. But we will discretize this age variable, that is, we will create three groups. So instructors who are 40 years and younger we put them in one group, those between 40 and 56.5 years of age, they are in another group, and those who are 57 years and older, you put them in the third group. So you have younger instructors, middle-age instructors and rather slightly older instructors, and the number of observations, taught by each group, is reported under N. What we also have here is the teaching evaluation score for each group, which is not deferring much. It's pretty much four for each group and for the older professors is slightly less at 3.9 in that respective standard deviations of it. So we have three groups and let's see what we are interested in is to determine if these three averages for the three respective age categories are statistically the same or they are different, so we use the one-way analysis of variance or ANOVA.



And using the ANOVA we use the F-distribution to compare the mean values for more than two groups.
Our null hypothesis is that samples in all groups are drawn from the same populations with the same mean values.
We fail to reject the null hypothesis if the P-value or the significance for the F-test is greater than 0.05 and we then infer equal means.

总结:

ANOVA in Python

Does beauty score for instructors differ by age?

Beauty score

	age_group	count	mean	std
0	40 years and younger	113	0.336196	0.913748
1	between 40 and 57 years	228	-0.035111	0.686637
2	57 years and older	122	-0.245777	0.740720

```
ratings_df.loc[(ratings_df['age'] <= 40), 'age_group'] = '40 years and younger'
ratings_df.loc[(ratings_df['age'] > 40)&(ratings_df['age'] < 57), 'age_group'] = 'between 40 and 57 years'
ratings_df.loc[(ratings_df['age'] >= 57), 'age_group'] = '57 years and older'
```

```
f_statistic, p_value = scipy.stats.f_oneway(forty_lower, forty_fiftyseven, fiftyseven_older)
print("F_Statistic: {0}, P-Value: {1}".format(f_statistic,p_value))
```

```
F_Statistic: 17.597558611010122, P-Value: 4.3225489816137975e-08
```

ANOVA

Does teaching evaluation score for instructors differ by age?

	age_group	count	mean	std
0	40 years and younger	113	4.002655	0.505763
1	between 40 and 57 years	228	4.030702	0.537923
2	57 years and older	122	3.933607	0.624250

```
f_statistic, p_value = scipy.stats.f_oneway(forty_lower_eval, forty_fiftyseven_eval, fiftyseven_older_eval)
print("F_Statistic: {0}, P-Value: {1}".format(f_statistic,p_value))
```

```
F_Statistic: 1.2226327996572204, P-Value: 0.29540894225417536
```

Let's say we are interested in determining if the beauty score for instructors differs by age. We have three groups, younger, middle aged, and older professors. We have the summary statistics for the standardized beauty scores. We see that there is a difference as the age goes up, the average value for the beauty score goes down. So let's run an ANOVA to see if the differences are statistically significant. Our null hypothesis will be, "Mean beauty scores for instructors don't differ with age," and the alternative hypothesis will be, "At least one of the means is different."
First, this variable does not exist in our data. We will need to group or been the continuous age data using the **.loc** function in Pandas. Then use the F underscore one wave function in the Scipy Stats Library to perform the ANOVA test. We will then print out the F statistics and the P-value. What we can see is that the P-value is 4.32 times ten raised to the power of negative eight, and that is less than 0.05. We will reject the null hypothesis as there is significant evidence that at least one of the means differ.

If I do the same tests for the teaching evaluation scores that we observe for the three groups, and we run ANOVA on these three mean values. We find out that the P-value is 0.295, which is greater than 0.05. We will fail to reject the null and infer equal means. That is, that the three means are not statistically different.

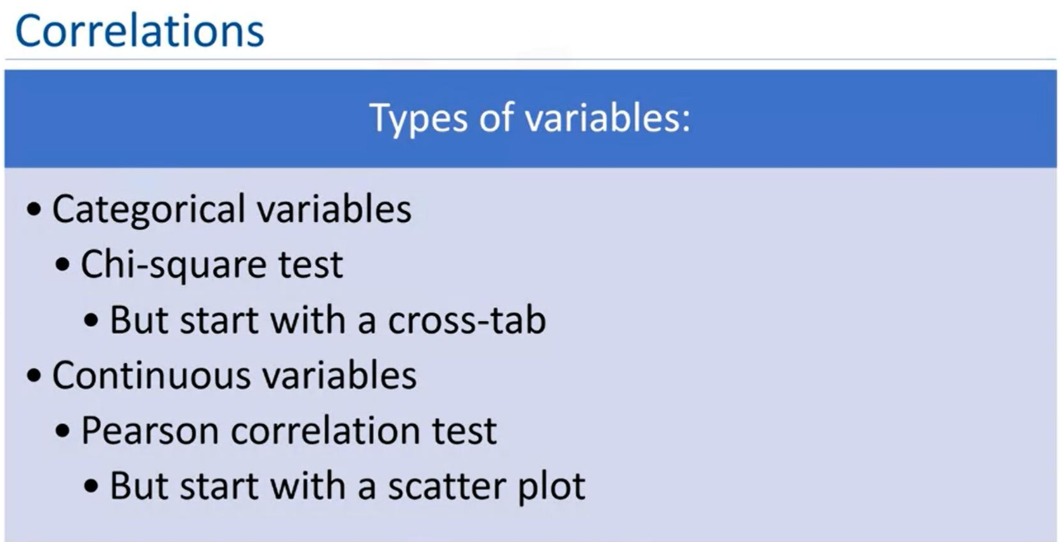
小结:

- Here we have the Analysis of Variants performed on two samples. One is the beauty score. We notice that the difference in means for beauty scores between the three groups is based on the significance value. This leads us to conclude that at least one mean is different and we reject the null hypothesis that states equal means.

- Here, because the P-value for teaching evaluation scores between the three groups is greater than 0.05, we fail to reject the null hypothesis. We believe that these three means are statistically equal.

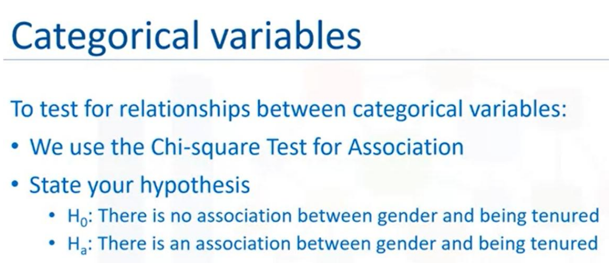
总结:

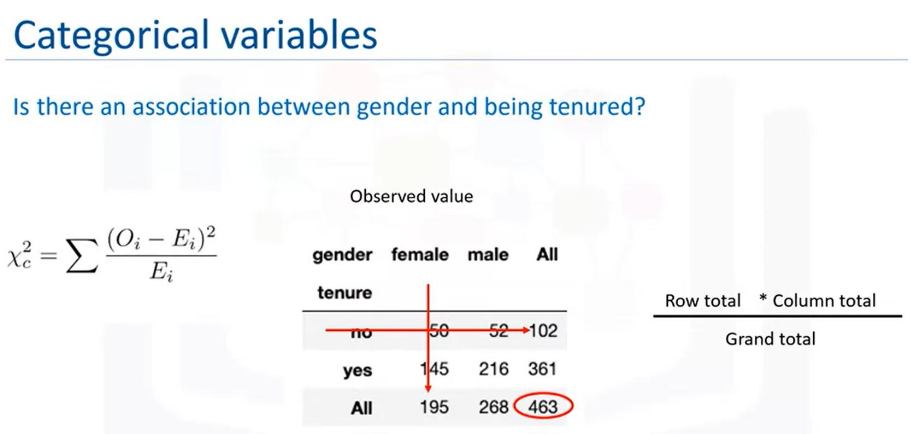
5. Correlation tests



Now, moving ahead from comparing the average values between two or more groups, we are looking at two variables. We want to know if there is a statistically significant correlation between these two variables and what is needed for this to happen. We would need to look back to the earlier definition of types of variables. <u>We generally define the variables in two groups, the categorical variables, an continuous variables</u>.

- So if we were to go back to our teaching ratings data, we have instructors who are male and female. Some instructors are visible minorities and some are Caucasian. So we have two variables, male, and female and the visible minority status. These two variables are examples of categorical variables, and if we're comparing or trying to determine the correlation between two categorical variables, we would use the Chi-square Ttest. We would begin with a cross tabulation between the two values.
- If we have two continuous variables, for example, the teaching evaluations score and the beauty score of an instructor, these are two continuous variables and they can assume any reasonable value within the range. In this case, we use a Pearson correlation test. We usually begin with a scatter plot to see what's the nature of the relationship between the two variables.





$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Let's start with categorical variables. We'll use the Chi-square Test for Association. First, we state our hypothesis. We will test the null hypothesis that gender an tenureship are independent against the alternative hypothesis that they're associated. Let's begin with a cross tabulation between gender male and female, and tenure. That is, tenured profs then followed by a Chi-square test.

So we do the tabulations. In the rows we have tenured no versus tenured yes, and female instructors and males. We would like to eyeball these numbers before we turn them into percentages. Looking at instructors who are non-tenured, we notice that 50 of the instructors are female versus 52 who are male. But for the instructors who are tenured, 145 of them are female, and 216 of them are male. So within the tenured group we see greater probability for males to be tenured, but in the untenured group the distribution between males and females look similar. Before we go to Python, let's do this by hand to understand the concept. The formula for Chi-square is given as follows. The summation of the observed value, i.e the counts in each group minus the expected value, all squared, divided by the expected value. Expected values are based on the given totals. What would we say each individual value would be if we did not know the observed values? So to calculate the expected value of untenured female instructors, we take the row total, which is 102 multiplied by the column total 195, divided by the grand total of 463. This will give you 42.96. (下一页继续)

总结:

Categorical variables

$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

gender	female	male	All
tenure			
no	50	52	102
yes	145	216	361
All	195	268	463

Observed value

Expected value

Row total * Column total

Grand total

tenure \ gender	female	male	
no	42.96	59.04	102
yes	152.04	208.96	361
	195	268	

Categorical variables

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Degree of freedom = (row-1)*(column-1)

$$\chi_c^2 = 2.557$$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x ²							
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92

P-value > 0.05, we fail to reject the null hypothesis that the two variables are independent and conclude that a systematic association *does not* exist between gender and tenure.

If we do the same thing for tenured male instructors, we will take the row total 361 multiplied by the column total 268 divided 463, we get 208.96.. If we repeat the same procedure for all of them, we get these values. If we take the row totals, column totals, and grand total, we will get the same values as the totals as the observed values.

Now going back to this formula, if we take a summation of all the observed minus the expected values, all squared, divided by the expected value, we will get a Chi-square value of 2.557. And the degree of freedom will be 1. On the Chi-square table, we check the degree of freedom equals row one and find the value closest to 2.557. Here we can see that 2.557 will most likely fall in between a p-value of 0.1 and 0.25. Therefore, we can say that the p-value is greater than 0.1. Since the p-value is greater than 0.05, we fail to reject the null hypothesis that the two variables are independent and therefore we will conclude that the alternative hypothesis that there is an association between gender and tenureship does not exist.

Categorical variables

gender	female	male
tenure		
no	50	52
yes	145	216

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

P-value of 0.109 > 0.05, we fail to reject the null hypothesis that the two variables are independent and conclude that a systematic association *does not* exist between gender and tenure.

```
scipy.stats.chi2_contingency(cont_table, correction = False)
```

```
(2.557051129789522,
 0.10980322511302845,
 1,
 array([[ 42.95896328,  59.04103672],
        [152.04103672, 208.95896328]]))
```

To do this in Python we will use the Chi-square contingency function in the SciPy statistics package, that is a Chi-square test value of 2.557. And the second value is the p-value of about 0.11. And a degree of freedom of 1. If you remember, the Chi-square table did not give an exact p-value but a range in which it falls. Python will give the exact p-value. We can see the same results as on the previous slides. It also prints out the expected values, which we also calculated by hand.

Since the p-value is 0.11, which is greater than 0.05, we fail to reject the null hypothesis that the two variables are independent. And therefore we will conclude the alternative hypothesis that there is an association between gender and tenureship does not exist. This was an example of testing independence between two categorical variables.

总结:

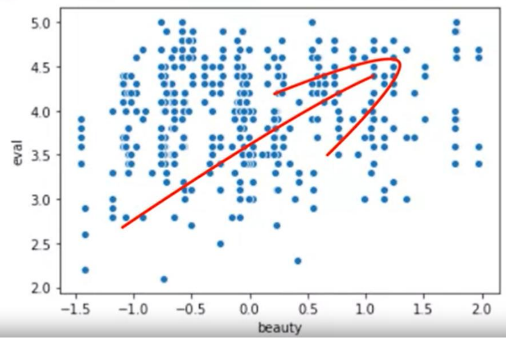
Continuous variables

To test for relationships between continuous variables:

- We use the Pearson Correlation Test
- State your hypothesis
  - H_0: There is no correlation between an instructor's beauty score and their teaching evaluation score.
  - H_a: There is a correlation between an instructor's beauty score and their teaching evaluation score.

Continuous variables

Is teaching evaluation score correlated with beauty score?



Now to continuous variables using a Pearson correlation test from the teaching ratings data. We will test the null hypothesis that there is no correlation between an instructor's beauty score and their teaching evaluation score against the alternative hypothesis that there is a correlation between both variables.
We had the normalized beauty score on the x-axis and the teaching evaluation score on the y-axis. You can eyeball a positive upward sloping curve, but let's run a Pearson correlation test to find out.

Pearson Correlation test

```
scipy.stats.pearsonr(ratings_df['beauty'], ratings_df['eval'])
```

```
(0.1890390908404521, 4.247115419812614e-05)
```

$$r_{xy} = \frac{\mathrm{cov}(x,y)}{\sqrt{\mathrm{var}(x)}\cdot\sqrt{\mathrm{var}(y)}}$$

Null Hypothesis: There is no association between an instructor's looks and teaching evaluation score.

Since the p-value (Sig. (2-tailed) < 0.05, we reject the Null hypothesis and conclude that there exists a relationship between beauty and teaching evaluation score.

Correlation coefficient varies between -1 and 1.

We will use the Pearson R package in the scipy.stats package and check for the correlation. We will get a coefficient value of how strong the relationship is and in what direction. Correlation coefficient values lie between -1 and 1. Where -1 means a strong negative correlation and visually represented by a downward sloping curve, and 1 means a strong positive relationship and visually represented by an upward sloping curve. In our case, we have a Pearson coefficient of 0.18, and a p-value of 4.25 times 10 raised to power -5. Since the p-value is less than 0.05, we reject the null hypothesis and conclude that there exists the relationship between an instructor's beauty score and teaching evaluation score.

总结:

2. For the following samples assume they follow a normal distribution and we assume equal variance, we will like to know if there is a difference between both sample means. If we perform a two-sample t-test for independent samples. What is the p-value for the test Statistics?

Sample1 = 9, 11, 10,11,10,12, 9,11,12, 9, 10

Sample2 = 10, 13, 10, 13, 12, 9, 11, 12, 12, 12, 13

- (●) 0.0384
- ( ) 2.21
- ( ) 0.0885
- ( ) 0.975

```
>>> from scipy.stats import ttest_ind
>>> s1 = [ 9, 11, 10,11,10,12, 9,11,12, 9, 10]
>>> s2 = [ 10, 13, 10, 13, 12, 9, 11, 12, 12, 12, 13]
>>>
>>> ttest_ind(s1, s2)
Ttest_indResult(statistic=-2.2164816032790386, pvalue=0.03841461541539729)
```

✓ **正确**

Correct!

3. What test is used to test the equality of variance

- ( ) t-test
- (●) Levene's test
- ( ) ANOVA
- ( ) z-test

✓ **正确**

Correct!

总结：

1. Using the teacher's rating data, is there an association between native (native English speakers) and the number of credits taught? What test will you use?

- ○ Z-test
- ◉ Chi-Square Test for Association
- ○ T-test
- ○ ANOVA

✓ **正确**

Correct!

2. If I wanted to test for association using chi-square test, whether there is an association between gender (Male or Female) and tenure-ship (tenured or not tenured), what will be my degree of freedom?

1

✓ **正确**

Formula for degree of freedom for chi-square if (r-1) *(c-1)

4.

Phone Brand	A	B	C	D
	24	26	18	27
	19	24	18	24
	22	20	20	22
	25	18	23	24
Mean	22.5	21.75	19.75	24.25
Std. dev.	2.64	10.91	2.36	4.25

Battery life of smartphones is of great concern to customers. A consumer group tested four brands of smartphones to determine the battery life. Samples of phones of each brand were fully charged and left to run until the battery died. The table above displays the number of hours each of the batteries lasted. What test will be be using to test the difference in means?

Z test 和 T test：可简单理解为检测两个变量之间是否有相同的 mean。对于大于或等于两个变量的，采用 ANOVA 检测。

- ○ Pearson Correlation Test
- ○ T-test
- ◉ ANOVA
- ○ Chi-square Test

Chi Square 和 Pearson Correlation：前者是 categorical 变量的相关性检测，后者是 continuous 变量的相关性检测。

✓ **正确**

Correct! there are more than two groups

总结：

5. A room in a laboratory is only considered safe if the mean radiation level is 400 or less. When a sample of 10 radiation measurements were taken, the mean value of the radiation was 414 with a standard deviation of 17. There are concerns that mean radiation is above 414. Radiation levels in the lab are known to follow a normal distribution with standard deviation 22. We will like to conduct a hypothesis test at the 5% level of significance to determine whether there is evidence that the laboratory is unsafe.

What will be the appropriate test?

- (◉) z-test
- ( ) t-test
- ( ) ANOVA
- ( ) Chi-square

✓ **正确**

Correct! We use a z-test when the population standard deviation is known

7. The P-value for a normally distributed right-tailed test is P=0.042. Which of the following is **INCORRECT**?

- ( ) We will reject H0 at α=0.05, but not at α=0.01
- ( ) The z-score test statistic is approximately z=1.73
- ( ) The P-value for a two-tailed test based on the same sample would be P=0.084
- (◉) The P-value for a left-tailed test based on the same sample would be P= -0.042

✓ **正确**

Correct! P-values are proportion and range from 0 to 1. The left-tail test for this will also be 0.042

8. The time X taken by a cashier in a grocery store express lane to complete a transaction follows a normal distribution with mean 90 seconds and standard deviation 20 seconds.What is the first quartile of the distribution of X (in seconds)?

- ( ) 88.0
- ( ) 73.8
- ( ) 81.2
- (◉) 76.6

对于 normal distribution，它的
mean = median = mode

✓ **正确**

Correct!

总结: