Two-Sample Tests

There are many problems in which we have to decide whether an observed difference between two means is attributable to chance or whether they represent different populations.

Suppose we are interested in cardiovascular risk factors in children and want to know if the average heart rate among newborns is different between whites and blacks.

When comparing two means, our objective is to decide whether the observed difference is statistically significant or whether the difference is attributable to chance fluctuation.

We let $\overline{X}_1 - \overline{X}_2$ represent the difference between the means of sample 1 and sample 2.

I. Test between Means: Large Samples

Suppose we selected two large independent samples of size n_1 and n_2 having the means $\overline{X}_1$ and $\overline{X}_2$.

We **assert** (rather than prove) that the difference $\overline{X}_1 - \overline{X}_2$ can be approximated by a standard normal curve whose mean and standard deviation are given by:

$$\mu=\mu_1-\mu_2$$

$$\sigma_{X1-X2} = \sqrt{\frac{\sigma_1^{\,2}}{n_1}+\frac{\sigma_2^{\,2}}{n_2}}$$

Where μ_1 and μ_2 are the population means from the respective populations and σ^2 represents the variance. We refer to σ_{X1-X2} as the *standard error of the difference between two means*.

If the population standard deviations are not known, we substitute S_1 for σ_1 and the same for S_2 and estimate the standard error of the difference between two means by:

$$\sqrt{\frac{S_1^{\,2}}{n_1}+\frac{S_2^{\,2}}{n_2}}$$

When comparing two means, the null hypothesis usually assumes the form:

H_o: $\mu_1=\mu_2$ or $\mu_1 - \mu_2 = 0$

Which implies that there is no difference between the means

The alternative for a two tailed test:

H_1: $\mu_1 \neq \mu_2$ or $\mu_1 - \mu_2 \neq 0$

If it is a one-tailed test:

H_1: $\mu_1 > \mu_2$ or $\mu_1 - \mu_2 > 0$

The Z value in this case (recall we can use the standard normal here since either we know the population standard deviation, or the sample size is large so that the normal distribution can be used as an approximate.

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_1^2 / n_1 + S_2^2 / n_2}}$$

Which is a random variable having the normal distribution.

Back to our example: Suppose we collect the following data on heart rate among newborns:

Race	Mean (beats per minute)	sd	n
White	129	11	218
Black	133	12	156

And we want to test the hypothesis that the average heart rate for white newborns is different than the average heart rate for black newborns.

H_o: $\mu_w = \mu_b$ or $\mu_w - \mu_b = 0$
H_1: $\mu_w \neq \mu_b$ or $\mu_w - \mu_b \neq 0$

$$Z = \frac{129 - 133}{\sqrt{11^2 / 218 + 12^2 / 156}} = \frac{-4}{\sqrt{.555 + .923}} = \frac{-4}{1.216} = -3.289$$

Using Excel: =normsdist(-3.289) gives the p-value associated with this: .00050.

There is only a .05% chance that we could get this large of a difference between the sample means if, indeed, the population means were equal. Thus we would reject the null hypothesis and conclude that there is evidence from our sample that the average heart rates are different between white and black newborns.

II. Test between Means: Small Samples

If our samples come from populations that can be approximated by the standard normal curve and if $\sigma_1 = \sigma_2 = \sigma$ (i.e., the standard deviations of the two populations are equal), a small sample test of the differences between two means may be based on the t distribution. Then our test statistic becomes:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\Sigma(x_1 - \overline{X}_1)^2 + \Sigma(x_2 - \overline{X}_2)^2}{n_1 + n_2 - 2} * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Note that this "kind of" looks like (X-μ)/σ/√n). In the denominator note that there is something like the variance of each variable in there, and we are dividing by n.

By definition $\Sigma(x_1\text{-}X_1b)^2 = (n_1\text{-}1)S_1^2$ and $\Sigma(x_2\text{-}X_2b)^2 = (n_2\text{-}1)S_2^2$
This equation can be "simplified" to be:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} * \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

This is the equation in your book (page 375) This is known as the "pooled variance t-test" for the difference between two means. Note that we are calculating something that looks like the two variances combined in the equation.

Since n_1-1 of the deviations from the mean are independent in S_1^2 and n_2-1 of the deviations from the mean are independent in S_2^2, we have (n_1-1) + (n_2-1) or (n_1+n_2-2) degrees of freedom. Thus the above equation represent the t distribution with (n_1+n_2-2) degrees of freedom.

Suppose we want to know if there is a difference in turnover between RNs who work in the ICU and those who work in the med/surg unit. We take a sample of both groups and calculate the average number of years spent on the job.

Type	Average Term Years	Sd	N
ICU	20	5	12
Med/Surg	22	4	15

Is there evidence that Med/Surg RNs stay on the job longer than ICU RNs?

H_o: $\mu_{icu}=\mu_{med}$ or $\mu_{icu} - \mu_{med} = 0$
H_1: $\mu_{icu} < \mu_{med}$ or $\mu_{icu} - \mu_{med} < 0$

$$t = \frac{20 - 22}{\sqrt{\frac{(12-1)5^2 + (15-1)4^2}{12+15-2}} * \sqrt{\left(\frac{1}{12} + \frac{1}{15}\right)}} = \frac{-2}{\sqrt{\frac{499}{25}} * \sqrt{.15}} = \frac{-2}{1.73} = -1.156$$

The critical value for α=.05 and 25 degrees of freedom is -1.7081 so we would fail to reject the null and conclude that there is no evidence from this sample that Med/Surg RNs stay on the job longer than ICU RNs.

Or using Excel: =tdist(-1.56, 25, 1) gives a pvalue of .129, so there is about a 13 percent chance of getting two sample means this far apart if the population means were really equal, so if we use α=.05 we would fail to reject the null and conclude there is no evidence that med/surg RNs stay on the job longer than ICU RNs.

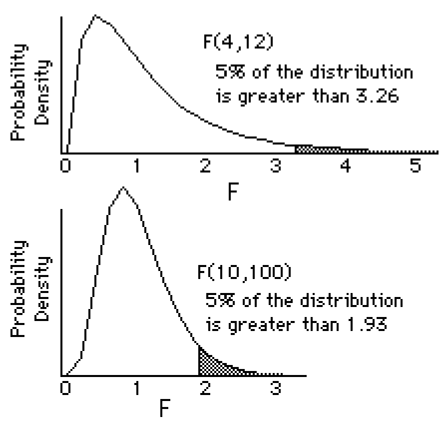
Note that in Excel under the Tools, data analysis, tabs there are canned formulas for the z-test for different sample means, t-tests under equal variances and t-tests under unequal variances. These are OK to use as a check on your work, but for this section of the homework, I want you to calculate the z/t values on your own. Afterwards you are free to use them.

III. F-Test for the difference between two variances

The above t-test made the assumption that the two samples have a common variance. In order to be complete this assumption must be tested. Assuming that independent random samples are selected from two populations, test for equality of two population variances are usually based on the ratio S_1^2/S_2^2 or S_2^2/S_1^2. If the two populations from which the samples were selected are normally distributed, then the sampling distribution of the variance ratio is a continuous distribution called the F distribution.

The F distribution depends on the degrees of freedom given by n_1-1 and n_2-1 where n_1 and n_2 are the sample sizes that S_1^2 and S_2^2 are based.

The F- distribution looks as follows:



It is skewed to the right, but as the degrees of freedom increases it becomes more symmetrical and approaches the normal distribution.

If you really have lots of time to waste and are curious you can go to the following web site and play with df and how the F changes shape.

http://www.econtools.com/jevons/java/Graphics2D/FDist.html

The easiest way to conduct the test is to construct the ratio such that the large of the two variances are on the top. That way you are always looking at the upper tail and do not need to worry about the lower tail.

Basically what is going on is you are comparing the two variances and asking "is one bigger than the other?" If the variances are equal then this will equal 1, if the variances are "different" then this ratio will be something greater than 1. So the question is how far away is too far away? That is what the F distribution tells us. Under the Null the variances are equal and so the ratio should equal one, but there is some probability that the ratio will be larger than 1, thus if it is too unlikely for the variances to be equal we reject the null and conclude the variances are not equal:

Suppose we are interested in examining the mean stays experienced by two populations from which we have drawn a random sample of size n_1=6 and n_2=8. Based on this random sample, suppose we found that S_1^2 =12 days and S_2^2 = 10 days. Before testing the significance of the observed difference between the two mean stays, we should examine the assumption that $\sigma_1^2 = \sigma_2^2$ by:

H_o: $\sigma_1^2=\sigma_2^2$
H_1: $\sigma_1^2 \neq \sigma_2^2$

F = 12/10 = 1.2

Using α=.05. Here there are 5 degrees of freedom in the numerator and 7 degrees of freedom in the denominator. Using the Excel function **FDIST(F,df1,df2)** or =FDIST(1.2,5,7)=.301. Thus if the null is true that the variances are equal, there is a 30 percent chance we could get two sample variances that were this far apart from each other. Thus, we would not reject our null hypothesis and conclude that there is no evidence that the two variances are different.

IV. Tests between two samples: Categorical Data

This section deals with how we can test for the difference between two proportions from different samples. We might be interested in the difference in the proportion of smokers and nonsmokers who suffer from heart disease or lung cancer, etc.

Our Null Hypothesis would be: H_o: P_1=P_2
Where P_1 and P_2 are the population proportions. Letting p_1 and p_2 represent the sample proportions, we can use $p_1 - p_2$ to evaluate the difference between the proportions.

1. **Z test for the difference between two proportions**

One way to do this test is using the standard normal distribution. Here our Z statistic is:

$$Z = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{P^*(1-P^*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where: x_1/n_1 the proportion of successes from sample 1 and x_1/n_1 is the sample of successes from sample 2

and

$$P^* = \frac{x_1 + x_2}{n_1 + n_2}$$

Note that this is similar to the formula we used to test hypotheses on a single sample proportion, the denominator is using the combined proportion to estimate the standard error of the difference in the sample proportions.

EX: in 1997 a sample of 200 low income families (less than 200 percent of the poverty line) was taken and it was found that 43 of them have children who have no health insurance. In 2003 a similar survey of 250 families was taken and 48 were found not to have insurance. Is there evidence from these samples that the proportion of low income children without insurance has changed?

H_o: $P_{97}=P_{03}$
H_1: $P_{97}\neq P_{03}$

$$P^* = \frac{43+48}{200+250} = .2022$$

So:

$$Z = \frac{\frac{43}{200} - \frac{48}{250}}{\sqrt{.2022(1-.2022)\left(\frac{1}{200} + \frac{1}{250}\right)}} = \frac{.215-.192}{\sqrt{.1613 * .009}} = \frac{.023}{.038} = .6052$$

Thus we would reject the null hypothesis and conclude that there is no evidence that the proportion of uninsured children from low income families has changed.

2. **χ^2 test for the difference between two proportions**

An alternative to the Z test, is the χ^2 (chi-square) test. This starts by laying out a contingency table of the outcomes:

	1997	2003	Total
Successes	43	48	91
Failures	157	202	359
Totals	200	250	450

We want to test the Null: $H_o: P_{1997} = P_{2003}$
Against $H_1: P_{1997} \neq P_{2003}$

The χ^2 test statistic takes the following form:

$$\chi^2 = \sum_{AllCells} \frac{(f_0 - f_e)^2}{f_e}$$

where f_0 is the observed frequency, in a particular cell of the 2x2 table
f_e is the theoretical, or expected frequency in a particular cell if the null hypothesis is true. This test approximately follows a chi-square distribution with 1 degree of freedom.

I will discuss the chi-square distribution in a minute.

To understand what f_e is, assume the null hypothesis is true, then the sample proportions computed from each of the two groups would differ from each other only by chance and would each provide an estimate of the common population parameter, p. In such a situation, a statistic that pools or combines these two separate estimates together into one overall estimate of the population parameter provides more information than either of the two separate estimates could provide by itself.

This statistic, P^*, is:

$$P^* = \frac{x_1 + x_2}{n_1 + n_2}$$

To obtain the expected frequency for each cell pertaining to successes (the cells in the first row of the table), the sample size for a group is multiplied by P^* to obtain the expected frequency for each cell. For each cell in the second row the sample size is multiplied by $(1-P^*)$.

In our example:

$$P^* = \frac{43 + 48}{200 + 250} = .2022$$

	1997 actual	1997 expected	2001 actual	2001 expected	Total
Successes	43	40.44	48	50.55	91
Failures	157	159.56	202	199.45	359

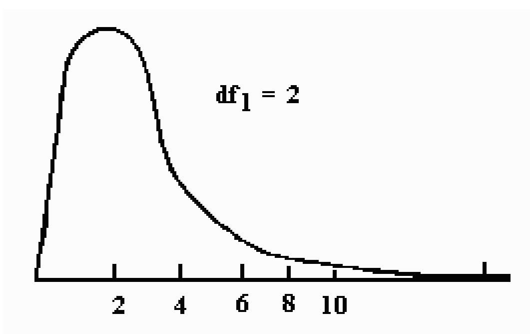
Totals	200	200	250	250	450

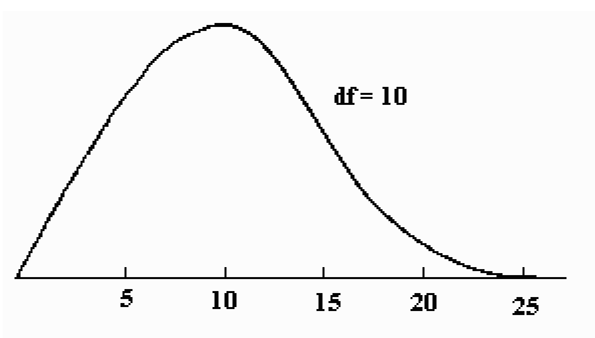
f_0	f_e	(f_0-f_e)	$(f_0-f_e)^2$	$(f_0-f_e)^2/f_e$
43	40.44	2.56	6.55	.1621
48	50.55	-2.55	6.50	.1286
157	159.56	-2.56	6.55	.0411
202	199.45	2.55	6.50	.0326
			Sum=	.3644

This is our test statistic, .3644, note that what this is doing is taking the sum of the squared difference between the actual proportion in each cell from the predicted proportion *assuming the null is true*. So if any of these are "way off" then the test statistic will be "large" while if they are not off the test statistic will be "small".

The question now is what is large and what is small. This is where the χ^2 distribution comes in. It turns out that if we have the sum of n statistically independent standard normal variables, then this sum follows the χ^2 distribution.

The probability density function for the chi-square distribution looks as follows:





The degrees of freedom are calculated as follows:
DF = (r-1)(c-1), where r is the number of rows in the table, and c is then number of columns.
So in our 2x2 table df = 1.

The rejection region is the right hand tail of the distribution, that is we reject H_o if $\chi^2 > \chi^2_1$

In this case the critical value of the we use =CHIDIST(x,df) or =CHIDIST(.3644,1)=.546 to get the pvalue. So if the null is true and the two proportions are equal, then there is a 54 percent chance of getting sample proportions this far from each other. Thus we fail to reject the null hypothesis and conclude that there is no evidence that the two populations have a different rate of uninsured.

The F-test and χ^2 tests will usually give the same answers; they are just two different ways of doing the same thing.

The nice thing about the χ^2 test is it is generalizable to more than two proportions. The logic is exactly the same. I won't go through this; you can do it on your own if you get bored.

V. χ2 test for independence

What we did above was show the χ^2 test for the difference between two (or more) proportions. This can be generalized as a test of independence in the joint responses to two categorical variables. This is something similar to the correlation coefficient (but applied to categorical data) in that it tells you if there is a relationship between two variables.

The null and alternative hypotheses are as follows:
H_o: the two categorical variables are independent (there is no relationship between them)
H_1: The variables are dependent (there is a relationship between them)

We use the same equation as before:

$$\chi^2 = \sum_{AllCells} \frac{(f_0 - f_e)^2}{f_e}$$

Suppose we are interested in the relationship between education and earnings for registered nurses. We collect data for 550 randomly chosen RNs and get the following table:

	Level of Education			
Earnings	High School	Some College	College grad+	Total
< 30k	80	65	44	189
30k-40k	72	34	53	159
40k+	48	31	123	202
Total	200	130	220	550

So our null would be that there is no relationship between education and earnings, and the alternative is that there is a relationship between the two.

To obtain f_e in this case we go back to the probability theory. If the null hypothesis is true, the multiplication rule for independent events can be used to determine the joint probability or proportion of responses expected for any cell combination.

That is, P(A and B) = P(A|B)*P(B) – the joint probability of events A and B is equal to the probability of A, given B, times the probability of B.
So, suppose we want to know the probability of drawing a Heart and a face card from a normal deck of cards, let Heart = A, Face =B, then
P(H and F) = P(H|F)*P(F), there are 4*4 or 16 face cards in a deck, 4 of them hearts so P(H|F) = 4/16 or 1/4. Then the probability of Face is 16/52=4/13.
So P(H and F) = 1/4*4/13 = 4/52 or 1/13.
But note that the events Heart and Face are independent of each other since P(H) = 13/52=1/4 and P(H|F)=1/4. That is, knowing the card is a face card does not change the probability of getting a heart. Then the joint probability formula simplifies to:
P(A and B) = P(A)*P(B) when A and B are independent.

But suppose we had a deck of cards that had an extra king of hearts and the king of spades was missing. Then P(H) ≠ P(H|F). Knowing that the card is a face card would change the probability of drawing a heart.

So back to the example above: assuming independence, if we want to know the probability of being in the top left cell – representing having a high school degree and earning less than 30k – this will be the product of the two separate probabilities:

P(high school and < 30k) = P(high school)*P(<30k)

P(high school) = 200/550 = .3636
P(<30k) = 189/550 = .3436

So if these two guys are independent then P(high school and < 30k)= .3636*.3436 = .1249. This is the predicted proportion.
The observed proportion is 80/550=.1455. Basically we want to compare the actual vs. predicted proportions for each cell.

If the predicted proportion was .1249, we would expect 550*.1249 = 68.7 RNs to be in this cell

To find the expected number in each cell:

f_e = (row total x column total)/n

For this example: 200*189/550 = 68.7

So to calculate the χ^2 statistic:

	f_0	f_e	(f_0-f_e)	$(f_0-f_e)^2$	$(f_0-f_e)^2/f_e$
HS/<30k	80	68.7	11.3	127.7	1.86
HS/30-40	72	57.8	14.2	201.6	3.49
HS/40k+	48	73.5	-25.5	650.3	8.85
Some/<30k	65	44.7	20.3	412.1	9.22
Some/30-40	34	37.6	-3.6	13.0	0.34
Some/40k+	31	47.7	-16.7	278.9	5.85
Coll/<30k	44	75.6	-31.6	998.6	12.21
Coll/30-40	53	63.6	-10.6	112.4	1.77
Coll/40k+	123	80.8	42.2	1780.8	22.04
				Sum=	65.63

In this case we have c=3 and r=3, so there are 2*2 =4 degrees of freedom.

The pvalue is very close to zero. In Excel: =chidist(65.63,2) returns 5.60566E-15
This is scientific notation.

Thus we would reject the null hypothesis and conclude that there is evidence of a relationship between earnings and education among RNs.

Note that just like the correlation coefficient this test does not tell you anything about the relationship between the two variables, just that one appears to exist.