Non-Parametric Tests

MATTHEW FRICKE

Version 1.1. Please send corrections to mfricke@unm.edu

Revisions

- 1.1:
  - Added a reference to a proof that independence of the sample mean and sample variance are necessary and sufficient conditions for the parent distribution to be Gaussian ().
  - Highlighted the typo in the contingency table example. Two figures were transposed (293 instead of 239) so the sum of the first row is incorrect.

Parametric Tests Revisited

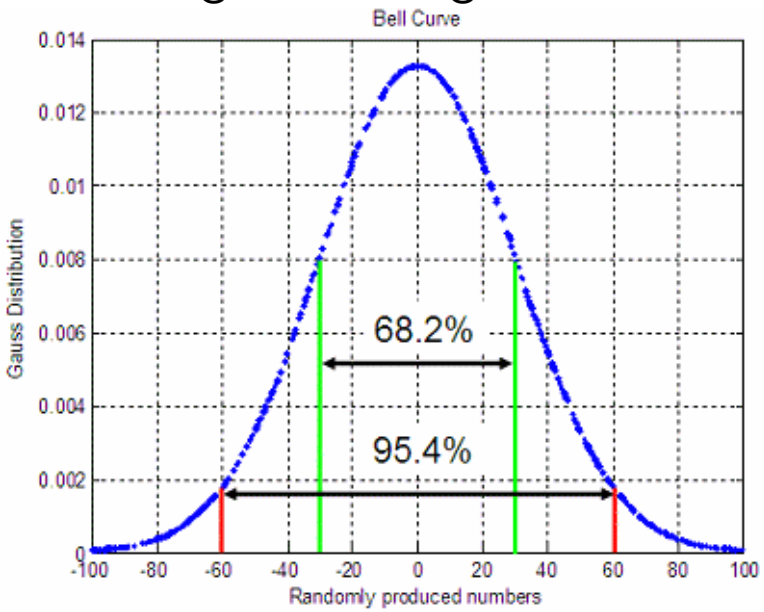
- *Parametric tests assume something about sampling distribution?*

Parametric Tests Revisited

- *Why do we make this assumption?*
- The sample, itself, does not provide enough information for us to estimate how wrong our guesses about the population mean are given the samples
- It gives us a start, but we still have to fill in certain blanks in order to derive the center, spread, and shape of the sampling distribution of the mean. In parametric statistics, we fill in the blanks concerning shape by assuming that the sampling distribution of the mean is normal.
- In the late 1700s de Moivre, Gauss, Laplace, and Maxwell all noticed that the normal distribution was showing up everywhere.

Interstate Road Map

Each city is a node, each interstate is
An edge. The degree distribution is Normal



$$P(k) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(k-\mu)^2/2\sigma^2}$$

$$P(k) \approx e^{-k^2}$$

Parametric Tests Revisited

- ***Why do we assume that the sampling distribution of the mean is normal, as opposed to some other shape?***
- The **First Known Property** of the Normal Distribution says that: *given random and independent samples of N observations each (taken from a normal distribution), the distribution of sample means is normal and unbiased (i.e., centered on the mean of the population), regardless of the size of N*.
- If a population has finite variance σ^2 and a finite mean μ and is normally distributed, then the distribution of sample means (from an infinite set of independent samples of *N* independent observations each) must be normally distributed (with variance σ^2/N and mean μ), regardless of the size of *N*.

Parametric Tests Revisited

- ***Why do we assume that the sampling distribution of the mean is normal, as opposed to some other shape?***
- Therefore, if the population distribution is normal, then even small N will produce a sampling distribution of the mean that is normal (by the First Known Property). As the population is made less and less normal (e.g., by adding in a lot of skew and/or messing with the kurtosis), a larger and larger *N* will be required.
- In general, the Central Limit Theorem "kicks in" at an N of about 30.
- As long as the sample is based on 30 or more observations, the sampling distribution of the mean can be "safely" assumed to be normal.

Parametric Tests Revisited

- ***Why do we assume that the sampling distribution of the mean is normal, as opposed to some other shape?***
- Worst-case scenario (i.e., a population distribution that is the farthest from normal); this is the exponential.
- If the population has an exponential distribution, how big does *N* have to be in order for the sampling distribution of the mean to be close enough to normal for practical purposes?
- Extensive computer simulation suggests that 30 is a good value.
- It has not been proved that 30 is sufficient; this rule-of-thumb was developed by having a computer perform Monte Carlo simulations for a couple of months.

Parametric Tests Revisited

- ***Why do we assume that the sampling distribution of the mean is normal, as opposed to some other shape? Let's review:***
- 1. Parametric statistics work by making an assumption about the shape of the sampling distribution of the characteristic of interest; the particular assumption that all of our parametric stats make is that the sampling distribution of the mean is normal.
  - i.e. We assume that if we took a whole bunch of samples, calculated the mean for each, and then made a plot of these values, the distribution of these means would be normal.)
- 2. As long as the sample size, *N*, is at least 30 and we're making an inference about the mean, then this assumption is true (by Central Limit Theory plus some simulations) ☺

Parametric Tests Revisited

- ***Why do we assume that the sampling distribution of the mean is normal, as opposed to some other shape?***
- The remaining problem is this: we want to make the same assumption(s) for all of our inferences even when we sometimes use samples that are smaller than 30.
- Parametric methods might not be warranted when samples are small.

Parametric Tests Revisited

- ***Why do we assume that the sampling distribution of the mean is normal, as opposed to some other shape?***
- *We are always safe for any N if the underlying distribution is normal.*
- *Researchers often "hope" this is true.*

Parametric Tests Revisited

- ***Why do we assume that the sampling distribution of the mean is normal, as opposed to some other shape?***
- *Thankfully there is another reason to believe the distribution is normal.*
- The **Second Known Property** of the Normal Distribution says that: *given random and independent observations (from a normal distribution), the sample mean and sample variance are independent*[†].
- *The only distribution with this property is the Gaussian distribution*[†].
- *But really – you should perform a normality test, such as the Shapiro-Wilk test*, if you are going to use a parametric test.*

*Shapiro, Samuel Sanford, and Martin B. Wilk. "An analysis of variance test for normality (complete samples)." *Biometrika*52.3/4 (1965): 591-611.

[†]Lukacs, Eugene. "A Characterization of the Normal Distribution." The Annals of Mathematical Statistics, Vol. 13, No. 1 (Mar., 1942), pp. 91-93

Contingency Table Test

- Consider qualitative data such as gender, occupation, party affiliation.
- What does the mean and standard deviation mean for this kind of data?
- Nothing.
- But there is variability defined by proportions. We can look at the size of a subset compared to the whole set. (e.g. how many males vs females in the class.)
- We still want to ask the same sort of question we have been asking throughout the course. Is there an association between two distributions?

Contingency Table Test

- Consider qualitative data such as gender, occupation, party affiliation.
- We still want to ask the same sort of question we have been asking throughout the course. Is there an association between two distributions?
- To tackle this for qualitative data (which are certainly not parametric) we use contingency tables.

- Consider a number of outcomes, O. For example an outcome might be being (tall, hairy), (short, bald). Recovered and treated, or recovered and not treated, or not-recovered and not-treated, or treated and not-recovered.

Contingency Table Test

- Consider qualitative data such as gender, occupation, party affiliation.
- Once we have estimated the probabilities we can compare them to the expected probabilities for independent random variables.

A contingency table with r rows and c columns representing two variables R and S with r and c possible outcomes, respectively.

	1	$\cdots$	c	$+$
1	O_{11}	$\cdots$	O_{1c}	O_{1+}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
r	O_{r1}	$\cdots$	O_{rc}	O_{r+}
$+$	O_{+1}	$\cdots$	O_{+c}	N

Contingency Table Test

- Consider qualitative data such as gender, occupation, party affiliation.
- Once we have estimated the probabilities we can compare them to the expected probabilities for independent random variables.
- The idea is simple. If there is no dependency or association between two distributions then the probabilities are independent and obey the multiplication rule: $P(R{=}i, C{=}j) = P(R{=}i)P(C{=}j)$
- This is easy to check by looking at the relative frequencies to estimate the probabilities.

$$P(R{=}i, C{=}j){=}O_{ij}/N \qquad P(C{=}j){=}O_{+j}/N$$

$$P(R{=}i){=}O_{i+}/N$$

Contingency Table Test

- Consider qualitative data such as gender, occupation, party affiliation.
- Once we have estimated the probabilities we can compare them to the expected probabilities for independent random variables.
- The idea is simple. If there is no dependency or association between two distributions then the probabilities are independent and obey the multiplication rule: $P(R{=}i, C{=}j) = P(R{=}i)P(C{=}j)$
- This is easy to check by looking at the relative frequencies to estimate the probabilities.
- Expected probabilities for independent random variables:

$$E_{ij} = O_{i+}O_{+j}/N$$

Contingency Table Test

- Consider qualitative data such as gender, occupation, party affiliation.
- Expected probabilities for independent random variables:

$$E_{ij} = O_{i+}O_{+j}/N$$

- Define a test statistic that measures how far the distributions are from the expected independent proportions:

$$Y = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Contingency Table Test

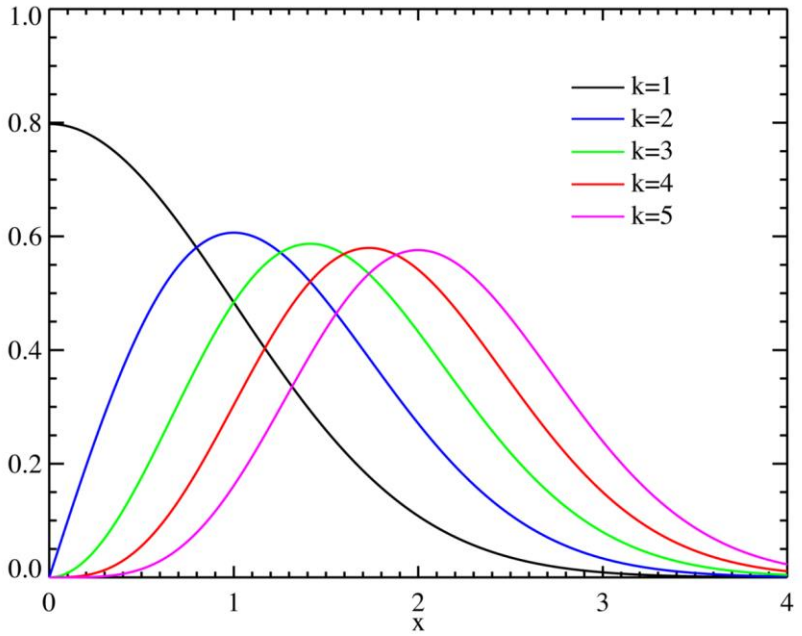
- So now we have this test statistic 'Y'
- We think it makes sense as a way to tell if qualitative data distributions are independent or not.
- Here we resort to the idea that things tend to be normal. Not the outcomes that wouldn't make sense.
- We are going to appeal to the distribution of $\frac{O_{ij} - E_{ij}}{E_{ij}}$ being pretty normal if we have enough outcomes. The argument is that these are counts of discrete events on a finite interval and are therefor well modelled by a Poisson distribution.

Contingency Table Test

Do you recall the distribution we have
Seen earlier in the class is the sum of
Squared normal deviates?

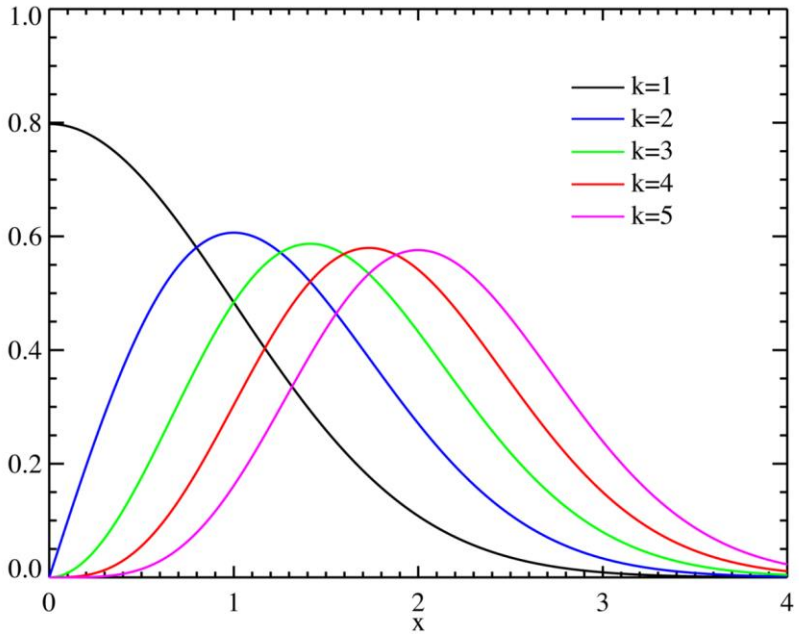
Contingency Table Test

Do you recall the distribution we have
Seen earlier in the class is the sum of
Squared normal deviates?



Contingency Table Test

Do you recall the distribution we have
Seen earlier in the class is the sum of
Squared normal deviates?

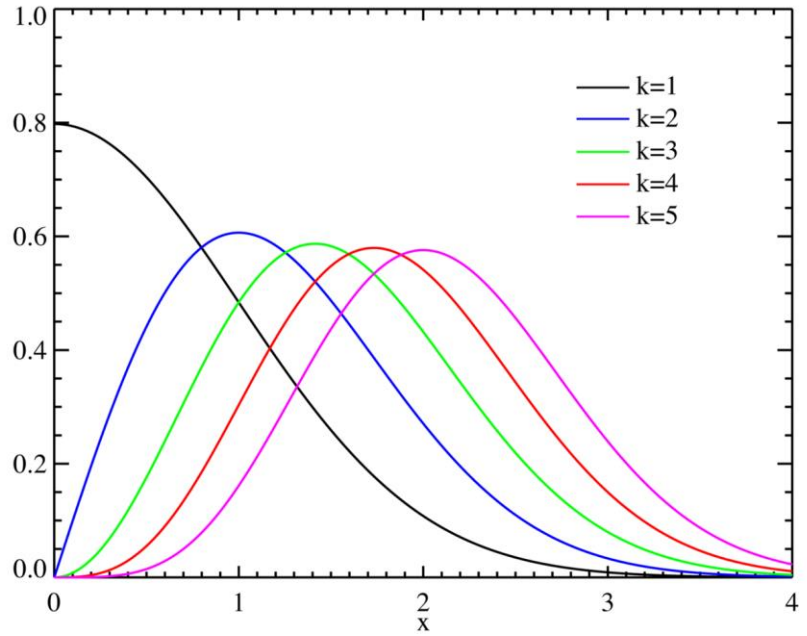


χ^2

Contingency Table Test

Do you recall the distribution we have Seen earlier in the class is the sum of Squared normal deviates?

$$Y = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2$$



Contingency Table Test

- As we have just seen, the χ^2 test is an approximate test
- (e.g. approximating Poisson variables by Normal variables). This approximation requires large samples to be good. In practice, the expected frequencies should be above 5, although it is common to allow 20% exceptions, but none of these expected frequencies may be less than 1.

Contingency Table Test

- Now we have everything we need to perform the Contingency Table Test:
  - A way to map divergence from the expected probabilities of independent qualitative distributions to a distribution for which we can define a significance level, the χ^2 distribution.

Contingency Table Test

- The algorithm:
- Null hypothesis H_0 : No association, i.e. for all i and j, $P(R{=}i, C{=}j) = P(R{=}i)P(C{=}j)$
- Alternative hypothesis H1 : Association
- Significance level: 5%
- Input Outcomes: $O_{ij},\ i = 1, \ldots, r,\ j = 1, \ldots, c$
- Step 1: Calculate the expected proportions if H_0 is true $E_{ij} = O_{i+}O_{+j}/N$
- Step 2: Calculate the test statistic: $Y = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2$
- Accept H_0 iff $Y \leq \chi^2_{\text{crit}}$ (from table)

Contingency Table Test

- The algorithm:
- Null hypothesis H_0 : No association, i.e. for all i and j, $P(R{=}i, C{=}j) = P(R{=}i)P(C{=}j)$
- Alternative hypothesis H1 : Association
- Significance level: 5%
- Input Outcomes: $O_{ij},\ i = 1, \ldots, r,\ j = 1, \ldots, c$
- Step 1: Calculate the expected proportions if H_0 is true
- Step 2: Calculate the test statistic: $E_{ij} = O_{i+}O_{+j}/N$
- Accept H_0 iff $Y \leq \chi^2_{\text{crit}}$ (from table) $Y = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2$
- **We need to know the degrees of freedom to use χ^2**

Example One

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	15	24	25	293	303
Mediterranean	7	14	8	273	302
Total	22	38	33	512	605

Should have been 239.

Despite this the conclusion is the same.

de Lorgeril, Michel, et al. "Mediterranean dietary pattern in a randomized trial: prolonged survival and possible reduced cancer rate." *Archives of Internal Medicine* 158.11 (1998): 1181-1187.

Example One

$$E_{ij} = O_{i+}O_{+j}/N$$

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA					303
Mediterranean					302
Total	22	38	33	512	605

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	15	24	25	293	303
Mediterranean	7	14	8	273	302
Total	22	38	33	512	605

Example One

$$E_{ij} = O_{i+}O_{+j}/N$$

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA					303
Mediterranean					302
Total	22	38	33	512	605

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	15	24	25	293	303
Mediterranean	7	14	8	273	302
Total	22	38	33	512	605

Example One

$$E_{ij} = O_{i+}O_{+j}/N$$

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	$\frac{22 \times 303}{?}$				303
Mediterranean					302
Total	22	38	33	512	605

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	15	24	25	293	303
Mediterranean	7	14	8	273	302
Total	22	38	33	512	605

Example One

$$E_{ij} = O_{i+}O_{+j}/N$$

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	$\frac{22 \times 303}{605}$				303
Mediterranean					302
Total	22	38	33	512	605

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	15	24	25	293	303
Mediterranean	7	14	8	273	302
Total	22	38	33	512	605

Example One

$$E_{ij} = O_{i+}O_{+j}/N$$

	Outcome				
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	$\frac{22 \times 303}{605}$	$\frac{38 \times 303}{605}$	$\frac{33 \times 303}{605}$	$\frac{512 \times 303}{605}$	303
Mediterranean	$\frac{22 \times 302}{605}$	$\frac{38 \times 302}{605}$	$\frac{33 \times 302}{605}$	$\frac{512 \times 302}{605}$	302
Total	22	38	33	512	605

	Outcome				
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	15	24	25	293	303
Mediterranean	7	14	8	273	302
Total	22	38	33	512	605

Example One

$$E_{ij} = O_{i+}O_{+j}/N$$

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	11.02	19.03	16.53	256.42	303
Mediterranean	10.98	18.97	16.47	255.58	302
Total	22	38	33	512	605

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	15	24	25	293	303
Mediterranean	7	14	8	273	302
Total	22	38	33	512	605

Example One

$$E_{ij} = O_{i+}O_{+j}/N$$

Calculate $Y = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$

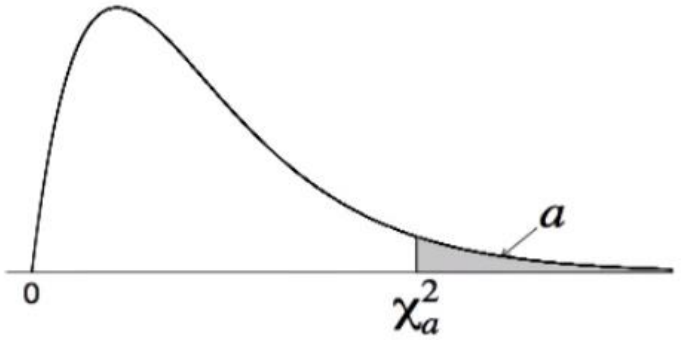
$$\frac{(15-11.02)^2}{11.02} + \frac{(24-19.03)^2}{19.03} + \frac{(25-16.53)^2}{16.53} + \frac{(293-256.42)^2}{256.42}$$
$$+ + \frac{(7-10.98)^2}{10.98} + \frac{(14-18.97)^2}{18.97} + \frac{(8-16.47)^2}{16.47}$$
$$+ \frac{(273-255.58)^2}{255.58} = 16.55 \backsim \chi^2_{crit}$$ with 3 degrees of freedom: $(r-1)(c-1) = 1 \times 3 = 3$

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	11.02	19.03	16.53	256.42	303
Mediterranean	10.98	18.97	16.47	255.58	302
Total	22	38	33	512	605

Outcome					
Diet	Cancer	Fatal Heart	Non-Fatal Heart	Healthy	Total
AHA	15	24	25	293	303
Mediterranean	7	14	8	273	302
Total	22	38	33	512	605

Example One

$16.55 \backsim \chi^2_{crit}$ with 3 degrees of freedom: $(r-1)(c-1) = 1 \times 3 = 3$



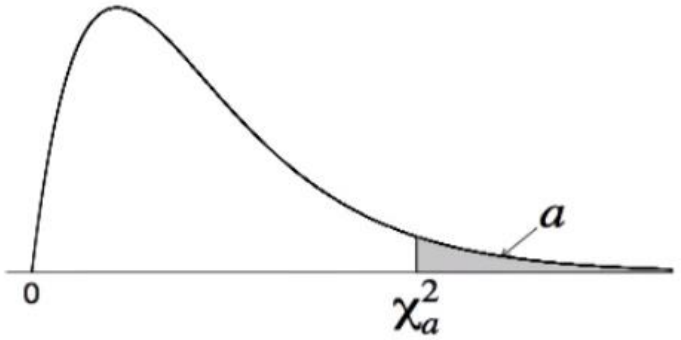
16.55

df	$\chi^2_{0.20}$	$\chi^2_{0.15}$	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$	$\chi^2_{0.01}$	$\chi^2_{0.005}$	$\chi^2_{0.001}$	$\chi^2_{.0005}$
1	1.642	2.072	2.706	3.841	5.024	6.635	7.879	10.828	12.116
2	3.219	3.794	4.605	5.991	7.378	9.210	10.597	13.816	15.202
3	4.642	5.317	6.251	7.815	9.348	11.345	12.838	16.266	17.730
4	5.989	6.745	7.779	9.488	11.143	13.277	14.860	18.467	19.997
5	7.289	8.115	9.236	11.070	12.833	15.086	16.750	20.515	22.105
6	8.558	9.446	10.645	12.592	14.449	16.812	18.548	22.458	24.103

Example One

$16.55 \backsim \chi^2_{crit}$ with 3 degrees of freedom:
$(r-1)(c-1) = 1 \times 3 = 3$

So H_0 is rejected with a p-value < 0.001.



df	$\chi^2_{0.20}$	$\chi^2_{0.15}$	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$	$\chi^2_{0.01}$	$\chi^2_{0.005}$	$\chi^2_{0.001}$	$\chi^2_{.0005}$
1	1.642	2.072	2.706	3.841	5.024	6.635	7.879	10.828	12.116
2	3.219	3.794	4.605	5.991	7.378	9.210	10.597	13.816	15.202
3	4.642	5.317	6.251	7.815	9.348	11.345	12.838	16.266	17.730
4	5.989	6.745	7.779	9.488	11.143	13.277	14.860	18.467	19.997
5	7.289	8.115	9.236	11.070	12.833	15.086	16.750	20.515	22.105
6	8.558	9.446	10.645	12.592	14.449	16.812	18.548	22.458	24.103

Example One

What does that mean?

If the null hypothesis associated with the null model that the outcomes (health status and diet) are independent is true, then we would expect to see this sample less than 10 times in 10,000 trials.

The (slightly imprecise) implication is that the alternate hypothesis is probably true, and diet does effect health.

Example Two

Eye Color					
Gender	Blue	Brown	Green	Hazel	Total
Female	370	352	198	187	1107
Male	359	290	110	160	919
Total	729	642	308	347	2026

Froelich, Amy G., and W. Robert Stephenson. "Does Eye Color Depend on Gender? It Might Depend on Who or How You Ask." *Journal of Statistics Education* 21.2 (2013).

Confirmation Bias

- Consider why there were objections to the conclusions from our analysis of Example 2, but none raised after our analysis of Example 1.
- That diet impacts health was the result we expected. That eye color is gender associated was not expected, so we were encouraged to dig deeper to find an error.
- This is particularly a problem in computer science. When a simulation or algorithm does what we expect we tend to stop debugging. When it has an unexpected result we put a lot more effort into finding the “bug”.

The Sign Test

- **Motivation:** If we want to test whether the means of two distributions are the same without assuming normality we can make use of the fact that half the area under the distribution will be less than the mean and half will be greater than the mean.
- **Data distribution:** The Sign test is a non–parametric test, so we do not assume that the data is normally distributed.
- **One sample:** We test to see if the number of outcomes less than a hypothesized mean is equal to the number of outcomes greater than the that mean.
- **Two sample:** Pair up outcomes from samples S1 and S2 and see if the number of times $S1_i < S2_i$ equals the number of times $S1_i < S2_i$, discard $S1_i = S2_i$
- $\mathbf{H_0}$ is that there are an equal number of + and –, i.e. the population means are equal to the sample mean.

The Sign Test

- As with the Contingency Tables method we have come up with a reasonable measure of similarity between distributions.
- And again we have to map this into a probability distribution so we can understand the probability of seeing particular values for this measure.
- In the sign test this is the binomial distribution.
- Why?

The Sign Test

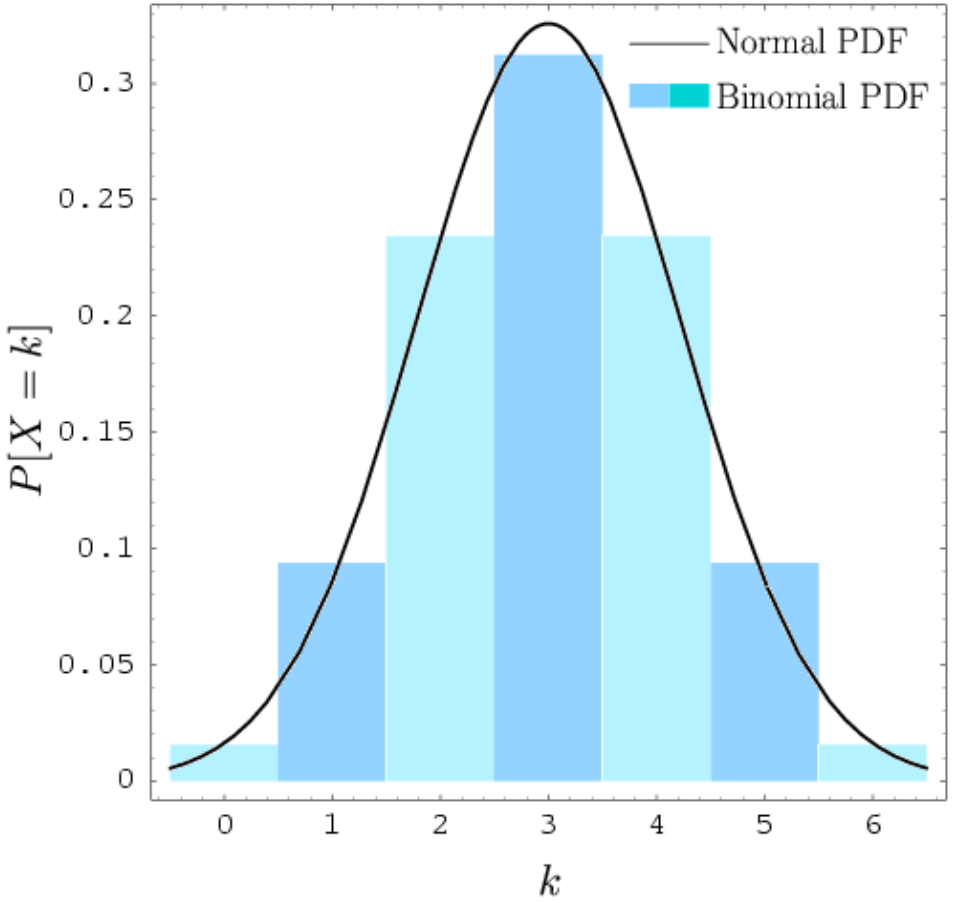
- As with the Contingency Tables method we have come up with a reasonable measure of similarity between distributions.
- And again we have to map this into a probability distribution so we can understand the probability of seeing particular values for this measure.
- In the sign test this is the binomial distribution.
- Why?
- Recall your discrete math class (CS261 at UNM). Because the **bi**nomial distribution is a measure of the probability of events when there are only 2 possibilities. (here greater than and less than).

The Sign Test

The binomial distribution tells us the probability of x "successful" outcomes in a sample
of size N, where p=0.5 is the probability of success.

$$P(x) = \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x}$$

Note: the binomial distribution converges to the normal distribution for large N.
As a practical matter when p=0.5 you need N>20.
Usually you can use an exact table or binomial calculator.



Example

A new chemotherapy treatment is proposed for patients with breast cancer. Investigators are concerned with patient's ability to tolerate the treatment and assess their quality of life both before and after receiving the new chemotherapy treatment. Quality of life (QOL) is measured on an ordinal scale and for analysis purposes, numbers are assigned to each response category as follows: 1=Poor, 2= Fair, 3=Good, 4= Very Good, 5 = Excellent. The data are shown below.

Patient	QOL Before Chemotherapy Treatment	QOL After Chemotherapy Treatment
1	3	2
2	2	3
3	3	4
4	2	4
5	1	1
6	3	4
7	2	4
8	3	3
9	2	1
10	1	3
11	3	4
12	2	3

Example

The question of interest is whether there is a difference in QOL after chemotherapy treatment as compared to before.

Step 1. Set up hypotheses and determine level of significance.

H_0: The median difference is zero versus

H_1: The median difference is not zero $\alpha=0.05$

Step 2. Select the appropriate test statistic.

The test statistic for the Sign Test is the smaller of the number of positive or negative signs.

Step 3. Set up the decision rule.

This will be based on the binomial distribution, our hypothesized p-value, and the number of samples.

Step 4. Compute the test statistic.

Example

Patient	QOL Before Chemotherapy Treatment	QOL After Chemotherapy Treatment	Difference (After-Before)	Sign
1	3	2	-1	-
2	2	3	1	+
3	3	4	1	+
4	2	4	2	+
5	1	1	0	-
6	3	4	1	+
7	2	4	2	+
8	3	3	0	+
9	2	1	-1	-
10	1	3	2	+
11	3	4	1	+
12	2	3	1	+

There are 3 negative and 9 positive outcomes. But is it significant at p=0.05 given only 12 outcomes?

Example

There are 3 negative and 9 positive outcomes. But is it significant at p=0.05 given only 12 outcomes?

Because we care
About positive and negative
Changes (two-tailed) and the
Binomial is symmetric we can just double P(x)

Patient	QOL Before Chemotherapy Treatment	QOL After Chemotherapy Treatment	Difference (After-Before)	Sign
1	3	2	-1	-
2	2	3	1	+
3	3	4	1	+
4	2	4	2	+
5	1	1	0	-
6	3	4	1	+
7	2	4	2	+
8	3	3	0	+
9	2	1	-1	-
10	1	3	2	+
11	3	4	1	+
12	2	3	1	+

$$P(x) = \frac{N!}{x!(N-x)!}p^x(1-p)^{12-3} = \frac{12!}{3!(12-3)!}\left(\frac{1}{2}\right)^3\left(1-\frac{1}{2}\right)^{12-3}$$

Example

There are 3 negative and 9 positive outcomes. But is it significant at p=0.05 given only 12 outcomes?

Because we care
About positive and negative
Changes (two-tailed) and the
Binomial is symmetric we can just double P(x).

Patient	QOL Before Chemotherapy Treatment	QOL After Chemotherapy Treatment	Difference (After-Before)	Sign
1	3	2	-1	-
2	2	3	1	+
3	3	4	1	+
4	2	4	2	+
5	1	1	0	-
6	3	4	1	+
7	2	4	2	+
8	3	3	0	+
9	2	1	-1	-
10	1	3	2	+
11	3	4	1	+
12	2	3	1	+

$$P(x) = \frac{N!}{x!(N-x)!}p^x(1-p)^{12-3} = \frac{12!}{3!(12-3)!}\left(\frac{1}{2}\right)^3\left(1-\frac{1}{2}\right)^{12-3}$$ x 2

Example

There are 3 negative and 9 positive outcomes. But is it significant at p=0.05 given only 12 outcomes?

Patient	QOL Before Chemotherapy Treatment	QOL After Chemotherapy Treatment	Difference (After-Before)	Sign
1	3	2	-1	-
2	2	3	1	+
3	3	4	1	+
4	2	4	2	+
5	1	1	0	-
6	3	4	1	+
7	2	4	2	+
8	3	3	0	+
9	2	1	-1	-
10	1	3	2	+
11	3	4	1	+
12	2	3	1	+

$$P(x) = \frac{12!}{3!(12-3)!}\left(\frac{1}{2}\right)^3\left(1-\frac{1}{2}\right)^{12-3} + \frac{12!}{2!(12-2)!}\left(\frac{1}{2}\right)^2\left(1-\frac{1}{2}\right)^{12-2} + \frac{12!}{1!(12-1)!}\left(\frac{1}{2}\right)^1\left(1-\frac{1}{2}\right)^{12-1}$$

x 2

Example

There are 3 negative and 9 positive outcomes. But is it significant at p=0.05 given only 12 outcomes?

Patient	QOL Before Chemotherapy Treatment	QOL After Chemotherapy Treatment	Difference (After-Before)	Sign
1	3	2	-1	-
2	2	3	1	+
3	3	4	1	+
4	2	4	2	+
5	1	1	0	-
6	3	4	1	+
7	2	4	2	+
8	3	3	0	+
9	2	1	-1	-
10	1	3	2	+
11	3	4	1	+
12	2	3	1	+

$$P(x) = \frac{129}{2048} = 0.0727$$

$$P(x) = \frac{12!}{3!(12-3)!}\left(\frac{1}{2}\right)^3\left(1-\frac{1}{2}\right)^{12-3} + \frac{12!}{2!(12-2)!}\left(\frac{1}{2}\right)^2\left(1-\frac{1}{2}\right)^{12-2} + \frac{12!}{1!(12-1)!}\left(\frac{1}{2}\right)^1\left(1-\frac{1}{2}\right)^{12-1}$$

x 2

Example

There are 3 negative and 9 positive outcomes. But is it significant at p=0.05 given only 12 outcomes?

$$P(x \leq 3) = \frac{129}{2048} = 0.0727$$ x 2

$$2 \times P(x \leq 3) = 0.1454$$

Patient	QOL Before Chemotherapy Treatment	QOL After Chemotherapy Treatment	Difference (After-Before)	Sign
1	3	2	-1	-
2	2	3	1	+
3	3	4	1	+
4	2	4	2	+
5	1	1	0	-
6	3	4	1	+
7	2	4	2	+
8	3	3	0	+
9	2	1	-1	-
10	1	3	2	+
11	3	4	1	+
12	2	3	1	+

Example

There are 3 negative and 9 positive outcomes. But is it significant at p=0.05 given only 12 outcomes?

Patient	QOL Before Chemotherapy Treatment	QOL After Chemotherapy Treatment	Difference (After-Before)	Sign
1	3	2	-1	-
2	2	3	1	+
3	3	4	1	+
4	2	4	2	+
5	1	1	0	-
6	3	4	1	+
7	2	4	2	+
8	3	3	0	+
9	2	1	-1	-
10	1	3	2	+
11	3	4	1	+
12	2	3	1	+

$$P(x \leq 3) = \frac{129}{2048} = 0.0727 \text{ x } 2$$

$$2 \times P(x \leq 3) = 0.1454$$

Step 3. Conclusion: We fail to reject the H_0 that the distributions are the same at the 0.05 significance level.

Wilcoxon Signed-Rank Test

- What if we have quantitative values not just binary choices?
- Instead of just looking at the category (>,<) we can add the direction and magnitude of the differences.
- The Wilcoxon signed-rank test statistic, W converges to a Normal distribution for large samples. For small samples we refer to a table of values.

Example

- A study is run to evaluate the effectiveness of an exercise program in reducing systolic blood pressure in patients. The blood pressure of 15 participants is measured before and after the exercise program:

Patient	Systolic Blood Pressure Before Exercise Program	Systolic Blood Pressure After Exercise Program
1	125	118
2	132	134
3	138	130
4	120	124
5	125	105
6	127	130
7	136	130
8	139	132
9	131	123
10	132	128
11	135	126
12	136	140
13	128	135
14	127	126
15	130	132

Example

- A study is run to evaluate the effectiveness of an exercise program in reducing systolic blood pressure in patients. The blood pressure of 15 participants is measured before and after the exercise program:

Example

- A study is run to evaluate the effectiveness of an exercise program in reducing systolic blood pressure in patients. The blood pressure of 15 participants is measured before and after the exercise program:

Question: Is there is a difference in systolic blood pressures after participating in the exercise program as compared to before?

•**Step1.** Set up hypotheses and determine level of significance.

H_0: The median difference is zero versus, H_1: The median difference is not zero $\alpha=0.05$

•**Step 2.** Select the appropriate test statistic.

The test statistic for the Wilcoxon Signed Rank Test is W, defined as the smaller of W+ and W- which are the sums of the positive and negative ranks, respectively.

•**Step 3.** Set up the decision rule.

The critical value of W can be found in the table of critical values. To determine the appropriate critical value from the table we need sample size (n=15) and our two-sided level of significance ($\alpha=0.05$). The critical value for this two-sided test with n=15 and $\alpha=0.05$ is 25 and the decision rule is as follows: Reject H_0 if $W \leq 25$.

•**Step 4.** Compute the test statistic.

Example

- A study is run to evaluate the effectiveness of an exercise program in reducing systolic blood pressure in patients. The blood pressure of 15 participants is measured before and after the exercise program:

Patient	Systolic Blood Pressure Before Exercise Program	Systolic Blood Pressure After Exercise Program	Difference (Before-After)
1	125	118	7
2	132	134	-2
3	138	130	8
4	120	124	-4
5	125	105	20
6	127	130	-3
7	136	130	6
8	139	132	7
9	131	123	8
10	132	128	4
11	135	126	9
12	136	140	-4
13	128	135	-7
14	127	126	1
15	130	132	-2

Example

- **Now we rank the absolute values of the differences.** Specifically, we assign ranks from 1 through n to the smallest through largest absolute values of the difference scores. When there is a tie we take the mean value. Then set the appropriate sign.

Observed Differences	Ordered Absolute Values of Differences	Ranks
7	1	1
-2	-2	-2.5
8	-2	-2.5
-4	-3	-4
20	-4	-6
-3	-4	-6
6	4	6
7	6	8
8	-7	-10
4	7	10
9	7	10
-4	8	12.5
-7	8	12.5
1	9	14
-2	20	15

Example

- **Now we rank the absolute values of the differences.** Specifically, we assign ranks from 1 through n to the smallest through largest absolute values of the difference scores. When there is a tie we take the mean value. Then set the appropriate sign.

Next: Sum the positive ranks (W+) and negative ranks (W-).

$$W+ = 89$$

$$W- = 31$$

Observed Differences	Ordered Absolute Values of Differences	Ranks
7	1	1
-2	-2	-2.5
8	-2	-2.5
-4	-3	-4
20	-4	-6
-3	-4	-6
6	4	6
7	6	8
8	-7	-10
4	7	10
9	7	10
-4	8	12.5
-7	8	12.5
1	9	14
-2	20	15

Example

- **Finally we consult a Wilcoxon table to find out to which p-value our test statistic W corresponds:**

Next: Sum the positive ranks (W+) and negative ranks (W-).

$$W+ = 89$$

$$W- = 31$$

Table of Critical Values for the Wilcoxon Sign Rank Test

Compare the smaller value obtained from the Wilcoxon test to the critical values in the table below.

Select the row corresponding to your sample size (N)

Your obtained value must be **equal to, or smaller than**, the critical value in order to be statistically significant at the given level.

	Two-Tailed Significance Level		
N	0.05	0.02	0.01
6	0	-	-
7	2	0	-
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14		16	13
15	25	20	16
16		24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61

Example

- **Finally we consult a Wilcoxon table to find out to which p-value our test statistic W corresponds:**

$$W = 31, 31 > 25$$

We cannot reject H_0 at the 0.05 significance level.

We cannot conclude that the changes
In blood pressure were statistically significant.

Table of Critical Values for the Wilcoxon Sign Rank Test

Compare the smaller value obtained from the Wilcoxon test to the critical values in the table below.

Select the row corresponding to your sample size (N)

Your obtained value must be **equal to, or smaller than**, the critical value in order to be statistically significant at the given level.

	Two-Tailed Significance Level		
N	0.05	0.02	0.01
6	0	-	-
7	2	0	-
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61