# BIO 682
# Nonparametric Statistics
# Spring 2010

Steve Shuster

http://www4.nau.edu/shustercourses/BIO682/index.htm

Lecture 2

## Getting it Right

| | Accepted | Rejected |
|---|---|---|
| **True** | Correct | Type I |
| **False** | Type II | Correct |

## What Does This Really Mean?

1. This concept is easier to understand with an example that shows:

a. Relative frequency distributions for samples deviating from null hypothesis by chance.

## So Let Us Suppose,

1. You are going to sample 17 insects from a large population.
2. What is the probability of drawing 3 females and 14 males?
3. In other words, how likely am I to draw this combination, *by chance alone*?

## If The Population Is Large,

1. It is reasonable to assume a 50:50 sex ratio on any individual draw.
2. Again, this is a ***null hypothesis***, and therefore:

a. The chance of getting female ($p$) is .5

b. The chance of getting male ($q$) is .5

c. And $(1 - p) = q$, because $(p + q) = 1$

c. Then, the probability distribution is generated by ***binomial expansion***.

## Binomial Expansion

1. If the population is large, the probability of obtaining 1-17 females can be found using:

$$\boldsymbol{(p + q)^N = 1}$$

where $N = 17$ draws from the population.

## Binomial Expansion

2. Example with smaller *N*:

a. If we let $N = 5$, then by expanding the equation, we obtain the probabilities of obtaining different combinations of female and male individuals:

$$(p + q)^5 = 1$$

$$p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5 = 1$$

## Binomial Expansion

$$p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5 = 1$$

$p^5$ = all 5 female

$5p^4q$ = 4 female, 1 male

$10p^3q^2$ = 3 female, 2 male

$10p^2q^3$ = 2 female, 3 male

$5pq^4$ = 1 female, 4 male

$q^5$ = all male

## Alternatively,

1. We can figure out the exact probability:

a. Let $k$ = the sum of counts of one class (# of females = 3.

b. Let $N$ = total number of opportunities to choose = 17.

c. The number of objects in $k$ and in $N$-$k$ is given by the equation for a binomial distribution.

## The Binomial Equation

$$P[k] = \binom{N}{k} p^k q^{N-k}$$

Where,

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}$$

## The Binomial Equation

1. With $p=q=.5$ and $k=3$, $N=17$, the exact probability is .0052, or .5% of the time.

   a. How likely is that?

   b. To find out, we need to figure out total distribution.

## Expected frequencies

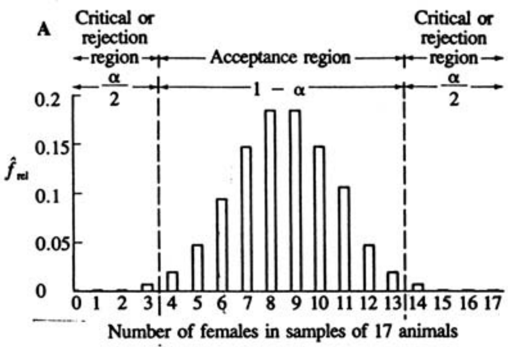
Note that each value in Column 3 is the expected exact probability.

We will see that when we are testing hypotheses, we need to add this value to *all more extreme* values; more on this later.

**Table 7.7** Relative expected frequencies for samples of 17 animals

| (1) ♀♀ | (2) ♂♂ | (3) $H_0: p_♀ = q_♂ = \frac{1}{2}$ $f_{rel}$ |
|---|---|---|
| 17 | 0 | 0.000,007,6 |
| 16 | 1 | 0.000,129,7 |
| 15 | 2 | 0.001,037,6 |
| 14 | 3 | 0.005,188,0 |
| 13 | 4 | 0.018,158,0 |
| 12 | 5 | 0.047,210,7 |
| 11 | 6 | 0.094,421,4 |
| 10 | 7 | 0.148,376,5 |
| 9 | 8 | 0.185,470,6 |
| 8 | 9 | 0.185,470,6 |
| 7 | 10 | 0.148,376,5 |
| 6 | 11 | 0.094,421,4 |
| 5 | 12 | 0.047,210,7 |
| 4 | 13 | 0.018,158,0 |
| 3 | 14 | 0.005,188,0 |
| 2 | 15 | 0.001,037,6 |
| 1 | 16 | 0.000,129,7 |
| 0 | 17 | 0.000,007,6 |
| Total | | 1.000,000,2 |

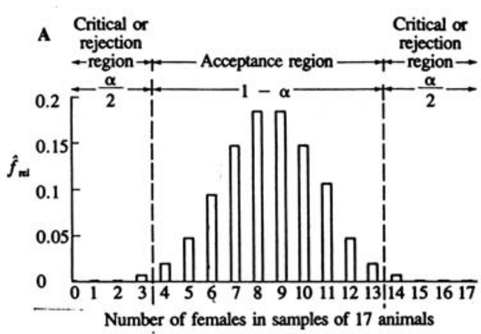
## The Distribution of Expected Values




## Note That:

a. The region of $H_o$ acceptance is 1 - α

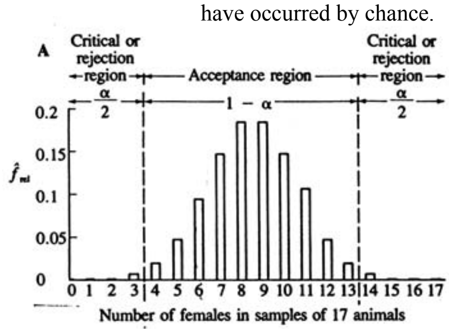
b. The tails include α/2




## Also Note:

c. Decreasing α *increases* the acceptance region.

d. Area in tails are cases too deviant to have occurred by chance.



## Statistical Significance

1. Usually, $\alpha < 0.01$ is considered significant.
2. Between .05 and .01 is up to experimenter
3. > .05 is difficult to justify.

- Why? Because you are making it easier to see "statistical significance."
- It gives the (justifiable) impression that you *want* to assign importance to deviant values.

## Example

1. Two methods of instruction are compared using pre- and post-method questionnaires.
2. The average attitude scores toward science in general, toward math anxiety, toward self-esteem gained by learning more about science, are different with $p = 0.07$.
3. The experimenter raises his $\alpha$ to 0.10, and concludes that his methods were a success.

## However,

1. Type I errors occur when observed frequencies *actually did* occur by chance.
2. If only happens 10% of time, this might be acceptable from a scientific point of view.
3. In general, 5% error is considered sufficiently conservative.
4. Thus, at least 95% of the time ($=1-\alpha$) you will correctly reject $H_o$.

## But There is More…

1. Recall that as $\alpha$ is *decreased*, the probability of Type II error *increases*.
   a. To figure out what this probability is, is necessary to define an *alternative hypothesis,* $H_A$, or in this case, $H_1$
   b. in this case, $H_A : p = 2q$.
2. That is, *females are twice as numerous as males*.

## The Distribution of Expected Values

1. Note that the area in curve B that falls within the critical region of curve A is $\beta$.

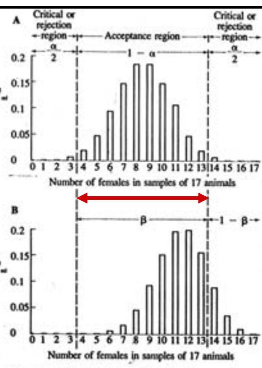



*Figure 7.11* Expected distributions of outcomes when sampling 17 animals from two hypothetical populations. A. $H_0: p_♀ = q_♂ = \frac{1}{2}$. B. $H_1: p_♀ = 2q_♂ = \frac{2}{3}$. Dashed lines separate critical regions from acceptance region of the distribution of Figure A. Type I error $\alpha$ equals approximately 0.01.

## Remember That,

1. $\beta$ is the probability of making a Type II error.
2. The area that falls outside is $(1 - \beta)$;
   a. This is the ***statistical power*** of this particular test.

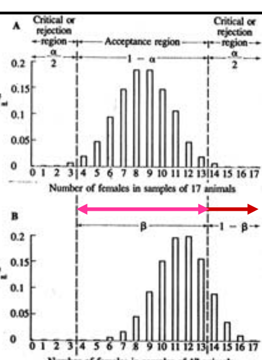



*Figure 7.11* Expected distributions of outcomes when sampling 17 animals from two hypothetical populations. A. $H_0: p_♀ = q_♂ = \frac{1}{2}$. B. $H_1: p_♀ = 2q_♂ = \frac{2}{3}$. Dashed lines separate critical regions from acceptance region of the distribution of Figure A. Type I error $\alpha$ equals approximately 0.01.

# Note:

1. How much of the distribution for $H_1$ overlaps that of $H_o$ (about 87%).

2. Thus, for this sample size and for the parameters of $H_1$, the two hypotheses are *virtually identical*.

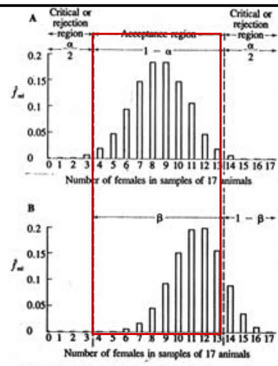



*Figure 7.11* Expected distributions of outcomes when sampling 17 animals from two hypothetical populations. A. $H_0: p_♀ = q_♂ = \frac{1}{2}$. B. $H_1: p_♀ = 2q_♂ = \frac{2}{3}$. Dashed lines separate critical regions from acceptance region of the distribution of Figure A. Type I error α equals approximately 0.01.

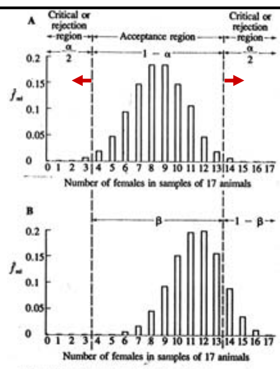



*Figure 7.11* Expected distributions of outcomes when sampling 17 animals from two hypothetical populations. A. $H_0: p_♀ = q_♂ = \frac{1}{2}$. B. $H_1: p_♀ = 2q_♂ = \frac{2}{3}$. Dashed lines separate critical regions from acceptance region of the distribution of Figure A. Type I error α equals approximately 0.01.

# Note Too That:

1. In this case, decreasing α (increasing acceptance region);

2. Also increases β, thus *decreasing* 1-β.

3. It defeats the purpose of maximizing POWER.

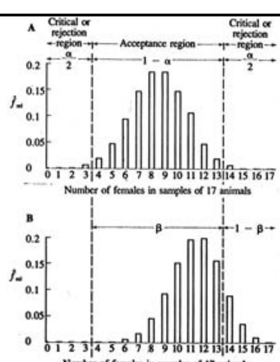



*Figure 7.11* Expected distributions of outcomes when sampling 17 animals from two hypothetical populations. A. $H_0: p_♀ = q_♂ = \frac{1}{2}$. B. $H_1: p_♀ = 2q_♂ = \frac{2}{3}$. Dashed lines separate critical regions from acceptance region of the distribution of Figure A. Type I error α equals approximately 0.01.

# Recall That:

1. If 1 - α is the portion of the possible data distribution in which $H_o$ is accepted.

2. β is the portion of the possible data distribution for $H_1$ that overlaps with 1 - α.

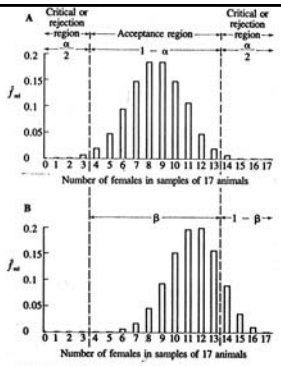



*Figure 7.11* Expected distributions of outcomes when sampling 17 animals from two hypothetical populations. A. $H_0: p_♀ = q_♂ = \frac{1}{2}$. B. $H_1: p_♀ = 2q_♂ = \frac{2}{3}$. Dashed lines separate critical regions from acceptance region of the distribution of Figure A. Type I error α equals approximately 0.01.

## So,

1. If statistical "power" is $1 - \beta$,

This is the portion of $H_1$'s distribution that *doesn't overlap* with $1 - \alpha$.

## How Can You Maximize Power?

Change your alternative hypothesis

## Increasing Power

1. The smaller β is (the less overlap), the more powerful the test is.
2. i.e., the more likely it is to *correctly* accept or reject $H_o$.

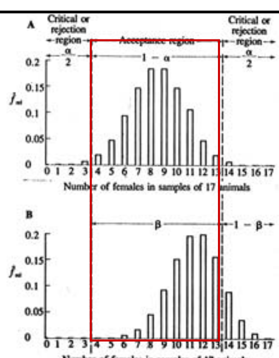



*Figure 7.11* Expected distributions of outcomes when sampling 17 animals from two hypothetical populations. A. $H_0: p_♀ = q_♂ = \frac{1}{2}$. B. $H_1: p_♀ = 2q_♂ = \frac{2}{3}$. Dashed lines separate critical regions from acceptance region of the distribution of Figure A. Type I error α equals approximately 0.01.

## Increasing Power

3. The power of a test is 1 - (the probability of wrongly accepting $H_o$).

a. Thus Power = the probability of *correctly* rejecting $H_o$.

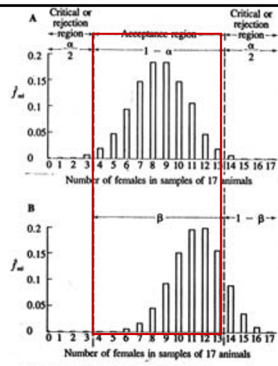



***Figure 7.11*** Expected distributions of outcomes when sampling 17 animals from two hypothetical populations. A. $H_0: p_♀ = q_♂ = \frac{1}{2}$. B. $H_1: p_♀ = 2q_♂ = \frac{2}{3}$. Dashed lines separate critical regions from acceptance region of the distribution of Figure A. Type I error $\alpha$ equals approximately 0.01.

## Statistical Power Includes

1. Minimizing the chance of having chance events bias data (Type I error).

2. Minimizing the chance of not seeing differences when they exist (Type II error).

## Power is Influenced By:

1. The parameters that describe the alternative hypothesis (i.e., how *different* $H_1$ is relative to $H_o$).

2. Sample size.

## Note:

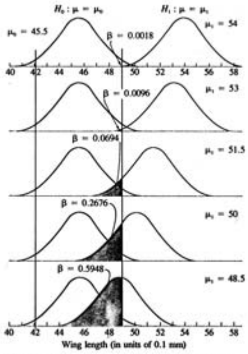



*Figure 7.13* Diagram to illustrate increases in type II error, $\beta$, as alternative hypothesis, $H_1$, approaches null hypothesis $H_0$—that is, $\mu_1$ approaches $\mu$. Shading represents $\beta$. Vertical lines mark off 5% critical regions ($2\frac{1}{2}$% in each tail) for the null hypothesis. To simplify the graph the alternative distributions are shown for one tail only. Data identical to those in Figure 7.12.

1. As $H_1$ becomes increasingly distinct from $H_o$ (i.e., as $\boldsymbol{\mu}_1$ becomes larger), 1-β increases.

2. Note that this is possible only when $H_1$ is ***extrinsic*** to the data (e.g., Fisher's expected sex ratio).

3. If hypothesis is ***intrinsic*** to data, this it is not always possible to manipulate 1- β.

## Sample Size

1. Note that for small sample sizes, the chance of rejecting $H_o$ is high *only* when large deviations are present.

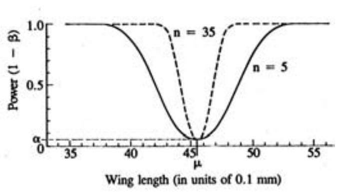



*Figure 7.14* Power curves for testing $H_0: \mu = 45.5$, $H_1: \mu \neq 45.5$ for $n = 5$ (as in Figures 7.12 and 7.13) and for $n = 35$. (For explanation see text.)

## The Effect of Increasing Sample Size

1. A larger N more clearly defines the shape of the distribution; i.e., it *decreases* the population variance.
2. This makes it *easier* to discriminate subtle differences between groups.
3. With larger sample sizes, even small deviations have a high probability of detection.

# Thus, In General,

1. The power of a test can be increased by increasing N.

2. This is an important relationship when choosing between parametric and nonparametric tests.

# For Example,

1. Test A may be more powerful than test B for a given sample size.
   a. i.e., test A is more likely to correctly reject Ho.
2. But if sample size is increased, test B may become more powerful than test A.

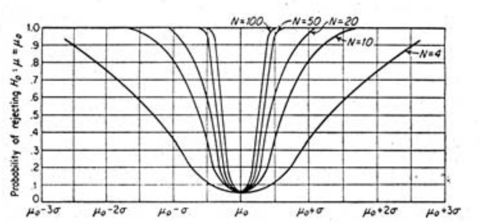



FIGURE 2.1
Power curves of the two-tailed test at $\alpha = .05$ with varying sample sizes.

# How To Do This?

3. We don't have time to consider how to calculate these differences, but as a general rule, even if small sample is available, power of test is increased with increased sample size.

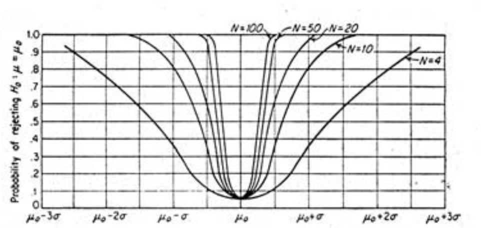



FIGURE 2.1
Power curves of the two-tailed test at $\alpha = .05$ with varying sample sizes.

## An Example

a. In the 1970s Dow Corning manufactured and marketed silicone implants.

b. Implants were considered safe until 1977, when several cases of ruptures were settled in court for $170K.

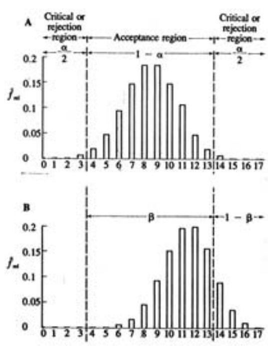
c. In 1988, San Francisco attorneys found DC documents indicating that testing may not have been sufficient to eliminate any chance of patient harm.

d. Plaintiffs were awarded $1.7M and more suits followed.

## What Might Have Happened

1. Dow Corning may have asked if there was a difference between immunological difficulties experienced by silicone implant recipients and those in general population.

a. Their desire may have been to avoid effect of rare events on decision to market their product.




## How To Explain a Rare Event..

1. One possible DC marketing decision may have been,

a. If implants did perform without difficulty in a sizable portion of the human population, the product *could* be profitable.

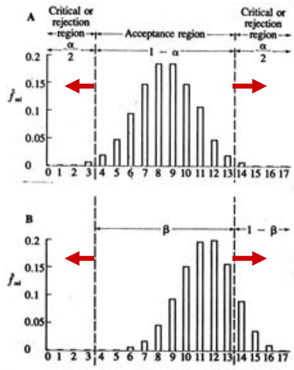
b. It may be possible to compensate individuals who did react negatively.

c. The way to determine whether to market the product would depend on the incidence of medical problems between Treatment and Control groups.

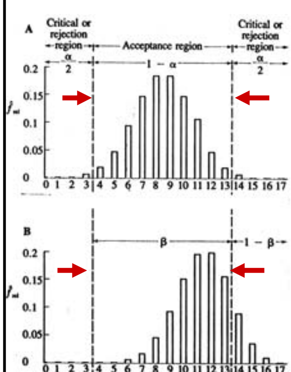
# Therefore,

1. Dow Corning may have set a *low* value on α (say 0.05), with a correspondingly *higher* value for β.
2. This is a *conservative approach* from the experimentalists point of view.

a. You don't want to report differences between experimental and control groups unless they exist.




# However,




2. This is likely to be a *different* perspective from the point of view of patients.

a. If there is a difference, *no matter how slight,* they'd want to know.

b. Patients would probably like to see a high α and therefore a smaller β.

# The Conflict

1. This is *less conservative* from an experimentalists perspective.
2. But it *more conservative* from patient's perspective.
3. Similar situation for AIDS drug research.

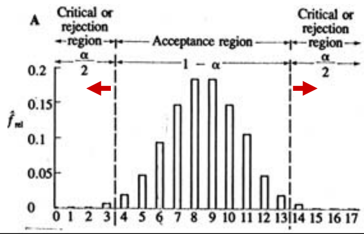
a. Patients want to see new more new drugs.

4. Researcher do too, but want to avoid false hope.

## What about β?

1. Dow Corning may have been willing to accept more difficulties (=lawsuits) when considering the apparent low frequency of no problem cases in their tests.

2. Therefore was willing to accept the relatively high β that went with low α.




## What about β?

2. This perspective is similar to that of perspective of journal editors.

a. They tend to be conservative in accepting evidence of an effect when there may be none.

2. Implant patients were clearly less willing to accept chances of failure.

a. Preferred a low β; and actually β=0.

b. The lack of this has led to near collapse of Dow Corning

('Better living through chemistry').

## This Expectation of Significant Results Where None May Exist Are Similar Among,

1. Patients attributing other illnesses to their implant surgery.
2. Attorneys seeking profitable settlements.
3. Politicians seeking justification for expensive international enterprises.
4. Eager young scientists seeking significant results in their experiments.