Aledo StatCat

Song Book

email: mmclellan@aledoisd.org

Mary McLellan
Aledo High School
Aledo, Texas

Updated May 2019

Experimental Design

Cluster Vs. Stratify
Tune: Hey Mickey

Hey Cluster you're a group, I will sample all of you, Hey Cluster, Hey Cluster

Hey Cluster you're a group, I'll take one or some of you, Hey Cluster, Hey Cluster

Hey Cluster you're a group, I'll take all from some of you, Hey Cluster, Hey Cluster

Hey Cluster you're a group, I'll take all from some of you, Hey Cluster, Hey Cluster

Stratify, Stratify, Homogeneous
Every single group is homogeneous
Stratify, Stratify, Homogeneous
I get all kinds cuz..
I get you and you and some of all of you…

Principles of Experimental Design
Tune: Spongebob Squarepants

You make an experiment
What do you do?
Control, Randomize, Repeat!

You make an experiment
What do you do?
Control, Randomize, Repeat!

Randomly assign to a
Treatment group so
You can claim the
Cause and Effect!

Confounding Variable
Tune: 99 Bottles of Beer

99 variables in the experiment
99 variables
Confounding variables cloud the issue
You can't claim a cause

98 variables in the experiment
98 variables
Confounding variables cloud the issue
You can't claim a cause

Lurking Variable
Tune: Happy

It might see crazy what I'm 'bout to say
There was a big fire last Saturday
Hot coals and plumes of smoke everywhere
They sent-- they sent a lot of -- firemen in there

Chorus- Lurking Variable makes it look like there's a link

All over the news about this and that
There's too much damage and you can't go back
If you send a lot to fireman to help you
Fireman will cause the damage, it's what they do!

Chorus-Lurking variable makes it look like there's a link

Randomly Select vs. Randomly Assign
Tune: All About that Bass

You know you randomly select, select,
no effect.
You know you randomly select , select,
no effect.
You know you randomly select, select,
no effect.
When your sample's like the population,
you can generalize.

When you make an experiment,
you must randomly assign.
To make the groups the same,
so your treatment is the why.

When you make an experiment,
you must randomly assign.
To make the groups the same,
so your treatment is the why.

Repeat

Randomize to Reduce Bias
Tune: Stayin' Alive

Well you can tell by the way that I design my
experiment I use my mind
Treatment A and Treatment B of the
experiment is what you see because
Subjects, they go in my big bag, I shake
them well
Subjects, they go in my big bag, I shake
them well

When you need to spread unknown traits
evenly, randomize, randomize
When you need to spread unknown traits
evenly, randomize, randomize
Randomize to reduce bias, reduce bias
Randomize to reduce biiiiiiiias

Blocking Reduces Variation
Tune: Good Vibrations

I see you have a trait there…
That may affect the way subjects respond
I see you have a trait there
That may affect the way subjects respond

What do you do?
Make homogeneous groups
What do you do?
Make homogeneous groups

Blocking reduces variation
Blocking reduces variation

Within each block is experimentation
Within each block is experimentation

SR*T ME CA/P T Design Paragraph
Tune: Shake, Shake, Shake

Start with the subjects that you have
Randomly assign them with a big bag

*CHORUS: SR*T ME CA/P
SR*T ME CA/P
Shake your bag up
Shake your bag up
(repeat)

Treatments, state your treatments- very well
Measure **E**ach subject's response to them

(Repeat chorus)

Compare compare their **A**verage or their
Proportion
To see, to see which treatment is really the
best!

(Use **BB**R*T ME CA/P for blocking. "BB"
stands for "**B**lock by <this trait in context>
Because the trait may affect the response
<context>)

Descriptive Data Songs

Independent
Tune: Oh my Darlin'

Independent, Independent, Independent
When the fact that you are one thing
Does NOT change the rate

Simpson's Paradox
Tune: Another one Bites the Dust

_ _ _ Simpson's Paradox
_ _ _ It does not rock
The conclusions you get from the
Separate groups
Are different when combined

Hey! Why does it do that?
They added those fractions wrong!

_ _ _ Simpson's Paradox
_ _ _ It does not rock

Boxplot song:
Tune: 10 Little Indians

Scale, Box, Fences, Whiskers
Scale, Box, Fences, Whiskers
Scale, Box, Fences, Whiskers
Put those outliers on!

Scale, Box, Fences, Whiskers
Scale, Box, Fences, Whiskers
Scale, Box, Fences, Whiskers
LABEL, LABEL, LABEL!!

Scale, Box, Fences, Whiskers
Scale, Box, Fences, Whiskers
Scale, Box, Fences, Whiskers
That's how you make a box plot!

Inter Quartile Range
Tune: Home on the Range

Inter Quartile Range
The middle 50%
You use it
To find
The upper and lower outlier fence

Lower Outlier Fence
It is Q1 minus
1 and a half IQR
Q1 - 1.5 IQR

Upper Outlier Fence
It is Q3 plus
1 and a half IQR
Q3 + 1.5 IQR

Outlier
Tune: Uptown Girl by Billy Joel

Outlier
Happens when the data that you have
Is more than 2 standard deviations
Above or below the mean, (mean that you have)

(repeat)

And if you use interquartile range rule
Then you check Q3 + 1.5 IQR
And you check Q1 – 1.5 IQR

Data above upper fence,
Or below the lower fence is an
Outlier…

Describe the Distribution
Tune: We are Young

Give me a second I, I need to get my words straight

I need to learn the vocab that is all about the data's shape

Uniform, symmetrical, skewed to the right and left

Will there be an outlier or is the shape regular and

I know I then choose the center, though
Do I pick the mean or the median
So that I don't get confused, the spread that I choose to use is
IQR or Standard Deviation
When the AP test approaches and I want to get a five,
I know what to do

Describe the Distribution
Shape and Outlier, Center and Spread

Describe the Distribution
Shape and Outlier, Center and Spread

When the AP test approaches and I want to get a five,
This song will help me to thrive!

Skew
Tune: Dancing Queen

If you have a skew
Then you must choose to use
Median and IQR

Ooh

Symmetry-
Use standard Deviation and mean

Transformation:
Tune: BINGO

When the data that you have goes through
A transformation-
Addition and Subtraction
Addition and Subtraction
Addition and Subtraction
Affects the center only

When the data that you have goes through
A transformation-
Multiply and Divide
Multiply and Divide
Multiply and Divide
Affects center and spread.

When to Use What Measure
Tune: Faithfully (Journey)

How do I know what measure to use?
Sometimes there's symmetry
Sometimes there's skew

When there's skew
When there's outliers
All I know is that I need resistance

The measure of center that is resistant to the skew
The measure of spread that is resistant, too
Are the measures that I should choose to use

I guess that I should choose then-
Interquartile range and median

Symmetry, Symmetry
Use Standard Deviation and Mean

Symmetry, Symmetry
Use Standard Deviation and Mean

Bivariate/ Linear Relationship

Interpret the Slope
Tune: The Addams Family

Interpret the slope (snap, snap)
Interpret the slope (snap, snap)
Interpret the slope, Interpret the
Slope, Interpret the slope (snap, snap)

For every 1 increase
In the "X"
The predicted "Y"
It does this…

For example –

For every 1°F increase in the temperature outside, the predicted energy use increases 256 Kwh.

Interpret the Y-Intercept
Tune: Wrecking Ball

Interpret the Y-Intercept
It happens when the X is 0
The PREDICTED Y is this
And it often does not make sense

For example –

When a student studies zero hours for a statistics test, the predicted test score is a 52.

Correlation Coefficient
Tune: Blue Suede Shoes

There is a strong
There is a weak
Positive or Negative
Linear
Correlation between "x" and "y"
Linear between "x" and "y"
-- the "r"

For example –

There is a strong positive linear correlation between the amount of hours studied and the grade in statistics.

Coefficient of Determination
Tune: Frosty the Snowman

Coefficient of Determination
The percent of variation that is found in the "y"
Coefficient of Determination
The percent of variation that is found in the "y"

It can be explained by the changes in the "x"
Or it can be explained by the model itself,

Oh
Coefficient of Determination
The percent of variation that is found in the "y"

For example:
About 84% of the variation in pumpkin cost is explained by the changes in the diameter of the pumpkin.

↑ this could be replaced by,
the linear relationship between diameter and cost of the pumpkin.

Linear Model
Tune: "Ice Ice Baby"

Yikes! Is it Linear, Listen…
Have you got a correlation coefficient
__ r is the strength of linearity
R^2 is % of variability
Can we tell from r- NO! You don't know
Residual plot- you must check
Look at residuals and see if there's a pattern
"Statistical Evidence" only with a p-value

Linear Model…. Linear Model….

Mean, Mean Point
Tune: Shake it Off

X- bar is the mean of X and
Y- bar is the mean of Y
And the mean, mean point is always
On the line, on the line

Residual Rap
R-A-P
Residual
R = A - P

$\text{Residual} = y - \hat{y}$

Actual y-value

Predicted y-value

Standard Deviation of Slope
Tune: Hello by Adele

Slope, varies
It's the standard deviation of the slope
You need to know
Slope, varies
It's the standard deviation of the slope
You need to know

If you sample many times
I'd expect the slope you find
Of the relation between "X" and "Y"
To vary this much
On average over time

For example:
If you sample many times, I'd expect the slope of the relationship between height and volume to vary about .38 $\frac{m^3}{inch}$, on average.

Standard Deviation of Residuals
Tune: Billie Jean

Standard Deviation
Standard Deviation
Of residuals

Average amount that
The actual "Y"
Actual "Y-value" point (ooh)
Varies from what's
Predicted to become
Varies from… the predicted "Y"

For example:
The average amount the actual volume varies from the predicted volume is $13.4m^3$.

The "r" Song
Tune: Play that Funky Music

Correlation Coefficient
Represented by the "r"
Correlation Coefficient
The variable that you use to represent it is the "r"

r" Doesn't Change
Tune: Hit Me With Your Best Shot

Well you can add, subtract, multiply, divide
You can even switch the X and Y
All but negative transformations fly
The "r" doesn't change

Influential Point
Song: You're the one that I want from Grease

If you omit a piece data
And it changes the slope
Of the equation of the line
Of the line

If you omit (ooh ooh ooh) a piece data and
(a piece of data and)
Changes equation of the line
If you omit (ooh ooh ooh) a piece of data and
Changes equation of the line
… especially slope of the line

Influential point
Influential point
Ooh ooh ooh honey
Influential point
Influential point
Ooh ooh ooh

That's what I mean- oh yes indeed ooh

Don't Extrapolate
Tune: Born in the USA

First we look at the data we get
Put in the best function that we think will fit
And then we try to use the function to predict
But the future data pattern may change quite a bit

Don't you extrapolate
Don't you extrapolate
Future of the function is a debate, so
Don't you extrapolate

High Leverage
Song- Bye Bye Miss American Pie

A point's leverage is high
When your X-value is far away from the mean
When your X-value is far away from the mean
Then your point's leverage is high
Potentially influence the line

Prelude:
When you have a point
that is so very far away from
the mean of X

Then I know that it can
have such an influence
If it's far away from the line

But if the point is in the line
then the correlation strengthens
And if the point's an outlier
The correlation weakens

And if the point is not so fine
then it can really change the line
Significantly change the slope and
Equation of the line

Transformation of a Function

Tune: Jeremiah was a Bullfrog

Transformation of a function
Happens if the X and Y
Have a pattern in the residual plot
So the linear doesn't apply
Oh the linear doesn't apply

Exponential Try
X and Log Y
Y equals a b to the x
Exponential model we check

Transformation of a function
Happens if the X, log Y
Has a pattern in the residual plot
So exponential doesn't apply
Oh exponential doesn't apply

Power Model Try
Log X and Log Y
Y equals a x to the b
Power Model's what you'll see!

Normal Model

Z Score
Tune: Go Big Red

X- Mean over Standard Deviation
X- Mean over Standard Deviation
(Repeat)
Inverse Norm of the Area Below
Inverse Norm of the Area Below
Inverse Norm of the Area Below
Is the Z at that point

When to Use NormCDF
Tune: Lego Movie Theme Song

Find Probability
When you find the proportion of
Observations
When you find percentage
Use NormCDF

The Percent is NOT the Z
Tune: The Champion

Percent is NOT the Z
It's NOT the Z
Use the percent to GET the Z
Use area below as the percent

Percent is NOT
the Z
It's NOT the Z
Use the percent to GET the Z
Use Inverse Norm to get the Z
Percent is NOT the Z

The Normal Curve
Tune: America

Normal Curve
Normal Curve
The Probability
Of finding X
Norm CDF
Low Z to high Z

Probability

Law of Large Numbers
Tune: Bear Necessities

If you do many trials, many, many, trials
The cumulative probability-
It will settle down to true probability
The Law of Large Numbers, you see!

Probability of At Least One
Tune: All I Wanna Do

Probability of At Least One
Listing ALL the combinations is NOT fun!

Probability of At Least One
Listing ALL the combinations is NOT fun!

Probability of At Least One
Easier with…
One minus the Probability of None

Expected Value
Tune: We Three Kings

Expected Value is the mean
Expected Value is the mean
Geometric: 1 over p
Binomial: n times p

Binomial Probability
Tune: Rock Around the Clock

Bi-no-mi-al probability
Bi-no-mi-al probability
Bi-no-mi-al probability
Out of "n" what do you need?

First you have combinations
"n" choose "k" successes
Count down n! Factorial
Over Success! times Failure! Factorial
Success! times Failure! Factorial

Then success probability
To the Number of successes that you need

Exactly this, Exactly that
...PDF is where it's at
Binomial PDF for that

Finally failure probability
To the Number of failures that you need

If you want At Most / At Least
acCumulate, yes indeed
Binomial CDF is what you need

Geometric vs. Binomial
Tune: I Wanna Soak Up The Sun
(Format: V, B1, C; V, B2, C; V, B3, C)
Verse:
Don't know the size of n?
Just trial where first success is

Then with success prob. "p"
Geometric Probability

Bridge1:
If you want the upper than
With 1 minus lower, you can
If you want the lower part
Than go ahead, ready to start. Go ahead ready to start

Bridge 2:
Geometric PDF
Exact place of the first success
Geometric CDF
Accumulate successes' places. 'Cumulate successes' places

Bridge 3:
Binomial PDF
Exact number of the successes
Binomial CDF
Collection of Bin-PDFs. Collection of Bin-PDFs

Chorus:
Binomial tells me
The probability of how—ma—ny
suc—ces—ses
Are out of n
Binomial tells me
The probability of how—ma—ny
suc—ces—ses
Are out of n
Are out of a set n

Disjoint Events
Tune: This is Gonna Be the Best Day of My Life

There are things down on the ground
and there are things up in the clouds
Oh oh oh oh oh oh oh
Oh oh oh oh oh oh oh

And the things up in the sky
Can't touch the ground at the same time
Oh oh oh oh oh oh oh
Oh oh oh oh oh oh oh

Never Intersect, no oh
They'll never intersect no, oh
Exclusive mutually

Oh, oh
Disjoint can't happen at the same time
Oh oh oh oh oh oh oh

Oh, oh
Disjoint can't happen at the same time

Simulations (ELUARSA)
Tune:Twinkle Twinkle

Each digit represents…
Let the #'s be…
Use a random # table, start here please.
A trial consists of, what do you ignore?
Your **R**esponse is critical, what do you look for?
Get your **S**tatistic, man-cuz I wanna know
According to your simulation, ABOUT what does this show

Inference

Central Limit Theorem
Tune: Jingle bell Rock

Central Limit theorem is about the shape
Of the sampling distribution and the population
Central Limit theorem is about the shape
Of the sampling distribution and the population

If n is large enough, it doesn't matter
The shape of the population,
It (sampling distribution) will still be normal

Central Limit theorem is about the shape
Of the sampling distribution
And the parent, and the parent
Parent population…

Assumptions Song
Tune: I Knew You Were Trouble

Randomly selected from the population or assigned
To the treatment group
Independent- n is less than 10%
Of the population
Large Enough is tricky 'cause it all depends
Mean or proportion
Mean is 30, Proportion is 10
And the Normal Model Applies Now!

Assumptions
Tune: The Christmas Song

Assumptions are our favorite things to do every time we run a test

We have to do all of our checks because the normal model is the best

So, we'll get RIL and start with RANDOM-sample's selected or assigned

If you don't have a representative sample, your test will face doom and gloom

Now independent is next… the sample size must be less
Than 10% of all the population so one sample doesn't change the chance

SO Large Enough is our final check… just how large, it depends
If you have a proportion or mean- will change the rule, you see

If you have a mean, n is greater than 30,

Proportions - np and nq are greater than 10 and now it's the end…

Normal Model, Normal Model, Normal Model for you!

Critical Value:
Tune: Shallow by Lady Gaga

Tell me something now
Aren't you tired of calculating wrong?
Well do you need more?
You just need to listen to this song

I'm trying
In every class time I find myself
Trying- to teach
And all the Z-stars you blame yourself

You need to find the critical value
While you may use the confidence level,
The confidence level is what's in the middle
Inverse norm of one tail

Inverse norm of one tail
Inverse norm of one tail
Inverse norm of one tail
The critical value now

What proportion to use to find n:
Tune- I can see clearly

The best it can be is if you use true **p**
If you don't have that, you can just use **p-hat**
Worst case scenario is those don't jive
Please don't cry, cry - just use **.5**
Please don't cry, cry- just use **.5**

Confidence Level
Tune: Smelly Cat

How many, How many
And what size interval?

How many, How many
Contain the true P?

Confidence Level
Tune: Can't Buy Me Love

Confidence level
Not an interval
Confidence level
Oh, oh, oh, oh

Let's say 95%, 95%
Of all 95%, 95%
Intervals of confidence,
They contain the TRUE

Parameter
Of Interest
Parameter
Oh, Oh, Oh, Oh

Let's say 98%, 98%
Of all 98%, 98%
Intervals of confidence,
They contain the TRUE
Parameter
Of interest
Parameter
Oh, Oh, Oh, Oh

Confidence Level Example:

About 95% of all 95% confidence intervals contain the population proportion of blue M&Ms the Mars factory makes.

Definition of P-Value
Tune: Brown- Eyed Girl

Do you remember when we used to sing…
The p-value is the probability
Of getting this sample
Or something more extreme
If the null is true

The p-value is the probability
Of getting this sample
Or something more extreme
If the null is true

P-Value is Low
Tune: Hound Dog

If the P-value is low (reject Ho)
If the P-value is low (reject Ho)
The null hypothesis has to go

If the P-value is low (reject Ho)
If the P-value is low (reject Ho)
The null hypothesis has to go

Statistically Significant
Tune: Star spangled Banner

Statistically Significant
Happens when your observed
Is so far away
Far away from expected

There's not really a way
That the sample I got
Happened because
Of chance, because of chance

Sampling Variability
Sampling Variability
Happens because you got
Different data than me

Oh if my observed sample's
In the rejection region
Then the data that I got's
Statistically Significant

Type 1 Error
Tune: We Will Rock You

Type 1, Type 1 Error
Type 1, Type 1 Error

If the null is really true and we get a sample
That causes us to reject the null,
That's a Type 1 Error

Type 1, Type 1 Error
Type 1, Type 1 Error

If the null is really true and we get a sample
That causes us to reject the null,
That's a Type 1 Error

Type II Error
Tune: Barbara Ann

Type II Error, Type II Error, Type II Error
Cause me to Fail to reject
The False null Hypothesis

Type II Error. Type II Error

No Evidence
No Evidence
No Evidence
Cause me to Fail to reject
The False null Hypothesis

Type II Error, Type II Error

Power
Tune: Hotel california

Power is the probability
To reject the null (to reject the null)
If it's false indeed

Power is the probability
To reject the null (to reject the null)
If it's false indeed

T, T,T, Not Z!

Tune: Who let the Dogs Out (2 groups of singers)

1st Group:	2nd Group:
Who's got a mean test?	T,T,T,
not Z!	
Sample Standard Deviation	T,T,T,
not Z!	
Who's got a mean test?	T,T,T,
not Z!	
Must include the df	T,T,T,
not Z!	
Who's got a mean test?	T,T,T,
not Z!	