V0F 2015 Schield SS Shortcuts 1

Statistical-Significance Shortcuts

by
Milo Schield

StatChat Feb 24, 2015

Slides at:
www.StatLit.org/pdf/
2015-Schield-StatChat-Slides.pdf

V0F 2015 Schield SS Shortcuts 2

Background & Goal

Statistical significance is one of statistics' big ideas.

For Z-scores, statistical significance is a single value.

For Chi-squared, student-T, the F-statistic, correlation and relative risk, statistical significance is complex.

To better understand statistical significance, students need to see it in different contexts.

Goal: To create "shortcut" formulas for statistical significance that are sufficient, memorable and apply to a wide variety of statistics.

V0F 2015 Schield SS Shortcuts 3

#1: Proportions Shortcut (SS)

If |p2 - p1| > 1/Sqrt(n), then
that difference is statistically significant

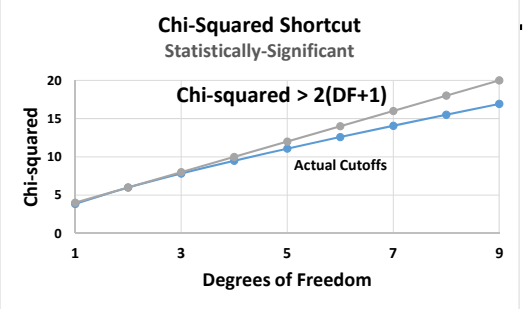
Q. Has anyone seen this shortcut? Where?

Yes! *Seeing Through Statistics*, Jessica Utts
Statistics: Art+Science of Data, Agresti/Franklin

Q. Anywhere else?

V0F 2015 Schield SS Shortcuts 4

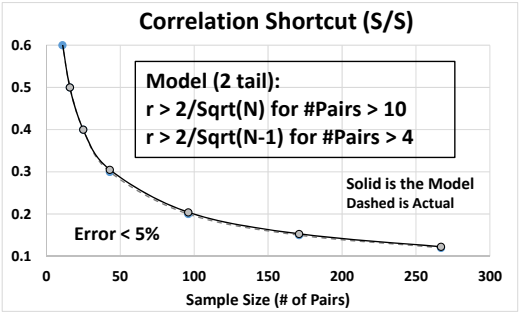
#2: Chi-Squared



Has anyone seen this shortcut anywhere?

V0F 2015 Schield SS Shortcuts 5

#3: Correlation



Has anyone seen this shortcut anywhere?

V0F 2015 Schield SS Shortcuts 6

#4: Relative-Risk

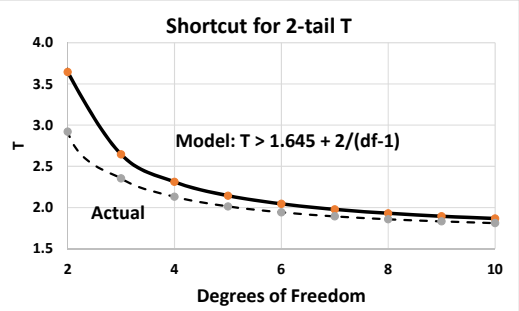
Consider two groups each of size n.
Relative Risk: RR = p2/p1

RR>1 is statistically significant if
RR-1 = 2/sqrt(k1)
where k1 = n*p1 >4

Has anyone seen this shortcut anywhere?

V0F 2015 Schield SS Shortcuts 7

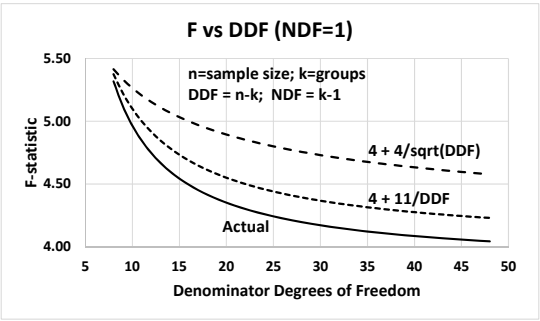
#5: T - Z



Has anyone seen this shortcut anywhere?

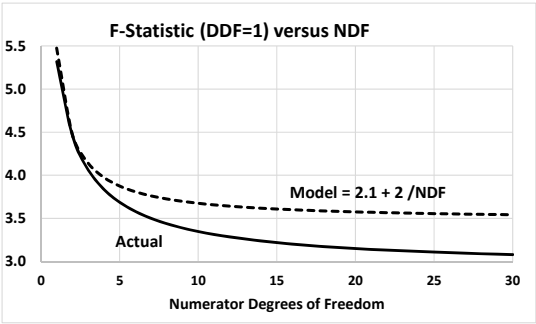
V0F 2015 Schield SS Shortcuts 8

#6a: F-statistic



V0F 2015 Schield SS Shortcuts 9

#6b: F-statistic DDF = N-k; NDF = k-1



V0F 2015 Schield SS Shortcuts 10

#6c: F-statistic Model DDF = n-k; NDF = k-1

N = sample size; K = # of groups

If 7 < DDF < 100 and 0 < NDF < 30, then Fcritical value (sufficient) =
2.1 + (11 / DDF) + (2 / NDF)

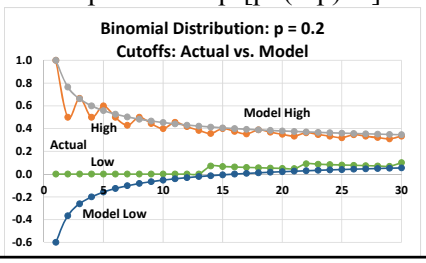
Error in this region (Model vs. actual):
Min 1.5%, Max 31%.

If n = 13 and k = 2, then ddf=11, ndf = 1, and Fsuff. = 2.1+1+2 = 5.1

V0F 2015 Schield SS Shortcuts 11

#7: Binomial Distribution If p*n > 5

High: k/n > p*n + 2*Sqrt[p*(1-p)*n]
Low: k/n < p*n - 2*Sqrt[p*(1-p)*n]



V0F 2015 Schield SS Shortcuts 12

Why don't we teach these shortcuts?

1. |p2-p1| > 1/sqrt(n)
2. Chi-squared: $X^2 > 2(df+1)$
3. Correlation: r > 2/sqrt(n-1) for n > 4
4. RRisk > 1+ 2/sqrt(k1): k1=n*p1, p1<p2
5. t-stat (2-tail): t > 1.645 + 2/sqrt(df-1)
6. F > 2.1 + 11/(n-k) + 2/(k-1)
7. Binomial: k/n > p+2√[p(1-p)/n] if p*n>5

Statistical-Significance Shortcuts

by
Milo Schield

StatChat Feb 24, 2015

Slides at:

www.StatLit.org/pdf/2015-Schield-StatChat-Slides.pdf

Background & Goal

Statistical significance is one of statistics' big ideas.

For Z-scores, statistical significance is a single value.

For Chi-squared, student-T, the F-statistic, correlation and relative risk, statistical significance is complex.

To better understand statistical significance, students need to see it in different contexts.

Goal: To create "shortcut" formulas for statistical significance that are sufficient, memorable and apply to a wide variety of statistics.

#1: Proportions Shortcut (SS)

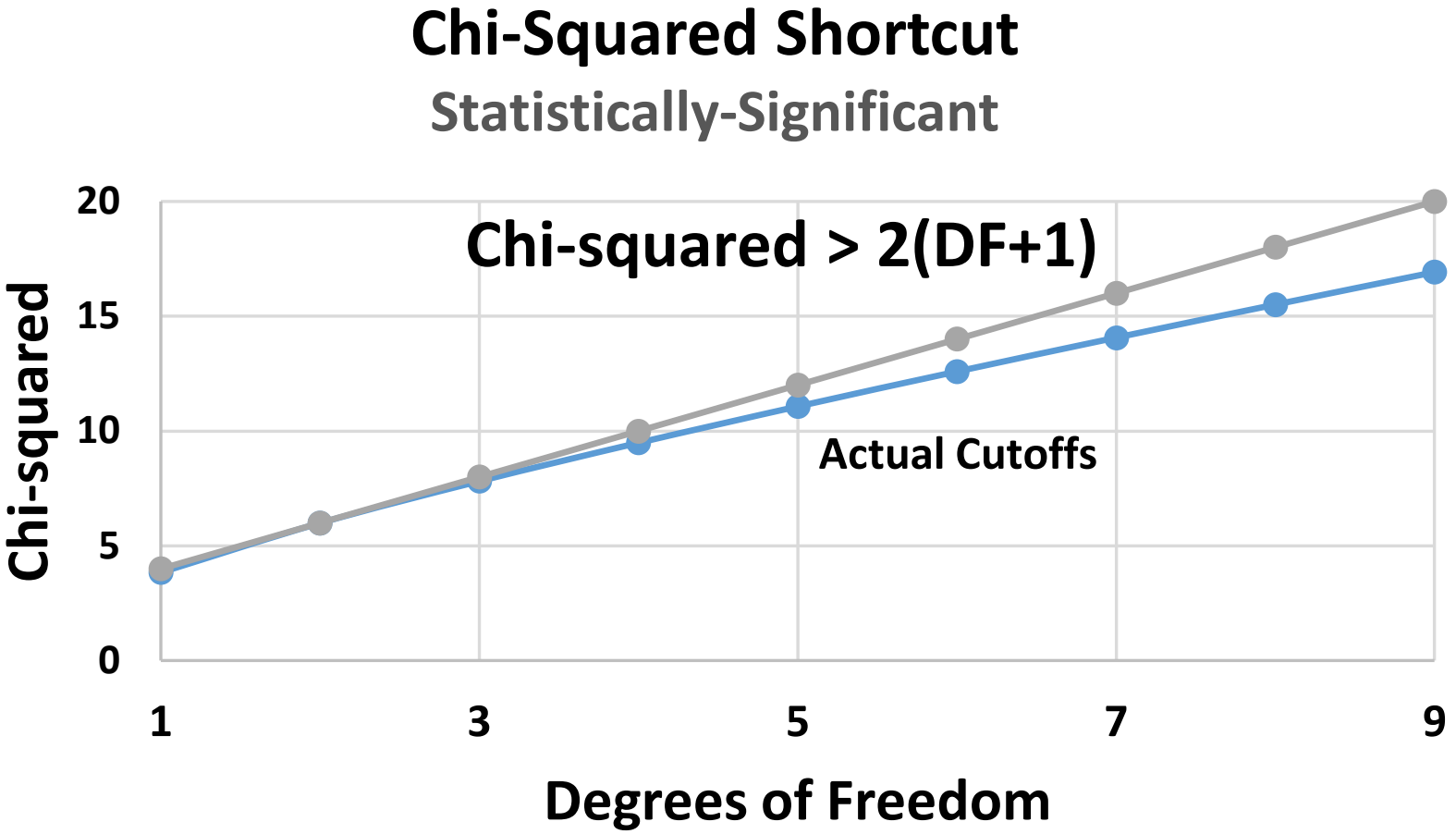
If |p2 - p1| > 1/Sqrt(n), then
that difference is statistically significant

Q. Has anyone seen this shortcut? Where?

Yes! *Seeing Through Statistics*, Jessica Utts
Statistics: Art+Science of Data, Agresti/Franklin

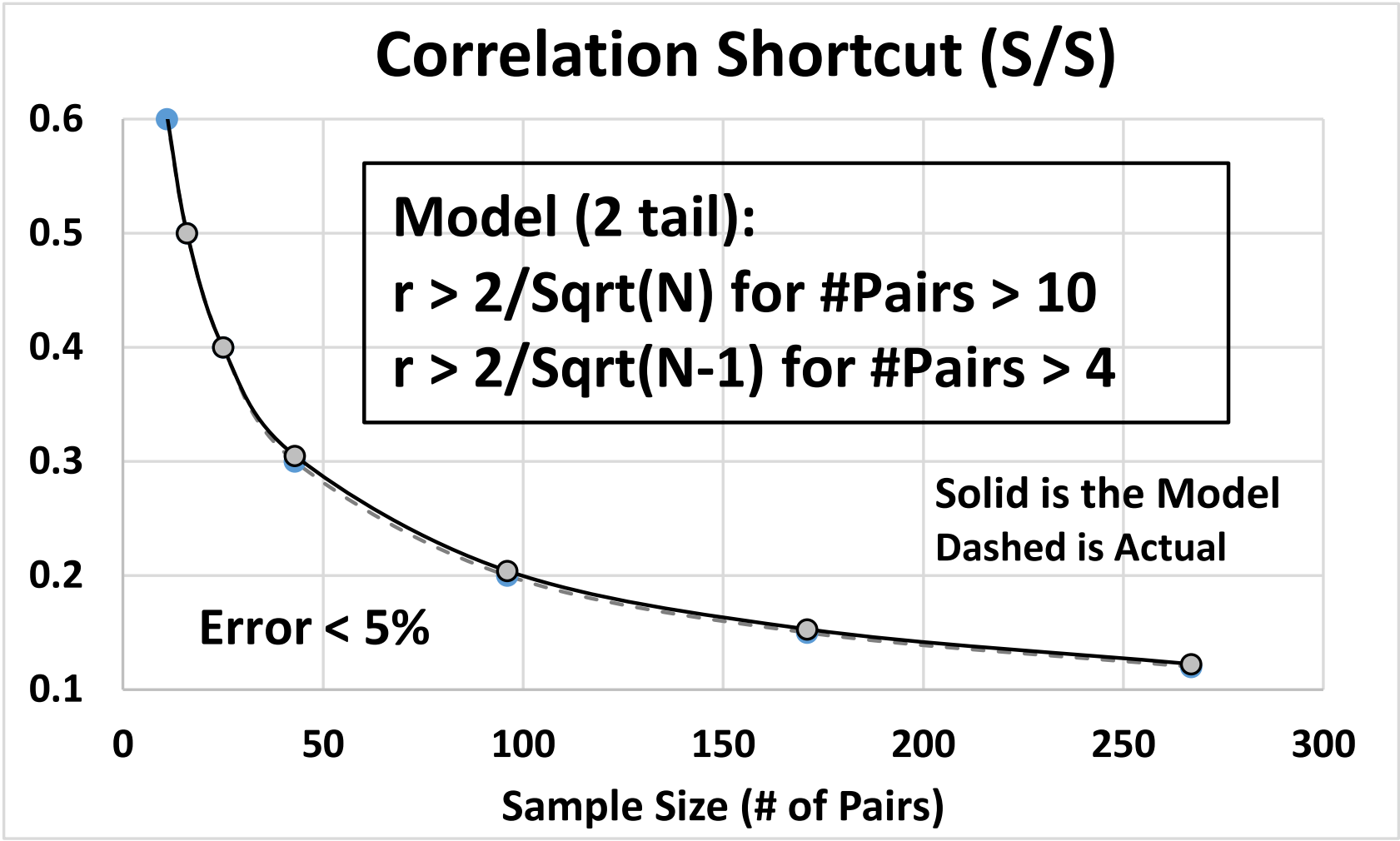
Q. Anywhere else?

#2: Chi-Squared



Has anyone seen this shortcut anywhere?

#3: Correlation



Has anyone seen this shortcut anywhere?

#4: Relative-Risk

Consider two groups each of size n.

Relative Risk: RR = p2/p1

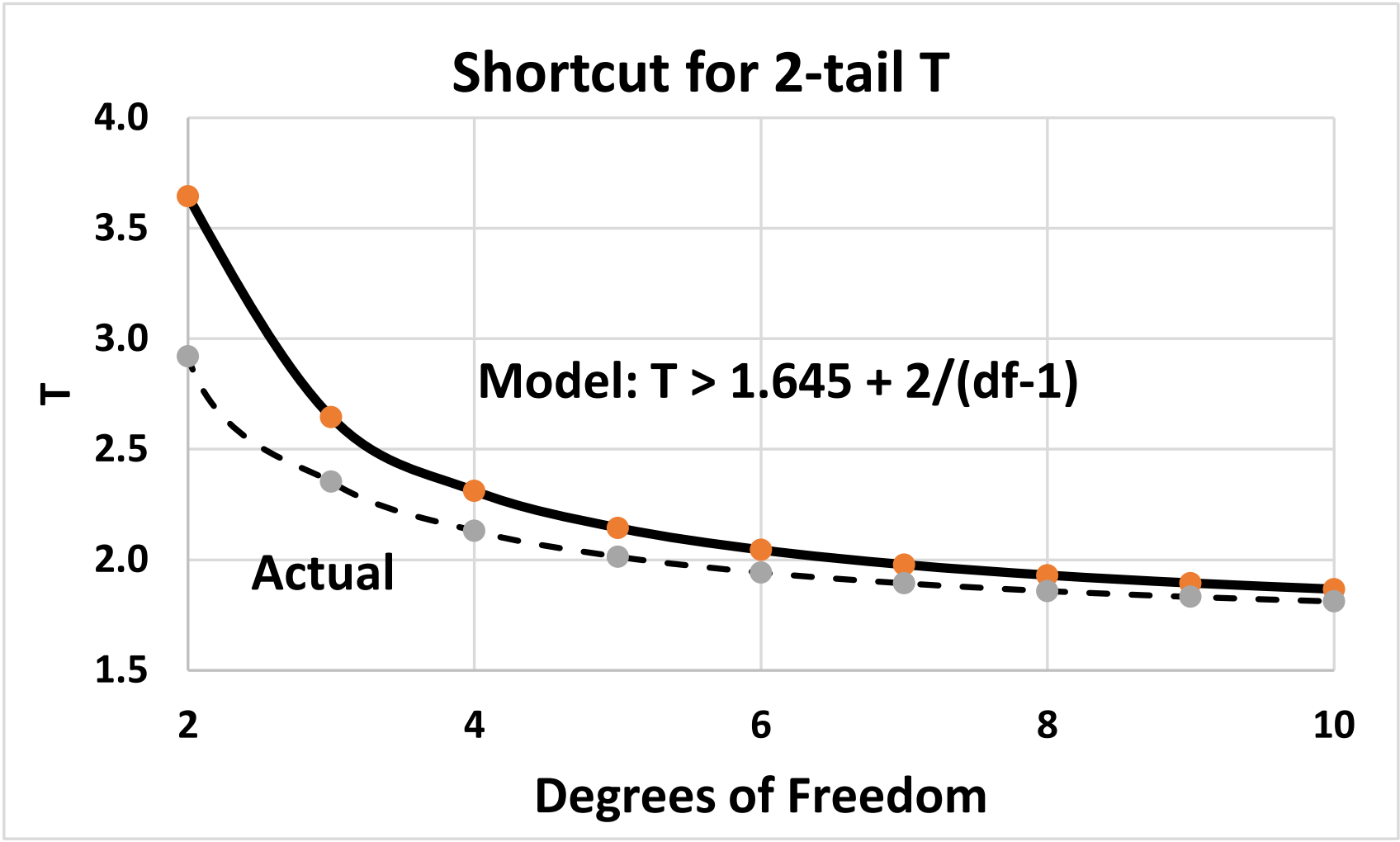
RR>1 is statistically significant if

$$RR-1 = 2/\sqrt{k1}$$

where $k1 = n*p1 > 4$

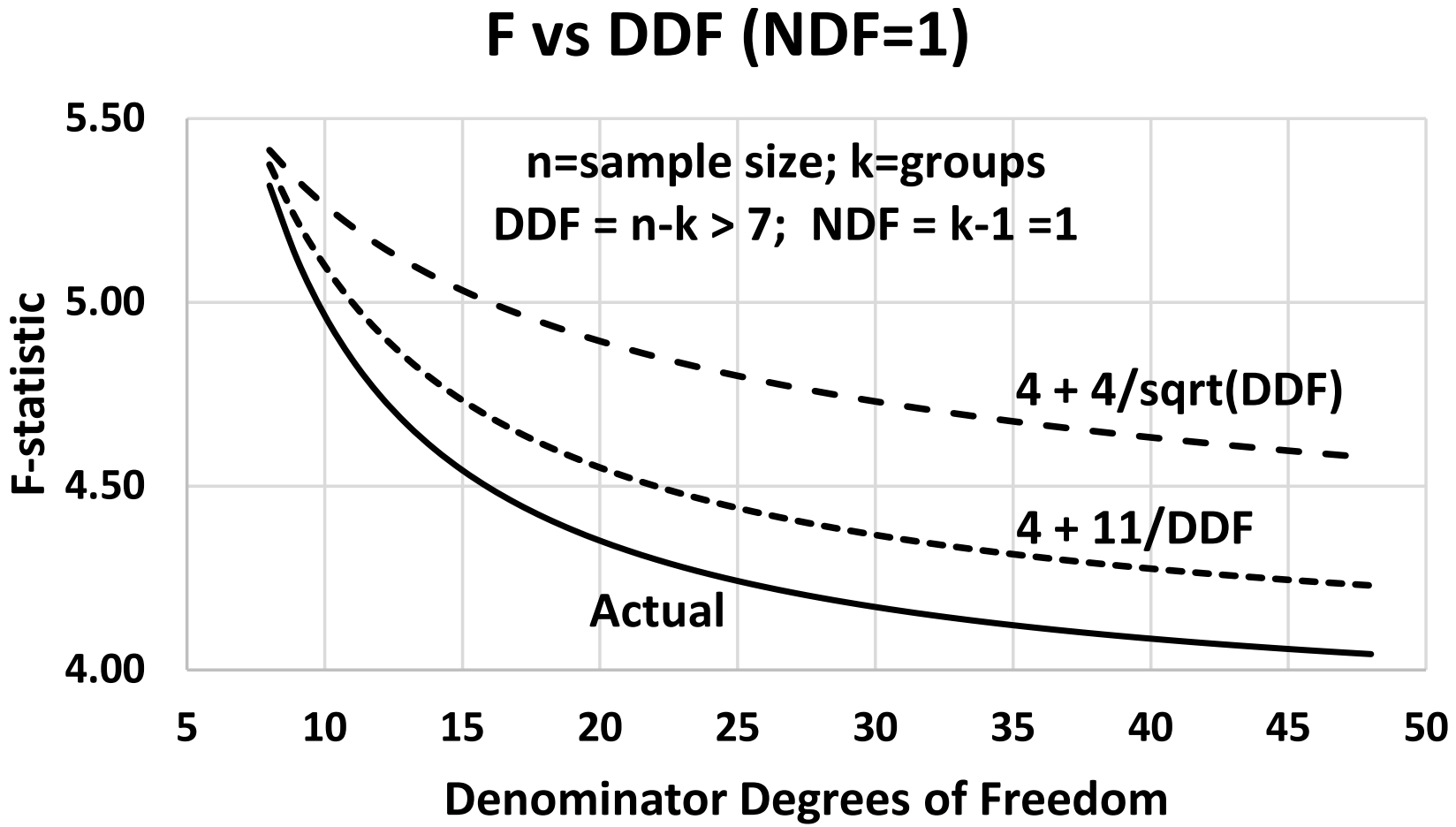
Has anyone seen this shortcut anywhere?

#5: T - Z

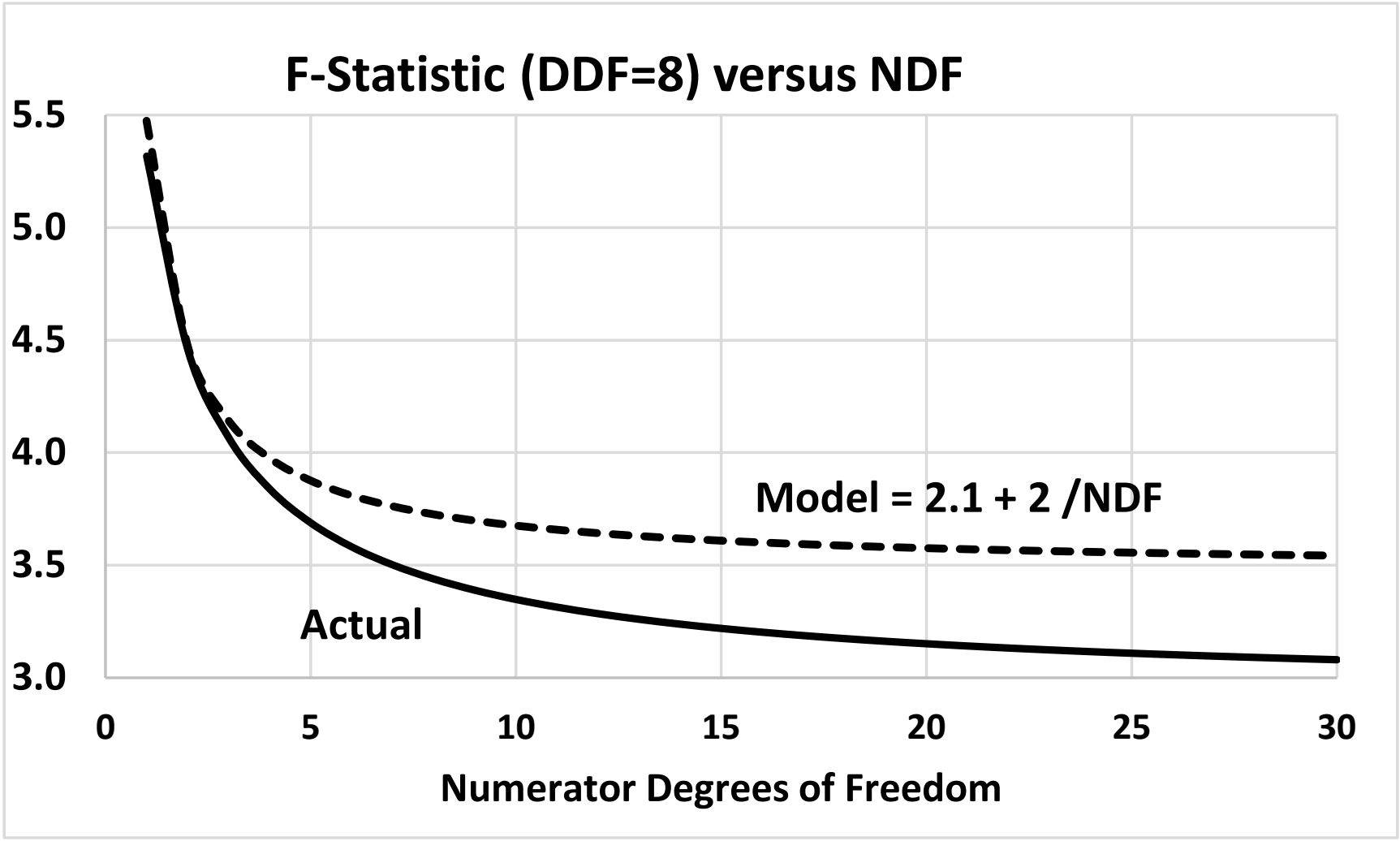


Has anyone seen this shortcut anywhere?

#6a: F-statistic



#6b: F-statistic
DDF = N-k; NDF = k-1



#6c: F-statistic Model DDF = n-k; NDF = k-1

N = sample size; K = # of groups

If 7 < DDF < 100 and 0 < NDF < 30, then Fcritical value (sufficient) =

2.1 + (11 / DDF) + (2 / NDF)

Error in this region (Model vs. actual):

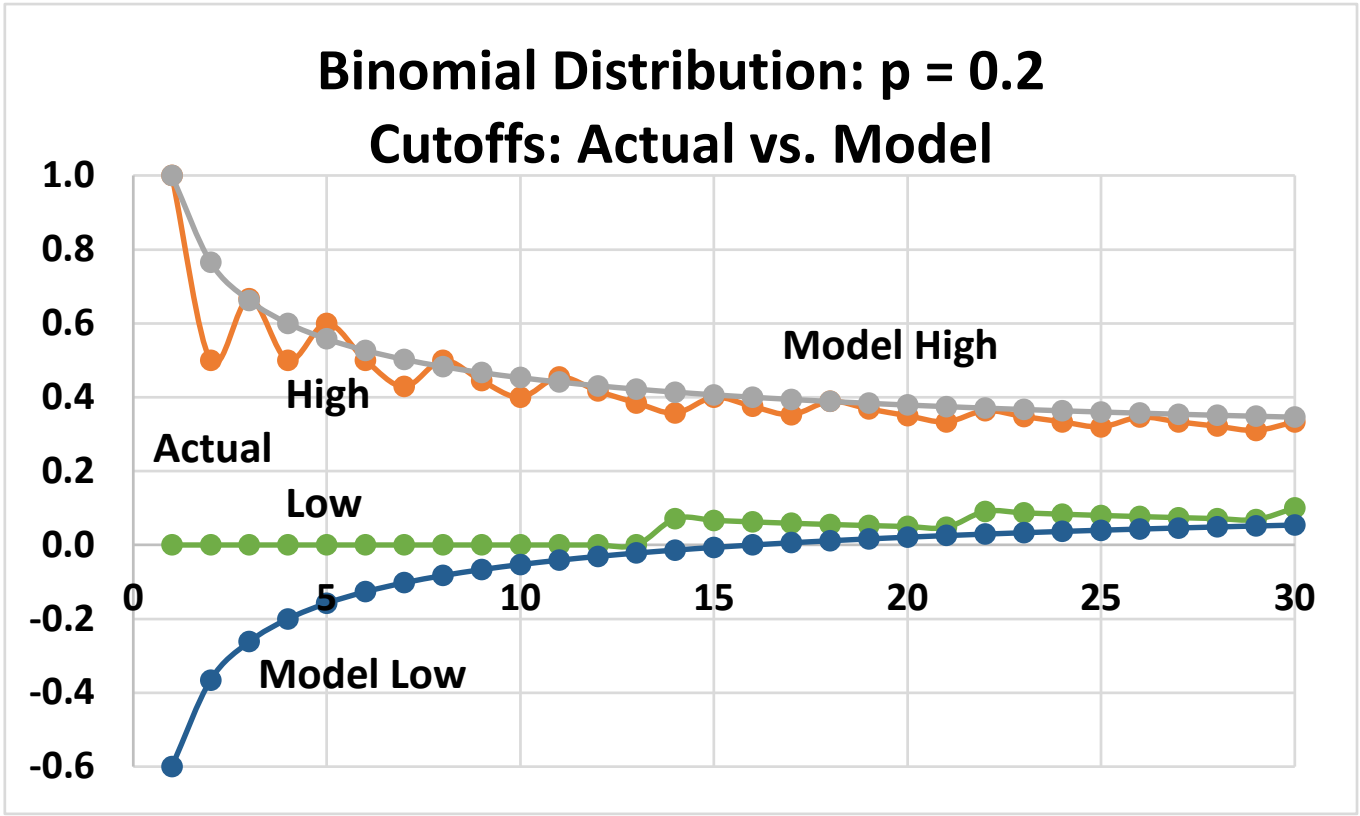
Min 1.5%, Max 31%.

If n = 10 and k = 2, then Fcrit = 5.5

#7: Binomial Distribution If p*n > 5

High: k/n > p*n + 2*Sqrt[p*(1-p)*n]

Low: k/n < p*n - 2*Sqrt[p*(1-p)*n]



Why don't we teach these shortcuts?

1. $|p2-p1| > 1/\sqrt{n}$
2. Chi-squared: $X^2 > 2(df+1)$
3. Correlation: $r > 2/\sqrt{n-1}$ for $n > 4$
4. RRisk > 1+ $2/\sqrt{k1}$: $k1=n*p1$, $p1<p2$
5. t-stat (2-tail): $t > 1.645 + 2/\sqrt{df-1}$
6. $F > 2.1 + 11/(n-k) + 2/(k-1)$
7. Binomial: $k/n > p+2\sqrt{[p(1-p)/n]}$ if $p*n>5$