4/1/24

X^2 Tests $\boxed{10.1}$

on calc, 2nd matrix, edit to the proper _ x _, enter your data, then stat → tests → C (x^2-test), dont edit anything, then enter

- Contingency tables are row x columns, count by rows first.

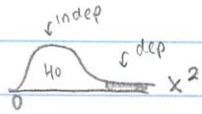
x^2 distr tests the independence of two factors, or their association

↳ it is not symmetrical, but as df ↑, becomes more symmetric

df formula: $\boxed{(\overset{\text{rows}}{r}-1)(\overset{\text{columns}}{c}-1)}$

to find cv, use df and lvl of significance

$\boxed{x^2 = \sum \frac{(O-E)^2}{E}}$ expected frequency... $\boxed{E = \frac{(\text{row total})(\text{column total})}{\text{grand total}}}$

- note: always positive. 

you'll always have $\boxed{H_0: x^2 > 0}$ right tailed always

- to check that all $E \geq 5$, chose smallest value. if it is >5, then all $E \geq 5$
- checks:
- calc: enter in matrix, then C

$\boxed{10.2}$ Goodness of Fit X^2

Ho: says now fits past pattern Ha: its different now

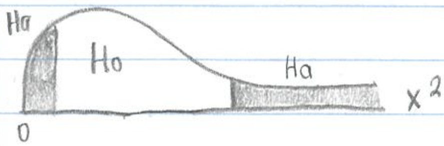
df = k-1 (k is # of categories)

$E = \% \times n$

$x^2 = \sum \frac{(O-E)^2}{E}$

10.3

Theorem: $\boxed{x^2_{scr} = \frac{df \cdot s^2}{\sigma^2}}$ → $df = n-1$



← still in family of x^2
same distribution

<u>only</u> x^2 that can have left-tailed or two-tailed tests

$H_0: \sigma^2 = K$
$H_a: \sigma^2 \ ? \ K$

to get p-value, do 2nd-vars: 8 (x^2cdf)

if right tail $\begin{cases} \text{lower is our } x^2 \text{ value} \\ \text{upper is big \#} \end{cases}$

if left tail $\begin{cases} \text{lower is 0} \\ \text{upper is } x^2 \text{ value} \end{cases}$

<u>Checks:</u>

- RRS
- Independent
  - $n \leq N/10$
- normally distr

10.1 Chi-Square Study Guide

1. Read pp. 624 through 630
2. Draw a 4 by 6 contingency table

title

label							total
	X						

rows first

3. Put a * in cell #7
4. Chi-squared distributions test the independence of two factors / association / relationship
5. T/F The graph of χ^2 is symmetrical

   0 χ^2 one-tailed test to right always

6. T/F As the d.f. increase, the χ^2 distribution becomes more skewed right

   no, symmetric

7. Where's the mode/high point on a χ^2 distribution? $n-2$
8. How does one calculate the critical value for χ^2?

9. Which is correct for finding the expected frequency of a cell?

$$\frac{\text{row total}}{\text{sample total}} * \frac{\text{column total}}{\text{sample total}} * \frac{\text{sample total}}{1} \quad \text{OR} \quad \boxed{\frac{\text{row total} * \text{column total}}{\text{sample total}}}$$

10. Find E for cell #9 table 10-2

    20

11. Symbolize observed frequencies O
12. χ^2 measures $O-E$ (A - P)
13. Use the χ^2 formula to justify the authors' claim in guided exercise 3 that χ^2 =13.31. Show your work.

$$\chi^2_{scr} = \sum \frac{(O-E)^2}{E}$$

14. Is that "it"? What else do we do?

15. How does one calculate the d.f.?

    $df = (r-1)(c-1)$ rows columns

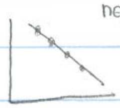
16. Find the d.f. for your contingency table in #2 above. 25

17. Do 10.1 (9)

χ^2 always > 0

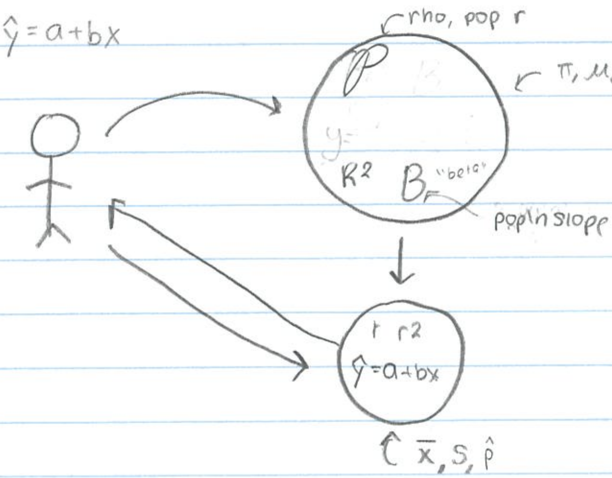
9.2 rvw

r=-1



negative perfect linear correlation

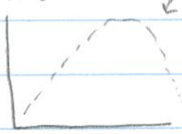
- r is correlation coefficient
- r^2 is coefficient of determination
- LSRL = $\hat{y} = a + bx$



r= .81

positive strong linear corelation

r=0



0 is not linear

9.3

on calc, test → F

$$r \to t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

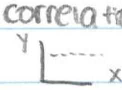
$$b \to t = \frac{b}{S_e}\sqrt{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}$$

will be the same

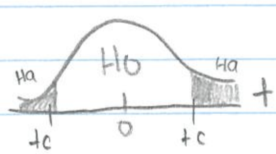
<u>df=n-2</u> *

name: T-test of rho (pop corr coeff) & beta (pop'n slope) <u>or</u> Lin Regr T-Test

Ho: ρ=0 <u>x</u> and <u>y</u> have no correlation
B=0 pop slope of LSRL is flat...



Ha: ρ≠0
B≠0



learn minitab

Confidence Interval for β (slope of popln LSRL)

Estimating

- r-sq (adj) is irrelevant (fyi)
- S_e = sd of dots (on minitab, just s)
- always on minitab, constant coeff = a, x-label + coef = b, can get $\hat{y} = a + bx$ from this
- x label + SE coeff is SE_b (standard error of slope)

$$\boxed{b \rightarrow t = \frac{b-\beta}{SE_b}}$$ (another formula, learned one previously)

$$\boxed{b \dashrightarrow \beta = b \pm t_c \cdot SE_b}$$

- $\mu_y = \alpha + \beta x$ — can check w/ residual plot, or visual of scatterplot
  - conditions: 1) real line has to be linear (relationship between x & y is linear)
    2) all σ_y are about equal
    3) the dots are normally distributed throughout

↳ checks:

- RRS
- Indep
  - $n \le N/10$
- normally distributed throughout (y values are normal for each x)
- all σ_y about equal
- the true relationship between x and y is linear

Sampling Distribution of Slope

- Shape: appx normal
- center: $\mu_b = \beta$
- spread: $\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$

Estimating σ

- use S_x to estimate σ_x
- use s to estimate σ

Name: T-confidence interval for popln slope β

- $SE_b = \frac{s}{s_x\sqrt{n-1}}$ (standard error)
- $df = n-2$ (use to get t_c)
- $E = t_c\left(\frac{s}{s_x\sqrt{n-1}}\right)$
- $b - E < \beta < b + E$

on calc - G