§11.1 Goodness of Fit

Ch. 8 $H_0 : P = P_0$
     1-Prop Z Test

         $H_0 : \mu = \mu_0$
         T Test

Ch. 9 $H_0 : P_1 = P_2$
     2-Prop Z Test

         $H_0 : \mu_1 = \mu_2$
         2 Samp T Test

Ch. 11.1 $H_0 : P_1 = P_{0_1} , P_2 = P_{0_2} , P_3 = P_{0_3}$
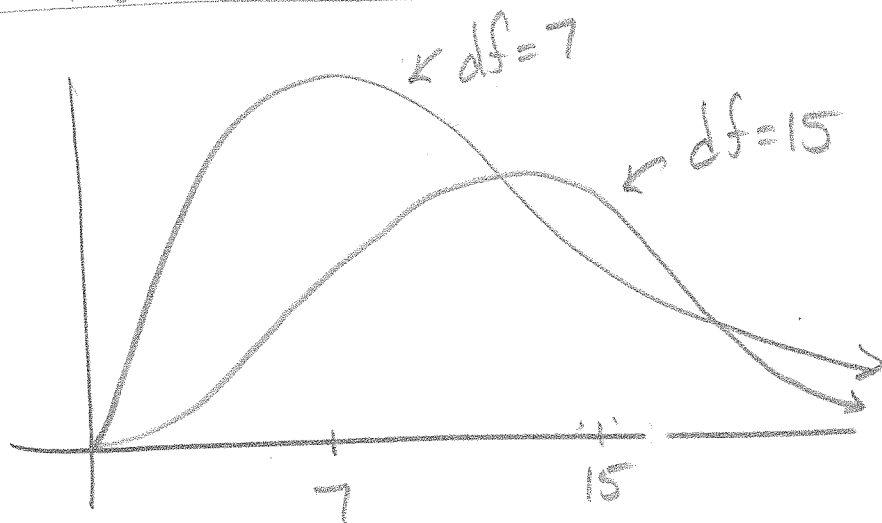
     11.2 $H_0 : O = E$ $P_1 = P_2 = P_3 = \cdots$

Ch. 12.1 $H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots$

Recall the X^2-distribution

① Curve depends on degrees of freedom, df.

$df = k - 1$
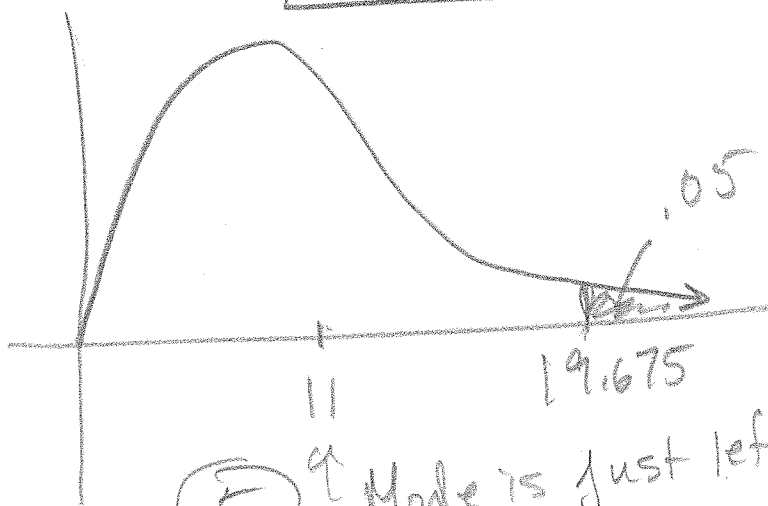
$k = \#$ of categories

② Always positive

③ Test will always be right Tailed.

④ [Prgm] INV X2

Ex If $k = 12$ and $\alpha = .05$ find X^2_R

$df = k - 1 = 11$

$SIG = .05 \Rightarrow$ $\boxed{CV \text{:} X^2_R = 19.675}$

⑤ Mode is just left of df

Ex A Loaded die

loaded ↓

 ← Do Not Put in calc.

Outcome	1	2	3	4	5	6
Observed	13	4	3	1	12	15
Expected	8	8	8	8	8	8

$$\underline{\frac{n = \Sigma obs}{K}} = Expected = \frac{13 + 4 + \ldots + 15}{6} = \overset{\text{Proportion for category}}{p \cdot \Sigma obs}$$

$$= \frac{1}{6} \cdot 48 = 8$$

$K = 6 = \#$ of categories

$$TS: \chi^2 = \frac{\Sigma(O-E)^2}{E}$$

$$= \frac{(13-8)^2}{8} + \frac{(4-8)^2}{8} + \ldots + \frac{(15-8)^2}{8}$$
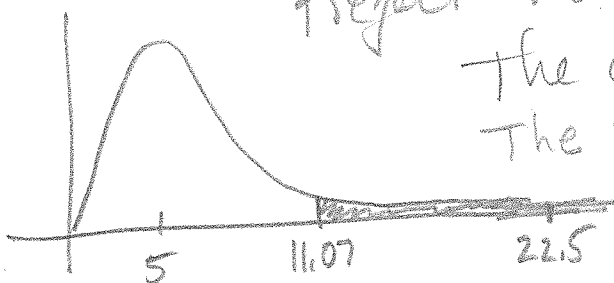
$$= \frac{1}{8}(25 + 16 + 25 + 49 + 16 + 49)$$

$$\boxed{TS: \chi^2 = 22.5}$$

$$CV: \chi_R^2 = Inv \chi2 (df = 5, \alpha = .05) = 11.07$$

Reject H_0: P_k are all equal
the die is Not fair
The proportion of each outcome
is Not the same.

5   11.07   22.5

GOF Test on Calculator

[STAT] ▷▷ [Tests] D: X^2 GOF-Test

Observed: L_1 ⟩ Do t-list first ⟩
Expected: L_2 Then run Test

[STAT] [Edit]

L_1	L_2
13	8
4	8
3	8
1	8
12	8
15	8

$H_0: O = E$ → proportions match
Given proportions

$H: O \neq E$ At least one proportion
is Not what was expected.

Output

P-value = 4.205×10^{-4} = .0004205

CNTRB = { 3.125 ≈ . . . .   6.125 }
         ↑   ↑
Contribution of category to TS
large values Means that category was
far from expected.

⟩ Meaning of p-value —
there is Very little chance of seeing proportion
this far from uniform. If the die is fair

Reject H_0 ⇒ Support H_A ⇒ the die is Not fair

M&M's

	Br	Rd	Bl	Y	Or	Gr
Ex. prop.	$P_1 = .15$	$.10 = P_2$	.10	.25	.25	.15
Obs.	22	31	56	38	64	30
Exp	36.15	24.1	24.1	60.25	60.25	36.15

Test the claim that the proportion of each color in the bag match the company proportions. at $\alpha = .05$.

① $\boxed{Exp = P_i \cdot \sum Obs} = P_i \cdot (22 + 31 + \cdots + 30) = P_i(241)$

② Put $L_1 = $ Obs. Values
$L_2 = $ Obs. Values $\rightarrow$ Not proportions

③ χ^2 GOF-Test $\qquad \chi^2 = 59.235$
$df = 6$ $\qquad$ P-Value $= 1.75 \times 10^{-11} \approx 0$

Change Proportions & Try again

Exp Prop	.1	.1	.25	.15	.25	.15
Obs.	22	31	56	38	64	30
Exp.	24	24	60	36	60	36

TS: $\chi^2 = 3.832$

P-Value $= .5738 > .05$

fail to Reject H_0

TINSE to Reject
that count fit the
Expected M&M Frequencies

df=5

3.832   11.071
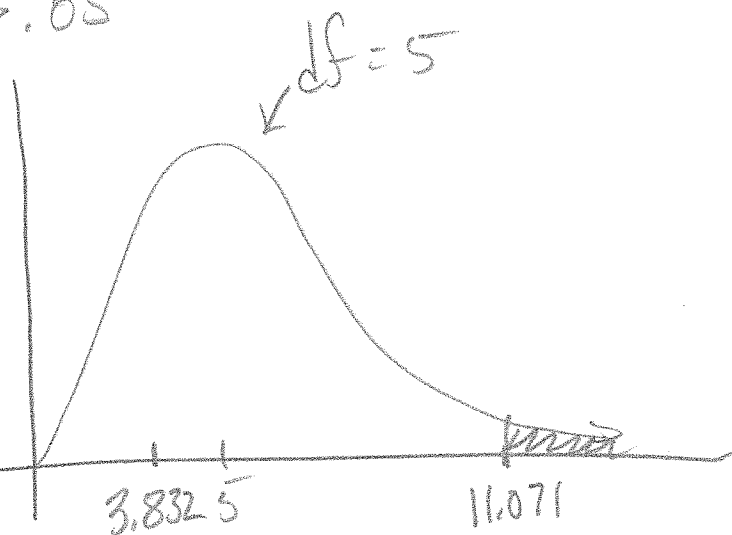
∴ The given prop. are close to the
observed proportions

Never say "Data fits"

Prop in #19 in Book

M&M	Gr .16	Or .20	Y .14	Bl .24	Rd .13	Br .13	
Obs	30	64	38	56	31	22	= Σ = 241
Exp							

§10.2 #6 Bear Chest Width vs Weight

r = Calculator = .879 $\hat{y} = -212 + 61.9 X$

r* = Table = .268 $\hat{y}(6.5) = -212 + 61.9(6.5) = 190.35$ lb

Yes, there is a linear correlation

$$Slope = M = \frac{Rise}{Run} = \frac{\Delta y}{\Delta x} \frac{lb}{inch}$$

h) Write a sentance to interpret the slope.

$\underline{y-var}$ changes $\underline{inc. or Dec.}$ by $\underline{M}$ for each one $\underline{\frac{unit}{of x}}$ in $\underline{x-var}$

The Weight of Bear increases by 61.9 lb for each 1 inch change in the width on picture.

i) Find r^2 and say what it means.

$r^2 = (.879)^2 = .7726$ or 77%

77% of the variation in weight of Bears (y-var) can be explained by the linear relationship with the overhead picture width (x-var).

1) (18 Points) Periodically during the last two and a half years my husband has gone out to do capacity runs to determine the range of his Electric Mustang. The paired data below consist of battery age in years and the range of my husband's EV in miles.

AGE (years)	0.2	0.3	0.8	1.0	1.2	1.5	1.7	1.8	2.0	2.5
Range (miles)	37	35	35	34	10	29	31	33	32	29

a) (3 Points) At the 5% level of significance, do the data provide sufficient evidence of an association between the age of the batteries and the range of the car? r = _____   r* =_____

Is there a significant linear corrolation?   Yes   No

b) (2 Points) Make a scatter plot of this data.

c) (1 Points) Find the equation for and graph the regression line._____

d) (1 Points) Based on the above data what is the best predicted range for the car after 3 years. _____

One of the batteries died and had to be exchanged. One of the capacity runs was done right before this happened. The data point that corresponds to this run is an outlier. Remove it and redo the test for correlation.

Years	0.2	0.3	0.8	1.0		1.5	1.7	1.8	2.0	2.5
Range	37	35	35	34		29	31	33	32	29

e) (3 Points) At the 5% level of significance, do the data provide sufficient evidence of an association between age and range.

    r = _____   r* =_____

Is there a significant linear corrolation?   Yes   No

f) (1 Points) Find the equation for the regression line._____

g) (1 Points) Using the data without the outlier, find the best predict ed range for the car after 3 years. _____

h) (3 Points) Interpret the slope of this regression line.

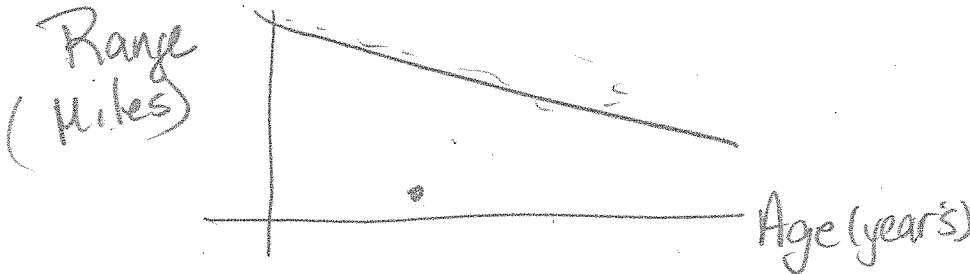i) (3points) Find r^2 and interpret what it means. r^2=_____

1) (18 Points) Periodically during the last two and a half years my husband has gone out to do capacity runs to determine the range of his Electric Mustang. The paired data below consist of battery age in years and the range of my husband's EV in miles.

AGE (years)	0.2	0.3	0.8	1.0	1.2	1.5	1.7	1.8	2.0	2.5
Range (miles)	37	35	35	34	10	29	31	33	32	29

a) (3 Points) At the 5% level of significance, do the data provide sufficient evidence of an association between the age of the batteries and the range of the car? $r =$ _____ $r^* =$ _____

Is there a significant linear corrolation?   Yes   No

b) (2 Points) Make a scatter plot of this data.

c) (1 Points) Find the equation for and graph the regression line._____

d) (1 Points) Based on the above data what is the best predicted range for the car after 3 years. _____

One of the batteries died and had to be exchanged. One of the capacity runs was done right before this happened. The data point that corresponds to this run is an outlier. Remove it and redo the test for correlation.

Years	0.2	0.3	0.8	1.0		1.5	1.7	1.8	2.0	2.5
Range	37	35	35	34		29	31	33	32	29

e) (3 Points) At the 5% level of significance, do the data provide sufficient evidence of an association between age and range.

$r =$ _____ $r^* =$ _____

Is there a significant linear corrolation?   Yes   No

f) (1 Points) Find the equation for the regression line._____

g) (1 Points) Using the data without the outlier, find the best predicted range for the car after 3 years. _____

h) (3 Points) Interpret the slope of this regression line.

$m = -3$ miles/year   ~> For each year the range decrease
$\frac{y \, units}{x \, units}$ by 3 miles

i) (3points) Find r^2 and interpret what it means. $r^2 =$ _____

$r^2 = .7198$ => 72% of the variation in Range can be predicted by the linear relationship with age. the other 28% is due to other factors