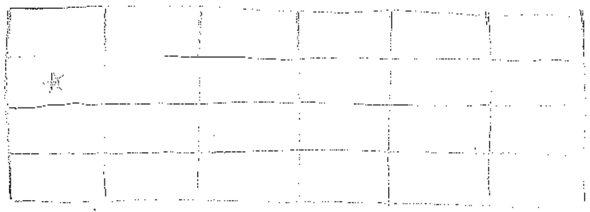
chi - square test

total



total

4. independence of two factors ← categorical

observed − expected ← measure of irony

5. F

6. F

7. where n = 2   when n ≥ 3

?. $E = \dfrac{(\text{row total})(\text{column total})}{\text{sample size}}$

$\chi^2 = \sum \dfrac{(O-E)^2}{E}$

d.f. = (rows − 1)(columns − 1)

√'s

RRS

$E \geq 5$ (all)

↳ if not, 80% should be > 5

none of E can = 0

$\boxed{\chi^2 \text{ test of independence}}$

11.1 (9, 10, 15, 16, 19)

9. table size: 3x2

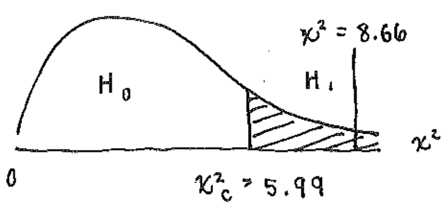
$\chi^2 = \sum \dfrac{(O-E)^2}{E}$

O	E	O − E	$(O-E)^2$	$(O-E)^2/E$
62	49.02	12.98	168.48	3.44
45	57.98	−12.98	168.48	2.91
68	74.22	−6.22	38.69	.5213
94	87.78	6.22	38.69	.4407
56	62.76	−6.76	45.70	.7282
81	74.24	6.76	45.70	.6136

$\Sigma = 8.66$

$H_0: \chi^2 = 0$ occupation and personality are independent

$H_1: \chi^2 > 0$



d.f. = 2

$\alpha = .05$

$\chi^2 = 8.66$

~~H_0~~ H_1 $\alpha = .05$

∃x sufficient statistical evidence suggesting that occupation and personality are dependent

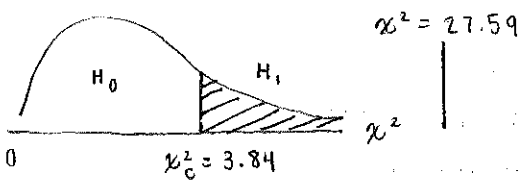
$.010 < p\text{-value} < .025 < .05$ (α)   "SEVN" ex. smoking and risk

O	E	O − E	$(O-E)^2$	$(O-E)^2/E$
45	26.33	18.67	348.57	13.24
46	64.68	−18.68	348.57	5.39
36	54.68	−18.68	348.57	6.37
153	134.33	18.67	348.57	2.59
				$\Sigma = 27.59$

√'s
RRS
all $E \geq 5$

$H_0: \chi^2 = 0$ smoking status and risky tendencies are not associated

$H_1: \chi^2 > 0$



~~H_0~~ H_1 $\alpha = .05$

∃x sufficient statistical evidence suggesting that SS and RT are associated

p-value	<	α	"SEVN"
< .005	<	.05	

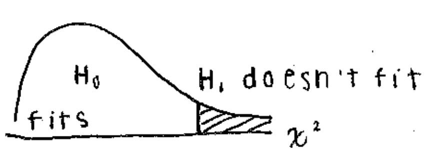
calc: $\boxed{\text{2nd}}$ → $\boxed{\text{Matrix}}$ → $\boxed{\text{edit}}$

goodness of fit test 11.2 (6, 10, 13, 14) $\boxed{\chi^2 \text{ goodness of fit test}}$

O = observed H_0 and H_1
E = expected χ^2
d.f. = k − 1 ; k = # of categories / rows
E = n × p = (sample size)(probability)

√'s
RRS
E all ≥ 5
↳ 80% beat 5, others ≥ 1

$H_0: \chi^2 = 0$ population fits the distribution predicted by …
$H_1: \chi^2 > 0$



ex. pine park USD

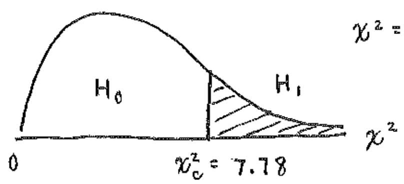
$n = 137$

	'11	'21	E	(O − E)	$(O-E)^2$	$(O-E)^2/E$
V	4%	3	5.48	−2.48	6.15	1.12
$	65%	77	89.05	−12.05	145.2	1.63
S	13%	9	17.81	−8.81	77.62	4.36
H	12%	41	16.44	24.56	603.19	36.69
O	6%	7	8.22	−1.22	1.49	.1813
		n=137				$\sum = 43.98 = \chi^2$

$\chi^2 = 0$

H_0: pine park USD staff in '21 fits how they felt in '11

H_1: $\chi^2 > 0$



$\chi^2 = 43.98$ df = 4 $\alpha = .1$

~~H_0~~ H_1 $\alpha = .1$

$\exists_x$ sufficient statistical evidence suggesting that the 2021 staff doesn't fit opinions in 2011

p-value		α
<.005	<	.10 "SEVN"

put O and E into a list

$\boxed{\text{stat}} \rightarrow \boxed{\text{test}} \rightarrow \boxed{\chi^2 \text{ GOF - test}}$

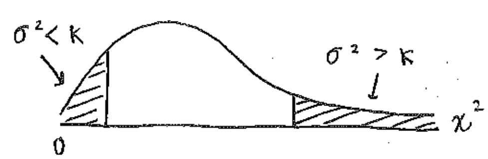
χ^2 test of variance

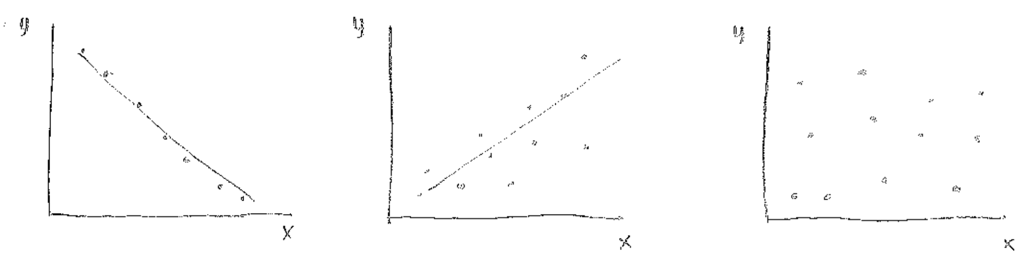
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

d.f. = n − 1

H_0: $\sigma^2 = K$

H_1: $\sigma^2 \gtrless K$, $\sigma^2 \neq K$





pg 164 # 15

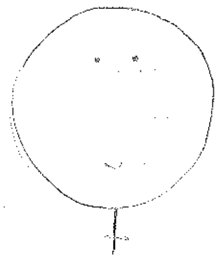
$\bar{x} = 18.8$ LSRL → $\hat{y} = -17.20 + 1.2024x$

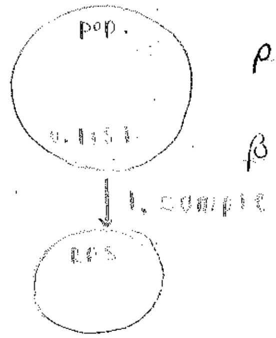
$\bar{y} = 5.4$

x	y
18.8	5.4
0	-17.20

11.4 (5, 7 - 9, 12)

working with t distribution





ρ = population correlation coefficient

β = slope of pop. LSRL

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

d.f. = n - 2

example 5:

1. use a calculator to find r ($r \approx .887$)

$H_0: \rho = 0$ → no linear correlation

$H_1: \rho > 0$ → positive linear correlation

2. use t equation, n = 6 r = .887

$t = 3.84$

example 7:

$H_0: \beta = 0$ → slope is positive

$H_1: \beta > 0$ → slope is positive

r → t
b → t
same t-value

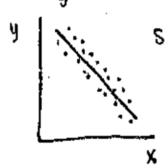
calculator: $\boxed{\text{stat}}$ → test → $\boxed{\text{F}}$

√'s

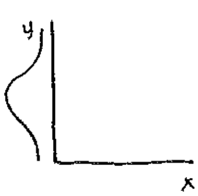
each (x, y) RRS

each (x, y) independent

all $s_y \approx$



standard dev. is relatively the same



Inference Conditions/Assumptions

Means:

Sigma known: 1-sample *z*-test: 1) SRS (and/or data is from a randomized experiment)
2) Sampling distribution of xbar is normal, meaning n $\geq$ 30 or
population
is normal.

2-sample *z*-test: 1) independent SRS's (and/or data is from a randomized experiment)
2) Sampling distributions of xbars are normal, meaning both n's $\geq$ 30
or
populations are normal.

Sigma unknown: 1-sample *t*-test: 1) SRS (and/or data is from a randomized experiment)
2) Sampling distribution of xbar is normal, meaning n $\geq$ 40, or n $\geq$
15
with no outliers or strong skewness, or population is normal.

2-sample *t*-test: 1) independent SRS's (and/or data is from a randomized experiment)
2) Sampling distribution of xbars are normal, meaning n's $\geq$ 40, or n's $\geq$ 15 with no outliers or strong skewness, or population are normal.

Matched pairs *t*-test: 1) SRS (and/or data is from a randomized experiment)
2) Sampling distribution of xbar of differences is normal, meaning n $\geq$ 40, or n $\geq$ 15 with no outliers or strong skewness, or
population
is normal.

Proportions

1-sample *z*-test
1) SRS
2) population is at lest 10 times larger than sample.
3) np and n(1-p) $\geq$ 10

2-sample *z*-test
1) independent SRS's
2 population is at lest 10 times larger than sample.
3) for both samples, np and n(1-p) $\geq$ 5

Chi-Square

1) independent SRS's (and/or data is from a randomized experiment)
2) all expected counts > 0
3) no more than 20% of expected counts < 5

Linear Regression

Linear regression *t*-test: 1) observations are independent
2) relationship between variables is linear
3) standard deviation of *y*'s is the same for all values of *x*
4) *y* varies normally for all values of *x*