Regression

a) exp = Fat (g)
Resp = calories

b) $\hat{y} = 210.95 + 11.056x$

$\hat{y} = a + bx$
Variables: x, y
Coeff: a, b

variables
x = fat(g)
$\hat{y}$ = pred. calories

c) $r = \sqrt{.923} = .961$

corr. coeff.

d) There's a strong, positive, linear assoc. between fat + calories.

r^2 = 92.3% of the variability in (y) calories explained by LSRL (line) with (x) fat.

e) $\hat{y} = 210.95 + 11.056(33)$

$\hat{y} = 575.79$ calories

pred.

f) ~~3X~~ 33 g and 580 calories

actual

Resid = actual − pred.
= 580 − 575.79
= 4.21

g) line under-predicts
actual > pred
positive Resid

Sum of Resids = 0

$(\bar{x}, \bar{y})$ on line

$\hat{y} = a + bx$

$b = r \cdot \frac{s_y}{s_x}$

$a = \bar{y} - b\bar{x}$

Area/ Price

C.I. for $\boxed{\text{slope:}}$

Statistic $\pm$ (crit. value) (st. dev. of stat)

$91.36 \pm (2.306)(13.52)$

df = n-2
= 10-2

(60.18, 122.54)

95%

95% conf. the actual slope is in C.I.

t-test for slope

pop. slope

$H_0: \beta = 0$ (no assoc.)

$H_a: \beta > 0$ (is a pos. assoc.)

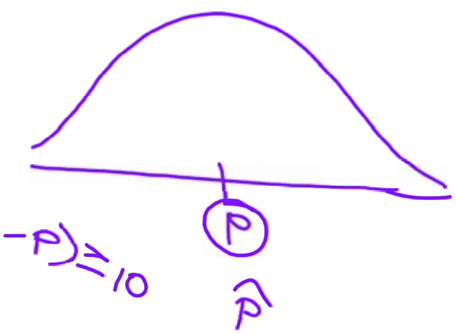
$t = \frac{91.36 - 0}{13.52}$ 6.76

p-value = .0001

with a p-value ≈ 0, this is sign. @ any reasonable level. Reject H_0. There's evid. that there is a pos. assoc. size & price of house.

$\boxed{\text{CLT}}$

CLT says the sampling dist. of $\hat{p}$ will be:



- approx. normal
  if $np \geq 10$ $n(1-p) \geq 10$
- Center: $\mu_{\hat{p}} = p$
- Spread: $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$

CLT

sampl. dist. of $\bar{x}$:

- approx. normal if $n \geq 30$ or pop. is normal
- $\mu_{\bar{x}} = \mu$
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

1 prop

n(1

2 prop

1 - prop

- $np \geq 10$
- $n(1-p) \geq 10$
- SRS from pop. of int.

2 - prop

(or 5)

- $n_1 p_1 \geq 10$
- $n_1(1-p_1) \geq 10$
- $n_2 p_2 \geq 10$
- $n_2(1-p_2) \geq 10$
- 2 ind SRS from pop...

1 - mean

- $n \geq 30$
  or pop. is normal
  (graph sample, comment shape, if normal pop. is reasonable)
- SRS from pop.

2 - mean

- n_1 & $n_2 \geq 30$ or both normal pop.
- 2 SRS from pop.