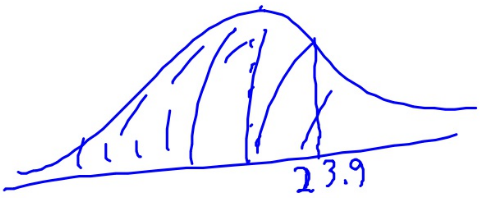
Warm Up

The army reports that the distribution of head circumference among the male soldiers is approximately Normal with mean 22.8 inches and standard deviation 1.1 inches.

A male soldier whose head circumference is 23.9 inches would be at what percentile? Show your work.

$$z = \frac{23.9 - 22.8}{1.1} = 1.0$$



$$P(z < 1) = 84\%$$

~~Tue~~. Oct. ~~13~~

Objective: SWBAT create scatterplots from given data
SWBAT calculate linear regressions and evaluate their fitness

Agenda:
- Warm Up
- Refresh from Fri.
- Notes
- Practice
- Reflection

Refresh from Friday

where we left off...

scatterplots: - direction
- outliers
- form (linear or not)
- strength (how tightly grouped)

correlation
r - correlation coefficient, from -1 to +1
close to ± 1: strong corr.
close to 0: weak corr.
close

Notes: Regression Lines

Not only do we want to know about a correlation, we want to predict things. We will use a regression line with the equation

$$\hat{y} = a + bx \quad \text{or}$$

$$\hat{y} = b_0 + b_1 x$$

where
- $\hat{y}$ (y-hat) is the predicted value of the response y for a given value of x
- a is the y-intercept (on the AP test, written as b_0)
- b is the slope, how much y is predicted to change when x increases by 1 (on the AP test, written as b_1)

ex) $\hat{y} = \$1.04 + \$0.13x$

Example

both examples on pg. 166

1) $\widehat{\text{fat gain}} = 3.505 - 0.00344(\text{NEA change})$

$a = 3.505$ kg → predicted fat gain if NEA doesn't change

$b = -0.00344$ → predicted fat gain drops by 0.00344 kg for each calorie of NEA increase

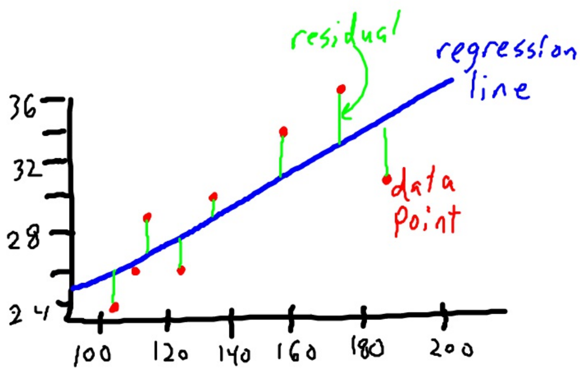
2) $\widehat{\text{fat gain}} = 3.505 - 0.00344(400) \approx 2.13$ kg

Notes: Residuals

Because life doesn't always match predictions, we have errors. For regressions, we call this a residual.

A residual is the difference between an observed value and the value predicted by the regression line.

residual = $y - \hat{y}$



we want to keep these small!

Notes: Regression in the Calculator

The calculator does what's called "least squares regression," which means it's finding the line that makes the sum of the squared residuals, $(y - \hat{y})^2$, as small as possible.

How do we get it?

1. put your data in a list

2. make sure the stat diagnostics are on

3. use LinReg(a + bx)

ex)Technology Corner on pg. 170

Notes: r^2

There are two ways of thinking about r^2:

1. the percent of variation in predicted y-values that can be explained by the regression of y on x

2. the percent of error for predictions of y that is removed by using the regression of y on x, *compared to using the mean*

r^2 goes from 0 to 1, and bigger is better

Practice

Analyzing Bivariate Data Practice from AP Stats Monkey

Reflection

How do we know if two variables are associated with each other?