2/4/15

Announcements

No office hours today 2/4

Extra office hours Monday 2/9 from 3:30-5:30

Homework due on Tuesday (slide under door if not there, will be collected Wed. morning)

Notes

Today: Best-fit lines, error, higher power exponents

- Linear: $y_i = B_0 + B_1x_i$
- Quadratic: $y_i = B_0 + B_1x_i + B_2x_i^2$
- Interval prediction: prediction w/ error: $y = B_0 + B_1x_i \pm \varepsilon$
- **IMPORTANT:** These are all naïve prediction intervals (for now)
    - We are assuming that we know the actual values B_0 and B_1 when we are really only calculating them from a sample (more uncertainty)

- We will be start dealing much more with the correlation coefficient *r* and its ability to predict variation using r^2

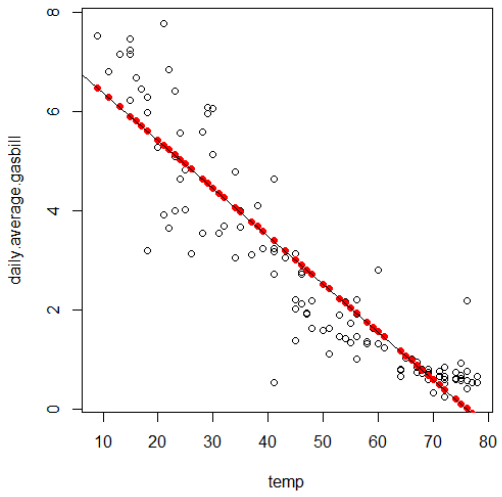
$$r^2 = \frac{Var(\hat{y})}{Var(y)} = \frac{Var(Part)}{Var(Total)} = \frac{[\sigma(\hat{y})]^2}{[\sigma(y)]^2}$$

- Related to Pythagorean theorem:
    - $c^2 = a^2 + b^2$
    - $[\sigma(y)]^2 = [\sigma(\hat{y})]^2 + [\sigma(\varepsilon)]^2$

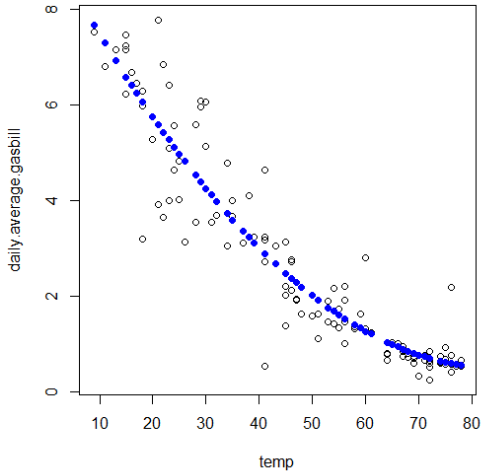
In-class datasets

utilities.csv

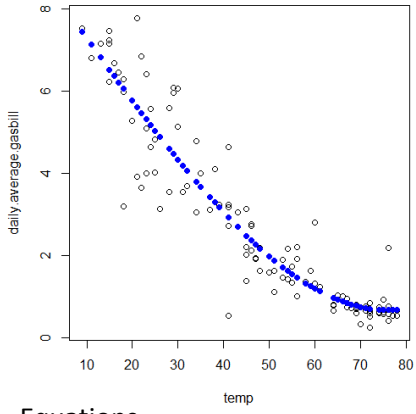
Linear fit:



Quadratic fit:



Cubic fit:



Equations

Linear

Quadratic

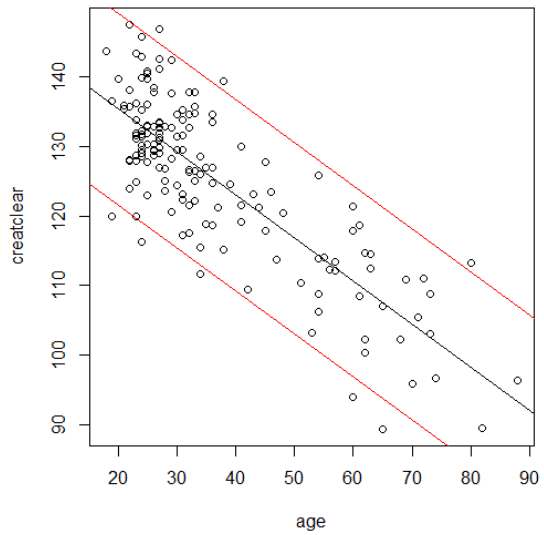
Cubic

```
lm1=lm(daily.average.gasbill ~ temp, data=utilities)
lm2=lm(daily.average.gasbill ~ temp + I(temp^2), data=utilities)
lm3=lm(daily.average.gasbill ~ temp + I(temp^2) + I(temp^3), data=utilities)
```

Plot w/ color:

```
points(fitted(lm2)~temp, data=utilities, col='green', pch=19)
```

creatinine.csv



Find error sigma) and coefficients (betahat)

```
betahat=coef(lm1)
newx=50
yhat=betahat[1]+betahat[2]*newx
yhat

sigma=sd(resid(lm1))
sigma
yhat - 2*sigma
yhat + 2*sigma
```

Plot w/ 2*sigma error margins

```
# Plot the data and show the straight-line fit
plot(creatclear~age, data=creatinine)
abline(betahat[1], betahat[2])
# Now shift the intercept of fitted line up and down to get the interval bounds
abline(betahat[1] + 2*sigma, betahat[2], col='red')
abline(betahat[1] - 2*sigma, betahat[2], col='red')
```

Quantile function:

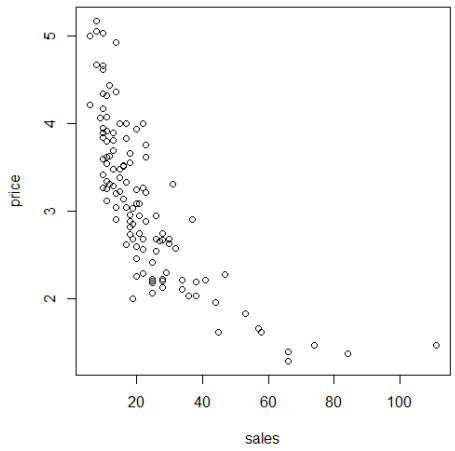
```
> quantile(creatinine$creatclear, probs=c(0.025, 0.975))
  2.5%  97.5%
96.34 143.44
```

milk.csv

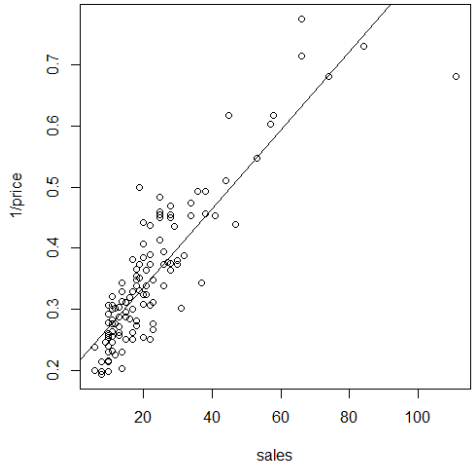
Objective: find best price to charge (most profitable) for a gallon of milk

First make plot linear (double logs can also be used):

```
plot(price~sales,data=milk)
```




```
plot(1/price~sales,data=milk)
lm1=lm(1/price~sales,data=milk)
abline(lm1)
```



Continue working on data outside of class, will come back to it on Monday