Class 2

- Power
- Estimation of subpopulation means
- Prediction intervals
- Confidence band for regression line

Power

- The *power* of a significance test is the probability that the null hypothesis is to be rejected when, in fact, it is false.
- This probability depends on the particular value of the parameter in the alternative space.

Power for β_1 (1)

- H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$
- $t = b_1/s(b_1)$
- $t^* = t(1-\alpha/2, n-2)$
- for α=.05 , we reject H_0 when $|t| \geq t^*$
- so we need to find $P(|t| \geq t^*)$ for arbitrary values of $\beta_1 \neq 0$
- when $\beta_1 = 0$, the calculation gives α

Power for β_1 (2)

- $t \sim t(n-2,\delta)$ – noncentral *t* distribution
- $\delta = \beta_1/\sigma(b_1)$ – noncentrality parameter
- We need to assume values for
- $\sigma^2(b_1)=\sigma^2/\Sigma(X_i - \bar{X})^2$ and n

Example of Power for β_1

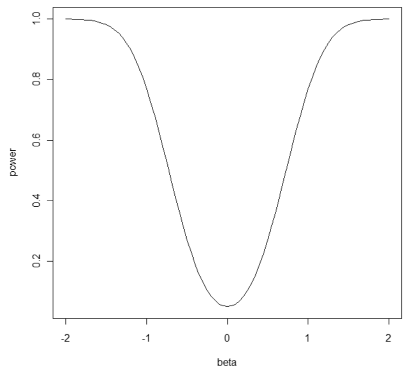
- we assume σ^2=2500 , n=25
- and $\Sigma(X_i - \bar{X})^2$ =19800
- so we have $\sigma^2(b_1)=\sigma^2/\Sigma(X_i - \bar{X})^2$= 0.1263

Example of Power (2)

- consider $\beta_1 = 1.5$
- we now can calculate $\delta = \beta_1/\sigma(b_1)$
- $t \sim t(n-2,\delta)$, we want to find $P(|t| \geq t^*)$
- we use a function that calculates the cumulative distribution function for the noncentral t distribution

```
sig2<-2500;
n<-25;
ss<-19800;
sigb2<-sig2/ss;
sigb<-sqrt(sigb2);
beta<-1.5;
delta<-beta/sigb;
alpha<-0.05;
cr<-qt(1-alpha/2,n-2);
power<-pt(-cr,n-2,delta)+1-pt(cr,n-2,delta);
power= 0.981209
```

```
beta<-seq(from=-2, to=2, by= 0.05)
delta<-beta/sigb;
power<-pt(-cr,n-2,delta)+1-pt(cr,n-2,delta);
plot(power~beta, type="l");
```



Estimation of $E(Y_h)$

- $E(Y_h) = \mu_h = \beta_0 + \beta_1 X_h$, the mean value of Y for the subpopulation with $X=X_h$
- we will estimate $E(Y_h)$ by
- $\hat{\mu}_h = b_0 + b_1 X_h$

Theory for Estimation of $E(Y_h)$

- $\hat{\mu}_h$ is normal with mean μ_h
- (it is an unbiased estimator)
- and variance $\sigma^2(\hat{\mu}_h)$=

$$\sigma^2\left[\frac{1}{n}+\frac{\left(X_h-\overline{X}\right)^2}{\sum\left(X_i-\overline{X}\right)^2}\right]$$

Theory for Estimation of $E(Y_h)$ (2)

- The normality is a consequence of the fact that $\hat{\mu}_h = b_0 + b_1 X_h$ is a linear combination of Y_i's

Application of the Theory

- we estimate $\sigma^2(\hat{\mu}_h)$ by
- $s^2(\hat{\mu}_h) = s^2\left[\frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum (X_i - \overline{X})^2}\right]$
- it follows that $t = \frac{\hat{\mu}_h - E(Y_h)}{s(\hat{\mu}_h)}$ ~ t(n-2)
- details for confidence intervals and significance tests are consequences

95% Confidence Interval for $E(Y_h)$

- $\hat{\mu}_h \pm t^* s(\hat{\mu}_h)$
- where t^* = t(.975, n-2)
- and $s(\hat{\mu}_h) = \sqrt{s^2(\hat{\mu}_h)}$

```
time<-read.table('CH01TA01.txt',
col.names=c("size", "hours"))
u<-order(time$size);
time<-time[u,];
time <- data.frame (time)
reg1<-lm(hours~size, time)
new <- data.frame(size = c(65,100));
c1<-predict.lm(reg1, new, se.fit=TRUE,
interval='confidence');
```

```
$fit
      fit      lwr      upr
1 294.4290 273.9129 314.9451
2 419.3861 389.8615 448.9106

$se.fit
       1         2
 9.917579 14.272328
```

Prediction of $Y_{h(new)}$

- $Y_h = \beta_0 + \beta_1 X_h + \xi$ the value of Y for a new observation with $X=X_h$
- $Var(Y_h - \hat{\mu}_h) = Var\ Y_h + Var\ \hat{\mu}_h = \sigma^2 + Var\ \hat{\mu}_h$
- $S^2(pred) = s^2\left[1 + \frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum (X_i - \overline{X})^2}\right]$

$(Y_h - \hat{\mu}_h)/s(pred)$ ~ t(n-2)

Prediction of Y_h (notes)

- Procedure can be modified for the mean of *m* observations at $X=X_h$

```
c2<-predict.lm(reg1, new, se.fit=TRUE,
interval='prediction')


$fit
      fit       lwr       upr
1 294.4290 191.3676 397.4904
2 419.3861 314.1604 524.6117

$se.fit
       1         2
 9.917579 14.272328
```

Notes

- The standard error (`Std Error Mean Predict`) given in this output is the standard error of $\hat{\mu}_h$, $s^2(\hat{\mu}_h)$, not $s^2(pred)$
- The prediction interval is wider than the confidence interval

Confidence band for regression line

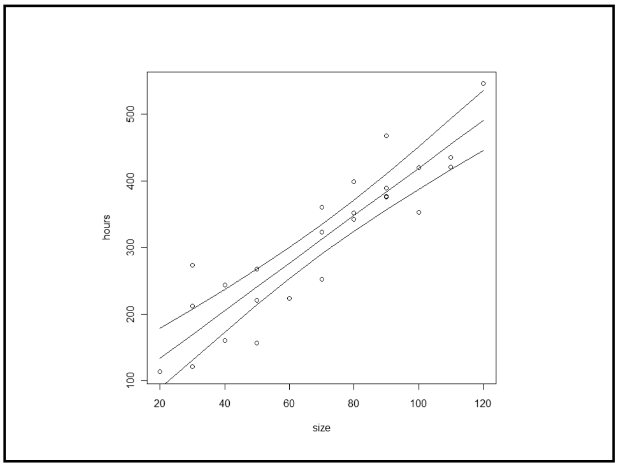
- $\hat{\mu}_h \pm Ws(\hat{\mu}_h)$
- where $W^2=2F(1-\alpha; 2, n-2)$
- This gives intervals for *all* X_h
- Boundary values define a hyperbola

Confidence band for regression line

- Theory comes from the joint confidence region for (β_0, β_1) which is an ellipse
- We can find an alpha for t^* that gives the same results
- We find W^2 and then find the alpha for $t^*=W$.

```
n<-25;
alpha<-.1;
dfn<-2;
dfd<-n-2;
w2<-2*qf(1-alpha,dfn,dfd);
w<-sqrt(w2);
level1<-1-2*(1-pt(w,dfd));
 level1= 0.966260
```

```
c1<-predict.lm(reg1, se.fit=TRUE,
interval='confidence', level=level1)
plot(hours~size, time)
lines(c1$fit[,1]~size, time)
lines(c1$fit[,2]~size,time)
lines(c1$fit[,3]~size,time)
```



hours
size
100
200
300
400
500
20
40
60
80
100
120