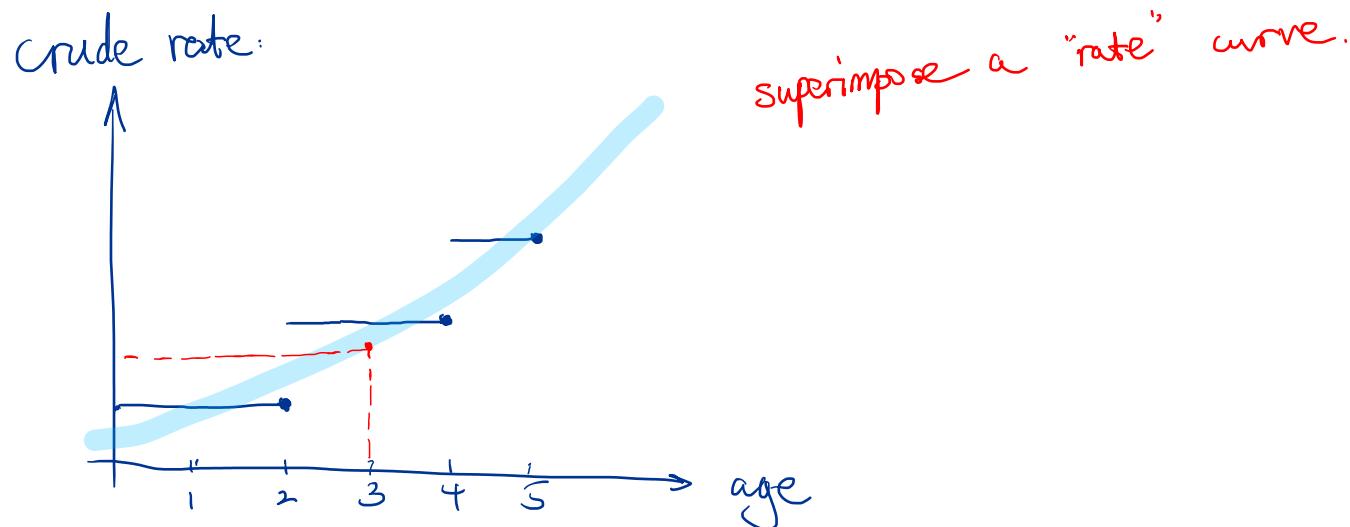


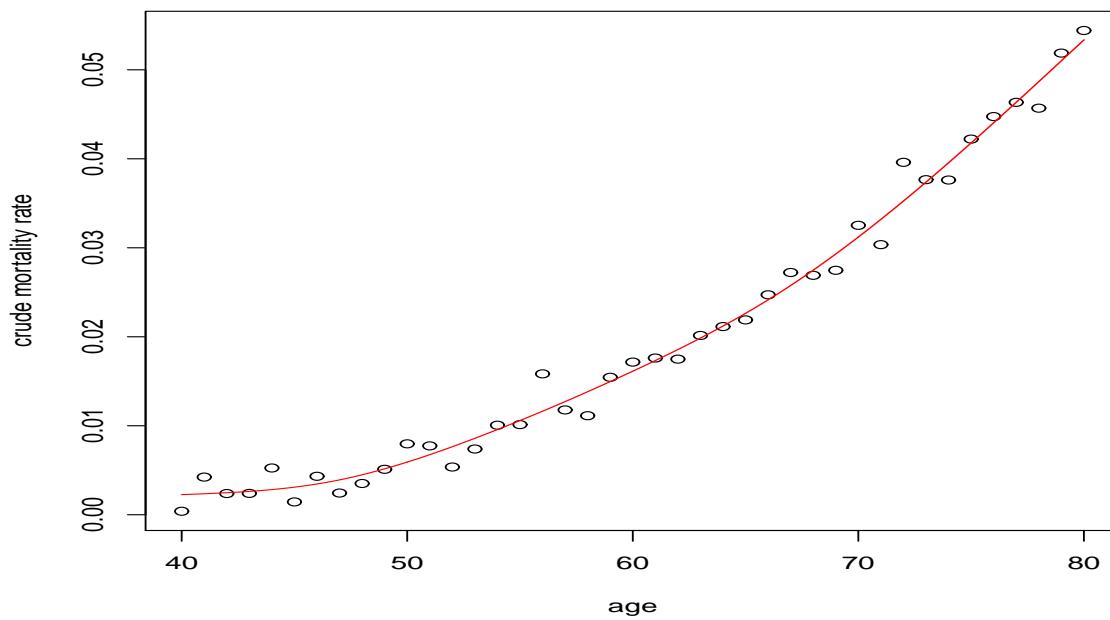
Survival Models: Week 9

Graduation and Statistical Tests

- Graduation is the process of **smoothing** a set of crude mortality rates.
- The crude mortality rates are estimated using observed data.
- Graduated rates are smoothed rates of mortality which can be used for calculating premium rates and for many other purposes in insurance.
- Why is graduation important?



Graduation and Statistical Tests



*lots of ways
to do so.
(next week)*

Figure 1: Plot of crude mortality rates. A possible set of graduated values are superimposed (red line)

Reasons for Graduation

1. Pooling of information.
2. Belief is that mortality rates are a smooth function of age.
3. Practical reasons - pricing of insurance.

Graduation Trade-off

When performing a graduation we are essentially trying to find a compromise between *smoothness* and *adherence* to the crude mortality rates. In statistics this is known as the “bias-variance trade-off”.

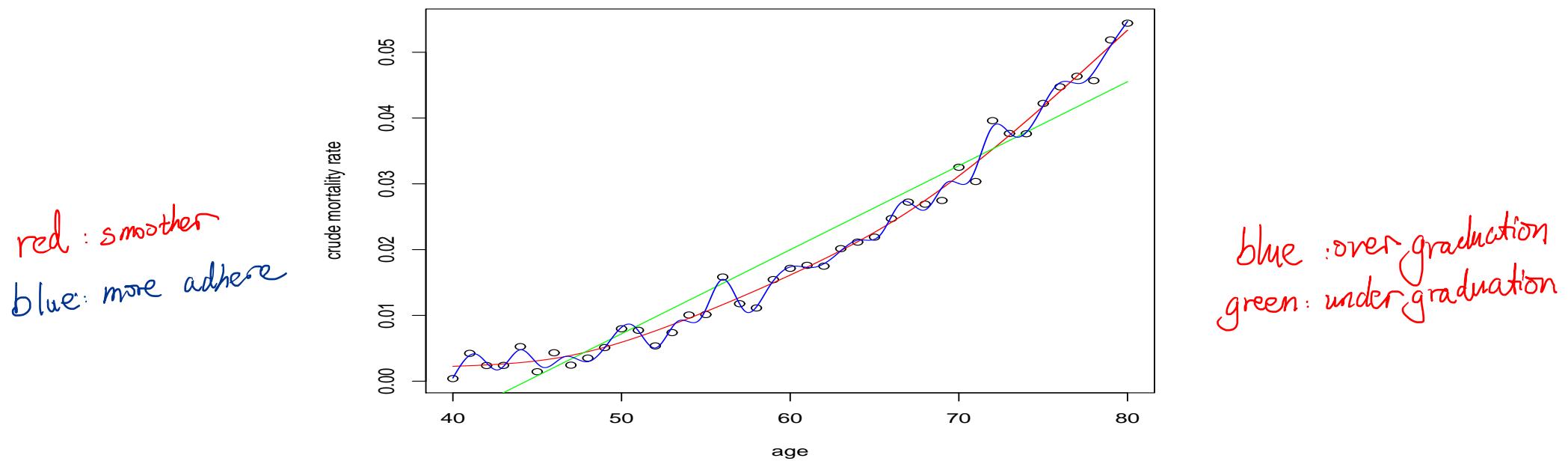


Figure 2: Examples of “over-” and “under-” graduation

Aside: Ridge Regression

not examinable

Note: Ridge Regression and Lasso Regression will not be examined.

Consider a standard linear regression model,

$$y_i = X_i^T \beta + \epsilon_i, \quad i = 1, 2, \dots, n$$

where y_i is the response variable on the i -th observation, $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ are the values of covariates, $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ are parameters. The parameter estimates for ridge regression solve the following:

$$\hat{\beta}^{\text{ridge}} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p (\beta_j)^2 \right\}.$$

The estimates are biased but can have much smaller variance. Ridge regression is particularly useful when multicollinearity is present.

Aside: Ridge Regression

```
#based on code from http://www.stat.sc.edu/~hitchcock
library(MASS)
body<-read.table("bodyfat.txt")
colnames(body)<-c('triceps', 'thigh', 'midarm', 'bodyfat')
round(cor(body),2)

> round(cor(body),2)
      triceps  thigh  midarm bodyfat
triceps    1.00   0.92   0.46   0.84
thigh      0.92   1.00   0.08   0.88
midarm     0.46   0.08   1.00   0.14
bodyfat    0.84   0.88   0.14   1.00
attach(body)
```

```

> select(lm.ridge(bodyfat ~ triceps + thigh + midarm,
lambda = seq(0,1,0.02)))
modified HKB estimator is 0.008505093
modified L-W estimator is 0.3098511
smallest value of GCV at 0.02 #optimal value of lambda

fit.ridge<-lm.ridge(bodyfat ~ triceps + thigh + midarm, lambda = .019)
fit.lm<-lm(bodyfat ~ triceps + thigh + midarm)
round(coef(fit.ridge),2)
round(coef(fit.lm),2)
> round(coef(fit.ridge),2)
      triceps    thigh   midarm
  43.84     2.12   -0.96   -1.02
> round(coef(fit.lm),2)
(Intercept)      triceps        thigh       midarm
  117.08        4.33       -2.86       -2.19

```

Aside: Lasso Regression

The parameter estimates for lasso regression solve the following:

$$\hat{\beta}^{\text{lasso}} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}.$$

The lasso has the attractive property that it sets many of the parameter estimates to exactly zero. This is a particularly useful property for high-dimensional data - data where p is greater than n .

Example: Lasso Regression

As an example, we will fit the lasso to a dataset of size $n = 50$ and $p = 1000$.

```
library(glmnet)
#generating data
x<-matrix(rnorm(1000*50),50,1000)
y<-2*x[,1]+3*x[,2]+4*x[,3]+10*x[,100]+rnorm(50)
#fitting the lasso
fit<-glmnet(x,y)
plot(fit,xvar="lambda")
fitcv<-cv.glmnet(x,y)
abline(v=fitcv$lambda.min)
```

Example: Lasso Regression

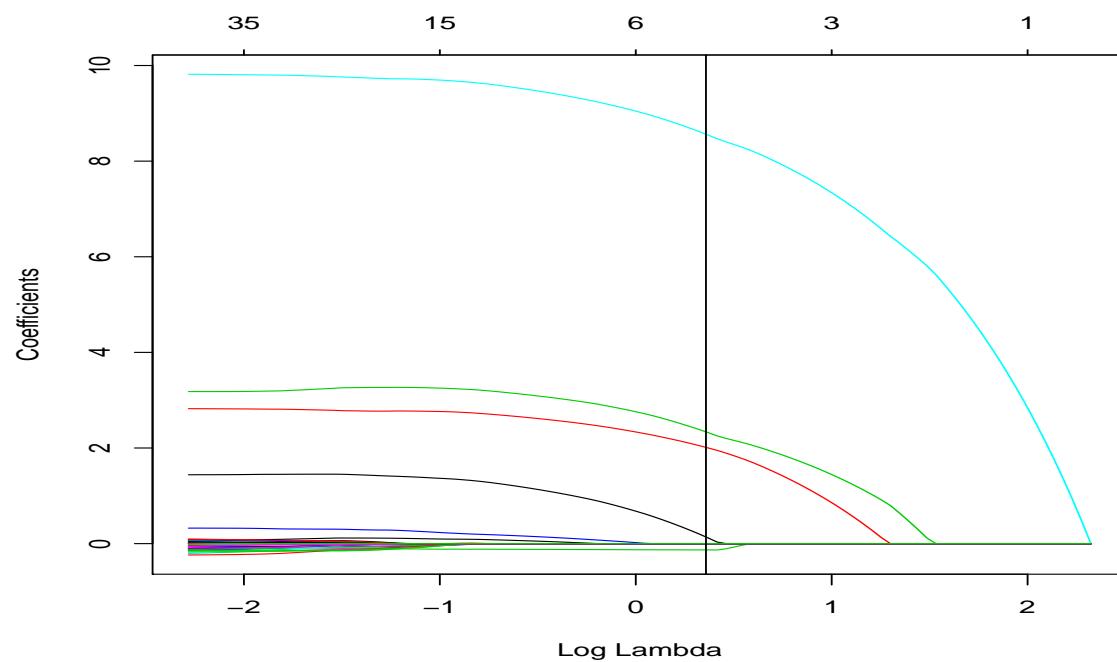
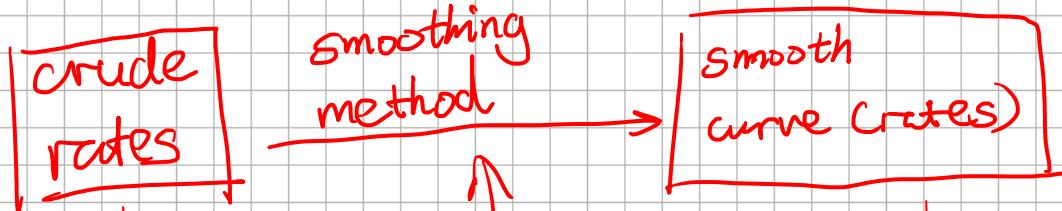


Figure 3: The lasso regularization path



week 9 & 10



the rates
you've
seen in
week 3

comparing these rates (week 9)

statistical tests.

(e.g. if the standard table
is describing the experience.)

Graduation and Statistical Tests

- We will talk more about how to perform a graduation later.
- Our focus now will be on statistical tests of whether a set of crude-mortality rates are consistent with a *standard table* or a set of smoothed (or graduated) rates.
- There will be minor modifications to some tests depending on whether we are using a standard table or graduated rates. We will focus on the case of using standard tables.

Standard Tables

- An insurance company may use a set of standard tables (ALT 2005-2007 or ELT14 (Males), for example) for pricing premiums.
- From time to time the company may want to see whether the standard table is describing the experience of their policy holders well enough.
- This process of “checking” the standard tables can be conducted by obtaining crude rates of mortality from the company’s policy holders and comparing these values to the standard table.

Standard Tables

Example: A particular company uses ALT 2005-2007 to price its insurance. To test the validity of ALT 2005-2007 the company observed the experience of its members over a given period. The full table is available on wattle:

X	E_x	Experience		Standard		
		Exposed	Deaths	ALT05-07	Expected	Variance
1	2	3	4		Deaths	
20	5000	2	0.00074	3.70	3.70	0.78
21	4500	2	0.00076	3.42	3.42	0.59
22	5500	6	0.00077	4.24	4.23	0.74
23	10000	10	0.00079	7.90	7.89	0.56
24	6250	3	0.00081	5.06	5.06	0.84
25	11000	13	0.00083	9.13	9.12	1.64
..

Chi-Square Test

The chi-square test provides an overall assessment of the “closeness” of the observed mortality and that expected given the standard table.

Null: The mortality rates from the standard table are appropriate.

Alt: The mortality rates from the standard table are not appropriate.

Under the null hypothesis the observed number of deaths is binomial(E_x, q_x).

The Chi-square test-statistic is: $\chi^2 = \sum_{\text{age groups}} \frac{(\theta_x - E_x q_x)^2}{E_x q_x}$.

Under the null hypothesis χ^2 has a chi-square distribution with degrees of freedom equal to the number of age groups. [Note: if we are using graduated rates a degree of freedom is lost for every parameter estimated in the graduation.]

chi-square test

$$D_x \sim \text{Bin}(E_x, g_x)$$

$$E(D_x) = E_x g_x \approx 1$$

$$\text{Var}(D_x) = E_x g_x (1 - g_x)$$

$$\approx E_x g_x$$

Normal Approximation (CLT)

$$Z_x = \frac{D_x - E(D_x)}{\sqrt{E_x g_x}} \sim N(0, 1)$$

D_x : observed deaths

E_x : expected deaths

$Z_i \sim N(0, 1)$

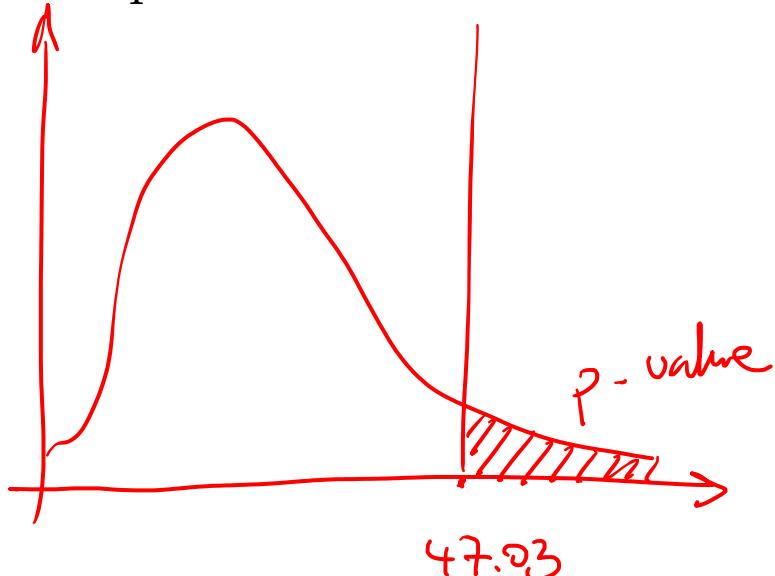
$$\chi^2_n = \sum_{i=1}^n Z_i^2 \sim \chi^2_n \quad n: \# \text{ of age groups}$$

Example - Chi-Square Test

For our example, $\chi^2(39) = 47.03$. The degrees of freedom is 39 due to the fact that there are 39 age groups. To compute the p-value for this test we need to compute the area to the right of 47.03 for a chi-square distribution with 39 degrees of freedom.

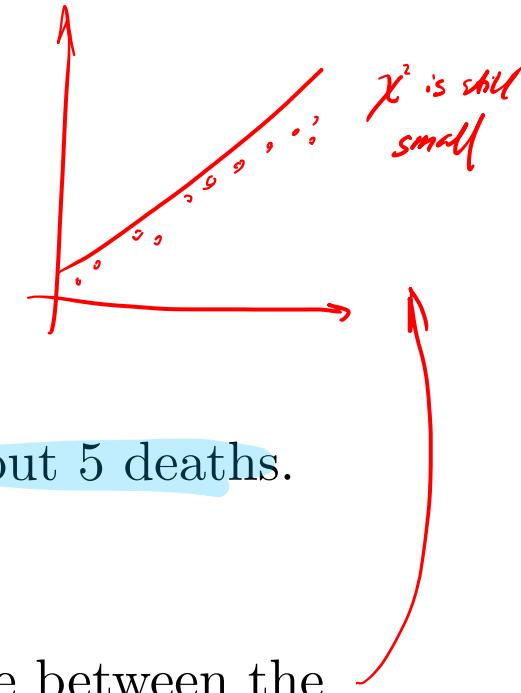
```
> 1-pchisq(47.03,39)  
[1] 0.1767697
```

The p-value of 0.18 means that we fail to reject the null hypothesis.

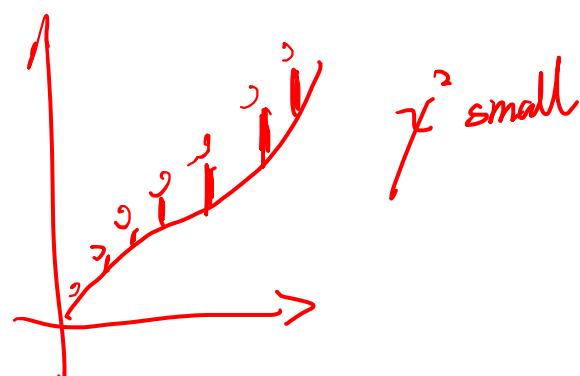


Comments - Chi-Square Test

- The individuals in each age group should be homogenous.
- A “rule of thumb” is that each age group should contain about 5 deaths.
- The χ^2 is an omnibus or overall test.
- *Upside* of χ^2 test is that it can detect even a single difference between the observed and expected deaths.
- *Downside* of χ^2 test is that it can fail to detect certain “issues”. For example, all the observed deaths are slightly greater than the expected deaths.
- May not detect large groups of positive or negative deviations.



Should use other
tests!



Standardized Deviation Test

Under the null hypothesis that the mortality rates from the standard table are appropriate, the quantity:

$$Z_x = \frac{\theta_x - E_x q_x}{\sqrt{(E_x q_x (1 - q_x))}},$$

has an approximate standard normal distribution. We can use this fact to see whether the observed values of Z_x behave like standard normal random variables. This can be done in the following way:

1. Decide on a set of intervals. For example: $(-\infty, -2]$, $(-2, 0]$, $(0, 2]$, $(2, \infty)$.
2. Compute the **observed number** of Z_x falling in each interval.
3. Compute the **expected number** of Z_x falling in each interval, under the null hypothesis. The expected value is the number of age ranges times the appropriate area from the standard normal.

4. Use a χ^2 test to compare the observed and expected values.

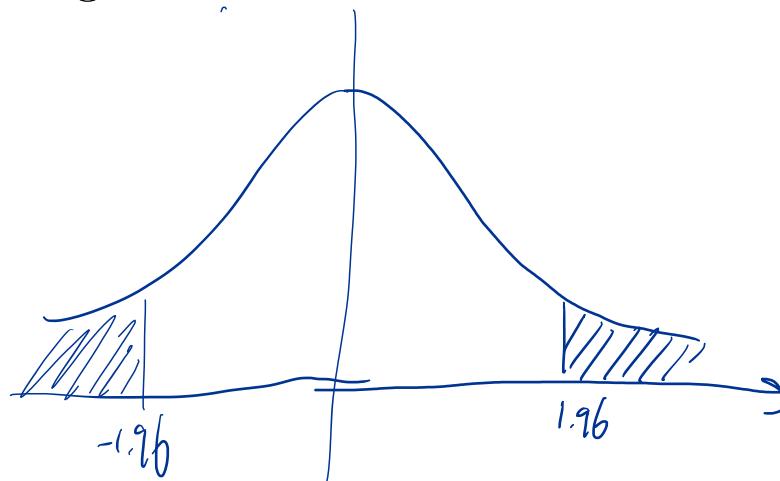
The number of degrees of freedom for this test is the number of intervals less one
- the total of the observed and the total of the expected must be the same.

This test can detect situations of a few “large” deviations being “masked” by an excessive number of “smaller” deviations.

Example Standardized Deviation Test

For our example the values falling in each interval are:

$(-\infty, -2]$	1	0.975
$(-2, -0]$	20	18.525
$(0, 2]$	17	18.525
$(2, \infty)$	1	0.975



$$\chi^2(3) = \frac{(1 - 0.975)^2}{0.975} + \frac{(20 - 18.525)^2}{18.525} + \frac{(17 - 18.525)^2}{18.525} + \frac{(1 - 0.975)^2}{0.975} = 0.34$$

The p-value for this test is 0.95. Again, we fail to reject the null hypothesis.

Note: ideally we would want the counts to be greater than 5. Can combine rows if necessary.

Sign Test

Under the null hypothesis that the mortality rates from the standard table are appropriate, we would expect there to be roughly a 50% chance that any given z_x is positive. Under this assumption, the number of positive deviations is Binomial($m, 1/2$), where m is the number of age ranges. The sign test is conducted by comparing the observed number of positive deviations to the Binomial($m, 1/2$) distribution. Typically, in order to find p-values we will approximate the Binomial distribution by the Normal($0.5m, 0.25m$) distribution.

This is a two-sided test - both too many or two few positive deviations are evidence against the null hypothesis.

The sign test can detect an excessive number of positive or negative deviations.

Example: Sign Test

For our example there are 18 positive deviations over the 39 age groups. Using the Normal approximation our test statistic is: $\frac{18-39/2}{\sqrt{9.75}} = -0.48$. Clearly, in this case our p-value will be greater than 0.05.

Note: More formally we should use the continuity correction when using the normal approximation here.

*Generally, you'll get the same conclusion
with sign test & standard deviation test.*

Cumulative Deviation Test

Under the null hypothesis that the mortality rates from the standard table are appropriate, and deaths at successive ages are independent, we have that the quantity:

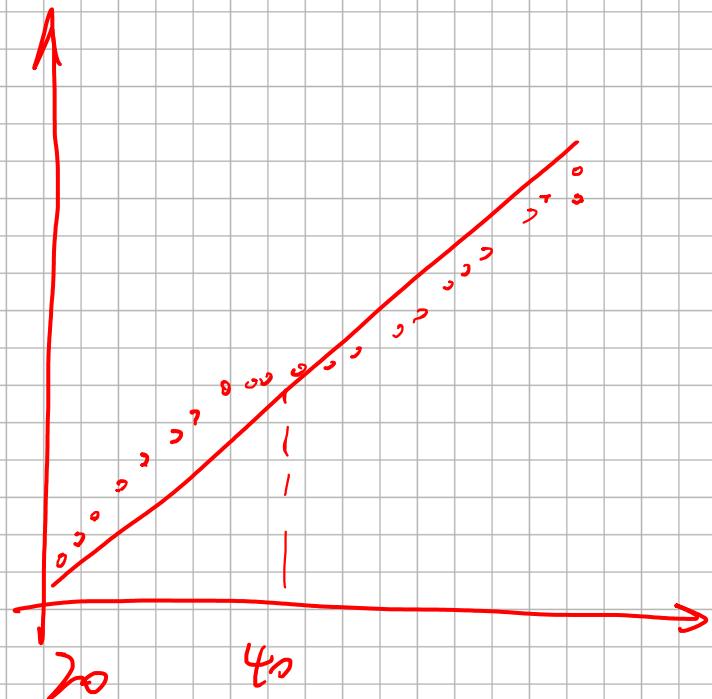
$[a, b]$ is an
age interval.

$$\frac{\sum_a^b (\theta_x - E_x q_x)}{\sqrt{\sum_a^b (E_x q_x p_x)}}$$

is approximately standard normal.

This test can be applied to the whole age range or over particular age ranges.

This is a two-sided test.



- sign test cannot say anything,
(+ as many as -)
- standardized ~~normal~~ ^{deviation} test cannot say
anything, generally fits normal distribution

USE cumulative deviation test.
↓

b/c cumulative is always positive,
not normal.

Example Cumulative Deviation Test

We will apply the cumulative deviation test over the ages ranges 20-50 and 51-65.

$$\frac{\sum_{20}^{40}(\theta_x - E_x q_x)}{\sqrt{\sum_{20}^{40}(E_x q_x p_x)}} = 0.66$$

$$\frac{\sum_{41}^{58}(\theta_x - E_x q_x)}{\sqrt{\sum_{41}^{58}(E_x q_x p_x)}} = -0.66$$

The cumulative deviations over both age ranges do not provide evidence against
the null.

Note: need to be careful when simultaneously conducting multiple hypothesis tests.

Aside: Example of Multiple Testing

Note: Multiple Testing will not be examined.

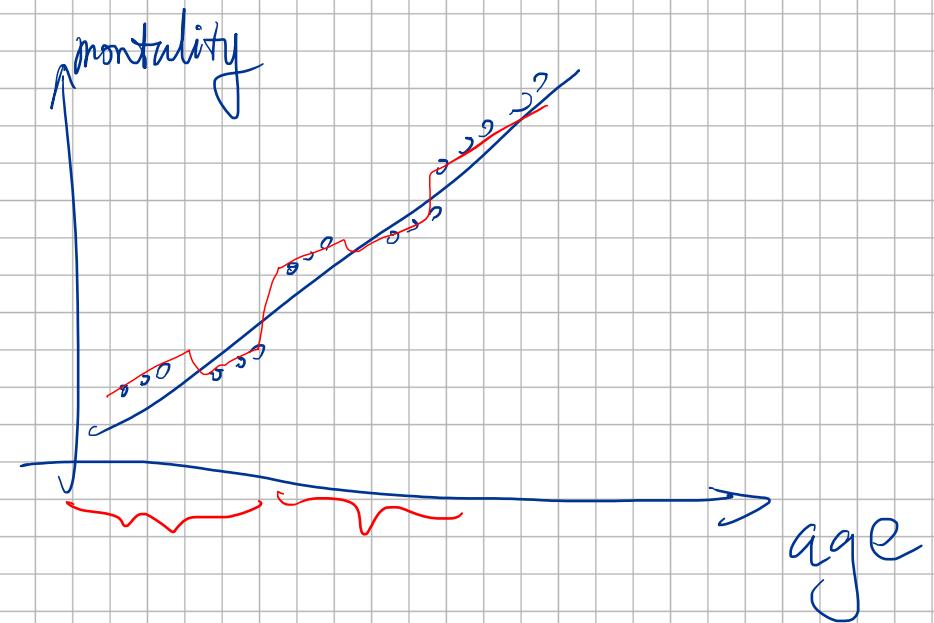
The R code below looks at how often we find a significant result when we simultaneously conduct 5 or 40 hypothesis tests at the 5% level of significance. In particular, we fit a regression model and test whether each of the covariates is related to the response. The data is generated such that *none* of the covariates are actually related to the response.

```
> set.seed(123)
> x<-matrix(rnorm(50*5),50,5)
> res<-rep(0,1000)
>
> for(i in 1:1000) {
+
+ y<-rnorm(50) #none of the 5 covariates related to response
+ fit<-lm(y~x)
```

```
+ temp<-summary(fit)$coef[-1,3] #looking at test
statistic for each covariate
+ temp<-ifelse(abs(temp)>1.96,1,0)
+ temp<-sum(temp)
+ res[i]<-ifelse(temp>0,1,0) #1 if at least one
covariate found to be sig at 5% level
+
+ }
>
> sum(res)/1000
[1] 0.23
> (1-0.95^5)
[1] 0.2262191
>
> #even more tests
>
> x<-matrix(rnorm(100*40),100,40)
```

```
> res<-rep(0,1000)
>
> for(i in 1:1000) {
+
+ y<-rnorm(100) #none of the 40 covariates related to response
+ fit<-lm(y~x)
+ temp<-summary(fit)$coef[-1,3] #looking at test
statistic for each covariate
+ temp<-ifelse(abs(temp)>1.96,1,0)
+ temp<-sum(temp)
+ res[i]<-ifelse(temp>0,1,0) #1 if at least one
covariate found to be sig at 5% level
+
+ }
>
> sum(res)/1000
[1] 0.812
> (1-0.95^40)
```

[1] 0.8714878



might need a less smooth curve
to describe the experience.

- chi-square test might not detect problem since sd is small.
- if appropriate cumulative deviation test

chosen, you might find the problem
but if choose like — —
You could not.
IT DEPENDS.

"grouping effect"

Runs Test

This test looks at whether there are excessive groupings of positive (or negative deviations). Again, if the standard tables are appropriate we would expect to see the number of positive deviations randomly scattered amongst the negative deviations (e.g. we should see a large number of runs). The runs test is based on the observed number of “runs” of positive deviations. For example, if we had deviations of $1, 2, -1, -2, 3, 1$ there are two “runs” of positive deviations. It can be shown that the probability of seeing t or less positive runs is:

$$\sum_{j=1}^t \frac{\binom{n_1-1}{j-1} \binom{n_0+1}{j}}{\binom{m}{n_1}},$$

where, n_1 and n_0 are the number of positive and negative deviations. The value m is the number of age groups. For large values of m a normal approximation can be used to compute the p-value, rather than the above quantity. For large m ,

Runs Test

Def of "Run"

of grouping of positive deviation

(1, 2), -1, -2, (3, 1)

2 runs (positive)

1 negative run

(0.1, 0.1, 0.2), -0.1, (0.1, 0.2), -0.1

2 runs

runs test \rightarrow # of + runs is large enough?

If # of + runs is too small, we reject H_0 .

e.g. (0.1, 0.1, 0.2, 0.3), -0.2, -0.5, -0.3

1 run

n_+ : # of positive deviations

n_- : # of negative deviations

m : # of age groups

N : # of positive runs

$$P(N=j) = ?$$

① $\binom{m}{n_+}$ # of ways to pick n_+ # of positive deviations

② $n_+ \rightarrow j$ groups

e.g. 1, 2, 3, 4

$n_+ - 4$ positive sd } how many ways?

$j=2$ groups }

$$\binom{n_+-1}{j-1}$$

'inserting sticks!'

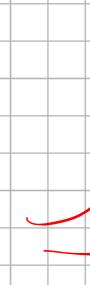


③ Insert j groups into n_- # of negative deviations

0-1, 0-2, 0-3 0

e.g. $j=2$, $n_+=4$, $n_-=3$

$$\binom{n_-+1}{j}$$



$$\begin{aligned} P\text{-value} &= P(N < ts) \\ &= P\left(\frac{N - E(N)}{\sqrt{\text{Var}(N)}} < 0\right) \end{aligned}$$

$$P(N=j)$$

$$= \frac{\binom{n_+-1}{j-1} \binom{n_-+1}{j}}{\binom{m}{n_+}}$$

$N \sim$ hypergeometric dist

$$E(N) = \frac{n_+(n_-+1)}{n_++n_+}$$

$$\text{Var}(N) = \frac{(n_+ n_-)^2}{(n_++n_+)^3}$$

t is approximately normal with mean $n_1(n_0 + 1)/(n_0 + n_1)$ and variance $(n_0 n_1)^2/(n_0 + n_1)^3$.

The runs test is a one-sided test - a small number of runs is evidence against the null hypothesis.

Example Runs Test

For our example $m = 39$, $n_1 = 18$, and $n_0 = 21$. Using the normal approximation we have the following test-statistic:

$$\frac{9 - 10.2}{\sqrt{2.4}} = -0.77,$$

the p-value is given by $p(Z < -0.77)$ which is clearly greater than 0.05. Again, we fail to reject the null hypothesis.

χ^2 -test $\circlearrowleft \chi_{(m)}^2 \rightarrow$ one-sided

std-dev-test $\circlearrowleft \chi_{(v)}^2 \rightarrow$ one-sided

under Normal assumption

Cum-dev-test $\circlearrowleft \rightarrow$ separate groups

(2-sided)

sign-test $\circlearrowleft \rightarrow$ two sided

runs test $\circlearrowleft \rightarrow$ one sided