Was the fall in Scottish life expectancy in 2014 a chance event?

rm(list = ls())  
  
pacman::p\_load(  
 tidyverse  
)  
  
dta\_e0 <- read\_csv("hmd\_explorer/data/hmd\_e0.csv")

## Parsed with column specification:  
## cols(  
## code = col\_character(),  
## year = col\_integer(),  
## gender = col\_character(),  
## e0 = col\_double()  
## )

# Introduction

* Risk of rare ecological events, such as a severe flood, are often expressed in terms of frequencies rather than probabilities. For example, a 1-in-100 year flood means that a flood of that severity or greater is expected to have a 1% probability of occurring in any particular year. However, this expectation is dependent on the models used to predict such risks, and these models are dependent on assumptions about how much the past is a reliable guide to the future.
* It is now becoming clear that many 1-in-100 year ecological events are occurring much more often than once every 100 years. For example [examples here] . The appearance of many very rare events in quick succession suggests that the models are wrong: what used to be a 1-in-100 year event is now a 1-in-20 year or 1-in-10 year event instead. The fundamentals have shifted, and the past is a less reliable guide to the future than it used to be.
* In recent years there have been concerns about life expectancies in Scotland either falling from the previous year, or not improving. This is against a backdrop of life expectancies tending to improve almost every year for many generations. [EXAMPLES]
* From 2014 to 2015 there was a notable fall in life expectancy in Scotland, by 0.30 years (3.6 months) for females and by 0.37 years (4.4 months) for males. This is in contrast to an average annual gain in life expectancy of 0.23 years (2.8 months) per year for females, and 0.23 years (2.7 months) per year for males, in Scotland since 1855; and an average gain of 0.18 years per year (2.2 months per year) for females and 0.20 years per year (2.4 months per year) for males, in the 60 year period between 1955 and 2014.
* The aim of this short report is to explore how likely it was that Scotland’s 2014-15 fall in life expectancy was simply a ‘chance event’, a bad outlier like a particularly severe flood, that ‘just happens’ once every few decades, without the long-term fundamental drift towards ever longer lives really having changed. We conclude by asking whether the fall in 2014-15, and (lack of) improvement in gains subsequently, should give us concern that, like ‘1-in-100 year’ floods occurring surprisingly frequently in the last decade or so, rare events like the 2014-15 life expectancy fall are no longer as rare as they used to be, and that the kinds of gains in longevity experienced by previous generations will not be experienced by subsequent generations.

dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(year == 2015) %>%   
 mutate(change\_e0\_months = change\_e0 \* 12)

## # A tibble: 2 x 6  
## # Groups: gender [2]  
## code year gender e0 change\_e0 change\_e0\_months  
## <chr> <int> <chr> <dbl> <dbl> <dbl>  
## 1 GBR\_SCO 2015 female 81.0 -0.300 -3.60  
## 2 GBR\_SCO 2015 male 77.0 -0.370 -4.44

# Methods

* Data on life expectancy at birth () in Scotland was extracted from the Human Mortality Database for all available years (First year 1855, last year 2016). The annual change in life expectancy () was then calculated for each year by sex. Change in life expectancy was then plotted against time, to help understand how unusual the 2014-15 fall in life expectancy was in a longer term context. Based on visual exploration of the data, the 60 year period 1955-2014 was selected as providing a range of observations both long-term enough, and recent enough, for the 2015 annual fall to be compared against.
* The rarity, probability and estimated frequency of the fall in 2014-15 was estimated using the data from 1955-2014 in two ways: firstly by counting the number of years where annual falls in life expectancy were as or more severe than that seen in 2014-15, and secondly by using sample means and standard deviations of the life expectancy changes to produce a normal distribution function for each sex, and calculating the cumulative probability of falls in life expectancy as or more severe than the 2014-15 fall from the 1955-2014 distribution of annual changes.

# Results

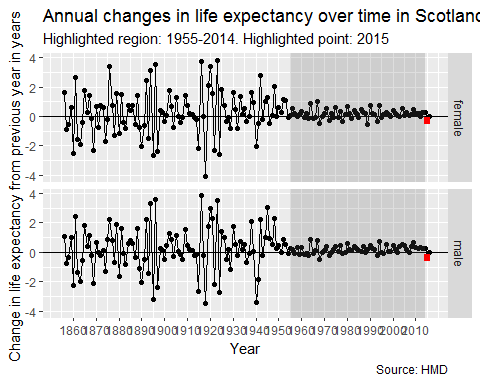
## Figure 1: Long term trend of change

* The figure below shows annual change in life expectancy in Scotland over time by sex. The point marking the 2014-15 fall is highlighted with a red point. Throughout the latter half of 19th century, and first half of the 20th century, there were much greater swings in annual changes in life expectancy than in the latter half of the 20th century and start of 21st century. Because high annual variability has not been observed for many decades, only data from 1955-2014 (highlighted grey) are used in subsequent analyses. [Though results with all data could be included in an appendix?] We can see from this figure that, though the 2014-15 fall in life expectancy is rare, it is not unprecedented in the 1955-2014 period.
* Some summary numbers?

dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 ggplot(aes(x = year, y = change\_e0)) +   
 geom\_line() + geom\_point() +  
 facet\_grid(gender ~ .) +   
 geom\_hline(yintercept = 0) +   
 scale\_x\_continuous(breaks = seq(1860, 2010, by = 10)) +   
 annotate("rect", xmin = 1955, xmax = 2014, ymin = -Inf, ymax = Inf, alpha = 0.2) +   
 labs(  
 title = "Annual changes in life expectancy over time in Scotland, 1855-2016",  
 subtitle = "Highlighted region: 1955-2014. Highlighted point: 2015",  
 caption = "Source: HMD",  
 x = "Year",   
 y = "Change in life expectancy from previous year in years"  
 ) +   
 geom\_point(  
 aes(x = year, y = change\_e0),  
 size = 2, shape = 15, colour = "red",  
 data = dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(year == 2015)  
   
 )

## Warning: Removed 1 rows containing missing values (geom\_path).

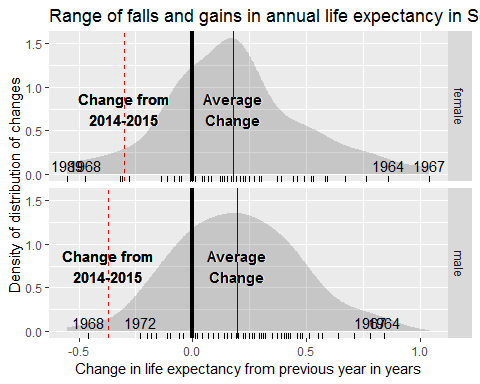
## Warning: Removed 2 rows containing missing values (geom\_point).



## Figure 2: Distribution

* Figure 2 below shows the distribution of annual changes in life expectancy over the 1955-2014 period. Each observation is represented as a vertical ‘tick’, and the years corresponding to the two largest annual increases and declines are labelled. The black vertical line indicates no change from the previous year, the thinner blue vertical line shows the average annual change over the period, and the dashed red line indicates the change observed in 2014-15. The grey shape above the points shows the density of the distribution.
* We can see from this that for males, only one year, 1967-68, has seen a sharper annual fall in life expectancy than the change in 2014-15. We can also see that 1967-68 was preceeded by an especially high rate of improvement the previous year, 1966-67. Over the 60 year period, only one year saw a faster fall in life expectancy for males, whereas for females four years saw greater annual declines than occurred in 2014-15.

dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(between(year, 1955, 2014)) %>%   
 ggplot(aes(x = change\_e0)) +   
 stat\_density(alpha = 0.2) +  
 geom\_rug() +   
 facet\_grid(gender ~ .) +  
 geom\_vline(xintercept = 0, size = 1.5) +  
 geom\_vline(  
 aes(xintercept =change\_e0),  
 data = dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(year == 2015),  
 colour = "red",  
 linetype = "dashed"  
 ) +  
 geom\_vline(  
 aes(xintercept =mean\_change\_e0),  
 data = dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(between(year, 1955, 2014)) %>%   
 group\_by(gender) %>%   
 summarise(mean\_change\_e0 = mean(change\_e0)) %>%   
 ungroup(),  
 colour = "blue"  
 ) +   
 labs(  
 title = "Range of falls and gains in annual life expectancy in Scotland, 1955-2014",  
 x = "Change in life expectancy from previous year in years",  
 y = "Density of distribution of changes"  
 ) +  
 geom\_text(  
 aes(x = mean\_change\_e0), y = 0.75,   
 label = "Average\nChange", fontface = "bold", colour = "white",  
 nudge\_y = 0.01, nudge\_x = 0.0025,  
 data = dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(between(year, 1955, 2014)) %>%   
 group\_by(gender) %>%   
 summarise(mean\_change\_e0 = mean(change\_e0)) %>%   
 ungroup()  
 ) +  
 geom\_text(  
 aes(x = mean\_change\_e0), y = 0.75,   
 label = "Average\nChange", fontface = "bold",  
 data = dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(between(year, 1955, 2014)) %>%   
 group\_by(gender) %>%   
 summarise(mean\_change\_e0 = mean(change\_e0)) %>%   
 ungroup()  
 ) +  
 geom\_text(  
 aes(x = change\_e0), y = 0.75,   
 label = "Change from\n2014-2015", fontface = "bold", colour = "white",  
 nudge\_y = 0.01, nudge\_x = 0.0025,  
 data = dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(year == 2015)  
 ) +  
 geom\_text(  
 aes(x = change\_e0), y = 0.75,   
 label = "Change from\n2014-2015", fontface = "bold",  
 data = dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(year == 2015)  
 ) +  
 geom\_text(  
 aes(x = change\_e0), y = 0.75,   
 label = "Change from\n2014-2015", fontface = "bold",  
 data = dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(year == 2015)  
 ) +  
 geom\_text(  
 aes(x = change\_e0, label = year), y = 0.1,  
 data = dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(between(year, 1955, 2014)) %>%   
 mutate(rank\_ch = rank(change\_e0)) %>%   
 filter(rank\_ch <= 2 | rank\_ch >= 59)  
 )



## Modelling

## How rare was the 2014-15 change? The simple approach

* As there was one event as or more severe than the 2014-15 fall for males, and four events as more severe than the 2014-15 fall for females, and there were 60 observations for each sex, the simple estimates for the 2014-15 fall are 1/60 (around 1.7%) for males and 4/60 (around 6.7%) for females.
* If needed, credible intervals for the above could be produced through bootstrapping. (i.e. repeatedly resampling from the same data).

## How rare was the 2014-15 change? A slightly more complicated approach

The table below shows the mean and standard deviation of the annual changes that occurred between 1955-2014. The average (mean) annual improvement was 0.18 years per year for females and 0.20 years per year for males, with a standard deviation of 0.31 for females and 0.27 for males. These values are used to estimate the probability, and so expected frequency, of the 2014-15 decline, using a normal distribution. This approach suggestes there is a 6.0% change of an event as or more severe than the 2014-15 decline for females, and a 1.7% probability of observing an event as or more severe than the 2014-15 decline for males. This suggests the 2014-15 fall was a 1-in-17 year event for females, and a 1-in-60 year event for males.

summary\_stats <- dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(between(year, 1955, 2014)) %>%   
 group\_by(gender) %>%   
 summarise(  
 mean\_ch\_e0 = mean(change\_e0),  
 sd\_ch\_e0 = sd(change\_e0)  
 ) %>%   
 ungroup()  
  
summary\_stats

## # A tibble: 2 x 3  
## gender mean\_ch\_e0 sd\_ch\_e0  
## <chr> <dbl> <dbl>  
## 1 female 0.179 0.309  
## 2 male 0.197 0.266

summary\_stats %>%   
 left\_join(  
 dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 filter(year == 2015) %>%   
 select(gender, change\_in\_2015 = change\_e0)  
 ) %>%   
 mutate(  
 prob\_2015\_chance = pmap\_dbl(  
 list(  
 q = change\_in\_2015,  
 mean = mean\_ch\_e0,  
 sd = sd\_ch\_e0  
 ),  
 .f = pnorm  
 ),  
 one\_in = 1/ prob\_2015\_chance,  
 prob\_no\_improvovement = pmap\_dbl(  
 list(  
 q = c(0,0),  
 mean = mean\_ch\_e0,  
 sd = sd\_ch\_e0  
 ),  
 .f = pnorm  
 )  
 )

## Joining, by = "gender"

## # A tibble: 2 x 7  
## gender mean\_ch\_e0 sd\_ch\_e0 change\_in\_2015 prob\_2015\_chance one\_in  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 female 0.179 0.309 -0.300 0.0605 16.5  
## 2 male 0.197 0.266 -0.370 0.0166 60.2  
## # ... with 1 more variable: prob\_no\_improvovement <dbl>

# Discussion

* The simple analyses above has indicated that the fall in life expectancy in Scotland from 2014-15 was rare but not unprecedented. It was a rare event for both sexes, but more so for males than females. Considering the ups and downs in life expectancy changes in Scotland from 1955-2014, a fall in life expectancy as or more severe than that seen in 2014-15 could be expected to occur once every 17 years on average for females, and once every 60 years on average for males.
* Limitations
* Comparison with other findings
* This result might be taken to suggest that the possibility that the 2014-15 was simply a ‘chance event’, rather than the start of a long-term slowdown or stagnation in improving life expectancy, cannot be ruled out. Variation in mortality rates from one year to the next can occur due to random seasonal variation in temperature and infectious diseases, especially influenza. [REFS]
* However, perhaps more unusual and concerning than the fall in life expectancy from 2014-15 was the lack of any substantial increase in life expectancy which occurred in the following year. This should be concerning because, unlike the very simple modelling approach used for illustration and estimation above, changes in life expectancy over a year are likely to be *negatively correlated* (rather than *uncorrelated*) with changes in the previous year.
* Analysis of annual mortality changes *does* suggest that the data for Scotland exhibits this negative correlation pattern. (See appendix?) This means the lack of improvement in life expectancy from 2015-16, after the substantial fall in life expectancy in 2014-15, is particularly unusual, and that recent trends in changes in life expectancy should be monitored and analysed very carefully to determine whether 2014-15 represents a change in the ‘population climate’, a chance event, or a combination of both.
* Final summary statement

# Appendix? Quick exploration: negative autocorrelation

The model above assumes that the change in life expectancy in any year is not dependent on that in the previous year. Put another way, the model is ‘memoryless’, and has not predictor terms (i.e. ‘right hand’ terms in regression formula). The next simplest model approach would be to build a model that includes the previous year’s change as a predictor variable. So, a one year lag term.

The prior hypothesis - consistent with ‘harvesting’ - is that the change in one year will be negatively correlated with the previous year.

This can be done by building a linear regression model with the previous year as a predictor variable

mdls <- dta\_e0 %>%   
 filter(code == "GBR\_SCO") %>%   
 filter(gender != "total") %>%   
 group\_by(gender) %>%   
 arrange(year) %>%   
 mutate(change\_e0 = e0 - lag(e0)) %>%   
 mutate(last\_change\_e0 = lag(change\_e0)) %>%   
 filter(between(year, 1955, 2014)) %>%   
 group\_by(gender) %>%   
 nest() %>%   
 mutate(mdl\_nolag = map(data, ~lm(change\_e0 ~ 1, data = .))) %>%   
 mutate(mdl\_lag = map(data, ~lm(change\_e0 ~ last\_change\_e0, data = .))) %>%   
 mutate(anova = map2(mdl\_nolag, mdl\_lag, anova)) %>%   
 mutate(p\_autocorrelation = map\_dbl(anova, ~.[2,6])) %>%   
 mutate(model\_lag\_tidy = map(mdl\_lag, broom::tidy))  
  
lag\_mdl\_summary <- mdls %>%   
 select(gender, model\_lag\_tidy) %>%   
 unnest()  
   
p\_autocorr\_female <- mdls %>% filter(gender == "female") %>% pull("p\_autocorrelation")  
p\_autocorr\_male <- mdls %>% filter(gender == "male") %>% pull("p\_autocorrelation")

Two model specifications have been built.

* mdl\_nolag is effectively the model specification used above, which has no predictor terms.
* mdl\_lag includes the previous year’s change as a predictor variable.

As mdl\_nolag is the same as mdl\_lag with one of its terms set to zero, mdl\_nolag is a restricted version of mdl\_lag, and so ANOVA can be used to test whether the move from the unrestricted model (mdl\_nolag) to the restricted (mdl\_lag) is justified in terms of improved fit to the data. In effect, the ANOVA test gives us an estimated probability that the data contains autocorrelation. These probabilities are 1.410^{-4} for females, and 0.02706 for males, and so provide strong evidence that autocorrelation exists in the data.

The specific type of autocorrelation of interest is negative autocorrelation. Negative autocorrelation would be present if the estimate column for the term last\_change\_e0 were negative. The coefficient of this term is statistically significant at p < 0.05 if the corresponding row in the column p.value is below 0.05. We can see from the table below that both of these conditions are true, though the magnitude of the negative autocorrelation is larger for females than males, and the associated p value for females much smaller than for males.

lag\_mdl\_summary

## # A tibble: 4 x 6  
## gender term estimate std.error statistic p.value  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 female (Intercept) 0.260 0.0407 6.40 0.0000000294  
## 2 female last\_change\_e0 -0.469 0.115 -4.08 0.000140   
## 3 male (Intercept) 0.251 0.0409 6.14 0.0000000806  
## 4 male last\_change\_e0 -0.283 0.125 -2.27 0.0271

One conclusion we might draw from this evidence that the data contains negative autocorrelation is that the most troubling aspect of changes in life expectancy in Scotland since 2014 is not necessarily the fall in life expectancy from 2014-15, but the lack of any notable recovery in life expectancy increases from 2015-16. Negative autocorrelation would imply that a year showing a substantial fall in life expectancy is more likely to be followed by a substantial improvement in life expectancy, above the long-term average improvement, the following year. Instead, life expectancy improvement in 2015-16 was close to zero.