Lecture 9: Normal Approximation

Chapter 3.2

Homeworks

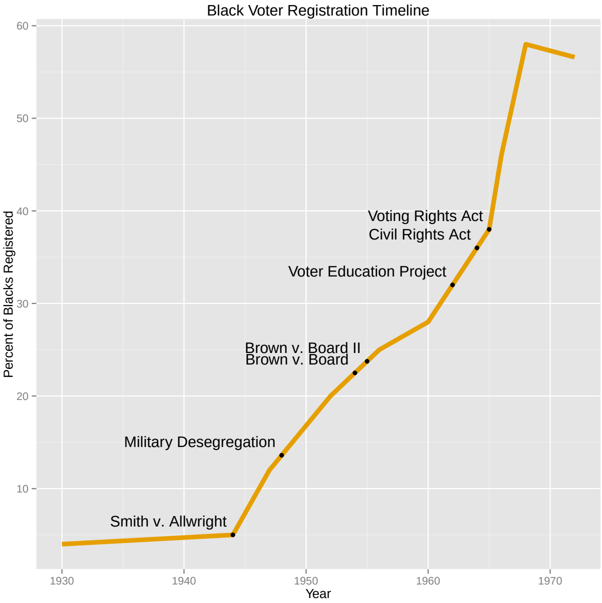
- Please mark you name and lab section (not class section) on all your written AND lab homeworks.
- For this week's lab, in order to get reproducible random values consider the `set.seed()` command. `runif()` generates random numbers between 0 and 1.

```
> set.seed(11)
> runif(3)
[1] 0.2772497942 0.0005183129 0.5106083730
> set.seed(11)
> runif(3)
[1] 0.2772497942 0.0005183129 0.5106083730
> runif(3)
[1] 0.01404791 0.06468978 0.95484923
```

African American Voter Registration

A friend of mine (Loren Collingwood) is a Political Science prof at UC Riverside. He made the following plot for his girlfriend as a Valentine's present and because it was Black history month...

African American Voter Registration



Normal Probability Tables

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
	Second decimal place of Z									
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

From textbook:

- Red case: We are given a z-score of 0.43. A lookup tells us the area to the left of z=0.43 is 0.664, i.e. the 66th %'ile
- Blue case: We want the z-score that is the 80th %'ile. We do a reverse lookup: the closest value on the table is 0.7995, which corresponds to a z-score of 0.84.

Normal Probability Tables

In practice after you've taken this class, use R to find such values:

- Red case: We are given a z-score of 0.43 and want the %'ile:

  ```
  > pnorm(0.43, mean=0, sd=1)
  [1] 0.6664022
  ```

- Blue case: We are given a %'ile of 80%=0.80 and want the z-score:

  ```
  > qnorm(0.80, mean=0, sd=1)
  [1] 0.8416212
  ```

Goals for Today

- Discuss how to find %'iles for negative values of z
- Examples
- Evaluating how "normal" certain data are.

Solving Normal Questions

Whenever solving questions of this sort ALWAYS draw a rough picture first and keep in mind:

1. The normal distribution/curve is symmetric
2. The total area under the curve is 1

Ex: From the table, a z-score of 1 corresponds to a %'ile/area of 0.84. What about a z-score of -1?

Normal Probability Tables

Alternatively, whereas

- the table on page 409 gives areas to the left of positive values of z.
- the table on page 408 gives ares to the left of negative values of z.

Exercise 3.12 on page 151: Speeding on I-5

The distribution of passenger vehicle speeds traveling on Interstate 5 Freeway (I-5) in California is nearly normal with a mean of 72.6 mph and a standard deviation of 4.78 mph.

a) What percent of passenger vehicles travel slower than 80 mph?
b) What percent of passenger vehicles travel between 60 and 80 mph?
c) How fast to do the fastest 5% of passenger vehicles travel?
d) The speed limit on this stretch of the I-5 is 70 mph. Approximate what percentage of the passenger vehicles travel above the speed limit on this stretch of the I-5.

Exercise 3.12 on page 151: Speeding on the I-5

a) What percent of passenger vehicles travel slower than 80 mph?

Exercise 3.12 on page 151: Speeding on the I-5

b) What percent of passenger vehicles travel between 60 and 80 mph?

Exercise 3.12 on page 151: Speeding on the I-5

c) How fast to do the fastest 5% of passenger vehicles travel?

Exercise 3.12 on page 151: Speeding on the I-5

d) The speed limit on this stretch of the I-5 is 70 mph. Approximate what percentage of the passenger vehicles travel above the speed limit on this stretch of the I-5.

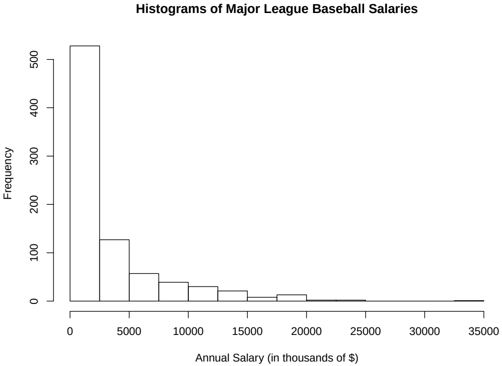
Switching Gears: Normal Approximation

Although we stated that many processes in the physical world look bell-shaped, i.e. roughly normal, we must keep in mind that this is an approximation.

Question: How do we verify normality?

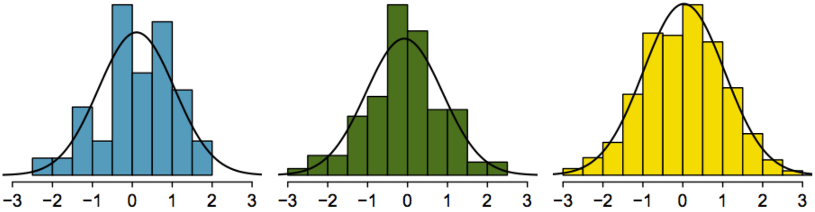
Normal Approximation

Consider the MLB salary data histogram:



Normal Approximation

What about these ones? How well do the histograms fit to the normal curve?



Normal Probability Plots

Normal probability plots (AKA quantile-quantile plots AKA QQ-plots) are a method for visually displaying how well data fit a normal curve.

The k^{th} $q-quantile$ is the value such that proportion $\frac{k}{q}$ of the observations fall below it. So

- The 4-quantiles are the quartiles.
  Ex: the $k = 2^{\text{nd}}$ 4-quantile is just the median.
- The 100-quantiles are the percentiles.
  Ex: the $k = 76^{\text{th}}$ 100-quantile is just the 76th percentile

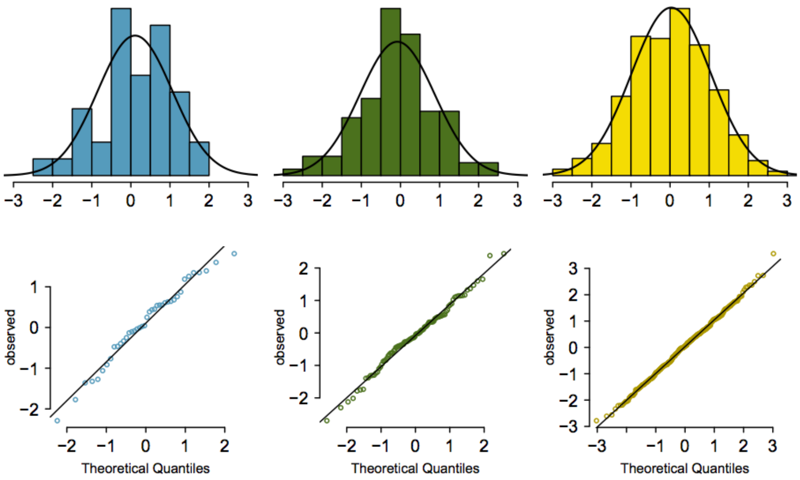
Normal Probability Plots

A normal probability plot compares:

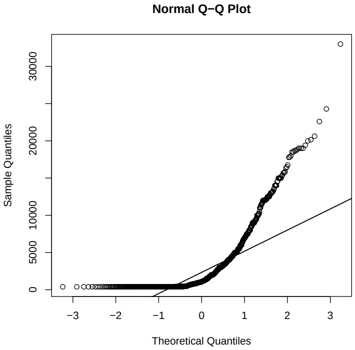
- The observed quantiles of a data set (on the y-axis)
- The theoretical quantiles that are exactly normal (on the x-axis)

The more “normal” the data is, the better the fit.

Normal Probability Plots



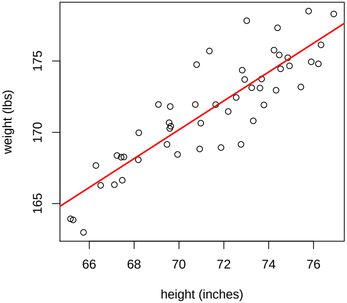
MLB Salary Normal Probability Plot




```
library(openintro)
data(MLB)
qqnorm(MLB$salary)
qqline(MLB$salary)
```

Linear Regression

Say we have a scatterplot/bivariate plot of height vs weight. We'll in Chapter 7 that linear regression involves finding the best fitting line between the two variables:



For inference from linear regression to be valid, there is a normality assumption, which we will verify with normal probability plots.

Next Time

- Introduce some of the more useful other distributions: Bernoulli, Geometric, and Binomial