13 ChPts K1K2

K1.14

Bernoulli Trials

For a 100 question multiple choice exam with 4 options per question, a student randomly guesses. Each guess is a Bernoulli trial with success probability $p = 1/4$. Thus, the number of correct answers S_{100} has a binomial distribution with

$$\text{mean } np = 100 \cdot \frac{1}{4} = 25 \text{ and standard deviation } \sqrt{np(1-p)} = \sqrt{100 \cdot \frac{1}{4} \cdot \frac{3}{4}} = \frac{5}{2}\sqrt{3} \approx \frac{13}{3}$$

A student has 7 correct answers. This has a z-score

$$z \approx \frac{7-25}{13/3} = \frac{54}{13} < -4$$

Did this student *try* to give incorrect answers?

Exercise. Find the exact z-score and use `pnorm` to estimate the probability of 7 or fewer correct answers. Compare this value to the value obtained using `pbinom`.

```
(z<-(7-25)/sqrt(75/4))
```

```
## [1] -4.156922
```

```
pnorm(z)
```

```
## [1] 1.612821e-05
```

```
pbinom(7, size=100, prob=1/4)
```

```
## [1] 2.988924e-06
```

K1.16

Example

You want to store 400 pictures on your smart phone. Pictures have a mean size of 450 kilobytes (KB) and a standard deviation of 50 KB. Assume that the size of the pictures are independent. S_{400}, the total storage space needed for the 400 pictures, has

$$\text{mean } 400 \times 450 = 180,000 \text{ KB and standard deviation } 50\sqrt{400} = 1000 \text{ KB.}$$

To estimate the space required to be 99% certain that the pictures will have storage space on the phone, note that

```
> qnorm(0.99,400*450,50*sqrt(400))
[1] 182326.3
```

So we need about 182.3 megabytes (MB).

Exercise. Give the storage space to be 95% certain to have the space for 300 pictures.

```
qnorm(0.95, 300*450, 50*sqrt(300))
```

```
## [1] 136424.5
```

Therefore, we need about 136.4 megabytes.

Intro to Propatation of Error/Delta Method

We're going to motivate propagation of error and the delta method by comparing the distribution of outputs of $g(x) = x^2$ to those of the linearization of $g(x)$ about x = 1, $\hat{g}(x) = 1^2 + 2(1)(x-1) = 2x - 1$.

1. Simulate 1000 uniformly distributed variables between 0.9 and 1.1 and apply $g(x)$ and $\hat{g}(x)$. How do their means compare? How do their standard deviations compare?

In this case, the means and the standard deviations are quite similar. We can see that the mean is the same up to 2 decimal points. In the case of the standard deviation, the value is a bit high as compared to the standard deviations that we get using rnorm. These too are similar upto atleast 3 decimal points.

```
x <- runif(1000,0.9,1.1)
g_nonlinear <- x^2  # type in g(x) here
g_linear <-  2*x-1 # type in hat g(x) here
mean(g_nonlinear); mean(g_linear)
```

```
## [1] 1.003311
```

```
## [1] 1.000006
```

```
sd(g_nonlinear); sd(g_linear)
```

```
## [1] 0.1150399
```

```
## [1] 0.1150281
```

2. Simulate 1000 normally distributed variables with a mean of 1 and a SD of $0.2/\sqrt{25 \cdot 12}$ and apply $g(x)$ and $\hat{g}(x)$. How do their means compare? How do their standard deviations compare?

Over here, we see that the means are much closer to each other compared to the previous example. They reach up to 4 digits after the decimal point of similarity. Here, we can also observe that the standard deviation is a magnitude of 10 lower when compared to the standard deviation in the previous example. Also, their values are similar upto 5 digits after the decimal point.

```
x <- rnorm(1000,1,0.2/sqrt(40*12))
g_nonlinear <- x^2  # type in g(x) here
g_linear <-  2*x-1 # type in hat g(x) here
mean(g_nonlinear); mean(g_linear)
```

```
## [1] 0.9995036
```

```
## [1] 0.9994214
```

```
sd(g_nonlinear); sd(g_linear)
```

```
## [1] 0.01812748
```

```
## [1] 0.01813587
```