

# Inference

*Kyle Payne*

*September 16, 2015*

## Stock portfolio diversification?

An investor with a stock portfolio worth several hundred thousand dollars sued his broker and brokerage firm because lack of diversification in his portfolio led to poor performance. The code below gives the rates of return for the 39 months that the account was managed by the broker. Create a histogram and a boxplot for these data.

a. Is the sample of stock portfolio returns approximately normally distributed?

b. Based on the boxplot, are there any apparent outliers?

```
stock_returns <- c(-8.36, 1.63, 6.82,
                  -2.35, -9.16, -1.25,
                  -0.65, 4.34, -2.14, -1.01,
                  -2.27, -2.93, -3.58, 6.13,
                  -1.22, -2.64, -1.03, 1.28,
                  -7.24, 12.22, -7.21, -1.41,
                  12.03, -2.70, 7.00, -5.11,
                  -10.27, -2.56, -2.93, -9.14,
                  -15.25, -8.66, -0.80, -1.44, 7.34,
                  5.04, 4.33, 2.35, -7.24)
```

An arbitration panel compared these returns with the average of the Standard and Poor's 500 stock index for the same period. Consider the 39 monthly returns as a random sample from the population of monthly returns the brokerage would generate if it managed the account forever.

c. Create a .95 confidence interval for the mean return of the S&P 500  $\mu = .95$

i.e. If the returns are normally distributed, use a z-interval test, assuming (incorrectly) that  $\sigma = s$ .

That is,

$$\bar{Y} \pm \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}$$

```
## calculate the mean and standard error
n <- length(stock_returns)
mean_return <- mean(stock_returns)
se <- sd(stock_returns)/sqrt(n)
pivot <- qnorm(.975)

CI <- mean_return + c(-1,1)*se*pivot
CI
```

```
## [1] -3.1875160 0.6213622
```

d. Is  $\mu = 0.95$  within the 0.95 confidence interval?

## Hypothesis Testing

Believe it or not, you just performed what is equivalent to a hypothesis test. As the Standard and Poor's percent return was not included in the confidence interval, we have .95 confidence that the interval does not include  $\mu = 0.95$ . That is, the true mean is not 0.95 with .95 confidence.

$$H_0 : \mu = 0.95$$

$$H_A : \mu \neq 0.95$$

e. Perform a Z-test for the preceding null hypothesis.

$$Z = \frac{\bar{Y} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

There are two equivalent approaches to coming to a conclusion for a hypothesis test.

- Compare a test statistic to a critical value
- Compare a p-value to an  $\alpha$

Critical value is the value(s) by which we state that if the test statistic is larger (or smaller) than, we will reject the test. This value is typically defined as a quantile on the null distribution.

*Remember: The null distribution is the distribution defined by the null hypothesis* For a two-sided test at the  $\alpha = 0.05$  level, the critical values are  $Z_{\alpha/2}$ , and  $Z_{1-\alpha/2}$  or the  $\alpha/2$  and  $1 - \alpha/2$  quantiles of the null distribution. Because the normal distribution is symmetric about the mean, the critical values are equal in magnitude. Thus, in our case, this would be the value

```
pivot <- qnorm(.975)
```

Thus, with  $Z_{.975} = 1.96$  and likewise we have  $Z_{0.025} = -1.96$ , we have

$$Z = \frac{(-1.28 - 0.95)}{0.972}$$

```
Z <- (mean_return-0.95)/(se)
if(Z < -1.96 || Z > 1.96){
  print('Reject the null hypothesis')
}else{
  print('Fail to reject')
}
```

```
## [1] "Reject the null hypothesis"
```

Likewise, the p-value of the test statistic is the probability of all test statistics that are 'more extreme' on the null distribution. If we observe that the probability that the test statistic or more extreme test statistics is smaller than the probability than  $\alpha$ , then we reject. As  $Z_{.975}$  is picked such that there is a  $\alpha$  probability of observing values more extreme, these two methods are equivalent.

## Statistical Errors

A type I error is a “false positive”, or when one rejects the null hypothesis when the null is actually true.

A type II error is a “false negative”, or when one fails to reject a false null hypothesis.

**For the previous test, calculate the probability of a type II error.**

$$\beta = P(Z < \frac{\bar{Y}_{crit} - \mu_a}{\sigma_{\bar{Y}}})$$

For  $Z = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} \leq -1.96$ , then for the null,

$\bar{Y} \leq -1.96\sigma/\sqrt{n} + \mu_0 = \bar{Y} \leq \bar{Y}_{crit}$  would result in a rejection of the null hypothesis.

```
crit <- -1.96*se + 0.95
```

If the true mean was  $\mu = -0.5$ , then

$$P(\bar{Y} \leq \bar{Y}_{crit} | \mu = -0.5) = P\left(\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \leq \frac{\bar{Y}_{crit} - \mu}{\sigma/\sqrt{n}}\right) = \beta$$