Week 3 Cheat Sheet

Statistics and Data Analysis with Excel, Part 2

Charlie Nuttelman

Here, I provide the mathematical equations and some of the important Excel functions required to perform various calculations in Week 3 of the course. The headings represent the screencasts in which you will find those calculations and concepts. Not all screencasts are referenced below – just the ones that have complex mathematical formulas or Excel formulas that are tricky to use.

My equations and analysis follow that of Montgomery and Runger in: *Applied Statistics and Probability for Engineers, 7th edition*, Douglas C. Montgomery and George C. Runger, Wiley (2018). This is an excellent text for applied probability and statistics and highly recommended if you need a supplementary text for the course.

Introduction to Hypothesis Testing (Parts 1 and 2)

In hypothesis testing, we set up a null hypothesis (H_0) and an alternate hypothesis (H_1). The alternate hypothesis is what you aim to prove (accept) or disprove (reject). The null hypothesis can never be accepted – only rejected.

Hypothesis tests can be lower-tailed, upper-tailed, and two-tailed. If you've already got the sample data, there's no sense in performing a two-tailed test since your sample data will point you in the direction of an upper- or lower-tailed test.

A lower-tailed test is one in which the alternate hypothesis involves a less than sign. For example,

$$H_0: \mu = 10$$

$$H_1: \mu < 10$$

An upper-tailed test is one in which the alternate hypothesis involves a greater than sign. For example,

$$H_0: \mu = 10$$

$$H_1: \mu > 10$$

And a two-tailed test (not as common, especially if you've already obtained the sample data) is one in which the alternate hypothesis involves a not equal sign. For example,

$$H_0: \mu = 10$$

$$H_1: \mu \neq 10$$

There are several ways to perform a hypothesis test: 1) using confidence intervals (not that common), 2) using test statistics (common), and 3) using P-values (becoming more common, and I would recommend this as the best way to make statistical conclusions).

Hypothesis Tests on the Mean Using Confidence Intervals (Variance Known)

I will work through an example of a lower-tailed test here. To perform a lower-tailed hypothesis test, we first set up the hypothesis to test:

$$H_0: \mu = \mu_0$$

$$H_1: \mu < \mu_0$$

Here, μ_0 is an actual value (number).

Then, using our sample statistics, we can put together a $(1-\alpha)$% confidence bound on the mean:

$$P\left[\mu > \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

If μ_0 lies in that confidence bound (i.e., if μ_0 is greater than $\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}$), then we can reject the alternate hypothesis. However, if μ_0 is not in that confidence bound (i.e., if μ_0 is less than $\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}$), then we accept the alternate hypothesis and reject the null.

A similar analysis is done for an upper-tailed test (μ_0 must be greater than $\bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$ for us to accept the alternate hypothesis) and for a two-tailed test (μ_0 must be less than $\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ or greater than $\bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ for us to accept the alternate hypothesis).

The Test Statistic Approach to Hypothesis Testing

The second, and more common, approach to hypothesis testing is the test statistic approach. We are essentially converting our sample average to a z-value (on the standard normal distribution):

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

z_0 is known as a "test statistic". We can create test statistics based on the T distribution (t_0), the chi-squared distribution (χ_0^2), and the F distribution (f_0), as we'll see later on.

Once we have our test statistic, we compare it to the critical value on the distribution of interest. For the standard normal distribution (a test on the mean when variance is known), we distribute our Type I risk (α) into the lower tail (for a lower-tailed test), the upper tail (for an upper-tailed test), or both tails (for a two-tailed test) of the distribution.

For a lower-tailed test, this critical value is $z_{1-\alpha} = -z_\alpha$. In order for us to accept the alternate hypothesis, z_0 must be less than $-z_\alpha$.

For an upper-tailed test, this critical value is z_α. In order for us to accept the alternate hypothesis, z_0 must be greater than z_α.

And for a two-tailed test, we have two critical values: -$z_{\alpha/2}$ and $z_{\alpha/2}$. In order for us to accept the alternate hypothesis, z_0 must be less than $-z_{\alpha/2}$ or greater than $z_{\alpha/2}$.

All of the above applies essentially the same to hypothesis tests on the mean when variance is unknown (based on the T distribution), but we use t_0 as the test statistic:

$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

We then compare t_0 to $-t_{\alpha,n-1}$ for a lower-tailed test, $t_{\alpha,n-1}$ for an upper-tailed test, and $-t_{\alpha/2,n-1}$ or $t_{\alpha/2,n-1}$ for a two-tailed test. Note that the T-distribution is dependent on degrees of freedom, which are $n-1$.

Using P-Values for Hypothesis Testing on the Mean (Parts 1 and 2)

The final way to perform hypothesis tests is to use P-values. The method is based on test statistics (calculated above) but we go one step further and we calculate how probable it is to get a z-value (or t-value, chi-squared value, F-value for other distributions) based on the sample statistics that we obtained.

For a hypothesis test on the mean with variance known, we calculate z_0, our test statistic, and then we calculate a P-value as the area in the tail to the right (for upper-tailed test), left (for lower-tailed test), or both tails (for two-tailed test).

For an upper-tailed test based on the standard normal distribution: P-value = $P[z > z_0]$

For a lower-tailed test based on the standard normal distribution: P-value = = $P[z < z_0]$

For a two-tailed test based on the standard normal distribution: P-value = $2 \cdot P[z < z_0]$ (if z is negative) or P-value = $2 \cdot P[z > z_0]$ (if z is positive).

Excel is a great tool to calculate P-values based on the standard normal distribution (NORM.S.DIST), the T distribution (T.DIST and T.DIST.RT), the chi-squared distribution (CHISQ.DIST and CHISQ.DIST.RT), and the F distribution (F.DIST and F.DIST.RT). See the various cheat sheets that I have posted related to Excel functions for these distributions.

P-values can be similarly obtained based on the T distribution, chi-squared distribution, and F distribution) and they provide us with a level of confidence in making a decision. For example, if α = 0.05 and we get a P-value of 0.049, we would accept the alternate hypothesis. But if we got a P-value of 0.007 we would also accept the alternate hypothesis but with such a small P-value, we would be a lot more confident in our conclusion in this latter case than with getting a P-value of 0.049.

Type I and Type II Errors

A Type I error is the probability that you accept the alternate hypothesis when the alternate hypothesis is false. The probability of making a Type I error is α, and a typical value for α is 0.05. α is typically chosen by the one performing the hypothesis test. If α is too small, then it is more difficult to accept the alternate hypothesis; if α is too big, then there is a higher probability of making a Type I error. Keep in mind that, even if α = 0.05, one will make a Type I error 1 in 20 times if the null hypothesis is false.

A Type II error is failing to accept the alternate hypothesis when it is true. The probability of making a Type II error is known as β, and it depends upon the shift in mean (or difference in mean) between the null and alternate hypotheses.

To calculate β, we must first calculate the critical sample average ($\bar{x}_c$). For an upper-tailed test, we would accept the alternate hypothesis if our sample average were greater than $\bar{x}_c$ and we would reject the alternate hypothesis if our sample average were less than $\bar{x}_c$. For a lower-tailed test, we would accept the alternate hypothesis if our sample average were less than $\bar{x}_c$ and we would reject the alternate hypothesis if our sample average were greater than $\bar{x}_c$.

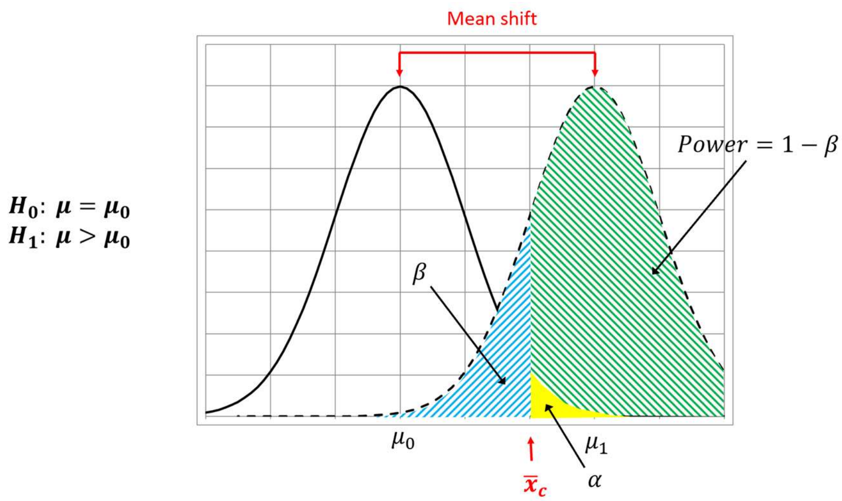
Once we have identified $\bar{x}_c$, which depends on α, we can calculate β by calculating the area underneath the alternate hypothesis sampling distribution to the left of $\bar{x}_c$ for an upper-tailed test and to the right of $\bar{x}_c$ for a lower-tailed test:

For an upper-tailed test: $\beta = P[\bar{X} < \bar{x}_c \mid \mu_1]$

For a lower-tailed test: $\beta = P[\bar{X} > \bar{x}_c \mid \mu_1]$

For both cases, "$\mid \mu_1$" indicates that we are operating under the alternate sampling distribution (we must know or assume a value for μ_1 in order to calculate a numerical value for β).

For an upper-tailed test (graphically):



Power of the Test

Power is simply $1 - \beta$. See diagram above. Power is the probability that we correctly identify that the alternate hypothesis is true (the mean has shifted) when it has, in fact, shifted. You want power to be high. Ways to increase power include increasing the acceptable shift in mean, increasing sample size (n), and increasing α. Increasing α is not a good idea since it increases the probability that you'll make a Type I error. The most straightforward way to increase power is to simply increase the sample size.

Choice of Sample Size and OC Curves (Variance Known)

Oftentimes, the question arises as to what size sample should be taken during routing quality control/sampling to guarantee that the power of the test is above some value. This sample size depends on the desired detectable shift in mean, δ, as well as α and β according to the following equation, where we have known variance (σ):

$$n = \frac{\left(z_\alpha + z\ \right)^2 \sigma^2}{\delta^2}$$

Before Excel and other computing tools allowed one to easily use the formula above, operating characteristic (OC) curves/charts could be used to estimate sample size based on the above parameters.

Choice of Sample Size, Variance Unknown

For choice of sample size when the variance is unknown, the analysis is a bit more complicated. There is no explicit equation for calculating sample size as in the case above when variance is known. The analysis is based on the noncentral T distribution, which is beyond the scope of basic statistics (i.e., beyond the scope of this course). Regardless, we can use operating characteristic (OC) curves that are published in various resources to estimate the sample size required to obtain a certain level of power based on a specified level of α.

Hypothesis Tests on the Variance

We can perform hypothesis tests on the variance, and this analysis is based on the chi-squared distribution. We can set up a lower-tailed hypothesis test, an upper-tailed hypothesis test, or a two-tailed hypothesis test on the variance (σ_0^2 just refers to an actual number, like 10).

Lower-tailed test on the variance:

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 < \sigma_0^2$$

Upper-tailed test on the variance:

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 > \sigma_0^2$$

Two-tailed test on the variance:

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 \neq \sigma_0^2$$

To utilize the test statistic approach or the P-value approach to hypothesis testing on the variance, we set up a test statistic, $\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$

In order for us to accept the alternate hypothesis for a lower-tailed test on the variance, $\chi_0^2 < \chi_{1-\alpha,n-1}^2$, in order for us to accept the alternate hypothesis for an upper-tailed test on the variance, $\chi_0^2 > \chi_{\alpha,n-1}^2$, and in order for us to accept the alternate hypothesis for a two-tailed test on the variance, $\chi_0^2 < \chi_{1-\alpha/2,n-1}^2$ or $\chi_0^2 > \chi_{\alpha/2,n-1}^2$.

A P-value for all three hypothesis tests can be calculated as the area in the tail (for lower- and upper-tailed tests) or tails (for two-tailed test) to the right/left of the test statistic:

For a lower-tailed test: P-value = $P[\chi_0^2 < \chi_{1-\alpha,n-1}^2]$

For an upper-tailed test: P-value = $P[\chi_0^2 > \chi_{\alpha,n-1}^2]$

For a two-tailed test: P-value = $2 \cdot P[\chi_0^2 < \chi_{1-\alpha/2,n-1}^2]$ or $2 \cdot P[\chi_0^2 > \chi_{\alpha/2,n-1}^2]$

Hypothesis Tests on a Binomial Proportion

We can perform a hypothesis test on a binomial proportion (the proportion of items that satisfy a certain constraint). As with other tests, these can take on lower-tailed, upper-tailed, or two-tailed versions. For example, a lower-tailed test on a binomial proportion:

$$H_0: p = p_0$$

$$H_1: p < p_0$$

Here, p_0 is just some proportion (number), like 0.75.

To perform the test, we collect sample data. X is the number of items out of the total sample size, n, that have the quality of interest. We can then calculate the proportion, p, as:

$$p = \frac{X}{n}$$

There are two ways to perform the hypothesis test on a binomial proportion. The first one is an approximation based on the standard normal distribution and was useful before the advent of computing tools like Excel. In order to use the normal approximation to the binomial distribution, the following two constraints must be satisfied:

$np \geq 5$ and $n(1-p) \geq 5$

If either of these is not satisfied, then we cannot use the normal approximation to the binomial distribution. If both of these are satisfied, then we can create the following test statistic:

$$z_0 = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$$

We then base the analysis exactly as we would with the standard normal distribution. For example, for a lower-tailed test, we would accept H_1 if $z_0 < -z_\alpha$. See the beginning of this document for the ways that we can perform hypothesis tests based on the standard normal distribution.

If either of the constraints $np \geq 5$ and $n(1-p) \geq 5$ are not satisfied, then we must use the direct binomial approach, which uses the binomial distribution equation (and can easily be done in Excel). We base this approach on the P-value approach to hypothesis testing. To calculate a P-value for a lower-tailed test based on the binomial distribution, we can use the BINOM.DIST function in Excel:

(P-value) =BINOM.DIST(X,n,p0,TRUE)

Here, X is X from above, n is the sample size (n), and p0 is the hypothesized proportion in the null/alternative hypotheses (p_0).

For an upper-tailed test, we can calculate the P-value as follows in Excel:

(P-value) =1-BINOM.DIST(X-1,n,p0,TRUE)

In either case, we then compare our P-value to α and make our conclusion.