Statistics 101B Kitu Komya
Homework 1 UID: 404-491-375, Dis 3B

1.3 Step 1: Recognition of the statement of the problem

We want to examine the effects of different conditions of sunlight, water, fertilizer, and soil conditions on the growth of garden flowers.

Step 2: Selection of the response variable

The response variable is the height of the garden flower, measured by a ruler in centimeters (continuous responses).

Step 3: Choice of factors, levels, and range

The factors studied will be sunlight, water, fertilizer, and soil conditions. The levels of each factor will be the following:

sunlight: completely dark, partially shaded, fully sunny
water: 0 mL, 100 mL, 300 Ml
fertilizer: compost, organomineral, inorganic
soil conditions: clay soil, sandy soil, organo-dirt soil

1.4 Step 1: Recognition of the statement of the problem

I would like to examine the effects of different conditions of nail polish brand, fingernail, and color on the durability of nail polish.

Step 2: Selection of the response variable

The response variable is the first day the nail starts to chip (discrete responses).

Step 3: Choice of factors, levels, and range

The factors studied will be nail polish brand, fingernail, and color. The levels of each factor will be the following:

nail polish brand: Sally Hansen, Sinful Colors, Revlon
fingernail: left foot big toe, right hand pink, left hand index finger
color: red, blue, black

1.8 Replication refers to the number of observations to which one specific treatment is applied. It allows for a more reliable estimate of the effect of each treatment which thus allows determination of whether the observed differences are statistically significant. This thus allows us to estimate the experimental error as well. Going back to the nail polish example, suppose I want to examine the effect of just nail polish brand on durability of nail polish. I could use my 10 fingers as 10 replicates, since they all will undergo the same treatment. Then, I could repeat my experiment 3 times, (re-painting my nails 3 times), hence having 3 repeated measurements.

1.9 Randomization is important because it removes bias by preventing subjective assignment. This can also help minimize the effects of irrelevant or unknown factors present. Only then can a valid and reliable conclusion be made.

2.5 (a) $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{31.2000 - 30}{0.3000} = 4$

Using the standard normal table, the p-value is 0.00006. We reject the null hypothesis at a confidence level of 95% since the p-value is smaller than 0.05 and thus conclude that the mean is not equal to 30.

(b) This is a two-sided test since in the computer output we are testing whether the mean is equal to 30 or not equal to 30 instead of just one side, so that allows for both tails.

(c) The formula for finding a confidence interval is the following: $\bar{x} \pm z * \frac{s}{\sqrt{n}}$, where
$\bar{x}$ is the sample mean = 31.2000
z is the z-score = 2.58 (found for a 99% confidence level from a standard z-table)
$\frac{s}{\sqrt{n}}$ is already given in the table as SE mean to be 0.3000

Plugging in these values, we get (30.42725, 31.97275).

(d) Since that would be a one-sided test, we simply divide the two-sided p-value by 2 to get 0.00003.

2.9 (a) $t = \frac{\overline{y_1} - \overline{y_2}}{StdError} = 2.01 = 2.35/\text{StdError}$
Standard error = 2.35/2.01 = 1.169

(b) Using the standard table, we see that a p-value of 0.0298 with a t-statistic of 2.01 is only achieved when it's a one-sided test. The p-value would be 0.0596 for a two-sided test.

(c) Because the p-value is smaller than 0.05, we can reject the null hypothesis at a 95% confidence level and conclude that there is a difference in means in the two samples.

(d) The formula for a confidence interval for 2 independent sample means is:

$$(\overline{y_1} - \overline{y_2}) \pm t^* \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}$$

$$2.35 \pm 1.734 * \sqrt{\frac{1.169^2}{19} + \frac{1.169^2}{19}}$$

(0.323, 4.377)