Day2

Olivia Wu

2024-02-08

Red shows the correct answer or where I should have added more detail

Problem 0.23

Choose: There is a categorical explanatory variable (Control or Incentive) and the response variable is quantitative.
Let μ_1 represent the population mean weight loss at seven months for people without financial incentives and μ_2 represent the population mean weight loss at seven months with financial incentives.
Potential models for the two groups are:

$$Y = (\mu_1 + \epsilon) \sim N(0, \sigma_1)$$

$$Y = (\mu_2 + \epsilon) \sim N(0, \sigma_2)$$

Fit: The means for the control and incentive groups are 4.639 and 7.8 pounds, respectively. Our models may be expressed as:

$$Y = (4.639 + \epsilon) \sim N(0, 7.8)$$

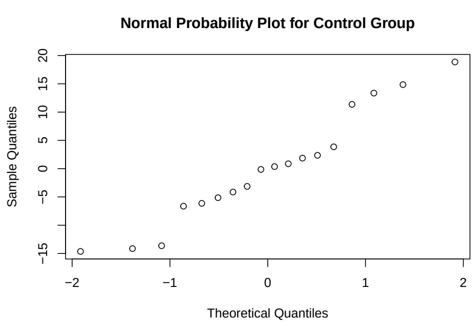
$$Y = (7.8 + \epsilon) \sim N(0, 12.06)$$

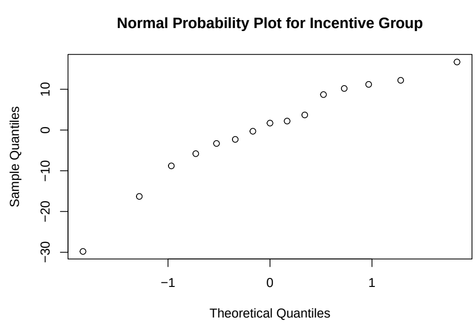
Assess: We conduct a two-sample t-test for the hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 < 0 \quad \mu_1 - \mu_2 \neq 0$$

Conditions We assume independence and random assignment is given. The following normal probability plots are roughly linear, so we assume normality.






```
##
##  Welch Two Sample t-test
##
```

```
## data:  WL7$Month7Loss[WL7$Group %in% "Control"] and WL7$Month7Loss[WL7$Group %in% "Incentive"]
## t = -0.8144, df = 26.994, p-value = 0.2113
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 3.450298
## sample estimates:
## mean of x mean of y
##  4.638889  7.800000
```

Since $p = 0.2113 > 0.05$ $p = 0.4226 > 0.05$, we fail to reject the null hypothesis.

Use: There is not enough evidence to suggest that the beneficial effects of financial incentives still apply to the weight losses at the seven-month point.

Problem 0.26

Choose: There is a categorical explanatory variable (Short or Long) and the response variable (Pace) is quantitative.
Let μ_1 represent the population mean pace for short runs and μ_2 represent the population mean pace for long runs.
Potential models for the two groups are:

$$Y = \mu_1 + \epsilon$$

$$Y = \mu_2 + \epsilon$$

Fit: The means for the short and long groups are 7.96 and 8.16 minutes per mile, respectively. Our models may be expressed as:

$$Y = 7.96 + \epsilon$$

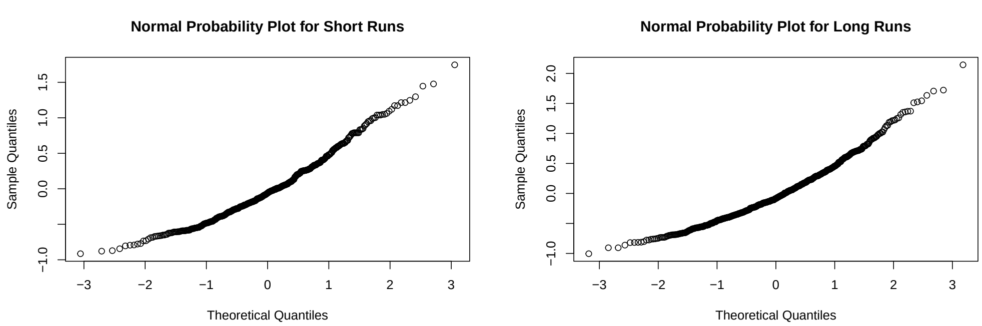
$$Y = 8.16 + \epsilon$$

Assess: We conduct a two-sample t-test for the hypotheses:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 < 0$$

Conditions We assume randomness and independence. There are more than 30 observations for both groups, so the normality conditions is satisfied. This is further supported by the roughly linear normal probability plots shown below.




```
##
##  Welch Two Sample t-test
```

```
##
## data:  short and long
## t = -6.6991, df = 962.52, p-value = 1.781e-11
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf -0.1481883
## sample estimates:
## mean of x mean of y
##  7.961188  8.157665
```

$p = 0 < 0.05$ so we reject the null hypothesis. There is a statistically significant difference in pacing for short and long runs.

Use: There is enough evidence to suggest that the runner tends to run faster on short runs than on long runs.

Problem 0.27

a) **Choose:** $Y = \mu_i + \epsilon$ where μ_1 represents the average running pace from 2002-2004, and μ_2 represents the average running pace from 2005-2006.

   Fit: $Y = 7.914 + \epsilon$ for 2002-2004
   $Y = 8.411 + \epsilon$ for 2005-2006

   Assess: We test for the hypotheses:

   $H_0 : \mu 1 - \mu 2 = 0$

   $H_a : \mu 1 - \mu 2 \neq 0$

   The conditions have been verified in Problem 0.26

```
##
##  Welch Two Sample t-test
##
## data:  young_pace and old_pace
## t = -18.715, df = 813.19, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.5488239 -0.4446252
## sample estimates:
## mean of x mean of y
##  7.914414  8.411139
```

$p \approx 0 < 0.05$ so we reject the null hypothesis. There is a statistically significant difference in pacing for runs before and after 2004.

Use: There is enough evidence to suggest that the runner had different pacings when he was young compared to when he was old.

b) **Choose:** $Y = \mu_i + \epsilon$ where μ_1 represents the average daily distance from 2002-2004, and μ_2 represents the average daily distance from 2005-2006.

   Fit: $Y = 7.492 + \epsilon$ for 2002-2004
   $Y = 6.629 + \epsilon$ for 2005-2006

   Assess: We test for the hypotheses:

$H_0: \mu1 - \mu2 = 0$

$H_a: \mu1 - \mu2 \neq 0$

The conditions have been verified in Problem 0.26

```
##
##  Welch Two Sample t-test
##
## data:  young_dist and old_dist
## t = 3.9261, df = 937.03, p-value = 9.265e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.4318012 1.2949073
## sample estimates:
## mean of x mean of y
##  7.492154  6.628800
```

$p \approx 0 < 0.05$ so we reject the null hypothesis. There is a statistically significant difference in distance ran per day before and after 2004.

Use: There is enough evidence to suggest that the runner ran different distances per day when he was young compared to when he was old.

Problem 0.30

a) effect size $= \frac{66.65-67.82}{11.31} = 0.103$, so there is a small difference in pulses between men and women

b) effect size $= \frac{|7.96-8.16|}{0.48} = 0.407$, so there is a moderate effect of changing the distance of runs on the pace.

Problem 0.31

a) effect size $= 0.228$, so there is a small effect of training before or after 2004 had on the mean miles per day for training.

b) effect size $= 1.072$, so there is a large difference in the guessing abilities between men and women.