Homework 3 Solution

Charlotte Wickham

2017-10-23

Submit your answers on canvas.

1. p-values

a) We usually think about a small p-value providing evidence against the null hypothesis. What else does the article imply a small p-value may cast doubt on?

   In addition to casting doubt on the null hypothesis, a small p-value may cast doubt on the assumptions of the procedure.

   (*Consider for example the Wilcoxon Signed Rank Test as a test of the population mean. A small p-value indicates that the data is inconsistent with the null hypothesis under the assumption the population is symmetric. If we conduct a the test and get a small p-value it's possible the population is symmetric and the the p-value suggests the null hypothesis isn't true, but it's also possible the population isn't symmetric, in which case the small p-value tells us nothing about the mean, but has arisen becasue of the assymetry in the population.*)

b) What is the primary argument for not basing scientific conclusions or policy decision solely on whether the p-value is below some threshold?

   Applying a black and white decision to a continuous measure is too big a simplification for important decisions.

c) What is *p-hacking*?

   Doing many tests and reporting only those that are significant, or using the data to suggest hypotheses to test.

d) Can a p-value measure the size of an effect? What can measure the size of an effect?

   No. Confidence intervals can measure the size of an effect. The article also suggests a few other approaches too.

e) Skim through the references in the "A brief p-Values and Statistical Significance Reference List", and shortlist three article titles that interest you. (*You may be required to read one of these in a future homework*)

   (*Chuan and Erin: could you keep track of any popular items.*)

2. Data analysis

The Behavioral Risk Factor Surveillance System (BRFSS) is a nationwide health-related survey of U.S. residents. For this question you can get a sample of responses from the 2003 survey by downloading an R data file from the class website:

```
library(tidyverse)
download.file("http://st551.cwick.co.nz/data/brfss.rds",
  "brfss.rds", mode = "wb")
```

Then load it into the variable `brfss` with:

```
brfss <- read_rds("brfss.rds")
```

The variables `weight_kg` and `wtdesire_kg` correspond to the responses to the questions:

- About how much do you weigh without shoes?
- How much would you like to weigh?

respectively converted to kilograms.

You can create a variable to represent the amount of weight a respondent would like to lose with:

```
brfss <- mutate(brfss, desired_loss = weight_kg - wtdesire_kg)
```

a) Find summary statistics (mean, standard deviation and number of observations) for `desired_loss` for both males and females in the sample.

```
sum_stat <- brfss %>%
  group_by(sex) %>%
  summarise(
    mean = mean(desired_loss, na.rm = TRUE),
    sd = sd(desired_loss, na.rm = TRUE),
    n_obs = sum(!is.na(desired_loss)))
```

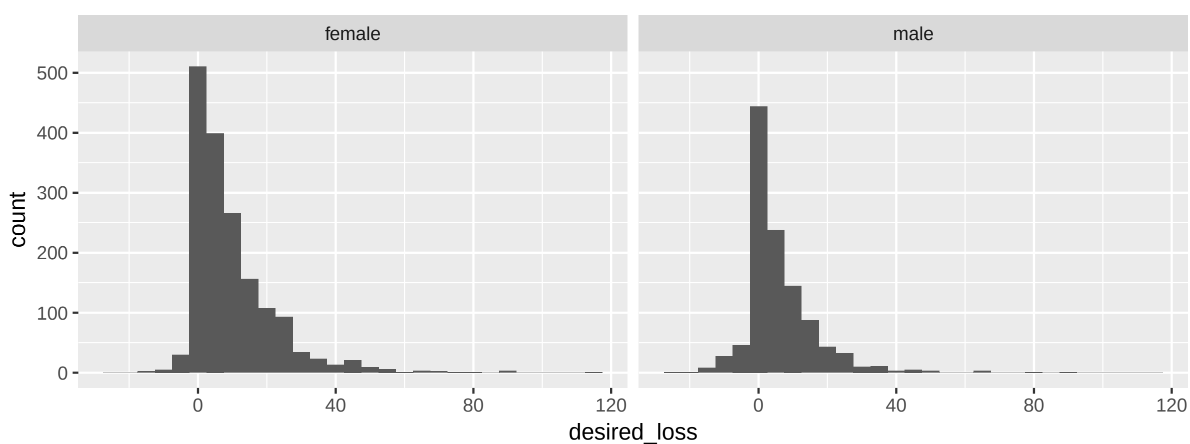
Table 1: Summary statistics for the BRFSS sample

Sex	Mean	Std. Dev	Sample size, n
female	9.62	12.1	1686
male	5.85	10.2	1109

b) Produce histograms of `desired_loss` for both males and females.

```
ggplot(brfss, aes(x = desired_loss)) +
  geom_histogram(binwidth = 5) +
  facet_wrap(~ sex)
```

```
## Warning: Removed 205 rows containing non-finite values (stat_bin).
```



Notice the asymmetry in the distributions. Most people in the sample want to lose weight (`desired_loss > 0`), and the tail extends a long way on the right.

c) Do US resident females, on average, want to lose weight (i.e. is the mean desired loss greater than zero)? **Conduct the appropriate analyses and write a statistical summary of your findings**

```
females <- filter(brfss, sex == "female")
(t_test_females <- t.test(females$desired_loss,
  mu = 0, alternative = "greater"))
```

```
##
##  One Sample t-test
##
## data:  females$desired_loss
## t = 32.559, df = 1685, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  9.134858      Inf
## sample estimates:
## mean of x
##  9.621183
```

```
(ci_females <- t.test(females$desired_loss,
  mu = 0)$conf.int)
```

```
## [1]  9.041593 10.200774
## attr(,"conf.level")
## [1] 0.95
```

Conduct a t-test (since population variance is unknown) of the hypothesis $H_0 : \mu_{\text{female}} = 0$, versus one sided alternative $H_A : \mu_{\text{female}} > 0$.

(Careful with n, it needs to be the number of **non-missing** observations).

$$t(\mu_0 = 0) = \frac{\overline{Y} - \mu_0}{\sqrt{s^2/n}} = \frac{9.6211834 - 0}{\sqrt{147.2244724/1686}} = 32.56$$

Statistical Summary

There is convincing evidence that the mean desired weight loss for US resident females is greater than zero (one-sample t-test, p $<$ 0.001).

We estimate the mean desired weight loss for US resident females is 9.6 kg.

With 95% confidence we estimate the mean desired weight loss for US resident females is between 9 and 10.2 kgs.

d) Do US resident males, on average, want to lose weight (i.e. is the mean desired loss greater than zero)? **Conduct the appropriate analyses and write a statistical summary of your findings**

```
males <- filter(brfss, sex == "male")
(t_test_males <- t.test(males$desired_loss,
  mu = 0, alternative = "greater"))
```

```
##
##  One Sample t-test
##
## data:  males$desired_loss
## t = 19.156, df = 1108, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  5.34716     Inf
## sample estimates:
## mean of x
```

```
##    5.84989
```

```
(ci_males <- t.test(males$desired_loss,
  mu = 0)$conf.int)
```

```
## [1] 5.250697 6.449084
## attr(,"conf.level")
## [1] 0.95
```

Conduct a t-test (since population variance is unknown) of the hypothesis $H_0 : \mu_{\text{male}} = 0$, versus one sided alternative $H_A : \mu_{\text{male}} > 0$.

(Careful with n, it needs to be the number of **non-missing** observations).

$$t(\mu_0 = 0) = \frac{\overline{Y} - \mu_0}{\sqrt{s^2/n}} = \frac{5.8498903 - 0}{\sqrt{103.4237786/1109}} = 19.16$$

Statistical Summary

There is convincing evidence that the mean desired weight loss for US resident males is greater than zero (one-sample t-test, p $<$ 0.001).

We estimate the mean desired weight loss for US resident males is 5.8 kg.

With 95% confidence we estimate the mean desired weight loss for US resident males is between 5.3 and 6.4 kgs.

3. Performance of t-test

Explore the Type I error rate of the t-test for a two-sided level $\alpha = 0.05$ test, for samples of size $n = 5, 10, 25, 50$, for **one** of the following population distributions:

- Uniform(0, 1)
- Chi-squared(1)
- Beta(.5, .5)
- Exponential(1)

Use at least 10,000 simulations for each scenario.

a) Provide a table of the estimated Type I error rate by sample size.

	5	10	25	50
Beta(0.5, 0.5)	0.0748	0.0549	0.0489	0.0473
Chi-squared(1)	0.1824	0.1341	0.0947	0.0829
Exponential(1)	0.1184	0.1073	0.0722	0.0626
Uniform(0, 1)	0.0602	0.0535	0.0519	0.0517

Students only need to have simulated from one distribution.

b) Write a short (3-5 sentence) summary of how the t-test performs: is it close enough to exact that you would be comfortable using it even when the underlying distribution is as far from normal as these distributions?

Any valid simulations get credit. Hopefully everyone sees that the type I error rate is pretty good regardless of population. Even Chi-Square and Exponential aren't too bad by $n = 50$.