STAT 390 A
Statistical Methods in Engineering and Science

Week 10 Lectures – Part 1 – Spring 2023

Testing Proportions & Comparison of Treatments

Alexander Giessing

Department of Statistics

University of Washington

May 31, 2023

Outline

Tests for the Population Proportion

- So far, we have only discussed how to construct tests for the population mean.
- In many cases the feature of interest is not the population mean:
  - population variance,
  - the rate parameter λ of the exponential distribution,
  - success probabilities of a multinomial distribution (population genetics),
  - ...

  If you take more advanced courses in statistics, you will learn how to construct tests for these quantities (i.e. *likelihood ratio tests*).
- The population proportion is a feature for which we can easily construct tests based on the machinery that we have already developed.

  Recall: For binary data (taking values 0 or 1), the sample proportion is simply the sample average, i.e. $\hat{p} = \bar{X}_n$.

Large Sample Test for the Population Proportion

Large Sample Test for the Population Proportion

Let $x_1, \ldots, x_n$ be a realization of a random sample of $Ber(p)$ random variables with unknown success probability p. Let $\alpha \in (0,1)$ and z_α be the critical value of the standard normal distribution and define

$$t = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}, \qquad \text{where} \qquad \hat{p} = \bar{x}_n.$$

(a) One-sided upper tail test at level α:

Reject H_0 in favor of H_1 whenever $t > z_\alpha$.

(b) One-sided lower tail test at level α:

Reject H_0 in favor of H_1 whenever $t < -z_\alpha$.

(c) Two-sided (tail) test at level α:

Reject H_0 in favor of H_1 whenever $|t| > z_{\alpha/2}$.

- Since the data is binary H_0 does not only determine the mean but also the variance. Therefore, under H_0 the variance is known!

Example with a large sample: Testing a new drug

A standard medication has cure rate 60%. A new drug is introduced and tested on 100 patients. Among those, 70 get cured. Is the new drug substantially better than the standard one? Formulate the hypotheses and answer the question at $\alpha = 5\%$.

- Let p be the proportion of people the new drug cures. Then, null and alternative hypothesis are

$$H_0 : p \leq 0.6 \qquad vs. \qquad H_1 : p > 0.6.$$

- The test statistic is

$$t = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{0.7 - 0.6}{\sqrt{0.4 \times 0.6/100}} = 2.041.$$

- The 5% critical value of the standard normal distr. is $z_{0.05} = 1.645$. Thus, we reject H_0 at level 5%.

- Alternatively, the p-value is $1 - \Phi(2.041) = 2.06\%$. Hence, we reject the null hypothesis at a 5% significance level.

Example with a small sample: Crash tests

A certain type of automobile is known to sustain no visible damage 25% of the time in 10-mph crash tests. A modified bumper design has been proposed to increase this percentage. Among 15 car crashes, 6 have no visible damage. Is the observed percentage increase real at the level of 5% or is it due to chance?

- Let p be the percentage of cars with no visible damage after the crash test. Then, null and alternative hypothesis are

$$H_0 : p \leq 0.25 \qquad vs. \qquad H_1 : p > 0.25.$$

- The sample size $n = 15$ is too small to use a normal approximation (i.e. the large sample test statistic).
- What can we do? $\Longrightarrow$ We need to reason from first principles.
- Recall the basic idea: We reject H_0 if the probability of observing a sample proportion $\hat{p} = 6/15$ or larger is very small when H_0 is true.

Example with a small sample: Crash tests

A certain type of automobile is known to sustain no visible damage 25% of the time in 10-mph crash tests. A modified bumper design has been proposed to increase this percentage. Among 15 car crashes, 6 have no visible damage. Is the observed percentage increase real at the level of 5% or is it due to chance?

- Note that under H_0 (i.e. when $p = 0.25$), $\sum_{i=1}^{15} X_i \sim Bin(15, 0.25)$. Hence, we can directly compute the p-value under H_0!

- p-value = "probability of obtaining a test statistic as extreme (or even more extreme) than the one observed given that H_0 is true", i.e.

$$\begin{aligned} P\left(\bar{X}_n \geq \frac{6}{15} \mid H_0 \text{ is true}\right) &= P\left(\sum_i^{15} X_i \geq 6 \mid p = 0.25\right) \\ &= 1 - P\left(\sum_i^{15} X_i \leq 5 \mid p = 0.25\right) \\ &= 1 - \texttt{pbinom}(5, 15, 0.25) \approx 0.148. \end{aligned}$$

- Since $14.8\% > 5\%$, we do not reject H_0 at a 5% level.

Outline

Comparison of Treatments

- We extend our results about confidence intervals and hypothesis testing of a single population distribution to situations involving two population distributions.

- We will only consider scenario of a **randomized controlled experiment**:

  - One randomly chosen group is assigned to receive a treatment, the other group receives no treatment.

  - The groups are called treatment group and control group, respectively.

  - We are interested in hypotheses concerning the difference of the populations means of treatment and control group.

Setup and Assumptions about Two Samples

- Treatment group: $X_1, \ldots, X_n$ random sample from distribution with mean μ_1 and variance σ_1^2.
- Control group: $Y_1, \ldots, Y_m$ random sample from distribution with mean μ_2 and variance σ_2^2.
- Feature of interest: $\Delta = \mu_1 - \mu_2$.
- Sample sizes m and n and variance σ_1^2 and σ_2^2 may be different.
- $X_1, \ldots, X_n$ and $Y_1, \ldots Y_m$ are mutually independent.
- Setup also applies to comparing two population proportions. *(Recall that proportions are averages of binary data...)*

A Natural Estimator for the Population Difference Δ

- An intuitive estimator for $\Delta = \mu_1 - \mu_2$ is the plug-in estimator

$$\widehat{\Delta} = \bar{X}_n - \bar{Y}_m.$$

(If the data are normally distributed, $\widehat{\Delta}$ is in fact the MLE of Δ.)

- Note that

$$\mathrm{Var}(\widehat{\Delta}) = \mathrm{Var}(\bar{X}_n - \bar{Y}_m) = \mathrm{Var}(\bar{X}_n) + \mathrm{Var}(\bar{Y}_m) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}.$$

- Hence, the standard error of $\widehat{\Delta}$ is

$$\mathrm{SE}(\widehat{\Delta}) = \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}.$$

- A plug-in estimate of the standard error is

$$\widehat{\mathrm{SE}}(\widehat{\Delta}) = \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}},$$

where s_1, s_2 are the sample standard deviations of treatment and control group.

Large sample CIs for the Population Difference

Large sample CIs for the population difference

Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be a realizations of random samples from treatment and control group and $z_{\alpha/2}$ the critical value of $N(0,1)$.

- If the variances σ_1^2 and σ_2^2 are known, a large sample (asymptotic) $100(1-\alpha)\%$ confidence interval for the population mean difference $\Delta = \mu_1 - \mu_2$ is

$$\left(\bar{x}_n - \bar{y}_m - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}},\, \bar{x}_n - \bar{y}_m + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right).$$

- If the variances σ_1^2 and σ_2^2 are unknown, a large sample (asymptotic) $100(1-\alpha)\%$ confidence interval for the population mean difference $\Delta = \mu_1 - \mu_2$ is

$$\left(\bar{x}_n - \bar{y}_m - z_{\alpha/2}\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}},\, \bar{x}_n - \bar{y}_m + z_{\alpha/2}\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}\right),$$

  where $s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$ and $s_2^2 = \frac{1}{n-1}\sum_{i=1}^{m}(y_i - \bar{y}_m)^2$.

Example: Confidence Interval for Treatment Effect

The following information has been reported from an experiment:

Group	**Sample Size**	**Sample Mean**	**Population SD**
Control	50	24.3	5.2
Treatment	50	27.0	5.8

- 95% Confidence interval for $\Delta = \mu_1 - \mu_2$.

$$(\bar{x} - \bar{y}) \ \pm \ z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

$$= (24.3 - 27) \ \pm \ 1.96 \cdot \sqrt{\frac{(5.2)^2}{50} + \frac{(5.8)^2}{50}} = -2.7 \ \pm \ 2.2$$

and therefore $(-4.9 \ , \ -0.5)$ is a 95% Confidence interval for Δ.

Configuration of Comparison of Treatments

Comparison of Treatments

Given a 'prior belief' $\Delta_0 = \mu_1^0 - \mu_2^0$ about the population difference between treatment and control group population mean we have the following three testing problems named after the alternative hypothesis that one hopes to establish:

(a) One-sided upper tail test:

$$H_0 : \Delta \leq \Delta_0 \qquad vs. \qquad H_1 : \Delta > \Delta_0.$$

(b) One-sided lower tail test:

$$H_0 : \Delta \geq \Delta_0 \qquad vs. \qquad H_1 : \Delta < \Delta_0.$$

(c) Two-sided (tail) test:

$$H_0 : \Delta = \Delta_0 \qquad vs. \qquad H_1 : \Delta \neq \Delta_0.$$

Comparison of Treatments (Large Sample)

COMPARISON OF TREATMENTS (LARGE SAMPLE, KNOWN VARIANCES)

Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be a realizations of random samples from treatment and control group. Let $\alpha \in (0,1)$ and z_α be the critical value of the standard normal distribution and define

$$t = \frac{\widehat{\Delta} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} = \frac{(\bar{x}_n - \bar{y}_m) - (\mu_1^0 - \mu_2^0)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}.$$

(a) One-sided upper tail test at level α:

Reject H_0 in favor of H_1 whenever $t > z_\alpha$.

(b) One-sided lower tail test at level α:

Reject H_0 in favor of H_1 whenever $t < -z_\alpha$.

(c) Two-sided (tail) test at level α:

Reject H_0 in favor of H_1 whenever $|t| > z_{\alpha/2}$.

Comparison of Treatments (Large Sample)

Comp. of Treat. (Large Sample, Unknown Variances)

Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be a realizations of random samples from treatment and control group with unknown variances. Let $\alpha \in (0,1)$ and z_α be the critical value of the standard normal distribution and define

$$t = \frac{\widehat{\Delta} - \Delta_0}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}},$$

where $s_1^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$ and $s_2^2 = \frac{1}{n-1} \sum_{i=1}^m (y_i - \bar{y}_m)^2$.

(a) One-sided upper tail test at level α:

Reject H_0 in favor of H_1 whenever $t > z_\alpha$.

(b) One-sided lower tail test at level α:

Reject H_0 in favor of H_1 whenever $t < -z_\alpha$.

(c) Two-sided (tail) test at level α:

Reject H_0 in favor of H_1 whenever $|t| > z_{\alpha/2}$.

Comparison of Treatments (Small Sample)

Comp. of Treat. (Small Sample, Unknown Variances)

Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be a realizations of random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with unknown $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$. Let $\alpha \in (0,1)$ and $t_{\alpha,\nu}$ be the critical value of the $t(\nu)$-distribution with degrees of freedom

$$\nu = \frac{\left(s_1^2 + s_2^2\right)^2}{\frac{s_1^4}{n-1} + \frac{s_2^4}{m-1}}.$$

and define $t = (\widehat{\Delta} - \Delta_0)/\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}$, where $s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$ and $s_2^2 = \frac{1}{n-1}\sum_{i=1}^{m}(y_i - \bar{y}_m)^2$.

(a) One-sided upper tail test at level α:

Reject H_0 in favor of H_1 whenever $t > t_{\alpha,\nu}$.

(b) One-sided lower tail test at level α:

Reject H_0 in favor of H_1 whenever $t < -t_{\alpha,\nu}$.

(c) Two-sided (tail) test at level α:

Reject H_0 in favor of H_1 whenever $|t| > t_{\alpha/2,\nu}$.

Example: Testing Treatment Effect (Known Variances)

45 units are divided at random into $n = 20$ control units and $m = 25$ treatment units. The average outcome of the control group is $\bar{x} = 29.8$, while the average response in the treatment group is $\bar{y} = 34.7$. Assume the outcomes for the two groups come from two independent normal distributions with $\sigma_1 = 4.0$ and $\sigma_2 = 5.0$, respectively.

- Hypothesis test: $H_0 : \mu_1 - \mu_2 = \Delta_0$ vs. $\mu_1 - \mu_2 \neq \Delta_0 = 0$.
- Test statistic:

$$z = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} = \frac{20 - 34.7}{\sqrt{\frac{4^2}{20} + \frac{5^2}{25}}} = -10.957$$

- Reject null hypothesis at all conventional significance levels!

Example: Testing Treatment Effects in Small Samples

The tensile strength (psi) of liner specimens with and without a certain fusion process is summarized as follows:

Treatments	**Sample Size**	**Sample Mean**	**Sample SD**
No Fusion	10	2902.8	277.3
Fused	8	3108.1	205.9

1. Is there any substantial evidence that the fused process increases the tensile strength?
2. Construct the 95% one-sided confidence interval for the mean difference.
3. Construct the 95% two-sided confidence interval for the difference.

Example: Testing TE in Small Samples (Cont.)

1. Is there any substantial evidence that the fused process increases the tensile strength?

- Denote the population means of "Fused" (treatment) and "No Fusion" (control) by μ_1 and μ_0, respectively. Write $\Delta = \mu_1 - \mu_2$.
- The hypothesis test is

$$H_0 : \Delta \leq 0 \qquad vs. \qquad H_1 : \Delta > 0.$$

- The estimated difference is $\widehat{\Delta} = 3108.1 - 2902.8 = 205.3$ with $\widehat{SE} = \sqrt{277.3^2/10 + 205.9^2/8} = 113.97$. The observed test statistic is

$$t = \frac{205.3}{113.97} = 1.801$$

- Using $s_2^2 = 227.3^2/10 = 7689.529$ and $s_1^2 = 5299.351$ we have

$$\frac{(5299.351 + 7689.529)^2}{5299.351^2/7 + 7689.529^2/9} = 15.94$$

rounded down to 15.
- We find that $t_{0.05,15} = 1.753$ and therefore reject H_0 at level 5%!

Example: Testing TE in Small Samples (Cont.)

2. Construct the 95% one-sided confidence interval for the mean difference.

- Since $205.03 - t_{0.05,15} \times 113.97 = 5.51$, the one-sided 95% CI corresponding to the testing problem is $[5.51, \infty)$. Since this interval does not cover 0, we reject H_0 (again).

3. Construct the 95% two-sided confidence interval for the difference.

- $205.03 + t_{0.025,15} \times 113.97 = 205.03 + 2.131 \times 113.97 = 447.786$ and $205.03 - t_{0.025,15} \times 113.97 = -37.726$. Thus, a two-sided 95% CI for the population difference is $[-37.726, 447.786]$.

Analyzing Paired Data

- Imagine that there is only one group of subjects and you apply treatment and control to the the same subject.
  - Comparison of two sleeping pills. Half of the randomly chosen subjects receive drug A and then drug B. The other half receive drug B and then drug A.
  - Effect of drug on hypertension. Measure blood pressure of patients before and after administering a drug.
  - Study progress. Measure study progress of students based on standardized tests that are conducted after certain time intervals.
- In these examples we record two measurements (treatment, control) $= (X, Y)$ per individual. The data are

$$(X_1, Y_1), (X_2, Y_2), \dots (X_n, Y_n),$$

  that is the treatment and control are <u>paired</u>.

Analyzing Paired Data (Cont.)

- Feature of interest: $\Delta = \mu_1 - \mu_2$.
- Differences: $D_i = X_i - Y_i$ are assumed to come from a distribution with mean Δ and variance σ_d^2.
- Statistical Inference: Proceed with the analysis in the same way as in the one-sample situation when testing the population mean.
- Advantages of paired analysis:
  - reduces confunding factors since treatment and control are administered to the same subject.
  - smaller sampling variability:

$$\mathrm{Var}(D) = \mathrm{Var}(X - Y) = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2.$$

    Two measurements on the same subject are often positively correlated, i.e. $\rho > 0$. Thus, the sampling variance of the sample mean differences of paired data is often smaller than the sampling variance of the mean difference of independent treatment and control groups, i.e.

$$\frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{n} \leq \frac{\sigma_1^2 + \sigma_2^2}{n}.$$

Conclusion

- Construction of CIs and hypothesis testing of the population mean difference follows the same principles as construction of CIs and hypothesis testing of the population mean.
- We distinguish between two situations:
  - unpaired data (sample sizes of treatment and control group are different, i.e. $n \neq m$)
  - paired data (treatment and control group are of the same size)
- If the data is paired, the standard error of the sample mean difference can be estimated more efficiently (i.e. pooled estimate)
- Bootstrap procedures for the population mean carry over to the population mean difference with the obvious (trivial) modifications.

Outline

Difference of Population Proportions

- Feature of interest: $p_d = p_1 - p_2$, where p_1, p_2 are the population proportions of treatment and control group, respectively.
- Estimator: $\hat{p}_d = \hat{p}_1 - \hat{p}_2$, the intuitive plug-in estimator of the difference.
- Note that

$$\mathrm{Var}(\hat{p}_d) = \mathrm{Var}(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}.$$

- Hence, a plug-in estimate of the standard error is

$$\widehat{\mathrm{SE}}(\hat{p}_d) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}.$$

Large Sample CIs for the Diff. of Population Props.

Large sample CIs for the difference of population proportions

Let $z_{\alpha/2}$ be the critical value of $N(0,1)$. A large sample (asymptotic) $100(1-\alpha)\%$ confidence interval for difference of the population proportions $p_d = p_1 - p_2$ is

$$\left(\hat{p}_d - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}},\ \hat{p}_d + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}\right).$$

- One-sided upper and lower CIs are defined analogously.
- Bootstrap CIs can also be constructed using the usual procedure.

Testing the Difference of Population Proportions

- Distinction from population mean difference:
  - we always need a large sample (i.e. normal approximation, there is no t-approximation)
  - under the null hypothesis $H_0 : p_1 = p_2 = p_0$ the standard error is computed differently:

$$SE(\hat{p}_d) = \sqrt{\frac{p_1(1-p_1}{n} + \frac{p_2(1-p_2)}{m}} = \sqrt{p_0(1-p_0)\left(\frac{1}{n} + \frac{1}{m}\right)}.$$

An intuitive estimate for p_0 is

$$\hat{p}_{\text{pool}} = \frac{n\hat{p}_1 + m\hat{p}_2}{n+m} = \frac{\sum_{i=1}^{n} X_i + \sum_{i=1}^{m} Y_i}{n+m},$$

Hence, we estimate

$$\widehat{SE}(\hat{p}_d) = \sqrt{\hat{p}_{\text{pool}}(1-\hat{p}_{\text{pool}})\left(\frac{1}{n} + \frac{1}{m}\right)}.$$

Testing the Diff. of Population Proportions (Cont.)

Testing the Diff. of Population Proportions

Let $z_\alpha, z_{\alpha/2}$ be critical values of $N(0,1)$ and define

$$t = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_{\text{pool}}(1-\hat{p}_{\text{pool}})\left(\frac{1}{n}+\frac{1}{m}\right)}}.$$

(a) One-sided upper tail test at level α:

Reject $H_0 : p_1 = p_2$ *in favor of* $H_1 : p_1 > p_2$ *whenever* $t > z_\alpha$.

(b) One-sided lower tail test at level α:

Reject $H_0 : p_1 = p_2$ *in favor of* $H_1 : p_1 < p_2$ *whenever* $t < -z_\alpha$.

(c) Two-sided (tail) test at level α:

Reject $H_0 : p_1 = p_2$ *in favor of* $H_1 : p_1 \neq p_2$ *whenever* $|t| > z_{\alpha/2}$.

Example: A Prison's Dilemma

Consider the data on defendants from the San Francisco county accused of robbery in 1981-1982:

	Number	**Sentenced to Prison**
Plead guilty	191	101
Plead not guilty	64	56

1. Do this data suggest that the proportion of all defendants in these circumstances who plead guilty and are sent to prison differs from the proportion who are sent to prison after pleading innocent and being found guilty?
2. Construct the 95% confidence interval for the population rate difference.

Example: A Prison's Dilemma (Cont.)

1. Do this data suggest that the proportion of all defendants in these circumstances who plead guilty and are sent to prison differs from the proportion who are sent to prison after pleading innocent and being found guilty?

- $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$.
- pooled estimate of standard error

$$\hat{p}_{\text{pool}} = \frac{101 + 56}{191 + 64} = 0.616.$$

- observed test statistic

$$t = \frac{0.529 - 0.875}{\sqrt{0.616 \times 0.384(1/191 + 1/64)}} = -4.83$$

- Thus, we reject H_0 at 5% level.

Example: A Prison's Dilemma (Cont.)

2. Construct the 95% confidence interval for the population rate difference.

- Important: The CI is not computed under H_0. Hence the standard error has to be estimated as

$$\widehat{\mathrm{SE}}(\hat{p}_d) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}} = 5.49\%.$$

- A 95% CI is given by

$$-34.6\% \pm 10.8\%.$$