MATH 147B Exam 3 Study Guide

David Collins and Thu Quan

November 23, 2022

1 Chapter 18

P-values

- P-value: The probability of seeing data at least as far from the hypothesized value as the observed value, given that the null hypothesis is true.

  In other words, P-value is a conditional probability of getting observed statistic value or even more extreme, given that the null hypothesis is true.

- P-value is not the probability that the null hypothesis is true.
- There are two conclusions for the null hypothesis, we base our decision on P-value:
  - Low enough P-value leads us to **reject** the null hypothesis.
  - When P-value is high, the null model doesn't appear to be false, we **fail to reject** the null hypothesis.

Alternatives

- One-sided alternative: Example: $p < 0.2$ or $p > 0.2$.
- Two-sided alternative: Example: $p \neq 0.2$.

One-proportion z-text

Assumptions and conditions:

- Independence Assumption: Individuals in the samples (whose proportions we're finding) must be independent of each other, knowing how one responses should not provide information about other responses.
- Randomization Condition: Individuals in sample should be representative of the population, or the sample should be drawn randomly.
- 10% Condition: The sample size should not be more than 10% of the population.
- Success/Failure Condition: Check for at least 10 success and 10 failures, or make sure $np \geq 10$ and $nq \geq 10$ where $q = 1 - p$.

One sample t-test for the mean

Assumptions and conditions:

- Randomization Condition: This condition is satisfied if the data arise from a random sample or it should be a randomly sampled data. The sample you have should be likely to be representative of the population you wish to learn about.
- Independence Assumption: Each response in the sample you have should be likely to be independent to others.
- Nearly Normal Condition (or Normal Population Assumption): The data comes from a distribution that is unimodal and symmetric.

Bootstrap Hypothesis Tests

- When we bootstrap histograms, the center of the histogram of our sampling distribution is at the sample mean, NOT the population mean.

Table 1: One Sample Hypothesis Tests

Proportion $n, \hat{p}$	**Mean** $n, \overline{x}, s$
$H_0 : p = p_0$ $H_A : p(>, <, \neq) p_0$	$H_0 : \mu = \mu_0$ $H_A : \mu(>, <, \neq) \mu_0$
$SE(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}}$	$SE(\overline{x}) = \frac{s}{\sqrt{n}}$
z=$\frac{\hat{p} - p_0}{SE(\hat{p})}$	$t = \frac{\overline{x} - \mu}{SE(\overline{x})}$

- Note that μ_0, p_0 are the initial assumed values. For example, if we assume the population proportion is 3%, then we say $H_0 : p = .03$.

2 Chapter 19

Alpha Levels, Statistical Significance

How small P-value has to be for you to reject the null hypothesis?

- Alpha level (α-levels): The threshold P-value that determines when we reject a null hypothesis. If **p-value** based on the null hypothesis is **less than** α, we reject that null hypothesis. If **p-value** falls **above** α, we **fail to reject** that null hypothesis.
  - Common α-levels are $0.1, 0.05, 0.01$, and 0.001.
  - The alpha level is also called the **significance level**. For example, we might say that we reject the null hypothesis "at the 5% level of significance". If we reject the null hypothesis, we call the result statistically significant.
  - In general, $\alpha = 1 - C$, where C is the confidence level. For a two-tailed significance level, the equation is $\frac{1-C}{2} = \alpha$.

Errors

- Type I error: The null hypothesis is true, but we mistakenly reject it. How often will a Type I error occur? It happens when the null hypothesis is true, but we've had a bad luck to reject it. To reject H_0, the P-value must fall below α, so we're setting the probability of Type I error to α.
- Type II error: The null hypothesis is false, but we fail to reject it. We assign the letter β to the probability of this mistake.
- Power: When the null hypothesis actually *is* false, we hope our test is strong enough to reject it. The probability of making that correct decision is called the **power** of the test. That's the probability that it *succeeds* in rejecting a false null hypothesis, so it's just $\mathbf{1} - \beta$.

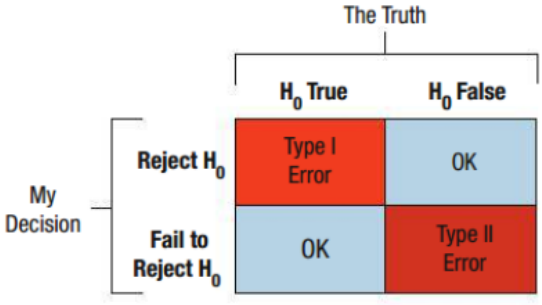

Figure 1: Type I and Type II errors

- Several important relationships:
  - Power $= 1 - \beta$. When we decrease the probability of a Type II error, it increases the power.
  - Reducing α to lower the chance of committing a Type I error, this will have the effect of increasing β, the probability of a Type II error, and correspondingly reducing the power

3 Chapter 20

The Two-Sample z-Test: Testing difference between proportions

Assumptions and Conditions

- Independence
  - Independence Assumption: *Within each group*, individual reponses should be independent of each other. Knowing one response should provide no information about other reponses.
  - Independent Groups Assumption: The responses in the two groups we're comparing must also be independent of each other. Knowing how one group responds should not provide information about other group.
- Randomization Condition: The responses should be selected with randomization.
- Success/Failure Condition: We should expect at least 10 successes and 10 failures. In other words, $np \geq 10$ and $nq \geq 10$, where $q = 1 - p$.

The two-sample t-Test: Testing for the difference between two means

Assumptions and Conditions

- Independence
  - Independence Assumption: *Within each group*, individual responses should be independent of each other. Knowing one response should provide no information about other responses.
  - Independent Groups Assumption: The responses in the two groups we're comparing must also be independent of each other. Knowing how one group responds should not provide information about other group.
- Randomization Condition: The responses should be selected with randomization.
- Nearly Normal Condition (or Normal Populations): The histograms for both groups should look unimodal, symmetric, without outliers.

Pooling

- For proportions, the null hypothesis assumes that two proportions are the same, then their variances will be the same as well. In that case, it is OK to combine or **pool** the data from two groups when estimating the standard error.
- We can ALWAYS pool two proportions. It is beneficial to do so since this increases the power of our hypothesis test.
- For means, the assumption that the two groups have the same variance is much less natural. The null hypothesis assumes the means are equal, but usually there's no reason to believe that the variances should be as well. This is incredibly rare so please do not pool means.

Table 2: Two Sample Hypothesis Tests

Proportion	**Mean**
$n_1, n_2, \hat{p_1}, \hat{p_2}$	$n_1, n_2, \overline{x_1}, \overline{x_2}, s_1, s_2$
$H_0 : p_1 = p_2$ $H_A : p_1(>, <, \neq)p_2$	$H_0 : \mu_1 = \mu_2$ $H_A : \mu_1(>, <, \neq)\mu_2$
$SE_{\text{pooled}}(\hat{p_1} - \hat{p_2}) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_1} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_2}}$	$SE(\overline{x_1} - \overline{x_2}) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
z=$\frac{\hat{p_1} - \hat{p_2} - 0}{SE_{\text{pooled}}(\hat{p_1} - \hat{p_2})}$	$t = \frac{\overline{x_1} - \overline{x_2} - 0}{SE(\overline{x_1} - \overline{x_2})}$

4 Chapter 21

- The point of pairing means is to reduce a two-sample t-test to a one-sample t-test of differences IF we can find an appropriate pairing. If you compare the formulas below to the one-sample t-test, you'll notice they're exactly the same, just with a subscript "d" on certain variables.

Assumptions/Conditions

- Paired data: Data are paired when the observations are collected in pairs or the observations in one group are naturally related to observations in the other. Paired data arise in a number of ways. Perhaps the most common way is to compare subjects with themselves before and after treatment.
- Independence: The groups are NOT independent, but our new data list of paired differences should be independent.
- Nearly Normal Condition: The histogram of pairwise differences should look unimodal and symmetric.
- The paired t-test: A hypothesis test for the mean of the pairwise differences of two groups. Mechanically, a paired t-test is just a one-sample t-test for the mean of these pairwise differences.
- Means cannot always be paired. There has to be a logical way to pair them, and the samples must be independent of each other. Example: there is no natural pairing between a bunch of cats and toasters in an experiment to find the difference of their weights.

Table 3: One Sample Paired t-test

Mean
$n, \overline{x}_d, s_d$
$H_0 : \mu_d = \Delta_0$ $H_A : \mu_d(>, <, \neq)\Delta_0$
$SE(\overline{x}_d) = \frac{s_d}{\sqrt{n}}$
$t = \frac{\overline{x}_d - \Delta_0}{SE(\overline{x}_d)}$

- We usually assume Δ_0 to be zero.

5 Chapter 22

Goodness-of-fit Tests

- Goodness-of-fit Test: A test of whether the distribution of observed counts in one single categorical variable matches the distribution predicted by a probability model.
- In this test, we start with one variable, one sample. The null hypothesis for this test is that the distribution of a variable predicted by the probability model is correct. The alternative hypothesis is that the model doesn't predict the probability model/distribution.

Chi-Square Test of Homogeneity

- A test for homogeneity: measures one categorical variable and asks if the distributions of that variable are the same across several groups.
- There are several groups, but still only one variable here.
- The null hypothesis for this test: Two (or several) groups which have the variable have the same distribution of that variable. The alternative hypothesis is that they don't have the same distribution.

Chi-Square Test of Independence

- A test for independence categorizes one group on two categorical variables and asks whether two variables measured are independent.
- There's just one sample here, but two variables.
- The null hypothesis for this test is that there is no association between two variables. The alternative hypothesis is that there is an association (a.k.a. the variables are dependent).

Assumptions and Conditions for these three tests:

- Counted Data Condition: The data must be *counts* for the categories of a categorical variable. Applying this method to proportions, percentages, or measurements is incorrect.
- Independence Assumption: The counts in the cells should be independent of each other. The easiest case to notice this condition is that the individuals who are counted in the cells are sampled independently.
- Sample Size Assumption (or Expected Cell Frequency Condition): We should expect to see at least 5 individuals in each cell.

Subtle difference between tests for Independence and for Homogeneity:

- Independence: Two categorical variables measured on a single population, and the question is "Are the variables independent?"
- Homogeneity: A single categorical variable measured on two or more groups, and we ask if the distribution of the variable was the same across all the groups. ("Are the groups homogeneous?/ Are the groups distributed the same?")
- IMPORTANT: The math for all Chi-Square tests is exactly the same. The only differences are the hypotheses and conclusion. Also, all Chi-Square tests are one-sided. This is because the χ^2 distribution is right skewed, unlike the Normal and Student's t distributions.
- There are two ways to calculate degrees of freedom.
  1) If you do NOT have a contingency table, df= number of categories -1.
  2) If you have a contingency table, the formula is

$$df = (\text{number of rows} - 1)(\text{number of columns} - 1) \tag{1}$$

- The most important step of calculating a Chi-Square statistic is finding the expected values. If all expected counts are greater than 5, then the Expected Cell Frequency condition is satisfied. If you have to find expected values from a contingency table, use this formula:

$$\frac{\text{Row total} \times \text{Column total}}{\text{Overall total}} \tag{2}$$

  For example, if your data is in row 2, column 3, you use the second row total times the third column total divided by the overall total count.

- $$\chi^2 = \sum_{\text{all categories}} \frac{(\text{Observed value} - \text{Expected value})^2}{\text{Expected value}} \tag{3}$$

6 Chapter 23.1-.4 and 23.6

Assumptions and Conditions for regression inference

- Linearity Assumption (or Straight Enough Condition): is satisfied if a scatterplot looks straight. Some examples are shown in Figure 2.

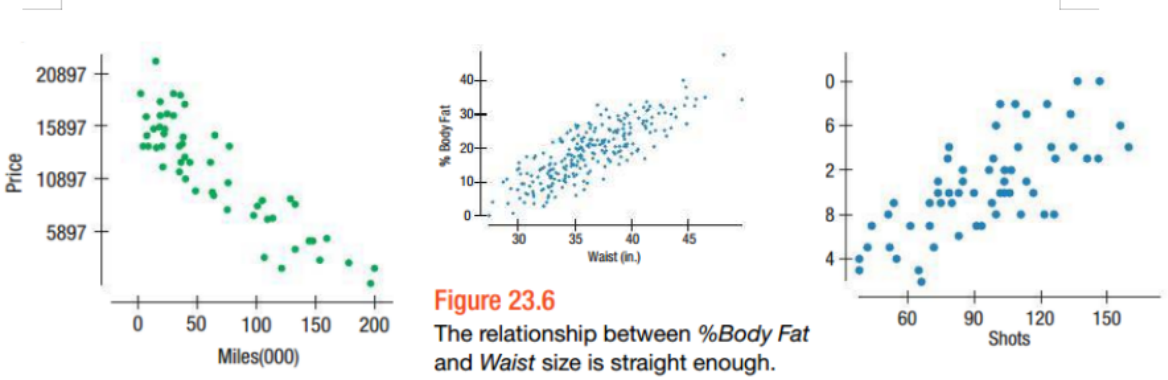

Figure 23.6
The relationship between *%Body Fat* and *Waist* size is straight enough.

Figure 2: Linear relationship

- Independence Assumption: We can check that the individuals are a representative (ideally, random) sample from the population. We can also check displays of the regression residuals, it should not have any patterns, trends, or clumping, because any of which would suggest a failure of independence, (see Figure 3 for some examples of independence).

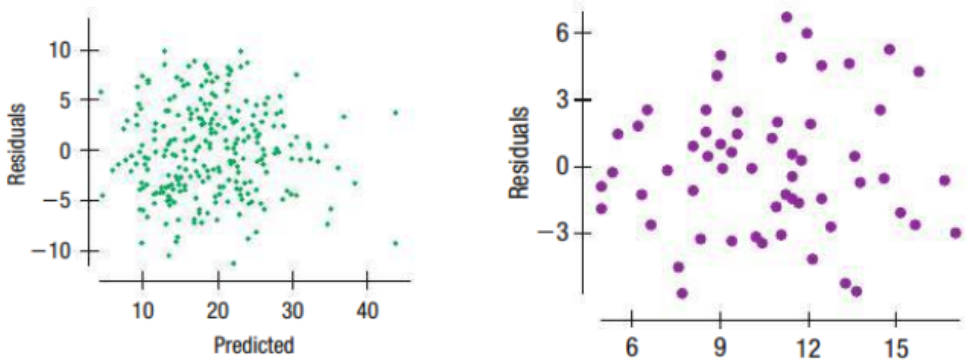

Figure 23.7
The residuals show only random scatter when plotted against the predicted values.

Figure 3: Scatterplots of residuals randomly plotted

- Equal Variance Assumption: The variability of y should be about the same for all values of the x's, or the scatter of the residuals is the same everywhere.

  This condition is also known as the Does the Plot Thicken? Condition. A scatterplot of the residuals offers a visual check for this condition. The residuals should show no patterns, so the data are independent, and the plot doesn't thicken. Be alert for a "fan" shape or other tendency for the variation to grow or shrink in one part of the scatterplot (as shown in Figure 4).

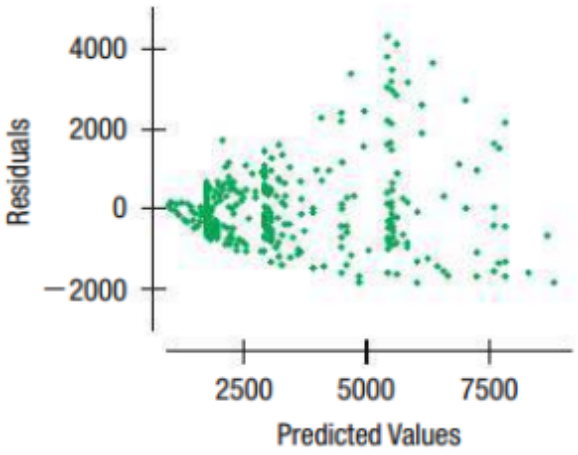

Figure 4: 'Fan' shape

- Normal Population Assumption: The histogram of residuals satisfies Nearly Normal Condition (bell-shaped, unimodal, symmetric) and without striking outliers.

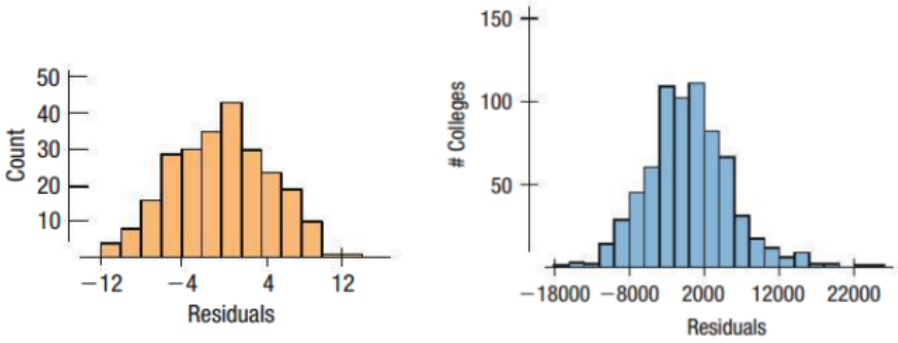

Figure 5: Histograms satisfy Normal Assumption

Table 4: Regression Inference Test

Slope
n, b_1, s_x, s_e
$H_0 : \beta = 0$ $H_A : \beta \neq 0$
$SE(b_1) = \frac{s_e}{\sqrt{n-1}s_x}$ $df = n - 2$
$t = \frac{b_1 - 0}{SE(b_1)}$

- All Regression Inference hypothesis tests are TWO-sided. We can find β the true slope of the line, which acts as our population mean.

Most importantly, good luck everyone! Study hard and you'll do great! - David and Thu